International Journal of Electrical and Computer Engineering (IJECE)
Vol. 14, No. 6, December 2024, pp. 6313~6321
ISSN: 2088-8708, DOI: 10.11591/ijece.v14i6.pp6313-6321

Traffic signs detection and prohibitor signs recognition in Morocco road scene

Imane Taouqi[1], Abdessamad Klilou[1], Kebir Chaji[1], Assia Arsalane[2]
[1]Department of Electrical Engineering MiSET Team, Faculty of Sciences and Technologies, Sultan Moulay Slimane University, Beni Mellal, Morocco
[2]Laboratory of Engineering and Applied Technologies, Higher School of Technology, Sultan Moulay Slimane University, Beni Mellal, Morocco

Article Info

Article history:

Received Feb 1, 2024
Revised Jul 11, 2024
Accepted Jul 17, 2024

Keywords:

Advanced driver assistance system
Image processing
K-means++
Recognition
Traffic sign detection
YOLOv7

ABSTRACT

Traffic sign detection is a crucial aspect of advanced driver assistance systems (ADAS) for academic research and the automotive industry. seeing that accurate and timely detection of traffic signs (TS) is essential for ensuring the safety of driving. However, TS detection methods encounter challenges like slow detection speed and a lack of robustness in complex environments. This paper suggests addressing these limitations by proposing the use of the you only look one version 7 (YOLOv7) network to detect and recognize TS in road scenes. Furthermore, the k-means++ algorithm is used to acquire anchor boxes. Additionally, a tiny version of YOLOv7 is used to take advantage of its real-time and low model size, which are required for real-time hardware implementation. So, we conducted an experiment using our proprietary Morocco dataset. According to the experimental results, YOLOv7 achieves 85% in terms of mean average precision (mAP) at 0.5 for all classes. And YOLOv7-tiny obtains 90% in the same term. Afterward, a recognition system for the prohibitive class using the convolutional neural network (CNN) is trained and integrated inside the YOLOv7 algorithm; its model achieves an accuracy of 99%, which leads to a good specification of the prohibitive sign meaning.

Corresponding Author:

Imane Taouqi
Department of Electrical Engineering MiSET Team, Faculty of Sciences and Technologies, Sultan Moulay Slimane University
523, Beni Mellal, Morocco
Email: imane.taouqi@usms.ma

1. INTRODUCTION

Traffic sign detection systems have a crucial role in reducing accidents and incidents, helping drivers anticipate and react to potentially dangerous situations. However, traffic signs (TS) detection algorithms are categorized into two groups: detection using traditional approaches and detection utilizing deep learning. Traditional detection algorithms for TS are structured into three steps: select the region, extract the feature, and classify the feature [1]. The main goal of the first step is to generate a diverse set of proposed regions where the object may localize in the feed image. This step can be performed by using the selective search algorithm [2]. Afterward, algorithms for feature extraction, such as histograms of oriented gradients [3], are employed to extract features of TS from proposed regions. Those extracted features are classified using supervised machine learning algorithms, for example, support vector machines [4] and random forests [5]. Nevertheless, the traditional object detection method has some limits in terms of real-time processing. It can also require frequent retraining if datasets or scenarios change, which can be costly in terms of time and hardware resources. Nowadays, the emergence of the artificial intelligence era on the computer vision side and technological advancements [6] such as sophisticated sensors and high-resolution cameras have considerably facilitated real-time interaction with the environment, enabling fast and more accurate road decisions. The only need is to be trained with a large amount of complex environment image data to obtain accurate detection and recognition results. In this term, object detection methods have introduced more integrated approaches that combine region selection, feature extraction, and classification into a single pipeline process; some of those algorithms are composed of two stages, and others only one stage [7]. The main difference here is that one-stage detectors directly predict bounding boxes and class labels in a single step without a prior region proposal step, which involves lower computation and faster speed. While the two-stage detection algorithms follow a two-step process, The first stage generates region proposals, and the second stage refines these proposals by predicting the final bounding box and class label, which leads to a higher computation cost and low rapidity. which is undesirable in the TS detection context [8]. Therefore, the primary challenge for a traffic sign detecting system is to understand and react to traffic signs correctly and instantly, even in the presence of factors like occlusion, complex backgrounds, and variable lighting conditions.

