International Journal of Engineering

Journal Homepage: www.ije.ir

# Traffic Scene Analysis and Classification using Deep Learning

Z. Dorrani*

*Department of Electrical Engineering, Payame Noor University (PNU), Tehran, Iran*

*PAPER INFO*

*Paper history:*
Received 28 August 2023
Received in revised form 09 October 2023
Accepted 12 November 2023

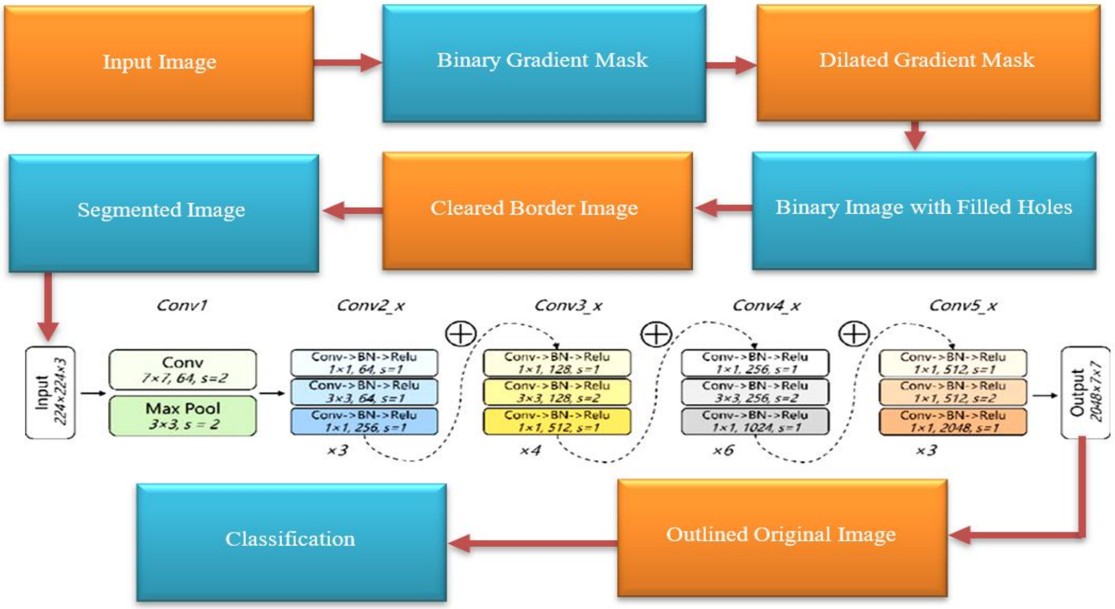
*Keywords*:
*Classification*
*Convolutional Neural Network*
*Deep Learning*
*ResNeXt*
*Traffic Scene Analysis*

*ABSTRACT*

In this study, we aim to use new deep-learning tools and convolutional neural networks for traffic analysis. ResNeXt architecture, one of the most potent architectures and has attracted much attention in various fields, has been proposed to examine the scene, and classify it into three categories: cars, bikes (bicycles/motorcycles), and pedestrians. Previous studies have focused more on one type of classification and reported only human-facial recognition or vehicle detection. In contrast, the proposed method uses precise architecture to perform the classification of three classes. The proposed plan has been implemented in several steps: the first stage is to divide the critical objects. In the next step, the characteristics of the obtained objects are extracted to classify the process into three classes. Experiments have been conducted on different and essential datasets such as high-traffic, low-quality, real-time scenes. Essential evaluation criteria such as accuracy, sensitivity, and specificity show that the performance of the proposed method has improved compared to the methods being compared. The accuracy criterion reached more than 92%, sensitivity about 89%, and specially to 90.25%. The proposed method can be used to implement intelligent cities, public safety, and metropolitan decisions and use the results in urban management, predictive modeling of lost data management, sequential data management, and generalizability.

