



**Volume 112**

**2021**

p-ISSN: 0209-3324

e-ISSN: 2450-1549

DOI: <https://doi.org/10.20858/sjsutst.2021.112.16>



**Silesian  
University  
of Technology**

Journal homepage: <http://sjsutst.polsl.pl>

---

**Article citation information:**

Trivedi, J., Devi, M.S., Dhara, D. Vehicle classification using the convolution neural network approach. *Scientific Journal of Silesian University of Technology. Series Transport*. 2021, **112**, 201-209. ISSN: 0209-3324. DOI: <https://doi.org/10.20858/sjsutst.2021.112.7.16>

**Janak TRIVEDI<sup>1</sup>, Mandalapu Sarada DEVI<sup>2</sup>, Dave DHARA<sup>3</sup>**

---

## **VEHICLE CLASSIFICATION USING THE CONVOLUTION NEURAL NETWORK APPROACH**

**Summary.** We present vehicle detection classification using the Convolution Neural Network (CNN) of the deep learning approach. The automatic vehicle classification for traffic surveillance video systems is challenging for the Intelligent Transportation System (ITS) to build a smart city. In this article, three different vehicles: bike, car and truck classification are considered for around 3,000 bikes, 6,000 cars, and 2,000 images of trucks. CNN can automatically absorb and extract different vehicle dataset's different features without a manual selection of features. The accuracy of CNN is measured in terms of the confidence values of the detected object. The highest confidence value is about 0.99 in the case of the bike category vehicle classification. The automatic vehicle classification supports building an electronic toll collection system and identifying emergency vehicles in the traffic.

**Keywords:** convolution neural network, vehicle classification, deep learning, intelligent transportation system

---

<sup>1</sup> Faculty of Electronics & Communication Engineering Department, Gujarat Technological University, Government Engineering College Bhavnagar-364002, Gujarat, India. Email: Trivedi\_janak2611@yahoo.com. ORCID: 0000-0002-8662-5153

<sup>2</sup> Principal, Ahmedabad Institute of Technology-380060, Gujarat Technological University, Gujarat, India. Email: saradadevim1@gmail.com. ORCID: 0000-0003-4904-1906

<sup>3</sup> Faculty of Electronics & Communication Engineering Department, Gujarat Technological University, Government Engineering College Bhavnagar-364002, Gujarat, India. Email: dave.dhara24888@gmail.com. ORCID: 0000-0002-7724-800X

## 1. INTRODUCTION

Vehicle classification, vehicle speed measurements, electronics toll connection system, vehicle counting, and vehicle recognition system are numerous ITS applications. The vehicle classification system helps to build an electronic toll collection system and vehicle recognition system in a smart city. Automatic vehicle classification plays a central role in developing an intelligent system traffic management system. The automatic extraction of a vehicle from a surveillance video is a hot research topic for a smart city's traffic organisation.

The practice of CCTV camera systems helps in managing traffic control in metropolitan areas. The vehicle classification can be done in two ways: using the human operator and the computer-vision system. Computer vision-based traffic management is more innovative and efficient compared to human operators.

Deep learning and machine learning revealed a new direction with high accuracy for an automatic vehicle recognition system in the field of computer-vision techniques. The well-known Convolution Neural Network (CNN) application, which is part of deep learning, automatically helps classify vehicles using image processing techniques. Deep learning works similarly like the human brain for the proposed system [9]. The layer-by-layer structure in deep learning extracts valuable information for the given dataset.

### 1.1. Motivation

The image processing techniques manage traffic control systems without prior installation of different sensors on the road. Vehicle classification is still a challenge due to meagre processing time for real-time applications, different resolution of images, illumination changes in the video, and other disturbances. There is main problems in vehicle classification using the traditional method – object's classification requires human supervision. The images are location variants for the defined method.

