Proceedings of CONF-MPCS 2024 Workshop: Quantum Machine Learning: Bridging Quantum Physics and Computational Simulations
DOI: 10.54254/2753-8818/52/2024CH0133

Research on Art Classification Based on the Improved Resnet50 Network

Yingkai Liang[1], Hangxi Ren[2,5], Ziwei Wan[3], Rui Zhou[4]

[1]Helen Way Klingler College of Arts and Sciences, Marquette University, Milwaukee,USA
[2]Software College, Xinjiang University, Urumqi, China
[3]College of Information Engineering, Sichuan Agricultural University, Sichuan, China
[4]College of Computer Science, Nanjing University of Post and Telecommunication, Nanjing, China

[5]20222501439@stu.xju.edu.cn

Abstract. In the field of computer vision, deep learning technology has made breakthroughs in image classification. ResNet50, an efficient deep residual network architecture, has demonstrated excellent performance in many image classification tasks. However, In the task of classifying artworks, ResNet50, still shows some limitations. It has difficulty identifying different artistic styles and is lacking in capturing the delicate details of the artwork. This article proposes an improved resnet50 model. By introducing SE blocks and extended convolution parameters, this article have made a deep revamp to ResNet50 to enhance its performance in art classification tasks. SE blocks increase the sensitivity of the network to differences in art styles by dynamically adjusting the dependencies between channels. Concurrently, the enhanced convolutional operation expands the model's sensory scope, so that the network can capture more delicate artistic details. The experimental results show that the improved model has achieved significant performance improvement on wikiart artwork classification datasets. In upcoming research, the paper plan to enhance art classification by broadening the dataset to encompass diverse art movements and forms, and by optimizing convolutional neural networks for large-scale data. This aims to improve the model's generalization and applicability across various visual arts, ensuring practical effectiveness and reliability.

Keywords: Computer vision, Image classification, Resnet 50, Residual network.

1. Introduction

Artwork classification is an important application field in computer vision [1], and automatic classification of artworks helps with their identification, management, and protection [2]. Traditional image classification methods often perform poorly due to the complexity and diversity of images when processing art images [3].

In traditional image classification methods, researchers usually need to manually extract image features and formulate classification rules. This method often appears to be unsatisfactory when processing large -scale and diverse image data, not only low efficiency, but also difficult to guarantee accuracy [4]. However, with the rapid development of deep learning technology, Particularly the

DOI: 10.54254/2753-8818/52/2024CH0133

efficient use of convolutional neural networks (CNNs) in the realm of image classification, these problems have been effectively resolved [5]. CNN has greatly improved the accuracy and efficiency of classification through the hierarchical structure characteristics of automatic learning images [6].

However, art images have unique characteristics, such as complex texture, diverse styles and delicate details. These characteristics make art classification tasks more challenging. Traditional ResNet50 may not be able to fully capture abstract information such as details and styles in art images when processing these features. Therefore, the goal of this experiment is to use the deep learning framework of PyTorch in the field of computer vision to complete the image classification tasks in specific scenarios by deep learning neural network ResNet, focusing on improving and optimizing some modules of the Resnet network [7]. To adapt to specific application scenarios to achieve higher classification accuracy and better generalization capabilities [8].

This experiment was developed based on the PyTorch library of Python. Pytorch is a widely used opensource machine learning library. Especially in the field of deep learning, it is favored by researchers and developers for its dynamic calculation diagram and ease of use. Pytorch provides rich APIs and tools, making the construction and training deep learning models intuitive and efficient.

During this experiment, the structure of the resnet network was first analyzed in depth to identify key modules that may affect classification performance. Then, try different improvement strategies, such as adjusting the depth and width of the network, introducing attention mechanisms, optimizing the initialization of the network, etc., in order to improve the expression ability and generalization of the model. In addition, the number of overpairies in data pre -processing, enhancement strategies, and training in the training process has been carefully adjusted and optimized.

2. Related works

Image classification is an important research direction in the field of computer vision, which aims to classify images into predefined categories. With the development of deep learning technology, the image classification has made significant progress [9].

