ISSN 1870-4069

# Proposal of a CNN-Based Approach for Traffic Signal Detection

Madaín Pérez Patricio[1], Carlos Alexis Ramírez Mendoza[1], Germán Rios Toledo[1], Juan Antonio de Jesús Osuna Coutiño[2]

[1] Tecnológico Nacional de México, Campus Tuxtla Gutiérrez, Mexico

[2] Instituto Nacional de Astrofísica, Optica y Electrónica, Mexico

{M14270620, madain.pp, german.rt}@tuxtla.tecnm.mx, osuna@inaoep.mx

**Abstract.** Traffic Signal Detection (TSD) is an important module in autonomous vehicles and Driver Assistance Systems (DAS). Although there are several approaches to TSD, in most cases, these are based on only the 2D localization, i.e., these systems do not provide their information to other drivers or future travel routes. On the other hand, some works only focus on a specific signal (traffic light) causing a bias into the signal set. To address these problems, an alternative is to use deep networks, image metadata, and Information Technology (IT). Motivated by the latter, we propose a methodology for traffic sign detection and geolocation using a CNN-based approach. This strategy combines the abstraction power of deep learning with IT and metadata information. For that, our methodology has three steps. First, traffic sign detection provides the location and classification of the road signs. Second, we use the image metadata to obtain the geolocation. Third, the information technology step presents the geospatial and classification information into an Application Programming Interface (API). Also, we evaluate this methodology in public images and a proposed dataset with metadata information. The quantitative experiments were conformed from the signal detection in two urbanized environments (open imagesV6 and proposed dataset). For that, we analyzed two labels of road signs (Traffic light, Traffic sign). Also, our road sign detection had an average recall of 0.89, i.e., considering the ground-truth, we recognized 89%.

**Keywords:** Convolutional neural network, image processing, trafic signal.

## 1 Introduction

The population in cities is increasing continuously, and they require more efficient services. The concept of smart cities try to deal with these requirements incorporating technologies as internet of things, artificial intelligence, cloud computing, among others. These technologies are combined to providing citizens the best place to live.

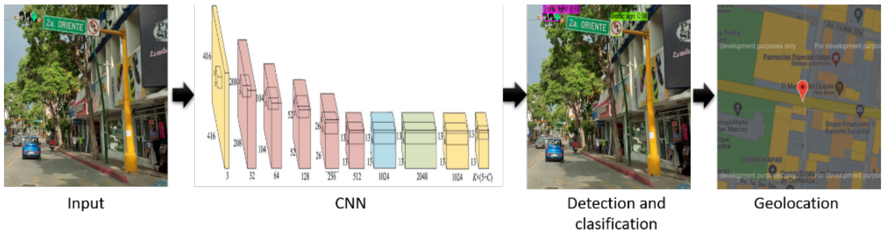



**Fig. 1.** Methodology.

Transport, health, and living are the most common research subjects in the literature related to smart cities [24]. For the transport area and autonomous vehicles, the main idea is the efficient traffic management to reduce time lost, fuel consumption, and risks of injury.

This problem has been treated from different perspectives. For example, vehicles have been equipped with Advanced Driver-Assistance Systems (ADAS) that includes powerful computer vision systems used to detect pedestrian, other vehicles and traffic signs. On the other hand, smart traffic control systems establish rules of control traffic to help vehicles as ambulances to reduce the lost time [11].

One of the most important objects of the urban infrastructure is the traffic signs. They are designated to regulate the traffic and to fulfill requirements of safety and comfort to drivers. So, an efficient method is required for automatic detection and positioning of the traffic signs. This subject has been extensively studied in the last few years. Autonomous driving and intelligent transportation systems also require the precise identification of traffic signs.

Automatic traffic signs detection and positioning is a challenging task because there are several problems such as: variable light conditions, non-standard form and size signals, and weather changes. An approach to deal with these problems proposes the use of smart traffic signs. They can send wireless messages to vehicles placed in the neighborhood of the sign.

This approach requires efficient protocols of communication between smart traffic signs and vehicles. Another approach uses computer vision based systems [27]. The algorithms and technologies used in computer vision based systems are efficient.

Three stages are generally involved in traffic sign detection: detection, tracking and recognition [16]. Nevertheless, the accuracy achieved in the traffic sign detection is 95.71%. For several applications, this error can be acceptable, but only one traffic sign non-detected can produce injury in the drivers [26].

