A Large and Diverse Dataset for Improved Vehicle Make and Model Recognition

Faezeh Tafazzoli, Hichem Frigui
Multimedia Research Lab
University of Louisville, KY
{faezeh.tafazzoli, h.frigui}@louisville.edu

Keishin Nishiyama
Cyber Security Lab
University of Louisville, Louisville, KY
keishin.nishiyama@louisville.edu

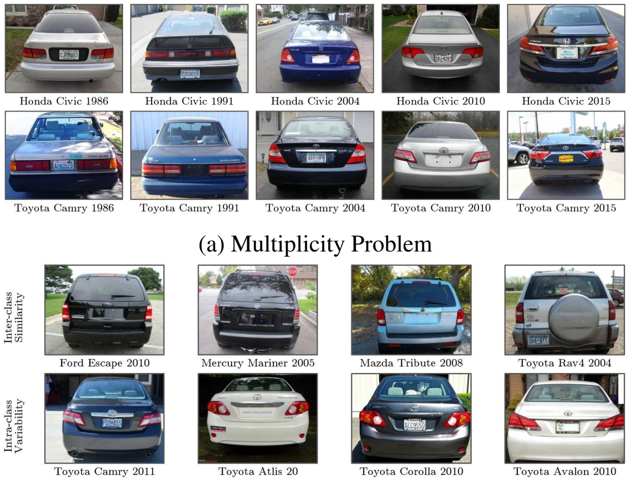
Abstract

Vehicle Make and Model Recognition (VMMR) has evolved into a significant subject of study due to its importance in numerous Intelligent Transportation Systems (ITS) and corresponding components such as Automated Vehicular Surveillance (AVS). A highly accurate and real-time VMMR system significantly reduces the overhead cost of resources otherwise required. The VMMR problem is a multi-class classification task with a peculiar set of issues and challenges like multiplicity, inter- and intra-make ambiguity among various vehicle makes and models, which need to be solved in an efficient and reliable manner to achieve a highly robust VMMR system.

In this paper, facing the growing importance of make and model recognition of vehicles, we present an image dataset[1] *with* $9,170$ *different classes of vehicles to advance the corresponding tasks. Extensive experiments conducted using baseline approaches yield superior results for images that were occluded, under low illumination, partial or non-frontal camera views, available in our VMMR dataset. The approaches presented herewith provide a robust VMMR system for applications in realistic environments.*

1. Introduction

Over the recent years, a deluge of innovative technologies and solutions are bringing Intelligent Transportation Systems (ITS) closer to reality. Identification and classification of vehicles is of great interest in such applications, due to elevated security concerns in ITS and demanding areas such as targeted advertisement, behavior analysis and understanding or surveillance for crime prevention and safety. Vehicles offer several unique properties compared to other objects. They provide a more diverse and challenging set of issues and facilitate a range of novel research topics in fine-grained image classification. The first set of issues stems from one vehicle model of the same make having different shapes and/or appearances in different years (Figure 1(a)), referred to as multiplicity. The second category of problems, called ambiguity, can be further classified into two types: (a) Inter-class similarity, and (b) Intra-class variability. The former ambiguity refers to the issue of vehicles of different manufactures having visually similar shape or appearance, i.e., two different make-model classes have similar front or rear views. The latter kind of ambiguity is a result of similarity between different models of the same make. Samples of such differences are depicted in Figure 1(b). Additionally, the considerably large number of car models, including different car manufactures and models depending on the year has made VMMR one of the most challenging fine-grained classification problems. This application, thus, can potentially foster more sophisticated computer vision models and algorithms.



(a) Multiplicity Problem

(b) Ambiguity Problem

Figure 1: VMMR Challenges

Traditional vehicle identification systems recognize makes and models of vehicles relying on manual human observations or automated license plate recognition (ALPR)

[1] The latest version of dataset will be available upon request by contacting the corresponding author.

systems [2, 7, 30]. Both approaches are failure-prone and have several limitations. It is practically difficult for human observers to remember and efficiently distinguish between the wide variety of vehicle makes and models. On the other hand, the AVS systems that rely on license plate suffer from several limitations. First, many surveillance cameras are not installed for license plate capturing, thus, plate recognition performance drops dramatically on images/video data captured by these cameras. Furthermore, license plates are easy to be forged, damaged, modified, occluded, or invisible due to uneven lighting conditions. Moreover, in some areas, it may not be required to have the license plate at the front or rear of vehicle. This could lead to retrieving the wrong information regarding make or model of the vehicle from the registry [31].

