

# Improved Method of ResNet50 Image Classification Based on Transfer Learning

Tao Shi

School of Computer Science and Engineering  
Xi'an Technological University  
Xi'an, 710021, Shaanxi, China  
E-mail: s19829601144@163.com

Zhiyi Hu

Engineering Design Institute  
Army Research Laboratory  
Beijing, 100042, China  
E-mail: 763757335@qq.com

Jun Yu

School of Computer Science and Engineering  
Xi'an Technological University  
Xi'an, 710021, Shaanxi, China  
E-mail: yujun@xatu.edu.cn

Kuncai Jiang

School of Computer Science and Engineering  
Xi'an Technological University  
Xi'an, 710021, Shaanxi, China  
E-mail: 2992165628@qq.com

**Abstract**—Aiming at the issues of high computational cost and limited generalization ability of ResNet50 in classifying images, this study advances an optimization strategy based on transfer learning. The model is initialized with transfer learning to reduce computational burden, and data augmentation techniques are employed to enhance generalization ability. Additionally, label smoothing is introduced to optimize the cross-entropy loss, thereby reducing sensitivity to noisy labels. The training process is further optimized using cosine annealing learning rate decay. Experimental findings reveal that the optimized ResNet50 model achieves a 6.25% improvement in classification accuracy on the CIFAR-10 dataset, validating the validity of the suggested methods.

**Keywords**-Transfer Learning; ResNet50; Data Augmentation; Image Classification

## I. INTRODUCTION

With the computer vision community's growing reliance on deep learning systems, CNNs have evolved into cornerstone solutions for visual classification problems [1-2]. Among them, ResNet (Residual Networks), as a deep residual network, has shown excellent performance in image classification tasks. Through its unique residual structure, it effectively addresses the issue of gradient vanishing and has achieved

satisfactory results in many scenarios. However, in practical applications, the ResNet50 model still exposes some problems, such as large computational load, long training time, and a tendency to overfit [3-5]. Especially in efficient classification tasks, how to ensure classification accuracy while improving computational efficiency has become an urgent challenge. Jiang Zhengfeng et al. proposed combining the attention mechanism with deep residual networks for the categorization of remote sensing image scene data [6], achieving an accuracy of 92.94%. Zhang Yizhuo advanced a hierarchical fusion strategy rooted in residual architectures for hyperspectral ground object classification [7], achieving an average overall accuracy of 98.75%. Fang Liang et al. proposed a categorization approach for rusted steel bars rooted in deep residual networks, using industrial camera images combined with data enhancement techniques to classify the rust levels of six datasets [8], with classification accuracies all above 93.2%, and the highest reaching 98.8%. Although these methods have achieved good results in their respective application scenarios, they still have certain limitations when dealing with multi-classification problems.

Rooted in this, this study advances an approach to improve the performance of ResNet50 image classification rooted in transfer learning. This approach exploits the idea of transfer learning to optimize a priorly trained model, reducing the huge computational cost of training from scratch and accelerating network convergence. At the same time, data enhancement techniques are introduced to effectively solve the issues of excessive fitting and inadequate generalization capacity in model training; the label smoothing method is used to modify the cross-entropy loss function to alleviate the oscillation of the model loss value; the cosine annealing decay method is used to train the model, further accelerating the network convergence speed and improving grouping exactness. Verified on the CIFAR-10 data set, the image grouping exactness of this model reaches 93.75%, fully demonstrating its good classification ability.

## II. RESNET NETWORK MODEL

ResNet (Residual Neural Network) was advanced by HE Kaiming. from Microsoft Research [9]. It serves as a key component in modern image classification systems. The proposal of ResNet's network structure has improved the training speed and model accuracy of neural networks. With deeper network architectures, the model's complexity grows significantly, and if the number of layers exceeds a reasonable range, the vanishing gradient issue may emerge. To solve this problem, scholars proposed the residual structure learning unit, as shown in Figure 1. This architecture aims to improve the approximation capability of the nonlinear stacked layers  $f(x)$  toward  $H(x)$  with increasing depth, where residual connections enable  $F(x)+x$  to approximate  $H(x)$  [10]. At the same time, by adding identity mapping in deep networks, shallow features can be directly transmitted to deep networks, avoiding the problem of gradient disappearance during backpropagation, thereby preventing network performance degradation and improving the performance of deep networks.