Jagadeesh and Sree [5] splits the TS recognition process into two stages: the first involves preprocessing the input image (filtering, color space conversion) and color thresholding; the second stage involves feature extraction using histogram of oriented gradients (HOG) and classification using random forest; this method achieves an accuracy score of 95.59%; however, there is no information regarding inference, despite the fact that inference is a crucial metric to assess the algorithm's performance; additionally, the methods employed may have some limitations in terms of processing each step independently in real-time, on the other hand, encounter various problems, such as occlusion and detection of the inappropriate sign as traffic sign (e.g. commercial sign). Srivastava *et al.* [9] used YOLOv4 by adjusting the relevant parameters and retraining the model on a manually annotated data set to detect TS. This method attained an accuracy of 90.7%. Tian [10] proposed a new model called feature fusion single shot detector (FFSSD), which improves the SSD algorithm [11]. By fusing features from low dimensions into features that are in high dimensions, inserting the SE model, and adopting the focal loss to improve and fix the original loss function, this method reaches a precision of 92.79%. both methods provided no information about inference. Selçuk *et al.* [12] used the YOLOv4 network to detect traffic speed limit signs. It is achieving an accuracy of 95% and 42 frames per second (FPS). Nonetheless, this method covers only one class of TS. Megalingam *et al.* [13] proposes the refined mask region-based convolutional neural network (mask R-CNN) for detection of traffic signs with an attaint precision of 97.08%. noting that the accuracy is higher using the series of R-CNN, because they use two stages for detection to get more accurate results; otherwise, the inference is not sufficient to get real-time responses in critical situations. Xia *et al.* [14] presents in his paper DSRA-DETR, a novel approach focused on improving multiscale detection performance, this method reaches an average precision of 76.13% which still be improved especially when we are in a complex environment or the signs are occulted, the accuracy will decrease.

Despite the improvement mentioned above, detecting, and recognizing traffic signs using object detection algorithms, still face challenges in achieving an effective trade-off between detection and recognition speed, and accuracy, because of different shapes and meanings, to tackle these concerns, we will investigate in this paper how to deal with the mentioned limits above, especially when we are in a complex environment, also kept the real-time response which is needed in critical situations and deployment into an embedded processor. The present article suggests the use of YOLOv7 [15], which is both precise and functions in real-time, with anchor box adaptation using k-means++ [16]. The proposed algorithm for detection is trained over our proprietary Morocco dataset; and split into 5 classes based on the shapes and colors to have a good generation ability for TS detection, in case we have images that do not belong to the trained datasets. Then, for the prohibitive class, a recognition module based on CNN is integrated into the YOLOv7 algorithm to recognize the meaning of the current detected prohibit sign. The major contributions are:
- Use the latest version of the YOLO algorithm to take advantage of their amelioration as a one-stage detector.
- Train the YOLOv7 algorithm on Moroccan highway scene datasets.
- Split the datasets into 5 classes derived from their shapes and colors for precise detection.
- Adapt the anchor box using k-means++ algorithm.
- Use the tiny version of YOLOv7 to perform real-time detection that is reliable for deployment in embedded hardware.
- Explore the result of YOLOv7, add a recognition module for the prohibitory signs class, which is the critical category that can limit road scene accidents.

The following sections of this paper are structured as follows: First section is introduction. In the second section, the proposed method, with a description of the used algorithm, model architecture, and methodology. The third section presents an experiment evaluation to discuss datasets, evaluation indicators, and test results using different parameters and environment complexity. At the end, a conclusion is provided to recapitulate this paper.

2. PROPOSED METHOD

The proposed system to detect and recognize traffic signs is accomplished in three major steps: first, the TS detection model has been built using the proposed algorithm YOLOv7 with k-means++ adaptation to raise the accuracy of the algorithm, then another model was added, so it will be responsible for TS recognition and classification according to the meaning of the current detected sign, and finally the association of detection and recognition models. Figure 1 describes the workflow of the proposed system, where the input image taken from the road scene environment is fed to the YOLOv7 network to localize the relevant TS by a bounding box. A different sign class may be detected depending on the road scene sign (e.g. mandatory and warning), if the detected signs are associated with prohibitory class, it will be cropped and feed to the second trained network to recognize the meaning of current detected critical prohibit sign.

