DEVELOPMENT OF WASTE DETECTION AND IDENTIFICATION SYSTEM

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF

MASTER OF ENGINEERING IN ELECTRICAL ENGINEERING

Submitted By

MD MANARUL SK
Registration No: **154009** of **2020-21**
Examination Roll No.: **M4ELE22019**

Under the guidance of

Prof. SAYANTAN CHAKRABORTY

Department of Electrical Engineering

Prof. SHELI SINHA CHAUDHURI

Department of Electronics & Telecommunication Engineering

Jadavpur University

Kolkata – 700032

July, 2022

DECLARATION OF ORIGINALITY AND COMPLIANCE OF ACADEMIC ETHICS

I hereby declare that this thesis contains a literature survey and original work by the undersigned candidate, as part of his Master of Engineering in Electrical Engineering in the Department of Electrical Engineering. All information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that I have thoroughly cited and referenced all material and findings that are not original to this research, as provided by these rules and conduct.

Name: Md Manarul Sk

Exam Roll No : M4ELE22019

Thesis Title: “ DEVELOPMENT OF WASTE DETECTION AND IDENTIFICATION SYSTEM”

Signature of Candidate

FACULTY OF ENGINEERING & TECHNOLOGY
JADAVPUR UNIVERSITY

CERTIFICATE

This is to certify that the thesis entitled — "DEVELOPMENT OF WASTE DETECTION AND IDENTIFICATION SYSTEM" has been carried out by Md Manarul Sk bearing Class Roll No: 002010802019, Examination Roll No.: M4ELE22019 and Registration No: 154009 of 2020-21, under my guidance and supervision, and be accepted in partial fulfillment of the requirement for the degree of Master of Engineering in Electrical Engineering in the Department of Electrical Engineering.

Prof. Sayantan Chakroborty
Assistant Professor
Department of Electrical Engineering
Jadavpur University
Kolkata-700032

Prof. Sheli Sinha Chaudhuri
Professor
Department of Electronics and
Telecommunication Engineering
Jadavpur University
Kolkata-700032

Prof. Saswati Mazumdar
Head of the Department
Department of Electrical Engineering
Jadavpur University
Kolkata-700032

Prof. Chandan Mazumdar
Dean
Faculty Council of Engineering & Technology
Jadavpur University
Kolkata-700032

FACULTY OF ENGINEERING & TECHNOLOGY
JADAVPUR UNIVERSITY

CERTIFICATE OF APPROVAL

The foregoing thesis titled − "**DEVELOPMENT OF WASTE DETECTION AND IDENTIFICATION SYSTEM**" is here approved as a creditworthy study of an engineering subject conducted and presented satisfactorily to warrant its acceptance as a precondition to the degree for which it was submitted. It is understood that the undersigned does not automatically support or accept any argument made, opinion expressed, or inference drawn on it by this approval, but only approves the thesis for the reason for which it was submitted.

Committee on Final Examination for
Evaluation of the Thesis

……………..……………………….

Signature of an External Examiner

Prof. Sayantan Chakroborty
Supervisor
Department of Electrical Engineering
Jadavpur University
Kolkata-700032

Prof. Sheli Sinha Chaudhuri
Supervisor
Department of Electronics and Telecommunication Engineering
Jadavpur University
Kolkata-700032

ACKNOWLEDGEMENT

This thesis, entitled “**DEVELOPMENT OF WASTE DETECTION AND IDENTIFICATION SYSTEM**” is the result of the work whereby I have been accompanied and supported by many people, my guide, my friends, and lab seniors. It is a pleasant aspect that now I have the opportunity to express my gratitude to all of them.

First and foremost, I would like to express my sincere gratitude to my thesis supervisors, Dr. Sayantan Chakraborty, Professor of the Department of Electrical Engineering, and Dr. Sheli Sinha Chaudhuri, Professor of the Department of Electronics and Telecommunication Engineering, Jadavpur University, for their valuable guidance, insightful suggestions, and support while conducting this thesis work as well as in the writing of this thesis. I have been very fortunate to have had a guide like them. Their positivity, confidence, and ideas help me to complete my thesis, and they guide me as a guardian.

I am thankful to my fellow project mates, friends, and technical and non-technical staff of Jadavpur University who have helped me directly or indirectly during the tenure of my thesis work.

I want to express my gratitude to my parents and family, as, without their sacrifices, I couldn't do anything. And their invaluable love, encouragement, and support made me, whatever I am today.

Md Manarul Sk.
Jadavpur University
Kolkata-32
West Bengal

ABSTRACT

One of the biggest environmental problems in the modern world is waste pollution. Recycling is significant for both ecological and economic reasons, and the sector demands great productivity. Due to the lack of standards and benchmarks for the often utilized measurements and data, current studies on autonomous waste detection are difficult to compare. The work carried out in this thesis addresses the issues to some extent by performing a comprehensive examination of more than twenty existing trash datasets as well as a brief but insightful evaluation of the available Deep Learning-based waste identification methods. This thesis work, conducted to create a standard baseline for litter detection, has compiled prior research and presented the findings of the experiments performed on the dataset. The datasets thus created can be used as a benchmark for future research work on waste detection and classification. These datasets have unified annotations covering all potential waste categories, including plastic, glass, metal, thermocol, nylon, paper,etc.

Waste segregation encourages the creation of energy from waste, the decrease of landfills, recycling, and waste reduction. Waste that is improperly disposed of contaminates recycled materials. The recycling sector faces a serious problem with contamination, but automatic computerized waste sorting can help. The availability of models or tactics that assist individuals in sorting garbage has grown to be crucial for the proper disposal of such garbage. Even though there are many different recycling categories, many individuals are still unsure of how to choose the appropriate trash can to dispose of each item of waste. Around the world, it is believed that waste management and the methodical sorting of waste play a key part in the development of a healthier environment. By reusing and recycling unwanted products, society may reduce waste and, consequently, solve environmental issues. In order to appropriately dispose of waste in the recyclable and non-recyclable bins, this project intends to develop an automated waste detection system utilizing a deep learning algorithm that will collect waste images or videos from a camera with object recognition, detection, and prediction.

Table of Contents

LIST OF FIGURES:

LIST OF TABLES:

CHAPTER 1

INTRODUCTION

The waste generated from human and animal activities is unwanted or useless material. There are many types of waste materials like plastic , glass, paper,organic, biomedical, metals, and others . Out of those, some waste materials are non-biodegradable or very slow to degrade. The amount of waste is increasing day by day at a very high rate and polluting the air, water, and soil, leading to an unhealthy environment for all living beings. This calls for having proper waste management systems for creating a sustainable life, but it is a very complex and challenging process because the system often does not have any segregation of waste rules. Improper waste management leads to severe effects on the atmosphere, human life, and natural resources, and effective waste management is expensive.

In this thesis work, I have tried to automatically classify the different types of waste by applying deep learning techniques and creating a database of images of waste materials and labeling them using a software called Labelme.

As the human population grows rapidly, it is needless to point out that the amount of waste generated is rapidly increasing day by day. Waste disposal has serious environmental and health hazards and is such an important issue globally that each and every person should be aware of it. Thus, proper waste management is very important in order to protect our mother earth and for the safety of not only humans but also other living beings. Besides having ecological reasons, proper waste management has economic reasons also, from the point of view of recycling.

The conventional , expensive and inefficient waste management systems as existing currently can be replaced by a smart waste management system with the incorporation of Internet of Things (IoT) and deep learning models. In this thesis, a smart waste management system is implemented by the analysis of many existing waste data sets and with the help of deep learning based on waste classification in an image containing plastic , paper, glass, metal, nylon,cloth, thermocol, etc.

In physics, waste is a substance that is abandoned after its primary use or is no longer useful, worthless, or defective. The by-product, on the other hand, is a jointly produced good of limited economic value. By inventing a process that raises a waste product's value above zero, it can become a joint product, a by-product, or a resource.

By managing waste, we are able to minimize the impact of waste on the environment, health, and so forth. Solid, liquid, gaseous, and hazardous waste can all be managed through proper waste management. By managing wastes, one aims at reducing the danger they pose to the environment and to people. Municipal solid waste is the waste created by businesses, industry, and households, which is a major part of waste generation.

Waste management in the municipal sphere can have adverse environmental effects such as creating infectious diseases, polluting soil and water, obstructing drains, and reducing biodiversity. While some waste eventually rots, not all do, and while this process is occurring, it may emit methane gas, which is explosive and contributes to global warming. When waste is dumped into the ocean, toxic substances are retained, soaking up all the oxygen from the water. As a consequence, marine life is deprived of adequate oxygen in their natural habitat, causing them to die and, in some cases, to become extinct.

CHAPTER 2

LITERATURE REVIEW

In the past few years, various studies have been conducted to limit the effects of improper waste disposal. In the past, neural networks and support vector machines have been used to classify images. Automated waste sorting can be supported by machine learning (ML). Recently, ML-based systems have been implemented that can help or fully cover the sorting procedure, resulting in a faster pace of this process.

2. OBJECT DETECTION:

The detection of objects in images or videos is done by using computer vision techniques. In order to produce meaningful results, object detection algorithms typically utilize machine learning or deep learning algorithms. Images and videos of objects of interest can be recognized and located within moments by humans when they look at them. In object detection, the goal is to simulate this intelligence with a computer.

Basically, the object detection methods are of two types, as discussed below:

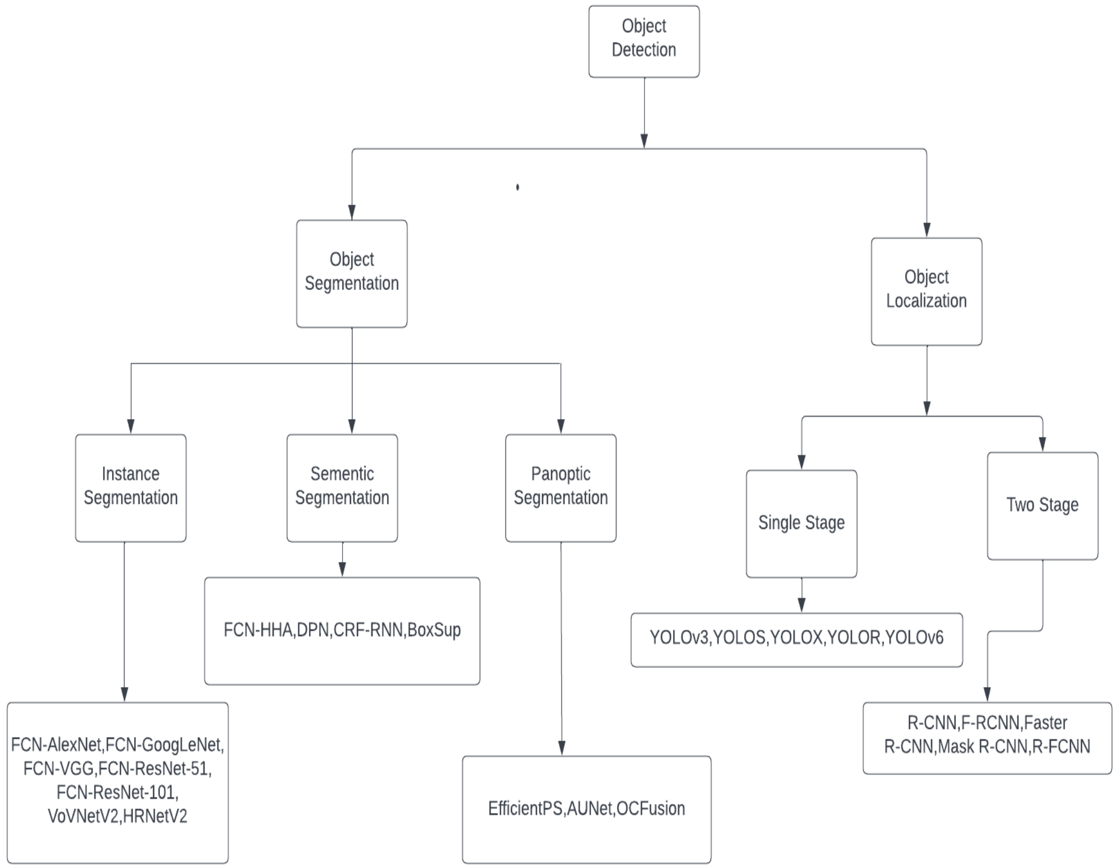


Fig 2.1
Taxonomy of Object Detection

2.1.a Image Segmentation: The process of segmentation[3] is used in the processing of digital images as well as in computer vision for separating different regions in the images and identifying objects. This can be again subdivided according to applications as shown:

- **Face Segmentation:** Computer Vision Systems [2] are helped by semantic segmentation to achieve tasks like expression recognition, prediction of age, gender, and race of individuals. These tasks are enabled by semantic segmentation[4], which marks the regions of the face into important attributes, including the mouth, chin, nose, eyes, and hair[1].

- **Medical Imaging:** Image segmentation is used for evaluating X-rays and MRI scans in medical diagnosis. As a result, semantic segmentation can enable diagnostic tests to be easier and simpler by identifying relevant areas within an image[5].

- **Self Driving:** The process of autonomous driving is extremely complex and requires real-time analysis, perception, and change. Objects such as automobiles, traffic signage, and regions such as highways and sidewalks are identified using semantic segmentation[7]. Autonomous driving uses instance segmentation to segment individual cars, pedestrians, signs, and other objects[6].

- **Satellite image processing :** Land area and surface structures are displayed in satellite images[8]. Analysis of these images gives us information about land use, deforestation areas, agricultural land,water bodies, coastal areas, etc. through semantic segmentation[4]. This information is useful for proper planning and development by different government bodies.

In digital image processing, segmentation algorithms divide the image into its components according to the requirements of the application. This division is implemented based on its properties and features of the components. Image segmentation has the primary purpose of simplifying the image for easier analysis. Each division or segment of the image is either the background or an object in the image. A technique for segmenting images allows us to select specific pixels from an image, group them, assign labels to these groups, and then classify further pixels according to these labels. Segmenting images is a very broad topic, and there are several ways to accomplish the task.

Two types of object segmentation are available, as presented below:

2.1.a.i) Semantic Segmentation: Humans have the intrinsic ability to recognize the things they observe in their environment. Our brain's visual cortex is capable of quickly and easily differentiating between a cat and a dog. This applies to practically all of life, not only cats and dogs. However, a computer cannot do this on its own because it is not as intelligent as a human brain. Through a particular class of artificial neural networks called Convolutional Neural Networks (CNNs)[9], deep learning[10] researchers have attempted to close this gap between the human brain and computers over the course of the last few decades [4]. Every pixel in an image is assigned a semantic label by semantic segmentation as opposed to classification, where all images are assigned a single label, which is an extremely different approach. As a result of semantic segmentation, multiple objects of the same class are considered to be one entity. Image segmentation[11] involves dividing a digital image into various segments (sets of pixels) in accordance with specific criteria. The goal is to transform the image into a more insightful, understandable representation.

- **Deep Parsing Network (DPN):** To simulate unary terms, DPN extends a CNN, and new layers are created to roughly resemble the mean field (MF) approach for paired terms[12]. To improve the outcome, non-deep-learning post processing was used in certain earlier methods, which was unnecessary when the network approximates the pairwise terms.

- **Conditional Random Fields Recurrent Neural Network (CRF-RNN)[13]:** One of the most effective graphical models for computer vision is the CRF. A Fully Convolutional

Network (FCN) has been discovered to produce extremely coarse segmentation results. Hence, many techniques, like DeepLabv1[14] and DeepLabv2[15], use CRF as post-processing steps to enhance the output semantic segmentation map acquired from the network in order to get more precise segmentation results. The CRF parameters, however, are not trained alongside FCN. In other words, during training, the FCN is not aware of CRF. This could reduce the network's capacity.

- **Box Sup[16]:** Deep convolutional networks trained using pixel-level segmentation masks that have been human-annotated are a key component of contemporary leading techniques for semantic segmentation. The performance of deep networks, which often benefit from more training data, is constrained by such pixel-accurate supervision, which necessitates expensive labeling work. The fundamental concept is to alternate between automatically generating region suggestions and convolutional network training. To enhance the networks, these two processes gradually regain segmentation masks and vice versa. When supervised by simple boxes, our approach, dubbed "BoxSup," achieves competitive results with strong baselines completely guided by masks in the same environment. BoxSup further produces state-of-the-art results on PASCAL VOC 2012[17] and PASCAL-CONTEXT[18] by utilizing a high number of bounding boxes.

2.1.a.ii) Instance Segmentation: One of the more significant, intricate, and difficult fields of research in machine vision is instance segmentation. It localizes several classes of object instances present in diverse images with the goal of predicting the object class label and the pixel-specific object instance mask[53]. Robotics[54], autonomous vehicles[52], surveillance, etc., are mostly helped by instance segmentation. Multiple objects of the same class are treated as distinct objects by instance segmentation. In most cases, instance segmentation is more challenging than semantic segmentation.

- **FCN-AlexNet[55]:** The initial five layers of AlexNet were convolutional; parts of them were followed by max-pooling layers, and the final three layers were fully connected. [2] It made use of the non-saturating ReLU activation function[56], which outperformed tanh and sigmoid in terms of training performance.

- **ResNet-101[57]:** The ResNet-101 model is returned by the resnet101 command. The ImageNet Large-Scale Visual Recognition Challenge uses a portion of the ImageNet database on which this model was trained, the Large Scale Visual Recognition Challenge (ILSVRC)[58]. The model can identify images into more than a thousand object categories, including keyboards, mice, pencils, and numerous animals, after being trained on more than a million images. The model has acquired extensive feature representations for a variety of images.

