*Electronic Engineering and Informatics*
*G. Izat Rashed (Ed.)*

*doi:10.3233/ATDE240095*

# Research on Image Classification Method Based on Dual Network Feature Fusion

Jinzheng JIANG[a1], Wenjing LUO[b]
[a] *School of Information Engineering, Nanjing University of Finance & Economics, 3 Wenyuan Ave., Nanjing, JiangSu, 210023, China*
[b] *School of computer science and engineering, Shenyang Jianzhu University, 25 Hunnan East Road., Shenyang, Liaoling, 110000, China*

**Abstract.** Image classification has always been an important research topic in the field of computer vision. By designing different CNN network models, an increasing number of image classification applications have undergone significant changes, such as crop species recognition in agriculture, medical image recognition in the medical field, and vehicle recognition in transportation. However, most existed CNNs only use single model and rigid classification module to encode features and classify objects in the images, which resulting in semantic wasting and trapped in a fixed feature extraction pattern. Based on this, this article focuses on how to solve the problem of extracting features from insufficient attention regions in CNN network models by using deep learning to solve image classification problems. A dual network feature fusion model (DNFFM) is proposed to improve image classification results. DNFFM has a dual backbone networks, which extracts complementary non-redundant information from the feature layer of the backbone network through the fusion module of DNFFM, so that the entire network model has a broader and richer effective attention area, thus improving the accuracy of classification. DNFFM has achieved better results on CIFAR10, CIFAR100 and SVHN than a single backbone network. Reached 97.6%, 85.7% and 98.1% respectively. Compared with the original single network with the same backbone network, 2.4%, 2.9%, 1.6% and 2.2%, 3.2%, 1.3% are improved respectively. DNFFM has the following advantages: it is an end-to-end network that can extract more feature information when the data are the same ones, and has better classification results than a single network.

**Keywords.** Deep Learning, image classification, feature extraction, feature fusion

## 1. Introduction

Image classification is an important issue in the field of computer vision. In today's society, image classification has various applications, such as medical diagnosis, traffic management, safety monitoring, product detection, etc. Therefore, the importance of image classification is self-evident. However, there are also some technical challenges to image classification. For example, achieving precise classification requires sufficient data and computing resources. Due to the complexity of image patterns, many machine learning methods cannot achieve good accuracy in classification results.

[1] Corresponding author: Jinzheng JIANG, School of Information Engineering, Nanjing University of Finance & Economics, e-mail: 2629710763@qq.com

With the development of deep learning technology, some deep neural network models (such as Convolutional neural network) have become the mainstream methods of image classification tasks. Convolutional neural network (CNN) is a special neural network structure for processing image data. It can automatically learn features from data and has high-performance in image classification. Many Convolutional neural network models applied to image classification, such as AlexNet, VGG, GoogLeNet and ResNet. Those networks have achieved excellent performance in image classification.

In image classification technology, deep learning technology has achieved very good results on standard datasets. Nowadays, many numerous high-performance networks emerge one after another, and network models play the role of feature extractors in classification tasks. Different network models focus on different regions when extracting image features, resulting in different extracted features. This paper is based on this theoretical foundation. A dual network feature fusion model (DNFFM) was proposed, which combines the attention regions of different networks and removes redundant attention regions to expand the entire image's attention region, thereby extracting more discriminate feature information from limited image data and improving the classification accuracy of the model.

## 2. Related Work

As one of the important research directions in the field of computer vision, image classification is to divide the input images into different categories and minimize classification errors as much as possible. With the rise of deep learning technology, deep neural networks, as one of the most commonly used models in image classification, have become an important research subject in image classification related work. For example, designing different deep learning network structures to achieve image classification; Using deep networks to extract image deep semantic features and using deep semantic features to achieve image classification; Using convolution kernels of different sizes of the depth learning network to obtain different receptive field, and the feature information of the image under different receptive field is extracted for image classification. One of the most important factors affecting image classification is the performance of feature extractors. In recent years, more and more models for optimizing feature extraction have emerged, and image classification has further developed.