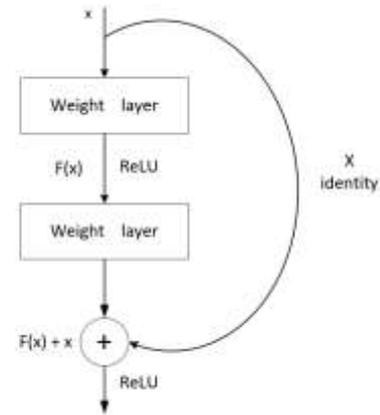


Figure 1. Residual unit structure diagram

As illustrated in Figure 1, the input  $x$  first passes through an initial weight layer to generate  $F(x)$ . After processing by a ReLU activation function, it proceeds to a second weight layer. A skip connection directly adds the original input  $x$  to this output, producing the final mapping  $F(x)+x$ . This architectural design enables direct feature propagation from shallow to deep layers, facilitating inter-layer information flow.

ResNet series network models are commonly employed across multiple fields, with significant adoption in processing medical images and analyzing satellite imagery. ResNet50 is a typical model in the ResNet series, indicating that the network has 50 layers. This article employs ResNet50 as the foundational architecture for investigation, the architectural details are presented in Table I.

TABLE I. RESNET50 ARCHITECTURE

<i>convolutional layer</i>	<i>output layer</i>	<i>ResNet50</i>
Conv-1	112×112	7×7, 64, S=2 3×3 maxpool, S=2
Conv2-x	56×56	$\begin{bmatrix} 1 \times 1 & 64 \\ 3 \times 3 & 64 \\ 1 \times 1 & 256 \end{bmatrix} * 3$
Conv3-x	28×28	$\begin{bmatrix} 1 \times 1 & 128 \\ 3 \times 3 & 128 \\ 1 \times 1 & 512 \end{bmatrix} * 4$
Conv4-x	14×14	$\begin{bmatrix} 1 \times 1 & 256 \\ 3 \times 3 & 256 \\ 1 \times 1 & 1024 \end{bmatrix} * 6$
Conv5-x	7×7	$\begin{bmatrix} 1 \times 1 & 512 \\ 3 \times 3 & 512 \\ 1 \times 1 & 2048 \end{bmatrix} * 3$
1×1	Average_pool,1000-dfc,Soft_max	
Flops		$3.8 \times 10^9$

### III. IMPROVED RESNET50 IMAGE CLASSIFICATION

This experiment uses the PyTorch deep learning framework, researching and optimizing the classification performance of ResNet50 on the CIFAR-10 dataset based on transfer learning. The research includes four modules: first, by introducing transfer learning, pre-trained models are used to accelerate convergence; second, data enhancement techniques are applied to advance the model's capacity to generalize to unseen data; third, the loss function is optimized, and the label smoothing method is used to reduce the model's sensitivity to noisy labels; finally, the cosine annealing learning rate decay is used to accelerate model convergence and improve classification accuracy. Through these optimizations, the model's performance on small-sample datasets has been significantly improved.

#### A. Introduction of Transfer Learning

Transfer learning improves model learning efficiency and generalization ability by employing large-scale previously trained models with subsequent task-specific refinement [11]. The network improvement part of this paper is based on this idea, using the original network structure and parameters of the ResNet50 model pre-trained on ImageNet and applying it to new image classification tasks. Transfer learning is used to save training time as much as possible, significantly lowering the quantity of training parameters, and fine-tuning the parameters for the new model.

Assuming our original field model parameters are  $\theta_s$ , our target field model parameters are  $\theta_t$ , and the fine-tuning objective can be defined as equation (1):

$$\theta_t = \arg \min_{\theta} L_{\text{target}}(f(x; \theta), y) \quad (1)$$

The last fully-interconnected layer of this ResNet50 model is designed for the 1000-class classification task of the ImageNet dataset, with an output dimension of 1000. However, the CIFAR-10 dataset has only 10 categories. Therefore, this experiment replaces the last fully-interconnected layer of the ResNet50 model with a linear layer with an output dimension of 10, it can be expressed as equation (2):

$$\text{model.fc} = \text{nn.Linear}(2048, 10) \quad (2)$$

where 2048 represents the number of input features of the fully-connected layer of ResNet50, and 10 represents the quantity of classes in the CIFAR-10 dataset. This modification ensures that the number of output categories of the model is consistent with the target task, and fully utilizes the feature extraction ability of the pre-trained model through transfer learning.

