Publisher: State and Provincial Joint Engineering Lab. of Advanced Network Monitoring and Control (ANMC)

Cooperate:

Xi'an Technological University (CHINA)
West Virginia University (USA)
Huddersfield University of UK (UK)
Missouri Western State University (USA)
James Cook University of Australia
National University of Singapore (Singapore)

Approval:

Library of Congress of the United States

Shaanxi provincial Bureau of press, Publication, Radio and Television

Address:

4525 Downs Drive, St. Joseph, MO64507, USA

No. 2 XueFu Road, WeiYang District, Xi’an, 710021, China

Telephone: +1-816-2715618 (USA) +86-29-86173290 (CHINA)

Website: www.ijanmc.org

E-mail: ijanmc@ijanmc.org

xxwlcn@163.com

ISSN: 2470-8038

Print No. (China): 61-94101

Publication Date: March 28, 2022

Table of Contents

Research on Image Super-resolution Reconstruction Based on Deep Learning

Jingyu Jiang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: jjy1030@126.com;

Yan Jiao
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: jiaoyan@st.xatu.edu.cn;

Li Zhao
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 332099732@qq.com;

***Abstract*—Image super-resolution reconstruction (SR) aims to use a specific algorithm to restore a low-resolution blurred image in the same scene into a high-resolution clear image. Due to its wide application value and theoretical value, image super-resolution reconstruction technology has become a research hotspot in the field of computer vision and image processing, and has attracted widespread attention from researchers. Compared with traditional methods, deep learning methods have shown better reconstruction effects in the field of image super-resolution reconstruction, and have gradually developed into the mainstream technology. Therefore, this paper classifies the image super-resolution reconstruction problem systematically according to the structure of the network model, and divides it into two categories: the super-division method based on the convolutional neural network model and the super-division method based on the generative confrontation network model. The main image super-resolution reconstruction methods are sorted out, several more important deep learning super-resolution reconstruction models are described, the advantages and disadvantages of different algorithms and the applicable application scenarios are analyzed and compared, and the different types of super-resolution algorithms are discussed. The method of mutual fusion and image and video quality evaluation, and a brief introduction to commonly used data sets. Finally, the potential problems faced by the current image super-resolution reconstruction technology are discussed, and a new outlook for the future development direction is made.**

Keywords-Image Super-Resolution Reconstruction; Deep Learning; Convolutional Neural Network; Generative Confrontation Network

I. INTRODUCTION

Image super-resolution reconstruction (SR) refers to the use of image processing and machine learning methods to reconstruct one or more low-resolution (LR) images in the same scene

DOI: 10.2478/ijanmc-2022-0001

with rich image details and High-Resolution (HR) image process with clear texture [1]. It has important application value in the fields of video, remote sensing, medicine and security monitoring. With the rapid development of machine learning in the field of computer vision, deep learning technology has been widely used to solve SR problems and achieved good reconstruction results, and has gradually become the mainstream.

Existing super-resolution reconstruction algorithms are usually divided into three categories: interpolation-based methods, which are simple but provide too smooth reconstructed images, lose some details and produce ringing effects; methods based on modeling, Compared with the interpolation method, this type of algorithm has a better reconstruction effect, but when faced with a large amount of calculation, the calculation process takes a long time, is difficult to solve and is greatly affected by the amplification factor; based on the learning method, this type of algorithm solves the problem of The scale factor is sensitive to the problem and the reconstruction effect is the best, which is the mainstream direction of current research [3].

Convolutional neural network (CNN) and generative adversarial network (GAN) are the current mainstream network models. When the scaling factor is large, using these two network models can restore the height of the image very well. Frequency information to make the output image closer to the original real image [3].

II. NETWORK MODEL OF SUPER-RESOLUTION RECONSTRUCTION METHOD

According to the different network model structure, the image super-resolution reconstruction method based on deep learning can be divided into the following two categories: ① Super-segmentation method based on Convolutional Neural Network (CNN) model; ② Based on generative adversarial network (Generative Adversarial Networks, GAN) model super-division method. In response to various requirements for image super-resolution, super-resolution network models with various characteristics have been produced (Figure 1) [4].

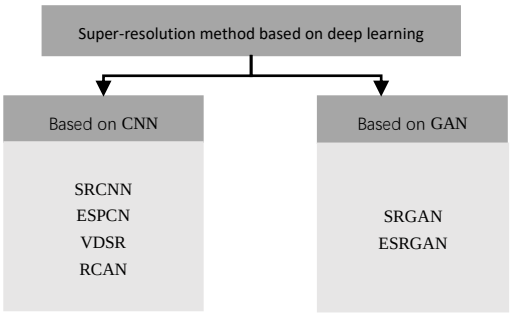


Figure 1. Super-resolution method based on deep learning

A. Super-division method based on CNN network model

1) SRCNN

SRCNN [5] uses the relationship between deep learning and traditional sparse coding as a basis, and divides the 3-layer network into feature extraction, nonlinear mapping, and final reconstruction. For a low-resolution image, as shown in Figure 2, the method first uses bicubic interpolation to enlarge it to the same size as the target, and then extracts and represents the image block, and then uses a three-layer convolutional network to make nonlinear The result of mapping and reconstruction is output as a high-resolution image. SRCNN introduces convolutional neural networks to SR tasks for the first time. Unlike the step-by-step processing of traditional SR algorithms, SRCNN integrates various stages into a deep learning model, which greatly simplifies the SR workflow and can be regarded as a super deep learning based Milestones in resolution methods [6].

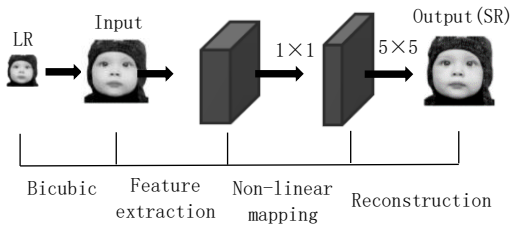


Figure 2. SRCNN network model

Although the network structure of SRCNN is simple in design and is superior to traditional super-resolution algorithms in terms of image reconstruction quality and speed, it has the following problems: ①Does not use any prior knowledge; ②It is only suitable for SR tasks with a single magnification factor. For different magnification factors, the network model needs to be trained again; ③Because the input image needs to be interpolated and magnified to the same size as the target, the entire image reconstruction process is performed in the HR space, which takes up a lot of memory space and increases the amount of calculation. , The error produced by the interpolation process will also have an impact on the reconstruction effect, the model convergence speed is slower, and the training time is longer; ④The number of network layers is less, and the receptive field of the convolution kernel is also small (13 × 13). Good application of image context-related information, resulting in unclear texture of the final reconstructed HR image and limited algorithm adaptability [7].

Initially, the smaller datasets Set5 and Set14 were used to train the SRCNN algorithm. After training, the knowledge learned is relatively small, and the image reconstruction performance is constrained. When the relatively large dataset BSD200 is used, the reconstruction performance is significantly improved. The reconstruction performance of the image is also greatly affected by the size of the data set [8].

After that, although experts have proposed various network models, SRCNN is still used as a benchmark experiment for evaluating the performance of other network models.

2) (2) ESPCN

In 2016, Shi et al. proposed an Efficient Sub-Pixel Convolutional Neural Network (ESPCN) network model based on pixel rearrangement [9], The core concept of ESPCN is a sub-pixel convolutional layer, which performs a convolution operation on the LR image to obtain LR image features, and then expands the features in the LR space to the HR space through the sub-pixel convolutional layer. Reorganize the HR feature map obtained after convolution to obtain an HR image [10].

The ESPCN network mainly improves the reconstruction layer of SRCNN. The LR image is used as the network input. The sub-pixel convolutional layer is used in the reconstruction layer to double the network training speed. The simple network structure and extremely high reconstruction speed make it very It is suitable for high-speed real-time systems that require relatively low reconstruction performance. In the ESPCN network, the interpolation function used for image size enlargement is implicitly included in the previous convolutional layer, which can be automatically learned. Since the convolution operation is performed on the size of the low-resolution image, the model efficiency is higher. The sub-pixel convolutional layer proposed by the ESPCN model has been widely used later. Compared with the deconvolutional layer proposed in the FSRCNN model, the learned nonlinear effect of upsampling from low-resolution images to high-resolution images is better. It is worth noting that the model also modified the activation function, replacing the

ReLU function with the tanh function, and the loss function is the mean square error.

3) (3) VDSR

VDSR (Very Deep CNN for SR) [11] is the first deep model that proposes to use the global residual learning idea to solve the image SR problem. It is an improved network based on SRCNN. It uses a multi-layer convolution kernel for deep convolution, which not only reduces the amount of parameters, but also makes the following the network layer has a larger receptive field, can better utilize the image context information of a larger area, and obtain a better reconstruction effect than SRCNN. The biggest feature of VDSR is the deep network layers, good image reconstruction effect, and faster training speed. Since the author found that the input LR image is very similar to the output HR image, that is, the low frequency information carried by the LR image is very similar to the low frequency information of the HR image [12], So only need to learn the high-frequency residual part between the HR image and the LR image. The VDSR network structure is shown in Figure 3. It will interpolate to obtain an LR image of the same size as the target and input it into the network, and then add this image and the residual error learned by the network to obtain the final HR reconstructed image. The adaptive gradient pruning strategy is to train the network with a higher learning rate. Although the architecture is huge, it can still speed up the convergence. Therefore, on the basis of increasing the network depth, combining residual network and adaptive gradient cropping to accelerate model training can improve network performance and achieve better reconstruction effects. At the same time, through mixed training of images of different scales, the VDSR network can achieve a single Multi-scale SR reconstruction of the model.

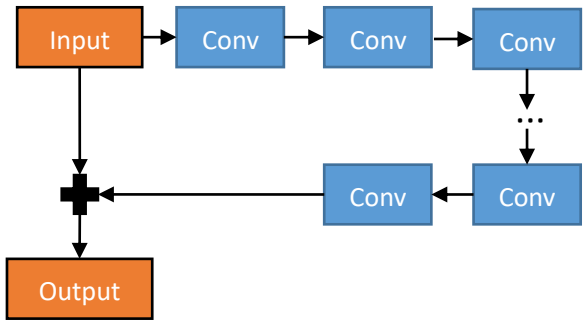


Figure 3. VDSR network structure

4) RCAN

In 2018, Zhang et al. [13] It is believed that the input image contains a large amount of low-frequency information, and the existing SR network treats all channels of this information equally, and lacks the ability to distinguish and learn these channels, which hinders the network's characterization ability. Therefore, the deep residual channel attention network RCAN is proposed [13]. RCAN is the first network to apply the attention mechanism to SR problems. The algorithm obtains a weight value by learning the importance of different channels, which is equivalent to modeling the relationship between channel features, adaptively adjusting each channel feature, thereby effectively strengthening useful feature channels while suppressing useless feature channels, to make fuller use of computing resources. The model uses a locally nested residual structure (residual in residual), which is composed of a residual group (RG) and a long jump connection (LSC). A deeper network is built by simply stacking residual blocks and passes between feature channels the dependence relationship of the selection contains more key information feature channels, and enhances the identification and learning ability of the entire network.

B. Super-division method based on GAN network model

1) SRGAN

The Generative Adversarial Network (GAN) was proposed by Goodfellow et al. It is inspired by the two-person zero-sum game in game theory. The two players in the GAN model are respectively composed of a generative model and a discriminant model (discriminative model) as [14]. SRGAN first applied adversarial training to the problem of image super-resolution reconstruction. The results show that the introduction of adversarial training can enable the network to generate finer texture details. GAN can complete many incredible generation problems, in the fields of image generation, speech conversion, and text generation. Occupy a very important position. As shown in Figure 4, SRGAN inputs the LR image to the generator G for image reconstruction. The discriminator D will train the generated image against the HR image, and finally output the image generated by the training [15]. The collaborative training of the generator and the discriminator enables the network to not only judge the similarity between the generated image and the actual high-resolution image in the pixel domain, but also pay more attention to its distribution similarity in the pixel space. Compared with the previous algorithm, although SRGAN is relatively low in objective evaluation indicators (such as PSNR), it has a better reconstruction effect in visual effects, image details and other intuitive aspects. This is related to its unique network structure and the loss function that combines perceptual loss and adversarial loss. Perceptual loss is a feature extracted by using convolutional neural networks. The generated image is compared with the target image. The feature difference after the convolutional neural network makes the generated picture and the target picture more similar in semantics and style. The confrontation loss is provided by GAN, and the network is trained according to whether the image can be successfully deceived. SRGAN is a milestone in the pursuit of the development of visual experience. SRGAN has significantly improved the overall visual quality of PSNR-based reconstruction.

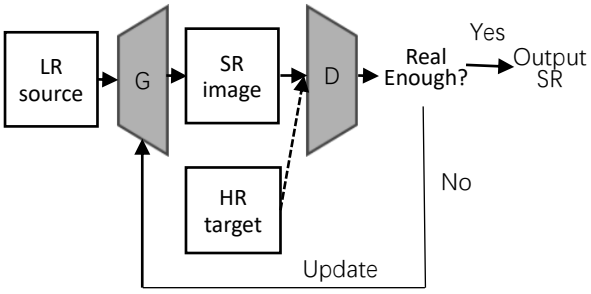


Figure 4. Schematic diagram of the basic structure of SRGAN

2) ESRGAN

In order to improve the reconstruction accuracy of the SRGAN model, on this basis, ESRGAN [15] model has been improved in three aspects: the generator architecture is changed, the batch normalization (BN) layer is removed to reduce the artifacts generated in the reconstructed image, and a denser with higher reconstruction accuracy is introduced. Residual block (RDDB) to improve the structure of the model, make it more capacity and easier to train, help to improve generalization ability, reduce computational complexity and memory usage; improve the perceptual domain loss function, use the VGG before activation Features. This improvement will provide clear edges and more visually consistent results, which can better maintain image brightness consistency and restore better detail texture; enhance the discriminator's ability to discriminate, and use relative discriminators to make them then there is the absolute value of true and false but the relative distance between the predicted generated image and the real image.

Compared with the SRGAN model, on most standard test images, the ESRGAN model can reduce the noise in the reconstructed image while maintaining better details. A large number of experiments have shown that enhanced SRGAN, called ESRGAN, is always better than the most advanced methods in sharpness and detail.

C. Summary

In recent years, deep learning technology has developed vigorously. The super-resolution reconstruction method based on deep learning has gradually become the mainstream of image super-resolution, and a series of network models based on CNN and GAN have been developed, which are introduced in sections 2.1-2.2. The super-division method based on CNN network model and GAN network model is presented. On the whole, the reconstruction performance of the algorithm is continuously improving, and the reconstructed image is getting closer and closer to the original image. The image texture details obtained by the algorithm reconstruction based on the GAN network model are better, the realism of the image is improved, and the human eye perception effect is better. In contrast, the algorithm based on the CNN network model has lower network complexity, lower training difficulty, and higher reconstruction result accuracy, but it will produce artifacts and the reconstruction speed is slower.

III. LOSS FUNCTION CONSTRUCTION

In the field of image super-resolution, the loss function is used to measure the difference between the generated image and the actual high-resolution image. At present, in the field of super-resolution, loss functions play an important role. Commonly used loss functions include pixel-based loss functions and perception-based loss functions.

A. Pixel loss function

Pixel loss mainly measures the pixel difference between the predicted image and the target image, mainly including L1 (mean absolute error) loss and L2 (mean square error) loss.

L1 loss, also known as the minimum absolute deviation (LAD) and the minimum absolute error (LAE), calculates the sum of the absolute difference between the actual value and the target value. In the supervised image super-resolution task, the goal is to make the generated image (SR) as close to the real high-resolution image (HR) as possible, and the L1 loss is used to calculate the value between the corresponding pixel positions of SR and HR The error. The basic expression of mean absolute value error (MAE) is shown in formula (1):

$$L_{MAE}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - F(X_i, \theta)\right| \qquad (1)$$

Among them, L(θ) represents the loss function that the network needs to optimize, n represents the number of training samples, θ represents the parameters of the deep neural network, F(Xi, θ) is the image reconstructed by the network, and Yi represents the corresponding HR image. L2 loss, also called mean square error loss (MSE), calculates the sum of the squares of the absolute difference between the actual value and the target value, which greatly improves the performance of the image super-resolution model based on deep learning, is the most commonly used loss function. However, MSE loss can punish large losses, but can do nothing for small losses, so it will produce blurred images. Its basic expression is shown in formula (2):

$$L_{MSE}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left\|Y_i - F(X_i, \theta)\right\|^2 \qquad (2)$$

The meaning of its basic symbols is the same as in formula (1). The application of the minimum mean square error effectively solves the problem of measuring the difference between the SR reconstructed image and the target HR image, making the image SR model based on deep learning a greater improvement than the traditional learning-based SR reconstruction model.

Compared with MSE, MAE has the advantage that it is more robust to outliers and more tolerant. It can be seen from equation (1) that MAE calculates the absolute value of error Yi-F(Xi, θ), so whether it is Yi-F(Xi, θ)>1 or Yi-F(Xi, θ)<1, there is no The effect of the square term, the punishment is the same, and the weight is the same.

Since these loss functions evaluate the class prediction of each pixel vector separately and then average all pixels, they assert that each pixel in the image has the same learning ability.

Compared with the L1 loss function, the L2 loss function will amplify the gap between the maximum error and the minimum error (such as 2*2 and 0.1*0.1), and the L2 loss function is also more sensitive to outliers.

Since the definition of peak signal-to-noise ratio (PSNR) (see 4.1) has a high correlation with pixel difference, and minimizing pixel loss is equivalent to maximizing PSNR, the pixel loss function has gradually become a commonly used loss function. But because it does not explore image quality issues (e.g., perceptual quality [16], Texture detail [17]), Therefore, the results often lack high-frequency details, resulting in too smooth texture details.

B. Perceptual loss function

For tasks such as image stylization and image super-resolution reconstruction, the L2 loss in the image pixel space was used in the early days, but the L2 loss does not match the image quality of the human eye perception loss, and the restored image details often perform poorly. In today's research, L2 loss is gradually replaced by human eye perception loss. The perceptual loss of the human eye is also called perceptual loss. The difference from the MSE loss that uses image pixels for difference is that the calculated space is no longer the image space. The loss function based on perception can recover more high-frequency information and make the reconstruction performance better. At present, the loss function based on perception mainly includes content loss and counter loss.

The content loss function is divided into feature reconstruction loss function and style reconstruction loss function. Bruna et al. [18] proposed the feature reconstruction loss function for the first time. The feature maps corresponding to the reconstructed SR image and the HR image in the feature space were extracted and compared through the pre-trained VGG19 network. The expressions are as follows:

$$L_{VGG/j} = \frac{1}{W_j H_j} \sum_{h=1}^{H_j} \sum_{w=1}^{W_j} (\phi_j(Y)_{x,y} - \phi_j(G_{\theta_G}(X))_{x,y})^2 \quad (3)$$

Among them, W_j and H_j represent the width and height of the jth feature map, ϕ_j represents the feature map obtained by the jth convolution in the VGG19 network, X represents the original LR image, Y is the reconstructed HR image, G_{θ_G} epresents SR image generated by the network. Gatys et al [19] .proposed a style reconstruction loss function based on the feature reconstruction loss function. This function defines a Gram matrix. The calculation formula is as follows:

$$G_j^{\phi}(x)_{c,c'} = \frac{1}{C_j W_j H_j} \sum_{h=1}^{H_j} \sum_{w=1}^{W_j} \phi_j(x)_{h,w,c} \phi_j(x)_{h,w,c'} \quad (4)$$

The size of the feature map extracted by the VGG network is Cj×Wj×Hj. Then the Euclidean distance difference of the Gram matrix is calculated in the corresponding layer and added to obtain the style reconstruction loss function, as follows:

$$L_{style/j}^{\phi} = \left\| G_j^{\phi}(Y) - G_j^{\phi}(G_{\theta_G}(x)) \right\|^2 \quad (5)$$

The SRGAN network proposes to combat loss for the first time, and its basic form is as follows:

$$L_{Gen}^{SR} = \sum_{n=1}^{N} -lbD_{\theta_D}(G_{\theta_G}(x)) \quad (6)$$

Among them, N represents the number of images, θ D epresents the parameters of the identi-fication network, θ G represents the parameters of the generating network, and represents the probability that the generated image is a real HR image. The final optimization goal of the network is a minimum-maximization problem:

$$\min_{\theta_G} \max_{\theta_D} E_{I^{HR} \square P_{train}(I^{HR})} \left[lbD_{\theta_D}(I^{HR}) \right] + E_{I^{HR} \square P_{G}(I^{HR})} \left[1 - lbD_{\theta_D}(G_{\theta_G}(I^{LR})) \right] \quad (7)$$

Among them, P train（IHR） represents the distribution of HR images, P G（ILR） represents the distribution of original LR images. The adversarial training makes the generated SR image highly similar to the original HR image, which makes it difficult for the discriminant network to distinguish, and finally obtains a fake SR image. The discrimination target of SRGAN is whether the input image is true or not. Unlike SRGAN, ESRGAN replaces the discriminator of the discriminant network with a relative average discriminator. The discrimination target is to predict the probability that the real HR image is more realistic than the generated SR image。The discriminant network is shown in equation (8) and equation (9):

$$\begin{aligned} D(x_r) &= \sigma(C(real)) \to 1 \\ D(x_f) &= \sigma(C(fake)) \to 0 \end{aligned} \quad (8)$$

$$\begin{aligned} D_{Ra}(x_r, x_f) &= \sigma(C(real) - E\left[C(fake)\right]) \to 1 \\ D_{Ra}(x_f, x_r) &= \sigma(C(fake) - E\left[C(real)\right]) \to 0 \end{aligned} \quad (9)$$

Among them, real represents the real HR image, fake represents the generated SR image, C (real) represents the judgment result of the discriminator, and E[C (fake)] represents the average value of the judgment result of the authentication network. Among them, σ represents the Sigmoid activation function, which helps the network learn sharper edges and more texture details by improving the discriminator.

IV. IMAGE SUPER-RESOLUTION QUALITY EVALUATION

The higher the similarity between the high-resolution image reconstructed by super-resolution technology and the real high-resolution image, the better the performance of the image super-resolution algorithm. Generally speaking, two objective quantitative indicators are mainly used for evaluation, including peak signal-to-noise ratio (PSNR) [20] and structural similarity (Structual similarity, SSIM) [21]. For the fairness of comparison, all PSNR (dB) and SSIM metrics are calculated on the y-channel with the center drooping. The higher the evaluation index value, the smaller the difference between the reconstruction result and the original image and the higher the fidelity.

PSNR is one of the most commonly used image reconstruction quality evaluation methods, based on the mean square error (MSE) of the image, MSE is shown in the following equation (10):

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[I(i,j) - K(i,j)\right]^2 \quad (10)$$

Among them, I represents the original high-resolution image, K represents the reconstructed image, and m and n represent the length and width of the image, respectively. PSNR is defined as follows:

$$PSNR = 10\log\left(\frac{MAX_I^2}{MSE}\right) \quad (11)$$

Among them, the logarithm base is the natural base e, which is used unless otherwise stated. MAXI is the maximum number of colors that can be represented by each pixel in the current picture, that is, the number of bits in the picture. t can be seen from equation (11) that minimizing MSE is equivalent to maximizing PSNR. PSNR calculates the error between pixels at the same position. In the calculation process, the influence of visual perception characteristics on the image quality is not considered, so occasionally, although the PSNR value is high, the image quality that people subjectively feel is not improved. Due to the inability to quantitatively analyze the image perception quality, PSNR is still the most commonly used evaluation method in the field of image super-resolution [15].

Structural similarity (Structural SIMilarity, SSIM) can simultaneously compare the similarity of image brightness, contrast, and structure. The value range of SSIM is [0,1]. The higher the value, the more similar the reconstruction result is to the original image structure.

$$\begin{aligned} l(x,y) &= \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \\ c(x,y) &= \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2'} \\ s(x,y) &= \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3} \end{aligned} \quad (12)$$

$$SSIM(x,y) = (x,y)^{\alpha}\cdot c(x,y)^{\beta}\cdot s(x,y)^{\gamma} \quad (13)$$

Among them,μx and μy are the mean values of the two images respectively,σx and σy respectively represent the variances of the two images,σxy represents the covariance between the two images.SSIM can better evaluate the image perception quality, so it is widely used in the field of image super-resolution.

V. DATA SET INTRODUCTION

Nowadays, there are many public data sets that can be used for image SR tasks. These data sets contain different image content and quantity, and can comprehensively test the performance of super-resolution reconstruction methods. Table 1 lists the commonly used data sets and briefly introduces them. Some data sets have been divided into fixed training sets, validation sets and test sets, and some larger data sets are often used as training sets, such as ImageNet [22], DIV2K [23] and Flickr2K [15]. Researchers can also divide the training set, validation set, and test set on the data set according to different usage requirements, or increase the training set through image rotation and other methods, or combine multiple data sets for training. In the experiment, the images in the data set need to be correspondingly cropped to adapt to different SR network training.

TABLE I. INTRODUCTION TO COMMONLY USED DATA SETS

Data set name	Number of pictures	Image Format	Brief description of the data set
Set5[24]	5	PNG	The pictures included are from babies, birds, butterflies, children's heads, and a lady.
Set14[25]	14	PNG	The included pictures come from characters, animals, insects, flowers, vegetables, comedians, etc.
Berkeley segmentation[26]	500	JPG	Referred to as BSD500. The pictures included are from animals, buildings, food, people and plants, etc. One of the 100 or 300 pictures is often used, which is called the BSD100 BSD300 data set.
Urban100[27]	100	PNG	The pictures included are mainly different types of urban buildings.
Manga109[28]	109	PNG	The pictures included are all from Japanese manga.
T91[29]	91	PNG	The included pictures come from vehicles, flowers, fruits and human faces. Often used as a training set.
General-100[30]	100	BMP	The pictures included are from animals, daily necessities, food, plants, people, etc. It is also often used as a training set.
DIV2K[23]	1000	PNG	A dataset of high-definition pictures, with pictures from natural environments, landscapes, handicrafts and people, etc. Among them, 800 pictures are often used as training sets.
Flickr2K[15]	2650	PNG	The included pictures come from people, animals, landscapes, etc., and are often used as larger training sets.

VI. SUMMARY AND OUTLOOK

With the development of deep learning, the network of super-resolution reconstruction algorithms is becoming more and more complex, and the reconstruction effect is getting better and better, and it has reached a very high level. The super-resolution method based on deep learning can automatically extract image features, acquire prior knowledge and learn from massive training data, and have a variety of distinctive training models and support from a large number of public data sets. The reconstructed image is in various evaluation indicators. All performed well. The rapid development of deep learning and the continuous improvement of hardware facilities provide very good development opportunities for the field of image super-resolution. Undoubtedly, it has become the most popular research direction in the field of super-resolution research.

Although the performance of existing deep learning image super-resolution reconstruction algorithms has been greatly improved compared to before, far surpassing traditional algorithms, there is still much room for improvement. Looking to the future, research on super-resolution can be carried out from the following aspects:

1) Improve network performance. Improving the image effect after reconstruction has always

been a hot issue for researchers, but for different usage requirements, the performance requirements of the network are also different. For example, in video surveillance images, the reconstructed image needs to have a good visual perception effect and high reconstruction efficiency; in medical image reconstruction, the reconstructed image needs to have better texture details, while ensuring authenticity and credibility. Therefore, improving the reconstruction efficiency, obtaining better visual perception effects, better texture details, higher magnification and other aspects are the focus of future research to continue to improve the performance of super-resolution networks.

2) Application of image super-resolution in various fields. Super-resolution has high application value in video surveillance, medical images, satellite remote sensing imaging, criminal investigation analysis, and face recognition. It optimizes the reconstruction effect in the corresponding scene and has a practical application value for improving image super-resolution. Significant

3) Model evaluation problem. For the problem of image super-resolution reconstruction, the evaluation index directly affects the model optimization measures. At present, PSNR and SSIM are usually used as objective evaluation indicators. Although the calculation is simple and convenient, the consistency with the visual perception effect is poor. Subjective evaluation indicators require high costs, consume a lot of manpower and material resources, and have greater limitations in practical applications. Therefore, according to the characteristics of different reconstruction methods and the needs of different scenes, corresponding evaluation indicators should be designed, which is of great significance for improving the application value of image super-resolution.

REFERENCES

[1] Tsai R. Multiframe image restoration and registration [J]. Advance Computer Visual and Image Processing, 1984, 1: 317-339.

[2] Viet Khanh Ha,Jin-Chang Ren,Xin-Ying Xu,Sophia Zhao,Gang Xie,Valentin Masero,Amir Hussain.Deep Learning Based Single Image Super-resolution:A Survey [J]. International Journal of Automation and Computing,2019, 16(04) :413-426.

[3] Huang Jian, Zhao Yuanyuan, Guo Ping, Wang Jing. A review of single image super-resolution reconstruction methods based on deep learning [J]. Computer Engineering and Applications, 2021, 57(18): 13-23.

[4] Zhang Kaibing, Zhu Danni, Wang Zhen, Yan Yadi. A review of super-resolution image quality evaluation [J]. Computer Engineering and Applications, 2019, 55(04): 31-40+47.

[5] Dong C, Loy C C, He K, et al. Learning a deep convolutional network for image super-resolution[C]//European conference on computer vision. Springer, Cham, 2014: 184-199.

[6] Xu Ran, Zhang Junge, Huang Kaiqi. Image super-resolution algorithm using dual-channel convolutional neural network [J]. Journal of Image and Graphics, 2016, 21(5): 9.

[7] Tang Yanqiu, Pan Hong, Zhu Yaping, Li Xinde. A review of image super-resolution reconstruction research [J]. Chinese Journal of Electronics, 2020, 48(07): 1407-1420.

[8] Liu Yuefeng, Yang Hanxi, Cai Shuang, Zhang Chenrong. Single image super-resolution reconstruction method based on improved convolutional neural network [J]. Computer Applications, 2019, 39(05): 1440-1447.

[9] Shi W , Caballero J , F Huszár, et al. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.

[10] Xie Haiping, Xie Kaili, Yang Haitao. Research progress of image super-resolution methods [J]. Computer Engineering and Applications, 2020, 56(19):34-41.

[11] Kim J, Lee J K , Lee K M . Accurate Image Super-Resolution Using Very Deep Convolutional Networks[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2016.

[12] Wang Jiaming, Lu Tao. Satellite image super-resolution algorithm based on multi-scale residual deep neural network [J]. Journal of Wuhan Institute of Technology, 2018, 40(04): 440-445.

[13] Zhang Y, Li K, Li K, et al. Image Super-Resolution Using Very Deep Residual Channel Attention Networks[C]// 2018.

[14] Kim J, Lee J K, Lee K M. Accurate Image Super-Resolution Using Very Deep Convolutional Networks[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2016.

[15] Wang X, Yu K, Wu S, et al. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks [J]. Springer, Cham, 2018.

[16] Johnson J, Alahi A, Fei-Fei L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution[C]// European Conference on Computer Vision. Springer, Cham, 2016.

[17] Mishiba K, Suzuki T, Ikehara M. Edge-adaptive image interpolation using constrained least squares[C]// IEEE International Conference on Image Processing. IEEE, 2010.

[18] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition [J]. Computer Science, 2014.

[19] Gatys L A, Ecker A S, Bethge M. Image Style Transfer Using Convolutional Neural Networks[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.

[20] Zhou W, Bovik A C. A universal image quality index [J]. IEEE Signal Processing Letters, 2002, 9(3):81-84.

[21] Zhou W, Bovik A C, Sheikh H R , et al. Image quality assessment: from error visibility to structural similarity [J]. IEEE Trans Image Process, 2004, 13(4).

[22] Russakovsky O, Deng J, Su H , et al. ImageNet Large Scale Visual Recognition Challenge [J]. International Journal of Computer Vision, 2014:1-42.

[23] Agustsson E, Timofte R . NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2017.

[24] Bevilacqua M, Roumy A, Guillemot C, et al. Low-Complexity Single Image Super-Resolution Based on Nonnegative Neighbor Embedding [J]. Bmvc, 2012.

[25] Zeyde R. On single im-age scale-up using sparse repre-sentation [J]. Curves&Surfaces, 2010.

[26] Martin D, Fowlkes C, Tal D, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics[C]// IEEE International Conference on Computer Vision. IEEE, 2002.

[27] Huang J B, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars[C]// IEEE. IEEE, 2015.

[28] Fujimoto A, Ogawa T, Yamamoto K, et al. Manga109 dataset and creation of metadata[C]// the 1st International Workshop. ACM, 2016.

[29] Yang, J.Wright, J.Huang, T.Ma, Y. Image Super-Resolution Via Sparse Representation [J]. IEEE Transactions on Image Processing, 2010, 19(11):2861-2873.

[30] Chao D, Chen C L, Tang X. Accelerating the Super-Resolution Convolutional Neural Network[C]// European Conference on Computer Vision. Springer International Publishing, 2016.

基于深度学习的图像超分辨率重建研究

蒋婧宇
计算机科学与工程学院
西安工业大学
西安，中国
邮箱：jjy1030@126.com;

焦炎
计算机科学与工程学院
西安工业大学
西安，中国
邮箱：jiaoyan@st.xatu.edu.cn;

赵莉
计算机科学与工程学院
西安工业大学
西安，中国
邮箱：332099732@qq.com;

***摘要：*图像超分辨率重建（Super-resolution,SR）旨在使用特定算法将同一场景中的低分辨率模糊图像恢复成高分辨率清晰图像。由于广泛的应用价值与理论价值,图像超分辨率重建技术成为计算机视觉与图像处理领域的一个研究热点,引起了研究者的广泛关注。与传统方法相比，深度学习方法在图像超分辨率重建领域展现出了更好的重建效果，已逐渐发展成为主流技术。因此，本文将图像超分辨率重建问题按照网络模型结构的不同进行系统分类，分为基于卷积神经网络模型的超分方法和基于生成对抗网络模型的超分方法两大类。梳理了主要的图像超分辨率重建方法，阐述了几种较为重要的深度学习超分辨率重建模型，分析比较了不同算法的优缺点及适应的应用场景，讨论了各不同类别超分辨率算法的互相融合和图像视频质量评价的方法，并对常用数据集进行了简单介绍。最后讨论了目前图像超分辨率重建技术所面临的潜在问题，并对未来的发展方向做出了全新的展望。**

关键字：图像超分辨率重建；深度学习；卷积神经网络；生成对抗网络

1. 引言

图像超分辨率重建(Super-resolution Reconstruction，SR)是指采用图像处理和机器学习方法，从同一场景中的一张或多张低分辨率(Low-Resolution，LR)图像重建具有丰富图像细节和清晰纹理的高分辨率(High-Resolution，HR)图像的过程[1]。其在视频、遥感、医学和安全监控等领域都有着重要的应用价值。随着机器学习在计算机视觉领域的迅猛发展，深度学习技术被广泛应用于解决 SR 问题中并取得很好的重建效果，如今已逐渐成为主流。

现有的超分辨率重建算法通常分为三大类：基于插值的方法，这类算法虽简单但提供过于平滑的重建图像，失去了部分细节，产生了振铃效应；基于建模的方法，相较于插值法该类算法重建效果较好，但当面临很大的计算量时，计算过程耗时长，求解困难且受放大因子的影响较大；基于学习的方法，该类算法解决了对尺度缩放因子敏感的问题且重建效果最好，是目前研究的主流方向[3]。

卷积神经网络（convolution neural network，CNN）和生成对抗网络（generative adversarial network，GAN）是目前主流的网络模型，当缩放因子较大时采用这两个网络模型都可以很好

的恢复图像的高频信息，使输出的图像更接近原始真实图像[3]。

2. 超分辨重建方法的网络模型

根据网络模型结构的不同，基于深度学习的图像超分辨率重建方法可以分为以下两大类：①基于卷积神经网络（Convolutional Neural Network，CNN）模型的超分方法；②基于生成对抗网络（Generative Adversarial Networks，GAN）模型的超分方法。针对图像超分辨率的各类需求，产生了具有各种不同特点的超分辨率网络模型（如图 1）[4]。

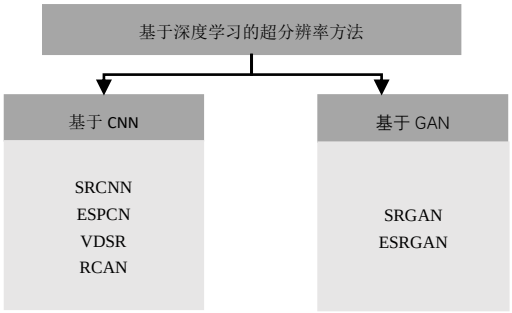


图 1 基于深度学习的超分辨率方法

2.1 基于 CNN 网络模型的超分方法

1) SRCNN

SRCNN[5]将深度学习与传统稀疏编码之间的关系作为依据，将 3 层网络划分为特征提取、非线性映射以及最终的重建。对于一个低分辨率图像，如图 2 所示，该方法先使用双三次（bicubic）插值将其放大至与目标等大小，然后进行图像块提取与表示，再通过三层卷积网络做非线性映射，重建得到的结果作为高分辨率图像输出。SRCNN 首次将卷积神经网络引入到 SR 任务中，与传统 SR 算法的分步处理不同，SRCNN 将各阶段整合到一个深度学习模型中，大幅简化了 SR 工作流程，可以视为基于深度学习的超分辨率方法的里程碑[6]。

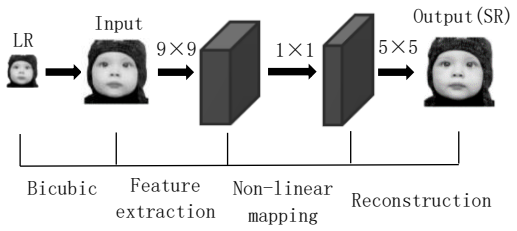


图 2 SRCNN 网络模型

虽然 SRCNN 的网络结构设计简单，在图像重建质量和速度上均优于传统的超分辨率算法，但存在以下问题： ①没有利用任何先验知识；②仅适用于单放大因子的 SR 任务，针对不同的放大因子，网络模型需要再次进行训练；③由于输入图像需要先经过插值放大至与目标等大小，图像重建全过程均在 HR 空间中进行，占用大量的内存空间，增加了计算量，同时，由插值过程产生的误差也会对重构效果产生影响，模型收敛速度较慢，训练耗时较长；④网络层数较少，卷积核感受野也较小(13 ×13)，不能很好地应用图像上下文相关信息，导致最终重建的 HR 图像纹理不清晰，算法适应性也会受限[7]。

最初是采用较小的数据集 Set5 和 Set14 训练 SRCNN 算法，训练后学习到的知识相对较少，图像重建性能受到约束，当采用相对较大的数据集 BSD200 后，重建性能显著提升，由此可见图像的重建性能受数据集大小的影响也较大[8]。

在此之后，虽然专家们又提出了各种网络模型，但是 SRCNN 仍作为一个用于评估其他网络模型性能的基准实验。

2) ESPCN

2016 年，Shi 等人提出了一种基于像素重排的 ESPCN（Efficient Sub-Pixel Convolutional Neural Network）网络模型[9]，ESPCN 的核心概念是亚像素卷积层(sub-pixel convolutional layer)，在 LR 图像上进行卷积操作来获取 LR 图像特征，再通过亚像素卷积层将 LR 空间中的特征扩充到 HR 空间，将卷积后得到的 HR 特征图进行通道重组，得到 HR 图像[10]。

ESPCN 网络主要是对 SRCNN 的重建层进行了改进，将 LR 图像作为网络输入，在重建层采用亚像素卷积层使得网络训练速度成倍提升，简单的网络结构和极高的重建速度使其非常适用于高速且对于重建性能要求相对较低的实时系统。在 ESPCN 网络中，图像尺寸放大使用的插值函数被隐式地包含在前面的卷积层中，可以进行自动学习得到。由于都是在低分辨率图像尺寸大小上进行卷积操作，因此模型效率较高。ESPCN 模型提出的亚像素卷积层在之后被广泛应用，与 FSRCNN 模型中提出的反卷积层相比，学习到的从低分辨率图像到高分辨率图像的上采样的非线性效果更好。值得注意的是，该模型还对激活函数进行了修改，采用 tanh 函数替代了 ReLU 函数，且损失函数为均方误差。

3) VDSR

VDSR(Very Deep CNN for SR)[11]是第一个提出用全局残差学习思想来解决图像 SR 问题的深度模型，是基于 SRCNN 的改进网络，其通过采用多层卷积核进行深层次卷积，既减少了参数量，又使得后面的网络层拥有更大的感受野，能够更好的利用更大区域图像上下文信息，与 SRCNN 相比，获得更好的重建效果。VDSR 的最大特点是网络层数深，图像重建效果好，训练速度比较快。由于作者发现输入的 LR 图像和输出的 HR 图像极其相似，即 LR 图像携带的低频信息与 HR 图像的低频信息极其相似[12]，所以只需要学习 HR 图像和 LR 图像之间的高频残差部分即可。VDSR 网络结构如图 3 所示，它将插值后得到与目标等大小的 LR 图像输入网络，再将这个图像与网络学到的残差相加得到最终的 HR 重建图像。自适应梯度裁剪策略是以更高的学习率来训练网络，尽管架构巨大，但仍可加快收敛速度。因此，在增加网络深度的基础上，结合残差网络和自适应梯度裁剪来加速模型训练，可以提升网络性能且重建效果更好，同时通过对不同尺度大小图像进行混合训练，VDSR 网络可以实现单一模型的多尺度 SR 重建。

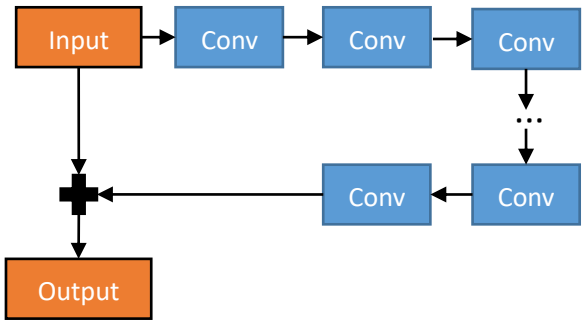


图 3 VDSR 网络结构

4) RCAN

2018 年，Zhang 等人[13]认为输入图像含有大量低频信息，现有的 SR 网络同等对待这些信息的所有通道，缺乏对这些通道的分辨学习能力，阻碍了网络的表征能力，因此提出了深度残差通道注意力网络 RCAN[13]，RCAN 是首个将注意力机制应用于 SR 问题的网络。该算法通过学习不同通道的重要性得到一个权重值，这相当于对信道间特征的相互关系进行建模，自适应调整每个信道特征，从而在有效地强化有用特征通道的同时抑制无用特征通道，更加充分利用计算资源。该模型使用局部嵌套残差结构(residual in residual)，该结构由残差组（RG）和跳远连接（LSC）组成，通过简单堆叠残差块来搭建更深的网络，并通过特征通道之间的依赖关系选择包含更多关键信息的特征通道，增强整个网络的辨识学习能力。

2.2 基于 GAN 网络模型的超分方法

1) SRGAN

生成对抗网络(Generative Adversarial Network，GAN）由 Goodfellow 等提出，它启发自博弈论中的二人零和博弈，GAN 模型中的两位博弈方分别由生成式模型（generative model）和判别式模型（discriminative model）充当[14]。SRGAN 首次将对抗训练应用于图像超分辨率重建问题中，结果显示引入对抗训练能够使网络生成更加精细的纹理细节，GAN 可以完成很多匪夷所思的生成问题，在图像生成、语音转换、文本生成领域均占有很重要地位。

如图 4 所示，SRGAN 将 LR 图像输入至生成器 G 中进行图像重建，由判别器 D 将生成图像与 HR 图像对抗训练，最后输出训练生成的图像。生成器和判别器的协同训练，使网络不仅在像素域判断生成图像与实际高分辨率图像的相似度，且更加关注其在像素空间中的分布相似度。与之前的算法相比，虽然 SRGAN 在客观评价指标（如 PSNR）上相对较低，但是在视觉效果、图像细节等直观方面重建效果更佳。这与其独特的网络结构以及将感知损失(perceptual loss)和对抗损失(adversarial loss)相结合的损失函数有关，其中感知损失是利用卷积神经网络提取出的特征，通过比较生成图片和目标图片经过卷积神经网络后的特征差别，使生成图片和目标图片在语义和风格上更相似。对抗损失由 GAN 提供，根据图像是否可以成功欺骗判别网络进行训练。SRGAN 是追求视觉体验发展中的一个里程碑，SRGAN 显著提高了基于 PSNR 的方法重建的整体视觉质量。

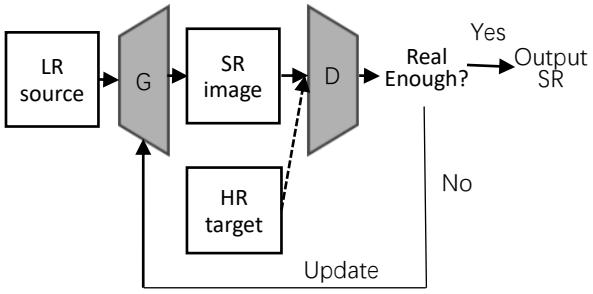


图 4 SRGAN 基本结构示意图

2) ESRGAN

为了提升 SRGAN 模型的重建精度，在此基础上，ESRGAN[15]模型在三个方面上进行了改进：对生成器架构进行了更改，通过去除批量归一化（BN）层以降低重建图像中产生的伪影，并引入一种有更高重建精度的密集残差块（RDDB）来提升模型的结构，使之具有更大容量且更易于训练，有助于提高泛化能力，降低计算复杂度和内存使用；改进感知域损失函数，使用激活前的 VGG 特征，这个改进会提供清晰的边缘和更符合视觉的结果，可以更好地保持图像亮度一致性和恢复更好的细节纹理；增强了判别器的判别能力，使用相对判别器，使其判别不再是真伪绝对值而是预测生成图像与真实图像的相对距离。

与 SRGAN 模型相比，在大部分的标准测试图像上，ESRGAN 模型能在保持较好细节的同时使重建图像中的噪声变少。大量实验表明，增强型 SRGAN，称之为 ESRGAN，在锐度和细节上始终优于最先进的方法。

2.3 小结

近年来，深度学习技术蓬勃发展，基于深度学习的超分辨率重建方法已逐步成为图像超分辨率的主流，并发展出了一系列以 CNN 和 GAN 为基础的网络模型，2.1-2.2 节分别介绍了基于 CNN 网络模型和基于 GAN 网络模型的超分方法。整体来看，算法的重建性能在不断提升，重建后的图像与原始图像越来越接近。基于 GAN 网络模型的算法重建得到的图像纹理细节较优，提升了图像的真实感，人眼感知效果较好。与之相比，基于 CNN 网络模型的算法网络复杂度更低，训练难度更小，重建结果精度更高，但会产生伪影且重建速度更慢。

3. 损失函数构建

在图像超分辨领域，损失函数是用来度量生成的图像与实际高分辨率图像之间的差异。目前，在超分辨领域中，损失函数发挥着重要的作用，常用的损失函数有基于像素的损失函数和基于感知的损失函数。

3.1 像素损失函数

像素损失主要是度量预测图像和目标图像之间的像素差异，主要包括了 L_1（平均绝对误差）损失和 L_2（均方误差）损失。

L_1 损失，也被称为最小绝对偏差（LAD）和最小绝对误差（LAE），计算的是实际值与目标值之间绝对差值的总和。在有监督的图像超分辨率任务中，目标是使得生成的图像（SR）尽可能接近真实的高分辨率图像（HR），使用 L_1 损失计算的则是 SR 和 HR 对应像素位置的值

之间的误差。平均绝对值误差（MAE）的基本表达式如式（1）所示：

$$L_{MAE}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - F(X_i,\theta)\right| \quad (1)$$

其中，L(θ)表示网络需要优化的损失函数，n 表示训练样本的数目，θ 表示深度神经网络的参数，$F(X_i, \theta)$为网络重建后的图像，Y_i 是表示对应的 HR 图像。L_2损失，也叫均方误差损失（MSE），计算的是实际值与目标值之间绝对差值的平方总和，使得基于深度学习的图像超分辨率模型在性能方面有了很大的提升，是最常用的损失函数。但是 MSE 损失能够对大的损失进行惩罚，对于小的损失上却无所作为，因此会产生模糊的图像。其基本表达式如式（2）所示：

$$L_{MSE}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left\|Y_i - F(X_i,\theta)\right\|^2 \quad (2)$$

其基本符号意义与式（1）中一致。最小均方误差的应用有效解决了 SR 重建图像与目标 HR 图像之间差值衡量问题，使得基于深度学习的图像 SR 模型相对传统基于学习的 SR 重建模型有了较大的提高。

与 MSE 相比，MAE 有个优点，即它对离群点有更好的鲁棒性，更具有包容性。由式（1）可知，MAE 计算的是误差 Yi-F(Xi, θ)的绝对值，所以无论是 Yi-F(Xi, θ)>1 还是 Yi- F(Xi, θ)<1，没有平方项的作用，惩罚力度都是相同的，所占权重相同。

由于这些损失函数分别对每个像素向量的类预测进行评估，然后对所有像素进行平均，因此它们断言图像中的每个像素都具有相同的学习能力。

与 L_1 损失函数相比，L_2 损失函数会放大最大误差和最小误差之间的差距（比如 2*2 和 0.1*0.1），另外 L_2 损失函数对异常点也比较敏感。

由于峰值信噪比（PSNR）的定义（见 4.1）与像素差相关度较高，且最小化像素损失等同于将 PSNR 最大化，所以像素损失函数逐渐成为被普遍使用的损失函数。但由于其并不探究图像质量问题(例如，感知质量[16]，纹理细节[17])，因此结果往往欠缺高频细节内容，产生过于平滑的纹理细节。

3.2 感知损失函数

对于图像风格化，图像超分辨率重建等任务来说，早期都使用了图像像素空间的 L_2 损失，但是 L_2 损失与人眼感知损失的图像质量并不契合，恢复出来的图像细节往往表现较差。如今的研究中，L_2 损失逐步被人眼感知损失所替代。人眼感知损失也被称为感知损失（perceptual loss），其与 MSE 损失采用图像像素进行求差的不同之处在于所计算的空间不再是图像空间。基于感知的损失函数可以恢复更多的高频信息，使重建性能更好。目前，基于感知的损失函数主要有内容损失和对抗损失。

内容损失函数又分为特征重建损失函数和风格重建损失函数。Bruna 等人[18]首次提出特征重建损失函数，通过预训练的 VGG19 网络分别提取重建 SR 图像与 HR 图像在特征空间中对应的特征映射且进行比较，其表达式如下：

$$L_{VGG/j} = \frac{1}{W_j H_j}\sum_{h=1}^{H_j}\sum_{w=1}^{W_j}\left(\phi_j(Y)_{x,y} - \phi_j(G_{\theta_G}(X))_{x,y}\right)^2 \quad (3)$$

其中，W_j 和 H_j 表示第 j 幅特征图的宽与高，ϕ_j 表示 VGG19 网络内第 j 次卷积获得的特征映射，X 表示原始 LR 图像，Y 为重建后的 HR 图像，G_{θ_G} 表示网络生成的 SR 图像。Gatys 等人[19]为了使重建 SR 图像与 HR 图像的纹理细节等保持一致，在特征重建损失函数的基础上又提出

了风格重建损失函数，该函数定义了一个 Gram 矩阵，计算公式如下：

$$G_j^{\phi}(x)_{c,c'} = \frac{1}{C_j W_j H_j} \sum_{h=1}^{H_j} \sum_{w=1}^{W_j} \phi_j(x)_{h,w,c} \phi_j(x)_{h,w,c'} \quad (4)$$

经过 VGG 网络提取到的特征图大小为 $C_j \times W_j \times H_j$。随后在对应层中计算 Gram 矩阵的欧式距离差并相加得到风格重建损失函数，如下式：

$$L_{style/j}^{\phi} = \left\| G_j^{\phi}(Y) - G_j^{\phi}(G_{\theta_G}(x)) \right\|^2 \quad (5)$$

SRGAN 网络首次提出对抗损失，其基本形式如下：

$$L_{Gen}^{SR} = \sum_{n=1}^{N} -lbD_{\theta_D}(G_{\theta_G}(x)) \quad (6)$$

其中，N 表示图像数量，θ_D 表示鉴别网络的参数，θ_G 表示生成网络的参数，表示生成图像，是真实的 HR 图像的概率。网络最终的优化目标为一个最小最大化问题：

$$\min_{\theta_G} \max_{\theta_D} E_{I^{HR} \square P_{train}(I^{HR})} \left[lbD_{\theta_D}(I^{HR}) \right] + E_{I^{HR} \square P_G(I^{HR})} \left[1 - lbD_{\theta_D}(G_{\theta_G}(I^{LR})) \right] \quad (7)$$

其中，P_{train}（I^{HR}）表示 HR 图像分布，P_G（I^{LR}）表示原始 LR 图像分布，对抗训练使得生成的 SR 图像与原始的 HR 图像高度相似，使得判别网络难以辨别，最终得到能够以假乱真的 SR 图像。SRGAN 的判别目标为输入图像是否为真，与 SRGAN 不同，ESRGAN 用相对平均判别器替代了判别网络的判别器，判别目标为预测真实 HR 图像比生成 SR 图像更真实的概率。判别网络如式（8）和式（9）所示：

$$\begin{aligned} D(x_r) &= \sigma(C(real)) \to 1 \\ D(x_f) &= \sigma(C(fake)) \to 0 \end{aligned} \quad (8)$$

$$\begin{aligned} D_{Ra}(x_r, x_f) &= \sigma(C(real) - E\left[C(fake)\right]) \to 1 \\ D_{Ra}(x_f, x_r) &= \sigma(C(fake) - E\left[C(real)\right]) \to 0 \end{aligned} \quad (9)$$

其中，real 表示真实 HR 图像，fake 表示生成 SR 图像，C(real)表示鉴别器判断结果，E[C(fake)]表示鉴别网络判断结果的平均值。其中 σ 表示 Sigmoid 激活函数，通过改进判别器帮助网络学习更敏锐的边缘和更多的纹理细节。

4. 图像超分辨质量评价

通过超分辨技术重构后得到的高分辨率图像，与真实高分辨率图像的相似度越高，则表示图像超分辨算法性能越好。一般来说，主要通过两种客观的量化指标进行评价，包括峰值信噪比(Peak signal-to-noise ratio, PSNR)[20]和结构相似度(Structual similarity, SSIM)[21]。为了比较的公平性，在中心下垂的 y 通道上对所有 PSNR（dB）和 SSIM 度量进行计算。评价指标值越高，说明重建结果和原始图像差异性越小，逼真度越高。

PSNR 是一种最常用的图像重构质量评价方式，基于图像的均方误差（MSE），MSE 如下式(10)所示：

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \left[I(i,j) - K(i,j) \right]^2 \quad (10)$$

其中，I 表示原始高分辨率图像，K 表示重建后的图像，m 和 n 分别表示图像的长和宽。PSNR 定义如下：

$$PSNR = 10 \log\left(\frac{MAX_I^2}{MSE}\right) \quad (11)$$

其中，对数底为自然底数 e，若不额外说明均采用此底数。MAX_I 为当前图片中每一个像素能表示的颜色最大个数，即图片的 bit 数。由式

(11)可以看出，最小化 MSE 相当于最大化 PSNR。PSNR 计算的是同位置像素之间的误差，在计算过程中未考虑可视化感知特性对图片质量的影响，所以偶尔会产生虽然 PSNR 值很高但人主观感受到的图像质量并没有提高的现象。由于无法对图像感知质量进行定量分析，PSNR 仍作为图像超分辨领域最常用的评价方式。

结构相似度（Structural SIMilarity，SSIM）可以同时比较图像亮度、对比度、结构这三方面的相似度。SSIM 的取值范围为［0，1］，值越高就越表明重建结果和原始图像结构越相似。

$$\begin{aligned} l(x,y) &= \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \\ c(x,y) &= \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2'} \\ s(x,y) &= \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3} \end{aligned} \tag{12}$$

$$SSIM(x,y) = (x,y)^{\alpha} \cdot c(x,y)^{\beta} \cdot s(x,y)^{\gamma} \tag{13}$$

其中，μ_x，μ_y分别为两张图像的均值，σ_X，σ_Y分别代表两张图像的方差，σ_{xy}表示两张图像之间的协方差。SSIM 能够较好地评估图像感知质量，所以在图像超分辨领域被广泛使用。

5. 数据集介绍

如今有很多公开数据集可用于图像 SR 任务，这些数据集包含的图像内容、数量等各不相同，可对超分辨重建方法的性能进行综合测试。表 1 列出了常用的数据集，并进行了简单介绍。一些数据集已被划分为固定的训练集、验证集和测试集，也有一些较大的数据集常被用做训练集，如 ImageNet[22]、DIV2K[23]和 Flickr2K[15]等。研究人员也可根据不同的使用需求在数据集上自行划分训练集、验证集和测试集，或者通过图像旋转等方式进行训练集扩增，或者联合多个数据集进行训练。在实验中，需要将数据集中的图像进行对应裁剪，以适应不同 SR 网络训练。

表 1 常用数据集介绍

数据集名称	图片数量	图片格式	数据集简单描述
Set5[24]	5	PNG	包含的图片分别来自婴儿、鸟、蝴蝶、小孩的头部和一个女士。
Set14[25]	14	PNG	包含的图片来自人物、动物、昆虫、花、蔬菜和喜剧演员等。
Berkeley segmentation[26]	500	JPG	简称 BSD500。包含的图片来自动物、建筑、食物、人和植物等。经常使用其中的 100 张或 300 张图片，称为 BSD100 BSD300 数据集。
Urban100[27]	100	PNG	包含的图片主要是不同类型的城市建筑物。
Manga109[28]	109	PNG	包含的图片均来自日本漫画。
T91[29]	91	PNG	包含的图片来自车辆、花、水果和人脸等。常被用作训练集。
General-100[30]	100	BMP	包含的图片来自动物、日常用品、食物、植物和人物等。也常用作训练集。
DIV2K[23]	1000	PNG	高清图片数据集，图片来自自然环境、风景、手工艺品和人物等。其中的 800 张图片经常作为训练集使用。
Flickr2K[15]	2650	PNG	包含的图片来自人物、动物和风景等，常作为较大规模训练集使用。

6. 总结与展望

随着深度学习的发展，超分辨率重建算法的网络越来越复杂，重建效果也越来越好，已经达到了很高的水平。基于深度学习的超分辨率方法能够自动提取图像特征，从海量训练数据中获取先验知识并进行学习，拥有各类各具特色的训练模型和大量公开数据集的支持，重建图像在各评价指标上都表现良好。深度学习的快速发展，以及硬件设施的不断完善，为图像超分辨领域提供了非常好的发展机遇。毋庸置疑，其已成为超分辨率研究领域的最热门研究方向。

虽然现有的深度学习图像超分辨率重建算法的性能较之前已经有了很大提升，远超越传统算法，但还有很大提升空间。展望未来，超分辨率的研究可以从以下几个方面开展：

（1）提升网络性能。提升重构后的图像效果一直是研究者们关注的热点问题，但对于不同的使用需求，对网络的性能要求也不同。例如，视频监控图像中，需要重建图像视觉感知效果好，重建效率高；医学图像重建中，需要重建图像具有较优的纹理细节，同时保证真实可信。因此，提升重建效率、获得更好的视觉感知效果、更优的纹理细节、更高的放大倍数等方面，是未来继续提升超分辨率网络性能的研究重点。

（2）图像超分辨率在各领域的应用。超分辨率在视频监控、医学图像、卫星遥感成像、刑侦分析和人脸识别等方面有很高的应用价值，实现对应场景下重构效果最优化，对于提升图像超分辨率的实际应用价值具有重大意义。

（3）模型评价问题。对于图像超分辨率重建问题，评价指标直接影响着模型优化举措。目前通常采用 PSNR 和 SSIM 作为客观评价指标，虽计算简单方便，但与视觉感知效果一致性较差。主观评价指标所需成本较高，花费大量人力、物力，在实际应用中局限性较大。因此，应针对不同重建方法的特点和不同场景需求，设计相应的评价指标，对于提升图像超分辨率的应用价值具有重要意义。

参考文献

[1] Tsai R. Multiframe image restoration and registration [J]. Advance Computer Visual and Image Processing, 1984, 1: 317-339.

[2] Viet Khanh Ha, Jin-Chang Ren, Xin-Ying Xu, Sophia Zhao, Gang Xie, Valentin Masero, Amir Hussain. Deep Learning Based Single Image Super-resolution:A Survey [J]. International Journal of Automation and Computing, 2019, 16(04):413-426.

[3] 黄健,赵元元,郭苹,王静.深度学习的单幅图像超分辨率重建方法综述[J].计算机工程与应用,2021,57(18):13-23.

