



# Automatic Detection of Sewage Defects, Traffic Lights Malfunctioning, and Deformed Traffic Signs Using Deep Learning

Khalid M.O. Nahar<sup>1,2</sup> and Firas Ibrahim Alzobi<sup>3</sup>

<sup>1</sup>Department of Computer Sciences, Faculty of Computer Sciences and Information Technology, Yarmouk University, Irbid, 21163, Jordan

<sup>2</sup>Faculty of Computer Studies, Arab Open University, P.O. Box 84901, Riyadh 11681, Saudi Arabia

<sup>3</sup>Information System and Networks Department, The World Islamic Sciences and Education University, Amman 11947, Jordan

Received 20 May 2024, Revised 5 October 2024, Accepted 25 October 2024

**Abstract:** Effective urban planning calls for quick diagnosis and correction of infrastructure problems, including malfunctioning sewage systems, broken traffic lights, and misaligned traffic signs. Often labor-intensive, prone to mistakes, and useless are traditional inspection methods. This paper provides a multi-task convolutional neural network (CNN) framework derived on YOLOv5 for autonomous identification of urban infrastructure problems. Using street-level imagery, our system concurrently detects sewage issues, malfunctioning traffic signals, and misaligned traffic signs. Nine-fold (X9) and three-fold (X3) augmentations, respectively, follow from the model's training on a sizable collection of 2,438 annotated photos of metropolitan settings acquired under various lighting conditions. Notwithstanding fluctuations in F1-scores for particular categories resulting from sample distribution differences, the YOLOv5-based system shows better performance than leading algorithms in extensive evaluations on real-world datasets, attaining an overall accuracy of 86% measured by the F1-score across all classes. Our studies show that the system is quite accurate and robust, which qualifies for scalable real-time deployment in metropolitan surroundings. Several advantages from this study could be higher commuter safety, financial savings for local governments, and better infrastructure maintenance operations helping to build smarter, more sustainable cities.

**Keywords:** Convolutional Neural Network (CNN), Deep Learning, Sewage Defects, Traffic Lights Malfunctioning, Manhole Damage, YOLOv5.

## 1. INTRODUCTION

Urban infrastructure maintenance is crucial for ensuring the functionality and safety of cities. However, detecting and addressing issues such as sewage defects, malfunctioning traffic lights, and deformed traffic signs in a timely manner is challenging due to the sheer scale and complexity of urban environments.

This research aims to create an automated, multi-task Convolutional Neural Network (CNN) system utilizing the YOLOv5 framework for the real-time detection of traffic signal malfunctions, sewage issues, and distorted traffic signs, employing street level imagery. The proposed system seeks to substitute labor-intensive human inspections with automated instruments, therefore enhancing the precision and efficiency of municipal infrastructure monitoring. This work enhances the field by introducing a flexible and scalable technique applicable to smart cities, yielding substantial accuracy improvements over existing models.

Traditional manual inspection methods are often inefficient and labor-intensive, necessitating the exploration of automated solutions to streamline these processes. Traditional manual inspection techniques are frequently labor-intensive and inefficient; therefore, it was necessary to look at automated ways to improve these procedures. Deep learning techniques have become increasingly potent tools in recent years for image recognition tasks, with the potential to automate the identification of anomalies in urban infrastructure. You Only Look Once (YOLO), a well-known object identification algorithm renowned for its quickness and accuracy in real-time applications, is one of these methods. Neural networks are a special type of computer algorithm that power YOLO. Their name stems from the fact that they are pattern recognition machines, just like human brains. YOLO is declared to be a type of CNN which is particularly good at seeing patterns in images and hence, in objects and similar items.

By using YOLO, the target is to create a recognition system that can automatically recognize from street-level imagery flaws in the sewage system, broken traffic lights, and distorted traffic signs. For YOLO to work well, YOLO algorithm divides an image into a grid and concurrently predicts bounding boxes and class probabilities for every grid cell. This method is ideal for real-time applications since it allows for effective object detection with just one neural network run. YOLO can be trained to reliably identify and locate these things by using an annotated dataset of a collection of photos for anomalies in urban infrastructure. In this study, we report our implementation and evaluation of YOLOv5, which is a YOLO algorithm variation, for autonomous identification of deformed traffic signs, malfunctioning traffic lights, and sewage issues. In this research, YOLOv5's architecture was discussed and how well it will be suited to targeted detection tasks. In addition, emphasizing its benefits in terms of deployment ease, speed, and accuracy. We tested the proposed approach on a variety of datasets that included annotated street-level photos taken in a range of environmental settings in order to verify functionality. Using these datasets, YOLOv5 was optimized and refined with extensive testing to assess its detection performance. Findings show how well YOLOv5 performs in real-world circumstances when it comes to precisely identifying and localizing anomalies in urban infrastructure. Moreover, possible uses and consequences of findings on public safety and urban management were discussed also. By the proposed methodology, the reaction time can be decreased, efficiency and safety of urban environments can be improved, and preventative maintenance plans can be enabled by automating the identification of sewage flaws, faulty traffic lights, and deformed traffic signs. In the subsequent sections of this paper, detailed insights into the methodology, experimental setup, results, and discussions of YOLOv5 results and effectiveness are provided. Traffic lights are light machines located at the cross of a road traffic flow and guarantee accident free on the road cross. Traffic lights are one of the most popular tools that maintain the safety of vehicles flowing on highways or internal roads. Traffic lights gained popularity all over the world due to the ease of use as they have three well-known colors (Green, Yellow, and Red) which safely tell the driver to either pass, standby, or stop respectively. Moreover, traffic lights play a significant role in protecting pedestrians on the road. The traffic light changes the lights every few minutes, as it at a certain time allows a certain direction to pass through the intersection, and it is mostly located at the main intersections that are crowded with cars, and it is linked to a computerized system which works when receiving electronic indicators from the sensors that are placed in the street which is also known as the smart traffic light [1][2]. Traffic lights have a wide range of styles, and there is a big difference in how they are mounted and placed. The majority of current systems only prioritize "circle" type or "arrow" type detection [3]. For safe driving, it's crucial to understand traffic signals. A motorist will have important information to comprehend the

road environment if it is feasible to identify and recognize a traffic signal [4]. The number of lives saved by light signals and prevented from accidents and injuries is estimated to be equivalent to 11,000 per year, which shows that their presence and continuous operation is the invention that preserves human safety the most and that they are very faithful to their intended purpose [5]. Sewage manholes: A separate underground conveyance system intended for transporting wastewater from homes and commercial buildings for treatment or disposal. Other sewers serving industrial areas also carry industrial wastewater. The sewage system is called the sewer network system, and the sewage is part of the water distribution network. This network concerns with the discharge of liquid waste from buildings and factories to the treatment plant or disposal sites [6].

