Journal of AI and Data Mining
Published online:

IRVD: A Large-Scale Dataset for Classification of Iranian Vehicles in Urban Streets

H. Gholamalinezhad and H. Khosravi*

Faculty of Electrical Engineering and Robotics, Shahrood University of Technology, Shahrood, Iran.

Received 15 May 2019; Revised 26 March 2020; Accepted 20 May 2020
**Corresponding author: hosseinkhosravi@shahroodut.ac.ir (H. Khosravi).*

Abstract
In recent years, vehicle classification has been one of the most important research topics. However, due to the lack of a proper dataset, this field has not been well-developed as other fields of intelligent traffic management. Therefore, the preparation of large-scale datasets of vehicles for each country is of great interest. In this paper, we introduce a new standard dataset of popular Iranian vehicles. This dataset, which consists of the images of the vehicles moving in urban streets and highways, can be used for vehicle classification and license plate recognition. It contains a large collection of vehicle images in different weather and lighting conditions with different viewing angles. It took more than a year to construct this dataset. The images were taken from various types of mounted cameras with different resolutions and at different altitudes. In order to estimate the complexity of the dataset, some classical methods alongside the popular deep neural networks were trained and evaluated on the dataset. Furthermore, two light-weight CNN structures are also proposed, one with three Conv layers and the other with five Conv layers. The 5-Conv model with 152K parameters reached the recognition rate of 99.09% and could process 48 frames per second on CPU, which is suitable for real-time applications.

Keywords: *Vehicle Dataset, Vehicle Classification, Deep Learning, IRVD.*

1. Introduction

In the modern world, transportation is an infrastructure sector of every society. One of the most important topics in the transportation sector is the traffic control. Having information about the vehicles moving on a road is helpful for the governments to improve the condition of that road. The first use of the traffic control systems dates back to the 1930s and 1940s; they were used to control the industrial processes [1, 2].

Nowadays, traffic control is more intelligent due to the use of the machine vision methods. Vehicle classification is a machine vision problem that is of interest in the today's traffic management systems; it is very important for roads, highways, parking lots, and toll stations.

Due to the enormous number of vehicles makes and models, vehicle classification is a complex task, and therefore, it requires more sophisticated and robust computer vision algorithms. Due to the diversity of the makes and models and different conditions during image acquisition, vehicle classification is more challenging than the other classification problems. Vehicles show large appearance differences in their unconstrained poses, which demand viewpoint-aware analyses and algorithms. This makes the image-based classification of vehicles a major challenge [3, 4] in addressing the issues related to the intelligent traffic management such as vehicular density estimation, utility specific lane maintenance, and load estimation.

Vehicle processing can be divided into two primary categories: vehicle location detection and vehicle identification.

The purpose of the first category is to just identify the vehicles from other stationary and moving objects in the image [5, 6]. In the second category, the purpose is to determine the type of vehicle, which can be considered in two aspects: identification of the general vehicle category and identification of the car model. A lot of research works have likewise been performed on the

second task. The subject of many traffic system users such as police and tolls is the recognition of the general class of vehicles, and identification of the car model is not important to them. Some of the most important and prominent research works conducted in this case are discussed below.

Liu *et al*. [7] were able to classify cars with the features extracted from the front window, two headlamps, and plate placement. In this method, a neural network with some histogram features are used to recognize the type of vehicle. The main weakness of this approach is that the process depends on the location of the vehicle, which can only be identified from the frontal face of the car. Messelodi *et al*. [8] have introduced a 3D model of moving objects. In this work, identification was made in five classes including cars, buses, pickups, bikes, and minibuses. The main weakness of this method is its sensitivity to the environmental conditions. For example, the presence of shadows or light reflections in images causes the vehicle to be misclassified. Jang *et al*. have proposed a new feature extraction method based on the SURF algorithm [9]; this method has a precision of 90%.

Kafai *et al*. [9, 10] have proposed a network-based approach using the features extracted from the rear parts of the vehicle. The feature vector is a set of geometric parameters of the car such as its width and height. In this method, a recognition rate of 95.7% can be achieved for a dataset of 169 cars in four classes.

Zhang *et al*. [11] have proposed a method for identifying the type of vehicle based on the structural error on the CVQ[1] network, reporting a 95% accuracy for a dataset of 2800 vehicle images of four classes.

