

# Enhancing Infrastructure Detection Using Drone Imagery: A Comparative Analysis Using Machine Learning Models with a Custom Bangladesh-Based Dataset for Improved Urban Planning

by

ATIKUR RAHMAN

20301440

SHUDEB GHOSH BARNO

20301222

NABILA ISLAM BORNO

20301117

MUSAIB IBN HABIB MIKDAD

20301356

MD. RAFIUL ISLAM MAHI

20301324

A thesis submitted to the Department of Computer Science and Engineering  
in partial fulfillment of the requirements for the degree of  
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering  
Brac University  
July 2024

© 2024. Brac University  
All rights reserved.

# Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

## Student's Full Name & Signature:



---

Atikur Rahman  
20301440



---

Shudeb Ghosh Barno  
20301222



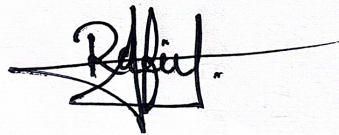
---

Nabila Islam Borno  
20301117



---

Musaib Ibn Habib Mikdad  
20301356



---

Md. Rafiul Islam Mahi  
20301324



# Approval

The thesis titled “Enhancing Infrastructure Detection Using Drone Imagery: A Comparative Analysis Using Machine Learning Models with a Custom Bangladesh-Based Dataset for Improved Urban Planning” submitted by

1. Atikur Rahman (20301440)
2. Shudeb Ghosh Barno (20301222)
3. Nabila Islam Borno (20301117)
4. Musaib Ibn Habib Mikdad (20301356)
5. Md. Rafiul Islam Mahi (20301324)

Of Spring, 2024 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on JULY 5, 2024

## Examining Committee:

Supervisor:  
(Member)



Dr. Md. Khalilur Rhaman, PhD  
Professor

Department of Computer Science and Engineering  
Brac University

Co Supervisor:  
(Member)



MR. Rafeed Rahman  
Lecturer

Department of Computer Science and Engineering  
Brac University

Thesis Coordinator:  
(Member)

---

Md. Golam Rabiul Alam, PhD  
Professor  
Department of Computer Science and Engineering  
Brac University

Head of Department:  
(Chair)

---

Sadia Hamid Kazi, PhD  
Chairperson and Associate Professor  
Department of Computer Science and Engineering  
Brac University

# Abstract

The growth of cities in different developing nations including Bangladesh is posing several challenges in the making of sustainable cities. Growth of infrastructure whilst maintaining the sustainability of urban sprawl is a major challenge in these areas experiencing rapid development. This research aims to address the issues of achieving the efficient and effective technology for an accurate and timely monitoring and cost-effective urban infrastructure through the utilization of airborne remote sensing technologies. This paper describes and compares several machine learning models, including YOLOv5, YOLOv7, YOLOv8 and a custom model developed by us that enhances the performance of object detection of infrastructure conditions on images captured by drones. Our study is based on a unique data set specifically developed for Bangladesh encompassing multiple aspects of urban features as well as the context of infrastructure expansion across different regions. This dataset is critical during the training and testing stage since the combination of features that works well for our models is tailored for optimal use in urban environments in Bangladesh. In the developed approach, the created machine learning models will be used to identify and recognize the exact building, the roads, and other urban infrastructure features in the images that are used for training. Although our comparison and experimental tests are devoted to the measurement of the detection accuracy and calculation speed of each model, we also dedicate a section for the assessment of detection appropriateness in varied urban scenarios as well as in various illumination conditions. There is a significant improvement in many of the conclusion and result as shown for YOLOv8 and our own custom model in terms of the ability of the model to perform more accurate prediction and the ability to handle the spam case at various conditions of spatial and temporal dimensions of urbanization. The above findings suggest that using drone imagery can indeed be effective in utilizing deep learning tools for urban growth in cities enhancing urban planning. It also demonstrates the possibility of using the approach presented in our work in cities with various populations and levels of development for the most efficient management of these dynamic urban processes. The use of temporal data also have enhanced our knowledge in that it is now possible to monitor temporal changes that occur in infrastructural systems and also in making urban design prediction. The former approach means avoiding the necessary conceptual development for more effective uses of such models in real-world development of cities and communities. In conclusion, this thesis is able to show that hybrid of the latest machine learning models for infrastructure detection shows a great potential for the application to urban areas within Bangladesh and similar developing areas with a high population density in a short time. Through surveys it provides a comparison not only of the strengths and weaknesses of each model, but also the framework for future development of more responsive and sustainable urban planning applications.

**Keywords:** Deep Learning, Machine Learning, YOLO, Object Detection, Urban Infrastructure, Drone Imagery, Aerial Analysis, Urban Planning, Bangladesh, Spatial-Temporal Analysis, Urban Development, Sustainability

## Acknowledgement

Firstly, all praise to the Great Allah for whom our thesis have been completed without any major interruption.

Secondly, to our supervisor Dr. Md. Khalilur Rahman Sir for his kind support and advice in our work.

And also, to Sayantan Roy Arko, RA of Dr. Md. Khalilur Rahman who helped us whenever we needed help.

Thirdly, to our co-supervisor Mr. Rafeed Rahman Sir for his kind support and advice in our work.

And finally to our parents without their throughout support, it may not be possible. With their kind support and prayer we are now on the verge of our graduation.

# Table of Contents

|                                                     |           |
|-----------------------------------------------------|-----------|
| Declaration                                         | i         |
| Approval                                            | ii        |
| Abstract                                            | iv        |
| Dedication                                          | v         |
| Acknowledgment                                      | v         |
| Table of Contents                                   | vi        |
| List of Figures                                     | viii      |
| <b>1 Introduction</b>                               | <b>2</b>  |
| 1.1 Aims and Objectives . . . . .                   | 3         |
| 1.2 Research Problem . . . . .                      | 4         |
| <b>2 Literature Review</b>                          | <b>5</b>  |
| <b>3 Prediction Modeling using Machine Learning</b> | <b>10</b> |
| 3.1 Working Plan . . . . .                          | 10        |
| 3.2 Machine Learning Models . . . . .               | 13        |
| 3.2.1 YOLOv5 . . . . .                              | 13        |
| 3.2.2 YOLOv7 . . . . .                              | 14        |
| 3.2.3 YOLOv8 . . . . .                              | 15        |
| 3.2.4 Custom Model . . . . .                        | 16        |
| 3.3 Dataset Collection . . . . .                    | 17        |
| 3.4 Input Data . . . . .                            | 18        |
| 3.5 Annotation . . . . .                            | 19        |
| 3.6 Data Pre-processing . . . . .                   | 20        |
| 3.7 Augmentation . . . . .                          | 25        |
| <b>4 Result and Discussion</b>                      | <b>26</b> |
| 4.1 YOLO V5 . . . . .                               | 27        |
| 4.2 YOLO V7 . . . . .                               | 31        |
| 4.3 YOLO V8 . . . . .                               | 33        |
| 4.4 Custom Model . . . . .                          | 36        |
| 4.5 Analysis . . . . .                              | 38        |
| 4.6 Categorization . . . . .                        | 38        |

|          |                                   |           |
|----------|-----------------------------------|-----------|
| 4.7      | Comparison . . . . .              | 40        |
| 4.8      | 3D Augmentation . . . . .         | 42        |
| 4.9      | Map Integration . . . . .         | 47        |
| <b>5</b> | <b>Conclusion and Future Work</b> | <b>48</b> |
| 5.1      | Future Work . . . . .             | 49        |
|          | <b>Bibliography</b>               | <b>53</b> |

# List of Figures

|      |                                                                                                     |    |
|------|-----------------------------------------------------------------------------------------------------|----|
| 2.1  | Graphical Abstract of UGS segmentation . . . . .                                                    | 6  |
| 2.2  | Example input images and the corresponding output images produced<br>by the trained model . . . . . | 7  |
| 2.3  | Semantic segmentation . . . . .                                                                     | 8  |
| 2.4  | Fast multiclass vehicle detection . . . . .                                                         | 9  |
| 3.1  | The flow chart of the proposed detection model . . . . .                                            | 11 |
| 3.2  | YOLO v5 Internal Architecture . . . . .                                                             | 14 |
| 3.3  | YOLO v7 Internal Architecture . . . . .                                                             | 15 |
| 3.4  | YOLO v8 Internal Architecture . . . . .                                                             | 16 |
| 3.5  | Custom Model Internal Architecture . . . . .                                                        | 16 |
| 3.6  | Classes . . . . .                                                                                   | 17 |
| 3.7  | Workflow of Dataset . . . . .                                                                       | 17 |
| 3.8  | Input Pictures from Drone . . . . .                                                                 | 18 |
| 3.9  | Manual Annotation . . . . .                                                                         | 19 |
| 3.10 | Picture of isolated building . . . . .                                                              | 21 |
| 3.11 | Picture of isolated solar panel . . . . .                                                           | 21 |
| 3.12 | Picture of isolated pond . . . . .                                                                  | 22 |
| 3.13 | Picture of isolated vehicles . . . . .                                                              | 22 |
| 3.14 | Original & resized image after scalling . . . . .                                                   | 23 |
| 3.15 | Auto-oriented image . . . . .                                                                       | 24 |
| 4.1  | Map, precision and recall metrics & Train-loss Graphs from V5 . . . . .                             | 27 |
| 4.2  | Precision-Recall Curve from V5 . . . . .                                                            | 28 |
| 4.3  | F1-Confidence Curve from V5 . . . . .                                                               | 28 |
| 4.4  | Confusion Metrics from V5 . . . . .                                                                 | 29 |
| 4.5  | V5 Instance . . . . .                                                                               | 29 |
| 4.6  | Predicted Output from V5 . . . . .                                                                  | 30 |
| 4.7  | Map, precision and recall metrics & Train-loss Graphs from V7 . . . . .                             | 31 |
| 4.8  | Precision-Recall Curve from V7 . . . . .                                                            | 31 |
| 4.9  | F1-Confidence Curve from V7 . . . . .                                                               | 32 |
| 4.10 | Confusion Metrics from V7 . . . . .                                                                 | 32 |
| 4.11 | V7 Custom Image Prediction Result . . . . .                                                         | 33 |
| 4.12 | Map, precision and recall metrics & Train-loss Graphs from V8 . . . . .                             | 33 |
| 4.13 | Confusion Metrics from V8 . . . . .                                                                 | 34 |
| 4.14 | Predicted Output from V8 . . . . .                                                                  | 34 |
| 4.15 | Custom Image for Detection in v8 model . . . . .                                                    | 35 |
| 4.16 | Result for v8 using Custom Image . . . . .                                                          | 35 |
| 4.17 | Custom Model Training Graph . . . . .                                                               | 36 |

|      |                                                |    |
|------|------------------------------------------------|----|
| 4.18 | Custom Loss . . . . .                          | 36 |
| 4.19 | Predicted Output from Custom Model . . . . .   | 37 |
| 4.20 | Catagorization . . . . .                       | 39 |
| 4.21 | Bangladeshi picture in our model . . . . .     | 40 |
| 4.22 | Foreign picture in our model . . . . .         | 40 |
| 4.23 | Foreign picture in foreign model . . . . .     | 41 |
| 4.24 | Bangladeshi picture in foreign model . . . . . | 41 |
| 4.25 | Densify Point Cloud . . . . .                  | 43 |
| 4.26 | Mesh Construction . . . . .                    | 43 |
| 4.27 | Taken picture . . . . .                        | 44 |
| 4.28 | 3D mapping . . . . .                           | 44 |
| 4.29 | Building height prediction . . . . .           | 45 |
| 4.30 | Long distance prediction . . . . .             | 45 |
| 4.31 | Short distance prediction . . . . .            | 45 |
| 4.32 | Exact Location With Map Integration . . . . .  | 47 |
| 5.1  | Prototype . . . . .                            | 49 |
| 5.2  | Predicted Output . . . . .                     | 49 |



# Chapter 1

## Introduction

Urban development is a changing process that impacts the appearance and livability of cities. Monitoring and comprehending this transformation for planning, environmental preservation and managing infrastructure is essential. Imagery is a valuable tool in achieving these goals, which offers a down perspective of urban areas and their evolution over time. This article presents an approach that effectively combines temporal analysis of aerial images to detect urban development.

In the decades, as cities have overgrown and resources have become more in demand, monitoring urban expansion and transformations has become crucial. Aerial imagery captured by satellites, drones, or aeroplanes presents an opportunity to observe these changes in detail. By analyzing these photographs, we can identify growth aspects, including the construction of new structures, the development of road networks, the presence of green spaces, and shifts in land usage patterns. Object detection consists of determining the location and scale of all object instances, if any, present in an image. Thus, the objective of an object detector is to find all object instances of one or more given object classes regardless of scale, location, pose, view concerning the camera, partial occlusions, and illumination conditions.

Change detection is the process of identifying differences by observing images at different times. In this context, spatial analysis involves examining photographs to identify distinct features and transformations. It involves utilizing computer vision techniques like object identification and segmentation to recognize and categorize elements within each image. However, when studying growth, it is crucial to consider the aspect. Urban environments change over time due to construction, demolition and other factors. Therefore, comprehending the mechanisms of expansion necessitates a perspective [8] [3] [9] [1].

Our approach employs a multifaceted strategy, utilizing machine learning techniques and deep neural networks to combine spatial and temporal data. By utilizing a time series of aerial pictures, our model effectively tracks changes in urban areas over time, revealing patterns, trends and anomalies in development. This integration of geographical and temporal information provides extensive insights into the progression of urbanization. Detecting urban development in both spatial and temporal aspects has considerable potential for urban planners, politicians, and academics, providing significant insights into the mechanics of urban expansion. This model can help decision-makers make better-informed decisions about sustainable urban growth, infrastructure expansion, and environmental preservation, eventually lead-

ing to more livable and resilient communities. In the subsequent sections, we will explore the technical aspects and practical applications of this groundbreaking approach to detecting urban development.

## 1.1 Aims and Objectives

This research aims to develop a model that utilizes aerial photos to accurately identify changes in a region over both time and space. The model will focus on detecting modifications in land use and cover. It will be trained using an extensive dataset of aerial photos, enabling it to compare and identify changes in land use and cover effectively.

We got the idea of this work from area-based development or changes in BRAC University New Campus, Merul Badda. Recently we know that our university will shift there within Fall semester. As a result when The university shifts there then there will automatically be a lot of changes. So we are thinking we will take aerial images or drone images from that place after 10-15 days back to back. As a result after 2-3 months we can get some changes and development based changes in the University area. As a result, we can identify the changes via detection and as per the goal we will develop a model for checking overall view and changes from the past aerial images. Our model will work together with detection and overall changes of the university area. Such as we can imagine that, that place will be more crowded and as far as we know the main road will be super busier than now because a lot of students will go there and will stay over there. Other Hand, Traffic jam will have occurred highly and there will be an increasing number of shop,building for helping students and business also. So we can see a variety of changes in this place at a same time. So if we can prepare a model with the thoughts of changes/development it will ensure us to know about any particular place's current condition both economic and natural progress with the environment. Furthermore, the model should be able to determine the size and rate at which urban development occurs. This information will be useful for tracking the growth of areas and identifying regions that are susceptible to urban sprawl. To ensure usability it is important that the model is user friendly and accessible to a range of individuals including planners, researchers and policymakers who want to improve any area for future betterment.

This research can also address many objectives in order to develop our model for detecting urban development or changes with any area based on aerial images for example:

- Gaining an understanding of the factors driving development.
- Assessing the social and economic impacts of development.
- Developing strategies for growth
- Contributing to advancements, in monitoring and managing development through technologies.
- Advancement in monitoring and managing development.
- Assessing social and economic impacts of urban planning.

## 1.2 Research Problem

Aerial picture semantic segmentation may be used to identify land use in any location, such as a city, municipality, or nation, without tying it to an official entity. By examining the segmentation data, socioeconomic indices such as population density, road network, and others may be calculated. Such a program, however, only assesses indicators for a single picture capture time. The socioeconomic dynamics should be evaluated by studying a time series of aerial photos before using the indicators in decision-making. The issue arises from the drastically varied quality of aerial photographs collected over time, as aerial image collecting is costly and time-consuming. Image preparation methods must be used to avoid the appearance of biased findings.

The process of urban growth is complicated and dynamic, making it difficult to monitor and follow. It is mostly labor-intensive work to eventually detect the development changes for which aerial imagery becomes a easy solution comparing to that which is why we decided to implement this. Aerial photographs are a valuable data source for detecting urban expansion, but several problems must be overcome to build accurate and effective techniques [5].

One difficulty is the vast number of aerial photos that must be analyzed. Another issue is the variety in image quality, which may be influenced by weather, time of day, and resolution. Furthermore, diverse forms of urban growth, such as new construction, building renovations, and land use changes, can be challenging to differentiate.

For both the environment and society, urban growth carries significant implications and ramifications. Therefore, keeping tabs on and comprehending the growth process are crucial in order to secure its long-term sustenance and prepare for its effects on the economy. Overcoming the challenge of discerning alterations in land use and cover over prolonged periods remains a complex task when it comes to keeping track of urban expansion. This is due to the ceaseless evolution of metropolises, posing a challenge in distinguishing between human-caused and natural changes.

Land use and cover can undergo diverse variations over time, both naturally and due to human activities. Erosion, floods, and deforestation constitute natural shifts, while human-induced transitions may comprise construction of buildings and infrastructure, and expansion of agricultural and urban areas. The environment, society and economy can all be significantly impacted by changes to land use and cover. These effects may be negative, as is the case with urbanization creating traffic congestion, air pollution and loss of wildlife habitat. Agricultural growth can lead to soil erosion, water pollution, and greenhouse gas emissions. Building infrastructure may cause changes in communities, noise pollution and more traffic congestion. For successful implementation of sustainable urban development policies and examination of impacts of urban growth, identification of alterations in land use and coverage over time is imperative.

