Research on Image Classification Based on ResNet

Yuheng Wang

Computer School, Beijing Information Science And Technology University, Beijing, 100101, China

yhwang@gotolianshun.com

Abstract. This paper introduces the importance of image classification in computer vision. It aims to classify input images into different categories. Traditional image classification methods use manual feature extraction or feature learning to describe images but it is difficult to reveal deep semantic abstract features. It also requires a lot of manual work. This paper proposes an improved ResNet image classification model that solves problems such as computational complexity and overfitting. The proposed method uses smaller convolutional kernels. Data augmentation techniques are also implemented to improve network performance. By doing so, the algorithm achieves higher accuracy. Results of experiments on CUB200-2011 dataset demonstrate that the improved ResNet model achieves a validation accuracy of 95.50%, significantly outperforming other models. However, some overfitting is observed, indicating the need for further research. The results show the capability of deep learning methods, especially for ResNet model in image classification tasks.

Keywords: ResNet, Image Classification, CNN, Data preprocessing, Deep Learning

1 Introduction

Image categorization is a significant challenge in the domain of visual computing. This field aims to classify input images into different categories[1].

Inspired by biological functions, neural networks[2] are combinations of certain numbers of neurons. They are applied in many fields such as face recognition and medical diagnosis. Neural networks have strong adaptability and learning ability,

P. Kar et al. (eds.), *Proceedings of the 2023 International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2023)*, Advances in Computer Science Research 108,
https://doi.org/10.2991/978-94-6463-300-9_98

non-linear robustness, and error correction ability.

Traditional image classification methods used manual feature extraction or feature learning to globally describe images and relied on classifiers[3] to determine whether an image contained a certain object. There was a lot of research in this area with algorithms such as SIFT-based image feature matching, Principal Component Analysis (PCA) and HOG-based matching in remote sensing images, SURF and RANSAC-based image matching[8].

Nevertheless, those approaches were limited to extracting basic characteristics like hue, pattern, and contours, which posed challenges in revealing profound, meaningful features and demanded considerable manual labor.

Convolutional Neural Networks (CNN)[5] became the mainstream method for processing image classification problems[2,4]. However, CNN networks[5] could suffer from long model training times, overfitting and imbalanced datasets, all of which could affect model performance and precision. In order to tackle these concerns, the ResNet model was proposed and widely applied in the field of image classification[6]. The ResNet model introduced residual blocks and skip connections, allowing the network to be deeper and wider, thereby improving model accuracy[6]. Unlike traditional CNN networks, each residual block[7] in the ResNet model contained skip connections, allowing information to be directly conveyed from the entrance layer to the outcome layer and reducing information loss within the network[2,6].

Given these challenges, further research and exploration on the ResNet model is required to improve model performance and accuracy.

2 Method

2.1 Convolutional Neural Network

The operating principle of CNN is performing convolution operations on input data through convolutional layers, extracting their feature information[5]. The convolution operation is achieved by convolving filters of different sizes (also known as convolution kernels). By using multiple convolutional layers, CNN can automatically discover deeper abstract features in input data[2]. After the convolutional layer, To reduce computational complexity, CNN utilizes a pooling layer to compress the feature map size. Pooling is usually achieved by down-sampling the characteristics in an area, such as average pooling or maximum pooling.

By using these components, CNN can automatically learn useful features of input data.

2.2 Residual Network

ResNet is a deeper convolutional neural network model. The core module of ResNet is Residual Block, which increases network depth by introducing residual connections.

In traditional convolutional neural networks, due to multi-layer convolutional operations, information transmission in deep networks becomes very difficult, and signals may disappear or explode, resulting in slow or unable convergence of the network. The innovation of ResNet lies in retaining information spanning multiple layers through residual connections, allowing the network to be deeper and perform better. Fig. 1 is a structure diagram of the traditional ResNet network.