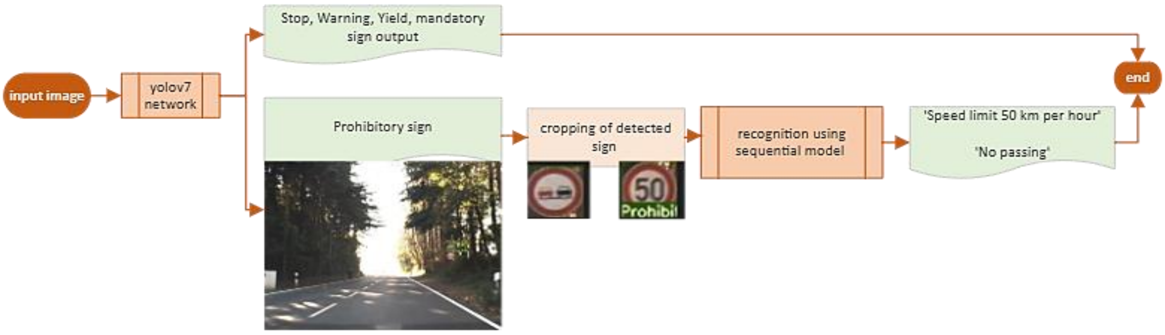

Figure 1. Workflow of TS detection and recognition system

3. METHOD

3.1. YOLOv7 algorithm

YOLO network comprises four main sections: input, backbone, neck, and head, followed by the prediction layer [17]. The input layer in YOLOv7 is responsible for receiving the image or video to be processed and ensuring uniform resizing of the input image to 640×640×3, thus meeting the specified input size criteria for the backbone, which is the second part of the network, which extracts crucial features from the input image. the backbone used in the YOLOv7 algorithm is based on Darknet-53, which is a deep convolution network with 53 layers. It uses residual blocks to facilitate deep learning and avoid the disappeared gradient problem. The backbone takes an input image and passes it through several convolution layers to extract features from different scales. Then YOLOv7 uses a neck to compile the retrieved feature maps and builds feature pyramids to merge the extracted features at different scales. The neck consists of several convolution and aggregation layers, which allow contextual information from different spatial resolutions to be integrated. Once the features have been merged by the neck [18], they are transmitted to the "head" of the network to predict bounding boxes and corresponding object classes. The head is composed of several convolutional and linear layers that generate the final predictions [19]. This last is performed using anchors and grids. Anchors are predefined bounding boxes of different sizes and shapes [20], which are used to predict the coordinates of bounding boxes. Grids divide the image into several regions, and each region is responsible for predicting the classes of objects present in that region. For each grid, the network generates predictions for several potential bounding boxes, using linear regressions to predict the coordinates (x, y, width, height) of each bounding box relative to the corresponding anchor. Then, a SoftMax activation function [21] is applied to the network outputs to get the probabilities of the different classes of objects. Finally, a confidence threshold is applied to predictions to filter detections with confidence less than a predefined threshold. The remaining detections are then grouped, and non-maximum suppression is applied to eliminate redundant or overlapping detections.

3.2. Anchors box adaptations using k-means++ algorithm

Existing anchor boxes that are used in YOLOv7 with COCO datasets are not suitable for our dataset, so before training our datasets using the YOLOv7 network, it is relevant to perform k-means++ clustering, to analyze and select the three most suitable anchor boxes for each scale and format to enhance model accuracy. The following points describe the performed steps by k-means++ algorithm:

– Initialization: randomly select k points in the data space as the initial cluster centers.
– Points allocation: each data point is assigned to the cluster with the closest center. This is usually based on the Euclidean distance between the point and the center of the cluster.
– Update cluster centers: recalculate the centers of each cluster by taking the average of the points belonging to this cluster.
– Repeat steps 2 and 3: repeat the steps for assigning points and updating centers until the cluster centers converge to a stable position.
– Convergence: the algorithm converges when cluster centers no longer change or change negligibly [22].

At the end of the algorithm, each data point is assigned to a specific cluster, forming separate groups. Cluster centers represent the central characteristics of each cluster. Once we have the centers of each partitioning, we classify them to group them into small, medium, and large sets of Anchor. This gives us three anchors:

– [4.11939038,3.90441712, 7.00466399,11.5332484, 13.51595867,5.38801812]
– [15.13099576,17.66111785, 20.26743326,9.84297648, 25.76287195,23.66278607]
– [27.33875207,13.42246189, 34.86696371,18.5344746, 42.62157509,23.67647803].