### A. Historical Overview of the YOLO

The YOLO algorithm is applied to the input data. The output of the algorithm is a class and a bounding box around the object. YOLOv1 is identified by the speed of recognizing objects, but with less accuracy than Fast\_RCNN. As a result, scientists have developed YOLOv2. Although YOLOv3 has made significant strides in both detection accuracy and speed, its ability in detecting tiny objects is still far from ideal. In April 2020, scientists developed a new version of YOLO, it is YOLOv4, a big difference between YOLOv3 and YOLOv4 was discovered when they were applied on the same dataset [2]. Five alternative model versions are included in the YOLOv5 release: YOLOv5s (the smallest), YOLOv5m, YOLOv5l, and YOLOv5x (the largest). Figure 1 shows the development of YOLOv5 models where the GPU latency in (ms) is shown in (x-axis) and COCO AP val is shown in (y-axis) and then they were compared.

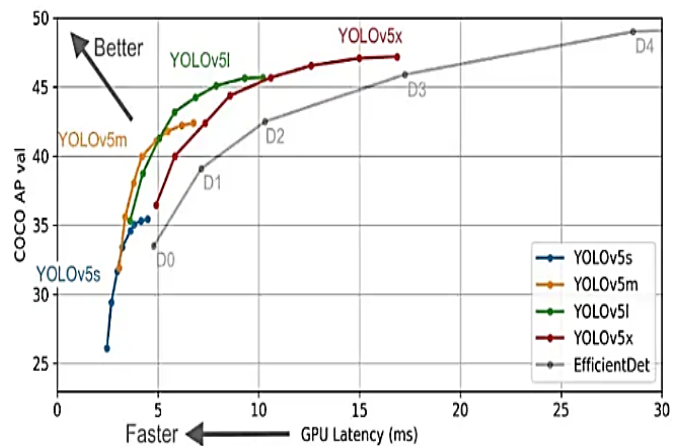


Figure 1. The evaluation of YOLOv5 model [7]

### B. Comparison Between YOLO Versions

Authors in [8] wrote about helping Anki Vector detect another Vector in its camera feed using multi-version of YOLO (YOLOv1 to YOLOv8) and presents an evaluation

between them, the performance depends on the use case and the KPIs targeted. The authors used a dataset comprised of 590 images for training, 61 images for validation, and 30 images for testing. The authors trained each model for 50 iterations and evaluate the accuracy based on used metrics. The Table 1 compares the YOLO versions based on: Accuracy, Speed (FPS), Best Usage [8].

TABLE I. Comparison of YOLO Versions

| Versions | Accuracy (mAP-Mean Average Precision) | Speed (FPS) | Best Usage Scenarios   |
|----------|---------------------------------------|-------------|--|
| v1       | 63.4% mAP on PASCAL VOC               | 45 FPS      | Basic object detection tasks with real-time needs                                    |
| v2       | 76.8% mAP on PASCAL VOC               | 40-67 FPS   | Real-time object detection with medium accuracy                                      |
| v3       | 57.9% mAP@0.5 on COCO (mean)          | 20-30 FPS   | High-speed applications requiring decent accuracy                                    |
| v4       | 65.7% mAP@0.5 on COCO                 | 62 FPS      | Faster real-time detection with enhanced accuracy (small)                            |
| v5       | 66% mAP on COCO                       | 140 FPS     | Versatile for different edge and cloud Easy to use, highly customizable, ultralytics |
| v6       | 52.3% mAP@0.5 on COCO                 | 123 FPS     | (tiny) Fast inference for low-computation devices                                    |
| v7       | 56.8% mAP@0.5 on COCO                 | 161 FPS     | (tiny) Best for real-time video analysis and resource-constrained environments       |
| v8       | 54.8% mAP@0.5 on COCO                 | 140 FPS     | Advanced computer vision tasks with high efficiency                                  |

In Table I, the Accuracy (mAP): Represents the model's mean average precision, a measure of how well the model detects and classifies objects. Speed (FPS) is the Frames Per Second, it shows how fast the model processes images, useful for real-time applications. The Best Usage Scenarios are the types of tasks and environments where each YOLO version performs best. Finally, the Key Features, Highlights major improvements or differentiating features of each version. From the Table I we conclude that there is no simple approach to choose the optimal machine learning model, it depends on the problem under study. YOLOv5 is to be used since it has the least training time compared to other versions, and it has the same inference time as the YOLOv7. This research contributes the combination of YOLOv5 in Multi-Classification CNN for autonomous identification of urban infrastructure problems like sewage system abnormalities, misaligned traffic signs, and faulty traffic lights. Unlike other studies using slower object recognition techniques or focused on identifying certain anomalies, this work uses a singular, extremely efficient real-time model to concurrently detect many infrastructure issues. Leveraging the improved speed and accuracy of YOLOv5, the system is built to interpret street-level imagery across various lighting conditions. This assures consistent performance in many real-world environments. Furthermore, the application of cutting-edge data augmentation techniques using Roboflow and Keras improves the generalization capacity of the model, therefore addressing a major obstacle in urban detection systems confronted with item variability resulting from lighting or angle differences. This integrated architecture offers a scalable, real-time means of improving municipal decision-making, lowering expenses, and increasing public safety. It supports smart city initiatives and greatly improves the front line of urban infrastructure monitoring. In the upcoming parts, the paper presents the related work, the proposed model, the model evaluation metrics used, the experimental results, and lastly, the conclusion and future work section.

## 2. RELATED WORK

This is an overview of previous work for the problem under study. In this review we will look at others contributions and research gaps in their work. The upcoming sections.

### A. Object recognition using YOLO

Deep learning methods have helped real-time object detection models for monitoring urban infrastructure become rather popular recently. One of the most often utilized algorithms in this industry, the YOLO (You Only Look Once) framework has witnessed significant development since its introduction. Because YOLOv5 strikes a mix between speed, accuracy, and adaptability, it has become a common choice for object recognition tasks. Wan et al. suggested a lightweight version of YOLOv5 to detect road deterioration, which produced faster and more accurate results due to an efficient design fit for infrastructure uses. Based on the RDDC dataset, the study revealed road defects

including potholes and fractures, therefore indicating the potential of YOLOv5 in urban monitoring systems [9]. In Analogously, Kristo et al. created a thermal object detection system using YOLO to identify anomalies in unfavorable conditions. This work shows the resilience of YOLO in real-world situations and offers information for its application in urban infrastructure monitoring [10].

