Vehicle Face Recognition Algorithm Based on Fusion of Siamese Neural Network

Chunhe Shi¹²*, Chengdong Wu¹ and Yuan Gao¹

¹College of Information Science and Engineering, Northeastern University, Shenyang, China
²School of Information Engineering, Shenyang University, Shenyang, China

*Corresponding author e-mail: schsydx@syu.edu.cn

Abstract. Because the license plate is easy to be altered, blocked and forged, the method of license plate recognition alone cannot accurately and quickly confirm the vehicle identity. This paper proposes a vehicle re-recognition method based on the fusion of Siamese deep neural network. The method is based on the huge differences in the face area of each car. First, the YOLOv3 algorithm is improved to detect the face area of the vehicle picture, and then the improved Siamese network algorithm recognition is used. Finally, the feature output of vehicle face image is mapped to Euclidean distance for vehicle face recognition. The YFSDNN (YOLOv3 Fusion Siamese Deep Neural Network) method is evaluated in data sets and experiments. The experimental results show that this method not only has high accuracy, but also greatly improves the detection and recognition speed, which can meet the real-time needs of vehicle face recognition.

Keywords: Vehicle Face Recognition, YFSDNN, YOLOv3, Siamese Network

1. Introduction

With the development of research technology, Vehicle recognition has attracted extensive attention. In the public safety system, the demand for vehicle recognition is increasing. As an important feature of vehicle appearance, license plate is easy to be altered, blocked and forged by lawbreakers. Therefore, the method of license plate recognition alone cannot accurately and quickly confirm the vehicle identity. Therefore, vehicle face recognition based on vision has high practical value in monitoring applications.

Although the research of vehicle face recognition has been for some years, most of them are realized with the help of hardware equipment, [1-4] and so on. Although these methods can improve the accuracy of identification, they are not conducive to promotion due to high cost, difficult maintenance and limited application scenarios. Traditional car face recognition methods mainly rely on LBP, Haar, HOG [5] and other methods to extract features. These methods are mostly based on experience, which has great limitations. Compared with the classical face recognition problem, vehicle face recognition is more challenging, because the same type of vehicle has a highly similar appearance. If you don't use the license plate, it’s difficult to distinguish between the same type of vehicles.
However, there are still some special marks that can be used to distinguish from other vehicles. For example, there are huge differences in the area of the car face, especially some information displayed on the windshield, including the location, color and quantity of the pasted marks, decorations and even vehicle scratches. With the rapid development and major breakthrough of computer vision theory of deep learning, convolutional neural network is widely used the network training model is invariant to image distortion such as scaling, translation and rotation, and has strong generalization, which solves the problems of manual design, function limitation and robustness of traditional image features. At present, the classic convolutional neural network models include AlexNet [6], ResNet [7], MobileNet [8], VGGNet [9], etc.

The YFSDNN method proposed in this paper is a framework specially designed for vehicle face recognition. In order to solve the problem of vehicle face recognition, the difference of vehicle face area is specially focused on. Experiments show that this method has the characteristics of high accuracy and good real-time performance.

The chapters of this article are as follows: the related work is summarized in the second part. The third part discusses the YSDNN algorithm for vehicle face recognition. The fourth part introduces the experimental results and analysis.

2. Related Work

Most detection and recognition work currently targets human or faces. Usually, given an image and multiple candidate images as a library, it is necessary to determine the same object. Although there is not much research on vehicle face recognition, it is as important as face recognition in practical application. More research on vehicles includes: vehicle classification [10] and vehicle verification [11]. However, these methods can only be used for preliminary identification and cannot identify whether it is the same vehicle.

There are fewer articles on vehicle face recognition [12]. The human identification method is used to complete the vehicle task, and good results are achieved [13]. This paper discusses how to ensure the accuracy of classification by using deep separable neural network under weak light condition [14]. The DCNN (Deep Convolution Neural Network) is used to extract feature descriptors for each proposed region. Finally, the SVM (Support Vector Machine) is used to score and classify the proposed regions, and good results are achieved according to different vehicle types.

Based on the above analysis, this paper proposes YFSDNN method, the difference in the face area of the car, especially the windshield area, is used as a more fine-grained feature to solve this problem. The innovation of this method has two aspects: (1) for vehicle images, the improved YOLOv3 is used for detection and recognition, which provides smaller scale and more abundant information for the subsequent Siamese network, and improves the network processing speed. In the case of little change in accuracy, it is 20% of the time of Faster R-CNN; (2) the improved Siamese network reduces the number of parameters while ensuring the accuracy rate, thereby reducing the calculation time, which is 75% lower than the original system, and the processing speed is increased, laying a foundation for the widespread use of the model.

