Athens Institute for Education and Research
ATINER

ATINER's Conference Paper Series
TRA2017-2275

CNN-based Vehicle Logo Classification for Vehicle Manufacturer Recognition

Mario Jimenez-Garcia
Research Assistant
University of Alcala
Spain

David Fernandez Llorca
Associate Professor
University of Alcala
Spain

An Introduction to ATINER's Conference Paper Series

ATINER started to publish this conference papers series in 2012. It includes only the papers submitted for publication after they were presented at one of the conferences organized by our Institute every year. This paper has been peer reviewed by at least two academic members of ATINER.

Dr. Gregory T. Papanikos
President
Athens Institute for Education and Research

This paper should be cited as follows:

Jimenez-Garcia, M. and Fernandez Llorca, D. (2017). **"CNN-based Vehicle Logo Classification for Vehicle Manufacturer Recognition",** Athens: ATINER'S Conference Paper Series, No: **TRA2017-2275.**

Athens Institute for Education and Research
8 Valaoritou Street, Kolonaki, 10671 Athens, Greece
Tel: + 30 210 3634210 Fax: + 30 210 3634209 Email: info@atiner.gr URL: www.atiner.gr
URL Conference Papers Series: www.atiner.gr/papers.htm
Printed in Athens, Greece by the Athens Institute for Education and Research.
ISSN: 2241-2891
02/10/2017

CNN-based Vehicle Logo Classification for Vehicle Manufacturer Recognition

Mario Jimenez-Garcia

David Fernandez Llorca

Abstract

Vehicle Manufacturer Recognition (VMR) has recently become a key feature to be used when license plate numbers cannot be detected or fake plate numbers are used in traffic control and management applications. One of the clearest indicators of a vehicle manufacturer is its vehicle logo. In this way, most of the vision-based VMR approaches are based on vehicle logo classification. In this work, we contribute to this topic by presenting an experimental comparison of different Convolutional Neural Networks (CNN) applied to vehicle logo recognition. A data set containing 26.289 logo images, corresponding to 30 car manufacturers and one non-logo class, is used to validate the proposed approach. The ability of the CNN to learn features robust to rotations is analyzed by artificially increasing the data set up to 446.913 adding rotated samples. The proposed methodology is also compared with a baseline classifier based on Histograms of Oriented Gradients (HOG) and Support Vector Machines (SVM). Experiments show an average accuracy of 99.87% in the validation set, which clearly demonstrate the potential of our CNN-based logo classifier to deal with VMR.

Keywords: Convolutional Neural Networks, SVM, Vehicle Logo Recognition, Vehicle Manufacturer Recognition.

Acknowledgments: This work was mainly supported by Research Grants DPI2014-59276-R (Spanish Ministry of Economy) and SEGVAUTO-TRIES-CM S2013/MIT-2713 (Community of Madrid).

Introduction

The use of License Plate Recognition (LPR) systems as the main tool to identify vehicles has been applied during more than two decades (Anagnostopoulos et al., 2008). However, there are two main drawbacks of using LPR as the unique identification method. First, they are unable to detect fake car plate numbers. Second, recognition errors cannot be automatically verified (manual inspection is needed). Accordingly, more sophisticated vehicle identification approaches are needed. The next step consists in providing the system with the ability of recognizing the car manufacturer. Vehicle Car Manufacturer (VMR) is a subject that has attracted the attention of the research community during the last years. Since the common approach to deal with VMR is to recognize the vehicle's logo, this problem is often named as Vehicle Logo Recognition (VLR). Note that one the car manufacturer has been detected, vehicle model recognition can be applied, considering all the potential models for the specific car make in some kind of hierarchical approach (Fernández-Llorca et al., 2014).

VLR is a challenging problem due to the nature of the images from which the logo has to be recognized. These images are usually captured by traffic cameras in outdoor environments, and that involves a high variance of both illumination conditions and logo pose. Accordingly, the method used to classify the logos of the vehicles has to be robust to these conditions. In this paper, we propose the use of a Convolutional Neural Network (CNN) to deal vehicle manufacturer recognition. The main contribution of the paper is the use of a new CNN-based architecture to successfully deal with logo recognition, including a comparison with other classification approaches (Fernández-Llorca et al., 2013) to explore the limits of the performance related with extreme rotations of the logos appearance. In addition, the number of classes of our dataset (31), and the number of samples (26.289), is considerably higher than the ones used in previous works. Finally, the ability of the CNN to learn features robust to rotations is analyzed by virtually increasing the data set up to 446.913, adding rotated samples. Experiments show an average accuracy of 98.33% in the validation set, which clearly demonstrate the potential of our CNN-based logo classifier to deal with VMR.

