Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 31 No. 2s (2024)

An Improved Deep Network and Handcrafted Feature-Based Scene Classification Convolutional Model for Self-Driving Cars

Sanjay P. Pande[1], Sarika Khandelwal[2]

[1]Department of Computer Science & Engineering, G.H. Raisoni University, Amravati, Maharashtra, India

[2]Department of Computer Science & Engineering, G.H. Raisoni College of Engineering, Nagpur, Maharashtra, India

[1]sanjaypande2001@gmail.com, [2]sarikakhandelwal@gmail.com

Article History:

Received: *28-03-2024*

Revised: *14-05-2024*

Accepted: *29-05-2024*

Abstract:

With growing urbanization, cars have populated the roads to great extent. Intelligent cars are the need to mitigate the traffic and improve the traffic efficiency. Scene classification in one of the important ingredient of self-driven cars, which acts as a key source for good decision making tasks. Scene complexities and diversities have made the problem more complex resulting in similarities in different classes and differences in same categories. To deal with such scenarios, the proposed work present a fusion of various effective features obtained from deep networks and descriptors to classify scenes in four different categories. Local Features are obtained using the YOLOV5m and the VGG19 networks. The YOLOV5m is used to detect the relevant objects in the scene and the VGG19 is used to lift the blind features of the objects detected. The global feature extraction module is used to extract the blind features using VGG19 network. For improving the classification accuracy, eight handcrafted features adding fine and coarse details of the image are fused with local and global features. A fully connected network comprising of five layers is finally used to differentiate the scenes in four categories viz crosswalk, highway, over pass/tunnel and parking. A self-generated dataset is constructed from four different publicly available datasets to evaluate the performance of the proposed scene classification model. The experimental results show that even under high correlation between classes, the system was able to classify the test samples with 86.79% accuracy which is higher than the state-of-the-art models.

Keywords: Intelligent cars, Scene classification, deep networks, local features, YOLOV5m, VGG19, global features and handcrafted features.

1. Introduction

Intelligent transportation is todays need with growing traffic especially in the urban and populated cities. Many countries flying from developing stage to developed stage are undergoing rapid development concerning social economy. The means of transportation had grown to extreme extent which includes cars and other vehicles and had greatly affected the traffic situations. The traffic scenario had become more severe and complex. Automated vehicles combines various neighbouring information from the vicinity and ground traffic to be efficient part of the worsened traffic involving congestions, accidents and other sorts of problems [1-3] affecting a convenient and safe ride [4-5]. An automated or self-driven vehicle rely on the surrounding environment and runs with no or some manual support. The decision by such indispensable part is a function of evaluation of environment perception which basically includes scene classification, lane identification, or-road object detection and others [6-7]. Today's research focusses on scene classification which is challenging and prominent in self-driving car systems due the complexity of traffic and numerous scene categories [8].

Radars, sensor and geographical positioning systems acquires surrounding information related to road conditions processed for current state conditions [9-10]. The driving strategy decisions and the path planning requires proper high level semantic data of its current coordinates for intelligence. The car should be slowed down when approaching a school or hospital, it should be accelerated along the highway, make use of anti-skid mechanism during rains and snowy conditions, stop at a red signal, turns on fog lights during dense fog atmosphere, automatically gets parked for drowsiness or drunk conditions, etc. [11-12]. Many researchers have suggested object detection and classification systems over the traffic network [13-17]. The objective of their work was to mitigate the road accidents and driver stress during driving vehicles [18]. The aim of their work was ultimately to replace driver decisions with autonomous vehicle system making effective use of perception and intelligent decisions. The intelligent autonomous vehicle system depends on the fast and accurate investigation of existing objects. Investigation involves nature of the object, size and location in the frame acquired. Analyzing warning signs is also important aspect which requires crucial attention. Overall, the intelligence of the autonomous vehicle is determined by proper analysis of various objects acquired using the vehicle scene acquisition system.

Earlier work proposed in [19] adopted Bayesian classifier using low-level visual descriptors to construct series of semantic tags. The accuracy of such scheme is low due to failure of such scheme to represent the complex visual contents using low-level one dimensional visual features. The scheme suggested in [20] made use of bag of words for scene recognitions to improve the generalization ability as a part of invariant features acquired locally and probabilistic latent space models. The authors failed to satisfy the requirement for self-driving cars as a fact that bag of words neglected the impression of synonyms and polysemous.

