

Deep Convolutional Neural Network and Its Application in Image Recognition of Road Safety Projects

Lingling Wang^{a,b}, Wenyin Gong^{a,*}, and Xiang Li^a

^a*School of Computer Science and Technology, China University of Geosciences, Wuhan, 430000, China*

^b*School of Information Engineering, Wuhan University of Engineering Science, Wuhan, 430000, China*

Abstract

Road safety projects constitute an important part of road safety facilities. Assessing the safety of these projects is important for assessing the safety of roads. In recent years, road safety problems have caused enormous losses to the country and its people. Traditional inspection and maintenance of road safety projects mainly involve manual inspection and on-site maintenance; however, manual inspection is time-consuming and laborious, and it cannot be used to identify safety issues in large areas. This paper focuses on the application of the deep convolutional neural network algorithm, a deep learning algorithm, for the recognition of road safety projects. Comparative analysis of the experimental results shows that both the convolutional neural network models VGG16 and InceptionV3 can identify the pre-processed data sets of the road safety projects; however, the accuracy of the test set model InceptionV3 is higher than that of VGG16, reaching 93.3%.

Keywords: deep learning; deep convolutional neural network; image recognition; road safety project

(Submitted on June 11, 2019; Revised on July 7, 2019; Accepted on August 12, 2019)

© 2019 Totem Publisher, Inc. All rights reserved.

1. Introduction

In recent years, deep convolutional neural networks (DCNN) have gained attention in the area of deep learning in image recognition and target detection because of its exceptional performance [1-8]. The first CNN was LeNet5, which was proposed by LeCun [9] in 1998. In 2010, Mamalet et al. [10] proposed an embedded facial image analysis framework based on concave neural networks (ConvNets). Their work proved that ConvNets can process the complete framework and its facial image processing library for mobile applications. In 2012, Krizhevsky et al. [11] proposed AlexNet based on LeNet5. AlexNet used corrected linear units as a LeNet5-based activation function and selectively ignored certain neurons using dropout and data augmentation techniques during training. This slows down the overfitting of the models and overlaps max pooling, which avoids the average pooling effect and reduces the training time. From LeNet5 in the 1990s to AlexNet, and then to GoogleNet [12], the accuracy of these network models is approaching that of human judgment.

At present, DCNNs are mainly used in road safety facilities for the identification of traffic signs, roads, and license plates. In 2017, by using the automatic feature extracting ability of the DCNN depth structure, Wang et al. [13] improved the detection rate of the existing non-scene-adaptive road segmentation algorithms by an average of 4.5%. Cao et al. [14] used a convolutional neural network to extract the convolution features for the recognition of license plates, replacing the traditional manual feature extraction process. Wu et al. [15] used the full convolutional neural network of 32 layers to effectively identify the category of each pixel in road scene images and performed smooth segmentation on the targets, providing valuable road environment information for self-driving cars. Guo et al. [16] used convolutional neural networks to classify license plates, attaining a recognition rate of 98.25%. Therefore, this paper proposes the use of deep learning methods for the image recognition and detection of road safety projects and provides a theoretical basis and practical guidance for road safety evaluation.

* Corresponding author.

E-mail address: winniay@163.com

2. Models and Features of Convolutional Neural Networks

The principle of neural networks is that the output value of the data of the input layer after passing through multiple neurons should be as close as possible to the expected value. If the difference between the output data and the actual value is too large, the weight value is repeatedly corrected until the two are within the error range. A CNN consists of an input layer, convolutional layers, pooling layers, and fully connected layers.

2.1. Convolutional Layer

The most important function of the convolutional layer is convolution, and the mathematical definition is described as follows in Equation (1):

$$f(t) = \int x(t-a)\omega(a)da \quad (1)$$

In a convolutional layer, the convolution formula is defined as follows in Equation (2):

$$s(i,j) = f\left(\sum_m \sum_n \omega_{m,n} \times x_{i+m,j+n} + b\right) \quad (2)$$

The convolutional layer is characterized by local connectivity and shared weights. In a fully connected network, the input is drawn into a vertical arrangement of neurons. A small number of input neurons are connected to the next hidden layer, which is called the local receptive field. The local receptive field is then moved in the input image with each move corresponding to one of the neurons in the hidden layer, causing all the neurons in the hidden layer to repeat. Each hidden neuron has a bias and weight that are connected to the local receptive field of the neuron and uses the same weights and offsets for all the neurons in that layer. Sharing means that all the neurons in the hidden layer detect the exact same feature and then enter different locations of the image. This shows that the convolutional network can adapt well to image translation invariance. Shared weights and offsets are called convolution kernels or filters, and they can greatly reduce the number of network parameters.

