Moving Vehicle Detection and Information Extraction Based on Deep Neural Network

Ihsan Ullah[1], Hyo Jong Lee [1, 2,*]
[1]Division of Computer Science and Engineering,
[2]Center for Advanced Image and Information Technology
*Corresponding author
Chonbuk National University, Jeonju 561-756, Korea
Ihsanullah736@gmail.com, hlee@chonbuk.ac.kr

Abstract - *In recent years, vehicle recognition has become an important application in intelligent traffic monitoring and management. Vehicle analysis is an essential component in many intelligent applications, such as automatic toll collection, driver assistance systems, self-guided vehicles, intelligent parking systems, and traffic statistics (vehicle count, speed, and flow). The main goal of our study is to extract the information from the moving vehicles like their make, model and type. We address the vehicle detection and recognition problems using Deep Neural Networks (DNNs) approach. Our proposed approach outperforms state-of-the-art method. We first detect the moving vehicle based on frame difference and then extract the frontal part of the vehicle based on symmetrical filter, the frontal part of the vehicle is fed into the deep architecture for recognition. The Top 1 accuracy of proposed VMMTR algorithm is 96.31%.Our method achieves promising results on image.*

Keywords: Vehicle Recognition, Deep Learning, Vehicles Analysis, Pattern Recognition, CNN.

1 Introduction

Deep learning is widely used in both academic and industry due to its impressive performance. In order to train deep network, large scale dataset is required, labelling of which is difficult and tedious. Thus, a cluster method is required for speedup the labelling process. For vehicle analysis, deep learning also achieved favorable performance, vehicle analysis is an essential component in many intelligent applications, such as automatic toll collection, driver assistance systems, self-guided vehicles, intelligent parking systems, and traffic statistics (vehicle count, speed, and flow). Specially, an electronic toll collection system can automatically collect tolls according to the identification of vehicle models. Also, the identification of vehicle models can provide valuable information to the police for searching suspect vehicles. The appearance of a vehicle will change under varying environmental conditions and market requirements. The shapes of vehicles between companies and models is very similar, which results in confusion in vehicle model recognition. This makes the vehicle model recognition a challenging task.

The development of digital image sensors and computer vision techniques over a great deal of advantages in enabling many important ITS applications and components such as Advanced Driver Assistance Systems (ADAS), Automated Vehicular Surveillance (AVS), traffic and activity monitoring, traffic behavior analysis, traffic management, etc. Identification and classification of vehicles is of great interest in these applications, owing to heightened security concerns in ITS.

AVS broadly includes Vehicle Detection, Identification, Classification, and Vehicle Tracking. Over the years, significant research has been done to solve challenges in vehicle identification, detection and tracking. However, to classify vehicles into fine categories such as makes and models, has gained attention only recently, and many challenges remain yet to be addressed [1-4]

LLorca et al [4] presented vehicle model recognition approach is presented modelling the geometry and appearance of car emblems (model, trim level, etc.) from rear view images. Thus, a generic methodology is defined to build a hierarchical structure of car-make-dependent vehicle model classifiers. The emblems location, size and variations are firstly learnt. Then, the appearance of each badge is modelled using a linear SVM binary classifier with HOG features and the outputs of each individual classifier are converted to an estimate of posterior probabilities. In [5] Image representation method has been applied to the application of vehicles make and model recognition. Each make, model class is represented as an over complete sub-lexicon of mid-level feature representation. The classification of vehicles is performed by comparing the visual words of probe image with the learned lexicon of training data using Euclidean distance. Jun et al [6] has represented a symmetric SURF descriptor applied to vehicle MMR system to detect vehicles and recognize their MMs extremely accurately. Zhang et al [7] presented, a scheme of cascade classifier ensembles has been proposed with rejection strategies. Rather than simply pursuing accuracy, the importance of reject option was stressed to minimize the cost of misclassifications.

Vehicle detection and classification are important parts of Intelligent Transportation Systems. They aid traffic monitoring, counting, and surveillance, which are necessary for tracking the performance of traffic operations. Existing methods use various types of information for vehicle detection and classification. The evolution of computer vision and image processing techniques, together with wide deployment of road cameras, facilitate image and video based vehicle detection and classification.

The traditional vehicle identification systems recognize makes and models of vehicles relying on manual human observations or automated license plate recognition (ALPR) systems, hardly meeting the real-time constraints. Both approaches are failure-prone and have several limitations. Firstly, it is practically difficult for human observers to remember and efficiently distinguish between the wide variety of vehicle makes and models. Secondly, it becomes a laborious and time-consuming task for a human observer to monitor and observe the multitude of screens and record the incoming or outgoing makes and models, or to even spot the make and model being looked for.