Literature Review

According to the literature, existing logo-based VMR systems can be divided into two main groups: hand-crafted features-based and CNN-based. Among the works contained in the first group we find approaches based on the use of Scale-Invariant Features Transform (SIFT) features (Psyllos et al., 2010), Chebyshev moments (Dai et al., 2009; Sam and Tian, 2012), Bag-of-Words (Yu et al., 2013), Histograms of Oriented Gradients (Fernández-Llorca et al., 2013) and Statistical Random Sparse Distribution (SRSD) (Peng et al., 2015). Since 2014, the use of CNNs became one of the main approaches used to deal with VLR due to its robustness to poor illumination conditions, logo viewpoint variation, noise, and to its lower dependence on precise logo detection. Thus, in (Thubsaeng et al., 2014), CNN was combined with a SVM. A CNN was used in a first stage to select candidate regions likely to be the manufacturer logo. Then, a second stage was applied where pyramid of histogram of orientation gradients and SVM were used to remove false logo regions. In (Huang et al., 2015) a two-layer CNN was proposed, where they included a pre-training strategy based on the use of principal components analysis, to improve the accuracy and the speed of the training. A multi-scale parallel CNN structure was proposed in (Zhang et al., 2016) where the multi-scale convolution kernel was used to extract the features from the original data in a parallel way. Finally, in (Xia et al., 2016) a combination of CNN with Multi-Task Learning was proposed, being capable of predicting up to nine binary attributes simultaneously (alphabet, symmetries, animal-like, etc.).

In order to have a better overview of the significance of the results of the different works, we present Table 1, where a summary of the state of the art methods is presented, containing the type of approach, the number of classes and number of samples of the dataset, the size of each sample and the average accuracy. As can be observed, our dataset is the richest one considering the number of samples and the number of classes. The average accuracy of our approach is at the same level of the best performances reported by state of the art methods.

Table 1. *Summary of the State of the Art*

Ref.	**Features / Classifier**	**# classes**	**# samples**	**size WxH**	**Avg. Acc. (*)**
(Dai et al., 2009)	Chebyshev moments / SVM	18	200	--	92.00%
(Psyllos et al., 2010)	SIFT / Feature Matching	10	1200	100x100	97.00%
(Sam and Tian, 2012)	Chebyshev moments / Adaboost	10	200	15x15	92.00%
(Yu et al., 2013)	Bag of Words /	14	840	Nx30	97.30%
(Fernández-Llorca et al., 2013)	HOG / SVM	27	3579	32x32	92.59%
(Peng et al., 2015)	SRSD / Nearest Neighbor	56	3370	50x50	97.21%
(Thubsaeng et al., 2014)	CNN + Pyramid HOG / SVM	20	7000	32x32	99.99%
(Huang et al., 2015)	CNN	10	11500	70x70	99.07%
(Zhang et al., 2016)	CNN	10	6600	38x38	98.8%
(Xia et al, 2016)	CNN	15	19780	64x64	98.14%
Ours	**CNN**	**31**	**26289**	**32x32**	**99.87%**

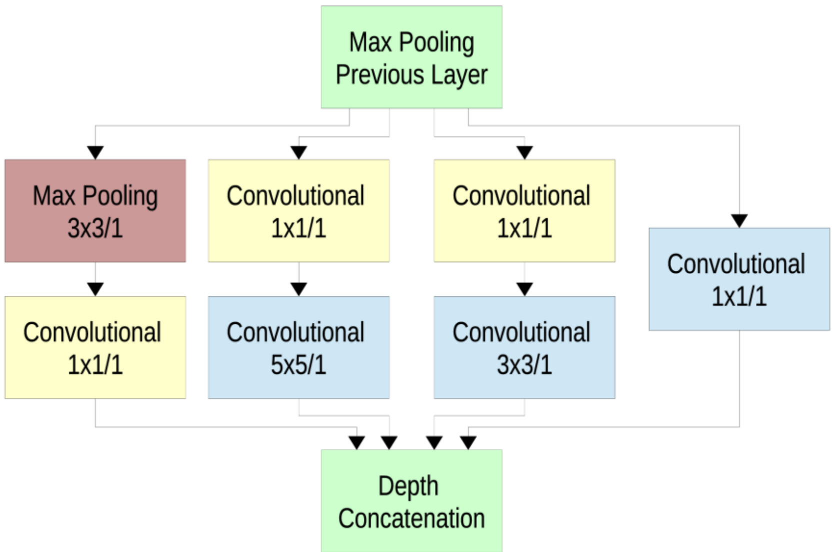
(*) Average accuracy in %.