# Chapter 2

## Literature Review

Urban development is a complicated process that affects facilities, population growth, and the utilization of land. Aerial photographs may be used to track and observe these changes over time, offering insightful information on the spatial and temporal patterns of urban development.

According to the spatial pattern, we have to detect the changes in urban form, expansion as well as density. Aerial images can be used in detecting the development in urban sections in various ways. We can detect so many materials in our surroundings which are related to urban areas including buildings, greenery, humans, traffic, etc. These all can be detected using aerial images. There is a common process to deal with these. We can use image classification to recognize different lands in urban areas such as roads, buildings, gardens, playgrounds, trees, etc. We can do this by taking the help of different technical systems such as supervised & unsupervised classification and also the image analysis based on objects. There is another approach related to spatial classification. That is the change detection algorithm which is used to do the comparison. The thing is actually to compare between two or more images taken from the same area but at different times. According to temporal patterns, once the uplift has been done, that can be observed over time to track how much development has been done in that area and how much needed to be in that. It's mostly similar to the change detection algorithm. Also here, comparison among different aerial images can occur to detect the growth, expansion and density of the urban areas.

Artificial Intelligence (AI) has been an innovative thing in the world, nowadays. This has reached the higher level of technologies. Every work in every section has been made easier by the use of AI to detect the greenery or the green spaces in our surroundings, the use of AI has no alternative. Artificial Intelligence can be utilized in many ways to track and observe the greenery in our surroundings. But it should be used in proper ways which lead to making the proper decision about the outcomes.

According to (Huerta et al., 2021), Urban Green Spaces (UGSs) provide different worships which are very necessary related to the environment for the benefits of ecologies and the societies. Due to global, communal, and financial transformations of urban areas, UGSs pose different challenges for management, specifically in the fast developing metropolitan areas. Using different convolutional neural network encoders on the U-Net architecture and very high resolution (VHR) imagery, this study compares two deep learning model methods for semantic segmentation of

UGS polygons in order to acquire revised data on UGS polygons at the urban level [16].

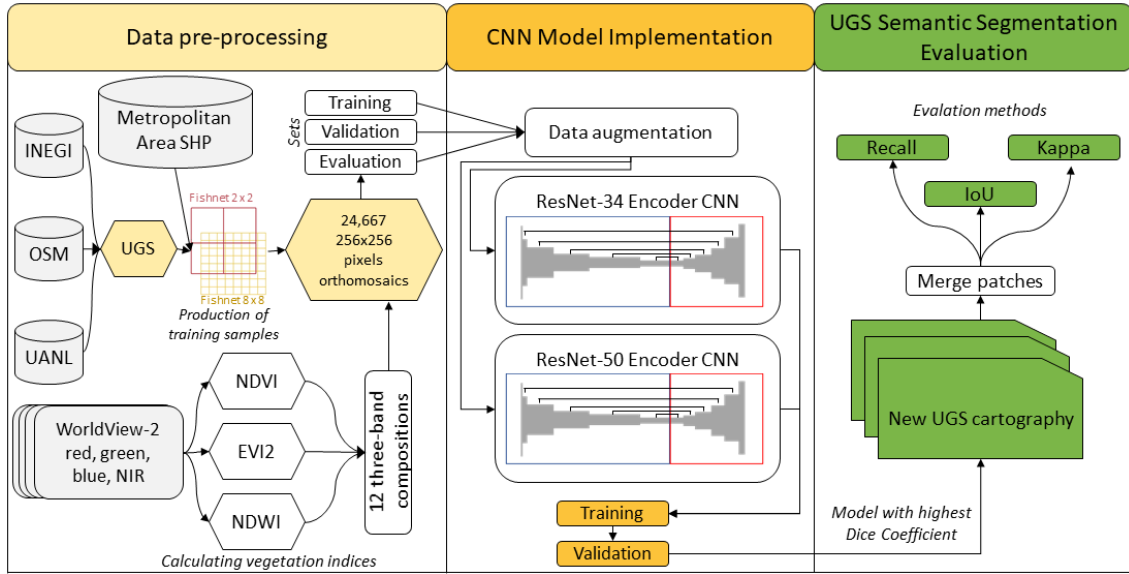


Figure 2.1: Graphical Abstract of UGS segmentation

In urban areas, dense pollution and mostly overcrowding have become one of most major concerns in the progress of modernization. Trees as well as green spaces are hardly important to keep everyone alive and gutsy. According to (Yadav et al., 2019) By using practical approaches, remote sensing technology based on Unmanned Aerial Vehicle (UAV) can effectively regulate and track green ecosystems. The majority of UAV-developed methods, however, are for certain tree species. UAVs are used as IoT nodes in the suggested approach to gather information in real time. The method or algorithm is very efficient to use. It performs the detection of green areas and can work for almost all types of trees [14].

The immensely unprivileged precision for detecting short items which are densely separated throughout a picture and the failings of semantic facts are the two largest technical hurdles facing aerial image to detect objects at the moment. According to (Hu et al., 2023) Efficient-Lightweight YOLO, they design and scrutinize three model architectures to set the focus of the models to detect small objects as well as in identifying the most effective network structure. They also used a design of efficient spatial pyramid pooling (ESPP) to augment the representation of features of small objects in aerial images. In addition they used Alpha Intersection over Union to deal with the balance issues between the positive and negative samples in aerial images [20].

Change detection algorithm as we discussed a little bit before plays an important role in remote sensing data. In changing remote sensing data, the approach related to unsupervised learning methods is very crucial. Yadav et al. (2019) proposed a supervised gaining knowledge of-based totally algorithm for greenery location detection the use of unmanned aerial vehicles. Change detection methods have been applied to satellite or aerial images which are taken within a certain period of time. According to Celik, a proposal was made for applying Principal Component Analysis (PCA)

and C-fuzzy to identify the changes between the two photos of the landsat database in 2007 and 2011 [13]. Xue et al.(2021) proposed a new unsupervised alternate detection method based totally on a Gaussian mixture version guided through saliency enhancement.As per C-fuzzy , the clusterization was done for vectors which are claimed from squares and blocks that are non overlapping of different images taken after PCA.The vectors were divided into two clusters and represented obstacles of image with or without change.In the proposed method which was made by Celik, the feature maps of Convolutional Neural Network(CNN) propagate the difference of two temporal images.This can be utilized to segment the input images semantically : The multiscale and multiresolution Gaussian-mixture-model guided by Saliency Enhancement [17].

In the picture [12] below, we can define the differences easily :

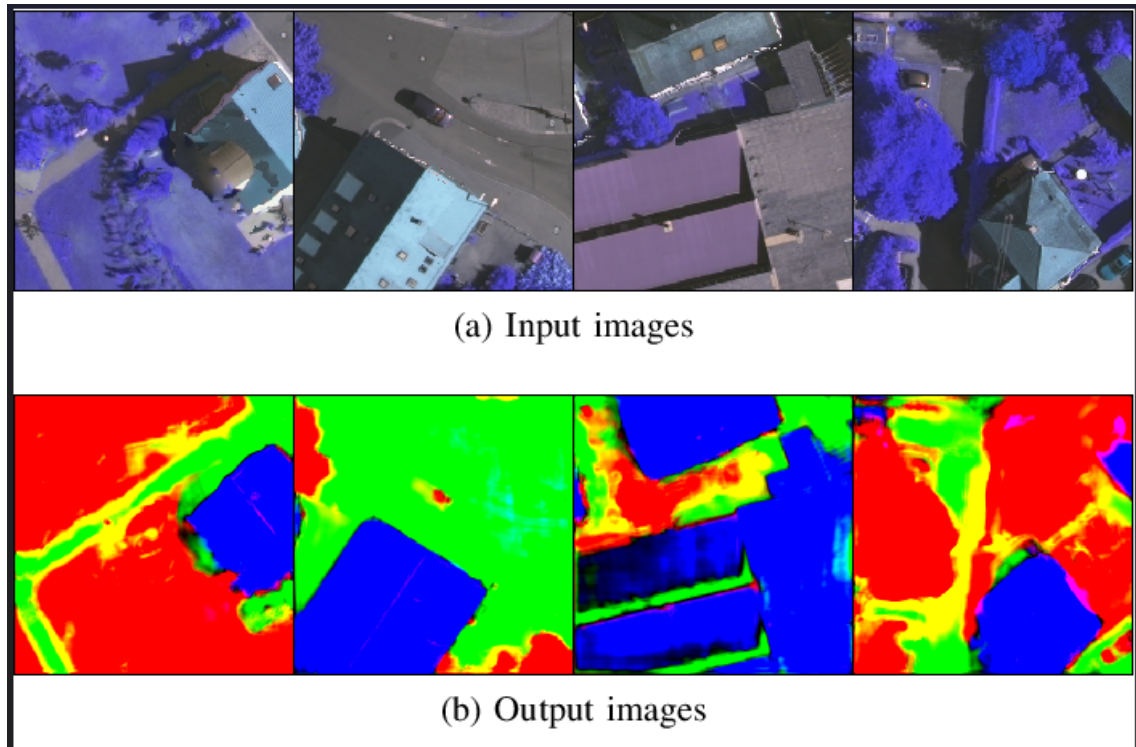


Figure 2.2: Example input images and the corresponding output images produced by the trained model

Here in the Figure 2.2, we can see the example input images and the corresponding output images produced by the trained model. There is another way for this detection. According to (Zhang, 2021), their work is based on observational learning and focuses on features to be searched in remote sensing data such as roads, buildings and trees change trees change detection based on deep learning is divided into two categories from aerial images: fusion late and early fusion [18].

In terms of building detection in aerial images can use an approach of a seeded region growing algorithm, photometric and geometric features, and numerical classification

to differentiate between building and non-building classes which was tested on a set of IKONOS images with a spatial resolution of 1.0m, and the results look comparably good to those of the tested aerial images [4]. However, By using invariant color features, color characteristics of objects can be extracted without being impacted by changes in the imaging environment, such as lighting, surface characteristics, highlights or shadows, or changes in the viewing angle. Because the computer can now concentrate on the color characteristics of the rooftops and the shadows, which are key indications for identifying buildings in aerial photos, the accuracy of building detection is increased. In that method they have used edge detection using boxes which eases the way of detecting any building of an irregular shape [6]. In contrast to conventional pixel-level classification methods, the CNN-based methodology suggested in this research is an object-level detection system. In contrast to semi-automatic methods, the study's end findings are the locations of buildings enclosed by bounding boxes and the kinds of roof shapes. Additionally, this technology is more precise and effective than conventional ones, and it may be used to recreate 3D buildings utilizing model-based methodologies [11].



Figure 2.3: Semantic segmentation

[19]

In recent times, vehicle detection through aerial image has become a crucial topic in surveillance, to observe the traffic in the roads. As per the addition of the authors in “Methods of analyzing traffic imagery collected from aerial platforms”, the process for collecting info and processing images that are utilized in the experiments monitoring traffic behavior. By the use of cameras attached to a helicopter, anyone can claim the imagery from the aerial platform. GPS and IMU are also required to imagery geo-referencing. Image processing is mainly used to search for and follow moving cars in real-time, taking into track the movement of the camera platform



which are needed. A process has been developed which is essential to calculate the traffic metrics [2]. It explores huge numbers of strategies to detect vehicles and traffic of the roads in aerial images. Support Vector Machine(SVM), Haar-like features, and deep neural networks are some methods which are executed to complete the detection. Explorations with integral channel features on the pictures from Munich are presented, along with reports about the channels that made up the landmarks and the process of the precision being affected by them and also the tests are recalled: Fast Multiclass Vehicle Detection on Aerial Images [10].



Figure 2.4: Fast multiclass vehicle detection

[15]

The first step of the process is to analyze the circumstance of the images at different close levels and remove it. Then a group of overlapping windows would be created. By calculating the mean amount of the width, height, and position of the center of every overlapping rectangle, the centroid rectangle of the group of windows can be obtained. The other rectangles are also modded executing the same process as before. The process continues until at sequent iterations, the center, width, and height of the norm of the centroid rectangle fall below a certain threshold: Vehicle detection from aerial imagery [7].



# Chapter 3

## Prediction Modeling using Machine Learning

### 3.1 Working Plan

The sole purpose of this model is to determine the changes that have happened over a period of time in a certain area which will eventually allow us to learn about the development of that area. From that information, we can eventually reach decisions like observing economical change, growth in population, traffic congestion and number of vehicles increased or decreased resulting in the ultimate change that has happened in that place. For that, the model needs a process which will collect data from the drone or satellite images, pre-processes and splits the data which is being analyzed properly and returns a proper decision based on the development that has happened. Subsequently, in figure 1, the model design is depicted at a significant stage. Our motive is to start working with the collection of data and after that we will work on our dataset making and collecting. As a result we can achieve a perfect dataset with aerial images for our model training and testing. Data preprocessing is also an important part for making a dataset perfect and we have to work in this part for a better dataset for train, testing. After that we will extract our specific feature for our model. After this we will split our data for training and testing and for this training part we can choose 80% of dataset and 20% for testing. After the training part of our model we can test our data which is not used for training or we can check with a random aerial image for checking the model accuracy and output. For the rest of the work of our model we will try to make a best model with high accuracy. When we will get high and a perfect accuracy then we will go to the development part and model will make decision that area is developed or not, With the development basis we can see some other things also such as economic condition of that area because when a great percentage of area is developed that means that area is economically developed and for this we can think about some factor also. After the successful model creation we will decide factors which can give us a perfect success ratio for our entire model.

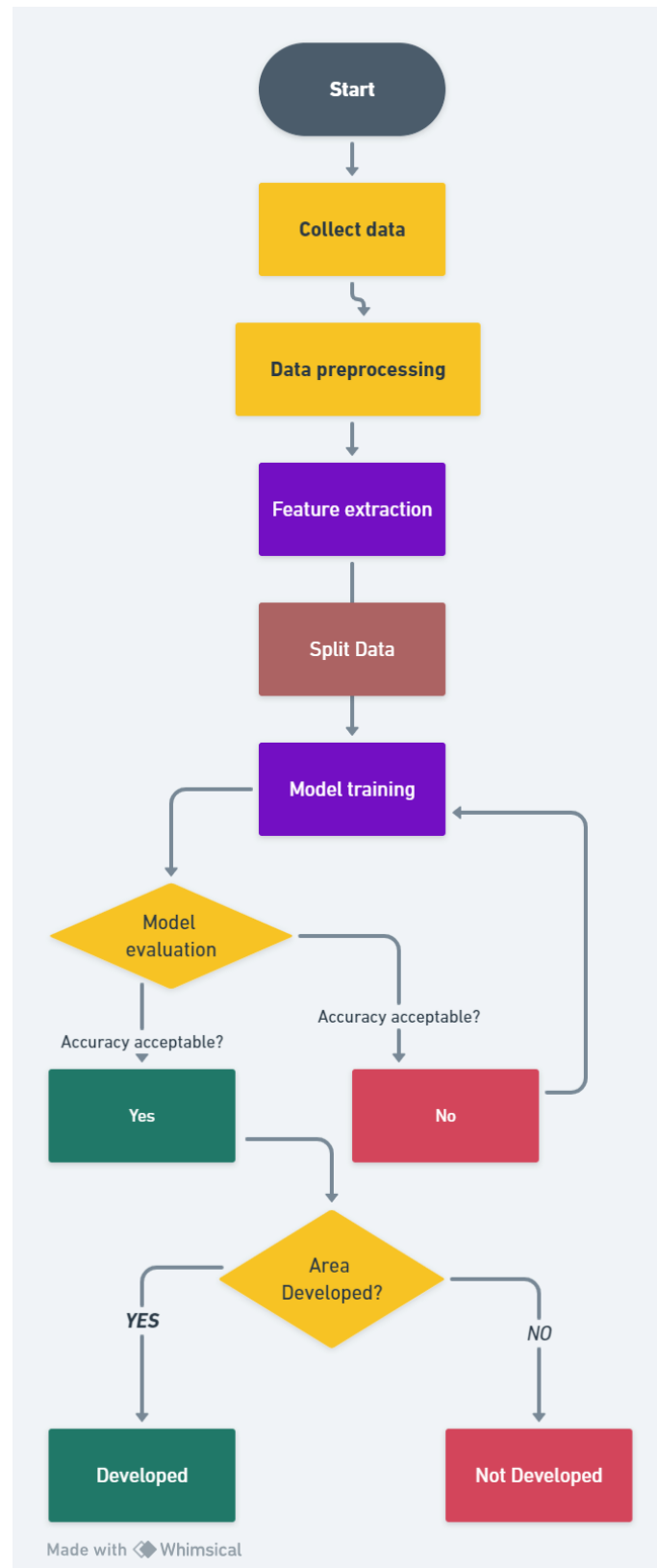


Figure 3.1: The flow chart of the proposed detection model

Aeronautical images captured by drone datasets can be used to really observe urban development and area-based changes. This approach enables a spatial and common evaluation of changes in metropolitan scenes. It is feasible to distinguish after some time examples and changes in foundation, land use, and improvement by using state of the art picture handling techniques like PC vision and AI. Drone symbolism's successive and high-goal information assortment empowers exact metropolitan elements observing, helping city organizers, policymakers, and analysts in settling on informed choices for reasonable metropolitan development and advancement. Here are some major stages related to this research:

- **Data Acquisition** : Drones are utilized to capture aerial images that provide high-resolution views of urban areas. This stage involves planned drone missions to cover the designated regions, ensuring comprehensive coverage for detailed analysis. The drone captures overlapping images from multiple angles and positions, enabling the reconstruction of 3D models and the detection of various objects and features.
- **Image Preprocessing** : Before they are analyzed, raw images are preprocessed in the case of drone-based images. Try to improve what is already there and probably correct these distortions that are likely to have occurred. Techniques like orthorectification and radiometric correction are applied that makes right spatial representation the reference of the element. over an extended period. This step is particularly important for the detection of objects from the given image as it helps in formulating an effective model detection and feature extraction.
- **Feature Extraction and Analysis** : Machine learning models help in finding and isolating different objects and characteristics on the images that are preprocessed. These may include the structures, roads, trees, car and other structures that are characteristic of the zone. There exist algorithms to detect objects in images such as the CNNs or region proposals for object detection. Moreover, image processing algorithms are employed for feature extraction aimed at determining the essential characteristics of the detected items, including the height of buildings, width of roads and so on.
- **Visualization and Interpretation** : The objects and features that are extracted from result are examined with the intention of making comparisons, drawing conclusions concerning particular patterns or trends, and study temporal differences. This analysis gives information on the condition of the cities and such aspects as construction work, infrastructure spatial structure and positioning, urban development, and use of the landscape. The extracted information can be depicted by maps, 3-D models, or any other graphical depiction of the problematic under consideration. In order to give recommendations to stakeholders such as urban planners, and decision makers that can support rational decision-making for sustainable urban creation.