CNN can overcome these limitations for object classification. The recent development in the area of CNN boosts the research direction in vehicle detection, identification, and classification. Vehicle classification is applicable in numerous applications, including robotics, automatic toll collection system, and recognises the emergency vehicle. The classification of a vehicle can be done online, offline, and in a real-time application.

The remaining part of this paper is organised as follows: Section two reports the literature survey. Section three presents CNN using the deep learning approach with different layer's explanations. Next, section four is devoted to experimental results. Finally, section five presents the discussion and conclusion and the future scope is drawn in section six.

## 2. LITERATURE SURVEY

The recent advancement of the deep learning approach using the CNN structure simplifies the use of ITS in traffic management. The traffic surveillance video with low-resolution images suffers low accuracy in its output results. The results are varied for daytime and nighttime, controlled environment conditions and uncontrolled environmental conditions. The Support Vector Machine (SVM) and Artificial Neural Network (ANN) both have registered low accuracy for the vehicle classification approach, according to Bautista et al. [3]. CNN has a high accuracy rate for vehicle recognition compared to SVM and ANN. Bautista et al. [3] explained CNN-based vehicle detection and classification using high-resolution images.

Huseyin et al. [5] discussed detecting, tracking, and classification of multiple moving objects in real-time. The image registration is used to align the two images that determine the key point of the image. For increasing the speed of operation, greyscale images are used in the image registration steps. A similar kind of operation is used in this article before resizing the original images. The moving object detection using the frame differencing method between two consecutive frames creates a problem with the image registration process. However, this can be solved using different filtering operations, like median filtering and mean filtering. The vehicles' classification has been demonstrated with the help of a deep learning-based CNN approach and a threshold value. The image was first resized to 227 x 227 to match the network layer.

The deep neural network approach-based vehicle detection is explained by Jun et al. [6]. The two light-weighted CNN structures, one to propose Region of Interest (ROI) and the other to make classification, are used. The input image size is 32 x 32 to light the CNN structure. Tang et al. [10] demonstrated vehicle detection using state of art method like YOLOv2.

Bing et al. [2] discussed object detection for the surveillance video using the CNN structure used in traceability applications. The system works in two-phase: target classification and target location. Zhou et al. [12] demonstrated multiple vehicle detection using a combination of two neural networks. One network detects vehicles with their type, and the other network predicts vehicle speed with the vehicle location in a real-time condition.

Suhao et al. [7] outlined three types of vehicle detection and classification using the deep learning method CNN. Ajeet et al. [1] highlighted different image datasets used in object detection and classification. The comparison of deep learning-based object detection methods with their features is explained. The different deep learning frameworks with their interface, features, and deep learning model are discussed with the developer and licence.

Zhao et al. [13] considered the vehicle detection method using CNN of the deep learning approach and multi-layer feature fusion. This method effectively detects vehicles and matches the real-time level speed, which helps in smart driving cars. Vehicle detection in a traffic-congested scene is not enough using this method, so it requires further study to improve this method. The input image size in this CNN structure is 312 x 312. Wenming et al. [11] developed a fast deep neural network for real-time video object detection; identify the ROI which contains a target vehicle with high confidence.

### 3. METHOD DESCRIPTION

In this article, three different vehicle classification (bike, car, and truck) and recognition was done using a deep learning-based CNN approach. The key procedure for vehicle classification using CNN can be summarised into four steps: define each image to network, prepare a layer, train, and create a network, test network with accuracy. In the proposed method, we involved only images of bikes, cars, and trucks. To develop a network using deep learning, we have collected around 3,000 images of bikes, 6,000 images of cars, and 2,000 images of trucks. Thus, a total of 11,000 images were used to classify and recognise three different categories of objects. The different images have been simulated to verify the proposed method. The training of a network depends on the network speed and processing power.

### 3.1. Convolution Neural Network

CNN is a type of neural network that is usually applied to image processing problems. The different CNN applications including identified any object in an image, driverless car, robotic. They are helpful for fast-growing areas, and that is also one of the main important factors for deep learning and artificial intelligence today. CNN is unique and different from the simple Neural Networks (NN). It is effective for processing and classifying images.