[4] 张凯兵,朱丹妮,王珍,闫亚娣.超分辨图像质量评价综述[J].计算机工程与应用,2019,55(04):31-40+47.

[5] Dong C, Loy C C, He K, et al. Learning a deep convolutional network for image super-resolution[C]//European conference on computer vision. Springer, Cham, 2014: 184-199.

[6] 徐冉,张俊格,黄凯奇.利用双通道卷积神经网络的图像超分辨率算法[J].中国图象图形学报，2016, 21(5):9.

[7] 唐艳秋,潘泓,朱亚平,李新德.图像超分辨率重建研究综述 [J]. 电子学报, 2020, 48(07):1407-1420.

[8] 刘月峰,杨涵晰,蔡爽,张晨荣.基于改进卷积神经网络的单幅图像超分辨率重建方法[J].计算机应用,2019,39(05):1440-1447.

[9] Shi W, Caballero J, F Huszár, et al. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.

[10] 谢海平,谢凯利,杨海涛.图像超分辨率方法研究进展[J].计算机工程与应用,2020,56(19):34-41.

[11] Kim J, Lee J K, Lee K M. Accurate Image Super-Resolution Using Very Deep Convolutional Networks[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2016.

[12] 汪家明,卢涛.多尺度残差深度神经网络的卫星图像超分辨率算法[J].武汉工程大学学报，2018, 40(04):440-445.

[13] Zhang Y, Li K, Li K, et al. Image Super-Resolution

Using Very Deep Residual Channel Attention Networks[C]// 2018.

[14] Kim J, Lee J K, Lee K M. Accurate Image Super-Resolution Using Very Deep Convolutional Networks[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2016.

[15] Wang X, Yu K, Wu S, et al. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks [J]. Springer, Cham, 2018.

[16] Johnson J, Alahi A, Fei-Fei L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution[C]// European Conference on Computer Vision. Springer, Cham, 2016.

[17] Mishiba K, Suzuki T, Ikehara M. Edge-adaptive image interpolation using constrained least squares[C]// IEEE International Conference on Image Processing. IEEE, 2010.

[18] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition [J]. Computer Science, 2014.

[19] Gatys L A, Ecker A S, Bethge M. Image Style Transfer Using Convolutional Neural Networks[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.

[20] Zhou W, Bovik A C. A universal image quality index [J]. IEEE Signal Processing Letters, 2002, 9(3):81-84.

[21] Zhou W, Bovik A C, Sheikh H R , et al. Image quality assessment: from error visibility to structural similarity [J]. IEEE Trans Image Process, 2004, 13(4).

[22] Russakovsky O, Deng J, Su H, et al. ImageNet Large Scale Visual Recognition Challenge [J]. International Journal of Computer Vision, 2014:1-42.

[23] Agustsson E, Timofte R. NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2017.

[24] Bevilacqua M, Roumy A, Guillemot C , et al. Low-Complexity Single Image Super-Resolution Based on Nonnegative Neighbor Embedding [J]. bmvc, 2012.

[25] Zeyde R. On single im-age scale-up using sparse repre-sentation [J]. Curves&Surfaces, 2010.

[26] Martin D, Fowlkes C , Tal D , et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics[C]// IEEE International Conference on Computer Vision. IEEE, 2002.

[27] Huang J B, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars[C]// IEEE. IEEE, 2015.

[28] Fujimoto A, Ogawa T, Yamamoto K, et al. Manga109 dataset and creation of metadata[C]// the 1st International Workshop. ACM, 2016.

[29] Yang, J. Wright, J. Huang, T. Ma, Y. Image Super-Resolution Via Sparse Representation [J]. IEEE Transactions on Image Processing, 2010, 19(11):2861-2873.

[30] Chao D, Chen C L, Tang X. Accelerating the Super-Resolution Convolutional Neural Network[C]// European Conference on Computer Vision. Springer International Publishing, 2016.

Application of A Capture Technique to the 3D Harbor Construction

D. C. Lo*
National Kaohsiung Marine University
Institute of Maritime Information and Technology
Kaohsiung, Taiwan
* E-mail: loderg@webmail.nkmu.edu.tw

Dong-Taur Su
National Kaohsiung Marine University
Department of Shipping Technology
Kaohsiung, Taiwan

Zi-Yao Wang
National Kaohsiung Marine University
Institute of Maritime Information and Technology
Kaohsiung, Taiwan

***Abstract*—The main purpose of this paper is to design an innovative ship handling simulation program by combining the development of Quest3D software and the Google earth harbor terrain software. The paper describes the usage of the Google earth program to retrieve the latitude and longitude elevation, through GIS conversion and 3D graphical model to generate the Quest 3D harbor construction blueprint. Projectors display the image with build-in formula calculated angle settings to achieve a better perspective for the projectors installation. A 3D virtual reality scene was built by using the Maya Model, and then we overlap with the harbor module to complete the famous world ports. The results will be tested in the ship handling simulator machine at the 3D virtual reality software development laboratory.**

Keywords-Component; Harbor Construction; Virtual Reality; Ship Handling Simulator

I. INTRODUCTION

In recent years S57 has been served as a standard format in operating the sea chart. It takes the advantage of its high accuracy in both latitudinal and longitudinal measurements as well as the precision of elevation to ensure the safety of ship docking guided by port lead ship. However the generation of such S57 nautical charts of the harbor and its 3D undersea terrain appears to be very complicated. This paper describes an alternative method of combining the Google Earth program with altitude intercept programs with virtual reality to produce a high accuracy of a complete virtual reality simulator containing both the land and sea port models with a build-in content of a harbor.

Several factors may result in a more realistic simulation allowing users to simulate the real situation with great experience, such as the rapid development of computer hardware, the maturation of the virtual reality technology, and virtual reality software with a hardware interface to form a human-computer interaction. In this study, we focus on developing the elevation matrix capture program to register records with a high degree of depth points and to further generate 3D models creating virtual reality simulator content. As computing power increases, along with the progress of web-based technology and the improvement of numerical algorithms render the advanced utilization of satellite positioning technology with Google Earth map and many are increasingly using GPS (Global Positioning System global Positioning system). To date GPS usage has reached to a level of high accuracy and reliability [5]. Currently using the commercial GPS software as a practical application has become a trend leading to the inevitable continuation of scientific development. This paper uses Google Earth software to retrieve data from the Google earth server that transmits values to the

DOI: 10.2478/ijanmc-2022-0002

software [6]. These data are processed with a geographic information system GIS (Geography Information System). Through multiple conversions these data generate 3D models and software and further apply in virtual reality Quest 3D, so that the simulator contains some depth and accuracy of terrain and texture, increasing the simulation of ship to maneuvering situations [2, 3].

The basic simulation system modules using 3D display technology have become very reliable and precised in the current simulator models on display, even though simulation also involves in the height and the depth of the model. However one draw back of the current system is that it exclusively depends on the S57 sea charts for the drawing of terrain elevation and collection of electronic as well as S57 charts is rather difficult. Further, S57 charts production depends on the accessibility of the company and other unfavorable factors.

The purpose of this study is to generate a 3-D virtual reality system to simulate the maneuvering of ships. The system will be housed in the Marine National Kaohsiung Marine University Institute of Information Technology Ships manipulate three-dimensional virtual reality software development laboratory to facilitate simulation of the operation of ship machinery hardware. Combining the hardware of simulators with virtual reality software to manipulate the ship was generated using Quest 3D software to build virtual scene models and simulator software system design. Centering on the functional module the function of the system can easily be expanded. Implementing the development of the elevation capture program and construction of the model terrain into the Quest 3D expandable modules, not only can enhance the visual effects but also improve the accuracy of the latitude and longitude and topographic characterization [7]. Ultimately, the utility of the system will greatly enhance the visibility of our research endeavor in the simulation systems and make significant academic contributions.

II. METHODS

As shown in figure 1, in the built-in features of the Google Earth software screenshots mouse can be moved to any specific location to reveal its precise latitude, longitude, altitude and elevation measurements thereby creating the map.

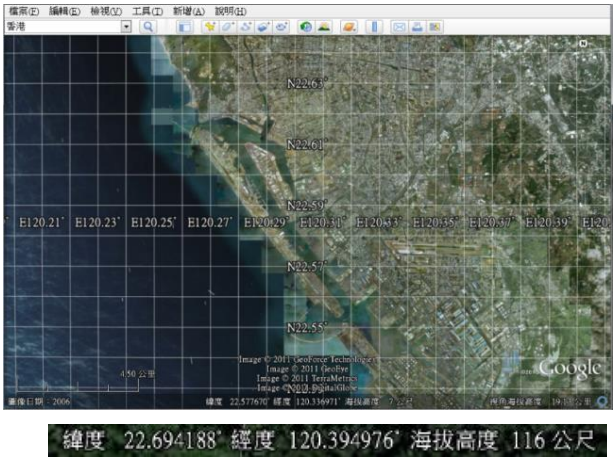


Figure 1. Range of latitude and longitude data

To develop the program with Google Earth API software the codes from both the upper left corner and lower right corner are set to make two points of latitude and longitude of the matrix, and then determine the horizontal and vertical granularity. Using the cumulative calculation to the set operation range boundary. After running the program the matrix will automatically calculate the point where all the coordinates of point range thus obtain data of all heights. After scheduling the query or interrupt data all information will be immediately stored in GeElevation.csv file name (those unfinished query due to sudden interruption can not be saved). The stored data can be directly opened in Notepad, or be imported into Excel and other applications. The data format in the order of latitude, longitude, height, will be displayed in the layout shown in Table 1

TABLE I. CAPTURE PROGRAM PLANNING TABLE PANEL

Enactment item	Enactment content
(Left Latitude)	_____ (Decimal)
(left Longitude)	_____ (Decimal)
(Lower right Latitude)	_____ (Decimal))
(lower right longitude)	_____ (Decimal)
(Horizontal spacing)	_____(Minimum accepted value of Google earth)
(vertical spacing)	_____(Minimum accepted value of Google earth)
(batch search) (cancel)	

After putting the latitude and longitude values into the parameters for the query: Latitude = 22.694188; into the parameters: longitude = 120.394976 the program will immediately display outgoing parameters: height = 116 (unit: m) access in order to avoid explosive action (thousands of queries a second issue), which is too frequent to be blocked by Google for the website limits to a single-threaded query sequence. The Google restricts a maximum of 2,500 queries and the computer network will be locked with over the limit query for 24 hours without access. If one wishes to check more items, one must copy the program to other computers with different network cards.

Figure 2. Capture program planning table panel.

Figure 2 demonstrated using C # [8] program to combine with API [9], in accordance with the original design of the interface elevation capture program which the horizontal and vertical spacing after the 6th decimal point to the most basic unit.

The calculation of the minimum sampling distance is about 25 cm and the maximum sampling distance is about two thousand five hundred kilometers. The calculated distance is in accordance with Google Earth layer for the image-based processing of precision. With the altitude closer to the ground, it yields more detailed and precise value to latitude: 11.362330 longitude:142.573144 Mariana trench perspective altitude 338.69 km of marked elevation of -8060 m. If we view the marked elevation 1000 meters above sea level is -10856 meters, the latitude for due to image slicing resulted latitude is offset by 0.01 and by 0.04 in longitude. Figures 3 and 4 show the latitude: 11.362330; longitude: 142.573144; altitude: -8060 m; perspective altitude: 338.69 km

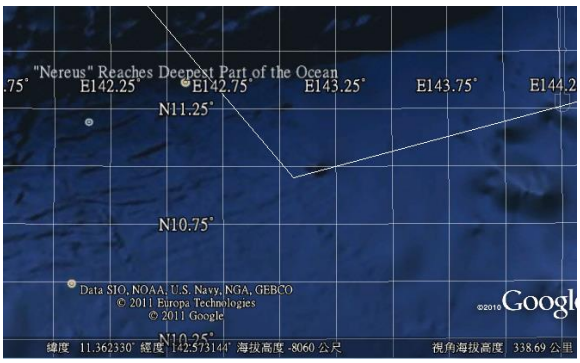


Figure 3. Capture Recognition

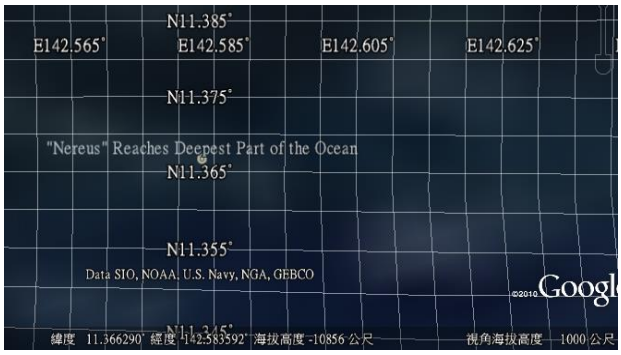


Figure 4. Capture Recognition

Latitude: 11.366290; Longitude: 142.583592; Altitude: -10856 meters; perspective altitude 1000 meters

While retrieving the data is set at a fixed elevation angle between 18 to 19 km. This captured data serves as the basic perspective of altitude to ensure under the circumstances for few variability in elevation value data thereby obtain the best capture range. Furthermore the spacing on both horizontal and vertical spacing will be adjusted by 0.01100 degree in both latitude and longitude on each capture. The Kaohsiung Harbor will be divided into 10 portions with a 2 x 11 rectangle representing each portion, which will then be assembled into a matrix for data collection, as shown in Figure 5.

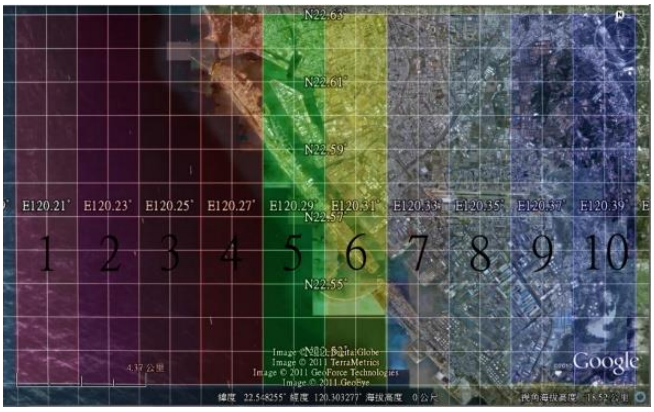


Figure 5. Capture range

The data obtained from the block 10 after a series of deductions are fed to 10 computers to perform capturing with each capture of 2360 points (maximum of 2500 points) altitude and other values, negative values represent depth of the sea water.

III. VALUE ELEVATION MODEL GENERATION

Integration of the acquired elevation data from 10 computers into one excel file revealed a total of 22,539 measurements

Through the New Geographic Coordinate System in Display XY Data WGS84 (the earth coordinate system) is set and converted to DEM as shown in Figure 6

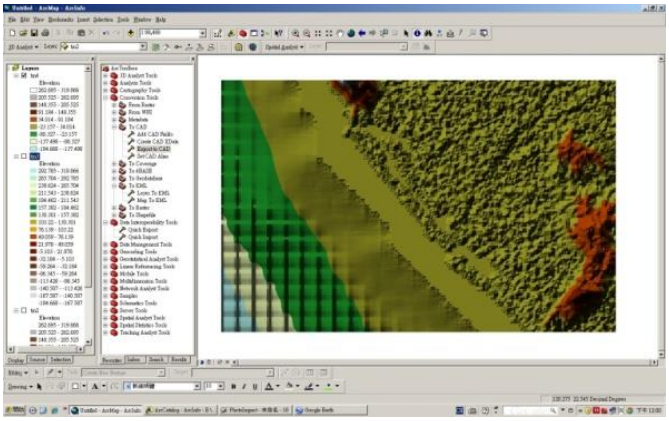


Figure 6. The DEM conversion

After conversion, the map is imported and overlaid with that of generated from Google Earth to ensure the accuracy and compatibility of the two maps as shown in Figure 7

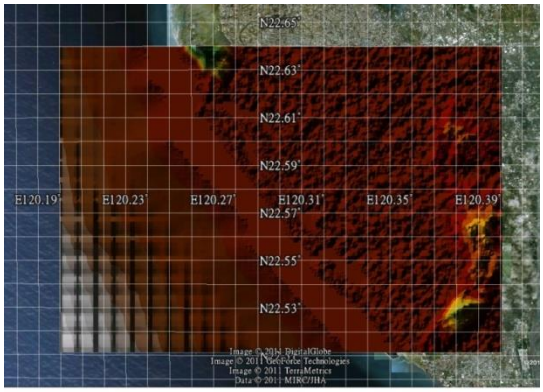


Figure 7. After making GIS import Google Earth to confirm

After confirming the precision of latitude and longitude data on the graphics, 3D models is then generated as shown in Figure 8,9 This method is used to generate the model Singapore harbor.

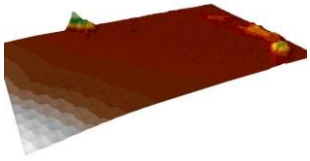


Figure 8. 3D terrain complete

Figure 9. 3D terrain complete

The majority of 3D software are able to import and export files with VRML format. Some search engines can directly upload and support a three-dimensional model of VRML format and the browser can display a Java applet providing a simple VRML experience [1.4]. To experience the full visual and interactive effects, generally it needs to install a separate browser plug-ins or stand-alone program. However using our newly developed this 3D software program enable the codes sharing to perform the DEM conversion and convert the model to other types of 3D software. This greatly improves the usage convenience, and versatility of the program. After the implementation of the VRML program, one can make good use of GIS program generated DEM files and save them as VRML compatible formats. These formats can be further converted into the post-production model called the wrl file as shown

in Figure 10, using the VRML page to confirm the implementation of GIS software generated DEM files.

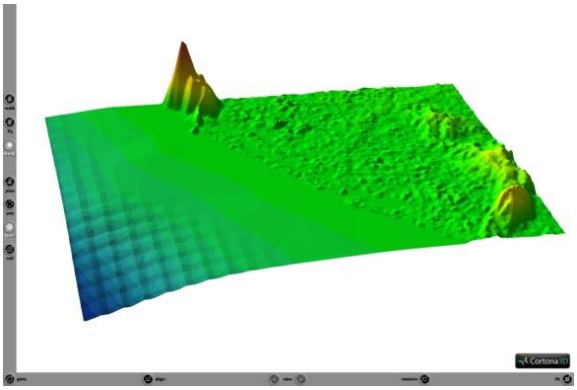


Figure 10. Vrml of the window and model

Finally, we utilized the 3D MAX 9 to convert wrl files. After several conversions to generate the port of Kaohsiung in a DEM model. Figure 11 shows both the regular map and the DEM elevation model of Kaohsiung harbor.

Figure 11. Terrain texture

To alignment texture and overlapped to complete Kaohsiung Harbour model shown in Figure 12, 13.

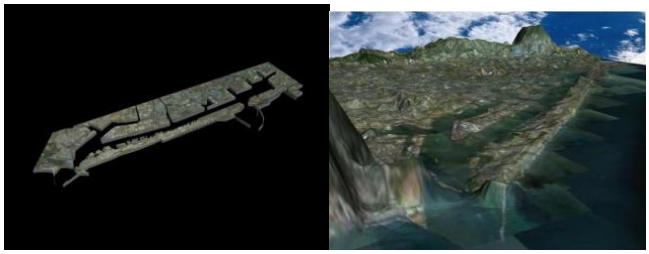


Figure 12. Kaohsiung port model

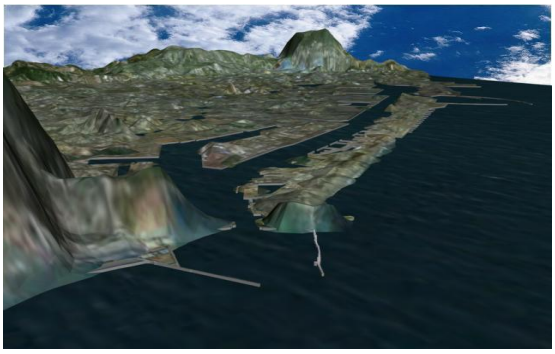


Figure 13. Build the terrain and Kaohsiung Harbor

Previously described the use of the software Google Earth virtual globe software to generate Singapore the electronic map with the following steps:

A. First, used the software in Google Earth ruler function to draw the map image of Singapore harbor as shown in Figure 14.

Figure 14. The scope of the map image to draw in Singapore

B. Use Google Earth to divide the above map into 26 parts as shown in Figure 15. Refine each part for highest resolution and precision because Google Earth satellite images were acquired at different time intervals, resulting in color difference among pictures. In order to obtain high-resolution Pixel the scale of each part was chosen at 1 km / 2,000 feet, and the height set at 12.21 km. After these manipulations all 26 segments were imported into Photoshop to generate a complete map.

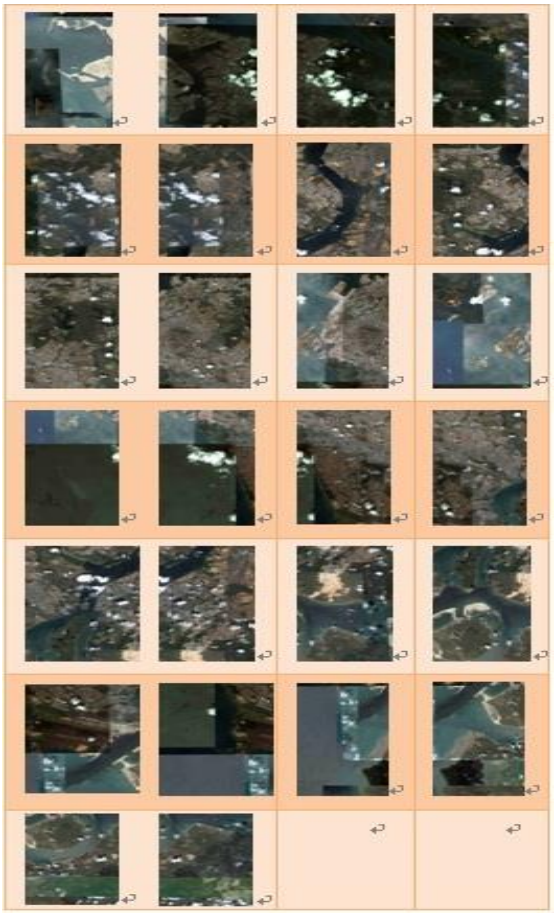


Figure 15. 26 intercept of Figure

C. Using Google Map to enlarge the vector-based maps to the same size of electronic charts as shown in Figure 16. The 26 segment maps were assembled into a rough map and used Illustrator to polished the map as shown in Figure 17.

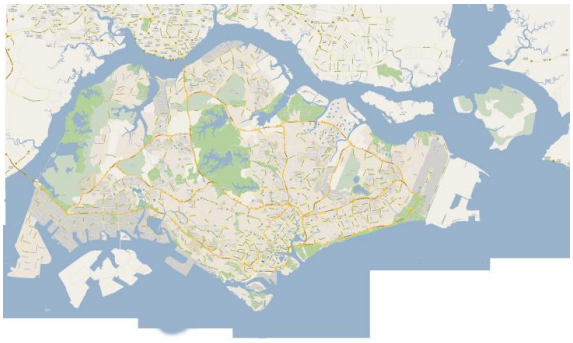


Figure 16. Vector collage map

Figure 17. Vector map transfer line of draft

D. Using the same method as described previously for building the map of Kaohsiung Harbor a good digital elevation model [7], to generate Singapore Harbour virtual 3D model is shown in the Figure 18.

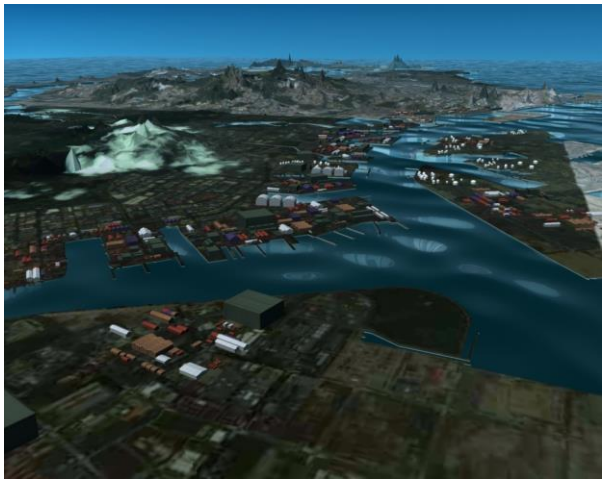


Figure 18. Singapore Harbor Model

IV. CONCLUSION

In this study, we use the Google Earth software to design elevation capture program. Through the Google Earth server to we can retrieve required range of data for any terrain in Google Earth database. These information can then be imported into GIS program and 3D drawing software to create 3D models. The process is rather rapid and with accuracy allowing the virtual simulation of port activities such as ship maneuvering. Although it is not as accurate as S57 charts however the ease of generation the 3D map and its compatible accuracy render it potentially a popular and widely use application.

Our innovative process of building 3D virtual model utilizing the satellite pictures from Google

Earth, gradually implementation of Maya software to build a scene with the surrounding areas and outer contours of the inner harbor, and production of electronic maps using Photoshop to generate the 3D virtual harbor maps is an effective method to save manpower and money without affecting the quality and accuracy of the map.

Therefore, the accuracy of the model or the latitude and longitude for positioning the accuracy of harbor the program features of this application to provide for future development and recommendations:

1) To further develop Pro version of the program allowing the production of latitude and longitude of the high-precision terrain elevation data to produce Harbor. The Institute of Ship Handling Simulator latitude and longitude positioning research will conduct a more accurate simulation of the terrain and build a more precise database for the latitude, longitude, and elevation for various harbors.

2) The use of 3D graphics software to build a variety of ship models and coordinate the information and date from ship models and the ship RS232 module corresponding to the real message to distinguish between types of ships, ship building simulator database.

3) To utilize the operating systems in ship simulator Quest 3D by adding RS232 communication, which receives the ship modules with a variety of terrains and latitude and longitude identification module the system allows accurate ship simulation in the harbor. Further the information of real vessels within the harbor will be input into the simulator to generate the real-time 3D simulation model system.

REFERENCES

[1] Ka-Hou Lam,「A Vrml/X3D-based Interactive Three Dimensional Spatial Information Mangement System」, 2005.

[2] Shan-Zhi Yang, 「Quality Assessment for Digital Elevation Models」, 2006.

[3] Yen-Nan Pan, 「Computer Aided Algorithm for Automatic Lineament Detection in Digital Elevation Model」, 2007

[4] Chun-Huang Wang, 「Research and Implementation for the Web3D Virtual Reality Based on Java3D」, 2007.

[5] Wan-Chien Chang, 「A Study of Maritime Mobile Navigation System」, 2008.

[6] Jeng-Wei Lai, 「The Study for E-learning Of High School's Geographic Education – Based On Google Earth」, 2009.

[7] Wen-Tzu Lin, 「The Application of 3D-GIS Techniques for Landscape Visualization」, 2009.

[8] Micorsoft Visual C# Developer Center http://www.microsoft.com/taiwan/vstudio/vcsharp/default.mspx

[9] Google Earth API Code http://code.google.com/intl/zhTW/apis/earth

Network Security System Design of Big Data Platform in Tai'an of Health Based on IPV9 Technology

Hongwen Zhao, Chao Lu, Yuyu Li, Guangli Li, Guotao Wen
Tai'an Branch of China Radio and Television Shandong Network Co., LTD
E-mail: tagdglcs@qq.com

***Abstract*—At present, IPv4 is the core of the Internet. When it was designed, security protection was hardly considered. Therefore, the Internet has many security loopholes, which cause information leakage or even breakdown. Compared with IPv4, IPv6 has been improved in terms of security, but IPv6 packets are not encrypted and verified by default, and the problem of network layer being attacked is still unresolved. So, the Internet based ipv6 is still faces the risk of data being monitored and tampered, which cannot effectively prevent malicious attacks. China Decimal Network Standard working group developed the future network system, which adopts the zero-trust security mechanism of verification before communication. Big Data Platform in Tai'an of Health is responsible for the unified management of all medical and health institutions in the platform, and for the management, communication and maintenance of all data. Therefore, the establishment of the network security system of this platform should pay more attention to effective, scientific and comprehensive requirements. In this paper, the Future Network (IPV9) with independent intellectual property rights in China and its encryption technology are introduced, and the network security system of Big Data Platform in Tai ' an of Health is designed, and the corresponding simulation test is carried out, which achieves the expected effect. The design of network security system of Big Data Platform in Tai'an of Health based on IPV9 can play a certain role and reference value in solving network security problems in Big Data Platform of Health.**

Keywords—Big Data; Tai'an Health Platform; Network Security Architecture; Future Network (IPV9)

I. INTRODUCTION

With the development of science and technology, the medical and health platform is gradually becoming electronic, which provides strong support for the development of the medical system by managing, communicating and maintaining the file information of all medical and health institutions in the region. In the daily use and maintenance of Health Big Data Platform, the security of information transmission has attracted more and more attention.

The network security of the Big Data Platform of Health means that the hardware, software and transmitted information in the platform can be effectively protected. The specific contents include: the platform runs reliably, continuously and stably, the network service is not interrupted, and the information in the platform is not damaged, changed or leaked due to accidental or malicious behavior. In this paper, taking Big Data Platform in Tai'an of Health as a sample, the network security architecture of Big Data Platform of Health is designed by using IPV9 technology.

Big Data Platform in Tai ' an of Health is based on IPV9 technology with national independent intellectual property rights, and a five-level

DOI: 10.2478/ijanmc-2022-0003

dedicated network for medical and health safety covering cities, districts, counties, towns, villages and families is built. Based on the medical and health information system, with the Big Data Platform in Tai'an of Health as the core, the unified management of all medical and health institutions in the platform and the unified scheduling among all business modules in the platform are realized.

II. THE SECURITY OF INFORMATION IN CHINA

With the continuous development of cyberspace and technology, network attack means emerge one after another, so it is urgent to improve the level of network security.

The rapid development of China's Internet and the continuous improvement of network application level have made great contributions to promoting economic development and social progress. At the same time, many new situations have emerged, and many new problems and challenges have been encountered in the process of network and business development, such as network security problems encountered in the application of cloud computing, big data and other new technologies. In addition, many core information technologies were invented and created by other countries, and there are certain security risks. Therefore, we should invent and popularize our own controllable information technologies. Specific analysis has the following three points:

1) The IPv4 addresses are mainly used in domestic networks.

2) The technology and means to identify the source of network attacks are insufficient.

3) The dependence on foreign information technology and products is too high.

III. THE ANALYSIS OF DEMAND

The transmission line of Big Data Platform in Tai ' an of Health is mainly wired transmission security, with the following security risks.

1) An attacker can install a stealing device on the transmission cable of the wired network to steal the data transmitted through the wired network, and analyze it by technical means, so as to obtain important information such as the account password input by the user, and even change or delete the transmitted data, which will affect the reliability and authenticity of the data.

2) After intercepting the wired data, the attacker can attack the IP address according to the IP address in the transmission information. This makes the server address or user address in the Big Data Platform in Tai ' an of Health directly exposed to the attacker.

The risk of this communication link seriously threatens the security of Big Data Platform in Tai'an of Health. Therefore, IPV9 and IPV9 encryption technologies are used to encrypt the transmission information and address at the same time in the process of communication link transmission to ensure the security of information transmission.

IV. THE KEY TECHNOLOGIES OF IPV9

The IPV9 protocol starts with basic algorithm and uses decimal algorithm to assign addresses. The address of IPv6 is 128 bits, while IPV9 is expanded to 256 bits on this basis. The IPV9 can be compatible with IPv4 and IPv6, and it can be deployed without changing the existing IPv4 or IPv6 network environment.

The protocol family of IPV9 mainly consists of the following parts: the header protocol of IPV9; Address model of IPV9 address, address text representation of IPV9, unicast address of IPV9, any on-demand address of IPV9, multicast address

of IPV9 and address required by IPV9 node; IPV9 digital message format protocol mainly includes message header extension, authentication and encryption; control message protocol of IPV9; Adjacent node detection protocol of IPV9; security architecture IPSEC protocol of IPV9; IPV9 mobile communication protocol; DNS extension protocol for digital domain name of IPV9; plug and play protocol of IPV9; Transition agreement of IPV9. IPv6 and IPV9 are both based on TCP/IP framework, but they are different. IPV9 integrates many telecommunication technologies, such as the concept of phone number, geographical location and so on.

V. THE SCHEME OF TECHNICAL

A. The encryption features of Future Network

The IPV9 can be compatible with the current IPv4 and IPv6 protocols at the same time, and at the same time, it is more in line with the requirements of future long-term network development, and it is a safe and reliable bridge to the future network.

Using IPV9 technology to set up IPV9 logical isolation private network has obvious advantages compared with the previous pure IPv4 private network. The communication address in the IPV9 private network uses the brand-new IPV9 address, which is compatible with IPv4 and IPv6 addresses. When using IPv4 address for data transmission, only the application layer data can be encrypted, but the address cannot be encrypted. Attackers can attack according to the address in the data. IPV9 protocol can not only encrypt the application layer of transmitted data, but also encrypt the IP address of the network layer. At the same time, the result of address encryption is different every time, which plays a better security effect.

B. The scheme of deployment

The user access terminal of the dedicated line of medical institutions uses the special equipment IPV9 router to ensure logical isolation from other network connection systems of users' office computers.

Use 1000M-IPV9 router at the user side of the hospital access terminal. 1000M-IPV9 router is connected to the hospital client as a dedicated terminal device. Because the device uses IPV9 address to access the network, it can act as a firewall to ensure isolation from other network-connected office systems on the hospital computer. In addition, at the data transmission level, the data is in the form of IPV9 messages, and the IPV9 private network is set up. By allocating IPV9 addresses, the network layer encryption is realized, which is more secure and reliable than IPv4 networks.

With the establishment of IPV9 private network, the networking mode and network topology of Health Tai 'an Big Data Platform are basically unchanged, except that IPv4/IPV9 routers and protocol conversion routers are added in the networking mode and network topology, and other functions such as users' online habits remain unchanged. IPV9 private network deployment is shown in Fig 1.

The core area consists of two data center-level switches, both of which use VRRP technology to back up each other. When one of the switches fails, the business on this switch will automatically switch to the other switch, which will not affect the normal business use and enhance the availability of the platform.

The access area is the access area for all health and medcal institutions. Now, six 10 Gigabit firewalls are used for isolation and convergence, one for each county.

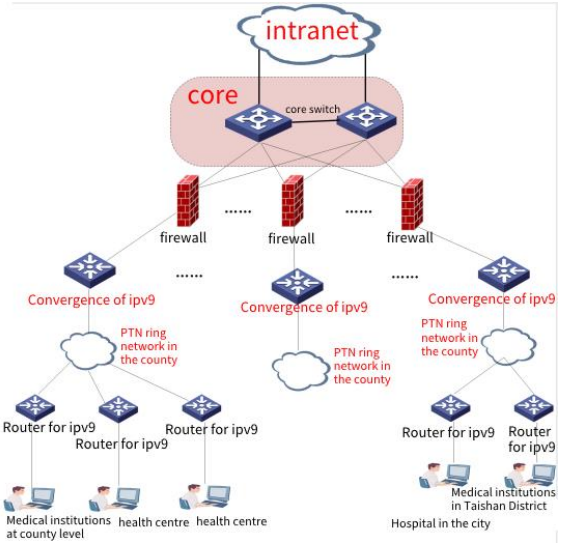


Figure 1. Deployment of IPV9 Private Network of Big Data Platform in Tai'an of Health

The network architecture of each county is the same, with IPV9 convergence router as the core, 10 Gigabit uplink data center firewall and PTN ring network as the link to ensure bandwidth and security. IPV9 convergence router encrypts and decrypts the address of the data transmitted by the access users, which makes the information transmission more secure, and accesses the PTN dedicated ring network of the county through the switch.

The IPV9 access routers with gigabit devices are deployed at the client side of all medical institutions. According to different network environments and security requirements of various institutions, strategies such as address translation, logical isolation and firewall protection are adopted.

The user still uses the address of IPv4. When the IPv4 message passes through the access router of IPV9, the header of IPV9 is automatically encapsulated. When the message passes through the IPV9 convergence router in the core area, the header of IPV9 is unsealed and restored to the IPv4 message. In this way, all IPv4 applications can run on IPV9, as shown in Fig 2. It can be seen that the network of IPV9 can be built without affecting and changing the use of IPv4 by existing terminals. Based on this, Big Data Platform in Tai'an of Health has gradually built the backbone network of IPV9.

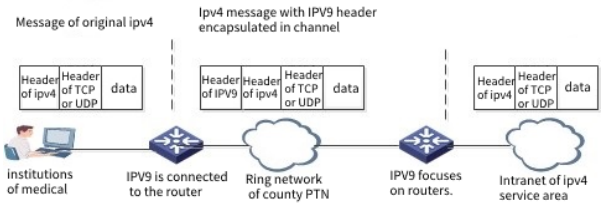


Figure 2. Encapsulation and decapsulation of IPV9 message

C. The scheme of encryption

In order to prevent the data of medical and health institutions in the access area from being tampered with or destroyed during the transmission to the core area, the Big Data Platform in Tai ' an of Health uses IPV9 encryption technology on the basis of IPV9 backbone network to encrypt the transmission. Different from the current single application encryption method, IPV9 innovatively designed address encryption, which extended the security protection to the network layer and greatly improved the security of information transmission.

In the Big Data Platform in Tai ' an of Health, the IPV9 convergence router and IPV9 access router deployed in the access area enable the conversion protocol between IPv4 and IPV9 at the same time, and enable the encryption technology of IPV9. The configuration is as follows (the eth0 network interface of two routers is the communication port of two routers):

set interface state eth0 up // Set the network interface of eth0

reverse enable eth0 4to9 // Set eth0 as the network interface supporting the conversion protocol between IPv4 and IPV9.

reverse keyset key// Set the key

The encryption principle of IPV9 in Big Data Platform in Tai ' an of Health is as follows: IPV9 convergence router A broadcasts the public key on the designated interface according to the configuration. When any device connected with the interface (such as access router B) sends data to the router, it can choose to use the public key to encrypt the data or address, and then send the encrypted IPV9 data to convergence router A and then decrypt it, thus ensuring the confidentiality of data transmission. Similarly, the same is true for IPV9 encrypted transmission from convergence router A to access router B. As shown in Fig 3.

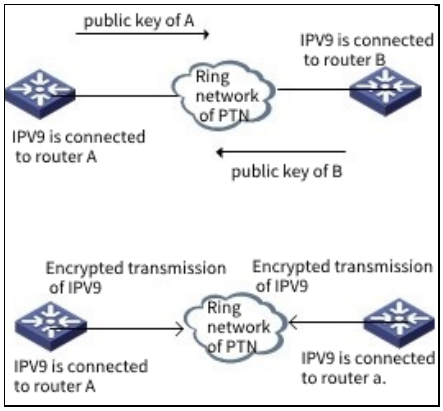


Figure 3. Key transmission and encrypted transmission of IPV9

IPV9 communication protocol has a reasonable message structure design and clear message item functions. IPV9 protocol is superior to IPv4 in address space, service quality and security. The address expression and header structure of the data message of IPV9 protocol are different from those of IPv4 or IPv6 protocol, so the data message header of IPV9 protocol will not be recognized by IPv4 or IPv6 system and will not be directly spread in these systems. At present, all hacker attacks and all online eavesdropping software are developed based on IPv4, and IPV9 routers and network cards will not release these eavesdropping and hacker attack data packets, thus building a Great Wall against hacker attacks and wanton stealing of online information. IPV9 enables China to realize the security and controllability of the underlying Internet protocol. And the integration of advanced design concepts such as address and data double encryption mechanism has greatly improved the security of network information.

VI. THE EXPERIMENT OF ENCRYPTION SIMULATION OF IPV9

In the design scheme of this paper, the transmission from the user-side IPV9 access router to the IPV9 convergence router adopts IPV9 encryption processing to ensure the transmission security at both ends of the transmission, including address encryption and data encryption of IPV9 transmission. In this experiment, the address encryption in IPV9 transmission is simulated and tested. The topology diagram is shown in Fig 4.

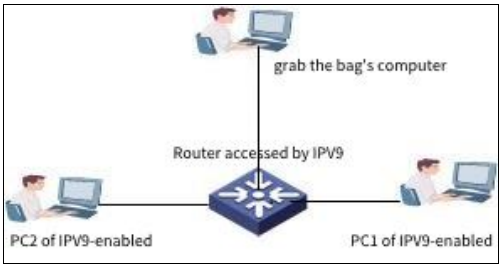


Figure 4. Packet capture test of IPV9 address encryption transmission

The process of the experiment:

1) Build a simulation test environment according to the above topology, connect two computers PC1 and PC2 that support IPV9 protocol under the IPV9 access router, and connect one computer to the IPV9 access router for wireshark packet capture analysis.

2) In order to compare the address encryption effect of IPV9, the address encryption function is enabled in the transmission direction from computer PC1 to computer PC2, but not enabled in the transmission direction from computer PC2 to computer PC1. In the test environment, the IPV9 address of PC1 is 32768[86[21[4]146, and that of PC2 is 32768[86[21[4]145.

3) Ping the IPV9 address of PC2 on PC1. Then, by grabbing the packets, we found the packets of ICMPv9 protocol. As shown in Fig 5.

Time	Source	Destination	Protocol	Length	Info
0.472183	ac:1f:6b:00:8f:10	AsustekC_73:45:ee	ICMPv9	150	ICMPv9 Request
0.472308	AsustekC_73:45:ee	ac:1f:6b:00:8f:10	ICMPv9	150	ICMPv9 Response

Figure 5. Wireshark grab package 1 of ICMPV 9

4) Analyze these two ICMPV9 bags in detail. Look at the response packet of ICMPV9 first, that is, the packet with IPV9 address encryption function not enabled in the transmission direction. As shown in Fig 6.

```
▶ Frame 5: 150 bytes on wire (1200 bits), 150 bytes captured (1200 bits)
▶ Ethernet II, Src: AsustekC_73:45:ee (2c:56:dc:73:45:ee), Dst: SuperMic_00:8f:10 (ac:1f:6b:00:8f:10)
▼ Internet Protolcol Version 9
  ▶ Flags: 0x90
    flow flag: 0x08324c
    len: 64
    next hdr: 0x9e
    hop limit: 0x40
  ▼ IPV9 Src addr: 0000800000000056000000150000000000000000000000000…
      saddr1: 32768
      saddr2: 86
      saddr3: 21
      saddr4: 0
      saddr5: 0
      saddr6: 0
      saddr7: 0
      saddr8: 0.0.0.145
  ▼ IPV9 Dst addr: 0000800000000056000000150000000000000000000000000…
      daddr1: 32768
      daddr2: 86
      daddr3: 21
      daddr4: 0
      daddr5: 0
      daddr6: 0
      daddr7: 0
      daddr8: 0.0.0.146
    Message Content: 810071914c2e00024e29d15cb1dc020008090a0b0c0d0e0f…
```

Figure 6. ICMPv9 response packet (address encryption direction is not used)

As can be seen from Figure 6, in the response packet of ICMPV9, the address of source IPV9 is 32768[86[21[4]145, that is, the address of computer PC2 IPV9; The destination IPV9 address is 32768[86[21[4]146, that is, the IPV9 address of computer PC1. That is to say, when the encryption function of IPV9 address is not enabled, the IPV9 addresses of both parties can be seen.

The ICMPV9 request packets are packets with IPV9 address encryption enabled in the transport direction. As shown in Fig 7.

Figure 7. ICMPV9 request packet (using address encryption direction)

In Fig 7, although there are source and destination IPV9 addresses for both sides of the transport, they are no longer the IPV9 addresses of PC1 and PC2. Note Address encryption is enabled for the transmission direction from PC PC1 to PC PC2.

5) Analyze the remaining ICMPV9 response packets in the packets captured by wireshark. As shown in Fig 8.

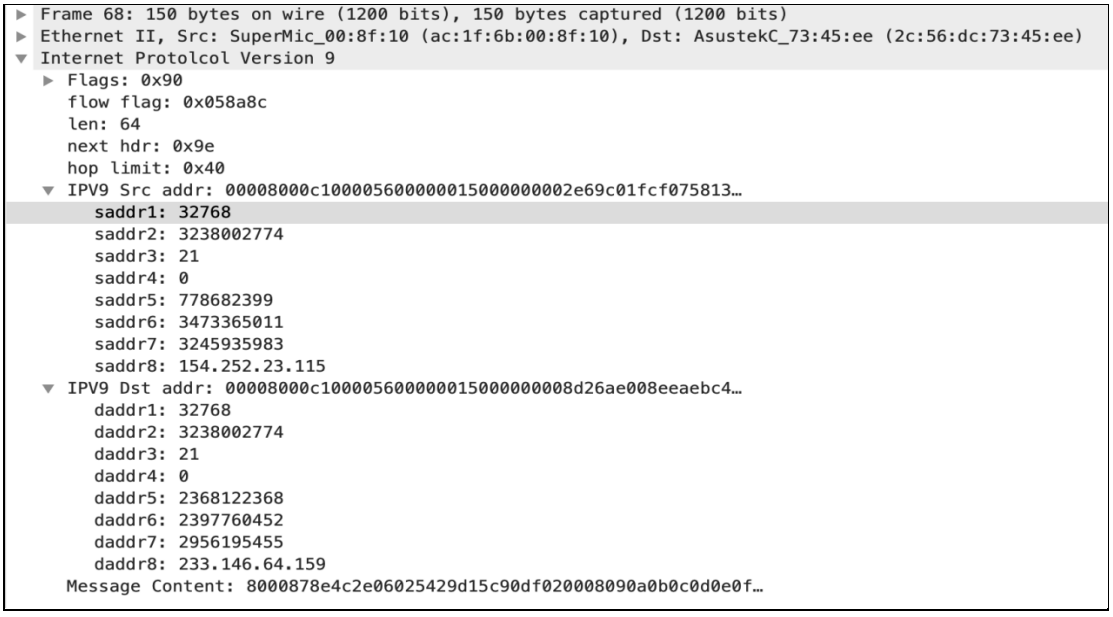


Figure 8. ICMPv9 Request Packet 2 (using address encryption direction)

In Fig 8, it can be seen that the IPV9 addresses of both parties are still encrypted, but the encrypted data is different from that in Fig 5.

6) By comparing the ICMP packet of IPv4, we can see that the address in the packet of IPv4 is not encrypted by default. As shown in Fig 9.

Time	Source	Destination	Protocol	Length	Info
4.025674	192.168.1.146	192.168.1.145	ICMP	74	Echo (ping) request
4.025805	192.168.1.145	192.168.1.146	ICMP	74	Echo (ping) reply

Figure 9. Request and corresponding packet of icmpv4

In Fig 9, we can see the IPv4 addresses of both parties. When the attacker intercepts this data packet, it can forge the source IP address or the destination IP address for address spoofing attack.

This experiment shows that the IPV9 address encryption function can encrypt the IPV9 addresses of both parties. And it is one encryption at a time, that is, the result of each encryption is different. With IPV9 address encryption, even if an attacker intercepts or listens to transmitted packets, it cannot identify the real IPV9 addresses of the sending parties or the source or destination of the packets, ensuring network security for both parties.

VII. SUMMARY AND OUTLOOK

In the design of the network security system of Big Data Platform in Tai ' an of Health, the IPV9 technology independently developed by China was used to build the network, and the IPV9 encryption technology was used to ensure the security of information transmission. This is the first application of IPV9 technology and IPV9 encryption technology in the field of health care, which demonstrates the independent innovation capability of national network information technology.

This paper studies the potential threats and harms of the network transmission process of Big Data Platform in Tai ' an of Health, including transmission information theft, source address attack, etc. By using IPV9 and its encryption technology, the network security architecture of the platform is designed and implemented in a targeted way, and a lot of basic work is carried out and corresponding simulation experiments are conducted. Practice proves that the network security architecture of the platform can really improve the security coefficient of the network transmission of Healthy Tai 'an Big Data Platform.

REFERENCES

[1] Decimal network working group. Decimal network address protocol [R]. http://www.em777.net/v9bt.html, 2010

[2] Decimal Network Working Group. Digital Domain Name Specification DDNS [R]. http://www.em777.net/1.html, 2010

[3] Decimal Network Working Group. Digital Domain Name Specification DDNS [R]. http://www.em777.net/1.html, 2010

[4] Li Guoling. IPV9 transition technologyresearch and test validation [D].Chongqing Universityof Post and Telecommunications, 2018

[5] Wang Zhongsheng, Xie Jianping. Technology and Application of Future Network [M], Tsinghua University Press, 2021.02

[6] Wang Zhongsheng, Xie Jianping. Technology and Application of Decimal Network [M], Publishing House of Electronics Industry [M], 2021.10

[7] Wang Wenfeng, Xie Jianping. Product and service digital identification format for information procession. SJ/T11603-2016, 2016. 06

[8] Xie Jianping etc. A method of assigning addresses to network computers using the full decimal algorithm [P]. CN: ZL00135182.6, 2004.2.6.

[9] Xie Jianping etc. Method of using whole digital code to assign address for computer [P]. US: 8082365, 2011.12.

[10] Information technology-Future Network- Problem statement and requirement-Part 2: Naming and addressing, ISO/IEC DTR 29181-2, 2014,

基于未来网络的健康泰安网络安全系统研究

赵洪汶 路 超 李浴宇 李广立 温国涛
中国广电山东网络有限公司泰安分公司

摘要: **目前因特网的核心是 IPv4，在当初设计时，几乎没有考虑安全防护方面的问题，因此基于 IPv4 的因特网存在诸多安全漏洞，极易受到攻击而造成信息泄露甚至网络瘫痪。相较 IPv4 而言，IPv6 在安全方面得到了一定的提升，但 IPv6 数据包默认情况下不加密和校验，网络层被攻击的问题仍然没有解决，因此基于 IPv6 的因特网仍面临数据被监听和篡改的风险，无法有效阻止被恶意攻击。中国十进制网络标准工作组开发了未来网络系统，建成了未来网络系统，该系统采用先验证后通信零信任安全机制。健康泰安大数据平台承载了泰安市所有医疗卫生机构的所有医疗健康数据的通讯、管理和维护，因此健康大数据平台的网络安全体系设计应该建立在有效、科学和全面的基础上。本文采用了我国自主知识产权的未来网络（IPV9）及其加密技术来设计健康泰安大数据平台的网络安全体系，并进行了相应的仿真模拟测试，达到了预期的效果。基于未来网络（IPV9）的健康泰安大数据平台网络安全体系设计，对解决健康大数据平台中网络安全问题能起到一定的借鉴作用和参考价值。**

关键词：健康大数据；网络安全；体系结构；未来网络（IPV9）

1. 网络安全现状

在安全方面，网络空间与现实空间融合发展使网络具有了超级基础设施的属性，网络安全问题已全面渗透到社会各个领域。网络主权已成为各国网络空间乃至国家安全与发展的重要保障。在发展方面，网络及其相关技术已经成为全球经济增长的主要驱动力，网络主权不仅关乎一个国家经济社会转型能否成功，还关系到该国未来的核心竞争力建设问题。

1.1 Internet 的安全性

目前，全球因特网的总枢纽、Web 总站、主干、建设与规划总部等机构实体全部在美国。美国通过对它们的控制而牢牢地控制了全球因特网，也控制了中国的互联网。从 2016 年 10 月 1 日起，ICANN 不是共管，而是永久性地划归美国管理。2017 年 12 月 14 日美国联邦通信委员会（FCC）正式废除网络中立法则，互联网具有了明显的政治色彩。2022 年 3 月俄乌之战，美国几乎所有的互联网企业对俄罗斯断网断服。

美国不仅控制了全球因特网和中国互联网，而且建立了美国军队赛博司令部和黑客队伍，研制了大量赛博（网络）武器，对敌对国家（尤其是中国）的网络安全造成巨大威胁。3 月 2 日，网络安全企业 360 公司发布《网络战序幕：美国国安局 NSA(APT-C-40)对全球发起长达十余年无差别攻击》的报告，外交部发言人汪文斌谴责报告曝光的恶意网络活动，再次强烈敦促美方作出解释，并立即停止此类活动。

从技术角度看，目前因特网的核心是 IPv4。IPv4 在当初设计时，几乎没有考虑安全防护方面的问题，因此基于 IPv4 的因特网存在诸多安全漏洞，极易受到攻击而造成信息泄露甚至网络瘫痪。相较 IPv4 而言，IPv6 在安全方面得到了一定的提升，但 IPv6 数据包默认情况下不加密和校验，同时其网络层被攻击的问题仍然没有解决，因此基于 IPv6 的因特网仍面临数据被监听和篡改的风险，同时无法有效阻止被恶意攻击。

工信部十进制网络标准工作组、上海十进制网络信息科技有限公司等多所高等院校和科研机构在自主可控网络方面进行了二十多年的研究，开发了整套十进制网络框架体系，完成了具有中国自主知识产权的自主可控网络 IPV9 系统。获得的专利得到包括中国、美国、英国、俄罗斯等多个国家的认可，同时得到中国工信部、国家标准委等部委的大力支持，取得了一系列创造性、国家战略性的新成果，建成了除美国之外的第二个互联网络体系。

1.2 网络安全面临的风险

网络攻击的技术和手段随着网络的不断发展层出不穷，提升网络安全水平刻不容缓。世界经济论坛《2018 年全球风险报告》中首次将网络攻击纳入全球风险前五名，成为 2018 年全球第三大风险因素。

《2017-2022 年网络安全设备市场监测与投资可行性研究报告》显示，中国互联网的快速发展和网络应用水平的不断提高，为推动经济发展和社会进步做出了巨大贡献。同时，也有许多新的情况,网络和业务发展的过程中遇到了很多新的问题和挑战，如云计算、大数据和其他新技术的应用中遇到的网络安全问题。

一是国内网络主要使用 IPv4 地址。但 IPv4 地址资源有限,特别是在 2019 年 11 月 25 日，欧洲网络协调中心（RIPE NCC）宣称 IPv4 地址都已分配。虽然国内很多地方都在推广 IPv4 地址到 IPv6 地址的过渡，但 IPv4 与 IPv6 是由国外主导的技术，存在一定的安全隐患。因此，需要保证地址资源充足的同时，也要推广我国自主可控的互联网协议，不能过度依赖 IPv4 协议与 IPv6 协议。

二是甄别网络攻击源头的技术和手段不足。特别是在使用 IPv4 地址进行网络通信时，只能针对应用层的传输数据进行加密，而不对网络层的 IP 地址进行加密。当攻击者截获数据时，可以分析出传输双方的 IPv4 地址。进而可以伪造其中一方的 IPv4 地址对另一方发起网络攻击，事后很难判断真正的网络攻击源头。

三是对国外信息技术和产品的依赖过高，内存、芯片等核心基础产品主要依靠进口，网络通信协议更是依赖于 IPv4/IPv6 的通信协议。然而，2018 年，美国封杀中兴；2019 年，美国封杀华为。越来越多的人意识到信息技术和产品能自主可控的重要性。迫切需要开发实用、易用的信息技术和产品；同时需要评估、支持和推广自主可控的产品和技术，尽快减少对国外信息技术和产品的依靠。

1.3 健康数据的重要性

健康大数据平台的网络安全是指平台内的硬件、软件及传输的信息受到保护。它包括平台的可靠、连续、稳定地运行,网络服务不中断,平台中的信息不因偶然的或恶意的行为而遭到破坏、更改或泄露。本文以健康泰安大数据平台为例，通过使用未来网络（IPV9）技术来实现健康大数据平台网络安全体系设计。

未来网络（IPV9）是由我国自主研发的通信协议，是对未来网络的一种积极探索。未来网络（IPV9）能同时兼容当下的 IPv4 与 IPv6 协议，同时更符合未来长期的网络发展要求，是既安全又可靠的未来网络的桥梁。

健康泰安大数据平台建设基于国家自主知识产权的未来网络（IPV9）技术，建设覆盖市、区县、乡镇、村、家庭五级医疗卫生安全专用网络。以医疗卫生信息系统为基础，以健康泰安大数据平台为核心，实现平台内各医疗卫生机构的统一管理、业务之间的统一调度。

2. 未来网络技术及安全

2.1 未来网络安全技术

未来网络协议从基本算法入手，使用十进制算法来分配地址。IPv6 的地址是 128 位，而未来网络在此基础上扩充到了 256 位。未来网络与 IPv4、IPv6 兼容，可以在不改变现有的 IPv4 或 IPv6 的网络环境来部署 IPV9。未来网络的协议族主要有以下几部分组成：包括未来网络报头协议;未来网络地址的寻址模型、未来网络的地址文本表示、未来网络单播地址、任意点播地址、组播地址以及未来网络节点需要的地址；

未来网络数报文格式协议,主要是报文首部扩展及认证、加密；控制报文协议；未来网络邻节点探测协议；未来网络的安全体系结构 IPSEC 协议；未来网络移动通信协议；未来网络数字域名 DNS 扩展协议；未来网络即插即用协议；未来网络过渡期协议。IPv6 与未来网络 IPV9 都是基于 TCP/IP 框架，但又有所不同。未来网络中融合了很多电信技术，例如电话号码概念、地理位置概念等。

在未来网络中，一台路由器 A 按照配置在指定接口广播公钥，任何与该接口相连的设备(如 B)向该路由器发送数据时，可选择使用该公钥进行地址加密（也可选不加密）。路由器根据地址中的是否加密标志自行判断是否进行解密，注意该过程仅影响单方向的数据流向(B->A)。双向加密就是上面概念中，增加把 B 和 A 互换的过程，即 B 广播公钥，A 使用接收到的 B 的公钥进行地址加密并将数据传送到 B(A->B)。

以下为地址加密系统中对地址编码产生影响的部分：

（1）地址中总长 32 个字节中的第 5 个字节必须为 0，该字节会用于标记地址是否加密以及相关标志。（2）地址加密系统仅对总长 32 个字节中的后 16 个字节进行加密。

2.2 健康数据传输安全研究

健康泰安大数据平台的传输线路主要是有线传输安全，主要存在以下安全隐患。

攻击者可以在有线网络的传输线缆上安装窃取装置，窃取通过有线网络传输的数据，并通过技术手段进行分析，可能会得到用户输入的账户密码等重要信息，甚至可能会更改或删除传输数据，影响数据的可靠性和真实性。攻击者截取到有线传输的数据后，可以根据传输信息中的 IP 地址，针对该 IP 地址发起攻击。这使得健康泰安大数据平台中的服务器地址或用户地址直接暴露在攻击者面前。

这种通信链路的风险严重威胁健康泰安大数据平台的安全性。为此，使用未来网络地址及未来网络加密技术，在通信链路传输的过程中同时对传输信息和地址的进行加密，来保障信息传输过程中的安全性。

3. 系统安全技术设计

未来网络 IPV9 能同时兼容现有的 IPv4 与 IPv6 协议，同时更符合未来长期的网络发展要求，是既安全又可靠的过渡到未来网络的桥梁。健康泰安大数据平台建设基于国家自主知识产权的未来网络技术，建设覆盖市、区县、乡镇、村、家庭五级医疗卫生安全专用网络。以医疗卫生信息系统为基础，以健康泰安大数据平台为核心，实现平台内各医疗卫生机构的统一管理、业务之间的统一调度。

利用未来网络技术组建逻辑隔离专网，与以前单纯 IPv4 专网区别如下：未来网络专网中的通信地址使用全新的十进制地址，同时可以兼容 IPv4 与 IPv6 的地址。使用 IPv4 的地址进行数据传输时，只能针对应用层数据加密，不能对地址进行加密，攻击者可以根据数据中的地址发起攻击。而未来网络协议不仅可以对传输数据的应用层加密，还可以对网络层的 IP 地址加密，而且每次地址加密的结果都不一样，起到了更好的安全效果。

医疗卫生机构专线用户接入端使用专用设备未来网络路由器，保证与用户办公电脑其他网络连接系统的逻辑隔离。在医院接入端的用户侧使用 1000M-未来网络（IPV9）路由器。1000M-未来网络路由器作为专用终端设备与医院客户端相连，由于该设备接入网络使用的是十进制地址，可以起到防火墙的作用，确保与医院电脑上其他网络连接办公系统相隔离。另外在数据传输层面，数据是未来网络报文形式，组建的是未来网络专网，通过分配十进制地址实现了网络层的加密，与 IPv4 网络相比，具有更高的安全性，更加可靠。

健康泰安大数据平台的组网方式、网络扑拓结构与原有网络基本相同，只是在组网方式、网络扑拓中增加了 IPv4/ IPV9 路由器和协议转换路由器，网络拓扑如图 1 所示。