## 3.2 Machine Learning Models

Artificial intelligence is a kind of algorithmic models that allow computers to learn from and predict or make decisions based on data they process. These models are developed to recognize trends and analysis in a larger collection of data and are able to perform tasks such as categorizing, grouping data, and identifying outliers without specifically programming the model for each of these. Unlike conventional models that are solved once and then applied as they are, ML models can get smarter over time as they are exposed to new data by using methods such as neural networks, decision trees, support vector machines, and more. These models are considered pivotal in most of the modern applications and provide a fundamental technology for the implementation of such advanced features as speech recognition, knowledge processing for the image and text, and self-driving of a vehicle. For example ML techniques are applied in healthcare to forecast the patient's health status, describe medical information also to individualize the therapy. Financial application works to detect fraudulent transactions and to automate assessments. Similarly, in the retail industry, these models aid organizations in controlling inventory and improving customer experience through presenting them with relevant recommendations. Usually machine learning models are nowadays a modern improvement in terms of easy completion of severe tasks which include categorization, object detection and so on. In recent times, after the AI sector has bloomed ML models has also come up to the picture for being the backbone of these works. As, we decided to do object detection so these models can be of huge help. Eventually, regarding converting our task to the object detection, we have selected YOLOv5, v7, v8 and one custom model.

### 3.2.1 YOLOv5

The backbone of YOLOv5 is based on CSPDarknet53, which uses CSP connections to enhance the gradient flow and reduce computation. Its neck uses a PANet for its pyramidal feature fusion to improve the feature pyramid information extraction ability. YOLOv5 head comprises of YOLO layers which predict bounding boxes, objectness score and class probability. In the dimension of performance, YOLOv5 is characterized by having a good trade-off between speed and accuracy. For the task of object detection on aerial images, YOLOv5 showed to be an effective model with high speed inference making it a great choice for real-time scenarios. It provided decreased costs for processing through its architecture but it suffered from a small object or a dense group of objects.

## Overview of YOLOv5

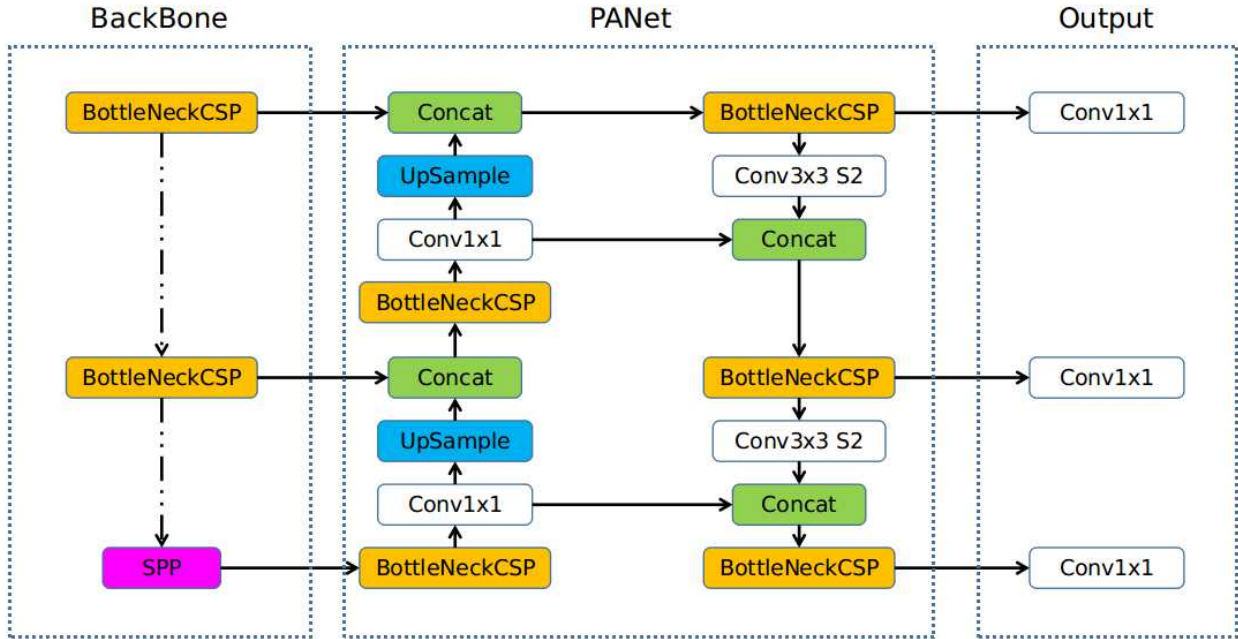


Figure 3.2: YOLO v5 Internal Architecture

### 3.2.2 YOLOv7

The new version of YOLOv7 suggests improving the extraction of features and efficiency in the base network and the CSPNet versions and E-ELAN in the base networks. When specifying the neck part of the architecture that is used in YOLOv7 it should be mentioned that although this part is well developed with regard to the special unit for filtering feature maps of different dimension the head is the semantic YOLO head that was subject to considerable alterations related to the anchor's boxes and the loss's function at the initial stage. Some of the examples are the complexity of images that may have many parts of the image that share the same objects or even objects overlapping over each other. It was of a good compromise in regards to its computational performance and its faith regarding enhancing its detection element and presented a good alternative for those cases where higher accuracy was granted when compared to some reduction in speed.

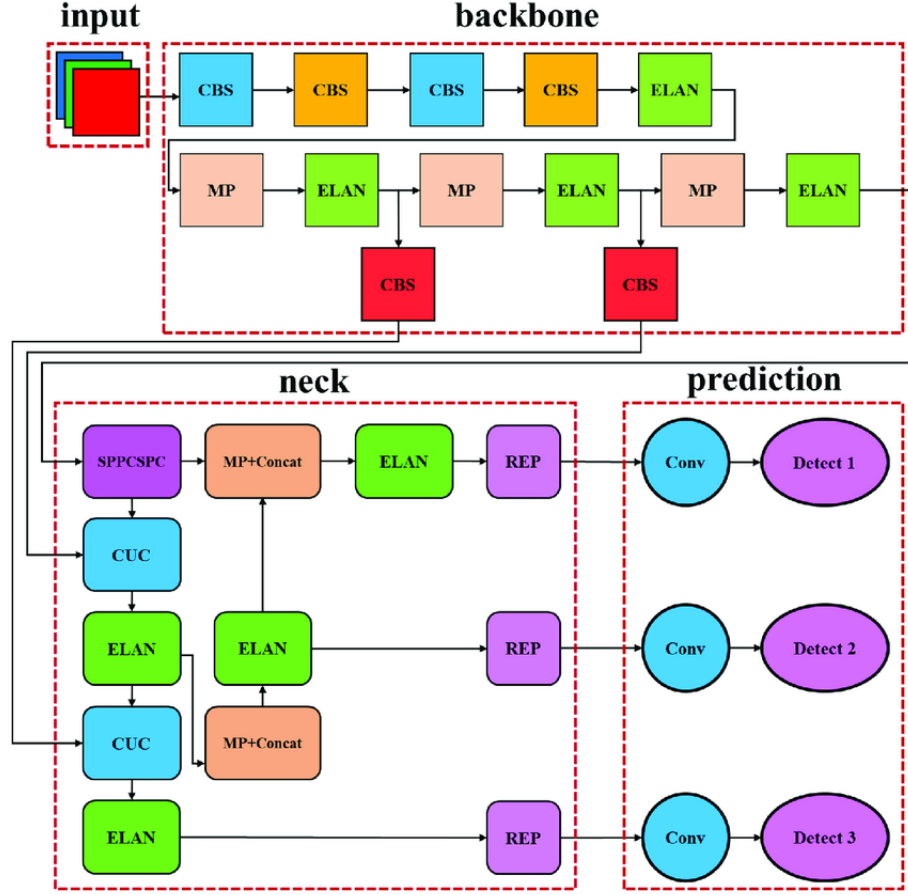


Figure 3.3: YOLO v7 Internal Architecture

### 3.2.3 YOLOv8

CSPNet is utilized as a pre-headed and deeper architecture in YOLO8 for enhancing the feature description. For the neck of the network, its architecture is also improved to be an improved and enhanced PANet to allow for the function of efficient multi-scale feature learning and the head module then follows the concept of the YOLOv5 network that was designed using the recent and best techniques on the design of anchors and how the loss function can be implemented and estimated. It has shown that the new YOLOv8 model worked effectively and efficiently in terms of detection and speed compared to the previous models. The deep model was more suitable in detecting very reduced and compact structures in aerial imagery. The enhanced Feature Extraction and Aggregation Methods used by YOLOv8 were the most effective Methods that resulted in quite higher number of Detections than the other models can boast of.

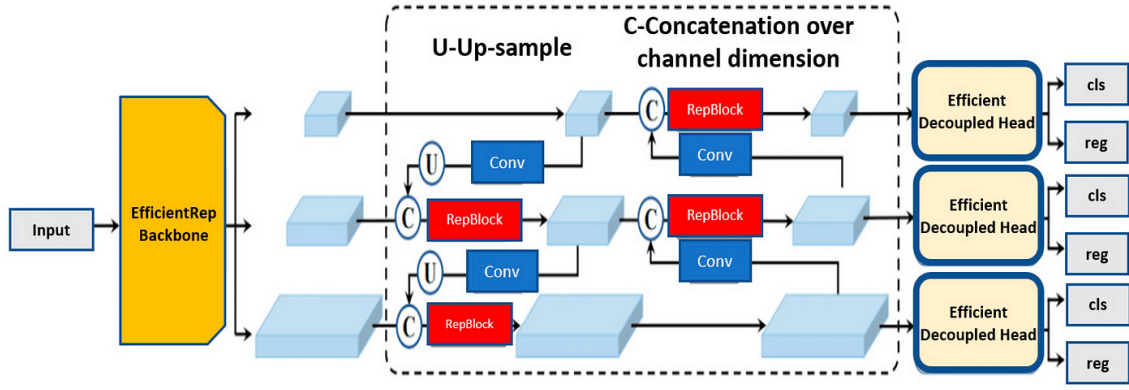


Figure 3.4: YOLO v8 Internal Architecture

### 3.2.4 Custom Model

As a result, we have completed the training of our custom object detection model through Roboflow, an AI-centered platform for training and deploying computer vision models. Therefore, using the strengths of the Roboflow platform to handle and add to datasets, we attained a small enhancement in the model's ability and efficacy. Thus, the training process involved using high spatial resolution aerial images to identify various targets and different categories of objects in the urban environment such as buildings, cars, roads etc. Our custom model based on the Efficient YOLO algorithm provides high accuracy and efficiency, which is especially important for object detection in the urban environment. Besides, this means that this tailored approach does not only improve the detection and assessment of changes, but it also provides significant evidence regarding the planning and building of cities and organisms. This will enable the faster and real-time gathering and analysis of data to enhance decision making hence support sustainable development of most urban cities.

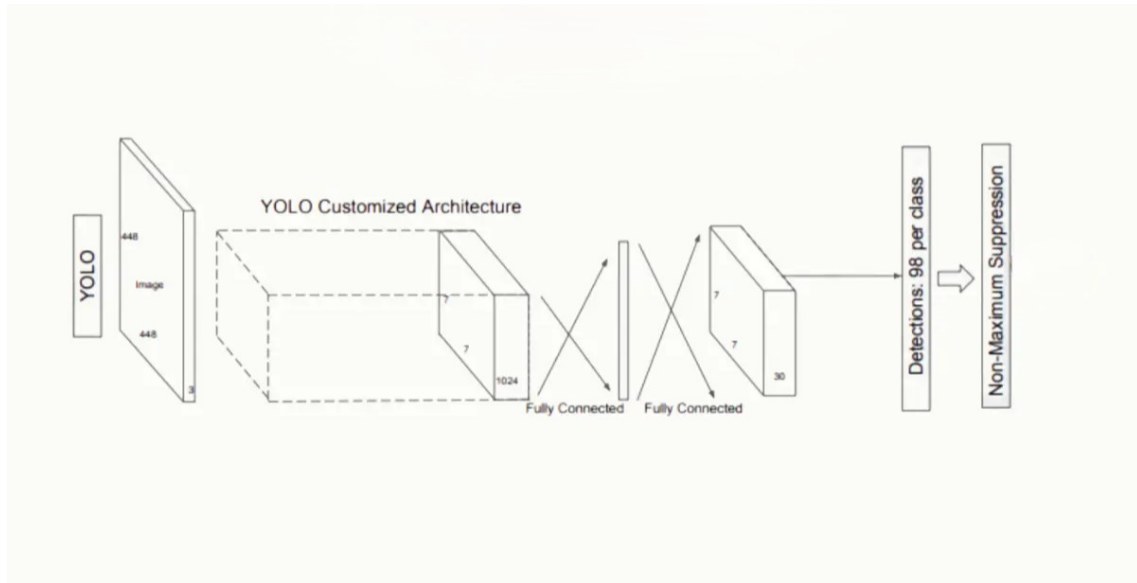


Figure 3.5: Custom Model Internal Architecture

### 3.3 Dataset Collection

During the data collection stage of our study, we carried out a systematic aerial imagery acquisition exercise by precisely shooting images from predetermined positions and latitudes in and around the Merul Badda sub-district of the Dhaka metropolitan area. After the necessary permissions were acquired, the drone were scheduled to maximize the exposure of the area to be photographed and comply with all the regulations. We were mainly focusing on BRAC University campus particularly with its new campus development images were acquiring in a regular manner through our efforts, so that we could able to generate the dataset that reflect the current images as well as show the action going on in the course of construction and expansion of the BRAC University campus area. With the aerial imaging initiatives we meant to make a movie like record covering a wide variety of objects from buildings, landscapes, trees, water bodies, roads, automobiles, persons, individuals with solar panels, tin-shed homes, etc. Our meticulously curated dataset, with a focus on the Merul Badda area and the BRAC University new campus development, captures the evolving landscape and transformations occurring within this region over time. Here is the simple work flow chart given.

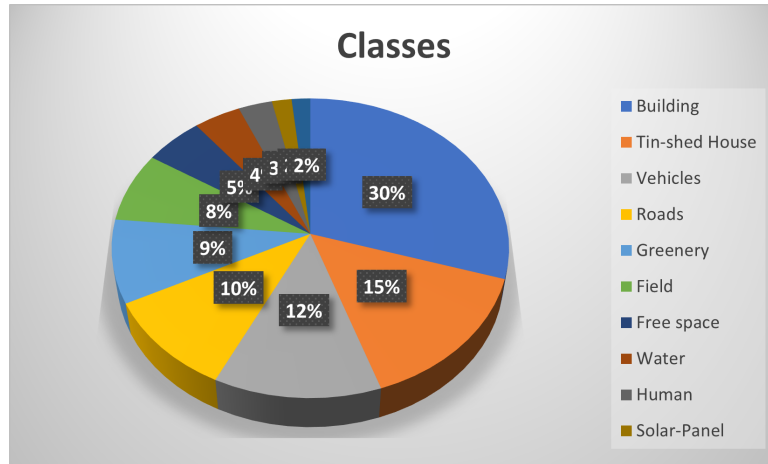


Figure 3.6: Classes

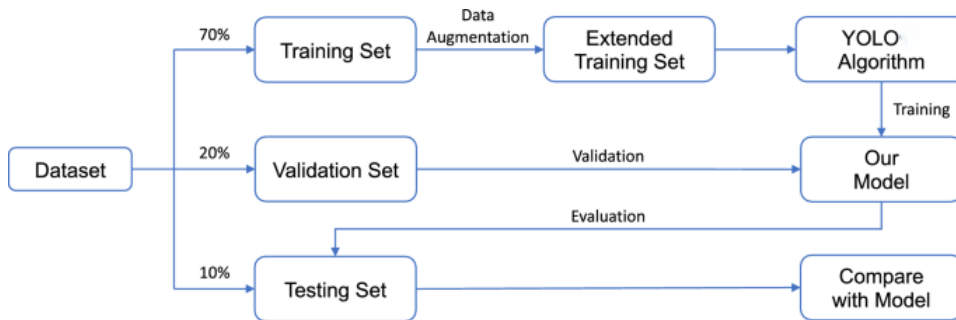


Figure 3.7: Workflow of Dataset



### 3.4 Input Data

Our study involved drone imagery as the essential wellspring of data to catch different scenes in this assessment endeavor. We have used around 400 raw drone images from different views of Brac University New Campus as well as the surroundings. Here is our taken drone pictures : <https://shorturl.at/Xdry6>. To update the common sense of our models, a custom dataset was carefully organized, involving significant solid areas for the cutoff points presented by an online platform Roboflow. Using this platform, we labeled the pictures with the elements present in the pictures. Then we annotated them to find the percentages of each element in those areas. This platform worked with the consistent naming and explanation of the obtained drone pictures, guaranteeing accuracy and accuracy in the plan cycle. Here is the link from where the dataset can be found: <https://app.roboflow.com/aerial-image-3> Here are some pictures which we have used to build our desired dataset :



Figure 3.8: Input Pictures from Drone

A comparison of different deep learning models, including CNN architectures, YOLO v7, YOLO v5, and YOLO v8, was then carried out. Utilizing each of these models' distinct strengths and skills, we were able to fully utilize the potential of our own dataset.

