Shallow Convolutional Neural Network for Image Recognition

Fangyuan Lei*, Xun Liu, Jianjian Jiang, Qingyun Dai, Hongyu Liu, Mengying Hu
School of Electronic and Information, Guangdong Polytechnic Normal University, Guangzhou, China.

* Corresponding author. Email: Leify@126.com
Manuscript submitted July 20, 2019; accepted September 8, 2019.
doi: 10.17706/ijcee.2019.11.4.192-197

Abstract: Deep convolutional neural networks (DCNNs) have achieved state-of-the-art results for image recognition. However, these DCNNs with complex structure consist of many layers like convolutional layers, which require high time and computational complexity for training. Therefore, we propose a novel shallow convolutional neural network (SCNND) with dropout to address the problems of the DCNNs for image recognition. The SCNND with 4 layers can fast learning the features of the images, using dropout technology between two convolutional layers to improve recognition performance. Compared to the SCNNs, our SCNND includes 4 layers, with low time complexity and parameters. Experimental results show that our SCNND outperforms shallow CNN methods on Fashion-MNIST dataset.

Key words: Deep convolutional neural networks, shallow convolutional neural network, dropout, image recognition.

1. Introduction

Lenet-5 [1] is the first convolutional neural network (CNN) which achieves excellent performance for handwritten character recognition. When the problem of gradient disappearance is overcome by Hinton [2] in the training of deep convolutional neural network (DCNN), the DCNN has been widely applied to various machine learning tasks like image recognition [3]. With abundant available data and computing resources, the DCNN can achieve higher accuracy via more layers. Starting from AlexNet [3], VGG19 [4] to ResNet34 [5] and DenseNet121 [6], the DCNN contains more layers and the structure of the DCNN is more complicated. Moreover, the training of these DCNNs leads to the issues of massive time resources and memory space.

In order to resolve the issues of the DCNNs, a set of shallow convolutional neural networks (SCNNs) with fewer layers and calculations have been proposed. Li *et al*. [7] applies a shallow CNN (SCNN) to face detection. Agarap [8] proposes an architecture combining shallow CNN of 4 layers and support vector machine (SVM). Compared to single SCNN, the architecture uses the combinations of CNN and SVM to achieve higher classification result. Lee et al. [9] introduces logarithmic group convolution module with 1×1 and 3×3 convolutions to reduce the size of model in SCNN. In [10]-[12], the authors apply batch normalization (BN) technology [13] to SCNN to speed up training and improve accuracy. However, in these SCNNs of [10]-[12], there is feature redundancy between the convolutional layers.

Motivated by these SCNNs of [10]-[12], we propose a new shallow CNN with dropout (SCNND) to address the issue of feature redundancy for image recognition. The SCNND utilizes the technologies of BN and dropout (each max-pooling layer and after fully connected layer) to prevent over-fitting and improve the generalization ability of the model. Our main contributions are the following.

1) We propose a new shallow convolutional neural network (SCNND) with 4 layers for image recognition.

2) The SCNND can reduce the feature redundancy between the convolutional layers via dropout technology to get more recognition performance.

3) Our model with 3.8M time complexity can achieve state-of-the-art results on image datasets.

2. Method

2.1. Dropout to Improve Accuracy

In the training of deep learning, if the training data is not enough whereas the complexity of the model is high, the model will be over-fitting. Dropout is an optimization technology proposed by Hinton [14] in 2012 to solve the over-fitting problem. At each training time, we use the dropout technique to temporarily inactivate the hidden neurons with a certain probability p in the forward propagation process, reducing the dependence on the hidden neurons. Therefore, improving the generalization ability of the network model. In [3], the author uses dropout technology between fully connected layers to prevent over-fitting problem and get more classification perform. To further prevent the over-fitting problem, we utilize the dropout technique (between convolutional layers and after the first fully connected layers) to achieve higher accuracy.