In 2020, Cen[1] et al. proposed using Enhanced Deep Feature Vectors (DFVs) to pretrain networks to improve classification accuracy. The original DFVs and pseudo DFVs form a set of augmented DFVs, where pseudo DFVs are generated by randomly adding differential vectors (DVs). On the ILSVRC2012 dataset, fine-tuning the ResNet50 network improved classification accuracy by 11.21%. In 2023, Wang[2] et al. highlighted the DS-ResNet50 model. This model adopts a dual scale hole convolution module, which extracts feature information from different scales and fuses them by cascading different hole convolution kernels. The classification accuracy on the ISIC2017 dataset improved by 0.88% compared to the ResNet50 model. In 2023, Gu[3] et al. proposed a residual network called DrANeT, which uses dynamic ReLU to automatically adjust parameters for feature maps, and embeds CBAM dual attention mechanism to improve the network's ability to extract useful special images. The experimental results show that the classification accuracy on the image dataset of benign lung images, Adenocarcinoma of the lung images and lung squamous cell carcinoma images reaches 100.00%, 99.96% and 99.96%. In 2023, Zhang[4] et al. highlighted the DC-DenseNet network model, integrating the extended convolutions on DenseNet into

dense blocks, thus achieving multiscale feature extraction. Compare with AlexNet, VGG-19, and DenseNet161 on the BACH dataset. Experimental results demonstrate that the proposed method effectively improves the classification accuracy of breast cancer, and the recognition rate reaches 94.10%. In 2023, Batchuluun[5] et al. proposed to train the Convolutional neural network utilizing the classification activation map as the ground truth to focus on specific areas in the input image in order to improve the classification accuracy. Batchuluun et al. used the Dongguk Thermal Image Database (DTh DB) for experiments, and the accuracy of using this method on DesseNet201 reached 97.64%, which is 1.64 percentage points higher than the original DesseNet201.

## 3. The Proposed Method

### *3.1 DNFFM network*

This paper proposes a dual network feature fusion model (DNFFM) to improve classification accuracy, as shown in Figure 1. The purpose of this network model is to extract more discriminative features for classification more effectively.

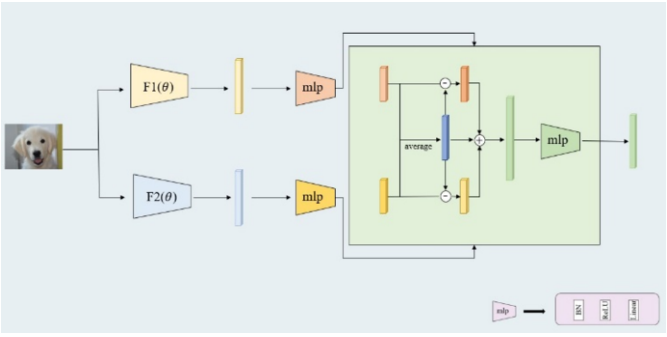


**Figure 1.** DNFFM Network Structure

DNFFM first selects the original backbone network, integrating CBAM[12] attention mechanism, fine-tunes the network structure of the backbone network, and obtains the F($\theta$)feature extractor. Using pretraining and Fine Tune methods to obtain the final feature extractor F($\theta$). Figure 2 displays the process of constructing a feature extractor.

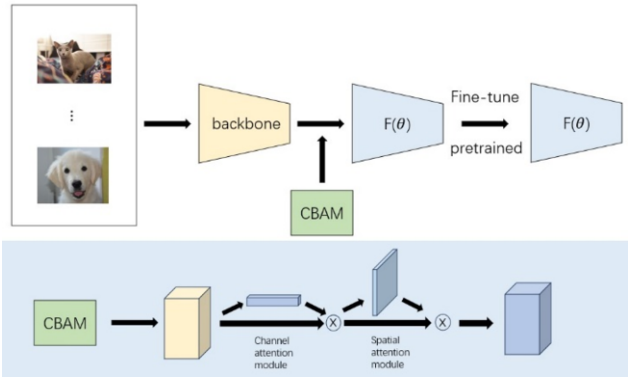


**Figure 2.** Construction process of feature extractor

### 3.2 DNFFM feature fusion module

Due to the rapid development of deep learning technology, different networks have emerged one after another, such as ResNet[6], VGG[7], DenseNet[8], Mobilenetv3[9], Regnet[10], Shufflenetv2[11], etc. These networks have already achieved good results in classification tasks. As is generally known, the features extracted by different networks from the same image are not the same, which means that the regions they pay attention to when extracting features are different. Therefore, each network has its own corresponding attention region. Based on the prior knowledge mentioned above, this paper proposes a dual network feature fusion method, using dual networks as two sets of feature extractors to extract two sets of features from the same image. Due to the possibility of redundancy in extracting features with discriminative from the same image using different feature extractors. DNFFM adopts the fusion method of taking the average vector $f_\theta^{avg}$ of two sets of feature vectors $f_\theta^1$ and $f_\theta^2$ as the redundant reference vector. The two sets of vectors remove the redundant reference vector and retain their unique feature information, and combine the redundant reference vector $f_\theta^{avg}$ with their own unique discriminative information vectors $f_\theta^1$ and $f_\theta^2$ to solve the problem of feature information redundancy. The feature fusion formulas are shown in Eqs. (1), (2), and (3):