To overcome the above shortcomings in traditional vehicle identification and classification systems, the make and model of the vehicle recognized by the VMMR system can complement the ALPR systems by providing a higher level of robustness against fraudulent use of license plates or poor image quality and consequently further enhance security. Furthermore, in applications such as electronic toll collection, vision-based AVS systems could serve as a complementary tool in improving efficiency of existing systems to apply different rates to different types of vehicles inexpensively and automatically. In traffic control or traffic monitoring, statistics of vehicle flow, associated with vehicle models, is more helpful in an intelligent transportation system.

In this paper, we present a new dataset covering most of the existing vehicle makes and models to help experiments in this direction by providing sufficient amount of data enriched by information automatically extracted to define each vehicle's make, model and production year.

2. Background

The task of VMMR is the most advanced use case of cars understanding, with high sensitivity on details, environment changes, rapid variations in manufacturer production and maintenance. However, the amount of relevant scientific literature is relatively small. Generally, existing work falls into three major approaches.

The first one is based on appearance and identifies cars by their inherent features including dimensions, shapes, and textures. These methods rely on the pose and position of the cameras [11, 19, 28]. The second approach is feature-based and classifies car models using local or global invariant features including low-level features such as edge-based [1, 24] and contourlet transform features [5] as well as high-level features, such as, Speeded Up Robust Features (SURF) [3, 4], Histogram of Gradient (HoG) [16], PHOG and Gabor features [33]. The third approach is model-based and follows the intuition that distinctive features of a fine-grained category are most naturally represented in 3D object space, representing both the appearance of the parts and their location with respect to the target object [13, 15, 17].

Many of the above methods have good performance when dealing with the classification of a few number of vehicle makes or models; but their performance usually cannot meet the requirements in realistic applications, dealing with a large number of classes [19]. Moreover, most of these works rely heavily on hand-crafted low-level features which might not be saliently distinctive among different subordinate-level categories that have extremely similar appearance. To address the fine-grained recognition problem more specifically, recently deep networks are being used to extract discriminative hierarchical features from the data [9, 10, 29].

2.1. Existing Datasets

Most research efforts on VMMR so far have been focused on medium-scale datasets. There are two main reasons for the limited effort on large-scale image based vehicle classification. First, there are only a few publicly available large-scale benchmark datasets for VMMR. This is mostly because class labels are expensive to obtain. In fact, most existing fine-grained image classification benchmark datasets only consist of a few thousands (or less) of training images. As of existing vehicle datasets, they either cover a subset of makes and models [24,32], or only categorize vehicles at a high level (i.e., SUV, Truck, Sedan) [6,20], and those usable mainly for vehicle related tasks such as detection and pose estimation [8, 17, 22]. Second, large-scale classification is difficult because it poses more challenges than its medium-scale counterparts. Having the appropriate set of training data can improve the performance of designed classifiers. Indeed, it is necessary to have a very large number of images for each class to cover the wide range of variations of view angles, lighting, as well as the fairly wild appearance difference within the same class.

The lack of public and standard datasets has moved researchers to use their own databases. Accordingly, it is very complicated to establish a performance comparison between the different approaches. A very recently published example is CompCars dataset [31]. This dataset consists of web-nature and surveillance-nature parts. The former is made of $136,727$ vehicles from 153 car makes with $1,716$ car models, taken from different viewpoints, covering many commercial car models in the recent ten years, most of which are Chinese, and the latter contains $44,481$ frontal images of vehicles taken from surveillance cameras. The CompCars dataset was originally used for fine-grained car classification, car attribute prediction and car verification. Sochor et al. collected and annotated the BoxCars dataset [29] containing vehicle images taken from surveillance cameras accompanied with their 3D bounding boxes. This dataset is composed of $21,250$ vehicles ($63,750$ images in diverse viewpoints) of 27 different makes, 102 make-model classes, 126 make-model-submodel classes, and 148 make-model-submodel-year classes. Lin et al. [17] published FG3DCar dataset including 300 images of 30 classes. The data provided in FGComp [15] includes $8,144$ images of cars, covering only 196 makes and models out of which 60% are 2012 models. A vehicle re-identification dataset, VehicleID, collected from multiple real-world surveillance cameras and including over $200,000$ images of about $26,000$ vehicles, was introduced in [18]. Almost $90,000$ images of $10,319$ vehicles in this dataset have been labeled with the vehicle model information. A summary of the existing datasets addressing the task of VMMR is featured in Table 1.