A simple NN has an input layer, a hidden layer, and an output layer. The input layers accept input in different forms, while the hidden layers perform calculations on these inputs. The output layer then delivers the outcome of the calculations and extractions. Each of these layers contains neurons connected to neurons in the previous layer. Each neuron has its weight, that is, no assumption is made about the data being fed into the network. This approach is not suitable for images or language.

CNN works treat data as spatial. Instead of neurons being connected to every neuron in the previous layer, they are only connected to the neurons close to it, and all have the same weight. The simplification in the connections means the network upholds the spatial aspects of the data set. The word 'Convolutional' refers to the filtering process that happens in this type of network. If an image is complex, like too many objects in one image, and thus challenging to identify the object, CNN can simplify it. Therefore, it can be better processed and understood. Like the simple NN, CNN is made up of multiple layers. There are different layers of CNN, which makes it unique. This is explained in section 3.2.

The flowchart of the proposed method is shown in Figure 1. Every neural network has a series of layers; the more layers it has, the more profound the network. Each layer takes in data from the previous layers, transforms the data, and then processes it. Thus, the first layer takes in the raw input image, and by the time it gets to the last layer, it will likely pick out the correct name of the vehicle in the original image.

### 3.2. Different layers

Here, the described layers learn and train using the Matlab software, and detail help is given in <https://www.mathworks.com/> [8].

(1) *Input image layer*: The input image layer specifies input image size properties. In this article, 28x28x1 (height, width, and a number of a channel) image size is used. The zero centre normalisation, which subtracts the mean, is used.

(2) *Convolution layer*: This layer is the most important. It works by placing a filter over an array of image pixels. Convolution layer with 3x3 filter size with stride [1 1] and padding [1 1 1]. Stride represents the filter's movements over the images, and padding '1' means several pixels added to the CNN. This process creates a convolved feature map.

(3) *Relu layer*: Relu layers acts as an activation function. A Rectified Liner Unit (Relu) carries out threshold operation to the input image element. When any value is less than zero, then those values are set to zero themselves.

(4) *Maxpool layer*: The pooling layer down samples or reduces the sample size. In this article, a maxpool layer with stride [2 2] and padding [0 0 0 0] is used, which makes processing much faster as it reduces the number of parameters the network needs to process.

(5) *Fully connected layer:* The fully connected layers allow us to perform classification on the dataset. A fully connected layer takes values from the relu layer process by multiplying the input by weight of matrix and then adding a bias vector. The fully connected layer with output size three is used for the classification of three different objects.

(6) *Softmax layer:* The softmax layer applies a softmax function to the input. The softmax transformation of a linear function was explained by C. M. Bishop [4] as shown in equation 1, where  $y_r(x)$  is a softmax function,  $a_r(x)$  is a linear function of  $x$ .

$$y_r(x) = \frac{\exp(a_r(x))}{\sum_{j=1}^k \exp(a_j(x))}, \quad 0 \leq y_r(x) \leq 1, \quad \text{and} \quad \sum_{j=1}^k y_j = 1 \quad (1)$$

Softmax is applied after the final fully connected layer and before the last classification layer.

(7) *Classification layer:* The classification layer is the last layer for multi-class classification problems. The last layer in the CNN indicates the number of classes from the previous layer computation. The classification layer follows the softmax layer. Classification layers use equation 2 from C. M. Bishop [4] book as mentioned below,

$$\text{Loss} = -\sum_{i=1}^N \sum_{j=1}^k t_{ij} \ln(y_{ij}), \quad (2)$$

where  $K$  and  $N$  indicate the number of classes and samples, respectively.

The first step is to separate and store each collected image with different labels in the system. The stored images have different sizes, so the next step is to resize all the images into 28 x 28 image size and convert all colour images into grey images. Thereafter, save this database for further processing.