In this paper, we propose a method suited for embedded processing. The system has two parts: a computer vision based system to traffic sign detection and a cloud based system to traffic sign positioning. The traffic sign detected is recorded in a database and can be updated over the time.

This information would be transmitted to drivers to reduce risks. More efficient services are required as the population in cities is increasing. With the use of traffic signs in the adequate positioning and number, the traffic can be reduced in the cities while the trash collection requires efficient management of trucks, workers, and truck rides.

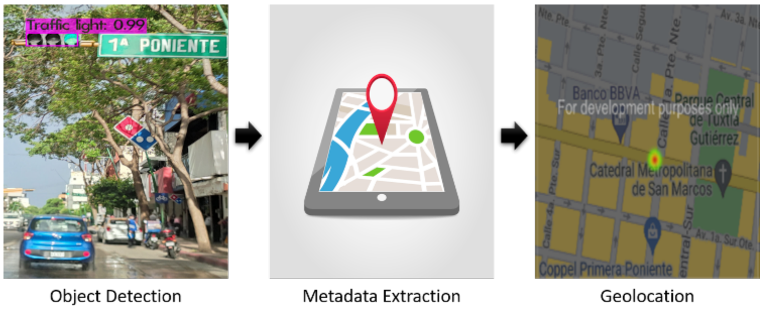



**Fig. 2.** Scheme of metadata extraction.

The city planners must have a detailed and complete description of the urban infrastructure to take well decisions. The information that city planners can take into count must include the kind and precise position of available urban infrastructure.

Several efforts have been realized to standardize traffic signs around the world. Nevertheless, each country and each city can define their particular traffic sign. Therefore, there is no universal method for traffic sign identification. Fig. 2 shows an example of geolocation using metadata extraction.

The proposed method uses a smartphone which includes a camera and a sensor of positioning. A dataset has been created and used to train a neural network. Images acquired by the smartphone are processed in a personal computer which yields a map where the traffic signs are identified and positioned. This method is suited for embedded systems.

## 2 Related Work

Variable light conditions, non-standard forms of the signs, changes of size of the observed signs because of distance from the sensor, partial occlusions, and weather changes make the traffic sign detection a challenging task. A huge quantity of algorithms has been proposed to avoid these problems. To test algorithms proposed by the community, several datasets have been created [9, 12, 21, 22]. As traffic signs are different by each country and each city, a particular dataset would be required for each city that uses this technology.

Traffic signs recognition can be performed using different methods. Methods as proposed in [26], are based on color or shape recognition. These methods take advantage of fact that the color or shape of traffic signs are highly visible and contrast the surrounding neighborhood [25]. This kind of proposal has a time of processing reduced but present weakness in presence of light changes, rotation, and viewing angle.

[19] try to deal with occluded and attached traffic signs by estimating the shape of the arc described by the contour of the traffic sign. In [4], the color information and object properties are used to identify regions of interest that reduces the processing times of a support vector machine based classification algorithm. Other methods are learning based, like the proposed by [15, 17]. They are based on the use of convolutional neural networks, where they use a huge number of images to train it.

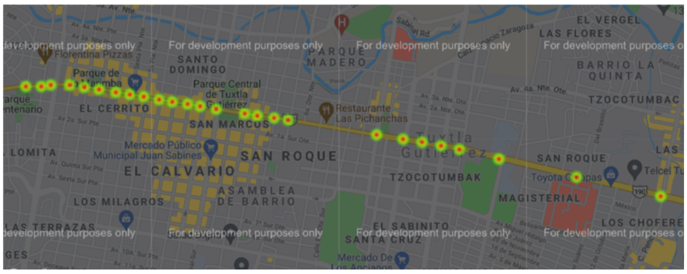



**Fig. 3.** Geolocation results.

For example, in [7], 40,000 images has been used to train the convolutional network. These methods are robust to light changes and changes of perspective but requires a high computational power. The use of models like the You Only Look Once (YOLO) yields results with a high throughput and real-time processing achieving 95.15% of precision [1]. To achieve traffic sign identification as long as possible, two cameras has been used in [6].

A wide-angle camera select traffic sign candidates while a narrow-angle camera is used to get a high-resolution image of the traffic sign candidate. A hardware implementation using Field Programmable Gate Arrays devices can yield results in real-time [20]. A preprocessing step has been applied to the image and traffic signs using a neural network. The use of simplified Gabor wavelets and convolutional neural networks can also achieve real-time performance [3, 18].