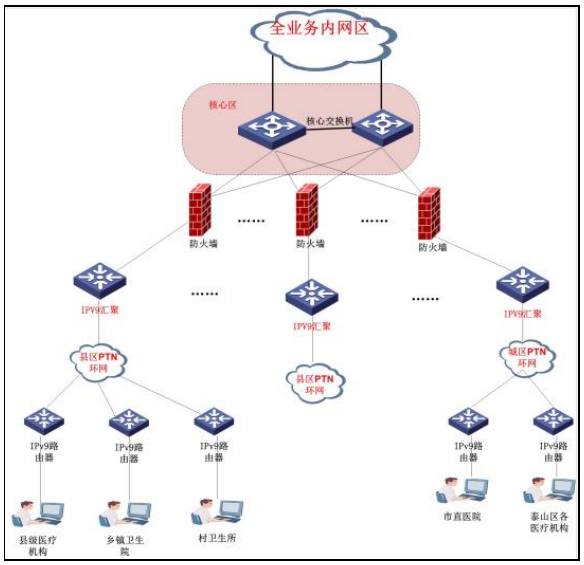


图 1 健康泰安大数据平台拓扑结构图

核心区为两台数据中心级交换机，并且都采用了 VRRP 技术进行互为备份，当其中一台交换机发生故障时，这台交换机上的业务会自动切换到另一台交换机上，不会影响正常业务的使用，增强了平台的可用性。接入区为所有卫生医疗机构接入区域，现使用 6 台万兆防火墙进行隔离及汇聚，每个县区各接入一台。各县区网络架构相同，核心为未来网络汇聚路由器，万兆上联数据中心防火墙，链路使用 PTN 环网，确保带宽及安全。

未来网络汇聚路由器针对接入用户传输的数据进行地址加密和解密，使得信息的传输具有更高安全性。通过交换机接入县区 PTN 专线环网。所有医疗机构均部署未来网络接入路由器，根据各机构的不同的网络环境及安全需求，采用地址转换、逻辑隔离、防火墙防护等策略。

未来网络接入路由器均采用千兆级别设备。用户侧仍使用 IPv4 的地址，在 IPv4 的报文经过未来网络的接入路由器时，自动封装上未来网络的报头，报文经过核心区的未来网络汇聚路由器时，再将未来网络的报头解封出来，还原成 IPv4 的报文。这种方式可以使所有 IPv4 应用都可在未来网络上运行。如图 2 所示。

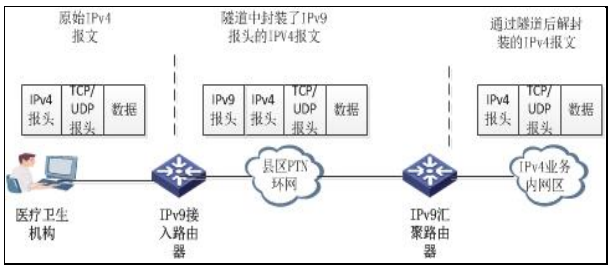


图 2 IPv4 报文封装未来网络（IPV9）与解封装

未来网络在实施中，可以在不影响和改变现有终端使用 IPv4 的情况下进行。健康泰安大数据平台以此为依据，逐步搭建了未来网络骨干网络。

为防止接入区医疗卫生机构的数据到核心区传输的过程中被篡改或破坏，健康泰安大数据平台在未来网络骨干网的基础上启用了未来网络加密技术，对传输进行加密。不同于目前的单一对应用加密的手段，未来网络创新的设计了地址加密，将安全保护延伸至网络层，极大的提高了信息传输安全性。

健康泰安大数据平台中，接入区部署的未来网络汇聚路由器和未来网络接入路由器上同时启用 IPv4 与未来网络的转换协议，并启用未来网络的加密技术。配置如下（以两台路由器的 eth0 网络接口为两个路由器的通信口）。

set interface state eth0 up //使能 eth0 网络接口

reverse enable eth0 4to9 //设置 eth0 为支持 IPv4 与未来网络的转换协议的网络接口

reverse keyset 密钥//设置加密密钥

健康泰安大数据平台中加密原理如下：未来网络汇聚路由器 A 按照配置在指定接口广播公钥，任何与该接口相连的设备(如接入路由器 B)向该路由器发送数据时，可选择使用该公钥进行数据或地址加密，然后将加密后的未来网络数据发送到汇聚路由器 A 后再解密，保障了数据传输的保密性。同理，汇聚路由器 A 到接入路由器 B 进行未来网络加密传输也是这样。如图 3 所示。

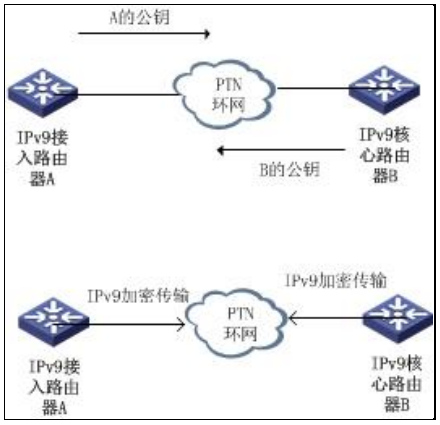


图 3 未来网络（IPV9）密钥传输与加密传输

未来网络通信协议报文结构设计合理，报文项目功能明确，未来网络协议在地址空间、服务质量、安全性等方面的设计优于 IPv4 协议。未来网络协议数据报文的地址表示方式与报文报头结构同 IPv4 或 IPv6 协议不同,所以未来网络协议的数据报文报头将不会被 IPv4 或 IPv6 系统识别，不会直接在这些系统中进行传播。目前所有黑客的攻击及所有网上窃听软件都是基于 IPv4 开发的。未来网络路由器及网卡对这些窃听及黑客的攻击数据包将不予放行，对黑客攻击及网上情报的肆意窃取筑起长城。未来网络使得我国实现了互联网底层协议的安全可控。地址和数据双重加密机制等先进设计理念极大的提高了网络信息的安全性。

4. 未来网络加密仿真测试

在本文的设计方案中，从用户测未来网络接入路由器到未来网络汇聚路由器之间的传输采用了未来网络加密处理，来保障传送两端的传输安全。其中包括未来网络传输的地址加密与数据加密。本实验针对未来网络传输中的地址加密进行仿真测试。测试拓扑图如下 4 所示。

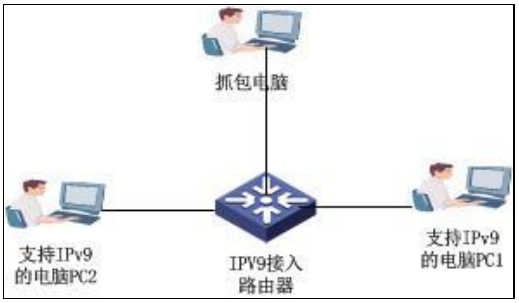


图 4 未来网络（IPV9）地址加密传输的抓包测试

测试过程如下。

（1）按上述拓扑搭建好仿真测试环境，未来网络（IPV9）接入路由器下接入两台支持未来网络协议的电脑 PC1 和 PC2，同时在未来网络接入路由器上连接一台电脑进行 wireshark 进行抓包分析。

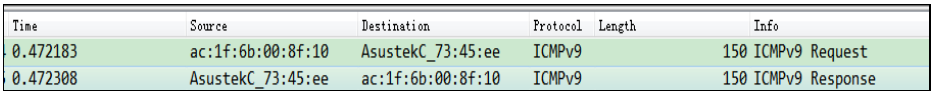
（2）为了对比未来网络的地址加密效果，在电脑 PC1 到电脑 PC2 的传输方向启用地址加密功能，在电脑 PC2 到电脑 PC1 的传输方向上不启用地址加密功能。其中测试环境中电脑 PC1 的未来网络地址为 32768[86[21[4]146，电脑 PC2 的未来网络地址为 32768[86[21[4]145。

（3）在 PC1 上 ping PC2 的未来网络地址。然后通过抓包,发现了 ICMPv9 协议的数据包。如图 5 所示。

（4）具体分析抓到的这两个 ICMPv9 的包。先看 ICMPv9 的响应包，即传输方向上没有启用地址加密功能的数据包。具体如图 6 所示。

从图 6 可以看出，在 ICMPv9 的响应包中，未来网络源地址为 32768[86[21[4]145，即电脑 PC2 的地址；目的地址为 32768[86[21[4]146，即电脑 PC1 的未来网络地址。也就是说，在未启用未来网络（IPV9）地址加密功能时，能看到通信双方的未来网络（IPV9）地址。

再看 ICMPv9 的请求包，即传输方向上启用未来网络（IPV9）地址加密功能的数据包。具体如图 7 所示。



Time	Source	Destination	Protocol	Length	Info
0.472183	ac:1f:6b:00:8f:10	AsustekC_73:45:ee	ICMPv9	150	ICMPv9 Request
0.472308	AsustekC_73:45:ee	ac:1f:6b:00:8f:10	ICMPv9	150	ICMPv9 Response

图 5 ICMPv9 的 Wireshark 抓包

```
▶ Frame 5: 150 bytes on wire (1200 bits), 150 bytes captured (1200 bits)
▶ Ethernet II, Src: AsustekC_73:45:ee (2c:56:dc:73:45:ee), Dst: SuperMic_00:8f:10 (ac:1f:6b:00:8f:10)
▼ Internet Protolcol Version 9
  ▶ Flags: 0x90
    flow flag: 0x08324c
    len: 64
    next hdr: 0x9e
    hop limit: 0x40
  ▼ IPV9 Src addr: 0000800000000056000000150000000000000000000000000…
      saddr1: 32768
      saddr2: 86
      saddr3: 21
      saddr4: 0
      saddr5: 0
      saddr6: 0
      saddr7: 0
      saddr8: 0.0.0.145
  ▼ IPV9 Dst addr: 0000800000000056000000150000000000000000000000000…
      daddr1: 32768
      daddr2: 86
      daddr3: 21
      daddr4: 0
      daddr5: 0
      daddr6: 0
      daddr7: 0
      daddr8: 0.0.0.146
    Message Content: 810071914c2e00024e29d15cb1dc020008090a0b0c0d0e0f…
```

图 6 ICMPv9 响应包（未使用地址加密方向）

```
▶ Frame 4: 150 bytes on wire (1200 bits), 150 bytes captured (1200 bits)
▶ Ethernet II, Src: SuperMic_00:8f:10 (ac:1f:6b:00:8f:10), Dst: AsustekC_73:45:ee (2c:56:dc:73:45:ee)
▼ Internet Protolcol Version 9
  ▶ Flags: 0x90
    flow flag: 0x058a8c
    len: 64
    next hdr: 0x9e
    hop limit: 0x40
  ▼ IPV9 Src addr: 00008000c10000560000001500000000031182a113cafd23d…
      saddr1: 32768
      saddr2: 3238002774
      saddr3: 21
      saddr4: 0
      saddr5: 823667217
      saddr6: 1018155581
      saddr7: 395588771
      saddr8: 94.247.116.174
  ▼ IPV9 Dst addr: 00008000c10000560000001500000000474936d7a1a70c51…
      daddr1: 32768
      daddr2: 3238002774
      daddr3: 21
      daddr4: 0
      daddr5: 1195980503
      daddr6: 2712079441
      daddr7: 2288582866
      daddr8: 236.216.220.25
    Message Content: 800072914c2e00024e29d15cb1dc020008090a0b0c0d0e0f…
```

图 7 ICMPv9 请求包（使用地址加密方向）

在上图中，虽然有传输双方的未来网络（IPV9）源地址和目的地址，但已经不是 PC1 的地址和 PC2 的地址。说明电脑 PC1 到电脑 PC2 的传输方向启用地址加密功能已经生效。

（5）再分析通过 wireshark 抓的包中其余的 ICMPv9 响应包。如图 8 所示。

在图 8 中，可以看出通信双方的未来网络（IPV9）地址仍然进行了加密，但加密后的数据与原图中的数据是不一样的。

（6）再对比 IPv4 的 ICMP 的包，可以看到 IPv4 的数据包中的地址默认是没有加密的。如图 9 所示。

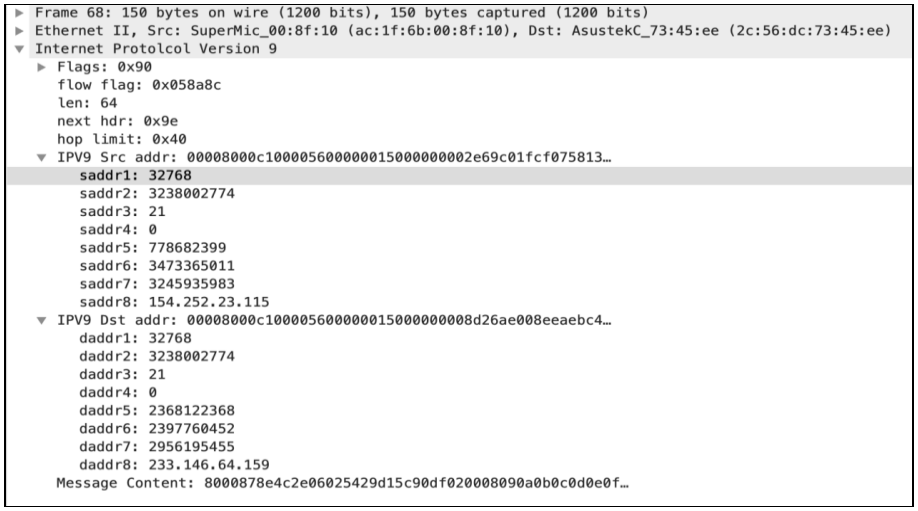
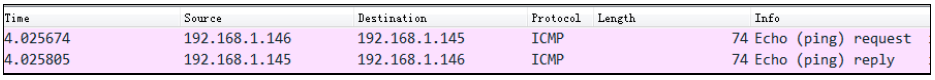


图 8 ICMPv9 请求包 2（使用地址加密方向）



Time	Source	Destination	Protocol	Length	Info
4.025674	192.168.1.146	192.168.1.145	ICMP	74	Echo (ping) request
4.025805	192.168.1.145	192.168.1.146	ICMP	74	Echo (ping) reply

图 9 ICMPv4 的请求与相应包

在上图中，可以看到通信双方的 IPv4 地址，当攻击者截获这个数据包后，可以伪造源 IP 地址或目的 IP 地址进行地址欺骗攻击。

通过本实验，可以看出未来网络的地址加密功能能对传输双方的地址进行加密，并且是一次一密，即每次加密的结果都不一样。通过未来网络（IPV9）地址加密，攻击者即使截获或侦听了传输的数据包，也无法判断发送双方的真实未来网络（IPV9）地址，无法判断信息来源或去往，保障了通信双方的网络安全性。

5. 结论

健康泰安大数据平台网络安全体系设计中，使用了我国自主研发的未来网络（IPV9）技术进行了网络建设，并使用了未来网络（IPV9）加密技术来保障信息传输安全。这是未来网络（IPV9）技术及未来网络（IPV9）加密技术首次在健康医疗领域的应用，彰显了国家网络信息技术的自主创新能力。本文研究健康泰安大数据平台的网络传输过程的潜在威胁和危害,包括传输信息窃取、源地址攻击等等，通过使用未来网络地址及其加密技术，有针对性地进行了平台网络安全体系架构设计与实现，开展了大量基础工作并进行了相应的仿真实验测试，实践证明该平台网络安全体系确实能提高健康泰安大数据平台的网络传输的安全系数，具有一定的应用价值。

参考文献

[1] 十进制网络工作组.十进制网络地址协议[R].http://www.em777.net/v9bt.html，2010

[2] 十进制网络工作组.十进制网络地址协议[R].http://www.em777.net/v9add.html，2010

[3] 十进制网络工作组.数字域名规范 DDNS[R].http://www.em777.net/1.html，2010

[4] 李国领. 十进制网络过渡技术研究及测试验证[D].重庆邮电大学,2018

[5] 王中生，谢建平. 未来网络技术及应用[M]. 北京：清华大学出版社，2021.

[6] 王中生，谢建平. 十进制网络技术及应用[M]. 北京：电子工业出版社，2021

[7] 王文峰，谢建平用于信息处理产品和服务数字识别格式. SJ/T 11603-2016, 2016. 06

[8] 谢建平. 联网计算机全十进制算法分配计算机地址的方法[P]. CN: ZL00135182.6, 2004.2.6.

[9] 谢建平等. 采用全数字编码为计算机分配地址的方法[P]. US: 8082365, 2011.12.

[10]信息技术-未来网络- 问题陈述与请求-Part 2: 命名与寻址, ISO/IEC DTR 29181-2, 2014,

Design and Research of Indoor Lighting Control System Based on the STM32

Chao Fan
Information and Telecommunication Branch, State Grid Zhejiang Electric Power Co., Ltd., Hangzhou, China
E-mail: ttla02@126.com

Lei Tian
School of Electronic Engineering
Xi'an University of Posts and Telecommunications
Xi'an, China
E-mail: tianlei@xupt.edu.cn

Qingmin Zhang
School of Communication Engineering
Xidian University
Xi'an, China
E-mail: qmzhang@stu.xidian.edu.cn

Guo Xukai
School of Humanities and foreign languages
Xi'an University of Posts and Telecommunications
Xi'an, China
E-mail: wxh2324@126.com

Kaifa Kang
Dept of Electronic Engineering
Xi'an University of Posts and Telecommunications
Xi'an, China
E-mail: 364945696@qq.com

***Abstract*—After the completion of the process of industrialization in the world, a large number of electrified equipment appeared, People's Daily life is also more and more dependent on electricity security, electricity consumption has risen sharply. At present, various lights driven by electricity play a crucial role in our daily life, which is also the main cause of electricity consumption. We often notice that even when the room is empty, or during the day, the lights are still on, which is a waste of electricity. The waste of power resources caused by traditional lighting equipment cannot be ignored. Therefore, intelligent home furnishing equipment is the favored object of people. So here for a kind of indoor lighting control system is designed, in order to meet the lighting needs on the basis of more humane, more importantly, can save power resources. With the rapid application of automation in daily life and the continuous development of computer technology, lighting control has become increasingly intelligent and automated. The control function of the system is realized by the STM32F103 chip, the illumination is automatically detected by the BH1750 module, and the time parameters are recorded and fed back by the DS3231 clock module. The infrared reflection sensor is used to detect whether there are people in the room. When there is no one, the light will be automatically**

DOI: 10.2478/ijanmc-2022-0004

turned off to save energy. If there are people, when the ambient light intensity is lower than the set value, the light will be automatically turned on to ensure the comfort of the ambient brightness. And can use OLED module real-time display indoor light intensity, number of people and working time. This indoor lighting control system can not only automatically turn on the light according to the environment, but also detect whether there is a human body in the current environment and turn on or off the light by itself, which has more practical significance.

Keywords-Lighting Control; STM32; Lighting Adjustment; Sensor; Intelligent

I. INTRODUCTION

Light consumption in daily life causes an unprecedented waste of electricity resources. Countries carry out research on this issue. Now, many products in the domestic market mainly use different step-down technologies to achieve energy saving, such as self-disaster transformers and magnetic saturation reactors, but these products have more or less problems. After the improvement, the equipment aging fast, eliminated products to the environment and other problems such as greater pollution.

This paper compares the control principle and main function structure of some indoor lighting control system, describes in detail the hardware structure of the indoor lighting control system and the design and implementation of the control system, and carries on various debugging [1].

II. THE OVERALL DESIGN

The traditional lighting control in China usually adopts manual control or timing control [2]. This method is time-consuming, inefficient, and will lead to a large degree of resource waste. The combination of camera monitoring and remote control greatly increases the cost of construction and will inevitably cause extravagance and waste of power resources [3].

Common domestic lighting control usually adopts some manual or timing control methods, such as the following common methods:

1: The lighting system is controlled manually, and the lighting device in a certain area is turned on or off by manual patrol, which is time-consuming and labor-intensive;

2: Use sound to control the lighting system, and use a sound sensor to judge whether the lighting device should be turned on. This method is inefficient, and all disturbances may cause the lighting device to work and cause a greater degree of resources waste;

3: Use optics to control the lighting system, mainly to control the lighting device according to the brightness in a certain area, this method is similar to method 2.

Although the traditional lighting device control methods have their own characteristics, they also have some shortcomings. Even some scenes with more lighting devices will be controlled by a combination of camera monitoring and remote control, which can indeed achieve better results, but this method greatly increases the cost of construction, and because this method is still manual Control, so it will inevitably still cause a lot of extravagance and waste of power resources [4].

In this paper, STM32F103 is used as the main control chip of the indoor light control system [4], and the design and implementation of the indoor light control system is carried out by combining DS3231, OLED, infrared detection and BH1750 light intensity detection modules. The BH1750 light intensity detection module detects the brightness of the current environment. DS3231 clock module is to realize the real-time display of

time and working time. TCRT5000 infrared reflection detection module realizes the function of human body induction [5]. By conducting infrared reflection induction in a specific area, it automatically controls the opening or closing of indoor light control system [6]. OLED reminds users of the current state of indoor lighting control systems. The system framework is shown in Figure.1 below.

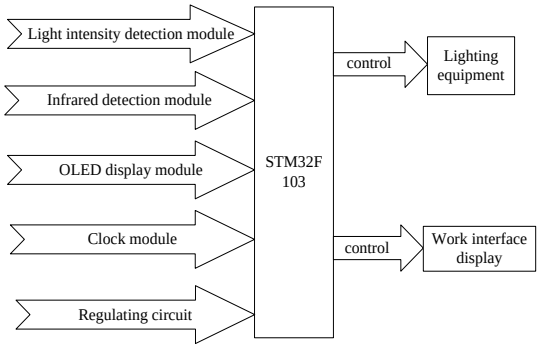


Figure 1. System block diagram

A. System control chip

In the indoor lighting control system, the position of the system control chip can be said to be very important [7], similar to the human brain, the instructions for the actions of each part of the body come from this. The control chip of the indoor lighting control system not only needs to control its own modules, but also needs to control some externally extended modules, so that the entire indoor lighting control system can operate [8].

Through the characteristics of 51 series, 32 series, 15 series and other types of single-chip microcomputers, combined with the requirements of the indoor lighting control system, the selected control chip has the characteristics of high speed, low power consumption and anti-interference. In the design of this article, the selected control chip It is a STM32F103 single-chip microcomputer chip. For this single-chip microcomputer, all the above-mentioned features are possessed, and it has more advantages. STM32F103 is widely used in many smart home control systems such as rice cookers, washing machines, refrigerators, and air conditioners, especially in the application of simple or single-function small micro-control systems is very common.

B. Environmental testing unit

The processing of sensors and brightness in the indoor lighting control system is also not negligible [9-10]. In the design of the system in this paper, TCRT5000 is used to detect the human body in the current environment; BH1750 detects the brightness in the current environment in real time.

For the TCRT5000 photoelectric sensor module, it is an infrared reflective photoelectric switch realized by the TCRT5000 infrared photoelectric sensor. For the sensor, it is composed of two parts, the first is an infrared photodiode, and the second is a high-sensitivity phototransistor [11]. The output control signal can be reshaped by Schmitt circuit, which is reliable, safe and reliable. The stability is very good. When in use, the built-in infrared emitting diode in the sensor will continuously emit infrared rays.

When the intensity of the emitted infrared rays is low or there is no emission at all, then the photosensitive transistor will always be in an off state. At this time, the output terminal corresponding to the module will be in a low level state, and at the same time, for the indicator diode, it will always be in a extinguished state; if the detected object is in the detection range, then the infrared will be Is emitted back and has a greater intensity at the same time, then the phototransistor will be in a saturated state [12]. At this time, the output terminal corresponding to the module is in a high-level state, and the indicator diode will be lit.

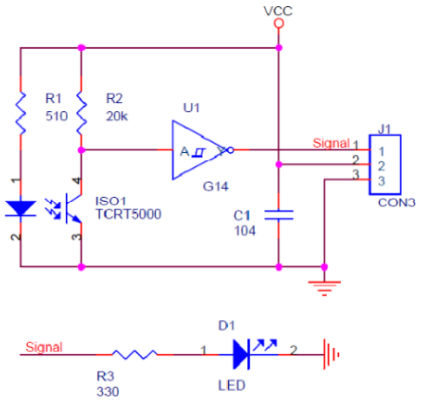


Figure 2. Infrared reflection module

According to the light intensity sensor model BH1750FVI, it is a digital light intensity sensor integrated chip in principle. The inside of the sensor is composed of four parts: ADC acquisition, photodiode, crystal oscillator and operational amplifier. For PD diodes, a photovoltaic effect is used to convert the input optical signal into an electrical signal, which is then amplified by an operational amplifier circuit, and then the voltage is collected by ADC, and then logic is used The circuit converts it into a 16-bit binary data, and then stores it in the internal register (it should be noted here that if the intensity of the light entering the light window is greater, the corresponding photocurrent will be greater.

At the same time, the voltage will be higher, so the voltage value can be used to judge the size of the light, but it should be noted that even if the voltage and the light intensity are corresponding, this relationship is not a proportional relationship, so the chip performs internally The corresponding linear processing, which also explains why not choose photodiodes, but choose integrated IC). At the same time, the sensor carries out the data line and the clock line. The microcontroller can use the I2C protocol to communicate with the sensor, can use the BH1750 working mode, and can also extract the illuminance data stored in the BH1750 register.

C. Clock module and display module

For the clock module model DS3231, it is an IPC real-time clock (RTC) in principle, with higher accuracy and lower cost, and an integrated temperature-compensated crystal oscillator (TCXO) is set inside. And crystals. For this device, there is a battery input terminal, after disconnecting the main power supply, it can still ensure that the timing is accurate enough. At the same time, through the use of an integrated crystal oscillator, the long-term accuracy of the device can be improved, and the number of components required in the production line can be reduced.

For, it can realize the provision of industrial grade and commercial grade temperature range, the number of pins used is 16 in total, and a 300mil so package is also used. RTC can maintain a variety of time parameters. If the number of days in the month does not exceed 31 days, the end of the month will be automatically adjusted and the leap year compensation function can also be realized.

The clock has two working formats, the first is 24 hours, and the second is 12 hours with AM\PM indication. Able to realize one programmable square wave and two programmable calendar alarm clock output. Both data and address can use the PC two-way bus to achieve a serial transmission, a voltage reference and a comparator that have completed temperature compensation to monitor the Vcc status, detect power failures, and provide reset output at the same time, In some necessary cases [13], it can realize automatic switching to the standby power supply. In addition, for the RST monitoring pin, it can also be used as a manual button to generate an external reset signal.

In the design of this article, the OLED display screen is chosen to design the display module [14]. This display screen has many names, and

sometimes it is also called an organic light-emitting semiconductor. This display screen was researched by a scholar in 1979. The characteristic of OLED display technology is that it can realize self-luminescence, has a wide viewing angle, fast reflection speed and low power consumption. The disadvantage is that the price is more expensive.

III. HARDWARE CIRCUIT DESIGN

The hardware circuit of the indoor lighting control system is mainly composed of STM32F103C8T6 MCU, DS3231 clock module, infrared detection module circuit, OLED display module and lamp tube circuit [15]. Combined with the requirements of indoor lighting control system control chip with high speed, low power consumption and anti-interference characteristics, this paper chooses the model of STM32F103 MCU chip.

The OLED display module and BH1750 light intensity sensor can be used together to obtain the brightness value of the current environment, so as to control the chip for the next processing judgment and issue corresponding instructions. In the clock module, users can adjust the indoor lighting control system with their own needs. Infrared detection module is set in the entrance and exit infrared reflection sensor, detection personnel in and out, so as to determine the number of people in the room. The lamp tube circuit is the main display part of the indoor lighting control system. The specific hardware circuit design is as follows.

A. Infrared Detection Module Circuit

Infrared detection module is composed of STM32 and CTRT5000 infrared detection module. After the indoor lighting control system is powered on, the CTRT5000 starts to work when it works normally. The infrared detection of the current environment is continuously carried out and the results are sent to the control chip. The control chip changes the state of the indoor lighting control system according to the returned results. The connection circuit between the infrared detection module and STM32 is shown in Figure 3.

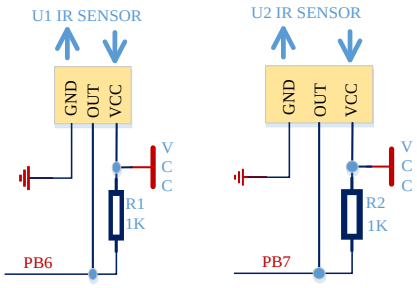


Figure 3. Connection circuit between IR sensor and STM32

B. Light Intensity Sensor Circuit

The light intensity sensor circuit of the indoor lighting control system is used to detect the light intensity. The BH1750 sensor is built with four components: an ADC acquisition, a photosensitive diode, a crystal oscillator, and an operational amplifier.

For PD diode, the use of light born v effect according to the input optical signals into electrical signals, then through operation amplifier circuit amplification processing, through the ADC to acquisition of voltage, after using logic circuit transformation, makes it a 16-bit binary data, and then stored in internal registers. This sensor carries on the data line and the clock line extraction, the micro-controller can use I2C protocol to realize the communication between the sensor, can use the BH1750 working mode, but also can extract the BH1750 register stored in the illuminance data. The BH1750 communicates in five steps: Step 1: Sends a power-on command. Step 2: Send the measurement command. Step 3: Wait for the measurement. Step 4: Read the data. Step 5: Calculate the result. Finally, the light

intensity data stored in the module is extracted by the master chip and displayed by the OLED.

C. OLED Display Circuit

The corresponding power supply input terminal of this module is VCC_IN, which supplies U1 and U4 circuits after depressurization by U2. A voltage of 7~7.5V is required to drive the OLED screen, which is provided by boosting the voltage through SSD1306's built-in charge pump. For SSD1306, it can support parallel, SPI and I2C serial protocols. This module is grounded by using BS0->BS2 pins, which have been set as SPI protocol and shared with U4 Chinese character library chip for SPI interface. During operation, the chip's corresponding chip pins are used to distinguish. During read and write operations, the CS pins corresponding to the chip will be set to a low-level state, which will be used as the corresponding display part of this project. OLED is connected to STM32 and BH1750 and DS3231. In addition to displaying normal working hours, it also displays brightness, number of people, current time, etc.

D. Rectifier Voltage Regulator Circuit

In the circuit of the lighting control system, AMS1117 series regulator has two versions, one is adjustable version, one is a variety of fixed voltage version, the role is to carry out 1A current output, while ensuring that the working pressure difference will not exceed 1V. The minimum differential pressure of the AMS1117 device is guaranteed to be less than or equal to 1.3V at the maximum output current, and to decrease as the load current decreases. SS34 is a patch Schottky diode used for instantaneous circuit rectification. The function of the rectifier circuit is to convert a kind of AC voltage in positive and negative changing state through a one-way conductivity of the diode, so that it becomes a one-way pulse voltage. Through the action of the AC power supply, the rectifier diode can achieve a periodic on or off state, by which the load can receive a pulse of direct current. Rectifier voltage regulator circuit is shown in Figure 4.

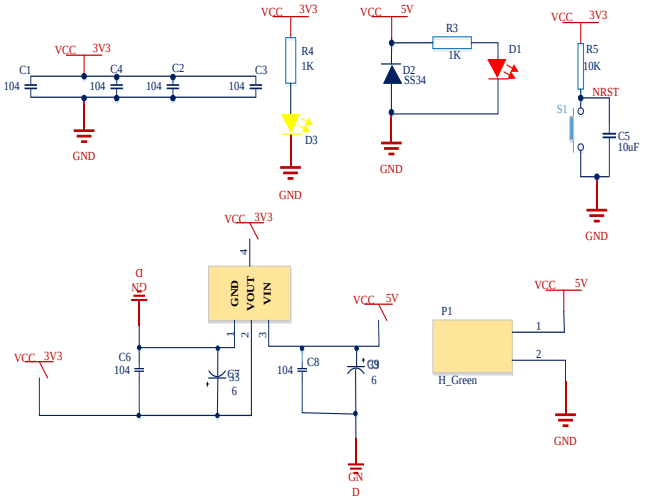


Figure 4. Rectifier voltage regulator circuit

IV. SOFTWARE PROGRAM DESIGN

A. System Software Design

In the research and development of any electronic product system, not only the design and implementation of hardware need to be emphasized, but the design and implementation of software programs are also equally important. At present, the development tools of various electronic systems can be described as a dazzling array of functions, but many senior developers still prefer to use KEIL for development, mainly because the pages are friendly and easy for developers to use since KEIL came out. Therefore, the software design part of the indoor lighting control system is also selected to be developed through KEIL and C language. Moreover, for the increasingly mature application and development of stm32 single-chip microcomputers, many basic codes and routines are also increasing, which makes it more convenient and simpler to develop

STM32 single-chip microcomputers based on KEIL in C language.

The research and development of the software part of the indoor lighting control system mainly adopts a process from a single sub-module to an overall realization, which facilitates the timely detection of errors and reduces the time for follow-up inspections:

(1) The design of the environmental brightness detection module: The BH1750 is mainly used to detect the current environmental brightness in real time and transmit the data to the program of the main control chip.

(2) OLED display module design: Mainly connected with the main control chip, it can realize the real-time display of the feedback data on the screen, so that the user can understand the current information program.

(3) Clock module design: DS3231 performs accurate time calculation, and feeds back the program of real-time time and current working time.

(4) Human body detection module design: It is mainly a program that detects the human body in the current environment and makes corresponding operation feedback.

Each subroutine relationship is shown as Figure 5.

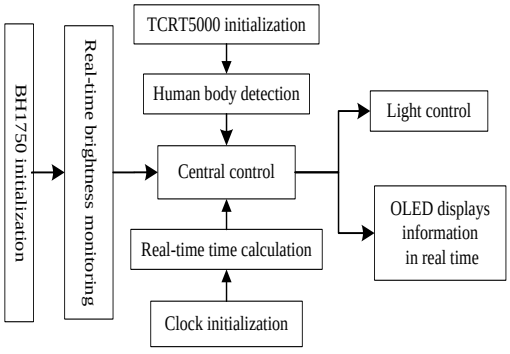


Figure 5. Relationship among subroutines

B. Ambient Brightness Detection

Intelligent lighting detection of the current environment brightness subroutine mainly uses BH1750 to detect the current environment brightness. The BH1750 light detection sensor needs to be initialized and the environment brightness is detected according to the instructions of the control chip. The environmental brightness detection subroutine flow is shown in Figure 6.

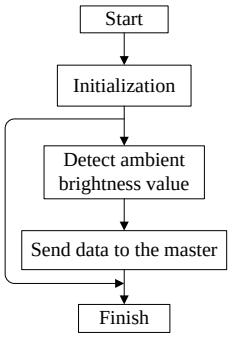


Figure 6. Brightness detection subroutine

When detecting the brightness of the current environment, this paper adopts the method of twice detecting, that is, continuously detecting the brightness of the current environment twice. When the two detection results are consistent, the collected ambient brightness value will be sent to the control chip; otherwise, "0" will be sent.

C. OLED Display Programming

The display module program of this design needs to be set for the IO connected by the MCU and the OLED module, and the IO port is set as output. Initialize the OLED module. Use functions to display characters and numbers on the OLED module.

D. Clock Subroutine Design

The design of the clock module program through STM32 analog I2C signal channel, then configure IO pin, and initialization function, in the main function to call the time data of the module.

E. Body Detection Subroutine

Human body detection subroutine is a characteristic function of intelligent lamp, which can intelligently detect whether there is human body in the current environment. The control chip will intelligently control the opening or closing operation of the intelligent lamp according to the detection results. The feature is very convenient and friendly, especially at night. The human body detection subroutine flow is shown in Figure 7.

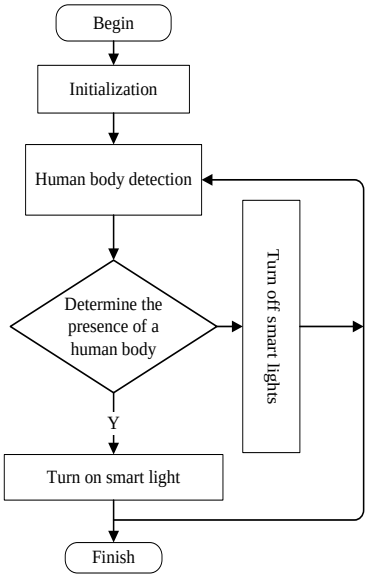


Figure 7. Human body detection subroutine diagram

Once powered on, the smart lamp circulates through the current environment. The intelligent light will be turned on automatically when someone is detected, and will be turned off when no human body is detected.

F. Main Programming

The main program of indoor lighting control system is to initialize and circularly call subroutines such as clock module, display module, human body detection module and light intensity detection module. In this way, the system can be ensured to run smoothly. The main program flow is shown in Figure 8.

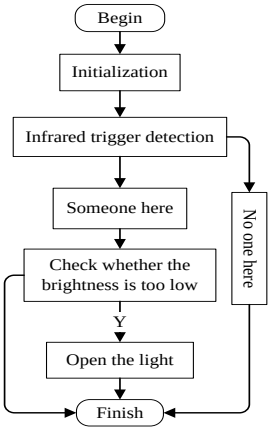


Figure 8. Main program flow

After the indoor lighting control system is powered on the initialization settings, mainly including the initialization of some sub-modules and the declaration of common functions and so on. Make circular calls to sub-modules and between sub-modules.

This chapter mainly uses KEIL and C language to research and develop the software part of the indoor lighting control system. In the process of designing and implementing subroutines such as clock module, display module, human body detection module, light intensity detection module, etc., not only the theoretical knowledge learned is consolidated, but also his practical ability has been strengthened.

V. FUNCTION IMPLEMENTATION AND TESTING

The selection of main modules and components of the indoor lighting control system and the design of software and hardware have been implemented in the previous chapters. The following is mainly to integrate and debug the completed indoor lighting control system

The indoor lighting control system is composed of STM32F103, human body detection module circuit (infrared detection), clock circuit, display circuit, light intensity detection circuit and lamp circuit.

The first thing to do during and after the realization of the electronic system is debugging. Debugging during the implementation process can help to find and correct the shortcomings in the implementation process as early as possible; and the debugging after the implementation can help to find and correct the shortcomings of the system in time. There are many ways to realize the debugging of electronic systems, including software and hardware debugging and module debugging, system joint debugging and simulation debugging.

This system adopts the method of single module debugging and simulation debugging. The indoor lighting control system is debugged through hardware debugging and software debugging respectively. Hardware debugging mainly uses multimeter, oscilloscope and other tools to check the hardware circuit of each submodule; software debugging mainly uses KEIL development tools to detect human body detection module subroutines, button module subroutines, various working mode subroutines, etc., as well as the entire indoor lighting The program of the control system is debugged and a file that can be recognized by the control chip is generated.

The whole system is shown in Figure 9.

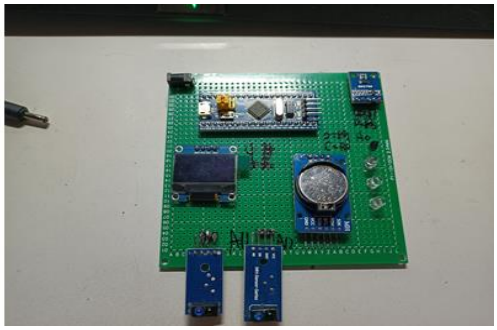


Figure 9. The photo of the system

After the power supply is connected, the system enters the standby state. The screen shows the current working status and ambient brightness with the number of detected people.

If the infrared detection module works properly and detects human body entering the room, the BH1750 module starts to work. If the ambient light is detected below the set value (500 lumens), the system will automatically turn on the light for brightness enhancement. If the infrared detection module detects a human body, but the ambient light is greater than the set value (500 lumens), the system will not turn on the light. The two cases are shown in Figure 10. and Figure 11 respectively.

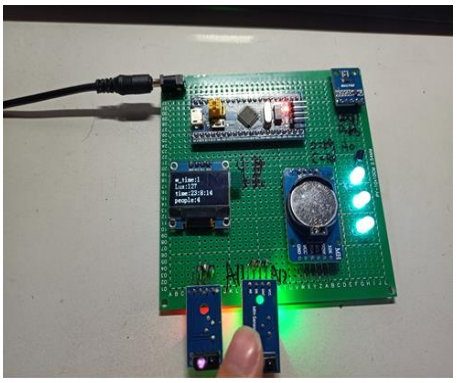


Figure 10. The ambient light is below the set value

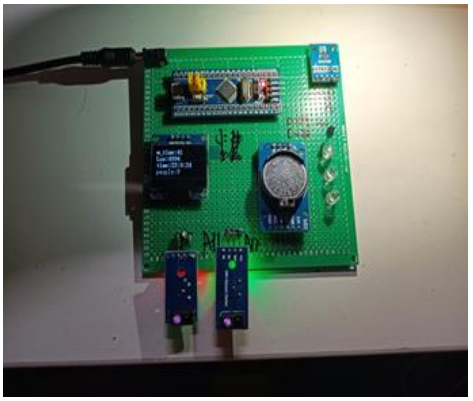


Figure 11. The ambient light intensity is too high

ACKNOWLEDGMENT

This work was partly supported by the Natural Science Basic Research Program of Shaanxi (Program No.2021JM-460) and the Scientific

Research Program Funded by Shaanxi Provincial Education Department (No.21JC033); the higher education scientific research project of Shaanxi higher education society (No. XGH21161) and the undergraduate innovation and entrepreneurship plan (No.202111664039, S202111664070 and 202111664019).

REFERENCES

[1] Bhartiya, G., and P. Pathak. "Intelligent Lighting Control and Energy Management System." 2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC) Feb. 2020, pp.86-89, doi:0.1109/PARC49193.2020.236563.

[2] L., Heng, et al. "Development of an Indoor Intelligent Lighting System Based on Occupants Positioning." International Journal of Pattern Recognition and Artificial Intelligence, 2021. doi:10.1142/S0218001421590412.

[3] As, A, et al. "Efficient photodetector placement for daylight-responsive smart indoor lighting control systems." Journal of Building Engineering, vol.42, Oct. 2021,pp.1-12, doi:10.1016/j.jobe.2021.103013.

[4] H. E.Shao, and X. C. Peng. "Research on Light Distribution Optimization Design of Road Lighting LED Lamps." Western China Communications Science & Technology,vol.3,Feb..2019, pp.148-152. doi:10.13282/j.cnki.wccst.2019.03.043.

[5] N. A. Binti Mohd Arifin and N. M. Thamrin, "Development of Automated Microcontroller-Based Lighting Control System for Indoor Room Implementation," 2018 4th International Conference on Electrical, Electronics and System Engineering (ICEESE), 2018, pp. 82-86, doi: 10.1109/ICEESE.2018.8703500.

[6] W. Byun, Y. Jin, Y. Kim and J. Lim, "Design of Lighting Control System Considering Lighting Uniformity and Discomfort Glare for Indoor Space," 2018 International Conference on Platform Technology and Service (PlatCon), 2018, pp. 1-6, doi: 10.1109/PlatCon.2018.8472750.

[7] A. Seyedolhosseini, N. Masoumi, M. Modarressi and N. Karimian, "Illumination Control of Smart Indoor Lighting Systems Consists of Multiple Zones," 2018 Smart Grid Conference (SGC), 2018, pp. 1-4, doi: 10.1109/SGC.2018.8777883.

[8] F.Y. Deng, Y.W. Qiang, Y.Q. Liu, et al. "Adaptive Parametric Dictionary Design of Sparse Representation Based on Fault Impulse Matching for Rotating Machinery Weak Fault Detection." Measurement Science and Technology , 2020, vol. 31, pp. 117-120. doi: 10.1088/1361-6501/ab6f2f.

[9] T., Tsagkatakis, P. P., Smirnakis and Tsakalides. "Adversarial Dictionary Learning for a Robust Analysis and Modelling of Spontaneous Neuronal Activity." Neurocomputing, 2020, vol. 38, pp. 188-201.

[10] doi: 10.1016/j.neucom.2020.01.041

[11] L. B. Wang, C., Chen, Y. J. Li. "Dictionary Learning Algorithm Based on Weighted Least Square." Systems Engineering and Electronics, vol. 33, 2011, pp. 1896-1900. doi:10.3969/j.issn.1001-506X.2011.08.41.

[12] Naderahmadian, Yashar, Beheshti, Soosan, Tinati and Mohammad: Correlation Based Online Dictionary Learning Algorithm. IEEE Transactions on Signal Processing, vol. 64, 2015, pp. 592-602,

[13] doi:10.1109/TSP.2015.2486743.

[14] U. I. Kizgut and Ersin. "Lossy Compressive Sensing Based on Online Dcitionary Learning." Computing and informatics, vol. 38, 2019, pp. 151-172. doi: 10.31577/CAI_2019_1_151.

[15] N. Takayuki, B., Yukihiro and Kiya, Hitoshi. "Secure Overcomplete Dictionary Learning for Sparse Representation." IEICE Transactions on Information and Systems, vol. E103D, 2020, pp. 50-58. doi: 10.1587/transinf.2019MUP0009.

[16] Li Yun, Fu Ai-Ping. "Designing Metadata for Chinese Dictionary Entries." Data Science Journal, vol. 6, 2007, pp. 853-856, doi: 10.2481/dsj.6.S853

[17] V. T. Huu and Monga Vishal. "Fast Low-Rank Shared Dictionary Learning for Image Classification." IEEE Transactions on Image Processing, vol. 26, 2017, pp. 5160-5175, doi: 10.1109/TIP.2017.2729885.

Style Transfer Based on VGG Network

Zhe Zhao
School of Computer Science and Engineering
Xi'an Technological University
No.2 Xuefu Middle Road, Weiyang district, Xi'an, Shaanxi, China
E-mai:zz_stony@163.com

Shifang Zhang
School of Computer Science and Engineering
Xi'an Technological University
No.2 Xuefu Middle Road, Weiyang district, Xi'an, Shaanxi, China
E-mai:zhangshifang2005@126.com

***Abstract*—With the rapid development of computer computing power, as an important method in the field of artificial intelligence, deep learning has amazing learning ability, especially in dealing with massive data, which makes deep learning in the fields of image recognition, image classification, natural language processing, data mining and unmanned driving, Has shown an extraordinary role. In previous studies, the style transfer algorithm has not developed well due to the poor computing power of Computer, the basic configuration of computer hardware can not meet the minimum requirements and the poor image effect after migration. However, with the development of computer hardware and the rapid change of GPU computing power, the style transfer network based on deep learning has become a hot issue in the study of style transfer in recent years. According to the research, although the traditional style transfer method can obtain the texture, color and other information of the style image, the model needs to be learned every time a new target image is generated, and the time cost during this period is very high. In this way, the trained model is not repeatable, and the generated image is often very random and can not get good results. Therefore, the emergence of style transfer methods based on deep learning solves the limitations of traditional style transfer methods. Style transfer methods based on deep learning are faster than traditional style transfer methods, and the generalization of the model is better. The style transfer algorithms of main neural networks are divided into two categories, Slow style transfer based on image iteration and fast style transfer based on model iteration. VGG network model can combine style image and content image, and greatly improve the style transfer efficiency of image.**

Keywords-VGG Network; Neural Network; Style Tr -ansfer

I. OVERVIEW

Image style transfer technology is to migrate the painting style, stroke, texture and other information of a style image to the content image, and re render the content image, so that the content image can change the color, texture and other information of the style image while retaining the content features. It is a technology with artistic creation and image editing. Style transfer can also be regarded as an extension of texture synthesis. Texture synthesis inputs a content image and a style image, and the generated image retains the structure of the original image and has the artistic style of the style image through the algorithm. The local texture is recorded by statistical model, and then the local texture is synthesized into the overall image texture. Texture based synthesis method is to combine texture and content image, so that the

content image has the texture and color of style image. The significance of style transfer technology is that an ordinary person without any skills can realize the desired style transfer of different images by using the model. In real life, style transfer technology is being applied in various commercial fields, such as Meitu software, animated film production, advertising design and so on. Inspired by the Convolutional Neural Networks (CNN) in visual perception task [1], in 2015, Gatys et al. [2] proposed using VGG network model to achieve the goal of image style migration, with ideal effect, and initiated the research on style transfer technology based on neural network.

II. Introduction of VGG Network

With the wide application of neural network in the field of image processing, convolutional Neural Networks also began to appear frequently in people's vision. Convolutional neural network is composed of multilayer neural network, which mainly includes five hierarchical structures, Input layer, conv layer, ReLU layer, Pooling layer and FC layer. The input layer processes the image, including normalization, resizing, de averaging and so on. Convolution layer is the most important step in convolution neural network. It connects the feature information of each layer of the image, The activation function is mainly used for nonlinear mapping of the output results of the convolution layer. The commonly used activation functions are ReLU function, Sigmod function, Tahn function, etc. The pooling layer is mainly used for image dimensionality reduction or dimensionality upgrading. The purpose of dimensionality reduction is to compress the number of parameters, reduce the over fitting of data and improve the training speed. Dimensionality upgrading is mainly to restore the original feature information of the image. The full connection layer concatenates the data elements after the operations of convolution layer, activate function and pooling layer to obtain the final classification result.

Visual Geometry Group Network (VGG) neural network model is a deep convolution neural network developed by the computer vision group of Oxford University and Google deep mind in 2014 [3]. It was originally born as an image classification network, Since its successful development, vgg-16 and vgg-19 models have been launched, the most commonly used VGG-16 and VGG-19.The VGG network model is shown in Figure 1 VGG Network structure.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figure 1. VGG Network structure.

VGG-16 and VGG-19 models are commonly used. There is no essential difference between them, but the depth of the network is different. The network model structure of VGG-19 is shown in Figure 2. VGG-19 contains 19 hidden layers, consisting of 16 convolution layers and 3 full connection layers. It adopts a continuous 3x3 convolution core, with stripe of 1 and padding of 0. The pool layer uses MaxPooling.

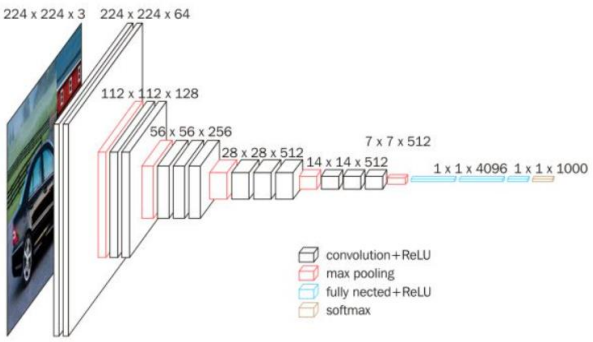


Figure 2. VGG-19 network model structure.

In the VGG network structure, multiple identical 3x3 convolution layers are stacked together. It shows that two 3x3 convolutions in series and one 5x5 convolution have the same receptive field, while three 3x3 convolutions in series and one 7x7 convolution have the same receptive field. This structural design method can reduce the amount of learning parameters and reduce over fitting, This makes the network more capable of learning features, which also makes it a good advantage to select VGG network structure for feature extraction of style migration.

III. IMAGE STYLE TRANSFER

In 2015, Gatys et al. [2] Divided the style transfer of images into two parts: content loss and style loss, and used VGG network as the style transfer network for the first time. The style transfer network of Gatys et al. belongs to the slow style transfer method based on image iteration. The stylized image is generated by pixel iteration on the noisy image, and the style matching is mainly carried out according to the global statistical information. Li and Wand's [9] style transfer method is based on regional fast similarity. The closer the shape of the content image is to the style image, the better the effect. The style transfer method of Johnson et al. [4] And Ulyanov et al. [13] is a fast style transfer method based on model iteration. The parameters of the model are obtained through the pre training of the feedforward network for style transfer. The image generated by this model is faster, but the image effect may not be very good. The style transfer method based on GAN network mainly converts the input image style through the game between generator and discriminator, represented by conditional generative adversarial networks (CGAN), CycleGAN and StarGAN. The advantage of this style transfer method is that the generated image is more realistic. This paper mainly studies the application and improvement of VGG neural network in the field of style transfer.

Establishment of experimental environment:

The experimental project uses Python as the programming language and tensorflow as the mainstream framework to realize VGG-19 neural network model. The processor of the experimental hardware platform is Intel (R) core (TM) i5-6300HQ, the main frequency is 2.50 GHz and the memory is 8.00 GB. The GPU is GTX960M. During the experiment, the selected content images are common landscape images, and the classic oil paintings with distinctive color and style are used as style images to carry out image style transfer experiments under different conditions.

Effects of different model parameters on image style transfer:

The landscape map of Taipei101 building is selected as the content image. The following is the experiment on the influence of different convolution layers on the image style. Figure 3 is the image with white noise, after using the convolution kernel of conv2_1 of VGG-19 network, it can be seen that the low-level convolution check of VGG network has obvious retention of the semantic information of the image, the dividing boundary between buildings is obvious. Using the convolution kernel of conv3_1,

we can see that the edge of the image has been blurred. After passing through the convolution kernel of conv4_1, the edge information has become difficult to identify, and after passing through the convolution kernel of conv5_1,the semantic information has been completely unrecognizable.

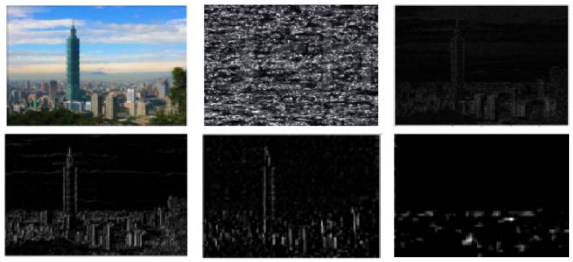


Figure 3. Convolution check of VGG Network and recognition ability of image semantic information.

VGG-19 network is composed of 16 convolution layers and 3 full connections. The low-level convolution layer can well retain texture and semantic information, while the high-level convolution layer often loses important semantic information and blurs the boundary between image objects, but the degree of style will be better. The image stylization algorithm mainly includes three most important parts: content reconstruction, style representation and style transformation. In this paper, the output of the middle and high-level activation function of VGG network is used to represent the content features of the image, mainly including its macro structure and contour, and then the Gram matrix is used to describe its style features. Image style transfer can be realized by minimizing the difference between the content features and style features of the generated image and the input image. The following is the definition of image content loss:

$$l_{content}(\vec{p},\vec{x},l)=\frac{1}{2}\sum_{i,j}(F_{i,j}^{l}-P_{i,j}^{l})^{2}. \quad (1)$$

On the left side of the equation, p represents the content image, x represents the stylized image, and l represents the l layer of the VGG network; $F_{i,j}$ and $P_{i,j}$, j on the right side of the equation represent the jth activation value of the ith feature mapping of the stylized image and the content image in layer l of the VGG network, respectively. The style loss function is defined as follows. The following is the definition of image style loss:

$$E_{l}=\frac{1}{4N_{l}^{2}M_{l}^{2}}\sum_{i,j}(G_{i,j}^{l}-A_{i,j}^{l})^{2}. \quad (2)$$

The following is the definition of Gram matrix:

$$G_{i,j}^{l}=\sum_{k}F_{ij}^{l}F_{jk}^{l}. \quad (3)$$

Where $G_{i,j}^{l}$ is the inner product of feature map i and j in layer l, Where F_{ij}^{l} represents the kth activation value of the ith feature map of the style image in the L layer of the VGG network.

The total loss function is defined as follows:

$$L_{total}(p,a,x)=\beta L_{style}(a,x)+\alpha Lcontent(p,x). \quad (4)$$

Among a, p and x represent style image, content image and generated image respectively; β and α is the weight of style loss function and content loss function in the total loss function.

Select the landscape map of Taipei 101 building as the content image and Van Gogh's star sky as the style image. The final target image generated after different iterations of the model is shown in the figure below：

Figure 4. Image after style transfer of Taipei 101 building.

Each training of such a network takes a lot of time, and the generated images vary greatly according to the number of iterations, which obviously can not meet the requirements of style migration. Therefore, the modification of VGG network structure is also a very important research direction.

A. Improved method of introducing residual block.

The training speed of images generated through VGG-19 network training is very slow, because the loss value needs to be calculated for each style transfer image. Such training requires a lot of computing resources of computers, and it is difficult for ordinary computers to train images with good results in a short time, so Johnson et al. [4] A method of training feedforward network with perceptual loss function is proposed to transfer image style. A feedforward convolutional neural network is trained in supervised mode, and the pixel by pixel gap is used as loss function to measure the gap between output image and input image. The advantage of this design is that only one feedforward is required to pass through the trained network, it greatly saves the time of image style migration. This style transfer method is improved based on the idea of residual network.

In 2014, GoogleNet [9] of Google and VGGNet [3] of visual geometry group of Oxford University once again achieved excellent results in using deep convolution neural network in ilsvrc that year, and was several percentage points better than alexnet in classification error rate, once again pushing the deep convolution neural network to a new peak. Compared with alexnet, these two network structures choose the strategy of continuing to increase the network complexity to enhance the feature representation ability of the network. Generally speaking, the learning degree of the deep convolution network is related to the depth of the network structure. The more layers of the network, the stronger the learning ability. The deep learning network designed based on this idea will have many convolution layers. Although the learning ability of the neural network is improved, the problem is that the parameters become miscellaneous and the speed of the training network will be slower. Therefore, some people put forward the problem of improving many parameters of deep convolution network and speeding up the training speed of convolution network. In 2015, he Kaiming and others from Microsoft Research Asia participated in the ilsvrc of that year using the residual network RESNET [10], and their performance in image classification, target detection and other tasks significantly exceeded the performance level of the competition of the previous year, and finally won the championship. The obvious feature of the residual network is that it has a considerable depth, from 32 layers to 152 layers, which is much deeper than the previously proposed depth network structure, and then a 1001 layer network structure is designed for small data. The depth of residual network RESNET is amazing, and the extremely deep depth makes the network have very strong expression ability.

The experiment of he Kaiming et al. [11] proved that in the same network structure, the deep network learning ability will be relatively improved. However, when the network is deep, continuing to improve the number of layers of the network will not improve the performance. As shown in Figure 5, in the same number of iterations, the training effect of the neural network with deep network layers decreases. With the increase of network layers, the learning ability of the network does not improve, but significantly degrades, and the training error also increases with the increase of layers. If this happens, we usually consider whether the data is over fitted, Whether different activation functions and normalization operations are required. However, such operation will make the network unable to go deeper. How can we ensure the depth of the network without the decline of training degree. So Deep Residual Learning was born.

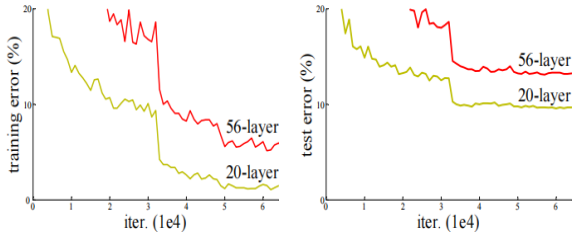


Figure 5. Training error (left) and test error (right) on CIFAR-10.

Let a hidden layer in the depth network be h(x)-x → f (x). If it can be assumed that the combination of multiple nonlinear layers can approximate a complex function, it can also be assumed that the residual of the hidden layer is approximate to a complex function. That is, we can express the hidden layer as H (x) = f (x) + x. In this way, we can get a new residual structure unit, as shown in Figure 6 It can be seen that the output of the residual unit is obtained by adding the output and input elements cascaded by multiple convolution layers (ensuring that the dimensions of the output and input elements of the convolution layer are the same), and then activated by relu. Connecting this structure, the residual network is obtained.

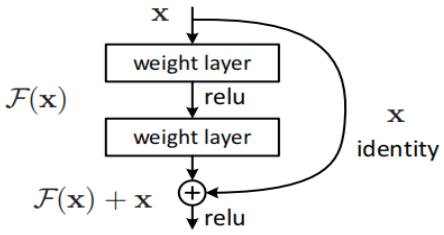


Figure 6. Residual structure.

It can be seen that the network with residual structure has better convergence performance and lower training error rate. As shown in Figure 7.

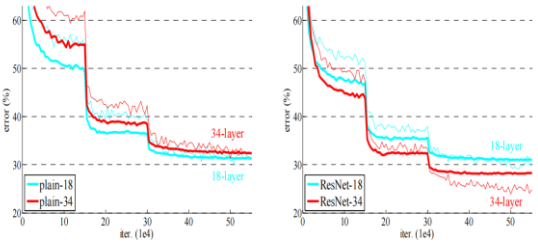


Figure 7. Training on ImageNet.

In this experiment, a pre trained VGG network is used as the classification network. The network layer is composed of convolution layer and residual block. The five-layer convolution layer in VGG-19 model is replaced by five-layer residual block. The last layer uses a scaled tahn function to ensure that the value of the output image is between 0 ~ 255. Except that the first layer and the last layer use 9x9 convolution core, the other layers are 3x3 convolution core. The residual structure is introduced into VGG network to better optimize the network, because its internal residual block uses jump connection, which alleviates the problem of gradient disappearance caused by increasing depth in depth neural network. Traditional neural network may have more or less information loss and loss during information

transmission, If the appropriate residual block is added to the VGG network, the input information can be directly bypassed to the output to protect the integrity of the information and improve the training speed of the network.