Through transfer learning, this experiment effectively reduces the time and resource costs required to train the model from scratch on the small dataset CIFAR-10, while improving the model's classification performance. This method fully utilizes the general features learned on the large-scale dataset ImageNet, enabling the model to quickly adjust to new assignments.

#### B. Introduction of Data Enhancement Techniques

The CIFAR-10 dataset has a relatively small number of images, containing only 60,000 images with a resolution of  $32 \times 32$ . If the original data is directly used to train deep neural networks, the model may not fully learn robust features, easily overfitting on the training set, leading to poor performance on the test set. Data enhancement can effectively increase the diversity of the dataset, making the model more generalizable [12-14].

Data enhancement can be mathematically viewed as perturbing the possibility distribution  $P(x)$  of the input data, simulating the changes of training data in real scenes. The enhanced data  $\tilde{x}$  satisfies the following formula (3):

$$\tilde{x} = T(x), T \sim \tau \quad (3)$$

Where  $T$  symbolizes a transformation function in the enhancement operation set  $\tau$ , and by applying different transformations  $T$  to the input  $x$ , diverse data samples can be generated.

For the characteristics of the CIFAR-10 dataset (resolution of  $32 \times 32$ , containing 10 categories), this experiment designs a series of enhancement operations to expand the diversity of training data.

### 1) Random Horizontal Flip

For natural images, left-right symmetrical structures are widely present. Therefore, flipping the image horizontally with a certain probability (set to 50% in this experiment) can enhance the model's adaptability to viewpoint changes.

Let the original image matrix be  $X \in R^{H \times W \times C}$ , where  $H$  and  $W$  are the height and width of the image, and  $C$  is the quantity of channels. The horizontal flip operation can be expressed as equation (4):

$$X'_{i,j,c} = X_{i,W-j-1,c} \quad \forall i \in [0, H-1], j \in [0, W-1], c \in [0, C-1] \quad (4)$$

Where  $X'$  is the flipped image. This operation rotates the image symmetrically along the vertical center axis, that is, the pixel values of column  $j$  in the picture are replaced by the pixel values of columns  $w-j-1$ .

Random horizontal flipping is able to generate samples that are symmetrical to the original image but have different viewing angles, thus enhancing the variability within the training dataset and avoiding model overfitting. By introducing the flipped image, the model is able to learn the feature representation under different viewing angles, thus improving the adaptability to changes in the viewing angle of the image. In natural scenes, the left-right symmetry of objects is common. Random horizontal flipping is able to better simulate the image distribution in the real world, making the model more stable in practical applications.

### 2) Random Crop

In image categorization assignments, the position of objects may vary in different images, so random cropping can improve the model's robustness to target position changes. Specifically, this experiment pads 4 pixels around the image and randomly crops it to the original size ( $32 \times 32$ ), simulating the offset of the target object's position.

Let the original image size be  $H \times W$ , and the size of the padded image be  $(H + 2p) \times (W + 2p)$ , where  $p$  is the number of padding pixels. The random cropping operation randomly selects the top-left corner coordinates  $(i, j)$  from the padded

image and crops out the target area, it can be expressed as equation (5):

$$X' = X [i : i + H, j : j + W, :] \quad (5)$$

Where  $i, j$  follow a uniform distribution  $i \sim U(0, 2P), j \sim U(0, 2P)$ .

### 3) Normalization

Normalization can adjust the numerical distribution of data, reduce the scale differences between input features, accelerate model convergence, and improve stability. Usually, we normalize each channel separately to have a value of 0 and a criterion deviation of 1, ensuring consistent data distribution across different channels.