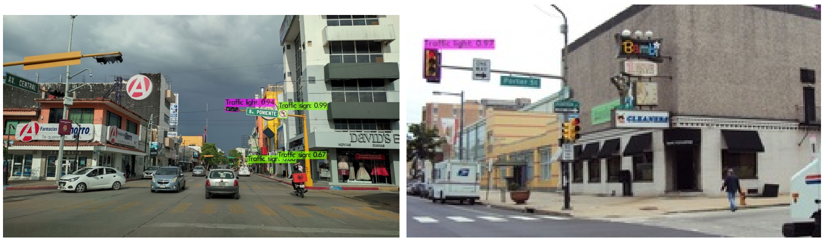
Reported results yields that the average processing time is 5.4 ms. Incorporating a Global Positioning System (GPS) device in the setup can yield interesting information for other kind of applications. In [13], road attributes has been inferred using traffic sign identification and GPS information. On the other hand, geo-tagged Google Street View images and a Gopro camera [27] has been used to traffic sign recognition a positioning.

GPS, inertial sensor, camera, and laser sensor are used in a van [10]. These results show that precision of detection is highly correlated to weather conditions. To reduce issues associated to traffic signs, wireless traffic signs has been developed [23]. They transmit information to road users that can be reproduced with auditive signals even if the traffic sign is not observed by the driver. Although there are several approaches to TSD, in most cases, these are based on only the 2D localization, i.e., these systems do not provide their information to other drivers or future travel routes.

On the other hand, some works only focus on a specific signal (traffic light) causing a bias into the signal set. In our case, unlike the related work we propose a methodology for traffic sign detection and geolocation using a CNN-based approach. This strategy combines the abstraction power of deep learning with IT and metadata information.

## 3 Proposed Method

This section presents the proposed methodology for traffic signal detection and geolocation. Our strategy combines the abstraction power of deep learning with Information Technology (IT) and metadata extraction.



(a) (b)

**Fig. 4.** The images show traffic signal detection using our methodology in the proposed dataset (a), and the open images V6 dataset (b).

For that, our methodology has three steps.First, traffic signal detection provides the location and classification of the road signs. Second, we use the image metadata to obtain the geolocation. Finally, the information technology step presents the geospatial and classification information into an Application Programming Interface (API). The schematic representation of the proposal is shown in Fig. 1.

### 3.1 Traffic Signal Detection

In this work, we use the Yolov4 architecture[2] for traffic signal detection. Although some works only focus on specific signal detection (traffic light). This approach causes a bias into detection, i.e., a variety of signals are omitted by the system. For that, our network learns two labels on urbanized environments (traffic light $\upsilon^1$ and signal $\upsilon^2$).

**Training set.** In the training set of traffic signal detection, we use urban images where the camera looks head-on to buildings. This set is formed of two datasets that provide different outdoor scenes (open images V6 [8], and a proposed dataset which can be accessed using the next link Dataset). In the open images V6 repository, we use 3000 images labeled with traffic lights and traffic signs.

On the other hand, our dataset has 2000 images labeled of the main avenues in Tuxtla city; between 13 street northwestern and Pencil street in Tuxtla, Chiapas, Mexico. For that, we use a Xiaomi Redmi Note 10 with a resolution of 4000$\times$3000 pixels. We apply data augmentation in the training step. For that, we transformed image set via mirroring.

On the other hand, we obtained the image number for training and test using the Pareto principle or 80/20 rule [**?**], (80% training, and 20% test). Finally, we use a Google colaboratory environment with 26 Gb of Ram, 150Gb storage and GPU Tesla P100-PCIE 16Gb.

**CNN.** The input of the CNN is an RGB image $\varPhi$. In this case, we train the YOLOv4 network to learn two labels on urbanized environments (traffic light $\upsilon^1$ and signal $\upsilon^2$). Also, our network uses a bounding box $\vartheta_i$ to delimit the elements in the image $\varPhi$. For that, we use two colors in the bounding box $\vartheta_i$. A purple bounding box $\vartheta_1$ delimits a traffic light $\upsilon^1$ On the other hand, a green bounding box $\vartheta_2$ delimits a traffic signal $\upsilon^2$, Fig. 4 shows examples of our detection. The training time of the CNN took roughly 10 hours.