The developments to YOLOv5 and later generations, notably YOLOv7 and YOLOv8, have kept stretching the limits of real-time object detection. Particularly in settings with limited resources, Olorunshola et al. found performance gains in terms of speed and accuracy in a research contrasting YOLOv5 with YOLOv7.

The study highlights YOLOv5's relevance to real-time applications, so it is a good choice for multitask detection of signs, traffic signals, and sewage problems in urban surroundings [8]. By means of its efficient design and higher inference speed, Dwivedi [7] compared YOLOv5 to higher R-CNN found that YOLOv5 performs better in real-time applications. This justifies its application in 4 infrastructure anomaly detection.

With artificial intelligence-based monitoring systems used more and more, urban infrastructure maintenance is a growingly crucial part of smart city projects. Emphasizing anomaly detection and real-time urban infrastructure, Zhang et al. created a multi-task learning framework for smart city applications. Their work, which emphasizes the benefits of multi-task models in increasing the efficacy of urban maintenance and safety [11], is very in line with the present research. Gallo et al. [12] investigated item identification in agricultural environments using YOLOv7, therefore offering insightful analysis of the adaptation of YOLO models for many sectors. Traffic signal and sign identification have advanced dramatically, thanks in great part to deep learning algorithms. Yang and Zhang proved YOLO's usefulness in dynamic environments by using it to instantly identify traffic signals [2]. Their study is quite pertinent to our investigation since one of the main purposes of the suggested model is to recognize traffic signs. Combining CNNs with YOLOv3, according to Novak et al., raised autonomous driving system classification accuracy [1]. At last, an interesting subject is AI-based sewage system and urban infrastructure monitoring of anomalies. The ensemble learning solution of Xu et al. for real-time environmental hazard monitoring has methodological similarities with sewage defect identification for anomaly detection in real-time [13].

These latest papers underline the relevance and importance of the research in the field since they reveal developments in object detection and deep learning for monitoring urban infrastructure. In [14], the author shows a vision system to determine what stationary items could obstruct a moving robot's route by Microsoft Kinect sensor using YOLO, the accuracy ratio was 96.36

### B. Traffic Sign Recognition

authors in [15], used CNN, a deep learning algorithm, to identify traffic signs and extract the main features of the images, and connect the output of all convolutional layers to

the Multilayer Perceptron (MLP), the author used the GTSB dataset, which has 43 classes; the data is divided into 34799 images for training, 4410 images for validation, and 12630 images for testing, and the author got the accuracy ratio of 97.1%.

in [16], authors used YOLO based on an active learning approach and real-time object detection to identify traffic signs, YOLOv2 gives the highest accuracy and reduces the size of the labeling dataset with an accuracy ratio of 97% in real-time object detection. In [1], the authors detect traffic signs using YOLOv3 as well as CNN, autonomous driving system required real-time detection, and the authors use pre-trained YOLO for the detection with 5 objects for classification (cars, trucks, pedestrians, traffic signs, and traffic lights), they used CNN for classifying traffic sign into 75 classes. The authors got accuracy ratio of 99.2% and used German traffic sign recognition benchmark dataset which consists of 120000 images and 75 categories.

in [17] the authors suggest a method to detect and classify traffic signs by image analysis, the color and shapes were chosen as features, in classification, the authors used the input pattern for a neural network.

In [18] the author creates a new benchmark for detecting and classifying traffic signs, he collected 10,000 images under different conditions, with more resolution than other datasets. Moreover, he trained two networks: one for detecting all traffic signs as one category, and the other detecting and classifying traffic signs as multiple categories.

In [19], the author suggests two algorithms for classification and detection, in detection, the author extract traffic sign proposal by using the color probability model and MSER region detector. The author used SVM and novel color HOG feature to filter out the false positives and classify the remaining proposal. The author detects and classifies the traffic sign at 6 fps on 1360\*800 and suggests using GPU to accelerate detection and classification.

In [20], the author used YOLOv2 and end-to-end learning to achieve fast detection of Chinese traffic signs in real time. authors have solved the small size of the traffic sign by a fine grid to divide the images, authors used the CTSD dataset which contains 1100 images with sizes 1024\*768 and 1280\*720. The architecture of the algorithm is visualized by figure 2.

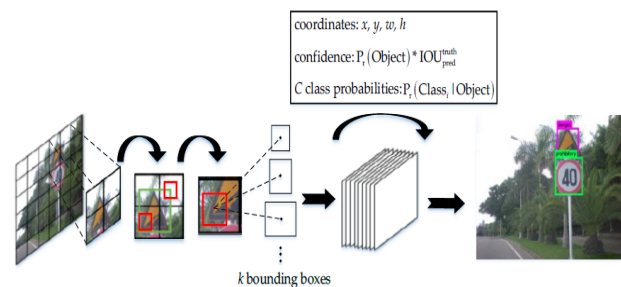


Figure 2. The Chinese Traffic Sign Detection Algorithm[20]

TABLE II. Summary of Related Work

| Ref  | Objective   | Dataset  | Approach                                     | Accuracy        |
|------|---|--|--|-----------------|
| [1]  | detect traffic signs using yolov3 extended CNN Identification and | (GTSRB)  | YOLOV3 and CNN for classification            | 99.2%           |
| [9]  | recognition of damaged roads                                      | RDDC   | YOLO-LRDD                                    | Precision 58.9% |
| [10] | Detect illegal immigrants people by a thermal camera              | The six databases mentioned above                              | Faster R-CNN, SSD, Cascade R-CNN, and YOLOv3 | 97.93% AP score |
| [14] | Object recognition Items can obstruct the path of a moving robot  | A special data set was collected of where the robot is walking | YOLOV2                                       | 96.36%          |
| [15] | identify traffic signs(CNN)                                       | (GTSRB)  | LeNet-5 network                              | 97.1%.          |
| [16] | real-time object detection traffic sign                           | (YOLOV2)   | (GTSRB)                                      | 97%             |
| [18] | detecting and classifying traffic signs                           | collected 10000 images (CTSD)                                  | FAST CNN                                     | 84% accuracy    |
| [19] | classification and detection of traffic sign                      | and GTSDDB)  | CNN  | 98.4% recall    |
| [20] | detection of Chinese traffic signs in real-time                   | CTSD   | YOLOV2                                       | 91.58% Recall   |

