1-1-2021

Convolutional Neural Network Based Vehicle Classification in Adverse Illuminous Conditions for Intelligent Transportation Systems

Muhammad Atif Butt
National University of Sciences and Technology Pakistan

Asad Masood Khattak
Zayed University

Sarmad Shafique
Mirpur University of Science and Technology

Bashir Hayat
Institute of Management Sciences

Saima Abid
Mirpur University of Science and Technology

Author First name, Last name, Institution

Muhammad Atif Butt, Asad Masood Khattak, Sarmad Shafique, Bashir Hayat, Saima Abid, Ki Il Kim, Muhammad Waqas Ayub, Ahthasham Sajid, and Awais Adnan

WILEY | Hindawi

Research Article

Convolutional Neural Network Based Vehicle Classification in Adverse Illuminous Conditions for Intelligent Transportation Systems

Muhammad Atif Butt,[1] Asad Masood Khattak [ID],[2] Sarmad Shafique,[3] Bashir Hayat [ID],[4] Saima Abid,[3] Ki-Il Kim [ID],[5] Muhammad Waqas Ayub,[6,7] Ahthasham Sajid,[8] and Awais Adnan[4]

[1]School of Electrical Engineering and Computer Science (SEECS), National University of Science and Technology (NUST), H-12, Islamabad 44000, Pakistan
[2]College of Technological Innovation, Zayed University, Abu Dhabi, UAE
[3]Control, Automotive and Robotics Lab Affiliated Lab of National Center of Robotics and Automation (NCRA HEC Pakistan), Department of Computer Science and Information Technology, Mirpur University of Science and Technology (MUST), Mirpur-10250, Pakistan
[4]Institute of Management Sciences, Peshawar, Pakistan
[5]Department of Computer Science and Engineering, Chungnam National University, Daejeon, Republic of Korea
[6]Department of Engineering, Lancaster University, Lancaster LA1 4YW, UK
[7]Department of Electrical Engineering, COMSATS University, Lahore 54000, Pakistan
[8]Department of Computer Science, Faculty of ICT, Balochistan University of Information Technology Engineering and Management Sciences, Quetta, Baluchistan, Pakistan

Correspondence should be addressed to Asad Masood Khattak; asad.khattak@zu.ac.ae

In step with rapid advancements in computer vision, vehicle classification demonstrates a considerable potential to reshape intelligent transportation systems. In the last couple of decades, image processing and pattern recognition-based vehicle classification systems have been used to improve the effectiveness of automated highway toll collection and traffic monitoring systems. However, these methods are trained on limited handcrafted features extracted from small datasets, which do not cater the real-time road traffic conditions. Deep learning-based classification systems have been proposed to incorporate the above-mentioned issues in traditional methods. However, convolutional neural networks require piles of data including noise, weather, and illumination factors to ensure robustness in real-time applications. Moreover, there is no generalized dataset available to validate the efficacy of vehicle classification systems. To overcome these issues, we propose a convolutional neural network-based vehicle classification system to improve robustness of vehicle classification in real-time applications. We present a vehicle dataset comprising of 10,000 images categorized into six-common vehicle classes considering adverse illuminous conditions to achieve robustness in real-time vehicle classification systems. Initially, pretrained AlexNet, GoogleNet, Inception-v3, VGG, and ResNet are fine-tuned on self-constructed vehicle dataset to evaluate their performance in terms of accuracy and convergence. Based on better performance, ResNet architecture is further improved by adding a new classification block in the network. To ensure generalization, we fine-tuned the network on the public VeRi dataset containing 50,000 images, which have been categorized into six vehicle classes. Finally, a comparison study has been carried out between the proposed and existing vehicle classification methods to evaluate the effectiveness of the proposed vehicle classification system. Consequently, our proposed system achieved 99.68%, 99.65%, and 99.56% accuracy, precision, and F1-score on our self-constructed dataset.

1. Introduction

With an exponential production of vehicles around the world, vehicle classification systems can play a significant role in the development of intelligent transportation systems, i.e., an automated highway toll collection, perception in self-driving vehicles, and traffic flow control systems. In the earlier times, laser and loop induction sensors-based methods have been proposed for the vehicle type classification [1–4]. These sensors have been installed under the pavement of the roads to collect and analyse the data to extract the relevant information regarding vehicles. However, the precision and stability of these methodologies are significantly influenced due to undesired weather conditions and impairment in the road pavement [5]. In step with the advancement in computer vision, image processing and pattern recognition-based vehicle classification systems have been proposed [6, 7]. Basically, computer vision-based classification system is a two-step procedure; in the first step, handcrafted extraction methods are utilized to obtain visual features from input visual frame. In the second step, machine learning classifiers are trained on the extracted features to perform classification on group-based data. Hand-crafted features are categorized into (i) global and (ii) local features to describe and represent the image data simultaneously [8]. These features are combined in the training of traditional machine learning classifiers to perform object recognition. Though these systems perform well in the specific controlled environment and are more convenient in terms of installation and maintenance than the existing laser and inductive-based schemes, these methods are trained on the limited handcrafted features extracted from the small datasets, whereas extensive prior knowledge is required to maintain accuracy time environment [9].

Recently, deep learning-based feature extraction and classification methods have been introduced, which demonstrated better applicability and adaptability than the traditional classification systems. Convolutional neural network (CNN) based classification systems have achieved significant accuracy on the large-scale image datasets due to their sophisticated architecture [10–12]. Though, the development of the graphical processing unit (GPU) has significantly increased the image processing capabilities of the computing machines. But the matter of fact is that CNN based classification system requires piles of data to sustain accuracy and ensure generalization. Until recently, to the best of our knowledge, no generalized benchmark dataset is available for the development and evaluation of vehicle classification systems. Consequently, available vehicle classification datasets are relatively small, based on limited classes of the specific regions, i.e., CompCars [13] and Stanford cars dataset [14]. Intelligent transportation systems of these regions can achieve significant results with these datasets; however, their performance is prejudiced in the occurrence of nonregional classes. To address the above-mentioned limitations in vehicle-classification systems, we have made the following contributions.

(i) Convolutional Neural Network (CNN) based generalized vehicle classification architecture is presented to improve robustness of vehicle classification systems for Intelligent Transportation Systems (ITS) in adverse illuminous conditions.

