Vehicle Make Recognition based on Convolutional Neural Network

Yongbin Gao
Division of Computer Science and Engineering
Chonbuk National University
Jeonju, South Korea
E-mail: gaoyongbin.sam@gmail.com

Hyo Jong Lee*
Division of Computer Science and Engineering
Center for Advanced Image and Information Technology
Chonbuk National University
Jeonju, South Korea
E-mail: hlee@chonbuk.ac.kr

***Abstract*—Vehicle analysis has been investigated for decades, which involves license plate recognition, intelligent traffic. Among these applications, vehicle make recognition is a challenging task due to the close appearance between car models. In this paper, we propose an architecture to recognize vehicle make based on convolutional neural network (CNN). The moving car is first localized by frame difference, the resultant binary image is used to detect the frontal view of a car by a symmetry filter. The detected frontal view is used to train and test the CNN. Experimental results show that our proposed framework achieves favorable recognition accuracy.**

Keywords-Moving car detection; Car model recognition; CNN;

I. INTRODUCTION

Vehicle analysis is critical for intelligent traffic, which involves license plate recognition, driver assistance systems, automatic toll collection, self-guided vehicles, intelligent parking, and traffic information. However, few researches has been done for car make recognition, which can be used for electronic toll collection system, the recognition can be also used for police to search suspect vehicles. While the difficulty of car make recognition lies on the close similarity between models that were made by various companies.

The detection and matching of interest points serve as the basis for many computer vision applications including image/ video retrieval, object categorization and recognition, and 3-D scene reconstruction directions. Two of the most widely used detectors are the Harris corner detector (widely used in Europe) and the Kanade–Lucas–Tomasi corner detector (used in the U.S.). However, these corner detectors are not invariant to scale and affine transformations. To address this problem, Lowe [1] approximated the Laplacian matrix of Gaussians with a DOG filter to propose a rotation-invariant descriptor called SIFT, which computes a histogram of locally oriented gradients around the interest point and stores the bins in a 128-D vector.

In this paper, we propose a framework to detect moving car and its make based on convolutional neural network. We first detect the moving car using frame difference. The resultant binary image is used to detect the frontal view of a car by a symmetry filter. The detected frontal view is used to identify a car based on CNN algorithm. The framework of the proposed system is presented in Fig. 1. The proposed CNN architecture composed of several convolutional layers, max-pooling layers, Relu layer and softmax loss layer. The max-pooling layer increase the tolerance of translation in the image, and the Relu layer gains the non-linear properties of the network.

The remainder of this paper is organized as follows. Section II presents some related works of vehicle detection and vehicle make recognition, Section III describes the proposed CNN architecture, the layer details are presented in this Section. Section IV applies the above algorithm to our car database, and presents the experiment results. Finally, we conclude this paper in Section V.

II. RELATED WORKS

Vehicle detection is the first step prior to the vehicle make recognition. There are numerous research for vehicle detection. Background subtraction [2-5] is extensively used to detect the motion features of moving vehicles from videos, while this motion feature cannot be extracted in static images. To solve this problem, Wu et al. [6] proposed to use wavelet transformation to extract texture features for localizing the candidate vehicles from the road. These candidate vehicles were verified by a PCA classifier. Tzomakas and Seelen [7] used the shadow area underneath a vehicle to detect vehicle. Ratan et al. [8] found that the vehicle wheels are the cue to localize the possible vehicles and further verified by a diverse density method.

Vehicle analysis can be performed in a number of applications. Chen *et al.* [9] classified the vehicles on the road into four classes, namely, car, bus, van, and bicycle/motorcycle based on SVM and random forests. Ma and Grimson [10] first extracted edge points and modified SIFT descriptors of vehicles and then used a constellation model to classify vehicles into two classes, namely, cars versus minivans and sedans versus taxis. In these applications, the inter-classes difference is quite large, while regarding the car model recognition; the appearances of various models are very similar. AbdelMaseeh *et al.* [11] proposed to incorporate the global cues with local cues to recognize car make and model. Hsieh *et al.* [12] used a symmetrical SURF for vehicle detection and vehicle make recognition.