In [12], a neural network has been used to classify the vehicles; they used the geometric features and achieved a precision of 69% for a dataset of 100 cars. The problem with this method is its low accuracy. Chen *et al*. [13] have offered a new method to improve the accuracy and reduce the computational complexity using the multi-branch and multi-layer features. In this way, each image is converted into several sub-images. Then using the proposed deep neural network, the local and global features are categorized. The dataset utilized in this research work was 57,600 images of 320 different classes (180 images of each class) at a resolution of 1920 × 1080. Based on this method, 53,120 images were used for training, and an accuracy of 94.88% was achieved. The main drawback of this method is the long response time of the recognition process for each vehicle.

In all the above methods, in order to determine the type of vehicle, it is necessary to extract the image feature, which will slow down the identification process. In [14], another method has been proposed based on the deep neural network (DNN); it does not require feature extraction. A DNN takes the actual images as the inputs and does not require the external feature extraction methods. This method has been compared with an SVM[2] classifier trained on the SIFT[3] features. Two different datasets (one containing the extracted images from a single camera and the other including the images received from two cameras at different locations) were used for evaluation. In the single-camera mode, the number of images examined was 1,500 and DNN achieved a recognition accuracy of 98.06%, while the SVM classifier achieved 97.35%. In the case of two cameras, the number of images analyzed was 6,555 images, and the recognition rates of 97.75% and 96.19% were achieved for DNN and SVM, respectively. Biglari *et al*. [15] have introduced new methods using a part-based model method. They reported an accuracy of 100% in the train data for recognition of the vehicle models. In these works, they used a small and clean dataset that included 4,858 fully annotated frontal images from 28 different makes and models.

There are very few datasets prepared for vehicle classification in the urban traffic and highway scenes. Some of these datasets are listed in table 1.

Due to the lack of suitable datasets for classification, sometimes the researchers use the datasets prepared for other goals. The i-LIDS datasets for event detection [16] have been used as one of the principal sources for collection of the vehicle-related data. The i-LIDS[4] datasets are licensed by the UK Home Office for the image research institutions and manufacturers. Each dataset comprises a 24 h of video sequences under a range of realistic operational conditions. The dataset appropriate for vehicle classification consists of videos of a busy urban road taken under different illumination and weather conditions. The vehicles move along two directions in the opposite lanes with almost no variations in the pose. This dataset has been used in [17-19].

[1]Classifier Vector Quantization

[2]Support Vector Machine

[3]Scale Invariant Feature Transform

[4]Imagery Library for Intelligent Detection Systems

Table 1. Some of the popular datasets used for vehicle classification.

Dataset	Urban/Highway	Classes	Lighting Condition	Number of Samples
iLids [16]	Urban	4 classes: Bike, Car, Van, Bus	Sunny, Overcast, Changing	24 h video with different frame rates
Yu Peng [20]	Highway	5 classes: Truck, Bus, Passenger Car, Minivan, Sedan	Daylight, Night	3,618 for day, 1,306 for night
BIT Vehicle [21]	Highway	6 classes: Truck, Bus, Microbus, SUV, Minivan, Sedan	Daylight, Night	9,850 images
Harish [22]	Urban	4 classes: Auto Rickshaw, Heavy, Light, Two Wheelers	Sunny, Moving shadows, Changing, Night Near IR	Day - 31,712 Evening - 33,764 Night NIR - 1,115
BVMMR [15]	Highway	28 classes	Daylight	5,991 images
BoxCars116k [23]	Highway	45 classes	Daylight	116,286 images from 27,496 cars
Asgarian [24]	Highway	9 classes	Daylight	7000 images
IRVD (ours)	Urban and Highway	5 classes: Bus, Heavy Truck, Medium Truck, Sedan, Pick-up	Sunny, Overcast, Night	110,366 images from distinct cars

Peng *et al*. [20] have collected a dataset for automatic license plate recognition. The dataset consists of 3,618 daylight images and 1,306 night time images. The images were taken on a highway at a resolution of 1600 × 1264. This dataset has been used in [20, 21, 25, 26].