As shown in Figure 1, twelve different layers are used to perform CNN for vehicle classification for the collected data and classify it into three different objects: bike, car, and truck. The twelve different layers include three convolution layers and relu layers, two maxpool layers, one image-input layer, a fully connected layer, a softmax layer, and finally, a classification layer after completing the CNN layers for classification of the different vehicles, next trained the network and test the network using different images for the three different vehicles.

#### 4. EXPERIMENTAL RESULT

The result section indicates recognition of the image and accuracy in terms of confidence values concerning the other two objects. The confidence values indicate the matching percentage to the respective objects. In Table 1, the highest confidence value of the bike image is about 99%, and the Figure with a recognition graph is shown in Figure 3a,b). The car-type vehicle is shown in Figure 2a, and its confidence value to the car, bike, and truck is shown in Figure 2b. The car-type vehicle's confidence value is above 85%, as shown in Figure 2b. The truck-type vehicle is shown in Figure 4a, and its confidence value to the car, bike, and truck is shown in Figure 4b.

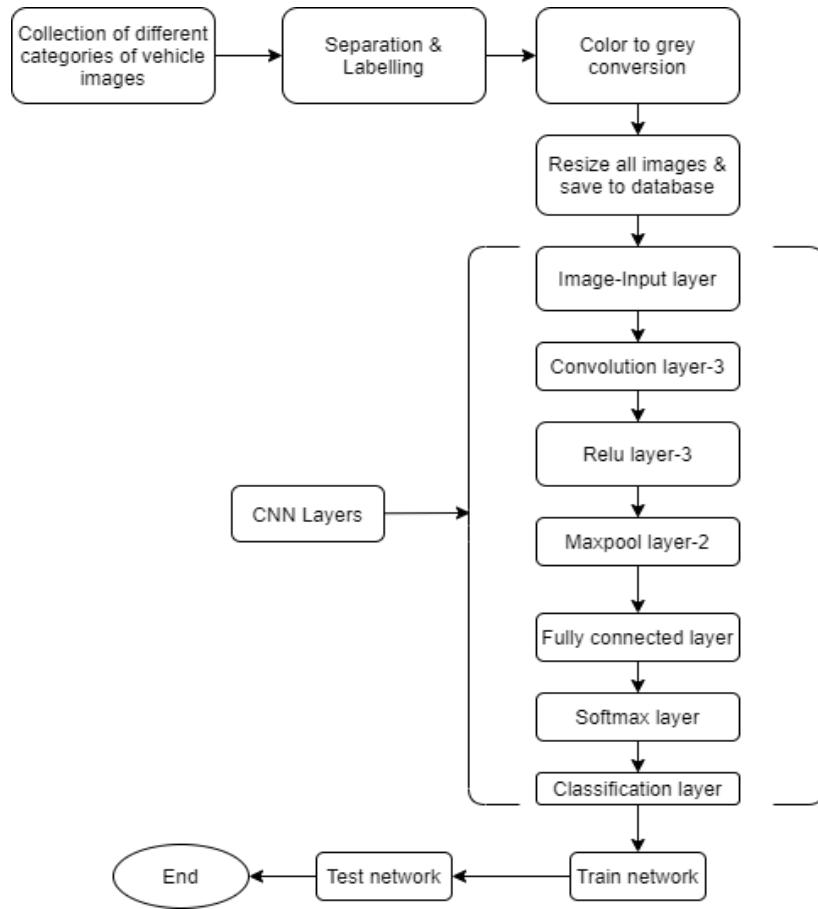


Fig. 1. Vehicle classification using CNN

Table 1 represents different images of bikes, and they recognise with higher accuracy compared to cars and trucks.