Figure 8 below shows the output image after the style transfer of Van Gogh's starry sky. Due to the simple structure of the residual block, it solves the problems of the degradation of the learning ability of the convolution neural network and the slow training speed of many parameters. In addition to the excellent classification ability of the VGG network, it can be seen that the image effect of the style transfer of the VGG network combined with the structure of the residual network is good, but there are also some other problems, For example, it can be clearly seen that the segmentation between image objects after style transfer is not obvious, and some image semantics are lost.

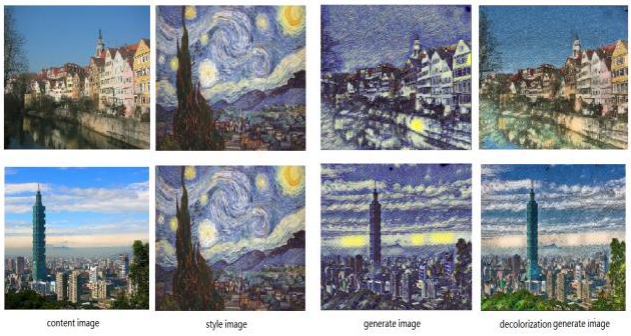


Figure 8. Output image after style conversion.

B. Improved method of introducing encoder -decoder.

The effect of transfer is ensured by calculating the content loss and style loss of the original image and style image, which leads to the need to train the corresponding network for each style, and the training Network is very time-consuming. Style loss and content loss still need to adjust the parameters of layer to get an area that matches the style image, so they can have a better effect. Moreover, this step needs to be retrained for different styles, so they need to be retrained, which will waste a lot of time in the process of a large number of parameters.

In order to solve the above problems, in 2017, Huang et al [4]. Proposed a multi style transfer network and introduced the encoder decoder structure. In 2018, Li et al. [12] Added whiting transform and coloring transform (WCT) operations to the reserved encoder decoder structure to carry out style transfer without training. The advantage of this model is that it can avoid the loss of time caused by model adjustment parameters, and better preserve the texture of the generated image.

The network of encoder decoder structure is an unsupervised learning technology, which uses neural network for characterization learning. It is a neural network that copies the input of the network to the output, compresses the input into a hidden space representation, and then outputs the reconstructed representation. The network consists of encoder and decoder. The encoder compresses the input into potential space, which can be represented by the coding function H = f (x). The decoder is to reconstruct the input from the hidden space, which can be represented by the decoding function r = g (H). The encoder decoder structure can also be understood as training multiple encoders with different layers, so that the input data can be reduced from the original multi-dimensional data to a smaller dimension, and then the reduced dimension data can be used for image classification respectively. In this way, the original big data classification problem will be transformed into a small-scale image classification problem. The encoder decoder structure is shown in Figure 9.

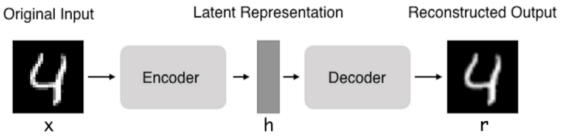


Figure 9. Encoder decoder structure.

The network structure is divided into two parts, generating network and calculating loss network. The generative network is a feedforward network, which is used for style conversion in the later stage. The computational loss network is used to constrain data during training.

The style transfer generation network is composed of encoder AdaIN decoder. The encoder part adopts the pre trained VGG network and only relu4_1. Turn the image space of the style image and the content image to the feature space, and then use the Adain module to normalize the content image. Adain is an adaptive instance normalization. In the feature space, the normalized mean and variance of each channel input of the content image are matched to the mean and variance of each channel input of the style image. Here, the input of content image and style image are feature space.

$$AdaIN(x,y)=\sigma(y)(\frac{x-\mu(x)}{\sigma(x)})+\mu(y). \quad (5)$$

Where x is the content image, y is the style image, is the standard deviation of the style image, is the average value of the content image, is the standard deviation of the content image, and is the average value of the style image.

The decoder part is a network that transforms the feature space into the image space. This part of the network generally adopts the network structure symmetrical to the encoder. What needs to be trained in the whole network is the parameter weight information of this part of the network. Generally, a pool layer is added between the convolution layers. In the process of image processing, the pool layer is mostly used to compress the image Compress the amount of data and parameters to reduce over fitting. The network structure diagram is shown in Figure 10.

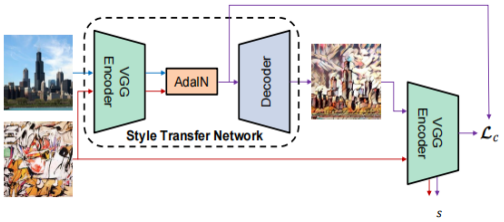


Figure 10. Encoder decoder style transfer network structure.

The style loss function consists of two parts: content loss and style loss. The overall style loss function consists of the sum of the two.

In this experiment, a pre trained VGG network is selected as the encoder to encode the input pictures, then a symmetrical decoder is designed for decoding, and a layer of adain is added between the encoder and the decoder for normalization. The following is the style transfer image obtained by extracting features from a simple encoder-decoder + VGG network. The generated image is shown in Figure 11.

Figure 11. Style transfer image.

It can be seen that the generated style transfer image works well in the image area with simple background, but it doesn't work well in the area with complex background and many objects. In

order to make the degree of stylization high and lose the semantic information of the original content image, many areas are easy to be ignored, and the edge boundary between objects is not obvious, the model does not understand which areas should be preserved and should be noticed when migrating the style image to the content image.

The effect of the style image generated by the style transfer network described above is not very good. Li et al. [12] Considered whether the image requiring style transfer can be improved in addition to the normalization processing, Therefore, it is proposed to color and decolor the image. Firstly, the image is input into the decoder, and then a symmetrical decoder is designed to color and decolor between the two networks, that is, the input feature map subtracts mean value, and then multiplies the inverse matrix of its own covariance matrix to control the centralization of the feature map to a whitening distribution space, That is, the features of the content image are extracted and the style color is removed. Then, the covariance matrix of the feature map is obtained for the style image, multiplied by the result of the whitening of the content image, and then added with the mean value, that is, the feature map after the whitening of the content image is transferred to the distribution of the style map. Before the output is passed into the decoder, the stylization degree can be controlled by adjusting parameters. The control style formula is shown in equation 6 below.

$$f_{cs} = \alpha f_{cs} + (1-\alpha) f_c \tag{6}$$

α is the stylization factor.

Firstly, this experiment trains multiple decoders, inputs the image into the pre trained VGG network, extracts different relu layer structures as the encoder output, trains the decoder for the results of conv layer, and designs multiple decoders for different relu1 to relu5 layers to restore the results of VGG convolution layer.

Figure 12 below shows the generated image obtained by adjusting the stylization parameters when selecting the landscape map and figure map as the content image. It can be seen that the higher the degree of stylization, the more obvious the style of the Image, and the appropriate adjustment parameters can make the fusion effect of content image and style image better.

Figure 12. Style images with different weights.

C. Improved method of introducing Generative Adversarial Network.

Generative adversarial network (GAN) [6] is a network proposed by Goodfellow et al. In 2014, at present, it has become one of the most important research directions in the field of deep learning. This technology is mainly used in the fields of image super-resolution, style transfer, image segmentation, text to image generation, natural language generation. GAN is based on the idea of two person zero sum game in game theory, in which both sides of the game are generators and discriminators in GAN. The function of the generator is to generate a sample similar to the real training data according to the input random noise. The purpose of discriminator is to distinguish between real data and generated data. The function of the generator is to generate a sample similar to the real training data according to the input random noise. The purpose of

discriminator is to distinguish between real data and generated data. In order to win the game, both generator and discriminator need to improve their generation and discrimination ability. The ultimate purpose is to find Nash equilibrium between generator and discriminator. Based on this principle, the generated countermeasure network can make the generated image close to the real image.

In recent years, many scholars have proposed a variety of improved GAN algorithms according to different application scenarios. Radford et al.[7] fused CNN (revolutionary neural network) and GAN and proposed deep convolution to generate countermeasure network, which makes the model training more stable and the generated images more diversified. Zhu et al. proposed CycleGAN [8] using bidirectional Gan, so as to control the learning of the model.

Compared with the traditional generation countermeasure network, CycleGAN has two main improvements: (1) The input of the traditional generation countermeasure network is random noise, so it can only generate pictures randomly, so the quality of the generated images can not be controlled. CycleGAN changes the input to the given picture data to control image generation. (2) In the past, the conversion between images, such as gray image to color image, image to semantic label, day image to night image, etc., required paired training data. However, in real life, the acquisition of paired data is difficult and expensive. CycleGAN can realize the conversion from the input image to the target image without paired training data. The main principle of CycleGAN is to introduce the cyclic consistent loss function based on the counter loss of GAN. The anti loss control generates an image close to the target image, and the cyclic consistent loss is used to preserve the content structure of the input image and the characteristics of the target image. When the image is generated, the potential relationship of multiple feature domains is found through training, so as to transform the relevant domain according to the input image. However, when the conversion degree is not constrained, the generation result of CycleGAN will have the obvious disadvantage of arbitrary change of irrelevant domain characteristics.

Figure 13 are CycleGAN demonstration pictures.

Figure 13. Night and Day switch.

This paper uses the method of combining VGG network and CycleGAN. The network structure is composed of encoder, decoder and converter.

Encoder: The images are input into the neural network in turn to extract different images type style. Convolution layer using VGG-19 network, the number of filters in the first convolution layer is 64. When input to the encoder, The size of the is 256 ×256, resulting in 256 ×64 ×64 feature map.

Converter Transform an image from one domain to another.

Decoder The decoder is the inverse process of the encoder. Also from the eigenvector, the original work of low-level features can wait for image generation.

Discriminator The discriminator predicts whether each image is the original image or the generated image formed image.

Figure 14 below shows the target image generated after training with the same content image and different style images. It can be seen that the image generated after using the CycleGAN structure is more realistic.

Figure 14. Generate different style images.

IV. SUMMARY

This paper takes image transfer as the main research content, extracts the characteristics of content image and style image through VGG network model, realizes the style transfer of generated image, and introduces some improvement measures for VGG network, which makes VGG network model more suitable for style migration. Experiments show that the image style transfer effect achieved by using the methods mentioned in this paper is good and the image generation speed is fast, but these methods also have some limitations, the generated image often loses some semantics, and the texture features and edge boundaries are fuzzy. In the future, we will continue to study and improve the image transfer algorithm, further improve the accuracy of stylized images, and extend the research results to practical product applications.

REFERENCES

[1] Krizhevsky A, Sutskever I, Hinton G E.Imagenet classification with deep convolutional neural networks. In: Proceedings of the 2012 Advances in Neural Information Processing Systems (NIPS). Lake Tahoe, USA: MIT Press, 2012. 1097–1105.

[2] GATYS L, ECKER A, BETHGE M.A Neural Algorithm of Artistic Style [J]. Journal of Vision, 2016, 16(12): 326.

[3] SIMONYAN K, ZISSERMAN A. Very deep con-v olutional networks for large-scale image recogn-itio n [J]. ArXiv preprint ar Xiv, 2014(9):1-14.

[4] JOHNSON J, ALAHI A, FEI-FEI L. Perceptual lo ss-es for real-time style transfer and super-resolutio n[C]//European Conference on Computer Vision, 20 16:694-711.

[5] LI Y J, FANG C, YANG J M, et al. Universal style transfer via feature transforms[C]//In Advances in Neural Information Processing Systems. California: NIPS, 2017: 386-396.

[6] Goodfellow I, Pouget-Abadie J, Mirza M, et al.Generative adversarial nets[C]//Advances in Neur-al Information Processing Systems, 2014: 2672-2680.

[7] Radford A, Metz L, Chintala S. Unsup-ervised repr esentation learning with deep convolutional generati ve adversarial networks [J]. arXiv : 1511.06434, 20 15.

[8] Zhu J Y Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//IEEE International Conference on Computer Vision, 2017: 2223-2232.

[9] Szegedy, Christian,et al. Going deeper wi-th convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

[10] He, Kaiming, et al. Deep residual learning f-or image recognition. arXiv preprint arXiv:1512.03385(2015).

[11] He, Kaiming, and Jian Sun. Convolutional neural networks at constrained time cost. Proceedin-gs of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

[12] HUANG X, BELONGIE S. Arbitrary style tra-nsferin real-time with adaptive instance normalization[C]//2017 IEEE International Conference on Com-puter Vision (ICCV). New York: IEEE Press, 2017:1510-1519.

[13] ULYANOV D, VEDALDI A, LEMPITSKY V. Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis[C]//Proceedingsof the IEEE Conference on Computer Vision and Pattern Recognition, 2017:6924-6932.

基于 VGG 网络的风格迁移方法

赵哲
计算机科学与工程学院
西安工业大学
中国陕西省西安市未央区学府中路 No.2 号
邮箱：zz_stony@163.com

张市芳
计算机科学与工程学院
西安工业大学
中国陕西省西安市未央区学府中路 No.2 号
邮箱：zhangshifang2005@126.com

***摘要*—随着计算机计算能力的飞速发展，深度学习作为人工智能领域的一种重要方法，具有惊人的学习能力，特别是在处理海量数据方面，使得深度学习在图像识别、图像分类、自然语言处理等领域得到了广泛的应用，数据挖掘和无人驾驶，已经显示出非凡的作用。在以往的研究中，由于计算机的计算能力差，计算机硬件的基本配置不能满足最低要求，迁移后的图像效果差，使得风格转换算法没有得到很好的发展。然而，随着计算机硬件的发展和 GPU 计算能力的快速变化，基于深度学习的风格迁移网络成为近年来风格迁移研究的热点问题。研究表明，传统的风格迁移方法虽然可以获得风格图像的纹理、颜色等信息，但每次生成新的目标图像时都需要学习模型，这一过程的时间开销非常大。在这种情况下，训练的模型是不可重复的，生成的图像往往是非常随机的，不能得到很好的结果。因此，基于深度学习的风格迁移方法的出现，解决了传统风格迁移方法的局限性。基于深度学习的风格迁移方法比传统的风格迁移方法速度快，模型的泛化性好。主要神经网络的风格迁移算法分为基于图像迭代的慢速风格迁移和基于模型迭代的快速风格迁移两大类。VGG 网络模型可以将风格图像和内容图像结合起来，大大提高了图像的风格迁移效率。**

关键词：VGG 网络；神经网络；风格迁移

1. 前言

图像风格迁移技术是将风格图像的绘画风格、笔划、纹理等信息迁移到内容图像中，并对内容图像进行重新渲染，使内容图像在保留内容特征的同时改变风格图像的颜色、纹理等信息。它是一种集艺术创作和图像编辑于一体的技术。风格转换也可以看作是纹理合成的延伸。纹理合成输入内容图像和风格图像，通过该算法生成的图像保留了原始图像的结构，具有风格图像的艺术风格。通过统计模型记录局部纹理，然后将局部纹理合成为整体图像纹理。基于纹理的合成方法是将纹理和内容图像相结合，使内容图像具有风格图像的纹理和颜色。风格迁移技术的意义在于，一个没有任何技能的普通人可以利用该模型实现不同图像的理想风格迁移。在现实生活中，风格迁移技术正被应用于各个商业领域，如美图软件、动画电影制作、广告设计等。受视觉感知任务[1]中卷积神经网络（CNN）的启发，2015 年，Gatys 等人[2]提出利用 VGG 网络模型实现图像风格迁移的目标，效果理想，并启动了基于神经网络的风格迁移技术的研究。

2. VGG 网络的介绍

随着神经网络在图像处理领域的广泛应用，卷积神经网络也开始频繁出现在人们的视觉中。卷积神经网络由多层神经网络组成，主要包括输入层、conv 层、ReLU 层、池层和全连接层五个层次结构。输入层处理图像，包括归一化、调整大小、去均值等。卷积层是卷积神经网络中最重要的一步。它连接图像各层的特征信息，激活函数主要用于卷积层输出结果的非线性映射。常用的激活函数有 ReLU 函数、Sigmod 函数、Tahn 函数等，池层主要用于图像降维或升级。降维的目的是压缩参数个数，减少数据的过拟合，提高训练速度。维数提升主要是恢复图像

的原始特征信息。全连接层通过卷积层、激活函数和池层的操作将数据元素连接起来，得到最终的分类结果。

Visual Geometry Group Network（VGG）神经网络模型是由牛津大学计算机视觉小组和谷歌 deep mind 于 2014 年开发的深度卷积神经网络[3]。它最初是作为一个图像分类网络诞生的，自其成功开发以来，最常用的是 vgg-16 和 vgg-19。VGG 网络模型如图 1 VGG 网络结构所示。

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figure 1. VGG 网络结构

通常使用的是 VGG-16 和 VGG-19 网络。它们之间没有本质上的区别，但网络的深度不同。VGG-19 的网络模型结构如图 2 VGG 网络模型所示。VGG-19 包含 19 个隐藏层，包括 16 个卷积层和 3 个全连接层。采用连续 3x3 卷积核，步长为 1，填充为 0。池化层使用最大池化（MaxPooling）。

在 VGG 网络结构中，多个相同的 3x3 卷积层堆叠在一起。结果表明，两个 3x3 串联卷积和一个 5x5 卷积具有相同的感受野，而三个 3x3 串联卷积和一个 7x7 卷积具有相同的感受野。这种结构设计方法可以减少学习参数的数量，减少过拟合，使网络更具有学习特征的能力，这也使得选择 VGG 网络结构进行风格迁移的特征提取具有很好的优势。

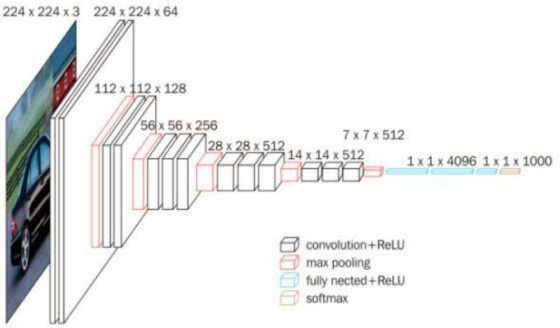


Figure 2. VGG 网络模型

3. 图像风格迁移

2015 年，Gatys 等人[2]将图像的风格迁移分为两部分：内容损失和风格损失，并首次使用 VGG 网络作为风格迁移网络。Gatys 等人的风格迁移网络属于基于图像迭代的慢速风格迁移方法。在噪声图像上通过像素迭代生成风格化图像，风格匹配主要根据全局统计信息进行。Li 和 Wand 的[9]风格转换方法基于区域快速相似性。内容图像的形状越接近风格图像，效果越好。Johnson 等人[4]和 Ulyanov 等人[13]的风格转换方法是一种基于模型迭代的快速风格转换方法。该模型的参数是通过前馈网络的预训练来获得的。该模型生成的图像速度较快，但图像效果可能不是很好。基于 GAN 网络的风格迁移方法主要通过生成器和鉴别器之间的博弈来转换输入图像的风格，以条件生成对抗网络（CGAN）、CycleGAN 和 StarGAN 为代表。这种风格迁移方法的优点是生成的图像更逼真。本文主要研究 VGG 神经网络在风格转换领域的应用和改进。

实验环境搭建：

该实验项目以 Python 为编程语言，以张量流为主流框架，实现 VGG-19 神经网络模型。实验硬件平台的处理器为 Intel(R)核心(TM)i5-6300HQ，主频率为 2.50GHz，内存为 8.00GB。GPU 是 GTX960M。在实验过程中，选取的内容图像为常见的景观图像，以色彩和风格独特的经典油

画作为风格图像，在不同条件下进行图像风格迁移实验。

不同模型参数对图像风格迁移程度的影响：

选择台北 101 楼的景观图作为内容图像。下面是关于不同卷积层对图像风格的影响的实验。图 3 为白噪声图像，使用 VGG-19 网络的 conv2_1 卷积核，可以看出 VGG 网络的低层卷积检查对图像语义信息保留明显，建筑物之间的边界明显。利用 conv3_1 的卷积核，我们可以看到图像的边缘已经被模糊了。通过 conv4_1 的卷积核后，边缘信息变得难以识别，通过 conv5_1 的卷积核后，语义信息已经完全无法识别。

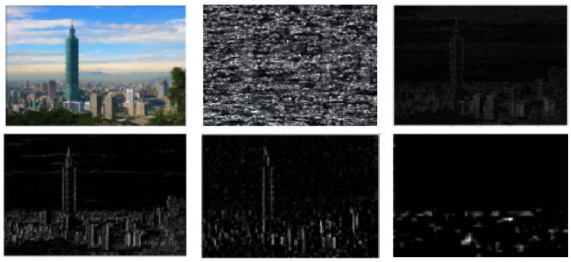


Figure 3. VGG 网络的卷积检查和图像语义信息的识别能力

VGG-19 网络由 16 个卷积层和 3 个全连接组成。低级卷积层可以很好地保留纹理和语义信息，而高级卷积层往往会丢失重要的语义信息，模糊图像对象之间的边界，但风格的程度会更好。图像风格化算法主要包括三个最重要的部分：内容重构、风格表示和风格转换。本文利用 VGG 网络的中、高级激活函数的输出来表示图像的内容特征，主要包括其宏观结构和轮廓，然后利用 Gram 矩阵来描述其风格特征。通过最小化生成的图像与输入图像的内容特征和风格特征之间的差异，可以实现图像风格迁移。以下是图像内容丢失的定义：

$$l_{content}(\vec{p},\vec{x},l)=\frac{1}{2}\sum_{i,j}(F_{i,j}^{l}-P_{i,j}^{l})^{2}. \quad (1)$$

在上述公式的左侧，p 表示内容图像，x 表示程式化图像，l 表示 VGG 网络的 l 层，方程右侧的 $F_{i,j}$，$P_{i,j}$ 和 j 分别表示 VGG 网络第 l 层中内容图像的 ith 特征映射的第 j 个激活值。风格损失函数的定义如下。以下是图像风格损失的定义：

$$E_l=\frac{1}{4N_l^2M_l^2}\sum_{i,j}(G_{i,j}^{l}-A_{i,j}^{l})^{2}. \quad (2)$$

Gram 矩阵的定义如下：

$$G_{i,j}^{l}=\sum_{k}F_{ij}^{l}F_{jk}^{l}. \quad (3)$$

$G_{i,j}^{l}$ 为 l 层特征映射 i 和 j 的内积，其中为 VGG 网络 L 层风格图像第 h 个特征映射的第 k 个激活值。

总损失函数的定义如下：

$$L_{total}(p,a,x)=\beta L_{style}(a,x)+\alpha Lcontent(p,x). \quad (4)$$

其中，a,p 和 x 分别表示风格图像、内容图像和生成图像， β 和 α 为风格损失函数和内容损失函数在总损失函数中的权重。

选择台北 101 大厦景观图为内容图像，以梵高的星控为风格图像。模型在经过不同迭代后生成的最终目标图像如下图所示：

Figure 4. 台北 101 大厦风格迁移后的图像

这种网络的每次训练都需要大量的时间，生成的图像随着迭代次数的不同而变化很大，显

然不能满足风格迁移的要求。因此，对 VGG 网络结构的修改也是一个非常重要的研究方向。

3.1 引入残差块的改进方法

通过 VGG-19 网络训练生成的图像的训练速度非常慢，因为需要计算每个风格图像的损失值。这样的训练需要大量的计算机计算资源，而普通计算机很难在短时间内训练出效果好的图像，所以 Johnson 等人[4]提出了一种具有感知损失函数的前馈网络训练图像风格的方法。在监督模式下训练前馈卷积神经网络，利用像素逐像素间隙作为损失函数来测量输出图像与输入图像之间的间隙。该设计的优点是只需要一个前馈就可以通过训练后的网络，大大节省了图像风格迁移的时间。基于残差网络的思想进行了改进。

2014 年，谷歌的 GoogleNet[9]和牛津大学视觉几何组的 VGGNet[3]再次在当年的 ilsvrc 中使用深度卷积神经网络取得了优异的成绩，分类错误率比 alexnet 高出几个百分点，再次将深度卷积神经网络推向了一个新的高峰。与 alexnet 相比，这两种网络结构选择了继续增加网络复杂度的策略，以增强网络的特征表示能力。一般来说，深度卷积网络的学习程度与网络结构的深度有关。网络层次越多，学习能力越强。基于这一思想设计的深度学习网络将有许多卷积层。虽然提高了神经网络的学习能力，但存在的问题是参数变得繁杂，训练网络的速度较慢。因此，有人提出了改进深度卷积网络的许多参数，加快卷积网络训练速度的问题。2015 年，微软亚洲研究院的何开明等人使用残差网络 RESNET[10]参加了当年的 ILSVRC，他们在图像分类、目标检测等任务上的表现明显超过了前一年的比赛表现水平，最终获得冠军。残差网络的明显特征是它具有相当大的深度，从 32 层到 152 层，比之前提出的深度网络结构要深得多，然后针对小数据设计了 1001 层网络结构。剩余网络 RESNET 的深度惊人，极深的深度使网络具有很强的表达能力。

何凯明等人的实验[11]证明了在相同的网络结构中，深度网络学习能力将得到相对提高。但是，当网络深度较深时，继续提高网络的层数并不会提高性能。如图 5 所示，在相同的迭代次数下，具有深度网络层的神经网络的训练效果降低。随着网络层数的增加，网络的学习能力并没有提高，但会显著下降，训练误差也随着层数的增加而增加。如果发生这种情况，我们通常会考虑数据是否被过度拟合，是否需要不同的激活函数和归一化操作。然而，这样的操作将使网络无法更深入。如何在不导致训练程度下降的情况下，保证网络的深度。所以，深度残差学习诞生了。

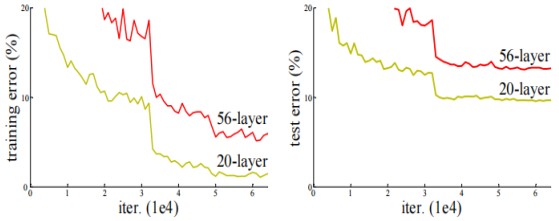


Figure 5. CIFAR-10 上的训练错误（左）和测试错误（右）

设深度网络中的一个隐含层为 h(x)-x→f(x)。如果可以假设多个非线性层的组合可以近似于一个复合函数，那么也可以假设隐层的残差近似于一个复合函数。也就是说，我们可以将隐层表示为 H(x)=f(x)+x。这样，我们可以得到一个新的残差结构单元，如图 6 所示，可以看出残差单元的输出是通过添加多个卷积层级联的输出和输入单元（确保卷积层的输出和输入单元的尺寸相同）得到的输出，然后由 relu 激活。连接该结构，得到了残差网络。

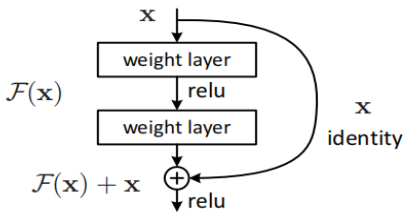


Figure 6. 残差结构

可以看出，具有残差结构的网络具有较好的收敛性能和较低的训练错误率。如图 7 所示。

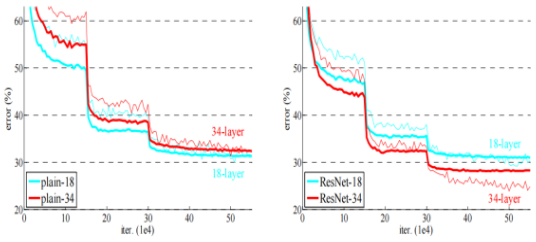


Figure 7. 在 ImageNet 上的训练

本实验采用预先训练好的 VGG 网络作为分类网络。网络层由卷积层和残差块组成。将 VGG-19 模型中的五层卷积层替换为五层残积块。最后一层使用一个缩放的 tahn 函数，以确保输出图像的值在 0~255 之间。除了第一层和最后一层分别使用 9x9 卷积核外，其他层均为 3x3 卷积核。将 VGG 网络引入残差结构，以更好地优化网络，其内部残差块采用跳转连接，缓解了深度神经网络深度增加导致的梯度消失问题。传统的神经网络可能有或多或少的信息损耗在信息传输，如果适当的残差块添加到 VGG 网络，输入信息可以直接绕过到输出保护信息的完整性和提高网络的训练速度。

下述的图 8 显示了梵高的星空风格迁移后的输出图像。由于残差块结构简单，解决了卷积神经网络学习能力下降和多参数训练速度慢等问题。除了 VGG 网络优秀的分类能力外，可以看出 VGG 网络的风格传输的图像效果很好，但也存在一些问题，例如可以清楚地看出风格传输后图像对象之间的分割不明显，一些图像语义丢失。

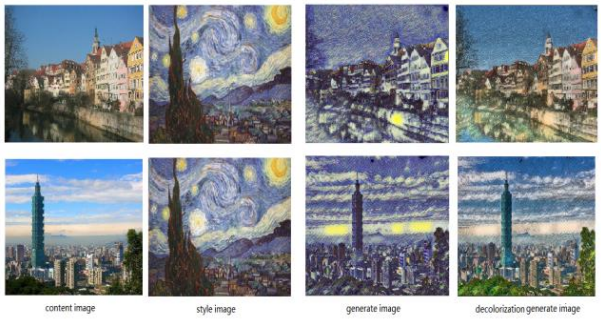


Figure 8. 风格迁移后的输出图像

3.2 引入编码器-解码器的改进方法

通过计算原始图像和风格图像的内容丢失和风格丢失来保证传输的效果，因此需要对每种风格进行相应的网络训练，并且训练网络非常耗时。风格丢失和内容丢失仍然需要调整图层的参数，才能得到与风格图像相匹配的区域，这样才能有更好的效果。此外，该步骤需要针对不同的风格进行再培训，因此需要进行再培训，这将在大量参数的过程中浪费大量时间。

为了解决上述问题，在 2017 年，Huang 等人[4]提出了一种多风格的传输网络，并介绍了编码器解码器的结构。2018 年，Li 等人[12]在预留的编码器解码器结构中增加了白变换和着色变换(WCT)操作，不经训练即可进行风格传输。该模型的优点是可以避免模型调整参数造成的时间损失，更好地保存所生成图像的纹理。

编码器解码器结构网络是一种无监督学习技术，它利用神经网络进行表征学习。它是一个神经网络，它将网络的输入复制到输出中，将输入压缩成一个隐藏的空间表示，然后输出重构的表示。该网络由编码器和解码器组成。编码器将输入压缩到潜在空间中，可以用编码函数 H=f(x)表示。解码器是从隐藏空间重构输入，可以用解码函数 r=g(H)表示。编码器解码器结构也可以理解为训练不同层的多个编码器，使输入数据从原始多维数据减少到更小的维数，然后将降维数据分别用于图像分类。这样，将将原始的大数据分类问题转化为一个小规模的图像分类问题。编码器解码器的结构如图 9 所示。

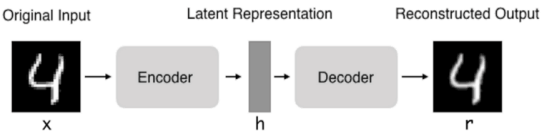


Figure 9. 编码器解码器结构

网络结构分为生成网络和计算损失网络两部分。生成网络是一个前馈网络，用于后期的风格迁移。计算损失网络用于训练过程中数据的约束。

风格迁移生成网络由编码器 AdaIN 解码器组成。编码器部分采用预先训练好的 VGG 网络，仅采用 relu4_1。将风格图像和内容图像的图像空间转到特征空间，然后使用 Adain 模块对内容图像进行归一化。Adain 是一种自适应的实例规范化。在特征空间中，将内容图像的每个通道输入的归一化均值和方差与风格图像的每个通道输入的均值和方差相匹配。在这里，内容图像和风格图像的输入是特征空间。

$$AdaIN(x,y)=\sigma(y)(\frac{x-\mu(x)}{\sigma(x)})+\mu(y). \quad (5)$$

其中 x 是内容图像，y 是风格图像，是风格图像的标准差，是内容图像的平均值，是内容图像的标准差，是风格图像的平均值。

解码器部分是将特征空间转换为图像空间的网络。这部分网络一般采用与编码器对称的网络结构。在整个网络中需要训练的是这部分网络的参数权重信息。通常，在卷积层之间添加一个池层。在图像处理过程中，池层主要用于压缩图像，压缩数据量和参数，以减少过拟合。网络结构图如图 10 所示。

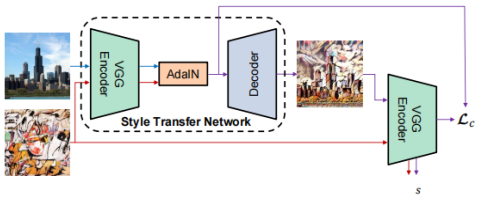


Figure 10. 编码器解码器式的传输网络结构

风格丢失功能包括内容丢失和风格丢失两部分。整体风格损失函数由两者的和组成。

本实验选择预先训练好的 VGG 网络作为编码器，对输入图像进行编码，设计对称解码器进行解码，在编码器和解码器之间添加一层 adain 进行归一化。以下是从一个简单的编解码器+VGG 网络中提取特征所获得的风格传输图像。生成的图像如图 11 所示。

Figure 11. 风格迁移图像

可以看出，生成的风格传输图像在背景简单的图像区域效果良好，但在背景复杂、对象多的区域效果不佳。为了使风格化程度高，失去原始内容图像的语义信息，许多区域容易被忽略，对象之间的边缘边界不明显，模型不了解哪些区域应该保留，在将风格图像迁移到内容图像时应该注意。

由上述风格传输网络生成的风格图像的效果不是很好。李彦等人。[12]考虑到除了归一化处理外，是否还能对需要风格转移的图像进行改进，因此，提出对图像进行着色和脱色。首先将图像输入解码器，然后设计对称解码器在两个网络之间着色和脱色，即输入特征图减去平均值，然后将其自身的协方差矩阵的逆矩阵相乘，控制特征映射的集中到白化分布空间，即提取内容图像的特征并去除风格颜色。然后，得到风格图像的特征图的协方差矩阵，乘以内容图像的白化结果，再加入平均值，即将内容图像的白化后的特征图转移到风格图的分布中。在输出迁移到解码器之前，可以通过调整参数来控制风格化程度。控制式公式如下式 6 所示。

$$f_{cs}=\alpha f_{cs}+(1-\alpha)f_c \quad (6)$$

α 是权重控制因子。

本实验首先训练多个解码器，将图像输入预先训练的 VGG 网络，提取不同的 relu 层结构作为编码器输出，训练解码器的结果，为不同的 relu1 设计多个解码器来对 VGG 卷积层的结果还原。

下图 12 显示了在选择风景图和人像图作为内容图像时，通过调整风格化参数而生成的图像。可以看出，风格化的程度越高，图像的风格越明显，以及适当的调整参数可以使内容图像与风格图像的融合效果更好。

Figure 12. 具有不同权重的风格图像

3.3 引入生成式对抗性网络的改进方法

生成性对抗网络（GAN）[6]是 Goodfellow 等人于 2014 年提出的一种网络，目前已成为深度学习领域最重要的研究方向之一。该技术主要应用于图像超分辨率、风格迁移、图像分割、文本图像生成、自然语言生成等领域。GAN 基于博弈论中的二人零和博弈思想，博弈双方都是 GAN 中的发生器和鉴别器。生成器的功能是根据输入的随机噪声生成与实际训练数据相似的样本。鉴别器的目的是区分真实数据和生成的数据。生成器的功能是根据输入的随机噪声生成与实际训练数据相似的样本。鉴别器的目的是区分真实数据和生成的数据。为了赢得比赛，生成器和鉴别器都需要提高它们的生成和识别能力。最终目的是在生成器和鉴别器之间找到纳什均衡。基于此原理，生成的对抗网络可以使生成的图像接近真实图像。

近年来，许多学者根据不同的应用场景提出了多种改进的 GAN 算法。Radford 等人[7]将 CNN（革命性的神经网络）和 GAN 融合，提出深度卷积生成对策网络，使模型训练更加稳定，生成的图像更加多样化。Zhu 等人提出了使用双向 Gan 的 CycleGAN[8]，以控制模型的学习。

与传统生成对抗网络相比，CycleGAN 有两个主要改进：（1）传统生成对抗网络的输入是随机噪声，只能随机生成图片，生成的图像质量无法控制。CycleGAN 更改给定图片数据的输入以控制图像生成。（2） 过去，图像之间的转换，如灰度图像到彩色图像、图像到语义标签、白天图像到夜晚图像等，都需要成对的训练数据。然而，在现实生活中，成对数据的获取既困难又昂贵。CycleGAN 可以实现从输入图像到目标图像的转换，无需成对的训练数据。CycleGAN 的主要原理是基于 GAN 的反向损耗引入循环一致损耗函数。防丢失控制生成接近目标图像的图像，循环一致丢失用于保持输入图像的内容结构和目标图像的特征。在生成图像时，通过训练发现多个特征域之间的潜在关系，从而根据输入图像对相关域进行变换。然而，当转换度不受约束时，CycleGAN 的生成结果将具有不相关域特征任意改变的明显缺点。

图 13 是 CycleGAN 的生成图像。

Figure 13. 夜间和日间转换

本文采用了 VGG 网络与 CycleGAN 相结合的方法。该网络结构由编码器、解码器和转换器组成。

编码器：将图像依次输入神经网络，提取不同的图像类型风格。卷积层使用 VGG-19 网络，第一卷积层的滤波器数为 64 个。当输入到编码器时，其大小为 256×256，从而得到 256×64×64 特征图。

转换器将图像从一个域转换到另一个域。

解码器解码器是编码器的反过程。同样从特征向量出发，低级特征的原始工作也可以等待图像的生成。

鉴别器预测每幅图像是原始图像还是生成的图像形成的图像。

下面的图 14 显示了使用相同内容图像和不同风格图像进行训练后生成的目标图像。可以看出，使用 CycleGAN 结构后生成的图像更加真实。

Figure 14. 生成不同风格图像

4. 总结

本文以图像的风格迁移为主要研究内容，通过 VGG 网络模型提取内容图像和风格图像的特征，实现生成图像的风格迁移，并介绍了 VGG 网络的一些改进措施，使 VGG 网络模型更适合风格迁移。实验表明，采用本文所述方法实现的图像风格迁移效果好，图像生成速度快，但也有一定的局限性，生成的图像往往会丢失一些语义，纹理特征和边缘边界模糊。在未来，我们将继续研究和改进图像传输算法，进一步提高程式化图像的准确性，并将研究结果扩展到实际的产品应用中。

参考文献

[14] Krizhevsky A, Sutskever I, Hinton G E.Imagenet classification with deep convolutional neural networks. In: Proceedings of the 2012 Advances in Neural Information Processing Systems (NIPS). Lake Tahoe, USA: MIT Press, 2012. 1097–1105.

[15] GATYS L, ECKER A, BETHGE M.A Neural Algorithm of Artistic Style [J]. Journal of Vision, 2016, 16(12): 326.

[16] SIMONYAN K, ZISSERMAN A. Very deep con-v olutional networks for large-scale image recogn-itio n [J]. ArXiv preprint ar Xiv, 2014(9):1-14.

[17] JOHNSON J, ALAHI A, FEI-FEI L. Perceptual los s-es for real-time style transfer and super-resolution [C]//European Conference on Computer Vision, 201 6:694-711.

[18] LI Y J, FANG C, YANG J M, et al. Universal style transfer via feature transforms[C]//In Advances in Neural Information Processing Systems. California: NIPS, 2017: 386-396.

[19] Goodfellow I, Pouget-Abadie J, Mirza M, et al.Generative adversarial nets[C]//Advances in Neur-al Information Processing Systems, 2014: 2672-2680.

[20] Radford A, Metz L, Chintala S. Unsup-ervised repr esentation learning with deep convolutional generati ve adversarial networks [J]. arXiv : 1511.06434, 20 15.

[21] Zhu J Y Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//IEEE International Conference on Computer Vision, 2017: 2223-2232.

[22] Szegedy, Christian,et al. Going deeper wi-th convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

[23] He, Kaiming, et al. Deep residual learning f-or image recognition. arXiv preprint arXiv:1512.03385(2015).

[24] He, Kaiming, and Jian Sun. Convolutional neural networks at constrained time cost. Proceedin-gs of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

[25] HUANG X, BELONGIE S. Arbitrary style tra-nsferin real-time with adaptive instance normalization[C]//2017 IEEE International Conference on Com-puter Vision (ICCV). New York: IEEE Press, 2017:1510-1519.

[26] ULYANOV D, VEDALDI A, LEMPITSKY V. Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis[C]//Proceedingsof the IEEE Conference on Computer Vision and Pattern Recognition, 2017:6924-6932.

Development of An Android-Based Visual Implementation of Student Project Allocation System

Abisha D
Assistant Professor
National Engineering College, K.R Nagar
Kovilpatti-628503, Tamil Nadu
Orcid Id: 0000-0002-7528-0698
E-mail: abisha_cse@nec.edu.in

Sneka G
Student, Department of Computer Science and Engineering
National Engineering College, K.R Nagar
Kovilpatti-628503, Tamil Nadu
Orcid Id: 0000-0002-7528-0698

Aishwaryalakshmi R K
Assistant Professor
National Engineering College, K.R Nagar
Kovilpatti-628503, Tamil Nadu
Orcid Id: 0000-0002-7528-0698
E-mail: abisha_cse@nec.edu.in

Deepitha N
Student, Department of Computer Science and Engineering
National Engineering College, K.R Nagar
Kovilpatti-628503, Tamil Nadu
Orcid Id: 0000-0002-7528-0698

***Abstract*—Technology is everyplace we go in moment's life. So, scholars of seminaries or sodalities, or universities bear an operation that supports smartphones to get all types of information related to examination, lecture notes, placement, systems regarding announcement, events, transportation, etc. Rather of calling systems because nearly all mobile druggies have smartphones currently. This being system takes a pupil list and allocates administrators to scholars. Using a manual system in the management and allocation of projects to students is characterized by many problems, including the Inability of the project Guide to know that a title has been approved already for a student. Difficulty and inappropriate documentation of allocated project topics. Ineffective in entering, updating, and retrieving records of allocated projects. Difficulty in accessing the project Guide for approval of the topic. Duplication in project topics approved for students. We designed an operation to attain the demand of scholars. The main ideal of the pupil design allocation system is to make a system that will give information for each pupil. Projects can efficiently be allocated to students without delay, and topic conflict between students in the same department will not arise here. Pupil updates can be fluently penetrated if the database system is enhanced. This design will give a fruitful way to manage data at a low cost. The Student Project Allocation contains colorful options similar as login/ logout, viewing and streamlining data, etc. It'll be secure. Data can be used by only those with an id and word while maintaining the data.**

Keywords-Android Studio; Java

I. INTRODUCTION

Mobile app development involves jotting software for small wireless computing bias, similar as smartphones. It frequently takes advantage of the unique features of a particular mobile device. Mobile app development is fleetly growing from retail, telecommunications, healthcare, and government associations across diligence. Educational Institutes give systems for a better understanding of the practical knowledge towards the subject in the real world, which involves a lot of tasks like abstract, thesis correction, and streamlining the proposed module with assigned administrators. Scholars elect a design in a given field of interest. Generally, a design is suitable for further than one pupil. Scholars' systems can efficiently be allocated to each or group of scholars without the problem of delayed design allocation from the companion or conflict of content between two individualities or groups of scholars in the same department.

Scholar's updates can be fluently penetrated if the database system is enhanced. Pupil design allocation is more secure, movable, fluently installed, and used on any mobile supporting android Zilches. Provides an interface that's easy to understand by the druggies and dramatically helps acclimatize to this operation's use. This application consists of Student login, staff login, and subject coordinator login and contains the details of the student project allocated, giving users an interactive experience. It will be secure that the data can be used by only those who have a username and password.

II. EXISTING SYSTEM

Vial, Peter James et al. proposed the software to help the subject coordinator quickly allocate students into optimal or suboptimal teams based on predetermined criteria. The actual use of team allocation software shows that it can find, in a short time, the solutions highly compliant to the team allocation criteria selected using a simple algorithm. The software reduces the time needed for allocation. On estimation, the software saves at least 8 hours of the coordinator's time. The disadvantage of this work is that the software must be installed to allocate and carry the system wherever we go.[1] Hasan, M. H. et al. proposed a system that uses an algorithm to allocate titles to students based on students' submissions of their ten most preferred titles in order of preference. Three algorithms were developed for this allocation, and they were evaluated by calculating and comparing the costs of their selection. The algorithm assured that each pupil got one of their ten preferred titles, and utmost of the scholars got largely ranked titles. The disadvantage of this system is several perpetration issues also need to be cleared out, particularly the ideal number of titles to be ranked by every pupil during the selection process. [4]. the deduplicate data conserve disc space. An essential notion among them is source-based Deduplication. [26].

Amadi et al. proposed a software system for students' projects allocation system. It will enable final year students to apply for projects and upload completed projects. Model View Controller (MVC) methodology used in designing the system. MySQL is used for the Backend, Used PHP for the front end. It develops a project allocation system that successfully allows students to apply for projects online. The system allocates a project title to the student based on the supervisor's highest score on the student's proposals. The disadvantage is they need to carry a system or laptop everywhere Animesh Tayal et al. work on SPAM (Student Project Allocation and Management) is a Web Application that automates the whole project management system through which we can view the documentation related to the project tasks. This system has overcome all the traditional processes of manually submitting the project abstracts, synopsis, or other documents. Also, it provides a platform where the guide can allot tasks to their respective group, and students can choose his group and choose his project guide. The project guide can allot Project-related tasks, and other faculties can only give reviews over it. Students can directly upload their proposed work and the documentation on this system to evaluate it. [24]

A. Adamu proposed a system developed using PHP ASP.NET to develop GUI and XAMPP for the database. The system is designed to run on Windows operating systems. The system can be used in any higher institution to replace the manual supervising final year students. It will reduce the challenges, energy, and time required to monitor and manage final-year student projects. The system can be used in any higher institution to replace the manual supervising final year students. It will reduce the challenges, energy, and time required to monitor and manage final-year student projects [2]. Srinivasan et al. proposed a student project allocation system as an effective fuzzy evolutionary algorithm is used for working the pupil design allocation problem. It presents a result frame for the pupil design allocation (Gym) problem grounded on evolutionary algorithms (EAs). Project is an engineering course that's conducted in universities. A list of projects must be selected in their final year of study. The EA-based project allocation system was implemented in a large university department to automate this process and enhance scholars' matching to their desired projects. [23].

Cheung, Yeung, et al. Proposed an "A dynamic project allocation algorithm for a distributed expert system. The manual processing sequence was very time-consuming and inconvenient to the parties involved in this system. A student had to do a lot of manual searches to find projects that were difficult to modify after submission. It was tough to assign FYPs to students manually. Each student will have a unique number used in the allocating process. Thus, a project best matches the student's personal preference with their ranking. [7]. Dimitar Kazakov et al. proposed " Collaboration of Student Project Allocation," The system not only performs design allocation but also allows academics to rate systems, observers to examine, scholars to propose their systems, scholars to submit the design, administrators to follow systems more nearly and allows systems fellow to have a view of the whole system. The system captures the preferences of observers as well as scholars. It allocates systems to them to maximize the number of scholars who get their choice in their preference list and keep administrators and observers' cargo within a reasonable range. The chance of scholars who attained their first choice is 82 on 30 systems proposed by 15 administrators for 11 brigades [9]. We could apply a location-based system to the advancement of numerous sciences, businesses, and vocations, among other things, to increase individual human needs. Markets, and even their components and placement, could be the causes of rescuing economically distressed countries as we analyze location-based system discussions. [28].

Tsvetelina Mladenova works on "A design operation system for time planning and coffers allocation" When it comes to the effective operation and planning of a company's design coffers and workload, numerous ultramodern businesses face a severe problem, which can be the cause for their failure. Lack of planning will affect dearth, reduced effectiveness, and detainments. A web-grounded result is proposed and is viewed through the prism of the whole business process and not just as an independent unit. However, the presented operation can be either a standalone system or a module of an ERP system, with the design operation being an operative module. Experimentally the design operation system is enforced in the surveyed company replacing a being PMS systems, the results of a fresh check are presented and anatomized [19]. H.Y. Chiang and B.M.T. Lin proposed a Decision Model for Human Resource Allocation in Project Management of Software Development," mortal resource allocation is critical not only for design success, including timely delivery and product quality, but also for cost estimation to a software company to decide to contract systems or not. Considering both the cost and the effectiveness, the performance is maximized to verge on the guests' prospects in software development design operation grounded on a real case. Before forming a platoon for a named design with well- defined ages and individualities with different places, the values of critical factors within the frame are calculated with an individual foundation of calculation. The proposed model helps the choice- making process of software companies for platoon conformation [8]. The security features of protocols suggested in LTE and LTE-A, such as EC-AKA2 and GR-AKA, were evaluated and confirmed using the ProVerif program in this study. Because of the encryption technique employed in GR-AKA, the results show that the protocol has better security qualities than the EC-AKA2 protocol. However, both protocols have achieved the requisite levels of security. The security analysis can also be expanded to include other protocols. [27].

P. J. Vial et al., work on "A Java Program for Automatic Team Allocation in Project-Based work," Universities around the world need project-based subjects and effective team allocation. Team allocation could be a long task in complicated project-based work subjects. This offers rise to the necessity of an automatic team allocation software system, which might facilitate the topic organizer quickly assigning students into optimum or sub-optimal groups supporting a collection of planned criteria. This paper details our team allocation software system developed in Java in 2012 for project-based engineering and was indeed enforced in 2 annual project-based subjects within the University of Wollongong within the years 2013 through 2017. The particular use of our

developed team allocation software system shows that this software system is in a position to search out. During an exceedingly short time, the solutions extremely compliant to the team allocation criteria elect to employ a straightforward algorithmic rule. Compared to manual allocation, our developed software system considerably reduces the time needed to create student groups [25]. V. Arumugam et al. proposed "Academic Project Information Management System," Academic projects are essential for any Engineering undergraduate course. It gives students a chance to illustrate all they have learned. This paper describes an automated system for conducting the final time systems process. Presently, numerous sodalities manage final time design- related data in an offline manner spreadsheet entries of all the groups, homemade group conformation, and administrator assignment, maintaining a hard dupe of the documents submitted by the scholars. The paper presents how a web- grounded automated system will amend all the issues and crimes being while maintaining them offline. The main ideal is to propose a system for managing groups, automated companion allocation, document sharing, smoothen the process of communication between attendants and scholars, maintaining a log of all the conditioning, and supervising scholars' design progress [3]. Hussain, S, et al. proposed “A Methodical Review of Project Allocation Styles in Undergraduate International Engineering Education". The final time design is one of the most critical factors of any undergraduate engineering program. Fair and effective design allocation procedures can be vital in icing a great pupil experience and exceptional literacy out of these systems, which could also shape scholars' unborn prospects. This paper reviews design allocation strategies used in colorful universities at undergraduate situations. We also concentrate on the design allocations in international education (TNE) surrounds, which inherit fresh allocation challenges. We punctuate these challenges and give recommendations to break them. We present and compare design allocation strategies espoused at two of China's most extensive TNE programs. We also present the factors impacting the design allocations, particularly TNE vittles. Eventually, we describe the challenges associated with the design allocations in the TNE script, along with proposing some possible results to address these challenges [14].

III. PROPOSED SYSTEM

We developed an application that can be used for project allocation and maintenance. In this student login to the system, they can enter project details and mention their team members. And it enables the student to view their marks for their projects, and they can also view the announcements given by the coordinator. Updation is also enabled to update their project details and team members. In the staff login, the staff can view the details of the students' project, review panel, and dates and enter the marks while reviewing their projects. And in the subject coordinator login, they can assign a guide to the team, enter announcements for review and view the marks details of students and the details of team members and the respective guide staff for the following student projects. This system aims to overcome the time required to search the student’s project details and marks. And they need not depend upon any other platform to manage the student’s project details and marks. First of all, the admin enters the student list and staff and gives them a Login password to enter the application and assign a coordinator. The student should enter the project details, such as the project title and domain. The coordinator will allot a guide for teams and prepare the review schedule.

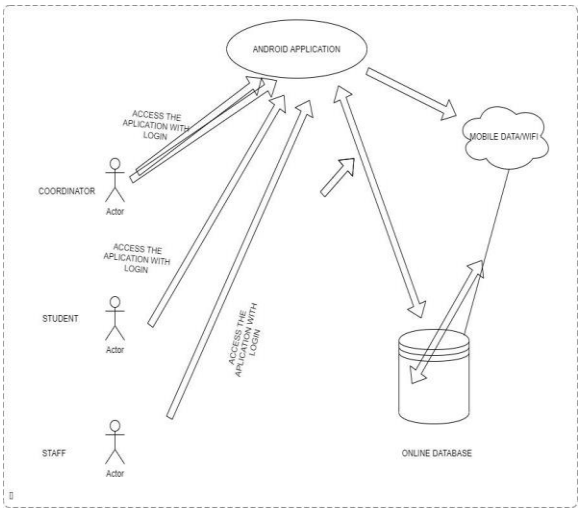


Figure 1. Architecture Diagram

The architecture diagram explains the accessibility of the application along with databases over the internet. Information can be fed and can also be retrieved from databases.

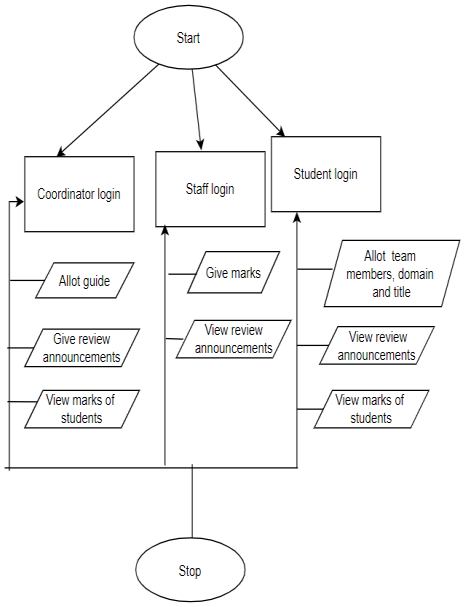


Figure 2. Dataflow Diagram

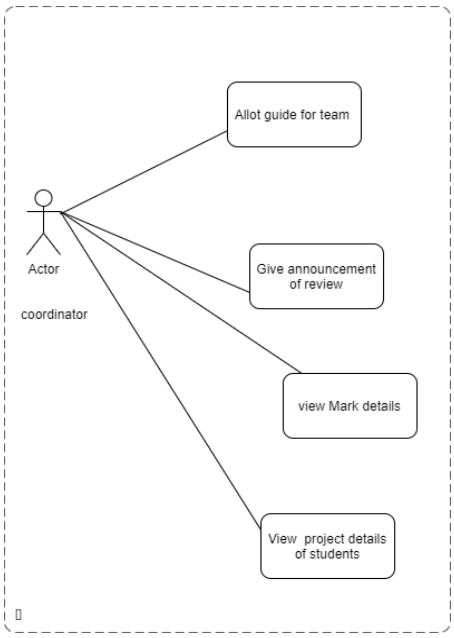


Figure 3. Fig Coordinator Module

The diagram explains all the four modules present in our system: coordinator login, staff login, student login. The coordinator will make the team with the leader and allot a guide for the students and give announcements for the review like the review date and the panel members. And the coordinator can view the marks given by the staff members for the students. In the staff login, the staff can accept or reject the team's title under their guidance, can also give marks for the particular panel students, and the staff can view the announcements given by the coordinator. And the student login, the student can enter the project details like the title of the project and the domain of their project and request a guide for the approval of the title, and the team leader can enter the team members of their project. The student can view the announcements given by the coordinator, and they can view the marks for the review given by the Reviewers. These are the modules and the features included in our project so that the students, staff, and coordinators can easily access the details of the project through this application. Fig 3 Depicts coordinator allot guide to the team of students and make review announcement of allotting review panel for the teams for reviewing their project work. The coordinator can view details of marks and view the project details of the students. Fig 4 describes that staff can view the announcement given by the coordinator. Accordingly, staff review the team about their project work and enter marks for the team. Staff can also view the project details of the students.

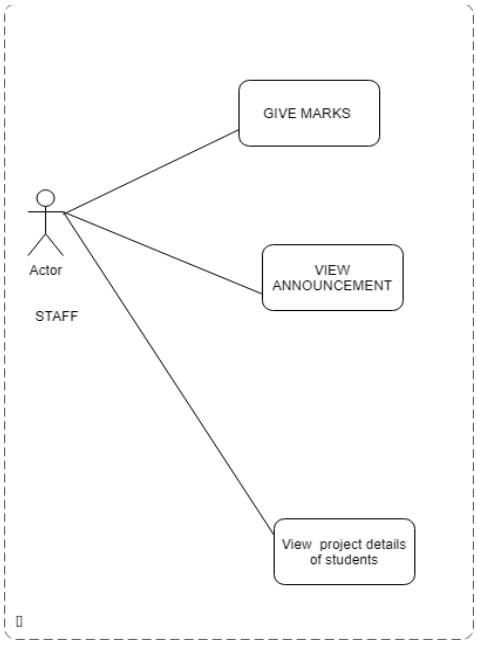


Figure 4. Staff Module

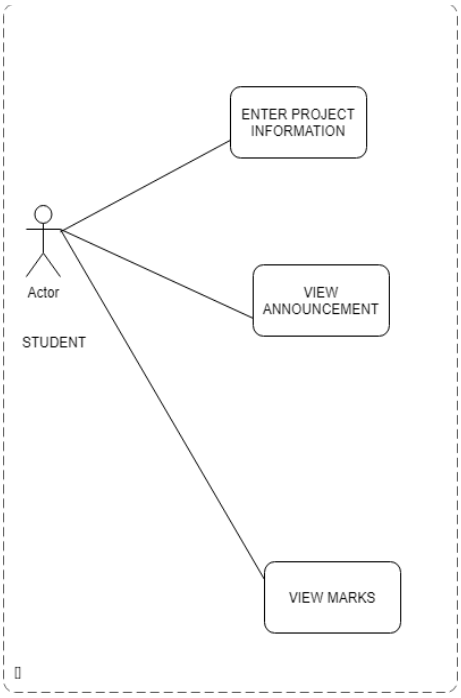


Figure 5. Student Module

Fig 5 depicts that students can enter the project details like the title of the project and the domain of their project, and the team leader can enter the team members of their project. The student can view the announcements given by the coordinator, and they can view the marks for the review given by the Reviewers.

IV. RESULTS

The proposed framework aims to address the previously mentioned concerns. The suggested framework's primary goal is to provide data in a flash as and when needed. The objective is to improve the efficiency and power of the Hostel data points of interest. This system should keep track of numerous data records so that information may be recovered quickly and effectively. The structure is exceptionally well-thought-out. It should ensure that processes are coordinated to the desired level, and different reports should be generated as needed. This framework should also ensure that the captured data does not repeat itself. Table 1 shows the testing method and its results. This software development research project was a success where the software developed may be deployed online for improved management of student project allocation after supervisors have been assigned.

TABLE I. TESTING RESULTS

Module	Unit Testing	Integration Testing	System Testing	Acceptance Testing
Login Page	Successfully executed and verified	Successfully executed	All the types of system testing was executed and verified	The application was accepted by the end user
Student Login		Integrated with login page and executed successfully		
Team Member Creation		Integrated with student login page and executed successfully		
Title and domain choosing		Integrated with Team Member Creation and executed successfully		
Viewing Announcements		Integrated with Title and domain choosing and executed successfully		
Viewing Marks		Integrated with Viewing Announcements and executed successfully		
View Panel		Integrated with Viewing Marks and executed successfully		
Assigning Marks		Integrated with View Panel and executed successfully		
Assigning the panel members for review		Integrated with Assigning Marks and executed successfully		

Students can propose subjects and have their supervisors review and assess them, while the system assigns the student to the topic with the most significant point total. The student uploads the completed.pdf file at the end of the project. A supervisor can also offer students research subjects via the supervisor website, which they can use in their proposals. If this software solution

is used, it will help students and supervisors communicate more effectively, resulting in better service delivery.

V. CONCLUSION

Our application assists in automating the existing manual system. It reduces the workforce required. Allows students, Coordinators, and staff easy and prompt access to data. Allocation of the project is essential in institutions for that many steps must be followed. Difficulties emerge due to the increasing number of students, increasing the complexity of allocating the projects. This project aims to save time by Allocating Guide and reducing the paperwork. The topic of the projects is too visible in this application to avoid the repetition of the same project by some other teams. This application is so secure that only the people responsible can allocate and view those data. And this application provides a well-organized platform to maintain all the project details.

REFERENCES

[1] Abdi, M. Reza, and Houssam A. Kaddoura. "Projects Management Office: a case study for best practices." In 2011 International Conference on Management and Service Science, pp. 1-5.IEEE, 2011.