Before the popularization of deep learning, image classification mainly depends on handmade characteristics extraction and traditional machine learning algorithms. Popular feature detection approaches encompass the Scale-Invariant Feature Transform(SIFT) method, Histogram of Oriented Gradients, Local Binary Patterns(LBP) and so on. These methods are represented by the local characteristics or global characteristics of the image, and then the image is represented as a feature vector, following that, algorithms such as Vector Support Machines (SVM), Nearest Neighbor (NN), and Stochastic Forests are applied for classification.

However, most of these traditional methods have these limitations: handmade feature extraction of the complexity and diversity of images cannot be completely captured; these methods have lower efficiency when dealing with large -scale data sets; these methods have noise and noise and noise in the image in the image. Deformation is more sensitive and has poor robustness [10].

In recent years, the emergence of deep learning methods, especially the convolutional neural network (CNNS), has greatly promoted the progress of image classification [11]. Deep learning methods can automatically learn the characteristics from the data, avoiding the limitations of handmade feature extraction. Typical deep learning models include Lenet, Alexnet, VGG, Googlenet and Resnet [12].

The RESNET is proposed by the Microsoft Research Institute [13] that the residual connection technique is employed to address issues of vanishing and exploding gradients within deep neural network architectures during the training process, so that the network can be deeper and easy to train. Resnet has achieved leading performance in multiple image classification benchmark tests.

In the field of art classification, researchers have proposed a variety of methods and technologies. Early studies primarily employed a method that relies on manual feature extraction and traditional machine learning approaches, including Support Vector Machines (SVM) and K-Nearest Neighbors (KNN), along with others. Although these methods have achieved certain effects in certain tasks, their classification performance is often not ideal due to the limitations of manual characteristics.

DOI: 10.54254/2753-8818/52/2024CH0133

With the development of deep learning technology, models based on convolutional neural networks have gradually become the mainstream method of art classification. Classic network structures such as VGG, Inception, and ResNet are widely used in various image classification tasks. Especially for resnet, by introducing residual connections, the disappearance of gradient disappearance in deep networks enables the network to extract image features deeper [14]. In the art classification task, many researchers also adopted Resnet as the basic model and improved on this basis.

However, directly applying standard ResNet50 models still have some problems in art classification tasks. For example, the model is easy to ignore the details of the details when processing artistic images with complex texture and diverse style. In addition, the ResNet50 framework underutilized the hierarchical feature nuances present in the image. Therefore, how to improve on the basin of ResNet50 to better adapt to artistic classification tasks and become a problem worth studying.

3. Research methods

Dilated Convolution, this study broadens the perspective by incorporating the parameter count or loss resolution within the convolutional core, thereby augmenting the model's feature extraction capabilities. The realization of expansion convolution is as shown in formulas (1):

$$Oi\text{dilated conv} = siOi\text{in} + 2pi - di \times (ki - 1) + 1 \tag{1}$$

Oidilated conv:Output size of the dilated convolution. si: Size of the input feature map. Oi: Size of the output feature map. pi: Padding size of the input feature map. di: Dilation rate. Ki: Dimension of the convolutional filter.

The convolution workflow and related parameters are shown in Figure 1.