3.3. Sequential model for prohibitory signs recognition

To recognize different signs of prohibited class, we used a sequential model for object recognition. It is a machine learning model that processes data sequentially, typically used in recognizing objects within images. The goal of object recognition is to identify and classify objects or patterns within the visual data. The architecture used is described in Figure 2 showing layer type, output shape, and number of parameters.

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 28, 28, 16)	448
conv2d_1 (Conv2D)	(None, 26, 26, 32)	4640
max_pooling2d (MaxPooling2D)	(None, 13, 13, 32)	0
batch_normalization (BatchNormalization)	(None, 13, 13, 32)	128
conv2d_2 (Conv2D)	(None, 11, 11, 64)	18496
conv2d_3 (Conv2D)	(None, 9, 9, 128)	73856
max_pooling2d_1 (MaxPooling2D)	(None, 4, 4, 128)	0
batch_normalization_1 (BatchNormalization)	(None, 4, 4, 128)	512
flatten (Flatten)	(None, 2048)	0
dense (Dense)	(None, 512)	1049088
batch_normalization_2 (BatchNormalization)	(None, 512)	2048
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 43)	22059

Total params: 1,171,275
Trainable params: 1,169,931
Non-trainable params: 1,344

Figure 2. CNN architecture for traffic sign classification

3.4. Training setup and datasets

All tests in this paper were executed under an identical experimental environment: Colab Google to train and test the YOLOv7, tiny YOLOv7 detector using GPU (Nvidia SMI, Cuda version 12 torch 2.0.1 + cu118 CUDA:0, Tesla T4). we used German traffic sign detection benchmark (GTSDB) dataset [23]; given that most of their road signs align with those used in Moroccan TS [24]. Then, we adapt the data containing 842 images, by removing images that do not correspond to data from Morocco and adding other different images captured in the Moroccan road scene in different weather conditions. We divided the signs into 5 classes: stop, mandatory, yield, warning, and prohibitory. We use Roboflow to label TS datasets. Subsequently, data augmentation techniques were employed on the labeled datasets, to finally have 2,163 images for training.

For the recognition module, we used German traffic sign recognition benchmark (GTSRB) datasets. that contain more than 40 classes, we adapted it for only prohibitory signs by deleting unwanted classes, to have 17 classes at the end with the following names: ‘Speed limit 20 km per hour’, ‘Speed limit 30 km per hour’, ‘Speed limit 50 km per hour’, ‘Speed limit 60 km per hour’, ‘Speed limit 70 km per hour’, ‘Speed limit 80 km per hour’, ‘End of speed limit 80 km per h’, ‘Speed limit 100 km per hour’, ‘Speed limit 120 km per hour’, ‘No passing’, ‘No passing vehicles over 3.5 tons’, ‘Stop’, ‘No vehicles’, ‘Vehicles over 3.5 tons prohibited’, ‘No entry’, ‘End speed and passing limits’, ‘End of no passing’, ‘End no passing vehicles over 3.5 tons’, to recognize and extract information from prohibitive class. The training parameters setup of the experiment’s training procedure of YOLOv7 detection and sequential model for recognition are shown in Tables 1 and 2 respectively with their values.

Table 1. Training parameters for YOLOv7

Parameter	Value
Epochs	160
Weight_decay	0.0005
Momentum	0.937
Learning rate	0.01
Batch size	16
Image size	640×640×3
Workers	8

Table 2. Training parameters for recognition

Parameter	Value
Epochs	30
Batch_size	32
Optimizers	Adam
Image size	30×30×3
Learning rate	0.001

4. RESULTS AND DISCUSSION

4.1. Evaluation indicators:

In this paper, three metrics recall, accuracy, and mean average precision (mAP) are employed to assess the model’s accuracy, while model size and FPS are used to evaluate the model’s speed [25].

- Accuracy(P): Accuracy is the degree of precision of the model in identifying only relevant objects. It corresponds to the ratio of true positives (TP) to all detections made by the model [25].

$$Accuracy = \frac{TP}{TP + FP}$$

- Recall: The recall measures the percentage of correct detections against all instances of real objects. It is a measure of the quality of both positive and negative predictions. The higher the recall, the fewer false negatives there are [26].

$$Recall = \frac{TP}{TP + FN}$$

- Inference or FPS: Inference measures the time taken to perform detection on a single image. It is a measure of the speed of the detection algorithm. A minor inference time is usually preferable for real-time applications.