For pixel  $x$ , the normalization process can be described as Equation (6):

$$\tilde{x} = \frac{x - \mu}{\sigma} \quad (6)$$

Where  $\mu$  and  $\sigma$  represent the value and criterion deviation of the data, separately. The normalized parameters used in this experiment can be expressed as equation (7).

$$\mu = (0.5, 0.5, 0.5), \quad \sigma = (0.5, 0.5, 0.5) \quad (7)$$

This setting scales the pixel values to the interval  $[-1, 1]$ , improving training stability.

Through data enhancement techniques such as horizontal flipping, random cropping, and normalization, this experiment markedly advances the model's generalization ability on the CIFAR-10 dataset. These enhancement strategies effectively expand the training data distribution and alleviate the model's overfitting problem, providing important support for deep learning model training on small-scale datasets.

### C. Improvement of Loss Function

In image classification, the cost function is employed to measure the goodness of the model's predictions, indicating the difference between predicted figure and the true figure. In the image classification problem of this paper, the multi-class cross-entropy cost function is generally used to measure the closeness between the real output mean and expected output mean. The closer the

actual output value is to the expected output value, the smaller the cross-entropy, and the higher the prediction accuracy [15-16]. The specific formulas of the multi-class cross entropy loss function are shown in equations (8), (9) and (10):

$$y_i = \begin{cases} 0, & i \neq c \\ 1, & i = c \end{cases} \quad (8)$$

$$Loss = -\sum_{i=0}^{c-1} y_i \log(p_i) = -\log(p_c) \quad (9)$$

$$Z_i^* = \begin{cases} +\infty, & i = c \\ 0, & i \neq c \end{cases} \quad (10)$$

In the formula:  $p = [p_0, p_1, \dots, p_{c-1}]$  represents a probability distribution, where  $p_i$  is the possibility of the  $i$ -th category of samples;  $y = [y_0, y_1, \dots, y_{c-1}]$  is the one-hot representation of the sample labels; and  $Z_i^*$  represents the optimal prediction probability distribution.

When the model processes one-hot labels, it is prone to overfitting, which cannot ensure the model's generalization capacity and leads to inaccurate predictions. So, this paper employs the label smoothing approach to modify the cross-entropy loss function. Specifically, it can be defined as Equations (11), (12), and (13):

$$y_i = \begin{cases} 1 - \varepsilon, & i = c \\ \varepsilon \div (c-1), & i \neq c \end{cases} \quad (11)$$

$$Loss_i = \begin{cases} (1 - \varepsilon) \times Loss, & i = c \\ \varepsilon \times Loss, & i \neq c \end{cases} \quad (12)$$

$$Z_i^* = \begin{cases} \log((c-1)(1 - \varepsilon) \div \varepsilon) + \alpha, & i = c \\ \alpha, & i \neq c \end{cases} \quad (13)$$

In the equations:  $\varepsilon$  is a relatively small constant;  $\alpha$  is an arbitrary real number; and  $c$  is the total quantity of classification types.

#### D. Improvement of Model Training

During the process of model training, the choice of a suitable learning rate is crucial for classification performance. This is because the learning rate decides whether the neural network

has the ability to converge towards the optimal value. The commonly used learning rate decay methods for training ResNet models include equally spaced adjusted learning rate and Exponential Decay Adjusted Learning Rate.

##### 1) Equal interval adjustment learning rate method

The model reduces the learning rate by a certain percentage after training for a certain number of iterations. This decay strategy is prone to oscillation when changing the learning rate due to the large amplitude of learning rate decay, making it difficult to converge to the optimal value.

##### 2) Exponential Decay Adjusted Learning Rate Method

The model dynamically changes the learning rate based on the current quantity of iterations, and the quantity of iterations determines the update frequency of the learning rate. When the decay rate is small, as the quantity of iterations gradually advances, the learning rate changes less, which can easily cause the model to converge too quickly.

This paper uses the cosine annealing decay method, which reactively changes the step size in learning over time during model training. Using the characteristics of the cosine function itself to decay to the optimal learning rate, it approaches the global minimum of the Loss value, thereby allowing the model to converge to the optimal value.

The specific renewal mechanism of cosine annealing attenuation can be defined as shown in Equation (14):

$$\eta_t = \eta_{min}^i + \frac{1}{2}(\eta_{max}^i - \eta_{min}^i) \left[ 1 + \cos\left(\frac{T_{cur}}{T_i} \pi\right) \right] \quad (14)$$

In the formula:  $i$  is the index value;  $\eta_{min}^i$  is the minimum learning rate;  $\eta_{max}^i$  is the maximum learning rate;  $T_{cur}$  is the current quantity of iterations;  $T_i$  is total number of iterations.