Methodology

Previous works related with logo-based VMR using CNN has proved to be very effective in practice using standard architectures. For example, LeNet5-based (LeCun et al., 1998) architecture of only two convolutional and two subsampling layers were used in (Thubsaeng et al., 2014; Huang et al., 2015; Zhang et al., 2016), whereas three convolutional and three sub-sampling layers were proposed in (Xia et al., 2016). In our case, we propose the use of a more complex architecture with a higher number of parameters taking into account the fact that our dataset contains a much larger number of samples than previous approaches.

A simplified version of the GoogleLeNet (Szegedy et al., 2015) is proposed as the main recognizer module of vehicle logos (SGoogleLNet). Compared to other complex CNN architectures such as AlexNet (Krizhevsky et al., 2012) or VGG (Simonyan and Zisserman, 2015), GoogleLeNet maintains an extremely low number of operations. It is based on two main ideas: the approximation of a sparse structure with spatially repeated dense components and using dimension reduction to keep the computational complexity in bounds but only when required. The optimal sparse structure is approximated with dense components using 1x1, 3x3 and 5x5 filters, and a pooling layer in the so-called *inception* module, adding 1x1 convolutions for dimensionality reduction (see Figure 1).

Figure 1. *Inception Module with Dimension Reductions (GoogleLeNet)*



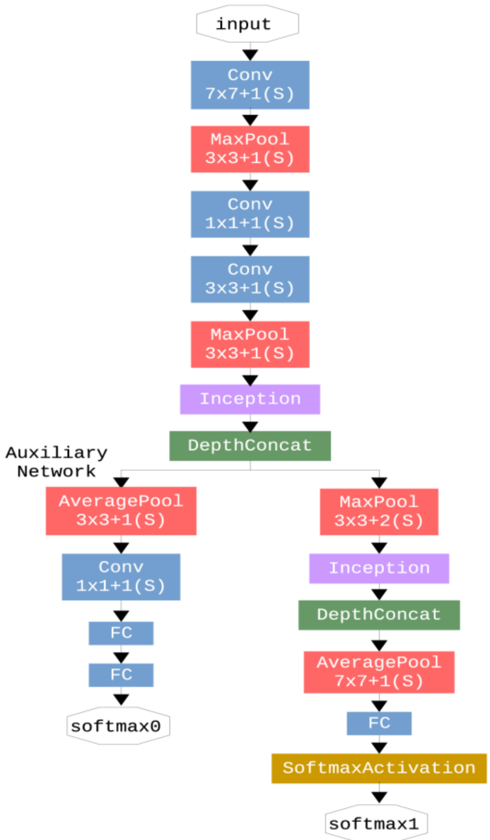
The original GoogleLeNet network is 27 layers deep counting both layers with parameters and pooling ones, from which 18 layers corresponds to 9 inception modules. In addition, the architecture introduces two auxiliary classifiers connected in the middle of the network to enhance the gradient signal that gets propagated back and provides additional regularization. These auxiliary networks are totally discarded when performing inference. In our case, we have simplified this network including two inception modules and only one auxiliary network.

We have maintained the depth of most of the original layers whereas the stride and padding parameters have been adapted to maintain the coherence with respect to the size of the input images. All the parameters of the different layers of our architecture can be seen in Table 2. In addition, a schematic view of our resulting network is depicted in Figure 2.