Deep learning techniques have been a boon in various fields of image and speech processing applications [21-24]. Recent researchers had made effective use of deep learning tools for scene classification problems especially using the convolutional neural networks. A cost function was estimated using different classification patterns of the scene which was used for training a multi-label neural network in [25]. The work carried out in [26] focussed on various dimension including the type of road, time period and the weather conditions which improved the classification accuracy of the semantic perception. GoogleNet was used for feature extraction at three different levels of the network and further the features were fused for final decision [27]. Features at small and large scales were extracted to collect visual structure using fine and coarse resolution CNN in [28]. As far as the available scene datasets are considered, two main aspects makes the scene classification problem a complex issue. The first aspects is the disparities among the same category images and the second is visual similarities among different categories of images.

To deal with the above problem, we considered scene images from different datasets with four different categories including crosswalk, highway, overpass-tunnel, and the parking. To enhance the complexity of the scene classification problem we combined the over pass and tunnel categories into a single class. Also, the road scene images considered involved scene considering all atmospheric and weather conditions such as day, night, rain, snowy and etc. The images were collected from four different datasets namely BDD100K [29], LabelMe [30], KITTI [31] and Places365 [32]. Different datasets have images with different dimension and diverse illumination conditions. This increases the

complexity for classification and requires more robust feature set at fine and coarse level. Also, the objects in the scene need to be located for better correlation among the classes. Based on these aspects, the proposed work extract features at local and coarse level including the handcrafted and blind features using deep networks. Lastly, the features are concatenated to realize scene classification with greater accuracy. Some of the sample images from all datasets belonging to different categories are shown in figure 1 below.

Figure 1 – Images from all four categories for scene classification. A crosswalk, highway, over pass and parking.

The work claims the following contributions:

1. Objects in the scene are detection using the YOLOV5 network and further features are extracted for every object using modified VGG19 network and summed up for dimension reduction.

2. Handcrafted low and high level features on gray scale image are also extracted to improve the disparities among the classes for better classification. **The features includes**, **wavelet based features, local binary pattern based features, gray level co-occurrence matrix features, histogram of gaussian features, and matched filter coefficients**.

3. Blind features using VGG19 directly on color image are extracted from the last layer of the network to cover the depth information of the images.

4. All the features from step 1 to 3 are lastly concatenated to represent an image from the set. A fully connected layer with softmax function is trained on 75% of images from each category. 10% images are used to validate the network while remaining 15% are used to test the trained network.

The organization of the article is as follows: Section II covers the review part of recent state-of-the-art work contributed to scene classification, Section III deals with datasets and proposed scene classification system, Section IV discusses the performance of the proposed system and the last section concludes the article.

2. Related Work

An object detection mechanism was incorporated using machine learning to investigate the presence of seven different objects in [33]. Images from two different datasets were considered for evaluating the performance. They pre-processed the data to handle missing values and used six different machine learning based classifiers to classify the objects. They studied the effect of imbalanced data, incomplete data and elimination of low views and redundant samples. With 80: 20 split, they evaluated the performance of their system using different metrics.

A hybrid approach incorporating features from two state-of-the-art object detection models viz. YOLO and Fast RCNN were used to improve the object detection accuracy and segmentation [34]. They used YOLO's capability of object detection and the bounding box selection characteristics and combined with Fast RCNN's region of interest pooling to improve the segmentation and the classification. They

also reduced the processing time of Fast RCNN by eliminating the region proposal network and trained the network over 10000 images. They obtained 5-7% increase in accuracy as compared to standalone YOLO network.

The work proposed in [35] used context based saliency detection algorithm for marking the visual attention regions (VAR) in the scene. The VAR is further enhanced by superimposing it on the original image. A deep CNN pre-trained on Places365 was used to extract features from the VAR, enhanced VAR and the original image. A classification model based on deep CNN was used to test the effectiveness of the model on four distinct datasets. They fused the output from layer 6 and layer 7 of the AlexNet network to generate the deep fusion features of the image. Combining the deep fusion features, they used random forest to train the classifier. The scene depth intrinsic characteristics and contextual semantic correlation helped to improve the classification accuracy.

Local semantic features were extracted by modifying the faster RCNN network and adding residual attention structure in [36] whereas global features were extracted using improved Inception network. The later network is presented with Leaky ReLU and the ELU function in order to reduce the redundancy of convolution kernel. Further, both the features were fused to realize the scene classification. The author built their own dataset to classify different scenes for self-driving cars. The work does not considered heterogeneous road agents.