- **GoogLeNet[59]:** A convolutional neural network with 22 layers is called GoogleNet. The network can be loaded in a pretrained state that has been trained on either the Places365 [2] or ImageNet [1] data sets. The network that was trained on ImageNet[60] divides images into over 1000 different object categories, including several animals, a keyboard, a mouse, and a pencil. Similar to networks trained on ImageNet, Places365 networks[61] classify images into 365 distinct place types, such as fields, parks, runways, and lobbies. For a variety of images, these networks have learned several feature representations. The input picture size for both of the pretrained networks is 224 by 224.

2.1.a.iii) Panoptic Segmentation[137]: Pan and optic are the roots of the word panoptic. Both pan and optic refer to all vision. Thus, the term "panoptic segmentation" essentially translates to "everything observable in a specific visual field." Panoptic segmentation in computer vision can be accomplished in three easy steps: those are dividing each picture item into distinct portions that stand alone from one another. Labeling involves painting a different color on each distinct component. categorizing the things.

As was indicated in the post's beginning, panoptic segmentation combines instance and semantic segmentation. To put it another way, we can find out using panoptic segmentation how many objects there are in each instance class (countable objects), bounding boxes, and instance segmentation. But semantic segmentation also enables us to determine to which class each pixel in the image belongs. This undoubtedly gives a situation a more complete understanding.

EfficientPS[137]: EfficientPS is a deep learning model that makes panoptic predictions at a low computational cost by using a backbone built upon EfficientNet architecture. It consists of:

A backbone network for feature extraction.

Two output branches: one for semantic segmentation and one for instance segmentation.

A fusion block that combines the outputs from both output branches.

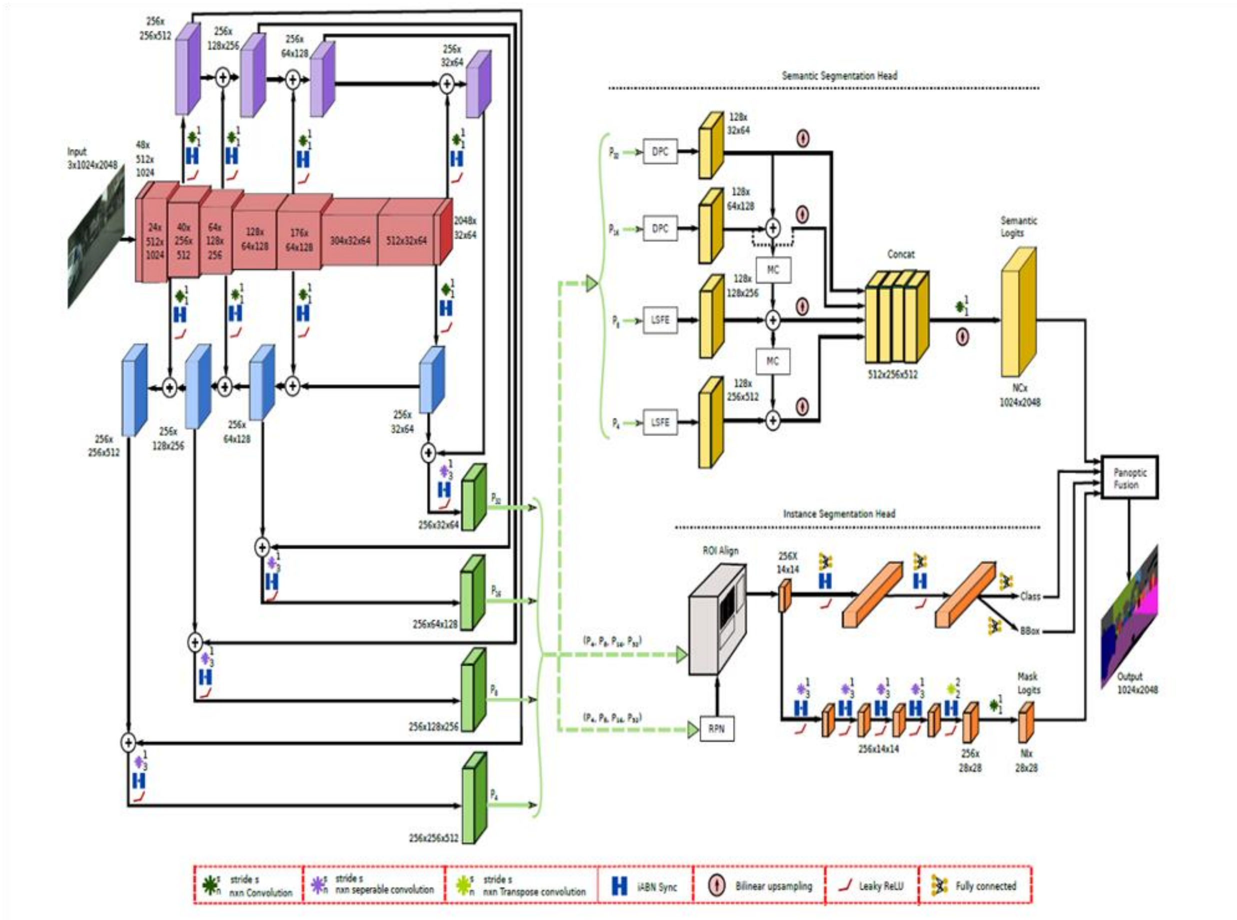


Fig 2.2
EfficientPS Model Architecture

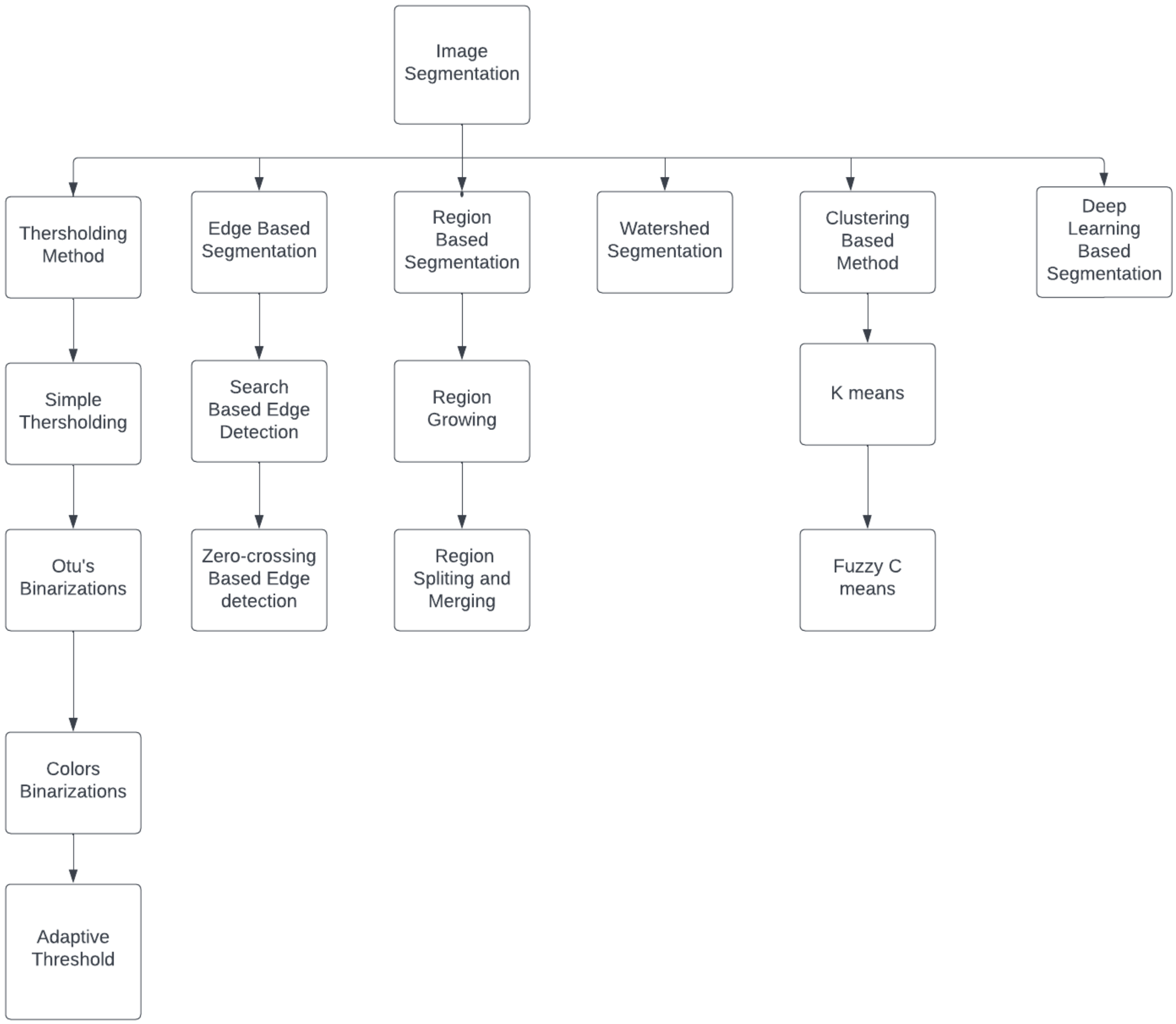


Fig 2.3
Taxonomy of Image Segmentation

2.1.b Image Segmentation Approach:

Image segmentation is the division of a digital image into several image segments, often referred to as image regions or image objects, in digital image processing and computer vision (sets of pixels). The purpose of segmentation is to reduce complexity and/or transform an image's representation into something more relevant and understandable. Image segmentation is frequently used to identify objects and boundaries in images (such as lines, curves, etc.). Image segmentation, in more exact terms, is the process of giving each pixel in an image a label so that pixels with the same label have specific properties.

Different Types of Image Segmentation Techniques:

2.1.b.i) Approach-Based classification[19]: Image segmentation is, at its most basic level, object recognition. Without first recognising an item, an algorithm cannot classify the various components. As a result, image segmentation approaches can be classified based on how computers recognize objects, which involves gathering similar pixels and separating them from dissimilar pixels. There are two methods for completing this task:

2.1.b.i.1)Region-Based Approach[20]: Similar pixels in the image are detected using this method, which includes region merging, region spreading, and region increasing. This strategy is used by clustering and other machine learning algorithms to find unknown traits and properties. This strategy is also used by classification algorithms to recognise features and separate image segments based on them.

2.1.b.i.2)Boundary-Based Approach[21]: The boundary-based method to object identification is the polar opposite of the region-based technique. Unlike region-based detection, which finds pixels with comparable properties, the boundary-based technique finds pixels that are distinct to one another. This method is used by algorithms like Point Detection[22], Edge Detection[23], Line Detection[24], and others, which detect the edges of distinct pixels and isolate them from the rest of the image.

2.1.b.ii) Thresholding Segmentation[31]: In image processing, the threshold approach is the simplest method of image segmentation. It divides an image's pixels by comparing the intensity of each pixel to a predetermined threshold value. This technique is very useful when the required object has a higher intensity. The threshold value (T) can be considered a constant, however it will only operate if the image contains very low noise (unnecessary information and data). The threshold value can be static or dynamic depending on the features of the image. The thresholding method divides a grayscale image into two segments and turns it to a binary image . We may categorize thresholding segmentation into the following types based on the different threshold values:

2.1.b.ii.1)Simple Thresholding[32]: In this process, the image's pixels are replaced with either white or black. If the intensity of a pixel at a certain location is less than the threshold value, it will be replaced with black. If it's higher than the threshold, however, you'd replace it with white. This is a basic thresholding method that is ideal for novices in image segmentation.

2.1.b.ii.2)Otsu's Binarization[33]: Thresholding is the technique used to distinguish foreground pixels from background pixels. The Otsu's method, put out by Nobuyuki Otsu, is one of the various methods for obtaining optimal thresholding. The threshold value where the weighted variance between the foreground and background pixels is the least is found using Otsu's method[1], which is a variance-based technique. The important thing is to measure the distribution of background and foreground pixels while iterating over all conceivable threshold settings. Locate the threshold at which the dispersion is the smallest.

2.1.b.ii.3)Colors Binarization[34]: A constant threshold value is chosen and utilized to accomplish image segmentation via thresholding. Various threshold values are taken and implemented to test the segmentation result. The threshold value that gives the best result is selected. Consider an image with a two-peak histogram, one for the foreground and one for the background. The approximate value at the midpoint of those peaks is taken as the threshold value using Otsu binarization.For bimodal images the threshold value from the histogram in Otsu binarization[33] method. Scanning papers, identifying patterns, and deleting extraneous colors from a file are all common uses for this method. It does, however, have a number of drawbacks. It can't be used with images that are not bimodal (images whose histograms have multiple peaks).

2.1.b.ii.4))Adaptive Threshold[35]: Having a single consistent threshold value may not be the best technique for every image. Different backdrops and settings alter the qualities of different photographs. As a result, rather than employing a single constant threshold value to segment the entire image, the threshold value can be made flexible. This method enables the user to assign distinct threshold settings for different parts of an image. This strategy works effectively for photographs that have shifting lighting conditions. The image needs to be divided into smaller portions and threshold values need to be calculated for each portion.

2.1.b.iii) Edge-Based Segmentation[36]: One of the most common types of image processing segmentation is edge-based segmentation. It focuses on recognising the edges of various things in an image. This is an important step since edges carry a lot of information that you can use to find the features of the various objects in the image. Edge detection[36] is popular because it aids in the removal of unwanted and unneeded data from images. It significantly reduces the image's size, making it easier to analyze. Algorithms used in edge-based segmentation identify edges in an image according to the differences in texture, contrast, gray level, color, saturation, and other properties. The quality of the results can be improved by connecting all the edges into edge chains that match the image borders more accurately. There are a variety of edge-based segmentation methods to choose from. It can be categorized into two groups:

2.1.b.iii.1)Search-based edge detection[37]: It is an approach that looks for local directed maxima of the gradient magnitude using a computed estimate of the edge's local orientation to find local directional maxima of the gradient magnitude.

2.1.b.iii.2)Zero Crossing-Based Edge Detection[38]: To discover the edges, zero-crossing-based edge detection algorithms search for zero crossings in a derivative expression extracted from the image.To reduce undesirable noise and make it easier to recognize edges, preprocessing of images is needed. Some of the most popular edge detection operators include Canny, Prewitt, Deriche, and Roberts Cross. They aid in the detection of discontinuities and the location of edges.

The purpose of edge-based detection is to get a partial segmentation minimum that allows you to group all local edges into a binary image. The edge chains in your newly formed binary picture must match the components of the image in question.

2.1.b.iv) Region-Based Segmentation[39]: The image is divided into sections with similar properties using region-based segmentation methods. The method finds these groups by first selecting a seed point, which could be a small chunk or a huge portion of the input image. A region-based segmentation method would either add more pixels to the seed points or decrease them so that they could be merged with other seed points when they were discovered. Region-based segmentation can be divided into two categories based on these two methods:

2.1.b.iv.1)Region Growing[40]: This method involves starting with a limited collection of pixels and then iteratively merging more pixels based on specific similarity criteria. A region-growing algorithm would select an arbitrary seed pixel in the image, compare it to neighboring pixels, and then begin expanding the region by finding matches to the seed pixel. When a region can no longer grow, the algorithm selects a new seed pixel that may or may not belong to any existing region. When a region has too many qualities, it can take up the majority of the image. To prevent making such a mistake, region-growing algorithms create numerous regions at once. For images with a lot of noise, you should utilize region-expanding algorithms instead of thresholding techniques because the noise makes it difficult to locate edges.

2.1.b.iv.2)Region Splitting and Merging[41]: A region splitting and merging focused approach, as the name implies, combines two actions: separating and merging regions of an image. It would divide the image into sections with similar properties first, then merge the adjacent portions that are comparable. The method in region splitting considers the entire image, whereas in region growth, the algorithm focuses on a single location. The process of area splitting and merging is based on a divide and conquer strategy. It separates the image into various parts and then matches them according to specified criteria. Split-merge algorithms are another name for the algorithms that execute this task.

2.1.b.v) Watershed Segmentation[42]: A watershed is a grayscale image change in image processing. It refers to a drainage divide or a geological watershed. The image would be treated as if it were a topographic map by a watershed method. It takes a pixel's brightness as its height and looks for the lines that run along the tops of the ridges. The Watershed algorithm[43] has a variety of technical definitions and applications. Apart from detecting the pixels' ridges, it concentrates on defining basins (the polar opposite of ridges) and flooding the basins with a0 markers until they meet the ridges' watershed lines. Because basins have more markers than ridges, the image is divided into numerous sections based on the 'height' of each pixel. Every image is converted into a topographical map using the watershed method. The topography of the watershed segmentation approach would be reflected in the gray values of their pixels. A landscape with slopes and hills, on the other hand, would undoubtedly have three-dimensional features. The watershed algorithm would take into account the image's three-dimensional representation and construct regions dubbed "catchment basins"[44] as a result. It has a wide range of medical applications, including MRI and medical imaging. Watershed segmentation is a key component of medical image segmentation. Thus, if someone wants to work in that field, he/she should concentrate on studying this method of image segmentation in particular.

2.1.b.vi) Clustering-Based Segmentation Techniques[45]: One could probably come across clustering algorithms if he/she researched classification algorithms. They are unsupervised algorithms that assist them in locating hidden data in an image that is not visible to the open eye. Clusters, structures, shadings, and other information are among the concealed data. A clustering method separates an image into clusters (disjoint groups) of pixels with comparable properties, as the name implies. It would divide the data items into clusters, with the components in one cluster similar to one another. Fuzzy c-means (FCM)[46], k-means[47], and enhanced k-means algorithms are some of the most common clustering techniques. The k-means clustering algorithm is commonly used in image segmentation because it is simple and efficient. The FCM method, on the other hand, divides pixels into separate groups based on their degree of membership. The following are the most important image segmentation and clustering algorithms:

> **2.1.b.vi.1)K-means Clustering[47]:** The K-means algorithm is a straightforward unsupervised machine learning algorithm. It divides an image into a certain number of clusters. It begins by splitting the image space into k pixels, each representing one of the k group centroids. Then, based on the distance between each object and the centroid, they assign each object to a group. The algorithm can move and reassign the centroids once it has assigned all pixels to all clusters.
>
> **2.1.b.vi.2)Fuzzy C Means[46]:** The pixels in an image can be grouped into numerous clusters using the fuzzy c-means clustering algorithm. This means that a pixel can be assigned to many clusters. Every pixel, on the other hand, would have varied degrees of similarity with each cluster. The optimization function in the fuzzy c-means method affects the correctness of the concerned output. Most image segmentation demands can be met by clustering methods.

