A Computer Vision Approach to Ambulance Classification in the Philippines using YOLOv5 Small

A Thesis Proposal

Presented to the

School of Information Technology

Mapúa University

In Partial Fulfillment

of the Requirements for the Degree

Bachelor of Science in Computer Science

Submitted by:

Feria, Steve Vincent C.

Franco, Jose Rafael A.

Laguna, Hans C.

Technical Adviser:

Sir. John Paul Tomas

Submitted to:

William T. Torres, PhD

October 2023

APPROVAL SHEET

This is to certify that I have supervised the preparation of and read the thesis report prepared by **Steve Vincent C. Feria, Jose Rafael A. Franco, and Hans C. Laguna,** entitled **A Computer Vision Approach to Ambulance Classification in the Philippines using YOLOv5 Small** and that the said thesis report has been submitted for final examination by the Oral Examination Committee.

John Paul Q. Tomas
Adviser

As a member of the Oral Examination Committee, we certify that we have examined this thesis report, presented on ____________, and hereby recommend that it be accepted as fulfillment of the thesis report requirement for the degree of Bachelor of Science in Computer Science.

Panel Member

Panel Member

Panel Member

This thesis report is hereby approved and accepted by the School of Information and Technology as a fulfillment of the thesis report requirement for the degree of Bachelor of Science in Computer Science.

Ariel Kelly D. Balan
Panel Member

ABSTRACT

This study outlines the creation of an object detection model utilizing YOLOv5 Small, designed to identify and categorize ambulances on the road, distinguishing them based on various types and characteristics. The research process included the assembly of a specific dataset of ambulance images from the Philippines, which comprised seven classes. The project made use of image augmentation, leading to a dissimilarity score of 0.1574 between the original and augmented images, signifying the successful introduction of variations. The process of hyperparameter tuning was carried out, with a batch size of 8 proving to be the most effective, resulting in a mAP@50 score of 93% for the detection and classification of ambulances and vehicles.

Keywords: Ambulance Detection, Ambulace Classification, Object Detection, YOLOv5 small, Computer Vision, Hyperparameter Tuning, Augmentation Techniques

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

Object detection is commonly used and applied in digital images and videos to detect or identify objects that are present within which can be in the form of animals, human beings, or vehicles. The distinct approaches for object detection are continuously evolving and adapting according to the needs of human studies [32]. YOLO, or You Only Look Once, is a widely recognized algorithm for real-time object detection. It is known for its speed and accuracy in identifying various objects in images or videos with just one forward pass through a neural network. YOLOv5 is one of the most popular versions, known for being fast, accurate, and reliable. It has unique architecture and training methods that make it suitable for a wide range of applications. It has three weights, small, medium, and large; they can be used for different use cases. The algorithm uses a unique loss function that improves the accuracy of object detection by applying different penalties for false positives and false negatives. This combination of speed and precision makes it an invaluable tool for those in fields such as computer vision, robotics, and autonomous systems. [33].

In the Philippines, traffic congestion not only wastes time but also hinders productivity and, in some circumstances, puts lives in danger. In this state, one of the aspects of the field of medicine that is greatly affected is the response time of ambulances in emergency situations. Despite the preferential treatment afforded to ambulances on the road, determined by their distinctive markings, vehicle classification, and the activation of an audible siren to alert other motorists, this prioritization frequently proves inadequate. In instances of significant road congestion, ambulances face significant operational challenges, rendering them unable to provide timely transportation for their patients to medical facilities [34].

The ambulance delay due to traffic congestion poses severe risks, including potentially fatal accidents for patients. Such issues demand prompt resolution to prevent these unfortunate events. With the aid of the established algorithm YOLOv5 and the small weight, this technology opens an opportunity to classify ambulances on the road and identify them in traffic. A main point for this model is that it uses a backbone of CSPDarknet52 architecture for features extraction, neck with PANet Architecture for feature refinement, head with class probabilities, ground truth comparison with less computation, optimizer with stochastic gradient descent, non-maximum suppression by refine bounding box predictions, for the validation. Towards classifying these ambulances, it will be easier for them to be prioritize in traffic and be able to act on their duties and responsibilities immediately. Moreover, through the use of a small weight version of YOLOv5, real-time ambulance classification and traffic identification by significantly reducing the computational load, while still maintaining a high level of accuracy. The science of computer vision and its applications in emergency services and traffic management will undergo a revolution with its technological edge.

Background of the Study

Traffic congestion in the Philippines is a long-time issue and continues to worsen ever year. According to the research [31], EDSA had 367,728 cars, Commonwealth Avenue had 272,720, while Circumferential Road 5 (C5) had 218,751 vehicles. Every year, the increasing number of vehicles causes a higher percentage of traffic congestion on major routes, which is not related to the area of the roads. [31]. This type of issue only causes one serious problem after another. As when the time comes when person or hospital personnel is under an emergency protocol, they won’t be able to attend to an urgent situation only due to heavy traffic [31]. This led the researchers to make a study on Ambulance Detection through computer vision object detection which aims to utilize an algorithm established for accurate detection that has faster inference speed, lower storage requirements, good portability, with less overfitting.

Further, with the previous research that is available online about prioritizing ambulances, most of it use different IOT devices for detecting and alerting officers about the ambulance vehicle that is stuck in traffic [6, 7, 8, 9, 10, 11, 12, 14]. These studies are effective in prioritizing ambulances and other emergency response vehicles. However, due to their costly nature [23], they may not be suitable for the Philippines, which is still a developing country. This research differs in various ways. One study is a vision based [13], others a combination of audio and vision based to detect ERV's [14]. The research used audio-based detection and can be said to be effective and successful [14]. However, this audio-based research cannot be applied locally due to the amount of noise produced by vehicles, people, and many more things on the road. Audio-based detection of ambulance sirens may not seem suitable in the bustling environment in the Philippines setting since audio receivers may make false positives due to the noise pollution. Other research utilizes networks to signal the driver of the ERV and the officers on what route they should take to lessen the time travelled going to the set destination [23, 28]. Apart from these studies, most of them focus on vehicle detection solely [20, 23, 28, 30]. This just demonstrates that the structure of their model is intended to recognize a vehicle plainly rather than distinguish between different vehicle kinds. On this occasion, the researchers saw a clear opportunity. They decided to further their study on distinguishing different types of vehicles, especially ambulances, using object detection with hyperparameter and augmentation techniques. Given this, the researchers propose YOLOv5 Small as an ideal model for efficient image processing, particularly for detecting labels and features on vehicles. This lightweight variant of YOLOv5 offers the advantage of ease of deployment compared to other heavier models, making it a practical choice for various applications. The object detection model used in this study will take in account the different physical outward appearance of ambulances found in the Philippines by training the model using a localized dataset. The reason for this is that ambulances from other countries are very much different compared to the local ones. In pursuit of this study, there is an opportunity to demonstrate the existence of alternative or superior methods for detection that do not necessitate the use of an excessive number of sensors, which adds to the operating costs of traffic management, especially with regard to ambulance prioritization [35]. In addition, the researchers will use this study as an opportunity to further broaden existing studies by performing the same framework on a dataset that has not been tested before. Towards this study, the field of computer vision will be extended.

YOLOv5 model is available in small, medium, and large weights and each have different model complexities. For applying in real-time vehicle detection, YOLOv5 small would be the best as it offers accuracy and speed in detection, it also has the smallest size among the three variations available [2]. The small weight also contains fewer parameters than medium and large weight that makes it faster and more efficient for real-time and real-world deployment as it provides the responsiveness and throughput needed for time-critical vehicle detection applications like in traffic monitoring systems. Its high performance, small size and minimal latency render it the most suitable choice among the YOLOv5 family for vehicle detection tasks.

Moreover, the researchers thought that it would be more suitable for the country mentioned if the method of detection used is the combination object detection and classification with hyperparameter tuning for accurate results. It is different to other studies available with its hyperparameter tuning, augmentation techniques, and with the dataset that is used for training and testing the model. A local dataset is necessary to the training stage of the proposed model due to ambulances from other countries having striking differences in appearance to ambulances found in the Philippines. According to the Department of Health (DOH), the Philippines has three standardized ambulance types: Type I, Type II, and Type III ambulances. Type I Ambulances are small trucks with a box-shaped patient compartment, typically used for Basic Life Support (BLS). Type II Ambulances are heavy-duty vans with minimal modifications, equipped for Advanced Life Support (ALS). Type III Ambulances are similar to Type I but use a heavy-duty van chassis instead of a truck for the patient compartment. Each type of ambulance must include features such as the siren and sticker labeling.

Overall, this study will be conducted by the researchers mainly to provide and contribute more knowledge to the world of computer vision. The study will utilize a modified YOLOv5 small weight model that is optimize through hyperparameter tuning and the use of 6 different augmentation techniques. With the study's proposed framework, the model will be able to identify and classify the ambulances on the road with accurate results. With the study, it may help to save more lives in the future through detection and classification of ambulances on the road.

Gaps/Opportunity

Currently, the study of ambulance detection towards computer vision is growing and spreading. However, in these studies, it focused with the ambulances internationally particular to European and American countries. Computer vision applications in the context of ambulance classification in the Philippines are relatively unexplored. There is a lack of existing research in this specific area. With the study being limited to the mentioned countries, there is an open opportunity to widen the field of computer vision in terms of ambulance detection and classification in the Philippines.

With the previous studies mentioned, one of their main similarities is the use of a computer model to detect vehicles in general [19, 23, 28, 30]. The prominent disparity became apparent, prompting the researchers to consider it as a potential avenue for an extended investigation into the application of the YOLOv5 small weight model for ambulance vehicle detection via image processing. Moreover, to unlock the model's full potential, a specialized Philippines ambulance dataset must be created due to the lack of a localized dataset online. The creation of this localized dataset presents an opportunity to align the proposed model with the specific context of the Philippines, ensuring accurate ambulance detection and classification within the country's distinct environment.

This study explores the viability of a computer vision-based ambulance detection model for detecting and classifying the ambulances on the road using a small weight version of the model. The small model is ideal for real-time vehicle detection due to its balance of speed and accuracy. Its compact size and fewer parameters make it efficient and responsive, perfect for time-sensitive applications like traffic monitoring. Among its variations, YOLOv5 small is the top choice for vehicle detection considering performance, size, and latency. The training process of this model employs a comprehensive approach with advanced neural network architecture and optimization techniques [2].

Another challenge lies in the area of model generalization. Crafting a model that effectively adapts to a variety of environmental conditions can be a daunting task, especially considering the diverse terrain, lighting, and weather conditions found in the Philippines. This situation opens up an opportunity to apply range of augmentation techniques and hyperparameter tuning to enhance the model’s accuracy. Augmentation techniques such as Flip, Hue, Saturation, Brightness, Blur, and Noise can be utilized in this process. Through the presence of hyperparameter tuning, it entails a systematic exploration of learning rates, batch sizes, and other model-specific parameters. Moreover, these adjustments are important to achieve the perfect equilibrium between the intricacy of the model, the pace of training, and the precision of detection.

Statement of the Problem

The purpose of this study is to use a newer version of YOLO from an existing detection framework and train it using a localized dataset. Although there have been previous studies on detecting vehicles through computer vision [19, 23, 28, 30], ambulance and ERV detection for road prioritization have mostly been done through the use of IOT devices. The model used in several studies are R-CNN, YOLOv3, YOLOv4, and the combination of Harr and HOG. All of these models only target the overall structure of the vehicle on the road. With these studies, it led the researchers to adapt a YOLOv5s model for ambulance detection and classification to take advantage of real-time inference, faster training, fast inference speed, and deployment flexibility. The study will also be mixed with the utilization of different hyperparameter tuning to have better and accurate results. In this case, ambulance will be better classified according to its type or classes and be easily detected on the road.

Specifically, it aims to do and answer the following:

1. Is the YOLOv5s model able to attain high performance metrics in detecting and classifying different types of Philippine ambulances and its features?
2. By using the Structural Similarity Index (SSIM) to evaluate the dissimilarity between clean and augmented images, what is the average SSIM score, on a scale from 0 to 1, indicating the degree of diversity and non-redundancy achieved in the dataset?

3. By applying hyperparameter tuning, what is the optimal set of hyperparameters that should be used to maximize the performance of the YOLOv5s model?

Objectives

This project aims to develop a customized vehicle dataset for ambulance detection in the Philippines and explore the scalability of an existing framework to create an improved YOLO model (YOLOv5s) for identifying ambulances different types and classes among standard vehicles on the road, facilitating their prioritization in traffic. The objective is to expedite ambulance travel times, enhancing patient care, and be able to utilize a model with hyperparameter tuning and augmentation techniques applied that is accurate for detecting and classifying the types of ambulances on the road. The researchers have created the following objective declarations to meet the research's goals:

1. Assess the performance metrics of the YOLOv5 model, specifically in the small weight, in detecting and classifying Philippines ambulances and their key features.
2. To assess the dissimilarity between clean and augmented images using SSIM in order to determine the average Structural Similarity Index (SSIM) score, which would reflect the diversity and non-redundancy attained in the dataset.
3. To identify the right set hyperparameters through hyperparameter tuning in order to maximize the performance of the YOLOv5 small model.

Significance of the Study

This study proposes an ambulance detection and classification system that is based on image-based object detection technology instead of the traditional methods of alerting traffic enforces about oncoming emergency response vehicle particular to ambulances. A YOLOv5 ambulance detection model has significant implications in the field of computer vision and machine learning, as it allows for real-time detection and localization of Philippine based ambulances in various settings. Adding a small weight of the model's version provides an improved efficiency, deployment flexibility, and faster inference speed [2]. Using computer vision technology may provide better performance and efficiency as compared to traditional methods that may have been criticized for ambulance delays and their ineffectiveness. Through this study, the detection and classification of ambulances on the road could be significantly enhanced and improved leading to addressing of the lack of prioritizing of emergency response vehicles on the road. This will also guide the investigation into optimizing hyperparameter tuning and applying augmentation techniques to achieve the best and most accurate results for detection. Furthermore, this will also help to determine whether object detection models can effectively differentiate distinct vehicles along the road and further detect ambulances, potentially serving as a more affordable substitute for costly Internet of Things (IOT) devices for prioritizing ambulance response. In return, the results and findings of this study may be used in future research. Moreover, being able to facilitate this research will further contribute to the added knowledge in the field of computer vision and prioritization will be given to ambulances which are beneficial for the following people:

1. Paramedics – These are the skilled drivers of the Emergency Response Vehicle or the ambulance. Prioritizing them along the road will lessen stress and accidents that could happen due to patients rushing towards the hospital.
2. Doctors - They are tasked with taking care of and curing patients that are feeling ill and sick. Bringing their patient at the right time and less delay will give them ample time to diagnose them properly and aid them with their needs.

3. Patients – As the paramedics bring their patients to the hospital fast and safely, while the doctors diagnose them with the proper aid, the patients will be able to be attended to by professionals and be given the right medication they need. Also, this will help to avoid worsening the situation of the patient due to their illness.

4. Image detection researchers – Towards the continuous expansion of human knowledge, this study will provide more knowledge about vehicle detection. It can provide information on what and how the performance of an existing framework will be using a localized dataset with the help of YOLOv5 small weight model and the application of hyperparameter tuning with the distinct augmentation techniques.

Scope and Limitations of the Study

The study primarily focuses on the detection of ambulance, specifically, in the Philippines using image processing techniques. The research builds upon an existing emergency response vehicle detection framework that uses a faster type of R-CNN model, but this study will utilize YOLOv5 with a small weight version together with the proper hyperparameter tuning and augmentation techniques particular to Flip, Hue, Saturation, Brightness, Blur, and Noise to refine the object detection model. The proposed framework will use a locally collected dataset that would be for training the YOLOv5s model.

The dataset that will be created by the researcher are images that will be downloaded online which will be localized ambulance in the Philippines. All of the ambulance images were only taken in the day without a specific angle and set up. The process will involve image scraping. The dataset will contain the three different types of ambulances in the Philippines. Type I, II, and III ambulances have unique features. Type I, a small truck with a box-shaped patient compartment, is for Basic Life Support (BLS). Type II, a heavy-duty van with minimal changes, is for Advanced

Life Support (ALS). Type III, like Type I but with a heavy-duty van chassis, is also for Basic Life Support. Essential features for all types include sirens and prominent labeling stickers. Further, these images will be utilized for ambulance vehicle detection and classification through image processing.

The researchers selected YOLOv5 small weight as the main model which will be applied with hyperparameter tuning and six different augmentation techniques to generate an accurate detection and classification model. The research does not extend to real-time implementation or other types of emergency vehicles. This study will conduct computer vision applications such as data gathering (image), data cleaning, data annotation, data augmentation, data augmentation using SSIM, model creation, hyperparameter tuning, validation and testing. The performance metrics of the model for validation and testing will include its precision/recall and mean average precision.