– Mean average precision-50 (mAP50): considers both accuracy and recall to reflect model performance. This metric is obtained by calculating the mean accuracy (AP) at the intersection over union (IOU) threshold 50 for each category, and then averaging all AP values across all categories [26].

AP is derived by computing the area under each precision-recall (P-R) curve, with precision plotted on the x-axis and recall on the y-axis. The formulas for calculating AP and mAP are as:

$$AP = \int_0^1 p(r)dr\,, \qquad mAP = \frac{\sum_{i=1}^{C} APi}{C}$$

Here, $p(r)$ represents the precision-recall curve and C denotes the number of categories.

4.2. Results discussion

This study investigated a relevant detection and recognition algorithm over different classes, environment severity, and occulted traffic signs, that can proficiently balance the speed and accuracy of the model. While earlier studies have explored TS detection on limited classes and focused on accuracy rather than real-time detection, also some research focuses only on detection or recognition separately. The proposed method in this study tended to use an efficient algorithm that combines both accuracy and real-time processing. And both detection and recognition of traffic signs. Our study suggests that the use of k-means++ increases the accuracy of the algorithm to generate relevant anchor boxes and then use a tiny version of YOLOv7 to take its real-time processing advantage, after that if the detected class is prohibitory, the recognition module will specify what the prohibited sign means. Although the training of the recognition module described in Figure 2 using the parameter of Table 2. gives an accuracy of 99%. For the detection module, different experiments were conducted using two backbone networks, to demonstrate and compare the backbone network's effects on detection accuracy and speed. the YOLOv7 backbone network which has 459 layers, and the tiny YOLOv7 backbone network which contains only 263 layers, also the effect of anchors box adaptation using k-means++ is tested. Table 3 displays the outcomes of the experiments. Table 4 shows a comparison between the proposed method conducted in this paper and other state-of-the-art TS detection.

Experiments showed that the use of tiny YOLOv7 with k-means++, trained on augmented datasets that consider different situations (blurring, complex environments), not only improved the accuracy of the original YOLOv7 algorithm but also significantly increased the speed of the model, reduced memory usage, and met the demand requirement for real-time gesture recognition. Also, by comparing our method to the previous research mentioned in Table 4, It can be seen that the use of YOLOv7 with k-means++ to adapt anchor box coordinates increases the accuracy detection to 97% compared to all mentioned methods in Table 4, although employing a tiny network as the foundational backbone greatly reduced the model size to 12.3 MB and inference to 11.2 (ms) compared to YOLOv4 and faster R-CNN [27]. However, Table 5 shows more details about the results of the proposed Tiny YOLOv7 algorithm in the five categories of TS.

The results show that employing tiny YOLOv7 in our context is more accurate regarding its execution time. Thus, embedded on a vehicle calculator can performed in real time because of the lower calculation charge of the tiny version. However future studies will be held to confirm the outcomes of this paper. By implementing the proposed algorithm in an embedded card which is limited in memory and resources not like the computer.

Table 3. Evaluation of YOLOv7-Tiny and YOLOv7 algorithm

Model used	Accuracy	Recall	mAP@.5	Model size	Inference (ms)
Default YOLOv7	0.852	0.833	0.866	74.8 MB	19.3
YOLOv7+k-means++	0.881	0.791	0.857	74.8 MB	22.5
Tiny YOLOv7	0.867	0.769	0.824	12.3 MB	11.5
Tiny YOLOv7+k-means and augmented data	0.969	0.802	0.903	12.3 MB	11.2

Table 4. Comparison to other state-of-art networks

Method	Accuracy	Recall	mAP@.5	Model size	Inference (ms)
FFSSD [10]	92.79%	83.52%	-	-	-
DSRA-DETR [14]	76;13%	-	98.12%	-	-
YOLOv5x [27]	95.1%	86.5%	75%	-	-
Faster R–CNN [27]	43.26%	-	-	-	166.67
YOLOv4 [28]	59.88%	-	-	-	28.57
Ours	97.00%	80%	90%	12.3MB	11.2

Table 5. Evaluation of YOLOv7-tiny over the 5 class

Classes	Labels	Accuracy	Recall	mAP@.5
All	146	96%	80%	90%
Prohibitory	27	95%	78%	86%
Mandatory	68	100%	97%	98%
Stop	3	99%	67%	86%
Warning	17	99%	82%	92%
Yield	31	92%	77%	88%

4.3. Test environment of traffic signs images

The results found while testing our proposed system, are displayed in Figure 3, which is correctly localized mandatory, prohibitory, warning, yield, and stop signs. Afterward, shows the label of prohibit signs if they exist. The system demonstrated impressive accuracy, successfully classifying prohibitory TS with high precision. The displayed images are taken at different parts of the day and in different Morocco cities. Which demonstrates the accurate result without dependency on specified cases.

