

Comparative Research on Performance of Image Recognition and Classification Using VGG16 with Different Features

Zhide Ren¹, Yuxian Wu² and Bicheng Zhang³

¹International Institute, Changsha University of Science and Technology, Changsha, 410004, China

²Faculty of Computer Science, University of Electronic Science and Technology of China,
Zhongshan Institute, Zhongshan, 528400, China

³Faculty of Mathematical Sciences, Luoyang Normal University, Luoyang, 471000, China

Keywords: Image Recognition and Classification, CNN, VGG16.


Abstract: Convolutional neural network (CNN) holds a prominent position in machine learning for image recognition and classification. To find out what affects the training efficiency and accuracy, this research primarily enumerates two types of test results based on a typical CNN model called Visual Geometry Group 16 (VGG16) using diverse data sets. At first, VGG16 model is separately introduced towards two data sets. Also, the research presents how VGG16 model works with two data sets. Then comes to the consequences of two data sets, the key points of this comparative research. This research primarily uses accuracy curve and learning curve as evaluation indicators. It not only highlights the challenges that CNN may encounter when dealing with complicated data sets, but also offers a framework for evaluating and comparing the performance of CNN. Throughout this paper, it is concluded that the accuracy rate of different features may present diverse performance. While dealing with data of more weakly correlated features, the VGG16 model may exhibit significantly different accuracy rates in different features compared to other accuracy rates. Furthermore, when compared to data with simple features, the VGG16 model presents lower accuracy in data containing more detailed and complex features.


1 INTRODUCTION


Within the realm of computer vision, CNN, based on deep learning, is widely utilized for image classification, object detection, face recognition and so forth. As CNN can extract hierarchical features from raw images, it has merged as the leading approach for tackling various image recognition and classification tasks. VGG16 model, one of the most popular CNN architectures used for image recognition and classification, is introduced by the Visual Geometry Group at Oxford University. This research mainly introduced the information related to CNN as well as VGG16 model. Also, it presents the application of CNN in image recognition and classification.

At present, worldwide research based on deep learning has achieved relatively more results, which

primarily involves the image identification, speech recognition, short-term traffic flow prediction, environmental awareness training and so on. For instance, the detail that illustrates the usage of CNN to carry out feature extraction and classification, selection of network structure, data processing, a training strategy and performance evaluation was presented (Bensedik et al., 2018). In addition, the application of compressed sensing technique and deep learning technique used to detect vehicle classification in vehicle network showed great reference (Li et al., 2018). This paper also learns from YOLO, an object detection algorithm based on CNN. As a representative example of one-stage target detection algorithm, the YOLO series algorithm has the advantages in speed and precision, and is commonly utilized for vehicle detection tasks. Sang et al. proposed an improved vehicle detection

^a <https://orcid.org/0009-0004-6667-9922>

^b <https://orcid.org/0009-0007-3697-5431>

^c <https://orcid.org/0009-0001-9294-6500>

algorithm called YOLOv2, which introduced the feature pyramid network (FPN) and multi-scale prediction techniques to improve the detection ability of different scales and different sizes of targets. Additionally, it also used batch normalization and residual connection techniques to improve the training effect and generalization ability of the network (Sang et al., 2018). In another research which is based on YOLOv4 algorithm, the optimization of YOLOv2 algorithm using convolutional spatial proposal network to predict more precisely as well as conditional instance segmentation to perform instance segmentation and category prediction at the same time, the research achieved not only the accurate classification of vehicles, but also the identification of License plate number (Park et al., 2022). After that, another research provided key insights into the employment of deep learning for vehicle type detection in traffic scenarios, laying the foundation for building more efficient vehicle classification systems (Li et al., 2018). Lastly, Njayou created a model which had been equipped with appropriate loss function and optimizer and ran it with two different methods (Faster R-CNN and YOLOv4) to make a comparison (Njayou, 2022).

Based on VGG16 model, Dhuri did a test in three different data sets using various VGG16 models with diverse feature sets. The performance is evaluated by different indexes like accuracy, precision, recall and F1 scores. The result finds that the selection of features has a notable impact on the performance of VGG16 in image recognition and classification tasks (Dhuri, 2021). What's more, Aytekin et al. suggests that VGG16 model, as a powerful deep learning model, performs well in image recognition and classification tasks, thereby providing new possibilities for driver dynamic detection (Aytekin and Khan, 2021).