(ii) A local dataset comprising of 10,000 images based on six classes (i.e., Car, Van, Truck, Motorbike, rickshaw, and Mini-Van) has been collected from traffic surveillance and driving videos. It is important to mention that these classes are unique in design and shape, which are not covered in the existing vehicle datasets.

(iii) Modified CNN has been employed and trained on the VeRi dataset, containing 50,000 images over six vehicle classes, to ensure generalization of the network.

(iv) Finally, an extensive comparison study has been carried out between the proposed and existing vehicle classification methods to demonstrate the effectiveness of the proposed classification network.

The rest of paper is organized as follows. In Section 2, the existing handcrafted and deep learning feature-extraction and vehicle-classification methods are discussed briefly. In Section 3, network architecture along with the preprocessing and dataset collection has been elaborated. The results and the comparison study are carried out in Section 4. Finally, the article is concluded in Section 5.

2. Related Work

In step with the rapid advancement in artificial intelligence, vision-based vehicle classification is considered as an important element in perception module of self-driving vehicles. In the existing research work [5], vision-based vehicle classification is categorized into two major categories: (i) handcrafted features-based and (ii) deep features-based methodologies.

In the early era of computer vision, handcrafted features-based vehicle classification methods have been proposed for intelligent transportation systems. In this regard, Ng et al. [15] have proposed HOG-SVM based handcrafted features method to train SVM classifier using HOG features with Gaussian kernel function. The proposed classifier has been evaluated on 2800-image dataset of surveillance videos, which classified the motorcycle, car, and lorries with 92.3% accuracy. In another research work, Chen et al. [16] have presented a classification method that extracts the texture and HOG features and classifies the vehicles using a fuzzy inspired SVM classifier. The presented classifier has been evaluated on dataset, comprising of 2000 images in which the proposed systems classified the cars, vans, and buses with 92.6% accuracy. Matos et al. [17] have proposed two-neural networks based combined method embedding the features, i.e., height, width, and bounding borders of the vehicles. Resultantly, the proposed classifier achieved 69% on the dataset of 100 images. Furthermore, Cui et al. [18] have

proposed Scale Invariant Feature Transform (SIFT) descriptors and Bad of Words (BoW) based combined model for the extraction of the features and utilized SVM to classify the dataset consisting 340 images of cars, minibuses, and trucks. In the results, it is shown that the proposed classifier achieved 90.2% accuracy on the provided dataset. Wen et al. [19] have proposed an AdaBoost based fast learning vehicle classifier to distinguish the data into vehicle and nonvehicle classes. Moreover, the authors have proposed an algorithm to extract Haar-like features for the rapid learning of classifiers. The presented classifier has been evaluated on the public Caltech dataset, in which the system achieved 92.89% accuracy.

To overcome the issues of the handcrafted features-based classifiers, deep features-based systems have been proposed. Dong et al. [20] have presented CNN based semisupervised classification method for real-time vehicle classification. A sparse-Laplacian filter-based method has been devised to extract relative vehicle information, and the softmax layer has been trained to calculate the class probability of the belonging vehicle. The presented method has been evaluated on the Bit-Vehicle dataset and achieved 96.1% and 89.6% accuracy in day and night images, respectively. In another research work, Wang et al. [21] have presented a Fast R–CNN based vehicle classification method for traffic surveillance in a real-time environment. A crossroad dataset consisting of 60,000 images has been collected and divided into training and tested data, on which the proposed method attained 80.051% accuracy. Cao et al. [22] have proposed CNN and an end-to-end combined architecture for the vehicle classification in the incontinent road environment. The proposed framework has been evaluated on the CompCars view-aware dataset, in which the proposed classifier achieved a 0.953 accuracy rate. Chauhan et al. [23] have proposed CNN based vehicle classification framework for vehicle classification and counting on highway roads. Authors have claimed that the proposed framework achieved 75% MAP on the collected dataset of 5562 CCTV camera videos of highway traffic. Jo et al. [24] have proposed a transfer learning-based GoogLeNet framework for vehicle classification of road traffic. The authors have shown that the presented classifier has achieved a 0.983 accuracy rate while experimenting on the ILSVRC-2012 dataset. Kim et al. [25] have proposed the PCANeT-HOG-HU based combined feature extraction method, which is provided to SVM as input data to train the classification model. Moreover, the authors have collected the dataset consisting of 13700 images of vehicles considering six-categories of vehicles (i.e., motorcycle, van, car, truck, mini-bus, and large-bus), extracted from the surveillance videos for the training and testing of the proposed classification network. Results demonstrated that the proposed light-weight classifier achieved 98.34% average accuracy on the provided dataset.

Though the deep feature-based approaches can enhance the accuracy of vehicle classification effectively, these methodologies need a huge amount of data to achieve significant accuracy in real-time ITS applications [26–29]. In the recent era, extensive research has been carried out in this field; however, the available public datasets for self-driving vehicles/intelligent transportation systems comprise modern vehicle types, which are common in well developed countries. Consequently, these classification systems are not feasible for the intelligent transportation systems in Asian countries, i.e., Pakistan, India, Bangladesh, and China. The above-mentioned issues are indication towards the need of a novel vehicle classification system along with the dataset that covers the common vehicles, i.e., traditional trucks, buses, cars, rickshaws, and motorbikes of Asian countries.

3. The Proposed Method

To address the above-mentioned issues, we present a new vehicle dataset comprising of 10,000 images having six classes based on the common road traffic vehicles, as elucidated in Figure 1. To enhance the performance of the proposed classification in real-time ITS applications, initially, the existing pretrained AlexNet [30], VGG [31], GoogleNet [32], Inception-v3 [33], and ResNet [34] are fine-tuned on self-constructed dataset to obtain the final network. Based on the performance of these models, the best performing model is selected for the fine-tuning to increase the classification accuracy of the network. To ensure generalization, the proposed network is further fine-tuned on public VeRi dataset for robust performance in the intelligent transportation system of different regions. The whole process is briefly discussed below in Figure 1.