Another dataset is BIT[1]-Vehicle introduced in [21]. This dataset has been specifically prepared for vehicle classification. The images have been captured from a road-camera installed on a highway, perpendicular to the direction of motion. The dataset consists of 6 vehicle categories including Bus, Microbus, Minivan, Sedan, SUV, and Truck with 150 images under each class. The dataset consists of the samples taken during daylight and night. The 'BIT-Vehicle dataset' has been used by [21] and [25] to report encouraging results for vehicle classification. However, the dataset contains only the frontal images of vehicles, limiting its usefulness as a dataset in developing the algorithms for practical urban installations. Fu *et al*. [26] have reported the results on a dataset of images with considerable variations in pose, illumination, and scales. It contains 8,053 images cropped from surveillance footage of four types of vehicles under 6 different views. Unfortunately, the dataset has not been released for public use or reproduction.

In [22], a new dataset has been introduced. In this article, the authors collected and organized a large-scale surveillance-nature image dataset for vehicle classification in practical installations. They also observed that there existed virtually no dataset of vehicles captured under extremely poor illumination conditions. Thus the emergence of day and night infrared cameras has extended the scope of vehicle classification to the near-infrared band of the electromagnetic spectrum.

In this paper, we introduced a new large-scale dataset for the Iranian vehicles collected in more than 1 year from different locations in Iran. In Section 2, we describe the details of the dataset[2]; the procedure we used for data collection, challenges of the dataset, and a study on its complexity are discussed here. In Section 3, we examine our dataset with some popular CNNs as well as two proposed networks and compare their accuracies and recognition times.

2. Description of IRVD dataset

Every vehicle identification system must be evaluated on a dataset to be compared with other systems. Thus the availability of an appropriate vehicle dataset is important for such systems. A standard dataset must accommodate different environmental conditions. More variations in the dataset make it more challenging.

Our dataset, called IRVD[3], was developed in different illuminations and air conditions including rainy weather, sunny, cloudy, moonlit night, and dark night. Furthermore, the road conditions were also diverse and include dry roads, wet roads, smart and clean asphalt, damaged asphalt, two-line and three-line roads, and two-way roads. The night images were taken in two ways, one using an LED projector and the other using the natural brightness of the moonlight

[1]Beijing Institute of Technology

[2]Available at https://shahaab-co.com/en/download/irvd

[3]Iranian Vehicle Dataset

or the weak light of the adjacent cars. The cameras were installed at different heights from 0.5m to 7.0m. After capturing the videos, the frames containing the vehicles were cropped, and finally, the images were normalized to 256 × 256 pixels. The details are described in Section 2.2.
Figure 1 shows the sample images in different lighting conditions and different viewing angles.

Figure 1. Sample images from the IRVD dataset in different lighting conditions and different viewing angles.

The dataset is divided into five classes:

- 1st class: Bus (urban and suburban buses)
- 2nd class: Heavy trucks and lorries
- 3rd class: Medium trucks, minibus, and van
- 4th class: Sedan cars
- 5th class: Pick-ups

Some sample images from each class in various conditions are shown in Figure 2.

Figure 2. Sample images of the IRVD dataset. From top row to bottom: Bus, Heavy truck, Medium truck, Sedan and Pick-up.

The number of images per class is depicted in table 2. As it can be seen, more than half of the dataset belongs to the Sedan cars that in real life exist more than the other vehicles.

Table 2. Number of images per class in IRVD dataset.

Class 1	Class 2	Class 3	Class 4	Class 5
Bus	Heavy truck	Medium truck	Sedan	Pick-up
5,024	16,715	8,255	71,355	9,017

As described earlier, we considered various conditions for collecting images of the IRVD dataset. These conditions among the number of samples in each condition are listed in table 3. It must be mentioned that all images were taken from a frontal view of the vehicles but the viewing angles are different, as shown in figures 1 and 2.

Table 3. Number of images in different conditions.