Chapter 2

REVIEW OF RELATED LITERATURE

YOLOv5

The study [1] on a novel and efficient method for a high-precision and rapid detection of smoky vehicles using embedded devices by recognizing the speed limitations of the existing target detection models on embedded hardware, this study also presented an improved lightweight YOLOv5 based network with the incorporation of an efficient Mobilenetv3-small backbone. Researchers discovered that by using this type of model, it will reduce the model's complexity to just 0.48M parameters which significantly lowers the computational load of the model. In enabling a precise vehicle exhaust detection, researchers have used a dedicated dataset of 6,102 images which were curated and have been augmented. The augmentation techniques used in this study were cutout and saturation transformation which were applied in order to overcome the challenges of the researchers because of shadows and occlusions. The results obtained in this study have an impressive 8.5% accuracy improvement due to data augmentation. Another is that the optimized model achieved a real-time detection speed of 12.5FPS on the embedded devices which doubled the YOLOv5 performance.

In this study [2], researchers have a comprehensive study on cutting-edge surveillance techniques which has a vital application on military and civilian domains. The study investigates the real-time object tracking solutions for Unmanned Aerial Vehicles using Deep Neural Networks which is specifically tailored for YOLOv5 Model. Researchers have modified the YOLOv5 architecture which includes the adoption of Rectified Linear Unit activation as well as a fine tuning hyperparameters. This study also employs the AU-AIR dataset for evaluating different YOLOv5 models based on depth for enhancing the training speed and accuracy. The overall performance of the model was examined using mean average precision as metric which has remarkable results. The modifications made by the researchers provided 2.9x performance gain over the original unmodified YOLOv5 model.

In the study [3], researchers have explored on the crucial task of object detection in Artificial Intelligence and Deep Learning, recognizing the limitations of traditional detection methods which are often produce a lot of false positives that reduces the accuracy of the model. This study also investigates a multi-objective real-time detection method which is made for vehicles using the YOLOv5 architecture. This study involves contructing a dedicated YOLOv5 based model in acquiring real-time vehicle video data and tracking or detecting multiple vehicles. Another is a comparative experiment to demonstrate the methods efficacy in reducing the false alarms compared to other models. Enhancing monitoring precision to meet Intelligent Transportation Command Center accuracy requirements for vehicle information acquisition.

Data Augmentation Technique

In this study [4], researchers have acknowledged the complexity of certain detection tasks, the goal of this study is to enable advancements for autonomous driving. This study utilizes Cascade R-CNN model with data augmentation. The strategy that the researchers sued have strong results which they won 4th place out of 187 preliminary participants in a competition. Researchers have implemented an open source “mmdetection” toolbox in PyTorch. The model has a selection process targeted for optimal performance on the COCO dataset. Cascade-RCNN that have been utilized were Cascade R-CNN with ResNeXt101 and FPN backbone for multi-scale object detection strengths, modified version of Cascade RCNN with a Global Context Block for enhancing contextual understanding, and a YOLOv5 model with ResNet50 backbone and deformable convolutions to optimize detection. Augmentation techniques were critical for robust training. Classical methods from Albumentations, like blur and color shifts, facilitated image-level changes without altering key annotations. Horizontal flips leveraged the consistent orientation of road signs. The systematic process resulted in a solution well-equipped to handle the complexes of real-world object detection. Vehicle identification and traffic statistics at traffic crossings are an efficient method for reducing traffic congestion and achieving intelligent traffic.

Emergency Vehicle Prioritization

Due to the issue in urban traffic management and traffic congestion, it has seen a big impact on responding emergency vehicles, significantly delaying the response times to emergencies. The study [5], researchers have explored the implementation of Artificial Intelligence for ambulance classification which is based on the features that identifies them like “Ambulance”, Siren, Red Cross Symbol, and the Star of Life Symbol. Implementing a Faster-RCNN framework on TensorFlow, it involves pre-processing of images techniques in order to enhance the accuracy of classification. The study demonstrates the effectiveness by having accuracies of 95.6% for the text “Ambulance,” 97.9% for the Siren, 99% for the Red Cross Symbol, and 98% for the Star of Life Symbol. This only shows that the systems capability to accurately classify and detect the incoming ambulance which contributes to a more efficient traffic management and the reduced response times to emergencies.

As the world's heavy traffic continues to increase, a big portion of an ambulance's duties are jeopardized. They must be given precedence on the route to ensure that they arrive at their location as soon as possible to provide relief for the emergency. This article [6] discusses signal light priority control, signal light cycle recovery, and signal light timing-recovery, and then builds three matching models to secure emergency response vehicles' pass priority. Following that, the paper employs double intersections as the research object in order to construct an SUMO simulation platform for simulation studies. In terms of time, the priority traffic strategy saves 11.29%, 18.48%, and 48.49 under normal traffic, light traffic congestion, and moderate traffic congestion, respectively [6].

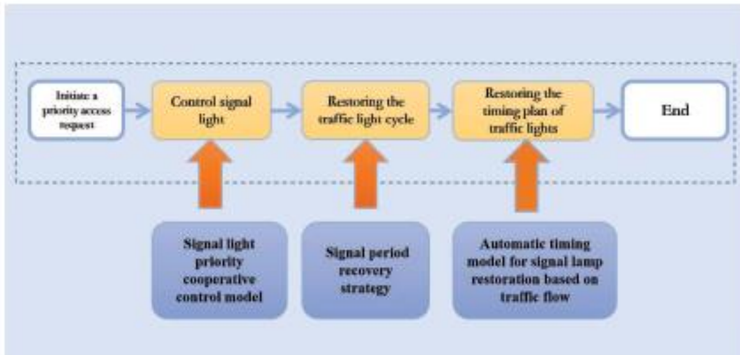


Figure 2. 1. Special Vehicle Priority Traffic Model.

The figure above features Special Vehicle Priority Traffic Model which is a system that helps emergency vehicles move through traffic more quickly. Seen in the figure the process starts with the sensors to detect emergency vehicles and communication technology to give them priority over other vehicles on the road. Once an emergency vehicle is identified, it sends a priority request to the SVP system. The SVP system then analyzes the request and prioritizes the vehicle's movement through traffic. This can involve changing traffic signal timings, activating traffic signal preemption, or creating a dedicated lane for the emergency vehicle. Once the necessary changes are made, the emergency vehicle can proceed through traffic with minimal delay. The SVP Traffic Model is an important tool for improving response times for emergency vehicles and enhancing overall traffic safety.

The goal of the research [7] on traffic control systems concerning density with emergency priority mechanisms is to create an automated traffic control system outline. In addition, this outline must also include programmable logic controller technology as well as a visual representation of these necessary adjustments when applied to SCADA software. To improve conventional traffic control systems, this study proposes the utilization of IR sensors as a convenient way for emergency vehicles to pass with fewer complications. The IR sensors have a huge role in vehicle detection since these sensors provide the exact number of vehicles on a certain road. Furthermore, SCADA, which is a centralized system responsible for controlling an area, is used to represent the actual model. Through this study [7], the researchers concluded that this system would avoid excessive waste of fuel cost and time. While SCADA software successfully demonstrated the entire process of the model, Python shows that this process may be smoothly utilized. This system may contribute unintentionally to improving the flow of traffic and avoiding congestion on the roads. This research helped to give the study an option on what software to use for simulation specialized for traffic in intersections.

The study [8] on the algorithm implemented on traffic management with priority for emergency vehicles focuses on providing an advantageous algorithm for an improved speed of emergency vehicles, particularly during heavy conditions. Concerning this, the traffic management algorithm was utilized for this study and as a guide to derive the information required to amend traffic appropriately. The smart traffic management system uses image processing, GPS tracking, and a cloud-based control center. Furthermore, the purpose of traffic monitoring data is to represent the capacity of road connections, detect incidents and monitor congestion. The transmission of data usually stops once a vehicle passes through an intersection. With this, traffic degree will be detected, enabling a directive of one intersection at a time and a controlled range for radio frequency signaling. Through this study, the researchers concluded that the proposed algorithm helps improve the movement of emergency vehicles during heavy road conditions through smart traffic management and efficient information [8].

There is always a possibility of unfortunate events on the road. In relation to this, reducing the clearance time of incidences, as well as imparting a clear pathway for emergency vehicles, is an efficient way to address these events. Despite this, emergency vehicles are assigned based on the severity and type of incident that took place. In response to this problem, the Emergency Vehicle Priority System (EVPS) was used to determine the rank and level of priority an emergency vehicle has, enabling an estimated number of signals. In this research, the algorithms used were the number of signal intervention estimations and estimated signal clearing time, respectively. Traffic conditions wherein an emergency vehicle arrives at the incident place were also monitored. The results of the paper [9] showed that the travel time of the emergency vehicle decreases along with an increasing number of interventions. However, considering the impact of other vehicles in cells that are monitored, the number of interventions is relative to the distance and occupancy rate of the emergency vehicle. The findings of this study [9] show that the system can recommend appropriate intervention numbers. Hence, the proposed system is significant in saving the life of a human and minimizing financial loss that may happen. The positive outcome of this research is that the suggested system can recommend appropriate intervention numbers, which can assist in saving lives and limit financial loss caused by traffic accidents.

The restriction of road intersections along with priority vehicles is important, specifically for their performance. The objective of this paper [10] is to implement Deep Reinforcement Learning in the traffic light control of a road intersection in relation to the existence of the three types of priority vehicles. Moreover, the paper [10] also emphasizes top-level dynamics or managing traffic. This paper [10] has two main objectives, which are to reach high performance and manage the three types of vehicle priority. Along with this is the stimulation of Urban mobility (SUMO), specifically a Traffic Control Interface (TraCI). The case study model specifies the light control system referring to the DRL. With this objective, the actions performed were the reward function and learning parameters. The researchers have concluded that the model used may be modified according to different intersection layouts allowing them to share data on traffic as well as self-driving vehicles. In this way, they can further communicate and cooperate, reducing the time allotted for travel. This study is an excellent starting place because it offers an innovative approach to a real-world problem, and the results are promising.

Vehicle Detection of different models and algorithms

Propose [11] an algorithm for dispatching emergency vehicles in smart cities to improve emergency management and reduce negative impacts of critical events. The algorithm uses sensors and real-time handling, prioritizes vehicles by optimizing traffic lights, and will be evaluated using simulation tools. The aim is to contribute to emergency management in smart cities. the dataset used by the researchers were results from combining multiple sensors and alert systems in detection for deploying emergency response vehicle. A sensor-based approach for emergency management in smart cities that uses IoT devices, sensors, and sensor networks to monitor urban environments and detect abnormal situations. The sensors collect scalar and visual data from multiple areas, process them to identify abnormal situations and release emergency alarms, which contains spatial (GPS coordinates) and temporal (timestamp) information of the emergency, as well as a numeric severity index computed using different data. The system can be configured according to the characteristics of any considered city, with a high degree of autonomy in its services. The alarms are processed to guide the dispatching of emergency vehicles or any other mitigation service. A JavaScript simulator was developed to generate random events on a map, using parameters such as the number of events, types of events, a central point, and a radius. The system generates the event location and type and sends an emergency alarm to an operations center for response. It uses fixed treatment centers and calculates routes using the Multi-Level Dijkstra algorithm. The simulation includes two types of vehicles, passenger and emergency, and settings such as broadcast interval and warning EV distance.

The study [12] focuses on the use of sensors-based monitoring systems and Internet of Things technologies in urban areas for emergency management, specifically the development and implementation of a dynamic algorithm for automatically assigning emergency vehicles in smart city scenarios in response to emergency alerts. The goal is to improve crisis management and potentially save lives. Due to urbanization and high population density, emergencies are growing more frequent in cities, which causes a sizable number of fatalities each year. Exploiting an existing sensors-based emergency warning system is one idea for solving this problem; this might potentially speed up rescue operations by promptly allocating special vehicles to respond to situations that are spotted. Researchers suggested method that limits the allocation of cars and lowers the number of vehicles that aren't accessible for subsequent emergencies by assigning no more than one vehicle of each type to each alert. Additionally, other intelligent transportation systems, like those that manage traffic lights, can make use of the route that the chosen cars travel. Researchers implemented an alarm system in which all emergencies are processed by the system, then a list of candidate vehicles are computed based on the mapping between type of emergencies and available emergency vehicles available in the area. If there are more alarms than vehicles, the closest vehicles are selected to each alarm using Dijkstra's algorithm. If there are more vehicles than alarms, the severity of the alarms is considered as a decision factor for assigning vehicles to alarms, using distance and emergency magnitude. The outcome is a group of ordered pairs indicating which vehicles should move to which destinations. In certain situations where multiple emergencies occur simultaneously, such as heavy rain or earthquakes, there may not be enough emergency vehicles to respond to all the alarms. To address this critical issue, the proposed approach includes a specific protocol to handle situations where there is a shortage of available emergency vehicles. The proposed approach for rapid emergency response in smart cities has the potential to save lives and reduce economic losses. However, validating it in real cities can be costly and complex, requiring a significant amount of time for quality assessments. Additionally, the COVID-19 pandemic made it difficult to safely conduct research in various locations, including in the target city of Porto, Portugal. As a result, we decided to initially validate the approach using simulated data, while reserving more complex analysis for future studies. The approach prioritizes critical alarms when there are fewer emergency vehicles available, which was the focus of our evaluations. Using two active Event Detection Units, we simulated emergency alarms with varying numbers of events in the city of Porto, retrieving the alarms via MQTT protocol and simulating the assignment of vehicles based on them.

The authors propose [13] a novel audio-vision emergency vehicle detection system (AV-EVD) that uses both image and audio data to detect emergency vehicles. They developed a modified YOLO model called YOLO-EVD for vision-based detection, which is tailored to the emergency vehicle detection problem. The YOLO-EVD achieves high detection accuracy and speed, surpassing results of two-stage counterparts. The reserchers also proposed a complete end-to-end model which is the WaveResNet intented audio-based detection. WaveResNet is designed for feature extraction and classification and achieved much higher accuracy compared to prior works. The two systems proposed are then integrated into one called AV-EVD, which is very useful when a highly reliable EVD is needed. AV-EVD meets the real-time requirements and can be used in private cars, autonomous vehicles, or smart roads. It achieved 95.5% mean average precision and 5ms processing per image.

This paper [14] is to propose a method for efficiently detecting an emergency vehicle in order to prioritize their movements and improve response times in emergency service workflows. The method uses the Shortest Job First scheduling algorithm to identify the sound of the EV siren using the Goertzel algorithm. The emergency vehicles and its type are detected from a roadside unit by comparing the magnitudes of tone calculated in the roadside units, and using this information to predict the EV position from the traffic signal. The study aims to prioritize individual EVs based on type, location, and traffic density in an IoT Intelligent Traffic System. This is to improve congestion and response times.The method outlined in this study requires the use of a DTMF tone detection circuit to identify the DTMF tone emitted by an emergency vehicle. The Goertzel algorithm is effective in this case due to hardware limitations. This algorithm involves determining the energy level of the specific frequency pair listed, using the incoming siren sound, and using this information to determine which frequency is present in the siren sound. The Arduino environment has pre-existing libraries, such as dtmgen.h for DTMF tone generation and Goertzel.h for frequency detection, making it possible to implement both the detection and generation circuits using an Arduino board.

Emergency Response Vehicle Detection

In this paper [15], researchers focused on implementation of emergency a vehicle signal pre-emption strategy on major road intersections. This study focuses on implementing a emergency vehicle signal pre-emption strategy at major road intersections in the Indian environment, focused on the Trivandrum city as the researchers area of study. The study gathered information on traffic volume, vehicle speeds, current traffic light timing, and road design for three intersections with traffic signals. The suggested approach looks at three different traffic situations, with the goal of clearing any traffic at intersections along a set emergency route ahead of an emergency vehicle. This is done while trying to minimize any extra delay from unnecessary activation of the traffic signal preemption system. The proposed design includes determining the detection distance for emergency vehicles. The design was simulated in a traffic network using a microscopic simulation model and was calibrated with the data collected from the network. The results show that the proposed method can reduce emergency vehicle delay at signalized intersections by up to 25-30%. The dataset used in this study were gathered for three signalized intersections, including information on traffic volume, spot speed, current signal timing, and geometric characteristics. Researchers considered three approaching scenarios so that they would understand the traffic queues at major intersections on a given emergency route for the emergency response vehicle, selection of the study location to be experimented in. Researchers also developed a simulation model in VISSIM and calibrated the network model in VISSIM. The process of creating a network layout for a study area using VISSIM software. Road links are created with the use of road geometry details and an Open Street Map base map. Vehicle input, speed data, and composition were inputted into the model using data collected from a survey. Queue counters were modeled and placed at each signal stop line to determine queue lengths. The existing signal details were programmed into the model to design the signal controller as a fixed time signal.

The scope of this study [16] is to review the latest traffic control strategies for reducing the response time of emergency vehicles during their travels. The study classifies traffic control strategies into different categories such as route optimization, signal pre-emption, lane reservation, and mixed traffic control strategies. It also presents a systematic literature review of traffic control strategies with different algorithms and classifies the articles by objective metrics such as response time, and other objective metrics. The study also reviews the limitations of existing emergency traffic control strategies and critically analyzes them, indicating core problems and proposing potential research areas for further exploration. Examination of 81 research articles from 2017 to 2022, researchers ensure real-time performance of this study. Additionally, researcher’s objective metrics referred to by the studies are classified for comparison and analysis. Emergency vehicles can reduce delays by utilizing traffic signal preemption strategies, which include rule-based methods, traditional optimization algorithms, intelligent optimization algorithms, and machine-learning-based algorithms. Route optimization algorithms are divided into several categories including traditional optimization algorithms, intelligent optimization algorithms, machine-learning-based algorithms, and other specific algorithms. The development trend in traffic control strategies is moving from using a single objective metric to using multiple objective metrics. This is because multi-objective metrics consider the needs and interests of all the different parties involved in a traffic road network. As a result, using multi-objective metrics improves the practical application of traffic control strategies in real-world scenarios.