3.1. Dataset. In deep learning-based classification systems, dataset is a key input that helps the algorithms learn the features to perform predictions based on the learned information. Currently, to the best of our knowledge, there is no generalized public vehicle dataset available that contains the images of the common vehicles to cater with the classification problems. For example, CompCars and Standford car datasets only contain the classes of modern cars of certain regions, which cannot be employed in the real-time classification systems of the other regions. Moreover, the proposed dataset is different from the existing datasets in terms of features and representations. Additionally, the existing vehicle classification systems are trained on relatively small datasets containing limited classes, which does not perform well in real-time intelligent transportation systems applications [35]. To address these issues, road surveillance and driving videos are collected from different regions to extract the images of the vehicles. Based on analyses, six common road vehicle classes are identified, and the dataset is formed through manual labelling using windows editing tool, as shown in Figure 2. The dataset comprises 10,000 images that have been categorized into six classes (i.e., car, bus, van, truck, motorbike, and rickshaw), and each class consists of 1670 images.

3.2. Data Augmentation. Equations data augmentation is the easiest and most common technique to mitigate overfitting from network by artificially expanding the dataset through label-preserving transformation methods [36]. To increase the diversity of our dataset, we employed four
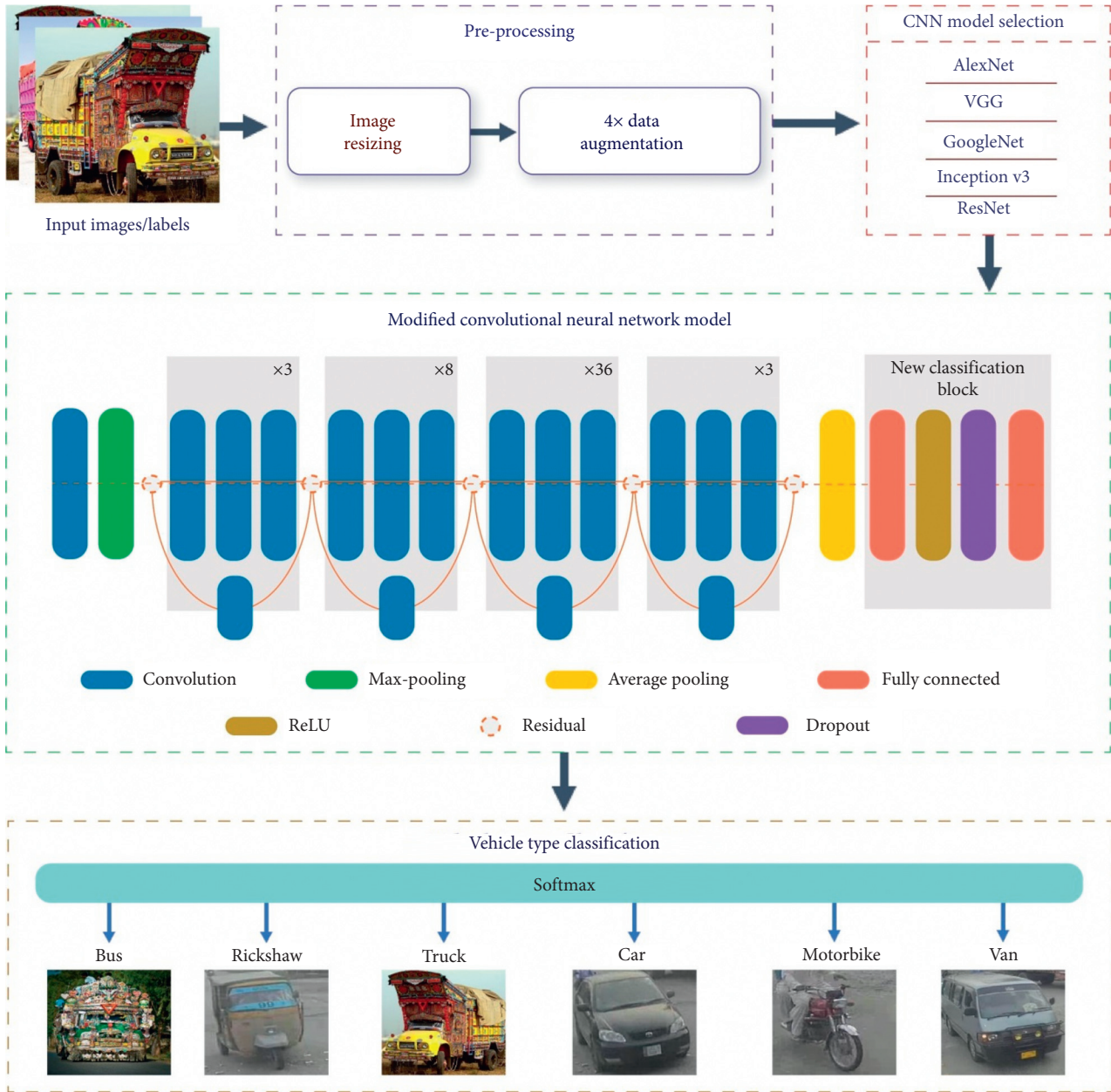
FIGURE 1: The proposed method.

distinct types of data augmentation: (i) Gaussian blur, (ii) rotation, (iii) horizontal flip, and (iv) Gaussian noise, as shown in Figure 3.

We utilized Gaussian blur with the 5×5 kernel to reduce the high-frequency noisy pixels while preserving the low spatial frequency through convolving the Gaussian kernel over the 224×224 size image. In the second type of data augmentation, i.e., rotation, we employed 10-degree rotation on the original dataset images to generate a diverse view of the original dataset. The third type of data augmentation includes the dataset generation through the horizontal flipping of the original dataset, whereas we used Gaussian noise as a fourth type of data augmentation to add some random luminous factor in the dataset. It is important to mention here that the horizontal flip, Gaussian blur, and

rotation are applied to the training dataset, whereas Gaussian noise is applied to the test dataset, as shown in Figure 3. The main purpose of applying Gaussian noise-based data augmentation to the test dataset is to validate the efficacy of the proposed classification network on the noisy data.