Condition	Description	#Samples	Percentage of total
Air Condition	Snowy	16,555	15%
	Rainy	22,073	20%
	Sunny	71,738	65%
Location	Urban	60,701	55%
	Street	49,665	45%
Lighting Condition	Day light	66,220	60%
	Night (moon light)	5,238	5%
	Night (light of near vehicles)	5,573	5%
	Night (LED projector light)	33,335	30%
Height of Camera	7.0 m	16,565	15%
	5.5 m	44,136	40%
	0.5–3 m	49,665	45%
Resolution	384 × 216, 640 × 480, 1280 × 720, 1280 × 736, 1000 × 750, 1600 × 1200, 1920 × 1080		

2.2. Collecting and cropping method

We spent more than a year (Sep 2017 – Dec 2018) on collecting data from different locations. The images were extracted from several cameras with different resolutions. These videos were then split into two categories. In one category, the distance and viewing angle of the camera is such that the license plate is clear and can be read with ANPR libraries. In the second category, the license plate quality is not suitable for OCR. Some frames of these two categories are shown in figure 3.
The first category's videos were analyzed by the SATPA library [27], and the license plate location was detected. According to the bounding box of the license plate, the surrounding rectangle of the vehicles was constructed and cropped. In the second category, the vehicle images were cropped using the YOLO algorithm [28]; YOLO is a unified model for object detection. This model is simple to construct and can be trained directly on full images. In this way, the vehicles are tracked, and when they are close enough, their bounding

boxes are cropped. In some cases, the bounding box of the vehicle was selected and cropped manually.

Figure 3. Some IRVD images suitable (bottom) and not suitable (top) for license plate recognition.

2.3. Applications of IRVD dataset

According to the various conditions presented in our dataset, it can be used in different applications. Some of the important cases are described in the following.

2.3.1. Vehicle classification

As described in Section 1, the most important goal for this dataset is car classification. IRVD is a fine-grained dataset and has several challenges, so it can be a good case for classification algorithms. We divided the dataset into two separate sets: training and test. The training subset contains 60% of the data and the test subset contains 40% of the data. In order to have an estimate of the complexity of the dataset, some classifiers were applied to it and the results obtained were presented in Section 3.

2.3.2. License plate recognition

As described earlier in Section 2, some images of our dataset are cropped according to the location of the license plate. These images have a good quality for OCR. Due to the different viewing angles and weather conditions, IRVD is also useful for the evaluation of the Persian ANPR systems. Some examples of such images are shown in figure 4.

Figure 4. Some IRVD images suitable for Persian license plate recognition systems.

2.3.3. Data mining

We collected our dataset from different places in Iran during the years 2017 and 2018. Thus another application is data mining for various purposes. For example, we can find the popular vehicles and market shares for each vehicle or using the image processing tools for vehicle appearance, we can understand the average economic level of the automobile community.

3. Dataset evaluation

In this section, we evaluate several classification methods on our dataset. We used four classic methods including the SVM, MLP, KNN, and Bayesian classifiers along with some popular CNN networks including VGG, ResNet, and DarkNet. Two deep structures are also proposed and evaluated.

For the classic methods, we used the Histogram of Oriented Gradients (HOG) as the input features. The HOG descriptor technique counts occurrences of gradient orientation in localized portions of an image or region of interest (ROI) [29]. We used these parameters for HOG extraction: block size = 32, cell size = 32, block stride = 16, and bins = 9. Using these parameters, 2,025 features were extracted from each image of the dataset.

MLP has two hidden layers of sizes 128 and 64 neurons. Considering its input and output neurons, its structure is 20225: 128: 64: 5, where 5 is the number of classes. It is implemented using PyTorch, and two activation functions were tried: Sigmoid and Tanh. SGD with momentum was selected as the optimizer.

SVM was implemented using the OpenCV library in Python, and its best result was achieved using a linear kernel with C (regularization parameter) equal to 1.

KNN and Bayesian were also implemented using OpenCV. Bayesian had no parameter to be adjusted but K in KNN was selected to be 5. The results of these classifiers along with other methods are listed in table 4 and will be discussed later in this section.

3.2. Convolution neural network

Nowadays, the application of neural networks has become very widespread, and a huge variation has

been found. In the traditional ANNs, the input was a vector of features. This means that the training (and test) data must be transferred to a feature space, and then fed into the neural network. This process differs from the process of learning in the human brain. The brain receives and learns the data without altering it. Research has, therefore, been carried out to improve the existing neural networks. DNNs are the outcome of these research findings. Although the initial idea of deep learning was presented several years ago [30], due to the weakness of the hardware in the past, it was not possible to implement this algorithm. With the advances made in the hardware, the implementation of these algorithms has already been realized. CNN is one of the most popular deep learning networks. This network consists of three main layers: the convolutional layer, the pooling layer, and the fully connected layer. Different layers perform different tasks.