**Table 1.** Routes of the proposed dataset.

| Route | Start | Finish |
|---|---|---|
| 1 | Ave. panamericana Km. 1080 | Blvd. Andres Sierra Rojas |
| 2 | Blvd. Andres Sierra Rojas | Rd. Juan Crispin |
| 3 | north fifth St | Libramiento norte |
| 4 | Libramiento norte | Blvd. Presa chicoasen |

**Table 2.** Signal detection evaluation.

| Images number | Precision | Recall | F-score |
|---|---|---|---|
| 1000 | 0.76 | 0.69 | 0.72 |
| 4000 | 0.87 | 0.86 | 0.86 |
| 6000 | 0.86 | 0.88 | 0.87 |
| 7000 | 0.88 | 0.89 | 0.88 |

### 3.2 Metadata Extraction

Metadata is defined as the information provided about one or more aspects of the data. In the case of the image metadata, this has a variety of information about the description of the picture, such as its origin, size, camera, GPS, aperture, shutter speed, among others.

In our methodology, we use the Global Positioning System (GPS) data of the images with signal detection, i.e., we use the metadata to assign geolocation of the signals (traffic light $\upsilon^1$ and signal $\upsilon^2$).

### 3.3 Information Technology

A geolocation system is an information technology solution that determines the location of an object in a physical or virtual environment. In our case, we use the Maps JavaScript [14] to show the geolocation in an API system. For that, we send to the API the GPS data of the images with signal detection (traffic light $\upsilon^1$ and signal $\upsilon^2$). This approach allows us to save and display the coordinates of traffic signals detected using the methodology.

Also, we can provide this information to other drivers or future travel routes. Fig. 3 shows an example of the traffic signal set detected in the JavaScript map. These traffic signals are located in the central avenue, between 13 street northwestern and Pencil street in Tuxtla city, Chiapas, Mexico.

## 4 Discussion and Results

In this section, we present the experiments of signal detection. For that, the problem was addressed as a classification problem (*Traffic light*, *Traffic sign*). On the other hand, we evaluate two datasets that provided different outdoor scenes (open images V6 [8] and proposed dataset). In our proposed dataset, we selected four routes of the main avenues in Tuxtla city, Chiapas, Mexico. Table 1 shows the different routes.

**Table 3.** Signal detection evaluation in the proposed dataset.

| Clase | Precision | Recall | F-score |
|---|---|---|---|
| Traffic light | 0.94 | 0.97 | 0.95 |
| Traffic sign | 0.82 | 0.81 | 0.81 |

The quantitative evaluation compares our detection with the ground truth of the signals detected. We used three measures ($recall$, $precision$ and $F - score$) **Eq. 1-3** based on the number of true positives ($Tp$), true negatives ($Tn$), false positives ($Fp$), and false negatives ($Fn$). The true positives $Tp$ count the number of signals whose label was predicted correctly w.r.t. the ground truth.

To count the number of true negatives $Tn$, we proceed as follows: suppose that we are interested in the traffic light, then all images with another classification than traffic light according to the ground truth, should have received any other predicted classification except traffic light; if that is the case, each of these detections is counted as true negatives.

The false positives $Fp$ correspond to all those detections whose label is incorrect. Finally, false negatives $Fn$ correspond to those image sections that should have received a specific label, but the prediction did not assign it correspondingly:

$$\text{recall} = \frac{Tp}{Tp + Fn}, \tag{1}$$

$$\text{precision} = \frac{Tp}{Tp + Fp}, \tag{2}$$

$$F - \text{score} = \frac{2}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}} = 2\frac{\text{recall} * \text{precision}}{\text{recall} + \text{precision}}. \tag{3}$$

In our first experiment, we analyzed the performance of our network in signals and traffic light detection. For that, we evaluate this experiment in our dataset. Table 3 shows the result of our approach for signal detection. The evaluation with traffic light showed a better performance since this signal has a shape standardized.

In our second experiment, we implement different training in our signal detection network. In the proposed dataset, we use 1000, 2000, 4000 images. In the open images V6 repository [8], we use 3000 images of the second to fourth training. On the other hand, we only use the proposed dataset in the first training. Also, in each case, we carry out 10,000 epochs. Table 2 shows the result of our approach for signal detection. The evaluation with 7,000 images showed the highest detection.