$$f_\theta^{avg} = \tau\left(\frac{f_\theta^1 + f_\theta^2}{2}\right) \tag{1}$$

$f_\theta^1$, $f_\theta^2$ are the feature vectors corresponding to feature extractors F1($\theta$)and F2($\theta$). $\tau$ is Proportional coefficient.

$$f_\theta^{1^s} = \lambda(f_\theta^1 - f_\theta^{avg}) \tag{2}$$

$$f_\theta^{2^s} = \gamma(f_\theta^2 - f_\theta^{avg}) \tag{3}$$

$f_\theta^{1^s}$, $f_\theta^{2^s}$ are the features vectors obtained by eliminating redundant reference vector for$f_\theta^1$ and $f_\theta^2$, which preserves the unique feature information of their respective feature extractors. $\lambda, \gamma$ are Proportional coefficients. Connect$f_\theta^{1^s}$, $f_\theta^{2^s}$, and $f_\theta^{avg}$ to obtain the final fusion vector.

Due to the different feature dimensions extracted by different feature extractors, DNFFM has designed three mlp modules, each consisting of BN layer, ReLU layer, and Linear layer, to solve the problem of feature dimension matching.

## 4. Experimental Results and Discussions

This paper selected three datasets (CIFAR10, CIFAR100, SVHN) for the experiment. Six networks with superior performance were selected for experimental comparison. The dual network module of DNFFM uses ResNet50 and DenseNet121 as the backbone network.

### *4.1 Experiments based on CIFAR10*

The first group of experiments was conducted using CIFAR10 as the dataset. The results are shown in Table 1. In terms of classification results, DNFFM has significant improvement compared to other networks, reached 97.6%. Compared with the independent network classification results of the backbone network ResNet50 and DenseNet121, DNFFM has increased by 2.4% and 2.2% respectively.

**Table 1.** Experimental Results on CIFAR10

| Method | Classification accuracy |
|---|---|
| ResNet50 | 95.2% |
| Vgg16 | 93.2% |
| DenseNet121 | 95.4% |
| MobilenetV3_large | 93.8% |
| Regnet_y_400MF | 93.9% |
| Shufflenet_v2_x0_5 | 88.4% |
| **Ours (DNFFM)** | **97.6%** |

### *4.2 Experiments based on CIFAR100*

The results of the second group of experiments are given in Table 2. This experiment uses CIFAR100 as the experimental dataset. Experimental data shows that DNFFM with ResNet50 and DenseNet121 as backbone networks performs best, reached 85.7%. Compared to the results of using ResNet50 and DenseNet121 classification alone, DNFFM improved by 2.9% and 3.2%.

**Table 2.** Experimental Results on CIFAR100

| Method | Classification accuracy |
|---|---|
| ResNet50 | 82.8% |
| Vgg16 | 76.2% |
| DenseNet121 | 82.5% |
| MobilenetV3_large | 77.2% |
| Regnet_y_400MF | 80.8% |
| Shufflenet_v2_x0_5 | 69.2% |
| **Ours (DNFFM)** | **85.7%** |

### *4.3 Experiments based on SVHN*

Table 3 demonstrates the results of the third group of experiments using the SVHN dataset. DNFFM has achieved better performance than other networks on this dataset, reached 98.1%. DNFFM increased by 1.6% and 1.3% compared to ResNet50 and DenseNet121.

**Table 3**. Experimental Results on SVHN

| Method | Classification accuracy |
|---|---|
| ResNet50 | 96.5% |
| Vgg16 | 95.4% |
| DenseNet121 | 96.8% |
| MobilenetV3_large | 95.8% |

| | |
|---|---|
| Regnet_y_400MF | 95.1% |
| Shufflenet_v2_x0_5 | 93.2% |
| **Ours (DNFFM)** | **98.1%** |

### *4.4 DNFFM performance analysis*

In order to verify the robustness and generalization ability of DNFFM, this paper analyzes the training and validation losses of DNFFM on three datasets (CIFAR10, CIFAR100, SVHN). As illustrated in Figure 3. Wherein, (a), (b) and (c) correspond to the Loss function graphs of DNFFM on CIFAR10, CIFAR100 and SVHN respectively. It is not difficult to see from the three loss curve graphs that DNFFM has not overfitting phenomenon, verifying its superior generalization ability.