2.1.b.vii) Deep learning based Segmentation[48,138]: In order to identify and process image data quickly and efficiently, convolutional neural networks are quite popular for image segmentation [48].Numerous deep learning-based methods have overrun the machine learning community. Diverse deep neural network architectures, including convolutional neural networks, recurrent networks, adversarial networks, and autoencoders, are effectively tackling many difficult computer vision tasks, including the detection, localization, recognition, and segmentation of objects in an unrestricted environment. Even though the object detection or recognition area has seen a lot of analytical research, various fresh deep learning methods for image segmentation have emerged. This essay takes an analytical approach to these numerous deep learning picture segmentation techniques. This research's main objective is to give readers a basic grasp of the key methods that have significantly advanced the field of picture segmentationThe article begins by discussing some of the conventional methods for segmenting images before showing how deep learning has changed this field. The majority of the key segmentation methods have since been logically grouped, with paragraphs devoted to each one's distinctive contribution. The reader should be able to better visualize the internal dynamics of these processes with the help of an abundance of intuitive explanations.

2.2. Object Localization[62]: The task of object localization involves locating specific objects within a given image by defining a tightly cropped bounding box centered on the object. Localization of images is a spin-off of the regular Convolutional Neural Network (RCNN) algorithm. Discrete classes are predicted by these algorithms. A bounding box is drawn around an object of interest based on a set of continuous numbers, namely a set of four coordinates, x, y, height, and width coordinates.

2.2.a) Single Stage: One-stage methods strive for real-time speed while retaining excellent performance. One of the earliest contemporary single-stage object detectors based on deep networks is OverFeat [125]. By skipping the region proposal generating step in favor of immediately predicting categorization scores and bounding box regression offsets, YOLO [126, 127] and SSD [128] have rekindled interest in one-stage techniques. Recently, Lin et al. noted that the performance is limited by the strong foreground-background class imbalance and proposed Focal Loss [129] to improve accuracy. In order to forecast scores and offsets at each localization, the majority of one-stage detectors use the sliding window approach and fully convolutional networks, which is advantageous because it reduces the computational complexity.

- **YOLOS[100] :** The model You Only Look Once at Sequence proposed by Yuxin Fang et al. to transform computer vision through object detection [ ]. Carion et al. proposed that the original vision transformer model be closely followed by YOLOS [12] which is also adjustable to various canonical transformer architectures that are available in computer vision and natural language processing [13]. In YOLOS, object detection is handled by the plain vision transformer, which is based on the detection transformer (DETR). In fact, a basic encoder-only transformer can also achieve 42 AP with COCO, just like DETR, which is a more complex architecture similar to Faster R-CNN.

ARCHITECTURE:

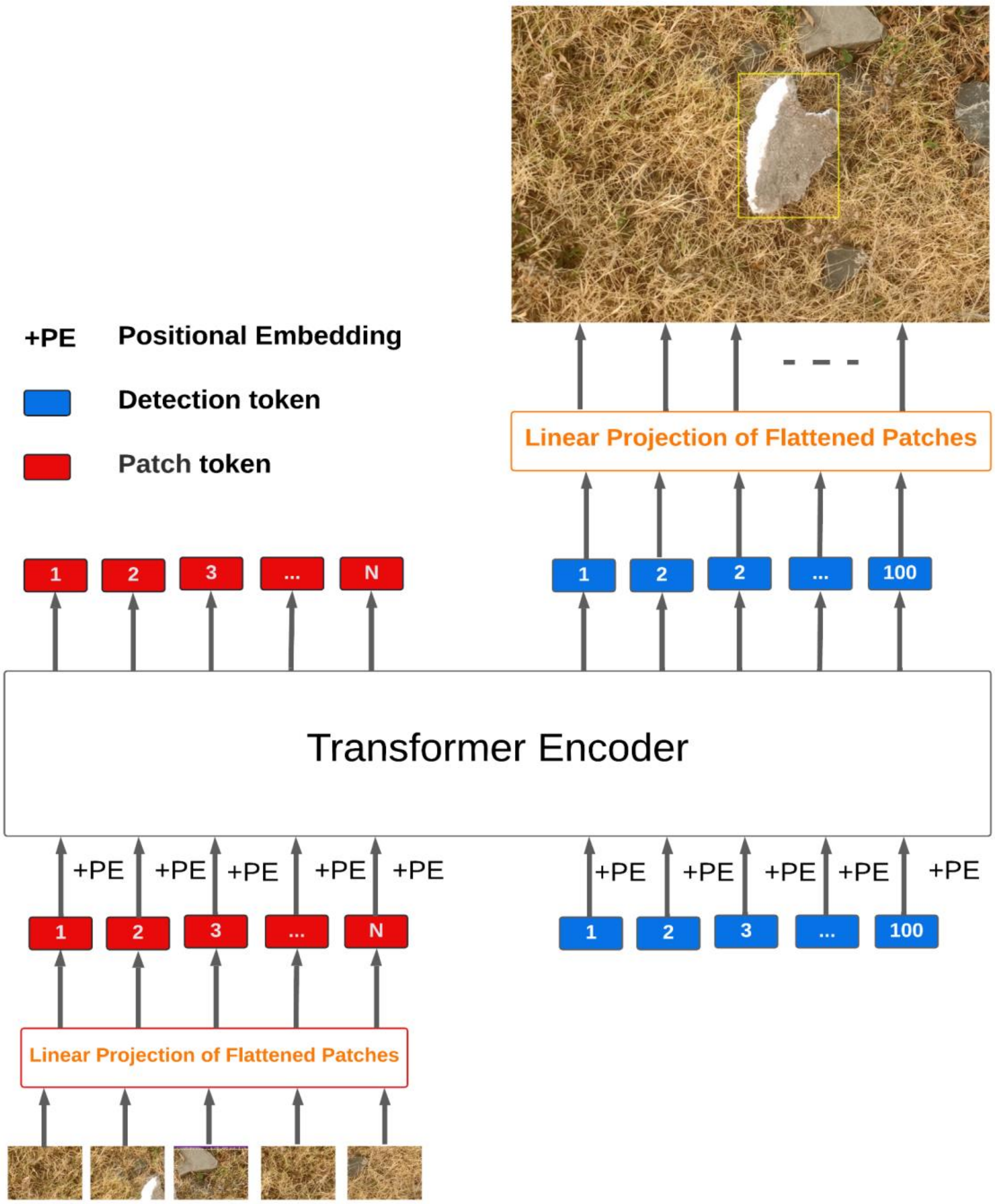


Fig 2.4
YOLOS Model Architecture

YOLOv5[101]: It is an object detection technique by which images can be divided into grids. Those grids are taken into account for object detection . YOLOv5 is smaller in size and easier to use in production. Due to its native implementation in PyTorch and not Darknet, it is straightforward to modify the architecture and deploy to many environments. The Pytorch community is larger compared to the Darknet community, so the Pytorch community may have more contributions and growth potential in the future. Using YOLOv5[105] is easier because the algorithm is written in Python, and for that reason, installation and integration are quite simple on IoT devices.

ARCHITECTURE:

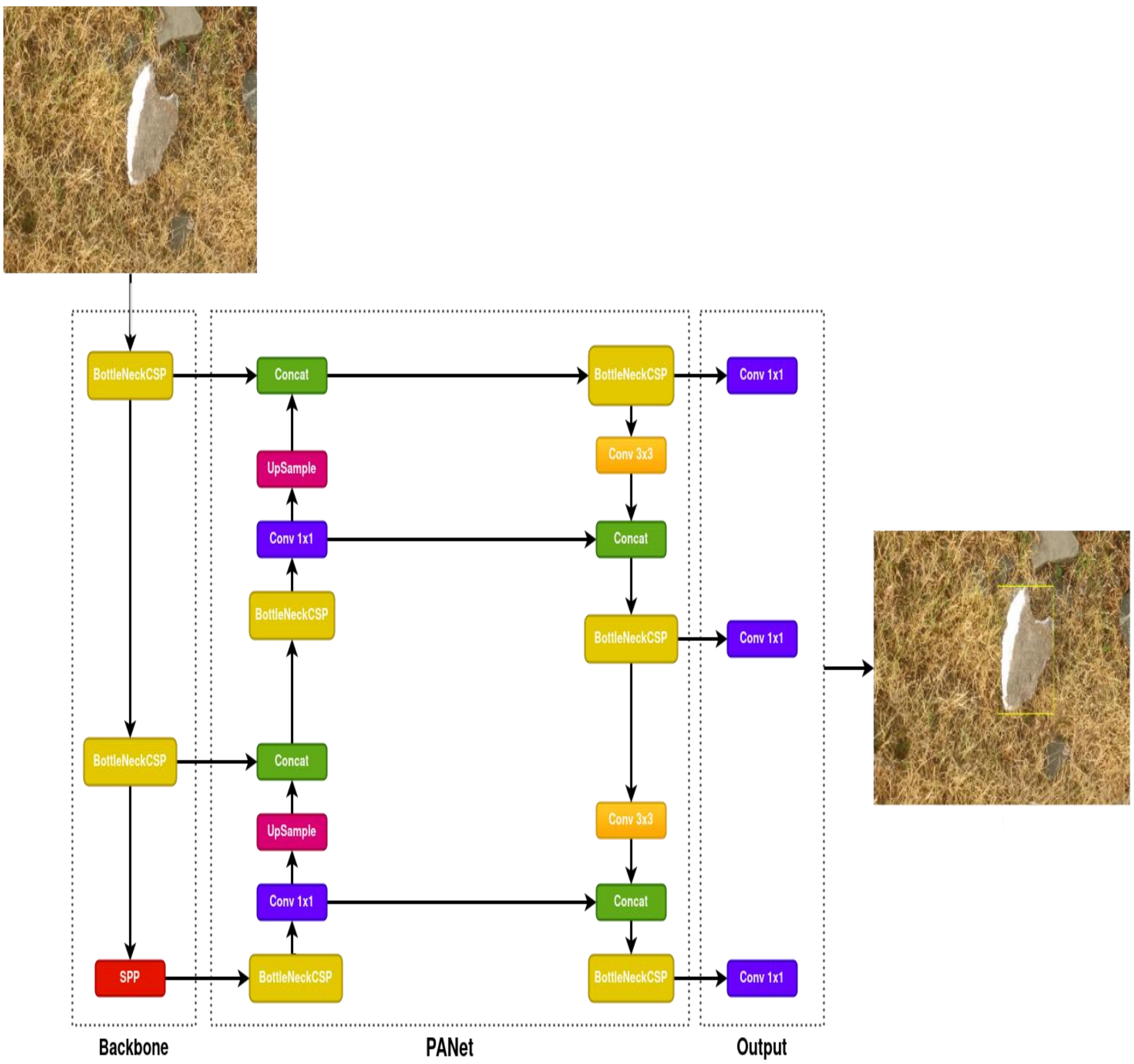


Fig 2.5
YOLOv5 Model Architecture

- **YOLOX[102]:** YOLOX is a new version of the YOLO model that pushes the accuracy and speed limits. It is the recent winner of the Streaming Perception Challenge. The computer vision model YOLOX is a modified version of YOLOv3 and the DarkNet53 backbone. With that model, we can detect the object in a single object. To be more specific, the YOLOS head has been decoupled. Initially, each FPN feature is reduced to 256 using a 1 x 1 convolution layer, and then two parallel branches with two 3 x 3 convolutional layers each are created to perform classification and regression tasks. YOLOX's augmentation strategies include Mosaic and MixUp. YOLOX is also anchor-free since the anchor mechanism is removed. In the last case, SimOTA simulates a top-k strategy for label assignment in which the label assignment problem is formulated as an optimal transport problem.

ARCHITECTURE:

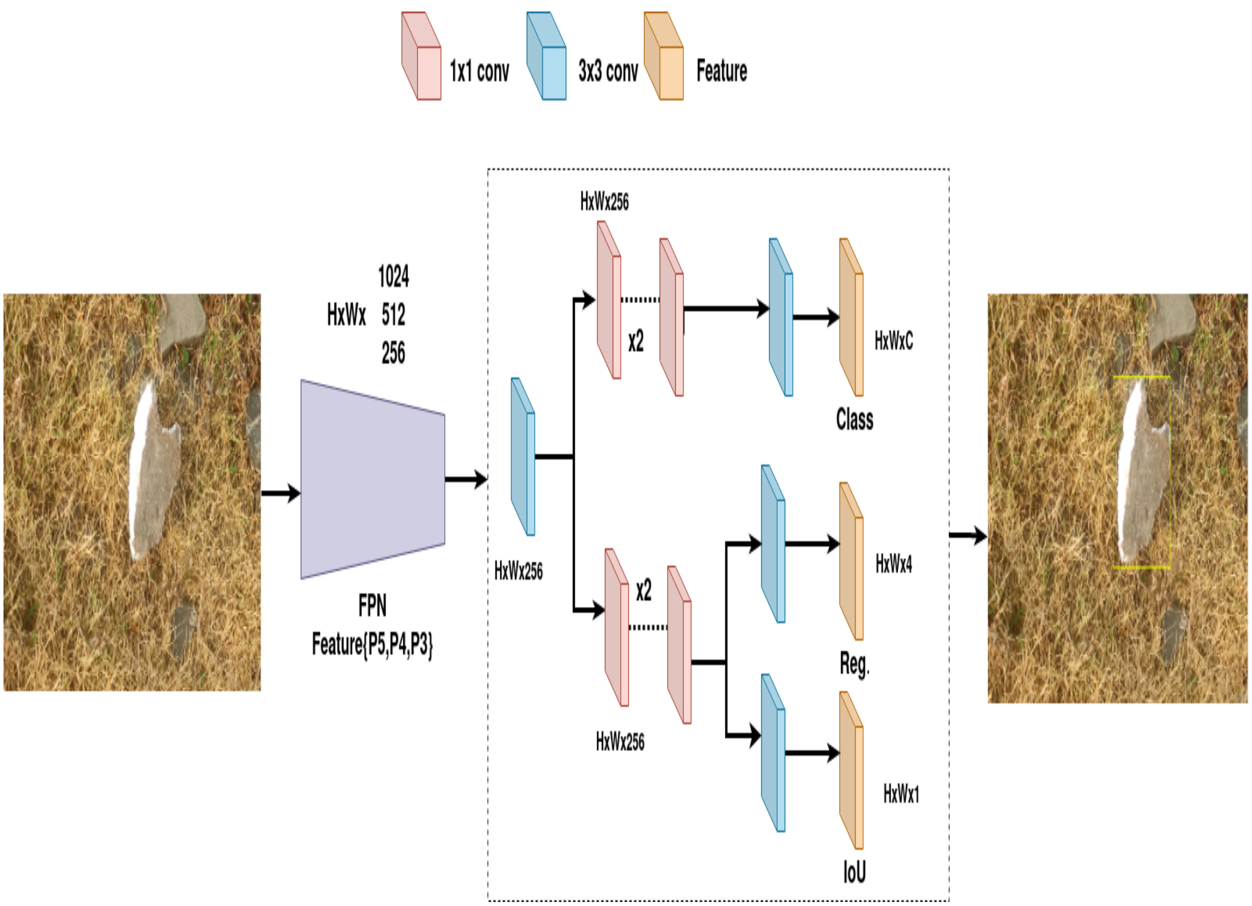


Fig 2.6
YOLOX Model Architecture

2.2.b.) Two Stage: Faster R-CNN [130], a representative of two-stage techniques, has been the dominant paradigm in the modern period of object detection, achieving the best results on a variety of benchmarks [131, 132, 133]. To improve performance, a number of improvements to this framework have been suggested, including the use of cascade methods [134], the adoption of multi-task learning schemes [135], and the construction of feature pyramids [136].
The field of computer vision has experienced a rapid revolution in object detection. But here we discuss some of them below.

- **Region-based Convolutional Neural Networks (R-CNN) :** RCNNs combine convolutional neural networks (CNN)[13] with region-based proposals. By using a deep network, R-CNN allows for the localization of objects and the training of a high-capacity detection model based only on a small amount of annotated detection data. A deep convolutional network is used to classify object proposals in order to achieve excellent object detection accuracy. In addition to its ability to scale to thousands of objects without the use of approximate techniques, such as hashing, R-CNN is fast. RCNN reached high object detection quality, but it has some significant flaws. Training in a multistage pipeline, for example, is slow and difficult since each stage must be trained separately. In addition, additional resources and time are required to train the SVM classifier and the BBox regressor, respectively. Finally, testing takes a long time since CNN features must be retrieved for each object proposal in each testing image without the use of shared computing. Because of the limitations of RCNN, additional techniques were developed, leading to the creation of enhanced detection frameworks such as Fast RCNN and Faster RCNN.

