Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-02 | June-2022

Traffic Participants Detection and Classification Using YOLO Neural Network

Fahmida Sultana Mim[1], S. M. Naimur Rhaman Sayam[2], Md. Tanvir Amin[3]
[1]Faculty, Rabindra Maitree University, Kushtia-Bangladesh. fahmida.bauet@gmail.com
[2]Faculty, Rabindra Maitree University, Kushtia, Bangladesh, sayam.bauet.eee@gmail.com
[3] Faculty, Rabindra Maitree University, Kushtia, Bangladesh, aminmd.tanvir@gmail.com

DOI: 10.5281/zenodo.7771342

ABSTRACT

One of the most important requirements for the next generation of traffic monitoring systems, autonomous driving technology and Advanced Driving Assistance Systems (ADAS) is the detection and classification of traffic participants. Although in the areas of object detection and classification research, tremendous progress has been made, we focused on a specific task of detecting and classifying traffic participants from traffic scenarios. In our work, we have chosen a Deep Convolutional Neural Networks – YOLOv4 (You Only Look Once Version 4), a object detection algorithm to detect and classify traffic participants accurately with fast speed. The main contribution of our work included: firstly, we generate a custom image dataset of traffic participants (Car, Bus, Truck, Pedestrian, Traffic light, Traffic sign, Vehicle registration plate, Motorcycle, Ambulance, Bicycle wheel). After that, we run K-means clustering on the dataset to design an anchor box that is utilized to adapt to various small and medium scales. Finally, we train the network for the mentioned objects and test it in several driving conditions (including daylight, low light, high traffic, foggy, rainy environment). The results showed cutting-edge performance with a mean Average Precision (mAP) of up to 65.95% and a speed of about 54 ms.

Keywords: Deep Convolutional Neural Networks, Traffic participants, YOLOv4, Object detection, Classification

Cite as: Fahmida Sultana Mim, S. M. Naimur Rhaman Sayam & Md. Tanvir Amin. (2022). Traffic Participants Detection and Classification Using YOLO Neural Network. *LC International Journal of STEM (ISSN: 2708-7123)*, 3(2), 9–18. https://doi.org/10.5281/zenodo.6844336

INTRODUCTION

Since the rebirth of the convolutional neural network in 2012, traffic participant detection and classification have become a prominent topic. It has a wide range of uses in computer vision, including traffic monitoring, autonomous driving technologies, ADAS, traffic density estimation, and many others. Deep CNN-based object detection algorithms have become more resilient and successful in recent years i.e. Faster R-CNN [1], SSD [2], CenterNet [3], YOLO [4], etc. Those algorithms are a new milestone of traffic participants detection and classification. Apart from the traditional algorithms, Deep CNN algorithms conduct representational learning on a large amount of data. At the same time, because the model is scalable, it is more flexible in practical application. There are two types of deep CNN algorithms available i.e. two-stage and one-stage. The two-stage object detectors, such as Faster R-CNN, Mask R-CNN, etc. use Region Proposal Network (RPN) and CNN together to detect and classify an object. So that high computing force is needed in two-stage detectors. But in one-stage

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-02 | June-2022

object detectors, such as SSD, YOLO, etc. can detect and classify objects directly through the network. As a result, one-stage detectors have a faster response time.

As an example of a one-stage object detector algorithm, the latest version of YOLO (YOLOv4) shows outstanding accuracy and speed of object detection tasks. It forecasts targets of various sizes on 3 different scales, which makes the object detection task more suitable for small objects. In our work, we built a custom image dataset of traffic participants (Car, Bus, Truck, Pedestrian, Traffic light, Traffic sign, Vehicle registration plate, Motorcycle, Ambulance, Bicycle wheel) from the Google open images dataset using its OIDv4 toolkit. We also run K-means clustering on the dataset to determine suitable priors automatically, rather than choosing anchor boxes by hand. After successfully trained the network with custom images, we render that the YOLOv4 algorithm can detect and classify traffic participants in several driving conditions i.e. daylight, low light, high traffic, foggy, etc. We also provide the evaluation of the qualitative performance of the network.