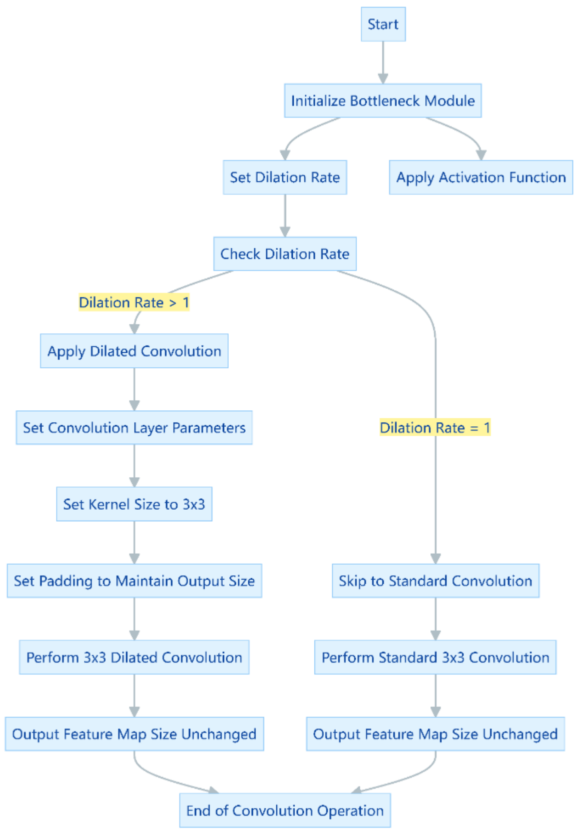

Figure 1. Convolution workflow and related parameters

Add Squeeze-And-EXCity Block. SE blocks adjust the weight of the characteristic channel through adaptive land, to ensure our models prioritize relevant features while discarding irrelevant ones, thereby

DOI: 10.54254/2753-8818/52/2024CH0133

enhancing the model's discriminative power exponentionally, while maintaining lower calculation costs and good versatility.

$$\mathrm{uc} = \mathrm{Fscale}(\mathrm{Fexcite}(\mathrm{Freduce}(X)), X) \tag{2}$$

X**:** Enter the feature diagram to contain multiple channels.

Freduce: Squeeze operation is usually achieved by global average pooling, which is used to compress the activation value of each channel into a single value to capture the activation distribution of the channel.

Fexcite: Excite operation is usually achieved through a full connection layer (or "squeeze" layer) to learn the dependence between channels and generate a weight for each channel.

Fscale: SCALE operation will multiply the weight of the operation generated by the input characteristics, so as to obtain the output feature diagram.

The SE block introduces an attention mechanism to the network in this way, so that the network can pay more attention to the more important feature channel for the current task, thereby improving the performance of the network.

Se block is shown in Figure 2.

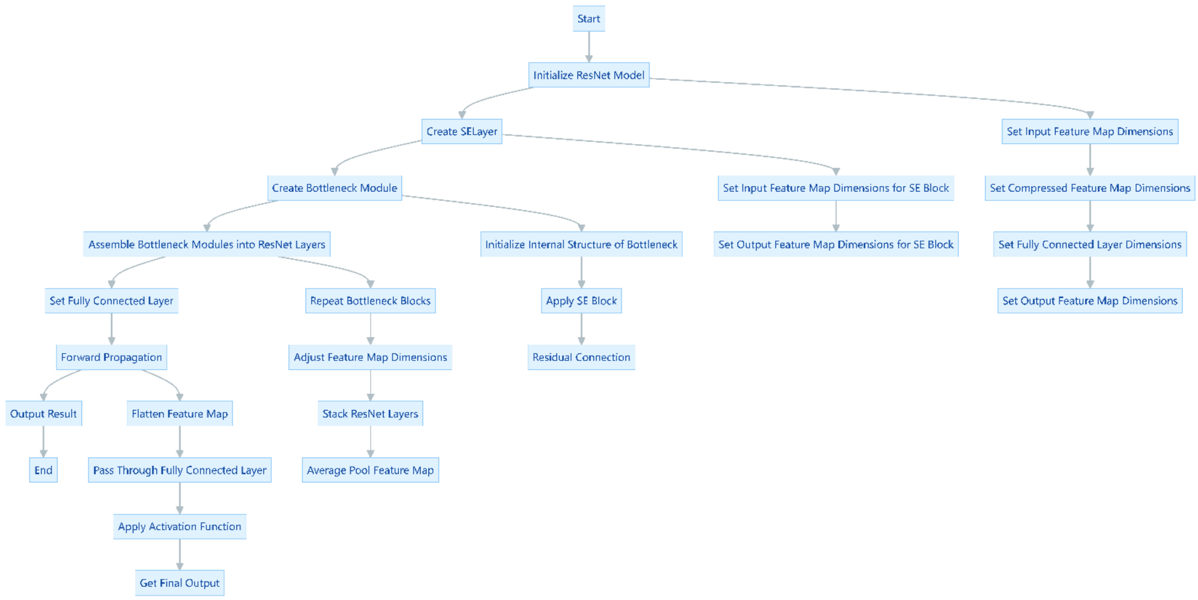

Figure 2. SE block

Data enhancement technology: The pre-training weight of the official website was used. The pre-training weight of the official website can not only significantly improve the ability to extract feature extraction, but also accelerate the training process, improve the accuracy and robustness of the model, and reduce overfitting risks and calculation costs.