ARCHITECTURE:

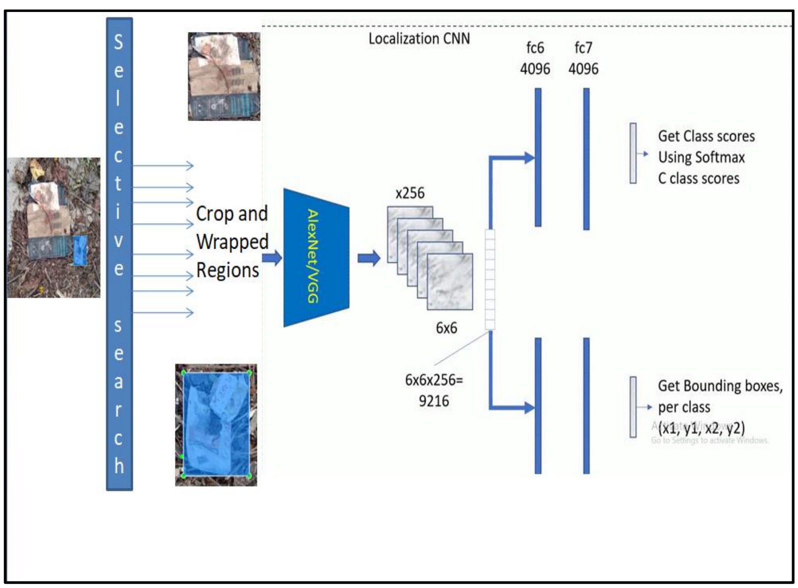


Fig 2.7
R-CNN Model Architecture

- **Fast R-CNN[13]:** Compared to R-CNN and Fast R-CNN, this algorithm utilizes the Region Proposal Network, which shares convolutional features throughout the image with the detection network. Fast RCNN [13] increased the object identification ability of RCNN by addressing some of its flaws [14]. The detector is trained from beginning to end in Fast RCNN[63]. It accomplishes this by speeding up the training process by learning the softmax classifier and the class-specific BBox regression at the same time, rather than training each component of the model separately like RCNN does. Fast RCNN[63] divides convolution computation between area proposals and then adds a ROI pooling layer between the last convolution layer and the first fully connected layer to extract features for each region proposal. To achieve image level warping, ROI pooling employs the concept of feature level warping. The characteristics of the ROI pooling layer are divided into a series of fully connected layers that branch out into two layers: object category prediction softmax probability and class proposal refinement offsets. Fast RCNN enhances efficiency significantly when compared to RCNN, for example, by three times the training speed and ten times the testing speed.

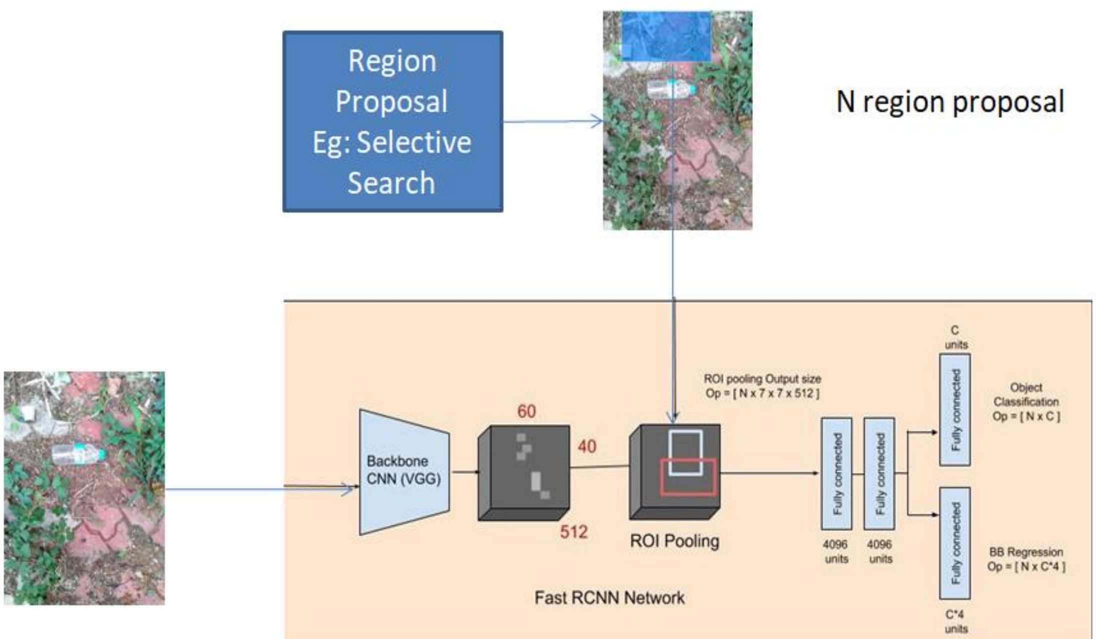


Fig 2.8
Fast R-CNN Model Architecture

- **Faster-RCNN:** Despite the fact that Fast RCNN increased detection speed significantly, it still relied on external region recommendations, which were the bottleneck in Fast RCNN [78]. Work [81,82] at the time showed that CNNs can localize objects in convolutional layers, but this ability is weakened in fully connected layers. Ren et al. [79] introduced a faster RCNN model that had a Region Proposal Network (RPN) for the creation of region proposals, which was efficient and accurate. To accomplish region proposal by RPN and region classification by Fast RCNN, the same backbone network was employed, with features from the final shared convolutional layer.

ARCHITECTURE:

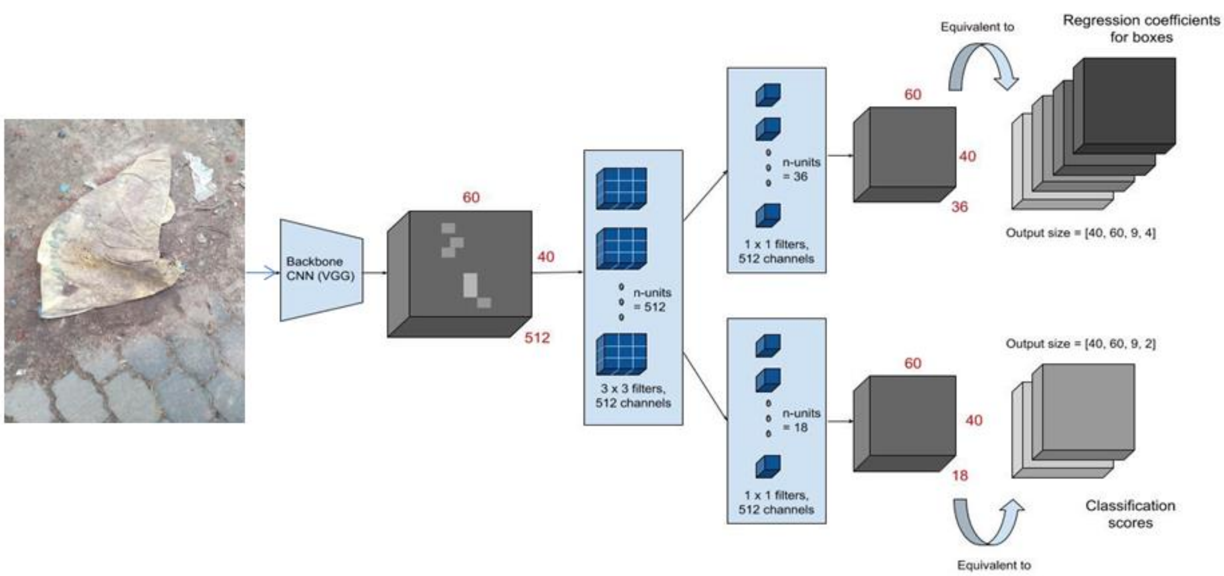


Fig 2.9
Faster R-CNN Model Architecture

- **Mask R-CNN[49]** is a deep learning architecture developed by Facebook AI Research (FAIR) that can construct pixel-wise masks for every object in a picture. The Faster R-CNN object detection architecture has been improved. For every object in an image, the Faster R-CNN uses two pieces of information: the bounding box coordinates and the object's class. I have gained an extra section in this process with Mask R-CNN. After performing segmentation, Mask R-CNN outputs the object mask. I should start by passing the input image to the ConvNet, which would then generate the image's feature map. The system then uses the region proposal network (RPN)[50] to generate object proposals with object ratings based on the feature maps. The Roi pooling layer is then applied to the suggestions to reduce them to a single size. The system then sends the proposals to the linked layer for classification, and generates an output with bounding boxes for each object.

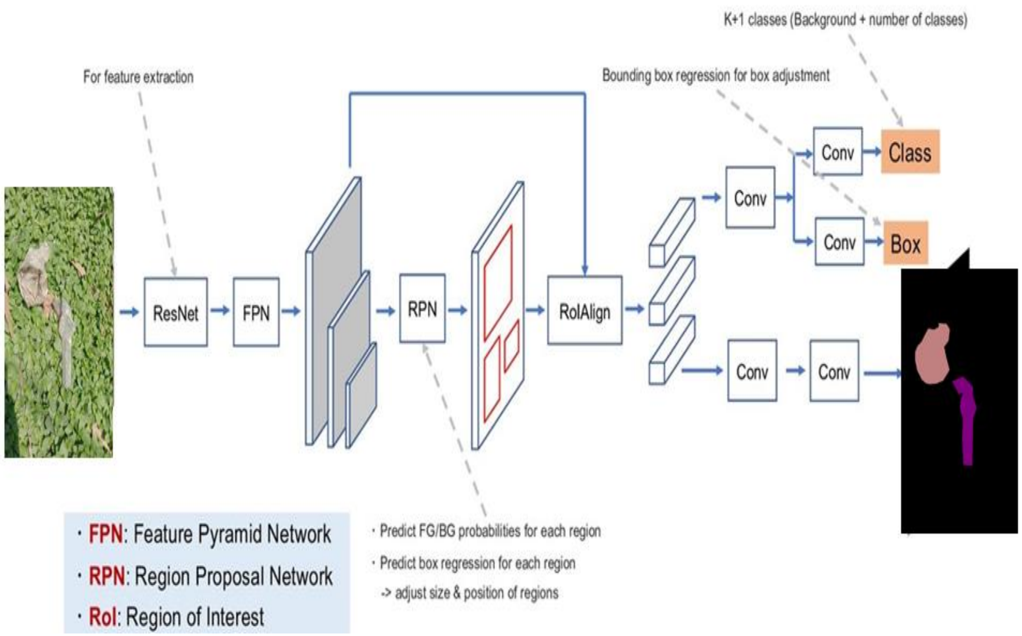


Fig 2.10
Mask R-CNN Model Architecture

- **Region Based Fully Convolutional Neural Network:** Using R-FCNN, one can determine a region-based fully convolutional neural network that detects objects. A region-based fully convolutional network called R-FCN [20] was first suggested for object detection. R-FCN builds a deeper fully convolutional network without raising the speed overhead by sharing processing on the entire image, in contrast to existing region-based detectors like Faster RCNN [20]. In contrast to the R-CNN module that was presented in

Faster R-CNN [20], R-FCN is built upon 101-layer ResNet [21], and it consists of a region proposal network (RPN) and an R-FCN module.The ResNet architecture serves as the feature extractor in R-FCN. The fact that ResNet builds a very deep network capable of extracting highly representative image attributes is well known.These characteristics have a substantially wider receptive field, allowing for the use of background information in microscopic face detection. RPN creates a batch of region of interests (RoIs) in accordance with the anchors from the feature maps that the basic ResNet outputs. To create class score maps and bounding box prediction maps, these RoIs are subsequently input into two sibling position sensitive RoI pooling layers in the R-FCN module. For aggregating the class scores, two global average pooling methods are applied at the end of R-FCN on both class score maps and bounding box prediction maps, respectively.

ARCHITECTURE:

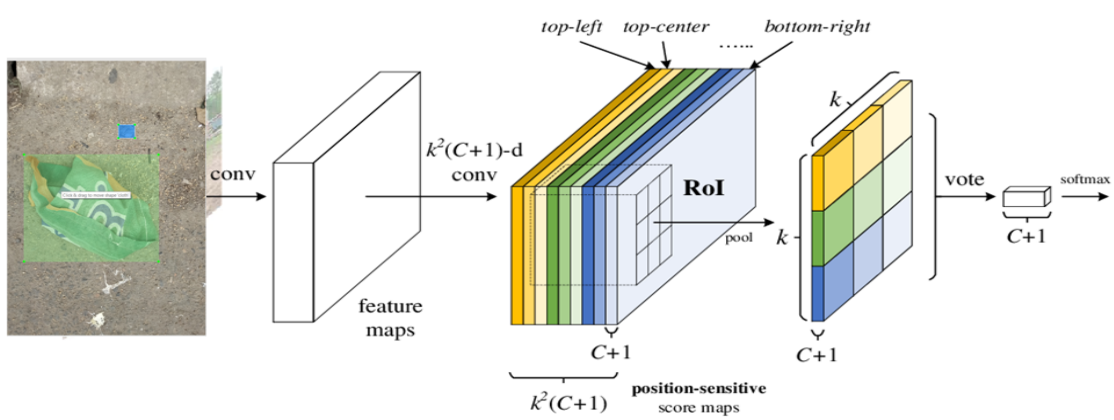


Fig 2.11
R-FCNN Model Architecture

A review on waste management , detection and identification system

In a paper by X. Wen et al.[86], the Chinese authorities tried to solve the e-waste management problems from technical aspects. Actually, they wanted to introduce and analyze the flow of electric and electronic waste management , recycling to use for other work. For environmental pollution control , disposal of waste is more important . In another paper by F. Yuan et al. [87], they tried to design waste information on the basis of WebGIS to improve management level and efficiency. Basically, this paper aims to manage solid waste through the use of information. technology. M. Yang and G. Thung proposed a method to classify single object waste images by support vector machine (SVM) with the use of scale-invariant feature transform (SIFT) and a CNN. [65] Also, they proposed using object detection networks like Fast R-CNN and low cost networks like Region Proposal Networks (RPNs)[64]. Mr. Zou proposed that computer vision has been successful at identifying objects by placing an Axis-Aligned Bounding Box (AABB)[69] on a particular part of the image and then labeling it. The detection of objects has traditionally been addressed with two techniques, namely one-stage and two-stage detection. Two stage detectors initialize their learning with class-agnostic object proposals, before classifying them into class specific objects at the next stage. The one-stage architecture provides both object locations and classes on a single stage. Object detection began with a two-stage method, which is used to slide the window technique. The image pyramid is used to initiate object proposals in numerous scales, and the next step was a cascade classifier, proposed by Viola and Jones in 2001[88].

By combining advanced communication methods with a gray level co-occurrence matrix (GLCM)[89], the author here proposed a method of waste classification and detection that speeds up the waste identification process. An integrated solid waste monitoring system is proposed with the use of Geospatial Information System (GIS)[90] , General Packet Radio System (GPRS) and Radio [89], Frequency Identification System (RFIS)[90]. In the obtained outcome , a k-nearest neighbor (KNN) classifier outperformed the multilayer perceptron (MLP) for waste segregation . Highlights from the GLCM were then input into a MLP for waste segregation. Using a deep learning approach combined with conventional recycling strategies [74], the author suggests a computerized framework that consolidates waste into four distinct recycling classes (paper,metal, glass, and plastic). In the best case scenario,the VGG-16 [75] methods achieve an accuracy rate of 93%, showing their productivity for this issue. An automatic garbage bin has been created at the TechCrunch Disrupt Hackathon that sorts garbage based on recycling and composting features, called "Auto Trash" [76]. To detect objects, the group built a layer on top of Google's TensorFlow AI engine, and for the framework they used a Raspberry Pi camera with a pivoting top.

Recently, deep learning techniques and computer vision have been used as instruments in waste classification, and they have achieved sensible results in controlled situations.With the help of the Faster R-CNN model, Oluwasanya Awe et al [77] investigated object identification in waste management, demonstrating reasonable outcomes. Using Faster Region-based Convolutional Neural Networks (Faster R-CNNs) [78] for region and object classification, the author proposed a strategy with an average precision of 68.3%.

Caruana, 1997 introduced the concept of transfer learning to the machine vision community, and it is widely used now. Zhang and Yang, 2018 [64]used that technique to improve the performance of an image classification task .Multiple DL structures have been implemented by Hafiz and Bhat, 2020[91] to solve the instance segmentation problem. These structures return the categories, bounding boxes, and segmentation masks of objects in the images.The MTL framework proposed in this thesis work can handle both the location of waste and the classification of waste using multi-labels. A fundamental and important aspect of computer vision is the detection of objects of defined categories in images. This was proposed by Zou et al., 2019. [69] Anchor-based detectors and anchorfree detectors are the two main types of DL-based detectors. Dhillion and Verma, 2020 proposed that the SOTA[92]methods of anchor-free detectors offer a better trade-off between precision and speed than traditional, handcrafted detectors. In terms of anchor-based detectors, the R-CNN and RetinaNet frameworks represent the most comprehensive methods of classifying and registering regions from images for the object detection problem. Those recent anchor-free approaches are Global Forecasting Models, which were proposed by Li et al., 2020[78] and AutoAssign Zhu et al., 2020[69] ; they introduce a center-weighting module to improve quality assessment, classification, and localization in the MTL framework. Rather than categorizing an image into one or more classes, multi-label classification generalizes the single-label image classification problem (Wu et al., 2020)[93]. A number of CNN-based classification schemes have been proposed and implemented since the rapid development of CNN. To train their classifier for the multi-label classification task, Hsu et al.[94] used the compressed sensing method . The use of Deep Learning for waste recognition has been explored in numerous important studies, as is evident by the work carried out by Hannan et al., 2015[95] .As an example, Adedeji et al. used a convolutional neural network to extract features from waste and classify it into four categories, including glass, metal, paper, and plastic, each of which are recyclable, and achieved an accuracy of 87%. DL-based convolutional neural networks (CNNs) were used to identify and classify e-waste and faster regions-based convolutional neural networks (Faster R-CNN) to locate and recognize waste in images by Nowakowski & Pamua, 2020[96]. Their self-collected dataset was used to train and evaluate their model, with a best accuracy of above 90%. An optimization method combined with the multi-label classification task in our study was utilized to train the multi-label classification and waste location tasks[94].