3.3. Convolutional Neural Network (CNN) Model. CNNs are supervised feed-forward networks that proved considerably significant performance on the large-scale object classification applications. The basic structure of the CNNs is stimulated by the key visual cortex of the human brain, which oversees the processing of visual information [37]. In the image classification, compared with the traditional

(a)	(b)	(c)	(d)	(e)	(f)

FIGURE 2: Sample dataset images representing each class: (a) car, (b) van, (c) truck, (d) motorbike, (e) bus, and (f) rickshaw.

handcrafted features extraction methods, the CNNs can automatically extract the learnable visual features from the large-scale dataset input images from the classes to perform the classification. One of the main superiorities of the CNNs over traditional classification methods is that, in CNNs, representation of the features and the classifier are employed in the same network to eliminate their dependencies. The architecture of the CNNs principally comprises three types of layers, (i) convolution layers, (ii) pooling layers, and (iii) the connecting layers, briefly discussed below.

3.4. Convolutional Layers.
Convolutional layers are considered as one of the most important layers in the CNNs, which consist of the defined set of learnable filters. The filters are spatially smaller than the input-size, which slides over the input image data during the forward pass to produce the two-dimensional activation map. The activation map indicates the location along with the strength of the detected visual features in an input image. The calculation of the features of the convolutional layers is obtained using

$$y_n^l = fl\left(\sum m \longrightarrow {}_n^l\, y_m^{l-1}\right), \tag{1}$$

where y_n^l is the nth feature map of l-layer, $m \longrightarrow {}_n^l$ is the C-kernel, while feature extraction from layer-l, and y_m^{l-1} is the Characteristic patterns linked to layer-l.

3.5. Pooling Layers.
Pooling layer is commonly used between consecutive convolution layers of the CNN structure to gradually minimize the spatial representation size to reduce computations while retaining useful information, which helps in controlling overfitting during the learning process. It is important to mention that there are two types of pooling layers being used in the existing state-of-the-art CNNs, i.e., a pooling layer having filter size = 2 along with stride = 3, which is called overlapping pooling; the other pooling layer

with filter size = 2 is having stride of 2. Besides, some other types of pooling, i.e., L2-norm pooling and average-pooling functions, have also been used in the existing CNNs. The pooling function can be performed through

$$y_n^l = fl\left(z_n^{l-1}\, x\, w_n^l + b_n^l\right), \tag{2}$$

where z_n^{l-1} is the value extracted from l - 1 convolution features, w_n^l is the map weight, and b_n^l is the offset value.

3.6. Drop-Out Layer.
In CNNs, regularization is a common way to avoid the effects of overfitting by adding a significant amount of penalty to the utilized loss function. In this regard, drop-out layer is added in the bottom of proposed network, so that the system does not learn interdependent weights of features.

3.7. Fully Connected Layer.
In the final section of structure of CNNs, neurons of fully connected layer are linked with all the activations of the previous layer to minimalize the feature dimensions. The final pooling layer of the CNNs flattens the convolutional layer, which is forwarded to fully connected nodes of the network. In the next step, the matrix multiplication is applied to compute these activations followed by a bias factor offset. Fully connected neurons can be computed using

$$y_n^l = fl\left(\sum_{m=1}^{N_{l-1}} y_m^{l-1}\, w_{m,n}^l + b_n^l\right), \tag{3}$$

where N_l is the No. of neurons of output-layer, y_m^{l-1} is the m characteristic pattern of layer l-1, and $w_{m,n}^l$ is the connected weights.

3.8. Selection of CNN Model.
In supervised learning, CNN based networks have demonstrated outstanding classification
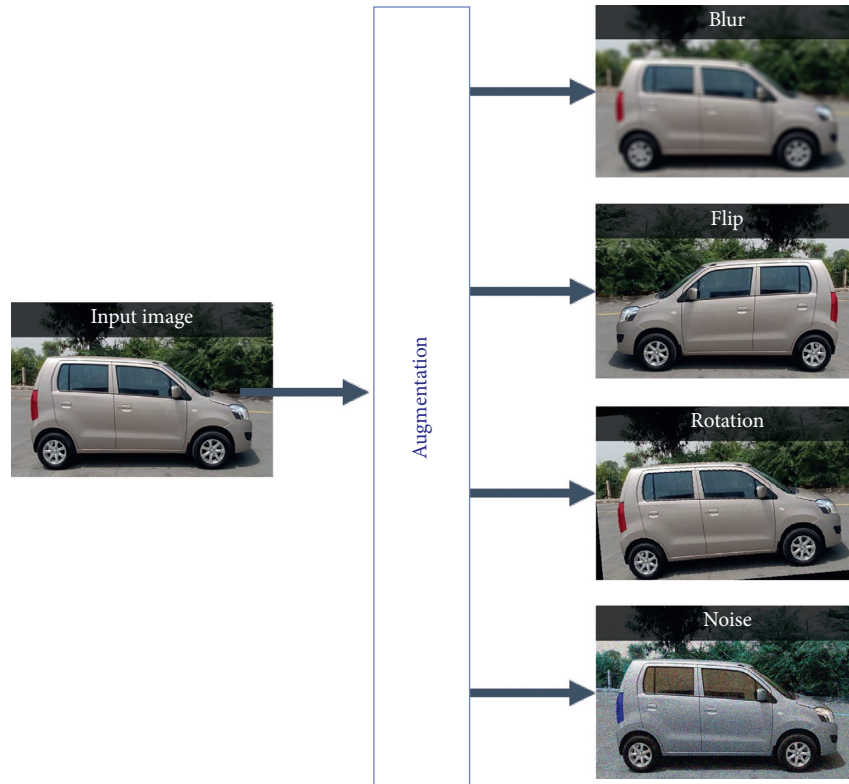
FIGURE 3: Data augmentation performed on the dataset images.

performance on large-scale datasets [38–41]. To choose the suitable CNN model, initially, we fine-tuned the existing state-of-the-art AlexNet, Inception-v3, GoogleNet, VGG, and ResNet models according to the classes of the collected dataset. In the next step, transfer learning is applied to these models to evaluate the self-constructed vehicle dataset. Resultantly, ResNet demonstrated better applicability in terms of convergence, response-time, and accuracy than the competitive networks (briefly discussed in Section 4). Consequently, the network architecture of the ResNet with 152 layers is improved and employed in the proposed vehicle classification system.

