# Open World Object Detection on Traffic Signs

Rusiru Thushara[1], Sudam Kalpage[1], Thilini Deshika[1], Gihan Jayatilaka[2], Salman Khan[3], and Roshan Ragel[1]

[1]Department of Computer Engineering, University of Peradeniya, Sri Lanka
[2]Department of Computer Science, University of Maryland, USA
[3]Mohamed bin Zayed University of AI, UAE

***Abstract*—In this research, we developed a system for accurately detecting and classifying traffic signs using advanced object detection, unsupervised clustering, and generative modeling techniques to improve road safety in autonomous driving systems. We trained a Faster R-CNN model with Feature Pyramid Networks (FPN) on the German Traffic Sign Detection Benchmark (GTSDB) dataset, focusing on only one class label "traffic-sign" to improve the model's generalization capabilities in localizing any traffic signal. The RPN and Fast R-CNN network of the Faster R-CNN model were used to generate a set of candidate object locations and refine RoIs. FPN was added as an extension to extract multi-scale features and detect traffic signs at various scales. After training the model on the GTSDB dataset, we performed object detection on a test set of traffic sign images and used a clustering model to group similar traffic signs. We also used unsupervised k-means clustering and Principal Component Analysis (PCA) for dimensionality reduction to improve the interpretability and efficient analysis of the model's outputs. Finally, we demonstrated the model's generalization capability by performing inference on the CeyRo dataset, without requiring the initially paired data, and used CycleGAN to create traffic sign manuals specific to a country. Overall, this approach improves the accuracy and generalization capabilities of the model for traffic sign detection, which is crucial for developing autonomous driving systems and enhancing road safety.**

***Index Terms*—Traffic sign detection, Computer Vision, Deep Learning**

## I. Introduction

Traffic sign detection can be identified as a vital application for autonomous driving systems because of its ability to provide necessary information for a vehicle to perceive the nature of the road and make decisions based on them. Having a proper understanding of the environment based on the critical information provided by the traffic signs is essential for safe and efficient driving.

Earlier approaches proposed for traffic sign detection [1]–[3] can be mainly identified as traditional image processing and classical machine learning-based approaches. Recently evolved deep learning-based approaches [4]–[6] have been able to outperform the traditional approaches in terms of their performance, adaptability, efficiency, and ability to handle complexity. However, achieving the generalization ability of the traffic sign detection systems is still a challenge, because of the vast diversity of traffic signs based on their geographical locations and road regulations.

In this work, we introduced a traffic sign detection system that tackles the generalization ability of the system according to different environments. We proposed an end-to-end deep learning-based traffic sign detection framework, which can approach real-world road scenarios. First, we developed a traffic sign detection model utilizing the German Traffic Sign Detection Benchmark (GTSDB) dataset [7], which is a current benchmark for traffic sign detection.

### A. GTSDB Dataset

For the training purposes of the proposed study, we utilized the GTSDB dataset, which is a widely used benchmark dataset for traffic sign detection. It was created by the Institute of Neuroinformatics at the University of Ulm in Germany and consists of more than 50,000 images of German traffic signs. The GTSDB dataset consists of images captured from German roads, featuring a variety of urban, suburban, and rural road scenes. The dataset contains 39,209 annotated images, each with a resolution of 1360 x 800 pixels. In total, the dataset includes 43,660 instances of traffic signs, belonging to 43 different traffic sign classes. The images are categorized into 11 distinct categories, including speed limit signs, stop signs, yield signs, and others. Overall, the GTSDB dataset provides a comprehensive and diverse set of images for training and evaluating traffic sign detection algorithms.

The GTSDB dataset has several innovative points that set it apart from other traffic sign detection datasets. First and foremost, the dataset tackles the rarely addressed challenges of existing traffic sign detection systems, such as detecting traffic signs in small-sized images, dealing with a large number of classes, and the complexity of road scenarios. Additionally, the dataset has been designed to capture a wide variety of road scenes, with a particular focus on different weather and lighting conditions, making it more diverse and representative of real-world situations. Another innovative aspect of the GTSDB dataset is the inclusion of a large number of traffic sign classes, with a total of 43 different classes, making it more challenging and useful for training and evaluating traffic sign detection algorithms. Finally, the dataset also provides annotations for different levels of detail, including bounding boxes and pixel-level segmentation, enabling a more fine-grained evaluation of detection and recognition performance.