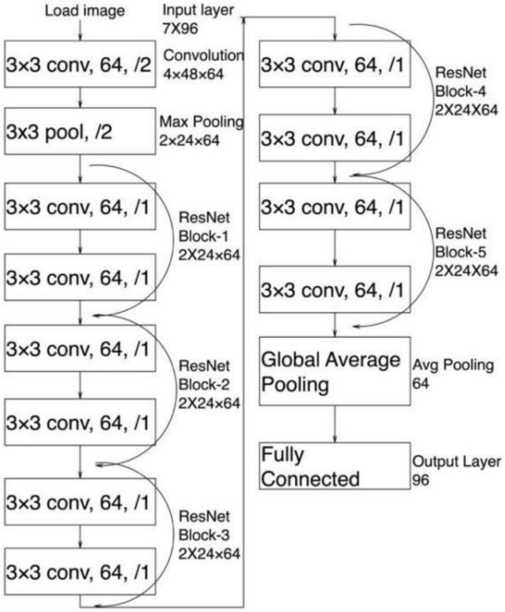

Fig. 1. Structure illustration of traditional Resnet network (Picture credit: Original)

In ResNet as shown in Fig. 1, each residual block holds a few convolutional layers. In the initial convolutional layer, the input feature map undergoes convolutional transformation to extract features across the extent and altitude of the picture. The second and third convolutional layers are used to further convolution the feature map and restore it to the same dimension as the input feature map. In order to realize residual connection, ResNet performs addition operations on the input and output between residual blocks, and performs Batch Normalization before activation function ReLU, which improves the robustness and transferability of the model.

Overall, the operating principle of ResNet is to introduce residual links to resolve the obstacles of gradient vanishing and exploding in deep networks, thereby achieving deep network expansion and excellent performance.

2.3 Data Preprocessing

This paper adopts a method of image data preprocessing to prepare image data sets for training and testing of deep learning models as shown in Fig. 2. The training set uses methods such as random rotation, random horizontal flipping, random cropping, and adding padding to enhance the diversity of the dataset, while the testing set only performs center cropping. This preprocessing method is aimed at improving training effectiveness of ResNet model and being able to recognize unseen images after training, which can be used in experiments on various classification tasks[9]. Fig. 2 is a schematic diagram of image types.

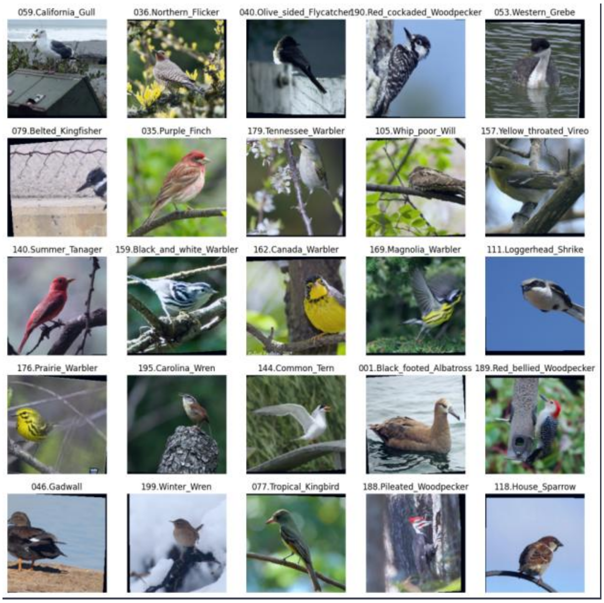

Fig. 2. Schematic diagram of image types (Picture credit: Original)

2.4 The Optimization Of Residual Units

In deep learning, larger convolutional kernels can extract more complex information in one step, but they also face the problem of slow model training and even inability to run on mobile devices due to a large amount of computation. Therefore, this paper replaces the large convolution kernel with smaller ones, as Fig. 3 shows.