From Table II and after reading the previous research, some researchers used YOLO (v2, v3, v5), CNN, YOLO\_LRDD [9][16][18] and the maximum accuracy was 97.1% using LetNet-5 but targeting only traffic signs. Meanwhile, this paper target traffic sign, traffic light and manholes damage using YOLOv5 for detection and recognition which means multiclass recognition. The YOLOv5 model is seen especially suitable since it achieves a balance between speed, precision, and versatility which is crucial for real-time applications in urban environments. Primarily, YOLOv5 exhibits accelerated processing speeds reaching up to 140 frames per second rendering it optimal for real-time identification in dynamic contexts such as infras-

structure inspection and traffic surveillance, where systems must rapidly evaluate frames from live video feeds. In contrast to its predecessors, YOLOv5's architecture has been optimized to enhance accuracy while preserving speed; it attained a mean average precision (mAP) of almost 66% on the COCO dataset. This degree of accuracy is essential for recognizing minor objects such as traffic signals and subtle flaws in sewage systems to avoid false positives or overlooked detections. Moreover, because of YOLOv5's extensive modularity and customizability, researchers can implement transfer learning and model optimization to refine the model for particular urban contexts. This is an optimal selection for on-site monitoring applications necessitating low latency and efficient hardware use, such as intelligent traffic cameras or drone inspections, as it accommodates smaller, resource-limited edge devices. Moreover, the model exhibits enhanced performance in recognizing objects of diverse sizes owing to its multi-scale predictions and superior management of anchor boxes. This is especially crucial for differentiating between larger, distorted traffic signs and smaller traffic signals. YOLOv5 is regarded as a state-of-the-art deep learning defect detection system for smart cities, owing to its accessible implementation through the Ultralytics framework and its features, especially when prompt and precise responses are essential for public safety and efficient infrastructure management.

In summary, the previous studies have made good use of deep learning models including YOLO to monitor urban infrastructure. Particularly in traffic sign recognition and road damage assessment, these research exposed improved accuracy and remarkable real-time detection capacity in specific fields. While Kristo et al. showed the efficiency of YOLO in demanding environmental conditions. Wan et al. and Yang et al. successfully identified road and traffic anomalies. But most of these models focus on single-object tasks or specific use cases, therefore restricting their application for multi-task detection in urban environments. Furthermore many older models lack the scalability required for smart city projects requiring real-time deployment. Our approach offers a more flexible multi-task convolutional neural network developed from YOLOv5, competent in spotting several irregularities in urban infrastructure, including traffic lights, traffic signs, and sewage problems, across different lighting conditions. While improving scalability and real-time capabilities for smart city uses, this combined approach guarantees high accuracy and durability [10] [9] [2].

### 3. RESEARCH METHODOLOGY

In this paper, we tackled the problem of identifying, Traffic lights, traffic signs, and sewage manholes using YOLOv5, which will be identified in real-time object detection from video streaming. More than 2,438 images depicting urban settings, such as sewage problems, broken traffic signals, and distorted signage, make up the dataset used in this investigation. The photos showed a variety of weather and lighting conditions and were originally from various online archives. There were a total of 3,118 pho-

tographs in the dataset; 2,118 for training, 117 for testing, and 203 for validation. To enhance model performance and reduce data imbalance, we utilized Jupyter Notebook for a nine-fold increase and the Roboflow tool for a three-fold augmentation, incorporating flipping, cropping, rotating, and zooming. The multi-task deep learning model was built and trained using the YOLOv5 framework. Roboflow's is an AI-assisted labeling tools were used to accomplish image annotations. To evaluate the model, we utilized GPU-accelerated hardware to shorten training and inference times, which allowed us to employ metrics like accuracy, precision, recall, and the F1-score.

The proposed methodology consists of the following steps:

1. Data Collection.
2. Data Augmentation.
3. Data Annotation.
4. Training, Validation, and Testing.
5. Results and Evaluation.

Figure 3 is the pictorial view of the proposed methodology.

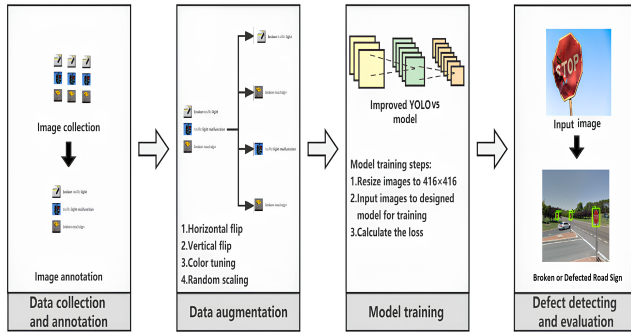


Figure 3. Methodology

#### A. Data collection

The dataset used in this paper was collected from random website, the images were all collected in a variety of circumstances, such as changes in lighting, color, and shape. A general, 2438 image has been captured and divided right into a training set, test set and validation set. Figure 4, shows

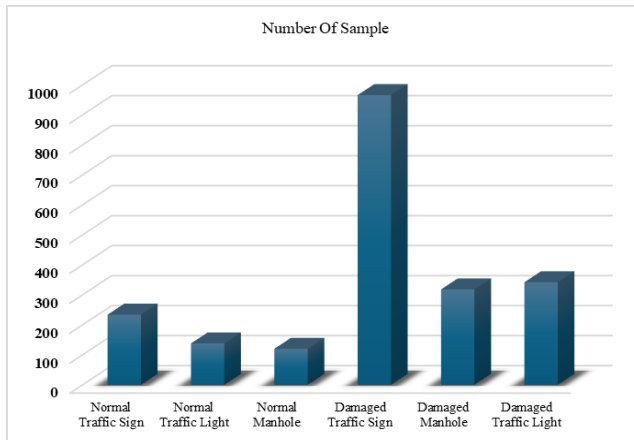

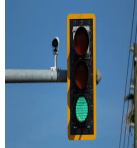






Figure 4. Data Distribution

the dataset distribution; it is clear that data is unbalanced (biased towards the damaged traffic sign), this is due to the characteristics of the damaged light signal and the normal light signal are similar in terms of shape and color, so the number of damaged light signal data was increased to increase the distinction between them. The training set consists of 2118 image, the test set consists of 117 image and The Val set consists of image 203. Table III shows some samples from the dataset. The dataset consists of 6 classes 1. Traffic light normal 2. Traffic light damage 3. Manhole normal 4. Manhole damage 5. Traffic sign normal 6. Traffic sign damage.