***doi****: 10.5829/ije.2024.37.03c.06*

**Graphical Abstract**


Input Image
Binary Gradient Mask
Dilated Gradient Mask
Binary Image with Filled Holes
Cleared Border Image
Segmented Image
Conv1
Conv2_x
Conv3_x
Conv4_x
Conv5_x
Input 224×224×3
Conv 7×7, 64, s=2
Max Pool 3×3, s = 2
Conv->BN->Relu 1×1, 64, s=1
Conv->BN->Relu 3×3, 64, s=1
Conv->BN->Relu 1×1, 256, s=1
×3
Conv->BN->Relu 1×1, 128, s=1
Conv->BN->Relu 3×3, 128, s=2
Conv->BN->Relu 1×1, 512, s=1
×4
Conv->BN->Relu 1×1, 256, s=1
Conv->BN->Relu 3×3, 256, s=2
Conv->BN->Relu 1×1, 1024, s=1
×6
Conv->BN->Relu 1×1, 512, s=1
Conv->BN->Relu 1×1, 512, s=2
Conv->BN->Relu 1×1, 2048, s=1
×3
Output 2048×7×7
Outlined Original Image
Classification

*Corresponding Author Email: dorrani.z@pnu.ac.ir (Z. Dorrani)

Please cite this article as: Dorrani Z. Traffic Scene Analysis and Classification using Deep Learning. International Journal of Engineering, Transactions C: Aspects. 2024;37(03):496-502.

**NOMENCLATURE**

| | | | |
|---|---|---|---|
| TP | true positive samples | **Greek Symbols** | |
| TN | true negative samples | η | Specially |
| FP | false positive samples | *φ* | Sensitivity |
| FN | false negative samples | | |

## 1. INTRODUCTION

The goal of machine vision is to imitate the capabilities of the human vision system (1). Computers do not have unique minds like humans, and human's previous knowledge makes them superior in this comparison. The significant weakness of the machine's vision compared to humans is the lack of extensive information and its flexibility in the real world. There are different ways to solve this challenge. One of these methods is deep learning based on Convolutional Neural Network (CNN) (2), which performs very well, especially for classification, and can reduce the error rate. The CNN can achieve end-to-end supervised training, without the need for unsupervised pre-training (3), and its various architectures have become the best tools for real-time object classification from image data (4).

Artificial neural networks, composed of many layers, drive deep learning. Deep Neural Networks (DNNs) are networks where each layer can perform complex operations such as representation and inference that make sense of images, audio, and text. Deep learning is the fastest-growing field in machine learning. Deep learning is used in a wide range of applications, such as the detection of researchers' communities (5), the detection of automatic modulation in radar signals (6), the monitoring of intelligent cities using facial recognition robots (7), and efficient classification of facial features (8). Deep learning has helped in image classification, language translation, and speech recognition. Deep learning can solve all pattern recognition problems without human intervention. Deep learning is used in object detection. Kurdthongmee et al. (9) used a YOLOv3 detector to detect parawood cross-sectional. This research seeks to solve the problem of needing data for detector training. It has a weakness due to the testing of the proposed method on specific data and the lack of comparison with previous methods.

Kurdthongmee et al. (10) studied the YOLOv7 method which was proposed for detection in cross-sectional images of a parawood log and pupil datasets. This study aims to identify critical points in these images.

Therefore, in this article, one of the robust architectures (ResNeXt) of this field is used to classify the images obtained from a street. The contribution of this study is the introduction of a scene analysis method and the diagnosis of the number of cars, pedestrians, and bikes based on the deep neural network. By using deep learning and pre-processing techniques, the proposed method performs better than existing methods. In this work, the identification of important goals in the traffic scene has been carried out in previous research, with limited and one-dimensional attention, and an essential gap in this field can be filled. The proposed method offers valuable insights and improvements in traffic scene analysis methods and offers potential applications for urban intelligence.

The pre-processing steps in the proposed method can lead to performance improvement so that it can be implemented for low-quality videos, and videos from high-traffic streets where there is much overlap between objects. In the following, the proposed method is explained first. In the results section, the performance of the proposed method has been checked on different videos, and the characteristics, sensitivity, and accuracy criteria have been calculated.