In this study [17], researchers propose and investigate an open IoT system for traffic monitoring and management of the city, with the goal of reducing the response time on emergencies. The system would use communication between emergency vehicles and the city's surveillance cameras network and traffic-light system to dynamically change their schedule and provide a green wave for the emergency vehicles. The study focuses on the theory behind the proposed system and presents results demonstrating that the subsystem used to identify and locate the emergency vehicles works properly. The study also discusses next steps for implementing all necessary subsystems to be integrated. Researcher created an open Internet of Things (IoT) system that would integrate with the current infrastructure of the city. This system would connect emergency vehicles with the network of surveillance cameras and traffic lights, allowing for real-time adjustments to traffic signals to create a "green wave" and ultimately decrease emergency response times. This paper examines the concepts and principles behind our proposed system. The data shows that the system for identifying and finding emergency vehicles is functioning correctly.

The study of this paper [18] implementing deep learning for vehicle detection and using TensorFlow with an emphasis on improving accuracy and speed. The study utilizes TensorFlow, and YOLO object detection algorithm for real-time detection of vehicles. The study compares the performance of the latest YOLOv4 algorithm to previous models using a custom dataset and the DeepSORT algorithm to aid in counting the number of vehicles in videos. The goal of the study is to determine that the YOLOv4 algorithm is the best model for vehicle detection, achieving state-of-the-art results with 82.08% AP50 at a real-time speed of 14 FPS on a GTX 1660ti. In the dataset used, utilizing the volume of vehicles as valuable data for detecting traffic congestion in which then benefits the traffic management. The method utilized by this paper implements a vehicle detection system using machine learning and the DeepSORT algorithm. To complete the project, the necessary dependencies and software, such as Git Bash, must be installed on a windows 10 machine. Python is the primary programming language used. The project also involves collecting and labeling images of vehicles for training a model. This process can be made easier by using the OIDv4 toolkit, which can automatically search for and label images in the YOLOv4 format. The use of Yolov4 and yolov4-tiny is shown to be efficient and faster than previous models.

With this study [19], researchers propose a method for vehicle detection and traffic statistics at complex traffic intersections using the YOLOV3 and DeepSORT algorithms. The goal is to solve traffic congestion and achieve intelligent traffic management by accurately identifying and counting vehicles at multi-mode traffic junctions. Dataset from traffic statistics method at complex traffic intersections. Researchers used DeepSORT tracking algorithm for detection result correlation in video sequence. The trained model will then be used to identify vehicles in video footage. The detection results will include the class, confidence score, and detection box of the target. The classes include car, bus, motorbike, and truck. The confidence score indicates the likelihood that the detected object belongs to a specific class. The higher the confidence score, the greater the probability that the detected object belongs to that class.

Researchers have developed a framework for tracking and estimating the speed of vehicles using lidar technology. The study [20] aims to improve the accuracy of vehicle speed estimation by implementing a centroid-based tracking flow and using an unscented Kalman Filter and a joint probabilistic data association filter. The framework is tested using lidar data from two different panoramic 3-D lidar sensors at a traffic light and a road intersection. The study evaluates the effectiveness of the proposed approach by comparing the results against reference data obtained from a test vehicle equipped with accurate positioning systems. The study shows that the framework can detect and track more than 94% of vehicles, with a mean speed accuracy of 0.22 m/s. in the dataset used by the researchers, Cars on the road are identified using point cloud observations, then a method called centroid-based tracking flow is applied to determine their initial positions and movements. A Vehicle tracking method that uses image matching to enhance the accuracy of vehicle speed estimation. The approach was evaluated using lidar data from two different panoramic 3-D lidar sensors, a RoboSense RS-LiDAR-32 and a Velodyne VLP-16, collected at a traffic light and a road intersection to account for real-world scenarios. Researchers have found that using high-precision 3-D lidar data can result in a speed accuracy of around 0.2 m/s, which can be used to improve the analysis of traffic flow and behavior.

The paper examines the effectiveness of an emergency vehicle detection system with the use of YOLOv3 in optimizing the response times to accidents and emergencies of responding vehicles. In this study, researchers have utilized a dataset of 300 images for training and evaluating the models. These images contain ambulances, fire trucks, and police cars and were annotated with bounding box in order to expand the data. With a total of 25 YOLOv3 models trained with different parameters, the results are evaluated based on their Mean Average Precision with their highest result of 0.9878 mAP in which results to a high accuracy in recognizing objects. With the utilization of YOLOv3, this verifies the study's accuracy when it comes to having an effective way in automating emergency vehicle detection. The paper provides a significant improvement when it comes to integration of detection systems in traffic management to reduce delays in emergency response times. A wider and larger dataset with the application of real-world testing would further improve the results.

YOLOv5 Detection

The paper [21] created a YOLOv5 model for detecting several classes of vehicles in four different categories by integrating the attention mechanism and the lightweighting method. Specifically, the researchers applied the SENet attention mechanism in the CSP1 residual block of the YOLOv5 backbone component and replace certain Relu functions with PRelu activation functions. The dataset from this paper consisted of a total of 10 hours' worth of traffic footage taken from the Beijing and Tianjin zones of China. With a resolution size of 960x540 pixels, over 140,000 frames of traffic footage were manually annotated and classified into cars, vans, buses, and other types of vehicles. The weather conditions were also classified into four categories: sunny, rainy, cloudy, and night. Several lightweight based YOLOv5 models such as YOLOv5s, Improved YOLO, Improved YOLO5, SE-YOLOv5, YOLOv5-DSC, SE-YOLOv5-DSC, and YOLOv4 model were created for comparison. Each model was trained using the created dataset with a learning rate of 0.001, a batch size of 8, and an epoch size of 100. The findings indicate that the YOLOv5 improvement algorithm proposed in this paper outperforms the original YOLOv5 network in terms of accuracy and speed. Specifically, the accuracy of this model is 3.1% higher than using YOLOv5 alone, without any decrease in speed. While other improved models are available, they exhibit lower speeds even though they have good accuracy results in comparison to this model. One of the good points of the paper is that it compares different lightweight based YOLOv5 models such as YOLOv5s, Improved YOLO, Improved YOLO5, SE-YOLOv5, YOLOv5-DSC, SE-YOLOv5-DSC, and YOLOv4 model for comparison.

This study [22] created a smoke detection model that detected smoke coming from a vehicles' exhaust. This was done by improving a lightweight YOLOv5s model using MobileNetv3 small in order to detect a vehicle's exhaust accurately. Comparing the proposed model to the base YOLOv5 model, there was reported increase of accuracy by 8.5%. The self-built dataset in this research consisted of a total of 6102 images of vehicle exhaust. Pre-processing techniques such as cutouts and saturation transformation were used in order to increase the count of the dataset to further increase the accuracy of the YOLOv5 model. The vehicle exhaust dataset was then used to train the YOLOv5s model with the applied MobileNet. In comparison to the base YOLOv5s model, the YOLOv5s model with MobileNet was able to perform its task with a reduced percentage of params, flops, and overall model size by 91.89%, 93.42%, and 93.45%, respectively.

The paper [23] reviews and evaluates different vehicle detection methods and suggests using a CNN-based method using YOLOv5s architecture and k-means algorithm in performing the anchor box optimization for different illumination levels. The dataset used in this model was self-built by taking 20-minute duration video recordings for each time scenario. The time scenarios included afternoon and evening. Approximately 750 images for each dataset were collected. Data augmentation was then used to help increase the accuracy of the model. The proposed solution utilized the k-means algorithm, which led to a significant improvement in the detection rate.

Specifically, it resulted in a 5.62% increase in the mean average precision (mAP) for the daytime dataset (A) and a 5.99% increase for the nighttime dataset. This demonstrates that optimizing the anchor box selection in object detection models such as YOLOv5 can lead to improved detection. One good point from this paper was its good use of data augmentation for its dataset that increased the performance of the model overall.

The paper [24] proposes a new vehicle detection and tracking model is in order to improve accuracy, reduce ID-switch and enhance anti-interference. The model is based on Attention-YOLOv5 and optimized DeepSort. Attention mechanism is added to the front-end target detector, YOLOv5, to improve feature extraction. The model is trained on a BDD dataset and the ECA-YOLOv5 had the highest detection accuracy. The DeepSort algorithm is also optimized by improving the re-recognition network structure, re-identification pre-training on Improved-VeRi dataset, and proposing a new loss function. The model was tested on highway traffic videos and results showed improved tracking performance compared to the benchmark algorithm. One aspect of this paper was its use of a tracker such as Deepsort in order to help the performance of the YOLOv5 model.

This paper [25] proposes a vehicle analysis system based on DeepSORT and YOLOv5 to improve real-time performance and accuracy of analyzation of traffic data as well as identifying a large number of small objects that is clustered together A dataset was created from a drone-captured video of high-speed road traffic flow, where 11,000 vehicle pictures were extracted. 150 of these pictures were chosen for the test set and the rest for the training set. The proposed system uses DeepSORT in detecting vehicles, a Kalman filter is then used to predict and update locations, and the Hungarian algorithm for matching trajectories. The system also uses inverse perspective mapping in calculating the vehicle speed. Experiments show the system has 96% accuracy for detecting vehicles, 98% accuracy for speed measurement, and an average model loss of 2%. This meets the need for real-time analysis.

This paper [26] presents a study on a small target vehicle detection and occlusion vehicle tracking. A new vehicle detection model which was (YOLOv5-NAM) developed based on the classical YOLOv5s model. A real-time small target vehicle tracking method (JDE-YN) was also proposed. The dataset used in this study is UA-DETRAC which is an open-source dataset for multi-vehicle detection models. Specifically, the dataset consisted of four classes: cars, buses, vans, and other vehicles. These classes were included in four different conditions: sunny, rainy, cloudy, and night in order to expand on the dataset. A total of 8250 vehicles were marked. the new method improved the mean average precision (mAP) by 1.6% compared to the YOLOv5s model, improved multiple object-tracking accuracy (MOTA) by 0.9% compared to JDE algorithm, and decreased the identity switching times of vehicles by 15%. The proposed method is effective in detecting small target vehicles, tracking multi-vehicles in real-time and efficient, and has a certain promotion effect in the field of vehicle detection and tracking.

This paper proposes an enhanced vehicle detection algorithm using YOLOv5s [27] in order to address issues with regards to false and missed detections due to the complex scenes and depending on the varying target sizes. The dataset used in this paper includes 10 hours of traffic footage from the Beijing and Tianjin zones of China. The footage was taken at a resolution of 960x540 pixels, and over 140,000 frames were manually annotated and classified into four categories: cars, vans, buses, and other types of vehicles. Additionally, the weather conditions during the footage were also classified into four categories: sunny, rainy, cloudy, and night. The algorithm adds a detection layer and replaces the Spatial Pyramid Pooling (SPP) module with the

Atrous Spatial Pyramid Pooling (ASPP) module to increase the receptive field of images of different sizes and extract multi-scale context information. The improved algorithm outperformed the original YOLOv5s algorithm, achieving a higher precision, recall and mean average precision. Specifically, the improved algorithm achieved 93.7%, 94.2% and 93.9% in precision, recall and mean average precision at 0.5, respectively, which are 0.8%, 1.9% and 2.3% higher than the original algorithm.

This article [28] presents an improved real-time object detection algorithm for autonomous vehicles based on YOLOv5s. The dataset used in this study is referred to as BDD100K. This dataset consists of 100,000 images of driving scenes set to a resolution of 1280x720. These images are all labeled except for the testing set of 20,000 images. The improved YOLOv5s utilizing a Non-Maximum Suppression trained from the 70/20/10 split of training, validating and testing BDD100K dataset. The algorithm uses shallow high-resolution features and changes the output feature map size to significantly improve detection of small objects. The algorithm's mean Average Precision on the BDD100K dataset increased by 3.2 percentage points and the average detection speed is 74.6 FPS, making it a suitable solution for various complex road scenes.

In this paper [29], a proposed YOLOv5 model is proposed in order to improve the performance of vehicle-pedestrian detection through adjusting the sizes of anchor boxes. The model uses an attention module to effectively collect discriminative features from dense contextual information, which improves the detection of distant objects. The YOLOv5 model was trained using a KITTI dataset which is often used for autonomous driving detection related studies and applications. The anchor box adjustment makes the model more familiar with the dataset. The results of experiments on the KITTI benchmark show that this model outperforms other state-of-the-art methods and is capable of running in real-time. The proposed model is useful for the vehicle and pedestrian detection task of autonomous driving environmental perception.

This paper [30] presents a vehicle detection technique that utilizes an enhanced version of Yolov5s. The method replaces the activation function in the Yolov5s backbone network with the SiLU activation function, which reduces the number of parameters and addresses the problem of gradient disappearance. The model is trained and evaluated on the KITTI dataset, which includes 3000 images vehicles. Additionally, the method employs a CBAM (Convolutional Block Attention Module) to enhance the feature extraction of small and medium-sized objects. The results indicate that this improved model outperforms the original Yolov5s model, as it is smaller, increases map 0.5 by 4.9% to 92.3%, and increases precision by 2.6% to 94.5%, making it more efficient and accurate in detecting vehicles on the road.

SSIM Image Augmentation Testing

In this study [37] researchers have investigated the critical domain of Artificial Intelligence-based Automated Optical Inspection using Convolutional Neural Networks for defect in detection in modern manufacturing with a specific focus on metal defect detection. Researchers enhance the performance of CNN-based AOI for metal defect detection, an approach by leveraging a generative AI technique for their data augmentation. The study assesses the performance of ten CNN models then comparing their efficacy when trained with data before and after augmentation by WGAN.

Synthesis

Studies have explored the impact of connected and automated vehicles (CAVs) on traffic operations and safety at intersections using microscopic traffic simulation. These models indicate

CAVs can enhance traffic flow and reduce delays, especially at lower traffic volumes [1]. Increasing CAV penetration rates may also significantly improve traffic movement and road capacity [2]. Optimization models for mixed traffic environments with CAVs can minimize average vehicle delay, especially in undersaturated conditions with prominent conflicting flows [3]. Adaptive traffic signal control methods like fuzzy logic and bio-inspired algorithms have also been proposed to optimize intersection traffic flow and reduce congestion [4-5].

Researchers have focused on optimizing traffic signals and implementing priority systems specifically for emergency vehicles to improve response times. Proposed priority systems include signal light priority control, timing recovery, and cycle recovery to give emergency vehicles priority intersection access [6]. Automated systems using sensors, software and cloud-based control have been suggested to reduce response time by optimizing signals [7-8]. Algorithms to determine emergency vehicle priority level have also been developed [9]. Reinforcement learning approaches for intersection signal control in the presence of different priority vehicle types have also been studied [10].

In terms of emergency vehicle detection, methods using sensors, real-time handling, and optimizing traffic lights have been proposed to improve dispatch in smart cities [11]. Dynamic algorithms that match vehicle availability to emergency type and location have also been suggested [12]. Audio-visual systems combining image and sound data have been designed to detect approaching emergency vehicles [13]. Technologies like sound recognition algorithms and shortest job first scheduling have been implemented to identify sirens and prioritize emergency vehicles [14].

Implementing emergency vehicle signal preemption has demonstrated potential for reducing intersection delays, with studies finding it can decrease delays by 25-30% [15]. Review papers have categorized and compared different traffic control strategies and preemption systems for emergency vehicles [16-17]. Open IoT systems integrating emergency vehicles, cameras, and traffic lights allow real-time signal adjustments to reduce response times [17]. Deep learning vehicle detection models like YOLOv5 can enable accurate real-time monitoring to improve traffic management [18-25]. Enhancements to YOLOv5 like attention mechanisms, optimized anchor boxes, and improved feature extraction have increased model accuracy and efficiency for vehicle detection and tracking [26-30].

Table 2. 1. Literature Review Summary Table.