For the detection model, the Faster R-CNN [8] model was used with Feature Pyramid Networks (FPN) [9] which enables multi-scale feature extraction which is required for traffic sign detection to detect small and distant signs

that may be difficult to detect with single-scale methods. The output of the detection model was then fed into an unsupervised clustering model which identifies clusters of traffic sign objects. To visualize the identified traffic sign clusters, Principal Component Analysis (PCA) was used as a dimensionality reduction method.

To demonstrate the generalization ability of our proposed work, we inference the model on CeyRo Traffic Sign and Traffic Light Dataset (ref) which constitutes images coming from a different context than the GTSDB dataset used for model training.

### *B. CeyRo Dataset*

For the inference purposes of the proposed study, we used the CeyRo Traffic Sign and Traffic Light Dataset [10] which consists of images taken from Sri Lankan road scenes. The CeyRo dataset can be considered a benchmark dataset for traffic sign detection, because of its composition of urban, suburban, and rural area scene images as mainly highlighted by the authors. There are 7,984 images in the dataset, each with a resolution of 1920 x 1080. These images include a total of 10,176 instances of traffic signs and traffic lights, belonging to 70 different traffic sign classes and 5 different traffic light classes. All the images are categorized into 7 superclasses, namely Danger Warning Signs (DWS), Mandatory Signs (MNS), Prohibitory Signs (PHS), Priority Signs (PRS), Speed Limit Signs (SLS), Other Signs Useful for Drivers (OSD), Additional Regulatory Signs (APR) and Traffic Light Signs (TLS).

The authors of the study have specifically focused on the rarely tackled challenges of existing traffic sign detection systems such as detecting traffic signs in small-sized images, dealing with a large number of classes, and the complexity of road scenarios. The dataset has also aimed to capture a huge variety of road scenes while focusing on different weather conditions and lighting conditions, which is an essential factor to consider for real-world applications of traffic sign detection systems. We particularly selected the CeyRo dataset for the inference, as it has different road infrastructures, traffic conditions, and cultural backgrounds compared to the GTSDB, which makes the proposed system closer to the open world settings.

Finally, the Cycle Generative Adversarial Network (CycleGAN) [11] was used to map the traffic sign objects localized by the model with its corresponding original images of the CeyRo dataset. This approach enabled the ability of training and evaluate a traffic sign detection model on CeyRo, without requiring the initially paired data.

The key contributions of our work are as follows.

1) The development of a traffic sign detection system improved with unsupervised clustering, and generative modeling techniques which can tackle real-world scenarios.
2) A Faster R-CNN and Feature Pyramid Networks (FPN) model architecture with improved accuracy and generalization capabilities for detecting traffic signs in different locations and environments.
3) The use of unsupervised k-means clustering and Principal Component Analysis (PCA) for dimensionality reduction, to improve the interpretability and efficient analysis of the model's outputs.
4) The application of the trained model to perform inference on the CeyRo dataset to demonstrate its generalization capability in the open world settings.
5) The use of a Cycle Generative Adversarial Network (CycleGAN) to create traffic sign manuals specific to a country, by mapping the localized traffic sign images to their corresponding original government traffic sign images.

## II. Methodology

In this study, we developed a system for accurately detecting and classifying traffic signs in real-world scenarios using cutting-edge object detection, unsupervised clustering, and generative modeling techniques. Our ultimate goal was to improve road safety by developing a system that can be used in autonomous driving systems.

We train a Faster R-CNN [8] model with Feature Pyramid Networks (FPN) [9] on the German Traffic Sign Detection Benchmark (GTSDB) dataset [7]. However, instead of using several class labels for the various traffic sign types, we focus on only one class label, "traffic-sign," to ensure that the model can generalize well for localizing any traffic signal.

The Faster R-CNN model is a popular object detection algorithm that utilizes two key components: a Region Proposal Network (RPN) and a Fast R-CNN network. The RPN generates a set of candidate object locations, or Regions of Interest (RoIs), and the Fast R-CNN network classifies and refines these RoIs.

The RPN operates by sliding a small network, called an anchor, over the convolutional feature map of the input image. At each position of the anchor, the RPN predicts two scores: the probability of an object being present and the coordinates of the bounding box around the object. The anchors with high objectness scores are selected as candidate RoIs.

The selected RoIs are then fed into the Fast R-CNN network, which consists of a set of fully connected layers that classify and refine the RoIs. The network takes the RoI and the convolutional feature map as input and outputs a class label and a refined bounding box.