Table 1: Summary of existing VMMR datasets

Ref. Year	Viewpoint	# Samples	# Classes
2004 [25]	Front	1132	77
2005 [23]	Front	180	5
2008 [5]	Front	830	50
2009 [26]	3D free	400	36
2011 [27]	Front	90	10
2011 [24]	Front	262	74
2012 [1]	Rear	400	10
2013 [15]	Mixed	16185	196
2014 [13]	3D free	190	8
2014 [14]	Front	6936	29
2014 [19]	Rear	1342	52
2015 [31]	Mixed	136727	1716
2016 [29]	3D free	63750	126

3. Proposed VMMR Dataset

Despite the ongoing research and practical interests, car make and model analysis attracts limited attention in the computer vision community, due to the aforementioned diversity and limitations of existing datasets. Thus, we collected a comprehensive dataset, VMMRdb, where each image is labeled with the corresponding make, model and production year of the vehicle.

The dataset used in our experiments contains images that were taken by different users, different imaging devices, and multiple view angles, ensuring a wide range of variations to account for various scenarios that could be encountered during testing, in a real-life scenario. The cars are not well aligned, and some images contain irrelevant background. The data was gathered by crawling web pages related to vehicle sales, mainly on *craigslist.com* and *amazon.com*, including 712 areas covering all 412 sub-domains corresponding to U.S. metro areas. Images were automatically annotated using the title and description the sellers had provided for each post. We developed a semi-automated process to prune the data and remove the undesired images belonging to interior parts of vehicles and noisy labels.

The VMMR dataset is much larger in scale and diversity compared with the existing car image datasets, containing **9,170** classes consisting of **291,752** images, covering models manufactured between 1950 to 2016. The distribution of images in different classes of the dataset is illustrated in Figure 2. Each circle is associated with a class, and its size represents the number of images in the class. The classes with labels are the ones including more that 100 images.

The dataset will be publicly available on our website in a few months. In the meantime, the dataset's latest version will be available upon request by contacting the corresponding author.

4. Impact of the Proposed VMMR Dataset

We compared the effect of multiple network architectures, including the ones evaluated in [31]. Here, we only report results using ResNet [12] as it outperforms other models. Considering the superior validation accuracy achieved by the model with 50 layers, referred to as ResNet-50, with respect to having less parameters, we choose this model for the evaluations. We used the model pre-trained on ImageNet, and fine-tuned on the datasets under experiment with the same mini-batch size, number of epochs, and learning rate.

In the first experiment, we analyze the importance of having diversity in data to handle real-world surveillance applications. We commence our evaluations by comparing the performance of our dataset with the CompCars dataset employing the same settings as the ones used in [31], using CNN learners. Despite having hierarchical labels of make, model and year in their dataset, Yang et al. [31] have merged all production years of each model to the same class in their experiments. This has resulted in 431 classes, many of which are Chinese manufacturers. To have a proper comparison, we choose only those classes existing in our dataset (125 classes) and following their approach, we use the labels at the make-model level only. We pick the exact year for which any image is included in CompCars. In the resulting 51 classes, with corresponding datasets referred to as CompCars-51 and VMMRdb-51, following the experimental settings in [31], we divide the images into two halves for training and testing. Table 2 details the number of images in each dataset.