## 2. RELATED WORKS

To create a intelligent city, there is a need to control traffic congestion so that city officials have real-time analysis of traffic flow information. Much research has been done in this regard. Using deep learning, traffic analysis has been done. In the research conducted by Zhzng et al. (11), the simulation is based on drone-based videos, where the moving objects of the film are identified. This process uses a pre-trained, real-time model for traffic analysis.

You Only One Once (YOLO) v5 model is proposed by Sharma et al. (12) to detect cars, traffic lights, and pedestrians in lousy weather conditions. Several different classifications have been done for video scenes.

CNN, One of the deep learning models, has been used to identify and classify objects on images taken by drones, including pedestrians, cars (13), and motorcycles/bicycles (14). Two data sets have been evaluated using three GoogleNet, VggNet, and ResNet50 architectures in this research, and the highest level of accuracy has been obtained with the ResNet50 architecture. Through object classification with sensors, the model of the object or location can be recognized more accurately (15). An optimal approach to improve face recognition with and without a mask is proposed using machine learning and deep learning techniques, which use three classifiers: SVM, KNN, and DNN. A network search with meta-parameter adjustment, and nested cross-validation is proposed in this method (16). According to the review of previous research, the difference in the current research includes several items.

- ResNeXt architecture has not been used for traffic scene analysis and classification.
    - The data set used in previous research mainly was images obtained from drones.
    - In this research, actual data, including low-quality and high-traffic videos, have been used.
- In the proposed scheme, pre-processing has been done on the examined data set, which plays a significant role in improving performance.
    - The simulation results show improved characteristics, sensitivity, and accuracy parameters.

## 3. PROPOSED METHOD

The flowchart of the proposed method is shown in Figure 1.

In the proposed method, the image is segmented first. Image segmentation (17), or object separation, is one of the critical processes in analyzing essential features in each segment. Segmentation can be divided into two parts: general segmentation and partial segmentation. The goal of general segmentation is to determine non-connected points in an image related to target objects. General segmentation algorithms are used in cases where the characteristics of the desired area are uniform and homogeneous. On the other hand, partial segmentation deals with distant points that are not directly related to the desired object. In this type of segmentation, the background segments can be separated from the image. The purpose of removing the background is to create a distinction between the desired object and the scene; it divides the image into structural points, such as background and foreground. In other words, removing the background should be done in such a way that the structure of the desired objects is not destroyed. The constancy of the background color usually makes a big difference with the desired object, which reduces the complexity of the processes.

The gray level threshold is one of the simple and widely used methods for removing the background of images. This method uses general information and is computationally fast and low-cost. This method converts a gray image into binary parts, including background and foreground. This process is done using Equation 1:

$$E_{k.l} = \begin{cases} 1 & p_{k,l}^{(N)} \geq T^s \\ 0 & p_{k,l}^{(N)} \leq T^s \end{cases} \quad (1)$$

where, $p_{k,l}^{(N)}$, and $T^s$ are the value of pixel k,l, and threshold, respectively. The method used for segmenting the background is that the pixels whose values are within a specific range are selected and kept as the background, and the rest of the pixels are assumed to be the background. It ignores the context.

After segmentation, feature extraction (18) is done using deep learning. At this stage, ResNeXt architecture, one of the most complete object recognition techniques, is used. The set of combined transformations can be signified as:

$$\mathcal{F}(x) = \sum_{i=1}^{c} \tau_i(x) \quad 2$$

where, $\tau_i(x)$ can be an arbitrary function. This architecture builds upon the inception and ResNet architectures (19) to provide a new and improved architecture. The inception module is a significant change from sequential architectures. In one layer, there are several types of feature extractors, layers that receive input values, and convert them into some kind of data for calculations, in a network that is learning itself and has to use different options to solve tasks, this type of layering indirectly helps the network to perform better. This module can use the inputs directly in its calculations, or sum them directly. Pre-training works like a transfer learning model so that patterns learned from Inception data can be transferred to a program that has its own set of training data samples. The retraining process uses existing parameters learned as part of the Inception classifier, and its use saves significant training time. In addition, with the help of this module, a high-accuracy classifier is built with less training data by exploiting the transfer learning paradigm. It is possible to classify vehicles, including cars and bikes, using physical characteristics and visual characteristics. These characteristics include visual distance from the ground, vehicle height, and the distance between the license plate and the rear bumper, which is extracted from the input using CNN.