2.2. Pooling Layer

The pooling layer is also known as the downsampling layer. The pooling layer further reduces the number of parameters by removing some unimportant samples from the feature map. The role of a pooling layer is mainly to reduce the resolution of the feature image, reduce the number of computations, and enhance network robustness. For example, max pooling, one of the pooling operations, has certain robustness towards image translation. The most common pooling operations are max pooling and average pooling. Max pooling effectively takes the largest maximum of $n \times n$ samples as the sample value, whereas average pooling calculates the average value of the image region as the cluster value of the region.

2.3. Fully Connected Layer

Typically, there are several fully connected layers after the last pooling layer. These layers are responsible for transforming the two-dimensional feature map into a one-dimensional feature vector, which contains hundreds of network model parameters that help convert the convolutional neural network into a fixed-length vector. This vector can be used as a feature vector for subsequent processing. Because a fully connected layer with multiple parameters can be computationally expensive, other techniques should be used to minimize the parameters of the connected layer.

3. Production and Preprocessing of Data Sets

Because of the wide variety of road safety projects, various safety facilities shall be combined with one another. Therefore, in this paper, the images of road safety projects are classified into seven large categories according to their shape, color, and texture, as shown in Figures 1 to 7.

The initial data set is mainly collected through the Internet and photographed using other devices. The image data set is segmented after preprocessing. First, various levels of brightness transformation are applied to the images of the initial road safety projects in each category. The transformed images then undergo Gaussian blurring. Finally, the images are cropped, and noise is randomly added to obtain a relatively balanced sample. Table 1 lists the data sets in each category.



Figure 1. Slope protection with grass



Figure 2. Slope protection with dry masonry



Figure 3. Slope protection with grid stones



Figure 4. Slope protection with grassed grid stones

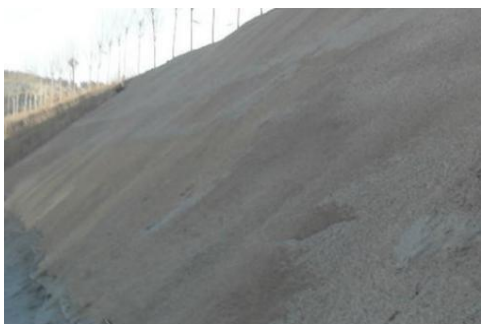


Figure 5. Slope protection with shotcrete



Figure 6. Slope protection with geotextile



Figure 7. Gabion for road

Table 1. Classified data set

Category	Initial data set	Pre-processed data set
Slope protection with grass	310	7440
Slope protection with dry masonry	289	6936
Slope protection with grid stones	257	6168
Slope protection with grassed grid stones	308	7392
Slope protection with shotcrete	122	7320
Slope protection with geotextile	104	6240
Gabion for road	224	6376
Total	1614	47872

4. Experiment and Result Analysis

The data set used in this experiment is a self-produced data set of road safety projects. The initial data set contains 1,614 images, whereas the pre-processed data set contains 47,872 images after brightness transformation, histogram equalization, Gaussian blurring, and image equalization expansion. The experiment mainly uses the VGG16 and InceptionV3 network models to train the initial data set and the pre-processed data set. By changing the network parameter learning rate, the number of iterations, and the number of samples, we evaluated the network models with respect to the accuracy, recall rate, and comprehensive evaluation index to obtain the optimal network model. The details are as follows.

4.1. Learning Rate

VGG16 and InceptionV3 are models of DCNNs. The learning rate determines the speed at which the parameters approach the optimal values.

In this experiment, the initial iteration number was set to 5,000, the sample number was proportionally divided into 10 parts, and the ratio between the initial training set and the test set was set to 7:3. To verify the robustness of the model, the test samples were divided into seven categories according to the safety project category. Each category has a certain proportion of negative samples that are added to observe the changes in the accuracy, recall rate, and comprehensive evaluation index F1 of the models at different learning rates. The experimental results are shown in Figures 8 and 9.