Table 2: Specifications of the overlap data between CompCars and VMMRdb datasets

Dataset	# Classes	# Train	# Test
CompCars-51	51	1527	1506
VMMRdb-51	51	1986	1984

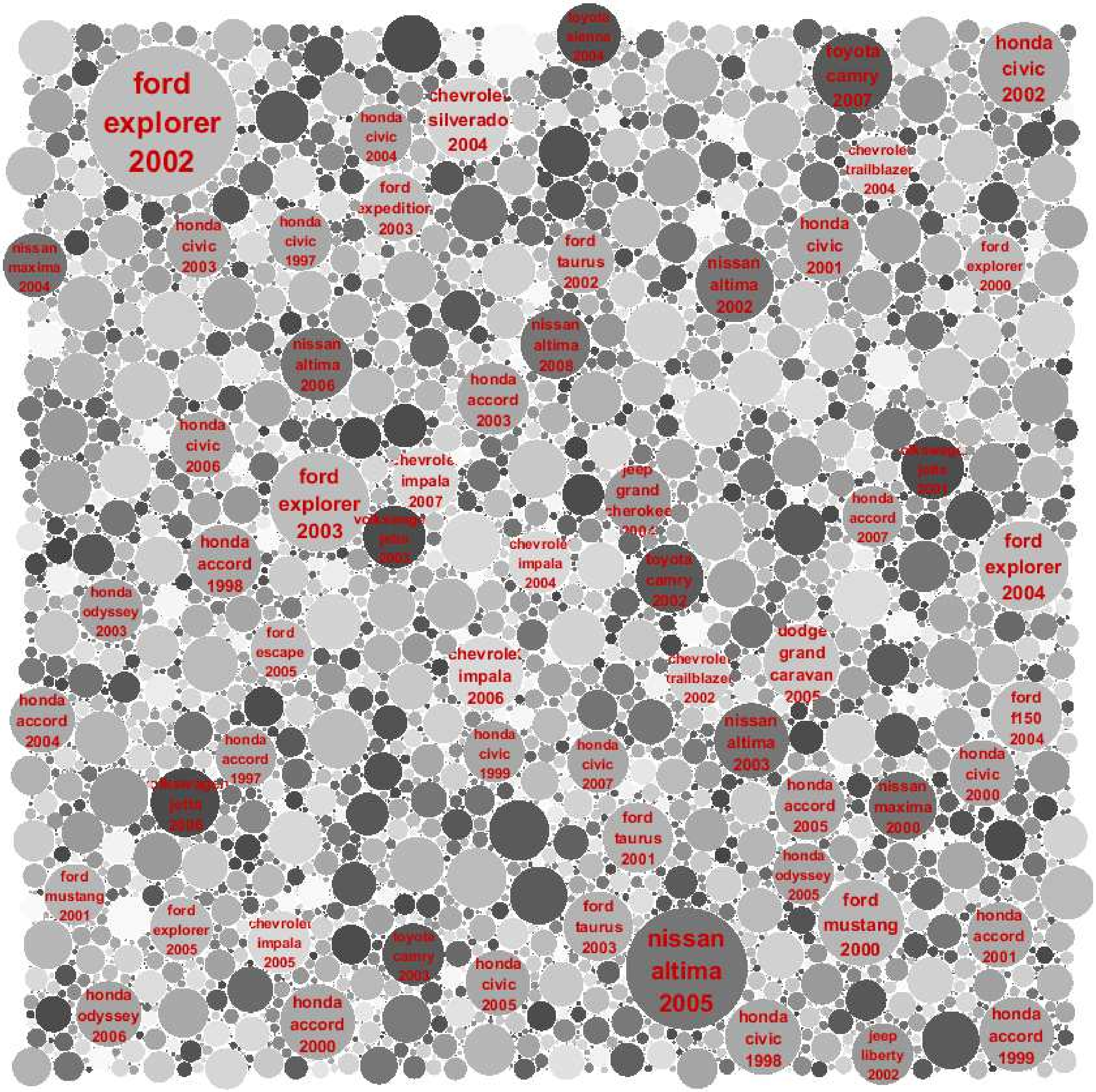


Figure 2: Distribution of images per class in VMMRdb

We investigated the classification accuracy of networks trained on each dataset in confronting with samples from other datasets. The nature of images provided in CompCars are very different from VMMRdb images in the sense that they are mostly captured in more controlled environment with much higher resolution. The purpose of these experiments is to see how well the model performs given images collected in more challenging scenarios. We, also, generate a third dataset which we refer to as CompCarsVMMRdb-51, by merging the discussed datasets. The performances of these experiments are summarized in Table 3. We report the Top-1 and Top-3 accuracies of car make-model classification, which denote the classification accuracy considering the first and up to three top matches, respectively, for each pair of train and test set.