Table 2. *Simplified GoogleLeNet Architecture (SGoogleLeNet)*

Type	Patch size / stride / padding	Output size
input	--	32 x 32 x 1
convolution	7 x 7 / 1 / 3	32 x 32 x 64
max pooling	3 x 3 / 1	30 x 30 x 64
convolution	1 x 1 / 1 / 0	30 x 30 x 64
convolution	3 x 3 / 1 / 1	30 x 30 x 192
inception	--	30 x 30 x 256
max pooling	3 x 3 / 1	28 x 28 x 256
inception	--	28 x 28 x 512
average pooling	7 x 7 / 1	22 x 22 x 512
linear	--	1 x 1 x 31
softmax	--	1 x 1 x 31

Figure 2. *Proposed Network Architecture Including Two Inception Modules and One Auxiliary Network (SGoogleLNet)*



Results

Dataset Description

In order to validate the proposed network architecture, we have created a dataset that contains vehicle logo images taken from different locations, using different types of cameras, in different lighting conditions, and with different image sizes. This dataset can be considered as one of the contributions of the current work, since up to now, it is the dataset that contains a larger number of samples including one of the largest set of different car manufacturers, as can be observed in Table 1. In Table 3, some details of our dataset are provided, including the number of samples for each car manufacturer, and some sample images. Note that the number of different vehicle logos that are contained in the dataset is 30 plus one class corresponding to no logo images. This is a key class that should contain representative samples of different parts of the vehicles fronts.

Table 3. *Dataset Description: Car Manufacturers, Number of Samples, and Sample Images*

car make	# samples	examples	car make	# samples	examples
Alfa Romeo	98		Mitsubishi	616	
Audi	960		Nissan	489	
BMW	613		Opel	829	
Chevrolet	313		Peugeot	957	
Citroen	2470		Renault	2574	
Dacia	230		Saab	142	
Fiat	531		Seat	1113	
Ford	985		Skoda	315	
Honda	836		SsangYong	296	
Hyundai	363		Subaru	21	
Kia	354		Suzuki	200	
Lancia	94		Toyota	1260	
Lexus	165		Volkswagen	2288	
Mazda	548		Volvo	561	
Mercedes	1612		None	4336	
Mini	120		TOTAL	26289	--

As can be observed in the provided sample images of Table 3, the logos do not always appear perfectly centered and bounded. The final appearance of the logo will depend on the method used to select the regions of interest (potential locations of the logo in the image) and the camera position with respect to the vehicles position. Accordingly, logos can appear considerably titled in the image, as can be observed in examples of Figure 3.

Although the exact localization of the license plate can help to work with canonical representations of the vehicle (Fernández-Llorca et al., 2014), it involves the use of a projective transformation that may also distort the final shape and appearance of the logo. Instead of trying to compensate rotations of the logo using computer vision techniques, we can design the logo recognition engine to deal with rotated versions of each logo.

Figure 3. *Examples from Different Systems where Logos Appear Considerably Rotated*

In order to test the robustness of the CNN to rotations, we artificially increase the data set up to 446.913 samples by adding 16 different rotations to each sample contained in the original dataset. Three different datasets have been created, as showed in Table 4.

Table 3. *Datasets Including Artificially Rotated Images (446.913 Samples Each)*

Name	Range of angles	Step
Dataset-R4-S0.5	[-4º, +4º]	0.5º
Dataset-R20-S2.5	[-20º, +20º]	2.5º
Dataset-R45-S5.625	[-45º,+45º]	5.625º

Classification Results and Discussion

Each dataset is divided into training (60%) and validation (40%) sets, making sure that no samples of the same vehicle are contained in both sets. We compare the current CNN-based approach with a hand-crafted features-based one (Fernández-Llorca et al., 2013) based on the use of Histograms of Oriented Gradients (HOG) and Support Vector Machine (SVM). We measure the accuracy of both methods using the three different datasets described in Table 4. The overall accuracy of both methods applied to each dataset is depicted in Table 5.

Table 4. *Overall Accuracy for Both CNN- and HOG-SVM-based (Fernández-Llorca et al., 2013) Methods on the Three Datasets*

Dataset	HOG-SVM (Fernández-Llorca et al., 2013)	SGoogleLNet
Dataset-R4-S0.5	98.33%	99.87%
Dataset-R20-S2.5	95.08%	99.35%
Dataset-R45-S5.625	88.78%	98.82%

In Figures 4-7 we depict the color-coded confusion matrices for each method, for Dataset-R4-S0.5 and Dataset-R45-S5.625.