The work presented in [37] used vehicle GPS location for scene classification. The image features were extracted using the VGG16 network pre-trained on Places365 dataset. They classified scenes related to urban, rural and the highways from the KITTI datasets. A novel object detection scheme was introduced by [38] for autonomous driving using the YOLOV5 network. The network was used with structural re-parameterization, small objects detection module, search module and coordinate attention strategy. The algorithm was tested on KITTI dataset with improved accuracy and low computational complexity. The algorithm suffered from noise and occlusion sensitivity and higher amount of training data.

3. Proposed Work

The scene classification model considers images of four different categories obtained from different datasets. **The aim is to increase the complexity of the problem under consideration and test the proposed classification scenario under the worst conditions**. The first challenge introduced is the number of samples from each category. We have manually **separated the images in four categories** pertaining to classes: Crosswalk, highway, over pass-tunnel and parking. Table I below shows the number of sample images considered for the proposed work. The second complexity is with the diverse nature of images since they are obtained from different source datasets (BDD100K, LabelMe, KITTI and Places365) which **possess different dimensions and different lighting conditions**. Figure 2 below shows some examples from the datasets. The third crucial parameter considered is related to objects which are not clear enough to be distinguished easily. Figure 3 shows the relevant objects to be detected from the scene which are not significant for the class under which they are considered as compared to other objects. As compared to figure 1, the main objects in figure 2 as per their categories which they belongs, are not easily identified. We have intentionally selected some images which belongs to more than one categories to increase the complexity.

Table I – Number of sample images considered in each category for scene classification

Class	Name of the Class	Number of images
1	Crosswalk	700
2	Highway	700
3	Over Pass-Tunnel	625
4	Parking	700

(a) Crosswalk (Night) (b) Highway (Without Vehicles)

(c) Over Pass/Tunnel

(d) Parking (Road Side and Parking Lot)

Figure 2 – Sample images from all four categories. (a) Crosswalk at night. (b) Highway without traffic. (c) Over pass/Tunnel. (d) Road side parking at night and parking lot.

(a) Non-Significant Crosswalk (b) Highway under Over Pass

(c) Non-Significant Over Pass (d) Parking at night

Figure 3 – images with Non-Significant features. (a) Crosswalk not seen clearly (b) Highway passing under an Over pass (c) An over pass identified by side pillars (d) Unidentified Road side parking on right side

The **proposed scene classification system** comprises of three distinct modules: the local feature extraction module, the global feature extraction module and the handcrafted feature extraction module for fine and coarse features. The framework for scene classification system is shown in figure 5.

The **local feature extraction module** is aimed to detect significant objects in the scene using the YOLOvV5m pre-trained network. The detected objects are then resized to 32x32 dimension for obtaining similar size features. The objects are fed to VGG19 network for feature extraction. The VGG19 deep network is used without the top layer to obtain 512 features from each of the detected object. The number of objects that can be detected using the YOLOV5m network varies with the contents of the scenes and therefore we added the features extracted from every object of the scene image. Higher the objects, greater the magnitude of 512 vector. It also helps to fill the missing values in the vector when object features are added up. Figure 4 shows the objects identified by the YOLOV5m network. Given an input image I, the YOLOV5m network detects N objects and provides their bounding boxes and class labels. Each detected object I_m can be extracted from the image based on its bounding box coordinates.

$$\text{Detection process: } \{(\mathrm{B}_1, \mathrm{C}_1), (\mathrm{B}_2, \mathrm{C}_2), \ldots, (\mathrm{B}_\mathrm{N}, \mathrm{C}_\mathrm{N})\} = \mathrm{YOLOV5m}(I)$$

$$\text{Extracting Objects from Bounding Boxes: } I_m = I[B_m]$$

$$\text{Constructing the Local Feature Vector: } F = \sum_{m=1}^{N} f_m;$$

where F= “final local feature vector”, N= “no. of detected objects”, f_m= “512-dimensional feature vector for the mth detected object ”.

(a) Original Image from the Crosswalk Category

(b) Cars located in the image by YOLOV5m

(c) Located keyboard, person and a traffic light by YOLOV5m

Figure 4 – Bounding Box results obtained using the YOLOV5m network

The following Algorithm 1 represents the steps to collect the features of objects identified by the YOLOV5m network. The value of 'N' in the algorithm will vary for each image since the number of detected objects depends on the scene image and the ability of YOLOV5m network.