When it comes to image identification, Scholars have primarily used methods based on CNN for utility in the research of image classification and recognition. For example, Chauhan et al. suggested a vehicle classification method derived from embedded CNN for the classification of vehicle types in non-lane traffic. This method used CNN to extract features from images and classified vehicle types through a classifier (Chauhan et al., 2019). Besides, the study by Gao and Xiao presented that the cascaded and CNN-based approach showed superior performance in real-time Chinese traffic warning sign recognition. This method illustrated that the cascaded classifiers were used for preliminary screening of images, and CNN was used for fine classification and recognition of candidate areas, which achieved efficient and

accurate recognition of traffic warning signs (Gao and Xiao, 2021). Liu and Wang also proved that CNN performed well in real-time anomaly detection like network traffic anomaly detection. Therefore, in the process of traffic image recognition and analysis, the CNN-based method has higher accuracy and real-time in real-time anomaly detection (Liu and Wang, 2023). The experience can be also learnt about in Qiao's research. Qiao proposed a recognition approach rooted in CNN, highlighting the disparities between traditional traffic sign recognition techniques and their prevalent problems, and demonstrated the recognition performance and generalization ability of CNN (Qiao, 2023). Eventually, according to the research of Chen et al., they achieved 97.88% accuracy in the classification of vehicle types within traffic surveillance videos by leveraging convolutional neural networks. This result not only proved the effectiveness of deep learning in processing image recognition and classification data, but also provided a useful starting point for the further research (Chen et al., 2017).

In conclusion, CNN has great significance and wide application prospects in image recognition and classification. Despite some achievements it has made, there are still a significant number of challenges, such as overfitting, under fitting, limitation of computing resource and over-dependence on image quality. To find out the factors which can affect the efficiency and accuracy of CNN, this paper mainly uses VGG16 model to test different data sets with diverse features.

2 METHODS

2.1 Data Source and Statement

CIFAR-10 is the first data set collected by Alex Krizhevsky and his colleagues. CIFAR-10 is a moderate-sized data set including 60,000 32x32 color images in 10 classes like different animals and machines. It is split into 50000 training images and 1000 testing image. And the other data set which is called BIT-Vehicle contains 9850 vehicle images. It is collected by Beijing university of science and technology. The images include changes in lighting conditions, scales and vehicle surface colors. All the vehicles are divided into 6 classes: bus, microbus, minivan, SUV and truck. In this research, data sets will be uploaded and normalized to the range [-1,1].

2.2 Index Selection and Description

The main evaluation metric in this research is accuracy curve and learning curve. In addition, loss function adjusts the model's parameters to optimize its performance and minimize prediction errors. The optimizer selected is stochastic gradient descent, with a learning rate set to 0.001 and a momentum of 0.9, enabling efficient parameter updates during the training process. Every 2000 batches of training, the current average loss is printed. Additionally, the learning rate scheduler is also used to multiply the learning rate by 0.1 every 5 epochs. This allows the model to dynamically adjust the learning rate during training to optimize the process.

2.3 Methodology Introduction

In this research, two different results based on VGG16 model will be compared to each other primarily for image classification. Then the consequences will be compared to another result based on the other data set.

For the VGG16 model, it is made up of 16 weight layers containing 13 convolutional layers and 3 full-connected layers. Firstly, the image is pre-processed by adjusting the input image size to 224*224. Then the previous convolution layers extract features such as edges, textures, and colors from the image. After that, the extracted features are sent to the full-connected layers for classification. The fully connected layer can be regarded as an ordinary neural network, which learns how to identify the category of the image based on the extracted features. Finally, the VGG16 model outputs a probability distribution representing the probability that the image belongs to each category. We can choose the category with the highest probability as the predicted category for the image.

VGG16 model is based on CNN, one of the most typical models in deep learning. Accordingly, the following formulas based on CNN are also fundamental to VGG16 model:

Convolutional layer slides on input data and run point multiplication through the convolution kernel to study local features of the input data. For a bidimensional input data I and a convolution kernel K , the convolution operation can be expressed as:

$$(I * K)(i, j) = \sum m \sum n I(m, n) K(i - m, j - n) \quad (1)$$

Among them, $*$ represents convolution operation, i and j denote the location of bidimensional

consequence, m and n represent position on the input data and convolution kernel.

Pooling layers are primarily used for dimensionality reduction to prevent overfitting. Usual pooling operation contains max pooling and average pooling. For a bidimensional input data I , max pooling and average pooling can be described by the equation below:

$$MaxPool(I) = \max(I) \quad (2)$$

$$AvgPool(I) = \frac{1}{N} \sum I \quad (3)$$

Where N is the number of elements in the pooling window.

The Full-connected layer is commonly used to classify or regress in the last few layers in network. The fully connected layer serves to link all the outputs from the preceding layer to each individual neuron in the present layer, ensuring comprehensive connectivity. For the input X and weight W , the output of full-connected layer is defined by the following equation:

$$Y = W^T X + B \quad (4)$$

Where Y represents the output of fully connected layer, W^T is the transposition of weight, B denotes bias item.