The Fisher Vector is a probability-based representation of VLAD [81] and can be characterized as a probabilistic version of VLAD . VLAD is a method for encoding residual vectors with respect to dictionaries with respect to images. The two types of shallow representation are powerful for both the retrieval and classification of images. It is shown that residual vectors [80] are more effective than original vectors when used for vector quantization.

A widely used method for solving partial differential equations (PDEs) in computer graphics and low-level vision [66] reformulates the system as multiple subproblems at different scales, each responsible for its own residual solution between two scales.Multigrid can be implemented using hierarchical basis preconditioning [97, 98], which uses variables representing residual vectors between two scales. These solvers converge much faster than conventional solvers that ignore residual nature, as shown by [66, 99].These methods suggest that optimization can be sped up or simplified with good reformulation or preconditioning.

It has been studied for a long time how shortcut connections might occur [2, 33, 48].Multi-layer perceptrons (MLPs) are initially trained by adding a linear layer between the input and output of the network [33,48]. As shown in [43, 24], augmentation classifiers address vanishing/exploding gradients by connecting a few intermediate layers directly to them. In their papers [38, 37, 31, 46], they propose shortcut connections as a method of analyzing layer responses, gradients, and and error propagation. A shortcut branch and a few deeper branches make up an “inception” layer.

Our identity shortcuts are parameter-free, whereas these gates have data-dependent parameters. Layers in highway networks represent non-residual functions when a shortcut gate is “closed” (approaching zero). We always learn residual functions from our formulation, because identity shortcuts are never closed, and all information is always passed through, which requires the learning of additional residual functions. Furthermore, the accuracy gains with increased depth (such as more than 100 layers) have yet to be seen with high-way networks.

An automated waste segregation system using the Internet of Things, with Arduino microcontrollers, is presented in this study[100]. The research proposed a system that separates waste into wet and dry categories. Once the bins are separated, the system monitors the level of garbage in each one, sending the level of the bin along with the unique ID provided for that bin when it reaches its threshold level. A gas sensor can be added in order to improve this system. Combustion of decayed waste releases toxic gases such as methane. This gas causes headaches and difficulty breathing. As soon as the gas sensor detects gas, a message will appear notifying the receiver to remove the trash from the bin.

U-net is a convolutional neural network[104] which is basically used for image segmentation. It works in two paths, which are the contracting path and the expansive path. A contracting path is also called an analysis path or encoder path. It is like a regular convolutional neural network and works for image classification. This path is made up of up-convolutions and concatenations of features that are derived from the contracting path. The network is able to learn localized classification information through this expansion. Decoding or synthesis is the next step, which combines these two paths to form a new classifier. Its almost symmetrical shape gives it a u-shaped appearance. In addition to increasing the resolution of the output, it also produces a segmented image. The UNET task was created by Ronneberger et al. [104] and is able to segment medical images much more accurately. In U-net, fully convolutional networks were used in accordance with the work of Long, J. et al. On the ISBI 2012 challenge, their implementation outperformed the previous best. In 2015[105], it beat the state of the art at the ISBI cell tracking challenge. By using a novel method called random elastic deformation on the training data[106], it learns how the shapes and sizes of images vary over time. Separating overlapping objects of the same class can also be challenging, and by using a weighted loss function, the model is penalized if it fails to separate them. Since U-net was established in 2015, its use in medical imaging has exploded. The purpose of this survey is to examine how image analysis and data analysis in medicine have been impacted by the Internet of Things[107].

Convolutional Neural Networks have a limitation that is facing problems in handling large amounts of data sets. But it has basically existed for a long time. Several visual recognition tasks have been significantly outperformed by deep convolutional networks. It was required to administer training of a large image net dataset with millions of parameters and training images. A large margin separated the winning network from the nearest competitors in the EM segmentation challenge at ISBI 2012[107].

The network can only see limited context in large patches and greater layers of max pooling reduce localization accuracy. In recent approaches [105, 106], multiple layers were used to classify the output of the classifier. This improves localization at the same time. My approach in this thesis is to utilize a more elegant architecture, referred to as the "fully convolutional network.".This architecture is modified and extended so that it can be used with very few images for training. There is less training data to analyze, and the segments are more precise. For localization, the contracting path's high-resolution[109] features are combined with upsampled output. Using this information, the subsequent convolution layer will be able to produce a more precise output. For localization, information about context is obtained from the network's feature channels, which can be propagated to higher resolution layers. This produces a u-shaped architecture as the expansive path is more or less symmetrical to the contracting path. An overlap-tile method of segmenting large images using a neural network has been developed. There are no fully connected layers in the network, and the segmentation map only contains pixels that have full context in the input image, i.e., the network uses only part of each convolution. For larger images, the strategy is crucial since otherwise the processor would limit the resolution. In addition, they applied elastic deformation to the training images in order to enhance the corpus with excessive data augmentation[110]. Hence, the model can learn that the deformations are invariant without the need to examine the annotated corporation. Biomedical engineers must understand how tissues change over time under a variety of conditions, so deformations are especially important in this context. The loss function we propose is weighted so that the background labels between touching cells are given a large weight. Additionally, we show results from the ISBI cell tracking challenge 2015 for segmenting cells in light microscopy images.

In this network, volumes are processed with 3D operations based on 3D inputs[111]. Convolutions in 3D, maximum pooling in 3D, and up-convolutions in 3D are among these operations. The network architecture is designed to avoid bottlenecks, and batch normalization is used to accelerate convergence.

To generalize to a third volumetric image, I trained the neural network on just two images in this thesis. The network can be trained with a few manually annotated slices, i.e., from sparsely annotated training data, with a weighted loss function and special data augmentation[113].

Xenopus kidneys[112] are complex structures, which limit the functionality of predefined parametric models. The proposed method was successfully applied to data from confocal microscopic images that were difficult to process.

Two-dimensional biomedical images are segmented[113] for generating the region of interest (ROI) using convolutional neural networks (CNN),with an accuracy that is close to human accuracy. Currently, CNNs are being employed to build up three-dimensional ROY. Milletari et al.[114] proposed a procedure for creating three-dimensional data with a Hough Voting approach. However, they only work for compact structures that resemble blobs, and their method is not end-to-end. Kleesiek et al. proposed [115]an end-to-end 3D CNN but it is not a deep learning network method and it has only one max pool layer after the first convolutional layer, so it can not be used to analyze the multiple structure scales. This work is based on a 2D U-net model which handles very good generalization performance from limited annotated samples. For semantic segmentation,up-convolutional networks can be used, which are similar to fully convolutional networks. In a paper by Tran et al. [116], it was demonstrated that the Up-convolutional architecture is applied to a video dataset for training. Researchers started from scratching sparsely annotated data and went on to work with huge amounts of data in a seamless tiling strategy using the above mentioned method.

LeCun et al.2015[117] and Konovalenko et al.2020[118] presented a study that machine learning and deep learning techniques are popular for face recognition software, text-to-speech conversion devices, vision for driverless cars, or object recognition in images and videos, etc. The main target is to sort the waste material by creating a machine learning-based recognition process, which is described by White et al. 2020, Sheng et al. 2020[143], and Glouche et al.2013.[119] Most of the approaches are based on deep learning algorithms and computer vision.

Object Segmentation Models: Some object segmentation models which I trained on my dataset are discussed below.

- **DeepLapv3+:** One of the most important computer vision tasks is semantic segmentation, which aims to assign semantic labels to each pixel in an image. In this instance, we constructed the DeepLabV3+ model for multi-class semantic segmentation, a fully convolutional architecture that achieves good results on benchmarks for semantic segmentation. A straightforward yet efficient decoder module is added to DeepLabv3+, a semantic segmentation architecture, to improve segmentation outcomes.

  Multiple CNN downsampling will result in a decreased feature map resolution, which will lower prediction accuracy and cause border information to be lost in semantic segmentation. Similar to how gathering context surrounding a feature aids in better segmenting it, atrous convolutions are used to do this. These problems can be resolved with DeepLabv3+.

ARCHITECTURE:

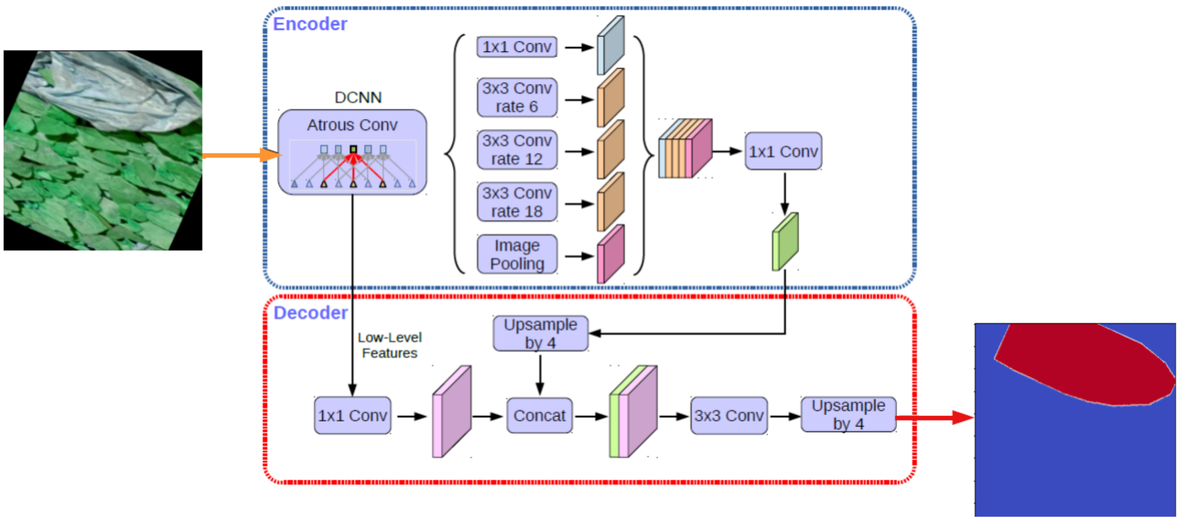


Fig 2.12
DeepLabv3+ Model Architecture

- **UNet:** Olag Ronneberger et al. developed the UNet architecture for biomedical image segmentation. The encoder and decoder were the two key components of the newly presented architecture. Covenant layers come first in the encoder, then pooling operations. It is used to extract the image's factors. Transposed convolution is used by the second part decoder to enable localisation. Once more, it is a network of layers connected by F.C. You may read U-Net: Convolutional Networks for Biomedical Image Segmentation, the original version of the paper. Also, read Understanding Semantic Segmentation with UNet, a publication that has more information on UNet architecture.

ARCHITECTURE:

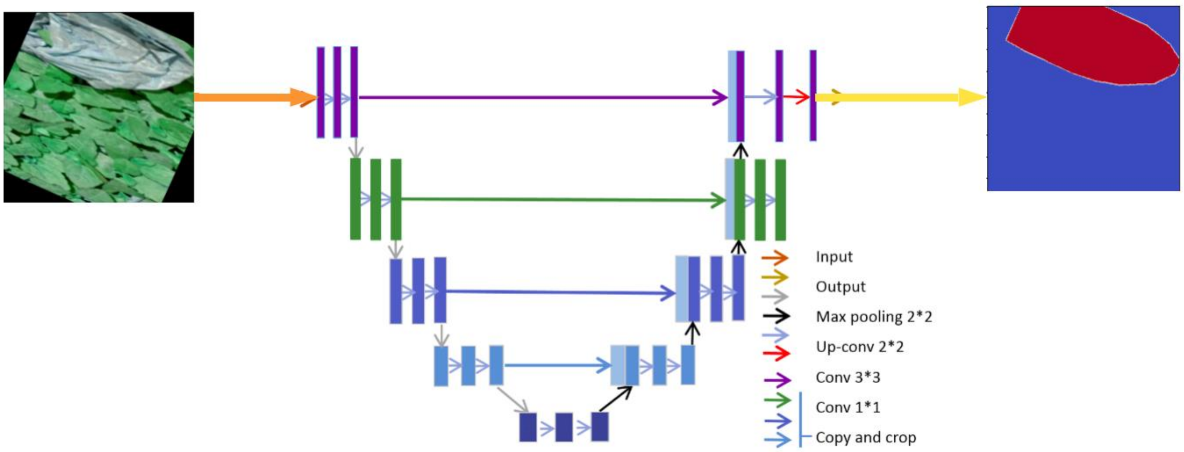


Fig 2.13
UNet Model Architecture

LinkNet: For the purpose of semantic segmentation, LinkNet is a light deep neural network architecture that can be applied to projects like self-driving cars and augmented reality. It can provide real-time performance on embedded devices like the NVIDIA TX1 and GPUs.

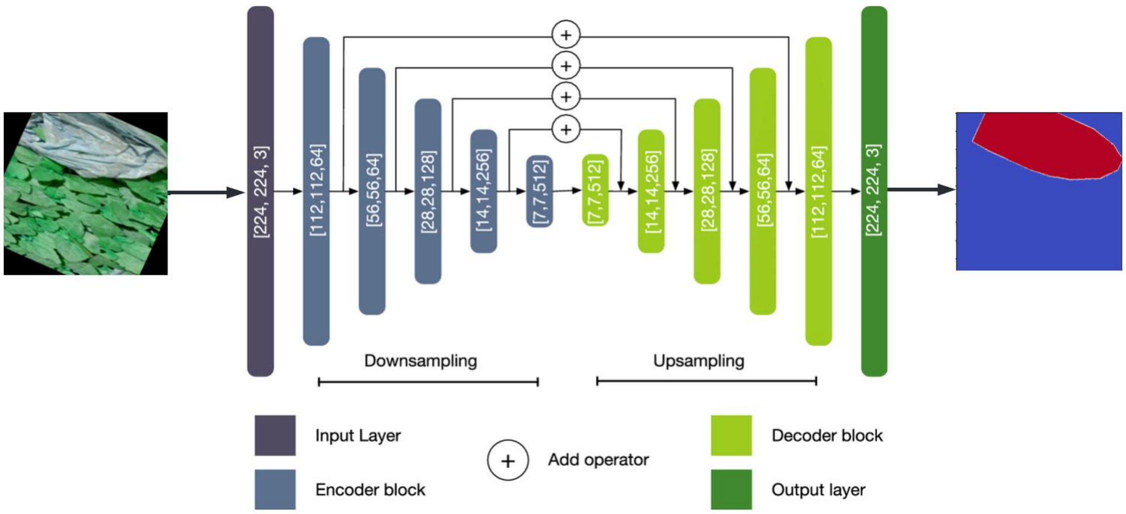


Fig 2.14
LinkNet Model Architecture

- **PSPNet:** Pyramid Scene Parsing Network, or PSPNet, is a semantic segmentation approach that makes use of a pyramid parsing module to take advantage of global context data through context aggregation depending on distinct regions. The final forecast is more reliable when the local and global hints are combined.

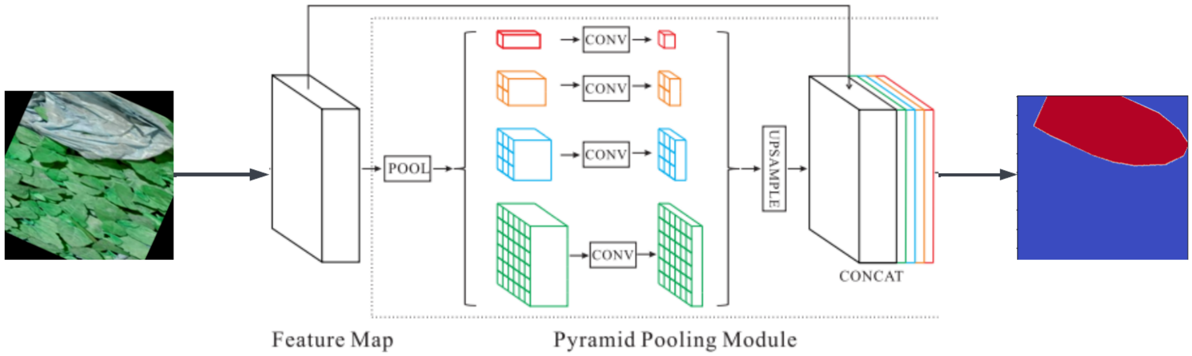


Fig 2.15
PSPNet Model Architecture

CHAPTER 3

THESIS MOTIVATION AND CONTRIBUTION

The improper management of waste contributes to environmental damage on a global scale. Massive manufacturing of disposable items in recent years has led to a huge increase in waste created, with 5.2 tonnes of rubbish being produced per EU resident in 2018, according to a report by the European household waste collection (Eurostat, 2018)[140]. Furthermore, the World Bank (WB) report (Kaza et al. 2018)[141] predicts that by 2050, there will be more than 3 billion tonnes of waste produced annually. According to the World Bank, just 13.5% of global waste gets recycled, while 33.5% of trash is thrown out in the open without being first sorted (Kaza et al. 2018)[141]. This results in the free dispersal of various kinds of trash in a range of situations. Plastic garbage is the most concerning issue since it is the most pervasive and causes long-term environmental harm (Li et al. 2016)[142]. Prompt action is required in order to assist in the careful collection and segregation of trash. This will help prevent additional environmental degradation and, as a result, will protect the existence of humans and other wild organisms.

- I have created a dense waste detection dataset for waste classification and localization which consists of 21 unique waste classes such as Plastic carry, paper, paper_cup, plastic, plastic_coantainer, paper_cardboard, cigarette_packet, plastic_bottle, newspaper, thermocol, Match_Box, metal, cloth, plastic_sack, medicine_wrapper, glass_bottle, glass, nylon, cotton, aluminium_foil, rubber. The dataset contains 6K high resolution images which were taken in different weather and daylight conditions.

- The dataset has been trained in various State-of the Art object detection models like YOLOv3, YOLOv4, Scaled YOLOv4, YOLOv5s, YOLOR, MT-YOLOv6 and YOLOv7s, and has created a noble benchmark for waste detection.