Figure 4. *Color-coded Confusion Matrix of the HOG-SVM Approach in Dataset-R4-S0.5*

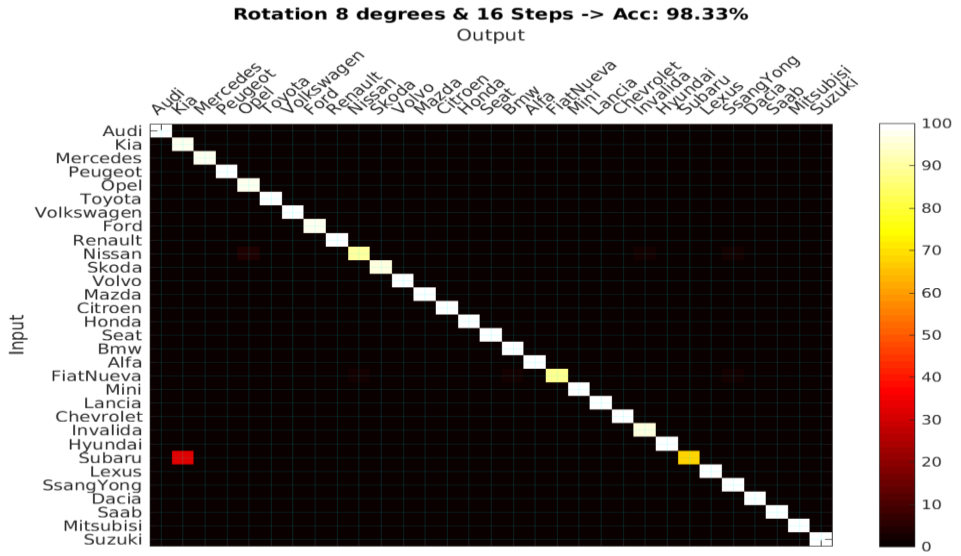


Figure 5. *Color-coded Confusion Matrix of the HOG-SVM Approach in Dataset-R45-S5.625*

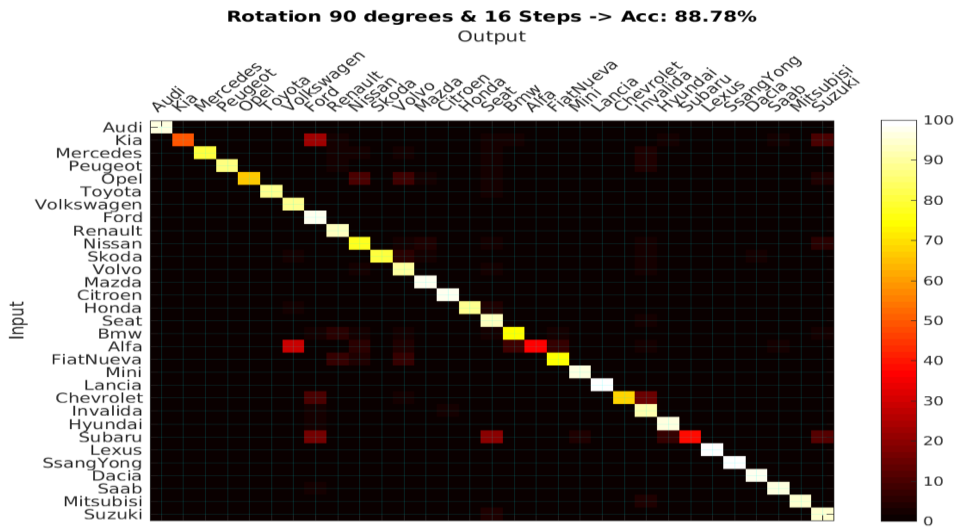


Figure 6. *Color-coded Confusion Matrix of the SGoogleLeNet Approach in Dataset-R4-S0.5*

