

Application of convolutional neural networks in image classification and applications of improved convolutional neural networks

Taoyu Liu

School of Computer Science and Engineering, Central South University, Changsha, China

8208210402@csu.edu.cn

Abstract. This paper reviews the application and improvement of convolutional neural networks (CNNs) in image classification. Firstly, a shallow CNN for interstitial lung disease image classification is presented. This model suppresses overfitting through a unique network architecture and optimisation algorithm. Next, the improved VGG16 architecture and MIDNet18 model are discussed and their superior performance in brain tumour image classification is demonstrated. Subsequently, a CNN-CapsNet model for cervical cancer image classification and its improvement are presented and the customised model is compared with the conventional VGG-16 CNN architecture in the paper. Next, the application of sparse convolutional kernels and hybrid sparse convolutional kernels (HDCs) in solving the problem of computational resource consumption is presented. Subsequently, methods for solving the problem of limited training data through transfer learning and network data augmentation techniques are discussed, as well as GAN-generated datasets for solving the overfitting problem. Finally, the effect of degraded images on the classification effectiveness of CNNs is explored. The results show that the improved CNN architecture and algorithms have significant effects in solving the problems of overfitting and computational resource consumption, and can significantly improve the accuracy and efficiency of image classification. And degraded images do adversely affect the accuracy of CNN for image classification.

Keywords: Convolutional neural networks, improved VGG16 architecture, MIDNet18 model, CNN-CapsNet, GAN.

1. Introduction

CNNs have their origins in the 1960's. It has gone through a number of important stages from theory to practice. In 1962, David Hubel and Torsten Wiesel put forward the idea of Receptive fields based on the idea of a cat's brain, and this discovery laid the foundation for subsequent neural network research. In 1980, Kunihiko Fukushima, a Japanese researcher, put forward a framework for deep learning based on convolution and pooling layers. These early works laid the foundation for the development of modern CNNs. In 1998, Yann LeCun and his colleagues developed a new method called LeNet-5, which applied the BP algorithm in order to train neural network structures, creating the foundation for modern CNNs. In 2012, AlexNet performed well in the ImageNet Image Recognition Contest, where it proposed a new deep structure and Dropout method to lower the error rate from than 25% to 15 %.

VGG proved that deeper network layers are an effective means to improve accuracy, but deeper networks tend to lead to gradient dispersion problems. ResNet alleviates the gradient vanishing problem by adding shortcut connecting edges.

Nowadays, CNNs have achieved great success in many applications such as image classification, and have shown high accuracy, even exceeding the accuracy of humans in some specific image classification problems. However, general Convolutional Neural Networks still have some shortcomings in image classification problems, such as the efficiency problem and the problem of further improving the prediction accuracy. The efficiency problem is partly attributed to a large number of weight parameters stored in CNN network, which can be effectively improved by optimising these parameters and making the CNN lighter; on the other hand, the insufficient amount of data is one of the main reasons for the poor prediction results of the CNN, and the prediction accuracy can also be improved by using different algorithms. In this paper, this paper will present several studies on CNNs for image classification applications, focusing on the overfitting problem in these articles by improving the network architecture and utilising migration learning, residual learning and spectral features, and generating datasets through GAN.

2. Application of CNN in Image Classification

The authors of the first paper designed a convolutional neural network to classify images of interstitial lung disease using shallow convolutional layers [1]. They also designed a unique network architecture as shown in Figure 1 that allows for the final classification based on neural auto-learning of features of the sample image. Because the samples of medical images are extremely limited and prone to overfitting problems, they suppressed the overfitting problem by cleverly designing the network architecture and using sensible algorithms. In terms of network architecture, they designed only one convolutional layer because the input images have no apparent structure, eliminating the need for multiple convolutional layers. Only one convolutional layer can help them to minimize the number of parameters that needs to be learnt, which can reduce the overfitting problem to some extent. In terms of optimisation algorithms, the authors used the drop-out algorithm to randomly select neurons to be disabled during training, effectively suppressing the overfitting problem. A combination of momentum weight update and weight decay is used to update the weights during backpropagation, which is used to speed up the machine learning. The authors also compared the effects of different activation functions on the CNN process and after testing concluded that the RELU activation function performed best in their framework.