- Finding the actual location of the waste in densely populated waste is difficult using the object detection model. So I have created another dataset for waste segmentation which consists of 52 high resolution images of congested waste which were classified into 14 classes, i.e plastic-container, plastic-bottle, thermocol, metal-bottle, plastic-cardboard, glass,thermocol-plate,plastic, paper, plastic-cup, paper-cup, aluminum-foil, cloth,nylon.

- For the proper segmentation of waste, I have developed a new object segmentation model that got 92% accuracy on our dataset. On the other hand, the dataset has been trained on State-of the Art segmentation models like UNet and DeepLabv3+ which give an accuracy of 82% and 86% respectively.

CHAPTER 4

PROPOSED WORK

In this chapter, the total work with methodology was stated . At first I was described as collecting the object detection dataset, after how to annotate those data,and also stated about dataset health. Then I wrote about the model training and validation. At last Iwas clarifying the result with accuracy.

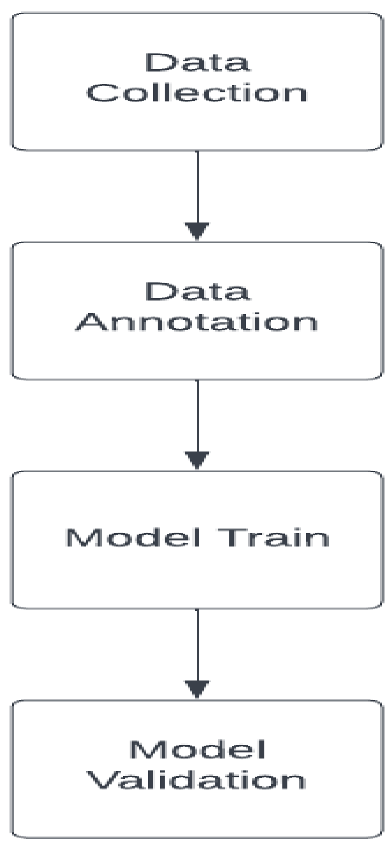


Fig 4.1
Flowchart of Thesis work

4.1) Dataset Collection: I have collected the images simply using my mobile and drone at Jadavpur University, Kolkata and my home town, Berhampore, West Bengal region. In that data set, many types of variables and conditions are there. I have taken the images in land surface like at bush, at municipality drain , at road, at station platform also taken at water surface like besides ponds, after rain, water culvert etc. I have collected the images in the daytime as well as at night.

In that data set I have taken some biodegradable and some non-biodegradable waste. The biodegradable wastes are paper, paper cups, paper cardboard, cloth, etc. and the non-biodegradable wastes are plastic, plastic bottles, glass, glass bottles, metal, plastic, thermocol objects, rubber, nylon, plastic containers, plastic sacks, etc. This data set is used for two different models: the segmentation model and the localization model.

4.1.a)Dataset for Object Segmentation : It is a process by which we can observe the image data information and can reduce the complexity. For this, we have to break the images into subgroups to facilitate analysis.

Here I have presented some of the images that I have collected for creating the segmentation dataset.

Table 4.1
Examples of object segmentation datasets.

The above images are collected from several places in the Indian environment . The waste in these images is present in a very congested manner compared to the images of waste presented in previous works. As the images are collected from waste dumps, they have many types of waste overlapping with one another. In my work I have considered 21 different classes, which are plastic containers, plastic bottles, thermocol, metal, thermocol, plastic cardboard, paper,glass, plastic cups, plastic paper cups, aluminum foil,etc.

4.1.b)Dataset for Object Localization: This type of data can also be similar to the previous dataset. But I have collected the new data in the same scenario. Here the number of dataset is much more compared with previous. The waste objects are labeled in a rectangular bounding box by a software called Labelimg.

Here I have presented some of the images that I have collected for creating the localization dataset.

Table 4.2
Examples of object localization datasets.

★ **Statistics of the classes in the images:**

Classes	Object Type	No. of object	Annotation Type	Environment	Instances
Plastic carry	Polyethylene	7443	Bounding box	Out Door	>3k
Paper	Paper	2574	Bounding box	Out Door	>1.5
Paper cup	Paper	1807	Bounding box	Out Door	>1k
Plastic	PVC,PET	914	Bounding box	Out Door	>500
Plastic container	PET	659	Bounding box	Out Door	>400
Paper cardboard	Paper	585	Bounding box	Out Door	>400
Cigarette packet	Paper	563	Bounding box	Out Door	>400
Plastic bottle	Polypropylene and polyethylene	335	Bounding box	Out Door	>250
Newspaper	Paper	323	Bounding box	Out Door	>250
Thermocol	Polystyrene	265	Bounding box	Out Door	>200
Match Box	Paper	215	Bounding box	Out Door	>150
Metal	Iron,Copper, Aluminum	210	Bounding box	Out Door	>150
Cloth	Cotton,Silk	171	Bounding box	Out Door	>100
Plastic sack	Polyethylene	114	Bounding box	Out Door	>80
Medicine wrapper	Tin,Aluminum, Led	104	Bounding box	Out Door	>80
Glass bottle	Sand,Soda ash, Limestone	103	Bounding box	Out Door	>80
Glass	Sand,Soda ash,Limestone	99	Bounding box	Out Door	>80
Nylon	Polyamide	95	Bounding box	Out Door	>50
Cotton	Cotton	84	Bounding box	Out Door	>50
Aluminium_foil	Aluminum	66	Bounding box	Out Door	>40
Rubber	Rubber	41	Bounding box	Out Door	>25

Table 4.3 Class Statics

Plastic containers: Plastic containers are used to carry food, water, different types of liquid, as well as to pack cosmetics, groceries, medical instruments, etc. Most of the containers are for one-time use , and some of them may be used a few more times .Because of this, plastic waste is hugely increasing day by day. To reduce this type of waste, we should reuse the containers.

Plastic Bottle: Plastic bottles are widely used for multiple purposes. We are so dependent on it for storing and serving water, beverages, syrups, liquid pesticides, liquid petroleum, etc.

Plastic Cardboard: Thick plastic PVC is used to cover something, make furniture, as well as kitchen accessories like tubs, buckets, baskets, dustbins, etc. As these types of plastic accessories have a longer lifetime, the waste produced from them is lower and they are also recyclable.

Paper: Paper is biodegradable waste. It is used for many types of work in our daily life, like food packing,wrapping,decorating,writing,cleaning, printing, toilet tissue, etc. We can also reuse some waste paper like newspapers, books, exercise books,etc. So paper waste is not that much harmful for the environment.

Paper Cup: It is used to drink tea,coffee,water, alcohol,etc. It is similar to paper waste. So we do not consider it separately.

Thermocol: Expanded polystyrene, also known as thermocol, is a material used for packaging expensive products such as televisions, refrigerators,computers, and many other electronic and electrical instruments. Thermocol is not recycled once it is used. So this is harmful waste.

Metal: Every day, many kinds of metals in enormous quantities are employed in industrial processes. Due to the mass production of commodities and the accompanying low unit price, our consumption levels have risen since the industrial revolution. Aluminum is the metal that is used the most throughout the world, followed by copper, zinc, lead, and nickel. Metal bottles are made of stainless steel or aluminum. Compared to plastic bottles metal bottles are durable,odor less, retain and lifetime also more. This type of waste is very slow to degrade and it is also recyclable.

<u>Aluminum Foil</u>: The use of aluminum foil is involved in the consumption of fossil fuels, polluting the environment, affecting health and producing greenhouse gasses that are higher than those of plastic wrap. It can be recycled if food residue is not present.

<u>Cloth</u>: Cloth waste is basically produced by the textile industry. As a result of every step of the textile manufacturing process, such as spinning, weaving, dyeing, finishing, and garment making, textile waste is produced. This type of waste is slowly degradable.

<u>Nylon</u>: Polyamides are repeating units linked together with amide links to form nylon, a family of synthetic polymers. Polyamide is a silk-like thermoplastic made from petroleum, which is commonly mass-produced into fibers, films, and shapes. Nylon is used for several applications, like in the textile industry, in rubber material, many injection materials for vehicles, and to make rope or thread, etc. Though it is a non-biodegradable waste, many companies are trying to reuse it.

<u>Glass</u>: Glass is a solid material that is non-crystalline, transparent, and amorphous. It is used for windows,tableware,packing,laboratory instruments, optics,art, glass walls , etc. It is a non-biodegradable waste. The primary source of glass found in municipal solid waste is beverage containers, such as beer and soft drink bottles, wine bottles, and bottles for food and cosmetics. The quality and purity of glass remain largely unchanged even after thousands of recycling cycles.

4.2) Datasets Annotation for Object Segmentation: The dataset for object segmentation is labeled by a software called Labelme. The waste object is labeled with a polygon of a matching shape by the software and mentions its name as discussed above. In the Indian scenario, the waste is more congested and all types of waste are mixed . That is why labeling is done more carefully, making it time-consuming work. Now I am showing the images that are labeled using Labelme Software.

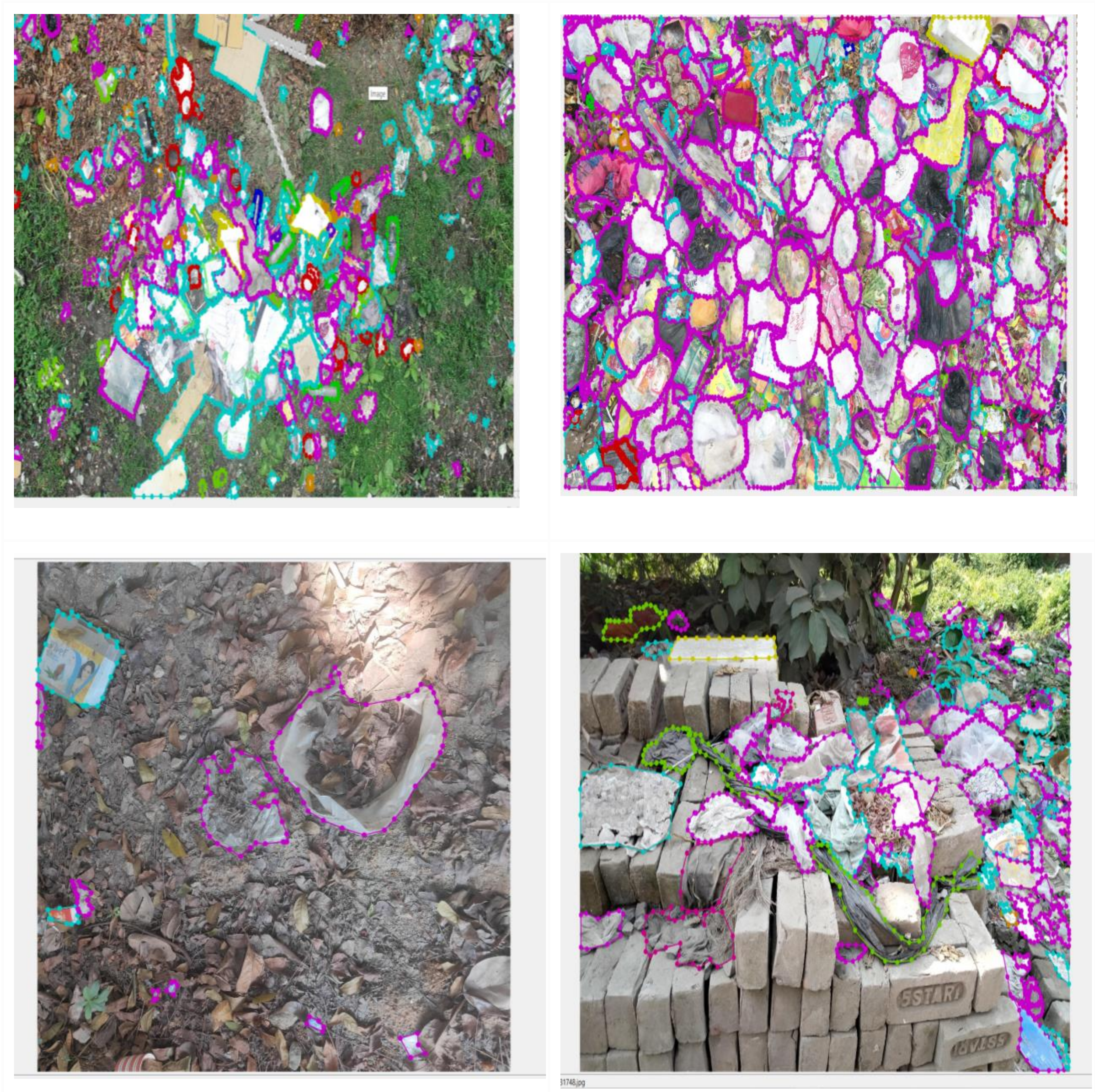


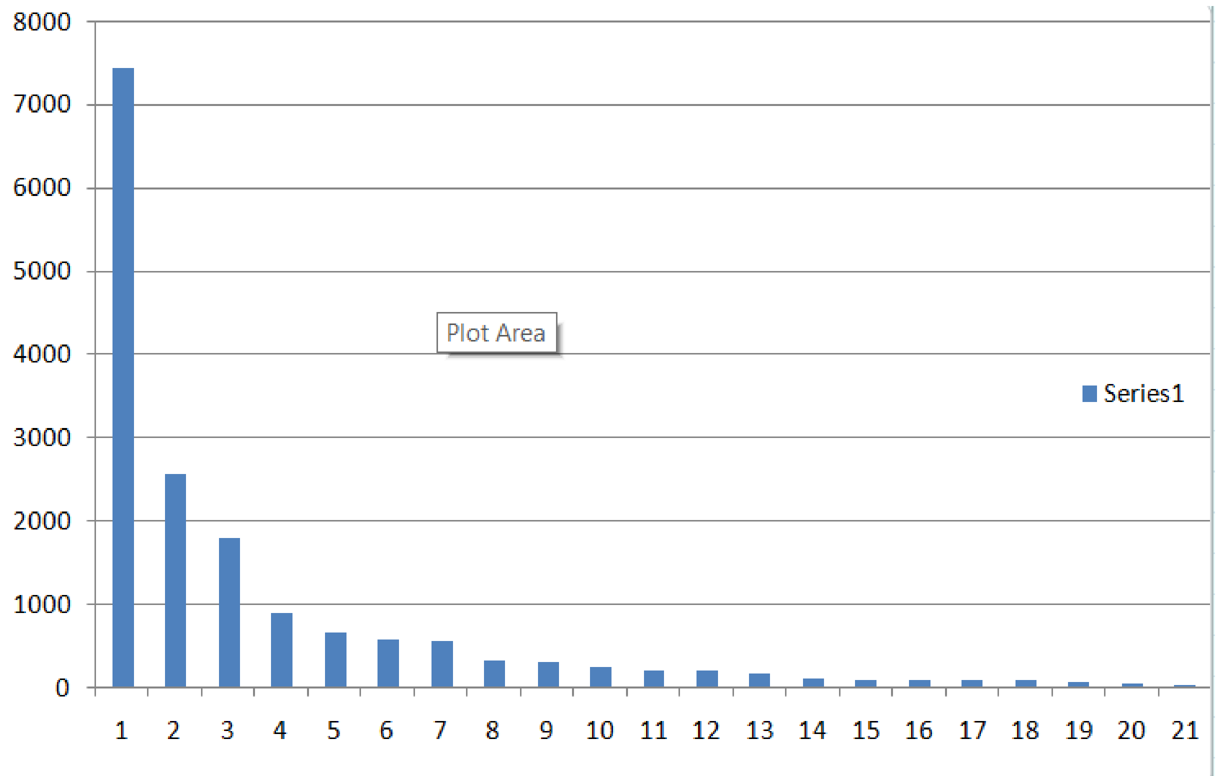
Table 4.4
Examples of annotated Object Segmentation dataset.

Currently, there are many types of deep learning algorithms for object detection. Some of them are used for waste classification and some pre-trained models of CNN like AlexNet, InceptionResNetV2, MobileNet, DenseNet, Xception, etc. are used as the backbone of the segmentation and classification models. Those models have an average range of accuracy of 22% to 98.2% [4,5,6]. The images are also segmented using some CNN like R-CNN, Faster R-CNN, Mask-RCNN, SSD and some YOLO models like YOLOV4, YOLOV5, YOLOX, YOLOR etc[7,8,9,10]. Using these methods, the accuracy value changes based on its varied architecture, resulting in an efficiency value of 15.9% for TACO with Mask R-CNN and 81% for Trash ICRA19 with Faster R-CNN.

4.3)Datasets Annotation for Object Localization: The dataset for object localization is labeled by a software called Labelimg. The waste object is labeled with a rectangular bounding box of a matching shape by the software and mentions its name as discussed above. This dataset type is also similar to the previous ones.

Table 4.5 Examples of annotated object localization datasets.

4.4)Dataset Health : We have collected a total of 6005 images to progress this thesis. Of those 6005 images, we had 16770 annotations across 21 classes, and the average number of annotations per image is 2.8. The average image size of those images is 5.78 mp and the median image ratio is 2044 X 2075. The dataset bar graphs are shown below.



(1) plastic carry (2) paper (3) paper_cup (4) plastic (5) plastic_container (6) paper_cardboard (7) cigarette_packet (8) plastic_bottle (9) newspaper (10) thermocol (11) Match_Box (12) metal (13) cloth (14) plastic_sack (15) medicine_wrapper (16) glass_bottle (17) glass (18) nylon (19) cotton (20) aluminium_foil (21) rubber

Fig 4.2 Dataset Bar Graph

- **Dimension Insights :** Every dot is a given image in our dataset's dimensions. If an image is a perfect square, it would extend 45 degrees from the origin of our graph (and fall right on the light gray line). If a dot is above this line, the image is taller than it is wide. The y-axis describes the height dimension of our images, and the x-axis the image width.