To improve the accuracy of the Faster R-CNN model, Feature Pyramid Networks (FPN) can be added as an extension. FPN is a multi-scale feature extraction network that enables the detection of objects at various scales. It creates a pyramid of feature maps with different resolutions by using a top-down pathway and lateral connections to merge high-resolution features with low-resolution features.

By training the model to detect only one class label, we simplify the problem and make the model more robust in

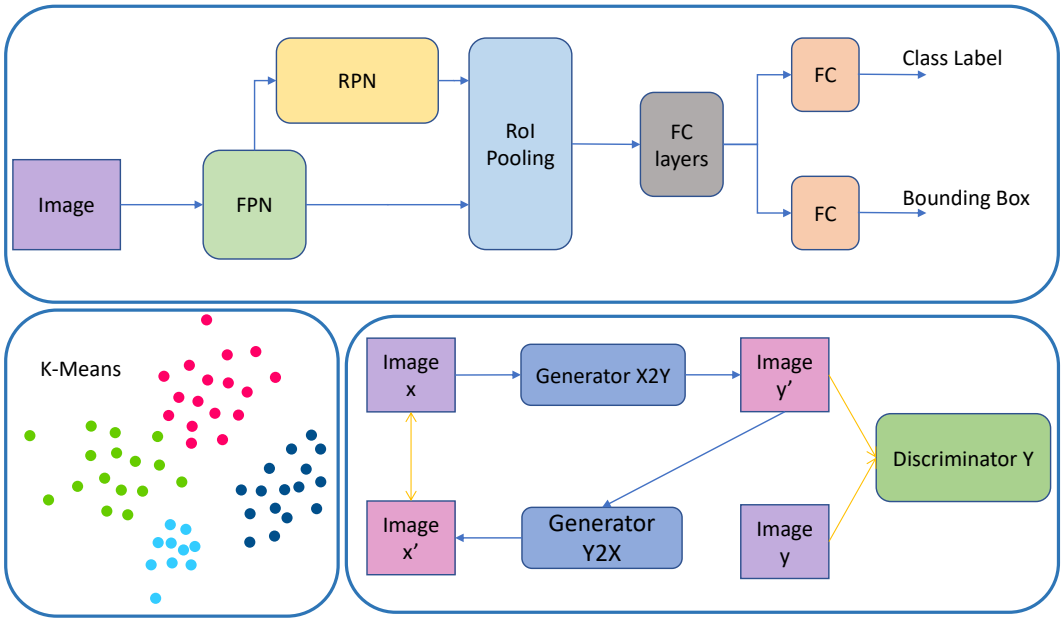


Fig. 1. Top: Faster R-CNN with FPN detection network that takes an image and outputs bounding boxes around objects of interest with class labels; Bottom-left: K-Means clustering; Bottom-right: CycleGAN to map road signs with original traffic sign.

detecting traffic signs in different locations and environments. This approach also enables the model to detect traffic signs that it may not have seen during training, which is important for practical use in real-world scenarios.

We utilize FPN, which allows the model to extract multi-scale features and detect traffic signs at various scales. This is particularly useful in detecting smaller traffic signs that may be harder to detect with a single scale feature extractor.

Overall, this approach to training a Faster R-CNN model with FPN on the GTSDB dataset with a simplified class label is aimed at improving the accuracy and generalization capabilities of the model for traffic sign detection, which is an essential component for developing autonomous driving systems and improving road safety.

After training the Faster R-CNN model with FPN on the GTSDB dataset, we performed object detection on a test set of traffic sign images. The detections generated by the model were then fed into a clustering model to group similar traffic signs together.

We utilized an unsupervised k-means clustering [12] approach, which is a popular technique for grouping data points into clusters based on their similarity. The k-means algorithm partitions the data into k number of clusters by minimizing the sum of the squared distances between each point and its assigned cluster center.

Once the clustering was completed, we used Principal Component Analysis (PCA) [13] for dimensionality reduction to visualize the clustered traffic signs in a 2D plot. PCA is a statistical technique that transforms the original high-dimensional data into a lower-dimensional representation while retaining the most important features of the data. By clustering the detected traffic signs, we aimed to improve the interpretability of the model's outputs and enable more efficient analysis of the detected traffic signs. Additionally, by using PCA for visualization, we were able to reduce the dimensionality of the data and provide a clear visual representation of the clustered traffic signs.