In order to further evaluate our dataset, we trained some popular CNNs on IRVD. These are VGG, ResNet, and DarkNet53. VGG [31] is one of the first popular structures that produced outstanding results in the ILSVRC-2014 competition. Its original structure cannot be trained from scratch due to the vanishing gradients problem [32]. However, today, with the help of batch normalization, its modified version can be trained from scratch. We used VGG11 with batch normalization and trained it on our dataset.

The second structure that we used was ResNet [33]. It has a novel structure, and having the new residual connections, can be trained on every dataset. It supports very deep layers and exists in different versions from ResNet18 to ResNet152 and even more. We trained three structures, ResNet18, ResNet50, and ResNet152, on IRVD.

Finally, we examined DarkNet [34] on our dataset. DarkNet is an open-source neural network written in C. It is the backbone of the YOLO detector [35].

These networks are so deep, and as a result, they are very slow, especially when using CPU. Thus we tried to construct shallower networks to be useful for real-time applications while producing acceptable results comparing with these structures. We tried two different structures. The first one had three convolution layers with 32,549 parameters (Figure 5). It consists of 16 filters (3 × 3) in the first Conv layer, 32 filters in the second, and 64 filters in the last Conv layer. Each Conv layer is followed by a batch normalization, a ReLU, and a max-pooling layer. The first pooling layer has a window size of 4 × 4 and a stride of 4. The other pooling layers have a window size of 2 × 2 and a stride of 2. A global average pooling layer connects the Conv layers to the fully connected (FC) layers. The first FC layer contains 128 neurons and the final layer contains 5 neurons equal to the number of classes.

The second structure contains five Conv layers with 151,133 parameters. The overall structure is the same as the previous model, except for the number of Conv layers and the number of filters. Its Conv layers consist of 16, 32, 64, 64, and 128 filters, respectively. All filters have a size of 3 × 3. Pooling layers also have the same size of 2 × 2 and a stride of 2. The fully connected layers are the same as the 3-Conv structure.

We conducted our experiments on a PC with a GeForce RTX-2080 GPU running at 1700 MHz and a Corei7-9700K CPU running at 4000 MHz.

Table 4 shows the results of our experiments. As shown here, among the traditional classifier trained with HOG, the normal Bayesian classifier produced the worse recognition rate of 66.57%. Looking at its confusion matrix, we saw that most of the samples were considered as a sedan car! because almost 65% of our data are sedans, and on the other hand, the Bayesian classifier is a statistical classifier, which is simply biased to the class with the most samples.

The best result among the traditional methods belongs to MLP with a 94.89% accuracy. The results of these classical methods show that the dataset has enough complexity.

Among the popular CNNs, the worse result belongs to ResNet152 with a 98.92% accuracy. This network is so deep and has 58 million parameters. Thus to have good results, the dataset must be very large. As a result, training it on IRVD did not produce fantastic results. To get better results, it must be used in a fine-tuned process, which is out of the scope of this paper.

The best performing CNN was ResNet18 with a 99.50% accuracy. We think that this behavior is due to the fact that its capacity (number of trainable parameters) is more suitable for IRVD than more deep structures that suffer from overfitting.

Among the proposed light structures, the best results achieved by the 5-Conv structure with 152K parameters that achieved a 98.63% accuracy and when data augmentation added, it reached a rate of 99.07%, which is very good. It is also suitable for real-time applications even on CPU. This model only takes 15.8 ms for prediction, and if we add the 5 ms required for image loading, the total process of the perdition is 20.8 ms, which means 48 FPS.