Tab. 1.  
Images of bikes and they recognise with higher accuracy compared to cars and trucks

Images	Bike	Car	Truck
1	0.88	0.1	0.02
2	0.92	0.07	0.01
3	0.86	0.13	0.01
4	0.84	0.12	0.04
5	0.87	0.12	0.01
6	0.99	0.01	0
7	0.93	0.06	0.01
8	0.92	0.07	0.01
9	0.86	0.1	0.04
10	0.92	0.06	0.02

## 5. DISCUSSION

The three types of vehicles: bike, car, and truck, classify and recognise using the deep learning-based CNN approach. CNN has a specific advantage over traditional vehicle classification methods. In CNN, the different layers share the relevant information, which invariance to a scalar, rotation, geometric transformation, and illumination changes. This approach can identify a fault and re-identify the object with another training of the network. The accuracy of the CNN can be improved with many image datasets with different conditions of images.

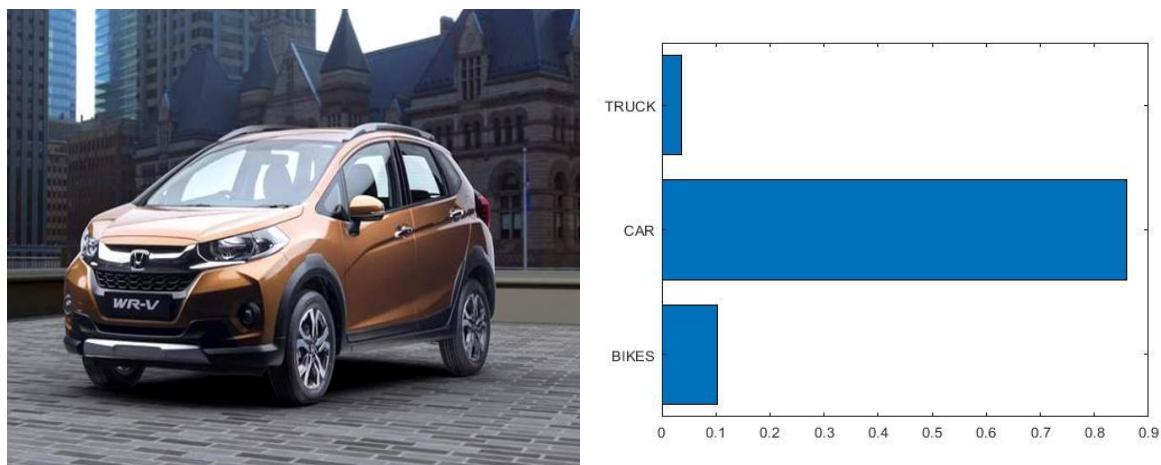


Fig. 2. (a) The car, (b) The car recognises with higher accuracy compared to truck and bike

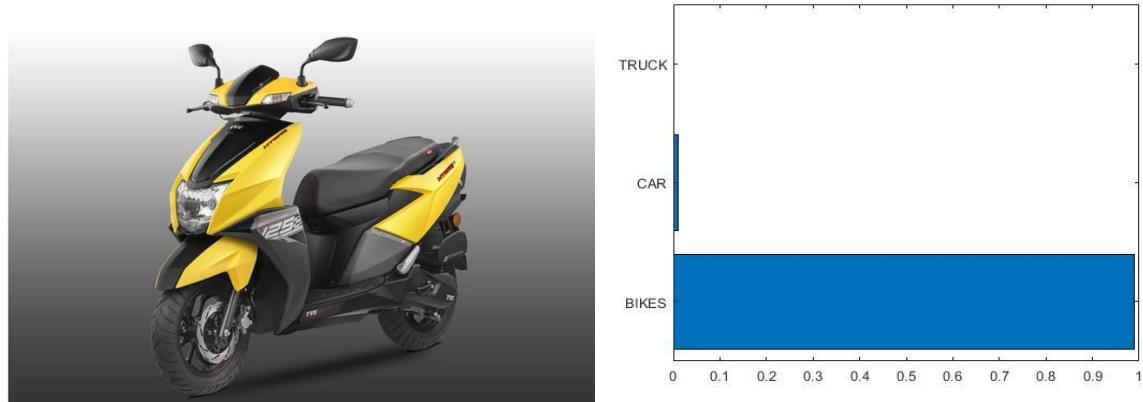


Fig. 3. (a) The bike, (b) The bike recognises with higher accuracy compared to car and truck