Reference	Paper Title	Authors	Methodology	Process	The Algorithm / Classifier / Framework used

[1]	**A High-Precision Fast Smoky Vehicle Detection Method Based on Improved Yolov5 Network**	Wang Chengpeng, Wang Huanqin, Yu Fajun, Xia Wangjin	Involves collection of aerial surveillance data from AU-AIR and employing Deep Neural Networks for object tracking.	Enhance the real-time object tracking from Unmanned Aerial Vehicles by modifying the YOLOv5 and evaluating its overall performance.	Deep Neural Network, Rectified Linear Unit, AU-AIR
[2]	**Deep Neural Network Based Multi-Object Detection for Real-time Aerial Surveillance**	Dey Rebanta, Pandit Binit Kumar, Ganguly Anirban	Used an improved lightweight network that is based on YOLOv5 for vehicle detection, data augmentation techniques were used such as cutout and saturation transformation were used to improve the models performance.	Developed a hig-precision and fast smoky vehicle detection using the improved lightweight network based on YOLOv5 with Mobilenetv3-small for a reduced parameters and calculations.	YOLOv5s, Mobilenetv3-small, TensorFlow

[3]	**Multi-objective real-time vehicle detection method based on yolov5**	Song Xuyang, Gu Wei	YOLOv5-based vehicle detection model is trained using the acquired vehicle image data and is applied to real-time vehicle detection in video streams or images.	Develop a multi-objective real-time vehicle detection method to accurately locate and classify vehicles in images or video streams. The focus is on reducing the number of false detections to improve accuracy.	YOLOv5, TensorFlow,
[4]	**Augmentation-Based Object Detection for Winter Time Applications**	Yagfarov Rauf, Ostankovich Vladislav, Gafurov Salimzhan	Used a Cascade R-CNN model in improving object detection in sever weather conditions specifically during winter	Implemented the Cascade R-CNN and data augmentation in improving the accuracy of detecting object in severe weather conditiond	Cascade R-CNN, TensorFlow
[5]	**Detection and Classification of Incoming Ambulance Vehicle using Artificial Intelligence Technology**	Sarapirom Tiramanat, Poochaya Settawit	Employing a faster R-CNN and implementation of AI-based system in classifying and detecting incoming ambulances.	A faster R-CNN model is trained on dataset gathered of ambulances in order to recognize and classify specified features.	Faster R-CNN, TensorFlow

[6]	**Research on the Model of Emergency Vehicles Passing Priority at Intersections Based on Vehicle-Road Cooperative System**	Qiusen Wang, Shenghong, Sanquiang Zheng	This takes double intersections as the research object to build a simulation to determine the emergency traffic control strategy. The data will be in a form of statistics which will be used for evaluation.	Will initiate a priority access request. This will be followed by a control signal light, restore traffic light cycle, then ends with the restoration of timing plan of traffic lights.	Special Vehicle priority traffic classifier, SUMO simulation
[7]	**Traffic Control System based on Density with Emergency Priority Mechanism**	Hitesh Agarwal, J. N. Rai	Used IR sensors and PLC for calculation of the number of vehicles on the road. Through vehicle density, traffic light is triggered to go or green.	This utilizes. the SCADA software to make demonstration. .	Emergency Priority Mechanism
[8]	**Algorithm for traffic management with priority for emergency vehicles**	E. Paunova-Hubenova, E. Trichkova-Kashamova	Used image processing, GPS tracking, and a cloud-based control center.	Start with making a model for the capacity of junctions and road links. Make an incident detection. Then, congestion monitoring.	Traffic Management Algorithm

[9]	**A Smart PriorityBased Traffic Control System for Emergency Vehicles**	Gour Karmakar, Joarder Kamruzzaman	Used IOT devices like RFID and sensors for gathering data.	Estimate type of incident occurred, collect information, determine priority level, controller makes a signal for priority.	SUMO simulation, EV Algorithm 1 and 2, EV Priority level
[10]	**Application of Deep Reinforceme nt Learning for Traffic Control of Road Intersection with Emergency Vehicles**	Giuseppe Benedetti, Maria Pia Fanti, Agostino Marcello Mangini, Fabio Parisi	Used Satellite images as dataset then through SUMO, simulations will be made to evaluate performance	Detected Vehicles will undergo Deep Reinforcement Learning Framework and its priority level will be identified. SUMO is also utilized for simulations.	Deep Reinforcement Learning Framework, Classes of Priority Vehicles, SUMO simulation
[11]	**An Optimization Approach for Emergency Vehicles Dispatching and Traffic Lights Adjustments in Response to Emergencies in Smart Cities**	E. O. Rangel, D. G. Costa, M. M. L. Peixoto	A sensor-based approach for emergency management in smart cities. It uses IoT devices, sensors, and sensor networks to monitor urban environments and detect abnormal situations.	Using IoT devices, sensors, and sensor networks to monitor urban environments and detect abnormal situations.	SUMO SImulation, OMNet++, Veins Simulation

[12]	**Automatic Assignment of Emergency Vehicles in Response to Sensors-based Generated Alarms in Smart City Scenarios**	D. G. Costa, F. Vasques, A. Aguiar, P. Portugal	System processes all emergencies, then computes a list of candidate vehicles based on mapping of emergency types to available emergency vehicles in the area	Dijkstra's algorithm is used to select the closest vehicles to each alarm	Dijkstra’s Algorithm
[13]	**Audio-Vision Emergency Vehicle Detection**	V. -T. Tran, W. -H. Tsai	System detects and classifies all vehicles in every image frame captured by a camera and outputs detection information to make a final decision on emergency vehicle detection.	A single-stage vision-based emergency vehicle detection system (SVEVD) which uses an end-to end object detector to detect and classify all vehicles in every image frame captured by a camera.	SV-EVD system, YOLO-EVD,WaveResN et, AV-EVD, VEVD, YOLOv3, YOLOv4
[14]	**Detection and Prioritization of Emergency Vehicles in Intelligent Traffic Management System**	K. Choudhury, D. Nandi	A method that prioritizes. emergency vehicles approaching traffic signals using the Shortest Job First scheduling algorithm.	Prioritizing emergency vehicles approaching traffic signals using the Shortest Job First scheduling algorithm.	SJF Algorithm, Goertzel Algorithm

[15]	**Emergency Vehicle Signal Pre-emption System for Heterogeneous Traffic Condition: A Case Study in Trivandrum City**	R. Anil, M. Satyakumar, A. Salim	Developed a simulation model in VISSIM and calibrated the network model by creating a network layout using VISSIM software, road geometry details and Open Street Map base map.	Creating a network layout for a study area used VISSIM software, where road links were created using road geometry details and an Open Street Map base map.	VISSIM model simulation
[16]	**State-of-theArt Review on Traffic Control Strategies for Emergency Vehicles**	W. Yu, W. Bai, W. Luan L. Qi	Categorizes route optimizatio n algorithms into traditional optimizatio n n algorithms, intelligent optimizatio n algorithms, machine learning based algorithms, and other specific algorithms.	Utilization of traffic signal preemption strategies for reducing delays for emergency vehicles.	Route Optimization Algorithm, preemption strategy, Intelligent Optimization algorithms, Machine learning-based Algorithms

[17]	**Leveraging Emergency Response System Using the Internet of Things. A Preliminary Approach**	L. A. Magre Colorado, J. Franco Ibañez, J. C. Martinez-Santos	IoT system that connects emergency vehicles with city's surveillance cameras and traffic lights to create a "green wave" and decrease emergency response times.	Integration Internet of Things system in the current infrastructure of the city specifically surveillance cameras and traffic lights	A* Algorithm, Dijkstra's Algorithm, Reinforcement Learning, Deep Convolutional Neural Networks, Markov Decision-making Models
[18]	**Vehicle Detection and Tracking using YOLO and DeepSORT**	M. A. Bin Zuraimi, F. H. Kamaru Zaman	Collecting and labeling images of vehicles for training the model, which can be simplified by using the OIDv4 toolkit to automatically label images in YOLOv4 format.	Utilizing OIDv4 toolkit for automation search and label images in YOLOv4 format	TensorFlow, DeepSORT Algorithm, YOLOv4, OIDv4 with Python

[19]	**Vehicle Identification and Traffic Statistics System in Traffic Intersection**	X. Guan, X. Sun	DeepSORT tracking algorithm for detection result correlation in video sequence.	Training the model in order to accurately detect and identify vehicles appropriately in video footage	DeepSORT algorithm, YOLOv3
[20]	**Vehicle Tracking and Speed Estimation from Roadside Lidar**	J. Zhang, W. Xiao, B. Coifman, J. P. Mills	Vehicle tracking method that uses image matching to enhance the accuracy of vehicle speed estimation.	Evaluating results detected by two different 3D lidar sensors specifically RoboSense RS-LiDAR-32 and Velodyne VLP-16 sensors	RoboSense, Velodyne, UKF, JPDAF, NNF, Euclidean Cluster Extraction Algorithm
[21]	**Road Traffic Vehicle Detection Method Using Lightweight YOLOv5 and Attention Mechanism**	Yunzhen Wang, Hongbing Ma, Liangliang Li,	Using the attention mechanism SENet in the CSP1 residual block of the Backbone component of YOLOv5. In order to lightweight the network using DSC placement into the PANet to reduce the number of	Placing the SENet in the CSP1 residual block of the Backbone component of the trained YOLOv5 model which would then be comparing the experimental data and test results to evaluate the model's performance.	YOLOv5, SENet, DSC, CSP1, PRelu, PANet

			parameters and improve performanc e.		
[22]	**A High-Precision Fast Smoky Vehicle Detection Method Based on Improved Yolov5 Network**	Chengpeng Wang, Fajun Yu, Huanqin Wang, Wangjin Xia	Using a self-built vehicle exhaust dataset in order to develop a lightweight YOLOv5 – MobileNet model.	Applying data augmentation to the self-built dataset created more instances within the dataset to improve the overall training of the smoke detection model. The trained model is tested and found to have an 8.6% increase in detection accuracy compared to the original model.	YOLOv5, YOLOv5s, MobileNetV3

[23]	**Vehicle Detection for Vision-Based Intelligent Transportation Systems Using Convolutional Neural Network Algorithm**	Othman O. Khalifa , Muhammad H. Wajdi, Rashid A. Saeed, Aisha H. A. Hashim, Muhammed Z. Ahmed, and Elmustafa Sayed Ali	Evaluating different vehicle detection methods based on their respective strengths and weaknesses. while also implementing a proposed vehicle detection through the YOLOv5s architecture, which was coupled with k-means to optimize the anchor boxes.	Took a look at various vehicle detection techniques that use the CNN architecture, and evaluated their strengths and weaknesses; then proposed a new method utilizing the YOLOv5 coupled with kmeans as an optimization method, and tested it against other research in the field as well as a standard YOLOv5s model, to see how well it performed.	YOLOv5, YOLOv5s K-Means
[24]	**Vehicle Tracking Method based on AttentionYOLOv5 and Optimized DeepSort Models**	Zhuang Li, Xincheng Tian, Yan Liu, Xiaorui Shi	Used a BDD dataset to create a vehicle detection and tracking model using YOLOv5 with attention mechanism and optimized DeepSort to improve accuracy, reduce IDSwitch and enhance anti interference	The vehicle detection and tracking model was trained and optimized by adding CBAM and ECA attention mechanisms to YOLOVS, training it on BDD dataset using transfer learning, optimizing the DeepSort algorithm, improving recognition network	YOLOv5, ECA-YOLOv5, Transfer Learning, DeepSORT

			ability for traffic management and autonomous driving.	structure and optimizing the VeRi reidentification dataset, and performing reidentification pre-training on the ImprovedVeRi dataset, resulting in significantly improved tracking performance compared to the benchmark algorithm.	
[25]	**Vehicle Analysis System Based on DeepSORT and YOLOv5**	Fuheng Guo, Yi Xu	YOLOv5 model was paired with a DeepSORT object tracking algorithm in order to optimize the tracking performance in congested roads for vehicle detection.	By means of inverse perspective mapping, the position of images is converted into latitude and longitude coordinates to detect vehicle speed, and a higher refresh frame rate is achieved for the processed video.	YOLOv5, DeepSORT
[26]	**A High-Precision Vehicle Detection and Tracking Method Based on the Attention Mechanism**	Jiandong Wang, Yahui Dong, Shuangrui Zhao, Zhiwei Zhang	Used a COCO Dataset and UADETRAC dataset in order to create a vehicle detection YOLOv5-NAM model.	A YOLOv5-NAM model was used as the vehicle detector, and a real-time small target vehicle tracking method (JDEYN) was implemented into the framework. The	YOLOv5s, JDE-YN, DeepSORT, SORT

				experiment results on the UA-DETRAC and COCO datasets demonstrated that the mAP value increased.	
[27]	**Vehicle Detection in Traffic Monitoring Scenes Based on Improved YOLOVS5s**	Liu Xiaomeng Feng Jun Chen Peng	An improved vehicle detection algorithm is proposed by using a UA-DETRAC dataset, to improve the accuracy of vehicle detection in traffic surveillance video by adding a small target detection layer.	The YOLOv5s vehicle detection model was able to use a small target detection layer to deepen the network depth and extract feature information from the deeper network and using the ASPP module to replace the SPP module, to increase the receptive field and obtain multi-scale context information.	YOLOv5s ASPP SPP

[28]	**Real-Time Object Detection Algorithm of Autonomous Vehicles Based on Improved YOLOv5s**	Baoping Xiao, Jinghua Guo, Zhifei He	Used the BDD100K dataset to make a YOLOv5s model for detecting vehicles for autonomous vehicles.	The training process involved adding shallow high-resolution features and adjusting the size of the output feature map to improve the detection ability of small objects.	YOLOv5s, CSP1 and CSP2
[29]	**Multiobjective real time vehicle detection method based on yolov5**	Xuyang Song Wei Gu	Used a KITTI dataset to create a vehicle and pedestrian detection model. YOLOv5 was the model used.	By adding an attention module to extract discriminative features from the full image using rich contextual information and adjusting anchor boxes to make the model more familiar with the car and pedestrian categories.	YOLOv5

[30]	**Research on Road Vehicle Detection based on improved Yolov5s**	Lei Shao, Zhenqiang Fan 1, Ji Li, Hongli Liu	The KITTI dataset was used to help create the vehicle detection model using YOLOv5s	By replacing the original activation function in Yolov5s backbone network with SiLU activation function to reduce the number of parameters in the network structure and solve the problem of gradient disappearance.	YOLOv5s, SiLU, CBAM
[31]	**An Intelligent Road Traffic Information System using Text Analysis in the Most Congested Road in Metro Manila**	Erika Bondoc, Francis Caparas, John Eddie Macias, Vileser Naculangga, Jheanel Estrada	Developed a real time road traffic information system using Twitter data from MMDA account and applying data preprocessing, extraction with LDA, and KNN classifier	Collected data through Twitter account of MMDA containing “TRAFFIC UPDATE”, tokenized tweets, and feature extraction using LDA, Data Forecasting, Data Classification, and Pattern Recognition	Twitter API, KNN, Naïve Bayes, Decision Tree, Random Forest, LDATwitter API, KNN, Naïve Bayes, Decision Tree, Random Forest, LDATwitter API, KNN, Naïve Bayes, Decision Tree, Random Forest, LDA

[32]	**Object Detection in 20 Years: A Survey**	Zhengxia Zou, Keyan Chen, Zhenwei Shi	Acceleration on object detection through optimizing detection pipeline, use of efficient detector backbones, and numerical acceleration techniques like integral images and vector quantization.	Implementation and evaluation of methods : lightweight network architecture, integral images, and vector quantization	Neural Architecture Search, Cascaded Detection, Network Pruning and Quantification
[33]	**Analysis of Live Video Object Detection using YOLOv5 and YOLOv7**	Thammi Konala, Anusha Nammi, Divya Sree Tella	Evaluation on YOLOv7 with the incorporation of instance segmentation through YOLOv7-mask, YOLOv7-pose using coarse losses for auxiliary head and fine losses for lead head.	Evaluation and comparison of YOLOv5 and YOLOv7 by training on COCO Dataset while analyzing the advantages and limitations of each model.	COCO Dataset, YOLOv5, YOLOv7

[34]	**Design and Development of an IoT-based Smart Ambulance System with Patient Monitoring**	Larry Valdez, Francis Earl Beran, Chester Azman, Angelino Pimentel, Renann Baldovino	Use of NodeMCU in connecting sensors on an ambulance in transmitting patient data to hospital.	Built website using PHP and web hosting in receiving and displaying transmitted data of NodeMCU, Arduino. Evaluation of performance metric with the use of statistical analysis (t-test) in comparing data	NodeMCU, Arduino, PHP, T-Test
[35]	**Efficient Dynamic Traffic Control System using Wireless Sensor Networks**	Bharadwaj, R., Deepak, J., Baranitharan, M., Vaidehi, V.	Development of dynamic traffic control system using proximity sensors in counting vehicles and RFID tags in identifying emergency vehicles	Designing of system architecture with traffic monitor units at intersections and a centralized traffic control server with the integration of proximity sensors, and RFID tags for emergency vehicles for identification	RFID Tags, IoT, Arduino

[36]	**Machine Vision System of Emergency Vehicle Detection System Using Deep Transfer Learning**	Kim Carol Maligalig, Albertson Amante, Ryan Tejada	LabelImg was utilized as an annotation tool, YOLOv3 was the main technique used in developing the detection system.	The mAP of each model is compared and evaluated with the integration of PyQt5 and ImageAI detection.	YOLO-EVD, WaveResNet, YOLOv3, RCNN
[37]	**GAN-based Data Augmentation for Metal Surface Defect Detection using Convolutional Neural Networks**	Tseng Ling-Shen, Wu Chih-Hung, Chen Yi Han, Tsai Chuing-Hui	Assesses the performance of ten CNN models then comparing their efficacy when trained with data before and after augmentation by WGAN	Employing Structural Similarity Index Measure to assess the degree of similarity between generated by WGAN and the original dataset	AI-AOI, CNN, SSIM, WGAN

Chapter 3

METHODOLOGY

This section discusses the steps needed to detect and classify ambulances in the Philippines using the proposed framework. As there is a lack of research on specific ambulance detection models, the researchers aim to expand current YOLO object detection models using a localized dataset of Philippines ambulances. As noted previously, Philippine ambulances differ from those in other countries, such as the USA and UK, in terms of their chassis, features, and coloring. In order to collect the dataset, parameters such as ambulance type, siren light, and label lettering will be considered.