Applying YOLO v7, v5, and v8 demonstrated how important it is to experiment with various versions of the YOLO framework. We evaluated these models, which are well-known for their speed and accuracy in real-time object detection, using our own dataset. This evaluation provided us with valuable feedback about how well the models were adapted and what kind of flexibility the model had in respect to the uncommon features of the camera footage we received from our drone.

Moreover, the introduction of CNN models into our experiments helped us to broaden the scope of this work and conduct a more complete analysis of the effectiveness of

deep learning architectures . We were able to uncover differences in the model’s performance through this general approach, which also assisted in highlighting specifically what makes our dataset different and difficult to process.

The beginning of these endeavors can state as an energetic final result for the perception of the connection between drone symbolism, creation of custom dataset, and the application of various profound learning models. This comprehensive study not only contributes to the advancement of item location methodologies but also has important empirical implications for wider fields such as computer vision, especially in the context of deep learning for drone applications.

### 3.5 Annotation

The process of annotation is a very significant milestone in the acquisition of raw aerial imagery to develop machine learning models like YOLO V5, V7, or V8. Annotation also relates to deeming an image and in assigning different tags to the objects in question to the model for identifying and understanding the said objects and features to it. It uses an annotated dataset to later learn to make predictions at the precise value.

We applied the Roboflow platform which is an effective online image annotation tool in our case study. This platform provided an easy to use interface and a set of tools that made it easy for the annotator to mark or label items in an aerial photo. Here we did manual annotation of 50,000 (approx.). Roboflow was used to pre-process the raw aerial photos taken by the drones, annotated the important objects in every picture by giving certain attributes to every object. All of this can be any of the application relevant units such as buildings, roads, ponds, or any other characteristics that are essential for the model to work. Here are some annotated pictures which we have created in Roboflow :



Figure 3.9: Manual Annotation

The interesting objects in every picture had to be tracked by the annotator and labeled. The model will be diligently informed of all the colors and features of the annotated items in terms of visual information. The annotations became meaningful since they reflected the real-world settings in their complexity and diversity and thus contributed to the improvement of the generalization capacity and the ability to make predictions for diverse cases.

Furthermore, the annotations contained useful metadata such as the object coordinates in the bounding box, information on the category of the item, and other data that was crucial for the YOLO model training. The bounding boxes helped the model to learn and identify objects by correctly describing their relationship to the other objects within the specified area of the image. We followed the same process for annotating our entire dataset by using the Roboflow platform to create the same annotation mechanism across all images. This is important because it removes the noise element and is effective in helping the model gain knowledge on patterns.

It is possible to say that annotation stage of our study was necessary to prepare unprocessed aerial photos to assign the YOLO model for training. Roboflow's platform is the creation of a primary dataset of extremely accurate and precise annotation that provides the models with the ability to accurately and precisely categorize and predict products in aerial imagery.

## 3.6 Data Pre-processing

Data in any field must undergo data pre-treatment to qualify it for the analysis; this pre-treatment also involves preparing and cleaning the data that will be later analysed. Pre-processing of the data was therefore one of the major contributions towards the improvement of quality of the dataset and its consistency towards our stated thesis. In this part, the precise procedures used to pre-process the photos from the dataset such as object isolation, image scaling, auto-orientation, and data filtering are described below :

**Object Isolation :** One of the methods of object isolation that we adopted was in connecting objects of importance isolating the objects for analysis. In the process of doing this we had to search and delete certain items within the given pictures in order for our study to concentrate on the relevant aspects. Aside from simplifying the dataset, object isolation also ensured that the analysis of the research variables is more focused and pertinent to a wider scope. The particular phase turned out to be in many ways beneficial especially in times when the photos consisted of several items or subjects for the analysis as it would assure that the attention would be on the specific elements that were being analyzed. In the figures 2,3 and 4 below, we can see some examples of pictures as an isolated object :



Figure 3.10: Picture of isolated building

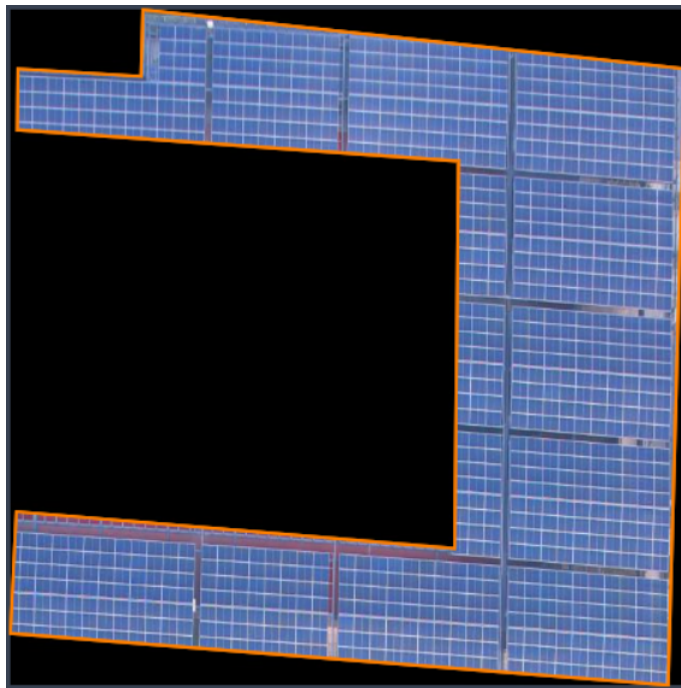


Figure 3.11: Picture of isolated solar panel



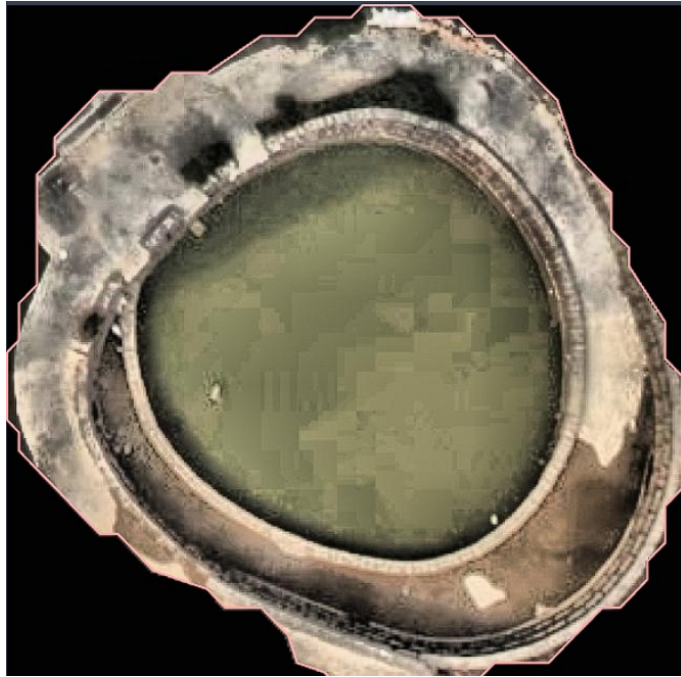


Figure 3.12: Picture of isolated pond



Figure 3.13: Picture of isolated vehicles

**Image scaling :** For the same reason the dimensions of their images need also to be of an equal size. For this reason, we had to recede or resize all of our photographs to 640 x 640 pixels because we needed to compare photographs. Moreover, this standardization aided in getting rid of differences such as dimension of the images used and differences that are attributable to the selected analytical models and led to easier computation processes for further research endeavors. The images were larger and this made the study better as this provided a better analytical technique. Here is an example of our resized picture below :

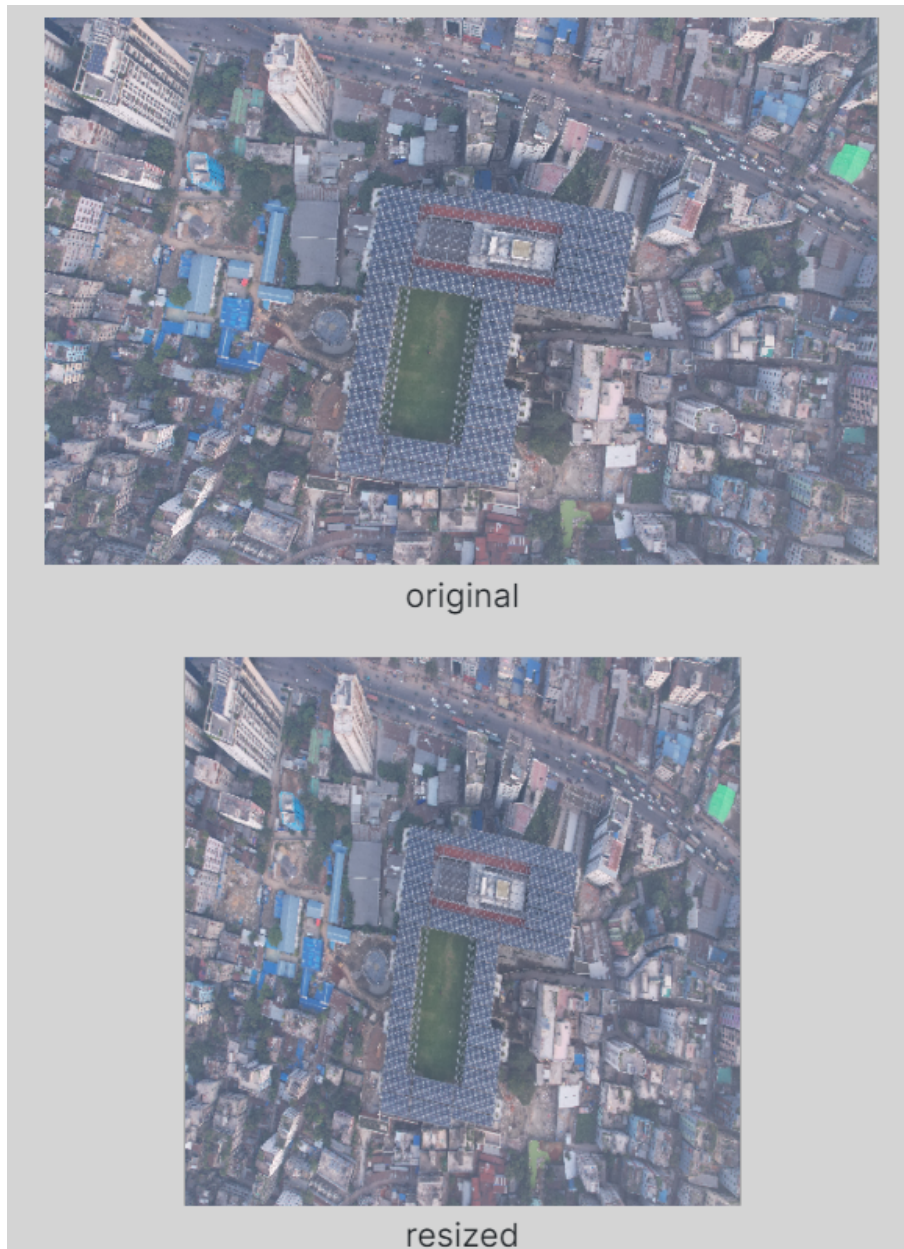


Figure 3.14: Original & resized image after scalling

**Auto-orientation :** Data pre-processing entailed some of the earliest manipulations of auto-orienting photos, where a photo would automatically be in a certain up or side wise orientation in a given photograph. Another example of such factors is the camera angles or the capturing techniques that might end up being distorted enough resulting in irregular orientation. The method that we deployed to achieve this is auto-orientation technique which assists in changing the directions of the image to a desired format. This phase was needed for the standardisation of the characteristics across all the photos as well as when trying to remove potential bias from the data set. Here is an example of our automatically oriented picture below :

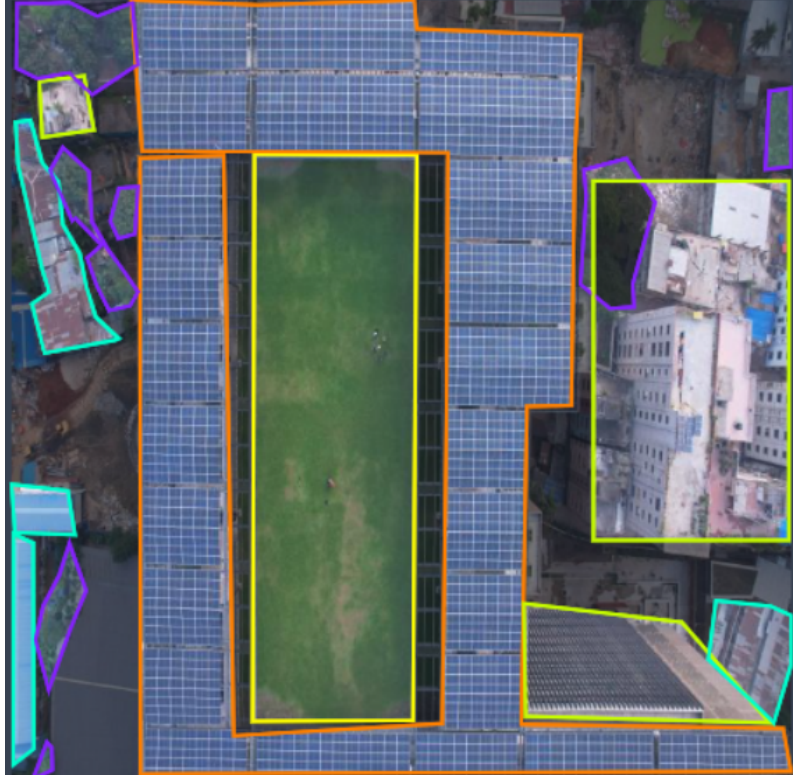


Figure 3.15: Auto-oriented image

**Data filtering :** One of the other concerns that affect the strength of research findings is the quality of the data that is introduced into a study. Some data was removed from the dataset by filtering out null or unneeded information. This step entailed the exclusion of the examples with either non-given relevance to the study objective or lack of critical information. We reduced the possibility of drawing incorrect conclusions by eliminating such data, improving the dataset's signal-to-noise ratio. By making sure that only relevant information was included in the final dataset, data filtering also helped to provide a more accurate and focused analysis. Our research's success was largely dependent on the data pre-processing procedures mentioned above. We prepared the raw dataset for rigorous analysis by auto-orienting photos, isolating important items, shrinking photographs, and filtering data. These pre-processing steps not only raised the dataset's general quality but also established the framework for precise and trustworthy study findings.

## 3.7 Augmentation

Augmentation is a technique for the most part used in the field of computer based intelligence, particularly in PC vision tasks, to work on the assortment of a dataset by applying various changes to the data. This capitalizes on the model's power and speculation. We have got around 5 outputs per training example with vertical flipping and the saturation in between -51% and +51% from the augmentations. These parameters play a crucial role in enhancing the overall performance of a machine learning model. Actually during the training process, each raw training picture is augmented to create 5 different variations. These variations are originated by engaging miscellaneous conversion to the input data, providing the model with a huge amount of examples to gain from. The expanded number of training cases adds to all the more likely model presentation and speculation. Vertical flipping is one of the augmentation techniques used to mimic object orientation variations. The model gains the capacity to identify objects from various angles by flipping images vertically, which helps it to generalize across many orientations. Representing another crucial augmentation element, Saturation level ranges from -51% to +51% is adjustable. Color changes substitute is the augmentation of pictures that the model can control in the case of colors, various lights, and their combinations. This helps avoid a situation where the model will be tuned towards certain lighting effects that were present during training and yet this is not what the practitioner will expect when using the model. It is a malicious condition in which a model gives a 100% performance on the training data but shows 0% performance on the new data that is not trained. Augmentation helps avoid this. More realistic situations are fed into the model by generating a variety of variants of original training samples; this reduces the likelihood of the model being trained to pick certain specific patterns in the dataset. To end generalize model it is necessary to address the saturation issue and provide multiple outputs per example. We can call a model general when it predicts well on data that it has never met before. As a result by using augmentation the model is bound to gain a large number of features which then translates to its ability to accommodate the minute details that might be involved in real world applications. Augmenting the samples would have been a good approach because we could easily increase the training set without having to collect another set of labeled samples. This is particularly important as it is challenging to obtain a sizable and diverse dataset in some cases. The model is able to pick instances that are non-trivial and use the improved instances in order to be able to notice patterns from it which will also be useful in a myriad of real world data sets. These augmentation parameters not only promote diversity but also help the model to resist over fitting and make it better in generalization. In other words, due to this, the model becomes flexible for the changes in the input data and simultaneously develops the capacity to predict numerous hypothetical outcomes in the real-world contexts.



# Chapter 4

## Result and Discussion

In our research, we have run with three different object detection models: YOLO v5, YOLO v7, YOLO v8 and custom models to compare their performances. A unique dataset, which was developed for our application, was used to train the models. The goal of the tests we conducted was to assess the overall quality of each model and its ability to detect and recognize objects in images and estimate their locations. These outcomes were very beneficial in identifying the applicability of each model for the intended purpose because there were just insignificant variations between the values of relative measures such as precision, recall, and the F1 measure. It also made it easier to compare the computational efficiency of each model since it quantified the accuracy given up per unit of computational speed for each model. These results lead to more opportunities in the area of computer vision as the object detection algorithms are examined and optimization of these algorithms takes place for utilization.