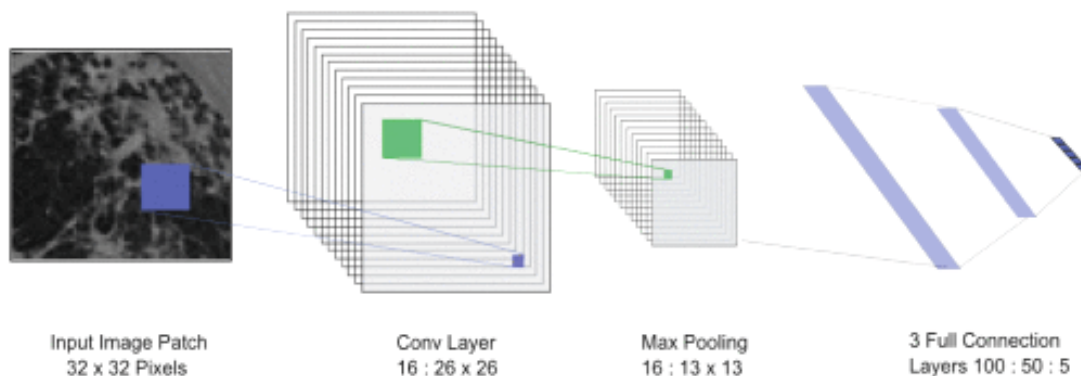


Figure 1. Network architecture [1].

3. Application of Improved CNN in Image Classification

However, early CNNs, such as AlexNet and VGGNet, had very good performance but they had strict rules about the size of the input image and required much computation to train and evaluate the network. Therefore, many researchers have improved this shortcoming of CNNs to achieve better algorithms.

The authors of the second paper have improved the VGG16 CNN architecture by designing a new MIDNet18 CNN architecture with 7 pooling, 4 dense layers and 1 classified layer [2]. In terms of results,

the authors have definitely succeeded, as the new CNN architecture he designed ended up with 98.7% accuracy, while the VGG16 architecture ended up with only 50% accuracy.

The author of this paper has also published another paper on this CNN-based brain tumour image classification problem, in which he compared the MIDNet18 CNN structure with the AlexNet CNN structure, and the final results showed that their, MIDNet18 have improved test precision in AUC, F1 score, precision, and recall [3].

Additional authors have other methods for optimising CNNs.

The aim of the forth paper is to complete the classification of cervical cancer images at different stages using a special CNN-CapsNet model as shown in Figure 2 [4].

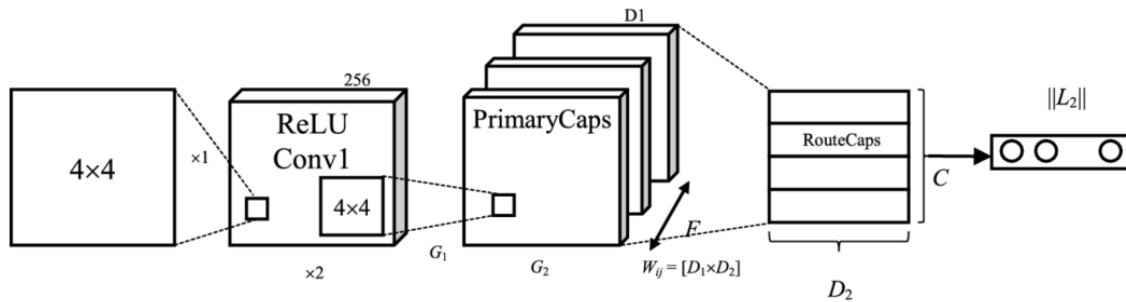


Figure 2. CNN-CapsNet model [4].

The authors apply the preprocessing technology to remove the noise and enhance the contrast when they import images, and then use the Capsnet model to deal with the images. In capsNet, every level receives inputs from all of the preceding levels, which is helpful in classifying the characteristics. This sequential input processing enhances feature recognition. Then, they estimate the hyperparameters and control the backpropagation gradient. In order to increase the performance of the capsNet, the remaining blocks are contained in the dense layers. At the end of the article, the authors compare the custom-built model with the conventional VGG-16 CNN architecture and the custom-built model classifies eight stages of cervical carcinoma with a precision of 90.28%, which is better than the conventional VGG-16 CNN architecture.