Overall, this approach of combining unsupervised clustering with dimensionality reduction using PCA is a useful tool for analyzing and visualizing the outputs of object detection models, such as Faster R-CNN with FPN. It can aid in improving the interpretability and efficiency of the model's outputs, which is important for practical use in real-world applications such as autonomous driving.

In addition to the experiments conducted on the GTSDB dataset, we also utilized the trained Faster R-CNN model with FPN to perform inference on a new dataset, the CeyRo dataset, which contains road sign data from Sri Lanka.

To cluster the detected traffic signs in the CeyRo dataset, we employed the same unsupervised k-means clustering approach as before. However, since the CeyRo dataset contains traffic signs from a different country, the clustering results were expected to differ from those obtained on the GTSDB dataset.

After clustering the traffic signs, we used a cycle generative

adversarial network (CycleGAN) [11] to map the localized traffic sign images to their corresponding original government traffic sign images of Sri Lanka. CycleGAN is a type of generative model that can learn the mapping between two domains without requiring paired training data.

Next, we leveraged the power of generative modeling with CycleGAN to map the localized traffic sign images to their corresponding original government traffic sign images of Sri Lanka. This enabled us to generate a high-quality dataset of traffic sign images for Sri Lanka that can be used for training and evaluating machine learning models for traffic sign detection.

What's particularly exciting about our approach is that it makes it easy for governments to create traffic sign manuals for their specific countries. By collecting random images of traffic signs in a particular country and giving specified labels for the clustered traffic sign instances from our model, governments can quickly and easily develop a comprehensive and accurate traffic sign manual.

Overall, our study demonstrates the potential of combining advanced computer vision and generative modeling techniques to solve real-world problems, such as improving road safety and developing autonomous driving systems.

## III. Results

Object detection is a crucial task in computer vision that involves identifying and localizing objects within an image or video. To evaluate the performance of an object detector, various metrics can be used, such as Average Precision (AP), Average Recall (AR), and F1 score. These metrics measure the precision and recall of the detector at different thresholds of intersection over union (IoU) between the predicted bounding boxes and ground truth bounding boxes.

Table I shows the evaluation metrics for the object detector used in the proposed traffic sign detection system. The metrics are computed for different sizes of objects and different IoU thresholds ranging from 0.50 to 0.95. The table provides a quantitative assessment of the performance of the detector and can be used to compare it with other similar systems.

| Area | Average Precision | Average Recall | F1 Score |
|---|---|---|---|
| Small | 0.618 | 0.751 | 0.678 |
| Medium | 0.576 | 0.711 | 0.636 |
| Large | 0.893 | 0.896 | 0.894 |

TABLE I
Results for IoU = 0.50:0.95

In addition to the object detector, the proposed traffic sign detection system also employs an unsupervised clustering mechanism to group traffic sign objects into clusters. Table II presents the evaluation metrics for the unsupervised clustering approach. The metrics include the clustering accuracy and the adjusted Rand index, which measures the similarity between the predicted clusters and the ground truth clusters. The table shows the evaluation metrics of a clustering algorithm on different numbers of initial clusters for a given task. The evaluation metrics used are accuracy, Rand score, adjusted Rand score, and normalized mutual info score.

The first row of the table shows the results for the maximum possible number of initial clusters, which is equal to the total number of classes (45) in the task. The second row shows the results for the optimal number of initial clusters (7) obtained from the initial elbow plot. The algorithm achieved lower scores for all evaluation metrics as compared to the maximum number of clusters. The third row shows the results for the number of known classes (19) in the task. The algorithm achieved better scores for all evaluation metrics as compared to the optimal number of clusters. The last row shows the results for the number of unknown classes (26) in the task. The algorithm achieved slightly better scores for all evaluation metrics as compared to the number of known classes. The table shows that the proposed clustering approach achieves high accuracy, indicating that it is effective in identifying clusters of traffic sign objects.

Figure 2 shows the clusters predicted from the KMeans algorithm for clustering visualization. The plot shows a 2D representation of the clusters, where each dot represents a traffic sign and is color-coded according to the cluster it belongs to. The plot allows us to visualize the grouping of traffic signs into clusters, which can help identify patterns and similarities among them.

Further Figure 2 shows the graphs generated using the Elbow method to identify Within Cluster Sum of Squares (WCSS) over the number of clusters for two different numbers of clusters, 19 and 7. The WCSS is a measure of the variation within each cluster, and the Elbow method is a technique used to determine the optimal number of clusters for a given dataset. The graphs show the WCSS on the y-axis and the number of clusters on the x-axis. The point on the graph where the curve starts to flatten out is known as the elbow and represents the optimal number of clusters. In Figure 2, for the number of clusters equal to 19, the elbow occurs at around 4 clusters, while for the number of clusters equal to 7, the elbow occurs at around 3 clusters. These values can be used to determine the appropriate number of clusters for the KMeans algorithm.