Machine Specification

This section provides a comprehensive overview of the various machines and specifications employed throughout the research process. In the initial stages of training and validating the YOLOv5 model, the researchers made use of Google Colab. The researchers relied on Python notebooks as their programming environment, seamlessly integrating them with Google Colab's resources. To evaluate the YOLOv5 model for ambulance detection in practical settings, a local workstation setup was used, relying on Visual Studio Code and Python notebooks for programming. This change enabled the researchers to work with locally stored video data.

Table 3. 1. Machine Specifications.

Category	Value
Processor	AMD Ryzen 5800H @ 3.20GHz
Graphics Card	NVIDIA GeForce RTX 3060
RAM	16 GB
Operating System	Windows 10 (64-bit)

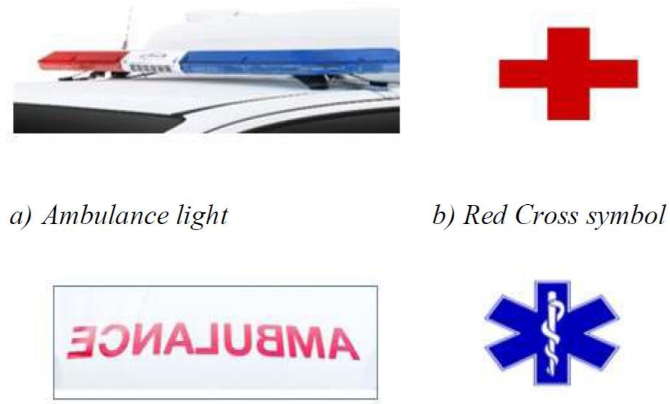
Theoretical Framework

This study is based on the theoretical framework, "Detection and Classification of Incoming Ambulance Vehicle using Artificial Intelligence Technology," that combines artificial intelligence (AI), computer vision, and optimizing emergency medical services (EMS) [5]. The primary issue tackled is the prompt arrival of emergency ambulances in situations of traffic congestion, with a focus on Thailand. The goal of the study is to enhance emergency medical response systems by utilizing advanced technologies, specifically by implementing the Faster R-CNN (Region-based Convolutional Neural Network) model in the TensorFlow framework.

The theoretical foundation utilizes Faster R-CNN as a deep learning model designed for object detection. TensorFlow, serving as a neural network library, forms the computational backbone for the implementation. The model undergoes training to identify distinct features associated with emergency ambulances, encompassing Text Ambulance, ambulance lights, the Red Cross symbol, and the Star of Life symbol.

The study incorporates various image pre-processing techniques to enhance the quality of input data. A dataset consisting of 289 ambulance images is divided into training (94 images) and testing (195 images) sets.

For labeling, a Python tool designed for image labeling is employed to mark regions of interest in each image. This process involves the identification and labeling of features such as Text Ambulance, ambulance lights, the Red Cross symbol, and the Star of Life symbol as seen in **Figure 3.1**. Subsequently, the labeled images are utilized to generate TFrecords, which function as inputs for the TensorFlow training model. These TFrecords encapsulate the essential information required to effectively train the Faster R-CNN model.



a) Ambulance light *b) Red Cross symbol*

c) Text Ambulance *d) Star of life symbol*

Figure 3. 1. Ambulance Features for Classification

The Labelmap.pbtxt file establishes a mapping between class names (e.g., Text Ambulance, ambulance lights) and unique IDs. This file also defines the format and parameters for training the model.

The Faster R-CNN model is structured into two networks: the Region Proposal Network (RPN) and the Object Detection Network (ODN). The RPN extracts areas of interest, and the ODN evaluates bounding boxes, utilizing features from convolutional feature maps.

The function of the Region Proposal Network (RPN) is comparable to extracting areas of interest from the feature map, represented numerically. Anchors assess proposals, producing outputs that indicate predicted bounding boxes. The Intersection over Union (IoU) metric plays a crucial role as an evaluator, measuring object detection accuracy by evaluating the overlap between ground-truth and predicted bounding boxes. To handle regions of varying sizes, the model utilizes Region of Interest (ROI) pooling. The final layer R-CNN takes charge of the classification phase.

The model's journey encompasses a training configuration spanning 500,000 steps. The total loss, serving as a metric for the model's performance, smoothly decreases to 0.0005062 at step 500,000. Thorough evaluation unfolds with 100 test images for each feature.

The detailed process and methodology employed for the training and classification of ambulances, along with the identification of key ambulance features, are comprehensively outlined in **Figure 3.2** as interpreted in the theoretical framework.

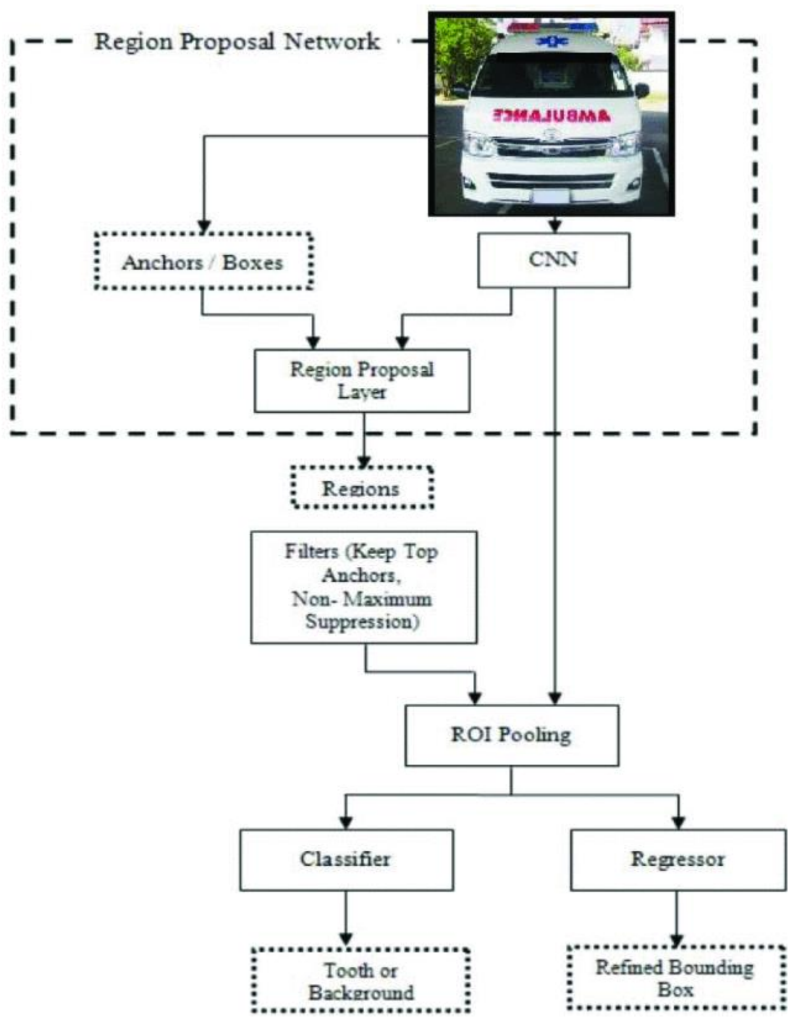


Figure 3. 2. Theoretical Framework [5].

Conceptual Framework

In the context of this study, the conceptual framework centers on the development of a YOLOv5-based model for ambulance detection in the Philippines setting. The dataset comprises localized images of Philippine ambulances used to train the YOLOv5 small model, followed by validation and testing phases. Figure 3.3 describes the proposed methods of developing an ambulance detection model using the proposed model. Subsequently, the conceptual framework encompasses critical stages such as data acquisition, image preprocessing, data classification, data augmentation, data augmentation testing, model training, hyperparameter tuning, validation and test performance. The proposed conceptual framework aims to advance the practicality of ambulance detection from a computer vision approach. **Figure 3.3** further illustrates the process flow for the conceptual framework of the study.

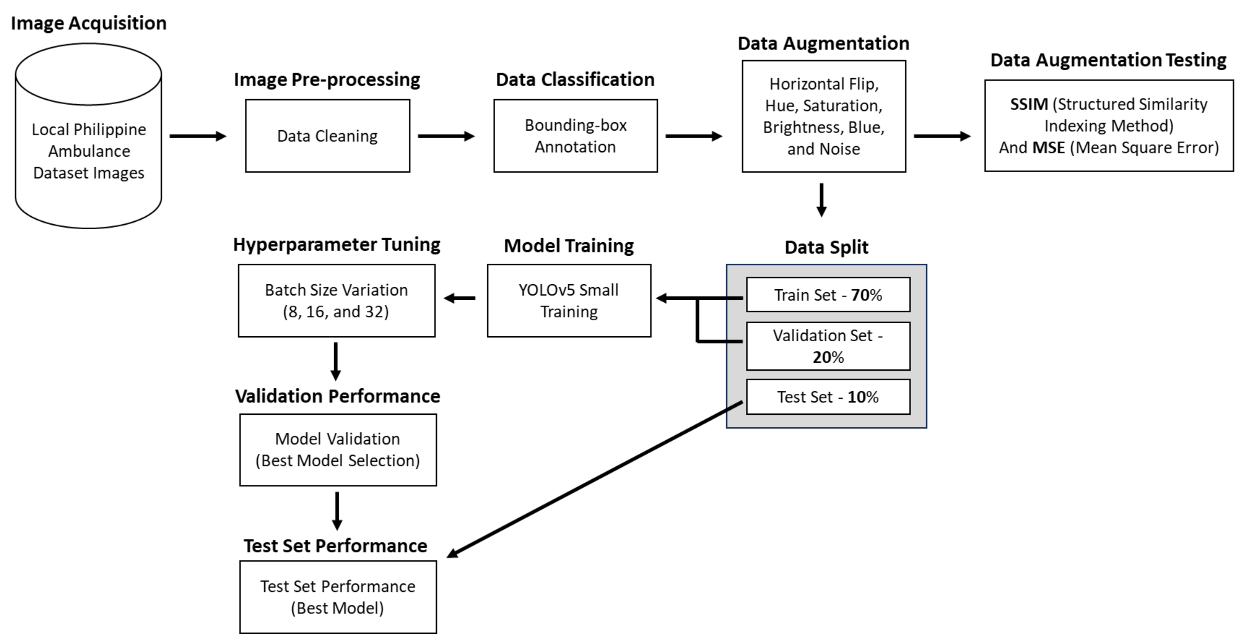


Figure 3. 3. Conceptual Framework.

Data Acquisition

The image acquisition phase followed a systematic approach called image scraping to gather a comprehensive dataset of ambulance images. Image scraping involves extracting and storing images from online sources, a task efficiently carried out in this study through a Python script utilizing the Selenium web automation library. The researchers developed a script tailored to interact with the Google Chrome browser, initiating a search for "Philippine Ambulance" on Google Images. After loading the search results page, the script prompted the user to scroll down until a satisfactory number of image containers were loaded. Upon user confirmation, the script automated the process, systematically opening each image container and downloading the images within. This method ensured the systematic and efficient compilation of a diverse and representative dataset for subsequent stages of analysis and model training. A study in vehicle detection using YOLOv5 proposed surpassing 5000 images after augmentation [1]. In alignment with this recommendation, the researchers collected 1840 clean images featuring Philippine ambulances along with common vehicles such as cars and vans. The count of 1840 clean images from the dataset will expand times 3 to 5520 images after data augmentation. The classification of vehicles will be further explained in detail in the Dataset Overview segment of the study. This image acquisition process can be found in Figure 3.3.

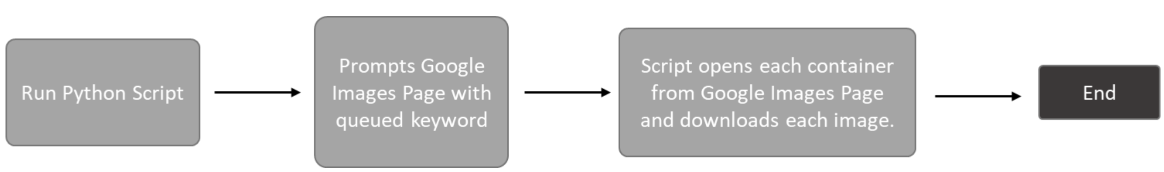


Figure 3. 4. Data Acquisition Process (Image Scrapping).

Dataset

The dataset primarily consists of images of ambulances exclusively from the Philippines. These ambulances classifications are representative of the Type I, Type II, and Type III ambulances commonly found in the country. They are equipped with distinctive sirens and feature prominent labels denoting "Ambulance." This specific focus on Type I, Type II, and Type III ambulances in the Philippines distinguishes the dataset and aligns it with the localized context of the research. According to the Department of Health or DOH in the Philippines, there are 3 types of standardized ambulances. These are the classifications of ambulances in the Philippines.

Type I Ambulance: Type I ambulances are essentially vehicles constructed from small trucks. The patient compartment, typically box-shaped, is mounted onto the chassis of the truck, with bodies commonly derived from pickup trucks and light-duty trucks. These ambulances are capable of providing Basic Life Support (BLS).

Figure 3. 5. Type I Ambulance Classification

Type II Ambulance: Type II ambulances are predominantly crafted from heavy-duty van units, undergoing minimal modification except for the incorporation of standard ambulance features like emergency vehicle lighting, sirens, and the patient compartment. Type II Ambulances are capable of providing Advance Life Support (ALS).

Figure 3. 6. Type II Ambulance Classification

Type III Ambulance: Type III ambulances share similarities with Type I units, with the distinction that in Type III ambulances, the box-shaped patient compartment is affixed to the chassis of a heavy-duty van rather than a truck.

Figure 3. 7. Type III Ambulance Classification

Each ambulance type must include the following key features in order to be added into the dataset (Siren and Label).

Siren: The dataset includes images of ambulances fitted with sirens. This feature is crucial for the model to learn the distinctive audio and visual cues associated with emergency vehicles.

Label: The ambulances in the dataset are equipped with labels that prominently display the word "Ambulance." These labels are positioned on the front, sides, and rear of the vehicles, making them easily identifiable by both humans and the model.

To prevent bias when training the dataset, the images collected all have equal distribution across all the vehicle types. Each vehicle type had a total of 368 images as seen in **Table 3.2**. Balancing the types of vehicles within the dataset would promote better generalization for the model and foster a more balanced learning environment.

Table 3. 2. Image Count for Vehicle Type

Vehicle Type	Image Count
Type I Ambulance	368
Type II Ambulance	368
Type III Ambulance	368
Common Cars	368
Common Vans	368
Total:	1,840 Images

The dataset comprises a total of 1,840 images of ambulances, common cars and vans, making it a substantial and diverse collection that allows for robust model training and evaluation. By creating and utilizing this localized dataset of ambulances from the Philippines, the researchers aim to develop a YOLOv5s model specifically tailored to detect and classify ambulance types in the unique context of the Philippines, considering their distinct visual characteristics and features.

By incorporating these additional classes, the YOLOv5s model can learn to distinguish between genuine ambulances and other vehicles that may share visual similarities. This strategy enhances the model's precision and reliability when deployed for real-world ambulance detection in the Philippines, reducing the likelihood of incorrectly identifying non-ambulance vehicles as ambulances.

Data Pre-Processing and Data Augmentation

The collected dataset will first be cleaned in order for the images to be suitable for further processing. This process will require the researchers to manually remove images that are considered to be ineffectual to the model training process. Images from the collected Philippine Ambulance dataset that are blurry, out of frame, or has the subject obstructed will be deleted manually.

Once the data cleaning process is finished, data labeling will be conducted. Data labeling involves adding labels to each individual training image. The researchers will use the labeling and annotation feature of Roboflow, a popular online platform used for image annotation. This annotation platform allows users to select an image and label specific features or subjects by drawing bounding boxes around them. This manual and supervised process will be done by the researchers, who will label Philippine ambulance types and their features, such as the Ambulance-like chassis, sirens, and the label "AMBULANCE" on the front, sides or rear of the vehicle. Once all images from the dataset have been labeled, the Roboflow annotation platform will export text files that will enable the machine learning classifier to identify where in the image certain subjects or features are located. Figure 3.8 displays the annotation process in Roboflow; labeling the type II ambulance and its specific features such as its siren and text.