It is possible to recognize pedestrians using methods based on the human body model using two-dimensional and three-dimensional information of different body parts. Among this information, can mention the location and movements of different parts of the body. In fact, in

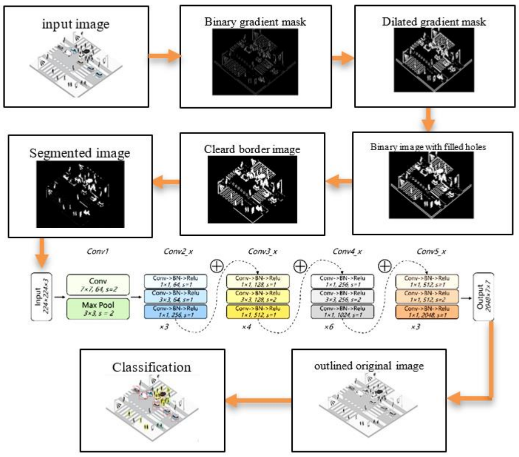

input image
Binary gradient mask
Dilated gradient mask
Binary image with filled holes
Cleard border image
Segmented image
Conv1
Conv2_x
Conv3_x
Conv4_x
Conv5_x
Input 224×224×3
Conv 7×7, 64, s=2
Max Pool 3×3, s = 2
Conv->BN->Relu 1×1, 64, s=1
Conv->BN->Relu 3×3, 64, s=1
Conv->BN->Relu 1×1, 256, s=1
×3
Conv->BN->Relu 1×1, 128, s=1
Conv->BN->Relu 3×3, 128, s=2
Conv->BN->Relu 1×1, 512, s=1
×4
Conv->BN->Relu 1×1, 256, s=1
Conv->BN->Relu 3×3, 256, s=2
Conv->BN->Relu 1×1, 1024, s=1
×6
Conv->BN->Relu 1×1, 512, s=1
Conv->BN->Relu 1×1, 512, s=2
Conv->BN->Relu 1×1, 2048, s=1
×3
Output 2048×7×7
outlined original image
Classification

**Figure 1.** The flowchart of the proposed method

this method, the way the human body stands is recognized by estimating the posture, and different parts of the body, and checking the connection points. In order to strengthen the proposed method against the change in the camera angle, and the change in the environment, a two-point perspective has been used based on the tracking of body connection points.

Therefore, The contributions of this paper can be expressed in a few cases: First, a proposed method that will achieve superior performance in scene analysis. Secondly, its application in public management plays a significant role in analyzing governance policies and features of intelligent cities and understanding the changing needs of intelligent cities. Thirdly, deep learning is vital in designing, decision-making, and implementation by understanding data patterns, classification, and prediction.

## 4. RESULTS

The use of deep learning in the proposed method makes it possible to consider environmental and background conditions with a higher degree of accuracy, which in turn helps to improve the accuracy of detection, localization, and classification. Camera images are converted to pixel-level data, which are processed to remove the foreground. These image data are segmented, and then all the segmented data are structured in the form of a tensor. This composite tensor is then further processed to optimize the bounding box. The obtained data use the deep learning algorithm through CNN. The work steps are shown in Figure 2.

The first stage is an image that includes cars, bikes (bicycles/motorcycles), and pedestrians. In the second step, objects are segmented. The object to be segmented is very different from the background image. Contrast changes can be detected by operators that calculate the gradient of an image. In the third stage, the binary gradient mask (20) of high-contrast lines is shown in the image. These lines do not entirely define the outline of the subject of interest, and compared to the original image, there are gaps in the lines around the object in the gradient mask, and the linear gaps should be removed. In the fourth stage, the internal gaps are filled, but there are still holes inside the cell that need to be filled. In the fifth step, the connected objects on the border are removed. In the sixth step, the object is smoothed, and finally, to make the segmented object look natural, the object is smoothed by eroding the image twice with a diamond structure element. In the seventh step, an alternative method to represent the segmented object is to place the outline around the image. In the eighth step, classification is done using CNN architecture, and three different classes are shown with red, green, and yellow boxes.