Subsequently, while evaluating the YOLOv7 detector on TS that does not belong to training datasets, or there is an occlusion, it attempts to identify their classes by looking for similitude with existing TS. because each class from the five classes has its characteristics. Nevertheless, our detector can detect signs that are not included in training datasets with an accuracy of 80%. In other terms, our trained algorithm detection demonstrates strong generalization ability by handling all found signs in road scenes as tested on different images in Figure 4.

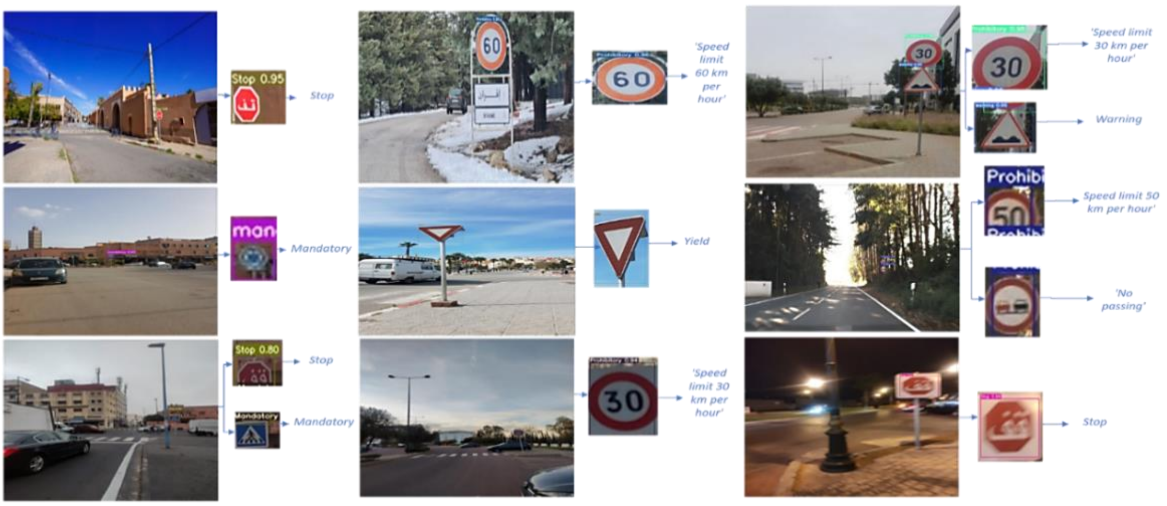

Figure 3. Different traffic sign detection and recognition results

Figure 4. Test on images does not belong the trained datasets

To summarize, the whole model detection and recognition system exhibited robust performance across different scenarios, showcasing its effectiveness in detecting, and categorizing a multitude of TS and classifying prohibitory classes, especially speed limit signs. The results underscore the reliability of the recognition system, making it a promising tool for enhancing road safety through improved sign detection and recognition.

5. CONCLUSION

This paper involved the detection of TS using the YOLOv7 and YOLOv7-tiny detection algorithm, to demonstrate its ability to deliver more accurate results in real-time, however as shown in the above results, it is capable of detecting signs in different angles, occlusion, and complex environments. Moreover, employing k-means++ improves the overall performance of the detection by adapting the anchor boxes, as they suggest potential locations for objects in the image. This facilitates the task of predicting bounding boxes and object classes while enabling the detection of different object sizes and shapes in the image. The experiments of the YOLOv7 algorithm were carried out using our own adapted Moroccan datasets. Also, the integration of a recognition module for critical prohibitory signs in road scenes. Hence, future research efforts will aim to deploy the trained algorithm into embedded hardware.

REFERENCES

[1] R. Ranjbarzadeh *et al.*, "MRFE-CNN: multi-route feature extraction model for breast tumor segmentation in mammograms using a convolutional neural network," *Annals of Operations Research*, vol. 328, no. 1, pp. 1021–1042, Sep. 2023, doi: 10.1007/s10479-022-04755-8.