Table III compares the performance of the clustering mechanism in the proposed traffic sign detection system using supervised learning and semi-supervised learning. The metrics used to evaluate the performance include Accuracy, Rand score, Adjusted rand score, and Normalized mutual info score

These metrics are used to evaluate the quality of the clustering results. Accuracy measures the percentage of instances that are correctly assigned to their respective clusters. Rand score is a measure of similarity between two sets of cluster assignments, taking into account both false positives and false negatives. Adjusted rand score is a variation of the rand score that corrects for chance agreement between the two sets of cluster assignments. The normalized mutual information score measures the mutual information between the true and predicted cluster assignments, normalized by the entropy of the two assignments.

In general, higher values for these metrics indicate better

| No. of Initial Clusters | Accuracy | Rand Score | Adjusted Rand Score | Normalized Mutual Info Score |
|---|---|---|---|---|
| 45 | 0.3939 | 0.5836 | 0.1140 | 0.2009 |
| 7 | 0.3158 | 0.4687 | 0.0414 | 0.0897 |
| 19 | 0.3564 | 0.5304 | 0.0740 | 0.1355 |
| 26 | 0.3578 | 0.5114 | 0.0853 | 0.1705 |

TABLE II
EVALUATION METRICS RESULTS FOR UNSUPERVISED APPROACH

| Clustering Method | Accuracy | Rand Score | Adjusted Rand Score | Normalized Mutual Info Score |
|---|---|---|---|---|
| Unsupervised | 0.050021 | 0.166492 | 0.050021 | 0.166492 |
| Semi-Supervised | 0.42406 | 0.67159 | 0.15610 | 0.245234 |

TABLE III
COMPARISON - UNSUPERVISED CLUSTERING VS SEMI-SUPERVISED CLUSTERING (CLUSTERS = 45)