4. Experimental results

4.1. data set

In order to verify the effectiveness of the model, and to improve the data quality, improve the processing efficiency and simplify the model training, the paper preprocessed the wikiart dataset and discarded some files that could not be decoded, so as to achieve the above goal.

Wikiart dataset shows the rich diversity and historical depth of art with its large scale of more than 81000 art images. Spanning a wide range of periods from the Renaissance to modern art, it records the evolution of artistic styles, covering 27 different artistic styles, such as abstract expressionism, Baroque, cubism and Impressionism. Each style is a unique chapter in the history of art. These works are from

DOI: 10.54254/2753-8818/52/2024CH0133

the hands of hundreds of artists, who belong to different cultural and historical backgrounds, and show a variety of artistic perspectives and innovative techniques.

The data set of more than 60 art genres has further enriched its content, including traditional portraits, landscape paintings and modern street art, which not only reflects the diversity of artistic expression, but also provides an opportunity to study the interaction and integration between different genres. The universality and application potential of Wikiart data set lies in its extensive art coverage. It is not only limited to the study of art history, but also provides training data for tasks such as image recognition, style classification, and art creation in the field of machine learning, becoming a bridge connecting art and technology. This interdisciplinary resource is a valuable asset for educators, students, artists, researchers and technology developers.

4.2. Evaluation index

4.2.1. Precision. Precision is the most intuitive evaluation index, which measures the proportion of the number of samples correctly predicted by the model to the total number of samples. The accuracy rate is very effective for the situation of uniform data distribution, but when the number of samples in different categories varies greatly, the accuracy rate may be misleading, because it will not distinguish whether the model performs well in a few categories.

4.2.2. Accuracy. The accuracy metric represents the ratio of the true positive instances among all instances predicted as positive. Precision is particularly crucial in scenarios where minimizing false positives (erroneously predicting the impending negative class as positive) is of paramount importance.

4.2.3. Recall rate. The recall metric quantifies the ratio of true positive instances that are accurately identified as such by the model. The recall is especially critical in scenarios where the minimization of false negatives (that is, the incorrect classification of positive instances as negative) is a primary concern.

4.2.4. F1-score. F1-score is the harmonic average of accuracy rate and recall rate. It tries to find a balance between accuracy rate and recall rate. When the accuracy rate and recall rate are both high, F1 score will also be high, which means that the model performs well in both aspects.

4.3. Experimental results and analysis

The experimental results show that the accuracy and other indicators of the improved resnet50 model are improved compared with other models. After a hundred rounds of training, the test indicators obtained are as follows:

Table 1. performance of the improved algorithm, the original algorithm and other algorithms on different datasets

Model name used	Training Accuracy	Validation Accuracy	Precision	Recall	F1 Score
Improved resnet-50	0.9998	0.6455	0.6303	0.6226	0.6206
Original resnet-50	0.9774	0.6071	0.5959	0.5828	0.5840
vgg-16	0.774	0.5365	0.5297	0.5163	0.5368
ConvNext	0.954	0.6032	0.5986	0.5878	0.5923

Table 1 shows the test performance of different neural networks after training. It can be seen that compared with the original resnet-50, vgg-16 and CNN networks, the performance of the improved resnet-50 has been improved to a certain extent, especially compared with the original resnet-50, the verification accuracy of the improved network has been improved by about 4%. In addition, compared with other neural networks, the improved resnet-50 has a training accuracy of 0.9998, which means that the model can be nearly completely correct on the training data set. The very high accuracy shows that

DOI: 10.54254/2753-8818/52/2024CH0133

the model performs well on the training set, which further shows that the model shows strong pertinence in predicting the data of known categories.

5. Conclusion

This paper proposes an improved resnet50 model to improve its performance in art classification tasks. Through the introduction of multi-level feature fusion mechanism, attention mechanism and a variety of data enhancement technologies, the improved model has shown significant performance improvement on multiple public art classification datasets. Specifically, the extended convolution mechanism enhances the feature capture ability of the model; Adding Se blocks enhances the attention of the model to important features and further improves the accuracy of classification; Data enhancement technology increases the diversity of training samples and improves the generalization ability and robustness of the model.