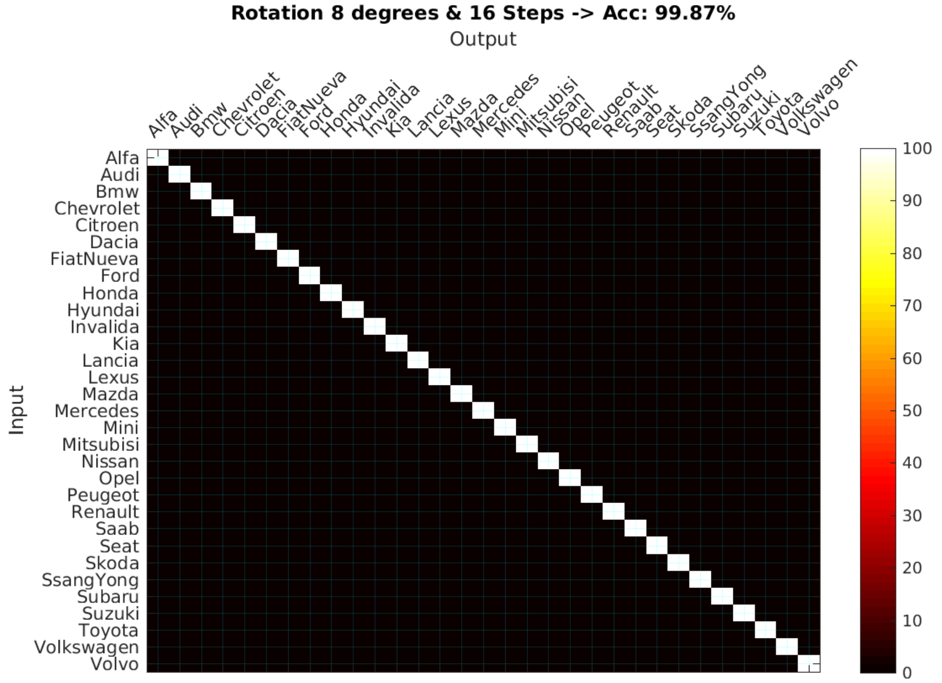
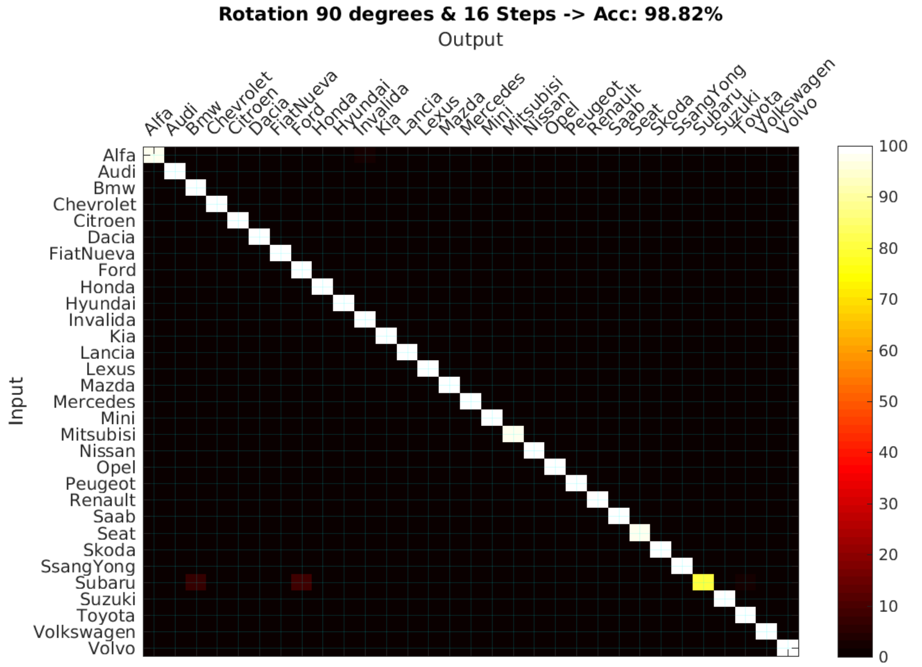


Figure 7. *Color-coded Confusion Matrix of the SGoogleLeNet Approach in Dataset-R45-S5.625*



As can be observed in Table 5 and Figures 4 and 5, the accuracy of the HOG-SVM method decreases proportionally with the range of rotation angles. The overall accuracy goes from 98.33% for rotations of up to ±4º to 88.78% for rotations of up to ±45º. Indeed, the gradient distribution gets mostly blurred when large rotations are allowed per each class. However, as depicted in Table 5 and Figures 6 and 7, the performance of our CNN architecture is robust to large rotations, obtaining an overall accuracy of 98.82% for rotations of up to ±45º. When small rotations are expected, the simplified GoogleLeNet architecture provides an overall accuracy of 99.87%.