#### IV. EXPERIMENTAL RESULTS AND ANALYSIS

##### A. Dataset and Experimental Configuration

The experimental dataset in this paper is CIFAR-10, the OS is Win11, the processor is Intel Core i7-10750H, and the video memory is 16GB. The deep-learned framework used is pytorch 2.0, GPU device is NVIDIA Tesla T4. This experiment uses transfer learning, with ResNet50

In an attempt to better objectively evaluate classification efficiency of the advanced ResNet50 network on the CIFAR-10 dataset, this paper adopts the following common evaluation metrics: classification accuracy (Accuracy), loss value (Loss), and confusion matrix (Confusion Matrix). These metrics measure the model's predictive performance from different perspectives.

Among them, classification accuracy is one of the most common evaluation metrics, which measures the ratio of accurately predicted samples to the total number of samples. For multi-classification tasks, accuracy calculation formula is shown in Equation (15):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

where TP represents the quantity of precisely predicted optimistic class samples, TN represents the quantity of precisely anticipated pessimistic class samples, FP represents the quantity of samples wrongly anticipated as optimistic class, and FN represents the quantity of samples incorrectly anticipated as pessimistic class. In this paper, the accuracy of each category is calculated, and the mean value is regarded as the last evaluation metric.

Loss value is an indicator used to guide model optimization during the training process, reflecting the distinction between the model-predicted results and real labels. This paper uses the label smoothing cross-entropy loss function, the calculation approach is presented in Equation (16):

$$Loss = -\sum_{i=1}^C q_i \log p_i \quad (16)$$

as the base network, and adjusts its structure to adapt to the CIFAR-10 dataset. The error function is the improved cross-entropy error function, where the training set is composed of 50,000 images, while the test set consists of 10,000 images, the batch size is 64, and the training is conducted for a total of 50 epochs.

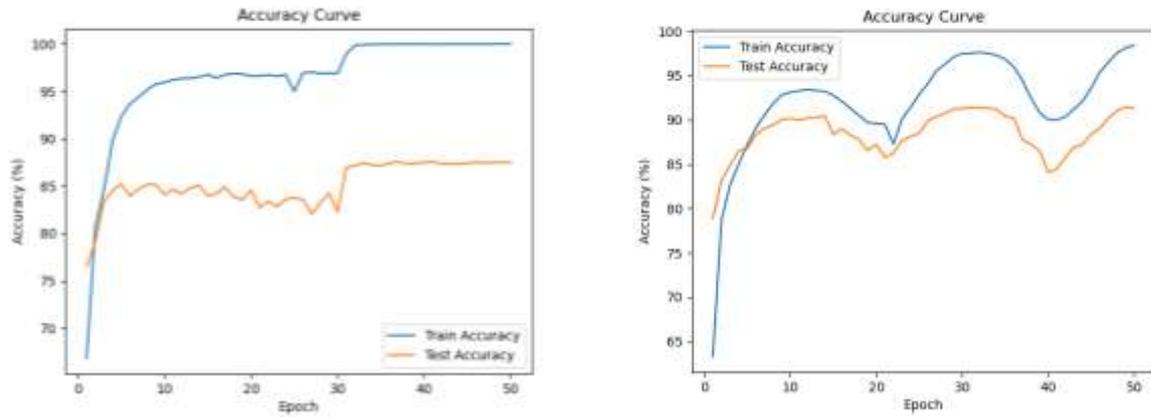
##### B. Assessment Criteria

Where C is the quantity of classes,  $q_i$  is target distribution after label smoothing, and  $P_i$  is the probability distribution predicted by the model. Label smoothing can reduce the model's overconfidence in certain categories, thereby advancing the model's generalization capacity. The fewer the lost-value, the greater the model's performance.

The confusion matrix is a visualization tool that intuitively displays the model's predictive performance in each category. It presents the relationship between the true labels and the predicted labels in matrix form. Diagonal values within the confusion matrix signify the quantity of correctly - labeled samples. In contrast, non - diagonal values represent the number of samples wrongly classified. In multi-classification tasks, the confusion matrix helps identify which categories are easily confused and provides guidance for model improvement.