An accurate sample of street videos was also analyzed to check the proposed method, and the obtained results are shown in Figure 3. In this figure, the proposed method has been checked on three different videos. Video A (21) contains 23435 frames with a low resolution of 480×320 pixels. This dataset is in an urban location. Video B shows the traffic of a three-way street in an urban area. Video C is a high-traffic city area where objects overlap a lot.

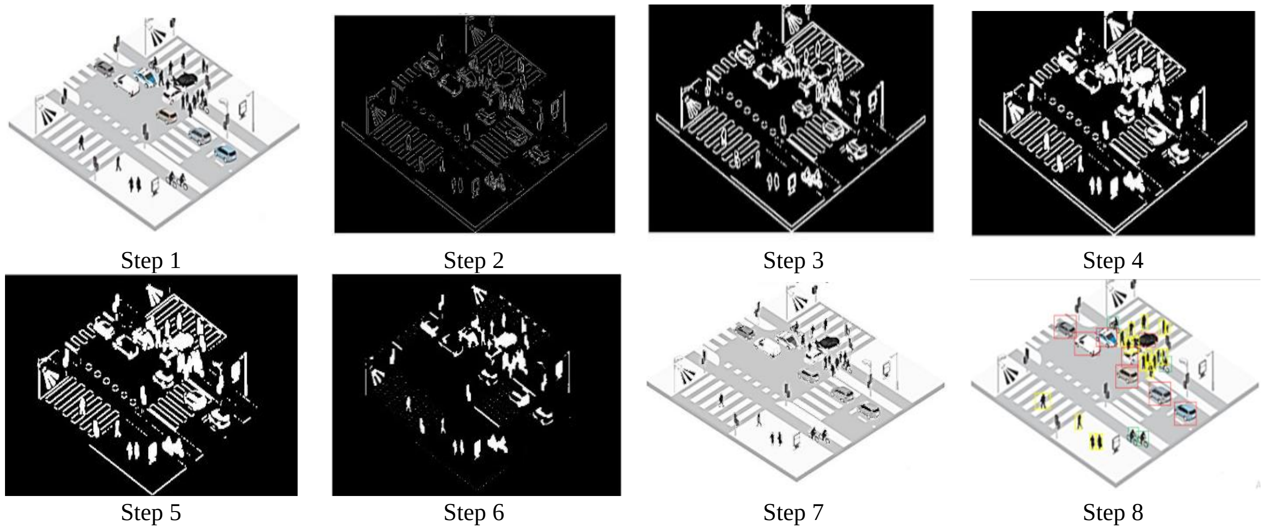

Step 1
Step 2
Step 3
Step 4
Step 5
Step 6
Step 7
Step 8

**Figure 2.** Proposed method teps: Step 1: Input; Step 2: Binary gradient mask; Step 3: Dilated gradient mask; Step 4: Binary image with filled holes; Step 5: Cleard border image; Step 6: Segmented image; Step 7: Outlined original image; Step 8: Classification

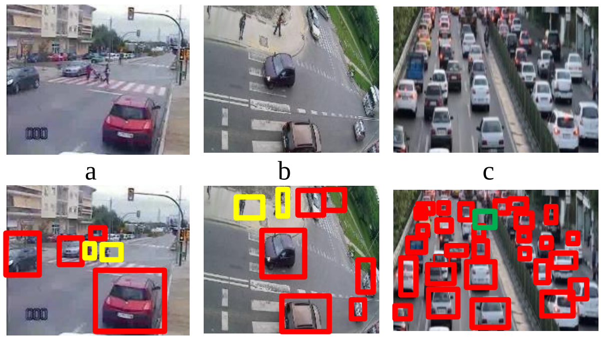


**Figure 3.** Traffic scene analysis and classification on the dataset

The evaluation of the proposed method is done using the existing criteria. Two critical criteria in order to recognize the correct classification, criteria The sensitivity and specially of classification, are obtained using the relations (2):

$$\varphi = \frac{TP}{TP+FN} \tag{3}$$

$$\eta = \frac{TN}{TN+FP} \tag{4}$$

where, TP is positive cases whose category is correctly diagnosed as healthy, TN is negative cases whose category is correctly diagnosed. FP is positive cases whose category is not correctly recognized, and FN is negative cases whose category is wrongly recognized.