For this task, we used the images provided to the drone directly and created a new aerial image set. Having set the environment for identifying objects, we proceeded to upload the images to Roboflow, an online platform where we annotated the images. We even categorized them into buildings, trees, water bodies, roads, vehicles, humans and the solar panels, they also posed to be taking pictures which we later on edited. In this case, YOLO v5, YOLO v7, and YOLO v8 architectures were deployed for the specific dataset that was generated after annotation on our end. We also trained a custom model on the task with the use of Roboflow. This section outlines the method for the proposed YOLO & Custom Models-based aerial picture object recognition strategy. The objective of this solution was to detect a range of targets from fields, buildings, trees, water bodies, roads, cars, and even people from basic drone photos. An annotated dataset was created by using Roboflow and the chosen model were trained on the Roboflow software. Then, the performance of all the latest versions including YOLOv5, YOLOv7, and YOLOv8 was tested.

Firstly, we gathered the low-level drone data that includes various contexts and environments we are interested in. The photographs were then labeled using the Roboflow platform using bounding boxes with the names of different classes like buildings, trees, water bodies, roads, cars, people, and solar panels. These raw drone images were used in creating a custom dataset for this analysis. To ensure that the model could distinguish and classify different classes, each object of interest

in photos was properly labeled. We ran a custom model along with YOLO v5, v7, and v8 models using the custom dataset. We obtained metrics including mAP, precision, recall, and train-loss graphs. We also generated graphs of precision-recall, confusion matrices, and F1 scores by implementing the custom model and the YOLO models.

Later, We provided arbitrary aerial photos to the models in order to further verify their performance, and they demonstrated notable detecting abilities. In precisely recognizing and categorizing objects in a range of aerial circumstances, this test proved the dependability and durability of the YOLO v5, v7, and v8 models in addition to the custom model. The graphs along with the results are given bellow for better understanding of which model performed better for our custom dataset.

## 4.1 YOLO V5

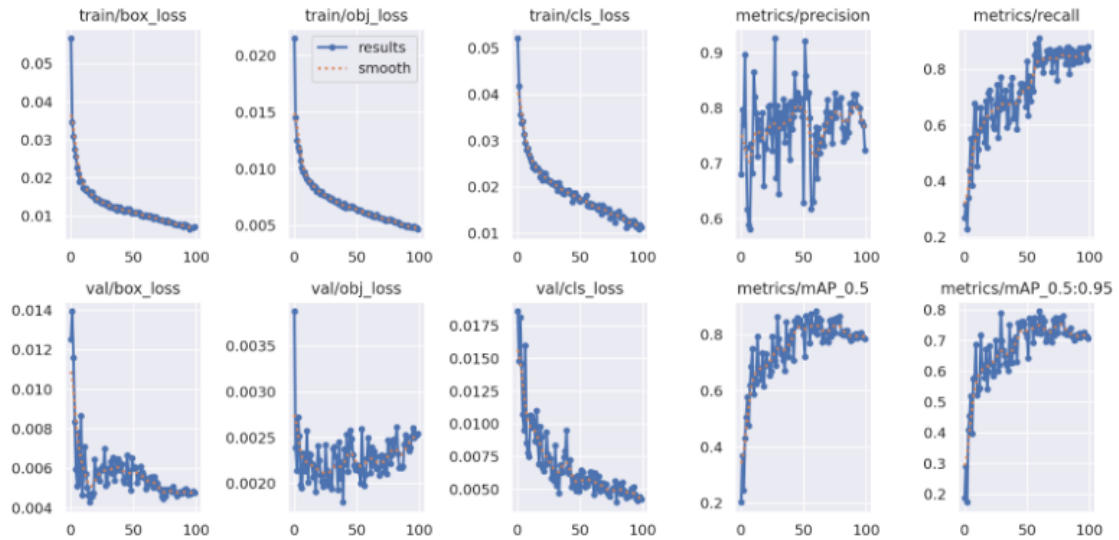


Figure 4.1: Map, precision and recall metrics & Train-loss Graphs from V5

The graphic above shows us a number of graphs that show the metrics our model uses for training and validation. Distinct features of the model's performance across the training epochs are depicted in each plot. Typically, the values of the corresponding metrics are displayed on the y-axis, while the x-axis indicates the number of epochs. The model is apparently learning and getting better over time, as evidenced by the notable decreasing trend in the loss functions (training and validation loss)

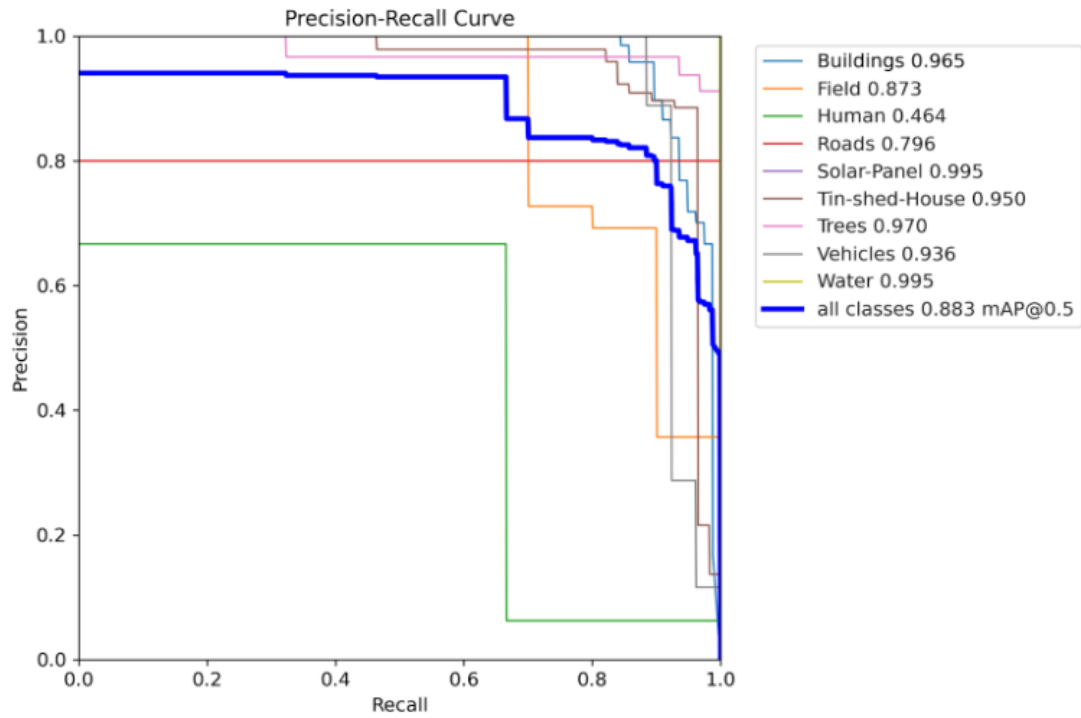


Figure 4.2: Precision-Recall Curve from V5

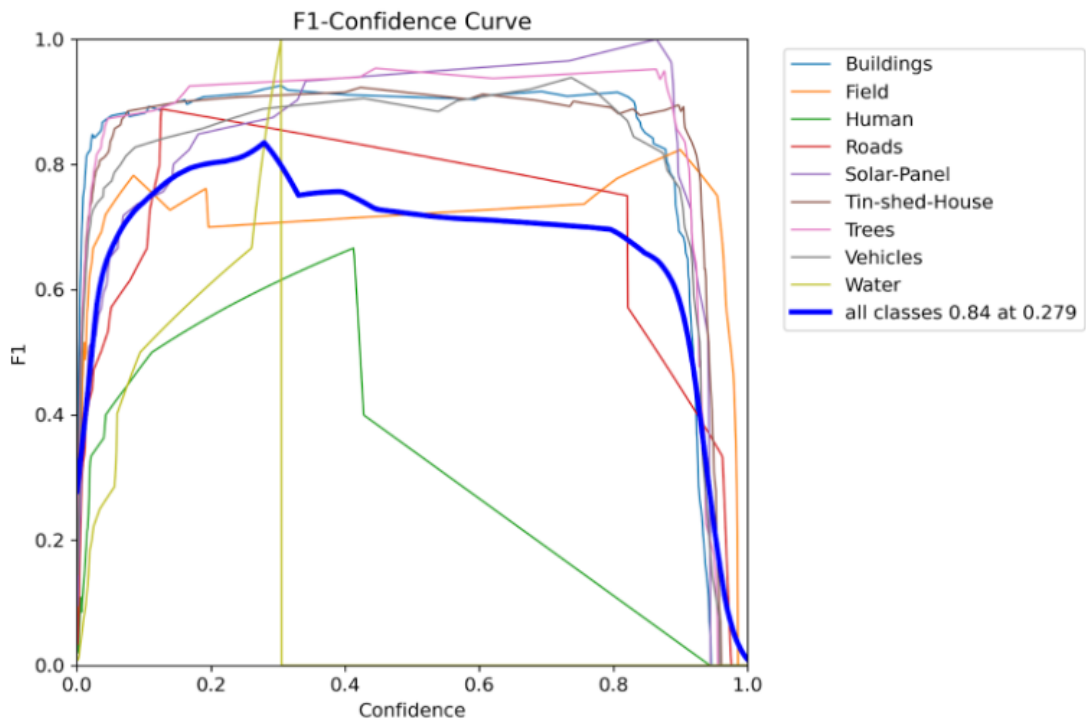


Figure 4.3: F1-Confidence Curve from V5

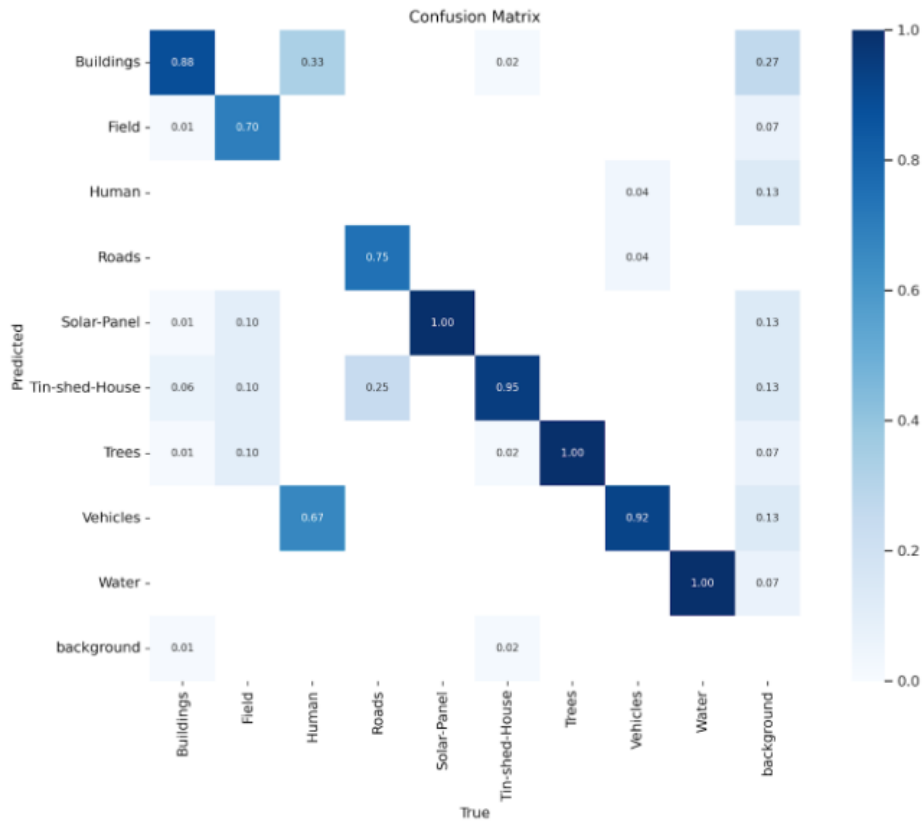


Figure 4.4: Confusion Metrics from V5

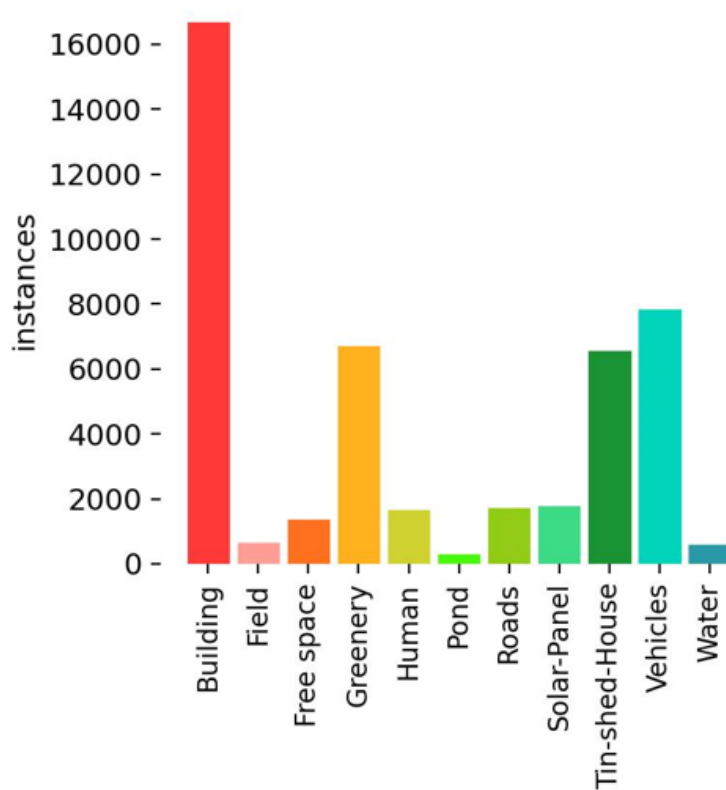
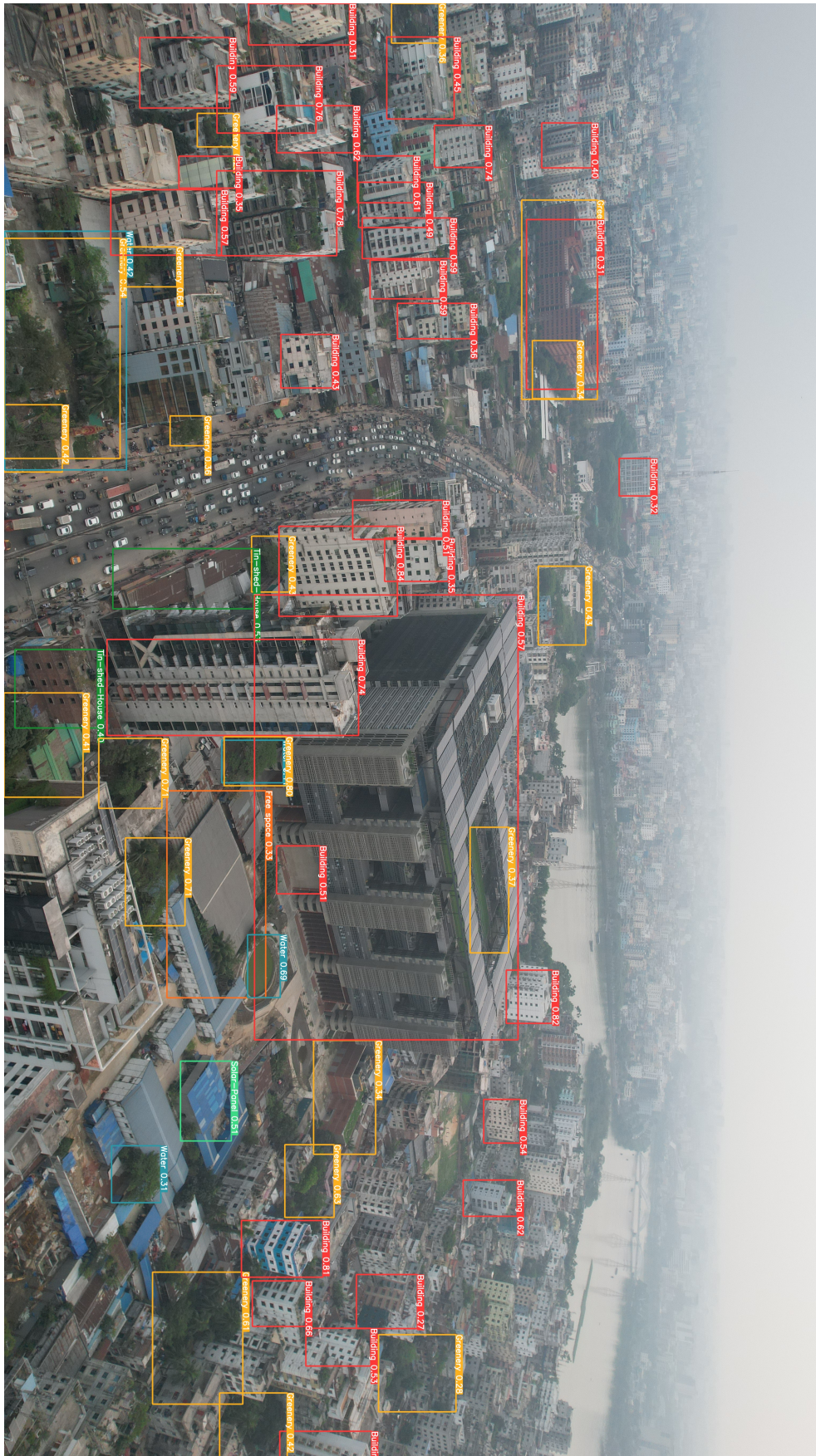