The residue of the paper is: Section 2 describes related work. Section 3 presents the methodology, which describes the methods, such as network architecture and the dataset we labeled. Section 4 describes the findings and analysis. Conclude with Section 5.

RELETED WORK

A noteworthy effort has been carried out in the past years of Traffic Participants detection and classification. Some of them are featured manually [5], which is less effective than deep learning algorithms [6]. Deep learning algorithms directly conduct feature extracting methods from original images. As mentioned in the previous section, researchers categorized Deep CNN algorithms into two classes and different researchers show the performance of different classes in their work. In 2014, R-CNN [7] was proposed by the researchers as the first region-based two-stage object detector. And they presented Fast R-CNN in 2015 [8]. We can use Fast R-CNN to do both object detection and bounding box regression in one network. In [1], researchers claimed the Faster R-CNN detector, which is a elevated version of Fast R-CNN and capable of performing end-to-end and almost real-time detection. Mask R-CNN [9], which included a section for calculate an object mask with Faster R-CNN, was recently introduced. Nowadays researchers have applied them in traffic scenarios to detect and classify individual traffic participants i.e. pedestrian, traffic sign, etc. [10] [11] [12] [13]. They have achieved high accuracy in two-stage object detectors, but the speed needs to be enhanced to perform in real-time. To overcome this limitation, one-stage object detectors were demonstrated. Researchers suggested YOLO [4], which advocates a grid-centric multiscale region. This method significantly improves detection efficiency while sacrificing accuracy for small objects to accomplish real-time requirements. In 2016, SSD [2] was introduced, it overcomes the drawbacks of YOLO and Faster R-CNN and achieve excellent accuracy and speed, especially for small objects. On the other hand, there have some limitations of SSD as follows: (1) As the resolution of the input image increases, so does the number of default bounding boxes. (2) Each layer of the network's boxes has a different scale and ratio, which should be manually adjusted. YOLOv2 [14] adds batch normalization after each convolution layer, performs multiscale training, and uses k-means clustering in the training dataset to automatically determine acceptable prior probabilities for detection efficiency. And further improve the accuracy. By updating the feature extractor of YOLOv2 from darknet19 to darknet53, researchers released YOLOv3 [15]. YOLOv3 enhances the detection accuracy for small objects in particular while keeping the speed. In recent years, YOLOv4 [16] is introduced as the latest version YOLO series. The main network part of YOLOv4, CSPDarknet53, is built on the Darknet53 and CSPNet [17]. The CSPDarknet53 network not only improves learning capabilities but also ensures detection accuracy while reducing computational time for faster processing. Researchers also applied these one-stage object detectors in traffic scenarios to detect and classify individual traffic participants and got better results than two-

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-02 | June-2022

stage object detectors [18] [19] [20]. But most of the researchers do not focus on the specific task of detecting and classifying all the traffic participants from traffic scenarios with a single network and how it performs in several driving conditions. We focus on that specific task and show how effectively YOLOv4 works.

METHODOLOGY

This section describes the overall detection and classification process of YOLOv4 on input images or videos. Also describes the overall network architecture of YOLOv4 and how it makes the model more efficient in traffic participants detection and classification. Dataset information is also mentioned in this section.

A. Flow Diagram

YOLOv4 follows a certain flow method to detect and classify traffic participants. Figure 1 shows the flow diagram of YOLOv4 algorithm. Firstly, YOLOv4 takes a video from the traffic scenario and extracts a single image from it or it can take a single image directly. After that, the model divides that image into 608×608 grids. Then, using that image as an input, CNN generates a tensor that represents:

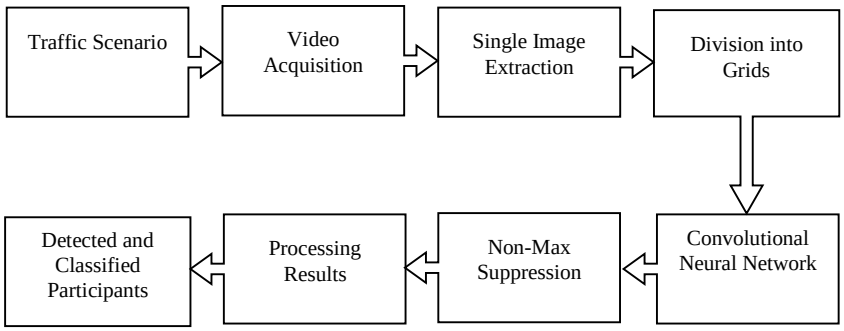

Fig. 1. Flow diagram of YOLOv4 Network

(1) Predicted bounding box coordinates and position (2) Probability that each bounding box contains an object (3) Probability that each object in the bounding box is a participant of a particular class. Let 'y' is that mentioned tensor,

$$y = \begin{bmatrix} p_c \\ b_x \\ b_y \\ b_h \\ b_w \\ c_1 \\ c_2 \\ c_3 \\ . \\ . \end{bmatrix} \qquad (1)$$

In equation (1), 'p_c' denotes the presence of an object in the grid and can take either a 0 or a 1 value, the coordinates of an object in a specific grid are defined by the variables 'b_x' and 'b_y', the percentage height and width of the whole grid cell are defined by 'b_h' and 'b_w' correspondingly and finally 'c_1', 'c_2', 'c_3' ... defines the classes i.e. car, bus, person, etc. After determining the class probability, the model determines the class-specific confidence scores. The confidence score represents the predicted accuracy of the bounding box.. The following is a definition of confidence:

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-02 | June-2022

$$C_{IJ} = Pr \times (object) \times IoU_{pred}^{truth} \quad (2)$$

Here, C_{IJ} denotes the confidence in the J^{th} bounding box of the I^{th} grid cell, whereas IoU_{pred}^{truth} denotes the correspondence between the references and predicted bounding boxes. When numerous bounding boxes identify the same target, YOLOv4 selects the optimal bounding box using the non-max suppression (NMS) method [22].

B. Dataset

We built a custom image dataset of 10 different participants (Car, Bus, Truck, Pedestrian, Traffic light, Traffic sign, Vehicle registration plate, Motorcycle, Ambulance, Bicycle wheel). We use Google open image dataset and its OIDv4 toolkit to collect labeled images for training the YOLOv4 network. By running K-means clustering on the training dataset, we got 9 anchor frames and select the anchor frame as ([15, 28], [38, 75], [60,174], [116, 94], [121,293], [224,173], [239,410], [425,283], [506,504]). These anchor frames not only affect the convergence but also enhance the performance of the network. Finally, to prepare the test dataset, images were observed and labeled manually. Different angles of the same object and complicated traffic situations were captured. We added all-weather images of traffic participants, which will ensure the accurate measurement of the performance of YOLOv4.

(a)

(b)

Fig. 2. Randomly chosen images from dataset (a) training images; (b) test images

In our dataset, 17317 labeled images are used for training (percentage of contribution 88.75%) and 2195 labeled images are used for testing (percentage of contribution 11.25%). Some randomly chosen train and test images are given in figure 2.

C. Network Architecture

Darknet53 is received by the combination and integration of ResNet and YOLOv3 accordingly. It helps to solve the deep network gradient problem. Based on Darknet53, YOLOv4 built CSPDarkNet53 as the backbone of the network, taking into account the excellent learning capabilities of the Cross-stage Partial Network (CSPNet) [17].

CSPNet not only improves YOLOv4's learning ability but also reduces computation and memory costs while assuring accuracy. It is done by integrating gradient changes into functional maps end-to-end.