It first categorised data into bike, car, and truck using a deep learning approach. The image dataset then resizes to 680 x 480 pixels. The larger and lower size of the image can affect the validation process. The higher image dataset's size requires a long computational time compared to the image dataset's smaller size. The image size should be such that it requires low computational time as well as high recognition accuracy. Afterwards, apply different layers of convolution neural layer to the desired image datasets.

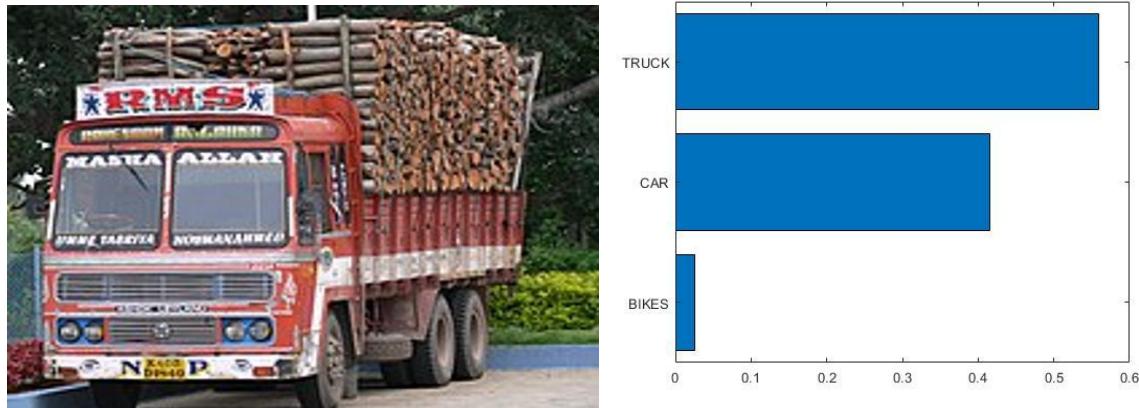


Fig. 4. (a) The truck, (b) The truck recognises with higher accuracy compared to bike and car.

In this article, about 11,000 images for the three types of vehicles: bike, car, and truck are trained using CNN. The highest confidence value is found in the case of a bike. The simulation result of the CNN approach is summarised in Table 1. From around 3,500 images of bikes, simulated 10 bikes images, the result is verified using the confidence matrix concerning the car and truck. Image no. six in Table 1 has a confidence value of 0.99 for bike prediction, 0.01 for car prediction, and 0.0 for truck prediction. This indicates the accuracy of simulation for bike prediction near 99%.

Similarly, from around 5,800 images of cars, we randomly checked the simulation result of 'car.' We identified 'car' as 'car' with the highest confidence values around 85%. From around 1,800 images of trucks, we identified 'truck' as 'truck' with the highest confidence values around '55%' .

## 6. CONCLUSION

In this paper, we presented a fast and robust vehicle classification using a deep learning-based CNN approach. This approach orders vehicles into three categories: bike, car, and truck. The accuracy of the classified object is measured using confidence values for selected images. The deep learning approach using CNN is the backbone for vehicle classification. The manual labelling for vehicle identification is time-consuming, laborious work that can be avoided using deep learning CNN.

Accordingly, a higher amount of efficiency is required to accept ITS in a traditional traffic management system. The bike is recognised accurately with the current image datasets. The car is recognised with higher accuracy than the truck but with lower accuracy than the bike. The least accuracy we received is the case of the truck. The truck recognition result is low due to the tested dataset containing a meagre amount of truck images compared to bike and car. We will try to overcome this limitation in our future work. CNN is less accurate for a smaller data training set but presents higher accuracy with an extensive image dataset.