3 RESULTS AND DISCUSSION

3.1 Analysis on CIFAR-10

According to the second step, the evaluation metrics of the VGG16 model include accuracy curve and learning curve. Therefore, the next step is to analyze these results.

This test shows great performance on the ship category judging by the results, which reach on 96% accuracy. The subsequent categories are frog and horse, with 95% and 92% accuracy. However, this model is relatively poorly on the 'cat' and 'bird', with accuracy rates of 70.5% and 74.5%. This possibly be thought to the significant variations in appearance these categories. Moreover, variety of colors and body types complicate the challenge for the models to accurately distinguish them. As a result, the model performs poorly on both categories. The specific data of VGG16 model on CIFAR-10 is shown in the Table 1 below.

Table 1: Accuracy rate of VGG16 model on CIFAR-10.

Category	Accuracy rate	Category	Accuracy rate
Plane	90.00%	Dog	75.00%
Car	92.00%	Frog	95.00%
Bird	74.50%	Horse	92.00%
Cat	70.50%	Ship	96.00%
Deer	84.00%	Truck	90.00%

Overall, the model has averagely accurate precision of 84.45% on the CIFAR-10. The test shows that VGG16 model has great performance on the CIFAR-10.

The precision rate of VGG16 model when tested on CIFAR-10 is depicted in Figure 1 that follows.

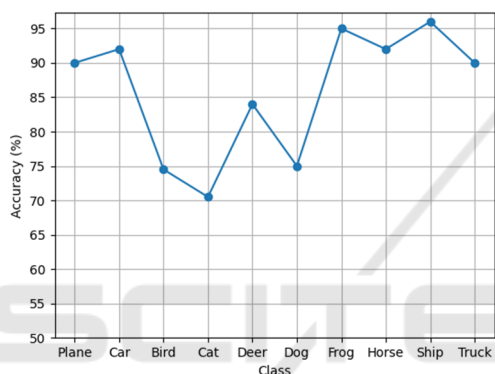


Figure 1: Accuracy of VGG16 CNN on CIFAR-10.

3.2 Analysis on BIT-Vehicle

For the BIT-Vehicle data set, the results not only show the accuracy rate of VGG16 of 77.41% overall, but also show that the sedan has the highest recognition accuracy of 77% in each category, while the microbus has the lowest recognition accuracy of 65 percent. It may be due to the fact that the microbus has more shapes and color variations, brings greater challenges to the recognition of the model.

The specific data of VGG16 model on BIT-Vehicle is shown in the following Table 2. The accuracy rate of VGG16 model on BIT-Vehicle is exhibited in Figure 2.

Table 2: Accuracy rate of VGG16 model on BIT-Vehicle.

Category	Accuracy rate	Category	Accuracy rate
SUV	71.00%	Microbus	65.00%
Sedan	77.00%	Truck	68.00%
Minivan	72.00%	Bus	64.00%

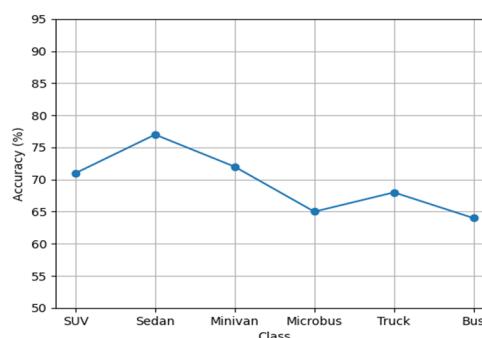


Figure 2: Accuracy of VGG16 CNN on BIT-Vehicle.

3.3 Comparison and Discussion

This research takes VGG16 model as an example, exploring the detection accuracy of convolutional neural networks in processing different kinds and spans of data sets. It is clearly to be seen from the learning curve which is shown in the following figure that the loss values decrease gradually, representing that VGG16 model can improve its ability of fitting for training data in the process of learning. It can optimize itself constantly and make its prediction of training data approaching the true value gradually.

Refer to Figure 3 below for the learning curve.

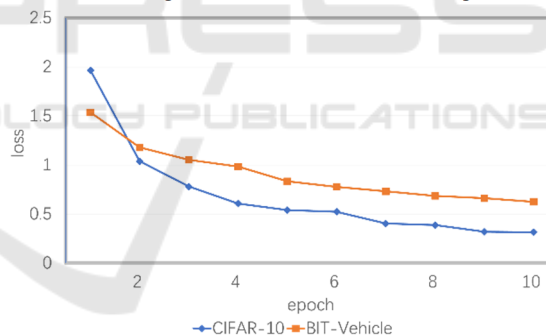


Figure 3: Learning curve of two data sets after training.