As we expected, the model trained on CompCars, despite its significant performance on the test set with images of similar nature, degrades considerably on the test images selected from VMMRdb-51. The performance of the model trained on VMMRdb-51 is just slightly better with respect to non-VMMR images. The merged dataset, however, outperforms both previous cases, proving the fact that employing additional training data can boost classification results by increasing data diversity in training examples.

Table 3: Classification results for the models trained on different datasets

Train	**Test** CompCars-51	 VMMRdb-51	 CompCars VMMRdb-51	
CompCars -51	96.88	36.10	62.23	*Top-1*
	97.88	50.05	70.69	*Top-3*
VMMRdb -51	40.28	90.26	68.22	*Top-1*
	52.85	93.48	75.93	*Top-3*
CompCars VMMRdb-51	96.61	94.10	95.16	*Top-1*
	97.48	96.47	96.91	*Top-3*

5. Fine-grained VMMR

We extend the CNN-based experiments to train a model for classification of vehicles make, model and production year on another subset of our dataset which we will refer to as VMMRdb-3036. In order to have enough data for training CNN, in this dataset, we have only considered the classes containing more than 20 images. It has 3036 classes and $246,173$ images. In each class we split the images into 70% and 30% for train and test, respectively. Figure 3 depicts the distribution of different classes based on the number of images. The distribution of images in the sub-classes of a selected make is visualized in Figure 4.

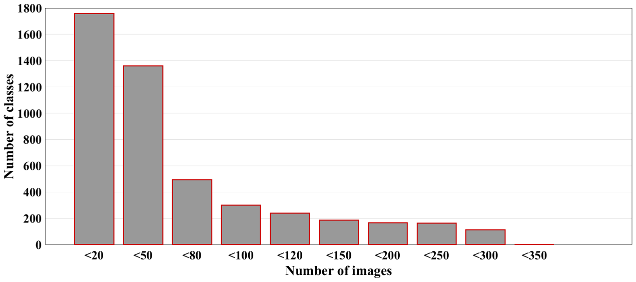


Figure 3: Distribution of number of images per class in VMMRdb-3036

Following the previous experiments, we used the model pre-trained on ImageNet, and fine-tuned it on VMMRdb-3036. We set the parameters to initial learning rate 0.01, and 200 training epochs. The learning rate decay was selectively applied after initial 30 epochs. In training, all inputs were color-normalized with the mean and standard deviation from ImageNet images after scale, aspect ratio, color, and horizontal flip augmentations. During the test phase, all detections were center-cropped and color-normalized by the system. We trained the models with a minibatch size of 32 within 110 hours on a NVIDIA GeForce GTX 1080 GPU using ~8 Gb of memory.

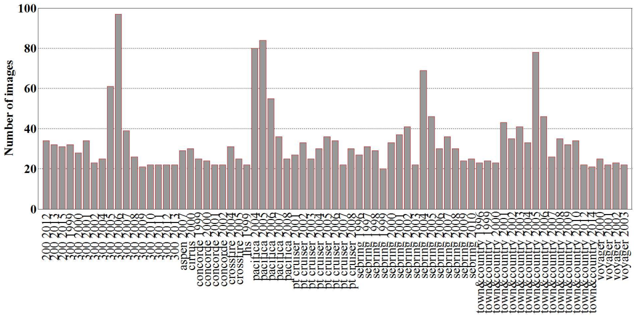


Figure 4: Distribution of number of images per model-year in VMMRdb-3036 for the sample make 'Chrysler'

The model trained on VMMRdb-3036 achieved the Top-1 and Top-5 accuracy of 51.76% and 92.90%, respectively. This decrease in accuracy compared to the results of Table 3, in addition to the considerable increase in the number of classes, illustrates the level of difficulty by going deeper in the hierarchy of fine-grained classification where we have included manufacture year as well as make and model.

Figure 5 displays some predictions, indicating that the model accounts for variations in viewpoints and lighting conditions. Below each image is the ground truth class and the probabilities for the Top-5 predictions with the matched class in the top bar. As we can see, the Top-1 match usually has a high confidence versus the rest of predictions. In these examples, the prediction matches the target make, model and production year. In Figure 6, however, few examples of images predicted to the true make and model, but incorrect year are displayed. These samples represent the multiplicity problem. The left example, shows an image of "Toyota Camry 2009" which has been matched to "Toyota Camry 2010". The interesting point is that in many of these samples, we are dealing with images which are partially occluded or have an uncommon viewpoint. A few samples of images incorrectly classified to a different model or make, representing the ambiguity problem, are illustrated in Figure 7.