  The last orienting component in this visualization is the purple area, denoting the "median" image size in a dataset. In our shaded areas on either axis, we note images that may fall into the "danger" territory of being too tall or too wide—where the longer dimension is 2x the shorter dimension. The median is less susceptible to outliers than the mean, and provides a central point to consider how far any given image is from our desired resize (again, likely square).

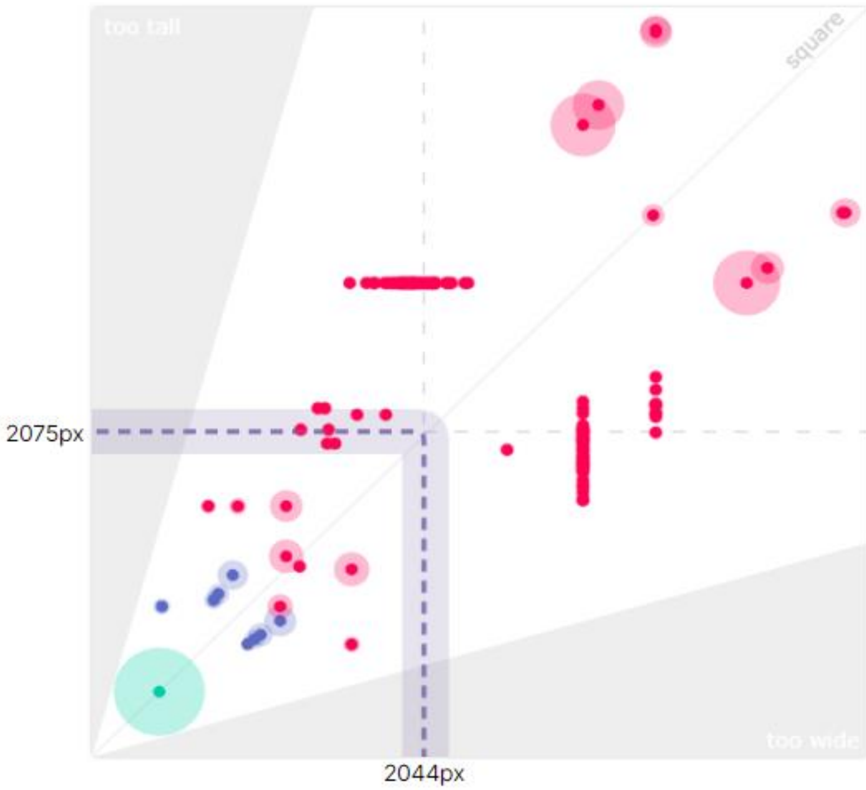


Fig 4.3
Dimension Insight of Dataset

- **<u>Annotation Heatmap</u> :** Heatmap annotations are important components of a heatmap that show additional information that is associated with rows or columns in the heatmap. The Complex Heatmap package provides very flexible support for setting annotations and defining new annotation graphics. The annotations can be put on the four sides of the map by top, bottom, left and right.

  The image command or imagesc command can be used to create heat maps. The difference between the two functions is that imagesc scales the colormap of an image to give the maximum range of colors, while imagesc gives the lowest possible resolution for a heat map.

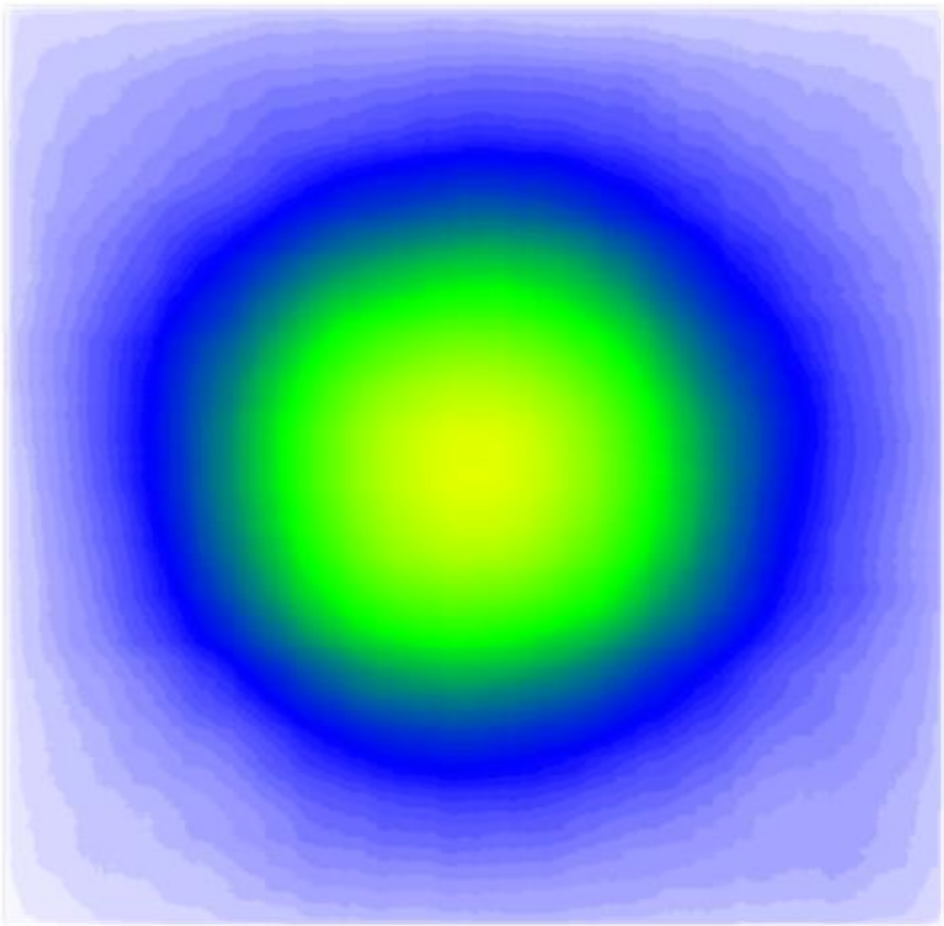
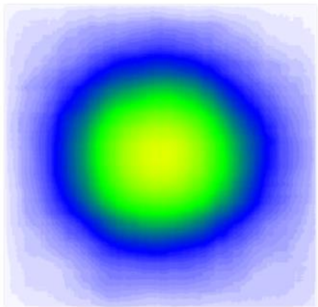
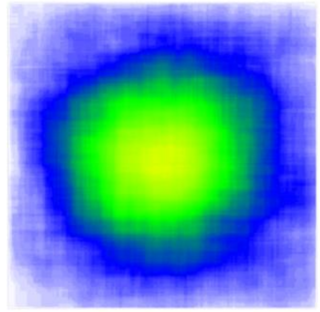


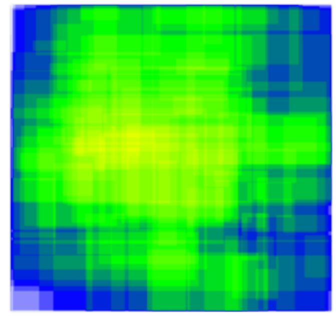
Fig 4.4
Annotation Heatmap of Dataset



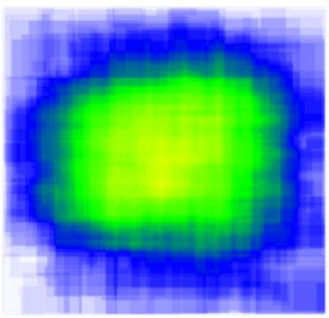
(a)



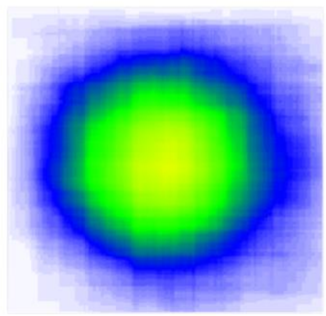
(b)



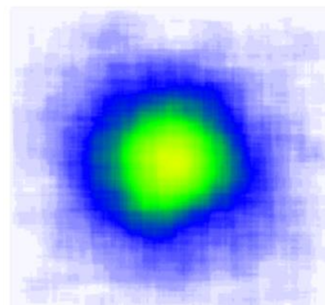
(c)



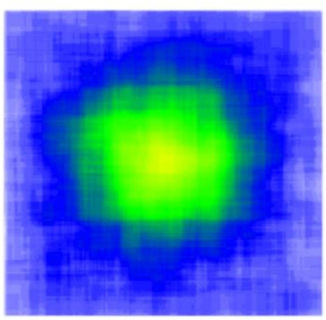
(d)



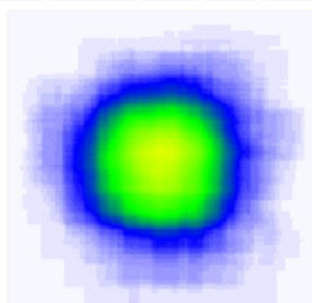
(e)



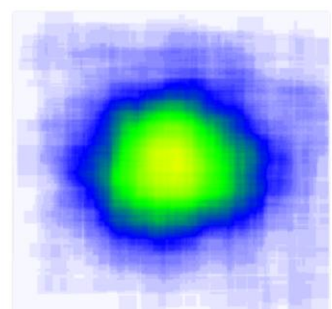
(f)



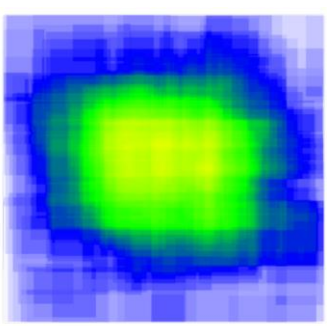
(g)



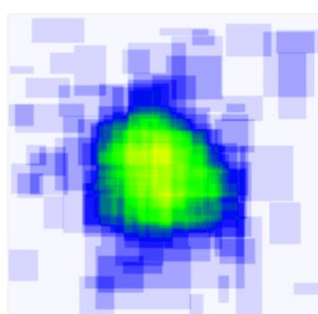
(h)



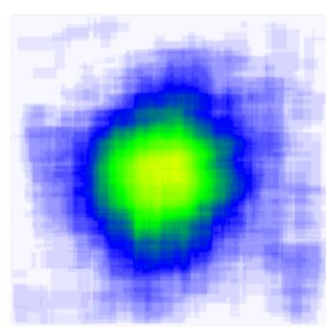
(i)



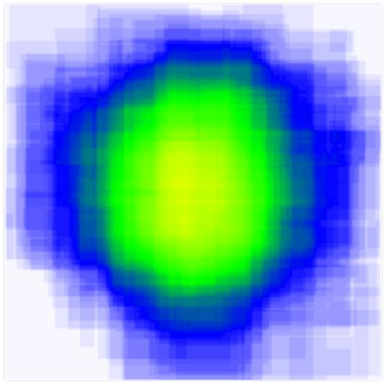
(j)



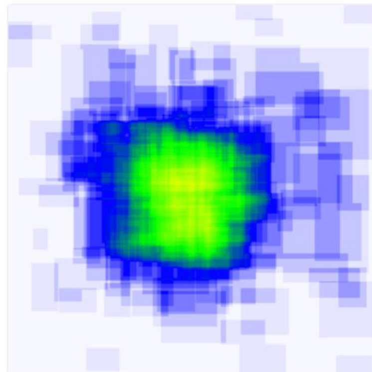
(k)



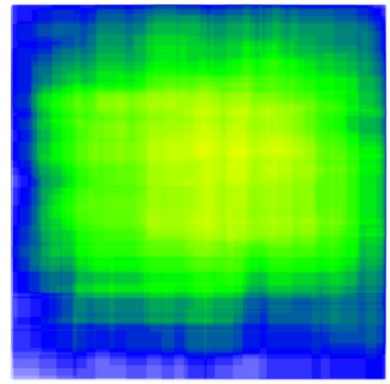
(l)



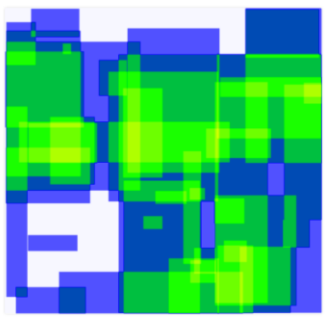
(m)



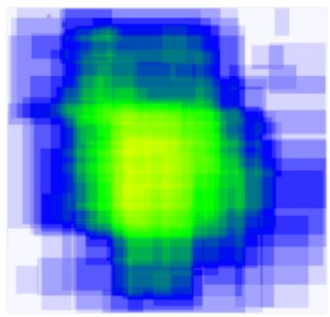
(n)



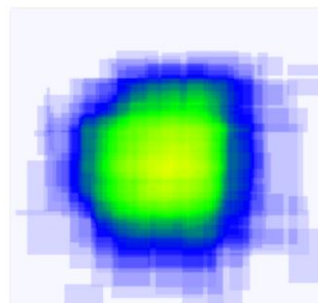
(o)



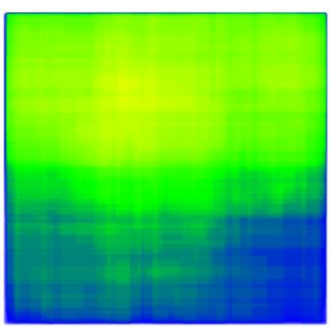
(p)



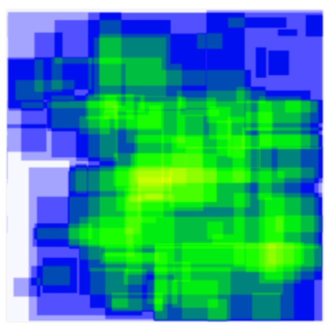
(q)



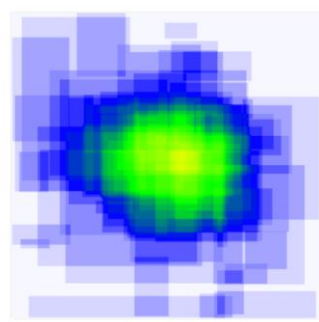
(r)



(s)



(t)



(u)

The annotation heatmap are : (a) plastic_carry (b) paper (c) paper_cup (d) plastic (e) plastic_container (f) paper_cardboard (g) cigarette_packet (h) plastic_bottle (i) newspaper (j) thermocol (k) Match_Box (l) metal (m) cloth (n) plastic_sack (o) medicine_wrapper (p) glass_bottle (q) glass (r) nylon (s) cotton (t) aluminium_foil (u) rubber

- **Histogram of Object Count by Image :** The object count histogram can be used to review data labeling quality. It is also a good way to get a better sense of the dataset. It can also help to understand why the models may be over/under performing in some categories.

The object count histogram lives on the Roboflow dataset health check page. We can access it by navigating to our dataset and clicking Dataset Health Check in the sidebar. For old datasets, we may need to regenerate our dataset at the bottom of the page.

Now I am showing the histogram of object count images.

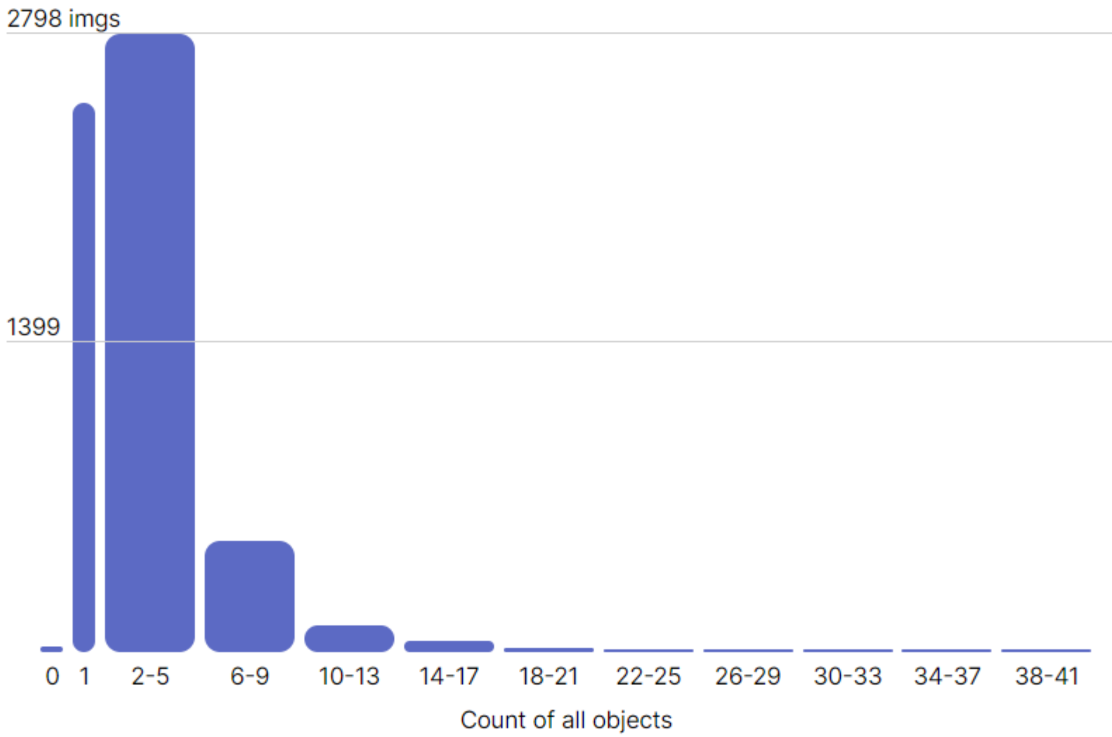
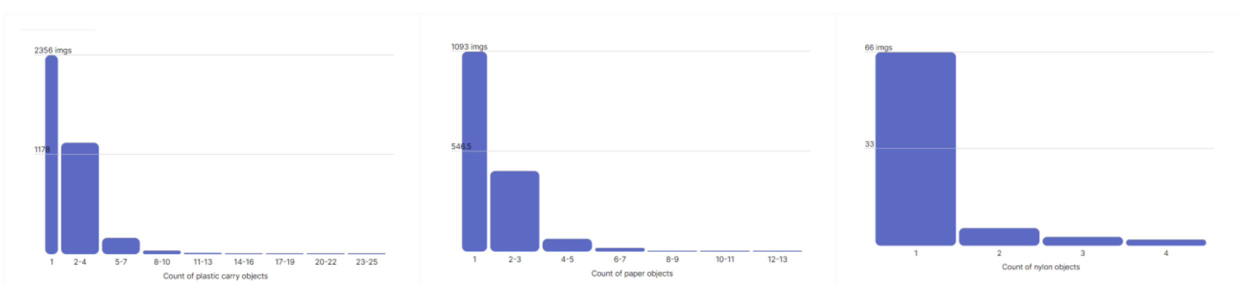
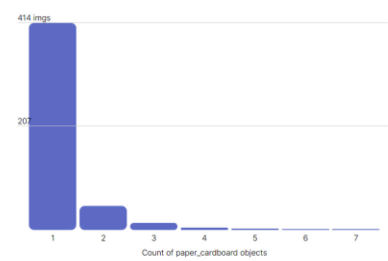
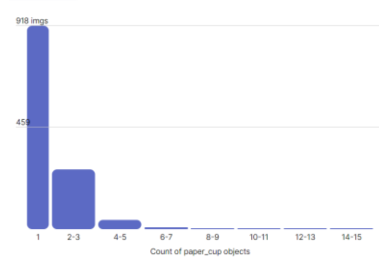