[2] Adamu, A. "Final Year Student Project Allocation Archiving and Management System."

[3] Arumugam, Vidya, Priyanka Singh, KajalPadhiyar, RiyaManek, and ShahziaSayyad. "Academic Project Information Management System."In 2021 Asian Conference on Innovation in Technology (ASIANCON), pp. 1-6.IEEE, 2021.

[4] Bibi, Nazia, Zeeshan Anwar, and Ali Ahsan. "Comparison of search-based software engineering algorithms for resource allocation optimization." Journal of Intelligent Systems 25, no. 4 (2016): 629-642.

[5] Bushuyeva, Natalia, Denis Bushuiev, VictoriiaBusuieva, and Igor Achkasov. "IT projects management driving by competence." In 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), vol. 2, pp. 226-229.IEEE, 2018.

[6] Carvalho, Marly M. "Communication issues in project management." In PICMET'08-2008 Portland International Conference on Management of Engineering & Technology, pp. 1280-1284.IEEE, 2008.

[7] Cheung, Yeung, GeckMeng Hong, and KokKeng Ang. "A dynamic project allocation algorithm for a distributed expert system." Expert Systems with Applications 26, no. 2 (2004): 225-232.

[8] Chiang, Hui Yi, and Bertrand MT Lin. "A decision model for human resource allocation in project management of software development." IEEE Access 8 (2020): 38073-38081

[9] De Oliveira, Elaine Cristina Batista, Luciana HazinAlencar, and A. P. C. S. Costa. "An Integrated Model for Classifying Projects and Project Managers and Project Allocation: A Portfolio Management Approach." International Journal of Industrial Engineering 22, no. 3 (2015): 330-342.

[10] El-Atta, Ahmed H. Abu, and Mahmoud Ibrahim Moussa. "Student project allocation with preference lists over (student, project) pairs." In 2009 Second International Conference on Computer and Electrical Engineering, vol. 1, pp. 375-379. IEEE, 2009.

[11] Gadge, Sonali, RuchiPahire, SnehaSuryawanshi, ShreyashChawhan, and AnimeshTayal. "Review on Student's Project Management System for Faculty of Engineering & Technology."

[12] Glisson, William Bradley, and Gobinda G. Chowdhury. "Design of a digital dissertation information management system." Program (2002).

[13] Gu, Mengyao, JuanerZheng, PengfeiHou, and Zhixi Dai. "Task allocation for product development projects based on the knowledge interest." In 2019 6th International Conference on Information Science and Control Engineering (ICISCE), pp. 600-604. IEEE, 2019.

[14] Hussain, Sajjad, Kelum AA Gamage, MdHasanuzzamanSagor, Faisal Tariq, Ling Ma, and Muhammad Ali Imran. "A systematic review of project allocation methods in undergraduate transnational engineering education." Education Sciences 9, no. 4 (2019): 258.

[15] Hussain, Fehmida, TeniaKyriazi, and Lynda Hyland. "Undergraduate students' attitudes towards research: lessons from an international branch campus in the UAE." International Journal of Teaching and Case Studies 9, no. 4 (2018): 382-395.

[16] Kudenko, Daniel, DimitarKazakov, and Eduardo Alonso. "Machine Learning for Agents and Multi-Agent Systems."In Intelligent Agent Software Engineering, pp. 1-26.IGI Global, 2003.

[17] Kumar, Ashok, and L. S. Ganesh. "Use of Petri nets for resource allocation in projects." IEEE Transactions on Engineering Management 45, no. 1 (1998): 49-56.

[18] Meade, Laura M., and Adrien Presley. "R&D project selection using the analytic network process." IEEE transactions on engineering management 49, no. 1 (2002): 59-66.

[19] Mladenova, Tsvetelina. "A project management system for time planning and resources allocation." In 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 1299-1303. IEEE, 2019.

[20] Paunovic, V., S. Tomic, I. Bosnic, and M. Zagar. "Fuzzy Approach to Student-Project Allocation (SPA) Problem." IEEE Access 7 (2019): 136046-136061.

[21] Scott, Carlton H., and Thomas R. Jefferson. "Allocation of resources in project management." International Journal of Systems Science 26, no. 2 (1995): 413-420.

[22] Siwani, Imran, and Miriam Capretz. "'fuzzy Project Manager'— Framework For Software Project Management Using Fuzzy Logic." International Journal of Innovation and Technology Management 1, no. 04 (2004): 435-453.

[23] Srinivasan, Dipti, and Lily Rachmawati. "Efficient fuzzy evolutionary algorithm-based approach for

solving the student project allocation problem." IEEE Transactions on Education 51, no. 4 (2008): 439-447.

[24] SonaliGadge, AnimeshTayal, ShreyashChawhan, SnehaSuryawanshi, RuchiPahire, AbhishekWakode, "Student's Project Management System for Faculty Of Engineering and Technology", International Journal Of Trend in Research and Development, Vol. 4, Issue 01, 2017

[25] Vial, Peter James, Timothy James Russell, David Stirling, Montserrat Ros, PrashanPremaratne, and Sasha Nikolic. "A Java Program for Automatic Team Allocation in Project-Based Coursework." In 2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), pp. 185-192. IEEE, 2018.

[26] R. N. Karthika; C. Valliyammai ; D. Abisha, "Data Deduplication and Fine-Grained Auditing on Big Data in Cloud Storage", Advances in Machine Learning and Data Science, Springer Nature Singapore, 2017.

[27] Abisha D, "Perlustration on Authentication Protocols in 4G (LTE/LTE-A) Using Pro-Verif", Journal of Network Security Computer Networks, Vol.3, Issue 3, 2017.

[28] Priyadharshini A, Gunaseelan K, Karthik S, Abisha D, "Perlustration on Techno Level Classification of Location Based Services", International Journal of Advance Research, Ideas and Innovations in Technology, Volume 4, Issue 1, 2018.

Research on Fatigue Classification of Flight Simulation Training

Lien Wang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: wanglien@st.xatu.edu.cn
*corresponding author

Changyuan Wang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: Cyw901@163.com

***Abstract*—Fatigue is an important factor affecting modern flight safety. It can easily lead to a decline in the pilot's operational capabilities, misjudgments and flight illusions. Moreover, it may even cause serious flight accidents. In this paper, a wearable wireless physiological device is used to obtain pilot electrocardiogram data in simulated flight experiments. Bioelectric signals have higher reliability than image information, and are not easily affected by the external environment (such as shooting angle and light intensity). On the other hand, neural networks have been widely used in various classification and regression tasks. In this study, the EEG was collected in the driving flight simulator, and after simple filtering and preprocessing, the time domain data was sent directly to the convolutional neural network, eliminating the need for additional feature extraction operations. We found that the convolutional neural network can effectively amplify the fluctuation details of the time domain data and train the pilot fatigue state recognition model. The results show that the recognition accuracy of the convolutional neural network model reaches 98%, which is 26% and 12% higher than the traditional k-nearest neighbor classification algorithm (KNN) and support vector machine (SVM) model, respectively. The recognition model based on convolutional neural network established in this paper is suitable for the recognition of pilot fatigue status. This has important practical significance for reducing flight accidents caused by pilot fatigue, and provides a theoretical basis for pilot fatigue risk management and the development of intelligent aircraft autopilot systems.**

Keywords-Fatigue Classification; Eeg; Cnn; Neural Network

I. INTRODUCTION

International Civil Aviation Organization (ICAO) statistics on global planned commercial flight accidents and casualties in the past ten years show that from 2013 to 2017, the annual number of flight accidents has not changed much, but the number of casualties has been there have been large fluctuations, and the overall number of casualties remains high. In addition, once an accident occurs, the direct loss caused is the production and manufacturing cost of an aircraft (the average value of Boeing's aircraft supplies is approximately $90 million), plus compensation for accident losses. Although the flight accident rate has been on a downward trend in recent decades, the injuries and losses caused by airplane accidents have not changed much. At the same time, every flight accident will cause fatal injuries to every family and indirectly cause national losses. Therefore, aviation safety issues need to be treated strictly for countries in all regions of the world.

According to the flight accident statistics of the Federal Aviation Administration (FAA) and NASA, only 12% of flight accidents are caused by problems with the aircraft itself, and more than 73% of accidents involve human factors. 67% of aircraft accidents caused by mistakes of the aircraft, the most important factor is the pilot's operation error, which accounts for about 51% of the total number of air crashes. The main reason for the pilot's operation error is that the driver is in a state of fatigue and his driving is alert. Degree drops. According to the statistics of road traffic accidents in my country, 90% of traffic accidents are caused by the driver's human factors, which are mainly due to the driver's dangerous

driving state, such as distracted driving, fatigue driving, etc.

The Current methods for detecting pilot alertness go in three directions, namely aircraft-based behavioural monitoring, pilot behavioural recording and pilot physiological signal measurement. Of these, the first two methods are more influenced by the external environment such as the aircraft model and driving environment, while the latter method depends only on the subject conditions; therefore, it shows a higher ability to detect driver drowsiness. Measurements of physiological signals include neuronal electrical activity from electroencephalography (EEG), eye movements from electrooculography (EOG), heart rate from electrocardiography (ECG), muscle activity from electromyography (EMG) and tissue oxygenation from near infrared spectroscopy (NIRS) [1]. However, of these signals, the EEG signal, which is less likely to be influenced by individual characteristics, has been widely used as the 'gold standard' for fatigue detection and has proved to be a promising method for studying drowsiness and changes in driver alertness.

EEG is the overall response of brain nerve cell electrophysiological activity on the scalp surface or cerebral cortex [2]. According to its frequency, it can be divided into 5 different bands: (1) δ wave (1~4 Hz), which generally only appears when adults are asleep; (2) θ wave (4~8 Hz), which mainly occurs in sleep State; (3) α wave (8~14 Hz), which generally occurs in a relaxed state; (4) β wave (14~30 Hz), the increase in the power spectrum of the β wave is closely related to the increase in alertness. (5) Gamma wave (30-49 Hz).

As early as the 1980s, studies on the correlation between EEG and brain fatigue have been carried out abroad. Relevant studies have shown that EEG is very sensitive to fluctuations in vigilance. EEG will change significantly with changes in vigilance. EEG It can predict the decline of brain performance caused by continuous mental work [3]; in the 1990s, EEG research was further deepened, and people began to pay attention to the changes in various bands of EEG during brain fatigue. Research found the correlation between reduced human alertness and fatigue It will be concretely reflected in different wave bands of EEG, among which the changes of theta wave and alpha wave are particularly obvious [4], which is specifically manifested in the power spectrum of theta wave and alpha wave when people are in a state of fatigue; enter 21 In the century, the research on the different bands of EEG in the state of brain fatigue is more detailed, the θ, α and β frequency bands in EEG and the combined parameters of different frequency bands. For example, Jap et al. [5] conducted a more comprehensive experiment. The four EEG bands of δ, θ, α, β and (θ+α)/β, α/β, (θ+ The evaluation of the four parameters of α)/(α+β) and θ/β showed that during the transition from non-fatigue state to fatigue state, the activities of δ wave and θ wave were relatively stable, and the activity of α wave decreased slightly. The β wave activity is significantly reduced; the values of the four combination parameters have increased, among which the increase in (θ+α)/β is more obvious, and the value of (θ+α)/β is also more obvious under different fatigue levels the difference. By combining EEG signals of different bands for analysis, and then proposing suitable combination parameters, not only can the respective characteristics of different bands of EEG signals be fully utilized, but also the detection results can be more accurate and comprehensive through the combination of parameters.

Many domestic scholars have also conducted research on the alertness of drivers by EEG. In addition to the above four combined EEG features, there are also time-domain analysis and frequency-domain analysis of brain waves, power spectral density (PSD), and entropy. Carry out the analysis and evaluation of the driver's alertness from an equal angle. EEG is also an important parameter in human factors engineering. EEG's brain fatigue detection is also widely used in human factors engineering. For example, using the EEG fatigue detection method to evaluate and guide the professional training and psychological adjustment of pilots, it was found that the level of positive emotions of pilots has been improved. Similarly, the brain fatigue detection using EEG can also study the effect of brain fatigue on

selective visual attention, and the results show that mental fatigue has a negative impact on the ability of selective visual attention.

At present, the use of EEG can be used to determine the degree of brain fatigue and brain fatigue. With the development of high-throughput EEG technology and the intelligence of EEG data analysis, traceability analysis of high-throughput EEG data is expected to be used for brain fatigue location analysis; on the other hand, the fusion of synchronous EEG and functional magnetic resonance imaging technology also provides technical support for brain fatigue location analysis. In short, EEG-based brain fatigue detection will develop in the direction of quantitative, precise, and accurate positioning, and the fatigue detection ability and credibility of drivers will also continue to improve.

II. MATERIALS

A. Experimental design

There is a certain gap between the difficulty and the degree of danger in simulated flight driving and real driving. In order to reduce the impact of new and veterans on driving task control ability and ensure the objectivity of test data, 20 graduate students are selected, aged 24-28 years old (Average age 26.8 years), healthy, all right-handed; no driving experience; no drugs taken during the test; no alcoholic foods were consumed 24 hours before the test, and no caffeine-containing beverages were consumed 12 hours before the test. Did not eat and exercise vigorously in the first 1 hour; in order to avoid the impact of the test period, the test was completed within a similar period of the human body's physiological cycle; 1 day before the test, the subjects were trained on the driving simulator operation for 20 minutes, and try to ensure that the samples are correct The sameness and equality of driving proficiency. Participants were informed of the specific content of the experiment to ensure that they fully understand how the physiological data collected during the study will be used. Every participant is willing to participate in the experiment; the experimental data must eliminate personal identification information, and only retain data that has a specific impact on the experiment. The research is carried out at noon (12:00-14:00), and the indoor temperature is controlled at 25±2℃ [6], the indoor humidity is 45±10% [7]. All participants were required to perform moderate mental work within 5 hours before the experiment to reduce nerve excitability. They are not allowed to participate in any form of physical labor to prevent changes in blood pressure and heart rate. Each participant had 8-9 hours of sleep before the experiment. The observer is set to record the state of the experimenter, including whether there are red blood streaks in the eyes and changes in blinking frequency [8]. The neural network model uses the Tensorflow GPU 2.4.0 framework, the CUDA version is 11.0, and the cuDNN version is 8.0. We use 4 NVIDIA Titan V graphics cards to accelerate the training process.

B. Experiment procedure

The experimental stimulus is presented on a 31.5-inch desktop curved display. The interactive objects on the screen are the "Microsoft Flight Simulator X: Steam Edition" game published by Dovetail Games and developed by aviation expert Jane Whittaker. This game allows non-professional players to feel the pilot's nervousness when encountering an emergency. The graphics and the degree of realism have reached the peak, and the various elements encountered in real-world flight, such as aerodynamics, weather, and geography The environment, flight control system, flight electronic system, combat flight weapon system, ground flight guidance, etc., are comprehensively simulated in the computer, and the flight simulation control and flight sensory feedback through external hardware equipment are used to feedback the fatigue expression of the pilot in the previous year. On the basis of, eye movement, line of sight, etc., complete the EEG acquisition and analysis of pilot training subjects (takeoff, landing), identify the emotional characteristics of pilots during training, complete multi-dimensional channel data fusion, and build pilot control response The mathematical model of time and attention distribution monitors and evaluates the effect of flight training and conducts control experiments. Before entering the simulated

control environment, the subject should wear the EEG device to ensure that the device is connected to the software to measure resistance (the Ergo software displays the port is green), and the Ergo software will display the connection status of the port. There are four types: 1. Green: The port is connected normally and the signal is stable 2. Orange: The port is connected normally and the signal is unstable 3. Red: The port is connected normally and the signal is weak 4. Gray: No signal at the port before the experiment, it is necessary to ensure that the device and the software are properly connected to measure the resistance (all ports are green), and blink and close the eyes to confirm that the device is receiving the eyeballs normally. After the test subjects enter the simulated flight environment, they need to continuously interact with the experimental materials and retrograde within half an hour to complete the take-off and landing of the fixed-wing aircraft, fly around the field five times, and fly on the designated route. In the formal test process, the subjects need to enter the simulated maneuvering scene (as shown in Figure 1), and the flight simulation platform simulates the experimental task of driving.

Figure 1. Simulated driving environment

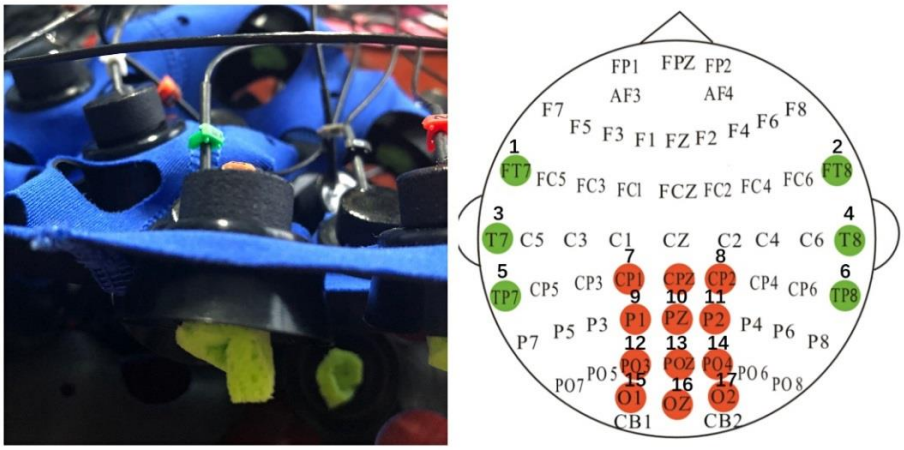


Figure 2. EEG acquisition equipment

During the execution of the task, random pop-up windows will appear within the subject's field of vision, and the subject needs to move the point of sight to the pop-up window to close the window. The time interval from when the pop-up window appears to when it closes is the reaction time. Based on the response time, the research team developed a mental label (fatigue/mental) for the EEG signals within 10s before the pop-up window appeared. One set of experiments is set for half an hour. Because the EEG cap will affect the subjects, too long time will increase the subject's eye height and cause extreme discomfort. Therefore, this topic divides a set of experiments into two Second, the specific timetable is shown in Table 1.

TABLE I. EXPERIMENT SCHEDULE

Participant ID	**001**	**002**	**003**
Experiment 101	9:00-9:30	10:00-10:30	11:00-11:30
Experiment 102	13:30-14:00	14:30-15:00	15:30-16:00

During the completion of the task, restrict the subjects' body and head to more vigorous movements. Moreover, before the start of the experiment, the recorder will tell the subjects that the eye-closing behavior should be spontaneous rather than deliberate, such as allowing eyes to be closed when they feel sleepy or uncomfortable. Since there is no actual threat to personal safety, subjects will go along with it when they are fatigued. These situations are difficult to obtain in a real manipulation environment. This helps us to investigate the characteristics of changes in physiological signals when the human body is fatigued.

III. TECHNICAL ROUTE

Carlo Matteucci et al [13]. First obtained the muscle nerve electrical signal with a galvanometer in 1881 and established the concept of neurophysiology. In the following nearly a hundred years, people gradually clarified the collection methods and standards of bioelectric signals. The non-invasive collection methods that have been widely used in the field of EEG signals include EEG, magnetic resonance imaging, near-infrared spectroscopy, and there are four kinds of magnetic EEG [14]. Among them, the multi-channel electrode EEG method, which integrates high time resolution, low cost, and non-invasive safety, is the most widely used. Thanks to the continuous advancement of technology, the 10-20 standard lead, 10-10 standard lead and 10-5 standard lead established by the American Society of Clinical Neurophysiology are the most common in clinical trials [15]. The three standard system guides are extensions of each other and keep the same overlap in the naming rules. This simplifies the EEG signal research process and reduces the difficulty of technical communication. It also clears the obstacles for the electrode naming rules for this study. Specific electrode naming and the spatial coordinates can be queried on the website of the American Society of Clinical Neurophysiology, and will not be listed in detail here. In this project, the electrode position recommended by the 10-10 standard lead system will be used as a benchmark to start the experiment.

The EEG signal is a sign of neural activity. The neural activity generated by any part of the human body will be more or less reflected in the EEG signal [16], and the research needs to focus on the "event potential", that is, the human brain because of a certain One or some physiological electrical signals generated by certain activities, so choosing a suitable reference electrode according to different research focuses will greatly reduce the research workload. The research of Lei Xu et al. [17, 18] showed that the reference electrode is the key to the study of EEG and event-related potentials. Since Yao et al. proposed the reference electrode standardization technology in 2001, REST has quickly become the first choice for most EEG research models. In addition, in some cases, the full electrode average can also be considered to have the same effect as REST. Essl et al. suggested choosing FCz electrode as a reference electrode when the results of the study are not clear. The consistency based on FCz electrode is higher than the consistency based on non-cranial reference, and their research has become part of REST. Yao Dezhong et al. studied the influence of different reference electrodes on

spectrum mapping. REST technology aims to build a bridge between traditional reference electrodes (such as scalp or average reference) and theoretical zero reference. The reference point at infinity has a theoretical neutral potential and is regarded as an approximate zero potential point in REST.

EEG signal is a 5-100μV low-frequency bioelectric signal, which needs to be amplified before it can be displayed and processed. An important operation in signal processing and interpretation is filtering. The main function of filtering is to remove interference signals from EEG signal data. Especially for high-frequency signals caused by the external environment, the filtering used in EEG signal processing is mainly divided into high-pass filtering, low-pass filtering and notch filtering. The filtered data can be analyzed for characteristics. There are two main characteristics of EEG data: spatial characteristics and time-frequency characteristics. High-pass filtering aims to pass signals above a limited frequency without attenuation, while blocking and attenuating signals below the limited frequency. Since the EEG is a signal of about 30 Hz, and the frequency above 50 Hz is only involved in the medical diagnosis of epilepsy and human brain physiology, high-pass filtering is rarely used in EEG signals, but some scholars choose to do it. The high-pass filter of 0.1~0.7Hz is designed to remove frequency components with extremely low interference such as breathing. If there is a problem of baseline drift in the signal, Alste et al. conducted a study on ECG and suggested using high-pass filtering to deal with such problems. Although high-pass filtering may be one of the means to solve the baseline drift, Acunzo et al. found that high-pass filtering can cause early ERP and ERF system deviations. High-pass filtering is used cautiously when dealing with the model of EEG and ERP signal fusion.

Low-pass filtering allows signals with frequencies below a certain range to pass, and signals above the critical frequency are blocked and attenuated. Because the EEG signal acquisition instrument is sensitive to weak electrical signals, and the frequency of the mains power in my country is around 50Hz, although there is still a certain frequency space below 30Hz from the target, low-pass filtering must be performed to reduce the impact of the mains signal on the data. This operation is also one of the normal operations in EEG signal processing.

McFarland et al. in 1997 proposed a spatial filter selection based on EEG, by selecting different filters to process the signal to obtain a clear EEG signal. When studying EEG signals, Higashi et al. found that the spatial filter based on the common space pattern method of electrode weights is very effective in the classification of EEG signals based on moving images, but the existing methods have certain limitations. For this reason, a discriminative filter bank is proposed to extract the bands related to the brain activity of moving images.

The time-frequency domain method is also a common EEG signal research method. Hjorh, Salinsky and Valdes discussed the reliability of EEG frequency domain analysis. An important step in time-frequency analysis is to convert time-domain signals into frequency-domain signals. If the signal is statistically stable, or there is a fixed law, then the finite length signal can be transformed into a frequency form by using a linear transformation.

IV. METHODS

EEG data analysis involves a variety of signal processing techniques, including but not limited to signal acquisition, preprocessing, and feature extraction. There are also a variety of methods that are widely used in data classification, such as KNN based on sample feature distance and VC theory. Linear SVM and models based on convolutional neural networks. This chapter will respectively introduce the key technologies of the above three fields involved in this research.

EEG signal processing is mainly composed of signal acquisition, conversion reference, filtering, artifact removal, segmentation, independent component analysis and other operations. In addition, you can also choose whether to downsample the collected data according to the actual situation, but you need to pay attention if it is down-sampling, it may be necessary to perform

linear or non-linear interpolation to complement the disappeared features. This article mainly extracts the four features of EEG δ, θ, α, β for fatigue classification.

A. K-Nearest Neighbor

The K-Nearest Neighbor (KNN) algorithm is one of the commonly used classification algorithms. When there is little or no prior knowledge of the data distribution, KNN should be the preferred method. Cover et al [9]. Made it clear that the upper limit of the classification error of the KNN algorithm is twice that of the Bayesian classification error. The algorithm aims to calculate the feature distance between the unknown sample and the known sample group, and infer the category of the unknown sample based on the distance. Common distances include Euclidean Distance, Minkowski Distance, Manhattan Distance, Chebyshev Distance, etc.

Euclidean distance is the most common measurement method, which measures the absolute distance between points in a multidimensional space, and is defined as follows.

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (1)$$

Since the Euclidean distance is calculated based on the absolute value of the features of each dimension, the premise of using the Euclidean distance is to ensure that the dimensions of the indicators have the same dimension. Different dimensions may cause the Euclidean distance to be invalid. The Mind distance is the Euclidean distance. The generalization of, the current p=2 in the following formula is the Euclidean distance.

$$d(x,y) = \left(\sum_{i=1}^{n}|x_i - y_i|^p\right)^{1/p} \qquad (2)$$

The Manhattan distance is derived from the city block distance. The result of summing the distances in multiple dimensions, that is, when p=1 in the formula, the Manhattan distance is obtained.

$$d(x,y) = \sum_{i=1}^{n}|x_i - y_i| \qquad (3)$$

In addition, feature conversion can also improve the accuracy of the model to a certain extent [10]. Commonly used feature transformations include standardization and fuzzification. Standardization eliminates the influence of different scales in the same dimension state, and fuzzification uses the uncertainty of eigenvalues to improve performance. The fuzzification of features in the field of EEG data analysis can show better performance in KNN. Yang et al [1]. Conducted a detailed discussion on the effect of distance measurement, and came to the conclusion: Compared with simply using Euclidean distance, designing the distance measurement according to the actual situation will greatly improve the accuracy of KNN classification.

B. Support Vector Machine

Support vector machines are one of the commonly used tools in machine learning. Compared with deep neural networks, support vector machines are particularly good at handling situations where feature dimensions are more than the number of samples. In the field of small samples, support vector machines are better than deep neural networks. Select [13]. The linear support vector machine aims to find a hyperplane far away from all types of samples. When the sample has random disturbances, the hyperplanes far away from the sample have a strong tolerance for the disturbance, making SVM not easy to over-fit combine. The essence of linear SVM is a convex quadratic programming problem.

$$\mathrm{argmax}_{w,b}\left(\min_i \frac{2}{\|w\|}|w^T x_i + b|\right) \qquad (4)$$

Among them, w, b are the vector of the weight and offset of the hyperplane. At this time, the learning goal of SVM is to find a suitable set of w, b values, so that the planning problem can be solved.

C. Convolutional Neural Network

According to the characteristics of EEG, a two-dimensional convolutional neural network is used in the design of the neural network. In the experiment, the data set was randomly divided into training set and test set at a ratio of 8:2, and then a neural network structure was established for training on EEG features. The training

iterations were 50 times, and the learning rate was set to 0.001. Through the analysis of the experimental results, the loss rate and accuracy curves of the training set and test set of EEG feature training are shown in Figure 3 and Figure 4. In the traditional fatigue detection method SVM, the average accuracy of KNN is 86%and 72% respectively. Compared with the traditional fatigue detection method, the convolutional neural network method has indeed improved a lot, especially the convolutional neural network method proposed in this paper can achieve 98%.

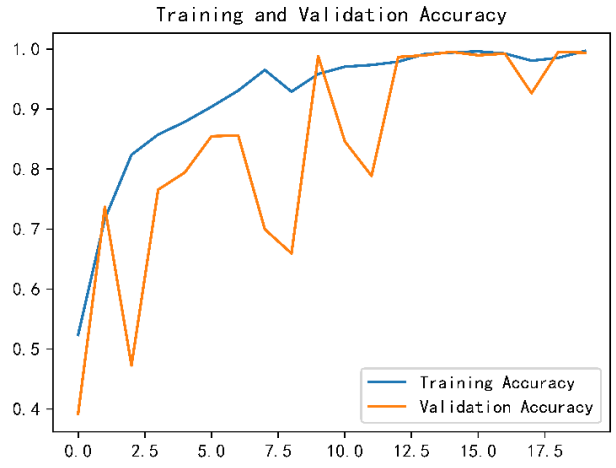


Figure 3. Accuracy of training set and test set

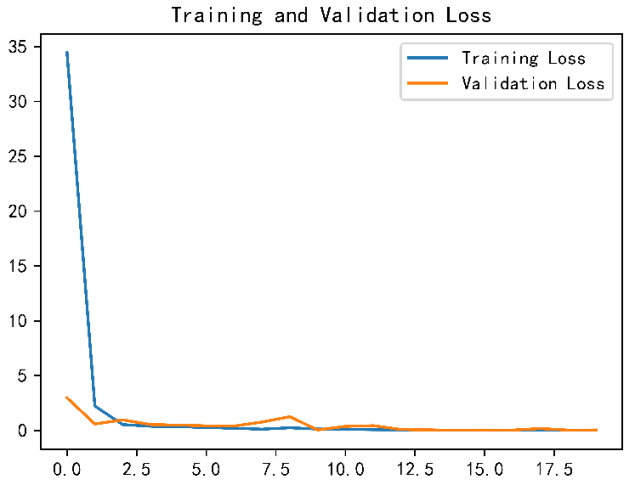


Figure 4. Loss rate of training set and test set

V. CONCLUSION

The detection of fatigue state is widely used in life, and the judgment of fatigue state is helpful to the safe operation of the operator, can reduce the occurrence of accidents, and protect the physical and mental health of the operator. In order to further verify the effectiveness of the method in identifying fatigue classification, the method in this paper is compared with traditional K nearest neighbors, support vector machines and the current more popular deep learning methods. The classification and recognition accuracy rates of KNN, SVM and CNN reached 72% respectively. , 86% and 98%, the accuracy of the method in this paper reaches 98%, which better realizes the operator's fatigue classification. Carry out operator fatigue classification, the classification accuracy rate reaches 98%, and the fatigue state classification is well realized. At the same time, the complex feature extraction process in traditional algorithms is avoided, which is beneficial to real-time and accurate detection of operator fatigue. In addition, applying the same model to different subjects, the classification accuracy rate of each subject's fatigue state exceeds 93%, which can better eliminate the influence of individual differences.

REFERENCES

[1] Nguyen,Thien.Utilization of a combined EEG/NIRS system to predict driver drowsiness. Scientific Reports. 10.1038/srep43933(2017)

[2] ZETTERBERG L H. Estimation of parameters for a linear difference equaiton with application to EEG analysis [J] . Math Biosci, 1965, 5(3-4): 227-275.

[3] MATOUSEK M, PETERSEN I. A method for assessing alertness fluctuations from EEG spectra [J] .Electroencephalogr Clin Neurophysiol, 1983, 55(1): 108-113.

[4] MAKEIG S, JUNG T P. Changes in alertness are a principal component of variance in the EEG spectrum [J] . Neuroreport, 1995, 7(1): 213-216.

[5] JAP B T, LAL S, FISCHER P, et al. Using EEG spectral components to assess algorithms for detecting fatigue [J] . Expert Syst Appl, 2009, 36(2): 2352-2359.

[6] Q.Zeng,G.Li,Y.Cui,G.Jiang,andX.Pan,"Estimating temperaturemortality exposure-response relationships and optimum ambient temperature at the multi-city level of china," International journal of environmental research public health, vol. 13, no. 3, p. 279, 2016.

[7] A. Baughman and E. A. Arens, "Indoor humidity and human health– part i: Literature review of health effects of humidity-influenced indoor pollutants," ASHRAE Transactions, vol. 102, pp. 192–211, 1996.

[8] Y. Du, P. Ma, X. Su, and Y. Zhang, "Driver fatigue detection based on eye state analysis," in 11th Joint International Conference on Information Sciences. Atlantis Press, Conference Proceedings.

[9] Keller JM, Gray MR, Givens JA. A fuzzy k-nearest neighbor algorithm. IEEE transactions on systems, man, cybernetics. 1985:580-5.

[10] Barker AL. Selection of distance metrics and feature subsets for K-nearest neighbor classifiers: University of Virginia; 1997.

[11] Yang L, Jin R. Distance metric learning: A comprehensive survey. Michigan State Universiy. 2006; 2:4.

[12] Codella N, Cai J, Abedini M, Garnavi R, Halpern A, Smith JR. Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images. International workshop on machine learning in medical imaging: Springer; 2015. p. 118-26.

[13] Oraii S. Eletrophysiology From Pants to Heart. USA: Books on Demand; 2012.

[14] Lakshmi MR, Prasad T, Prakash DVC. Survey on EEG signal processing methods. International Journal of Advanced Research in Computer Science Software Engineering. 2014; 4.

[15] Sanei S, Chambers J. EEG Signal Processing. England: John Wiley; 2007.

[16] Sundararajan A, Pons A, Sarwat AI. A generic framework for eeg-based biometric authentication. 2015 12th International Conference on Information Technology-New Generations: IEEE; 2015. p. 139-44.

[17] Lei X, Liao K. Understanding the influences of EEG reference: a large-scale brain network perspective. Frontiers in neuroscience. 2017; 11:205.

[18] Chella F, Pizzella V, Zappasodi F, Marzetti L. Impact of the reference choice on scalp EEG connectivity estimation. Journal of neural engineering. 2016; 13:036016.

飞行模拟训练的疲劳分类研究

王礼恩
计算机科学与工程学院
西安工业大学
西安,中国
e-mail: wanglien@st.xatu.edu.cn

王长元
计算机科学与工程学院
西安工业大学
西安,中国
e-mail: Cyw901@163.com

摘要—疲劳是影响现代飞行安全的重要因素。很容易导致飞行员操作能力下降、误判和飞行幻觉。而且，它甚至可能导致严重的飞行事故。本文采用可穿戴无线生理设备在模拟飞行实验中获取飞行员心电图数据。生物电信号比图像信息具有更高的可靠性，不易受外界环境（如拍摄角度、光照强度等）的影响。另一方面，神经网络已广泛用于各种分类和回归任务。本研究将脑电图采集在驾驶飞行模拟器中，经过简单的滤波和预处理后，将时域数据直接发送到卷积神经网络，无需额外的特征提取操作。我们发现卷积神经网络可以有效放大时域数据的波动细节，训练飞行员疲劳状态识别模型。结果表明，卷积神经网络模型的识别准确率达到了 98%，分别比传统的 k 近邻分类算法（KNN）和支持向量机（SVM）模型提高了 26%和 12%。本文建立的基于卷积神经网络的识别模型适用于飞行员疲劳状态的识别。这对于减少飞行员疲劳造成的飞行事故具有重要的现实意义，为飞行员疲劳风险管理和智能飞机自动驾驶系统的发展提供了理论依据。

关键词-疲劳分类; 脑电; 卷积网络;神经网络

1. 背景

国际民航组织(ICAO)过去十年全球计划商业飞行事故和人员伤亡的统计数据显示,2013 年至 2017 年,每年的飞行事故数量变化不大,但伤亡人数却一直很大波动较大,总伤亡人数居高不下。此外,一旦发生事故,造成的直接损失是飞机的生产和制造成本(波音飞机用品的平均价值约为 9000 万美元),加上事故损失的赔偿。尽管近几十年来飞行事故发生率呈下降趋势,但飞机事故造成的伤害和损失并没有太大变化。同时,每一次飞行事故都会给每个家庭造成致命的伤害,间接造成国家损失。因此,全球各个地区的国家都需要严格对待航空安全问题。根据美国联邦航空局(FAA)和美国宇航局的飞行事故统计,只有 12%的飞行事故是由飞机本身的问题造成的,超 73%的事故涉及人为因素。67%的飞机事故是由飞机失误造成的,其中最主要的因素是飞行员的操作失误,约占空难总数的 51%。飞行员操作失误的主要原因是驾驶员处于疲劳状态,驾驶时警觉。学位下降。据我国道路交通事故统计,90%的交通事故是由驾驶员的人为因素造成的,主要是由于驾驶员的危险驾驶状态,如分心驾驶、疲劳驾驶等。

目前检测飞行员警觉性的方法主要分为三个方向,即基于飞机的行为监测、飞行员行为记录和飞行员生理信号测量。其中,前两种方法受机种、驾驶环境等外界环境影响较大,后一种方法仅取决于主体条件;因此,它显示出更高的检测驾驶员睡意的能力。生理信号的测量包括来自脑电图(EEG)的神经元电活动、来自眼电图(EOG)的眼球运动、来自心电图(ECG)的心率、来自肌电图(EMG)的肌肉活动和来自近红外光谱(NIRS)的组织氧合[1]。然而,在这些信号中,不太可能受个体特征影响的 EEG 信号已被广泛用作疲劳检测的“黄金标准”,并已被证明是研究困倦和驾驶员警觉性变化的一种很有前途的方法.

脑电图是脑神经细胞电生理活动对头皮表面或大脑皮层的整体反应[2]。按其频率可分为 5 个不同波段:(1)δ 波(1~4Hz),一般只在成人睡着时出现;(2)θ 波(4~8Hz),主要发生在睡眠状态;(3)α 波(8~14Hz),一般出现在松弛状态;(4)β 波(14~30Hz),β 波功率谱的增加与警觉性的增加密切相关。

早在 1980 年代,国外就已经开展了脑电图与脑疲劳相关性的研究。相关研究表明,脑电图对警觉性的波动非常敏感。脑电图会随着警惕性的变化而发生显着变化。EEG 可以预测持续脑力劳动导致的大脑性能下降[3];1990 年代,脑电图研究进一步深入,人们开始关注脑疲劳时脑电图各波段的变化。研究发现人体警觉性降低与疲劳的相关性会具体体现在脑电图的不同波段,其中θ波和α波的变化尤为明显[4],具体表现在θ波的功率谱上和人处于疲劳状态时的α波;进入 21 世纪,对脑疲劳状态下脑电图不同频段的研究更加细致,脑电图中的 θ、α、β 频段以及不同频段的组合参数。例如,日本等人[5]进行了更全面的实验。δ,θ,α,β 和(θ+α)/β,α/β, (θ+α)/(α+β)和θ/β 四个参数的 EEG 四个波段的评估表明,在从非疲劳状态到疲劳状态的过渡,δ 波和 θ 波的活动相对稳定,α 波的活动略有下降。β 波活动明显降低;4 个组合参数的值都有所增加,其中(θ+α)/β 的增加更为明显,(θ+α)/β 的值在不同疲劳等级下的差异也更为明显。通过对不同频段的脑电信号进行组合分析,然后提出合适的组合参数,不仅可以充分利用不同频段的脑电信号各自的特点,而且通过参数的组合,检测结果可以更加准确和全面。

国内很多学者也对 EEG 对驾驶员的警觉性进行了研究。除了上述四种组合的脑电图特征外,还有脑电波、功率谱密度(PSD)和熵的时域分析和频域分析。对驾驶员的警觉性进行同等角度的分析评价。脑电图也是人因工程中的一个重要参数。EEG 的脑疲劳检测在人因工程中也有广泛的应用。例如,利用脑电疲劳检测方法对飞行员的专业训练和心理调整进行评估和指导,发现飞行员的积极情绪水平有所提高。同样,使用 EEG 进行脑疲劳检测也可以研究脑疲劳对选择性视觉注意的影响,结果表明精神疲劳对选择性视觉注意能力有负面影响。

目前使用脑电图可以判断脑疲劳程度和脑疲劳程度。随着高通量脑电技术的发展和脑电数据分析的智能化,高通量脑电数据的溯源分析有望用于脑疲劳定位分析;另一方面,同步脑电图与功能磁共振成像技术的融合也为脑疲劳定位分析提供了技术支持。总之,基于 EEG 的脑疲劳检测将朝着定量化、精准化、精准定位的方向发展,驾驶员的疲劳检测能力和可信度也将不断提升。

2. 实验

2.1 实验设计

模拟飞行驾驶与真实驾驶的难度和危险程度存在一定差距。为减少新老学员对驾驶任务控制能力的影响,保证测试数据的客观性,选取 20 名研究生,年龄 24-28 岁(平均年龄 26.8 岁),身体健康,都是右撇子;没有驾驶经验;测试期间没有服用任何药物;测试前 24 小时未食用酒精食物,测试前 12 小时未食用含咖啡因的饮料。前 1 小时没有吃东西和剧烈运动;为避免测试周期的影响,测试在与人体生理周期相近的周期内完成;考试前 1 天,对受试者进行驾驶模拟器操作 20 分钟的培训,并尽量保证样本正确、驾驶水平的相同性和平等性。参与者被告知实验的具体内容,以确保他们充分了解在研究期间收集的生理数据将如何被使用。每个参与者都愿意参与实验;实验数据必须剔除个人身份信息,只保留对实验有特定影响的数据。研究在中午(12:00-14:00)进行,室内温度控制在 25±2℃[6],室内湿度为 45±10%[7]。要求所有参与者在实验前 5 小时内进行适度的脑力劳动,以降低神经兴奋性。他们不允许参加任何形式的体力劳动,以防止血压和心率的变化。每个参与者在实验前都有 8-9 小时的睡眠时间。观察者被设置为记录实验者的状态,包括眼睛是否有红血丝和眨眼频率的变化[8]。神经网络模型采用 Tensorflow GPU 2.4.0 框架,CUDA 版本为 11.0,cuDNN 版本为 8.0。我们使用 4 块 NVIDIA Titan V 显卡来加速训练过程。

2.2 实验流程

实验刺激呈现在 31.5 英寸桌面曲面显示器上。屏幕上的互动对象是由 Dovetail Games 发行、航空专家 Jane Whittaker 开发的《微软飞行模拟器 X:Steam 版》游戏。这款游戏可以让非职业玩家在遇到紧急情况时感受到飞行员的紧张情绪。画面和逼真程度达到了巅峰,现实飞行中遇到的各种元素,如空气动力学、天气、地理环境、飞行控制系统、飞行电子系统、作战飞行武器系统、地面飞行制导等在计算机中进行综合模拟,通

过外部硬件设备进行飞行模拟控制和飞行感官反馈,反馈飞行员上一年的疲劳表现。以眼动、视线等为基础,完成飞行员训练科目(起飞、着陆)的脑电采集与分析,识别飞行员训练过程中的情绪特征,完成多维通道数据融合,构建飞行员控制响应时间和注意力分布的数学模型监控和评估飞行训练的效果并进行控制实验。进入模拟控制环境前,被试需佩戴脑电设备,确保设备与软件连接进行电阻测量(Ergo 软件显示端口为绿色),Ergo 软件会显示端口的连接状态.有四种:1.绿色:端口连接正常,信号稳定 2.橙色:端口连接正常,信号不稳定 3.红色:端口连接正常,信号弱 4.灰色:实验前端口无信号,需确保设备与软件连接正常,测量电阻(所有端口均为绿色),并眨眼闭眼确认设备正在接收眼球一般。测试对象进入模拟飞行环境后,需要在半小时内不断与实验材料交互并逆行,完成固定翼飞机的起降,绕场飞行五圈,继续飞行指定的路线。在正式测试过程中,受试者需要进入模拟机动场景(如图 1),飞行模拟平台模拟驾驶的实验任务。

图 1 模拟飞行环境

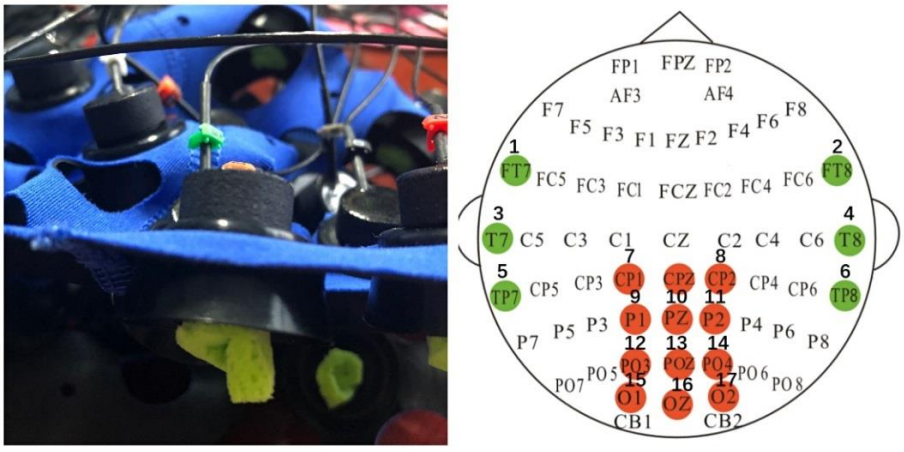


图 2 脑电设备

任务执行过程中,被试视野范围内会随机出现弹窗,被试需要将视点移至弹窗关闭窗口。从弹出窗口出现到关闭的时间间隔就是反应时间。根据响应时间,研究团队在弹出窗口出现前 10

秒内为脑电信号开发了一个心理标签(疲劳/心理)。一组实验设置半小时。因为脑电帽会影响被试,时间过长会增加被试眼高,造成极度不适。因此,本课题将一组实验分为两部分,具体时间表如表 1 所示。

表 1 实验计划表

Participant ID	001	002	003
Experiment 101	9:00-9:30	10:00-10:30	11:00-11:30
Experiment 102	13:30-14:00	14:30-15:00	15:30-16:00

在完成任务期间,限制受试者的身体和头部进行更剧烈的运动。而且,在实验开始前,记录仪会告诉被试,闭眼的行为应该是自发的,而不是刻意的,比如在感到困倦或不舒服的时候让眼睛闭上。由于对人身安全没有实际威胁,因此对象在疲劳时会顺其自然。这些情况在真实的操作环境中是很难获得的。这有助于我们研究人体疲劳时生理信号变化的特征。

3. 技术路线

Carlo Matteucci 等人[13]。1881 年首先用电流计获得肌肉神经电信号,确立了神经生理学的概念。在接下来的近一百年里,人们逐渐明确了生物电信号的采集方式和标准。已广泛应用于脑电信号领域的无创采集方法包括脑电图、磁共振成像、近红外光谱法,磁脑电图有四种[14]。其中,集时间分辨率高、成本低、无创安全于一体的多通道电极脑电图方法应用最为广泛。由于技术的不断进步,美国临床神经生理学会制定的 10-20 标准导联、10-10 标准导联和 10-5 标准导联在临床试验中最为常见[15]。这三个标准系统指南是彼此的扩展,并且在命名规则中保持相同的重叠。这简化了脑电信号研究过程,降低了技术交流的难度。也为本研究扫清了电极命名规则的障碍。具体的电极命名和空间坐标可以在美国临床神经生理学会网站上查询,这里不再详述。本项目以 10-10 标准导联系统推荐的电极位置为基准开始实验。

EEG 信号是神经活动的标志。人体任何部位所产生的神经活动都会或多或少地反映在脑电信号中,研究需要重点关注“事件电位”,即人脑因某一个或某些活动产生的一些生理电信号,因此根据不同的研究重点选择合适的参比电极将大大减少研究工作量。徐磊等人的研究。[17,18]表明参考电极是脑电图和事件相关电位研究的关键。由于姚等人。2001 年提出参比电极标准化技术,REST 迅速成为大多数脑电图研究模型的首选。此外,在某些情况下,全电极平均也可以认为具有与 REST 相同的效果。埃斯尔等人。建议在研究结果不明确时选择 FCz 电极作为参比电极。基于 FCz 电极的一致性高于基于非颅参考的一致性,他们的研究已成为 REST 的一部分。姚德中等。研究了不同参比电极对光谱映射的影响。REST 技术旨在在传统参考电极(如头皮或平均参考电极)和理论零参考电极之间架起一座桥梁。无穷远处的参考点具有理论中性电位,在 REST 中被视为近似零电位点。

脑电信号是 5-100μV 的低频生物电信号,需要经过放大才能显示和处理。信号处理和解释中的一个重要操作是滤波。滤波的主要作用是去除脑电信号数据中的干扰信号。特别是对于外部环境引起的高频信号,脑电信号处理中使用的滤波主要分为高通滤波、低通滤波和陷波滤波。可以分析过滤后的数据的特征。EEG 数据有两个主要特征:空间特征和时频特征。高通滤波旨在使高于有限频率的信号无衰减地通过,同时阻止和衰减低于有限频率的信号。由于 EEG 是 30Hz 左右的信号,而 50Hz 以上的频率只涉及癫痫的医学诊断和人脑生理,因此 EEG 信号很少使用高通滤波,但也有学者选择这样做。0.1~0.7Hz 的高通滤波器设计用于去除呼吸等干扰极低的频率成分。如果信号中存在基线漂移问题,Alste 等人。对心电图进行了研究,并建议使用高通滤波来处理此类问题。尽管高通滤波可能是解决基线漂移的方法之一,但 Acunzo 等人。发现高通滤波会导致早期的 ERP 和 ERF 系统偏差。在处理脑电图和 ERP 信号融合模型时,谨慎使用高通滤波。

低通滤波允许频率低于一定范围的信号通过,高于临界频率的信号被阻挡和衰减。因为脑电信号采集仪对微弱的电信号比较敏感,而我国市电频率在 50Hz 左右,虽然距离目标 30Hz 以下还有一定的频率空间,但必须进行低通滤波以

降低 电源信号对数据的影响。该操作也是脑电信号处理中的常规操作之一。

麦克法兰等人。1997 年提出了一种基于 EEG 的空间滤波器选择,通过选择不同的滤波器对信号进行处理,得到清晰的 EEG 信号。在研究 EEG 信号时,Higashi 等人。发现基于电极权重的公共空间模式方法的空间滤波器在基于运动图像的脑电信号分类中非常有效,但现有方法存在一定的局限性。为此,提出了一种判别滤波器组来提取与运动图像的大脑活动相关的频带。

时频域法也是一种常见的脑电信号研究方法。Hjorh、Salinsky 和 Valdes 讨论了 EEG 频域分析的可靠性。时频分析的一个重要步骤是将时域信号转换为频域信号。如果信号在统计上是稳定的,或者有一个固定的规律,那么可以通过线性变换将有限长度的信号变换为频率形式。

4. 研究方法

EEG 数据分析涉及多种信号处理技术,包括但不限于信号采集、预处理和特征提取。还有多种方法被广泛应用于数据分类,如基于样本特征距离的 KNN 和 VC 理论。线性 SVM 和基于卷积神经网络的模型。本章将分别介绍本研究涉及的上述三个领域的关键技术。

脑电信号处理主要由信号采集、转换参考、滤波、去伪影、分割、独立分量分析等操作组成。此外,还可以根据实际情况选择是否对采集的数据进行下采样,但需要注意的是,如果是下采样,可能需要进行线性或非线性插值来补充消失的特征。本文主要提取 EEG 的δ、θ、α、β四个特征进行疲劳分类。

4.1 K 最近邻法

K 最近邻(KNN)算法是常用的分类算法之一。当对数据分布知之甚少或没有先验知识时,KNN 应该是首选方法。Cover 等人[9]。明确了 KNN 算法的分类误差上限是贝叶斯分类误差的两倍。该算法旨在计算未知样本与已知样本组之间的特征距离,并根据距离推断未知样本的类别。常见的距离有欧几里得距离、闵可夫斯基距离、曼哈顿距离、切比雪夫距离等。

欧式距离是最常用的测量方法,它测量多维空间中点之间的绝对距离,定义如下。

$$d(x,y)=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2} \quad (1)$$

由于欧氏距离是根据各维度特征的绝对值计算的,所以使用欧氏距离的前提是要保证指标的维度具有相同的维度。不同的维度可能会导致欧式距离无效。心灵距离是欧几里得距离。概括来说,下式中的当前 p=2 就是欧几里得距离。

$$d(x,y)=(\sum_{i=1}^{n}|x_i-y_i|^p)^{1/p} \quad (2)$$

曼哈顿距离源自城市街区距离。多维距离求和的结果，即公式中 p=1 时，得到曼哈顿距离。

$$d(x,y)=\sum_{i=1}^{n}|x_i-y_i| \quad (3)$$

此外，特征转换还可以在一定程度上提高模型的准确性［１０］。常用的特征变换包括标准化和模糊化。标准化消除了同一维度状态下不同尺度的影响，而模糊化则利用特征值的不确定性来提高性能。ＥＥＧ数据分析领域的特征模糊化可以在ＫＮＮ中表现出更好的性能。杨等人［１］。对测距的效果进行了详细的讨论，得出结论：与单纯使用欧式距离相比，根据实际情况设计测距大大提高ＫＮＮ分类的准确率。

4.2 支持向量机

支持向量机是机器学习中常用的工具之一。与深度神经网络相比,支持向量机特别擅长处理特征维度大于样本数量的情况。在小样本领域,支持向量机优于深度神经网络。选择[13]。线性支持向量机旨在找到一个远离所有类型样本的超平面。当样本有随机扰动时,远离样本的超平面对扰动有很强的容忍度,使得 SVM 不容易过拟合组合。线性支持向量机的本质是一个凸二次规划问题。

$$\mathrm{argmax}_{w,b}\left(\min_i \frac{2}{\|w\|}|w^T x_i+b|\right) \quad (4)$$

其中,w,b 为超平面的权重和偏移量的向量。此时 SVM 的学习目标是找到一组合适的 w,b 值,从而解决规划问题。

4. 3 神经网络

根据脑电图的特点，在神经网络的设计中采用了二维卷积神经网络。实验中，将数据集以 8:2 的比例随机分为训练集和测试集，然后建立神经网络结构进行脑电特征训练。训练迭代次数为 50 次，学习率设置为 0.001。通过对实验结果的分析，EEG 特征训练的训练集和测试集的损失率和准确率曲线如图 3 和图 4 所示。在传统的疲劳检测方法 SVM 中，KNN 的平均准确率为 86%和 72%。与传统的疲劳检测方法相比，卷积神经网络方法确实提升了很多，尤其是本文提出的卷积神经网络方法可以达到 98%。

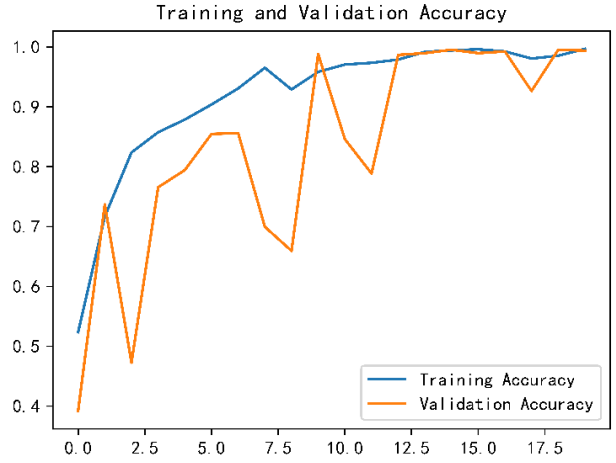


图 3 训练集和测试集的准确率

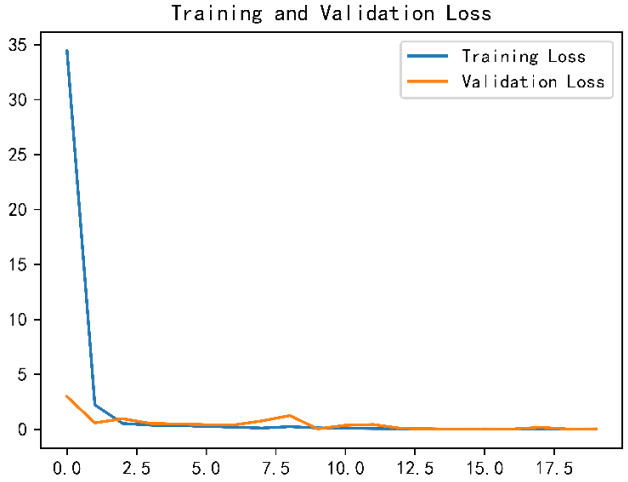


图 4 训练集和测试集的损失率

5. 结论

疲劳状态的检测广泛应用于生活中,疲劳状态的判断有助于操作者的安全操作,可以减少事故的发生,保护操作者的身心健康。为了进一步验证该方法在识别疲劳分类方面的有效性,本文将本文方法与传统的 K 近邻、支持向量机以及目前比较流行的深度学习方法进行了对比。KNN、SVM 和 CNN 的分类和识别准确率分别达到 72%。,86%和 98%,本文方法的准确率达到 98%,较好地实现了操作者的疲劳分类。进行操作员疲劳分类,分类准确率达到 98%,很好地实现了疲劳状态分类。同时避免了传统算法中复杂的特征提取过程,有利于实时、准确地检测操作者的疲劳。此外,将相同的模型应用于不同的受试者,每个受试者疲劳状态的分类准确率超过 93%,可以更好地消除个体差异的影响。

参考文献

[19] Nguyen,Thien.Utilization of a combined EEG/NIRS system to predict driver drowsiness. Scientific Reports. 10.1038/srep43933(2017)

[20] ZETTERBERG L H. Estimation of parameters for a linear difference equaiton with application to EEG analysis [J] . Math Biosci, 1965, 5(3-4): 227-275.

[21] MATOUSEK M, PETERSEN I. A method for assessing alertness fluctuations from EEG spectra [J] .Electroencephalogr Clin Neurophysiol, 1983, 55(1): 108-113.

[22] MAKEIG S, JUNG T P. Changes in alertness are a principal component of variance in the EEG spectrum [J] . Neuroreport, 1995, 7(1): 213-216.

[23] JAP B T, LAL S, FISCHER P, et al. Using EEG spectral components to assess algorithms for detecting fatigue [J] . Expert Syst Appl, 2009, 36(2): 2352-2359.

[24] Q.Zeng,G.Li,Y.Cui,G.Jiang,andX.Pan,“Estimating temperaturemortality exposure-response relationships and optimum ambient temperature at the multi-city level of china,” International journal of environmental research public health, vol. 13, no. 3, p. 279, 2016.

[25] A. Baughman and E. A. Arens, “Indoor humidity and human health– part i: Literature review of health effects of humidity-influenced indoor pollutants,” ASHRAE Transactions, vol. 102, pp. 192–211, 1996.

[26] Y. Du, P. Ma, X. Su, and Y. Zhang, “Driver fatigue detection based on eye state analysis,” in 11th Joint International Conference on Information Sciences. Atlantis Press, Conference Proceedings.

[27] Keller JM, Gray MR, Givens JA. A fuzzy k-nearest neighbor algorithm. IEEE transactions on systems, man, cybernetics. 1985:580-5.

[28] Barker AL. Selection of distance metrics and feature subsets for K-nearest neighbor classifiers: University of Virginia; 1997.

[29] Yang L, Jin R. Distance metric learning: A comprehensive survey. Michigan State Universiy. 2006; 2:4.

[30] Codella N, Cai J, Abedini M, Garnavi R, Halpern A, Smith JR. Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images. International workshop on machine learning in medical imaging: Springer; 2015. p. 118-26.

[31] Oraii S. Eletrophysiology From Pants to Heart. USA: Books on Demand; 2012.

[32] Lakshmi MR, Prasad T, Prakash DVC. Survey on EEG signal processing methods. International Journal of Advanced Research in Computer Science Software Engineering. 2014; 4.

[33] Sanei S, Chambers J. EEG Signal Processing. England: John Wiley; 2007.

[34] Sundararajan A, Pons A, Sarwat AI. A generic framework for eeg-based biometric authentication. 2015 12th International Conference on Information Technology-New Generations: IEEE; 2015. p. 139-44.

[35] Lei X, Liao K. Understanding the influences of EEG reference: a large-scale brain network perspective. Frontiers in neuroscience. 2017; 11:205.

[36] Chella F, Pizzella V, Zappasodi F, Marzetti L. Impact of the reference choice on scalp EEG connectivity estimation. Journal of neural engineering. 2016; 13:036016.

Spectral Efficiency Classification Schemes for Future Network Communications(SECS)

Zhongsheng Wang
State and Provincial Joint Engineering Lab. of Advanced Network, Monitoring and Control
Xi'an Technological University
No.2 Xuefu Middle Road, Weiyang district
Xi'an, Shaanxi, China
E-mail: 59483672@qq.com

Qingsong Zhang
State and Provincial Joint Engineering Lab. of Advanced Network, Monitoring and Control
Xi'an Technological University
No.2 Xuefu Middle Road, Weiyang district
Xi'an, Shaanxi, China

***Abstract*—Future Network is an project created and managed by ISO/IEC. The project has produced technical reports in ISO/IEC TR 29181 series and is in the process setting architectures and protocols. The project is known for its distinctive "clean slate design" approach and works on fundamental structural innovations to allow Future Network deliver its promises. Simultaneously, ISO/IEC Future Network should prepare itself for future breakthrough in SE technology and make plans to adapt Future Network to the fast changing "Post Shannon Era" technological revolutions. Using reference to the mechanism of radio frequency band classifications, this standard classifies the spectral efficiencies of the MCS systems, so as to facilitate the classification, discussion, evaluation and comparison of the efficiency of the spectrum of information systems.**

Keywords-Future Network Communications; Spectral Efficiency Classification Schemes（SECE）; International Standardization

I. INTRODUCTION

This paper is the first international standard proposal (draft) of The Ultra Limited Future Network. The development of this proposal (draft) is supported by JTC 1/SC 6/WG 7 working team. The application leader is Dr. Qingsong Zhang (Nanjing Bofeng), and the proposal is proposed by Professor Wang Zhongsheng (Xi 'an Technological University). Itu-r and ICAO acted as the focal points during the development process, while cooperating with ECMA, 3GPP and IEEE.