3. YSFDNN Vehicle Face Recognition

3.1. Improved Vehicle Face Detection Algorithm Based On YOLOv3

YOLO algorithm was originally a target recognition method based on regression proposed by REDMON et al. [15] in 2016, and it has been released in 2018. YOLOv3 is the target detection network with the most balanced speed and accuracy. It makes some adaptive improvements on the basis of YOLOv2, including multi-scale recognition, multi label classification and so on, and uses the improved DARKNet-53 network based on ResNet as the feature extractor, which improves the defect that YOLO series methods are not good at recognizing small objects. YOLOv3 still keeps the advantages of quick detection of YOLOv2, and the detection accuracy has been greatly improved.
Because the video based vehicle detection has high requirements for detection accuracy and real-time performance, although most vehicle detection methods can ensure the detection accuracy, the vehicle detection speed is far from meeting the requirements of video detection. Because YOLOv3 algorithm has fast detection speed and high accuracy, this paper uses YOLOv3 algorithm to realize the vehicle detection function based on image.

When training the YOLOv3 network model, the BN (batch normalization) layer can increase the speed of network convergence, thereby controlling overfitting, which is usually placed after the convolution layer. After the data is normalized by BN layer, the problem of gradient disappearance and gradient explosion can be solved effectively. Although the BN layer has a good effect during training, it adds computation in network forward reasoning, reduces the performance of the model and takes up a relatively large amount of memory or video memory. Therefore, it is necessary to merge the parameters of the BN layer into the convolutional layer to improve the speed of forward inference of the model. Insert the BN layer after the fully connected layer and before the nonlinear processing layer, namely: CNN-BN-ReLU. The calculation formula of convolution layer and BN layer before merging are shown in formula (1), (2).

$$\text{out}(j) = x(i) \times w(0) + x(i+1) \times w(1) + x(i+2) \times w(2) + \ldots + x(i+k) \times w(k) + b$$ \hspace{1cm} (1)

$$x_{out} = \frac{\alpha(\sum_{i=0}^{n} x_i w_i - \mu)}{\sqrt{\delta^2 + \varepsilon}} + \beta$$ \hspace{1cm} (2)

Where, $x_i$ is the image data, $w_i$ is the weight, $\alpha$ is the scaling factor, $\mu$ is the mean value, $\delta^2$ is the variance, $\beta$ is the bias, $\varepsilon$ is the minimum value (in order to prevent the denominator from being zero). The convolution formula is shown in formula (3).

$$x_{conv} = \sum_{i=0}^{n} (x_i w_i)$$ \hspace{1cm} (3)

In order to cancel the BN layer, formula 2 needs to be adjusted. The adjusted formula is shown in formula (4).

$$x_{out} = \sum_{i=1}^{n} (x_i \frac{\alpha w_i}{\sqrt{\delta^2 + \varepsilon}} - \frac{\alpha \mu}{\sqrt{\delta^2 + \varepsilon}} + \beta$$ \hspace{1cm} (4)

After adjustment, the new weight parameters are as follows formula (5):

$$w'_i = x_i \frac{\alpha w_i}{\sqrt{\delta^2 + \varepsilon}}$$ \hspace{1cm} (5)

After adjustment, the new bias show in formula (6).

$$\beta' = \beta - \frac{\alpha \mu}{\sqrt{\delta^2 + \varepsilon}}$$ \hspace{1cm} (6)

The final formula show in formula (7)

$$x_{out} = \sum_{i=0}^{n} x_i w'_i + \beta'$$ \hspace{1cm} (7)

After the above adjustment, the BN layer is cancelled, but the new weight parameter and offset are used.
Experimental data shows that using the optimized YOLOv3 algorithm, the accuracy of vehicle face detection is 93.4%, which meets the requirements of accuracy and calculation speed. Compared with the Faster R-CNN algorithm, the detection accuracy is 91.8%, but the detection time is 5 times of YOLOv3, which cannot meet the real-time requirements.

3.2. Improved Siamese Network Vehicle Recognition Algorithm

It’s important to identify specific vehicles and people, although there may be small differences between them. In theory, any two people’s appearance cannot be exactly the same, but for two cars, if their colors and models are the same, it is more difficult to distinguish them. In this paper, on the basis of extracting the windshield area above, the vehicle interior and annual inspection marks in the windshield window area are used as the identification of different vehicles to complete the vehicle identification for the same vehicle type. The car face recognition method refers to the recognition of different individuals by using the geometric shape and texture features of different facial features, as well as the position arrangement and high-level features of the whole face.