Figure 4.5: V5 Instance





## 4.2 YOLO V7

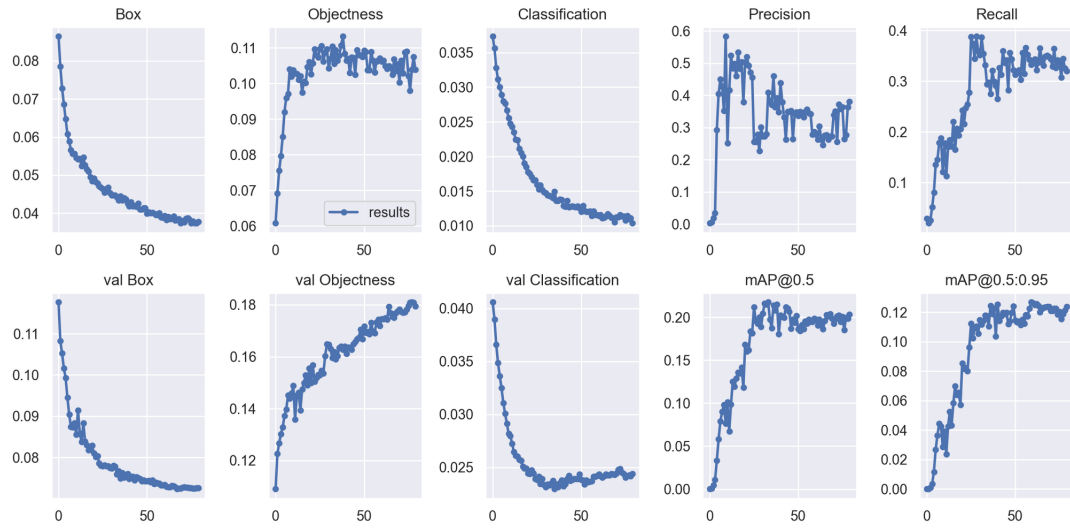


Figure 4.7: Map, precision and recall metrics & Train-loss Graphs from V7

The graphic above shows us a number of graphs that show the metrics our model uses for training and validation. Distinct features of the model's performance across the training epochs are depicted in each plot. Typically, the values of the corresponding metrics are displayed on the y-axis, while the x-axis indicates the number of epochs. The model is apparently learning and getting better over time, as evidenced by the notable decreasing trend in the loss functions (training and validation loss)

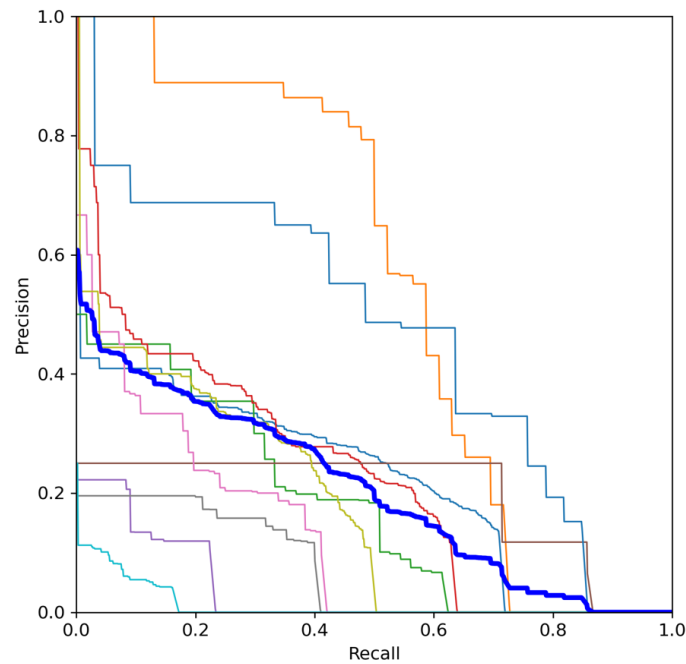


Figure 4.8: Precision-Recall Curve from V7

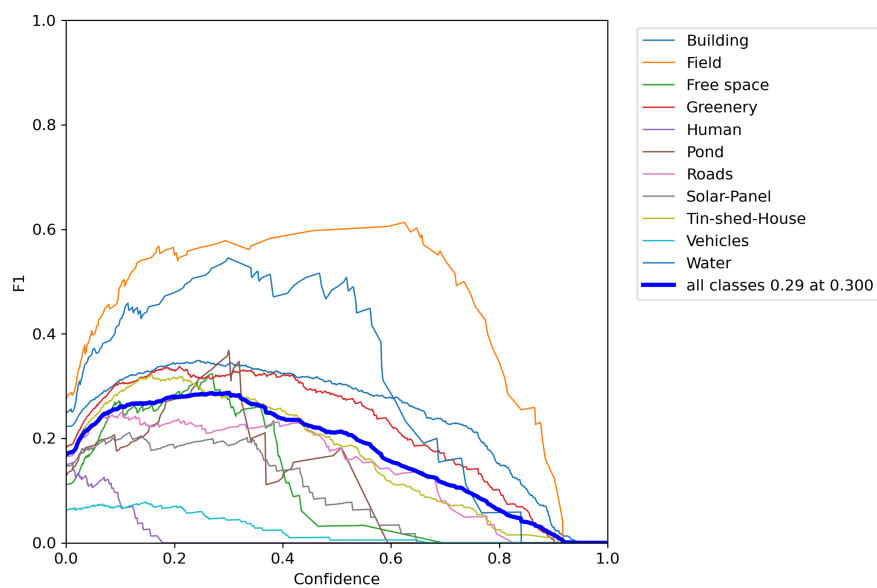


Figure 4.9: F1-Confidence Curve from V7

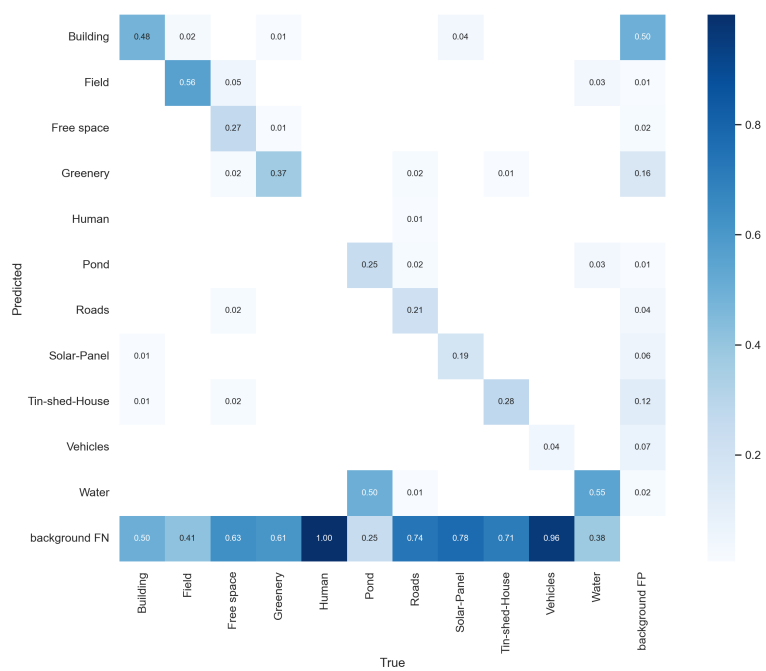


Figure 4.10: Confusion Metrics from V7





Figure 4.11: V7 Custom Image Prediction Result

### 4.3 YOLO V8

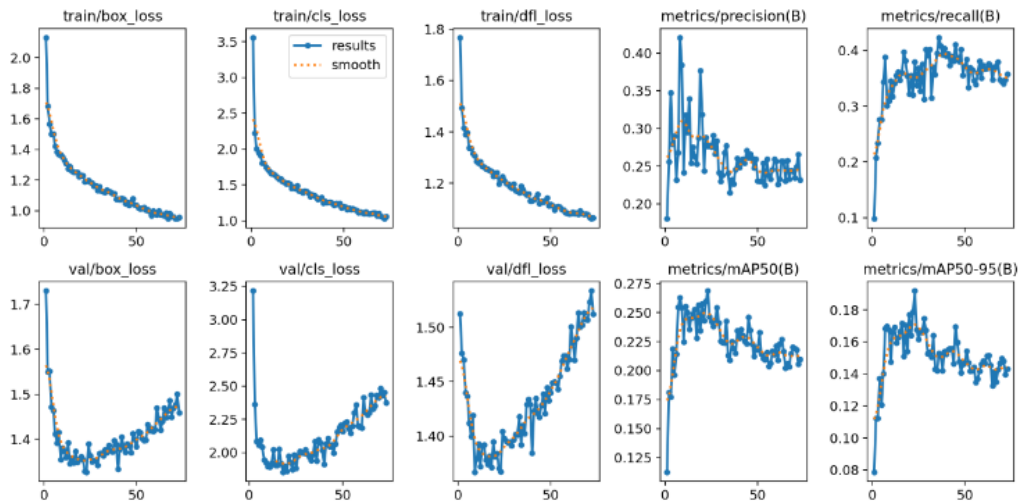


Figure 4.12: Map, precision and recall metrics & Train-loss Graphs from V8

From the above figure we can see different graphs explaining the training as well as validation of our model. The horizontal axes are training epochs and the vertical axes are accuracies, losses, cross-entropy, precision, and recall from each plot. On the x-axis, there are often the number of epochs or training iterations, and on the y-axis, there can be the values of the corresponding metrics. It can be noted that the loss functions (training and validation loss) decrease over the epoch iterations, hence showing an improvement on the training procedure of the model.