To observe the learned feature space of the models with the aforementioned challenges, 2D projection of the last fully connected layer for sample classes are visualized in Figure 8. We have employed t-Distributed Stochastic Neighbor Embedding (t-SNE) [21] for the projection of randomly selected images from the presented classes. The classes have been selected to represent the different chal-

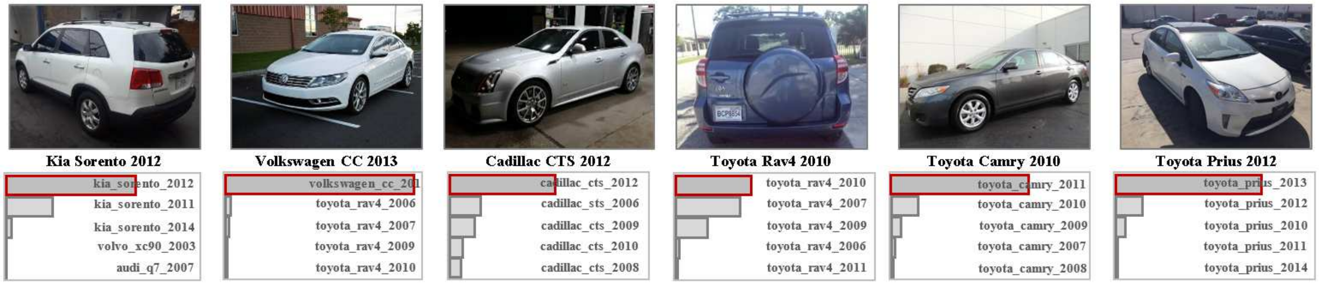


Figure 5: Top-5 predicted classes of ResNet-50 for sample images from VMMRdb-3036 classified correctly by the model

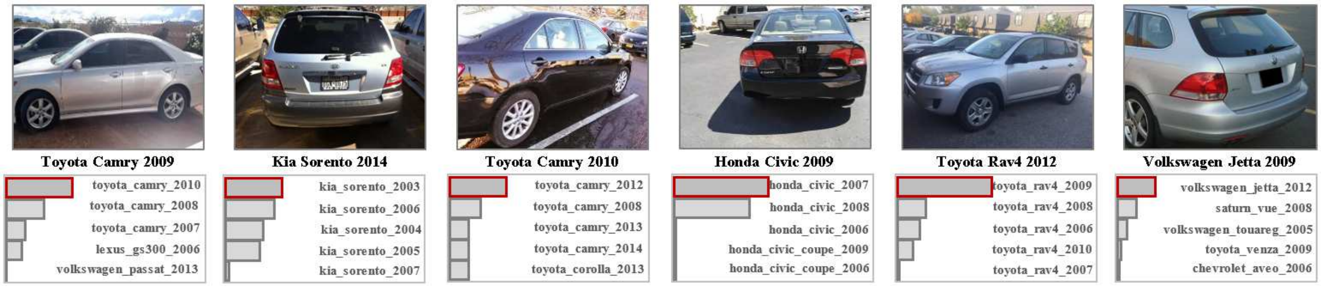


Figure 6: Top-5 predicted classes of ResNet-50 for sample images from VMMRdb-3036, incorrectly classified due to the multiplicity issue

lenges of VMMR depicted in Figure 1. We can see that features from the same model are closer to each other compared to the ones visually very different.

6. Conclusions and Future Work

In this paper we presented a very large dataset to address the problem of fine-grained classification of vehicles in hierarchies of make, model and manufacture year. For fine-grained recognition tasks, specifically, the challenge is in discovering and locating the regions that contain the discriminative details of each class. We make VMMRdb publicly available for future reference and benchmarking. Because of the natural environments and unconstrained image settings, our dataset can be used as a baseline for training a robust model in several real-life scenarios.

Our dataset can offer valuable situational information for law enforcement units in a variety of civil infrastructures. To demonstrate the effectiveness of our proposed approaches for VMMR, in our future work, we target an important real-life surveillance application where our system would be able to analyze video data acquired from multiple surveillance cameras to monitor and track vehicles under varying environment and capture conditions.