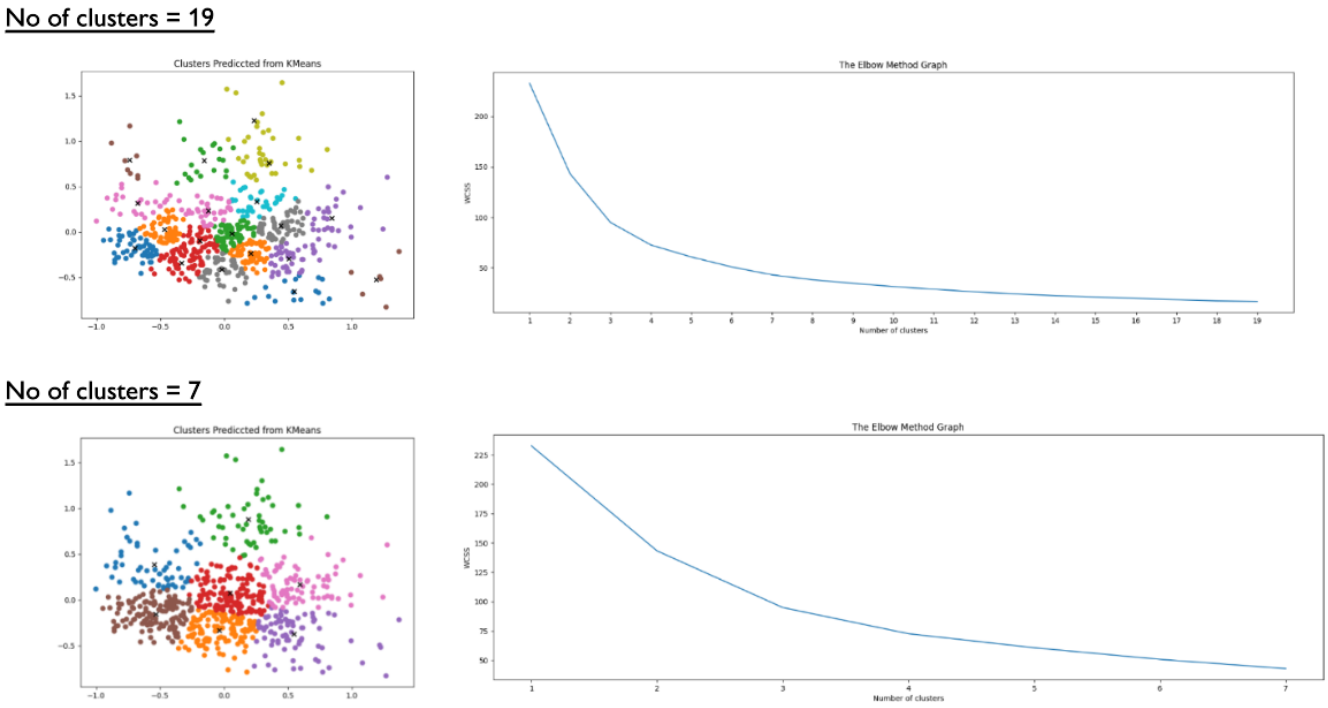


Fig. 2. Cluster Visualization for number of clusters 19 and 07

clustering performance. Therefore, This suggests that the semi-supervised approach is more efficient in utilizing the available labeled and unlabeled data and can improve the performance of the traffic sign detection system.

## IV. CONCLUSION

In this research work, we have presented a novel deep learning-based approach for traffic sign detection that aims to address the generalization ability of the system across different environments. To achieve this goal, we trained their model on the German Traffic Sign Detection Benchmark (GTSDB) dataset, which contains over 50,000 images of German traffic signs, using the Faster R-CNN model with Feature Pyramid Networks (FPN).

To identify clusters of traffic sign objects, we used unsupervised clustering with Principal Component Analysis (PCA). We then demonstrated the generalization ability of their model by testing it on the CeyRo Traffic Sign and Traffic Light Dataset, which consists of images taken from Sri Lankan road scenes. The results showed that their proposed model achieved high accuracy and could be used in different geographical locations, weather, and lighting conditions.

The proposed model can be helpful for autonomous driving systems, as it can provide essential information for the vehicle to perceive the nature of the road and make decisions based on it. The study contributes to developing more efficient and adaptable traffic sign detection systems, which are crucial for the safe operation of autonomous vehicles. We have demonstrated that their approach is effective in detecting traffic signs across different environments, which is a significant step toward making autonomous driving a reality.

## REFERENCES

[1] D. Barnes, W. Maddern, and I. Posner, "Exploiting 3d semantic scene priors for online traffic light interpretation," in *2015 IEEE intelligent vehicles symposium (IV)*. IEEE, 2015, pp. 573–578.

[2] M. B. Jensen, M. P. Philipsen, C. Bahnsen, A. Møgelmose, T. B. Moeslund, and M. M. Trivedi, "Traffic light detection at night: Comparison of a learning-based detector and three model-based detectors," in *Advances in Visual Computing: 11th International Symposium, ISVC 2015, Las Vegas, NV, USA, December 14-16, 2015, Proceedings, Part I 11*. Springer, 2015, pp. 774–783.

[3] G. Overett and L. Petersson, "Large scale sign detection using hog feature variants," in *2011 IEEE intelligent vehicles symposium (IV)*. IEEE, 2011, pp. 326–331.

[4] K. Behrendt, L. Novak, and R. Botros, "A deep learning approach to traffic lights: Detection, tracking, and classification," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 1370–1377.

[5] J. Li, X. Liang, Y. Wei, T. Xu, J. Feng, and S. Yan, "Perceptual generative adversarial networks for small object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1222–1230.

[6] Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li, and S. Hu, "Traffic-sign detection and classification in the wild," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2110–2118.

[7] S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing, and C. Igel, "Detection of traffic signs in real-world images: The german traffic sign detection benchmark," in *The 2013 international joint conference on neural networks (IJCNN)*. Ieee, 2013, pp. 1–8.

[8] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[9] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125.

[10] O. Jayasinghe, S. Hemachandra, D. Anhettigama, S. Kariyawasam, T. Wickremasinghe, C. Ekanayake, R. Rodrigo, and P. Jayasekara, "Towards real-time traffic sign and traffic light detection on embedded systems," *arXiv preprint arXiv:2205.02421*, 2022.

[11] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[12] J. MacQueen, "Classification and analysis of multivariate observations," in *5th Berkeley Symp. Math. Statist. Probability*. University of California Los Angeles LA USA, 1967, pp. 281–297.

[13] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley interdisciplinary reviews: computational statistics*, vol. 2, no. 4, pp. 433–459, 2010.