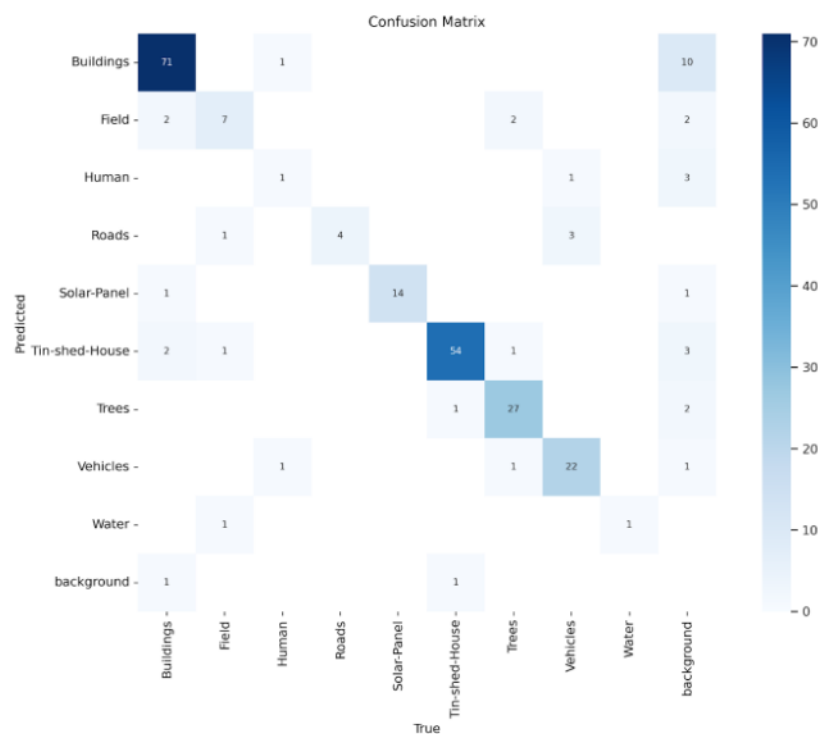


Figure 4.13: Confusion Metrics from V8



Figure 4.14: Predicted Output from V8

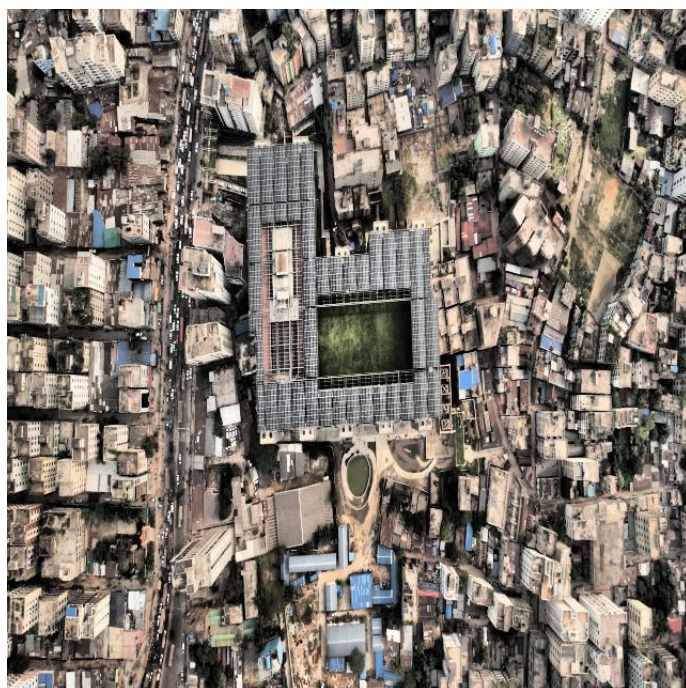


Figure 4.15: Custom Image for Detection in v8 model



Figure 4.16: Result for v8 using Custom Image

## 4.4 Custom Model

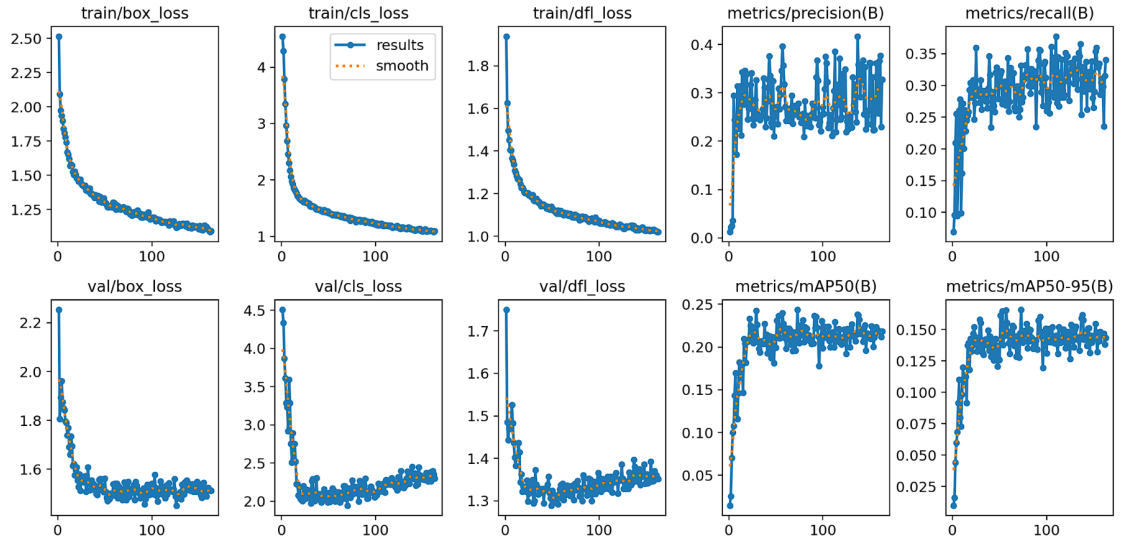


Figure 4.17: Custom Model Training Graph

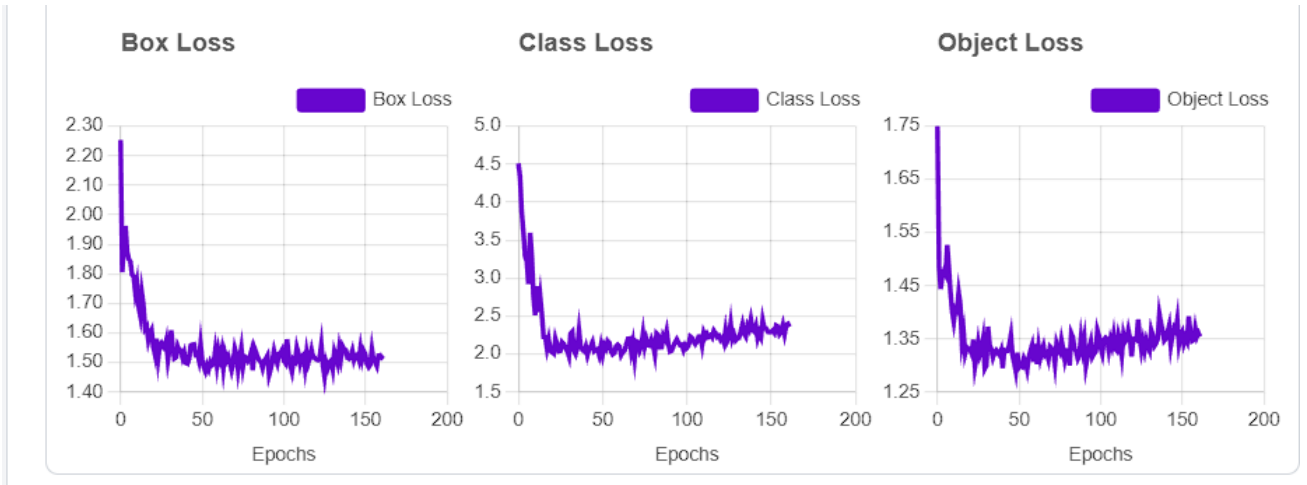


Figure 4.18: Custom Loss



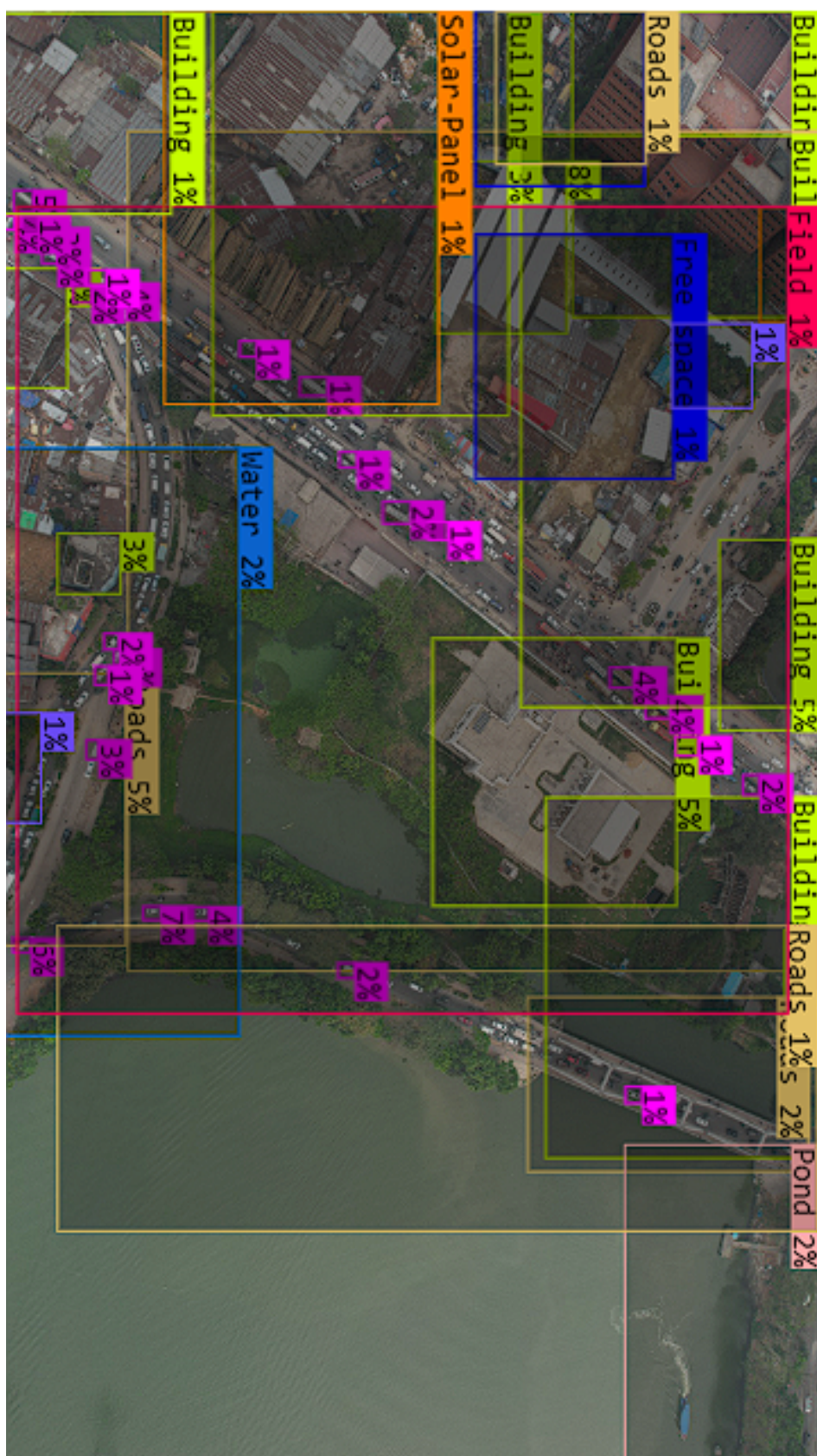


Figure 4.19: Predicted Output from Custom Model

## 4.5 Analysis

We used an aerial image dataset in our research to test the YOLO models V5, V7, and V8 using Roboflow. The results indicate a distinct hierarchy of performance: After comparing V8 to V5, V7 showed better outcomes. In aerial image object detection, the YOLO V8 model demonstrated superior performance compared to its predecessors. Improved model design, improved feature extraction, or improved optimization methods used in V8 could be the reason for this better performance. Applications for aerial image analysis should benefit greatly from the V8 model's enhanced accuracy and dependability. Conversely, the YOLO V5 model performed better than the V7 model even though it was not as well as the V8 model. As each iteration tends to bring about improvements in detection efficiency and accuracy, this shows a growth in the capabilities of the YOLO models. Depending on the particular use cases and resource limitations, V5's intermediate performance indicates that it finds a compromise between computing economy and accuracy, making it a feasible solution. The YOLO V7 model is also an efficient architecture with some limitations in comparison with the latter figures in the hierarchy. It may be architectural limitations or certain optimizations that are in V5 and V8 that are missing in the V6 architecture. Having an understanding of what causes them to perform differently in comparison to the other models would be beneficial when assessing how the V7 performs and which specific areas may need improvements. Further research should try to summarize the particular algorithm and architectural changes that enabled YOLO V8 to show such significant performance increase. Key approaches specific to V8 may include manipulations to feature maps, manipulations to model parameters, and V8 training techniques. One of the potentially useful further directions in exploring the characteristics of the dataset and the generalization to other aerial scenes of each model may give an indication on the robustness and viability of the models in different scenarios. For further optimizing training in specific cases potential for refinement can be considered in terms of what is possible, including further training, transfer learning, and/or perhaps a blended approach of various models. In future the detection accuracy can be further increased by further the investigation in incorporating additional data augmentation mechanisms or processes developed specifically for the aerial images. Overall, it is essential to maintain continuously updated YOLO models for aerial images, and this is showcased in the comparison of the results by the V5, V7, and V8 YOLO models. Future studies can further enhance object recognition capability in aerial image analysis by following through in these areas and applying more knowledge to better prepare models for practical applications.

## 4.6 Categorization

In our recent project of aerial image analysis we have been able to train four separate models for feature and object recognition for the given image. The models are designed to identify eleven specific classes within the images, namely: Building, Field, Free Space, Greenery, Human, Pond, Road, Solar Panel, Tin-shed House, Vehicle, and Water. These categories include virtually all the typical components that might exist in urban and rural settings, which allows making sensible and pre-

cise assessments of the inspected spaces by the standards of the best practices in the field. After detection, we use a classification to remove objects that are mostly detected randomly in urban or rural areas. This classification equally serves the purpose of urban planning, environmental assessment and resource management. Today, our method involves a threshold level that separates the class between the two possibilities. For example, if the tagged image contains a large amount of Buildings, Roads, and/or Vehicles, then the location is most likely an urban environment; conversely, the presence of Fields, Greenery, and/or Water signifies a rural setting. The thresholds used in this classification process are set a priori, but future studies with the professional community in the field of urban planning and with architects will help to improve the chosen classification thresholds. Through such synergism, we strive to make the categorization more objective while considering pragmatic, real-world criteria and pragmatic factors and situations. It has been considered as a substantial progress to enhance the automation and capability in analyzing the aerial images, with potential practical applications in urban planning policy-making. While the models and methodologies for analyzing complex environments are still being crafted and perfected, it can be expected that in the future, this process of automation would go through finer segmentation and classification schemes developed to deal with the heterogeneity found in environmental settings.

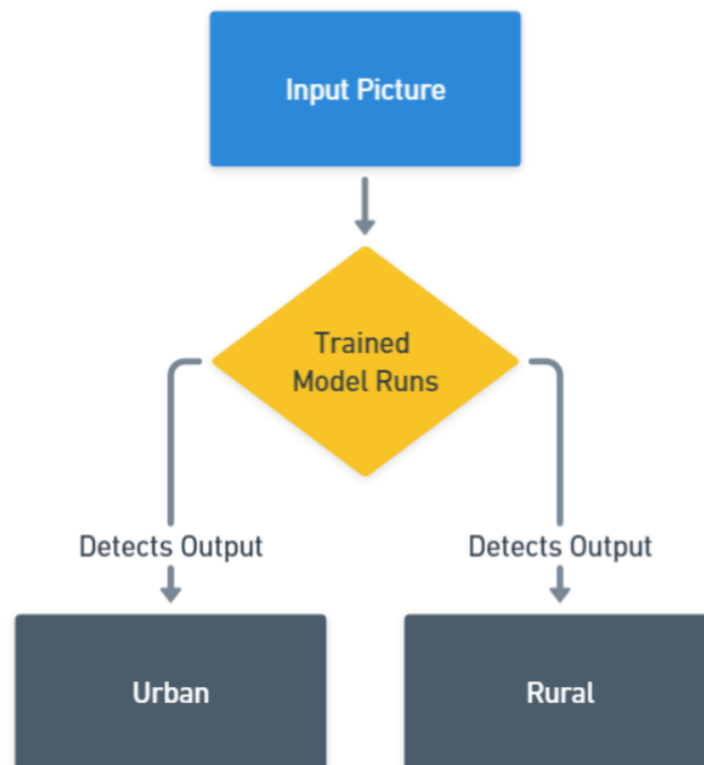


Figure 4.20: Catagorization

## 4.7 Comparison



Figure 4.21: Bangladeshi picture in our model



Figure 4.22: Foreign picture in our model

In between the above two pictures, we can see the comparison between our country picture and foreign picture used in our model. We can see that our picture performs better than foreign picture while we run our model. So, we can say that our model can detect well in our environment not in foreign environment as there is architectural differences.



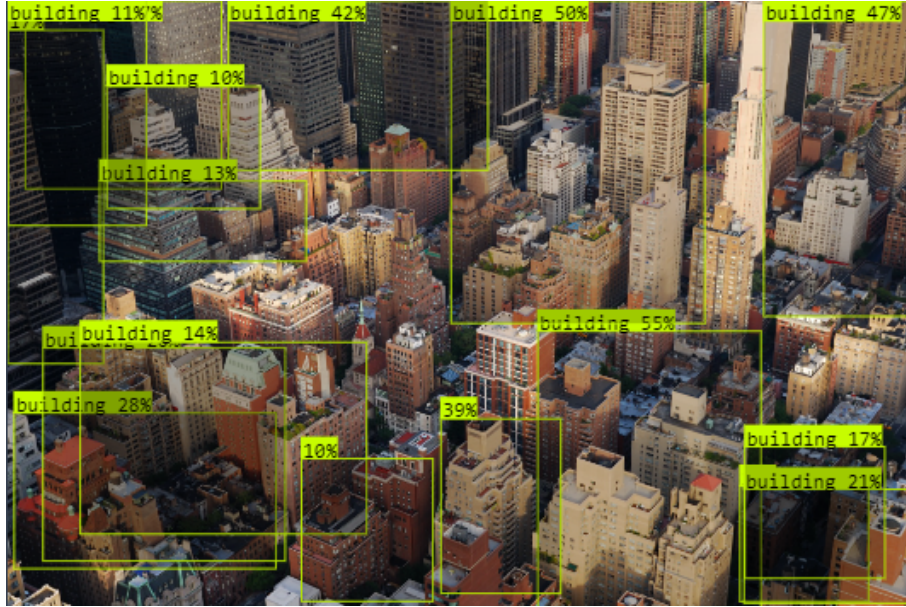


Figure 4.23: Foreign picture in foreign model



Figure 4.24: Bangladeshi picture in foreign model

In between the above two pictures, we can see the comparison between our country picture and foreign picture used in foreign model. We can see that Bangladeshi picture detection doesn't perform better while we run foreign model whereas for the foreign architecture it performs quite well. So, we can say that foreign model cannot detect well in our environment .

So, by analyzing the whole comparison, we can make the decision that our model performs better for Bangladesh based dataset. Furthermore, we trained the model with our custom dataset in a way so that it performs better in case of Bangladeshi scenario as there are severe architectural differences.

## 4.8 3D Augmentation

Photogrammetry is an advanced method that employs images from overlapping aerial photographs or other images to create 3D models, maps, measurements of the captured area object, and so on. Spatial imaging includes a type of imaging in which images are acquired from multiple positions and angles in order to derive spatial information to create high-definition maps. Photogrammetry is a new and rapidly developing approach to the digital depiction of the world around us that is used in urban planning, in architecture, and in research on the past and present of our planet, in the study of ecological systems, and in the assessment of disaster risks. Having discussed some aspects of the origin of photogrammetry, the importance of aerial imagery and the ability to derive highly accurate 3D models and maps of the captured terrain or urban landscape can prove essential for numerous applications. The limitation of photogrammetry also makes it possible to create zoning maps that are used to establish the boundaries of land that can be used for residential, commercial, as well as industrial purposes. It ensures that damages or deterioration on the structures around the urban area are detected and planned to be done to maintain and repair them in the future. Further, it aids in gauging the effects of new developments on the environment, which will consequently lead to sustainable urban development through the conservation of green spaces and the regulating of pollution sources. Photogrammetry has also been used to produce a detailed 3D model of the surveyed area from aerial images covering the Merul Badda Brac University area. It provides many opportunities for urban planning and development; influencing urban planners and decision-makers to know precisely about the current and developed cities and provide them with the ability to make the right decisions to develop cities the right way. One of the major benefits is the potential to create high-resolution 3D imagery of the structures, facilities, and landscape features. These realistic models present a full view of the urban landscape as it is today and can be used to assess different aspects of the built environment such as relationships of buildings, streets, and pedestrians. This data is very useful for decision makers in terms of planning the most suitable zoning regulations, usage of land planning, and even for the establishment of new construction of residential, commercial or industrial areas. In addition, photogrammetric data helps in analyzing and monitoring changes throughout the years when it comes to growth of an urban area as well as the growth of particular area that need attention or need to be controlled. Photo tracking can be used periodically taking aerial pictures and through analyzing changes on the physical structure of the city, planners can be able to review their urban planning policies and implement new policies in order to fit the new image of the city.