References

[1] M. AbdelMaseeh, I. Badreldin, M. F. Abdelkader, and M. El Saban. Car make and model recognition combining global and local cues. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 910–913. IEEE, 2012.

[2] C.-N. E. Anagnostopoulos, I. E. Anagnostopoulos, I. D. Psoroulas, V. Loumos, and E. Kayafas. License plate recognition from still images and video sequences: A survey. *IEEE Transactions on intelligent transportation systems*, 9(3):377–391, 2008.

[3] R. Baran, A. Glowacz, and A. Matiolanski. The efficient real-and non-real-time make and model recognition of cars. *Multimedia Tools and Applications*, 74(12):4269–4288, 2015.

[4] L.-C. Chen, J.-W. Hsieh, Y. Yan, and B.-Y. Wong. Real-time vehicle make and model recognition from roads. In *2013 12th Conference on Information Technology and Applications in Outlying Islands*, pages 1033–1040, 2013.

[5] X. Clady, P. Negri, M. Milgram, and R. Poulenard. Multi-class vehicle type recognition system. In *Artificial Neural Networks in Pattern Recognition*, pages 228–239. Springer, 2008.

[6] Z. Dong, Y. Wu, M. Pei, and Y. Jia. Vehicle type classification using a semisupervised convolutional neural network. page in press, 2015.

[7] S. Du, M. Ibrahim, M. Shehata, and W. Badawy. Automatic license plate recognition (alpr): A state-of-the-art review. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(2):311–325, 2013.

[8] M. Everingham, S. A. Eslami, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. The pascal visual object classes

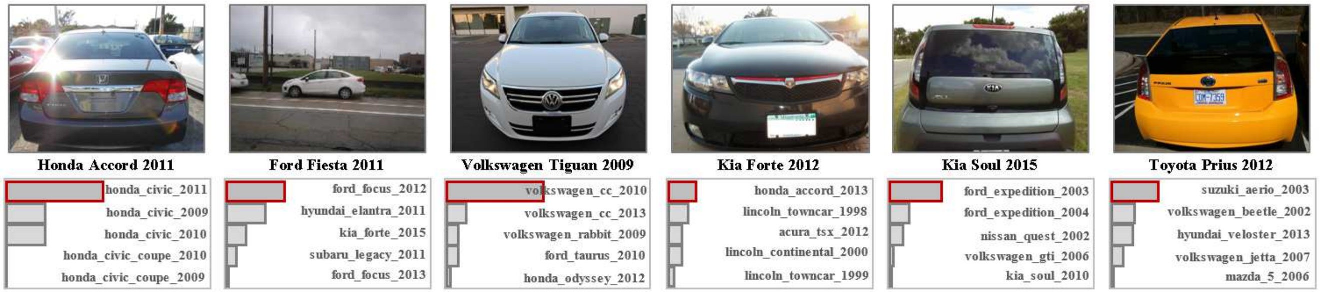


Figure 7: Top-5 predicted classes of ResNet-50 for sample images from VMMRdb-3036, incorrectly classified due to the ambiguity issue

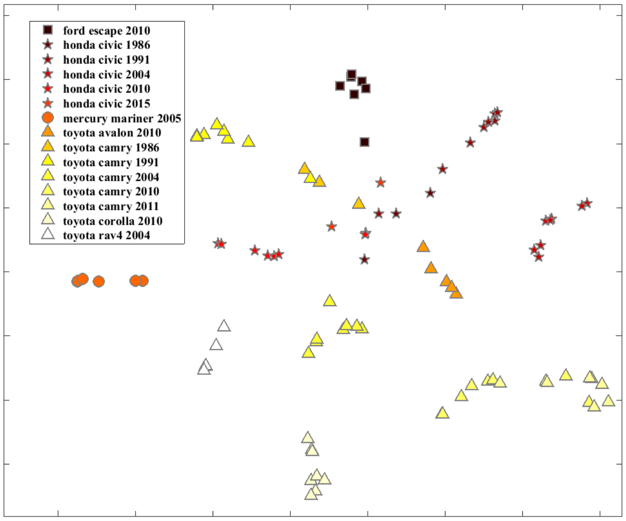


Figure 8: Features of sample car models projected to a 2D embedding using multi-dimensional scaling

challenge: A retrospective. *International Journal of Computer Vision*, 111(1):98–136, 2015.