The authors of the fifth paper, to solve the problem of excessive consumption of computational resources by CNN in image classification problems, present an extended CNN model which is based on the dilated convolutional kernels, which is tested on the Mnist Hand-Written Numeral Identification Data Set [5]. Secondly, to address the detail loss problem in the dilated CNN model, the hybrid dilated CNN (HDC) is constructed by stacking dilated convolutional kernels with different dilated rates, and the HDC model is tested on the earth topography broadband remote sensing image dataset.

According to this paper, the new CNN model can reduce the training time by 12.99% on average, than the conventional CNN model. Additionally, the training precision is increased by 2.86% on average.

With respect to the precision of training, as shown in this paper, HDC is an average improvement of 14.15 percent compared to the extended CNN. In the case of 100, 200 and 350, the HDC model improved significantly, but there was no change at 500 times.

As far as the testing precision is concerned, the HDC model has an average increase of 15.35% compared with the expanded CNN model, and the HDC has an increase of 32.98%, 19.95%, 8.48%. This is due to the fact that the HDC and the expanded CNN model have a tendency to be overloaded, that is, a simple addition of a few training cycles will not improve the model significantly.

In summary, the authors of this paper concluded that the extended CNN model outperforms the conventional CNN model. Additionally, HDC model outperforms both extended CNN model and CNN model on the RS dataset. The results show that expanding CNN and HDC can greatly increase the performance of image categorization.

Because of the complexity of CNNs, there is a need for a lot of tagged samples, which has constrained the use of several outstanding CNNs (for example, AlexNet, VGG, ResNet) in data constrained cases.

In order to solve this issue, authors of the sixth paper present a two stage method that integrates CNN migration learning and network data enhancement [6]. This approach can be used to transfer useful features from a pretrained neural network to a new object, and add a new one to the source data set. Not only does this not require a lot of training data, but it also expands the training set efficiently. These two methodological features significantly reduce overfitting of CNNs on small datasets. Furthermore, Bayesian optimisation is used to tackle hyperparameter tuning problem in network fine-tuning as well in this task. Experimental results demonstrate that the proposed approach can significantly improve the performance of CNNs (especially ResNet) in the context of six public small datasets. The results show that the proposed approach is a useful approach to solving the problem of CNNs in the context of a small data set.

For small datasets, a commonly used solution is parametric migration learning. The proposed method is divided into two phases: the first phase builds powerful classifiers by migration learning, and secondly, a new model is developed based on the Web image and the classifier. This results in a reduction of overfitting and an improvement in the precision of classification. Using Bayesian optimisation simplifies the hyperparameter tuning process. Although complex and time-consuming to perform, the potential of this approach will be further enhanced with the growth of image resources on the Internet.

Conventional shallow approaches to image classification are superior to those based on deep neural networks. Deep CNNs trained from scratch are less accurate due to so many deep CNN parameters and small training dataset. The combination of pretraining and fine tuning is an efficient approach to translation learning. The tuning of the neural net allows for a fairly high level of precision. Network data augmentation leads to better performance of all CNNs with more than 2% improvement in accuracy compared to traditional augmentation techniques which only improve by 1% or 0.5%.

Deeper networks are much more efficient with equal weights. There are similar weights for AlexNet and ResNet-152, but the precision of ResNet-152 is better for all data sets. In this paper, the authors propose a two-stage classification method that utilises migration learning and network data augmentation techniques. This method reduces the need for large amounts of training data, builds powerful classifiers, increases the diversity of training dataset, enhances generalisation ability and mitigates overfitting. Although enhancement process is time-consuming, the method has significant advantages in dealing with deep CNN applications on small datasets.

Similarly, to address the problem of training large-scale networks for hyperspectral image classification on a limited amount of data, the authors of the seventh paper propose another improved algorithm for CNNs [7]. They designed a deep convolutional neural network that is deeper and wider than existing networks for hyperspectral image classification, called a context-depth CNN, which optimally exploits local contextual interactions by jointly exploiting local spatial-spectral relationships of neighbouring individual pixel vectors. They utilise the concept of "residual learning" to improve training efficiency on relatively sparse samples. The results show that the residual structure of the network can significantly increase the depth and width of the network, thus enhancing the learning capability and ultimately improving the generative performance.