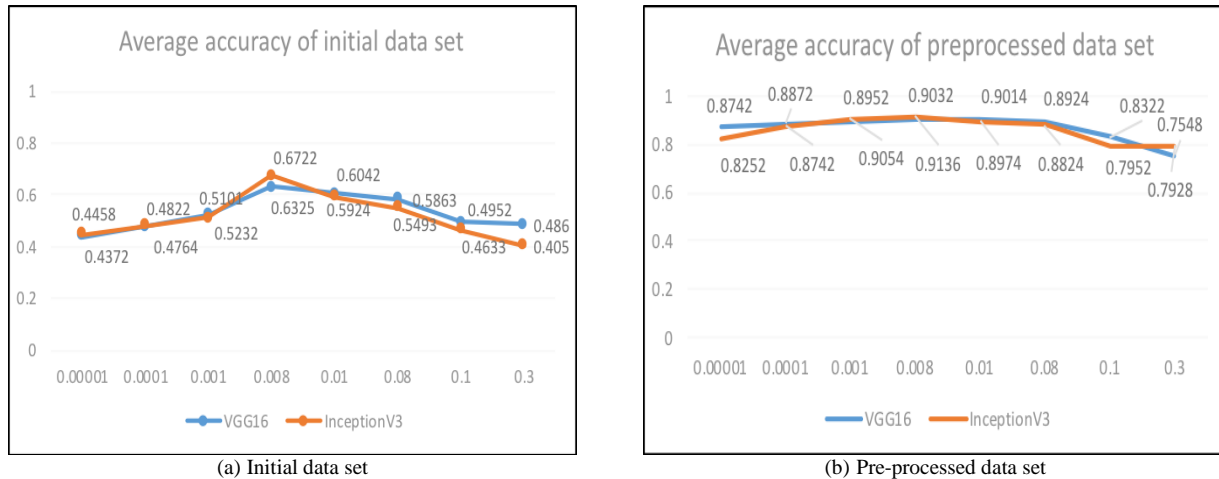


Figure 8. Variation in average accuracy of models using initial data set (a) and pre-processed data set (b) at different learning rates

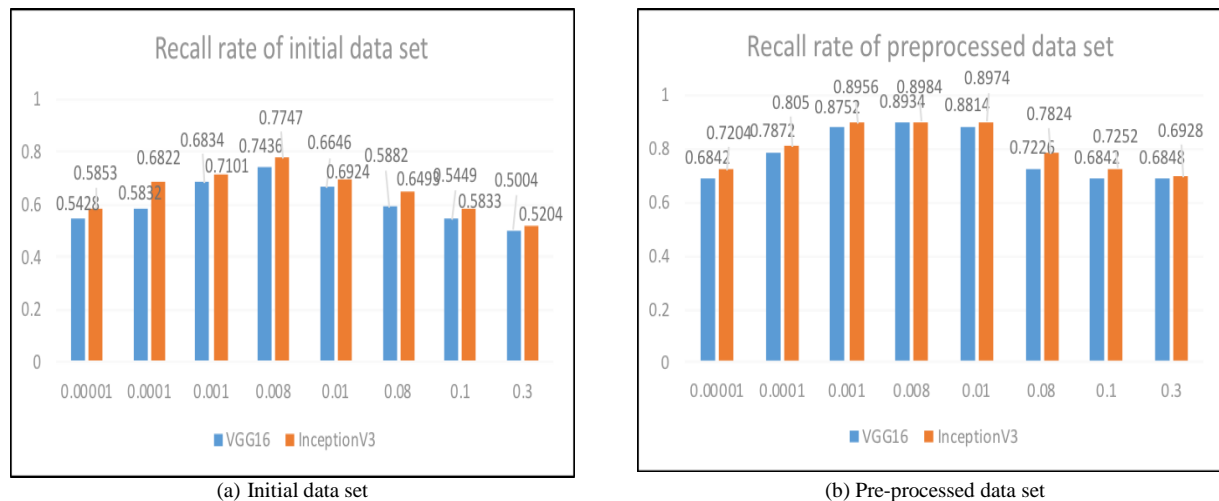


Figure 9. Variation in recall rate of models using initial data set (a) and pre-processed data set (b) at different learning rates

Figure 8 shows that both the VGG16 model and the InceptionV3 model have the highest classification accuracy when the learning rate is 0.008, and the accuracy of the former is lower than that of the later. Figure 9 shows that the VGG16

model and the InceptionV3 model have the highest recall rate of the data set when the learning rate is 0.008, and the rate of the former is lower than that of the later.