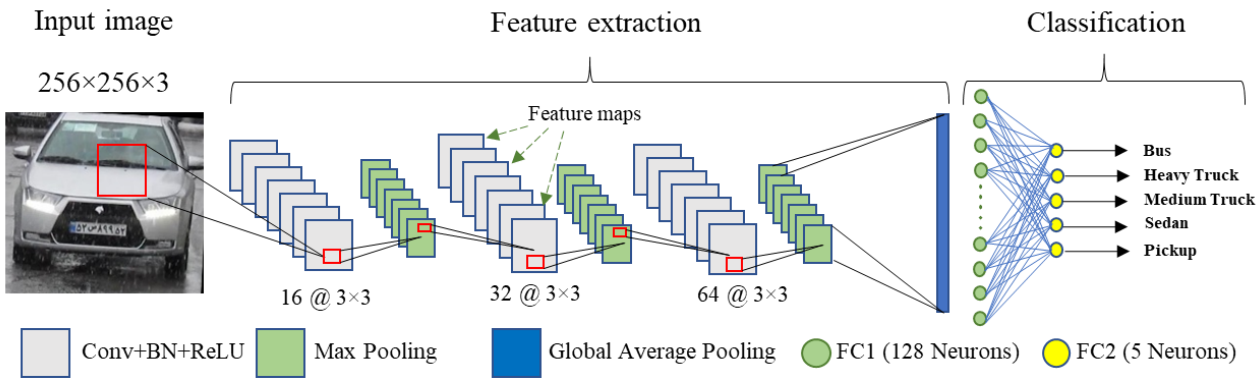

Figure 5. A light weight 3-Conv structure using BN and GAP layers proposed for vehicle classification.

Table 4. Comparison of the proposed method with some traditional classifiers and some popular CNNs. Prediction times do not include image loading, which takes about 5 ms and is always performed in CPU.

	Model	No. of Parameters	Prediction time CPU	Prediction time GPU	Test Rec. rate	Test loss
Traditional classifiers trained on 2025 HOG features	Bayesian Classifier	-	3.802ms	-	66.57	-
	MLP + Tanh MLP + Sigmoid	267,909	0.864ms	-	**94.89** 93.10	0.1537 0.2029
	SVM	-	**0.861ms**	-	77.79	-
	KNN (K=5)	-	13.922ms		93.00	-
Popular convolutional neural networks	VGG11 + BN	128,792,325	110.01ms	**0.207ms**	99.24	0.0346
	ResNet18	11,179,077	**39.67ms**	0.355ms	**99.50**	0.0253
	ResNet50	23,518,277	97.92ms	0.863ms	99.11	0.0476
	ResNet152	58,154,053	220.49ms	2.441ms	98.92	0.0519
	DarkNet53	40,591,078	139.77ms	0.877ms	99.38	0.0328
Proposed networks	3-Conv	32,549	**6.33ms**	**0.081ms**	95.70	0.1420
	3-Conv + BN	32,773	8.42ms	0.100ms	96.30	0.1126
	5-Conv + BN	152,133	15.80ms	0.141ms	98.63	0.0709
	5-Conv + BN + Aug	152,133	15.80ms	0.141ms	**99.07**	0.0392

3.3. Complexity of dataset

According to the results of the traditional methods on the IRVD, shown in table 4, it is clear that IRVD has an acceptable complexity. However, there are also some statistical methods for the complexity assessment of the dataset. The simplest one is to compute the mean of each class and visualize it. Figure 6 shows the mean images for five classes of the dataset. As shown here, these images are very blurred, which is a sign of the within-class variations.

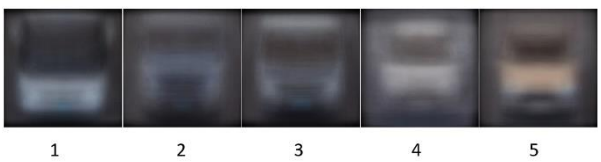

Figure 6. Mean image of different classes of the IRVD dataset.

A recent method for a complexity measure has been proposed in [36]. It is called the cumulative spectral gradient (CSG) and has been claimed to be strongly correlated with the test accuracy of the convolutional neural networks. This method computes a so-called W-Matrix (Eq. 1) and makes a 2D plot of the dataset that shows the inter-class distances.

$$w_{ij} = 1 - \frac{\sum_{k}^{K} \left| S_{ik} - S_{jk} \right|}{\sum_{k}^{K} \left| S_{ik} + S_{jk} \right|} \quad (1)$$

Here, S_{ik} indicates the similarity measure between classes i and k and W_{ij} is the distance between the likelihood distributions of classes C_i and C_j. For more details refer to [36].

Figure 6 shows a 2D plot of the W Matrix computed from Eq. 1. As it can be seen, heavy truck and medium truck are too close to each other. The same is for sedan and pick-up.

However, the bus class is far from the other classes.