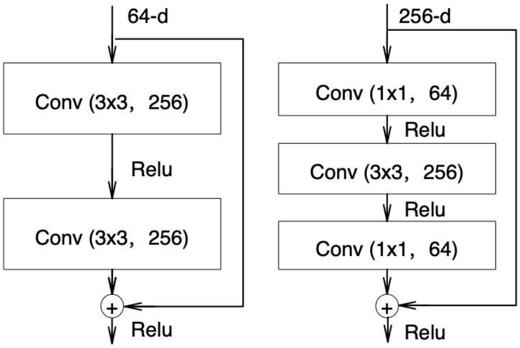

Fig. 3. shows the before and after network modification (Picture credit: Original)

Before the input passes through net1, an 1x1 convolution layer can be used to reduce the input dimension, greatly reducing the number of parameters to be calculated in net2. Using 3x3 convolutional kernels for computation in Net2 can extract richer feature information, although the computational complexity is relatively high. Before the output passes through net3, an 1x1 convolution layer can also be used to restore the output dimension. In this way, when adding the output and input, the problem of dimensional mismatch is avoided.

3 Results

3.1 Datasets

The CUB-200-2011 dataset [10], used in this study, is a collection of images featuring 200 different bird species. The dataset comprises 11,788 images, with approximately 60 images per bird species. In this paper, the data is partitioned into training and validation datasets with a proportion of 18 to 2. The validation set is employed to examine the training performance trends during each iteration.

3.2 Experimental Settings

This experiment uses four deep learning models to learn the proposed image representation, namely the VGG model, ALEXNET model, standard RESNET model and improved RESNET model proposed.

To reach highest experimental effect, This study employs discriminant fine-tuning to establish the learning rate of the model during the training phase. This is a technology of transfer learning. The latter layer in the model has a higher learning rate than the previous layer. The rate of learning found using the learning rate finder is used as the maximum learning rate, while the learning rate of the other layers is lower and gradually decreases with input. Fig. 4 shows the learning rate curve.

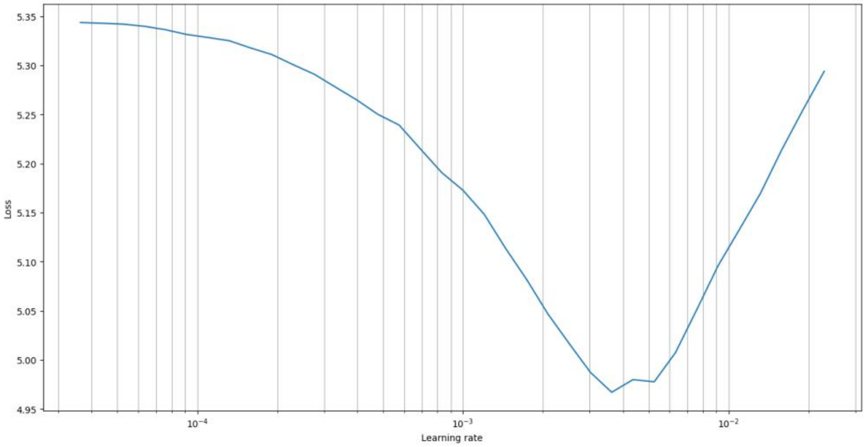

Fig. 4. Learning rate trajectory (Picture credit: Original)

According to Fig. 4, the loss reaches a minimum at around 0.003. A good learning rate to choose would be the middle of the steepest downward curve - which is around 0.001.

3.3 Image Classification Results

The test is conducted on 942 images, results shown in Table 1. This experiment uses four evaluation indicators: Train Loss, Train Accuracy, Validation Loss, and Validation Accuracy. In Table 1, it shows that compared to AlexNet, VGG, and ResNet, the improved ResNet this paper proposed has a higher recognition rate.

Table 1. The results of four models

model	Train Loss	Train Accuracy	Validation Loss	Validation Accuracy
AlexNet	0.637	78.10%	0.727	75.96%
VGG	0.082	97.25%	0.240	93.61%
ResNet	0.278	93.76%	0.396	91.07%
The modified ResNet	0.073	99.98%	0.733	95.50%

The performance of four models on the training and validation sets is demonstrated in Table 1. From the table, the AlexNet model has a relatively low training and validation accuracy, with values of 78.10% and 75.96%, respectively. This may be because the AlexNet model is relatively shallow and does not learn complex features well.