Their novel Fully Convolutional Network (FCN) shown in Figure 3 has multiple convolutional layers for HSI classification. Firstly, there is a 'multi-scale filter pool', and then there are two convolution layers associated with residual learning. The final three convolution layers are similarly to AlexNet's fully connected layers for classification, using local features for classification.

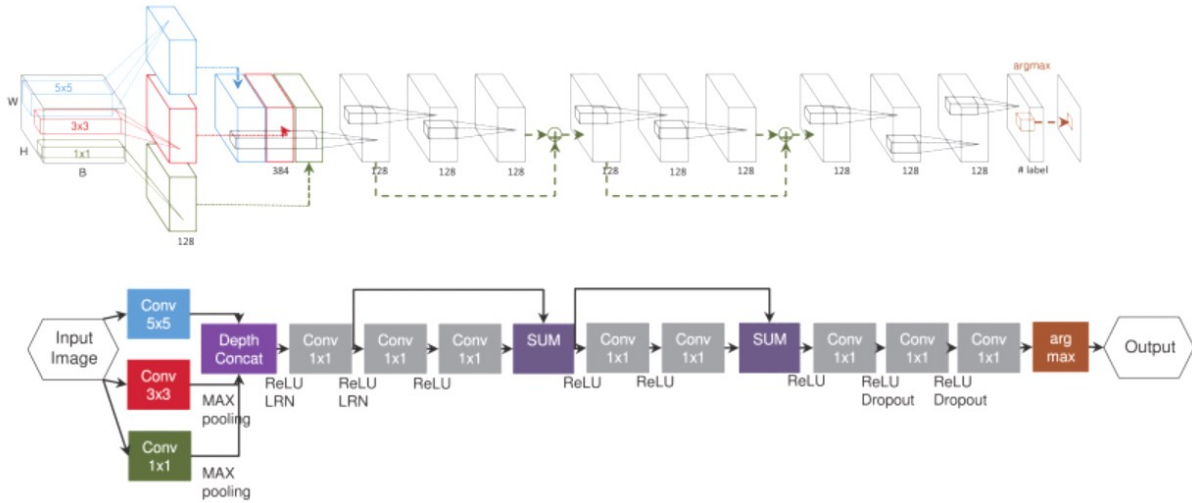


Figure 3. Fully Convolutional Network (FCN) [7].

In order to achieve optimal performance for HSI classification, they wanted to utilise both spectral and spatial features, and they designed a multiscale convolutional filter bank to scan local regions of the sample image to generate original spatial-spectral characteristics maps. The multi-scale filter pool is composed of three layers of convolution filters of varying dimensions: 2 (3×3 and 5×5) are utilized to extract the local space dependence, whereas (1×1) filter is employed for the resolution of spectral correlation. Then, the original space and spectral characteristic maps are merged into a common space and spectrum characteristic map, which includes abundant spatial spectral characteristics. Then, the combined feature maps are fed into the next level, and then the tags of the respective hyperspectral pixel vectors are obtained.

Medical training sets always suffer from overfitting problems due to insufficient number of samples when using CNN architectures. The authors of the eighth paper aimed to develop a Generative Adversarial Network (GAN)-based image enhancement technique to augment the dataset in order to optimise the performance of a custom lightweight convolutional network for Chest X-ray (CXR) classification [8].

To achieve this goal, the researchers used Progressive Generative Adversarial Networks (PGGANs) to create synthetic and enhance data to complement existing datasets. Additionally, they propose two novel convolutional neural network (CNN) architectures for multi-class classification of Covid-19, healthy, and pneumonia-affected chest X-rays. The superiority of these new architectures is evaluated by comparison with the most advanced models and migration learning methods. All models were trained with the enhancement of the X rays and compared based on classification metrics.

The original structure of CNN was founded on ResNet. The second structure of the research is based on DenseNet. The proposed models perform well on the classification metrics, reaching 98.78% and 99.2% of the tested accuracies, respectively, while the training parameters are reduced by 40% compared to the most advanced models.

The significance of this study lies in the proposal of a GAN-based progressive data enhancement method that complements the training dataset by generating synthetic data, As well as two new CNN-based Covid-19 screening strategies, both of which have classification accuracies of more than 98%. In the study, a progressive GAN framework was utilised to generate high resolution images, which resulted in effective generation of high quality 1024x1024 pixel images. Although other research has explored the need for ROC scores, this research indicates that there is no need for segmentation in this situation since the non-segmented images also offer sufficient classification capability because of the utilization of high-resolution composite images.