978-1-4673-8611-1/15/$31.00 ©2015 IEEE

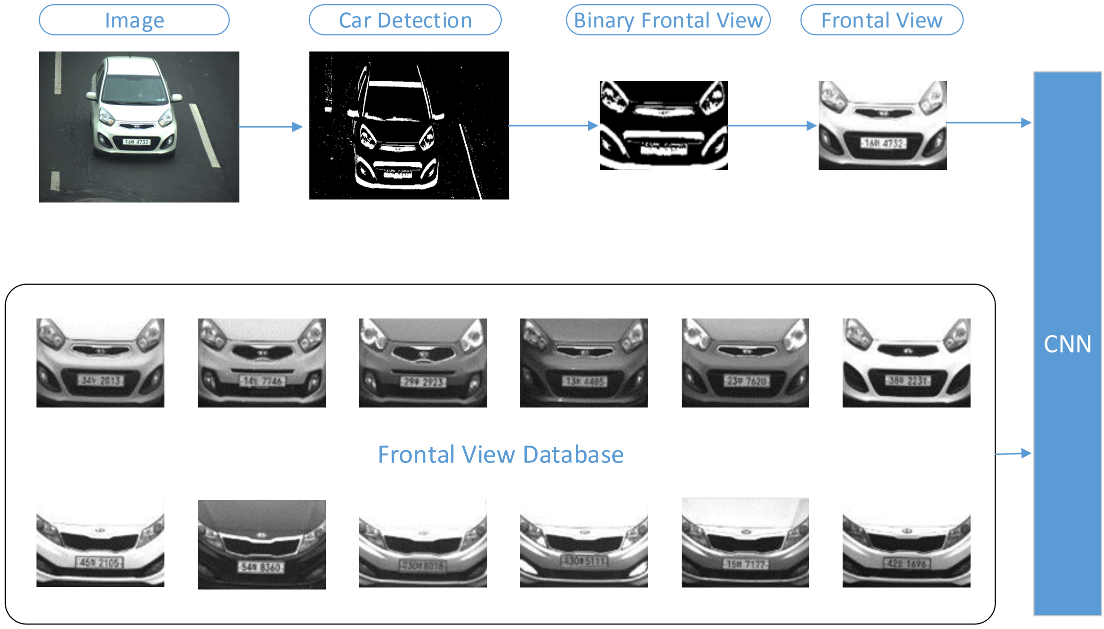

Fig. 1 Framework of proposed car detection and model recognition system based on CNN. Moving car is detected by frame difference; the resultant binary image is used to detect the frontal view of a car by a symmetry filter. The detected frontal view is used to identify a car based on CNN.

III. Proposed Convolutional Nerual Network

There are billions of parameters involved in the neural network, which requires much training time and large memory size. To avoid this limitation, convolutional neural network is developed whereby small portions of an image share the weights. This is effective since the spatial relationship between adjacent pixels. With different design of network architecture, CNN is able to learn various presentation of original input image. Fig. 2 presents the architecture of proposed convolution neural network. The input image was resized into size of 67x67, and fed into a convolutional layer by a filter of 10x10 following by a max-pooling layer, which results in 20 feature maps. This process was repeated one more time and produced 50 feature maps. After that, 500 neurons were generated by performing a convolutional layer. These neurons were forward to a Relu (Rectified Linear Units) layer to gain the nonlinearity. Finally, the neurons are full connected and forwarded to softmax loss to produce the 107 classes. These layers are interpreted as follows:

A. Convolutional Layer

Convolutional layer is the main layer of the network, where a kernel is defined to filter the input data. The input data is filtered by various kernels and results in different feature map. Support that the size of input image is $N \times N$, and we use a $m \times m$ filter as kernel and the number of resultant feature map is *K*, the *k*-th feature map at a given layer h^k is calculated as follows:

$$h_{ij}^{k} = \tanh\left((W^k * x)_{ij} + b_k\right)$$

where W^k and b_k are the weights and bias of *k*-th feature map, and the size of resultant feature map is *N-m+1*. The tanh(.) is a nonlinear function to fulfil the non-linearity property of convolutional neural network. In this way, the input image is characterized by different feature map using different filter. Through layer by layer convolutional operation, we are able to learn progressive level of features. For example, the first layer is low-level features, such as edges, lines and corners.