## 5 Conclusions

In this work, we have introduced a new methodology for traffic signal detection and geolocation using a CNN-based approach. Our strategy was to combine the abstraction power of deep learning with IT and metadata information.

For that, our methodology has three steps. First, traffic signal detection provides the location and classification of the road signs. Second, we use the image metadata to obtain the geolocation.

Third, the information technology step presents the geospatial and classification information into an API. The quantitative experiments were conformed from the signal detection in two urbanized environments (open images V6 [8] and proposed dataset). For that, we analyzed two labels of road signs (*Traffic light*, *Traffic sign*). Also, our road sign detection had an average $recall$ of 0.89, i.e., considering the ground-truth, we recognized 89%.

On the other hand, the road sign detection had an average $precision$ of 0.88, i.e., considering the classification, we classify 88.0% correctly. We should note that the proposed methodology involves a combination of techniques based on CNN with IT and metadata information. This methodology explores the abstraction power of deep learning and the information that provides geolocation technologies. In our opinion, our approach brings the best of the two worlds to address the difficult problem of traffic signal geolocation.

On the other hand, the results obtained in this research have demonstrated the feasibility of traffic sign detection and geolocation. Based on these results, we propose as future work to implement the proposed methodology with an informatic system to provide vehicle routes. In our opinion, this can increase the accuracy of route prediction time.

## References


1. Arief, R. W., Nurtanio, I., Samman, F. A.: Traffic Signs Detection and Recognition System using the YOLOv4 Algorithm. In: International Conference on Artificial Intelligence and Mechatronics Systems, pp. 1–6 (2021). DOI: 10.1109/AIMS52415.2021.9466006.
2. Bochkovskiy, A., Wang, C. Y., Liao, H. Y. M.: YOLOv4: Optimal speed and accuracy of object detection (2020). DOI: 10.48550/arXiv.2004.10934.
3. Cao, K., Wei, C., Gaidon, A., Arechiga, N., Ma, T.: Learning Imbalanced Datasets with Label-distribution-aware Margin Loss. In: 33rd Conference on Neural Information Processing Systems NeurIPS, pp. 1–12 (2019). DOI: 10.48550/arXiv.1906.07413.
4. Coțovanu , D., Zet, C., Foşalău, C., Skoczylas, M.: Detection of Traffic Signs Based on Support Vector Machine Classification using Hog Features. In: International Conference and Exposition on Electrical and Power Engineering, pp. 518–522 (2018). DOI: 10.1109/ICEPE.2018.8559784.
5. Craft, R. C., Leake, C.: The Pareto Principle in Organizational Decision Making. Management Decision, pp. 729–733 (2002). DOI: 10.1108/00251740210437699.
6. Gu, Y., Yendo, T., Tehrani, M. P., Fujii, T., Tanimoto, M.: A New Vision System for Traffic Sign Recognition. In: IEEE Intelligent Vehicles Symposium, pp. 7–12 (2010). DOI: 10.1109/IVS.2010.5548005.
7. Slam, M. T.: Traffic sign detection and recognition based on convolutional neural networks. In: International Conference on Advances in Computing, Communication and Control, pp. 1–6 (2019). DOI: 10.1109/ICAC347590.2019.9036784.
8. Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., Kamali, S., Popov, S., Malloci, M., Kolesnikov, A., Duerig, T., Ferrari, V.: The Open Images Dataset v4: Unified image classification, object detection, and visual relationship detection at scale. International Journal of Computer Vision, (2020). DOI: 10.1007/s11263-020-01316-z.