The vehicle make and model recognition (VMMR) systems that rely on license plates has the following disadvantages. License plates are easy to be forged, damaged, modified, or occluded. Also, there are some license-plates that can be ambiguous (e.g., between “0” and “O”). Moreover, in some areas, it may not be required to bear the license-plate at the front or rear. If the ALPR system is not equipped to check for license-plates at both (front and rear) views of the vehicle, it could fail. Consequently, when ALPR systems fail to correctly read the detected license-plates due to the above issues, the wrong make-model info could be retrieved from the license-plates registry or database.

To overcome the above shortcomings in traditional vehicle identification and classification systems, automated VMMR techniques have recently gained attention, but hardly meeting real-time processing speed requirements. The make and model of the vehicle recognized by the VMMR system can be cross-checked with the license-plate registry to check for any fraud. In this way, vision-based automated VMMR techniques, such as the ones proposed in this work, augment traditional ALPR-based vehicle identification and classification systems to further enhance security.

Vehicle detection is prerequisite of VMMR. In literature background subtraction [3, 5, 6, 8] is widely used for moving objects and vehicle detection in videos. However this feature is not applicable for still images. In order to address this problem Wu et al.[9] used wavelet transformation to extract texture features to locate possible vehicle ROI. Sun et al.[10] they proposed Gabor filters for vehicle feature extraction and support vector machines (SVM) for vehicle detection. Tzomakas and Seelen [1] used shadow in order to detect the vehicles. Various approaches for vehicle detection and classification have been proposed recently. Sivaraman and Trivedi [11] use active learning to learn from front part and rear part vehicle images, and achieves 88.5% and 90.2% precision respectively. Chen et al. [2] use a combination of Measurement Based Features (MBF) and intensity pyramid based HOG (IPHOG) for vehicle classification on front view road images. A rear view vehicle classification approach is proposed by Kafai and Bhanu [12]. They define a feature set including tail light and plate position information, then pass it into hybrid dynamic Bayesian network for classification.

The commonly used Deep Networks are convolutional neural networks (CNN) [13] and restricted Boltzmann Machine (RBM) [14]. In addition many advanced techniques are engineered into CNN structures, such as max pooling, average pooling, dropout and maxout. Recently many new layers are proposed in CNN in order to increase the accuracy and reduce the processing time such fire module and inception module. By going deeper in convolutional networks CNN perform better in various applications, such as AlexNet [15], Overfeat [16], GoogleNet [17], SqueezNet [18] and ResNet [19].

The focus of this paper is on developing novel approach to address the challenges in real-time and automated Vehicle Make, Model and Type Recognition (VMMTR), utilizing state-of-the art vision-based techniques.

The problem of automated vehicle classification into makes and models is an important task for AVS and other ITS applications. We provide the general architecture of vehicle classification systems in Figure 1. Most works first adopt a Vehicle Detection step which produces Regions of Interest (ROIs) containing the vehicles’ faces (front), segmented from the background. The Vehicle Classification systems then work on these ROIs. Depending on the granularity of classification, vehicle classification systems could be classified into three categories: Type, Make (Logo), or Make, Model and Type recognition, as depicted in Figure 2 The focus of our work is on automated Vehicle Make, Model and Type Recognition (VMMTR), which basically comprises of three steps: (1) Vehicle Detection, (2) Front Part Extraction, and (3) Classification based on Convolutional Neural Network (CNN), as shown in Figure 2. To specify, this paper proposes and investigates unexplored Deep Neural Network Classification approaches for VMMTR, to effectively tackle the issues and challenges therewith.

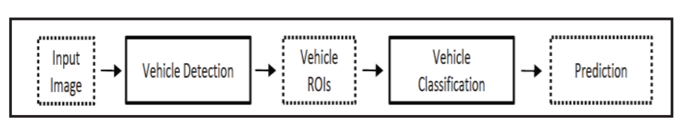

Figure 1 General Architecture of Vehicle Classification Systems

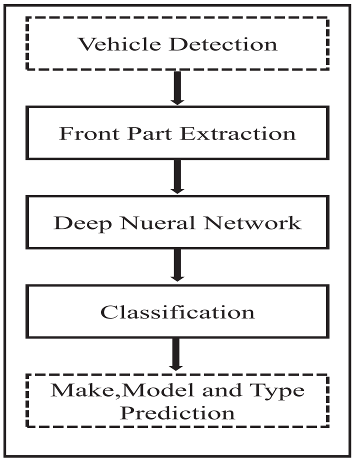

Figure 2 A Flow Chart of Vehicle Make Model and Type Recognition (VMMTR) Approach.