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-02 | June-2022

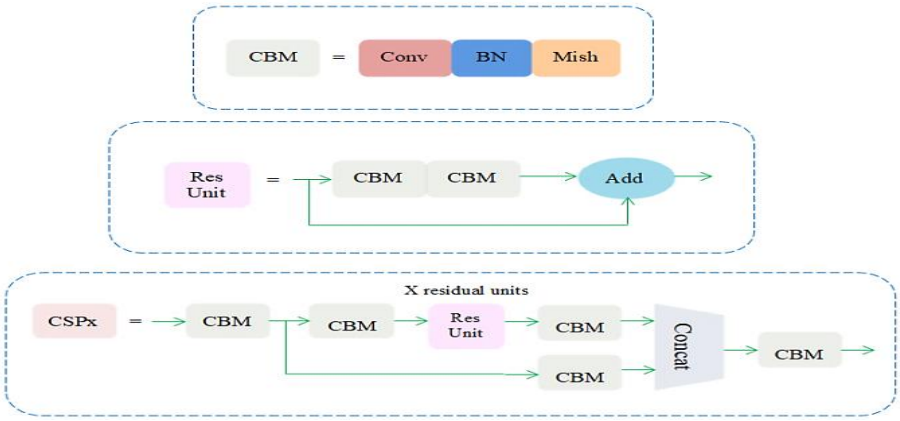

Fig. 3. CSPDarknet53 Module Structure

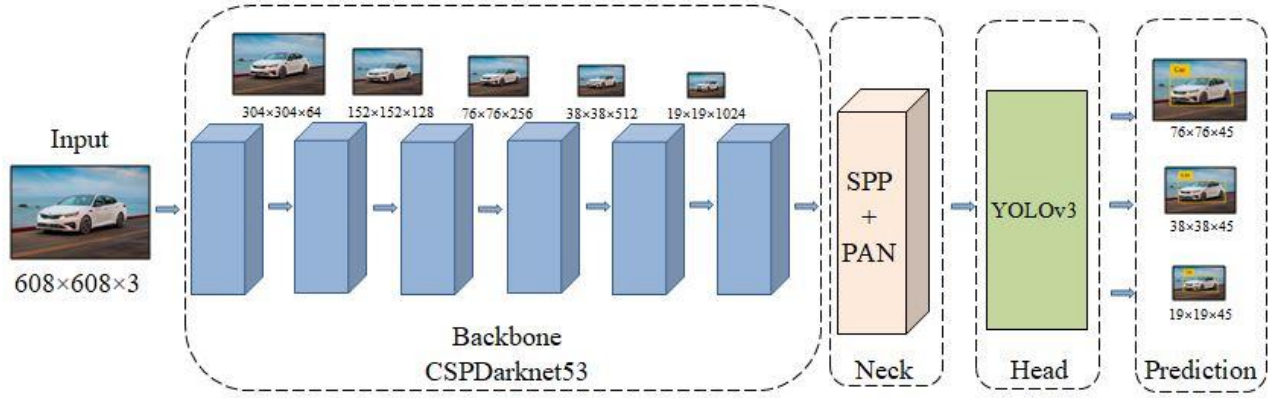

Fig. 4. YOLOv4 Network Architecture

As shown in Figure 3, each convolutional layer contains Convolution, Batch Normalization (BN), and Mish. Also, each residue module contains one shortcut and two convolutional layers. In the layers, there are multiple duplicate residual modules.

In addition, as a neck, YOLOv4 inserts SPP+PAN between the core network and the output layer. SSP blocks increase the acceptance fields, extract the most relevant features, and slightly slow down network operations. Also, instead of FPN, PANet [21] is employed, which helps shorten the information path, accurately localize low-level signals, enhance the feature pyramid and improve the overall feature layer. Finally, YOLOv3 is utilized in the head of the network, which improves the mean average precision and object detection ability, especially for small partcipants. Figure 4 shows the network architecture of the model.

RESULT AND ANALYSIS

We trained the network for 13000 iterations. After,every 100 iterations, checked the performance metrics (mAP, precision, recall, f1 score and avg IoU) and got a better result at 8000th iteration. After 8000 iterations the model started overfitting. The training results at different iteration are shown in Table 1. And, figure 5 shows that our model is improving the detection accuracy untill 8000th iteration.