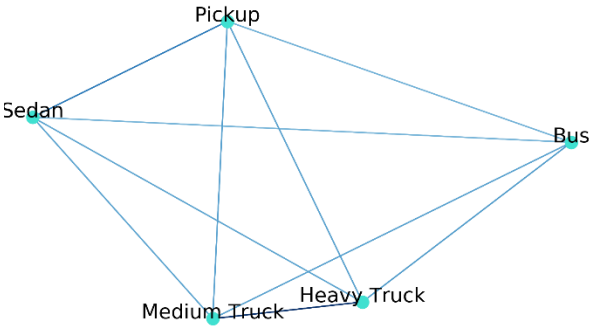

Figure 7 Two-dimensional plot of the W matrix, showing the inter-class distances.

4. Conclusion

We introduced a new image dataset of Iranian vehicles called IRVD. It is useful for both the vehicle classification and license plate recognition. Our dataset contains images of various conditions so it can be used for real-world applications. The procedure of data collection and cropping car area was discussed and some other applications of our dataset were introduced. IRVD is separated into two groups: 60% for training and 40% for testing. We evaluated our dataset with some traditional and deep learning methods. In the first one, we extracted the HOG features from our images and fed them into the SVM, MLP, Bayesian, and KNN classifiers. The best results were achieved by MLP with a 94.89% accuracy for the test data. In the second category, we used some popular CNNs among the two proposed structures. The best results were achieved by ResNet18 with a 99.50% accuracy, while our 5-Conv structure achieved a 99.07% accuracy. However, considering the required time for prediction, our method performs much faster and can process 48 frames per second on CPU, which makes it suitable for the real-time applications.

Acknowledgment

This work was supported by the research unit of Shahaab company (https://shahaab-co.com/en/). The authors would like to thank the staff of this company for their help.

References

[1] Webster, C. W. R., Töpfer, E., Klauser, F. & Raab, C. D. (2011). Revisiting the surveillance camera revolution: Issues of governance and public policy. Introduction to part one of the Special Issue. Information Polity, vol. 16, no. 4, pp. 297-301.

[2] Agustina, J. R. & Clavell, G. G. (2011). The impact of CCTV on fundamental rights and crime prevention strategies: The case of the Catalan Control Commission of Video surveillance Devices. Computer law & security review, vol. 27, no. 2, pp. 168-174.

[3] Krause, J., Deng, J., Stark, M. & Fei-Fei, L. (2013) Collecting a large-scale dataset of fine-grained cars. presented at the CVPR-FGCV2, Portland, Oregon.

[4] Yang, L., Luo, P., Change Loy, C. & Tang, X. (2015). A large-scale car dataset for fine-grained categorization and verification. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3973-3981.

[5] Sun, Z., Bebis, G. & Miller, R. (2006). On-road vehicle detection: A review. IEEE transactions on pattern analysis and machine intelligence, vol. 28, no. 5, pp. 694-711.

[6] Li, X.& Guo, X. (2013). A HOG feature and SVM based method for forward vehicle detection with single camera. in Intelligent Human-Machine Systems and Cybernetics (IHMSC), 5th International Conference on, vol. 1: IEEE, pp. 263-266.

[7] Liu, Y., You, Z., Cao, L. & Jiang, X. (2004). Vehicle detection with projection histogram and type recognition using hybrid neural networks. in Networking, Sensing and Control, 2004 IEEE International Conference on, vol. 1: IEEE, pp. 393-398.

[8] Messelodi, S., Modena, C. M. & Zanin, M. (2005). A computer vision system for the detection and classification of vehicles at urban road intersections. Pattern analysis and applications, vol. 8, no. 1-2, pp. 17-31.

[9] Jang, D. M.& Turk, M. (2011). Car-Rec: A real time car recognition system. in applications of computer vision (WACV), 2011 IEEE Workshop on: IEEE, pp. 599-605.

[10] Kafai, M.& Bhanu, B. (2012). Dynamic Bayesian networks for vehicle classification in video. IEEE Transactions on Industrial Informatics, vol. 8, no. 1, pp. 100-109.

[11] Zhang, B., Zhou, Y. & Pan, H. (2013). Vehicle classification with confidence by classified vector quantization. IEEE Intelligent Transportation Systems Magazine, vol. 5, no. 3, pp. 8-20.