ALGORITHM 1 : Object based Feature Extraction

Input – Object Images identified by YOLOV5m network
Output – Feature Vector of all objects

N – Number of identified Objects using YOLOV5m
F = (512,) zero vector ; Array to store the feature vector
For m = 1 to N
    Resize the object image ; 32x32
    Calculate features (512,) vector using the VGG19 network
    Add the features to F
end

The **global feature extraction module** uses VGG19 network without the top layer directly on the input image resized to 256x256. The number of feature extracted using the VGG19 network are 512. The blind features thus extracted depends on the scene information and ability of the network. This is to ensure that region not belonging to the objects detected using the local feature extraction module contributes to the feature set. The only problems with such features are too many missing values which depends on the quality of the image.

Even though the local features are considered using the two stage deep network framework using the YOLOV5 and the VGG19 networks, the resizing stage for the detected objects may suffers from information loss. A size of 32x32 is considered to uplift the fine features with respect to small objects but **objects greater than 32x32 would suffer data loss. Therefore, we added fine and coarse features to the local and global features to improve the classification accuracy.** A total of 2310 handcrafted features are extracted using various feature descriptors which includes wavelet based, matched filter based, local binary pattern (LBP) based texture, gray level co-occurrence matrix

(GLCM) based, and the histogram of gaussian (HOG) features. Use YOLOV5 to detect objects in the image and extract features using VGG19 for each detected object.

$$F_{local} = \sum_{i=1}^{N} VGG19(Resize(YOLOV5(I_i), 32 \times 32))$$

Use VGG19 to extract features from the entire resized image

$$F_{global} = VGG19(Resize(I, 256 \times 256))$$

The following traditional features depicted in Table II are extracted using 128x128 gray scale scene image.

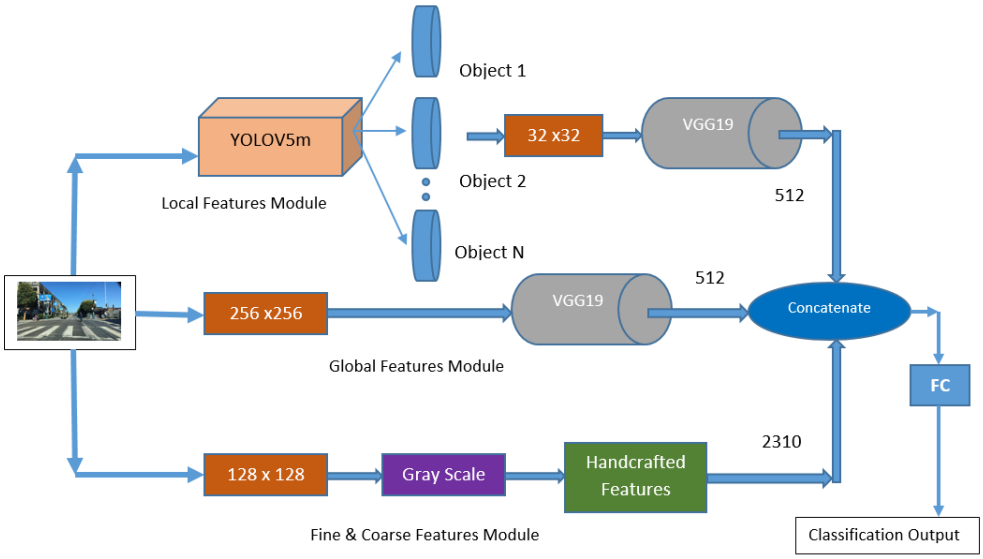

Figure 5 – The framework for the scene classification system

Table II – Fine and Coarse Handcrafted Features with Count

Name of the Feature	Type of feature	Number of features extracted
Matched Filter – Edge Kernel-based	Fine	256
Matched Filter – Orientation based	Coarse	32
Wavelet – Six wavelets ['bior3.1', 'bior3.5', 'bior3.7', 'db3', 'sym3', 'haar']	Coarse	24
Wavelet - Haar	Fine	1024
GLCM – [Contrast, Energy, Homogeneity, Correlation, ASM and Dissimilarity]	Coarse	6
LBP	Fine	256
LBP – Texture (Averaging 5x5 block)	Fine	676
HOG	Coarse	36
Total Feature length		**2310**

Detailed statistical descriptions of the handcrafted features using matched filter, wavelet, and LBP texture features can be found in [39]. These features strengthen the ability of blind features extracted using the CNN-based VGG19 network to improve the disparities of classes and assist the classifier.