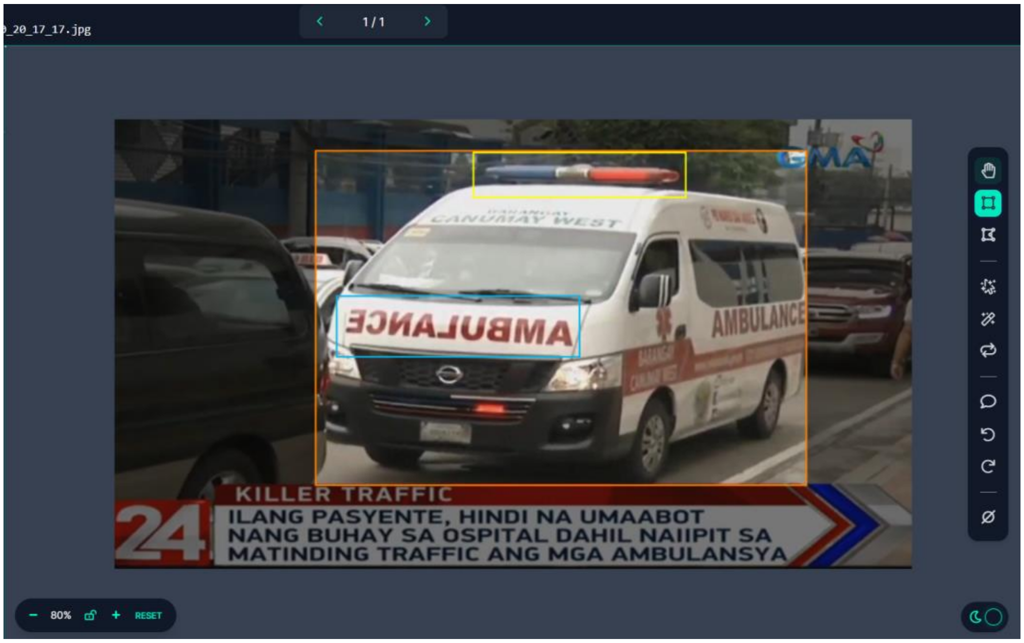


Figure 3. 8. Roboflow Manual Data Annotation of a Type II Ambulance.

The researchers applied data augmentation to the dataset in order to increase its size artificially. Data augmentation is often used to improve the overall performance of machine learning models such as YOLOv5. Several data augmentation techniques were applied to the ambulance dataset. A study on vehicle detection prompted the use of the following augmentation techniques on a custom dataset in order to improve the performance of their model [4]. The applied data augmentation techniques play a crucial role in strengthening the robustness and diversity of the dataset, a key factor for effective model training. Each augmentation technique contributes uniquely to enhancing the model's ability to generalize and make accurate predictions across various real-world scenarios [4]. The following data augmentation techniques were deliberately selected and implemented:

Outputs per Training Example: Setting three outputs per training example significantly enriches the dataset. This involves generating multiple variations of each image during training, providing the model with diverse perspectives, and aiding in capturing a broader range of visual features.

Horizontal Flipping: This involves mirroring the image along its vertical axis. This augmentation helps the model become invariant to left-right orientation, ensuring it can effectively recognize and classify ambulances regardless of their directional orientation.

Hue: Between -50° and +50°: Altering the hue within a range of -50° to +50° introduces variations in color tones. This augmentation simulates different lighting conditions, preparing the model to recognize ambulances in diverse environments with varying hues.

Saturation: Between -50% and +50%: Modifying saturation levels within a range of -50% to +50% adds variability to the color intensity of the images. This enables the model to adapt to scenarios with differing levels of color saturation, contributing to its resilience in various lighting conditions.

Brightness: Between -40% and +40%: Adjusting brightness levels between -40% and +40% simulates varying degrees of illumination. This augmentation prepares the model to detect ambulances under different lighting intensities, ensuring its performance in both well-lit and low-light conditions.

Blur: Up to 4px: Introducing blur to images, with a maximum deviation of up to 4 pixels, mimics the effect of motion or out-of-focus scenarios. This augmentation aids the model in recognizing ambulances in dynamic situations or instances with less clear visual cues.

Noise: Up to 10% of Pixels: Adding noise to a portion of the image, up to 10% of pixels, enhances the model's resilience to visual clutter and imperfections. This augmentation simulates real-world scenarios where images may contain pixel-level noise, ensuring the model's adaptability to such conditions.

Data Augmentation SSIM Testing

Augmentation techniques play a vital role in introducing diversity into the dataset, ensuring that the model encounters a broad spectrum of scenarios and variations. This ultimately enhances the model's capacity to generalize. However, images that are too similar to the original dataset may not contribute significantly to the model's learning. In this context, the incorporation of Structural Similarity Index (SSIM) testing becomes paramount. A study on metal defect detection uses SSIM as the filter to control the diversity of metal images generated from WGAN-GP. This method of comparing images and setting a numerical value to it would be beneficial in testing the generated augmented images.

SSIM is a robust metric quantifying image similarity with scores ranging from 0 to 1, where 1 indicates perfect similarity. The formula intricately evaluates means (μx, μy), standard deviations (σx, σy), and covariance (σxy) of pixel values in two images. Constants (c1, c2) prevent division by zero, set to small values for robustness across diverse datasets (e.g., c1 = $(0.01 \cdot L)^2$, c2 = $(0.03 \cdot L)^2$, where L represents the dynamic range of pixel values). The SSIM score is computed as follows:

$$\mathrm{SSIM}(\mathrm{x}, \mathrm{y}) = \frac{(2 \cdot \mu_x \cdot \mu_y + c_1) \cdot (2 \cdot \sigma_{xy} + c_2)}{\left(\mu_x^2 + \mu_y^2 + c_1\right) \cdot \left(\sigma_x^2 + \sigma_y^2 + c_2\right)}$$

Since the ambulance dataset has been augmented by setting three outputs per image, the SSIM testing process will have to compare images in triads. This means that images in three will be compared for their average SSIM scores. Once all the images from the augmented dataset have been tested for their SSIM scores, the average SSIM score of the dataset would be evaluated. A value closest to 0 would be preferred in this scenario in order to prevent images similar to one another being part of the dataset.

Data Split

After the dataset underwent meticulous cleaning and augmentation to ensure its quality and diversity, the next pivotal step in the model development pipeline was the strategic division of the dataset into three distinct subsets: training, validation, and testing. To achieve this, the dataset was partitioned in a well-balanced manner, adhering to a widely accepted and effective split ratio. Specifically, it was allocated as follows: 70% of the data was designated for the training set, 20% for the validation set, and the remaining 10% for the testing set.

YOLOv5s Model Training

The proposed model for this study on Philippine ambulance detection is YOLOv5s. This model is based on a convolutional neural network (CNN) algorithm that allows for real-time object detection from images and videos. YOLOv5 uses the PyTorch framework. This model was chosen for this study because of its ability to detect small objects. The study aims to expand on existing YOLO detection frameworks by testing it on a new dataset and a newer model. The conceptual framework for this study will the similar process but with a different model in detecting ambulances from the base paper [5].

YOLOv5 models comes in small, medium, and large weights, each with distinct complexities. For real-time vehicle detection, YOLOv5 small is optimal, offering a balance of accuracy and speed with the smallest size. Its fewer parameters make it faster and more efficient for real-world deployment, providing the responsiveness needed for time-critical vehicle detection applications, such as traffic monitoring systems. In terms of performance, size, and latency, YOLOv5 small stands out as the most suitable choice among its variations for vehicle detection tasks. The YOLOv5s model training process involves a systematic and multi-faceted approach, leveraging advanced neural network architecture and optimization techniques [3]. The YOLOv5s model will be trained using the following parameters followed from another study which performed hyperparameter tuning [2]:

Table 3. 3. YOLOv5s Hyperparameters

Hyperparameter	Value
Epoch Count	150
Batch Size	8, 16, and 32
Learning Rate	0.001

YOLOv5s Structure

YOLOv5, short for You Only Look Once version 5, leads in computer vision and object detection. With a robust architecture including CSPDarknet53, PANet, and an efficient detection head, YOLOv5s excels in real-time object identification. Tailored for discerning recognition, it incorporates Stochastic Gradient Descent (SGD) optimization and strategic Non-Maximum Suppression (NMS), positioning it as a sophisticated tool. YOLOv5s promises precision and operational efficiency, reshaping object detection with a keen focus on critical entities like ambulances. The following figure illustrates the flow of the model's training process.

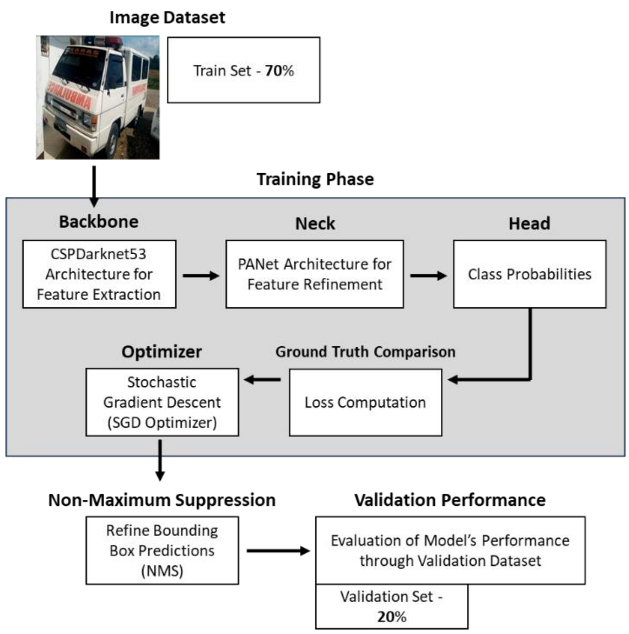


Figure 3. 9. YOLOv5 Training Process

Backbone: The structure of the YOLOv5s model is comprised of the backbone, neck, and head. The model's backbone is composed of CSPDarknet53, a powerful convolutional neural network designed for feature extraction. The ambulance image training set will pass through the backbone for its initial feature extraction [3].

Neck: As the features from the ambulance dataset progress through the neck, PANet refines and integrates these features, facilitating the model's comprehension of spatial relationships within the images. The neck plays a crucial role in augmenting the model's ability to recognize intricate patterns and contextual information, enabling more informed predictions during training. It acts as a bridge between the foundational features extracted by the backbone and the nuanced insights necessary for accurate object detection [3].

Head: The refined features from the neck are input into the detection head. At this stage, the model, having assimilated hierarchical and spatial features, interprets these representations to make predictions tailored to the characteristics of the image dataset. In the context of an ambulance detection task, the head discerns the presence of ambulances, classifies their types, and identifies key features such as sirens and labeling stickers. The head synthesizes the abstracted information into actionable insights, constituting the output of the model [3].

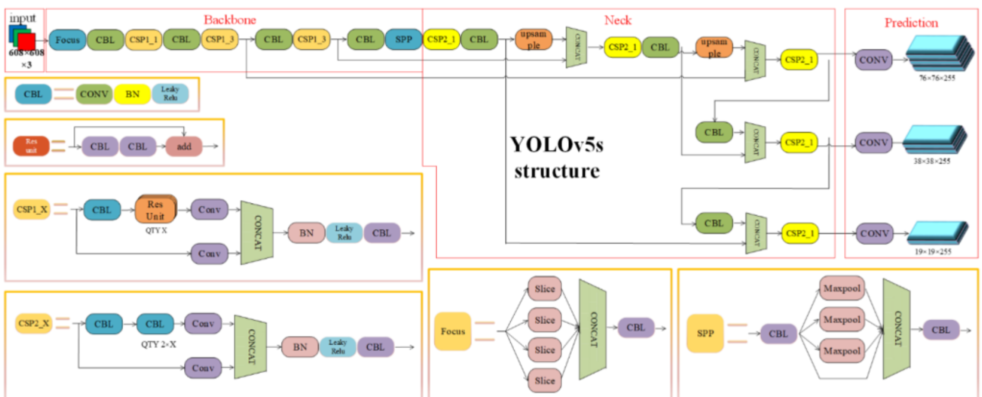


Figure 3. 10. YOLOv5s Model Network Structure.

Ground Truth Comparison: After the ambulance dataset has passed through the backbone, neck, and head, it would go through ground truth comparison and loss computation in order to compare the model's predictions with the ground truth labels which were the annotated information. The loss would be computed which represents the discrepancy between predicted and actual values (classification, localization, and confidence loss).

SGD Optimizer: As the ambulance training dataset traverses through the backbone, neck, head and ground truth comparison, the Stochastic Gradient Descent (SGD) optimizer comes into play during the training phase. It fine-tunes the weights and biases of the model based on the calculated loss. The optimizer's settings, including learning rate, momentum, and weight decay, collectively influence the adjustments made to the model's parameters. The learning rate determines the step size in parameter updates, momentum facilitates faster convergence, and weight decay helps prevent overfitting [3]. For this study, the learning rate of the SGD optimizer was set to 0.001, momentum was set to 0.9, while the weight decay was set to 1e-5.

Table 3. 4. SGD Optimizer Parameter Settings

Parameter	Value
Learning Rate	0.001
Momentum	0.9
Weight Decay	1e-5

Non-Maximum Suppression: Following the model's predictions on the image dataset, Non-Maximum Suppression (NMS) comes into play during the post-processing phase. NMS evaluates the predicted bounding boxes, considering their confidence scores and overlapping areas (IoU). Bounding boxes with scores below a specified threshold are discarded, and overlapping boxes are refined to retain only the most accurate and non-redundant predictions. This process ensures that the final output of the model is optimized for precision and eliminates duplicate detections.

Model Validation

For validating the performance of the model, the dataset was divided into a 70 – 20 - 10 split for both training, validation, and test sets respectively. The validation split from the created dataset is used as the input. The metrics for validating the results of the model’s performance is its mAP@50, precision, and recall.

The mAP@50 (mean Average Precision at Intersection over Union (IoU) of 0.5) scores of a model is a metric used to evaluate the performance of information retrieval or object detection models. It measures how well a model ranks and retrieves relevant items or objects from a list of candidates, and it considers both precision and recall. It provides a single number that compares the performance of different object detection models, considering both precision and recall, but only counting positive detections if they overlap with the ground truth by 50% or more.

$$\boldsymbol{mAP@50 = \frac{1}{N}\sum_{i=1}^{N} AP_i}$$

The model's precision measures how well the model is able to correctly identify ambulances while avoiding false positives. A high precision means that the YOLOv5 model was able to accurately detect ambulances with a low rate of false positives. Precision is calculated as the ratio of true positive instances to all positive instances of objects in the detector, based on the ground truth.

$$\boldsymbol{Precision = \frac{TP}{TP + FP}}$$

The model's recall is a measure of its ability to detect all the ambulances within a dataset. It is defined as the number of true positives divided by the number of true positives plus the number of false negatives. This means that the recall of the model measures how many of the actual positive cases were correctly identified by the model.

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

Hyperparameter Tuning

In this study, hyperparameter tuning will be performed in order to achieve the best model performance result. This involves batch processing, wherein the model is trained with varying batch sizes (8, 16, 32). Three models will be trained, and its performance metrics would be computed. The model with the highest overall mAP@50 score would be retained for the testing phase of the framework. This process helps identify the optimal batch size, which has a significant impact on the model's convergence and computational efficiency.

Model Testing

The model attaining the highest validation scores is chosen, indicating excellent performance and generalization capabilities. This ultimate model undergoes evaluation on a dedicated test set to assess its robustness and real-world applicability. This step offers a comprehensive measure of the model's efficacy and readiness for deployment.

Using consistent validation metrics, including mAP@50, precision, and recall, the model undergoes tests itself against the test split of the dataset. The mAP@50 score is a crucial benchmark, reflecting the model's ability to rank and retrieve relevant items, providing a holistic assessment of precision and recall. Precision measures the model's accuracy in identifying ambulances while balancing against false positives. Recall assesses the model's proficiency in detecting all ambulances in the dataset. This comprehensive evaluation ensures that the model excels not only in controlled environments but also demonstrates resilience in real-world scenarios, affirming its utility in detecting and classifying ambulances in the Philippines.

Chapter 4

RESULTS AND DISCUSSION

This chapter will present the results of the research, which revolves around the extensive study on ambulance detection and classification in the Philippines. The findings will encompass the utilization of the ambulance detection model based on YOLOv5 small weight model with the incorporation of data augmentation testing using SSIM and the application of hyperparameter tuning. This chapter will also provide the detailed analysis and discussion regarding the results and outcomes, underscoring the performance of the YOLOv5s model in detecting and classifying the different types of ambulances in the Philippines.

Data Augmentation Dissimilarity SSIM Score

The dataset was augmented with the generation of three outputs per original image, resulting in a total of three images per augmentation. To assess the diversity and dissimilarity introduced by these augmentations, SSIM scores were computed by comparing images in triads. The SSIM score, a metric ranging from 0 to 1 where 1 indicates perfect similarity, was chosen to evaluate the structural differences among augmented images.

Average SSIM Score

The average SSIM score for the entire augmented dataset was computed to be **0.1574**. This value suggests a moderate level of dissimilarity among the augmented ambulance and vehicle images, indicating that the augmentation process has effectively introduced variations into the dataset.

The obtained average SSIM score of **0.1574** signifies that, on average, the augmented ambulance and vehicle images are structurally distinct from one another. This aligns with the goal of data augmentation, which is to provide the model with diverse examples for improved generalization. A lower SSIM score is desirable in this context, as it indicates greater dissimilarity among the images, preventing redundancy in the dataset.