The remainder of this paper is organized as follows: Section 2 presents the proposed framework of MMR system which includes detection of vehicles, frontal part extraction and deep neural architecture; Section 3 describes the experiment results and discusses its performance. Conclusions is outlined in Section 4.

2 Framework of Proposed MMR System

The proposed MMR system involves the following three main steps: (1) Moving vehicle detection from a video, (2) extraction of frontal view of vehicles and (3) vehicle make and model recognition (MMR).

2.1 Moving Vehicle Detection

We first detect the moving vehicle based on frame difference it is effective in our proposed system because of the fact that the camera is fixed on the road. The frame differencing algorithm is simple and perform better in this fixed camera scenario and this scenario enables us to use this algorithm as part of real time application. It detected all most all vehicles in a video

2.2 Vehicles Frontal view extraction

The frontal view of a vehicle is basically symmetrical in shape, thus we used a symmetrical filter to extract frontal part of the vehicles. After the binary image of the vehicle is calculated from the frame difference, the symmetrical filter is used to extract the symmetrical regions of the binary image. As a result symmetrical region is regarded as frontal part of vehicles.

2.3 Vehicle Recognition

A deep neural network architecture of our proposed algorithm is shown in Figure 3 which is trained with extracted frontal part dataset. The architecture of GoogleNet is shown in Figure 5. The googleNet has 22 layer deep neural network, and almost 12x less parameters AlexNet. The googleNet has convolutional, relu, maxpooling, dropout and fully connected layers. In the googleNet a new module is introduced called inception module

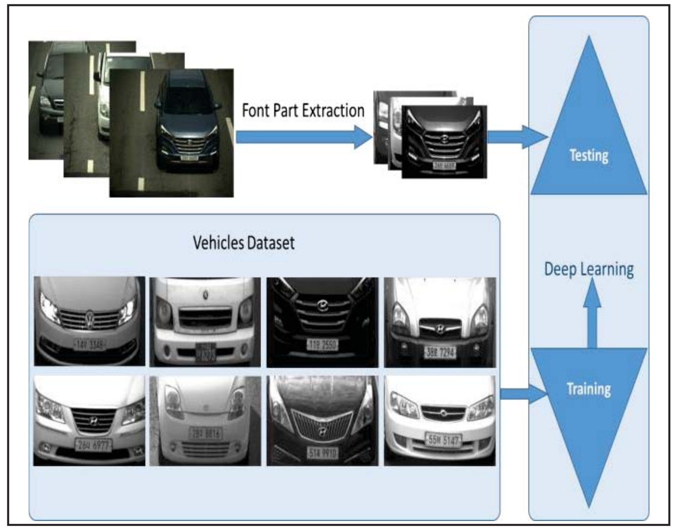

Figure 3. Our proposed framework.

2.3.1 Inception Modules

The basic building block of GoogLeNet, the inception module as shown in the Figure 4, is a set of convolutions and pooling at different scales, each done in parallel, then concatenated together. Along the way, 1×11×1 convolutions are used to reduce the dimensionality of inputs to convolutions with larger filter sizes. This approach results in a high performing model with drastically fewer parameters. GoogLeNet, in fact, has a factor of 12 times fewer parameters than AlexNet.

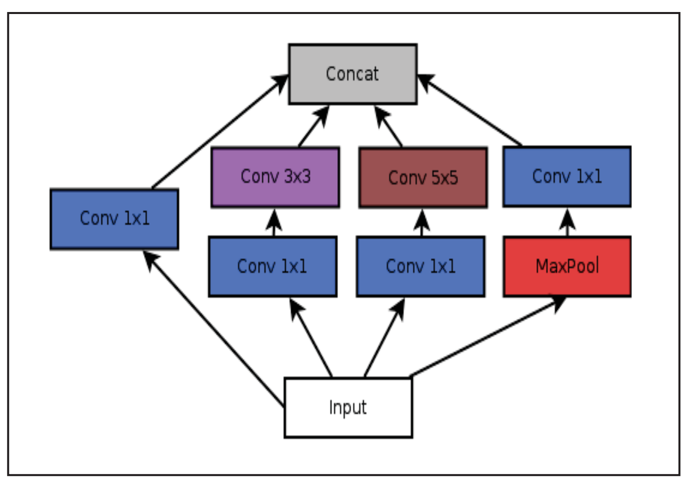

Figure 4. Inception Module.