[2] P. N. Chowdhury, A. K. Bhunia, A. Sain, S. Koley, T. Xiang, and Y.-Z. Song, "What can human sketches do for object detection?," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2023, pp. 15083–15094, doi: 10.1109/CVPR52729.2023.01448.

[3] B. Bhattarai, R. Subedi, R. R. Gaire, E. Vazquez, and D. Stoyanov, "Histogram of oriented gradients meet deep learning: a novel multi-task deep network for 2D surgical image semantic segmentation," *Medical Image Analysis*, vol. 85, 2023, doi: 10.1016/j.media.2023.102747.

[4] A. Kurani, P. Doshi, A. Vakharia, and M. Shah, "A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting," *Annals of Data Science*, vol. 10, no. 1, pp. 183–208, Feb. 2023, doi: 10.1007/s40745-021-00344-x.

[5] B. Jagadeesh and D. V. Vidhya Sree, "Detection and recognition of traffic sign boards using random forest classifier," *Review of Computer Engineering Research*, vol. 9, no. 3, pp. 135–149, 2022, doi: 10.18488/76.v9i3.3109.

[6] S. Wang and X. He, "The development and application of computer vision technology in the era of artificial intelligence," *Journal of Artificial Intelligence Practice*, vol. 5, no. 3, 2022, doi: 10.23977/jaip.2022.050301.

[7] J. Kang, S. Tariq, H. Oh, and S. S. Woo, "A survey of deep learning-based object detection methods and datasets for overhead imagery," *IEEE Access*, vol. 10, pp. 20118–20134, 2022, doi: 10.1109/ACCESS.2022.3149052.

[8] Y. Wu, F. Meng, Y. Qin, Y. Qian, F. Xu, and L. Jia, "UAV imagery based potential safety hazard evaluation for high-speed railroad using Real-time instance segmentation," *Advanced Engineering Informatics*, vol. 55, Jan. 2023, doi: 10.1016/j.aei.2022.101819.

[9] V. Srivastava, S. Mishra, and N. Gupta, "Automatic detection and categorization of road traffic signs using a knowledge-assisted method," *Procedia Computer Science*, vol. 218, pp. 1280–1287, 2023, doi: 10.1016/j.procs.2023.01.106.

[10] P. Tian, "FFSSD: Feature fusion SSD algorithm with SE on traffic sign detection," *Procedia Computer Science*, vol. 208, pp. 338–346, 2022, doi: 10.1016/j.procs.2022.10.048.

[11] Q. Liu, J. Bi, J. Zhang, X. Bu, and N. Hanajima, "B-FPN SSD: an SSD algorithm based on a bidirectional feature fusion pyramid," *The Visual Computer*, vol. 39, no. 12, pp. 6265–6277, Dec. 2023, doi: 10.1007/s00371-022-02727-4.

[12] S. Selçuk, S. Beker, and Ö. F. Boyraz, "Development of a traffic speed limit sign detection system based on Yolov4 network," *Advances in Artificial Intelligence Research*, vol. 3, no. 2, pp. 66–75, Oct. 2023, doi: 10.54569/aair.1184569.

[13] R. K. Megalingam, K. Thanigundala, S. R. Musani, H. Nidamanuru, and L. Gadde, "Indian traffic sign detection and recognition using deep learning," *International Journal of Transportation Science and Technology*, vol. 12, no. 3, pp. 683–699, Sep. 2023, doi: 10.1016/j.ijtst.2022.06.002.

[14] J. Xia, M. Li, W. Liu, and X. Chen, "DSRA-DETR: an improved DETR for multiscale traffic sign detection," *Sustainability*, vol. 15, no. 14, Jul. 2023, doi: 10.3390/su151410862.

[15] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," *arxiv.org/abs/2207.02696*, Jul. 2022.

[16] H. Li and J. Wang, "Collaborative annealing power k-means++ clustering," *Knowledge-Based Systems*, vol. 255, Nov. 2022, doi: 10.1016/j.knosys.2022.109593.

[17] F. A. Kateb, M. M. Monowar, M. A. Hamid, A. Q. Ohi, and M. F. Mridha, "FruitDet: Attentive feature aggregation for real-time fruit detection in orchards," *Agronomy*, vol. 11, no. 12, Nov. 2021, doi: 10.3390/agronomy11122440.