The object-based local features, the global features, and the handcrafted features are concatenated for classification.

The features were normalized using the Max-Normalization. The samples were arranged in rows and maximum values across the columns were found. The data was normalized by dividing values in columns by its column-maximum value. The problem of missing values in the feature set was solved using the mean values across columns. The normalization of the concatenated feature vector F_c (comprising local, global, and handcrafted features) is performed using max-normalization.

$$F_{cnorm} = \frac{F_c}{\max(F_c)}$$

where, max (F_c)= "maximum value across the columns of the feature vector F_c"

Missing values in the feature vector F_c are imputed by replacing them with the mean of the respective columns.

$$F_c(i,j) = \begin{cases} F_c(i,j) & if\ F_c(i,j) is\ not\ missing \\ \frac{1}{n}\sum_{k=1}^{n} F_c(k,j) & if\ F_c(i,j) is\ missing \end{cases}$$

where, $F_c(i,j)$ = "(i, j) element of the feature matrix", n = "total no. of samples".

4. Results and Discussion

The proposed scene classification system was developed using Python 3.9 on Spyder 5, Windows 11 Environment, i5 Processor, 16 GB RAM, and 512 GB SSD. The samples were separated in ratio 75:10:15% for training, validation and test. The following parameters indicated in Table III were initialized for the last classification layer for training the samples. The Fully connected layer comprised of one-dimensional convolutional layer with filter size=8, kernel size=3 with activation='linear. It was followed by a Batch Normalization Layer (BNL), a Rectified Linear Unit Activation Layer (ReLU) with alpha=0.25, one-dimensional Max Pooling layer (MPL) with pool_size=3, stride=3 and a Dropout layer (DL) with rate=0.25. Figure 6 below shows the fully connected layer used for classification.

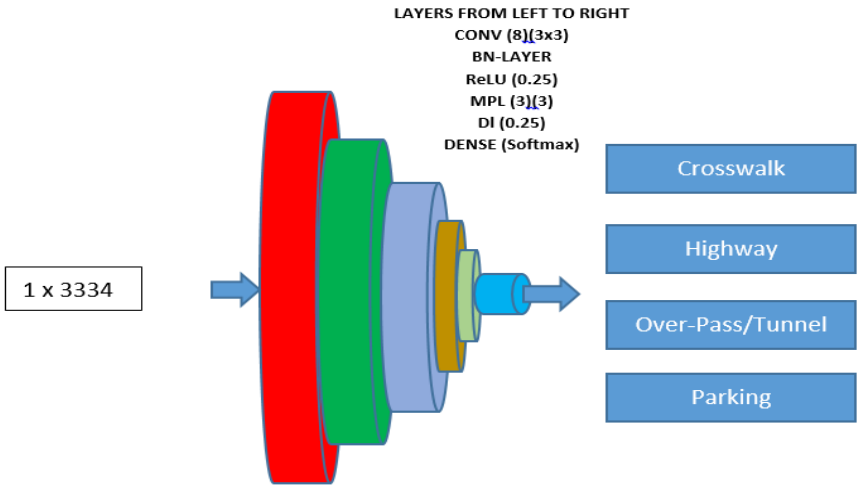

Figure 6 – Classification layer for scene categories

TABLE III – Network Parameters

Parameters	Value
Batch Size	25
Epoch	1000
Iterations	10
Activation	'Softmax'
Optimizer	Adam
Activation for FC	'selu'

We used random state to distinguish the train, validation and test samples and iterated the training-testing sequence for 10 times and noted the average classification accuracy. We conducted experiments considering combination of all classes from minimum two to maximum four. The aim was to evaluate the correlation between different classes and the effect of change in samples for training, validation and testing. The following Table IV shows the evaluation for all combinations in terms of classification accuracy. Table V indicates the classification accuracy when three classes were considered. As seen from Table IV and Table V, Crosswalk and Over Pass/Tunnel class are closely related due to the fact that many scene images considered in the dataset contains crosswalk prints and over pass too as depicted in the Figure 7. Also, we have considered scene with dimmed crosswalks or very few strips of crosswalk which are correlated with the lane strips. Some of the scene contents in other classes as shown in Figure 8 are also closely correlated which amounts to lower classification accuracy.