Therefore, it can be observed that both convolutional network models maintain a high accuracy and recall rate when the learning rate is 0.008, whereas both models have low accuracy and recall rate for the initial data set. This is because the data sets are relatively small and do not contain a large variety of situations; therefore, the model obtained by training is not reliable. It also proves that the convolutional neural network model requires a large number of samples for training to obtain the optimal network model. One of the criteria for evaluating the optimal network model is the comprehensive evaluation index F1. The average accuracy and recall rate corresponding to the initial data set and the pre-processed data set can be used to obtain the F1 value of VGG16 and InceptionV3. The F1 value varies with the learning rate, as shown in Figure 10.

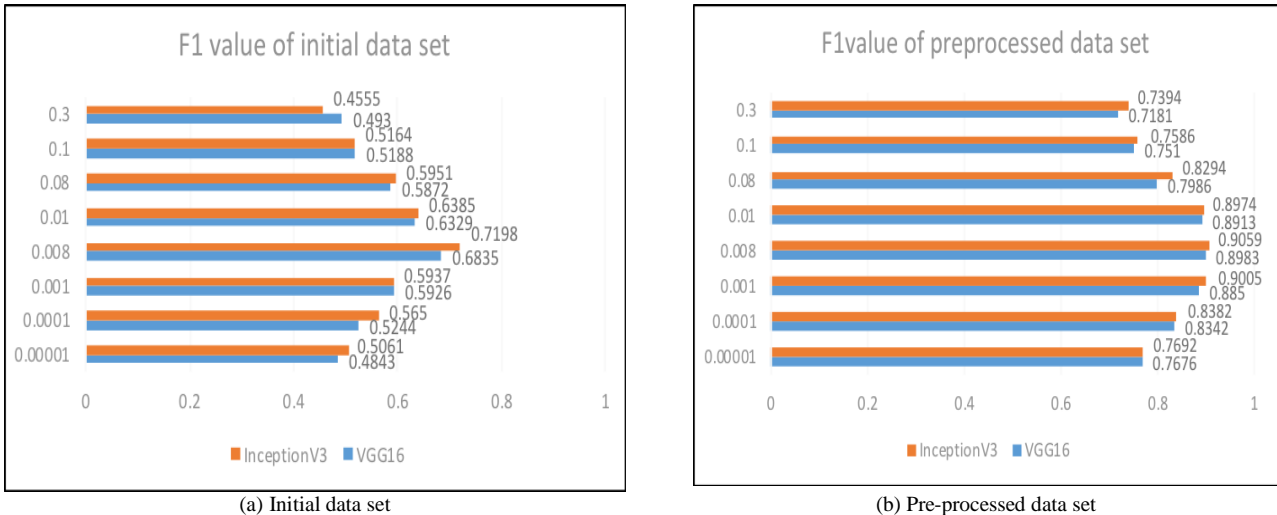


Figure 10. Variation in F1 value of models using initial data set (a) and pre-processed data set (b) at different learning rates

Figure 10(a) illustrates that the F1 value of VGG16 corresponding to the initial data set attains a maximum of 68.35% and that of InceptionV3 attains a maximum of 71.98%. Figure 10(b) shows that the highest F1 value of VGG16 corresponding to the pre-processed data set is 89.83% and that of InceptionV3 attains a maximum of 90.59%. This also proves the robustness of the network model on the data set after brightness transformation, Gaussian blur, and image equalization expansion processing.

4.2. Number of Iterations

From Section 4.1, it can be observed that the average accuracy, recall rate, and comprehensive evaluation index F1 are the highest when the learning rate is set to 0.008 where the classification effect of the two network models is the best. Therefore, the learning rate is set to 0.008 in the following experiment, and the training set and the test set still maintain a ratio of 7:3. The number of iterations is changed to observe the change in the average accuracy, recall rate, and comprehensive evaluation index F1. The experimental results are shown in Figures 11 to 14.