TABLE III. Samples From the Dataset

| Class                 | Sample  | Number of Samples |
|-----------------------|---|-------------------|
| Normal Traffic Sign   |    | 234               |
| Normal Traffic Light  |   | 138               |
| Normal Manhole        |  | 120               |
| Damaged Traffic Sign  |  | 966               |
| Damaged Manhole       |  | 318               |
| Damaged Traffic Light |  | 342               |

#### B. Data Augmentation

To achieve high accuracy, data augmentation was implemented by "image generation" using keras, and four geometric transformations were used, they are:

- i. Shear:  $\pm 15^\circ$  Horizontal,  $\pm 17^\circ$  Vertical.
- ii. Saturation: Between -8% and +8%.

- iii. Bounding Box Flip: Horizontal, Vertical.
  - iv. Bounding Box Crop: 0% Minimum Zoom, 20% Maximum Zoom.
  - v. Bounding Box Rotation: Between  $-15^\circ$  and  $+15^\circ$ .
- Images were augmented twice, firstly, by increasing number of images by  $\times(9)$  for all categories, this was done by Jupyter Notebook. Secondly, Only the training data was augmented using the Roboflow tool, in which the data was incremented by  $\times(3)$ .

Figure 5 shows a sample of and augmented image based on the described augmentation process, the image is scaled in and out, rotated, and flipped.

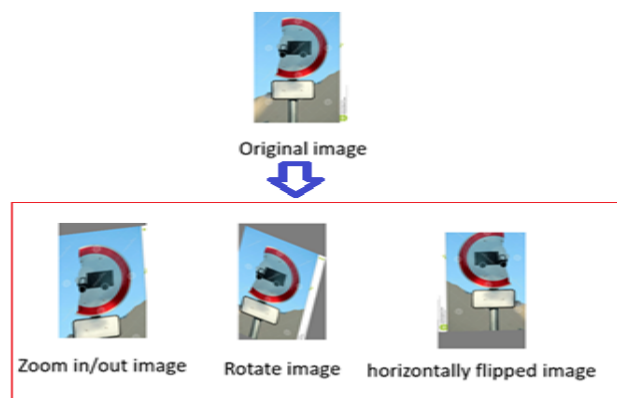


Figure 5. Sample of Data Augmentation

### C. Data Annotation

the annotation process was done by resizing the images to be  $640 \times 640$  pixel, and then locate the traffic lights, traffic signs, and sewage manholes by placing a rectangular box (Bounding boxes) that starts at the top left corner and ends at the bottom right corner and assigning each image a class name. Figure 6 show Data annotation process.

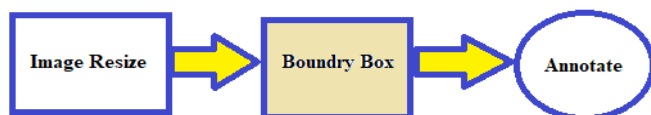


Figure 6. Data Annotation Process

Figure 7 shows an example of the process of image labeling using Roboflow tool. Roboflow, is an annotation tool used to build a computer vision applications. The bounding box and the name of the class are AI-sisted[12]. Roboflow is characterized by ease of use and flexibility because it uses a user friendly GUI. In the right part of the screen, we see the toolbar as shown in Figure 7. The toolbar consists of 9 small icons, the Drag tool that pans the image or select annotation, Bounding box icon that draw annotation around box, Polygon icon that freeform draw annotation for more precise shapes. Smart polygons use

an intelligent assistant to draw your polygon, label assist icon used the prediction from a Roboflow train model as a starting point, it will repeat the process on all images for full annotation up to the last image.

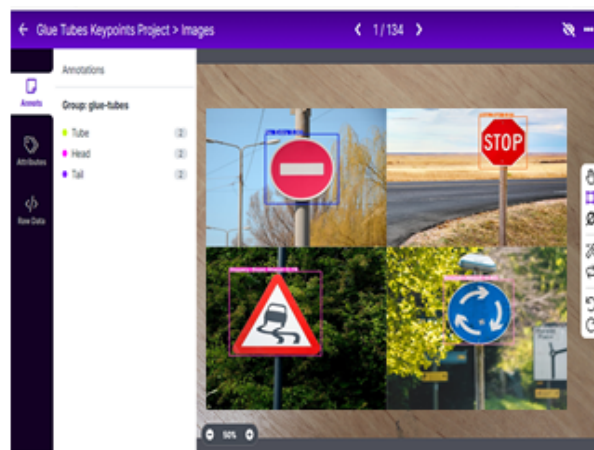


Figure 7. Annotating Images Using Roboflow Tool

This research employed YOLOv5 in conjunction with a methodical and scientific approach to accurately detect anomalies in urban infrastructure. The dataset utilized for training, validation, and testing consisted of 2,438 annotated photos obtained from various metropolitan environments under differing lighting conditions. We addressed the dataset's imbalance by geometric transformations, including rotation, flipping, and zooming, utilizing data augmentation techniques up to  $\times(9)$  augmentation using Jupyter Notebook. Meanwhile up to  $\times(3)$  augmentation was done using Roboflow tool. The photos were scaled to  $640 \times 640$  pixels, and utilizing Roboflow's AI-powered annotation capabilities, bounding boxes were designated for each object, including traffic lights, sewage manholes, and traffic signs. Following 40 epochs of training, we evaluated the model utilizing accuracy, precision, recall, and F1 score. These measures were utilized to evaluate the model's effectiveness across different item categories. To verify the model's calibration for authentic urban circumstances, a confusion matrix was created to visually evaluate classification accuracy and errors.

## 4. YOLO 5 ARCHITECTURE

Three parts made up the YOLOv5's network architecture: YOLO Layer for the head, PANet for the neck, and CSPDarknet for the backbone. The data are initially sent to CSPDarknet for feature extraction, and then they are transmitted to PANet for feature fusion. Subsequently, YOLO Layer outputs the detection results (class, score, position, size) [13].

The features of YOLOv5 that were absent from the previous version include: implementing the CSPNet strategy on the PANet model; replacing the SPP block in the model neck with the SPPF block; and integrating the Focus layer

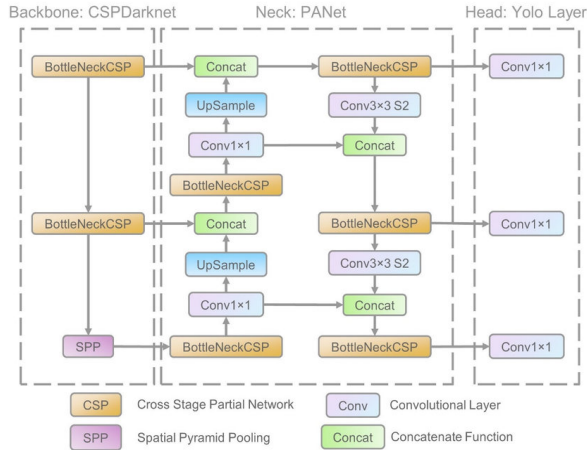


Figure 8. Yolo5 Architecture[21].

to the CSP-Darknet53 backbone. After addressing the issue of grid sensitivity, YOLOv5 and YOLOv4 are now able to recognize bounding boxes with center points in their edges with ease. Lastly, YOLOv5 is faster and lighter than earlier iterations [13]. Shape and color categorization in YOLO is done by dividing the image into an  $n \times n$  grid. After that, several bounding boxes are predicted by class probability map to give the final bounding boxes and objects classes, as shown in Figure 9 which is borrowed from [22].