Conclusions

In this paper, we have contributed to the topic of Vehicle Manufacturer Recognition (VMR) by means of vehicle logo classification, by proposing a new CNN architecture based on a simplified version of the GoogleLeNet (SGoogleLeNet). In addition, we have presented a new dataset that contains 26.289 logo images corresponding to 30 car manufactures plus one non-logo class. The ability of the SGoogleLeNet to learn features robust to rotations is analyzed by artificially increasing the data set up to 446.913 adding rotated samples up to angles of ±45º. Contrary to hand-crafted HOG-SVM-based approach, where the accuracy clearly decreases for large rotations (almost a 10% of accuracy reduction for rotations up 45º), the proposed CNN-based method provides a robust accuracy that only decreases a 1% for cases where the logo can appear rotated up to 45º. When the expected logo rotations are lower than ±4º, the proposed SGoogleLeNet architecture provides an average accuracy of 99.87%, which clearly validates and supports the use of this architecture to deal with vehicle logo recognition. Future works will be related with the integration of the proposed classifier with a LPR system and a region of interest selection mechanism to deal with VMR from traffic images.

References

Anagnostopoulos, N. E., Anagnostopoulos, I. E., Psoroulas, I. D., Loumos, V. and Kayafas, E. “License plate recognition from still images and video sequences: A survey,”, *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, pp. 377-391, 2008.

Dai, S., He, H., Gao, Z., Li, K. and Xiao, S. “Vehicle-logo Recognition Method Based on Tchebichef Moment Invariants and SVM,” In Proceedings of the *World Congress on Software Engineering,* 2009.

Fernández-Llorca, D., Arroyo, R. and Sotelo, M. A. “Vehicle logo recognition in traffic images using HOG and SVM”. In Proceedings of the *IEEE 16th International Conference on Intelligent Transportation Systems*, 2013.

Fernández-Llorca, D., Colás, D., Daza, I. G., Parra, I. And Sotelo, M. A. “Vehicle model recognition using geometry and appearance of car emblems from rear view images”, In Proceedings of the *IEEE 17th International Conference on Intelligent Transportation Systems*, 2014.

Huang, Y., Wu, R., Sun Y., Wang, W. and Ding, X. “Vehicle Logo Recognition System Based on Convolutional Neural Networks With a Pretraining Strategy”, *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, 2015.

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

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. "Gradient-based learning applied to document recognition". In Proceedings of the IEEE, 86(11), 2278-2324. 1998.

Peng, H., Wang, X., Wang, H. and Yang, W. "Recognition of Low-Resolution Logos in Vehicle Images Based on Statistical Random Sparse Distribution", *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, 2015.

Psyllos, A. P., Anagnostopoulos, C-N. E. and Kayafas, E. "Vehicle Logo Recognition Using a SIFT-based Enhanced Matching Scheme", *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, 2010.

Sam, K. T. and Tian, X. L. "Vehicle Logo Recognition Using Modest Adaboost and Radial Tchebichef Moments," In Proceedings of the *4th International Conference on Machine Learning and Computing*., 2012.

Simonyan, K., Zisserman, A. "Very Deep Convolutional Networks for Large-Scale Image Recognition", In Proceedings of the Computer Vision and Patter Recognition Conference, 2015.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. "Going Deeper with Convolutions", In Proceedings of the Computer Vision and Patter Recognition Conference, 2015.

Thubsaeng, W., Kawewong, A. and Patanukhom, K. "Vehicle Logo Detection Using Convolutional Neural Network and Pyramid of Histograms of Oriented Gradients", In Proceedings of the *11th International Joint Conference on Computer Science and Software Engineering*, 2014.

Xia, Y., Feng, J. and Zhang, B. "Vehicle Logo Recognition and Attributes Prediction by Multi-task Learning with CNN", In Proceedings of the *12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*, 2016.

Yu, S., Zheng, S., Yang, H. and Liang, L., "Vehicle Logo Recognition Based on Bag-of-Words," In Proceedings of the *10th IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2013.

Zhang, S-W., Zhang, Y-h., Yang, J. and Li S-b. "Vehicle-logo Recognition Based on Convolutional Neural Network with Multi-scale Parallell Layers". In Proceedings of the *International Conference on Computer, Mechatronics and Electronic Engineering*, 2016.