In evaluating the proposed method, TP and TN, are the most critical important values that should be maximized in a problem. As summarized in Table 1, these two criteria are close to 0.9 for cars and bikes and lower than 0.9 for pedestrians. Considering the low quality of video a and the high density and traffic in video c, acceptable results have been obtained. These results show that even for videos with high traffic and low quality, the performance of the proposed method is favorable. On the other hand, since the results were obtained from real samples, they can be used for actual applications.

Another vital criterion is accuracy, which is obtained by thefollowing relationship (19):

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{5}$$

Accuracy is one of the most important criteria to determine the efficiency of a classification method. This criterion is the most famous and most common efficiency calculation criterion. It is a classification algorithm that shows how many percent of the entire set of test records is correctly classified. Table 2 shows a comparison between the proposed method and the combined method.

Table 2 shows the comparison of accuracy. From this table, it can be concluded that the classification accuracy for cars, and bicycles is above 90%, while for pedestrians, it is below 90%. The average accuracy for the combined method is 85.8% and for the proposed method is 90.2%, which is a significant improvement.

## 5. DISCUSSION

The proposed solutions to the research questions and discussion of the traffic scene analysis are as follows:

Can preprocessing improve the performance? The performed pre-processing leads to the removal of additional information, restoration of extraction objects, and, as a result, increasing the segmentation quality. Due to the large number of calculations, CNN can be a good moderator for this purpose.

**TABLE 1.** Sensitivity and specially values

| | **Video a** | **Video b** | **Video c** | **Total** | $\varphi$ | η |
|---|---|---|---|---|---|---|
| **Car** | 125 | 41 | 223 | 389 | 0.91 | 0.89 |
| **Bike** | 9 | 7 | 86 | 102 | 0.95 | 0.92 |
| **Pedestrian** | 16 | 18 | 23 | 57 | 0.85 | 0.87 |
| **Total** | 150 | 66 | 332 | 53 | 0.9 | 0.89 |

**TABLE 2.** Accuracy values

| | **Car** | **Bike** | **Pedestrian** | **Total** | **Accuracy (%)** | |
|---|---|---|---|---|---|---|
| | | | | | **Ref. (5)** | **Proposed Method** |
| Car | 20 | 0 | 0 | 20 | 90.9 | **92.4** |
| Bike | 0 | 0 | 16 | 16 | 100 | **100** |
| Pedestrian | 0 | 5 | 12 | 17 | 66.7 | **78.2** |
| Total | 20 | 5 | 28 | 53 | 85.8 | **90.2** |

Can ResNeXt architecture increase the accuracy of the proposed method? One of the essential features of this architecture is its very high accuracy, and this has been proven in Pre-processing. This precise architecture contributes significantly to increase accuracy, sensitivity and specificity.

## 6. CONCLUSION

Traffic scene analysis is one of the essential components of intelligent cities. In traffic scenes, there are cars, bikes, and pedestrians in the street. The correct identification of each object in the scene, and then their correct classification can provide essential information to city managers and people who travel on that route. In this article, the classification is based on a robust ResNeXt architecture that contributes to increasing accuracy and improving diagnostic and classification performance. The preprocessing process makes these goals more attainable and helps to improve the performance. It shows the essential criteria for evaluating the calculation performance and the improvement results of the accuracy criterion. The proposed method reached over 90% on different data such as urban points, high traffic points, and actual data of implementation and evaluation criteria.