Sample Image SSIM Scores

Sample triad images were calculated for its SSIM score before averaging the entire dataset. These sample images range from all the vehicle types found from the dataset which include type I ambulances, type II ambulances, type III ambulances, common vans, and cars.

Figure 4.1 provides a visual representation of the differences between the clean image of a type I ambulance (positioned on the left) and its two augmented counterparts. Both augmented images undergo horizontal flipping, hue changes, and grain adjustments, yet nuanced differences emerge. The middle image exhibits a heightened brightness, creating a mirrored effect, while the right image takes a divergent route with a darker brightness. Notably, these intentional variations contribute to a deliberately low Structural Similarity Index (SSIM) score of 0.187536, indicating substantial dissimilarity from the original clean image.

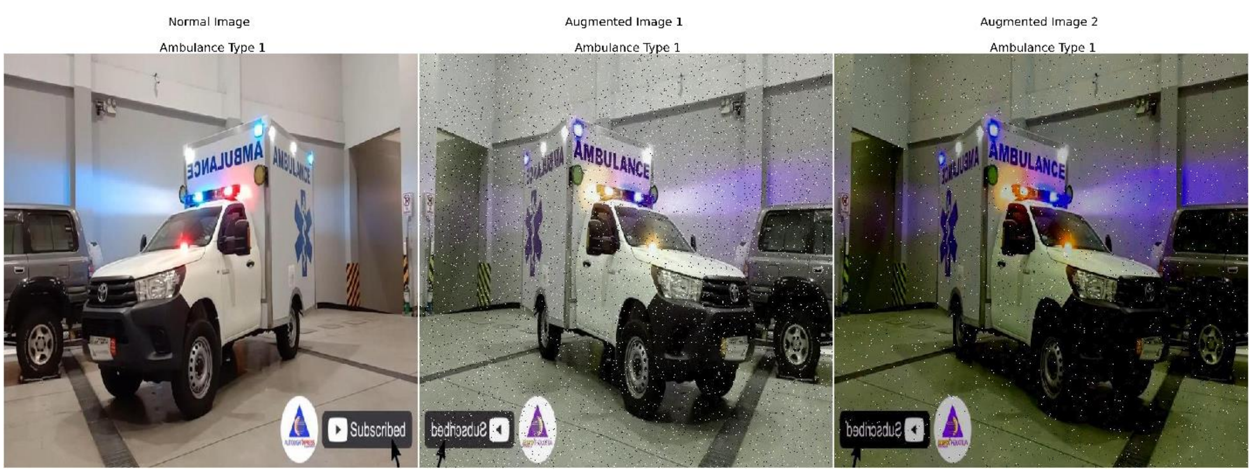


Figure 4. 1. Type I Ambulance Image Augmentation Sample

In **Figure 4.2**, the comparison involves augmented type 2 ambulance images, with the clean image on the left. Both augmented images feature horizontal flipping and added grain. The middle image is brighter, creating a mirrored effect, while the right image is darker. Notably, these deliberate modifications lead to a low Structural Similarity Index (SSIM) score of 0.08750, emphasizing the substantial dissimilarity from the original.

Figure 4. 2. Type II Ambulance Image Augmentation Sample

In **Figure 4.3**, the comparison focuses on augmented images of type 3 ambulances, where the unaltered image is positioned on the left. Both augmented images feature added grain, while the image on the right side undergoes horizontal flipping. The middle image exhibits increased brightness, producing a mirrored effect, while the right image presents a darker tonality. The deliberate introduction of dissimilarity in these augmentations is evident in the low Structural Similarity Index (SSIM) score of 0.22924, underscoring the significant difference from the original image.

Figure 4. 3. Type III Ambulance Image Augmentation Sample

In **Figure 4.4**, the comparison centers around augmented images of common cars, featuring the unaltered image on the left. Both augmented images integrate added grain, with the middle image additionally undergoing horizontal flipping. The middle image displays heightened brightness, introducing a mirrored effect, while the right image embraces a darker tonality. The intentional introduction of these dissimilarities results in a discernible Structural Similarity Index (SSIM) score of 0.12563, underlining a substantial dissimilarity from the original.

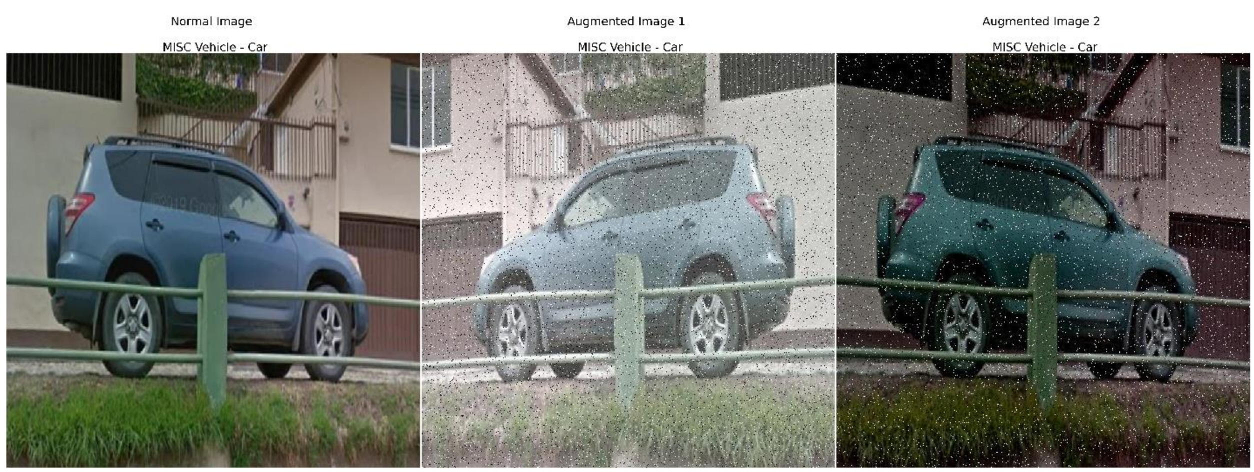


Figure 4. 4. MISC Vehicle - Car Image Augmentation Sample

In **Figure 4.5**, the comparative analysis is directed towards augmented images of common vans, showcasing the unaltered image on the left. Augmentations include the addition of grain in both images, while the right image also undergoes horizontal flipping. Notably, the middle image introduces heightened brightness, creating a mirrored effect, whereas the right image adopts a darker tonality and features increased saturation. The intentional incorporation of these dissimilarities yields a distinctive Structural Similarity Index (SSIM) score of 0.06922, underscoring a pronounced departure from the original.

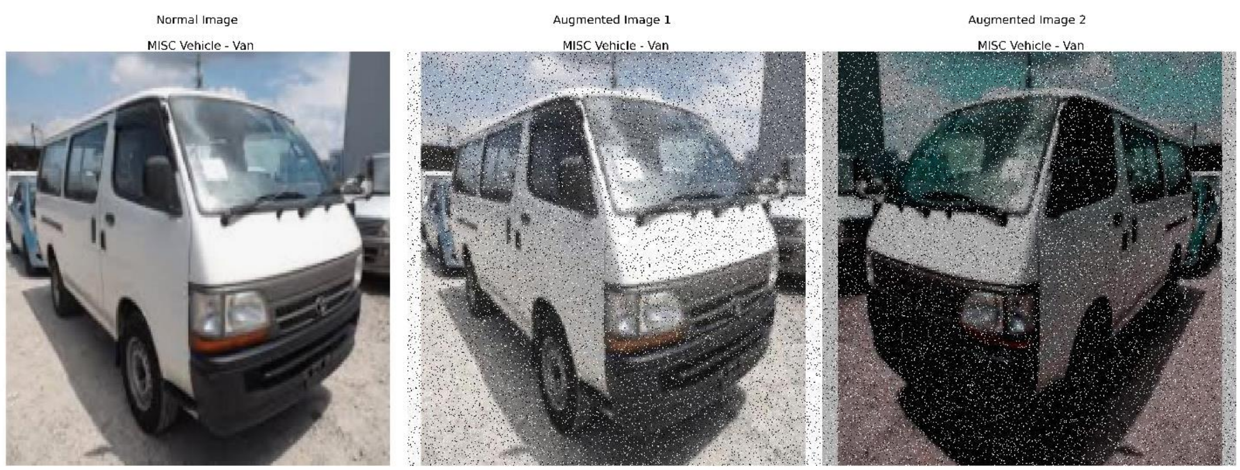


Figure 4. 5. MISC Vehicle - Van Image Augmentation Sample

In vehicle classification based on the sample images, The SSIM scores for vehicle classification depict the dissimilarity introduced during augmentation. Type I Ambulances exhibit noticeable dissimilarity with a score of 0.187536. Type II Ambulances also feature deliberate variations but with a lower score of 0.08750. Type III Ambulances achieve a higher score of 0.22924, emphasizing the successful introduction of diversity. In the category of miscellaneous vehicles, Cars display a moderate SSIM score of 0.12563, while Vans exhibit the lowest score at 0.06922, signifying deliberate differences. These scores collectively validate the efficacy of introducing intentional diversity through augmentation, essential for training adaptable vehicle classification models across real-world scenarios.

Table 4. 1. Sample Images Vehicle SSIM Scores

Vehicle Classification	SSIM Score
Type I Ambulance	0.187536
Type II Ambulance	0.08750
Type III Ambulance	0.22924
MISC Vehicle - Car	0.12563
MISC Vehicle - Van	0.06922

YOLOv5s Hyperparameter Tuning Validation Performance Metrics

mAP@50 score Results

The model was trained after applying the data augmentation to the ambulance dataset. The study has a setup for hyperparameter tuning. Each are different according to their batch sizes which are 8, 16, and 32. The YOLOv5 Small model was trained with the localized dataset and obtained a mean average precision (mAP@50) score of 0.928 for the batch size 8, batch size 16 had 0.927, while batch size 32 had 0.925. The mAP score of each class was obtained and recorded. These classes included the following: Ambulance - Type 1, Ambulance - Type 2, Ambulance - Type 3, Label, Misc Vehicle - Car, Misc Vehicle - Van, and Siren. The mAP@50 score for each of the classes may be found in Table.

The model, when trained with a batch size of 8, demonstrated impressive performance in object detection across various categories. The model's precision in identifying specific types of ambulances was particularly noteworthy, with Ambulance – Type 1 scoring 0.986, Ambulance – Type 2 scoring 0.955, and Ambulance – Type 3 scoring 0.975 in terms of mean Average Precision (mAP) at 50 IoU. The model also excelled in detecting Misc Vehicle – Car (0.944) and Misc

Vehicle – Van (0.982). While the mAP scores for Label (0.778) and Siren (0.824) were relatively lower, they still represent a satisfactory level of detection accuracy. The overall performance of our model was 0.928, demonstrates its versatility in detecting a wide range of objects. These encouraging results highlight the model's potential for practical use in object detection applications.

Table 4. 2. mAP@50 performance of YOLOv5s model Batch Size - 8

Class	mAP@50
Ambulance - Type 1	0.986
Ambulance - Type 2	0.955
Ambulance - Type 3	0.975
Label	0.778
Misc Vehicle - Car	0.994
Misc Vehicle - Van	0.982
Siren	0.824
All	0.928

In the table below, Batch 16 mAP@50 performance is shown. It is noticeable that Ambulance – Type 1 gained the highest score among the three types with a score of 0.986 while Ambulance – Type 2 gained the lowest score of 0.961 and Ambulance – Type 3 gained a score of 0.974. This shows that the system can identify various ambulance types with high precision even if the Ambulance – Type 2 got the lowest score. The "Label: class gained a score of 0.781 which is slightly higher than the Batch 8's result. For the "Misc Vehicle – Car" which has the lowest score among all classes of 0.99 and "Misc Vehicle – Van" have 0.983 score which is also slightly higher than Batch 8's result. The overall performance of Batch 16 provides an average score of 0.927 which represents the system's performance across all classes.

Table 4. 3. mAP@50 performance of YOLOv5s model Batch Size - 16

Class	mAP@50
Ambulance - Type 1	0.986
Ambulance - Type 2	0.961
Ambulance - Type 3	0.974
Label	0.781
Misc Vehicle - Car	0.99
Misc Vehicle - Van	0.983
Siren	0.815
All	0.927

Table 4.4 illustrates the performance of the model with a batch size of 32. It's worth mentioning that the model demonstrated a high degree of precision in identifying various types of ambulances (Type 1, Type 2, and Type 3), with mAP scores exceeding 0.95. This highlights the model's reliable performance in detecting these vital objects. Although the mAP scores for Label (0.777) and Siren (0.818) were somewhat lower, they still indicate a satisfactory level of detection accuracy for these classes. The "All" category, with an average mAP of 0.925, suggests that the model's overall object detection capabilities remain strong, even when trained with a larger batch size of 32.

Table 4. 4 mAP@50 performance of YOLOv5s model Batch Size - 32

Class	mAP@50
Ambulance - Type 1	0.975
Ambulance - Type 2	0.962
Ambulance - Type 3	0.967
Label	0.777
Misc Vehicle - Car	0.99
Misc Vehicle - Van	0.984
Siren	0.818
All	0.925

The YOLOv5s with batch size 8 proposed framework's mAP score is noticeably higher compared to the mAP score of the other models with batch sizes 16 and 32. This shows that YOLOv5s with batch size 8 is the model with the best performance particular to its accuracy.

Precision/Recall Results

In the evaluation of the proposed YOLOv5 Small model, the precision and recall performance for various classes were assessed. Notably, the batch size 8 model achieved high precision scores which signifies its ability to accurately capture instances.

Table 4.5 below presents the precision and recall scores of Batch size 8. Ambulance – Type 1 have a result of 0.965 precision and a recall of 0.988 while Ambulance – Type 2 have a precision of 0.888 which is lower than the precision of Ambulance – Type 1 and a recall of 0.952 while the Ambulance – Type 3 have a precision score of 0.993 which is higher than the Ambulance – Type 2 and a recall of 0.939 which is the lowest among the three types. For the "Label" precision score had 0.828 and a recall of 0.701 which is reasonably good, but it might have some error due to lower recall score. Misc Vehicle – Car has a precision of 0.959 and recall of 0.993 and Misc Vehicle – Van has a precision score of 0.986 and a recall of 0.95 which both indicates that the system have high accuracy in detecting them. The “Siren” also has a reasonable precision and recall of 0.863 and 0.764 but not as high as other classes. The overall performance of Batch size 8 have a precision of 0.926 and recall of 0.898 suggests that the system have an effective and accurate detection in identifying and locating objects.

Table 4. 5. Precision and Recall Results of YOLOv5s model Batch 8

Class	Precision	Recall
Ambulance - Type 1	0.965	0.988
Ambulance - Type 2	0.888	0.952
Ambulance - Type 3	0.993	0.939
Label	0.828	0.701
Misc Vehicle - Car	0.959	0.993
Misc Vehicle - Van	0.986	0.95
Siren	0.863	0.764
All	0.926	0.898

Table 4.6 displayed the results for the precision and recall of the model with a batch size of 16. Ambulance type 1 achieved a 0.976 precision while its recall achieved a 0.979 value. Type 2 had 0.986 and its recall had 0.957. Ambulance type with 0.98 and 0.945 for its precision and recall. In the table, the lowest classes are the label and siren since they are usually relatively small in size. It is hard to detect and identify in an image. For the other classes, it still received a high-performance rating making it a reliable model.

Table 4. 6. Precision and Recall Results of YOLOv5s model Batch 16

Class	Precision	Recall
Ambulance - Type 1	0.976	0.979
Ambulance - Type 2	0.896	0.957
Ambulance - Type 3	0.98	0.945
Label	0.818	0.708
Misc Vehicle - Car	0.948	0.98
Misc Vehicle - Van	0.986	0.95
Siren	0.857	0.761
All	0.923	0.897

Table 4.7 displayed the results for each class under the model trained with Batch size 32. The ambulance type 1 had a precision rate of 0.988 and recall of 0.962. The ambulance type 2 had a precision rate of 0.884 and recall for 0.965 The ambulance type 3 had a precision rate of 0.994 and recall for 0.938. These results simply showed that all batch sizes are accurate. However, all became similar with the detection result for label and siren since it is a little bit harder to detect due to its size. Furthermore, all classes are highly accurate.

Table 4. 7. Precision and Recall Results of YOLOv5s model Batch 32

Class	Precision	Recall
Ambulance - Type 1	0.988	0.962
Ambulance - Type 2	0.884	0.965
Ambulance - Type 3	0.994	0.938
Label	0.81	0.715
Misc Vehicle - Car	0.954	0.98
Misc Vehicle - Van	0.976	0.964
Siren	0.873	0.75
All	0.926	0.896

Model Testing Results

With a batch size of 8, the group effectively trained the YOLOv5s model, and it has stood out by achieving the best scores during the validation process. Given its excellent performance, it's well-suited for comprehensive testing and evaluation using the test dataset. In the next phase, researcher will be evaluating the model's capacity to generalize and consistently detect a wide variety of object classes accurately.