##### C. Experimental Results and Analysis

This method uses transfer learning to apply the pre-trained basic network structure, weights, and bias parameters of ResNet50 to the New-ResNet50 model, efficiently saving training time and advancing the model's generalization ability. New model is trained and tested on the information-rich CIFAR-10 dataset with 10 categories of partial image data, and finally, the model is verified to achieve better image classification ability, with a classification accuracy of 93.75% on the test set. The precision and lost-value of the original ResNet50 network and the New-ResNet50 network during testing are compared, as shown in Figure 2 and Figure 3.



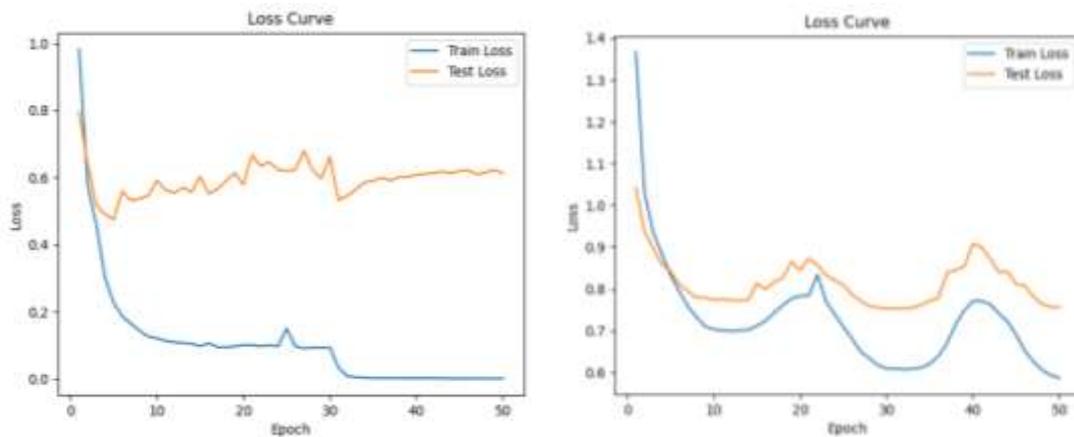
(a) ResNet50 network accuracy

(b) New-ResNet50 network accuracy

Figure 2. Comparison of accuracy between the two networks during training

Figure 2 illustrates the trend of validation accuracy during the coaching process for the initial ResNet50 and the improved New-ResNet50, with the horizontal axis representing the training epochs and the vertical axis representing the classification accuracy. Overall, the accuracy curve of New-ResNet50 consistently remains above that of the original ResNet50, indicating that the improved model exhibits stronger learning capability and generalization performance during training. Specifically, New-ResNet50 demonstrates significantly faster convergence in the early stages of training compared to the original ResNet50, achieving a test accuracy of 93.75% at 50 epochs, which is a 6.25% improvement over

the original ResNet50's 87.50%. This validates the effectiveness of the transfer learning strategy, improved loss function, and cosine annealing decay method. However, New-ResNet50 exhibits some fluctuations between the 20th and 40th epochs, which may be related to the learning rate adjustment strategy or increased model complexity. Despite these fluctuations, the overall performance of New-ResNet50 is significantly better than that of the original ResNet50, demonstrating the effectiveness of the proposed improvements. Future work could focus on optimizing the learning rate scheduling strategy or incorporating regularization techniques (such as Dropout or weight constraints) to further enhance model stability.



(a) ResNet50 network loss value

(b) New-ResNet50 network loss value

Figure 3. Comparison of loss values between the two networks during training

Figure 3 shows the trend of loss values during the training process of the original ResNet50 and the improved New-ResNet50. The x-axis represents the number of training epochs, while the y-axis symbolizes lost-value. Overall, the training loss of ResNet50 converges faster and reaches a lower final value, indicating better fitting ability on the training set. However, overfitting is relatively evident, as the validation loss increases slightly in the later stages. In contrast, the training loss of New-ResNet50 fluctuates more significantly, showing less stable training. Nevertheless, its validation loss remains at a lower level throughout the process, with a relatively lower degree of overfitting. This suggests that the improved model has enhanced generalization ability. Specifically, the training

loss of ResNet50 drops rapidly in the first 20 epochs and stabilizes after 50 epochs, with a low final training loss value. However, the validation loss increases slightly in the later stages, indicating potential overfitting. In contrast, the training loss of New-ResNet50 fluctuates more, especially between the 20th and 40th epochs. Yet, its validation loss is consistently lower than that of ResNet50 and reaches a lower level by the 50th epoch, confirming the effectiveness of the improvements in reducing overfitting.