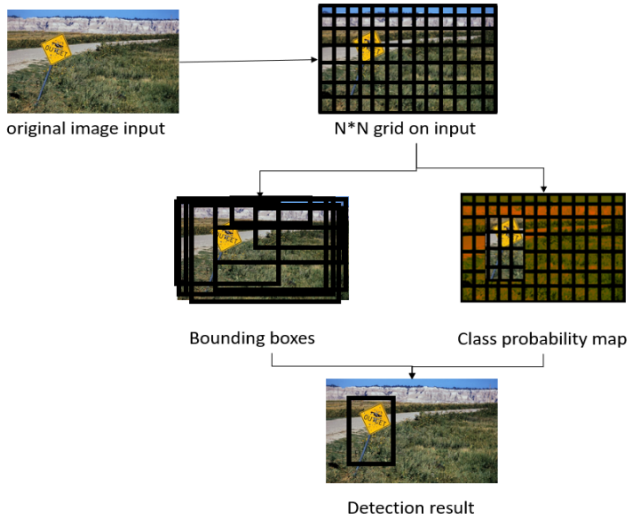


Figure 9. YOLO Detection Process.

## 5. MODEL EVALUATION METRICS

For any learning model, evaluating the model is a crucial step. The accuracy score metric is one of the most popular and useful measures for assessing the model. At times, relying solely on the accuracy score statistic is insufficient. As a result, additional metrics like recall, precision, and F1 score were employed. These measures are explained in the following sections[23].

### A. Accuracy Metric

The most widely used performance metric is accuracy. It is referred to as the ratio of all observations to accurately predicted observations. Only when the dataset is balanced will accuracy provide us with a strong indicator and assessment. We used the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) counts to gauge the accuracy because our data is not balanced [23]. Equation 1 represents the accuracy based on prior counts when the dataset is unbalanced, while Equation 1 represents the definition of accuracy.

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

### B. Precision Metric

Precision was used to test the classifier's capacity to return only relevant examples. Equation 2 represents this measure. To put it simply, it's the ratio of the number of positive outcomes produced by the algorithm to the number of accurately anticipated positive outcomes.

$$PRECISION = \frac{TP}{TP+FP} \quad (2)$$

### C. Recall Metric

Recall, sometimes referred to as sensitivity, is used to determine how well the classifier can recognize every pertinent event. Equation 3 is the formula that was utilized to compute it. It is calculated by dividing the total number of relevant samples by the number of correct positive outcomes.

$$RECALL = \frac{TP}{TP+FN} \quad (3)$$

### D. F1-Measure

The F1 score is a crucial statistic in the field of traffic sign recognition and identification since it offers a consolidated value that precisely represents the trade-off between recall and precision. The harmonic mean of two measurements is used to calculate the F1 score, which provides a comprehensive evaluation of algorithm performance. It considers the accuracy of positive forecasts as well as the system's capacity to precisely identify all pertinent circumstances. When the datasets are unbalanced or the class distributions are different, this statistic is quite helpful. The F1 score provides scholars and industry experts with a thorough assessment of the model's performance, enabling the creation of resilient and flexible traffic sign recognition systems that optimize precision and recall [23]. The equation that symbolizes F-Measure is given by Equation 4.

$$F1\text{-Measure} = 2 * \frac{PRECISION * RECALL}{PRECISION + RECALL} \quad (4)$$

Another important visual metric is the confusion matrix which summarizes the performance of a classification algorithm in a visualized matrix [24].



## 6. RESULTS AND DISCUSSION

A sample test data was plugged into the model for recognition, Figure 10, shows the recognized tested data surrounded by a boundary box with its class name indicated.



Figure 10. Test Data Recognition Using YOLO 5.

Table 4 below collects the important metrics values for each targeted class, the most precision ratio was in light damage and light normal class, and the less precision ratio was in sign normal class due to the small data size, the most recall ratio was in light normal class, and the less recall ratio was in manholes normal class.

Due to the unbalanced dataset event of augmentation, the focus was on the harmonic mean value between metrics, mainly F1-Measure in this case. Taking the F1-Measure as the representative value of all classes, it is clear that the overall accuracy achieved is 85.3%

TABLE IV. Measurement Values from Testing Experiments.

| Type           | Precision | Recall | F1-Measure |
|----------------|-----------|--------|------------|
| All classes    | 78.40%    | 93.70% | 85.37%     |
| Light damage   | 100.00%   | 98.50% | 99.24%     |
| Light normal   | 100.00%   | 98.90% | 99.45%     |
| Manhole damage | 85.10%    | 94.50% | 89.55%     |
| Manhole normal | 58.30%    | 87.40% | 69.94%     |
| Sign damage    | 95.80%    | 93.20% | 94.48%     |
| Sign normal    | 31.20%    | 89.70% | 46.30%     |

For clear representation of Table IV, a pictorial view for this table is showed in Figure 11.

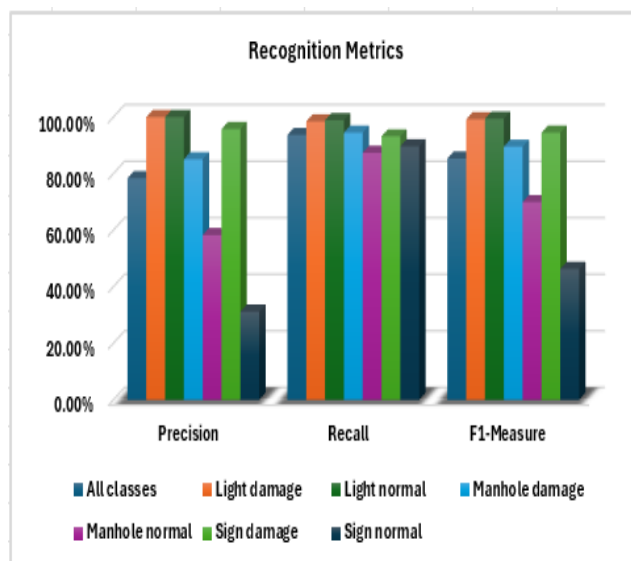


Figure 11. Test Data Recognition Using YOLO 5.