2.2. SCNND Model

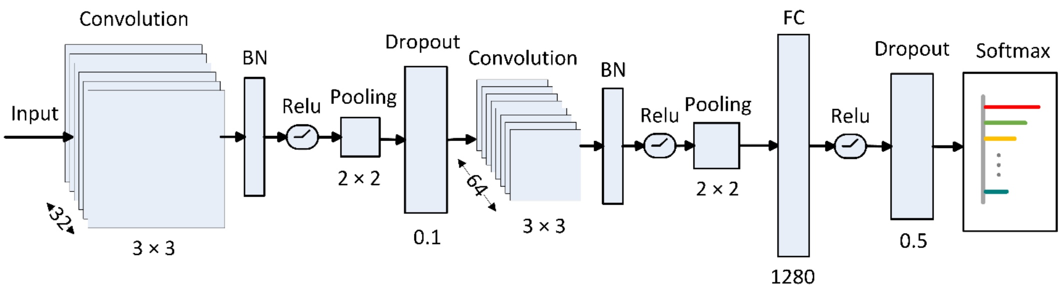

Fig. 1. SCNND model.

We use a pair of technologies BN [13] and dropout to construct a shallow convolutional neural network (SCNND) model. Fig. 1 shows our SCNND model. The SCNND model has 4 layers (regardless of BN and Pooling layers), including 2-layers convolutional layers, a fully-connected layer, and a softmax layer. The model first utilizes 3 × 3 convolution with 32 filters to extract the features of the images. After the extracted features, BN is used to normalize the extracted features and accelerate training, and Relu improves the nonlinearity of the SCNND. Then 2 × 2 max-pooling is used to reduce the dimension of the features. The max-pooling layer followed by dropout to prevent features redundancy and improve the learning ability of the model. Next the SCNND extracts deeper features via 3 × 3 convolution with 32 filters. After the second convolutional layer, we use BN and Relu to improve accuracy and nonlinearity. Next the model uses 2 × 2 max-pooling to further reduce the dimension of the features. Then 2 × 2 max-pooling is introduced to reduce the dimension. The second max-pooling layer followed by a fully connected layer with 1280 neurons to combine the extracted features. Next dropout is introduced to further reduce over-fitting and increase the learning ability of the model. Finally, the model achieves multi-classification by softmax output layer.

2.3. Time Complexity Assessment

The time complexity of CNN is related to pooling layers, fully connected layers, and convolutional layers.

In [15], the author points out that the pooling and fully connected layers only consume 5-10% of the training time. Namely, the training time of the model is mainly determined by the convolutional layers. We calculate the time complexity according to [11], as follows.

$$O = \left(\sum_{k=1}^{n} N_{k-1} \cdot S_w \cdot S_h \cdot N_k \cdot M_w \cdot M_h \right) \tag{1}$$

Here, k is the index of convolutional layer with n layers, and N_k denotes the number of filter in the i-th layer. S_w and S_h are the width and height of the filters, and M_w and M_h are the output feature map respectively.

3. Experiment

3.1. Datasets

MNIST [1] and Fashion-MNIST [16] are two benchmark datasets for image recognition. The MNIST dataset consists of 10 handwritten digital grayscale images from 0 to 9, and the Fashion-MNIST dataset contains grayscale images of 70000 different fashion items in 10 categories. The two datasets include 60000 training images and 10000 test images. The size of all images is 28×28 pixels. In our experiments, we randomly flip the 70000 fashion images with 0.5 probability on Fashion-MNIST.

3.2. Experimental Parameters

In our experiments, we use the following sets on two image recognition datasets: 0.000005 (L2 regularization), 300(epochs) and 0.02(learning rate). We update the learning rate by stochastic gradient descent (SGD) with the momentum of 0.9.

3.3. Results

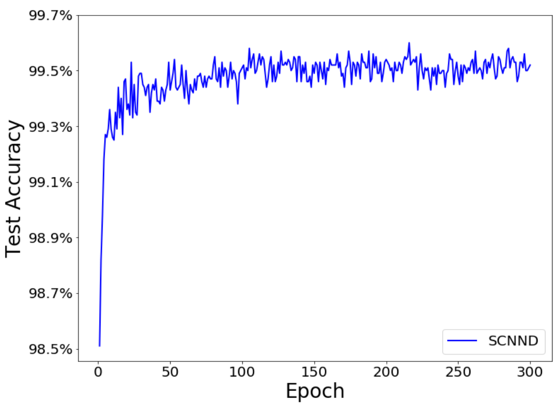

Fig. 2. The test accuracy of SCNND on MNIST.