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-02 | June-2022

After successfully trained the network, we tested it at several driving conditions. The robustness of the YOLOv4 network is noticed, it detects and classifies each class accurately with high speed.

TABLE I. TRAINING RESULTS

Iteration	mAP	Precision	Recall	F1 score	Avg IoU
1000	41.49%	0.49	0.34	0.40	34.69%
1500	53.84%	0.57	0.46	0.51	44.03%
2900	62.61%	0.61	0.52	0.56	47.67%
3900	61.95%	0.58	0.52	0.55	46.93%
5000	64.19%	0.63	0.49	0.55	51.33%
6000	65.97%	0.57	0.58	0.57	46.56%
7000	63.17%	0.57	0.54	0.56	45.13%
8000	65.95%	0.62	0.53	0.57	51.48%

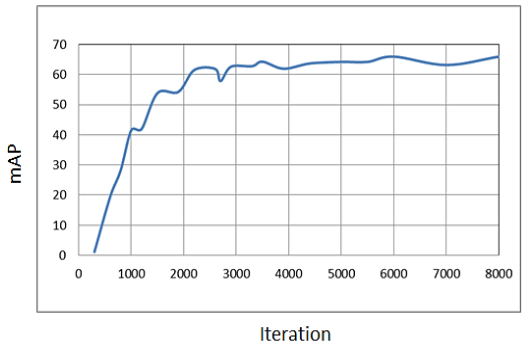

Fig. 5. Mean average precision (mAP) value is gradually increasing with iteration

(a)

```
/content/headlight-vs-foglight-1024x683.jpg: Predicted in 54.239000 milli-seconds.
car: 100%
car: 97%
```

(b)

Fig. 6. Performance at foggy condition (a) detection and lebelling (b) detection speed and confidence score

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-02 | June-2022

(a)

```
/content/HNL-rain.jpg: Predicted in 54.066000 milli-seconds.
car: 78%
car: 37%
traffic light: 92%
car: 42%
traffic light: 81%
car: 99%
car: 99%
car: 100%
```

(b)

Fig. 7. Performance at rainy condition (a) detection and labeling; (b) detection speed and confidence score

In figure 6 & 7, there are few participants in foggy weather and rainy weather, YOLOv4 detects them with comprehensive confidence levels and high computation speed. In order to verify the performance of YOLOv4, it was compared with Faster R-CNN and SSD algorithm. Both the algorithms were trained with the same dataset mentioned above. After a successful comparison, several points could be noticed. The confidence score is higher in Faster R-CNN, but it takes too much time in computing and bounding boxes were smaller than the object. In SSD algorithm, it gives a lower confidence score and sometimes detection errors were occurred. YOLOv4 managed to detect objects with higher confidence scores and accurate bounding boxes.

(a)

(b)

(c)

(d)

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-02 | June-2022

(e) (f)

Fig. 8. Detection results (a,b) YOLOv4; (c,d) Faster R-CNN; (e,f) SSD

Figure 8 shows the comparative detection results. The superiority of YOLOv4 is demonstrated. The scenario changes significantly when we switch from image to video inputs. The coordinates of objects in a video will now change in real-time. Even so, objects are continuously detected and labeled by YOLOv4 with a high level of confidence.

CONCLUSION

One of the main goals of this paper was to assess the appropriateness of You Only Look Once (YOLO) version 4 for application in traffic participants detection activities at several driving conditions i.e. high traffic conditions, low traffic conditions, foggy weather conditions, etc. Most of the researchers do not focus on the specific task to detect and classify traffic participants but on object detection in general. Our motive is to focus on that specific task by using the Deep CNN-based framework YOLOv4 for better accuracy and speed. Despite this, it is a considerably new approach or concept that we used in this study, which carried out mAP 65.95% with a better speed of around 54 milliseconds. As there was an excessive number of small objects in the test dataset and we couldn't manually label them properly, also, few objects were blocked by other identified items in the test dataset, and the precision and recall scores did not reach the expected high levels.

REFERENCES

[1] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, June 2017.