Using graphical charts, one approach to assess and see how well any machine learning model is performing. The accuracy of training and validation for every epoch is displayed in Figure 12. We stopped training the CNN model at epoch number 40 based on the plot and many experiments, as the model begins overfitting the data beyond these values. The training and validation losses for every period

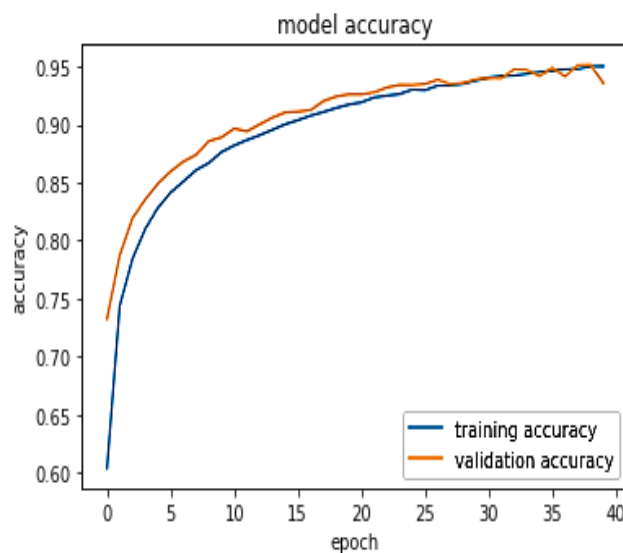


Figure 12. Training and Validation Accuracy

are shown in Figure 13. While the CNN model is operating tell epoch number forty, where the validation and testing loss is getting smaller. Additionally, it is evident that the validation loss begins to rise after 40 epochs, necessitating the termination of CNN training.

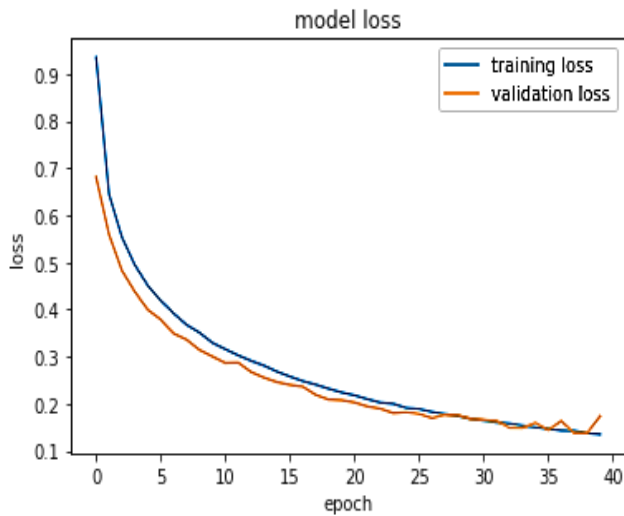


Figure 13. Training and Validation Loss

Confusion matrix, which characterizes a classifier's performance on test instances, often the test portion of a dataset, was a significant result. The majority of the classifications from Figure 14, were centered on the main diagonal, indicating that the classification had a high degree of accuracy.

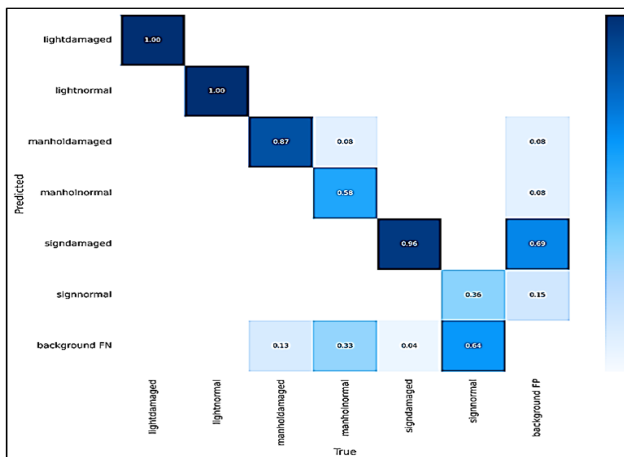


Figure 14. Confusion Matrix

The study's results have considerable implications for applications including smart city surveillance and infrastructure administration. The high accuracy and real-time detection capabilities of our YOLOv5 based system indicate that AI-driven solutions can significantly diminish the necessity for labor-intensive manual inspections, hence enhancing the efficiency and scalability of infrastructure maintenance operations. Automating the identification of faults in sewage systems, malfunctioning traffic signals, and distorted traffic signs enables municipalities to decrease long-term expenses and respond to possible hazards more

swiftly. This also allows municipalities to establish preventative maintenance programs. Moreover, our model's flexibility in response to varying illumination conditions and urban settings illustrates its appropriateness for application in diverse real-world contexts. This study enhances the existing research on urban management by integrating AI, hence promoting the development of smarter, safer, and more sustainable cities. As urban areas globally expand, the capacity for ongoing infrastructure monitoring will become progressively vital for ensuring public safety and resource allocation.

The proposed YOLOv5 based method yields compelling results; nonetheless, many limitations must be acknowledged. The model's application to entirely novel or unforeseen environments may be constrained by the very limited dataset, even after enhancements, when juxtaposed with extensive metropolitan contexts. Severe weather conditions or significant discrepancies in item appearances, such as extensively damaged or obscured infrastructure components, may also affect system operation and were not comprehensively included in the dataset. A further drawback is the computing requirements for real-time detection; although YOLOv5 operates effectively, its deployment in resource-constrained contexts may necessitate further optimization. Future investigations should aim to enhance the dataset by integrating a wider array of diverse and extensive photos, encompassing severe situations and edge cases. Moreover, employing more sophisticated data augmentation techniques, leveraging transfer learning from pre-trained models, and examining other YOLO iterations (such as YOLOv7) could enhance detection speed and precision. Ultimately, incorporating other sensory inputs like as LiDAR or thermal imaging could enhance system robustness and reliability, facilitating the development of completely autonomous urban monitoring systems.