4. The Architecture

In the proposed system, we have employed ResNet architecture to perform vehicle classification, one of the most ground-breaking CNN architectures proposed by He et al. [34], which demonstrated outstanding performance in object recognition and classification by securing first place in ILSVRC-15 with 3.57% Top 5 error-rate [34]. In the previous deep learning networks, increasing network layers can cause a vanishing gradient problem, due to which the model was unable to converge at its best. In ResNet network architecture, a novel skip connection-based technique was introduced, where each input from the previous layer is accumulated to output of next layer. Since network goes deeper, a bottleneck design was also adopted to mitigate time-complexity of this CNN architecture. We have employed a transfer learning approach where a model trained for some specific task can be tuned to perform another task by simply learning new weights. This approach can be effective if

we have a lower amount of data, which is insufficient for training from scratch.

In this work, we have deployed a pretrained ResNet-152 network for vehicle classification, as shown in Table 1. This network has a depth of 152 layers, which was achieved by replacing each 2-layer block in the original ResNet with the 3-layer bottleneck block [34]. The input layer of this network takes an RGB colour image with a size of 224×224 pixels. In Table 1, it can be observed that the structure of the presented method uses 64 convolution kernels of 7×7 with the stride of 2 in the first layer, and max-pooling layer of 3×3 with the stride of 2 is used to the first convention layer. Further, convolutional blocks, i.e., 2–5, are organized in the form of three-layer bottleneck blocks having several filters to 128, 256, 1024, and 2048 followed by an adaptive-average pooling layer, respectively. To perform transfer learning, last fully connected layer was removed from the network, which was pretrained to perform the classification of 1000 natural categories. Besides, we append a new classification block consisting of a fully connected layer having a feature vector of 1024 neurons superseded by the average-pooling layer and the ReLU layer to learn new visual features from the training dataset. In the bottom of network, drop-out layer is in place to overcome vanishing gradient problem. Based on the classification block, a new fully connected layer is inserted to perform six types of vehicle classification, where each unit in the last layer is linked to six-class output probability by utilizing softmax function. To ensure that these new layers learn higher-level visual features from the dataset, we have increased the learning rate of these layers as

compared to the previous layers whose learn rate remains unchanged. We have set batch-size and total epochs to 64 and 100, respectively. Network training was performed on a heavy computing machine equipped with the RTX 2080TI, 11 GB DDR5 GPU, core i9 – 9900k CPU along with 32 GB RAM, which took 8 hours to complete training.

5. Experiments and Results

The proposed vehicle classification method is assessed on the dataset-based platform setup. The experiments are performed on the heavy computing machine equipped with the RTX 2080TI, 11 GB DDR5 GPU, core i9 – 9900k CPU along with 32 GB RAM having a 64 bit windows 10 operating system.

5.1. Training of the Proposed Classification System. The whole training process is categorized into three steps: (i) data pre-processing, (ii) training, and (iii) evaluation. In the first step, the dataset images are distributed into training, validation, and testing data, normalized to the size of 224×224 according to standard input size of CNN architectures. The training and testing images are randomly split by an 80–20% ratio of the total dataset images, and the validation set is formed by a random selection of 20% images from the training set. Pytorch 1.4.0 library and MATLAB 2019a are utilized for the implementation (i.e., data preprocessing and organization, training, evaluation, and modification of the network) of the proposed classification system. The experiments have been categorized into three types, (i) evaluation of the networks without fine-tuning, (ii) evaluation of fine-tuned model on self-constructed vehicle dataset, and (iii) evaluation of fine-tuned model on public VeRi dataset, which are briefly discussed below.

5.2. Evaluation of the State-of-the-Art CNNs without Fine-Tuning. To evaluate the CNNs, AlexNet, Inception-v3, GoogleNet, VGG, and ResNet are loaded from Pytorch resources. The training of these networks is performed using the Pytorch framework; a stochastic gradient descent (SGD) optimizer is employed for the parameter learning with momentum, learning rate, and batch size of 0.9, 0.001, and 128, respectively. Cross-entropy, a commonly used loss function, is utilized to accumulate loss during the whole process, and validation is performed after every epoch to evaluate the learning while training the network. The comparative accuracy of these networks is shown in Figure 4.

Discussion: It can be observed from Figure 4 that ResNet with 152 layers demonstrated better accuracy than the 19-layer VGG, 22-layer GoogLeNet, AlexNet having 25 layers, and Inception-v3 with an average difference of 1.7%. Consequently, it can be assumed that the ResNet can achieve better accuracy after fine-tuning the architecture.

5.3. Evaluation of the Modified Network on Self-Constructed Vehicle Dataset. Based on the performance of the networks, discussed in the above section, the architecture of the ResNet is improved by adding a new classification block in the base of the network. The new classification block comprises fully connected layers, followed by an average-pooling and Relu layers, respectively. However, to find the best fitting feature vectors of fully connected layers, ResNet with 152 layers has been evaluated on a self-constructed dataset with the multiple combinations of the feature vectors of fully connected layers in classification block to improve robustness of network.

To apply transfer learning, feature extraction is set to the newly added classification block to learn the optimal weights and biases from the input dataset. SGD optimizer along with the same parameters, i.e., learning rate and momentum, is utilized in the training and evaluation of the proposed classification system. It can be observed from Table 2 that the proposed network with two fully connected containing higher feature vectors achieved significantly higher accuracy among other FC layers with the low feature vectors. Fully connected layer-1 is passing 2048 features down the classification block. Concurrently, the other FC layer with the higher feature vectors has been added to the classification block of the network.

In the next step, the proposed network with the different depth layers, i.e., 18, 34, 50, 101, and 152, has been evaluated, and the performance of these networks in terms of accuracy has been shown in Table 3.

Discussion: Table 3 demonstrates the influence of the network depth on the performance in terms of the accuracy of the self-constructed vehicle dataset. It can be seen that the performance of the ResNet is increased with an increase in the depth of the network. Consequently, ResNet with the 152 layers achieved better accuracy in overall classes of the dataset as compared with the ResNet with lower depth layers. The detailed performance matrices of the ResNet with 152 layers are shown in Table 4.

5.4. Evaluation of the Modified CNN on VeRi Dataset. Based on the performance of the fine-tuned networks demonstrated in Table 4, the proposed classification network has been fine-tuned on the public VeRi dataset [42, 43] to ensure generalization. The dataset involving 50,000 images has been categorized into six classes, i.e., bus, MPV, pickup, sedan, truck, and van, distributed into test and train set over 80 : 20 ratio. It is important to mention that these classes are selected based on the variation in data. The performance matrices shown in Table 5 elaborate the effectiveness of the presented classification system.