[18] B. R. Chang, H.-F. Tsai, and C.-W. Hsieh, "Accelerating the response of self-driving control by using rapid object detection and steering angle prediction," *Electronics*, vol. 12, no. 10, May 2023, doi: 10.3390/electronics12102161.

[19] G. D. Raj and B. Prabadevi, "Steel strip quality assurance with YOLOV7-CSF: a coordinate attention and SIoU fusion approach," *IEEE Access*, vol. 11, pp. 129493–129506, 2023, doi: 10.1109/ACCESS.2023.3333894.

[20] R. Zhang, Y. Tian, Z. Xu, and D. Liu, "Design of anchor boxes and data augmentation for transformer-based vehicle localization," *Journal of Visual Communication and Image Representation*, vol. 90, Feb. 2023, doi: 10.1016/j.jvcir.2022.103711.

[21] R. Mahima, M. Maheswari, S. Roshana, E. Priyanka, N. Mohanan, and N. Nandhini, "A comparative analysis of the most commonly used activation functions in deep neural network," in *2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Jul. 2023, pp. 1334–1339, doi: 10.1109/ICESC57686.2023.10193390.

[22] M. Zubair, M. A. Iqbal, A. Shil, M. J. M. Chowdhury, M. A. Moni, and I. H. Sarker, "An improved K-means clustering algorithm towards an efficient data-driven modeling," *Annals of Data Science*, Jun. 2022, doi: 10.1007/s40745-022-00428-2.

[23] V. Sichkar, "Traffic signs detection by YOLOv3, OpenCV, Keras," *Kaggle*, 2021. Accessed: Aug. 19, 2023. [Online], Available:

https://www.kaggle.com/code/valentynsichkar/traffic-signs-detection-by-yolo-v3-opencv-keras/data
[24] I. Taouqi, A. Klilou, K. Chaji, and A. Arsalane, "YOLOv2 implementation and optimization for Moroccan traffic sign detection," 2023, pp. 837–843.
[25] K. Jiang *et al.*, "An attention mechanism-improved YOLOv7 object detection algorithm for hemp duck count estimation," *Agriculture*, vol. 12, no. 10, Oct. 2022, doi: 10.3390/agriculture12101659.
[26] R. Padilla, S. L. Netto, and E. A. B. da Silva, "A survey on performance metrics for object-detection algorithms," in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, Jul. 2020, pp. 237–242, doi: 10.1109/IWSSIP48289.2020.9145130.
[27] N. Youssouf, "Traffic sign classification using CNN and detection using faster-RCNN and YOLOv4," *Heliyon*, vol. 8, no. 12, Dec. 2022, doi: 10.1016/j.heliyon.2022.e11792.

BIOGRAPHIES OF AUTHORS

Imane Taouqi received the master's degree in communication system and embedded electronic from the Abdelmalek Essaadi University, Morocco, in 2020. Department of Electrical Engineering. Her area of research includes artificial intelligence, computer vision, image processing and embedded systems. She is currently a Ph.D. student in the fields of embedded systems and artificial intelligence in Faculty of Sciences and Technology, University of Sultan Moulay Slimane, Beni Mellal Morocco. She can be contacted at email: imane.taouqi@usms.ma.

Abdessamad Klilou received an engineer's degree in 2010 and a Ph.D. degree in 2016 from the University of Cady Ayyad, Marrakech, Morocco. Since 2017, he is a professor at the Department of Electrical Engineering in the Faculty of Sciences and Technology, University of Sultan Moulay Slimane, Beni Mellal Morocco. His area of research includes on parallel and real time optimization of signal processing algorithms on multi-core and multi-processors parallel machine. He can be contacted at email: a.klilou@usms.ma.

Kebir Chaji is a professor of Electrical Engineering at Faculty of Sciences and Technology, University of Sultan Moulay Slimane, Beni Mellal Morocco, since 1996. He can be contacted at email: kebir.chaji@gmail.com.

Assia Arsalane received an engineer's degree in electrical engineering from the National School of Applied Sciences of Khouribga in 2014 and a Ph.D. degree in 2019 from the University of Hassan I, Settat, Morocco. Since 2020, she is an assistant professor in the Department of Mechatronics at the Higher School of Technologies, University of Sultan Moulay Slimane, Beni Mellal Morocco. Her area of research includes artificial intelligence, machine vision, image processing, data analysis, and embedded systems. She can be contacted at email: arsalan.assia@gmail.com.