Figures 11 and 12 illustrate that with the initial data set, VGG16 achieves a maximum average accuracy of 65.22% and a maximum recall rate of 78.44% when the number of iterations is 12,000, whereas InceptionV3 achieves a maximum average accuracy of 65.42% and a maximum recall rate of 75.46% when the number of iterations is 18,000. Figures 13 and 14 illustrate that with the pre-processed data set, VGG16 achieves a maximum average accuracy of 92.98% and a maximum recall rate of 90.49% when the number of iterations is 12,000, whereas InceptionV3 achieves a maximum average accuracy of 93.3% and a maximum recall rate of 90.49%.

From this, we can observe that VGG16 maintains a high average accuracy and recall rate when the learning rate is 0.008 and the number of iterations is 12,000, whereas InceptionV3 has the optimal classification accuracy and recall rate when the learning rate is 0.008 and the number of iterations is 18,000.

Therefore, the comprehensive evaluation index F1 of VGG16 and InceptionV3 at a different number of iterations can be obtained using the average accuracy and recall rate corresponding to the initial data set and the pre-processed data set, as shown in Figure 15.

Figure 15(a) illustrates that with the initial data set, VGG16 has a maximum F1 of 69.97% and InceptionV3 has a maximum F1 of 71.34%. Figure 15(b) shows that with the pre-processed data set, the highest value of F1 of VGG16 is 93.13% and the highest value of F1 of InceptionV3 is 90.31%.