5.5. Comparison with Existing State-of-the-Art Vehicle Classification Methods. The performance of the presented classification method is compared with the traditional vehicle classification methods to prove the applicability of the proposed system in terms of class-wise and average accuracy, as shown in Table 6. The existing classification systems [11, 44–46] have been implemented in the MATLAB 2019a, which has been trained and evaluated on the self-constructed vehicle dataset.

Discussion. The proposed classification system has been compared with the existing vehicle classification systems [11, 44–46] to validate the efficacy of the proposed network. The existing networks have been reproduced on the proposed dataset. Zhuo et al. [44] have presented the GoogleNet

TABLE 1: Architecture of modified CNN.

| Layer name	Output size	Layers
Conv 1	112 × 112	Kernel size = 7 × 7, number of kernels = 64, stride = 2
Pooling	56 × 56	Kernel size = 3 × 3, stride = 2
Conv 2	56 × 56	$\begin{bmatrix} 1x1 & 64 \\ 3x3 & 64 \\ 1x1 & 256 \end{bmatrix} \times 3$
Conv 3	28 × 28	$\begin{bmatrix} 1x1 & 64 \\ 3x3 & 64 \\ 1x1 & 256 \end{bmatrix} \times 3$
Conv 4	14 × 14	$\begin{bmatrix} 1x1 & 64 \\ 3x3 & 64 \\ 1x1 & 256 \end{bmatrix} \times 3$
Conv 5	7 × 7	$\begin{bmatrix} 1x1 & 64 \\ 3x3 & 64 \\ 1x1 & 256 \end{bmatrix} \times 3$
Pooling	1 × 1	Adaptive-average-pooling 2d
Proposed classification block		fc1: In-features = 2048, out-feature = 1024 Relu (in-place) drop-out(0.5) fc2: In-features = 1024, out-features = 6 Softmax() Classification output (cross entropy)
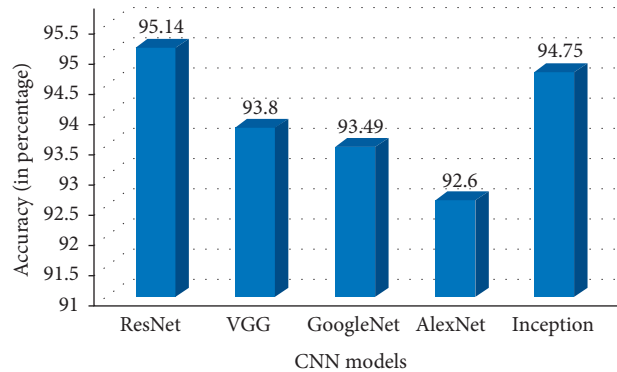
FIGURE 4: Results of the state-of-the-art CNNs without fine-tuning on a self-constructed dataset.

TABLE 2: Classification accuracy of the proposed system with different dimensional feature vectors on self-constructed dataset.

No. of FCs	fc1 out-features	fc2 in-features	fc2 out-features	Accuracy
1	6	—	—	95.14%
2	1536	1536	6	99.74%
2	1024	1024	6	99.68%
2	768	768	6	99.51%

architecture-based vehicle classification method with having 22-layer depth network. On the other side, Gao et al. [45] have introduced AlexNet based vehicle classification system containing 5 convolutional and 3 fully connected layers in the network. Shivai et al. [46] introduced a self-proposed CNN based vehicle classification system having 13 convolutional layers and one fully connected layer, which is followed by max-pooling and dropout layers, whereas Zakria et al. [11] have presented the inception architecture-based classification system. Though these systems demonstrated good performance on their datasets, one of the main reasons

for the difference in the accuracy is that these existing systems [11, 44–46] are comprised of the small depth networks which do not converge on the large-scale datasets. Besides, the existing systems [11, 44–46] are trained on the limited classes, which do not cover the common road traffic vehicles. Resultantly, these systems do not perform well in real-time classification applications. Moreover, it is important to mention here that these methods were trained on unbalanced datasets, which is also an influential factor in real-time performance of vehicle classification systems. Consequently, the performance of these existing systems is

TABLE 3: Class-wise accuracy of the fine-tuned network with different depth layers on our self-constructed dataset.

Depth	Accuracy (in percentage)						Overall accuracy
	Bus	Car	Motorbike	Rickshaw	Truck	Van	
18	95.89	99.68	97.77	92.82	97.37	95.56	96.52
34	97.83	98.17	98.66	95.39	97.13	96.92	97.35
50	97.31	98.93	97.45	94.57	97.04	98.70	97.33
101	98.51	99.02	98.34	96.51	98.93	97.27	98.10
152	99.48	99.68	99.08	100	100	99.65	99.65

TABLE 4: Performance metrics of proposed system on self-constructed dataset.

Class	Accuracy	Error rate	Specificity	Precision	Recall	F1 score
Bus	99.48	0.52	99.90	99.4852	99.13	99.3072
Car	99.68	0.32	99.94	99.6832	99.68	99.6816
Motorbike	99.08	0.92	99.82	99.0832	99.08	99.0816
Rickshaw	100.0	0.00	100.0	100.000	99.83	99.9149
Truck	100.0	0.00	100.0	100.000	99.48	99.7393
Van	99.65	0.35	99.94	99.6525	99.65	99.6512
Total:	**99.65**	**0.35**	**99.93**	**99.65**	**99.48**	**99.56**

TABLE 5: Performance matrices of modified CNN on VeRi dataset.

Class	Accuracy	Error rate	Specificity	Precision	Recall	F1 score
Bus	100.0	0.00	100.0	100.000	98.60	99.29507
MPV	94.19	5.86	99.82	94.1914	96.87	95.51192
Pickup	92.76	7.24	99.53	92.7637	99.04	95.79916
Sedan	99.59	0.41	99.91	99.5971	91.52	95.38787
Truck	96.01	3.99	99.18	96.0118	93.64	94.81107
Van	93.79	6.21	98.73	93.7947	98.35	96.01835
Total:	**96.06**	**3.95**	**99.53**	**96.06**	**96.34**	**96.14**

TABLE 6: Comparison of the proposed method with existing state-of-the-art vehicle classification methods.