The experimental results show that the improved ResNet50 model outperforms the original ResNet50 model and other existing models in terms of classification accuracy, precision, recall, and F1 score on the WikiArt, Rijksmuseum, and Art500K datasets, validating the effectiveness of the improved method.

In addition, this study also has certain practical application value. The improved model can be applied to fields such as museums, art galleries, and art trading platforms to assist in automatic classification and management of artworks, improving work efficiency. At the same time, this model can also provide technical support for art appraisal and protection, promoting the digital protection and inheritance of cultural heritage.

The experimental results show that the ResNet model built using the PyTorch framework has achieved satisfactory training results in both local environments and cloud platforms. The model has demonstrated high accuracy in image classification tasks in specific scenarios, verifying the effectiveness of the improved strategy. Meanwhile, the model also demonstrated good generalization ability when faced with new data, providing a solid foundation for practical applications.

In summary, this experiment successfully improved the performance of image classification tasks through the application of deep learning and PyTorch framework, providing new ideas and methods for research and application in the field of computer vision. In the future, it may be possible to continue exploring deeper network structures and more advanced optimization strategies in order to achieve greater breakthroughs in the field of image classification.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Surapaneni, S., Syed, S., & Lee, L. Y. (2020). Exploring themes and bias in art using machine learning image analysis. In *2020 Systems and Information Engineering Design Symposium (SIEDS)* 1-6.
[2] Tan, W. R., Chan, C. S., Aguirre, H., & Tanaka, K. (2019). Improved ArtGAN for Conditional Synthesis of Natural Image and Artwork. IEEE Transactions on Image Processing, 28(1), 394-409.
[3] Qian, X. C., & Mi, S. (2019). Color style feature extraction of thangka based on deep neural network. Computer Science and Application, 9, 2129.
[4] Cetinic, E., Lipic, T., & Grgic, S. (2018). Fine-tuning convolutional neural networks for fine art classification. *Expert Systems with Applications*, *114*, 107-118.
[5] Koonce, B., & Koonce, B. (2021). ResNet 50. Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization, 63-72.
[6] Zhao, Z. C., Luo, Z., Wang, P. Y., & Li, J. (2019). A review of image classification algorithms based on deep residual networks. *Computer System Applications*, *29*(1), 14-21.
[7] Sandoval, C., Pirogova, E., & Lech, M. (2019). Two-stage deep learning approach to the classification of fine-art paintings. *IEEE Access*, *7*, 41770-41781.

DOI: 10.54254/2753-8818/52/2024CH0133

[8] Li, Q., Ji, Y. B., Guo, C. W., & Liu, K. X. (2022). Pattern recognition and classification of women's shirts based on deep learning. *Advanced Textile Technology*, *30*(4).
[9] Zhu, J. (2022). SECure: SE block-based classification of buildings post hurricane. In International Conference on Cloud Computing, Performance Computing, and Deep Learning (CCPCDL 2022) 12287, 284-288.
[10] Guo, Y. X., Yang, W., Liu, Q., & Wang, Y. (2020). A Review of Residual Network Research. *Application Research of Computers/Jisuanji Yingyong Yanjiu*, *37*(5).
[11] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) 770-778.
[12] Yang, H. J., Wang, E., Sui, Y. X., Yan, F., & Zhou, Y. (2022). Simplified residual structure and fast deep residual network. *Journal of Jilin University (Engineering Edition)*, *52*(6), 1413-1421.
[13] Cheng, D. Q., Wang, Y. C., Kou, B. Q., Fu, X. Z., Chen, L. L., & Zhao, K. (2021). Mine image classification based on improved deep residual network. *Application Research of Computers/ Jisuanji Yingyong Yanjiu*, *38*(5).
[14] Li, X., Su, J. N., Zhang, Z. P., & Zhou B. Y. (2024). Fine recognition of product style based on deep learning. *Computer integrated manufacturing system*, *30*(3), 1011.