## References

1. Ajeet Ram Pathak, Manjusha Pandey, Siddharth Rautaray. 2018. „Application of Deep Learning for Object Detection”. *Procedia Computer Science*: 1706-1717. DOI: 10.1016/j.procs.2018.05.144.
2. Bing Tian, Liang Li, Yansheng Qu, Li Yan. 2017. „Video Object Detection for Tractability with Deep Learning Method”. *IEEE Computer Society*: 397-401. DOI: 10.1109/CBD.2017.75.
3. Carlo Migel Bautista, Clifford Austin Dy, Miguel Iñigo Mañalac, Raphael Angelo Orbe, Macario Cordel. 2016. „Convolutional neural network for vehicle detection in low-resolution traffic videos”. *IEEE Region 10 Symposium*: 277-281. DOI: 10.1109/TENCONSpring.2016.7519418.
4. Christopher M. Bishop. 2006. *Pattern recognition and machine learning*. Springer. ISBN: 0-387-31073-8.
5. Hüseyin Can Baykara, Erdem Büyükkaya, Gamze Güler, Deniz Onural, Ahmet Safa Öztürk, Ilkay Yıldız. 2017. „Real-Time Detection, Tracking and Classification of Multiple Moving Objects in UAV Videos”. *IEEE Computer Society*: 945-950. DOI: 10.1109/ICTAI.2017.00145.
6. Jun Hu, Wei Liu, Huai Yuan, Hong Zhao. 2017. „A Multi-View Vehicle Detection Method Based on Deep Neural Networks”. *9th Intr. Conf. on Measuring Technology and Mechatronics Automation*: 86-89. DOI: 10.1109/ICMTMA.2017.27.
7. Li Suhaao, Lin Jinzhao, Li Guoquan, Bai Tong, Wang Huiqian, Pang Yu. 2018. „Vehicle type detection based on deep learning in traffic scene”. *Procedia Computer Science*: 564-572. DOI: 10.1016/j.procs.2018.04.281.
8. Mathworks. Available at: <http://www.mathworks.com>.
9. Staniak Marcin, Czech Piotr. 2016. “Self-correcting neural network in road pavement diagnostics”. *Automation in Construction* 96: 75-87. DOI: 10.1016/j.autcon.2018.09.001.
10. Tianyu Tang, Zhipeng Deng, Shilin Zhou, Lin Lei, Huanxin Zou. 2017. „Fast Vehicle Detection in UAV Images”. *International Workshop on Remote Sensing with Intelligent Processing (RSIP)*. DOI: 10.1109/RSIP.2017.7958795.
11. Wenming Cao, Jianhe Yuan, Zhihai He, Zhi Zhang, Zhiqian He. 2018. „Fast Deep Neural Networks with Knowledge Guided Training and Predicted Regions of Interests for Real-Time Video Object Detection”. *IEE Access*: 8990-8999. DOI: 10.1109/ACCESS.2018.2795798.
12. Yi Zhou, Li Liu, Ling Shao. 2017. „Fast Automatic Vehicle Annotation for Urban Traffic Surveillance”. *IEEE Trans. On Intelligent Transportation Systems*: 1-12. DOI: 10.1109/TITS.2017.2740303.
13. Zhao Min, Jia Jian, Sun Dihua, Tang Yi. 2018. „Vehicle Detection Method Based On Deep Learning and Multi-Layer Feature Fusion”. *30th Chinese Control & Decision Conference*: 5862-5867. DOI: 10.1109/CCDC.2018.8408156.

Received 18.03.2021; accepted in revised form 19.06.2021



Scientific Journal of Silesian University of Technology. Series Transport is licensed under a Creative Commons Attribution 4.0 International License