B. Max-pooling Layer

In order to increase the robustness of CNN to handle translations, pooling layer is usually used after the convolutional layer. The max value or the average value of a local feature map is widely used as a pooling technical. In our architecture, we use max-pooling layer to ensure that the same result can be achieved even there are translations existed. Support that the size of local region is $n \times n$, the output layer size will be $\frac{N}{n} \times \frac{N}{n}$.

C. Relu Layer

The Relu layer is used to gain the non-linearity of the network. This layer uses the non-saturating function as $f(x) = max(0, x)$, which has the non-linear property and has no affection of the receptive fields of the convolution layer.

D. Softmax Loss Layer

Softmax loss layer is used to predict the probability of K mutually exclusive classes. The softmax loss layer is usually designed at the final layer as follows:

$$\mathcal{L}(y,z) = -\log\left(\frac{e^{z_y}}{\sum_{j=1}^{m} e^{z_j}}\right)$$

where *y* and *z* are the class and predicted value of input data, respectively. $\mathcal{L}(y,z)$ is the probability of predicted value to be class *y*.

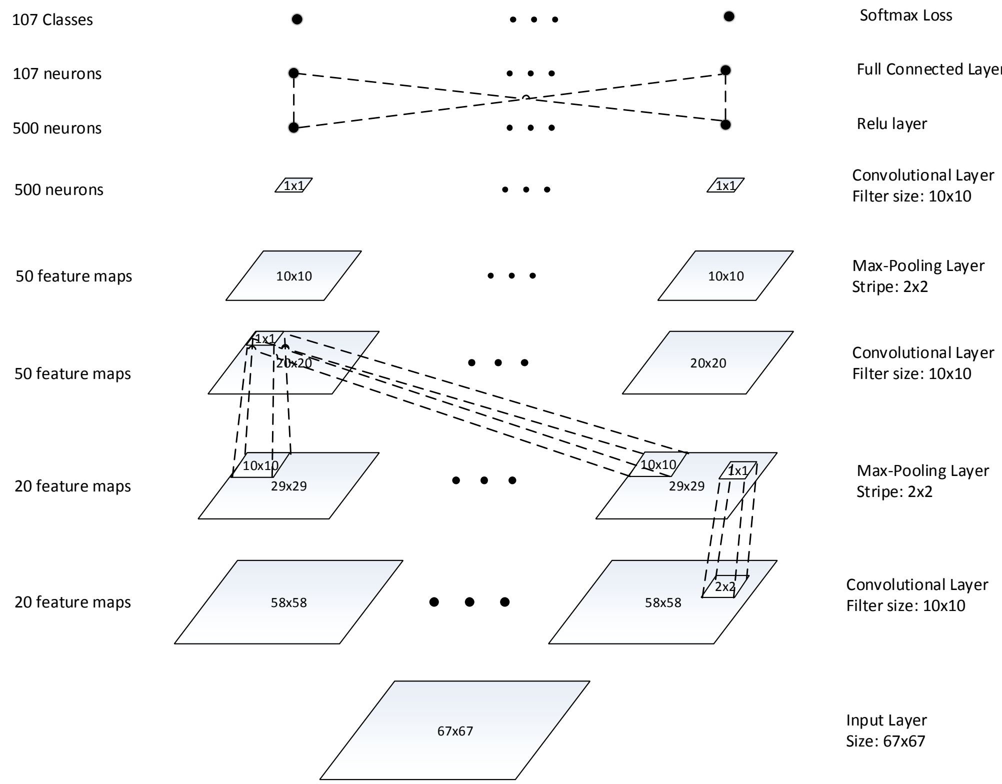

Fig. 2 Architecture of proposed convolutional neural network

IV. RESULTS

Moving car detection is the prerequisite of car analysis. In this paper, we use frame difference to detect the moving car on account the fact that our camera is fixed on the street. Therefore, we use frame difference to detect the moving car, which is simple and sufficient that enables real time implementation. The 3D characteristic of a car increases the difficulty to recognize the car make, thus, we proposed to use frontal view of a car instead of entire vehicle for vehicle make recognition. As we can see, the structure of the frontal view of a vehicle is basically symmetrical, to this end, a symmetrical filter is applied to the frontal view of a car. As a result, the symmetrical region is regarded as a frontal view of a car.

A car dataset was collected to evaluate the performance of our proposed framework. The dataset consists of 3210 car images with varying companies and models, which consists of 107 car models with 30 images for each model. We are apt to use image instead of video, because it is easy to measure the accuracy. However, it is necessary to generate another frame to apply frame difference to image, we shifted each image with 10 pixels to create a neighboring image. The difference image was generated by an original image and its shifted image.