To visually prove that the improved New-ResNet50 model has better image classification accuracy, this experiment also obtained the confusion matrix results of the two models during training, as shown in Figure 4.

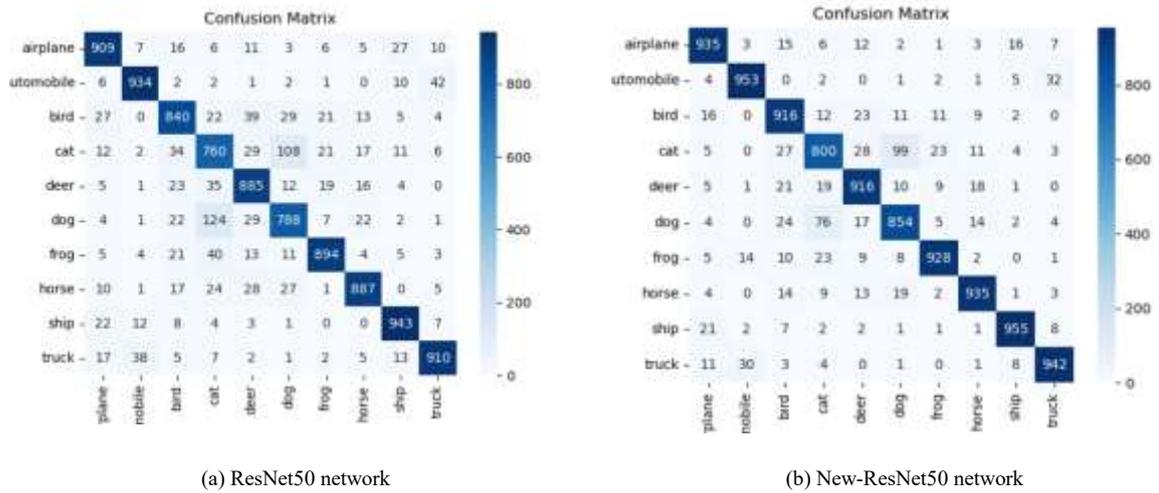


Figure 4. Comparison of confusion matrices between the two networks during training

Among them, the diagonal elements represent the number of correctly classified samples in each category, and the non-diagonal elements represent the misclassification situation. Overall, the New-ResNet50 network performs better on the confusion matrix than the ResNet50 network, with an increase in the number of correctly classified samples in multiple categories and a reduction in misclassification, indicating that the improvement measures have enhanced the model's classification accuracy to some degree. However, both models still have misclassification

problems in some similar categories, and future work can further optimize the model, such as by increasing data diversity and improving feature extraction methods, to enhance the capacity of the model to distinguish similar categories.

### V. CONCLUSIONS

This paper addresses the issues of low computational efficiency, large parameter volume, and the difficulty in balancing performance and efficiency in traditional ResNet50 models for image classification tasks

by proposing a ResNet50 image classification algorithm based on transfer learning. By employing transfer learning strategies for model initialization, introducing data augmentation techniques, and improving the loss function along with using cosine annealing decay for model training, the categorization accuracy of the model has been significantly enhanced. Experimental findings demonstrate that the suggested method improves the classification accuracy by 6.25% compared to the traditional ResNet50 image classification algorithm, achieving a classification precision of 93.75%, thereby validating the effectiveness of the proposed approach. This research not only demonstrates the significant role of transfer learning in image classification tasks but also provides new insights for optimizing deep learning models. By refining the loss function and incorporating cosine annealing decay, the study further explores optimization strategies during model training, offering theoretical references for related fields. In practical applications, the proposed method reaches a great harmony ranging from computational efficiency to classification accuracy, exhibiting extensive applicability in areas like medical image analysis, autonomous driving, and security surveillance, where it can significantly enhance system real-time performance and accuracy. Future research could further explore model lightweighting, cross-domain transfer, self-supervised learning, and multitask learning to improve model efficiency, generalization capability, and applicability, thereby promoting the application and development of deep learning technologies in more scenarios. It can be concluded that optimizing foundational models is highly necessary.