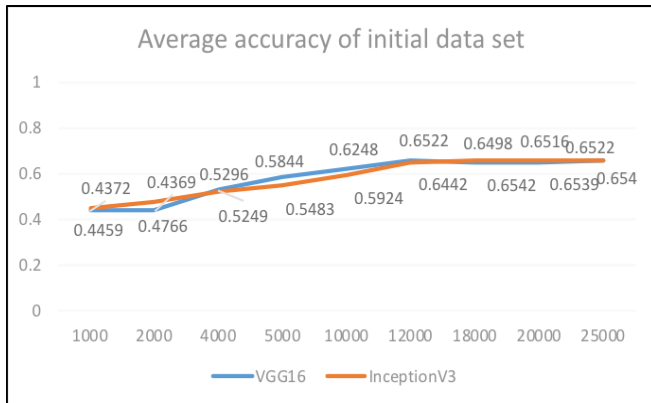


Figure 11. Variation in average accuracy of models

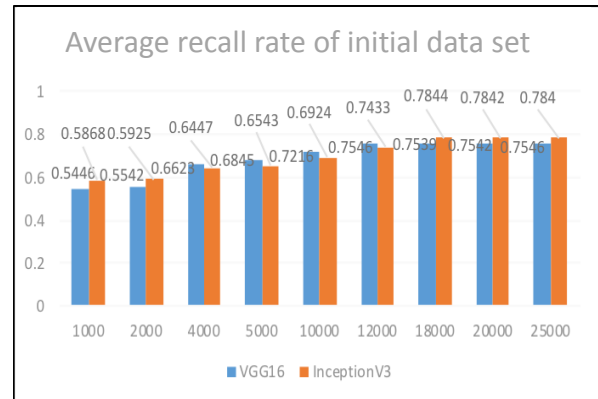


Figure 12. Variation in average recall rate of models

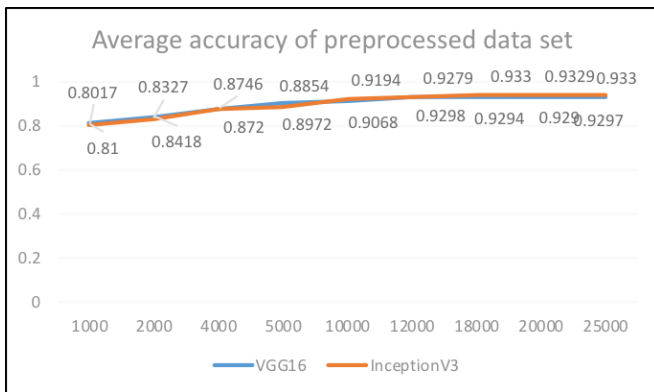


Figure 13. Variation in average accuracy of models

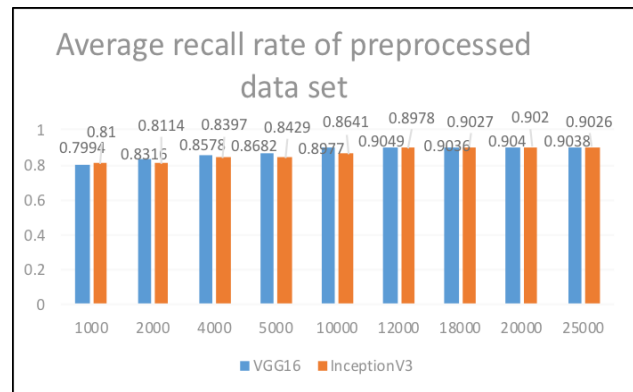
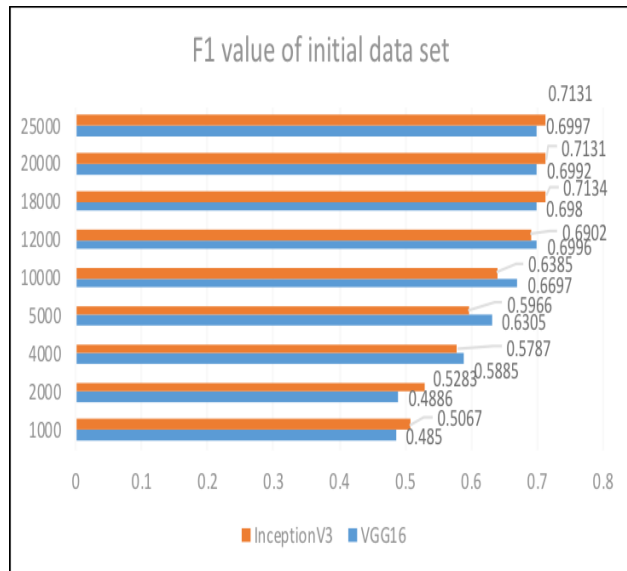
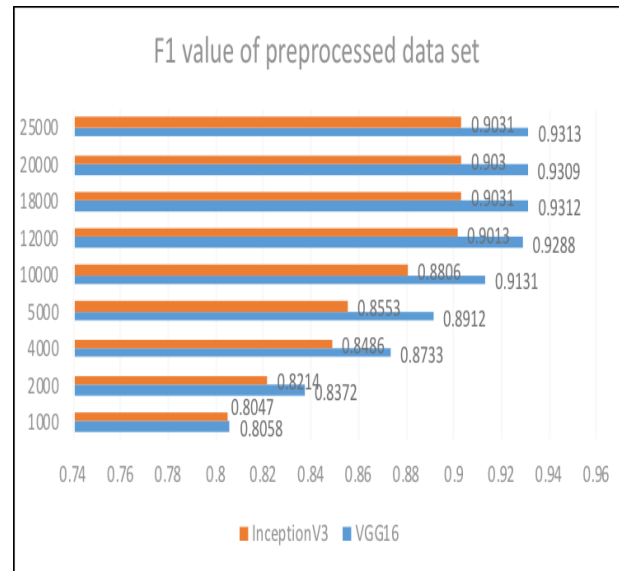


Figure 14. Variation in average recall rate of models



(a) Initial data set



(b) Pre-processed data set

Figure 15. Variation in F1 value of models using initial data set (a) and pre-processed data set (b) at different number of iterations

4.3. Sample Size

The sample size is one of the important factors affecting the convolutional neural network models. If the training sample

size is insufficient, the overfitting phenomenon will occur. If the training set is too large, the training time will be relatively long. Therefore, when the feature quantity is constant, the number of training samples should be as small as possible.

From Sections 4.1 and 4.2, it can be observed that when the learning rate is set to 0.008 and the number of iterations is set to 12,000 in this experiment, the average accuracy, recall rate, and comprehensive evaluation index F1 of VGG16 are the highest. In contrast, for InceptionV3, the average accuracy, recall rate, and comprehensive evaluation index F1 are optimal when the learning rate is set to 0.008 and the number of iterations is set to 18,000. Therefore, in the following experiment, the learning rate was set to 0.008 and the sample size was altered to observe the changes in the average accuracy, recall rate, and comprehensive evaluation index F1. Because the initial data set was small, the equal-scale control experiment was not performed; however, the pre-processed data set was equally divided into 10 groups, and a comparative experiment was conducted on the training sample and the test sample at different ratios.

The experimental results are shown in Figures 16 and 17.