Fig. 3 shows the frontal view detection results, which includes four car models of different companies. Left column exhibits the original car image with detected frontal region marked by red box. Right column presents the binary image of frontal view. The results indicate that our system is able to detect the frontal view of each car image accurately. We test the detection algorithm on all 3210 images in the system, and achieved 100% accuracy in terms of the detection accuracy. Thus, the detection algorithm is effective and fast in our system.

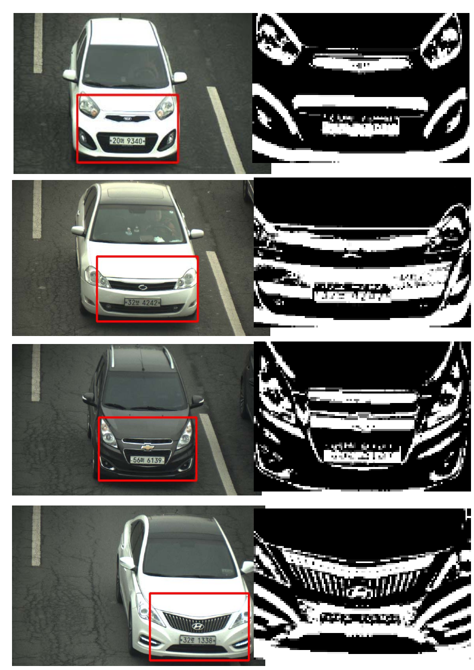

Fig. 3 Results of frontal view extraction on five car images with four companies and five models.

Once the frontal view of a car was obtained, we fed into our proposed CNN model to train and test the algorithm. In our experiments, 29 images of each car model are used for training, and the left one was used for test. We compared our CNN model to the following widely used features: local binary pattern (LBP) [14], local Gabor binary pattern (LGBP) [15], and scale-invariant feature transform (SIFT) [1]. Experimental results exhibit that our CNN model is able to improve over 10% accuracy towards the prestigious feature methods.

TABLE 1. PERFORMANCE COMPARISON OF CAR MODEL RECOGNITION WITH PRESTIGIOUS METHODS.

Algorithm	Accuracy (%)
LBP	46.0
LGBP	68.8
SIFT	78.3
CNN	88.4

V. CONCLUSION

In this paper, we proposed architecture to identify a car based on CNN algorithm. The proposed CNN architecture composed of several convolutional layers, max-pooling layers, Relu layer and softmax loss layer. The max-pooling layer increase the tolerance of translation in the image, and the Relu layer gains the non-linear properties of the network. Through layer by layer convolutional operation, we are able to learn progressive level of features. Experimental results exhibit that our CNN model achieved 10% accuracy more than the prestigious feature methods. The future work of our study is to extend our algorithms to deal with the variation in illumination for vehicle make recognition.