[2] W. Liu et al., "SSD: Single Shot MultiBox Detector," in *European Conference on Computer Vision (ECCV)*, 2016, pp. 21-37.

[3] K. Duan et al., "CenterNet: Keypoint Triplets for Object Detection," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), 2019, pp. 6568-6577.

[4] J. Redmon, S. Divvala, R. Girshick, and A. Farhad, "You only look once: Unified real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 779-788.

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-02 | June-2022

[5] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005.

[6] Yu-nan Dong and Guang-sheng Liang, "Research and Discussion on Image Recognition and Classification Algorithm Based on Deep Learning," in *International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, Taiyuan, China, 2019.

[7] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based Convolutional Networks for Accurate Object Detection and Segmentation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015.

[8] R. Girshick, "Fast R-CNN," in *IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, 2015.

[9] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 386-397, 1 Feb 2020.

[10] D. Tabernik and D. Skočaj, "Deep Learning for Large-Scale Traffic-Sign Detection and Recognition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 4, pp. 1427-1440, April 2020.

[11] Z. Zuo, K. Yu, Q. Zhou, X. Wang, and T. Li, "Traffic Signs Detection Based on Faster R-CNN," in *2017 IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW)*, Atlanta, GA, USA, 2017, pp. 286-288.

[12] H. Zhang et al., "Pedestrian Detection Method Based on Faster R-CNN," in *2017 13th International Conference on Computational Intelligence and Security (CIS)*, Hong Kong, China, 2017, pp. 427-430.

[13] A. Datta et al., "Road Object Detection in Bangladesh using Faster R-CNN: A Deep Learning Approach," in *2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)*, 2020, pp. 348-351.

[14] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 6517-6525.

[15] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," in *arXiv:1804.02767*, 2018.

[16] A. Bochkovskiy, C.Y. Wang, and H.Y. Mark, "YOLOv4: Optimal Speed and Accuracy of Object Detection," in *arXiv:2004.10934*, 2020.

[17] C. Wang et al., "CSPNet: A New Backbone that can Enhance Learning Capability of CNN," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2020, pp. 1571-1580.

[18] H. Zhang et al., "Real-Time Detection Method for Small Traffic Signs Based on Yolov3," *IEEE Access*, vol. 8, pp. 64145-64156, 2020.

[19] A. Ćorović, V. Ilić, S. Đurić, M. Marijan, and B. Pavković, "The Real-Time Detection of Traffic Participants Using YOLO Algorithm," in *2018 26th Telecommunications Forum (TELFOR)*, 2018, pp. 1-4.

[20] F. Ahmad, L. Ning, and M. Tahir, "An Improved D-CNN Based on YOLOv3 for Pedestrian Detection," in *2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP)*, 2019, pp. 405-409.

[21] S. Liu, Q. Lu, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 8759–8768.

[22] A. Neubeck and L. V. Gool, "Efficient Non-Maximum Suppression," in *18th International Conference on Pattern Recognition (ICPR'06)*, 2006.

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-02 | June-2022

BIOGRAPHY

[1]*Fahmida Sultana Mim completed her bachelor's program in Electrical and Electronic Engineering from Bangladesh Army University of Engineering and Technology (BAUET). Presently she is serving as a Lecturer at Rabindra Maitree University located at Kushtia, Bangladesh.*
E-mail: ***fahmida.bauet@gmail.com***

[2]*S. M. Naimur Rhaman Sayam completed his bachelor's program in Electrical and Electronic Engineering from Bangladesh Army University of Engineering and Technology (BAUET). Presently he is serving as a Lecturer at Rabindra Maitree University located at Kushtia, Bangladesh.*
E-mail: ***sayam.bauet.eee@gmail.com***

[1]*Md. Tanvir Amin completed his master's program in Information and Communication Technology from Bangladesh University of Professionals. He completed his bachelor's program in Electrical and Electronic Engineering from International University of Business Agriculture and Technology. Presently he is serving as a Lecturer at Rabindra Maitree University located at Kushtia, Bangladesh.*
E-mail: ***aminmd.tanvir@gmail.com***