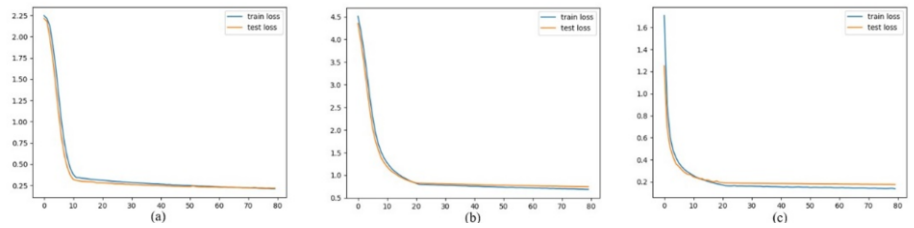


**Figure 3**. Loss function of DNFFM on (CIFAR10, CIFAR100, SVHN)

## 5. Conclusion

This paper investigates the problem of image classification based on public databases. This paper proposes a dual network feature fusion model (DNFFM) based on the idea of feature fusion and prior knowledge of attention regions and feature extraction relationships. DNFFM combines the features extracted from different attention regions of the dual backbone network, eliminates redundancy and extracts useful feature information

This paper uses CIFAR10, CIFAR100, and SVHN datasets. They reached 97.6%, 85.7%, and 98.1% respectively. The classification accuracy of DNFFM is 2.4%, 2.9%, 1.6% higher than that of the single backbone network ResNet50 of DNFFM, and 2.2%, 3.2%, 1.3% higher than that of the single backbone network DenseNet121 of DNFFM. And the model has good generalization ability.

## Acknowledgements

This work was supported by the Natural Science Foundation of Jiangsu Province Graduate Research and Practice Innovation Project in 2022 (Grant No. KYCX22_1700).

## References

[1] Cen Feng, XZ, WL, GW., "Deep feature augmentation for occluded image classification - ScienceDirect." Pattern Recognition 111 (2020).

[2] Wang Shiwei, Chen Jun, Yi Caijian., "Skin lesion image classification based on improved ResNet50." Software Engineering 26.06: 50-54 (2023).
[3] Gu Yu, Li Simin, Sharon Cheung, Yang Lidong, Lv Xiaoqi, Zhang Xiangsong, Jia Chengyi, He Qun., "Lung cancer pathological image classification based on improved residual network and dynamic ReLU" Laser Journal 44.05: 154-161 (2023).
[4] Zhang Miaolin, Shuai Renjun., "Pathological image classification of breast cancer based on DC-DenseNet" Computer Applications and Software 40.04: 116-121 (2023)
[5] Batchuluun Ganbayar, Choi Jiho, Park Kang Ryoung., "CAM-CAN: Class activation map-based categorical adversarial network." Expert Systems With Applications 222 (2023).
[6] He, Kaiming , X Zhang， S Ren， J Sun., "Deep Residual Learning for Image Recognition." IEEE DOI:10.1109/CVPR.2016.90 (2016).
[7] Simonyan, Karen, A. Zisserman., "Very Deep Convolutional Networks for Large-Scale Image Recognition." Computer Science DOI:10.48550/arXiv.1409.1556 (2014).
[8] Huang, Gao , Z Liu， VDM Laurens， KQ Weinberger., "Densely Connected Convolutional Networks." IEEE Computer Society DOI:10.1109/CVPR.2017.243 (2016).
[9] Howard, Andrew, et al. "Searching for MobileNetV3." 2019 IEEE/CVF International Conference on Computer Vision (ICCV) IEEE DOI:10.48550/arXiv.1905.02244 (2020).
[10] Radosavovic, I., Kosaraju, R. P., Girshick, R., He, K., "Designing Network Design Spaces." Computer Vision and Pattern Recognition IEEE DOI:10.1109/CVPR42600.2020.01044 (2020).
[11] Ma, N., Zhang, X., Zheng, H.-T., Sun, J., "ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design." Lecture Notes in Computer Science, 122–138. DOI:10.1007/978-3-030-01264-9_8 (2018).
[12] Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S., "CBAM: Convolutional Block Attention Module". Lecture Notes in Computer Science, 3–19. doi:10.1007/978-3-030-01234-2_1 (2018)