Paste marker of images has the following characteristics: relatively fixed position; limited color type; similar area and so on. Therefore, the quantity, content, arrangement and distribution of pasted signs become the focus of our attention as shown in Fig. 1.

![Image 1](image1.png)

**Figure 1.** Image of vehicle face and paste marker

This paper designs an improved Siam network to complete the task of vehicle face recognition. The Siamese network was first used to generate facial feature vectors for face recognition. The Siamese network learns a function that maps the input pattern to the potential space, in which the similarity measure is small for the same object and large for different objects. So it is very suitable for vehicle face recognition tasks.

To determine whether the cars in different images are the same, the above-mentioned pictures processed by YOLOv3 are input into two branches respectively, and each branch extracts the features of the car. The structure of the network is shown in Fig. 2.

![Image 2](image2.png)

**Figure 2.** YSDNN network structure diagram
During the training, two images are input to different CNN. After forward propagation, the output of CNN is merged into the contrast loss layer, and the loss of the model is calculated. Then, the sharing weight of CNN is optimized by back propagation. The data of two car images are mapped into the potential metric space, and the similarity between the images is calculated.

$W$ is the sharing weight of Siamese network. Give two pictures, called image1 and image2, the data is mapped to the potential space called $G_w(X_1)$ and $G_w(X_2)$. The energy function formula (8) measures the similarity of image1 and image2.

$$E_w(X_1, X_2) = ||G_w(X_1) - G_w(X_2)||$$  \hspace{1cm} (8)

Using the energy function, the contrast loss can be expressed as formula (9).

$$L(W,(X_1, X_2, y)) = (1-y)\max(m - E_w(X_1, X_2), 0) + y \cdot E_w(X_1, X_2)$$  \hspace{1cm} (9)

Where $(X_1, X_2, y)$ is two samples with $y$, and $m$ is the positive sample margin. In the test phase, Euclidean method is used to calculate the similarity of paired images.

VGGNet network model has the following characteristics: (1) small convolution kernel. Because the image size of the parts to be detected is small, using a smaller convolution kernel can reduce the amount of parameters and retain more underlying features. In this paper, we replace the convolution kernel with $3 \times 3$. (2) Small pool kernel. The pooling layer was replaced by $2 \times 2$. (3) The number of layers is deeper and the feature map is wider. VGGNet is an extension of CNN model AlexNet.

This article is improved on the basis of VGGNet-16 network model, and the network structure of improved is shown in Table 1. By changing the number of neurons in the last two fully connected layers in the VGGNet-16 network model from 4096 to 2048, the parameters and calculations of the fully connected layer are reduced, and nonlinear problems can be better solved to achieve classification. The full join layer can be regarded as a special convolution layer, the upper layer is $1 \times 1 \times 4096$, and the lower layer is $1 \times 1 \times 4096$. The convolution kernel of $1 \times 1$ is used for convolution operation, the calculation amount is $1 \times 1 \times 4096 \times 1 \times 4096 = 16777216$, and the parameter amount is $1 \times 1 \times 4096 \times 4096 = 16777216$. After the modification, the number of neurons in the upper and lower layers are both 2048, the calculation amount is $1 \times 1 \times 2048 \times 1 \times 2048 = 4194304$, the amount of computation and parameters are reduced by 75%, and the speed of neural network model training is accelerated.

| Number | Type     | Kernel size | Parameters | Number | Type     | Kernel size | Parameters |
|--------|----------|-------------|------------|--------|----------|-------------|------------|
| 1      | Conv_1   | $3 \times 3$ | 64         | 11     | Conv_8   | $3 \times 3$ | 512        |
| 2      | Conv_2   | $3 \times 3$ | 64         | 12     | Conv_9   | $3 \times 3$ | 512        |
| 3      | Pool_1   | $2 \times 2$ | 13         | 13     | Conv_10  | $3 \times 3$ | 512        |
| 4      | Conv_3   | $3 \times 3$ | 128        | 14     | Pool_4   | $2 \times 2$ |           |
| 5      | Conv_4   | $3 \times 3$ | 128        | 15     | Conv_11  | $3 \times 3$ | 512        |
| 6      | Pool_2   | $2 \times 2$ | 16         | 16     | Conv_12  | $3 \times 3$ | 512        |
| 7      | Conv_5   | $3 \times 3$ | 256        | 17     | Conv_13  | $3 \times 3$ | 512        |
| 8      | Conv_6   | $3 \times 3$ | 256        | 18     | Pool_5   | $2 \times 2$ |           |
| 9      | Conv_7   | $3 \times 3$ | 256        | 19     | FC_1     |              | 2048       |
| 10     | Pool_3   | $2 \times 2$ | 20         | 20     | FC_2     |              | 2048       |