#### REFERENCES

[1] Gu Ruifan, Li Xiang, Ren Weimin. Research on Image Classification Based on Improved ResNet50 Model

- [J]. *Modern Electronics Technique*, 2023, 46(04): 107-112.
- [2] Liu Hongda, Sun Xuhui, Li Yibin, Han Lin & Zhang Yu. A Review of Deep Learning Models for Image Classification Based on Convolutional Neural Networks. *Computer Engineering and Applications*, 1-29.
- [3] Wang Xiuju, Fu Zhumu, Zhai Kunming, et al. Road Surface Condition Recognition Algorithm Based on Improved ResNet [J]. *Science Technology and Engineering*, 2024, 24(32): 14033-14040.
- [4] Liu Yanru, Wu Xiaohong, He Xiaohai, et al. Research on Core Image Classification Based on Improved ResNet50 [J]. *Intelligent Computer and Applications*, 2025, 15(02): 10-16.
- [5] Wu Di, Xiao Yan, Shen Xuejun, et al. Fruit Image Classification Based on Improved Res2Net and Transfer Learning [J]. *Journal of University of Electronic Science and Technology of China*, 2025, 54(01): 62-71.
- [6] Jiang Zhengfeng, He Tao, Shi Yanling, et al. Remote Sensing Image Classification Based on Convolutional Attention Mechanism and Deep Residual Network [J]. *Laser Journal*, 2022, 43(4): 76-81.
- [7] Zhang Yizhuo, Xu Miaomiao, Wang Xiaohu, et al. Hyperspectral Ground Object Classification Based on Hierarchical Fusion of Residual Networks [J]. *Spectroscopy and Spectral Analysis*, 2019, 39(11): 3501-3507.
- [8] Fang Liang, Zhou Yun, Tang Zhiquan. Rusted Steel Bar Image Classification Based on Optimized Residual Network [J]. *Journal of Northeastern University (Natural Science)*, 2021, 42(11): 1625-1633.
- [9] Pan Renyong, Zhang Xin, Chen Xiaoyulong, et al. Apple Leaf Disease Recognition Method Based on DTS-ResNet [J]. *Foreign Electronic Measurement Technology*, 2022, 41(09): 142-148.
- [10] Mahjoubi A M, Lamrani D, Saleh S, et al. Optimizing ResNet50 performance using stochastic gradient descent on MRI images for Alzheimer's disease classification [J]. *Intelligence-Based Medicine*, 2025, 11100219-100219.
- [11] Sun H, Zhou W, Yang J, et al. An Improved Medical Image Classification Algorithm Based on Adam Optimizer [J]. *Mathematics*, 2024, 12(16): 2509-2509.
- [12] Martins M F, González M V, Villar R J, et al. Inception networks, data augmentation and transfer learning in EEG-based photosensitivity diagnosis [J]. *Machine Learning: Science and Technology*, 2025, 6(1): 015034-015034.
- [13] Chatterjee S, Byun C Y. Leveraging generative adversarial networks for data augmentation to improve fault detection in wind turbines with imbalanced data [J]. *Results in Engineering*, 2025, 25103991-103991.
- [14] Yang Suorong, Yang Hongchao, Shen Furao, et al. A Review of Image Data Augmentation for Deep Learning[J/OL]. *Journal of Software*, 1-23 [2025-01-17].
- [15] Chu Yuezhong, Wang Jiaqing, Zhang Xuefeng, et al. Image Classification Algorithm Based on Improved Deep Residual Network [J]. *Journal of University of Electronic Science and Technology of China*, 2021, 50(02): 243-248.
- [16] Hu Kun, Wu Guoqing, Hu Zuhui, et al. Research on Metal Surface Defect Image Classification Based on Improved VGG16 Network [J]. *Computer Applications and Software*, 2024, 41(06): 175-180.