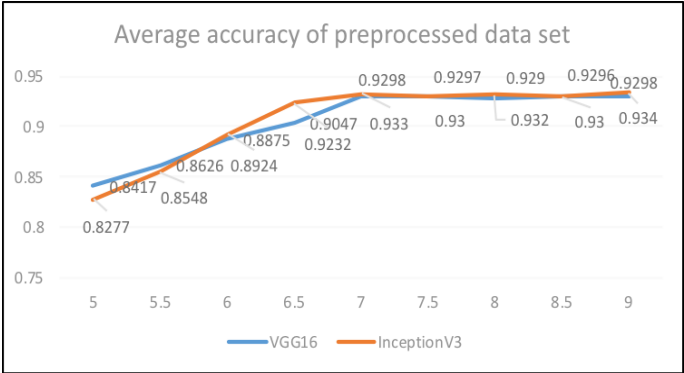


Figure 16. Variation in average accuracy of models

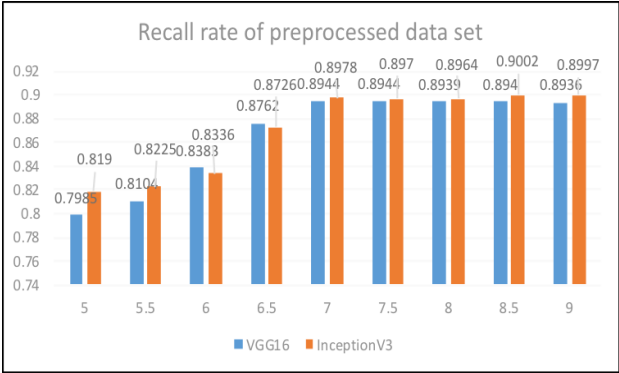


Figure 17. Variation in recall rate of models

Figures 16 and 17 illustrate the change in the average accuracy and recall rate of the models using the pre-processed data set as the sample size varies. It can be observed that both the convolutional network models VGG16 and InceptionV3 have an optimal average accuracy and recall rate when the ratio between the training sample and the test sample is 7:3. The variation in the F1 value can thereby be obtained for the pre-processed data set as the sample size varies, as shown in Figure 18.

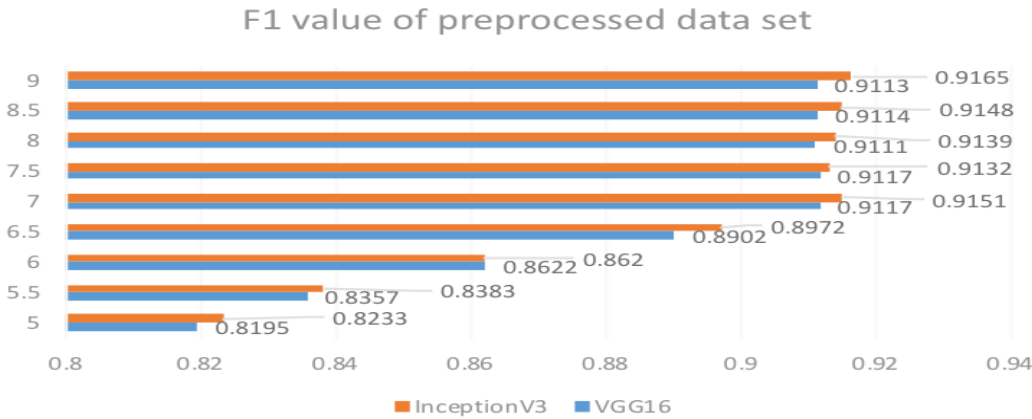


Figure 18. Variation in F1 value of models using pre-processed data set with different sample sizes

Figure 18 shows that when the ratio between the training sample and the test sample is 7:3, the F1 value of VGG16 corresponding to the pre-processed data set is 91.17% and the maximum F1 value of InceptionV3 corresponding to the pre-processed data set is 91.63%. By combining the learning rate, the number of iterations, and other evaluation criteria, we obtained two optimal network models for the image recognition of road safety projects.

5. Conclusions

In summary, the convolutional network models VGG16 and InceptionV3 can effectively identify the images of road safety projects. When using VGG16, the optimal convolutional network model for image recognition of road safety projects can be

obtained when the learning rate is set to 0.008, the number of iterations is set to 12,000, and the ratio between training samples and test samples is set to 7:3; the average accuracy can reach 92.98%. When all the other parameters are unchanged and the number of iterations is set to 18,000, the optimal InceptionV3 network model can be obtained and the average accuracy with the tested images of road safety projects can reach 93.3%. The two convolutional neural network models used in this paper achieved good results in the image recognition of road safety projects. Compared with VGG16, InceptionV3 has higher accuracy, recall rate, and comprehensive evaluation index F1, suggesting better overall robustness in image recognition of road safety projects.