TABLE IV – Training and test accuracy for two classes

Class 1	Class 2	Accuracy - Training	Accuracy – Test
Crosswalk	Highway	100	97.32
Crosswalk	Over Pass/Tunnel	99.90	**89.78**
Crosswalk	Parking	100	98.18
Highway	Over Pass/Tunnel	100	94.82
Highway	Parking	100	98.66
Over Pass/Tunnel	Parking	100	91.80

Figure 7- Examples for closely correlated objects. Crosswalks with over pass, crosswalks with tunnel, misleading lane as crosswalk with tunnel.

Figure 8- Examples for closely correlated objects in other classes. Crosswalks with parking, crosswalks under tunnel, crosswalk on highway with parking.

TABLE V – Training and test accuracy for three classes

Class 1	Class 2	Class 3	Accuracy - Training	Accuracy – Test
Crosswalk	Highway	Over Pass/Tunnel	99.67	83.70
Crosswalk	Highway	Parking	99.75	88.54
Crosswalk	Over Pass/Tunnel	Parking	99.93	**81.10**
Highway	Over Pass/Tunnel	Parking	99.93	86.26

Finally we tested the network on all four classes and found training accuracy of 99.52% and 86.79% accuracy on test samples. However, very few research are based on manual categorization of scene images, the proposed scene classification system can be compared with other likewise recent approaches as shown in Table VI. The comparison is based on average classification accuracy. The comparison clearly shows that the proposed scene classification model outperforms the Improved Deep Network suggested in [36]. The model suggested in [36] performed well on self-generated datasets. The accuracy 75.99% is claimed on BDDK-100 dataset from where we have used most of the images for our dataset.

Table VI – Comparative test results based on Average Accuracy

Method	Number of classes	Classes	Average Accuracy %
DNCVA [35]	2	Indoor, Outdoor	85.06
Improved DN [36]	5	Crosswalk, Gas Station, Highway, Parking Lot and Street	75.99
Ours	4	Crosswalk, Highway, Parking, and Over Pass/Tunnel	86.79

5. Conclusion

The scene classification based on local, global and fine and coarse features is studied and an efficient network is presented in this article. In the proposed deep network system, local features are obtained using the YOLOV5m and VGG19 networks through object detection and blind features respectively. Global features are obtained using the VGG19 network directly from resized image. And the handcrafted features are extracted to improve the classification accuracy owing to the complexity of classifying correlated classes. The handcrafted features used in this work are more than double of the blind features. The self-generated dataset uses four different standard datasets for categorizing the images into four classes. The problem of missing values in the feature set is solved using the mean values to improve the accuracy. In addition, various experiments are carried out to find the correlation between different classes through classification accuracy. The result obtained shows that the proposed scene classification model outperformed the work conducted in [36] for the BDDK-100 dataset. The only difference between work carried in [36] and ours persists in number of classes. In addition, we considered roadside parking along with the parking lot, dimmed crosswalks, crosswalk along highways, crosswalk under tunnels, roadside parking under tunnels, diminishing tunnels etc. However, there are some limitations which basically includes filling missing values in the feature set where the zero entries were considered for calculating the mean values. The number of samples in each of the categories considered are less especially for the over pass/tunnel class where only 625 samples were

used in total. The fine and coarse features were part of the gray scale conversion. YOLOV5m is not trained to detect crosswalk. Thus prior transfer learning would improve the accuracy.

In future, the dataset should be increased with samples and categories including gas stations, lane recognition, school vicinities, etc. Other deep network other than VGG19 should be considered for feature evaluation. The final layers should be substituted for better feature vectors. The object features obtained in the local feature extraction module were added for uniform dimension. Instead other measure should be incorporated to handle such situations.

References

[1] M. Huang, X. Yan, Z. Bai, H. Zhang, and Z. Xu, "Key technologies of intelligent transportation based on image recognition and optimization control," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 34, no. 10, Sep. 2020, Art. no. 2054024.

[2] A. Mammeri, T. Zuo, and A. Boukerche, "Extending the detection range of vision-based vehicular instrumentation," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 4, pp. 856–873, Apr. 2016.

[3] M. Karaduman and H. Eren, "Smart driving in smart city," in *Proc. 5th Int. Istanbul Smart Grid Cities Congr. Fair (ICSG)*, ˙Istanbul, Turkey, Apr. 2017, pp. 115–119.