[12] Matos, F. M. d. S. & de Souza, R. M. C. R. (2013). Hierarchical classification of vehicle images using nn with conditional adaptive distance. in International Conference on Neural Information Processing: Springer, pp. 745-752.

[13] Chen, C., Cai, X., Zhao, Q., Lv, L. & Shu, H. (2017). Vehicle Type Recognition Based on Multi-branch and Multi-layer features.

[14] Huttunen, H., Yancheshmeh, F. S. & Chen, K. (2016). Car type recognition with deep neural networks. in Intelligent Vehicles Symposium (IV), 2016 IEEE: IEEE, pp. 1115-1120.

[15] Biglari, M., Soleimani, A. & Hassanpour, H. (2018). Using discriminative part for vehicle make and model recognition. Signal and data processing, vol. 15, pp. 41-54.

[16] Branch, H. O. S. D. (2006). Imagery library for intelligent detection systems (i-lids). in 2006 IET Conference on Crime and Security: IET, pp. 445-448.

[17] Buch, N., Orwell, J. & Velastin, S. A. (2010). Urban road user detection and classification using 3D wire frame models. IET Computer Vision, vol. 4, no. 2, pp. 105-116.

[18] Buch, N., Orwell, J. & Velastin, S. A. (2009). 3D extended histogram of oriented gradients (3DHOG) for classification of road users in urban scenes.

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

[20] Peng, Y., Jin, J. S., Luo, S., Xu, M. & Cui, Y. (2012). Vehicle type classification using PCA with self-clustering. in Multimedia and Expo Workshops (ICMEW), 2012 IEEE International Conference on: IEEE, pp. 384-389.

[21] Dong, Z., Pei, M., He, Y., Liu, T., Dong, Y. & Jia, Y. (2014) Vehicle Type Classification Using Unsupervised Convolutional Neural Network. presented at the Proceedings of the 2014 22nd International Conference on Pattern Recognition.

[22] Bharadwaj, H. S., Biswas, S. & Ramakrishnan, K. (2016). A large scale dataset for classification of vehicles in urban traffic scenes. in Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing, Indian ACM, p. 83.

[23] Sochor, J., Špaňhel, J. & Herout, A. (2018). Boxcars: Improving fine-grained recognition of vehicles using 3-d bounding boxes in traffic surveillance. IEEE transactions on intelligent transportation systems, vol. 20, no. 1, pp. 97-108.

[24] Asgarian Dehkordi, R. & Khosravi, H. (2020). Vehicle Type Recognition based on Dimension Estimation and Bag of Word Classification. Journal of AI and Data Mining, doi: 10.22044/jadm.2020.8375.1975.

[25] Dong, Z., Wu, Y., Pei, M. & Jia, Y. (2015). Vehicle type classification using a semisupervised convolutional neural network. IEEE transactions on intelligent transportation systems, vol. 16, no. 4, pp. 2247-2256.

[26] Fu, H., Ma, H., Liu, Y. & Lu, D. (2016). A vehicle classification system based on hierarchical multi-SVMs in crowded traffic scenes. Neurocomputing, vol. 211, pp. 182-190.

[27] SATPA license plate recognition library (2020), Available: https://shahaab-co.ir/license-plate-recognition-library/.

[28] Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. (2016). You only look once: Unified, real-time object detection. in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779-788.

[29] Dalal, N. & Triggs, B. (2005). Histograms of oriented gradients for human detection. in international Conference on computer vision & Pattern Recognition (CVPR'05), vol. 1: IEEE Computer Society, pp. 886--893.

[30] Dechter, R. (1986). Learning while searching in constraint-satisfaction problems. University of California, Computer Science Department, Cognitive Systems.

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

[32] François, C. (2017) Deep learning with Python. Manning Publications Company.

[33] He, K., Zhang, X., Ren, S. & Sun, J. (2016). Deep residual learning for image recognition. in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778.

[34] Redmon Darknet, J.: Open Source Neural Networks (2016), Available: https://github.com/pjreddie/darknet.

[35] Redmon, J. & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

[36] Branchaud-Charron, F., Achkar, A. & Jodoin, P.-M. (2019). Spectral metric for dataset complexity assessment. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3215-3224.