References

1. Z. H. Zhao, S. P. Yang, and Z. Q. Ma, "License Plate Character Recognition based on Convolutional Neural Network LeNet-5," *Journal of System Simulation*, Vol. 22, No. 3, pp. 638-641, 2010
2. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Proceedings of International Conference on Neural Information Processing Systems*, pp. 1097-1105, Nevada, USA, December 2012
3. S. Liu, Y. Zhu, and L. Xue, "Remote Sensing Image Super-Resolution Method using Sparse Representation and Classified Texture Patches," *Wuhan Daxue Xuebao*, Vol. 40, No. 5, pp. 578-582, 2015
4. F. Hu, G. S. Xia, J. Hu, and L. P. Zhang, "Transferring Deep Convolutional Neural Networks for the Scene Classification of High-Resolution Remote Sensing Imagery," *Remote Sensing*, Vol. 7, No. 3, pp. 14680-14707, 2015
5. G. Y. Chen, X. Li, and L. L. Wang, "Identification and Classification of Adverse Geological Body based on Convolution Neural Networks," *Geological Science and Technology Information*, Vol. 35, No. 1, pp. 205-211, 2016
6. C. Dong, L. C. Chen, and K. He, "Image Super-Resolution using Deep Convolutional Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 38, No. 2, pp. 295-307, 2016
7. J. T. Zhang, A. L. Wang, N. An, and Y. J. Iwahori, "Superresolution Approach of Remote Sensing Images based on Deep Convolutional Neural Network," *International Journal of Performability Engineering*, Vol. 14, No. 3, pp. 463-472, 2018
8. G. Y. Chen, Z. H. Cai, and X. Li, "Recognition and Classification of High Resolution Remote Sensing Image based on Convolutional Neural Network," *International Journal of Performability Engineering*, Vol. 14, No. 11, pp. 2852-2863, 2018
9. Y. Lecun and Y. Bengio, "Convolutional Networks for Images, Speech, and Time Series," *The handbook of brain theory and neural networks*, MIT Press, 1998
10. F. Marmaret, S. Roux, and C. Garcia, "Embedded Facial Image Processing with Convolutional Neural Networks," in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, 2010
11. A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems*, Vol. 1, pp. 1097-1105, 2012
12. P. Tang, H. Wang, and S. Kwong, "G-MS2F: GoogLeNet based Multi-Stage Feature Fusion of Deep CNN for Scene Recognition," *Neurocomputing*, Vol. 225, pp. 188-197, 2017
13. H. Wang, Y. F. Cai, Y. Y. Jia, L. Chen, and H. B. Jiang, "Scene Adaptive Road Segmentation Algorithm based on Deep Convolutional Neural Network," *Journal of Electronics and Information Technology*, Vol. 39, No. 2, pp. 263-269, 2017
14. S. Y. Cao, Y. H. Liu, and X. Z. Li, "Vehicle Target Detection based on Fast R-CNN," *Journal of Image and Graphics*, 2017
15. Z. S. Wu and W. P. Fu, "Road Scene Comprehension based on Deep Convolutional Neural Network," *Computer Engineering and Applications*, Vol. 53, No. 22, pp. 8-15, 2017
16. K. Y. Guo, H. J. Gu, and X. L. Guo, "Application of Convolutional Neural Network in License Plate Classifier," *Computer Engineering and Applications*, 2017

Lingling Wang is a doctoral student in the School of Computer Science and Technology at the China University of Geosciences. She received her master's degree from the China University of Geosciences in 2013. Her current research interests include computer networks, security, and machine learning.

Wenyin Gong is a professor and doctoral supervisor. His main research interests include intelligent computing and its applications. He is currently the deputy secretary general of the Hubei Computer Society and the editor of the international SCI journal *Memetic Computing*. He has presided over two projects of the National Natural Science Foundation and one of the new Teachers' Funds for Doctoral Programs under the Ministry of Education.

Xiang Li is an associate professor at the China University of Geosciences and studies geoscience information engineering.