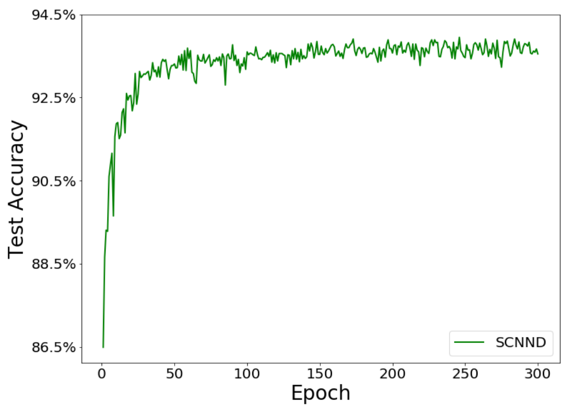

Fig. 3. The test accuracy of SCNND on Fashion-MNIST.

Fig. 2 and Fig. 3 show the test accuracy of our SCNND on MNIST and Fashion-MNIST datasets. The results of Fig. 2 and Fig. 3 prove that the SCNND model can effectively learn the characteristics of the data.

Table 1 shows the comparison results of SCNND and deep CNN methods [3]-[5], [17]-[19]. In Table 1, we can clearly see that the SCNND outperforms the methods of [3], [5], achieving high accuracy of 99.60% on MNIST dataset. Compared with the state-of-the-art method of [18] on MNIST, our SCNND includes two 3×3 convolutions with 32 and 64 filters, whereas the model of [18] uses the larger size of convolution (such as 7×7) and more filters (such as 419) to extract features.

On Fashion-MNIST dataset, our SCNND achieves 94.19% accuracy, which is lower than [5], [18], [19]. In

[18], the network includes 7 convolutional layers with more filters. The convolutional layers of [5], [19] are 6 times more than the SCNND. Compared with other deep CNN methods, the SCNND with less layers reaches better result.

Table 1. Comparison Test Accuracy of SCNND and Deep CNN Methods

Deep CNN Methods	MNSIT Test Accuracy(%)	Fashion-MNIST Test Accuracy(%)	Convolution/fully-connected Number
AlexNet [3]	98.81%	84.43%	5/3
VGG16 [4]	-	92.89%	13/3
VGG19 [4]	-	92.90%	16/3
VGG16 H-CNN [17]	-	93.52%	≥13/≥3
VGG19 H-CNN [17]	-	93.33%	≥16/≥3
ResNet [5]	99.37%	94.39%	≥17/1
Ma et al. [18]	99.72%	94.60%	3/4 for MNIST 7/4 for other
Zeng et al. [19]	-	97.66%	13/3
SCNND (ours)	99.60%	94.19%	2/2

The comparison results of our SCNND and shallow CNN methods [8], [10], [20]-[23] are shown in Table 2. On MNIST dataset, our SCNND achieves high accuracy of 99.60%, outperforming these results of [8], [21], [23]. Compared to [22], our SCNND achieves competitive test accuracy with lower time complexity. Compared to all shallow CNN methods, the SCNND achieves the state-of-the-art accuracy with 3.8M time complexity.

The comparison results of Table 1 and Table 2 demonstrate the effectiveness of the proposed method.

Table 2. Comparison Test Accuracy of SCNND and Shallow CNN Methods

Shallow CNN Methods	MNSIT Test Accuracy(%)	Fashion-MNIST Test Accuracy(%)	Time Complexity
Bhatnagar et al. [10]	-	92.54%	2M
Hossain et al. [20]	-	93.41%	54.1M
Agarap et al. [8]	99.17%	90.91%	-
Poernomo et al. [21]	99.59%	92.03%	-
Jain et al. [22]	99.66%	-	4.5M
Gorokhovatskyi et al. [23]	98.52%	-	1.2M
SCNND (ours)	99.60%	94.19%	3.8M

4. Conclusions

In this paper, we propose a new dropout-based shallow convolutional network (SCNND) framework for image recognition. The SCNND includes 4 layers, utilizing dropout technology to improve the learning ability of the SCNND. The classification results and comparisons with deep CNN and shallow CNN demonstrate the validity of the model on two benchmark image datasets.

Acknowledgment

This work was partly supported by the National Natural Science Foundation of China (U170120078), the Guangdong Provincial Key Laboratory Project (2018B030322016), the Scientific and Technological Projects of Qingyun (the grant number 170809111721249, 170802171710591), and the Scientific and Technological Projects of Guangdong Province (2019A070701013).