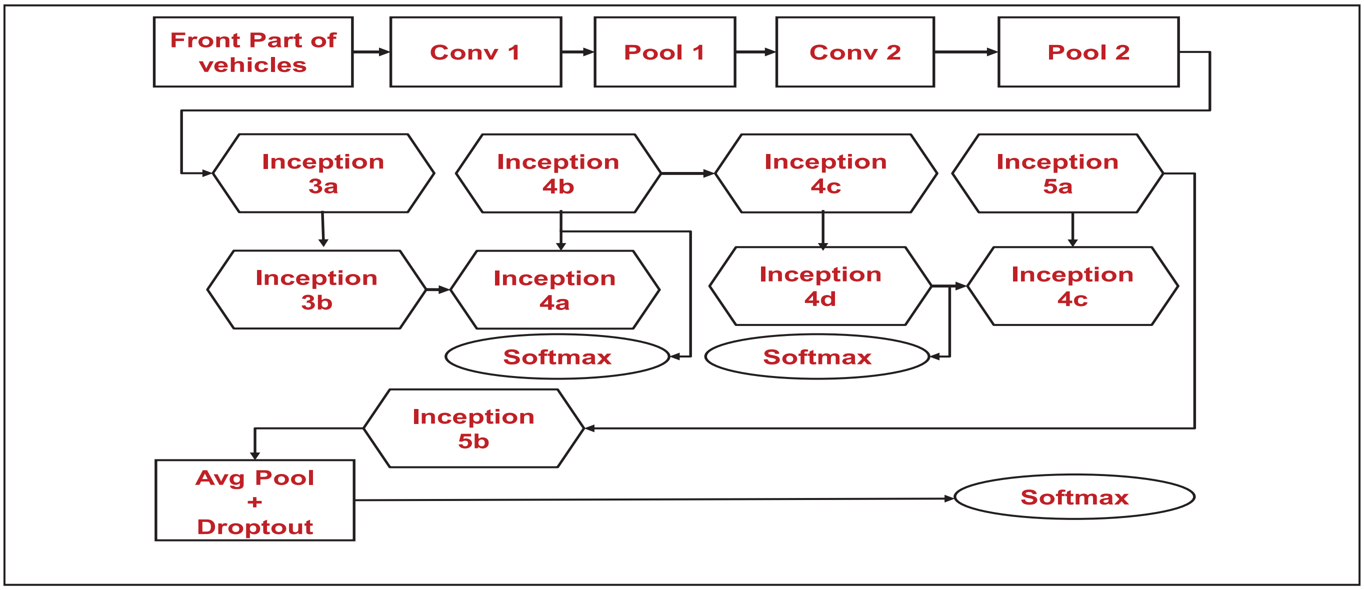

Figure 5. GoogleNet Architecture

2.3.2 Convolutional Layer

Convolutional layers consist of a rectangular grid of neurons. It requires that the previous layer also be a rectangular grid of neurons. Each neuron takes inputs from a rectangular section of the previous layer; the weights for this rectangular section are the same for each neuron in the convolutional layer. Thus, the convolutional layer is just an image convolution of the previous layer, where the weights specify the convolution filter. The first convolution layers will obtain the low-level features, like edges, lines and corners. The more layers the network has, the higher-level features it will get.

2.3.3 Max-pooling Layer

After each convolutional layer, there may be a pooling layer. The pooling layer takes small rectangular blocks from the convolutional layer and subsamples it to produce a single output from that block. There are several ways to do this pooling, such as taking the average or the maximum, or a learned linear combination of the neurons in the block.
The operation performed by this layer is also called 'down-sampling', as the reduction of size leads to loss of information as well.

2.3.4 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.

2.3.5 Softmax Layer

Fully connected layers convert input to output vectors and then the vector is fed into the softmax layer, the softmax function as the last layer computes the categorical probabilities.

3 Experimental Results

To evaluate the performance of our proposed framework, we built out vehicle database, which contain 291602 images from 766 image categorizations. These images were taken from real street CCTV. We are apt to use image instead of video, because it is easy to measure the accuracy. Our system is able to detect the frontal view of each vehicle image accurately. We performed the detection algorithm on all images in the system, and got 100% accuracy with respect to detection accuracy. Thus, our detection algorithm proved to be effective and fast in the desired system.