Method(s)	Accuracy (in percentage)						Total
	Bus	Car	Motorbike	Rickshaw	Truck	Van	
Zhuo et al. [44]	95.76	94.70	96.67	95.11	95.45	92.25	95.49
Gao et al. [45]	91.78	95.28	97.03	98.77	93.35	91.43	92.61
Shivai et al. [46]	88.17	90.42	91.78	83.37	89.71	89.11	88.96
Zakria et al. [11]	90.55	91.26	97.73	93.36	94.76	88.99	92.77
Proposed method	99.48	99.68	99.08	100.0	100.0	99.65	**99.68**

prejudiced while evaluating our proposed self-constructed balanced dataset, whereas, on the other side, the proposed vehicle classification system is trained on the self-constructed vehicle dataset comprising of 10,000 images, which cover the common road traffic classes, and further fine-tuned on the public VeRi dataset, which contains 50,000 images, to ensure the generalization of the proposed classification system. Consequently, our proposed classification system achieved higher accuracy than the existing vehicle classification systems.

6. Conclusion

In this paper, a CNN-based vehicle classification system is proposed to improve the effectiveness of intelligent transportation systems. A new dataset containing 10,000 images

of six classes is constructed to train classification system. Initially, five state-of-the-art CNNs, i.e., AlexNet, Inception-v3, GoogleNet, VGG, and ResNet, are trained on the collected dataset to validate the performance. Based on the effectiveness, ResNet with the 152 layers is improved by adding a new classification block in the original network through transfer learning. To ensure generalization, the proposed classification system is fine-tuned on public VeRi dataset. Results demonstrate that the proposed classification system achieved higher accuracy, i.e., 99.68% and 97.66%, on self-constructed and VeRi dataset, respectively, which is significantly higher than that of the existing state-of-the-art classification systems. In the future, we are aiming to extend our work to develop fine-grained classification system to improve the effectiveness of proposed method in intelligent transportation systems.

Data Availability

All the data used to support the findings of the study are available in the manuscript.

Disclosure

Muhammad Atif Butt and Asad Masood Khattak are the joint first authors to this work.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

The research conceptualization and methodology were done by Muhammad Atif Butt, Asad Masood Khattak, and Sarmad Shafique. The technical and theoretical framework was prepared by Sarmad Shafique and Saima Abid. The technical review and improvement were performed by Ahthasham Sajid, Muhammad Waqas Ayub, Bashir Hayat, and Awais Adnan. The overall technical support, guidance, and project administration were done by Muhammad Atif Butt, Asad Masood Khattak, Saima Abid, and Ki-Il Kim.

Acknowledgments

References

[1] H. M. Abdelbaki, K. Hussain, and E. Gelenbe, "August. A laser intensity image based automatic vehicle classification system," in *Proceedings of the ITSC 2001. 2001 IEEE Intelligent Transportation Systems*, pp. 460–465, IEEE, Oakland, CA, USA, 2001.

[2] C. Harlow and S. Peng, "Automatic vehicle classification system with range sensors," *Transportation Research Part C: Emerging Technologies*, vol. 9, no. 4, pp. 231–247, 2001.

[3] F. Nashashibi and A. Bargeton, "June. Laser-based vehicles tracking and classification using occlusion reasoning and confidence estimation," in *Proceedings of the 2008 IEEE Intelligent Vehicles Symposium*, pp. 847–852, IEEE, Eindhoven, Netherlands, 2008.

[4] Y.-K. Ki and D.-K. Baik, "Vehicle-classification algorithm for single-loop detectors using neural networks," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 6, pp. 1704–1711, 2006.

[5] H. Shokravi, H. Shokravi, N. Bakhary, M. Heidarrezaei, S. S. Rahimian Koloor, and M. Petrů, "A review on vehicle classification and potential use of smart vehicle-assisted techniques," *Sensors*, vol. 20, no. 11, p. 3274, 2020.

[6] P. K. Bhaskar and S. P. Yong, "June. Image processing-based vehicle detection and tracking method," in *Proceedings of the 2014 International Conference on Computer and Information Sciences (ICCOINS)*, pp. 1–5, IEEE, Kuala Lumpur, Malaysia, 2014.

[7] Y. U. Rong, W. A. N. G. Guoxiang, J. Zheng, and W. A. N. G. Haiyan, "Urban road traffic condition pattern recognition based on support vector machine," *Journal of Transportation Systems Engineering and Information Technology*, vol. 13, no. 1, pp. 130–136, 2013.

[8] L. Nanni, S. Ghidoni, and S. Brahnam, "Handcrafted vs. non-handcrafted features for computer vision classification," *Pattern Recognition*, vol. 71, pp. 158–172, 2017.

[9] F. Fooladgar and S. Kasaei, "A survey on indoor RGB-D semantic segmentation: from hand-crafted features to deep convolutional neural networks," *Multimedia Tools and Applications*, vol. 79, no. 7-8, pp. 4499–4524, 2020.

[10] J. Chang, L. Wang, G. Meng, S. Xiang, and C. Pan, "Vision-based occlusion handling and vehicle classification for traffic surveillance systems," *IEEE Intelligent Transportation Systems Magazine*, vol. 10, no. 2, pp. 80–92, 2018.

[11] J. Cai, J. Deng, M. S. Khokhar, and M. U. Aftab, "December. Vehicle classification based on deep convolutional neural networks model for traffic surveillance systems," in *Proceedings of the 2018 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, pp. 224–227, IEEE, 2018.

[12] M. A. Hedeya, A. H. Eid, and R. F. Abdel-Kader, "A super-learner ensemble of deep networks for vehicle-type classification," IEEE Access, 2020.

[13] 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*, pp. 3973–3981, Seattle, WA, USA, 2015.

[14] 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*, pp. 554–561, Venice, Italy, 2013.

[15] L. T. Ng, S. A. Suandi, and S. S. Teoh, "Vehicle classification using visual background extractor and multi-class support vector machines," in *Proceedings of the the 8th International Conference on Robotic, Vision, Signal Processing \& Power Applications*, pp. 221–227, Springer, Singapore, 2014.

[16] Y. Chen and G. F. Qin, "Video-based vehicle detection and classification in challenging scenarios," *International Journal on Smart Sensing and Intelligent Systems*, vol. 7, no. 3, 2014.