[9] J. Fang, Y. Zhou, Y. Yu, and S. Du. Fine-grained vehicle model recognition using a coarse-to-fine convolutional neural network architecture. *IEEE Transactions on Intelligent Transportation Systems*, 2016.

[10] Y. Gao and H. J. Lee. Deep learning of principal component for car model recognition. In *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV)*, page 48. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2015.

[11] H.-Z. Gu and S.-Y. Lee. Car model recognition by utilizing symmetric property to overcome severe pose variation. *Machine vision and applications*, 24(2):255–274, 2013.

[12] 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*, pages 770–778, 2016.

[13] E. Hsiao, S. N. Sinha, K. Ramnath, S. Baker, L. Zitnick, and R. Szeliski. Car make and model recognition using 3d curve alignment. In *Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on*, pages 1–1. IEEE, 2014.

[14] 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*, 15(1):6–20, 2014.

[15] J. Krause, M. Stark, J. Deng, and L. Fei-Fei. 3d object representations for fine-grained categorization. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 554–561, 2013.

[16] S. Lee, J. Gwak, and M. Jeon. Vehicle model recognition in video. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 6(2):175, 2013.

[17] Y.-L. Lin, V. I. Morariu, W. Hsu, and L. S. Davis. Jointly optimizing 3d model fitting and fine-grained classification. In *Computer Vision–ECCV 2014*, pages 466–480. Springer, 2014.

[18] H. Liu, Y. Tian, Y. Yang, L. Pang, and T. Huang. Deep relative distance learning: Tell the difference between similar vehicles. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2167–2175, 2016.

[19] 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*, pages 3094–3099. IEEE, 2014.

[20] X. Ma and W. E. L. Grimson. Edge-based rich representation for vehicle classification. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 2, pages 1185–1192. IEEE, 2005.

[21] L. v. d. Maaten and G. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008.

[22] K. Matzen and N. Snavely. Nyc3dcars: A dataset of 3d vehicles in geographic context. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 761–768, 2013.

[23] D. T. Munroe and M. G. Madden. Multi-class and single-class classification approaches to vehicle model recognition from images. *Proceedings of IEEE AICS*, 2005.

[24] G. Pearce and N. Pears. Automatic make and model recognition from frontal images of cars. In *Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on*, pages 373–378. IEEE, 2011.

[25] V. S. Petrovic and T. F. Cootes. Analysis of features for rigid structure vehicle type recognition. In *BMVC*, pages 1–10, 2004.

[26] J. Prokaj and G. Medioni. 3-d model based vehicle recognition. In *Applications of Computer Vision (WACV), 2009 Workshop on*, pages 1–7. IEEE, 2009.

[27] A. Psyllos, C.-N. Anagnostopoulos, and E. Kayafas. Vehicle model recognition from frontal view image measurements. *Computer Standards & Interfaces*, 33(2):142–151, 2011.

[28] D. Santos and P. L. Correia. Car recognition based on back lights and rear view features. 2009.

[29] J. Sochor, A. Herout, and J. Havel. Boxcars: 3d boxes as cnn input for improved fine-grained vehicle recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3006–3015, 2016.

[30] Y. Wen, Y. Lu, J. Yan, Z. Zhou, K. M. von Deneen, and P. Shi. An algorithm for license plate recognition applied to intelligent transportation system. *IEEE Transactions on Intelligent Transportation Systems*, 12(3):830–845, 2011.

[31] L. Yang, P. Luo, C. Change Loy, and X. Tang. A large-scale car dataset for fine-grained categorization and verification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3973–3981, 2015.

[32] I. Zafar, E. A. Edirisinghe, S. Acar, and H. E. Bez. Two-dimensional statistical linear discriminant analysis for real-time robust vehicle-type recognition. In *Electronic Imaging 2007*, pages 649602–649602. International Society for Optics and Photonics, 2007.

[33] B. Zhang. Reliable classification of vehicle types based on cascade classifier ensembles. *IEEE Transactions on Intelligent Transportation Systems*, 14(1):322–332, 2013.