The VGG model has high training and validation accuracy, at 97.25% and 93.61%, respectively. This indicates that the VGG model can learn image features well, but its training and validation losses are relatively low, suggesting that there may be some overfitting.

The ResNet model has higher training and validation accuracy than the AlexNet model, at 93.76% and 91.07%, respectively. This indicates that the ResNet model can learn image features well, but compared to the VGG model, its training and validation losses have increased, suggesting that there may be some underfitting.

The modified ResNet model performs very well on both training sets and validation sets, with the training accuracy of 99.98% and the validation accuracy of 95.50%. Compared with the traditional ResNet model, VGG model, and AlexNet model, the modified model achieves a 25%, 2%, and 4.8% improvement in validation accuracy, respectively. However, its training loss is very low, but the validation loss is relatively high, suggesting that there may be some overfitting. Thus, techniques such as data augmentation and regularization may be employed to avoid overfitting.

The improved ResNet model replaces large convolution kernels with 7 convolution cores in all four stages, and achieves the same feature propagation by using y convolution kernels in the pooling layer. Data augmentation techniques such as randomly rotating, flipping horizontally, and cropping are added during data preprocessing. Compared to the original ResNet model, the proposed method effectively improves recognition efficiency. Primarily, this is due to the fact that data augmentation methods can enhance the model's ability to generalize and identify patterns even in intricate environments. Regarding convolution kernels, although large convolution kernels can extract the desired feature map size at once, they produce excessive computational complexity. The small convolution kernels used in this study can achieve the desired feature size through multiple feature extractions. In this stage, due to the deep iteration of small convolution kernels, some weakly correlated features are preserved, which greatly strengthens feature description and further improves network classification accuracy.

4 Conclusion

This paper propose a upgraded ResNet network to classify images. It achieves significant improvements in accuracy compared to traditional models such as AlexNet, VGG, and ResNet. The use of data augmentation techniques and smaller convolution kernels is explored to improve network performance and reduce computational complexity. The outcomes of the experiments conducted on the CUB-200-2011 dataset indicate that the ResNet model, after modification, attains a validation accuracy of 95.50%, surpassing other models by a substantial degree.

References

1. Hong, D., Han, Z., Yao, J., Gao, L., Zhang, B., Plaza, A., Chanussot, J.: SpectralFormer: Rethinking hyperspectral image classification with transformers. IEEE Transactions on Geoscience and Remote Sensing 60(1), 1-15(2021).
2. Hancock, J., Khoshgoftaar, T.: Survey on categorical data for neural networks. Journal of Big Data 7(1), 1-41 (2020).
3. Siddiqui, M., Morales-Menendez, R., Huang, X., Hussain, N.: A review of epileptic seizure detection using machine learning classifiers. Brain informatics 7(1), 1-18 (2020).
4. Shorten, C., Khoshgoftaar, T., Furht, B.: Deep Learning applications for COVID-19. Journal of Big Data 8(1), 1-54 (2021).
5. Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S.: Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing 173, 24-49 (2021).
6. Wen, L., Li, X., Gao, L.: A transfer convolutional neural network for fault diagnosis based on ResNet-50. Neural Computing and Applications 32(21), 6111-6124 (2020).

7. Fang, W., Yu, Z., Chen, Y., Huang, T., Masquelier, T., Tian, Y.: Deep residual learning in spiking neural networks. Advances in Neural Information Processing Systems 34, 21056-21069 (2021).
8. Bell, P., Fainberg, J., Klejch, O., Li, J., Renals, S., Swietojanski, P.: Adaptation algorithms for neural network-based speech recognition: An overview. IEEE Open Journal of Signal Processing 2(1), 33-66 (2020).
9. Wang, S., Celebi, M., Zhang, Y., Yu, X., Lu, S., Yao, X., Zhou, Q., Miguel, M., Tian, Y., Gorriz, J., et al.: Advances in data preprocessing for biomedical data fusion: An overview of the methods, challenges, and prospects. Information Fusion 76, 376-421 (2021).
10. The Caltech-UCSD Birds-200-2011 Dataset, https://authors.library.caltech.edu/27452/, last accessed 2023/6/14.