4. Experimental Results and Analysis
The data set used in this experiment is 103028 vehicle images collected from 22 traffic checkpoints in a city of China and the time information collected. Among them, nearly 90,000 images of 10,319 vehicles in the data set have been marked with license plate information, so they can be used for vehicle face recognition. The data is allocated according to the proportion of training set, test set and validation set, which is 8:1:1.
In this paper, in the task of vehicle face recognition, we first use the vehicle face image detected by YOLOv3, input the deep convolution network of Siamese network to extract the normalized features, and then compare the differences between the two car’s images by measuring the Euclidean distance. AUC (Area Under Curve) index and accuracy index to evaluate the quality of the model and the classification results. As show in formula 10.

$$AUC = P(p_{positive} > p_{negative})$$ (10)

The AUC is 0.9692 and the accuracy is 0.9624.

Then, using the same datasets, the recognition algorithm in [14] is selected for comparison, and the experimental comparison results as shown in Table 2.

**Table 2.** Experimental results comparison of different algorithms

| Algorithm | AUC       | Accuracy  |
|-----------|-----------|-----------|
| [14]      | 91.35%    | 92.17%    |
| Ours      | 96.92%    | 96.24%    |

5. Conclusion

This paper proposed a YFSDNN method to solve the problem of vehicle face recognition. Firstly, the improved YOLOv3 algorithm is used to quickly and accurately detect the face area of the image, and then the improved two branch deep convolution network algorithm is used to recognize the vehicle face, and the difference of the windshield area and other feature outputs are mapped to Euclidean distance to judge whether the two images belong to the same vehicle. Experiments show that the method proposed in this paper has high accuracy and AUC value, and also greatly improves the detection and recognition speed, has a certain degree of robustness, and has a good real-time effect of vehicle face recognition.

Acknowledgments

This work is supported by National Natural Science Foundation of China under Grant U1713216, 61973063.

References

[1] R. O. Sanchez, C. Flores, R. Horowitz, R. Rajagopal and P. Varaiya, "Vehicle re-identification using wireless magnetic sensors: Algorithm revision, modifications and performance analysis," Proceedings of 2011 IEEE International Conference on Vehicular Electronics and Safety, Beijing, 2011, pp. 226-231.

[2] WU H, MENDEL J M. Classifier designs for binary classifications of ground vehicles Unattended Ground Sensor Technologies and Applications V. International Society for Optics and Photonics. New York: IEEE, 2003, 5090: 122-134.

[3] J. M. Ernst, J. V. Krogmeier and D. M. Bullock, "Non-linear compensation of vehicle signatures captured from electromagnetic sensors with application to vehicle re-identification," 13th International IEEE Conference on Intelligent Transportation Systems, Funchal, 2010, pp. 923-928.

[4] URAZGHILDIEV I, RAGNARSSON R, RIDDERSTROM P, et al. Vehicle Classification Based on the Radar Measurement of Height Profiles. IEEE Transactions on Intelligent Transportation Systems, 2007, 8(2):245-253.

[5] A. Arunmozhi and J. Park, "Comparison of HOG, LBP and Haar-Like Features for On-Road Vehicle Detection," 2018 IEEE International Conference on Electro/Information Technology (EIT), Rochester, MI, 2018, pp. 0362-0367.

[6] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012, 25: 1097-1105.

[7] HEK, ZHANG X, REN S, et al. Deep residual learning for image recognition 2016 IEEE
Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE Computer Society, 2016: 770-778.

[8] HowarD A G, ZHU M, CHEN B, et al. Mobile Nets: efficient convolutional neural networks for mobile vision applications. arXiv preprint, 2017, 32(9): 100-108.

[9] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[10] M. A. El-Khoreby and S. Abd Rahman Abu-Bakar, "Vehicle detection and counting for complex weather conditions," 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Kuching, 2017, pp. 425-428.

[11] Maqbool S, Khan M, Tahir J, et al. Vehicle detection, tracking and counting 2018 IEEE 3rd International Conference on Signal and Image Processing (ICSIP). IEEE, 2018: 126-132.

[12] X. Yang, C. Lang, P. Peng and J. Xing, "Vehicle Re-Identification by Multi-Grain Learni," 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp. 3113-3117.

[13] Shi, C.; Wu, C.; Gao, Y. Research on Image Adaptive Enhancement Algorithm under Low Light in License Plate Recognition System. Symmetry 2020, 12, 1552.

[14] Adu-Gyamfi Y O, Asare S K, Sharma A, et al. Automated vehicle recognition with deep convolutional neural networks. Transportation Research Record, 2017, 2645(1): 113-122.

[15] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91