Once the frontal view of a vehicle was obtained, we fed it into trained deep model as input; the output label is used to recognize the vehicle model. In order to keep the environment constant for comparison we trained the deep architectures for 30 epoch. During the training the base learning rate was set to 0.01 and the training images and test images were same for both deep architectures.

In our experiments, images of 766 models were used for training and the left over 58314 images for testing which account for 20 percent of our entire dataset. The experimental results are shown in Table 1. The results show that our framework can achieve impressive performance compared with other methods.

Table 1

Methods	Top1 Accuracy	Top 5 Accuracy
Proposed	96.31	99.47
AlexNet	92.54	98.02

4 Conclusions

In this paper, we propose a framework for vehicle MMR that is based on CNN. We first detected moving vehicles using frame difference; the detected frontal view was used to identify a vehicle based on symmetrical filters. The frontal view of the vehicle images were first extracted and then fed into the deep network for training and testing. The results show that our framework achieved the best performance in comparison to other existing methods. Our framework outperforms the feature extraction based methods significantly. The Top 1 accuracy of proposed MMR algorithm is 96.31%, and the accuracy of our MMR algorithm shows 4.19% higher than other deep learning approaches.

Acknowledgment

This work was supported by the Brain Korea 21 PLUS Project. This research was also supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (GR 2016R1D1A3B03931911). This research was also supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2017-R0992-15-1023) supervised by the IITP(Institute for Information & communications Technology Promotion).

5 References

Uncategorized References

[1] C. Tzomakas and W. von Seelen, "Vehicle detection in traffic scenes using shadows," in *Ir-Ini, Institut fur Nueroinformatik, Ruhr-Universitat*, 1998: Citeseer.

[2] Z. Chen, T. Ellis, and S. A. Velastin, "Vehicle detection, tracking and classification in urban traffic," in *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on*, 2012, pp. 951-956: IEEE.

[3] L.-C. Chen, J.-W. Hsieh, Y. Yan, and D.-Y. Chen, "Vehicle make and model recognition using sparse representation and symmetrical SURFs," *Pattern Recognition,* vol. 48, no. 6, pp. 1979-1998, 2015.

[4] D. Llorca, D. Colás, I. Daza, I. Parra, and M. Sotelo, "Vehicle model recognition using geometry and appearance of car emblems from rear view images," in *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*, 2014, pp. 3094-3099: IEEE.

[5] M. Fraz, E. A. Edirisinghe, and M. S. Sarfraz, "Mid-level-representation based lexicon for vehicle make and model recognition," in *Pattern Recognition (ICPR), 2014 22nd International Conference on*, 2014, pp. 393-398: IEEE.

[6] 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 Transactions on Intelligent Transportation Systems,* vol. 15, no. 1, pp. 6-20, 2014.

[7] B. Zhang, "Reliable classification of vehicle types based on cascade classifier ensembles," *Ieee Transactions on Intelligent Transportation Systems,* vol. 14, no. 1, pp. 322-332, 2013.

[8] V. Varjas and A. Tanács, "Car recognition from frontal images in mobile environment," in *Image and Signal Processing and Analysis (ISPA), 2013 8th International Symposium on*, 2013, pp. 819-823: IEEE.

[9] J. Wu, X. Zhang, and J. Zhou, "Vehicle detection in static road images with PCA-and-wavelet-based classifier," in *Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE*, 2001, pp. 740-744: IEEE.

[10] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection using evolutionary Gabor filter optimization," *IEEE Transactions on Intelligent Transportation Systems,* vol. 6, no. 2, pp. 125-137, 2005.

[11] S. Sivaraman and M. M. Trivedi, "Real-time vehicle detection using parts at intersections," in *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on*, 2012, pp. 1519-1524: IEEE.

[12] M. Kafai and B. Bhanu, "Dynamic Bayesian networks for vehicle classification in video," *IEEE Transactions on Industrial Informatics,* vol. 8, no. 1, pp. 100-109, 2012.

[13] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE,* vol. 86, no. 11, pp. 2278-2324, 1998.

[14] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proceedings of the 26th annual international conference on machine learning*, 2009, pp. 609-616: ACM.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.

[16] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using

convolutional networks," *arXiv preprint arXiv:1312.6229*, 2013.

[17] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1-9.

[18] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size," *arXiv preprint arXiv:1602.07360*, 2016.

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

ISBN: 1-60132-464-2, CSREA Press ©