9. Lau, M. M., Lim, K. H., Gopalai, A. A.: Malaysia Traffic Sign Recognition with Convolutional Neural Network. In: IEEE International Conference on Digital Signal Processing, pp. 1006–1010 (2015). DOI: 10.1109/ICDSP.2015.7252029.
10. Lee, J. S., Yun, D. G.: The road traffic sign recognition and automatic positioning for road facility management. International Journal of Highway Engineering, vol. 15, no. 1, pp. 155–161 (2013). DOI: 10.7855/IJHE.2013.15.1.155.
11. Lee, W. H., Chiu, C. Y.: Design and Implementation of a Smart Traffic Signal Control System for Smart City Applications. Sensors, vol. 20, no. 2, pp. 508 (2020). DOI: 10.3390/s20020508.
12. Mathias, M., Timofte, R., Benenson, R., Van Gool, L.: Traffic Sign Recognition—how Far are We From the Solution? In: International Join Conference on Neural Networks, pp. 1–8 (2013). DOI: 10.1109/IJCNN.2013.6707049.
13. Mèneroux, Y., Le-Guilcher, A., Saint-Pierre, G., Hamed, M. G., Mustière, S., Orfila, O.: Traffic Signal Detection from In-vehicle GPS Speed Profiles using Functional Data Analysis and Machine Learning. International Journal of Data Science and Analytics, vol. 10, no. 1, pp. 101–119 (2020). DOI: 10.1007/s41060-019-00197-x.
14. Googel Platform: Maps javascript api www.maps-backend.googleapis.com. (2021)
15. Radu, M. D., Costea, I. M., Stan, V. A.: Automatic Traffic Sign Recognition Artificial Inteligence-deep Learning Algorithm. In: 12th International Conference on Electronics, Computers and Artificial Intelligence, pp. 1–4 (2020). DOI: 10.1109/ECAI50035.2020.9223186.
16. Ruta, A., Li, Y., Liu, X.: Real-time Traffic Sign Recognition from Video by Class-specific Discriminative Features. Pattern Recognition, vol. 43, no. 1, pp. 416–430 (2010). DOI: 10.1016/j.patcog.2009.05.018.
17. Sermanet, P., LeCun, Y.: Traffic Sign Recognition with Multi-scale Convolutional Networks. In: International Joint Conference on Neural Networks, pp. 2809–2813 (2011). DOI: 10.1109/ IJCNN.2011.6033589.
18. Shao, S., McAleer, S., Yan, R., Baldi, P.: Highly Accurate Machine Fault Diagnosis using Deep Transfer Learning. IEEE Transactions on Industrial Informatics, vol. 15, no. 4, pp. 2446–2455 (2018). DOI: 10.1109/TII.2018.2864759.
19. Soendoro, D., Supriana, I.: Traffic Sign Recognition with Color-based Method, Shape-arc Estimation and SVM. In: International Conference on Electrical Engineering and Informatics, pp. 1–6 (2011). DOI: 10.1109/ICEEI.2011.6021584.
20. Souani, C., Faiedh, H., Besbes, K.: Efficient Algorithm for Automatic Road Sign Recognition and its Hardware Implementation. Real-time Image Processing, vol. 9, no. 1, pp. 79–93 (2014). DOI: 10.1007/s11554-013-0348-z.
21. Stallkamp, J., Schlipsing, M., Salmen, J., Igel, C.: The German Traffic Sign Recognition Benchmark: a Multi-class Classification Competition. In: International Join Conference on Neural Networks, pp. 1453–1460 (2011). DOI: 10.1109/IJCNN.2011.6033395.
22. Stallkamp, J., Schlipsing, M., Salmen, J., Igel, C.: Man vs. Computer: Benchmarking Machine Learning Algorithms for Traffic Sign Recognition. Neural Networks, vol. 32, pp. 323–332 (2012). DOI: 10.1016/j.neunet.2012.02.016.
23. Toh, C. K., Cano, J. C., Fernandez-Laguia, C., Manzoni, P., Calafate, C. T.: Wireless digital traffic signs of the future. The Institution of Engineering and Technology Networks, vol. 8, no. 1, pp. 74–78 (2019). DOI: 10.1049/iet-net.2018.5127.
24. Toh, C. K., Sanguesa, J. A., Cano, J. C., Martinez, F. J.: Advances in Smart Roads for Future Smart Cities. In: Proceedings of The Royal Society A, vol. 476 (2020). DOI: 10.1098/rspa.2019.0439.
25. Wali, S. B., Abdullah, M. A., Hannan, M. A., Hussain, A., Samad, S. A., Ker, P. J., Mansor, M. B.: Vision-Based Traffic Sign Detection and Recognition Systems: Current Trends and Challenges. Sensors, vol. 19, no. 9, pp. 2093 (2019). DOI: 10.3390/s19092093.

26. Wali, S. B., Hannan, M. A., Hussain, A., Samad, S. A.: An Automatic Traffic Sign Detection and Recognition System Based on Colour Segmentation, shape matching, and SVM. Mathematical Problems in Engineering (2015). DOI: 10.1155/2015/250461.
27. Wu, Z.: Computer Vision-Based Traffic Sign Detection and Extraction: A Hybrid Approach using GIS and Machine Learning. Electronic Theses and Dissertations (2019)