Deep learning has significant advantages, but its disadvantages cannot be easily overcome. One of its disadvantages is the high computational cost, and future research can try to reduce it by using creative methods. Another aspect of research can be adding other applications to the proposed method, Such as classifying cars separately, separating bicycles and motorcycles, and separating pedestrians into adults and minors. Tracking is another important application that can be considered after classification.

## 7. REFERENCES

1. Al-Mallahi A, Natarajan M, Shirzadifar A. Development of robust communication algorithm between machine vision and boom sprayer for spot application via ISO 11783. Smart Agricultural Technology. 2023;4:100212. https://doi.org/10.1016/j.atech.2023.100212
2. Dorrani Z. Road Detection with Deep Learning in Satellite Images. Majlesi Journal of Telecommunication Devices. 2023;12(1):43-7. https://doi.org/10.30486/mjtd.2023.1979006.1024
3. Long F, Yao T, Qiu Z, Li L, Mei T, editors. PointClustering: Unsupervised Point Cloud Pre-Training Using Transformation Invariance in Clustering. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2023.
4. Ranjith CP, Hardas BM, Mohideen MSK, Raj NN, Robert NR, Mohan P. Robust deep learning empowered real time object detection for unmanned aerial vehicles based surveillance applications. Journal of Mobile Multimedia. 2023:451–76-–76. https://doi.org/10.13052/jmm1550-4646.1925
5. Torkaman A, Badie K, Salajegheh A, Bokaei M, Fatemi Ardestani S. A Hybrid Approach to Detect Researchers' Communities Based on Deep Learning and Game Theory. International Journal of Engineering, Transactions B: Applications. 2023;36(11):2052-62. 10.5829/IJE.2023.36.11B.10
6. Aslinezhad M, Sezavar A, Malekijavan A. A noise-aware deep learning model for automatic modulation recognition in radar signals. International Journal of Engineering, Transactions B: Applications. 2023;36(8):1459-67. 10.5829/IJE.2023.36.08B.06
7. Medjdoubi A, Meddeber M, Yahyaoui K. Smart City Surveillance: Edge Technology Face Recognition Robot Deep Learning Based. International Journal of Engineering, Transactions A: Basics. 2024;37(1):25-36. 10.5829/IJE.2024.37.01A.03
8. Rohani M, Farsi H, Mohamadzadeh S. Deep Multi-task Convolutional Neural Networks for Efficient Classification of Face Attributes. International Journal of Engineering, Transaction B: Applications. 2023;36(11):2102-11. 10.5829/IJE.2023.36.11B.14
9. Kurdthongmee W, Suwannarat K, Kiplagat J. A Framework to Create a Deep Learning Detector from a Small Dataset: A Case of Parawood Pith Estimation. Emerging Science Journal. 2022;7(1):245-55. 10.28991/ESJ-2023-07-01-017
10. Kurdthongmee W, Suwannarat K, Wattanapanich C. A Framework to Estimate the Key Point Within an Object Based on a Deep Learning Object Detection. HighTech and Innovation Journal. 2023;4(1):106-21. 10.28991/HIJ-2023-04-01-08
11. Zhang H, Liptrott M, Bessis N, Cheng J, editors. Real-time traffic analysis using deep learning techniques and UAV based video. 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS); 2019: IEEE. 10.1109/AVSS.2019.8909879
12. Sharma T, Debaque B, Duclos N, Chehri A, Kinder B, Fortier P. Deep learning-based object detection and scene perception under bad weather conditions. Electronics. 2022;11(4):563. 10.3390/ELECTRONICS11040563
13. Dorrani Z, Farsi H, Mohammadzadeh S. Edge Detection and Identification using Deep Learning to Identify Vehicles. Journal of Information Systems and Telecommunication (JIST). 2022;3(39):201. 10.52547/JIST.16385.10.39.201
14. Cengiz E, Yilmaz C, KAHRAMAN H. Classification of human and vehicles with the deep learning based on transfer learning method. Düzce Üniversitesi Bilim ve Teknoloji Dergisi. 2021;9(3):215-25. 10.29130/DUBITED.842394
15. Dheekonda RSR, Panda SK, Khan N, Al-Hasan M, Anwar S. Object detection from a vehicle using deep learning network and future integration with multi-sensor fusion algorithm. 2017. https://doi.org/10.4271/2017-01-0117
16. Thanathamathee P, Sawangarreerak S, Kongkla P, Nizam DNM. An Optimized Machine Learning and Deep Learning Framework for Facial and Masked Facial Recognition. Emerging Science Journal. 2023;7(4):1173-87. 10.28991/ESJ-2023-07-04-010
17. Fooladi S, Farsi H, Mohamadzadeh S. Segmenting the lesion area of brain tumor using convolutional neural networks and fuzzy k-means clustering. International Journal of Engineering, Transaction B: Applications. 2023;36(8):1556-68. 10.5829/IJE.2023.36.08B.15
18. Lu S, Ding Y, Liu M, Yin Z, Yin L, Zheng W. Multiscale feature extraction and fusion of image and text in VQA. International Journal of Computational Intelligence Systems. 2023;16(1):54. 10.1007/S44196-023-00233-6
19. Dorrani Z, Farsi H, Mohamadzadeh S. Deep Learning in Vehicle Detection Using ResUNet-a Architecture. Jordan Journal of Electrical Engineering. 2022;8(2):166-78. 10.5455/JJEE.204-1638861465