[4] F. Duarte, "Self-driving cars: A city perspective," *Sci. Robot.*, vol. 4, no. 28, Mar. 2019, Art. no. eaav9843, doi: 10.1126/scirobotics.aav9843.

[5] N. A. Greenblatt, "Self-driving cars and the law," *IEEE Spectr.*, vol. 53, no. 2, pp. 46–51, Feb. 2016.

[6] Q. Xu, M. Wang, Z. Du, and Y. Zhang, "A positioning algorithm of autonomous car based on map-matching and environmental perception," in *Proc. 33rd Chin. Control Conf. (CCC)*, Nanjing, China, Jul. 2014, pp. 707–712.

[7] Q. Zou, H. Jiang, Q. Dai, Y. Yue, L. Chen, and Q. Wang, "Robust lane detection from continuous driving scenes using deep neural networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 41–54, Jan. 2020.

[8] Y. Parmar, S. Natarajan, and G. Sobha, "DeepRange: Deep-learningbased object detection and ranging in autonomous driving," *IET Intell. Transp. Syst.*, vol. 13, no. 8, pp. 1256–1264, Aug. 2019.

[9] J.-R. Xue, J.-W. Fang, and P. Zhang, "A survey of scene understanding by event reasoning in autonomous driving," *Int. J. Autom. Comput.*, vol. 15, no. 3, pp. 249–266, 2018.

[10] Y. Yang, F. Chen, F. Wu, D. Zeng, Y.-M. Ji, and X.-Y. Jing, "Multi-view semantic learning network for point cloud based 3D object detection," *Neurocomputing*, vol. 397, pp. 477–485, Jul. 2020.

[11] H. Xu and G. Srivastava, "Automatic recognition algorithm of traffic signs based on convolution neural network," *Multimedia Tools Appl.*, vol. 79, nos. 17–18, pp. 11551–11565, May 2020.

[12] Perumal, P.S.; Sujasree, M.; Chavhan, S.; Gupta, D.; Mukthineni, V.; Shimgekar, S.R.; Khanna, A.; Fortino, G. An Insight into Crash Avoidance and Overtaking Advice Systems for Autonomous Vehicles: A Review, Challenges and Solutions. Eng. Appl. Artif. Intell. **2021**, 104, 104406.

[13] Bachute, M.R.; Subhedar, J.M. Autonomous Driving Architectures: Insights of Machine Learning and Deep Learning Algorithms.Mach. Learn. Appl. **2021**, 6, 100164.

[14] Al-refai, G.; Al-refai, M. Road Object Detection Using Yolov3 and Kitti Dataset. Int. J. Adv. Comput. Sci. Appl. **2020**, 11, 1–7.

[15] Qaddoura, R.; Bani Younes, M.; Boukerche, A. November. Predicting traffic characteristics of real road scenarios in Jordan and Gulf region. In Proceedings of the 17th ACM Symposium on QoS and Security for Wireless and Mobile Networks, Alicante, Spain, 22–26 November 2021; pp. 115–121.

[16] Qaddoura, R.; Younes, M.B. Temporal prediction of traffic characteristics on real road scenarios in Amman. J. Ambient. Intell. Humaniz. Comput. **2022**, 1–16.

[17] Kajiwara, S. Evaluation of Driver Status in Autonomous Vehicles: Using Thermal Infrared Imaging and Other Physiological Measurements. Int. J. Veh. Inf. Commun. Syst. **2019**, 4, 232–241.

[18] C. Shen *et al.*, "Multi-receptive field graph convolutional neural networks for pedestrian detection," *IET Intell. Transp. Syst.*, vol. 13, no. 9, pp. 1319–1328, Sep. 2019.

[19] A. Vailaya, M. A. T. Figueiredo, A. K. Jain, and H.-J. Zhang, "Image classification for content-based indexing," *IEEE Trans. Image Process.*, vol. 10, no. 1, pp. 117–130, Jan. 2001.

[20] M. V. Latte, S. Shidnal, B. S. Anami, and V. B. Kuligod, "A combined color and texture features based methodology for recognition of crop field image," *Int. J. Signal Process., Image Process. Pattern Recognit.*, vol. 8, no. 2, pp. 287–302, Feb. 2015.