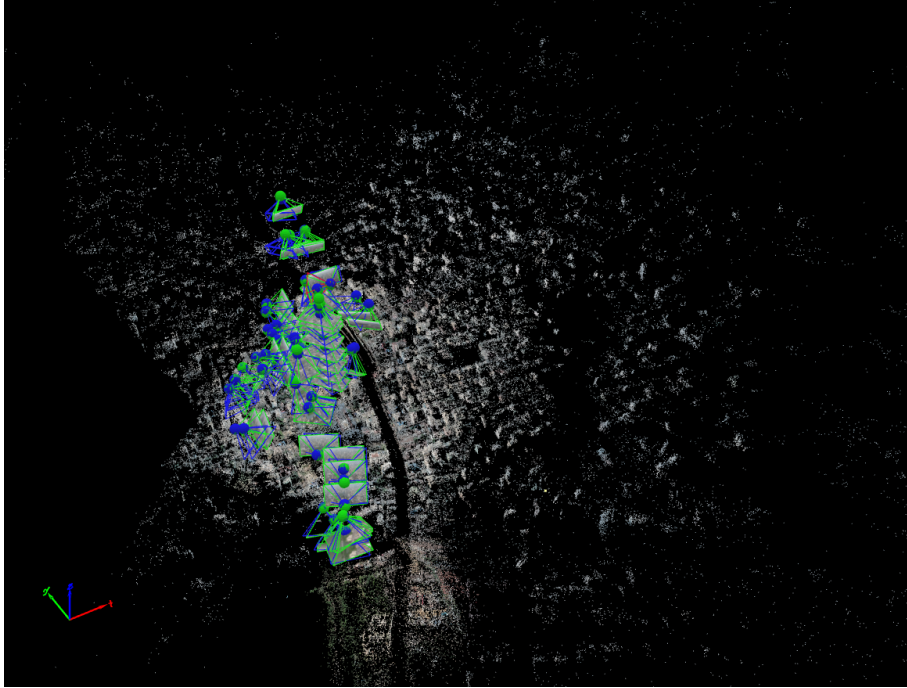


Figure 4.25: Densify Point Cloud

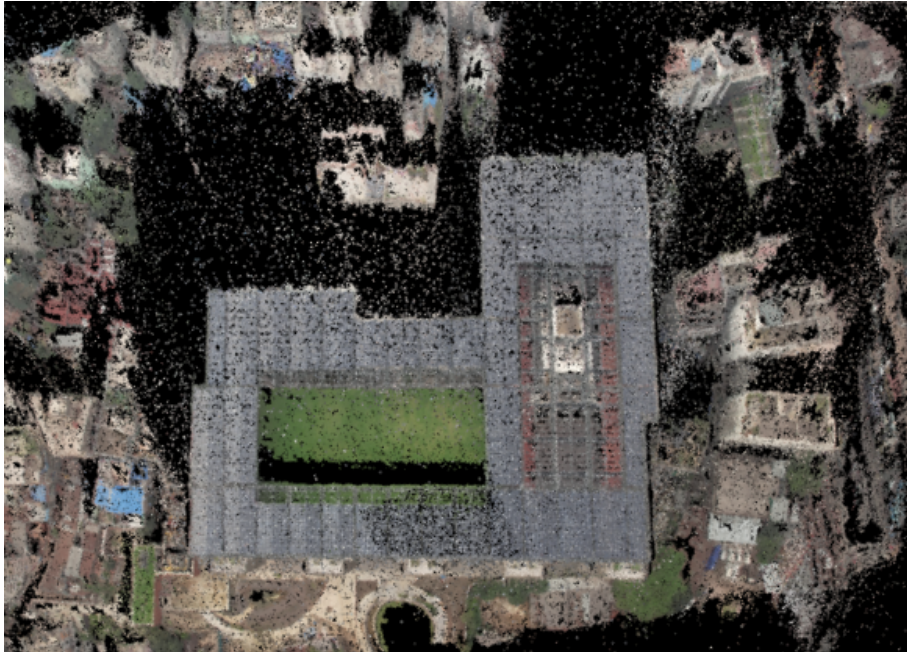


Figure 4.26: Mesh Construction





Figure 4.27: Taken picture

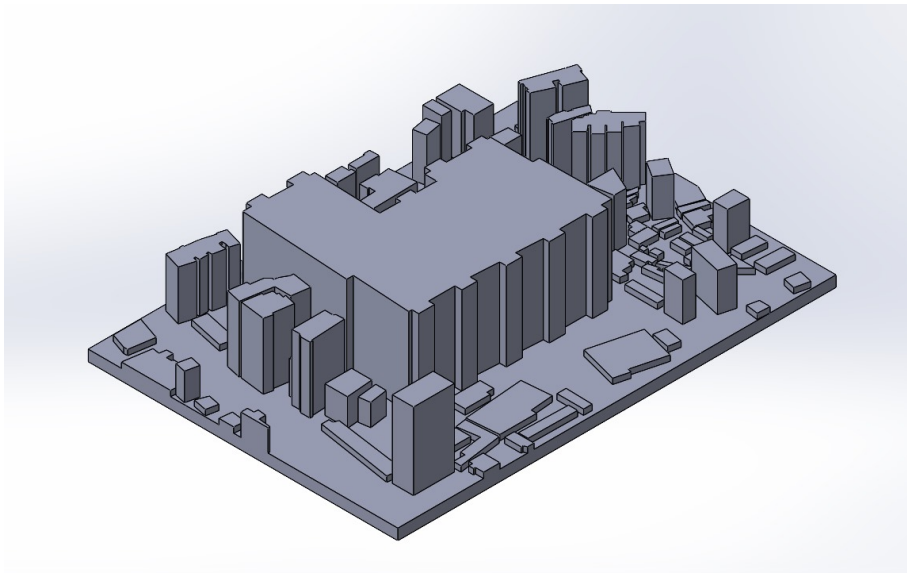


Figure 4.28: 3D mapping

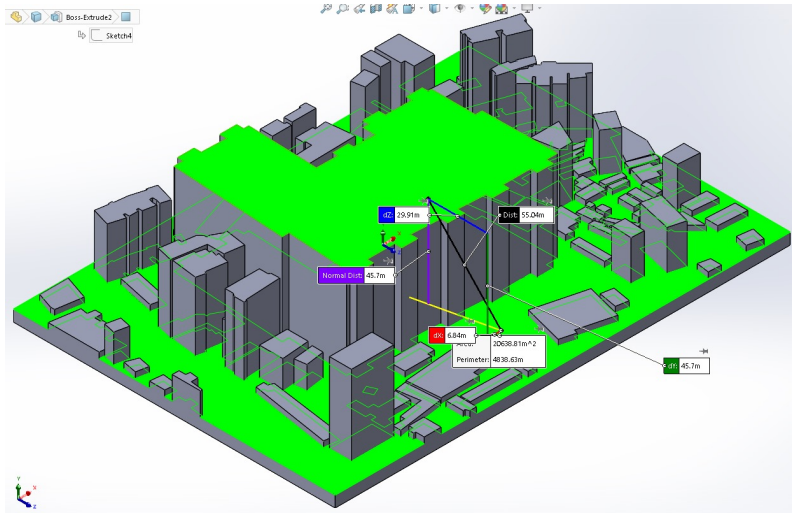


Figure 4.29: Building height prediction

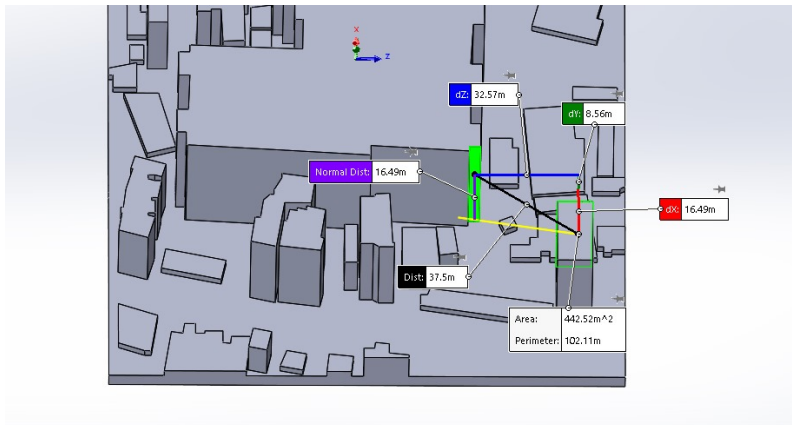


Figure 4.30: Long distance prediction

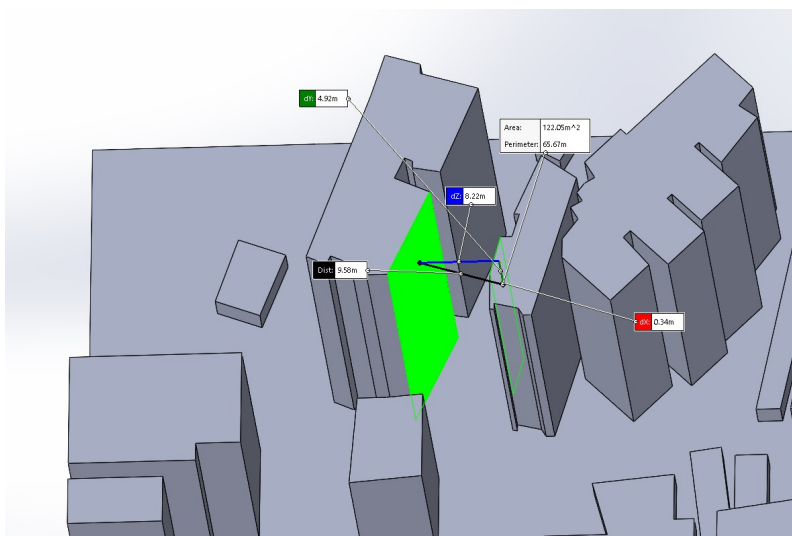


Figure 4.31: Short distance prediction

In our related work, we expanded ourselves to the field of 3D augmentation and gave emphasis on constructing a detailed 3D map that originates from a raw set of drone imagery that we captured. This initiative was particularly focused on the new campus of the Brac University, which was used as the base and from where further comparisons were made. The primary goals were to estimate building heights and to calculate inter-building distances, which are critical factors to estimate buildings' seismic risk.

The first-stop included aerial photography using drones to obtain high resolution images. It was covering the entire new campus of Brac University and had offered a data set ideal for doing 3D modeling. These pictures were then compiled and the data fed through the photogrammetry software to obtain a 3D point cloud. This one here contains the X, Y, Z coordinates of the area captured as well as the nature of surface area. To enrich the further details and adherence to the appearances of the real-world object for the 3D model, we used certain augmentations. These include correcting for lens distortions, up sampling the point cloud density, and accurate geo-referencing. Not only this, the end-product, which is the 3D map, represents the topography and the architectural layout of the campus, but also plays an essential part in the following predictions.

#### **Building Height Prediction :**

Having decided the layout of the ground and the upper floors in 3D, we moved on to forecasting the heights of the buildings. it would be possible to determine those extents of structures which could be described by the vertical range of the point cloud. This required determining the specific coordinates of the base and roof of individual buildings and finding the difference in their altitudes. These early predeterminations were adjusted using machine learning algorithms to allow for the fact that building designs and construction materials may vary. The obtained results were proved by further ground-truth measurement of selected buildings to ensure high accuracy of the predictions.

#### **Building to Building distance analysis :**

Subsequently, we shifted attention to the measurement of distances among the buildings. Thus, the use of the three dimensional map helped in correct measurement of distances in the horizontal plane from one building to the other. It is significant in expounding on the dispersion of the urban populace and effects of seismicity. Nearness of structures often leads to rise in forces considered during an earthquake and, therefore, a higher possibility of the building collapsing. Thus, by estimating these distances, we are in a position to deduce the likely hypertensive zones that may be more susceptible during an earthquake.

For this purpose, our inspection revealed high-risk zones , that is, locations that may need rehabilitation or bolstering to prepare against an earthquake. Moreover, they included geological factors that provided more detailed information on the probability of an area being struck by an earthquake. This multiple plane approach offered a broader perspective on the overall seismic risk since, along with structural and spatial considerations, it included the key premise of subsoil properties and the historic seismological activity in the region. The findings derived from this support the application of sophisticated GIS and machine learning in the area of mitigation and planning of urban regions.



## 4.9 Map Integration



Figure 4.32: Exact Location With Map Integration

Here, by uploading a custom drone image, user can get the exact location from the map which will ultimately help the user to pin point the analysis further. It will eventually help to exactly find the geo location of an image through which exact location can be inferred from an unknown aerial image providing further assistance in analysis.

# Chapter 5

## Conclusion and Future Work

In a larger perspective of learning algorithms' great influence in urban improvements and changes, the application of a spatio-temporal version for identifying urban improvements and changes with the help of aerial imagery can be viewed as a marker. These algorithms include the demand for devices with deep learning technologies like YOLO (You Only Look Once) to urban planning. This combination of algorithms geared toward future advances and urban improvement research will be very useful to the sustainable urbanization process in the future. Computers have now learned the skills to recognize things in a scene from large volumes of visual data. For example, YOLO can be considered as a great object recognition algorithm in terms of speed and accuracy. This framework enables the model to achieve an awesome ability to recognize multiple types of urban changes simultaneously. It covers a large number of urban changes that can be observed using high-resolution aerial imagery such as newly constructed buildings, changes in the existing structures of buildings, demolition of buildings, modifications to infrastructure, and other significant occurrences of the built environment. Utilizing Time Scales and Visualization with Real Time YOLO Modeling for Spatial Analysis of Future Traffic Congestion using Vision-Based Traffic Prediction. It may go further and observe the transformation over time. The dynamic character of interrelated processes illustrated by the YOLO model means that the experts can follow the subjects and objects of the research of urbanization with high speed and as accurately as possible. Models with buildings and cars, as well as yolo, are a tool for urban planning that could be useful for decision-makers, urbanists, and researchers. The capacity to gather real-time data on the processes of urban change at the micro-level allows communities to respond to demand for additional growth in the place of urban space itself, the need to invest in the infrastructure and solutions to the challenges of sustainability with help of real-time dynamic supplementary information. Now we have YOLO designing with machine learning for urban designs that are responsive to communities in change. This model presents a strong reasoning form to derive the spatial-temporal outcomes in urban change, and particularly improve the results of the detection of infrastructure that will be useful to improve urban planning and policy in the future.

## 5.1 Future Work

### Web based GUI

**Dashboard**

Upload your image:  No file chosen

Choose model:

Image will appear here after processing.

Figure 5.1: Prototype



Figure 5.2: Predicted Output

The long-term vision is to have a complete web application that would allow the stakeholders to upload aerial images of cities to monitor and analyze changes over time. This platform will utilise a sophisticated and the large quantity of the existing machine learning repositories would be also created alongside with the ones specifically tuned for the geographic and urban context of Bangladesh. The most important and comprehensive abilities of the web app would be to detect structural changes, categorising by the amount of detected changes as well as specifying 3D

augmentation results. Furthermore, it will incorporate mapping capabilities to identify the precise coordinates of documented shifts and to establish whether the region is urban or rural.

This system should be capable of detecting numerous elements and alterations; it is possible to address the consequences of both human interference, for instance, building or altering some constructions, and natural phenomena occurring when waters flood or erode coastlines. Since the information accumulated will be fresh and most appropriate by applying aerial imagery, the application shall be of great benefit to RAJUK and other urban town planner. It will help them in taking informed decisions, providing them with a time edge in tackling the future issues that may arise and could help them wisely plan for the urban development projects. In the long run, this technology will enhance effective delivery of services especially to the large cities in Bangladesh by facilitating monitoring and achievement of sustainable development by the various governmental departments. The application will be effective in helping the nation to meet its urban development agendas through efficient resource utilization and qualitative improvement of the standards of people.

# Bibliography

- [1] A. Singh, “Review article digital change detection techniques using remotely-sensed data,” *International Journal of Remote Sensing*, vol. 10, pp. 989–1003, 1989. DOI: 10.1080/01431168908903939. [Online]. Available: <https://www.tandfonline.com/doi/abs/10.1080/01431168908903939>.
- [2] A. Angel, M. Hickman, P. Mirchandani, and D. Chandnani, “Methods of analyzing traffic imagery collected from aerial platforms,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 4, no. 2, pp. 99–107, 2003. DOI: 10.1109/TITS.2003.821208. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/1253218>.
- [3] P. Coppin, I. Jonckheere, K. Nackaerts, B. Muys, and E. Lambin, “Review article digital change detection methods in ecosystem monitoring: A review,” *International Journal of Remote Sensing*, vol. 25, pp. 1565–1596, 2004. DOI: 10.1080/0143116031000101675. [Online]. Available: <https://www.tandfonline.com/doi/full/10.1080/0143116031000101675>.
- [4] D. W. Zaum, *Robust building detection in aerial images*, CMRT05. IAPRS, Vol. XXXVI, Part 3/W24, 2005. [Online]. Available: [https://www.isprs.org/proceedings/XXXVI/3-W24/papers/CMRT05\\_Mueller\\_Zaum.pdf](https://www.isprs.org/proceedings/XXXVI/3-W24/papers/CMRT05_Mueller_Zaum.pdf).
- [5] G. Gienko, J. G. Terry, and L. V. Tokalauvere, “Time series aerial photographs to detect spatiotemporal changes in coastal areas of fiji,” *The South Pacific Journal of Natural and Applied Sciences*, 2008. DOI: 10.1071/sp08009. [Online]. Available: <https://doi.org/10.1071/sp08009>.
- [6] B. Sirmacek and C. Ünsalan, “Building detection from aerial images using invariant color features and shadow information,” in *Department of Electrical and Electronics Engineering, Yeditepe University*, 2008. DOI: 10.1109/iscis.2008.4717854. [Online]. Available: <https://doi.org/10.1109/iscis.2008.4717854>.
- [7] J. Gleason, A. V. Nefian, X. Bouysseounousse, T. Fong, and G. Bebis, “Vehicle detection from aerial imagery,” in *2011 IEEE International Conference on Robotics and Automation*, Shanghai, China, 2011, pp. 2065–2070. DOI: 10.1109/ICRA.2011.5979853. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/5979853>.
- [8] L. Bruzzone and F. Bovolo, “A novel framework for the design of change-detection systems for very-high-resolution remote sensing images,” *Proceedings of the IEEE*, vol. 101, no. 3, pp. 609–630, 2013. DOI: 10.1109/JPROC.2013.2247551. [Online]. Available: <https://ieeexplore.ieee.org/document/6948368>.

- [9] S. Jin, L. Yang, P. Danielson, C. Homer, J. Fry, and G. Xian, "A comprehensive change detection method for updating the national land cover database to circa 2011," *Remote Sensing of Environment*, vol. 132, pp. 159–175, 2013. DOI: 10.1016/j.rse.2013.01.001. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0034425713000242>.
- [10] K. Liu and G. Mattyus, "Fast multiclass vehicle detection on aerial images," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 9, pp. 1938–1942, 2015. DOI: 10.1109/LGRS.2015.2439517. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/7122912>.
- [11] F. Alidoost and H. Arefi, "A cnn-based approach for automatic building detection and recognition of roof types using a single aerial image," *Pfg – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, vol. 86, no. 5-6, pp. 235–248, 2018. DOI: 10.1007/s41064-018-0060-5. [Online]. Available: <https://doi.org/10.1007/s41064-018-0060-5>.
- [12] K. L. Jong and A. S. Bosman, "Unsupervised change detection in satellite images using convolutional neural networks," in *2019 International Joint Conference on Neural Networks (IJCNN)*, 2018, pp. 1–8. [Online]. Available: <https://www.semanticscholar.org/unsupervised-change-detection>.
- [13] K. De Jong and A. Sergeevna Bosman, "Unsupervised change detection in satellite images using convolutional neural networks," in *2019 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2019, pp. 1–8. DOI: 10.1109/IJCNN.2019.8852236. [Online]. Available: <https://ieeexplore.ieee.org/document/5196726>.
- [14] D. Yadav, M. Choksi, and M. A. Zaveri, "Supervised learning based greenery region detection using unnamed aerial vehicle for smart city application," in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kanpur, India, 2019, pp. 1–7. DOI: 10.1109/ICCCNT45670.2019.8944548. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8944548>.
- [15] M. Won, "Intelligent traffic monitoring systems for vehicle classification: A survey," *IEEE Access*, pp. 1–1, 2020. DOI: 10.1109/ACCESS.2020.2987634. [Online]. Available: [https://www.researchgate.net/publication/340615750\\_Intelligent\\_Traffic\\_Monitoring\\_Systems\\_for\\_Vehicle\\_Classification\\_A\\_Survey](https://www.researchgate.net/publication/340615750_Intelligent_Traffic_Monitoring_Systems_for_Vehicle_Classification_A_Survey).
- [16] R. E. Huerta-García, F. D. Yépez, D. F. Lozano-García, *et al.*, "Mapping urban green spaces at the metropolitan level using very high resolution satellite imagery and deep learning techniques for semantic segmentation," *Remote Sensing*, vol. 13, no. 11, p. 2031, 2021. DOI: 10.3390/rs13112031. [Online]. Available: <https://doi.org/10.3390/rs13112031>.
- [17] D. Xue, T. Lei, X. Jia, X. Wang, T. Chen, and A. Nandi, "Unsupervised change detection using multiscale and multiresolution gaussian-mixture-model guided by saliency enhancement," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 1796–1809, 2021. DOI: 10.1109/JSTARS.2021.3052547. [Online]. Available: <https://ieeexplore.ieee.org/document/9305230>.



- [18] Y. Zhang, L. Fu, Y. Li, and Y. Zhang, “Hdfnet: Hierarchical dynamic fusion network for change detection in optical aerial images,” *Remote Sensing*, vol. 13, no. 8, p. 1440, 2021. DOI: 10.3390/rs13081440. [Online]. Available: <https://doi.org/10.3390/rs13081440>.
- [19] N. Coghe, *Detecting buildings in aerial and satellite images using semantic segmentation*, 2022. [Online]. Available: <https://sigblog.hexagon.com/detecting-buildings-in-aerial-and-satellite-images-using-semantic-segmentation>.
- [20] M. Hu, Z. Li, J. Yu, X. Wan, H. Tan, and Z. Lin, “Efficient-lightweight yolo: Improving small object detection in yolo for aerial images,” *Sensors*, vol. 23, no. 14, p. 6423, 2023. DOI: 10.3390/s23146423. [Online]. Available: <https://doi.org/10.3390/s23146423>.