[17] F. M. S. Matos and R. M. C. R. de Souza, "Hierarchical classification of vehicle images using NN with conditional adaptive distance," in *Proceedings of the International Conference on Neural Information Processing*, pp. 745–752, Springer, Berlin, Germany, 2013.

[18] Y. Y. Cui, *Research on Vehicle Recognition in Intelligent Transportation*, University of Electronic Science and Technology of China, Chengdu, China, 2013.

[19] X. Wen, L. Shao, Y. Xue, and W. Fang, "A rapid learning algorithm for vehicle classification," *Information Sciences*, vol. 295, no. 1, pp. 395–406, 2015.

[20] Z. Dong, Y. Wu, M. Pei, and Y. Jia, "Vehicle type classification using a semisupervised convolutional neural network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2247–2256, 2015.

[21] X. Wang, W. Zhang, X. Wu, L. Xiao, Y. Qian, and Z. Fang, "Real-time vehicle type classification with deep convolutional neural networks," *Journal of Real-Time Image Processing*, vol. 16, no. 1, pp. 5–14, 2019.

[22] J. Cao, W. Wang, X. Wang, C. Li, and J. Tang, "October. End-to-End view-aware vehicle classification via progressive CNN learning," in *Proceedings of the CCF Chinese Conference on Computer Vision*, pp. 729–737, Springer, Singapore, 2017.

[23] M. S. Chauhan, A. Singh, M. Khemka, A. Prateek, and R. Sen, "January. Embedded CNN based vehicle classification and

counting in non-laned road traffic," in *Proceedings of the Tenth International Conference on Information and Communication Technologies and Development*, pp. 1–11, Ahmedabad, India, 2019.

[24] S. Y. Jo, N. Ahn, Y. Lee, and S. J. Kang, "November. Transfer learning-based vehicle classification," in *Proceedings of the 2018 International SoC Design Conference (ISOCC)*, pp. 127-128, IEEE, Daegu, South Korea, 2018.

[25] J. Kim, J. Kim, G.-J. Jang, and M. Lee, "Fast learning method for convolutional neural networks using extreme learning machine and its application to lane detection," *Neural Networks*, vol. 87, pp. 109–121, 2017.

[26] Y. Tian, J. Gelernter, X. Wang et al., "Lane marking detection via deep convolutional neural network," *Neurocomputing*, vol. 280, pp. 46–55, 2018.

[27] J. Feng, Y. Wu, and Y. Zhang, "Lane detection base on deep learning," in *Proceedings of the 2018 11th International Symposium on Computational Intelligence and Design (ISCID)*, pp. 315–318, Hangzhou, China, December 2018.

[28] B. D. Brabandere, W. V. Gansbeke, D. Neven, M. Proesmans, and L. V. Gool, "End-to-end lane detection through differentiable least-squares fitting," *Computer Science*, http://arxiv.org/abs/1902.00293, 2019.

[29] L. Jiang, J. Li, L. Zhuo, and Z. Zhu, "August. Robust vehicle classification based on the combination of deep features and handcrafted features," in *Proceedings of the 2017 IEEE Trustcom/BigDataSE/ICESS*, pp. 859–865, IEEE, Sydney, Australia, 2017.

[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 1097–1105, 2012.

[31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, http://arxiv.org/abs/1409.1556.

[32] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, Seattle, WA, USA, 2015.

[33] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818–2826, Seattle, WA, USA, 2016.

[34] 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*, pp. 770–778, Seattle, WA, USA, 2016.

[35] F. Codevilla, E. Santana, A. M. López, and A. Gaidon, "Exploring the limitations of behavior cloning for autonomous driving," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 9329–9338, Corfu, Greece, 2019.

[36] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in *Proceedings of the AAAI*, pp. 13001–13008, 2020.

[37] C. M. Bautista, C. A. Dy, M. I. Mañalac, R. A. Orbe, and M. Cordel, "May. Convolutional neural network for vehicle detection in low resolution traffic videos," in *Proceedings of the 2016 IEEE Region 10 Symposium (TENSYMP)*, pp. 277–281, IEEE, Bali, Indonesia, 2016.

[38] V. K. Kiran, P. Parida, and S. Dash, "Vehicle detection and classification: a review," *Journal of Information Assurance \& Security*, vol. 15, no. 2, 2020.

[39] Y. Du, Y. Yan, S. Chen, and Y. Hua, "Object-adaptive LSTM network for real-time visual tracking with adversarial data augmentation," *Neurocomputing*, vol. 384, pp. 67–83, 2020.

[40] T. H. Vu, J. Boonaert, S. Ambellouis, and A. T. Ahmed, "February. Vehicles tracking by combining convolutional neural network based segmentation and optical flow estimation," in *Proceedings of the International Conference on Advanced Concepts for Intelligent Vision Systems*, pp. 529–540, Springer, Cham, Switzerland, 2020.

[41] L. Mou, H. Xie, S. Mao, P. Zhao, and Y. Chen, "Vision-based vehicle behaviour analysis: a structured learning approach via convolutional neural networks," *IET Intelligent Transport Systems*, 2020.

[42] X. Liu, W. Liu, T. Mei, and H. Ma, "October. A deep learning-based approach to progressive vehicle re-identification for urban surveillance," in *Proceedings of the European Conference on Computer Vision*, pp. 869–884, Springer, Cham, Switzerland, 2016.

[43] X. Liu, W. Liu, T. Mei, and H. Ma, "Provid: progressive and multimodal vehicle reidentification for large-scale urban surveillance," *IEEE Transactions on Multimedia*, vol. 20, no. 3, pp. 645–658, 2017.

[44] L. Zhuo, L. Jiang, Z. Zhu, J. Li, J. Zhang, and H. Long, "Vehicle classification for large-scale traffic surveillance videos using convolutional neural networks," *Machine Vision and Applications*, vol. 28, no. 7, pp. 793–802, 2017.

[45] H. Gao, B. Cheng, J. Wang, K. Li, J. Zhao, and D. Li, "Object classification using CNN-based fusion of vision and LIDAR in autonomous vehicle environment," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 9, pp. 4224–4231, 2018.

[46] N. Shvai, A. Hasnat, A. Meicler, and A. Nakib, "Accurate classification for automatic vehicle-type recognition based on ensemble classifiers," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 1288–1297, 2019.