[21] G. Guo and N. Zhang, "A survey on deep learning based face recognition," Comput. Vis. Image Understand., vol. 189, Dec. 2019, Article ID 102805.

[22] S. Ren, K. Sun, C. Tan, and F. Dong, "A two-stage deep learning method for robust shape reconstruction with electrical impedance tomography," IEEE Trans. Instrum. Meas., vol. 69, no. 7, pp. 4887–4897, Jul. 2020.

[23] Z. Wang, K. Liu, J. Li, Y. Zhu, and Y. Zhang, "Various frameworks and libraries of machine learning and deep learning: A survey," Arch. Comput. Methods Eng., pp. 1–24, Feb. 2019, doi: 10.1007/s11831-018-09312-w.

[24] E. Mutabazi, J. Ni, G. Tang, and W. Cao, "A review on medical textual question answering systems based on deep learning approaches," Appl. Sci., vol. 11, no. 12, p. 5456, Jun. 2021.

[25] L. Chen, W. Zhan, W. Tian, Y. He, and Q. Zou, "Deep integration: A multi-label architecture for road scene recognition," IEEE Trans. Image Process., vol. 28, no. 10, pp. 4883–4898, Oct. 2019.

[26] K. Zheng and H. A. H. Naji, "Road scene segmentation based on deep learning," IEEE Access, vol. 8, pp. 140964–140971, 2020.

[27] P. Tang, H. Wang, and S. Kwong, "G-MS2F: GoogLeNet based multistage feature fusion of deep CNN for scene recognition," Neurocomputing, vol. 225, pp. 188–197, Feb. 2017.

[28] L. Wang, S. Guo, W. Huang, Y. Xiong, and Y. Qiao, "Knowledge guided disambiguation for large-scale scene classification with multi-resolution CNNs," IEEE Trans. Image Process., vol. 26, no. 4, pp. 2055–2068, Apr. 2017.

[29] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 2636–2645, 2020

[30] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 10, pp. 1915–1926, 2012.

[31] R. McCall et al., "A taxonomy of autonomous vehicle handover situations," Transp. Res. A, Policy Pract., vol. 124, pp. 507–522, Jun. 2019.

[32] C. Szegedy et al., "Going deeper with convolutions," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Boston, MA, USA, Jun. 2015, pp. 1–9.

[33] Majd Alqarqaz, Maram Bani Younes and Raneem Qaddoura, "An object classification approach for autonomous vehicles using machine learning techniques," World Electric Vehicle Journal, 2023, 14, 41.

[34] Sajjad Ahmad Khan, Hyun Jun Lee and Huhnkuk, "Enhancing object detection in self-driving cars using a hybrid approach," Electronics, 2023, 12, 2768.

[35] Jing Shi. Hong Zhu, Yuxing Li, YangHui Li and Sen Du, "Scene classification using deep networks combined with visual attention," Journal of Sensors, 2022, Vol. 2022, Article ID 7191537.

[36] Jianjun Ni, Kang Shen, Yinan Chen, Weidong Cao and Simon X. Yang, "An improved deep network-based scene classification method for self-driving cars," IEEE Transaction on Instrumentation and Measurement, 2022, Vol. 71, 5001614.

[37] Roman Prykhodchenko and Pawal Skruch, "Road scene classification based on street-level images and spatial data," Array 15, 2022, 100195.

[38] Jia X., Tong Y., Qiao H., Li M., Tong J. and Liang B., "Fast and accurate object detector for autonomous driving based on improved YOLOv5," Sci. Rep. 2023, 13, 9711.

[39] Mahadeo D. Narlawar and D. J. Pete, "Occluded Face Recognition: Contrast correlation & edge preserving enhancement based optimum features on CelebA dataset," Journal of Harbin Engineering University, 2023, Vol. 44, No. 8, pp. 1192-1204.

[40] Thanh, H.V.T., Ngoc, H.N., Duc, P.N., "Identifying inverse source for Diffusion equation with Conformable time derivative by Fractional" Tikhonov method (2022) Advances in the Theory of Nonlinear Analysis and its Applications, 6 (4), pp. 433-450.

[41] Goyal, Dinesh , Kumar, Anil , Gandhi, Yatin & Khetani, Vinit (2024) Securing wireless sensor networks with novel hybrid lightweight cryptographic protocols, Journal of Discrete Mathematical Sciences and Cryptography, 27:2-B, 703–714, DOI: 10.47974/JDMSC-1921.