## 7. CONCLUSIONS AND FUTURE WORK

Traffic signals, signs, and sewage manholes are broken. We present an architecture for multi-task convolutional neural networks that detects and classifies anomalies in urban infrastructure on their own. Our method is quite flexible for complex urban settings since it can spot sewage anomalies, broken traffic signals, and misaligned traffic signs from street-level images. Training on a large, annotated collection of urban photos shot under various lighting conditions assured the model's robustness. On real-world datasets, comprehensive testing and evaluation show that our system outperforms current approaches by showing better accuracy and dependability in spotting previously recorded abnormalities. Furthermore, our model guarantees constant performance over several environmental conditions by means of strong generalizing ability. Beyond only technological advancements, this discovery has important consequences for urban government. Preemptive maintenance helps to maximize infrastructure management, thus improving commuter safety, maybe lowering municipal costs, and so enabling more efficient and effective urban maintenance activities.

## ACKNOWLEDGMENT

As the first and corresponding Author of this research, I would like to express my sincere gratitude to Yarmouk University for granting me a sabbatical leave, which provided invaluable support and the necessary time to dedicate to my research. This leave enabled me to advance my studies, pursue innovative ideas, and ultimately achieve success in my published work. I am deeply appreciative of the university's commitment to fostering academic growth and its unwavering support for research endeavors. Thank you for facilitating an environment where knowledge can flourish.

## REFERENCES

- [1] B. Novak, V. Ilić, and B. Pavković, "Yolov3 algorithm with additional convolutional neural network trained for traffic sign recognition," in *2020 Zooming Innovation in Consumer Technologies Conference (ZINC)*. IEEE, 2020, pp. 165–168.
- [2] W. Yang and W. Zhang, "Real-time traffic signs detection based on yolo network model," in *2020 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*. IEEE, 2020, pp. 354–357.
- [3] Z. Shi, Z. Zou, and C. Zhang, "Real-time traffic light detection with adaptive background suppression filter," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 3, pp. 690–700, 2015.
- [4] M. Omachi and S. Omachi, "Fast detection of traffic light with color and edge information," *Journal of the Institute of Image Electronics Engineers of Japan*, vol. 38, no. 5, pp. 673–679, 2009.
- [5] U. S. Kholmatov, S. J. Zingirov *et al.*, "Causing factors of road transport incidents in traffic," *International Journal of Education, Social Science & Humanities*, vol. 12, no. 5, pp. 1524–1534, 2024.
- [6] S. J. Burian, S. J. Nix, R. E. Pitt, and S. R. Durran, "Urban wastewater management in the united states: Past, present, and future," *Journal of Urban Technology*, vol. 7, no. 3, pp. 33–62, 2000.
- [7] P. Dwivedi, "Yolov5 compared to faster rcnn. who wins?" *Towards data science. com*. (accessed: May 7, 2021). <https://towardsdatascience.com/yolov5compared-to-faster-rcnn-who-wins-a771cd6c9fb4>, 2020.
- [8] O. E. Olorunshola, M. E. Irhebhude, and A. E. Ewwiekpaefe, "A comparative study of yolov5 and yolov7 object detection algorithms," *Journal of Computing and Social Informatics*, vol. 2, no. 1, pp. 1–12, 2023.
- [9] F. Wan, C. Sun, H. He, G. Lei, L. Xu, and T. Xiao, "Yolo-lrdd: A lightweight method for road damage detection based on improved yolov5s," *EURASIP Journal on Advances in Signal Processing*, vol. 2022, no. 1, p. 98, 2022.
- [10] M. Krišto, M. Ivacic-Kos, and M. Pobar, "Thermal object detection in difficult weather conditions using yolo," *IEEE access*, vol. 8, pp. 125 459–125 476, 2020.
- [11] Y. Zhang, Y. Yang, W. Zhou, H. Wang, and X. Ouyang, "Multi-city traffic flow forecasting via multi-task learning," *Applied Intelligence*, vol. 51, no. 10, pp. 6895–6913, 2021.
- [12] I. Gallo, A. U. Rehman, R. H. Dehkordi, N. Landro, R. La Grassa, and M. Boschetti, "Deep object detection of crop weeds: Performance of yolov7 on a real case dataset from uav images," *Remote Sensing*, vol. 15, no. 2, p. 539, 2023.
- [13] R. Xu, H. Lin, K. Lu, L. Cao, and Y. Liu, "A forest fire detection system based on ensemble learning," *Forests*, vol. 12, no. 2, p. 217, 2021.
- [14] D. H. Dos Reis, D. Welfer, M. A. De Souza Leite Cuadros, and D. F. T. Gamarra, "Mobile robot navigation using an object recognition software with rgb-d images and the yolo algorithm," *Applied Artificial Intelligence*, vol. 33, no. 14, pp. 1290–1305, 2019.
- [15] D. Yasmina, R. Karima, and A. Ouahiba, "Traffic signs recognition with deep learning," in *2018 International Conference on Applied Smart Systems (ICASS)*. IEEE, 2018, pp. 1–5.
- [16] S. Saleh, S. A. Khwandah, A. Heller, A. Mumtaz, and W. Hardt, "Traffic signs recognition and distance estimation using a monocular camera," in *6th International Conference Actual Problems of System and Software Engineering.[online] Moscow: IEEE*, 2019, pp. 407–418.
- [17] A. De La Escalera, L. E. Moreno, M. A. Salichs, and J. M. Armingol, "Road traffic sign detection and classification," *IEEE transactions on industrial electronics*, vol. 44, no. 6, pp. 848–859, 1997.
- [18] 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.
- [19] Y. Yang, H. Luo, H. Xu, and F. Wu, "Towards real-time traffic sign detection and classification," *IEEE Transactions on Intelligent transportation systems*, vol. 17, no. 7, pp. 2022–2031, 2015.
- [20] J. Zhang, M. Huang, X. Jin, and X. Li, "A real-time chinese traffic sign detection algorithm based on modified yolov2," *Algorithms*, vol. 10, no. 4, p. 127, 2017.
- [21] R. Khanam and M. Hussain, "What is yolov5: A deep look into the internal features of the popular object detector," *arXiv preprint arXiv:2407.20892*, 2024.
- [22] G. Liu, J. C. Nouaze, P. L. Touko Mbouembe, and J. H. Kim, "Yolo-tomato: A robust algorithm for tomato detection based on yolov3," *Sensors*, vol. 20, no. 7, p. 2145, 2020.
- [23] K. M. Nahar, F. Al-Omari, N. Alhindawi, and M. Banikhalaf, "Sounds recognition in the battlefield using convolutional neural network," *International Journal of Computing and Digital Systems*, vol. 11, no. 1, pp. 189–198, 2022.
- [24] J. Görtler, F. Hohman, D. Moritz, K. Wongsuphasawat, D. Ren, R. Nair, M. Kirchner, and K. Patel, "Neo: Generalizing confusion matrix visualization to hierarchical and multi-output labels," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022, pp. 1–13.