This standard is prepared in accordance with the provisions of the following three documents :

ISO/IEC TR 29181：Future Network：Problem Statement and Requirements, parts 1-9.

ITU-R SM.856-1：New Spectrally Efficient Techniques and Systems (1992-1997).

ITU-R SM.1046-3：Definition of Spectrum Use Efficiency of a Radio System (2017-09-06).

As an international standard proposal in the field of "industry, Innovation and Infrastructure" , This standard provides classification schemes for MCS Spectral Efficiencies including:

Definition of Modulation and Coding Scheme Spectral Efficiency (MCS SE).

- Method for classification of MCS SE.
- Naming system of MCS SE.

- Examples of the recommended use of SECS.
- Potential impact on Future Network Standardization.
- The differences between SE and Spectral Utilization Efficiency.

II. DEVELOPEMNT BACKGROUND

Spectrum efficiency is a key index to measure the level of development of information and communication systems, which can reflect or affect many key performances, including the efficient use of spectrum resources, higher information transmission rates and greater channel capacity and so on. The higher the spectrum efficiency, higher the utilization rates of resource-constrained spectrum resources, higher transmission rates, and greater the information throughput. Therefore, improving spectrum efficiency is one of the most critical objectives of ICT innovation.

With the development of digital information and computer science and technology, the level of spectrum efficiency of information and communication systems is also improving. 20 years ago at the turn of the century, the spectral efficiency of communication systems was still at a very low level of 1-2 bits. After 20 years of development, some areas of technology are already using 10-bit spectrum efficiency technology. Some areas have incorporated 12-bit spectrum efficiency into next-generation technical standard planning.

Based on historical experience and technical characteristics, the speed of improvement of the spectrum efficiency level of information and communication systems will slow down in the future, reaching 16 bits in 20 years and 20 bits in 45 years, which is the result of the congenital limitations of M-QAM modulation and demodulation technology, a key mechanism for improving spectrum efficiency. However, it cannot be ruled out that the emergence of new theories and new technologies in the basic layer of information and communication physics may lead to rapid improvement of spectrum efficiency. An article published in IEEE Access in 2018 by Chinese scientist Professor Li Daoben shows that information systems with up to 2000 bits spectrum efficiency can be achieved using the overlapping multi-domain multiplexing technology (OVXDM) he invented.

High spectrum efficiency will be the main manifestation of ICT levels in the post-Shannon era. In the next ten to twenty years, the discussion and evaluation criteria on the spectral efficiency of communication technology will exceed 20 bits and enter the category of hundreds of bits or even thousands of bits. Communication products will be increasing spectrum efficiency as the main sign of technical level and service capability.

In the existing ITU international standards, RSM.1046-3 provides for the definition of spectrum utilization efficiency and the evaluation methods for the utilization of various systems spectrum, but does not provide a mechanism for classifying MCS spectrum efficiency. Such a mechanism is necessary to discuss, analyze, evaluate, select and manage the spectrum efficiency of future communication systems.

For example, in some technical or policy documents, the discussion of "low spectrum efficiency" and "high spectrum efficiency" can often be seen, but there are no technical specifications to define and interpret these two concepts. How many bits of spectrum efficiency is "low spectrum efficiency"? How many bits of spectrum efficiency is "high spectrum efficiency"? Some documents refer to a 10-bit system as "high spectrum efficiency", so what category does 16-bit,

20-bit, 32-bit, or even 128-bit spectrum efficiency fall into? Therefore, the two-level classifications of "low spectrum efficiency" and "high spectrum efficiency" cannot meet the future development trend and the need for more accurate spectrum efficiency classification.

In radio spectrum management, there are many classification schemes for spectrum resources. One scheme is dividing frequency resources (RE) into kilohertz, megahertz, gigahertz, and terahertz. Another method is based on wavelength, dividing RE into categories such as Ultra-Long Wave, Long Wave, Medium Wave, Short Wave, Ultra-Short Wave, Microwave, etc. There is also a classification method by frequency, VLF, Low Frequency, Medium Frequency, Medium High Frequency, High Frequency, VHF, UHF, UHF, UHF, UHF and so on. Another band division method marked by the English Alphabet, dividing RE into L-Band, S-Band, C-Band, X-Band, Ku-Band, K-Band, Ka-Band, and so on.

Using reference to the mechanism of radio frequency band classifications, this standard classifies the spectral efficiencies of the MCS systems, so as to facilitate the classification, discussion, evaluation and comparison of the efficiency of the spectrum of information systems.

III. Abbreviations and Terms

A. Abbreviations

TABLE I. Abbreviations of terms

Abbreviations	Full Name
MCS	Modulation and Coding Scheme
MIMO	Multiple Input Multiple Output
FN	Future Network
OFDM	Orthogonal Frequency Division Multiplexing
OCCS	Over-Capacity Communication Systems
OVXDM	Overlapped X Domain Division Multiplexing
OVTDM	Overlapped Time Domain Division Multiplexing
QAM	Quadrature Amplitude Modulation
RE	Resource Element
RSE	Relative Spectral Efficiency
SE	Spectral Efficiency
SEI	Spectral Efficiency Index
SUE	Spectrum Utilization Efficiency
TSEI	Typical SE Indicator

B. Terms and Definitions

Over-Capacity Communication：Exchange of information that has higher capacity than the Shannon Limit.

OVXDM: An innovative way of modulation and coding that utilizes multiple domains such as time, frequency, spatial and coding overlapping and multiplexing to achieve higher SE, no coding overhead, higher coding gain and low decoding complexity.

MCS Spectral Efficiency： The maximum amount of useful information sent in one second and per Resource Element (RE in Hz) through a communication system based on its modulation and Coding schemes.

Spectral Utilization Efficiency：the product of the frequency bandwidth, the geometric (geographic) space, and the time denied to other potential users: $U = B \cdot S \cdot T$

Shannon Limit: Also known as Shannon Capacity, defined by Claude Shannon in the 1940s setting the limit of theoretically highest rate of information transmission under certain noise levels for a single channel.

Future Network: An International Standard project developed and managed by ISO/IEC for a new network system based on the clean slate design approach. Publications include ISO/IEC TR 29181 and ISO/IEC 21558-21559.

IV. MCS SE Classification Principles

A. Differentiating SE from USE

The SE (Spectral Efficiency) defined in this standard shall not be confused with the Spectrum Efficiency in ITU-RSM.1046-3 (2017). The TABLE II lists these differences:

TABLE II. Comparison of this standard with ITU standards

	Comparison Objective	This Standard	ITU
1	Source	ISO/IEC	ITU-RSM.1046-3
2	Term	Spectral Efficiency	Spectrum Efficiency
3	Abbreviation	MCS SE	USE
4	Factor	bps/Hz	$U = B \cdot S \cdot T$
5	Considering factors	● Capacity ● Resource element (Hz) ● Time (second)	● Bandwidth ● Geometric space (area) ● Time
6	Improvement method	● Modulation ● Channel Coding	● Antenna Directivity ● Geographical Spacing ● Frequency Sharing ● Orthogonal Frequency use ● Time-sharing ● Time division
7	SE Gain potentials	Sky is the limit	Limited potential
8	Perspective	Communication system	User
9	Service	For all	Denying others

The MCS spectrum efficiency defined in this standard refers to the number of bits of valid information transmitted per second per hertz frequency resource through technical means such as modulation and channel coding.

The value of MCS spectrum efficiency is relatively fixed. So long as we know the modulation mechanism and channel coding method used, we can deduct the performance level of the theoretical MCS spectrum efficiency. Because of the small variety of modulation mechanism and channel coding methods, some mainstream technology applications are very broad, such as M-QAM technology in the field of modulation and Turbo code and LDPC code in the field of channel coding. Therefore, MCS spectrum efficiency can be used as a general and important index to assess the performance level of ICT in different fields.

B. Deciding the Range of SE

Currently, communication systems having SE no higher than 10 bps/Hz, some systems may reach 12 bps/Hz in about five years from now. At such a low SE rate, there is no need for a standard classify SE levels.

However, standards are expected to identify future trends, provide directions for technological development, and to have market relevance lasting decades. Since there have been technical trends indicating potential breakthrough in spectral efficiency, this standard takes into account of SE in the hundreds and thousands bps/Hz range.

C. MCS SE Classification Architecture

MCS SE classification system contains three schemes described in TABLE III.

TABLE III. MCS SE CLASSIFICATION SYSTEM DESCRIPTION SCHEME

Scheme	Feature	Format	Purpose
A	Two Letter	#-SE	Indicating specific product SE capabilities
B	Three Letter	VSE	Group SE into category of levels
C	Four letter	DDSE	Provide an alternative and simpler classification of SE
D	Two levels	Lower	Make broader range

V. MCS SPECTRAL EFFICIENCY CLASSIFICATION SCHEMES

A. MCS SE Classification A: Two Letter Scheme

The two letter MCS SE classification system uses only two letters "SE" with numbers indicating specific bps/Hz. It is used not for referring to a level or class, but rather to indicate specific SE performance of a product.

Expression description: number of bits (omitting "s/Hz") with"-"followedby"SE", indicating "spectral efficiency at specific bps/Hz"."

Example:

"56-SE", which means spectral efficiency rate at 56 bps/Hz.

"256-SE", which means spectral efficiency rate at 256 bps/Hz.

"1008-SE", which means spectral efficiency rate at 1008 bps/Hz.

B. MCS SE Classification B: Three Letter Scheme

TABLE IV. THREE LETTER SCHEME IN MCS SE CLASSIFICATION SCHEME

SE Index	Index name	Full Title	SE Range（bps/Hz)	TSEI* （bps/Hz)
SEI 1	BSE	Basic Spectral Efficiency	0.1~2.0	2
SEI 2	LSE	Low Spectral Efficiency	2.1~5.9	5
SEI 3	MSE	Medium Spectral Efficiency	6~10.9	10
SEI 4	HSE	High Spectral Efficiency	11~15	15
SEI 5	VSE	Very-High Spectral Efficiency	16~20	20
SEI 6	USE	Ultra-High Spectral Efficiency	21~32	32
SEI 7	SSE	Super Spectral Efficiency	33~64	64
SEI 8	OSE	One-hundred level spectral efficiency	65~128	128
SEI 9	ESE	Extreme Spectral Efficiency	129~256	256
SEI 10	DSE	500 Spectral Efficiency	257~512	512
SEI 11	JSE	Jump Level spectral efficiency	513~999	768
SEI 12	1-KSE	1K Spectral efficiency	1000~1999	1024
SEI 13	2-KSE	2K Spectral efficiency	2000~2999	2048
SEI 14	3-KSE	3K Spectral efficiency	3000~3999	3072
SEI 15	4-KSE	4K Spectral efficiency	4000~4999	4096
SEI 16	XSE	X Spectral efficiency	5000~6999	6144
*TSEI is the Typical SE indicator for its class.				

As Spectral Efficiency increases, the gaps among the Three Letter Scheme also expand. In SEI 4 and SEI 5, for example, there are only 4 bits differences separating the high from the low. In SEI 9, the gaps are over 200 bits and in SEI 12, the gaps grow to one thousand.

It is anticipated that there will be need for more accurate SE references or comparisons for the upper part of the Three Letter Schemes. In that case and when technological development requires such changes, the Three Letter Scheme may use the following extension Scheme.

Rule 1. No extension needed for indexes SEI 1~6.

Rule 2. The extension in grouped into two index tables, one for SE lower than 1000 bps/Hz （TABLE V） and the other is for SE above 1000 bps/Hz（TABLE VI）.

Rule 3. A single double digit decimal number is added to index name to indicate extension numbers.

Rule 4. For SSE and OSE indexes, 5bps/Hz is used as bases for extension unit.

Rule 5. For ESE and DSE indexes, 10bps/Hz is used as bases for extension unit.

Rule 6. For JSE indexes, 20bps/Hz is used as bases for extension unit.

Rule 7. For KSE indexes, 50bps/Hz is used as bases for extension unit.

Rule 8. For XSE indexes, 100bps/Hz is used as bases for extension unit.

TABLE V. EXTENDED INDEX OF SE LOWER THAN 1000 BPS/HZ

SEI 7 EXT Index	33-64 SE RANGE	SEI 8 EXT Index	65~128 SE RANGE	SEI-9 EXT Index	129-256 SE RANGE	SEI 10 EXT Index	257~512 SE RANGE	SEI 11 EXT Index	513~999 SE RANGE
SSE 1	33-38	OSE 1	65-69	ESE 1	129-139	DSE 1	257-269	JSE 1	513-539
SSE 2	39-43	OSE 2	70-74	ESE 2	140-149	DSE 2	270-279	JSE 2	540-559
SSE 3	44-49	OSE 3	75-79	ESE 3	150-159	DSE 3	280-289	JSE 3	560-579
SSE 4	50-55	OSE 4	80-84	ESE 4	160-169	DSE 4	290-299	JSE 4	580-599
SSE 5	56-60	OSE 5	85-89	ESE 5	170-179	DSE 5	300-319	JSE 5	600-619
SSE 6	61-64	OSE 6	90-94	ESE 6	180-189	DSE 6	320-329	JSE 6	620-639
		OSE 7	95-99	ESE 7	190-199	DSE 7	330-339	JSE 7	640-659
		OSE 8	100-104	ESE 8	200-209	DSE 8	340-349	JSE 8	660-679
		OSE 9	105-109	ESE 9	210-219	DSE 9	350-359	JSE 9	680-699
		OSE 10	110-114	ESE 10	220-229	DSE 10	360-369	JSE 10	700-719
		OSE 11	115-119	ESE 11	230-239	DSE 11	370-379	JSE 11	720-739
		OSE 12	120-124	ESE 12	240-249	DSE 12	380-389	JSE 12	740-759
		OSE 13	125-128	ESE 13	250-256	DSE 13	390-399	JSE 13	760-779
						DSE 14	400-409	JSE 14	780-799
						DSE 15	410-419	JSE 15	800-819
						DSE 16	420-429	JSE 16	820-839
						DSE 17	430-439	JSE 17	840-859
						DSE 18	440-449	JSE 18	860-879
						DSE 19	450-459	JSE 19	880-899

DSE 20	460-469	JSE 20	900-919				
DSE 21	470-479	JSE 21	920-939				
DSE 22	480-489	JSE 22	940-959				
DSE 23	490-499	JSE 23	960-979				
DSE 24	500-512	JSE 24	980-999				

TABLE VI. EXTENDED INDEX OF KSE AND XSE（SE ABOVE 1000 BPS/HZ）

SEI 12 EXT Index	1000-1999 SE RANGE	SEI 13 EXT Index	2000-2999 SE RANGE	SEI 13 EXT Index	3000-3999 SE RANGE	SEI 14 EXT Index	4000-4999 SE RANGE	SEI 15 EXT Index	5000-6999 SE RANGE
1KSE 1	1000-1049	2KSE 1	2000-2049	3KSE 1	3000-3049	4KSE 1	4000-4049	XSE 1	5000-5099
1KSE 2	1050-1099	2KSE 2	2050-2099	3KSE 2	3050-3099	4KSE 2	4050-4099	XSE 2	5100-5199
1KSE 3	1100-1140	2KSE 3	2100-2140	3KSE 3	3100-3140	4KSE 3	4100-4140	XSE 3	5200-5299
1KSE 4	1150-1199	2KSE 4	2150-2199	3KSE 4	3150-3199	4KSE 4	4150-4199	XSE 4	5300-5399
1KSE 5	1200-1249	2KSE 5	2200-2249	3KSE 5	3200-3249	4KSE 5	4200-4249	XSE 5	5400-5499
1KSE 6	1250-1299	2KSE 6	2250-2299	3KSE 6	3250-3299	4KSE 6	4250-4299	XSE 6	5500-5599
1KSE 7	1300-1349	2KSE 7	2300-2349	3KSE 7	3300-3349	4KSE 7	4300-4349	XSE 7	5600-5699
1KSE 8	1350-1399	2KSE 8	2350-2399	3KSE 8	3350-3399	4KSE 8	3350-4399	XSE 8	5700-5799
1KSE 9	1400-1449	2KSE 9	2400-2449	3KSE 9	3400-3449	4KSE 9	4400-4449	XSE 9	5800-5899
1KSE 10	1450-1499	2KSE 10	2450-2499	3KSE 10	3450-3499	4KSE 10	4450-4499	XSE 10	5900-5999
1KSE 11	1500-1549	2KSE 11	2500-2549	3KSE 11	3500-3549	4KSE 11	4500-4549	XSE 11	6000-6099
1KSE 12	1550-1599	2KSE 12	2550-2599	3KSE 12	3550-3599	4KSE 12	4550-4599	XSE 12	6100-6199
1KSE 13	1600-1649	2KSE 13	2600-2649	3KSE 13	3600-3649	4KSE 13	4600-4649	XSE 13	6200-6299
1KSE 14	1650-1699	2KSE 14	2650-2699	3KSE 14	3650-3699	4KSE 14	4650-4699	XSE 14	6300-6399
1KSE 15	1700-1749	2KSE 15	2700-2749	3KSE 15	3700-3749	4KSE 15	4700-4749	XSE 15	6400-6499
1KSE 16	1750-1799	2KSE 16	2750-2799	3KSE 16	3750-3799	4KSE 16	4750-4799	XSE 16	6500-6599
1KSE 17	1800-1849	2KSE 17	2800-2849	3KSE 17	3800-3849	4KSE 17	4800-4849	XSE 17	6600-6699
1KSE 18	1850-1899	2KSE 18	2850-2899	3KSE 18	3850-3899	4KSE 18	4850-4899	XSE 18	6700-6799
1KSE 19	1900-1949	2KSE 19	2900-2949	3KSE 19	3900-3949	4KSE 19	4900-4949	XSE 19	6800-6899
1KSE 20	1950-1999	2KSE 20	2950-2999	3KSE 20	3950-3999	4KSE 20	4950-4999	XSE 20	6900-6999

C. MCS SE Classification C: Four Letter Scheme

TABLE VII. FOUR LETTER SCHEME IN MCS SE CLASSIFICATION SCHEME

	Title	Whole Title	SE (bps/Hz)	Relative B categories
1	SDSE	Single Digits Spectral Efficiency	0-9	BSE，LSE，MSE
2	DDSE	Double Digits Spectral Efficiency	10-99	HSE，VSE，USE，SSE，OSE
3	TDSE	Triple Digits Spectral Efficiency	100-999	ESE，DSE，JSE
4	QDSE	Quadruple Digits Spectral Efficiency	1000-9999	M-KSE，XSE

D. MCS SE Classification D: Comparative Scheme

TABLE VIII. COMPARISON SCHEME OF MCS SE CLASSIFICATION SCHEME

	Title	Whole Title	SE (bps/Hz)	Relation with other categories
1	L	lower	none	All levers below a specific class
2	H	higher	none	All levers above a specific class

VI. SAMPLE OF MAKING REFERENCES

A. Making reference to the standard

This standard is giving the original title "OCC-STD 21001", established by the developer institution. When adopted into other standard systems such as China's Industry standard, National Standard, ISO standard and ITU standard, the title and number may be reassigned. Before then, "OCC-STD 21001" is the only source for MCS SE classifications.

In the future, when making references to the classification schemes, it is recommended that a note is included in the document that the SE classifications are defined in "OCC-STD 21001 (2021)" developed by Nanjing Bofeng Communication Technologies Ltd.

B. Examples Referring Specific Classification Levels

- In 2023, the company is expected to deliver communications systems utilizing innovative modulation schemes that can provide VSE level spectral efficiency defined in "OCC-STD 21001".
- Comparing MCS spectral efficiency, the two products belong to two generations with Sample A is only at the VSE level while Sample B contains USE modulation technology.
- Industry consensus is that the KSE level spectral efficiency technology is only a few years away.
- Some experts anticipate that entering the next decade, communication systems can reach the QDSE level Spectral Efficiency as defined in the Four Letter Classification System in "OCC-STD 21001".

C. Recommended Use of Comparative Scheme

- When use these two expressions, they shall be accompanied with reference to a specific SE class level.
- Example:
- So far, the most advanced wireless communication system have MCS spectral efficiencies lower than the VSE level as defined in "OCC-STD 21001".
- It is expected that products with higher spectral efficiency than the VSE level will enter service by 2025.
- The new system has backward compatibility design providing continuous support to MCS SE levels of.

D. Referring Specific SE Rate

When referring specific SE rate, the following statement are examples:

- "Bofeng.com offers two radio systems that operate at SSE level spectral efficiency as defined in "OCC-STD 21001". Radio A system has 48-SE modulation scheme and Radio B system has 64-SE capabilities.
- Product specification: MCS SE: 32-SE, 48-SE and 64-SE. This description

indicates the system contains three types of MCS supporting three SE rates.

VII. Potential Impact on Future Network Standardization

Future Network is an International Standardization project created and managed by ISO/IEC. The project has produced technical reports in ISO/IEC TR 29181 series and is in the process setting architectures and protocols. The project is known for its distinctive "clean slate design" approach and works on fundamental structural innovations to allow Future Network deliver its promises.

Making MCS-SE classification system a Future Network standardization item can benefit the project in many ways. Firstly, ISO/IEC Future Network will become the first international standards adopting a MCS SE classification system; Secondly, Other standardization bodies may adopt this system or make normative reference to this standard; Thirdly, ISO/IEC Future Network becomes the first standard indicating future trends in MCS SE development; Fourthly, the inclusion of three digits and four digits MCS SE in Future Network standards will reflect the huge potential of network performance and capabilities; and finally, the successful adoption of this standard will open the door of ISO/IEC Future Network standards to future technologies that achieve higher and higher MCS SE.

ISO/IEC Future Network should prepare itself for future breakthrough in SE technology and make plans to adapt Future Network to the fast changing "Post Shannon Era" technological revolutions.

References

[1] T. M. Cover and J. A. Thomas, Elements of Information Theory. Hoboken, NJ, USA: Wiley, 2006.

[2] L. Daoben, Waveform Coding Theory of High Spectral Efficiency-OVTDM and Its Application. Beijing, China: Scientific, 2013.

[3] L. Daoben, ''A novel high spectral efficiency waveform coding–OVFDM,'' China Commun., vol. 12, no. 2, pp. 61–73, Feb. 2015.

[4] S. G. Wilson, Digital Modulation and Coding. Englewood Cliffs, NJ, USA: Prentice-Hall, 1996.

[5] L. Daoben, ''A novel high spectral efficiency waveform coding-OVTDM,'' Int. J. Wireless Commun. Mobile Comput., vol. 2, nos. 1–4, pp. 11–26, Dec. 2014.

[6] L. Daoben, Statistical Theory of Signal Detection and Estimation, 2nd ed. Beijing, China: Scientific, 2005

[7] J. G. Proakis, Digital Communications. New York, NY, USA: McGraw-Hill, 2001.

[8] G. J. Foschini, ''Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas,'' Bell Labs Tech. J., vol. 1, no. 2, pp. 41–59, 1996.

[9] G. J. Foschini and M. J. Gans, ''On limits of wireless communications in a fading environment when using multiple antennas,'' Wireless Pers. Commun., vol. 6, no. 3, pp. 311–335, Mar. 1998

[10] S. Wu, L. Kuang, Z. Ni, J. Lu, D. D. Huang, and Q. Guo, ''Low-complexity iterative detection for large-scale multiuser MIMO-OFDM systems using approximate message passing,'' IEEE J. Sel. Topics Signal Process., vol. 8, no. 5, pp. 902–915, Oct. 2014.

[11] N. Wu, W. Yuan, H. Wang, Q. Shi, and J. Kuang, ''Frequency-domain iterative message passing receiver for faster-than-Nyquist signaling in doubly selective channels,'' IEEE Wireless Commun. Lett., vol. 5, no. 6, pp. 584–587, Dec. 2016.

[12] J. CØspedes, P. M. Olmos, M. SÆnchez-FernÆndez, and F. Perez-Cruz, ''Expectation propagation detection for high-order high-dimensional MIMO systems,'' IEEE Trans. Commun., vol. 62, no. 8, pp. 2840–2849, Aug. 2014.

[13] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, ''Millimeterwave massive MIMO: The next wireless revolution?'' IEEE Commun. Mag., vol. 52, no. 9, pp. 56–62, Sep. 2014.

[14] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, ''Energy and spectral efficiency of very large multiuser MIMO systems,'' IEEE Trans. Commun., vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[15] Y. S. Cho, J. Kim, W. Y. Yang, and C. G. Kang, MIMO-OFDM Wireless Communication Technology With MATLAB. Beijing, China: PublishingHouse of Electronics Industry, 2013.

[16] Y. D. Zhang, M. G. Amin, and B. Himed, ''Altitude estimation of maneuvering targets in MIMO over-the-horizon radar,'' in Proc. IEEE 7th IEEE Sensor Array Multichannel Signal Process. Workshop (SAM), Jun. 2012, pp. 257–260

[17] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, ''Massive MIMO for next generation wireless systems,'' IEEE Commun. Mag., vol. 52, no. 2, pp. 186–195, Feb. 2014.

[18] U. Gustavsson et al., ''On the impact of hardware impairments on massive MIMO,'' in Proc. IEEE Global Telecommun. Conf. Workshops (GC Wkshps), Austin, TX, USA, Dec. 2014, pp. 294–300.

[19] E. Björnson, M. Matthaiou, and M. Debbah, ''Massive MIMO with nonideal arbitrary arrays: Hardware scaling laws and circuit-aware design,'' IEEE Trans.

Wireless Commun., vol. 14, no. 8, pp. 4353–4368, Aug. 2015.

[20] J. E. Mazo and H. J. Landau, ‘‘On the minimum distance problem for faster-than-Nyquist signaling,’’ IEEE Trans. Inf. Theory, vol. 34, no. 6, pp. 1420–1427, Nov. 1988.

[21] F. Rusek and J. B. Anderson, ‘‘CTH04-1: On information rates for faster than Nyquist signaling,’’ in Proc. IEEE GLOBECOM, Nov./Dec. 2006, pp. 1–5.

[22] F. Rusek and J. B. Anderson, ‘‘Multistream faster than Nyquist signaling,’’ IEEE Trans. Commun., vol. 57, no. 5, pp. 1329–1340, May 2009.

[23] J. B. Anderson, F. Rusek, and V. Öwall, ‘‘Faster-than-Nyquist signaling,’’ Proc. IEEE, vol. 101, no. 8, pp. 1817–1830, Aug. 2013.

[24] A. Prlja and J. B. Anderson, ‘‘Reduced-complexity receivers for strongly narrowband intersymbol interference introduced by faster-than-Nyquist signaling,’’ IEEE Trans. Commun., vol. 60, no. 9, pp. 2591–2601, Sep. 2012.

[25] S. Sugiura, ‘‘Frequency-domain equalization of faster-than-Nyquist signaling,’’ IEEE Wireless Commun. Lett., vol. 2, no. 5, pp. 555–558, Oct. 2013.

[26] J. Fan, S. Guo, X. Zhou, Y. Ren, G. Y. Li, and X. Chen, ‘‘Faster-thanNyquist signaling: An overview,’’ IEEE Access, vol. 5, pp. 1925–1940,2017.

[27] K. Takeuchi, M. Vehkapera, T. Tanaka, and R. R. Muller, ‘‘Large-system analysis of joint channel and data estimation for MIMO DS-CDMA systems,’’ IEEE Trans. Inf. Theory, vol. 58, no. 3, pp. 1385–1412, Mar. 2012.

[28] D. Dasalukunte, V. Öwall, F. Rusek, and J. B. Anderson, Faster than Nyquist Signaling: Algorithms to Silicon. Dordrecht, The Netherlands: Springer, 2014.

[29] E. Bedeer, M. H. Ahmed, and H. Yanikomeroglu, ‘‘A very low complexity successive symbol-by-symbol sequence estimator for faster-than-Nyquist signaling,’’ IEEE Access, vol. 5, pp. 7414–7422, 2017.

[30] A. D. Liveris and C. N. Georghiades, ‘‘Exploiting faster-than-Nyquist signaling,’’ IEEE Trans. Commun., vol. 51, no. 9, pp. 1502–1511, Sep. 2003.

[31] Y. J. D. Kim and J. Bajcsy, ‘‘Iterative receiver for faster-than-Nyquist broadcasting,’’ Electron. Lett., vol. 48, no. 24, pp. 1561–1562, Nov. 2012.

[32] Y. J. D. Kim, J. Bajcsy, and D. Vargas, ‘‘Faster-than-Nyquist broadcasting in Gaussian channels: Achievable rate regions and coding,’’ IEEE Trans. Commun., vol. 64, no. 3, pp. 1016–1030, Mar. 2016.

未来网络通讯的频谱效率分类方案

王中生
新型网络与检测控制国家实验室
西安工业大学
中国陕西省西安市未央区学府中路 No.2 号

张庆松
新型网络与检测控制国家实验室
西安工业大学
中国陕西省西安市未央区学府中路 No.2 号

***摘要：*作为一个由 ISO/IEC 创建和管理的国际标准化项目，未来网络已产生了 ISO/IEC TR 29181 系列的技术报告，并正处于设置体系架构和协议的阶段。该项目以其独特的“全新设计”方式而闻名，并致力于基本的结构创新，以使得未来网络实现其所承诺的目标。同时，未来网络的标准化工作将为 SE 技术的未来突破做好准备，并制定计划，使未来网络适应快速变化的“后香农时代”技术革命。本标准参考射频频带分类机制，对 MCS 系统的频谱效率进行分类，便于对信息系统的频谱效率进行分类、讨论、评价和比较。**

关键词：未来网络通讯； 频谱效率分类方案； 国际标准

1. 介绍

本文是超限未来网络的第一项国际标准提案（草案），该提案（草案）的开发，依托于 JTC 1/SC 6/WG 7 工作团队，申请负责人是张庆松博士（南京博峰），提案的提议人是王中生教授（西安工业大学），开发过程中作为联络人的团队是 ITU-R 和 ICAO，同时联合 ECMA、3GPP 和 IEEE 三个组织进行合作开发。

本文按照以下三份文件的规定进行拟定：

ISO/IEC TR 29181：未来网络：问题陈述和要求，第 1-9 部分；

ITU-R SM.856-1：新频谱效率技术和系统（1992-1997）；

ITU-R SM.1046-3：无线电系统频谱使用效率的定义(2017-09-06)。

本文作为一份“工业、创新和基础建设”领域的国际标准提案（草案），提供了调制编码方案频谱效率 MCS SE（Modulation and Coding Scheme Spectral Efficiency）的分类方案，包括以下六个方面：

MCS SE 的定义；

MCS SE 的分类方法；

MCS SE 的命名体系；

SECS 推荐使用的示例；

对未来网络标准化的潜在影响；

频谱效率 SE 和频谱利用效率之间的不同点。

2. 开发背景

频谱效率是衡量信息通信系统发展水平的一个关键指标，它可以反应甚至影响诸多关键性能指标，包括频谱资源的有效利用率、信息传输率、信道容量等等。频谱效率越高，资源受限频谱的资源利用率就越高，同时传输速率越高，那么信息吞吐量也就越大。因此，提高频谱效率是信息通信技术（ICT）创新最关键的目标之一。

随着数字信息和计算机科学技术的发展，信息通信系统的频谱效率水平也在不断提高。20 年前的通信系统频谱效率还处于很低的水平，仅有 1 到 2 个比特。经过 20 年的发展，部分技术领域已经采用了 10 比特的频谱效率技术，更有一些领域已经将 12 比特频谱效率纳入下一代技术标准规划中。

但是，上述这一发展速度并不是直线上升的。根据历史经验和技术特点分析，由于多进制正交调幅（M-QAM）调制和反调制技术的先天限制，未来的信息和通讯系统的频谱效率水平的发展速度将放缓，预计在 20 年内达到 16 比特，45 年内达到 20 比特，而突破这一限制也是提高频谱效率的关键机制。与此同时，也不能排除由于信息通信物理基础层新理论和新技术的出现，而导致频谱效率的快速提高的可能性。

在后香农时代，高频谱效率将会成为信息通讯技术（ICT）水平的主要表现形式。在接下来的十到二十年内，对通信技术频谱效率的讨论和评价标准将超过 20 比特，达到几百比特甚至几千比特。通讯产品也将会促进频谱效率作为主要技术水平及服务能力的标志。

在现有的国际电信联盟（ITU）标准 RSM.1046-3 中规定了频谱利用效率的定义和使用各种系统频谱的评估方法，但没有提供调制编码方案（MCS）频谱效率的分类机制。而这种机制，对未来通信系统频谱效率的讨论、分析、评估、选择和管理是必需的。

例如，在一些技术或政策文件中，经常可以看到对“低频谱效率”和“高频谱效率”的讨论，但没有技术规范来定义和解释这两个概念。多少比特的频谱效率是“低频谱效率”？多少位的频谱效率是“高频谱效率”？相当一部分的划分方式中，将 10 比特系统称为“高频谱效率”，那么 16 比特、20 比特、32 比特甚至 128 比特的频谱效率如何归类呢？因此，仅仅只有“低频谱效率”和“高频谱效率”这两级分类标准是不能满足未来发展趋势和更精确的频谱效率分类的需要的。

在无线电频谱管理中，对频谱资源有多种分类方案，这里我们列举其中四种方案：①将频率资源(RE)划分为千赫兹、兆赫、千兆赫兹和太赫兹；②基于波长，将频率资源（RE）分为超长波、长波、中波、短波、超短波、微波等类别；③按频率分类的方法，即超低频（VLF）、低频、中频、中高频、高频、极高频（VHF）、超高频（UHF）等；④以英文字母为标志，将频谱资源 RE 分为 L-波段、S-波段、C-波段、X-波段、Ku-波段、K-波段、Ka-波段等。

本标准参考射频频带分类机制，对 MCS 系统的频谱效率进行分类，便于对信息系统的频谱效率进行分类、讨论、评价和比较。

3. 术语规定

3.1 缩写

表 1 术语缩写示例

缩写	全称	中文名称
MCS	Modulation and Coding Scheme	调制编码方案
MIMO	Multiple Input Multiple Output	多入多出技术
FN	Future Network	未来网络
OFDM	Orthogonal Frequency Division Multiplexing	正交频分复用
OCCS	Over-Capacity Communication Systems	超容量通讯系统
OVXDM	Overlapped X Domain Division Multiplexing	重叠 X 分域复用系统
OVTDM	Overlapped Time Domain Division Multiplexing	重叠时分域复用系统
QAM	Quadrature Amplitude Modulation	正交调幅
RE	Resource Element	资源元素
RSE	Relative Spectral Efficiency	相对频谱效率
SE	Spectral Efficiency	频谱效率
SEI	Spectral Efficiency Index	频谱效率指标
SUE	Spectrum Utilization Efficiency	频谱使用效率
TSEI	Typical SE Indicator	经典频谱效率指示器

3.2 术语定义

超容量通讯：信息交换的容量超过香农限制。

OVXDM：一种创新的调制编码方式，利用时间、频率、空间等多个域以及编码的重叠和多路复用，以达到更高的频谱效率，无编码开销，更高的编码增益和解码的低复杂度。

调制编码方案的频谱效率：根据通信系统的调制和编码方案，每一资源单元（RE 以 H 为单位）每秒发送的最大有用信息量。

频谱使用率：频率带宽、几何（地理）空间和时间的乘积，不包含其他潜在因子：$U = B \cdot S \cdot T$

香农限制：又称香农容量，在十九世纪四十年代由 Claude Shannon 定义，为单一信道在一定噪声水平下设定理论上最高的信息传输速率的极限。

未来网络：一个由 ISO/IEC 开发和管理的国际标准项目，它是基于全新的网络系统设计方法。出版物包括 ISO/IEC TR 29181 和 ISO/IEC 21558-21559。

4. MCS SE 分类原则

4.1 频谱效率（SE）和超高频谱效率（USE）的区分

本标准中定义的 SE（频谱效率）不应该与 ITU-RSM.1046-3 (2017)中的频谱效率混淆。下面的表格所列的是它们之间的区别：

表 2 本标准与 ITU 标准对比

	比较对象	本标准	ITU
1	来源	ISO/IEC	ITU-RSM.1046-3
2	术语	Spectral Efficiency	Spectrum Efficiency
3	缩写词	MCS SE	USE
4	因子	波特率 bps/Hz	U=B·S·T
5	考虑因素	● 容量 ● 资源元素 (赫兹 Hz) ● 时间 (秒)	● 带宽 ● 几何空间 (空间) ● 时间
6	改进方法	● 调制 ● 通道编码	● 天线指向性 ● 地理空间 ● 频率公用 ● 正交频率使用 ● 分时技术 ● 时间划分
7	SE 增益潜力	空域限制了 SE 增益	SE 增益是有限的
8	视角	通讯系统	用户
9	服务	对所有开放	拒绝其他的使用

本标准定义的 MCS 频谱效率是指通过调制和信道编码等技术手段，而获得的每赫兹频率资源每秒传输的有效信息位数。

MCS 的频谱效率值是相对固定的，只要知道所使用的调制机制和信道编码方法，就可以推导出 MCS 频谱效率在理论上的性能水平。由于调制机制和信道编码方法种类较少，于是一些

主流技术的应用非常广泛，如调制领域中的 M-QAM 技术以及信道编码领域中的 Turbo 码和 LDPC 码。因此，MCS 频谱效率可以作为衡量 ICT 在不同领域性能水平的一个通用且重要的指标。

4.2 确定 SE 的范围

目前，通信系统的 SE 不高于 10bps/Hz，一些系统可能在大约 5 年后达到 12bps/Hz。在如此低的 SE 比率下，没有必要对 SE 水平进行标准分类。

然而，人们期望这样的一个标准能够指向未来的趋势，为技术发展提供方向，并具有持续数十年的市场相关性。由于有技术趋势表明在频谱效率方面有潜在的突破，本标准考虑了在数百和数千 bps/Hz 范围内的频谱效率。

4.3 MCS SE 分类架构

MCS SE 分类系统包括下表中描述的三种方案：

表 3 MCS SE 分类系统描述方案

方案	特征	版本	目的
A	两个字母	#-SE	表示特定产品 SE 能力
B	三个字母	VSE	将 SE 按级别类别分组
C	四个字母	DDSE	提供一个可替代的、更简单的 SE 分类
D	二级	Lower	更广泛的级别

5. MCS 频谱效率分类方案

5.1 MCS SE 类 A：两个字母方案

两个字母的 MCS SE 分类系统只使用两个字母“SE”和数字显示特定的 bps/Hz. 它不是用于引用级别或类，而是用于指示产品的特定 SE 性能。

表达式描述:带“-”后跟“SE”的比特数(省略“s/Hz”)，表示“特定 bps/Hz 下的频谱效率”。

例如：

“56-SE”表示 56bps/Hz 的频谱效率。

“256-SE”表示 256bps/Hz 的频谱效率。

“1008-SE”表示 1008bps/Hz 的频谱效率。

5.2 MCS SE 类 B：三字母名称方案

表 4 MCS SE 分类方案中的三字母方案

SE 索引	**索引名**	**全称**	**SE 范围（bps/Hz)**	**TSEI* (bps/Hz)**
SEI 1	BSE	Basic Spectral Efficiency 基本频谱效率	0.1~2.0	2
SEI 2	LSE	Low Spectral Efficiency 低频谱效率	2.1~5.9	5
SEI 3	MSE	Medium Spectral Efficiency 中频谱效率	6~10.9	10
SEI 4	HSE	High Spectral Efficiency 高频谱效率	11~15	15
SEI 5	VSE	Very-High Spectral Efficiency	16~20	20

		极高频谱效率		
SEI 6	USE	Ultra-High Spectral Efficiency 超高频谱效率	21~32	32
SEI 7	SSE	Super Spectral Efficiency 超级频谱效率	33~64	64
SEI 8	OSE	One-hundred level spectral efficiency 100 级频谱效率	65~128	128
SEI 9	ESE	Extreme Spectral Efficiency 极大频谱效率	129~256	256
SEI 10	DSE	500 Spectral Efficiency 500 频谱效率	257~512	512
SEI 11	JSE	Jump Level spectral efficiency 跃迁能级频谱效率	513~999	768
SEI 12	1-KSE	1K Spectral efficiency 1000 频谱效率	1000~1999	1024
SEI 13	2-KSE	2K Spectral efficiency 2000 频谱效率	2000~2999	2048
SEI 14	3-KSE	3K Spectral efficiency 3000 频谱效率	3000~3999	3072
SEI 15	4-KSE	4K Spectral efficiency 4000 频谱效率	4000~4999	4096
SEI 16	XSE	X Spectral efficiency X 频谱效率	5000~6999	6144
*TSEI 该等级下的典型 SE 指标 .				

随着频谱效率的提高，三字母方案之间的差距也随之扩大。例如，在 SEI4 和 SEI5 中，只有 4 字节差将高电平和低电平分开。在 SEI9 中，间隙超过 200 字节，而在 SEI12 中，间隙增加到 1000 字节。

对于三个字母方案的上部，预计需要更加准确的 SE 参考资料或是比较方式。在这种情况下，当技术发展需要这样的改变时，三字母方案可以使用下列八条规则进行拓展：

规则 1. SEI1～6 不需要拓展；

规则 2. 拓展的分组被分为两张索引表，一张是如表 5 所示的 SE 低于 1000bps/Hz 的情况，另一张是如表 6 所示的 SE 高于 1000bps/Hz 的情况；

规则 3. 一个简单的两位十进制数被添加到索引名字中来表示拓展数字；

规则 4. 对于 SSE 和 OSE 索引，5bps/Hz 作为扩展单元的基础；

规则 5. 对于 ESE 和 DSE 索引，10bps/Hz 作为扩展单元的基础；

规则 6. 对于 JSE 索引，20bps/Hz 作为扩展单元的基础；

规则 7. 对于 KSE 索引，50bps/Hz 作为扩展单元的基础；

规则 8. 对于 XSE 索引，100bps/Hz 作为扩展单元的基础。

表 5 低于 1000SE 的拓展

SEI 7 EXT Index	33-64 SE RANGE	SEI 8 EXT Index	65~128 SE RANGE	SEI-9 EXT Index	129-256 SE RANGE	SEI 10 EXT Index	257~512 SE RANGE	SEI 11 EXT Index	513~999 SE RANGE
SSE 1	33-38	OSE 1	65-69	ESE 1	129-139	DSE 1	257-269	JSE 1	513-539
SSE 2	39-43	OSE 2	70-74	ESE 2	140-149	DSE 2	270-279	JSE 2	540-559
SSE 3	44-49	OSE 3	75-79	ESE 3	150-159	DSE 3	280-289	JSE 3	560-579
SSE 4	50-55	OSE 4	80-84	ESE 4	160-169	DSE 4	290-299	JSE 4	580-599
SSE 5	56-60	OSE 5	85-89	ESE 5	170-179	DSE 5	300-319	JSE 5	600-619
SSE 6	61-64	OSE 6	90-94	ESE 6	180-189	DSE 6	320-329	JSE 6	620-639
		OSE 7	95-99	ESE 7	190-199	DSE 7	330-339	JSE 7	640-659
		OSE 8	100-104	ESE 8	200-209	DSE 8	340-349	JSE 8	660-679
		OSE 9	105-109	ESE 9	210-219	DSE 9	350-359	JSE 9	680-699
		OSE 10	110-114	ESE 10	220-229	DSE 10	360-369	JSE 10	700-719
		OSE 11	115-119	ESE 11	230-239	DSE 11	370-379	JSE 11	720-739
		OSE 12	120-124	ESE 12	240-249	DSE 12	380-389	JSE 12	740-759
		OSE 13	125-128	ESE 13	250-256	DSE 13	390-399	JSE 13	760-779
						DSE 14	400-409	JSE 14	780-799
						DSE 15	410-419	JSE 15	800-819
						DSE 16	420-429	JSE 16	820-839
						DSE 17	430-439	JSE 17	840-859
						DSE 18	440-449	JSE 18	860-879
						DSE 19	450-459	JSE 19	880-899
						DSE 20	460-469	JSE 20	900-919
						DSE 21	470-479	JSE 21	920-939
						DSE 22	480-489	JSE 22	940-959
						DSE 23	490-499	JSE 23	960-979
						DSE 24	500-512	JSE 24	980-999

表 6 KSE 和 XSE 拓展索引

SEI 12 EXT Index	1000-1999 SE RANGE	SEI 13 EXT Index	2000-2999 SE RANGE	SEI 13 EXT Index	3000-3999 SE RANGE	SEI 14 EXT Index	4000-4999 SE RANGE	SEI 15 EXT Index	5000-6999 SE RANGE
1KSE 1	1000-1049	**2KSE 1**	2000-2049	**3KSE 1**	3000-3049	**4KSE 1**	4000-4049	**XSE 1**	5000-5099
1KSE 2	1050-1099	**2KSE 2**	2050-2099	**3KSE 2**	3050-3099	**4KSE 2**	4050-4099	**XSE 2**	5100-5199
1KSE 3	1100-1140	**2KSE 3**	2100-2140	**3KSE 3**	3100-3140	**4KSE 3**	4100-4140	**XSE 3**	5200-5299
1KSE 4	1150-1199	**2KSE 4**	2150-2199	**3KSE 4**	3150-3199	**4KSE 4**	4150-4199	**XSE 4**	5300-5399
1KSE 5	1200-1249	**2KSE 5**	2200-2249	**3KSE 5**	3200-3249	**4KSE 5**	4200-4249	**XSE 5**	5400-5499
1KSE 6	1250-1299	**2KSE 6**	2250-2299	**3KSE 6**	3250-3299	**4KSE 6**	4250-4299	**XSE 6**	5500-5599
1KSE 7	1300-1349	**2KSE 7**	2300-2349	**3KSE 7**	3300-3349	**4KSE 7**	4300-4349	**XSE 7**	5600-5699

1KSE 8	1350-1399	**2KSE 8**	2350-2399	**3KSE 8**	3350-3399	**4KSE 8**	3350-4399	**XSE 8**	5700-5799
1KSE 9	1400-1449	**2KSE 9**	2400-2449	**3KSE 9**	3400-3449	**4KSE 9**	4400-4449	**XSE 9**	5800-5899
1KSE 10	1450-1499	**2KSE 10**	2450-2499	**3KSE 10**	3450-3499	**4KSE 10**	4450-4499	**XSE 10**	5900-5999
1KSE 11	1500-1549	**2KSE 11**	2500-2549	**3KSE 11**	3500-3549	**4KSE 11**	4500-4549	**XSE 11**	6000-6099
1KSE 12	1550-1599	**2KSE 12**	2550-2599	**3KSE 12**	3550-3599	**4KSE 12**	4550-4599	**XSE 12**	6100-6199
1KSE 13	1600-1649	**2KSE 13**	2600-2649	**3KSE 13**	3600-3649	**4KSE 13**	4600-4649	**XSE 13**	6200-6299
1KSE 14	1650-1699	**2KSE 14**	2650-2699	**3KSE 14**	3650-3699	**4KSE 14**	4650-4699	**XSE 14**	6300-6399
1KSE 15	1700-1749	**2KSE 15**	2700-2749	**3KSE 15**	3700-3749	**4KSE 15**	4700-4749	**XSE 15**	6400-6499
1KSE 16	1750-1799	**2KSE 16**	2750-2799	**3KSE 16**	3750-3799	**4KSE 16**	4750-4799	**XSE 16**	6500-6599
1KSE 17	1800-1849	**2KSE 17**	2800-2849	**3KSE 17**	3800-3849	**4KSE 17**	4800-4849	**XSE 17**	6600-6699
1KSE 18	1850-1899	**2KSE 18**	2850-2899	**3KSE 18**	3850-3899	**4KSE 18**	4850-4899	**XSE 18**	6700-6799
1KSE 19	1900-1949	**2KSE 19**	2900-2949	**3KSE 19**	3900-3949	**4KSE 19**	4900-4949	**XSE 19**	6800-6899
1KSE 20	1950-1999	**2KSE 20**	2950-2999	**3KSE 20**	3950-3999	**4KSE 20**	4950-4999	**XSE 20**	6900-6999

5.3 MCS SE 类C：四个字母方案

表 7 MCS SE 分类方案中的四字母方案

	标题	完整标题	SE (bps/Hz)	相对与 B 类
1	SDSE	Single Digits Spectral Efficiency 单字节频谱效率	0-9	BSE，LSE，MSE
2	DDSE	Double Digits Spectral Efficiency 双字节频谱效率	10-99	HSE，VSE，USE，SSE，OSE
3	TDSE	Triple Digits Spectral Efficiency 三字节频谱效率	100-999	ESE，DSE，JSE
4	QDSE	Quadruple Digits Spectral Efficiency 四字节频谱效率	1000-9999	M-KSE，XSE

5.4 MCS SE 类D：比较方案

表 8 MCS SE 分类方案中的比较方案

	标题	完整标题	SE (bps/Hz)	与其他范畴的关系
1	L	低	无	所有低于特定类别的等级
2	H	高	无	所有高于特定类别的等级

6. 对本标准的描述

对本项标准内容的引用描述，可以参考以下形式。

6.1 参考标准

本标准的原名是“OCC-STD 21001”，由开发机构制定。在纳入中国行业标准、国家标准、ISO 标准、国际电联标准等其他标准体系时，

可对标题和编号进行重新分配。在此之前，“OCC-STD 21001”是 MCS SE 分类的唯一来源。

今后在参考分类方案时，建议在文件中注明 SE 分类在南京博丰通信技术有限公司开发的“OCC-STD 21001(2021)”中进行定义。

6.2 参考特定分类级别的示例

（1）在 2023 年，该公司有望交付使用创新调制方案的通信系统，可以提供“OCC-STD 21001”中定义的 VSE 级频谱效率。

（2）比较 MCS 的频谱效率，两个产品属于两代，A 样本仅在 VSE 水平，B 样本包含 USE 调制技术。

（3）行业共识是，KSE 级的频谱效率技术只需要几年的时间。

（4）一些专家预计，进入下一个十年，通信系统可以达到“OCC-STD 21001”四字母分类系统中定义的 QDSE 级频谱效率。

6.3 推荐使用的比较方案

当使用这两个表达方式时，它们应该伴随着一个具体的 SE 分类水平，以便于比较。

（1）到目前为止，最先进的无线通信系统的 MCS 频谱效率低于“OCC-STD 21001”中定义的 VSE 水平。

（2）预计到 2025 年，频谱效率高于 VSE 水平的产品将投入使用。

（3）新系统具有向后兼容设计，提供对 MCS SE 级的持续支持。

6.4 用以参考的专用 SE 比率

当参考专用 SE 比率时，下面两种情况为例：

（1）“Bofeng.com”提供两个无线电系统，按照“OCC-STD 21001”中定义的 SSE 级频谱效率运行。射频 A 系统具有 48-SE 调制方案，射频 B 系统具有 64-SE 调制能力

（2）产品规格:MCS SE:32-SE、48-SE 和 64-SE。说明系统包含三种类型的 MCS，支持三种 SE 速率。

7. 对未来网络标准化的潜在影响

未来网络是一个由 ISO/IEC 创建和管理的国际标准化项目。该项目已经产生了 ISO/IEC TR 29181 系列的技术报告，并正处于设置体系架构和协议的阶段。该项目以其独特的“全新设计”方式而闻名，并致力于基本的结构创新，以使得未来网络实现其所承诺的目标。

将 MCS-SE 分类系统作为未来网络标准化项目，对项目的建设有多方面的好处。首先，ISO/IEC 未来网络将成为第一个采用 MCS SE 分类系统的国际标准；第二，其他标准化机构可以采用该体系或对本标准进行规范引用；第三，ISO/IEC 未来网络成为第一个预示 MCS SE 未来发展趋势的标准；第四，未来网络标准中包含三比特和四比特的 MCS SE 将反映出网络表现和容量的巨大潜力；最后，该标准的成功采用将打开 ISO/IEC 未来网络标准的大门，以实现越来越高的 MCS SE 的未来技术。

ISO/IEC 未来网络应该为 SE 技术的未来突破做好准备，并制定计划，使未来网络适应快速变化的“后香农时代”技术革命。

参考文献

[1] T. M. Cover and J. A. Thomas, Elements of Information Theory. Hoboken, NJ, USA: Wiley, 2006.

[2] L. Daoben, Waveform Coding Theory of High Spectral Efficiency-OVTDM and Its Application. Beijing, China: Scientific, 2013.

[3] L. Daoben, ‘‘A novel high spectral efficiency waveform coding–OVFDM,’’ China Commun., vol. 12, no. 2, pp. 61–73, Feb. 2015.

[4] S. G. Wilson, Digital Modulation and Coding. Englewood Cliffs, NJ, USA: Prentice-Hall, 1996.

[5] L. Daoben, ‘‘A novel high spectral efficiency waveform coding-OVTDM,’’ Int. J. Wireless Commun. Mobile Comput., vol. 2, nos. 1–4, pp. 11–26, Dec. 2014.

[6] L. Daoben, Statistical Theory of Signal Detection and Estimation, 2nd ed. Beijing, China: Scientific, 2005

[7] J. G. Proakis, Digital Communications. New York, NY, USA: McGraw-Hill, 2001.

[8] G. J. Foschini, ‘‘Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas,’’ Bell Labs Tech. J., vol. 1, no. 2, pp. 41–59, 1996.

[9] G. J. Foschini and M. J. Gans, ‘‘On limits of wireless communications in a fading environment when using multiple antennas,’’ Wireless Pers. Commun., vol. 6, no. 3, pp. 311–335, Mar. 1998

[10] S. Wu, L. Kuang, Z. Ni, J. Lu, D. D. Huang, and Q. Guo, ‘‘Low-complexity iterative detection for large-scale multiuser MIMO-OFDM systems using approximate message passing,’’ IEEE J. Sel. Topics Signal Process., vol. 8, no. 5, pp. 902–915, Oct. 2014.

[11] N. Wu, W. Yuan, H. Wang, Q. Shi, and J. Kuang, ‘‘Frequency-domain iterative message passing receiver for faster-than-Nyquist signaling in doubly selective channels,’’ IEEE Wireless Commun. Lett., vol. 5, no. 6, pp. 584–587, Dec. 2016.

[12] J. CØspedes, P. M. Olmos, M. SÆnchez-FernÆndez, and F. Perez-Cruz, ‘‘Expectation propagation detection for high-order high-dimensional MIMO systems,’’ IEEE Trans. Commun., vol. 62, no. 8, pp. 2840–2849, Aug. 2014.

[13] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, ‘‘Millimeterwave massive MIMO: The next wireless revolution?’’ IEEE Commun. Mag., vol. 52, no. 9, pp. 56–62, Sep. 2014.

[14] [14] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, ‘‘Energy and spectral efficiency of very large multiuser MIMO systems,’’ IEEE Trans. Commun., vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[15] Y. S. Cho, J. Kim, W. Y. Yang, and C. G. Kang, MIMO-OFDM Wireless Communication Technology With MATLAB. Beijing, China: PublishingHouse of Electronics Industry, 2013.

[16] Y. D. Zhang, M. G. Amin, and B. Himed, ‘‘Altitude estimation of maneuvering targets in MIMO over-the-horizon radar,’’ in Proc. IEEE 7th IEEE Sensor Array Multichannel Signal Process. Workshop (SAM), Jun. 2012, pp. 257–260

[17] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, ‘‘Massive MIMO for next generation wireless systems,’’ IEEE Commun. Mag., vol. 52, no. 2, pp. 186–195, Feb. 2014.

[18] U. Gustavsson et al., ‘‘On the impact of hardware impairments on massive MIMO,’’ in Proc. IEEE Global Telecommun. Conf. Workshops (GC Wkshps), Austin, TX, USA, Dec. 2014, pp. 294–300.

[19] E. Björnson, M. Matthaiou, and M. Debbah, ‘‘Massive MIMO with nonideal arbitrary arrays: Hardware scaling laws and circuit-aware design,’’ IEEE Trans. Wireless Commun., vol. 14, no. 8, pp. 4353–4368, Aug. 2015.

[20] J. E. Mazo and H. J. Landau, ‘‘On the minimum distance problem for faster-than-Nyquist signaling,’’ IEEE Trans. Inf. Theory, vol. 34, no. 6, pp. 1420–1427, Nov. 1988.

[21] F. Rusek and J. B. Anderson, ‘‘CTH04-1: On information rates for faster than Nyquist signaling,’’ in Proc. IEEE GLOBECOM, Nov./Dec. 2006, pp. 1–5.

[22] F. Rusek and J. B. Anderson, ‘‘Multistream faster than Nyquist signaling,’’ IEEE Trans. Commun., vol. 57, no. 5, pp. 1329–1340, May 2009.

[23] J. B. Anderson, F. Rusek, and V. Öwall, ‘‘Faster-than-Nyquist signaling,’’ Proc. IEEE, vol. 101, no. 8, pp. 1817–1830, Aug. 2013.

[24] A. Prlja and J. B. Anderson, ‘‘Reduced-complexity receivers for strongly narrowband intersymbol interference introduced by faster-than-Nyquist signaling,’’ IEEE Trans. Commun., vol. 60, no. 9, pp. 2591–2601, Sep. 2012.

[25] S. Sugiura, ‘‘Frequency-domain equalization of faster-than-Nyquist signaling,’’ IEEE Wireless Commun. Lett., vol. 2, no. 5, pp. 555–558, Oct. 2013.

[26] J. Fan, S. Guo, X. Zhou, Y. Ren, G. Y. Li, and X. Chen, ‘‘Faster-thanNyquist signaling: An overview,’’ IEEE Access, vol. 5, pp. 1925–1940,2017.

[27] K. Takeuchi, M. Vehkapera, T. Tanaka, and R. R. Muller, ‘‘Large-system analysis of joint channel and data estimation for MIMO DS-CDMA systems,’’ IEEE Trans. Inf. Theory, vol. 58, no. 3, pp. 1385–1412, Mar. 2012.

[28] D. Dasalukunte, V. Öwall, F. Rusek, and J. B. Anderson, Faster than Nyquist Signaling: Algorithms to Silicon. Dordrecht, The Netherlands: Springer, 2014.

[29] E. Bedeer, M. H. Ahmed, and H. Yanikomeroglu, ‘‘A very low complexity successive symbol-by-symbol sequence estimator for faster-than-Nyquist signaling,’’ IEEE Access, vol. 5, pp. 7414–7422, 2017.

[30] A. D. Liveris and C. N. Georghiades, ‘‘Exploiting faster-than-Nyquist signaling,’’ IEEE Trans. Commun., vol. 51, no. 9, pp. 1502–1511, Sep. 2003.

[31] Y. J. D. Kim and J. Bajcsy, ‘‘Iterative receiver for faster-than-Nyquist broadcasting,’’ Electron. Lett., vol. 48, no. 24, pp. 1561–1562, Nov. 2012.

[32] Y. J. D. Kim, J. Bajcsy, and D. Vargas, ‘‘Faster-than-Nyquist broadcasting in Gaussian channels: Achievable rate regions and coding,’’ IEEE Trans. Commun., vol. 64, no. 3, pp. 1016–1030, Mar. 2016.

Research on the Estimation of Gaze Location for Head-eye Coordination Movement

Qiyou Wu
School of Computer Science and Engineering Xi'an Technological University
Xi'an, China
E-mail: wu314650592@163.com

Changyuan Wang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: Cyw901@163.com

***Abstract*—Sight is the main source for humans to obtain information from the outside world. Due to the structure of the human eye [1], the range of human sight is limited. For this reason, people need to constantly move their line of sight when observing the surrounding environment and the target, and the movement of the sight is based on the coordinated movement of the head and the eye[2]. Therefore, the key issue for gaze research is how to correctly establish the relationship between head-eye movement and gaze movement. Taking the simulated flight environment as the research background, this paper collects a large number of head-eye images through the designed "three-camera and eight-light source" head-eye data acquisition platform, and proposes a gaze estimation method based on the combination of appearance and features, which effectively combines The relationship of head-eye coordination movement. Then, the ResNet-18 deep residual network structure and the traditional BP neural network structure are used to complete the effective fusion of the head pose and human eye features in the process of capturing the sight target, so as to realize the accurate estimation of the sight drop point, and its average accuracy up to 89.9%.**

Keywords-Head-Eye Coordination; Gaze Estimation Method；Experimental Platform Design; Deep Residual Networks

1. INTRODUCTION

Head-eye coordination is the process of coordinating and combining head and eye movements and synthesizing a unified action to complete the shift of sight to the target [3]. In the research of visual impact point estimation, it is necessary to establish the relationship between head movement, eye movement and gaze movement in order to obtain a more accurate gaze point location. Therefore, how to establish the head-eye-line of sight relationship is the key issue of research.