ACKNOWLEDGMENT

This work was supported by the Brain Korea 21 PLUS Project, National Research Foundation of Korea. This work was also supported by Business for Academic-industrial Cooperative establishments funded Korea Small and Medium Business Administration in 2014 (Grants No. C0221114).

REFERENCES

[1] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," Int. J. Comput. Vis., vol. 60, no. 2, pp. 91–110, Nov. 2004.

[2] A. Faro, D. Giordano, and C. Spampinato, "Adaptive background modeling integrated with luminosity sensors and occlusion processing for reliable vehicle detection," IEEE Trans. Intell. Transp. Syst., vol. 12, no. 4, pp. 1398–1412, Dec. 2011.

[3] H. Unno, K. Ojima, K. Hayashibe, and H. Saji, "Vehicle motion tracking using symmetry of vehicle and background subtraction," in Proc. IEEE Intell. Veh. Symp., 2007, pp. 1127–1131.

[4] A. Jazayeri, H.-Y. Cai, J.-Y. Zheng, and M. Tuceryan, "Vehicle detection and tracking in car video based on motion model," IEEE Trans. Intell. Transp. Syst., vol. 12, no. 2, pp. 583–595, Jun. 2011.

[5] G. L. Foresti, V. Murino, and C. Regazzoni, "Vehicle recognition and tracking from road image sequences," IEEE Trans. Veh. Technol., vol. 48, no. 1, pp. 301–318, Jan. 1999.

[6] J. Wu, X. Zhang, and J. Zhou, "Vehicle detection in static road images with PCA-and-wavelet-based classifier," in Proc. IEEE Conf. Intell. Transp. Syst., Aug. 25–29, 2001, pp. 740–744.

[7] C. Tzomakas and W. Seelen, "Vehicle detection in traffic scenes using shadow," Inst. Neuroinf., Ruhtuniv., Bochum, Germany, Tech. Rep. 98-06, 1998.

[8] A. Lakshmi Ratan,W. E. L. Grimson, andW. M.Wells, "Object detection and localization by dynamic template warping," Int. J. Comput. Vis., vol. 36, no. 2, pp. 131–148, Feb. 2000.

[9] Z. Chen, T. Ellis, and S. A. Velastin, "Vehicle type categorization: A comparison of classification schemes," in Proc. 14th Int. IEEE Conf. Intell. Transp. Syst., Oct. 2011, pp. 74–79.

[10] X. Ma, W. Eric, and L. Grimson, "Edge-based rich representation for vehicle classification," in Proc. IEEE Int. Conf. Comput. Vis., 2005, pp. 1185–1192.

[11] M. AbdelMaseeh, I. BadreIdin, M. F. Abdelkader and M. EI Saban, "Car Make and Model recognition combining global and local cues," in Proc. IEEE Int. Conf. Pattern Recognition, 2012, pp. 910-913.

[12] J. W. Hsieh, L. C. Chen and D. Y. Chen, "Symmetrical SURF and Its Applications to Vehicle Detection and Vehicle Make and Model Recognition," IEEE Trans. Intell. Transp. Syst., vol. 15, no. 1, pp. 6-20, Feb. 2014.

[13] J. Beis, and D. G. Lowe, "Shape indexing using approximate nearest neighbor search in high dimensional spaces," In Proc. Computer Vision and Pattern Recognition, Puerto Rico, 1997, pp. 1000-1006.

[14] Hough, P.V.C.. Method and means for recognizing complex patterns. U.S. Patent 3069654. 1962.

[15] T. Ahonen, A. Hadid, M. Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 12, pp. 2037-2041, 2006.

[16] W. Zhang, S. Shan, W. Gao, X. Chen, H. Zhang, "Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A Novel Non-Statistical Model for Face Representation and Recognition," in Proc. IEEE International Conference on Computer Vision (ICCV), 2005, pp. 786-791.