The CIFAR-10 data set includes lots of data in poor correlation. However, the average accuracy of CIFAR-10 data set is higher than that of BIT-Vehicle data set. It's probably due to the more sophisticated data with more details used in BIT-Vehicle data set. By comparing the two test results, it's clear that the accuracy difference between the highest one (ship) and the lowest one (cat) is about 25.5%. However, the result of the second group of data shows that the accuracy difference between the highest one (Sedan) and the lowest one (bus) is about 13%. It can be clear that the VGG16 model has a large deviation between the test accuracy of each feature when processing data with more complex, wide content distribution and

weak correlation features (like cats and cars in the first data set), while when processing data with strong correlation features (such as different models between cars in the second data set), the deviation between the test accuracy of each feature is small.

4 CONCLUSION

This research discusses the performance of VGG16 model in dealing data sets of different complexity and kinds through two experiments. The results of experiments show that VGG16 model has great performance on the simpler CIFAR-10 with exactness rates of 84.45%, however the performance of the VGG16 model decreased when it deals with complex BIT-Vehicle data set, and the training accuracy dropped to 77.41%. The accuracy of all categories in BIT-Vehicle data set are less than 80%.

The results of research show that the performance of CNN may be influenced in dealing datasets with different kinds and more span range. It may need more training samples and deeper network structure to extract effective features especially facing data sets with high complexity. Furthermore, the recognition accuracy of different categories exists big differences in the BIT-Vehicle data set. This may be due to the sample amounts of different categories in the data set is imbalance, or the complexity of the different categories is too high.

This research not only presents the challenges CNN may face while dealing with complicated data sets, but also provides a framework for evaluation and comparison of CNN. It provides a significant reference for subsequent image recognition in practical applications.

In general, this research provides a preliminary framework for evaluating and comparing the performance of CNN when processing data sets with different features. In future research, the exploration of optimizing the parameters and structure of the CNN model to improve its performance on different data sets will be promoted further. Moreover, it is necessary to design a more efficient model which can fit with further and deeper data or use more data to train and test while facing with diverse type of features.

AUTHORS CONTRIBUTION

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

REFERENCES

- Aytekin, A., Mençik, V., 2022. Detection of Driver Dynamics with VGG16 Model. *Applied Computer Systems*, 27(1): 83-88.
- Bensedik, H., Azough, A., Meknasssi, M., 2018. Vehicle type classification using convolutional neural network, *In 2018 IEEE 5th International Congress on Information Science and Technology (CiSt). Institute of Electrical and Electronic Engineers*, 313-316.
- Chauhan, M.S., Singh, A., Khemka, M. et al., 2019. Embedded CNN Based Vehicle Classification and Counting in Non-Laned Road Traffic. *Information and Communication Technologies and Development*, 1-11.
- Chen, Y., Zhu, W., Yao, D., Zhang, L., 2017. Vehicle Type Classification based on Convolutional Neural Network. *In 2017 Chinese Automation Congress (CAC). Institute of Electrical and Electronic Engineers*, 1898-1901.
- Dhuri, V., Khan, A., Kamtekar, Y. et al., 2021. Real-Time Parking Lot Occupancy Detection System with VGG16 Deep Neural Network using Decentralized Processing for Public, Private Parking Facilities. *In 2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, 1-8.
- Gao, Y., Xiao, G., 2021. Real-time Chinese Traffic Warning Signs Recognition Based on Cascade and CNN. *Journal of Real-Time Image Processing*, 18(3): 669-680.
- Li, S., Lin, J., Li, G. et al., 2018. Vehicle Type Detection Based on Deep Learning in Traffic Scene. *Procedia Computer Science*, 131: 564-572.
- Li, Y., Song, B., Kang, X. et al., 2018. Vehicle-Type Detection Based on Compressed Sensing and Deep Learning in Vehicular Networks. *Sensors*, 18(12): 4500.
- Liu, H., Wang, H., 2023. Real-Time Anomaly Detection of Network Traffic Based on CNN. *Symmetry*, 15(6).
- Njayou, Y., 2022. Traffic Sign Classification Using CNN and Detection Using Faster-RCNN and YOLOV4. *Heliyon*, 8(12).
- Park, S.H., Yu, S.B., Kim, J.A. et al., 2022. An All-in-One Vehicle Type and License Plate Recognition System Using YOLOv4. *Sensors*, 22(3): 921.
- Qiao, X., 2023. Research on Traffic Sign Recognition based on CNN Deep Learning Network. *Procedia Computer Science*, 228: 826-837.
- Sang, J., Wu, Z., Guo, P. et al., 2018. An Improved YOLOv2 for Vehicle Detection. *Sensors*, 18(12): 4272.