For the research on the relationship between head-eye-line of sight, the initial method is to limit the free movement of the human head and only track the eye movement [4], so that a slight movement of the head will cause a large systematic error, and it is not suitable for in practical application scenarios. For this reason, the mechanism of head-eye movement has become a hot research topic at that time. In 2008, Freedman [5] used physiological methods to study the relationship between eye movement and head-eye movement in rhesus monkeys, which are similar to human head-eye movement mechanisms. The experimental results show the relationship between the head-eye movement and the line of sight: when the target appears in a larger field of view, the eyes will first move towards the target before the head, and then the head starts to move in the same direction. Due to the fast movement of the eyes, the sight can quickly complete the target acquisition and stop moving. However, the head movement was relatively slow, and Ren did not stop moving in the target direction. At this time, under the action of the vestibular function, the tendency of the eyes to move in the opposite direction at a certain speed is used as a vestibulo-ocular reflex (VOR) [6] to compensate for the head movement, so as to ensure that the target exists stably in the line of sight. Inside. It can be seen that although the contribution of head movement to the target capture process is small, it also directly affects the direction of sight movement.

Gaze estimation is a study of the subject's current gaze direction or gaze location using existing detection technologies such as mechanical, electronic, and optical [7]. The early research on eye sight estimation benefited from the development of medicine and psychology, and researchers recorded the relevant information of eye movement by direct observation. Based on the different devices used for gaze estimation, gaze estimation research can be divided into wearable and non-wearable [8]. With the development of sight estimation technology, wearable sight estimation methods such as contact lens and electrooculography (EOG) have appeared. Although the influence of head movement is reduced, it is more disturbing to the subjects. , not suitable for long-term wear. With the development of image processing and computer vision technology, the advantages of video-based line-of-sight estimation methods are convenient and non-wearable. However, for gaze estimation research, how to efficiently integrate head motion data still needs further research.

In recent years, under the research upsurge of deep neural network, new progress has been made in the research of gaze estimation algorithm based on head-eye data fusion, which can be mainly divided into gaze target estimation, gaze location estimation and gaze direction estimation. In 2016, Recasens [9] et al. designed a deep neural network model composed of two branches, which were used to extract the head pose and gaze direction of human images respectively, and to estimate the gaze target by judging the saliency of the target; Kyle Krafka [10] et al. took mobile phones and tablet computers as the research objects, and designed a deep neural network composed of four branches, respectively inputting left and right eye images, face images and face positions, and realized a two-dimensional plane. line-of-sight estimate. The idea of sharing and processing the parameter weights of the left and right eye image branches in this study has been used for reference by many subsequent studies; in 2019, the Google [11] team further improved the above model and changed the line of sight estimation model to three The branch and the coordinate positions of the four corners of the eyes are used to replace the face image and the face position, and finally a good estimation effect of the line of sight is achieved; for the estimation of the line of sight, it is usually represented by a direction vector formed by two horizontal and vertical angles. , Zhang [12]'s research in 2015, spliced the head pose data of the input image with the eye features, and used a shallow network structure similar to LeNet [13] to estimate the line of sight direction. The way of data fusion has greatly inspired the follow-up research in this paper.

2. MATERIALS

1) Experiment preparation)

The data collection experiment in this paper recruited 8 male graduate student volunteers as subjects, aged 23-30 years old, in good health and with good eyesight. Before the experiment, each subject was familiar with the specific content and precautions of the experiment, and they all participated in the experiment voluntarily. Each subject signed a written commitment and informed letter to ensure the legitimacy of the experiment in this paper. In order to exclude external interference, after the preparation for the experiment, each experiment was completed by only one subject alone.

The experimental equipment includes a DELL computer, an inertial sensor (MTI-G-700), three industrial cameras, three high-definition displays (resolution 2560× 1440), and eight infrared point light sources. Among them, the inertial sensor was worn to about the position of the occipital bone behind the subject's head to measure the Euler angles (Pitch, Roll, Yaw) of the subject's head posture when capturing the target. Since the MTI-G-700 inertial sensor is only used to measure the change of the head posture in this paper, it will be referred to as the head posture instrument in the following. In this study, three Point Grey GS3-U3-41C6NIR-C industrial cameras were selected, and the resolution of the collected images was 2048×2048 and the chromaticity was near-infrared (NIR). Three cameras were installed above the three high-definition monitors, and were used to simultaneously capture the head motion images

and eye motion images captured by the subjects during the experiment.

For the data collection experiment of head-eye coordination movement, this study innovatively built a non-wearable sight-drop data collection platform of "three eyes and eight light sources". This platform not only expands the subject's head movement range, but also effectively reduces the impact of changes in lighting conditions. It is mainly composed of three industrial cameras mounted on three monitors and eight near-infrared light sources evenly distributed on the border of the monitors, which are used to record the head-eye images of the subjects when the target is captured. The schematic diagram of the platform deployment is shown in Figure 1.

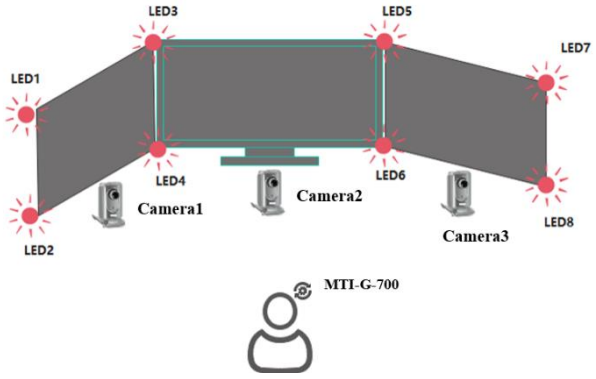


Figure 1. Deployment of the "Three Cameras and Eight Lights" platform

2) Maintaining the Integrity of the Specifications

The head-eye movement data collection experiment designed in this study mainly refers to the process of the subjects performing visual interaction with the randomly appearing objects on the three screens through coordinated head-eye movement. Before starting the experiment, the infrared light source, camera, head attitude meter and other equipment should be calibrated to ensure that each equipment is in normal operation. The subjects were required to wear the head posture meter, and adjust the horizontal distance between the sitting position and the middle screen to save about 60cm to ensure that they were within the best focal length of the three cameras. At the beginning of the experiment, the subjects' eyes need to face the center of the middle screen, and press the record button to calibrate the initial Euler angle of the head posture. After the calibration is successful, the target to be captured appears randomly on the three screens in the form of a red circle with a radius of 30 pixels. The subject uses the head-eye movement to aim at the target, and press the record button to complete the target capture process. . During the experiment, there are no other requirements for the subjects, and the head can move freely in a large range. After the experiment, the program will record the Euler angle of the head pose, head image, eye image and the coordinates of the center point of the target each time the subject captures the target, and set it as a set of data. The experimental process is shown in Figure 2.

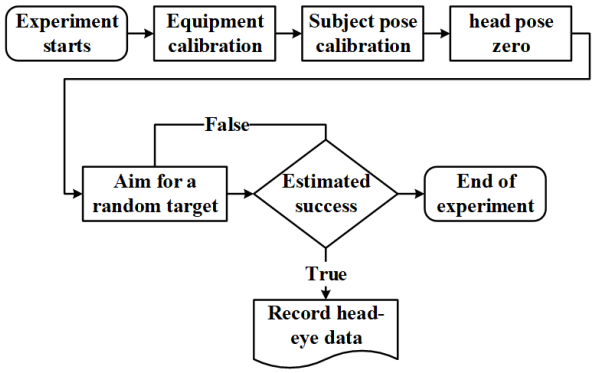


Figure 2. Flow chart of head-eye movement data collection

Considering that the experimental operation is relatively simple, in order to ensure the experimental status of the subjects and the quality of the experimental data, the duration of a single experiment is set to 20 minutes in this study. During the experiment, the equipment was deployed on a six-axis full-motion simulated flight platform, as shown in Figure 3. Due to the long-term use of the camera, the performance will be effectively degraded, and there may be cases of missed shots, so simple manual screening is required after each experiment. Finally, after screening unqualified samples, a total of 31507 groups of head-eye movement data were collected in this paper.

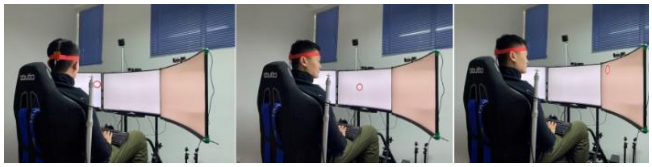


Figure 3. The experimental process of head-eye movement data collection

3. METHODS

In this study, the appearance-based line of sight estimation method is used to estimate the location of the subject's line of sight through head-eye images captured by a non-wearable multi-eye camera. Although the appearance-based line-of-sight estimation method has strong robustness, it also has some problems, such as a great restriction on the free movement of the subject's head and a great influence by the change of lighting conditions. The "three eyes and eight light sources" platform built in this paper can not only expand the subject's head movement range through the strapdown of three monitors; Feature points are added to the external image to reduce the influence of lighting. For this reason, this research uses the method of image processing and feature extraction, fuses head features and eye features, establishes a neural network model for line of sight estimation, and then realizes the research of line-of-sight drop estimation.

1) Head feature extraction

In this paper, the posture measuring instrument was worn on the back of the subject's head and used as the three-axis reference point for head movement. The head attitude data mainly includes: pitch angle (Pitch), yaw angle (Yaw), roll angle (Roll), namely looking up, shaking head and turning head, through these three Euler angles, the head position can be estimated more accurately. space pose. In this study, a right-handed Cartesian coordinate system is used, and the three-axis positions of X, Y, and Z in space and the corresponding Euler angles of the head posture are shown in Figure 4.

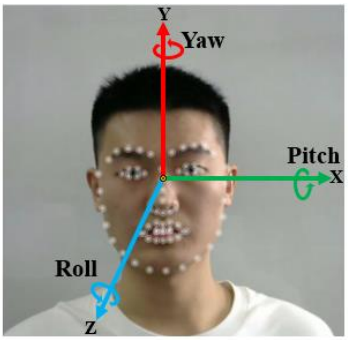


Figure 4. Euler angle of head posture

In order to facilitate the statistics and analysis of the head pose data, this study visualized the recorded data. The three-axis pose data of the head is shown in Figure 5. Through the three-axis Euler angle deflection angle, it can be intuitively observed that the head yaw angle (Yaw) and the pitch angle (Pitch) change greatly during the target acquisition process, while the roll angle (Roll) changes less.

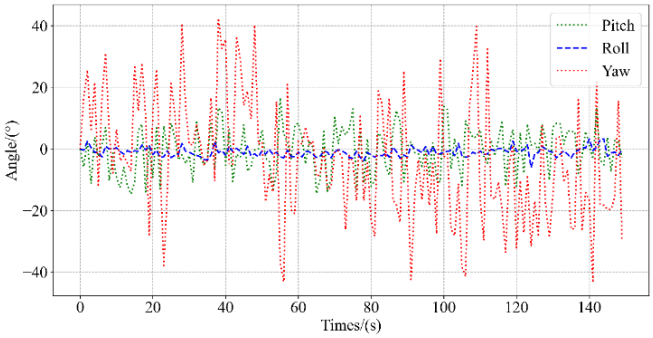


Figure 5. Three-axis attitude data record

2) Eye feature extraction

The human eye detection method used in this study mainly uses the AdaBoost cascade classifier combined with Haar-like features to first detect the face area, and then detects and intercepts the human eye area in the face area. Based on the pre-trained classifiers for faces, eyes, etc. included in OpenCV, this study carried out face detection and eye detection on the front view image collected by the camera corresponding to the target appearing screen. The recognition results are shown in Figure 6. According to the results of human eye detection, the monocular area of the subject is intercepted at a resolution of 64×64, and the left and right eye images are obtained as the input of the next convolutional neural network model.

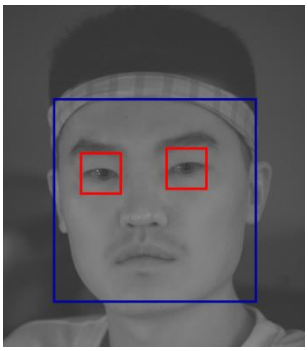


Figure 6. Face detection, Eye detection results

Aiming at the requirement of lighting conditions in appearance-based visual estimation method, this study combines the idea of feature-based visual estimation method, and places eight near-infrared point light sources equidistantly on the boundary of three screens. The Purkinje formed by the reflection are used as the feature points of the eye image. Since the camera is a near-infrared camera, the brightness of the Purkinje formed by the reflection of the infrared point light source through the corner of the eye is not affected by external light. To sum up the above assumptions, this study performed threshold processing on the intercepted left and right eye images before training the line of sight placement model, and obtained left and right eye images with more obvious Purkinje spots, which were used as the control group input by the convolutional neural network model. Its influence on the estimation result of the line-of-sight landing point.

In this study, the three monitors (resolution: 2560×1440) are numbered in the order of left (0), middle (1), and right (2). degree. This study takes the five target points on the left (0) screen as an example for analysis, and the specific positions are shown in Figure 7. Among them, the first two coordinates of each point are the position of the target center pixel on the screen, and the third coordinate is the screen number.

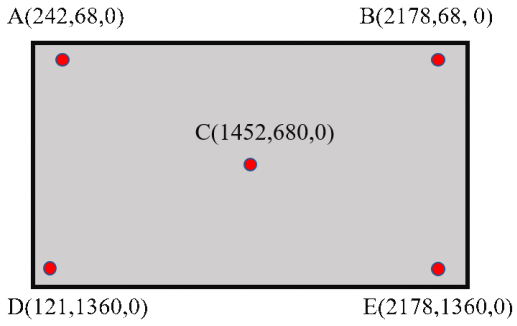


Figure 7. Location of the target center point

The eye images collected at the five points A, B, C, D, and E in the above figure are processed in the order of binarization thresholding, truncation thresholding, and super-thresholding zero. The detection effect is shown in Figure 8. It can be observed from the figure that for different fixation points, the number and positional relationship of Purkinje spots formed by the subjects' eyes are different. Treatment can detect 6-7 Purkinje. Among them, the Purkinje after the truncated thresholding process is more obvious, which can effectively eliminate the reflected light spots on the cornea of other external light sources, and retain the original eye image. Therefore, this paper will use the truncation thresholding method to process the cropped left and right eye images, and detect the left and right eye images with obvious Purkinje as the input of the next convolutional neural network model.

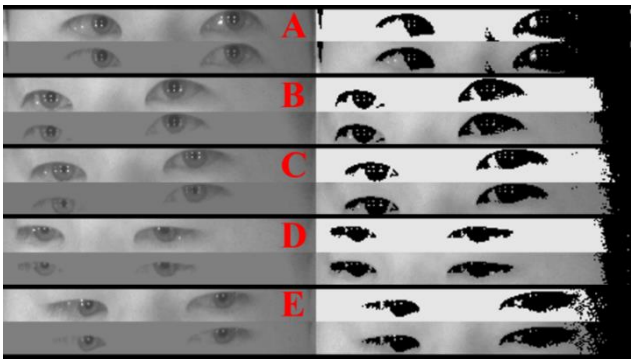


Figure 8. Purkinje detect results

For the deep convolutional neural network model, this paper refers to the deep residual network structure proposed by researchers such as He in 2015. At present, several commonly used ResNet networks mainly include: ResNet-18, ResNet-34, ResNet-50 and other variants. Although increasing the depth of the network can improve the accuracy of the model, the shallower residual network (ResNet-18) also has good accuracy in practical applications, and its model is small, which provides faster convergence speed and facilitates parameter optimization. Moreover, based on the short-circuit operation of the ResNet model, the combination of features of different resolutions can be realized, and it has a better feature extraction effect for the eye image input in this paper. Therefore, in this study, ResNet-18 is used as the estimation model of the human eye line of sight, and the network structure of ResNet-18 is shown in Table 1. By comparing different types of binocular image inputs (with and without thresholding to detect Purkinje), analyze the accuracy of the output on the screen where the sight falls. Since traditional residual neural networks are mostly used for classification tasks, this paper is inspired by the Google team's

research on line-of-sight drop estimation in 2019. Multiple fully-connected layers are connected after the hidden layer of the neural network to return the line-of-sight drop coordinates.

For the estimation model of human eye gaze point in this paper, the input is the grayscale images of the left and right eyes when the subject is facing the capture target in the middle of the screen. Due to the low resolution requirement of the eye image, this paper adopts the resolution size of 64×64 to capture the monocular image. Compared with the input size of the traditional ResNet-18 network RGB image (224×224×3), the image input in this paper is smaller (64×64×1), which improves the computational speed of the model. This model is divided into two branches with the same structure. The input layer is the cropped left and right eye images, and the main structure of the hidden layer of each branch is built according to the ResNet-18 network structure. It consists of 17 convolutional layers, 8 residual blocks and 2 pooling layers. Relu is used as the activation function of all convolutional layers to achieve feature extraction for left and right eye images. Finally, the left and right eye images are extracted through 4 fully connected modules. The feature maps of the eyes are fused and the estimated line-of-sight coordinates G(x, y, n) are output. Figure 9 shows the structure of the network model for the estimation of the human eye gaze point in this paper.

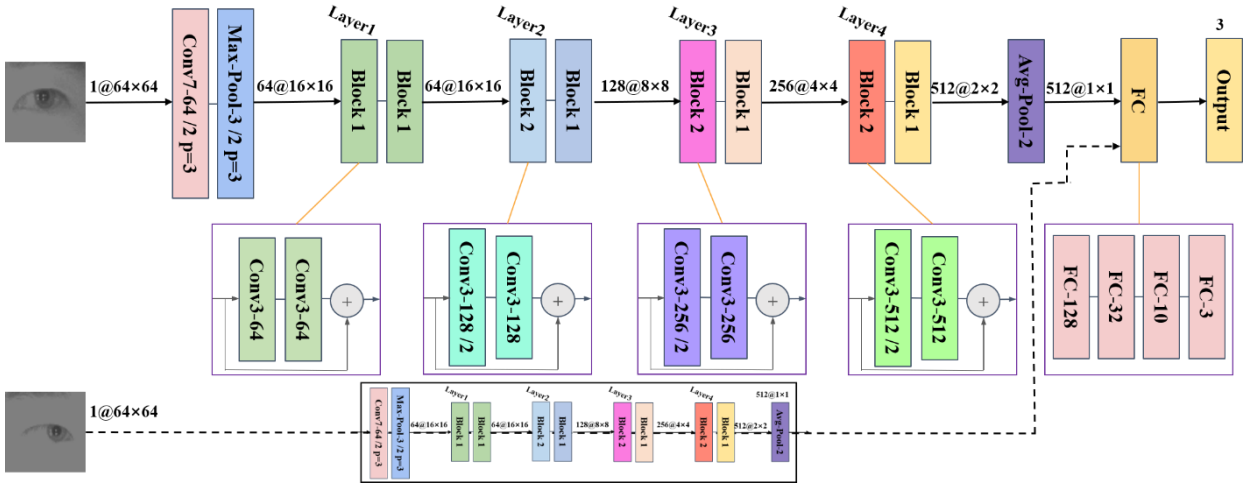


Figure 9. Eye gaze point estimation network structure

3) Eye feature extraction

In order to fuse the head pose data in the line-of-sight drop estimation, this study draws on the idea of using the traditional BP neural network, and uses multiple fully connected layers as the network branch of the head pose data feature extraction, mainly including: an input layer and a Hidden layer composition. Among them, the input layer is the subject's head posture Euler angles (Roll, Pitch, Yaw); the hidden layer is composed of three fully connected layers, the number of neuron nodes is 100, 16, 16 respectively, and Relu is used as the activation function uses the feature vector extracted by the last fully connected layer as the output.

Based on the principle of feature layer fusion, this research first preprocesses the head-eye data to complete feature extraction. Image feature extraction; for the head image, the Euler angle of the head pose is used to output the feature vector with the same dimension as the eye feature through the head feature extraction network to complete the dimension registration. The features of the two parts are fused through multiple fully connected layers to form a line-of-sight estimation model structure fused with head-eye movements. The network structure of the line of sight estimation model in this study includes three branches. The input layer inputs the left and right eye images and the Euler angle of the head pose when the subject is capturing the target. For the left and right eye branches, three fully connected

layers are used for feature extraction, and the number of neuron nodes is 128, 32, and 16 respectively; for the head pose branch, a head feature extraction network is used, and finally two neurons are used for feature extraction. The fully connected modules with the number of nodes are 16 and 3 to complete the feature fusion of the data of the three branches, and realize the regression of the landing point of the three screens, as shown in Figure 10.

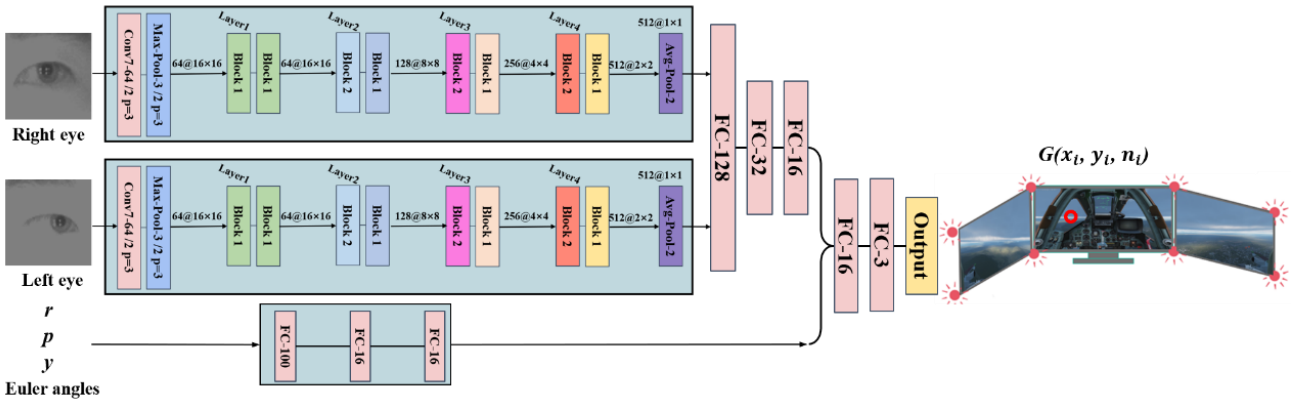


Figure 10. Gaze estimation model with head-eye movement fusion

4. RESULT

In order to improve the estimation performance of the human eye gaze point estimation model in the learning process, this paper uses the mean squared error (MSE) as the loss function. MSE represents the mean value of the sum of squares of the point errors corresponding to the predicted data and the original data:

$$MSE = \frac{1}{n}\sum_{i=1}^{m}\omega_i(y_i - \hat{y}_i)^2 \qquad (1)$$

where n is the number of samples, y_i is the original data, and $\hat{y}_i$ is the predicted data. When the MSE value is closer to 0, it indicates that the fitting ability of the model is stronger, and the estimation of the line of sight is more accurate. Based on the three-screen experimental platform in this paper, each screen is a two-dimensional plane, and the Euclidean distance is used to calculate the difference between the calibration point and the estimated point. The predicted receptive field is a circular area with the calibration point as the center and a radius of 30 pixels. In the training process of the model, this paper uses adaptive moment estimation (Adam) as the optimizer, which can adjust different learning rates for different parameters; the learning rate of the network is set to 10^{-3}, and the batch size is set is 32, and the training epoch is 100.

In this study, through the screening of the original data, after removing the images of the subjects with eyes closed, 30,000 images were selected from a total of 30,000 images and cropped to a size suitable for the model input, of which 70% were used for model training and 30% were used for model training. Performance Testing. At the same time, two different line-of-sight estimation models were constructed and used as a control experiment to compare and analyze the final prediction accuracy and other performances based on whether or not Purkinje detection was performed before inputting the original data.

The experimental results show that when only the original eye image is used as the model input, the model (Eye) needs to extract fewer features and the convergence speed is faster. At about 200 epochs, the model basically converges, and its average accuracy can reach 85.6%; when the input is the eye image after Purkinje detection, although the model (Eye & Purkinje) has a slower convergence speed, it is basically at about 400 epochs. Convergence, but its average accuracy can reach 87.7%.

By comparing the performances of the two models, the (Eye & Purkinje) model makes the Purkinje patch features more obvious through

thresholding before input, which increases the complexity of feature extraction in the hidden layer, thereby increasing the convergence time of the model, but it is relatively slow compared to the Eye model. The average accuracy of line-of-sight location estimation is increased by 2.1%, and the loss curve is relatively stable, and the stability of the model is better. Therefore, in this study, the (Eye & Purkinje) model is selected as the model for estimating the human eye gaze, and it is verified that the input of eye images with significant Purkinje spots can add feature points to the image, reduce the influence of lighting conditions, and improve the estimation accuracy of the model. Accuracy. The accuracy and loss curves of the training of the two models are shown in Figure 11.

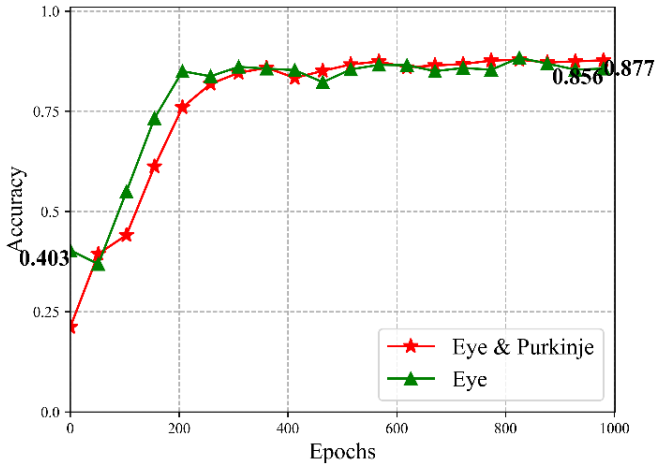


Figure 11. Comparison of the performance of different models

In order to evaluate the performance of the gaze point estimation model fused with head-eye motion, this paper compares it with two models that only use eye images. As shown in Figure 12, in the first 200 epochs, the performance of the model after adding head pose is better than that of the (Eye & Purkinje) model, but due to the need to fuse head-eye features, parameter optimization takes a long time, and the accuracy is not as good as the Eye model. From the analysis of the convergence speed of the model, although the (Eye & Purkinje & Head) model converges slowly, the model tends to be stable after 600 epochs, and the accuracy of the model is high. Enter the model, which compresses and correlates head-eye coordination motion data by fusing multi-dimensional head-eye data, and its accuracy is improved by up to 4.3%.

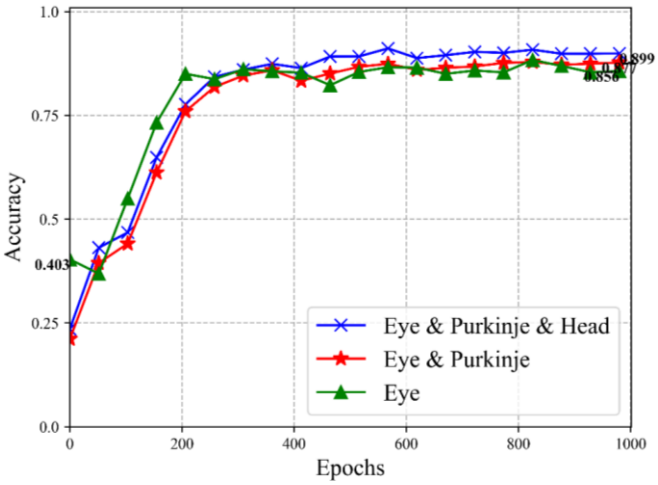


Figure 12. Comparison of the performance of three gaze estimation models

5. Conclusion

For the research on gaze point estimation, this paper uses three models for comparative analysis, namely, the gaze point estimation model using only eye images, the gaze point estimation model using eye features, and the fusion head-eye motion feature. Line-of-sight estimation model. By comparison, the line of sight estimation model (EPH) that fuses head motion and eye feature point images has a high test accuracy for the prediction results of the test set, and the estimated line of sight is basically within the prediction receptive field of the target to be captured. And there is no over-fitting phenomenon, and the average accuracy of the line of sight estimation can reach 89.9%. However, this paper also finds that the accuracy of the general estimation of these three models is not high. This problem is a common problem in the estimation method of line-of-sight placement based on appearance, which needs to be further studied and prospected.

To sum up, this paper is based on a line-of-sight estimation method that combines appearance and features, which effectively integrates the head motion and eye motion when the line of sight moves. Accurate estimation of the line-of-sight placement in a two-dimensional screen is achieved. A practical and effective research method is put forward for the estimation of sight drop point.

REFERENCES

[1] Atchison D A, Smith G, Smith G. Optics of the human eye[M]. Oxford: Butterworth-Heinemann, 2000.

[2] Wang Changyuan, Li Jingjing, Jia Hongbo, et al. Research methods and progress of head-eye movement[J]. Journal of Xi'an University of Technology, 2012, 32(3): 173-182.

[3] Mao Xiaobo. Research on Modeling and Control of Bionic Robot Eye Movement System[D]. Zhengzhou: Zhengzhou University, 2011.

[4] Lei Zhihui, Yu Qifeng. A new method to determine eye movement translation[J]. Experimental Mechanics, 2003, 18(4): 564-568.

[5] Freedman E G. Coordination of the eyes and head during visual orienting[J]. Experimental brain research, 2008, 190(4): 369-387.

[6] Mao Xiaobo, Chen Tiejun. A bionic model of head-eye coordination motion control[J]. Journal of Biomedical Engineering, 2011, 28(5): 895-900.

[7] Liu Jiahui, Chi Jiannan, Yin Yixin. Review of feature-based gaze tracking methods [J]. Journal of Automation, 2021, 47(2): 252-277.

[8] Zhang C, Chi J N, Zhang Z H, et al. Gaze estimation in a gaze tracking system[J]. Science China Information Sciences, 2011, 54(11): 2295-2306.

[9] Recasens ,A R C. Where are they looking?[D]. Massachusetts Institute of Technology, 2016.

[10] Krafka K, Khosla A, Kellnhofer P, et al. Eye tracking for everyone[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2176-2184.

[11] He J, Pham K, Valliappan N, et al. On-device few-shot personalization for real-time gaze estimation[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019: 0-0.

[12] Zhang X, Sugano Y, Fritz M, et al. Appearance-based gaze estimation in the wild[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 4511-4520.

[13] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

基于头眼协调运动的视线落点估计研究

吴其右
西安工业大学计算机科学与工程学院
中国，西安
E-mail: wu314650592@163.com

王长元
西安工业大学计算机科学与工程学院
中国，西安
E-mail: Cyw901@163.com

***摘要：*视线是人类获取外界信息的主要来源。由于人眼的构造[1]，人的视线范围是有限的。为此，人在观察周围环境与目标时需要不断地移动视线，而视线的移动又是基于头眼协调运动同时构成的，即当人大范围转移视线时，需要依靠头部的运动以扩大视野范围[2]。因此，对于视线研究的关键问题是如何正确地建立头眼运动与视线移动的关系。本文以模拟飞行环境为研究背景，通过所设计的“三目八光源”头眼数据采集平台采集大量的头眼图像，并提出一种基于外观与特征相结合的视线估计方法，有效地结合了头眼协调运动的关系。随后，利用 ResNet-18 深度残差网络结构与传统的 BP 神经网络结构完成对视线目标捕获过程中的头部姿态与人眼特征的有效融合，实现对视线落点的精确估计，其平均准确度可达 89.9%。**

关键字：头眼协调；实验平台设计；深度残差网络；视线估计方法

1. 介绍

头眼协调运动是协调组合头部、眼部运动并合成统一动作从而完成视线转移向目标的过程[3]。在视觉落点估计的研究中，需要建立头部运动、眼部运动和视线移动的关系，才能得到较为准确的视线落点位置。因此，如何建立头眼-视线关系是研究的关键问题。

对于头眼-视线关系的研究，最初的方式是通过限制人的头部自由运动，仅对眼部运动进行跟踪[4]，使得头部稍有移动便造成很大的系统误差，且不适应于实际应用场景。为此，头眼运动机制成为了当时研究的热点。2008 年，Freedman E G[5]使用生理学方法对与人类头眼运动机制相似的猕猴作为研究对象，测得视线移动与头眼运动的关系。其实验结果表明了头眼运动与视线的关系：当目标出现于较大视野范围中时，双眼会率先于头部朝目标进行移动，随后头部开始同向移动。由于双眼运动速度较快，视线可快速地完成目标捕获并停止运动。但头部运动较为缓慢，任未停止向目标方向移动。此时，在前庭功能的作用下，趋势双眼以一定的速度反向移动，作为补偿头部运动的前庭动眼反射(vestibulo-ocular reflex, VOR)[6]，保证目标稳定地存在于视线范围内。由此可得，虽然头部运动对目标捕获过程的贡献较小，但其也直接影响了视线移动的方向。

而视线估计是利用机械、电子、光学等现有检测技术对受试者当前视线方向或视线落点的研究[7]。早期的视线估计研究得益于医学与心理学的发展，研究者多以直接观察的方式记录眼睛运动的相关信息。基于视线估计所使用设备的不同，可将视线估计研究划分为穿戴式与非穿戴式[8]。随着视线估计技术的发展，出现了如接触镜(Contact Lens)和眼电图(EOG)等穿戴式的视线估计方法，虽然减小了头部运动的影响，但对受试者干扰较大，不适于较长时间的佩戴。伴随图像处理与计算机视觉技术的发展，使得基于视频的视线估计方法便捷、非穿戴式的优势逐渐显现出来，并在医疗诊断、辅助驾驶和人机交互等多个领域得到普及与应用。但对于视线估计研究而言，如何高效地融入头部运动数据仍然有待进一步的研究。

近些年，在深度神经网络的研究热潮下，基于头眼数据融合的视线估计算法研究有了新的进展，主要可分为对注视目标的估计、视线落点估计与视线方向估计。2016 年，Recasens[9]等人设计了一个由两支路组成的深度神经网络模型，分别用于提取人物图像的头部姿态与注视方向，通过目标显著性判断实现对注视目标

的估计；Kyle Krafka[10]等人以手机和平板电脑等设备为研究对象，设计了由四个支路构成的深度神经网络，分别输入左右眼图、人脸图像和人脸位置，实现了于二维平面的视线落点估计。此研究对左右眼图像支路的参数权值进行共享处理的思想，受到此后很多研究的借鉴；2019 年，Google[11]团队对上述模型做了进一步改进，将视线落点估计模型改为三支路并由四个眼角的坐标位置代替人脸图像与人脸位置，最终取得了不错的视线落点估计效果；对于视线方向的估计，通常由水平和竖直两个角度所构成方向向量表示，Zhang[12]于 2015 年的研究中，将输入图像的头部姿态数据与眼部特征进行拼接，使用类似于 LeNet[13]的浅层网络结构实现对视线方向的估计，此研究对头眼数据融合的方式对本文后续研究起到了很大启发。

2. 实验

1) 实验准备

本文的数据采集实验招募了 8 名男性研究生志愿者作为被试，年龄在 23-30 岁之间，身体健康，视力良好。在实验前，每位受试者已熟悉实验的具体内容和注意事项，且均为自愿参与实验。每位被试都签署了书面的承诺与知情书，保证本文实验的合法性。为排除外界干扰，在做好实验准备后，每次实验仅由一名被试单独完成。

实验设备包括一台 DELL 计算机、一部惯性传感器(MTI-G-700)、三部工业相机、三台高清显示屏(分辨率 2560×1440)和八个红外点光源。其中，惯性传感器佩戴至被试头部后方大约枕骨位置，用于测量被试在进行目标捕获时的头部姿态欧拉角(Pitch、Roll、Yaw)。由于在本文中 MTI-G-700 惯性传感器仅用于测量头部姿态的变化，下文将其简称为头部姿态仪。本研究选用了三部 Point Grey GS3-U3-41C6NIR-C 工业相机，其采集图像的分辨率为 2048×2048，色度为近红外光谱(NIR)。三部相机分别安装在三台高清显示器的上方，用于同时拍摄被试在实验过程中进行目标捕获的头部运动图像和眼部运动图像。

对于头眼协调运动的数据采集实验，本研究创新性地搭建了“三目八光源”非穿戴式的视线落点数据采集平台。此平台不仅扩大了被试的头部运动范围，还能有效减小光照条件变化的影响。其主要由三部挂载在三台显示器上的工业相机和八个均匀分布于显示器边界的近红外光源组成，用于记录被试在目标捕获时的头眼图像，平台部署示意图如图 1 所示。

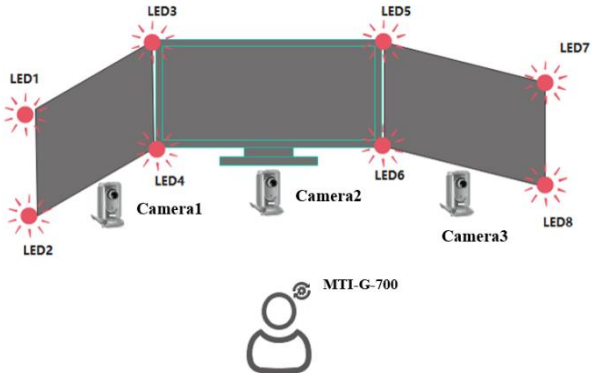


图 1 “三目八光源”平台部署示意图"

2) 实验设计

本研究设计的头眼运动数据采集实验主要是指：被试对三块屏幕上随机出现的目标通过头眼协调运动进行视觉交互的过程。开始实验前，要对红外光源、相机、头部姿态仪等设备进行校准，确保各设备出于正常运行状态。被试需佩戴好将头部姿态仪，调整坐位与中间屏幕的水平距离保存 60cm 左右，以确保自身出于三台相机的最佳焦距范围内。实验开始时，被试双眼需要正视中间屏幕的中心位置，按下记录键来标定头部姿态的初始欧拉角。标定成功后，待捕获目标以半径为 30 像素的红圈形式随机出现于三块屏幕上，被试通过头眼运动将视线落点瞄准在目标上，并按下记录键完成一次目标的捕获流程。实验过程中对被试无其他的要求，头部可以较大范围的自由移动。实验结束后，程序将记录每次被试在捕获目标时的头部姿态欧拉角、头部图像、眼部图像和目标中心点的坐标，并设置为一组数据。实验流程如图 2 所示。

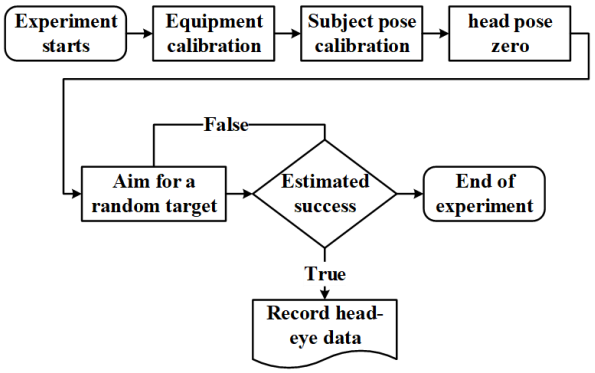


图 2 头眼运动数据采集流程图

考虑到实验操作较为单一，为确保被试的实验状态，保证实验数据质量，本研究设置单次实验时长为 20 分钟。实验时设备部署于六轴全动模拟飞行平台上，如图 3 所示。由于长时间使用相机性能会有效下降，存在多拍漏拍的情况，每次实验后需进行简单人工筛选。最终，通过对不合格样本的筛选后，本文共采集了 31507 组头眼运动数据。

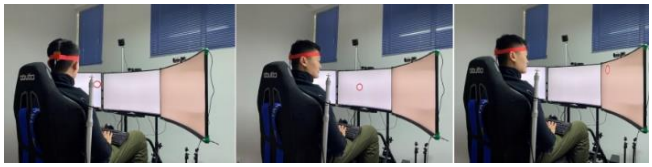


图 3 头眼运动数据采集实验过程

3. 方法

本研究采用基于外观的视线估计方法，通过非穿戴式多目相机所拍摄的头眼图像，实现对被试的视线落点估计。虽然基于外观的视线估计方法有较强的鲁棒性，但也存在着对被试头部自由运动限制较大、受光照条件变化影响较大等问题。而本文搭建的“三目八光源”平台通过三部显示器的捷联，不仅可以扩大被试的头部运动范围；并通过红外光源在人眼角膜上的反射光斑（普尔钦斑），为眼部图像增加了特征点，减小光照的影响。为此，本研究通过图像处理与特征提取的方法，融合头部特征与眼部特征，建立视线估计神经网络模型，进而实现视线落点估计的研究。

1) 头部特征提取

本文将姿态测量仪佩戴至被试的后脑勺位置，并将其作为头部运动的三轴基准点。头部姿态数据主要包括：俯仰角(Pitch)、偏航角(Yaw)、滚转角(Roll)，即抬头、摇头和转头，通过这三个欧拉角可以较为准确地估计出头部的空间姿态。本研究采用右手笛卡尔坐标系，其空间 X、Y、Z 三轴位置与对应的头部姿态欧拉角如图 4 所示。

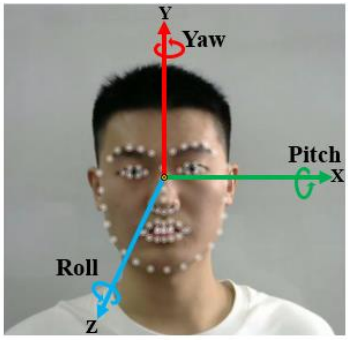


图 4 头部姿态欧拉角

为了便于对头部姿态数据的统计和分析，本研究对所记录数据进行了可视化处理，头部的三轴姿态数据如图 5 所示。通过三轴欧拉角偏转角度，可直观地观察出头部偏航角(Yaw)和俯仰角(Pitch)在目标捕获过程中的变化幅度较大，而滚动角(Roll)的变化幅度较小。

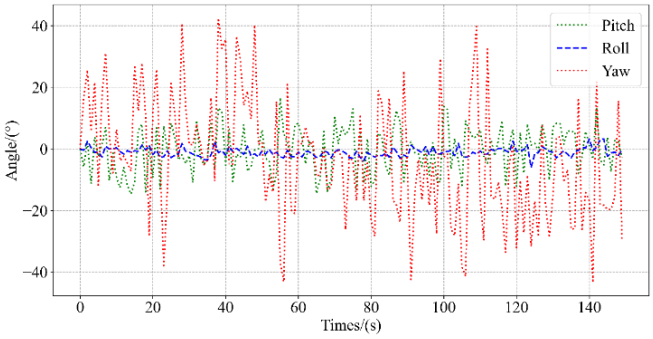


图 5 三轴姿态数据记录

2) 眼部特征提取

本研究使用的人眼检测方法主要是利用 AdaBoost 级联分类器结合 Haar-like 特征先检测出人脸区域，再在人脸区域中检测出人眼区域并进行截取。基于 OpenCV 所包含针对面部、眼部等进行过预训练的分类器，本研究对

目标出现屏幕对应相机所采集的正视图像进行了人脸检测、人眼检测，识别结果如图 6 所示。通过人眼检测结果，对被试的单眼区域按 64×64 的分辨率进行截取，得到左右眼部图像作为下一步卷积神经网络模型的输入。

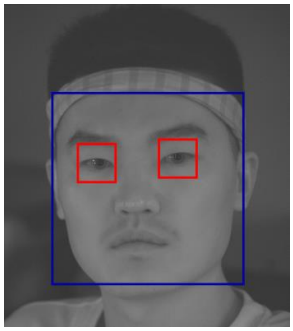


图 6 人脸检测、人眼检测结果

针对基于外观的视觉估计方法所存在对光照条件的要求，本研究结合基于特征的视觉估计方法的思想，在三个屏幕的边界等距地安放了八个近红外点光源，通过其在角膜的反射所形成的普尔钦斑作为眼部图片的特征点。由于相机属于近红外相机，因此红外点光源经眼角反射所形成普尔钦斑的亮度不受外界光照影响。综上述设想，本研究在对视线落点模型训练前，对所截取的左右眼图像进行阈值处理，得到含有较为明显普尔钦斑的左右眼图像并作为卷积神经网络模型输入的对照组，对比其对视线落点估计结果的影响。

本研究对三台显示器(分辨率：2560×1440)按左(0)、中(1)、右(2)的顺序编号，为了更加直观地对比几种阈值处理的效果与普尔钦斑的明显程度。本研究以左(0)屏的五个目标点位为例进行分析，具体位置如图 7 所示。其中，每个点前两个坐标为目标中心像素点于屏幕上的位置，第三个坐标为屏幕号。

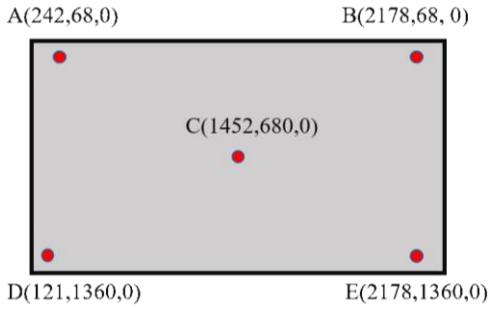


图 7 目标中心点位置

对上图 A、B、C、D、E 五点的所采集的眼部图像按照二值化阈值、截断阈值化和超阈值化零的顺序进行处理，对比几种阈值处理方法对普尔钦斑的检测效果，其结果如图 8 所示。由图中可观察出，对于不同的注视点，被试眼部所形成的普尔钦斑的数量与位置关系都有所不同，当注视左屏(0)屏幕中心位置时(C)，通过阈值处理可检测出 6-7 个普尔钦斑。其中，截断阈值化处理后的普尔钦斑较为明显，能有效地消除了其他外界光源的在角膜的反射光斑，保留了原始眼部图像。因此，本文将采用截断阈值化处理法对裁剪后的左右眼图像进行处理，检测含有明显普尔钦斑的左右眼图像作为下一步卷积神经网络模型的输入。

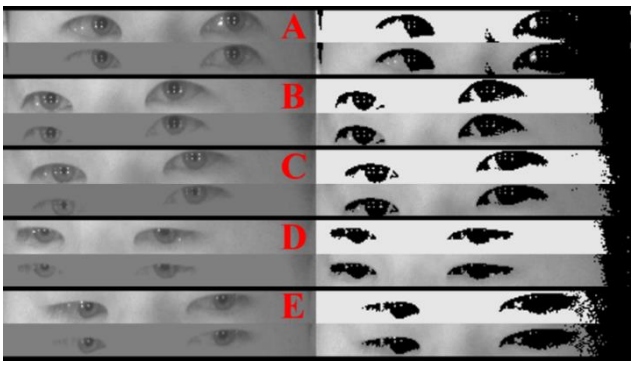


图 8 普尔钦斑检测结果

对于深度卷积神经网络模型，本文参考了 2015 年由何凯明等研究者提出的深度残差网络结构。目前，常用的几种 ResNet 网络主要包括：ResNet-18、ResNet-34、ResNet-50 以及其他变种，虽然增加网络的深度可以提升模型的准确率，但较为浅层的残差网络(ResNet-18)在实际应用中同样有良好的准确性，同时其模型较小，提供了更快的收敛速度，便于参数的优化。而且基于 ResNet 模型短接的操作，可实现对不同分辨率特征的组合，对于本文输入的眼部图像有较好的特征提取效果。因此，本研究以 ResNet-18 作为人眼视线落点估计模型，ResNet-18 网络结构如表 1 所示。通过对比不同类型的双眼图像输入（有无通过阈值处理检测普尔钦斑），分析其输出位于屏幕上视线落点的精度。由于传统的残差神经网络多用于分类任务，本文受 2019 年 Google 团队对视线落

点估计的研究启发，在神经网络的隐含层后连接多个全连接层，用于回归出视线落点坐标。

本文的人眼视线落点估计模型，输入为被试正视屏幕中带捕获目标时的左、右眼部灰度图像。由于眼部图像的分辨率要求较低，本文采用 64×64 的分辨率大小截取单眼图像。相对于传统 ResNet-18 网络 RGB 图片的输入大小(224×224×3)，本文的图像输入较小(64×64×1)，使得模型的计算速度提高。而本模型分为两条结构相同的支路，其输入层为裁剪后的左、右眼图图像，而每条支路隐含层的主要结构是按着 ResNet-18 网络结构搭建的，主要由 17 层卷积层、8 个残差块构成和 2 个池化层构成，使用 Relu 作为所有卷积层的激活函数，实现对左右眼图像的特征提取，最后通过 4 个全连接模块将左右眼的特征图进行融合并输出所估计的视线落点坐标 G(P_x, P_y, S_n)，本文的人眼视线落点估计网络模型结构如图 9 所示。

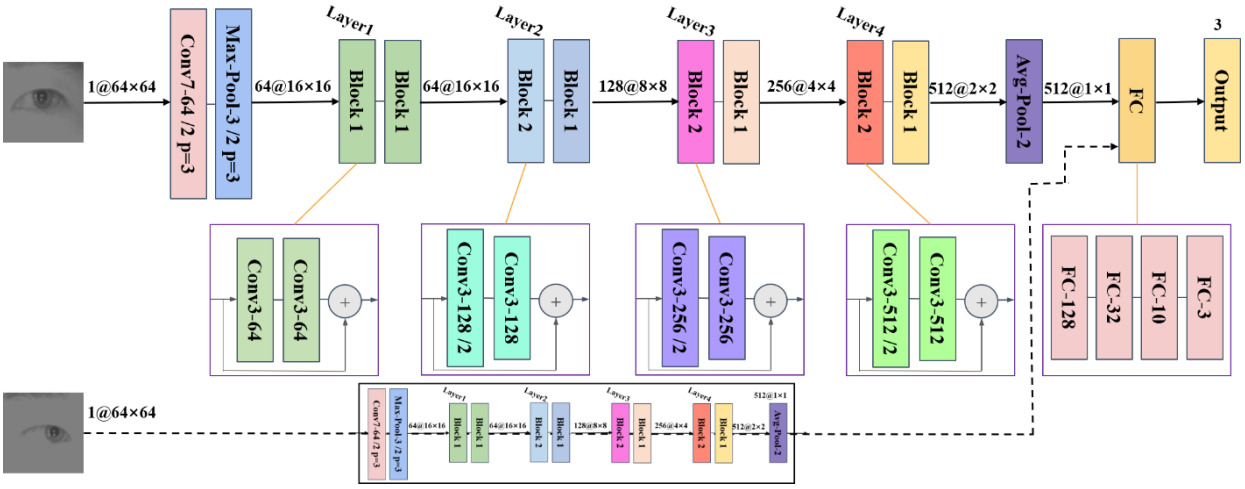


图 9 人眼视线落点估计网络结构

3) 融合头眼的视线落点估计

为了在视线落点估计中融合头部姿态数据，本研究借鉴采用传统的 BP 神经网络的思想，通过多个全连接层作为头部姿态数据特征提取的网络分支，主要包括：一个输入层和一个隐含层构成。其中，输入层为被试的头部姿态欧拉角(Roll, Pitch, Yaw)；隐含层由三个全连接层组成，神经元节点个数分别为 100，16，16，并使用 Relu 作为激活函数，将最后一层全连接层所提取的特征向量作为输出。

基于特征层融合的原理，本研究先对头眼数据进行预处理以完成特征提取，对于眼部图像由上文中所提及的眼部视线落点估计模型的输入层和隐含层实现左右眼部图像特征的提取；对于头部图像则是由头部姿态欧拉角，通过头部特征提取网络输出与眼部特征维度相同的特征向量，完成维度配准。两部分特征通过多个全连接层实现特征融合，构成融合头眼运动的视线落点估计模型结构。本研究的视线落点估计模型的网络结构共包括三个支路，输入层分别输入被试在捕获目标时的左右眼图像与头部姿态欧拉角。对于左右眼支路，使用三个全连接层进行特征提取，神经元节点个数分别为 128、32、16；对于头部姿态支路，采用头部特征提取网络，最后通过由两个神经元节点个数为 16 和 3 的全连接模块完成对三条支路数据的特征融合，实现对三屏实现落点的回归，如图 10 所示。

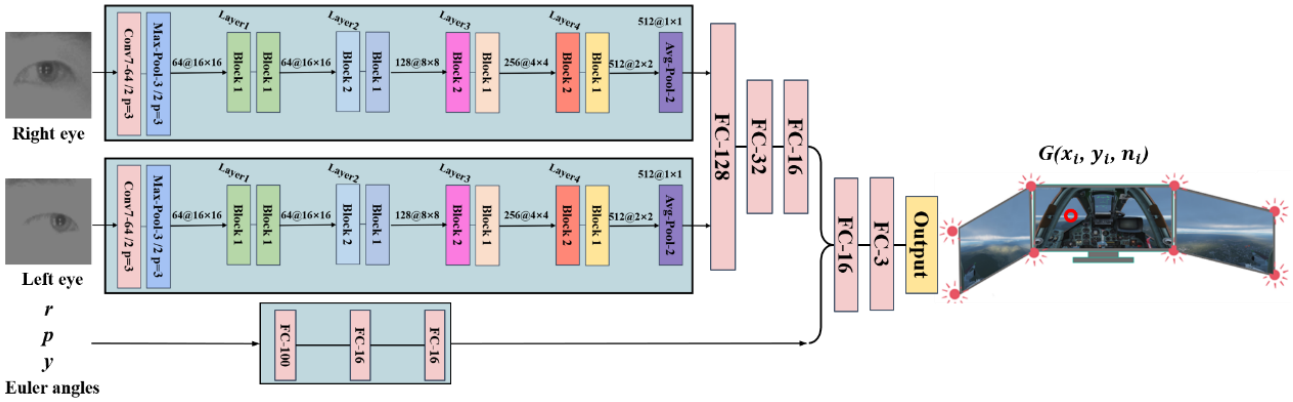


图 10 融合头眼运动的视线落点估计模型

4. 结果

为了使人眼视线落点估计模型在学习中朝向更精确的估计性能改进，本文使用均方方差(MSE)作为损失函数，MSE 表示预测数据与原始数据所对应点误差的平方和均值：

$$MSE = \frac{1}{n}\sum_{i=1}^{m}\omega_i(y_i - \hat{y}_i)^2 \qquad (1)$$

其中，n是样本的个数，y_i是原始数据，$\hat{y}_i$是预测数据。当 MSE 值越接近 0 时，说明模型的拟合能力越强，其视线落点估计也越准确。基于本文的三屏实验平台，每个屏幕为一个二维平面，采用欧式距离计算标定点与估计点间的差值，预测接受域为以标定点为圆心，半径 30 像素的圆形区域。在模型的训练过程中，本文使用自适应矩估计(Adam)作为优化器，其能对不同的参数调整不同的学习率；网络的学习率设定为10^{-3}，批处理量设置为 32，训练周期为 100。

本研究通过对原始数据的筛选，去除被试闭眼的图像后，从中共选取了 30000 张图像并将它们裁剪为适合模型输入的尺寸，其中 70%用于模型训练，30%用于模型的性能测试。同时，以输入原始数据前是否进行普尔钦斑检测的两种情况，构成两种不同的视线落点估计模型并作为对照实验，对比分析最终的预测准确率和其他性能。

实验结果表明，当仅使用原始眼部图像作为模型输入时，模型(Eye)所需提取的特征较少，收敛速度较快。在 200 个 epoch 左右模型基本收敛，其平均准确度可达到 85.6%；而当输入为普尔钦斑检测后的眼部图像时，虽然模型(Eye&Purkinje)的收敛速度较慢，在 400 个 epoch 左右基本收敛，但是其平均准确率可达到 87.7%。通过对比两种模型的性能，Eye&Purkinje 模型在输入前通过阈值处理使普尔钦斑特征更加明显，增加了隐含层特征提取的复杂度，从而增加了模型的收敛时间，但其相对于 Eye 模型的视线落点估计平均准确率提升了 2.1%，且损失曲线较为平稳，模型的稳定性更好。因此，本研究的选用 Eye&Purkinje 模型作为人眼视线落点估计的模型，并验证了输入含有显著普尔钦斑的眼部图像，可为图像增加特征点，减小光照条件的影响，提高模型估计的准确率。两种模型训练的准确率和损失变化曲线，如图 11 所示。

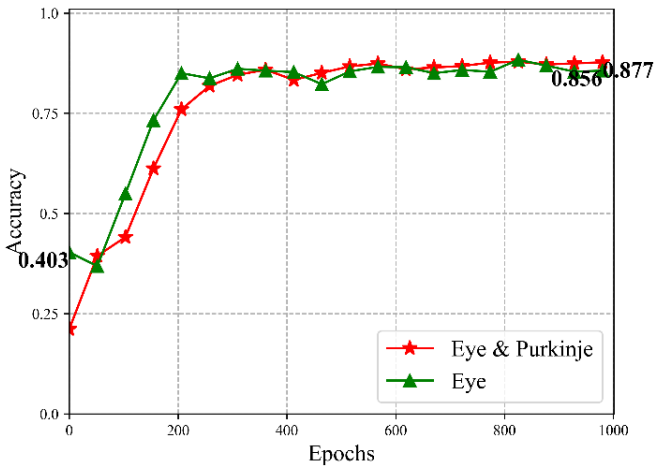


图 11 Eye 和 Eye&Purkinje 模型性能对比分析图

为了评价融合头眼运动的视线落点估计模型的性能，本文将其与仅使用眼部图像的两种模型进行了对比分析。如图 12 所示，在前 200 个 epoch 中，加入头部姿态后的模型性能已优于 Eye&Purkinje 模型，但由于需要融合头眼特征，参数优化所需时间较长，准确率不如 Eye 模型。从模型的收敛速度分析，虽然 Eye&Purkinje&Head 模型收敛较慢，但在 600 个 epoch 后模型趋于平稳，模型的精度较高，其平均准确率可达 89.9%，相对于仅使用单维度的眼部图像输入模型，该模型通过融合头眼多维度的数据，实现对头眼协调运动数据的压缩与关联，其准确率最多提升了 4.3%。

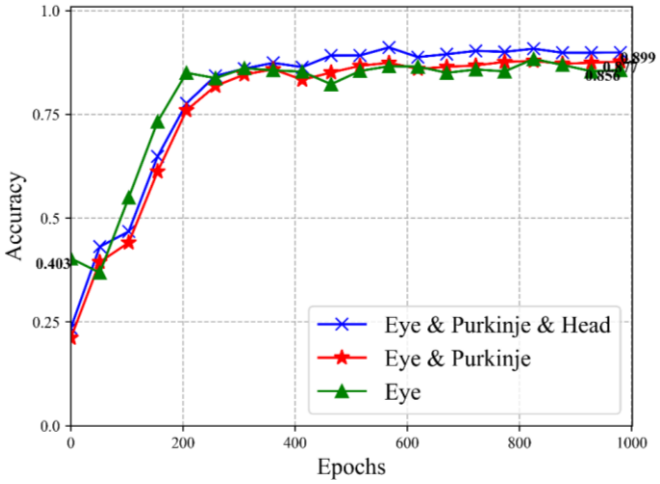


图 12 三种视线落点估计模型性能对比分析图

5. 总结

对于视线落点估计的研究，本文分别采用了三种模型进行了对比分析，即仅使用眼部图像的视线落点估计模型、使用眼部特征的视线落点估计模型与融合头眼运动特征的视线落点估计模型。通过对比，融合头部运动与眼部特征点图像的视线落点估计模型(EPH)对测试集的预测结果有较高的测试准确度，所估计的视线落地基本在待捕获目标的预测接受域内且未出现过拟合的现象，其对视线落点估计的平均准确率可达 89.9%。但本文也发现这三个模型普遍估计的精度不高，该问题是基于外观的视线落点估计方法存在的普遍问题，有待后续进一步的研究与展望。

综上，本文基于一种外观与特征相结合的视线估计方法，有效地融合了视线移动时的头部运动与眼部运动，将头部姿态数据与眼部特征通过深度卷积神经网络结构最终实现了对二维屏幕中视线落点的精确估计。为视线落点估计研究提出了一种切实有效的研究方法。

参考文献

[14] Atchison D A, Smith G, Smith G. Optics of the human eye[M]. Oxford: Butterworth-Heinemann, 2000.

[15] Wang Changyuan, Li Jingjing, Jia Hongbo, et al. Research methods and progress of head-eye movement[J]. Journal of Xi'an University of Technology, 2012, 32(3): 173-182.

[16] Mao Xiaobo. Research on Modeling and Control of Bionic Robot Eye Movement System[D]. Zhengzhou: Zhengzhou University, 2011.