References

[1] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86(11)*, 2278-2324.

[2] Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313(5786)*, 504-507.

[3] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 1097-1105.

[4] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. Retrieved from https://arxiv.org/abs/1409.1556

[5] He, K., Zhang, X., Ren, S., & Sun, J. (2016, June). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770-778).

[6] Huang, G., Liu, Z., Maaten, L., & Kilian, Q. W. (2017, July). Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4700-4708).

[7] Li, H., Lin, Z., Shen, X., Brandt, J., & Hua, G. (2015). A convolutional neural network cascade for face detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 5325-5334).

[8] Agarap, A. F. (2017). An architecture combining convolutional neural network (CNN) and support vector machine (SVM) for image classification. Retrieved from https://arxiv.org/abs/1712.03541

[9] Lee, T. K., Baddar, W. J., Kim, S. T., & Ro, Y. M. (2018, February). Convolution with logarithmic filter groups for efficient shallow CNN. *Proceedings of the International Conference on Multimedia Modeling* (pp. 117-129). Springer, Cham.

[10] Bhatnagar, S., Ghosal, D., & Kolekar, M. H. (2017, December). Classification of fashion article images using convolutional neural networks. *Proceedings of the 2017 Fourth International Conference on Image Information Processing (ICIIP)* (pp. 1-6).

[11] Lu, L., Yang, Y., Jiang, Y., Ai, H., & Tu, W. (2018). Shallow Convolutional neural networks for acoustic scene classification. *Wuhan University Journal of Natural Sciences*, *23(2)*, 178-184.

[12] Ide, H., & Kurita, T. (2018, December). Convolutional neural network with discriminant criterion for input of each neuron in output layer. *Proceedings of the International Conference on Neural Information Processing* (pp. 332-339). Springer, Cham.

[13] Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. Retrieved from https://arxiv.org/abs/1502.03167

[14] Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. Retrieved from https://arxiv.org/abs/1207.0580

[15] He, K., & Sun, J. (2015). Convolutional neural networks at constrained time cost. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 5353-5360).

[16] Xiao, H., Rasul, K., & Vollgraf, R. (2017). Fashion-mnist: A novel image dataset for benchmarking machine learning algorithms. Retrieved from https://arxiv.org/abs/1708.07747

[17] Seo, Y., & Shin, K. S. (2019). Hierarchical convolutional neural networks for fashion image classification. *Expert Systems with Applications, 116,* 328-339.

[18] Ma, B., & Xia, Y. (2018). Autonomous deep learning: A genetic DCNN designer for image classification. Retrieved from https://arxiv.org/abs/1807.00284

[19] Zeng, S., Zhang, B., Zhang, Y., & Gou, J. (2018). Collaboratively weighting deep and classic representation via L2 regularization for image classification. Retrieved from https://arxiv.org/abs/1802.07589

[20] Hossain, M. M., Talbert, D. A., Ghafoor, S. K., & Kannan, R. (2018). FAWCA: A flexible-greedy approach to find well-tuned CNN architecture for image recognition problem. *Proceedings of the Int'l Conf. Data Science* (pp. 214-219).

[21] Poernomo, A., & Kang, D. K. (2018). Biased dropout and crossmap dropout: Learning towards effective dropout regularization in convolutional neural network. *Neural Networks, 104*, 60-67.

[22] Jain, S., & Chauhan, R. (2018, April). Recognition of handwritten digits using DNN, CNN, and RNN. *Proceedings of the International Conference on Advances in Computing and Data Sciences* (pp. 239-248). Springer, Singapore.

[23] Gorokhovatskyi, O., & Peredrii, O. (2018, August). Shallow convolutional neural networks for pattern recognition problems. *Proceedings of the 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)* (pp. 459-463).

Fangyuan Lei received the B.S degree from Northwest Polytechnic University, Xi'an, China, the M.S. degree from Lanzhou University of Technology, Lanzhou, China, and the Ph.D. degree from Northwest Polytechnic University, China in 1998, 2001 and 2004, respectively. He is currently an associate professor with the School of Electronic and Information, Guangdong Polytechnic Normal University, Guangzhou, China. Graph convolutional network, image processing and sensor network have been of particular interest for him over recent years.