Our YOLOv5s model, which was trained using a batch size of 8, has demonstrated remarkable object detection abilities across a wide range of classes, according to the test results. The mean Average Precision (mAP@50) scores are consistently high, with Ambulance - Type 3 topping the list with an impressive 0.989, followed closely by Misc Vehicle - Car with a score of 0.991. These scores indicate the model's precision in recognizing specific objects within the tested classes.

The precision scores indicate robust model performance, with Ambulance - Type 3 achieving an extraordinary score of 0.983. The recall scores are also significant, with Ambulance - Type 3 achieving an impressive 0.998. However, the “Label” class has relatively lower scores in both precision and recall, indicating that there is potential for enhancing detection performance for this class.

Table 4. 8. Model Testing Results

Class	mAP@50	Precision	Recall
All	0.93	0.929	0.903
Ambulance – Type 1	0.974	0.961	0.987
Ambulance – Type 2	0.947	0.908	0.958
Ambulance – Type 3	0.989	0.983	0.98
Label	0.823	0.832	0.736
Misc Vehicle - Car	0.991	0.971	0.97
Misc Vehicle - Van	0.941	0.974	0.905
Siren	0.842	0.874	0.786

Chapter 5

SUMMARY, CONCLUSION, AND RECOMMENDATION

Summary

This study aimed to develop and utilize an object detection model that can help detect and classify ambulances on the road according to its type. The researchers focused on the utilization of the YOLO algorithm, specifically the YOLOv5 Small, which can aid in efficient image processing, particularly in detecting and classifying types and features of ambulances. As it is lightweight, it provides convenience and flexibility to apply the model to different devices. The group created a localized dataset that is composed of ambulance images in the Philippines with different classifications. This dataset was created using the method of image scraping on Google, consisting of 1840 images. In these images the group identified and considered 7 classes which are the following:

- Type I Ambulance
- Type II Ambulance
- Type III Ambulance
- Siren
- Label
- MISC Vehicle – Car
- MISC Vehicle – Van

The calculated average SSIM score for the complete augmented dataset is **0.1574**, indicating a degree of dissimilarity between the clean and augmented ambulance and vehicle images. This suggests that the augmentation process has successfully introduced variations into the dataset.

Once the dataset has been tested, the researchers performed hyperparameter tuning to find the best set of hyperparameters for the model. The model was trained three times with varying batch sizes of 8, 16, and 32. All of the trained models produced positive outcomes. However, among these instances, the batch 8 model had the best performance result based on the validation results and received a mAP@50 score of 92.8%. The model with the hyperparameter of batch size 8 was then tested against the test set of the dataset and achieved a total mAP@50 score of 93% in detecting and classifying ambulances and vehicles. The results of the configuration led the group to the model performance seen below:

Table 5. 1. Model Test Results

Class	mAP@50	Precision	Recall
Ambulance - Type 1	0.974	0.961	0.987
Ambulance - Type 2	0.947	0.908	0.958
Ambulance - Type 3	0.989	0.983	0.98
Label	0.823	0.832	0.736
Misc Vehicle - Car	0.991	0.971	0.97
Misc Vehicle - Van	0.941	0.974	0.905
Siren	0.842	0.874	0.786
All	0.93	0.929	0.903

Conclusion

To classify ambulances according to its different types namely type 1, type 2, and type 3, this study had several objectives. The primary objective of this study was to evaluate the performance metrics of the YOLOv5 model, with a specific focus on its small weight configuration. The task involved the detection and classification of Philippines ambulances, along with the identification of key features. The achieved validation score of 92.8% mAP and a test score of 93% indicate a robust and reliable performance of the YOLOv5 model in meeting the specified objectives. The high accuracy in both validation and test phases highlights the model's proficiency in detecting and classifying Philippines ambulances, thus fulfilling the first research objective.

The second objective aimed to assess the dissimilarity between clean and augmented images using the Structural Similarity Index (SSIM). The obtained SSIM score of 0.1574 indicates a low structural similarity between the original and augmented images. This low SSIM score suggests that the augmentation process introduced significant diversity and non-redundancy into the dataset. A low SSIM score is desirable in this context, as it reflects the successful augmentation of the dataset, meeting the intended purpose of creating diverse and non-redundant data.

The third objective involved the identification of the right set of hyperparameters for the YOLOv5 small model through hyperparameter tuning. The experimentation with different batch sizes (8, 16, and 32) while maintaining a constant epoch of 150 revealed that a batch size of 8 yielded the best results. The selected batch size achieved a validation score of 92.8%, showcasing the effectiveness of hyperparameter tuning in maximizing the model's performance. This finding affirms the importance of fine-tuning hyperparameters for optimal results in object detection tasks using the YOLOv5 model.

Recommendation

This study has seen promising evidence that implementation of an ambulance vehicle detection and classification. However, further research is needed to build, strengthen, validate, and to advance these findings. Implementing advanced optimization techniques may further enhance the YOLOv5 model's performance. Exploring methods such as fine-tuning, transfer learning, or leveraging more sophisticated architectures could lead to improved accuracy and efficiency in detecting and classifying Philippine ambulances. Continuous refinement of the model through optimization is essential for ensuring its reliability in practical deployment.

Recognizing the potential real-world impact of the model, it is recommended to explore applications in prioritizing ambulances at intersections. This would involve studying the model's performance in identifying and prioritizing emergency vehicles in traffic, thereby contributing to more efficient emergency response systems in the Philippines.

To ensure robust performance across various real-world scenarios, efforts should be made to create a more uniform dataset. This includes capturing images under different lighting conditions, weather scenarios, and traffic densities. A uniform dataset will contribute to a more generalized model that can effectively handle the diverse conditions encountered in the Philippines setting.

Recognizing that real-world scenarios are dynamic; it is crucial to establish a system for continuous monitoring and updates. Regularly updating the model with new data and adapting it to evolving conditions will ensure its continued effectiveness in the ever-changing environment of Philippine traffic management.

REFERENCES

[1] C. Wang, H. Wang, F. Yu and W. Xia, "A High-Precision Fast Smoky Vehicle Detection Method Based on Improved Yolov5 Network," 2021 IEEE International Conference on Artificial Intelligence and Industrial Design (AIID), Guangzhou, China, 2021, pp. 255-259, doi: 10.1109/AIID51893.2021.9456462.

[2] R. Dey, B. K. Pandit, A. Ganguly, A. Chakraborty and A. Banerjee, "Deep Neural Network Based Multi-Object Detection for Real-time Aerial Surveillance," 2023 11th International Symposium on Electronic Systems Devices and Computing (ESDC), Sri City, India, 2023, pp. 1-6, doi: 10.1109/ESDC56251.2023.10149866.

[3] X. Song and W. Gu, "Multi-objective real-time vehicle detection method based on yolov5," 2021 International Symposium on Artificial Intelligence and its Application on Media (ISAIAM), Xi'an, China, 2021, pp. 142-145, doi: 10.1109/ISAIAM53259.2021.00037.

[4] R. Yagfarov, V. Ostankovich and S. Gafurov, "Augmentation-Based Object Detection for Winter Time Applications," 2019 3rd School on Dynamics of Complex Networks and their Application in Intellectual Robotics (DCNAIR), Innopolis, Russia, 2019, pp. 178-180, doi: 10.1109/DCNAIR.2019.8875557.

[5] Xiaohui Guan and Xinxin Sun. 2022. Vehicle identification and traffic statistics system in traffic intersection. *2022 4th International Conference on Intelligent Control, Measurement and Signal Processing (ICMSP)* (2022). DOI:http://dx.doi.org/10.1109/icmsp55950.2022.9858960

[6] Qiusen Wang, Shenghong Xu, and Sanqiang Zheng. 2022. Research on the model of emergency vehicles passing priority at intersections based on vehicle-road cooperative system. *2022 14th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)* (2022). DOI:http://dx.doi.org/10.1109/icmtma54903.2022.00196

[7] Hitesh Agarwal and J.N. Rai. 2022. Traffic control system based on density with emergency priority mechanism. *2022 International Conference on Electronics and Renewable Systems (ICEARS)* (2022). DOI:http://dx.doi.org/10.1109/icears53579.2022.9751869

[8] E. Paunova-Hubenova and E. Trichkova-Kashamova. 2022. Algorithm for traffic management with priority for emergency vehicles. *2022 XXXI International Scientific Conference Electronics (ET)* (2022). DOI:http://dx.doi.org/10.1109/et55967.2022.9920275

[9] Gour Karmakar, Abdullahi Chowdhury, Joarder Kamruzzaman, and Iqbal Gondal. 2021. A smart priority-based traffic control system for emergency vehicles. *IEEE Sensors Journal* 21, 14 (2021), 15849–15858. DOI:http://dx.doi.org/10.1109/jsen.2020.3023149

[10] Giuseppe Benedetti, Maria Pia Fanti, Agostino Marcello Mangini, and Fabio Parisi. 2021. Application of deep reinforcement learning for traffic control of road intersection with

emergency vehicles. *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (2021). DOI:http://dx.doi.org/10.1109/smc52423.2021.9658968

[11] Elivelton O. Rangel, Daniel G. Costa, and Maycon M. Peixoto. 2021. An optimization approach for emergency vehicles dispatching and traffic lights adjustments in response to emergencies in Smart Cities. *2021 XI Brazilian Symposium on Computing Systems Engineering (SBESC)* (2021). DOI:http://dx.doi.org/10.1109/sbesc53686.2021.9628243

[12] Daniel G. Costa, Francisco Vasques, Ana Aguiar, and Paulo Portugal. 2020. Automatic assignment of emergency vehicles in response to sensors-based generated alarms in smart city scenarios. *2020 IEEE International Smart Cities Conference (ISC2)* (2020). DOI:http://dx.doi.org/10.1109/isc251055.2020.9239062

[13] Van-Thuan Tran and Wei-Ho Tsai. 2021. Audio-Vision Emergency Vehicle Detection. *IEEE Sensors Journal* 21, 24 (2021), 27905–27917. DOI:http://dx.doi.org/10.1109/jsen.2021.3127893

[14] Krishnendu Choudhury and Dalia Nandi. 2021. Detection and prioritization of emergency vehicles in Intelligent Traffic Management System. *2021 IEEE Bombay Section Signature Conference (IBSSC)* (2021). DOI:http://dx.doi.org/10.1109/ibssc53889.2021.9673211

[15] R. Anil, M. Satyakumar, and Ameena Salim. 2019. Emergency vehicle signal pre-emption system for heterogeneous traffic condition : A case study in Trivandrum City. *2019 4th International Conference on Intelligent Transportation Engineering (ICITE)* (2019). DOI:http://dx.doi.org/10.1109/icite.2019.8880151

[16] Weiqi Yu, Weichen Bai, Wenjing Luan, and Liang Qi. 2022. State-of-the-art review on Traffic Control Strategies for Emergency vehicles. *IEEE Access* 10 (2022), 109729–109742. DOI:http://dx.doi.org/10.1109/access.2022.3213798

[17] Luz A. Magre Colorado, Jorge Franco Ibanez, and Juan Carlos Martinez-Santos. 2020. Leveraging emergency response system using the internet of things. A preliminary approach. *2020 17th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)* (2020). DOI:http://dx.doi.org/10.1109/cce50788.2020.9299175

[18] Muhammad Azhad Bin Zuraimi and Fadhlan Hafizhelmi Kamaru Zaman. 2021. Vehicle detection and tracking using Yolo and DeepSORT. *2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)* (2021). DOI:http://dx.doi.org/10.1109/iscaie51753.2021.9431784

[19] Xiaohui Guan and Xinxin Sun. 2022. Vehicle identification and traffic statistics system in traffic intersection. *2022 4th International Conference on Intelligent Control, Measurement and Signal Processing (ICMSP)* (2022). DOI:http://dx.doi.org/10.1109/icmsp55950.2022.9858960

[20] Jiaxing Zhang, Wen Xiao, Benjamin Coifman, and Jon P. Mills. 2020. Vehicle Tracking and Speed Estimation from Roadside Lidar. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13 (2020), 5597–5608. DOI:http://dx.doi.org/10.1109/jstars.2020.3024921

[21] Yunzhen Wang, Hongbing Ma, and Liangliang Li. 2022. Road traffic vehicle detection method using lightweight yolov5 and attention mechanism. *2022 7th International Conference on Image, Vision and Computing (ICIVC)* (2022). DOI:http://dx.doi.org/10.1109/icivc55077.2022.9886525

[22] Chengpeng Wang, Huanqin Wang, Fajun Yu, and Wangjin Xia. 2021. A high-precision fast smoky vehicle detection method based on improved Yolov5 Network. *2021 IEEE International Conference on Artificial Intelligence and Industrial Design (AIID)* (2021). DOI:http://dx.doi.org/10.1109/aiid51893.2021.9456462

[23] Othman O. Khalifa, Muhammad H. Wajdi, Rashid A. Saeed, Aisha H. Hashim, Muhammed Z. Ahmed, and Elmustafa Sayed Ali. 2022. Vehicle detection for vision-based intelligent transportation systems using convolutional neural network algorithm. *Journal of Advanced Transportation* 2022 (2022), 1–11. DOI:http://dx.doi.org/10.1155/2022/9189600

[24] Zhuang Li, Xincheng Tian, Yan Liu, and Xiaorui Shi. 2022. Vehicle tracking method based on attention-yolov5 and optimized DeepSort models. *2022 IEEE 11th Data Driven Control and Learning Systems Conference (DDCLS)* (2022). DOI:http://dx.doi.org/10.1109/ddcls55054.2022.9858395

[25] Fuheng Guo and Yi Xu. 2022. Vehicle Analysis System based on DeepSORT and yolov5. *2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA)* (2022). DOI:http://dx.doi.org/10.1109/cvidliccea56201.2022.9824363

[26] Jiandong Wang, Yahui Dong, Shuangrui Zhao, and Zhiwei Zhang. 2023. A high-precision vehicle detection and tracking method based on the attention mechanism. *Sensors* 23, 2 (2023), 724. DOI:http://dx.doi.org/10.3390/s23020724

[27] Liu Xiaomeng, Feng Jun, and Chen Peng. 2022. Vehicle detection in traffic monitoring scenes based on improved Yolov5s. *2022 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI)* (2022). DOI:http://dx.doi.org/10.1109/icceai55464.2022.00103

[28] Baoping Xiao, Jinghua Guo, and Zhifei He. 2021. Real-time object detection algorithm of autonomous vehicles based on improved Yolov5s. *2021 5th CAA International Conference on Vehicular Control and Intelligence (CVCI)* (2021). DOI:http://dx.doi.org/10.1109/cvci54083.2021.9661149

[29] Xuyang Song and Wei Gu. 2021. Multi-objective real-time vehicle detection method based on Yolov5. *2021 International Symposium on Artificial Intelligence and its Application on Media (ISAIAM)* (2021). DOI:http://dx.doi.org/10.1109/isaiam53259.2021.00037

[30] Lei Shao, Zhenqiang Fan, Ji Li, and Hongli Liu. 2022. Research on road vehicle detection based on improved Yolov5s. *2022 IEEE International Conference on Mechatronics and Automation (ICMA)* (2022). DOI:http://dx.doi.org/10.1109/icma54519.2022.9856288

[31] Erika Ritzelle Bondoc, Francis Percival Caparas, John Eddie Macias, Vileser T. Naculangga, and Jheanel E. Estrada. 2018. An intelligent road traffic information system using text analysis in the most congested roads in Metro Manila. *2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology,Communication and Control, Environment and Management (HNICEM)* (2018). DOI:http://dx.doi.org/10.1109/hnicem.2018.8666416

[32] Keyan Chen and yUHONG gUO. 2023. Object Detection in 20 Years: A Survey. (January 2023). Retrieved September 21, 2023 from https://ieeexplore-ieee-org.mapua.idm.oclc.org/document/10028728

[33] Thammi Reddy Konala, Anusha Nammi, and Divya Sree Tella. 2023. Analysis of Live Video Object Detection using YOLOv5 and YOLOv7. (July 2023). Retrieved September 21, 2023 from https://ieeexplore-ieee-org.mapua.idm.oclc.org/document/10169926

[34] Larry Valdez, Francis Earl Beran, and Angelino Pimentel. 2023. Design and development of an IoT-based smart ambulance system with patient monitoring. (March 2023). Retrieved September 21, 2023 from https://ieeexplore-ieee-org.mapua.idm.oclc.org/document/10059568

[35] J. Deepak and M. Baranitharan. 2014. Efficient dynamic traffic control system using wireless sensor networks. (June 2014). Retrieved September 21, 2023 from https://ieeexplore-ieee-org.mapua.idm.oclc.org/document/6844280

[36] Kim Carol Maligali, Albertson D. Amante, and Ryan R. Tejada. 2022. Machine Vision System of Emergency Vehicle Detection System Using Deep Transfer Learning. (May 2022). Retrieved September 21, 2023 from https://ieeexplore-ieee-org.mapua.idm.oclc.org/document/9765002

[37] L. -S. Tseng, C. -H. Wu, Y. H. Chen and C. -H. Tsai, GAN-based Data Augmentation for Metal Surface Defect Detection Using Convolutional Neural Networks. (2023). Ieeexplore.ieee.org. Retrieved October 12, 2023, from https://ieeexplore.ieee.org/document/10219520