[17] Lei Zhihui, Yu Qifeng. A new method to determine eye movement translation[J]. Experimental Mechanics, 2003, 18(4): 564-568.

[18] Freedman E G. Coordination of the eyes and head during visual orienting[J]. Experimental brain research, 2008, 190(4): 369-387.

[19] Mao Xiaobo, Chen Tiejun. A bionic model of head-eye coordination motion control[J]. Journal of Biomedical Engineering, 2011, 28(5): 895-900.

[20] Liu Jiahui, Chi Jiannan, Yin Yixin. Review of feature-based gaze tracking methods [J]. Journal of Automation, 2021, 47(2): 252-277.

[21] Zhang C, Chi J N, Zhang Z H, et al. Gaze estimation in a gaze tracking system[J]. Science China Information Sciences, 2011, 54(11): 2295-2306.

[22] Recasens ,A R C. Where are they looking?[D]. Massachusetts Institute of Technology, 2016.

[23] Krafka K, Khosla A, Kellnhofer P, et al. Eye tracking for everyone[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2176-2184.

[24] He J, Pham K, Valliappan N, et al. On-device few-shot personalization for real-time gaze estimation[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019: 0-0.

[25] Zhang X, Sugano Y, Fritz M, et al. Appearance-based gaze estimation in the wild[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 4511-4520.

[26] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

Research on Real-Time Fusion Technology of Range Telemetry Data

Hanghang Zhou
School of Computer and Engineering
Xi'an Technological University
Xi'an, China

Yuchao She
Information Technology Center of
Xi'an Technological University
Xi'an, China

Xiaofeng Rong
School of Computer and Engineering
Xi'an Technological University
Xi'an, China
E-mail: xiaofengrong@126.com

Linjuan Fan
School of Computer and Engineering
Xi'an Technological University
Xi'an, China

Fangyuan Ma
China Institute of Experimental Testing and
Research for the Weapons Industry
Weinan, China

Mingjie Zhao
School of Computer and Engineering
Xi'an Technological University
Xi'an, China

Qianshi Yan
School of Computer and Engineering
Xi'an Technological University
Xi'an, China

***Abstract*—Telemetry data is the important data for the ground station to obtain the working status and environmental parameters of the aircraft system. Its fusion processing is the key technology to select the best selection of multiple channels of data and improve the reliability and accuracy of the entire recording. Due to the large amount of telemetry data, there is a delay error in the transmission process, and the phenomenon of frame loss and code error is accompanied by severe challenges for the alignment and optimization of data fusion. The research starts from the application background of fusion technology, introduces and analyzes the characteristics of telemetry data and the difficulties of fusion technology; outlines the current alignment and optimization related research results and development process, according to engineering requirements and technical points, from real-time and after-event From the perspective, the alignment algorithms and quality evaluation algorithms involved are classified and analyzed in detail; finally, the shortcomings of the existing methods are summarized, and the future development direction is looked forward to provide references for related researchers.**

Keywords-Telemetry Data; Data Fusion; Alignment; Optimization

I. INTRODUCTION

The telemetry system is an important part of modern aircraft and aviation weapon launch tests [1]. Over the years, with the continuous running-in and improvement of test tasks, a set of telemetry data processing procedures based on multi-station measurement and control has been formed [5]. As a key technology in the processing process, data

DOI: 10.2478/ijanmc-2022-0010

fusion mainly selects the best multi-channel data to improve the reliability and accuracy of the whole process [3, 4].

The traditional method of fusion technology is to manually count the errors of a piece of data after the fact, and select the best for splicing [19]. Based on the development of telemetry/computer systems, in 2009, the automatic processing of fusion technology use the "three judgments and two principles", which is a byte-by-byte comparison algorithm for full-frame data (F-frame). Meanwhile, it also exposed a series of problems such as transmission delay, frame loss, and bit errors in the processing process. Establish a standard and unified data quantification method based on the literature [17]. In 2014, literature [13] used theoretical ballistics to characterize the data transmission delay, and divided the fusion technology into two parts: alignment and optimization. Since then, various algorithms in the field of fusion technology have been proposed one after another [7-17], aiming to overcome the difficulties encountered in practical tasks and to develop in a more versatile, efficient and precise direction.

At present, many results have been achieved in the research of fusion technology, but the research content is relatively scattered and fragmented. Therefore, the research starts from the two parts of alignment and optimization involved in data fusion, summarizes the current research status of real-time and after-the-fact fusion processing methods, and elaborates the key algorithms involved. Strive to provide feasible research ideas for the development of data fusion processing technology.

II. TELEMETRY DATA

Telemetry data is usually measured using the PCM system, and the sampling period is milliseconds. The parameters in the same cycle are collected as a frame of PCM data stream, collected again after a certain time interval, and cyclically form the final PCM telemetry data stream, which is recorded in a binary data stream file, and the amount of data is relatively huge.

The processing flow of telemetry data is shown in Figure 1. The Ground Stations (GS) will send the received ciphertext data (C-data) to the command center in real time. Firstly, Passing the decryption pre-processing equipment (DPE) to complete the data encryption independently in parallel, and output the plaintext data (P-data), Then, through the data fusion processing server (DFPS), the multiple channels of plaintext data are clipped to form an optimal whole data stream (O-data), Finally, The central computer system divides the parameters of the whole measurement data (WM-data) flow, calculates, processes and displays the results data(R-data) .

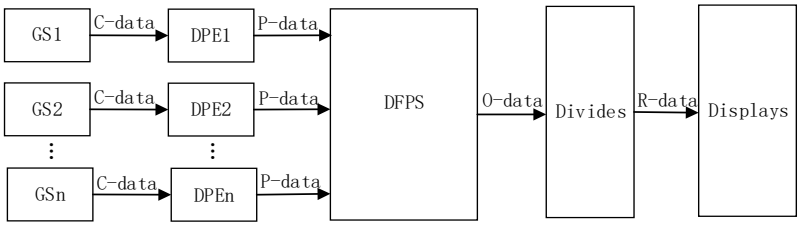


Figure 1. Data processing flowchart

In the data fusion processing, affected by the spatial geographic location of the GS, the data sent by the target at the same time is parsed to the DFPS at a different time, that is the transmission delay. Therefore, the first priority of data fusion is the alignment operation, to match the same-origin frame data sent by different ground stations to receive targets. The main difficulty of alignment lies in the real-time dynamic change of the distance between the target and the GS during the flight, which causes the change of the transmission delay and the possible frame loss and error during data transmission. After alignment, it is necessary to screen the parts with better record quality and complete the whole WM-data splicing, which is called optimization. The key is how to accurately evaluate the quality of the same-origin frames from different GS.

In summary, the data volume of telemetry data is huge, the information processing process is cumbersome, and it is accompanied by frame loss and error and transmission delay. The problems faced by data alignment and optimization processing are complex, and the algorithm is slightly delayed, there will be a cumulative waiting phenomenon, and the fusion time will double, which undoubtedly brings severe challenges to the alignment and optimization of fusion technologies.

III. REAL-TIME ALIGNMENT TECHNOLOGY

Real-time alignment processing must make timely judgments and choices on the currently received limited telemetry data, and the alignment algorithm requires high real-time and reliability. Real-time alignment is often rough, mainly including flag alignment (FA), time code matching alignment (TCMA), and error control alignment (ECA). The following is a key analysis of the real-time alignment algorithm.

A. Flag alignment (FA)

The most classic algorithm in FA is the alignment method based on frame count proposed in Literature [21]. Frame counting is a part of telemetry parameters, has continuity and unity at the same time, and its calculation amount is relatively small, the algorithm time complexity is relatively low, and it has become the first choice for real-time alignment technology. But the frame count error is fatal to the algorithm.

Literature [8] optimizes the receiving buffer and the judgment conditions when the frame count is wrong, and reduces the use of computer resources. When the frame count is inconsistent, the smaller frame count is selected as the alignment result data. However, when frame count errors occur continuously in multiple channels of data, this method will cause accumulated frame count errors in the fusion result. As shown in Table I, the frame counts of No. 1 and No. 4 of GS1 have errors, and GS2 Frame count consecutive errors, including data frames No. 2,3, and 4. When frame count alignment is used, the fusion result is 28586, 65467, 65356, 65523 cumulative errors.

TABLE I. ACCUMULATIVE ERROR ALIGNMENT PROCESS

Frame number	GS1	GS2	Fusion result	Correct
1	28586	65530	28586	65530
2	65531	65467	65467	65531
3	65532	65356	65356	65532
4	65523	65533	65523	65533

B. Time code matching alignment (TCMA)

The first step of real-time TCMA is performs time code correction, and then, uses the time code matching to align S-frame or F-frame. Reference [21] revises other stations with reference to the time code of the master station frame. This method is easy to implement, but because the time delay of the master station time code in the data transmission process is not considered, the time accuracy of the data alignment result is lost.

C. Error control alignment (ECA)

The main idea of real-time ECA is to calculate the time error range in the frame data transmission process, which is called the time delay error range. Based on the frame time code of any station, the corresponding data of the frame time code within the time delay error range is determined as the frame at the same time. As shown in formula:

$$T + \Delta t > T > T - \Delta t \tag{1}$$

T is the reference time, Δt is the time delay error range.

Literature [14] uses theoretical ballistics as equation (2) to accurately calculate the radio wave transmission delay.

$$\Delta t_i = \frac{R_i}{C} \tag{2}$$

C is the speed of light. R_i is the distance between the target at the time of t_i and the ground station, Δt is the time delay of the electric wave transmission at t_i which realizes the alignment of the F-frame data. The time code differences of adjacent S-frame of the same F-frame at different

stations are all less than the calculation delay, and the S-frame cannot be uniquely aligned.

Literature [2] determines the delay error of the S-frame period through the telemetry equipment indicators and the code rate technical indicators. The S-frame time difference within this range can be considered as the S-frame from the same time, and the maximum utilization of the S-frame is realized. This method relies too much on device index values and fails to solve the actual problems caused by frame loss and error codes.

Table II summarizes and analyzes the existing problems of the real-time alignment algorithm according to the document serial number. At present, the alignment algorithm can complete data alignment with different accuracy, but different alignment algorithms still have corresponding problems. The engineering needs to be further combined with actual needs. Analysis and optimization.

TABLE II. LITERATURE CORRESPONDENCE ALIGNMENT ALGORITHM ANALYSIS

Approach	Problem	Literature number
FA	Frame count errors have a greater impact	[21]
	Cumulative misalignment	[8]
TCMA	The transmission delay of the master station is not considered	[21]
ECA	S-frame time code delay calculation is not resolved	[14]
	Depends on device index value	[2]

IV. POST-MORTEM ALIGNMENT TECHNOLOGY

The post-alignment processing is aimed at the entire telemetry data file record, and the processing process is fine. Researchers pay more attention to post-processing methods. The existing post-alignment methods include: Post-event flag bit alignment (P-FA), post-event time code matching alignment (P-TCMA), post-event error control alignment (P-ECA).

A. P-FA

Compared with the real-time method, it pays more attention to the error correction of the flag bit. The literature [13] uses the time difference of the F-frame (S-frame) frame header divided by the S-frame sampling period to obtain the difference in the number of sub-frames between the F-frames (S-frame count difference).); Starting from the first frame, the frame count is accumulated frame by frame, and the frame count is restored. It overcomes the situation that the frame count and frame data are not one-to-one corresponding to the frame count caused by the clearing of the frame count and the error code. However, when a frame loss occurs in the data record file of one of the measurement stations, the F-frame count difference at the position of the lost frame will increase exponentially, and the alignment algorithm cannot solve the problem of matching the frame data and the frame count at this time.

B. P-TCMA

The focus of P-TCMA is timecode refinement correction. Literature [19] first selects two data streams of the same length, combines the characteristics of the sensor signal, and calculates the delay using the third-order mutual cumulant estimation method. To accurate time delay estimation. Assuming p is the expected maximum delay, Delay D as integer, The measurement signal $y(n)$ is the $AR(p)$ process. The calculation method satisfies:

$$\begin{cases} y(n)=\sum_{i=-p}^{p} a(i)x(n-i)+w(n) \\ a(i)=0, i\neq D, a(D)=1 \end{cases} \tag{3}$$

Where $a(i)$ is the coefficient of AR , $w(n)$ is Gaussian white noise, When it is maximum of $|a(i)|$, the i is the required delay. The specific formula is calculated as follows:

$$\begin{cases} c_{yxx}(\tau,\rho)=E\{y(n)x(n+\tau)x(n+\rho)\} \\ c_{xxx}(\tau,\rho)=E\{x(n)x(n+\tau)x(n+\rho)\} \\ c_{yxx}(\tau,\rho)=\sum_{i=-p}^{p} a(i)c_{xxx}(\tau+i,\rho+i) \\ C_{xxx}a=C_{yxx} \end{cases} \tag{4}$$

This method is relatively cumbersome to calculate, the algorithm is difficult to implement,

and the time complexity is high, which is not conducive to popularization and application. Literature [18] uses the transmit zero time plus a multiple of the number of data positions to refill the frame time code. As shown in formula: $T_i = T_0 + i * \Delta t$, Where T_0 is the moment when zero occurs. Δt is the number of data positions, T_i is the time corresponding to the data of the first frame. There are three ways to calculate the number of data positions. One is that the program reads the data of the same number of bits to find the corresponding number of data positions based on the same number of bits occupied by the frame data; The difference is divided by the frame period to obtain the number of data positions. I.e. formula $(T_i - T_0) / \Delta t$ .This centralized method reduces the computational complexity of the algorithm, but when frame loss occurs, the problem of the same number of data positions and different data contents has not been resolved. Literature [12] provides a local frame time code correction method, which uses the frame time interval to correct the time code. The specific process is: taking four adjacent frames in the same data recording file, using the principle of "the time difference between adjacent frames is the same", and correcting time codes with different differences. This partial correction method overcomes the problem of frame counting errors, but obviously does not consider the data transmission delay between multiple stations.

C. P-ECA

In the A. P-ECA, the calculation of the delay error range is the most critical problem to be solved. The traditional calculation method is to use the difference between the frame time code of the reference station and the time codes of the adjacent frames before and after other stations. Reference [16] sets its size based on the target test model. However, the relevant values are often not given in practice. Reference [10] calculates the current F-frame theoretical time according to formula:

$$T_n = T_0 + (C_n \times P) \tag{5}$$

Among them T_0 is the time zero point, C_n is the frame count value, and P is the frame period. The time delay error range is 20 milliseconds by analyzing the time delay of the time system link and the time difference between the data demodulated by the telemetry station. The calculation of this method relies on the frame count value. When there is a bit error, the theoretical time will be calculated incorrectly, resulting in data loss. Literature [4] determines the allowable error of S-frame sampling according to the code rate technical index, which is used as the time delay error range. Obviously, only the influence of the time difference of the ground station on the data delay is considered, and the transmission delay is not considered. The theoretical trajectory of literature [3] estimates the time delay and corrects the time code as shown in formula, and corrects and number the adjacent F-frame time codes. The F-frame with the same number is the aligned data.

Table III summarizes and analyzes the alignment algorithm after the fact according to the document serial number from the accuracy of the algorithm, the advantages of the algorithm, the existing problems, and whether to consider the transmission delay and frame error. The current alignment algorithm can complete data alignment with different accuracy. , But different alignment algorithms still have corresponding problems, and the engineering needs to be further analyzed and optimized in combination with actual needs.

TABLE III. LITERATURE CORRESPONDENCE ALIGNMENT ALGORITHM ANALYSIS

Approach	Problem	Literature number
P-FA	The problem of matching the frame data and the frame count when the frame is lost is not solved	[13]
	Large amount of calculation, not easy to promote	[19]
P-TCMA	The number of data positions is wrong when the frame is dropped	[18]
	Data transmission delay is not considered	[12]
P-ECA	Depends on device index value	[16]
	Frame count error has not been resolved	[10]
	No consideration of transmission delay	[4]
	Large amount of calculation, not easy to promote	[3]

V. REAL-TIME OPTIMAL TECHNOLOGY

The core of the real-time optimization technology is the QEA. The difficulty of the QEA is to reduce the complexity as much as possible on the premise of ensuring the accuracy of the evaluation. The exploration of the algorithm from selection to F-frame and S-frame marks the inevitable trend of the optimization technology to leap to refinement. The exploration of real-time quality assessment algorithms is the most challenging research problem of real-time data fusion. The existing real-time quality assessment algorithms are based on F-frame and S-frame, which will be described in detail below.

A. Based on F-frame

F-frame QEA (FF-QEA) was proposed and tested in the literature [14]. The specific process is: first check whether the frame time code is continuous and whether the frame synchronization code is correct as the basis for priority selection, then, compare it byte by byte According to the data, the best F-frame is evaluated according to the method selected by the three-judgment principle. Obviously, the system overhead of this method is relatively large, and as the amount of data increases, the problem that the information is too late to process is prone to appear. Moreover, in the process of implementing the three-judgment-two principle, when the three frame byte data in the multi-channel F-frame are all different, the algorithm only relies on the frame synchronization code for quality evaluation, and the accuracy is low.

In order to improve the processing efficiency of the FF-QEA, literature [8] sets a delay buffer window and improves the evaluation strategy. In practical applications, use the frame counter number and sampling correlation value (signal-to-noise ratio) with a small amount of data calculation for quality evaluation, and accurately calculate the buffer size by formula:

$$\frac{L_{FIFO}}{t_r - t_w} > \frac{D}{t_w} \tag{6}$$

L_{FIFO} is the depth of the FIFO buffer, t_w is the bit rate at which data is written to the fusion processing server, t_r is the rate at which the optimal algorithm reads the buffer, and D is the amount of data with the size of the transmission delay. This method effectively reduces the time and space complexity of the algorithm, but when a frame count error occurs, the signal-to-noise ratio alone cannot accurately evaluate the F-frame quality.

B. Based on S-frame

S-frame QEA (SF-QEA) improves the degree of refinement of data processing and maximizes utilization of S-frame. Literature [2] extracts S-frame from the F-frame; prioritizes the algorithm evaluation of S-frame normality, integrity, action period, and characteristic parameters; selects preferred S-frame according to the priority S-frame by S-frame. The algorithm time complexity of this method increases, but it provides a more accurate, reliable and efficient quality evaluation method for the SF-QEA.

Table IV summarizes and analyzes the real-time quality evaluation algorithm according to the document serial number. The current algorithm achieves quality evaluation with different accuracy based on the F-frame and S-frame. The time and resource overhead required for its operation are different. The engineering can be based on different actual conditions. Need to select and improve the appropriate quality assessment algorithm.

TABLE IV. LITERATURE CORRESPONDING QEA ANALYSIS

Approach	Problem	Literature number
FF-QEA	The system overhead is large, and when the amount of data increases, the information is too late to process.	[14]
	Accurate quality evaluation when errors occur in unresolved frame counts	[8]
SF-QEA	Increased algorithm time complexity	[2]

VI. POST-MORTEM OPTIMAL TECHNOLOGY

The ex post selection technology focuses on more accurate quality assessment, which provides a reference for real-time selection. The following is a specific introduction to the post-selection technology based on segment selection (PS- QEA), F-frame (PFF-QEA), and subframe (PSF-QEA).

A. PSF-QEA

PSF-QEA is also known as multi-site selection and splicing method [20] or time reference method [17]. The purpose of algorithm evaluation is to find the connection point of the selection. The specific method is to select the docking point in the critical point (T_1, T_2, T_3) of the measurement area of the ground station and compare and verify the N frames of data before and after. Literature [18] selects two stations to check the following 10 subframes before and after the node: (1) Check whether the frame length meets the predetermined size; (2) Whether the BCD code sequence meets the maximum allowable error range of the sampling period. This method requires a large amount of calculation, and data errors do not affect the frame length, but cause quality misjudgments. Literature [6] puts forward the concept of S-frame loss-of-lock rate [6], that is, the docking point is determined by the S-frame synchronization code error rate in a period of time. As shown in formula: $E = M / N \times 1000‰$, where M is the number of S-frames in the selection, and N is the number of S-frame data synchronization code errors. This method greatly improves the efficiency of selecting butt joints.

B. PFF-QEA

PFF-QEA was first proposed in the literature [20]. Compared with the method of segment selection, it obviously improves the utilization rate of the F-frame data and improves the accuracy of the processing result. However, the actual received data format is changeable, and there are errors and frame loss. The adaptability of this method is relatively poor. The classic quality evaluation method is proposed in [10], that is, the integrity of all subframe synchronization codes in the F-frame is used as the basis for evaluation. This not only guarantees the reliability of the quality assessment, but also improves the calculation efficiency of the quality assessment. Literature [3] uses the classic quality evaluation method, the difference is that a necessity check is performed, as shown in Figure 9: before the quality evaluation, the number of F-frames participating in the evaluation is judged. If there is only one F-frame, select it directly without performing quality evaluation. In this way, under the premise of ensuring the reliability of the algorithm, the calculation amount of the algorithm is reduced.

C. PSF-QEA

Literature [5] uses 3 kinds of constraint conditions to select the F-frame, and the F-frame that meets the constraint conditions will be determined as qualified. The constraints are calculated as follows:

$$\begin{cases} \left| T_{ki} - T_j^g \right| < \varepsilon \\ C_j^g = C_{ki} + \Delta C \\ C_{ki} = C_{k(i-1)} + 1 \end{cases} \tag{7}$$

k is the station number, T_{ki} is the F-frame BCD time code, T_j^g is the F-frame BCD time code of the j-th period, ε is allowable error for F-frame header time, C_j^g is the global frame count after the j-th period is corrected, C_{ki} is frame count of the i-th F-frame, ΔC is the correction value caused by frame count overflow or clearing, $C_{k(i-1)}$ is the frame count for the i-1th F-frame. This algorithm has high requirements for data preprocessing, discarding incomplete F-frame data, which is not conducive to full use of data. The S-frame-based post-mortem quality evaluation algorithm is characterized by a high degree of refinement and a large computational complexity. Literature [11] formatted the S-frame data as shown in the formula:

$$D = \left(T, a, A, F^{'} \right) \tag{8}$$

The T vector represents the time code, the a vector represents the S-frame data, the A vector represents the S-frame synchronization code, and the F represents the identifier; the quality evaluation method is as follows:

a) Check calculation for subframe structure: F data frame error evaluation value is 1, the synchronization code is normal synchronization code is 0, the synchronization code is inverted code is -1, and its value is assigned to the structure check vector value of δ_1;

b) Check the subframe count: the difference between the frame counts is 1, and the frame count check evaluation value is 1, otherwise it is 0. Assign the evaluation value in the entire matrix to the frame count check vector of δ_2;

c) S-frame time code verification: the time code difference between adjacent S-frames is 0 within the allowable error range of the time code; otherwise, it is 1. The verification result is recorded as a vector of δ_3;

d) Inverted code period check: the length between the positions where adjacent inverted codes appear is equal to the length of the whole frame, which is 0, otherwise it is unqualified and its value is 1. The inverted code period check result is recorded as a vector of δ_4;

e) Quality evaluation value: assign weight to the above four check vectors (w_1, w_2, w_3, w_4), Calculate the overall evaluation value. As shown in the formula:

$$\delta = w_i \bullet \delta_{ij} \tag{9}$$

The literature [7] evaluates the S-frame quality based on the principle of nearest neighbor clustering. The better the S-frame quality is mapped to the higher the similarity, the closer the distance from the cluster center. First, the S-frame data at the same time is subjected to a standardized metric value to reflect the degree of dispersion of the S-frame data, that is, the standard metric value formula is obtained by the average value of the absolute deviation:

$$\begin{cases} S_{vi} = \dfrac{1}{N}\sum_{n-1}^{N}\left|v_n^i - m_{vi}\right| \\ Z_{vi}^n = \dfrac{v_n^i - m_{vi}}{S_{vi}}(n = 1, 2, \cdots, N) \end{cases} \tag{10}$$

Among them, i is the station number, and N is the length of the subframe. Then calculate the Manhattan distance of the normalized metric of the subframe data:

$$d_{ij} = \sum_{n-1}^{N}\left|v_n^j - v_n^i\right| \tag{11}$$

d_{ij} is the difference value of the S-frame data between station i and station j. Finally, set the cluster center radius, classify the corresponding S-frame data, and select the S-frame closest to the cluster center as the optimal result. This similarity-based method provides a new idea for the quality evaluation algorithm, which is worthy of attention and in-depth study by researchers.

Table V summarizes and analyzes the real-time and post-event quality evaluation algorithms according to the document number. Among them, the complexity of the algorithm principle and the accuracy of the algorithm processing results include 5 levels from high to low, high, normal, low, and low. The resource overhead used by the algorithm is divided into large, large, general, small, and small from large to small. The current algorithm achieves quality evaluation with different accuracy based on selection, F-frame, and subframe. The time and resource overhead required for its operation are different. In engineering, suitable quality evaluation algorithms can be selected and improved according to different actual needs.

TABLE V. LITERATURE CORRESPONDENCE ALGORITHM ANALYSIS TABLE

Approach	Problem	Literature number
PSF-QEA	Data error does not affect the frame length, causing quality misjudgment.	[18]
	To solve the judgment error caused by dropped frames	[6]

PFF-QEA	Unsolved the problem of different comparison results caused by dropped frames and errors	[20]
	The assessment basis is relatively simple	[10]
	Not given due to frame loss	[3]
	Data preprocessing requirements are high, and the incomplete F-frame data is discarded, which is not conducive to the full use of data.	[5]
PSF-QEA	The calculation is cumbersome and not easy to promote	[11]
	Cluster center radius is not easy to choose	[6]

VII. SHORTCOMINGS AND PROSPECTS OF EXISTING METHODS

At present, it is difficult for fusion technology to take into account the intricacies of the actual situation at the same time. The research on generalized, high-efficiency, and high-precision processing methods for data fusion has increased the difficulty, and there are problems that need to be further studied and improved.

Data frame loss is an inevitable interference factor that affects fusion accuracy, and is a key issue faced by alignment and optimization algorithms. Once frame loss occurs, the algorithm will run delayed alignment, wrong alignment, and invalid selection. It will definitely affect the calculation time and accuracy of the fusion result. Researchers use a method based on pseudo-S-frames (PS-frame) to fill in the missing frame data and solve the related difficulties of the alignment technology [7]. However, there are few frame parameters in the PS-frame, which are quite different from the actual frame signal, and when participating in the optimization, the accuracy of the resultant data is reduced. In subsequent research, the method of pattern recognition and parameter estimation can be used to predict the S-frame data [25], so that the data participating in the selection is closer to the actual value.

The design of the key algorithms for alignment and optimization in the fusion technology mainly uses specific parameters such as frame count, time code, synchronization word, and the phenomenon of bit errors in this part of the data is bound to have a certain impact on the operation of the algorithm, especially for those that rely too much on specific parameters [21]. Algorithms are often fatal. Although fusion processing uses related technologies to repair specific parameters [3], the repair method will also fail when encountering more complex situations. Therefore, in the selection technique, researchers try to measure the similarity by calculating the Manhattan distance metric based on the standard metric value of the entire frame of data to achieve the purpose of selection [8]. Based on this idea, we can learn from the machine learning method for time series data mining technology [23, 24], accurately calculate the similarity, and match the entire frame of data to achieve the purpose of fusion.

VIII. CONCLUSION

Continuously improving the key algorithms of the fusion technology in practice is an effective way to ensure the accuracy and reliability of the test data. The study introduced the structural characteristics of telemetry data and the information processing flow, analyzed the actual problems of transmission delay and frame loss and error encountered in data processing, and pointed out the technical difficulties in data fusion. Starting from actual engineering requirements and technical points, the development process of real-time alignment and selection is explained, and the alignment algorithms and QEA involved are classified and analyzed in detail. According to the shortcomings of the existing methods, the next step of the algorithm will focus on the direction of frame loss prediction and the matching of the whole frame, and design pattern recognition and machine learning algorithms to improve the accuracy of the fusion process. With the continuous advancement of computer technology, real-time data fusion technology is bound to burst into new vitality.

ACKNOWLEDGMENT

Thank teacher for his careful guidance in thesis reading and summary writing. This work is supported by the research and development project of wireless network and Intelligent System Laboratory of Xi'an Technological University and real-time fusion of telemetry data.

REFERENCES

[1] Yu, K. (2017) The development and trend of real-time telemetry data ground station., 15(08):15-17.

[2] Jia, H.Y., Wang, S.H. (2020) Research on real-time fusion method of multi-station telemetry data based on the best sub-frame quality. J. Electronic Measurement Technology., 43(10):74-77.

[3] Yang, J., Zhang, D. (2019) Multi-station telemetry data fusion method based on full frame optimization. J. Electronic Measurement Technology., 42(17):101-105.

[4] Yu, C.H., Xu, S.T., Shi, Y.H. (2018) Research on Data Fusion Method in Multi-station Telemetry Data Processing. J. Journal of Telemetry, Tracking and Command., 39(01):47-52+56.

[5] Lu, Z.G. (2017) Intelligent multi-station telemetry data processing system. J. Journal of Telemetry, Tracking and Command., 38(04):9-19.

[6] Zhang, J., Zhang, X.X., Li, Z.F. (2017) Design and Realization of Accurate Mosaic Method of Multi-station Telemetry Data. J. Journal of Telemetry, Tracking and Command., 38(02):22-26.

[7] Zhu, X.F. (2016) Multi-station telemetry data fusion method based on nearest neighbor cluster analysis. J. Journal of Ballistics., 28(02):93-96.

[8] Wu, Y., Huo, J.H., Guo, S.W. (2016) Real-time telemetry data selection technology for multi-site networking. J. Measurement & Control Technology., 35(06):60-63+67.

[9] Han, N. (2015) Improved telemetry data docking method and error elimination. J. Measurement & Control Technology., 22(03):14-16.

[10] Du, P. (2015) Design and Realization of Software for Fast Fusion of Multi-station Telemetry Data. J. Computer Measurement & Control., 23(06):2218-2219+2240.

[11] Lu, N. (2015) Research on the technology of telemetry transmission and ground data fusion for high dynamic aircraft. D. National University of Defense Technology.

[12] Cheng, H.Y., Shu, C.H., Cui, J.F. (2015) Methods to improve the accuracy of telemetry through data processing. J. Radio Engineering., 45(08):10-14.

[13] Shu, C.H., Sun, X., Cui, J.F., Cheng, H.Y. (2014) Method for automatic connection of telemetry multi-station data based on the best single frame quality. J. Journal of Telemetry, Tracking and Command., 35(06):50-55+66.

[14] Liu, G.S., Gao, S., Gui, Y., He, J.J., Li, T.B. (2014) Research on real-time docking method of multi-station telemetry data. J. Radio Engineering., 44(11):34-37.

[15] Jia, H.Y., Yu, R.N. (2020) Multi-task telemetry data real-time processing system. J. Journal of Detection & Control., 42(05):102-106.

[16] [15] Liu, Y.N., Chen, L., Chang, S.L., Dai, Y.C. (2012) Design and Realization of Telemetry Data Fusion Software. J. Modern Electronics Technique., 35(04):136-138+144.

[17] [16] Zhu, X.F. (2011) Precise docking method for multi-station telemetry raw data. J. Tactical Missile Technology., (06):112-115.

[18] Zhang, D., Wu, X.L. (2011) Research on the Method of Missile Telemetry Data Preprocessing. J. Information Technology., 30(03):65-68.

[19] [18] Zhu, X.F., Han, N.(2009) Time delay estimation method of telemetry data based on mutual cumulant. J. Journal of Telemetry, Tracking and Command., 30(03):65-68.

[20] Xu, H.Z., Liang, H., Han, C.Z. (2009) Research on data extraction of telemetry preprocessing under multi-station measurement system. J. Information Technology., 33(02):26-29.

[21] Liang, H., Chen, L., Li, Y.L. (2007) Research on the Fusion Mechanism of Real-time Telemetry Data. J. Journal of Telemetry, Tracking and Command., 30(03):65-68.

[22] Ding, F., Jiang, Q.X., Zhang, N. (2007) Review and Prospect of the Development of Multi-sensor Data Fusion. J. Shipboard Electronic Countermeasure., (03):52-55+73.

[23] Chen, H.Y., Liu, C.H., Sun, B. (2017) A Survey of Similarity Measures in Time Series Data Mining. J. Control and Decision., 32(01):1-11.

[24] Wu, Y., Liang, J., Peng, Y. (2021) Similarity based telemetry data recovery for enhancing operating reliability of satellite., J. Microelectronics Reliability., 0026-2714

[25] Cui, G.L. (2017) Research on Spacecraft Telemetry Data Prediction Algorithm Based on Time Series., D. Xi`an Technological University.

靶场遥测数据实时融合技术研究

周行行
计算机科学与工程学院
西安工业大学
西安, 中国

折宇超
信息技术中心
西安工业大学
西安, 中国

容晓峰
计算机科学与工程学院
西安工业大学
西安, 中国
e-mail: xiaofengrong@126.com

范琳娟
计算机科学与工程学院
西安工业大学
西安, 中国

马方远
中国兵器工业试验测试研究院
渭南, 中国

赵明婕
计算机科学与工程学院
西安工业大学
西安, 中国

闫谦时
计算机科学与工程学院
西安工业大学
西安, 中国

***摘要*—遥测数据是地面站获取飞行器系统工作状态和环境参数的重要数据资料，其融合处理是对多路数据进行择优筛选，提高全程记录可靠性和准确性的关键技术。由于遥测数据量较大，传输过程存在时延误差，并伴随丢帧和误码的现象，为数据融合的对齐和选优带来了严峻挑战。研究从融合技术的应用背景出发，介绍并分析了遥测数据的特点及融合技术的难点；概述目前对齐、选优的相关研究成果以及发展历程，根据工程需求和技术要点，分别从实时和事后的角度出发，对其中涉及的对齐算法、质量评估算法进行了详细的分类、分析；最后，归纳总结了现有方法的不足，展望未来发展的方向，以此为相关研究者提供参考。**

关键词-遥测数据; 数据融合; 对齐; 选优

1. 介绍

遥测系统是现代飞机和航空武器发射试验中的重要内容[1]。多年来，随着试验任务之间的不断磨合和改进，形成了一套基于多站测控的遥测数据处理流程[5]。数据融合作为处理流程中的关键技术，主要通过对各站接收的多路数据进行择优筛选，以提高全程记录可靠性和准确性[3, 4]。

融合技术的传统方法是事后人工统计一段数据的出错情况，择优进行拼接[19]。基于遥测/计算机系统的发展，2009 年徐洪洲等人基于“三判二原则”，全帧数据逐字节对比算法，实现了融合技术的自动化处理。用时，也暴漏出了处理过程中存在的传输时延、丢帧、误码等一系列问题。基于文献[17]建立标准、统一的数据量化方法。2014 年，刘桂生等人利用理论弹道刻画了数据传输时延，并且，将融合技术分为了对齐和选优两部分进行研究。，之后融合技术领域的各种算法相继被提出[7-17]，旨在克服实际任务中遇到的难题，向着更加通用、高效、精准的方向不断发展。

目前，对于融合技术的研究已经取得较多成果，但研究内容相对分散、不成体系。因此，研究从数据融合主要涉及的对齐、选优两个部分出发，归纳总结实时和事后融合处理方法的研究现状，并对其中涉及的关键算法进行阐述。力求为数据融合处理技术的发展提供可行的研究思路。

2. 遥测数据

遥测数据通常使用 PCM 体制进行测量，采样周期为毫秒级。采集同一周期内的参数作为一帧 PCM 数据流，经过一定的时间间隔再次采集，循环往复形成最终的 PCM 遥测数据流，记录在表 1 所示的二进制数据流文件中，其数据量较为庞大。

遥测数据的处理流程如图 1 所示，各地面测站将接收的密文数据实时发送到指控中心。先经过解密预处理设备以及解密器并行独立完成数据密，并输出明文数据；再经过数据融合处理服务器，多路明文数据剪辑拼成一条最优的全程数据流；最后，中心计算机系统对全程测量数据流进行参数分路并计算处理和结果供显示 。

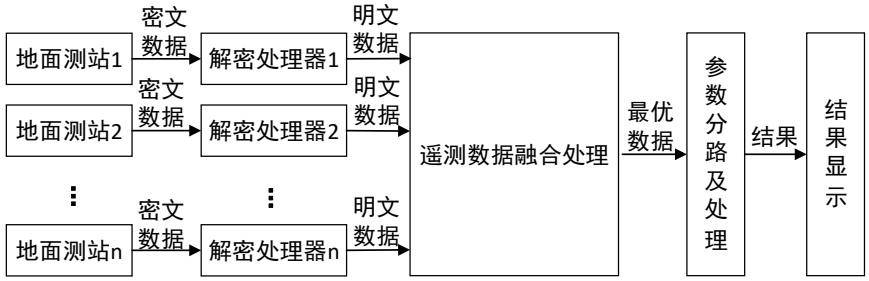


图 1 数据处理流程图

数据融合处理中，受测站空间地理位置的影响，目标同一时刻发送的数据解析到融合处理服务器的时刻不同，即传输时延。因此，数据融合的第一要务是对齐操作，匹配不同地面测站接收目标发出的同源帧数据。对齐的难点主要在于，目标飞行中与测站的距离实时动态变化引起传输时延的变化以及数据传输过程中可能出现的丢帧误码现象等。对齐之后，需要筛选记录质量较好的部分，完成全程数据的拼接，称为选优。而选优的关键，是如何对不同测站的同源帧进行准确的质量评估。

综上所述，遥测数据的数据量庞大，信息处理流程繁琐，且伴随丢帧误码以及传输时延。数据对齐、选优处理面临的问题错综复杂，并且算法稍有延时，将出现累积等待现象，融合时间成倍增加，这无疑给融合技术的对齐、选优带来了严峻挑战。

3. 实时对齐技术

实时对齐处理必须对当前接收的有限帧数据做出及时判断、选择，对齐方法实时性、可靠性要求高。实时对齐往往较为粗略，主要包括实时标志位对齐、实时时码匹配对齐和实时误差控制对齐。

3.1 *实时标志位对齐*

实时标志位对齐中最经典的算法是文献[21]中提出的基于帧计数的对齐方法。帧计数属于遥测参数的一部分,同时具备连续性和统一性，并且其计算量相对较小，算法时间复杂度相对较低，成为实时对齐技术的首选。但帧计数误码对算法来说是致命的。

文献[8]优化了接收缓冲区以及帧计数误码时的判决条件，降低了计算机资源的使用，当其出现帧计数不一致时，选择帧计数较小一路作为对齐结果数据。 然而，多路数据连续出现帧计数误码时，这种方法会导致融合结果出现累积的帧计数错误，如表 1 所示，测站 A 的 1，4 号帧计数产生误码，测站 B 帧计数连续误码，包括 2，3，4 号数据帧，采用帧计数对齐时，融合结果为 28586、65467、65356、65523 的累积错误。

表 1 累积错误对齐过程表

帧序号	测站 A	测站 B	融合结果	正确
1	28586	65530	28586	65530
2	65531	65467	65467	65531
3	65532	65356	65356	65532
4	65523	65533	65523	65533

3.2 *实时时码匹配对齐*

实时时码匹配对齐首先进行时码修正，然后利用时码匹配对齐子帧或全帧。文献[21]参照主站帧的时码对其他测站进行修正。这种方法便于实现，但由于未考虑主站时码在数据传输过程中存在的时延问题，数据对齐结果时间精度有所损失。

3.3 *实时误差控制对齐*

实时误差控制对齐的主要思路是计算帧数据传输过程中的时间误差范围，称为时延误差范围。以任意测站帧时码为基准，处于时延误差范围内的帧时码对应数据确定为同一时刻的帧。如公式（1）所示：

$$T+\Delta t>T>T-\Delta t \tag{1}$$

T 为基准时间，Δt 为时延误差范围。

文献[14]利用理论弹道如式（2），精确计算电波传输时延。

$$\Delta t_i=\frac{R_i}{C} \tag{2}$$

C 为光速；R_i 为 t_i 时刻目标与地面测站的距离；Δt 为 t_i 时刻电波传输延时，实现了全帧数据的对齐。不同测站同一全帧的相邻子帧时码差值均小于计算时延，未能唯一地对齐子帧数据。

文献[2]通过遥测设备指标及码速率技术指标确定子帧周期的延时误差，子帧时差在这一范围内可认为是来自同一时间的子帧，实现了子帧数据的最大化利用。这种方法过于依赖设备指标值，且未能解决丢帧误码带来的实际问题。

表 2 对实时对齐算法按照文献序号对现有实时对齐方法进行了汇总分析。

表 2 实时对齐算法统计表

处理方法	问题	文献编号
实时标志位对齐	帧计数误码影响较大	[21]
	累积错误	[8]
实时时码匹配对齐	未考虑主站传输时延	[21]
实时误差控制对齐	子帧时码延时计算未解决	[14]
	依赖设备指标值	[2]

4. 事后对齐技术

事后对齐处理针对的是全程遥测数据文记录件，处理过程精细。研究者多关注于事后处理的方法，现有的事后对齐方法有事后标志位对齐、事后时码匹配对齐和事后误差控制对齐。

4.1 *事后标志位对齐*

事后标志位对齐相比实时方法更加注重标志位的误码修复，文献[13]利用全帧（副帧）帧头时间差除以子帧采样周期，得到全帧之间的子帧个数差（子帧计数差）；再从首帧开始，逐帧进行帧计数累加，进行帧计数修复。克服了帧计数清零及误码带来的帧计数与帧数据不一一对应的情况。但当其中一台测站的数据记录文件中发生丢帧时，丢帧位置的子帧计数差将成倍增加，对齐算法未能解决此时帧数据与帧计数的匹配问题。

4.2 *事后时码匹配对齐*

事后时码匹配对齐的重点是时码精细化修正。文献[19]首先选取两段相同长度的数据流，结合传感器信号特点，利用 3 阶互累积量估计的方法计算延时，旨在充分利用现有数据之间相互关联以及充分考虑传感器噪声，做到精确时延估计。假设 p 为期望的最大时延，延迟 D 为整数，测量信号 $y(n)$ 为 $AR(p)$ 过程。如公式（3）所示：

$$\begin{cases} y(n)=\sum_{i=-p}^{p}a(i)x(n-i)+w(n) \\ a(i)=0, i\neq D, a(D)=1 \end{cases} \tag{3}$$

其中 $a(i)$ 为 AR 系数，$a(i)=0, i\neq D, a(D)=1$，$w(n)$ 是高斯白噪声。延时估计的计算过程是将（3）式代入（4）式和（5）式中，得到（6）式，对于不同的 (τ,ρ)，获得（7）式的线性方程组，当 $|a(i)|$ 最大时，i 即为所求的时延。具体的公式计算如下：

$$\begin{cases} c_{yxx}(\tau,\rho)=E\{y(n)x(n+\tau)x(n+\rho)\} \\ c_{xxx}(\tau,\rho)=E\{x(n)x(n+\tau)x(n+\rho)\} \\ c_{yxx}(\tau,\rho)=\sum_{i=-p}^{p} a(i)c_{xxx}(\tau+i,\rho+i) \\ C_{xxx}a=C_{yxx} \end{cases} \tag{4}$$

这种方法计算相对繁琐，算法实现的难度较大，时间复杂度较高，不利于推广应用。文献[18]使用发射零点时刻加上数据位置数的倍数对帧时码进行重新填补。如公所示：$T_i=T_0+i*\Delta t$，式中 T_0 为发生零点的时刻。Δt 对于数据位置数，T_i 即为第 i 帧数据对应的时刻。对于数据位置数，有三种方法计算方法，其一，根据帧数据所占的位数相同，程序读取相同位数的数据求出对应的数据位置数；其二，利用数据时刻与发射零点的差值除以帧周期取得数据位置数。即公式 $(T_i-T_0)/\Delta t$。这种集中方法降低了算法的计算量，但当发生丢帧时，数据位置数相同数据内容不同的问题尚未解决。文献[12]中给出了一种局部帧时码修正方法，即利用帧时间间隔修正时码。具体的过程为：取同一数据记录文件中相邻四帧，采用“相邻帧时间差相同”的原理，修正存在差值不同的时码。这种局部修正方法，克服了帧计数发生误码的问题，但显然没有考虑多个测站之间到数据传输时延。

4.3 *事后误差控制对齐*

事后误差控制对齐中，时延误差范围计算是需要解决的最关键的问题。传统的计算方法是利用基准测站帧时码与其他测站前后临近帧时码的差值得出。文献[16]根目标试验型号设置其大小。但实际中往往并未给出相关数值。文献[10]根据式（5）计算当前全帧理论时间：

$$T_n=T_0+(C_n\times P) \tag{5}$$

其中 T_0 为时间零点，C_n 为帧计数值，P 为帧周期。时延误差范围通过分析时统链路的延时和遥测站解调数据的时间差为 ± 20 毫秒。这种方法计算时依赖帧计数值，当其出现误码时，理论时间将计算出错，造成数据缺失。文献[4]根据码速率技术指标确定子帧采样的允许误差，以此作为时延误差范围。显然仅考虑了地面测站的时间差异对数据延时的影响，而未考虑传输时延。文献[3]理论弹道，如式（2）所示估算时延、修正时码，并将相邻对全帧时码进行了修正并编号，相同编号的全帧即为对齐后的数据。

表 3 对事后的对齐算法按照文献序号从存在的问题进行了汇总分析。

表 3 事后对齐算法统计表

处理方法	问题	文献编号
事后标志位对齐	未解决丢帧时的帧数据与帧计数的匹配问题	[13]
	计算量大，不易于推广	[19]
事后时码匹配对齐	丢帧时数据位置数出错	[18]
	未考虑数据传输时延	[12]
事后误差控制对齐	依赖设备指标值	[16]
	帧计数误码尚未解决	[10]
	未考虑传输时延	[4]
	计算量较大，不易于推广	[3]

5. 实时选优技术

选优技术的核心是质量评估算法。质量评估算法的难点是保证评估准确性的前提下尽可能降低复杂程度。算法从选段到全帧和子帧的探索，标志着选优技术向精细化跨越的必然趋势。实时质量评估算法的探索，是数据实时融合最具有挑战性的研究问题。现有的实时质量评估算法基于全帧和子帧，下面进行具体的详细介绍。

5.1 *基于全帧*

基于全帧的实时质量评估算法是由文献[14]提出并进行试验，具体的过程为：首先检查帧时码是否连续，帧同步码是否正确作为优先级选择的依据；然后，逐字节比较数据，按照次数最多相同的数据，即三判二原则选定的方法评定出最优的全帧。显然，这种方法系统开销较大，随着数据量的增大，容易出现信息来不及处理的问题。并且，在执行三判二原则过程中，当多路全帧中三个帧字节数据均不相同的时，算法仅靠帧同步码进行质量评估，准确度较小。

为了提高算法的处理效率，文献[8] 设置了时延缓存窗口并改善了评估策略。在实际应用中，使用数据计算量较小的帧计数字和采样相关值（信噪比）进行质量评估；并通过公式（6）精确计算缓冲区大小：

$$\frac{L_{FIFO}}{t_r - t_w} > \frac{D}{t_w} \tag{6}$$

其中 L_{FIFO} 为 FIFO 缓冲区的深度，t_w 为数据写入融合处理服务器的位速率，t_r 为选优算法读取缓冲区的速率，D 为传输时延大小时间的数据量。这种方法有效降低了算法的时间和空间复杂度，但当帧计数出现误码时，仅靠信噪比不能准确进行全帧质量评估。

5.2 *基于子帧*

基于子帧的实时质量评估提高了数据处理的精细化程度，子帧数据得到最大化利用。文献[1]从全帧中进行子帧提取；通过对子帧数据规范性、完整性、作用时段以及特征参数的算法评估进行优先级排序；逐子帧对照优先级进行择优选取子帧。这种方法的算法时间复杂度 2 增加了，但为基于子帧的实时质量评估算法提供了更加精准、可靠且高效的质量评判方法。

表 4 对实时选优算按照文献序号从存在的问题进行了汇总分析，目前选优算法在不同精度上可以完成数据选优，但不同的选优算法仍然存在相对应的问题，工程上需要进一步结合实际需求进行分析优化。

表 4 实时选优算法统计表

处理方法	问题	文献编号
基于全帧的实时质量评估算法	系统开销大，数据量增大时，信息来不及处理	[14]
	未解决帧计数出现误码时的精确质量评估	[8]
基于子帧的实时质量评估算法	算法时间复杂度增加	[2]

6. 事后选优技术

事后选优技术关注的是更为精准的质量评估，这为实时选优提供了借鉴的依据。下面对基于选段、基于全帧、基于子帧的事后选优技术进行具体介绍。

6.1 *基于选段*

基于选段的事后质量评估算法，又称为多站位选段拼接法[20]或时间基准法[17]。算法评估的目的是找出选段对接点。具体的做法是在地面测站测量区临界点 (T_1, T_2, T_3) 中选取对接点并前后 N 帧数据进行对比校验。文献[17]选取两站对节点前后 10 条子帧，进行如下检查：（1）检查帧长度是否符合既定大小；（2）BCD 码顺序是否符合采样周期最大允许误差范围。这种方法的计算量较大，且数据误码并不影响帧长度，反而造成质量误判。文献[6] 提出了子帧失锁率[5]的概念，即通过一段时间中子帧同步码错误率确定对接点。如式所示：$E = M / N \times 1000‰$，其中，M 为选段中的子帧个数，N 为子帧数据同步码错误的个数。这种方法大大提高了选定对接点的效率。

6.2 *基于全帧*

基于全帧的事后质量评估算法是文献[19]首次提出的，该算法采用逐字节三判二原则的方法进行计算。相对选段的方法，显然提高了全帧数据的利用率，提升了处理结果的准确性。然而实际接收的数据格式多变, 且存在误码、丢帧的情况,该方法的适应能力相对较差。经典的质量评估方法是文献[10]提出的，即利用全帧中的所有子帧同步码完整性作为评估的依据。既保证了质量评估的可靠性，又提高了质量评估的计算效率。文献[3]使用经典质量评估方法，不同的是，进行了必要性检查，在质量

评估之前，判断参加评估的全帧个数。如果仅有一条全帧则直接选定，不进行质量评估。这样在保证算法可靠性的前提下，降低了算法的计算量。

文献[5] 使用 3 种约束条件对全帧进行遴选，符合约束条件的全帧将被确定为质量合格。约束的计算如下：

$$\begin{cases} \left|T_{ki}-T_j^g\right|<\varepsilon \\ C_j^g=C_{ki}+\Delta C \\ C_{ki}=C_{k(i-1)}+1 \end{cases} \tag{7}$$

其中， k 为测站序号，T_{ki}为全帧 BCD 时码，T_j^g为第 j 个周期的全帧 BCD 时码，ε为全帧帧头时间允许误差；C_j^g为第 j 周期修正后全局帧计数，C_{ki}第 i 个全帧的帧计数，ΔC为帧计数溢出或清零导致的修正值；$C_{k(i-1)}$为第 i-1 全帧的帧计数。这种算法对数据预处理要求较高，丢弃残缺的全帧数据，不利于数据的充分利用。

6.3 *基于子帧*

基于子帧的事后质量评估算法的特点是结果精细化程度高，计算复杂度大。文献[10]将子帧数据进行如式 8 所示的格式化处理：

$$D=\left(T,a,A,F^{'}\right) \tag{8}$$

式中 T 向量代表时码，a 向量代表子帧数据，A 向量代表子帧同步码，F 代表标识字；质量评估的方法如下：

6.3.1 对子帧结构校验计算：F 数据帧错误评估值取为 1，同步码为正常同步码取值为 0，同步码为反码取值为-1，将其值赋给结构校验向量值 δ_1；

6.3.2 对子帧计数校验：帧计数之间的差值为 1，帧计数校验评估值为 1，否则为 0。将整个矩阵中的评估值赋给帧计数校验向量δ_2；

6.3.3 子帧时码校验：相邻子帧时码差值介于时码允许误差范围内为 0，否则为 1。校验结果记为向量δ_3；

6.3.4 反码周期校验：相邻反码出现的位置之间的长度等于全帧长度为 0，否则为不合格其值为 1.反码周期校验结果记为向量 δ_4；

6.3.5 质量评估值：将以上四个校验向量赋予权重(w_1,w_2,w_3,w_4)，计算总体评估值。如式 17 所示：

$$\delta=w_i\bullet\delta_{ij} \tag{9}$$

文献[7]子帧质量评估的依据最近邻聚类实现原理，将子帧质量越好映射为相似度越高，距离聚类中心的距离越近。首先将同一时刻子帧数据进行标准化度量值，用来反映子帧数据的离散程度，即通过绝对偏差的平均值求出标准度量值式：

$$\begin{cases} S_{vi}=\dfrac{1}{N}\sum_{n-1}^{N}\left|v_n^i-m_{vi}\right| \\ Z_{vi}^n=\dfrac{v_n^i-m_{vi}}{S_{vi}}(n=1,2,\cdots,N) \end{cases} \tag{10}$$

其中，i 为站位编号，N 为子帧长度。然后计算子帧数据标准化度量值的曼哈顿距离：

$$d_{ij}=\sum_{n-1}^{N}\left|v_n^j-v_n^i\right| \tag{11}$$

式中d_{ij}是测站 i 和测站 j 之间的子帧数据相异度值。最后，设置聚类中心半径，将对应的子帧数据进行分类，并选取距离聚类中心最近的子帧为最优结果。这种基于相似度的方法为质量评估算法提供了一种新的思路，值得研究人员关注和深入研究。

表 5 对事后的选优算法按照文献序号进行了汇总分析.目前算法基于选段、全帧、子帧达到了不同精度的质量评估，其运行需要的时间以及资源开销各不相同，工程上可根据不同的实际需求选择并改进合适的质量评估算法。

表 5 事后选优算法统计表

处理方法	问题	文献编号
基于选段的事后质量评估算法	数据误码并不影响帧长度，造成质量误判。	[18]
	为解决丢帧引起的评判失误	[6]
基于全帧的事后质量评估算法	未解决丢帧误码引起的对比结果均不相同的问题	[20]
	评估依据较为单一	[10]
	未给出丢帧引起的	[3]
	数据预处理要求较高，丢弃残缺的全帧数据，不利于数据的充分利用。	[5]
基于子帧的事后质量评估算法	计算繁琐，不易推广	[11]
	聚类中心半径不易选取	[6]

7. 现有方法的不足与展望

目前，融合技术在面临错综复杂的实际情况中很难同时兼顾，常常顾此失彼。对开展数据融合通用化、高效率、高精度处理方法的研究增加了难度，存在需要进一步研究和完善的问题。

数据丢帧是影响融合精度不可避免的干扰因素，是对齐、选优算法主要面临的关键问题。一旦发生丢帧，算法运行将出现延时对齐、错误对齐以及无效选优等。必将影响融合结果的计算时间和精度。研究人员使用基于伪子帧的方法，补齐丢失的帧数据，解决了对齐技术的相关难点[7]。然而，伪子帧中帧参数较少，与实际帧信号相差较大，参与选优时，降低了结果数据的精度。在后续的研究中，可以采用模式识别和参数估计的方法对子帧数据进行预测，使得参与选优的数据更加接近实际的数值。

融合技术中对齐、选优关键算法的设计主要利用帧计数、时码、同步字等特定参数，而这部分数据产生误码的现象势必对算法运行产生一定的影响，尤其对过于依赖特定参数的算法来说[21]，常常是致命的。虽然融合处理使用相关技术对特定参数进行修复，然而当遇到更为复杂的情况时，修复方法也将失效。因此，选优技术中，研究者试图通过整帧数据的标准度量值计算曼哈顿距离度量进行相似度衡量，达到选优的目的[8]。基于这种思想，可借鉴机器学习方法对时间序列的数据挖掘技术[23,24]，精确计算相似度，对整帧数据进行匹配，达到融合的目的。

8. 结论

实践中不断完善融合技术关键算法，是保证试验数据准确可靠的有效途径。研究介绍了遥测数据的结构特点以及信息处理流程，分析数据处理中遇到的传输时延和丢帧误码的实际问题，指出数据融合中的技术难点。从实际工程需求以及技术要点出发，阐述对齐、选优的发展历程，对其中涉及的实时和事后对齐和质量评估算法进行了详细的分类、分析。根据现有方法存在的不足，下一步算法将针对丢帧预测与整帧匹配的方向展开研究，设计模式识别和机器学习算法提高融合处理的精度。随着计算机技术的不断进步，数据融合技术必将迸发出新的活力。

致谢

感谢老师在论文阅读和总结写作中的细心指导。本课题研究得到了遥测数据实时融合系统开发项目的支持。

参考文献

[26] Yu, K. (2017) The development and trend of real-time telemetry data ground station., 15(08):15-17.

[27] Jia, H.Y., Wang, S.H. (2020) Research on real-time fusion method of multi-station telemetry data based on the best sub-frame quality. J. Electronic Measurement Technology., 43(10):74-77.

[28] Yang, J., Zhang, D. (2019) Multi-station telemetry data fusion method based on full frame optimization. J. Electronic Measurement Technology., 42(17):101-105.

[29] Yu, C.H., Xu, S.T., Shi, Y.H. (2018) Research on Data Fusion Method in Multi-station Telemetry Data Processing. J. Journal of Telemetry, Tracking and Command., 39(01):47-52+56.

[30] Lu, Z.G. (2017) Intelligent multi-station telemetry data processing system. J. Journal of Telemetry, Tracking and Command., 38(04):9-19.

[31] Zhang, J., Zhang, X.X., Li, Z.F. (2017) Design and Realization of Accurate Mosaic Method of Multi-station Telemetry Data. J. Journal of Telemetry, Tracking and Command., 38(02):22-26.

[32] Zhu, X.F. (2016) Multi-station telemetry data fusion method based on nearest neighbor cluster analysis. J. Journal of Ballistics., 28(02):93-96.

[33] Wu, Y., Huo, J.H., Guo, S.W. (2016) Real-time telemetry data selection technology for multi-site networking. J. Measurement & Control Technology., 35(06):60-63+67.

[34] Han, N. (2015) Improved telemetry data docking method and error elimination. J. Measurement & Control Technology., 22(03):14-16.

[35] Du, P. (2015) Design and Realization of Software for Fast Fusion of Multi-station Telemetry Data. J. Computer Measurement & Control., 23(06):2218-2219+2240.

[36] Lu, N. (2015) Research on the technology of telemetry transmission and ground data fusion for high dynamic aircraft. D. National University of Defense Technology.

[37] Cheng, H.Y., Shu, C.H., Cui, J.F. (2015) Methods to improve the accuracy of telemetry through data processing. J. Radio Engineering., 45(08):10-14.

[38] Shu, C.H., Sun, X., Cui, J.F., Cheng, H.Y. (2014) Method for automatic connection of telemetry multi-station data based on the best single frame quality. J. Journal of Telemetry, Tracking and Command., 35(06):50-55+66.

[39] Liu, G.S., Gao, S., Gui, Y., He, J.J., Li, T.B. (2014) Research on real-time docking method of multi-station telemetry data. J. Radio Engineering., 44(11):34-37.

[40] Jia, H.Y., Yu, R.N. (2020) Multi-task telemetry data real-time processing system. J. Journal of Detection & Control., 42(05):102-106.

[41] [15] Liu, Y.N., Chen, L., Chang, S.L., Dai, Y.C. (2012) Design and Realization of Telemetry Data Fusion Software. J. Modern Electronics Technique., 35(04):136-138+144.

[42] [16] Zhu, X.F. (2011) Precise docking method for multi-station telemetry raw data. J. Tactical Missile Technology., (06):112-115.

[43] Zhang, D., Wu, X.L. (2011) Research on the Method of Missile Telemetry Data Preprocessing. J. Information Technology., 30(03):65-68.

[44] [18] Zhu, X.F., Han, N.(2009) Time delay estimation method of telemetry data based on mutual cumulant. J. Journal of Telemetry, Tracking and Command., 30(03):65-68.

[45] Xu, H.Z., Liang, H., Han, C.Z. (2009) Research on data extraction of telemetry preprocessing under multi-station measurement system. J. Information Technology., 33(02):26-29.

[46] Liang, H., Chen, L., Li, Y.L. (2007) Research on the Fusion Mechanism of Real-time Telemetry Data. J. Journal of Telemetry, Tracking and Command., 30(03):65-68.

[47] Ding, F., Jiang, Q.X., Zhang, N. (2007) Review and Prospect of the Development of Multi-sensor Data Fusion. J. Shipboard Electronic Countermeasure., (03):52-55+73.

[48] Chen, H.Y., Liu, C.H., Sun, B. (2017) A Survey of Similarity Measures in Time Series Data Mining. J. Control and Decision., 32(01):1-11.

[49] Wu, Y., Liang, J., Peng, Y. (2021) Similarity based telemetry data recovery for enhancing operating reliability of satellite., J. Microelectronics Reliability., 0026-2714

[50] Cui, G.L. (2017) Research on Spacecraft Telemetry Data Prediction Algorithm Based on Time Series., D. Xi`an Technological University.