20. Dorrani Z, Farsi H, Mohamadzadeh S. Image edge detection with fuzzy ant colony optimization algorithm. International Journal of Engineering, Transactions C: Aspects 2020;33(12):2464-70. 10.5829/IJE.2020.33.12C.05

21. Wang Y, Ban X, Wang H, Wu D, Wang H, Yang S, et al. Detection and classification of moving vehicle from video using multiple spatio-temporal features. IEEE Access. 2019;7:80287-99. 10.1109/ACCESS.2019.2923199

## Persian Abstract

**چکیده**

در این مطالعه، هدف ما استفاده از ابزارهای جدید یادگیری عمیق و شبکه عصبی کانولوشن برای تجزیه و تحلیل ترافیک است. معماری ResNeXt که یکی از قدرتمندترین معماری‌هاست وتوجهات زیادی را در عرصه‌های مختلف به خود جلب کرده، برای بررسی صحنه و طبقه‌بندی آن در سه دسته‌ی ماشین، موتور (موتورسیکلت/دوچرخه) و عایر پیاده پیشنهاد شده است. مطالعات قبلی بیشتر روی یک نوع طبقه‌بندی متمرکز شده و فقط تشخیص انسان و چهره، یا فقط تشخیص وسایل نقلیه را گزارش کردند. در مقابل، روش پیشنهادی از معماری دقیقی استفاده کرده تا دسته‌بندی سه کلاسه را انجام دهد. طرح پیشنهادی شامل چند مرحله است: مرحله اول بخش‌بندی است که اشیاء مهم جدا می‌شوند. در مرحله‌ی بعد ویژگی‌های اشیای بدست آمده استخراج می‌شود تا فرآیند دسته‌بندی به سه کلاس موردنظر انجام گیرد. آزمایش‌هایی روی مجموعه داده‌های مهم و متفاوتی مانند صحنه‌های پرترافیک، کیفیت پایین و واقعی انجام شده است. معیارهای مهم ارزیابی مانند دقت، حساسیت و ویژگی به دست آمده، نشان می‌دهد عملکرد روش پیشنهادی نسبت به روش‌های مورد مقایسه بهبود یافته است. معیار دقت به بیش از ۹۲٪، حساسیت حدود ۸۹٪ و ویژگی به ۹۰.۲۵٪ رسیده است. روش پیشنهادی می‌تواند برای اجرای هوشمندسازی شهرها، ایمنی عمومی، تصمیمات کلان شهری و استفاده از نتایج بدست آمده در مدیریت شهری، مدل‌سازی پیش‌بینی کننده از مدیریت داده‌های از دست رفته، مدیریت داده‌های متوالی و تعمیم‌پذیری مورد استفاده قرار گیرد.