4. Impact of Degraded Images on CNN Effectiveness

While the previous study focused on enhancing datasets with synthetic high-resolution images, the ninth paper addresses the issue of image degradation in practical applications [9]. The authors have selected nine different degraded image types: haze images, motion blurred images, fish-eye images, underwater images, low-resolution images, salt and pepper images, images with white Gaussian noise, Gaussian blurred images, and out-of-focus images. Figure 4 shows some examples of degraded images. In order to measure the quality of classification at various stages of deterioration, the authors combined a lot of data with their own physics models and gathered actual blurry pictures from the Internet.



Figure 4. Some examples of degraded images [9].

The authors implemented CNN models using VGGNet, AlexNet, and ResNet on the Caffe platform and used these models to classify images with different levels of degradation. In addition, they investigated whether image degradation removal helps CNN-based image classification.

The results show that if a clear image is used to train a CNN classifier and then a blurred image is classified, the accuracy drops from 81.0% (trained and tested on a clear image) or 67.8% (trained and tested on a blurred image) to 33.0%. Overall, VGGNet-16 achieves better classification precision compared to that of AlexNet.

When blurred images are included in training dataset and then clear images are classified, classification accuracies for clear images are 79.7%, 77.9%, 78.7%, 80.2%, 76.2%, 76.9%, 74.8%, 78.8%, and 77.0%, respectively, which are lower than the 81% when training with only clear images. This suggests that including degraded images in the training affects the classification result of clear images.

The authors also attempted to perform image degradation removal on some of the degraded images, but the final results were still unsatisfactory. These experiments show that haze significantly affects CNN-based image classification accuracy, even when test images are preprocessed using a defogging algorithm.

5. Conclusion

In conclusion, this paper reviews the application and improvement methods of convolutional neural networks in image classification. By designing new network architectures, employing optimisation algorithms, introducing migration learning and data enhancement techniques, and generating datasets using Generative Adversarial Networks (GANs), researchers have made significant progress in solving problems such as overfitting and computational resource consumption. These approaches have not only improved the classification accuracy and efficiency of CNNs, but also expanded their application prospects on small datasets and degraded images. Future research can further explore the application of these improved methods in other image classification tasks and continuously optimise the performance of CNNs to meet more challenges in real-world applications.

References

- [1] Li Q, Cai W, Wang X, Zhou Y, Feng D D, Chen M 2014 In: 13th International Conference on Control Automation Robotics & Vision (ICARCV) Singapore 844-848
- [2] Mohan R, Ganapathy K, A R 2022 Brain tumour classification of magnetic resonance images using a novel CNN-based medical image analysis and detection network in comparison to VGG16 J. Popul. Ther. Clin. Pharmacol. 28 113-125
- [3] Ramya M, Kirupa G, Rama A 2022 Brain tumor classification of magnetic resonance images using a novel CNN-based medical image analysis and detection network in comparison with AlexNet J. Popul. Ther. Clin. Pharmacol. 29 97-108
- [4] Cibi A, Rose R J 2023 Classification of stages in cervical cancer MRI by customized CNN and transfer learning Cogn. Neurodyn. 17 1261-1269
- [5] Lei X, Pan H, Huang X 2019 A Dilated CNN Model for Image Classification IEEE Access 7 124087-124095
- [6] Han D, Liu Q, Fan W 2018 A new image classification method using CNN transfer learning and web data augmentation Expert Syst. Appl. 95 43-56
- [7] Lee H, Kwon H 2017 Going Deeper With Contextual CNN for Hyperspectral Image Classification IEEE Trans. Image Process. 26 4843-4855
- [8] Gulakala R, Markert B, Stoffel M 2023 Rapid diagnosis of Covid-19 infections by a progressively growing GAN and CNN optimisation Comput. Methods Programs Biomed. 229 107262
- [9] Pei Y, Huang Y, Zou Q, Zhang X, Wang S 2021 Effects of Image Degradation and Degradation Removal to CNN-Based Image Classification IEEE Trans. Pattern Anal. Mach. Intell. 43 1239-1253