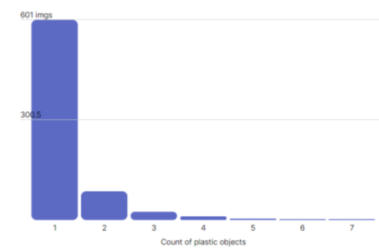
Fig 4.5
Annotation Heatmap of Dataset



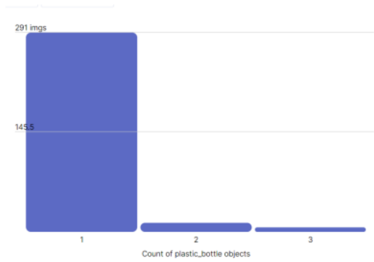
(a) (b) (c)

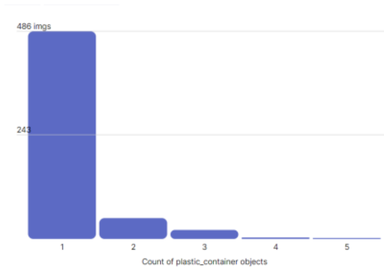


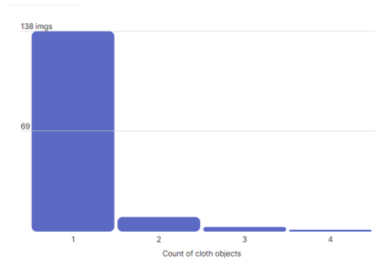




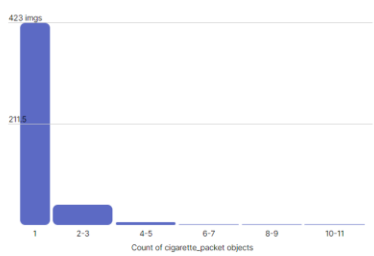
(d) (e) (f)

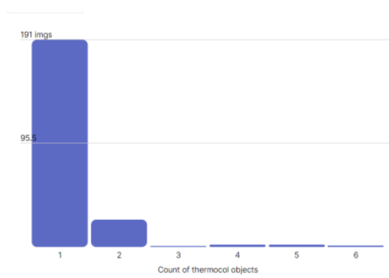


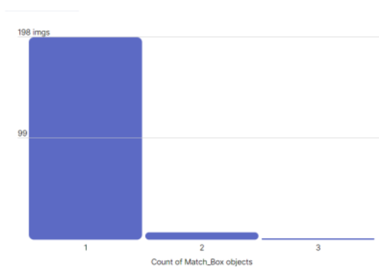




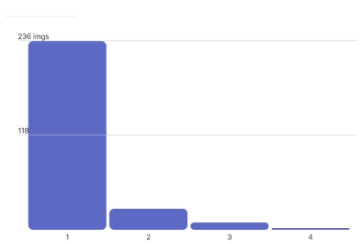
(g) (h) (i)

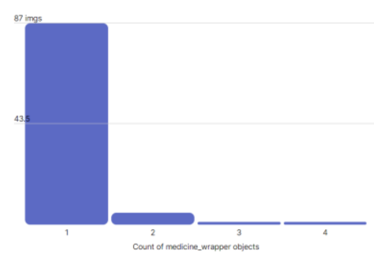


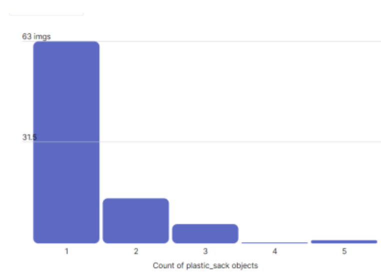




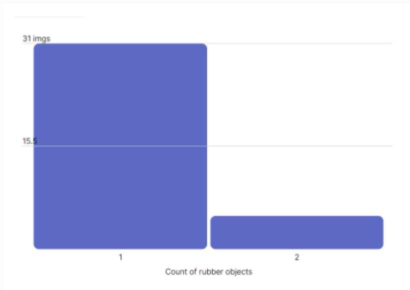
(j) (k) (l)



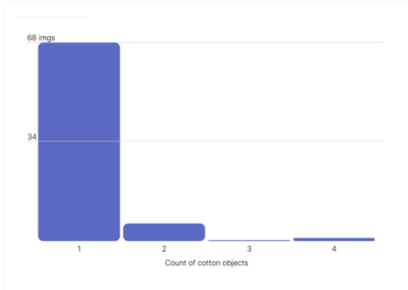




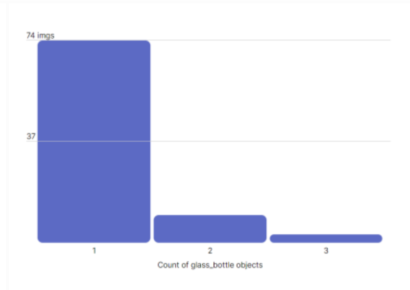
(m) (n) (o)



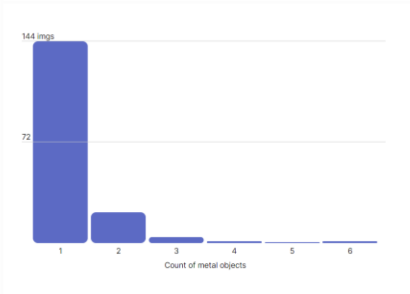
(p)



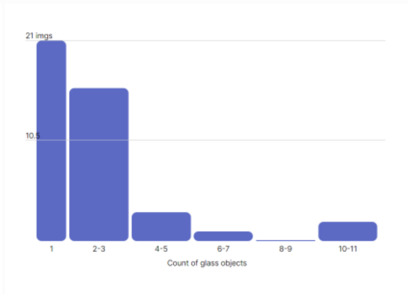
(q)



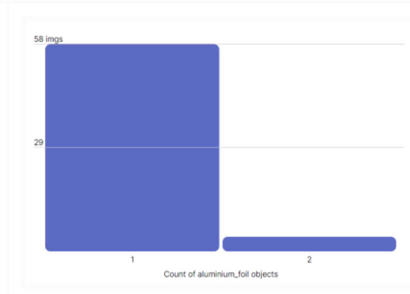
(r)



(s)



(t)



(u)

The histogram of object count for each classes are (a) plastic carry (b) paper (c) paper_cup (d) plastic (e) plastic_container (f) paper_cardboard (g) cigarette_packet (h) plastic_bottle (i) newspaper (j) thermocol (k) Match_Box (l) metal (m) cloth (n) plastic_sack (o) medicine_wrapper (p) glass_bottle (q) glass (r) nylon (s) cotton (t) aluminium_foil (u) rubber

4.5)Proposed Methods: In this thesis work, two different types of image classification have been done; one of them is object segmentation, and the other one is object localization.All the experiments have been done on a CoLab Notebook with 51 GB of RAM and a 32 GB Nvidia Tesla P100 GPU.

4.5.a)Object Localization Models: It is the primary task of computer vision to localize objects . In 2015, the YOLO techniques were born from a different approach. Object detection was taken care of as a regression problem, which was performed by a single neural network. I ran some of the models on annotated datasets that were collected .

In recent years , with the increasing use of computer applications , most machines are automated and the complex work is done by machines using artificial intelligence, machine learning, and deep learning algorithms . Those models are tested with huge amounts of data. To classify the objects ,models are trained and extract the valuable features. Computer vision has long dealt with the classification of images based on their object characteristics, a complex task.

Now I am discussing the training metric for all the models.

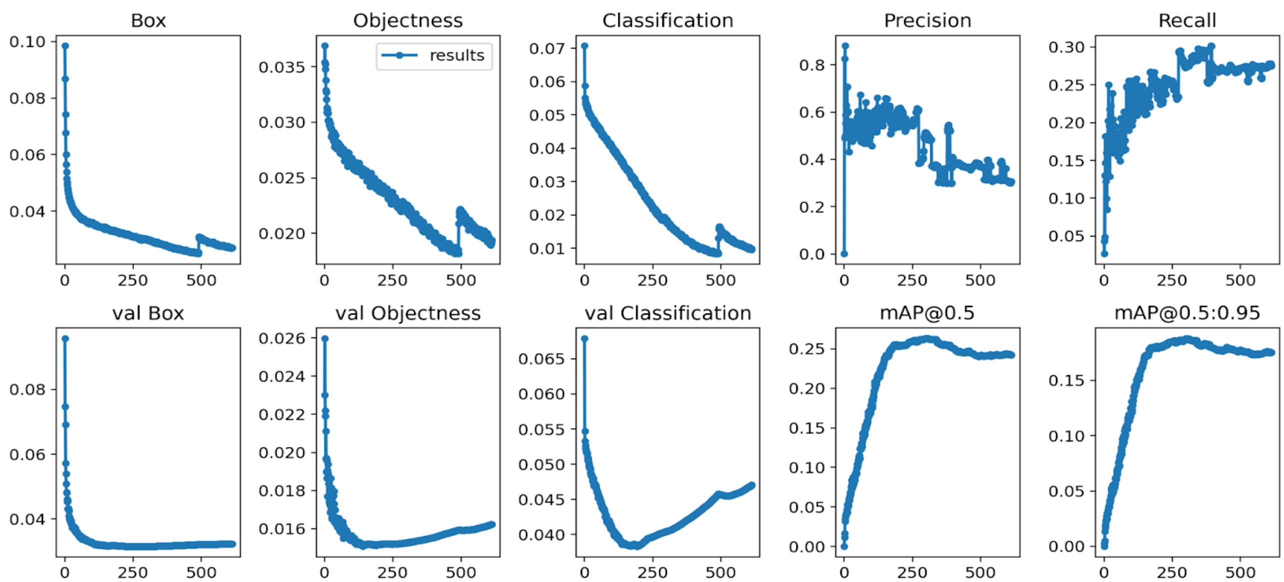


Table 4.6

Different training metrics of the YOLOv5 model

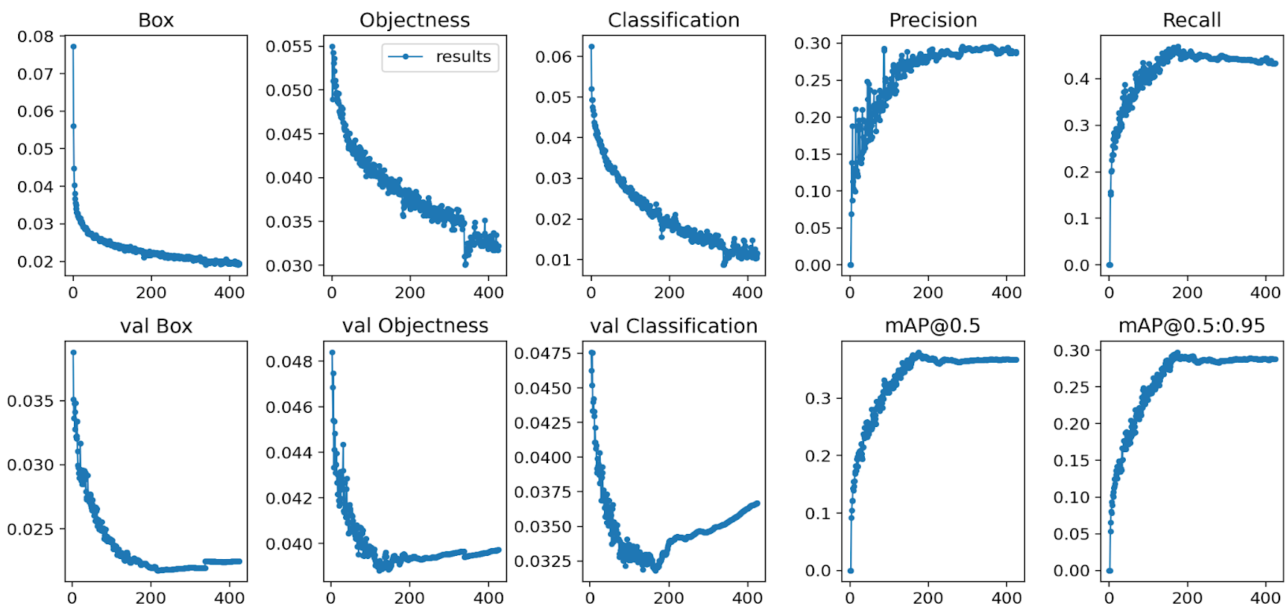


Table 4.7
Different training metrics of the YOLOR model

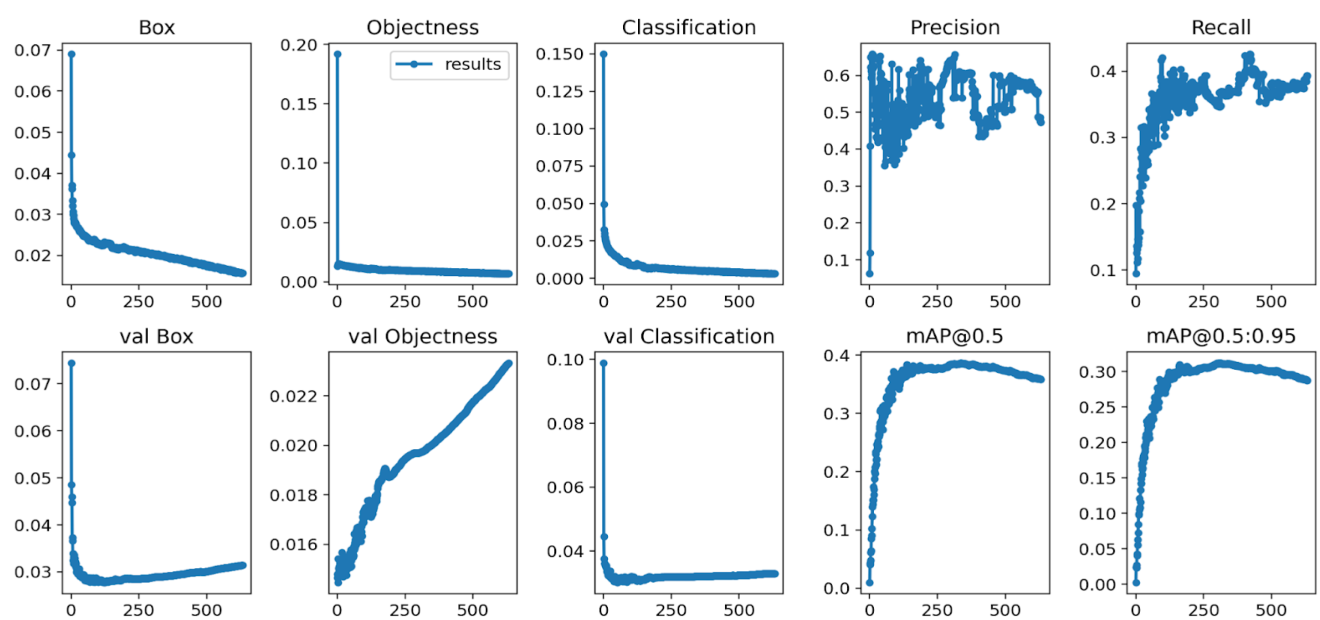


Table 4.8
Different training metrics of the YOLOv7 model

- **Box:** A loss function is used to forecast the bounding box and calculate the difference between the predicted and actual bounding box. Box loss can be found by using Smooth L1 and Intersection over the union. The model is trying to decrease the box loss for better detection. In YOLOv5 , YOLOR, and YOLOv7,the training box loss is decreasing over the epochs. However, validation box loss is increasing in YOLOv7 after 50 epochs, which means the model is overfeeding the data.

- **Objectness:** The probability that an object exists in a suggested zone of interest is basically measured by objectness. If our objectness score is high, an object is probably present in the image window. This enables us to swiftly remove suggested image windows that are empty of things. Objectness, which allows for the pruning of proposed zones of interest that don't contain any objects, generally calculates the likelihood that an object will appear in an image. The model is trying to decrease the objectivity for better detection. In YOLOv5 , YOLOR, and YOLOv7,the training objectness is decreasing over the epochs. However, validation objectness is rapidly increasing in YOLOv7 .

- **Classification Loss:** The most typical loss function in classification issues is this one. As the predicted probability approaches the true label, the cross-entropy loss diminishes. It evaluates how well a classification model performs when it predicts an outcome that has a probability between 0 and 1. The model is trying to decrease the classification loss for better detection. In YOLOv5 , YOLOR, and YOLOv7,the training classification is decreasing over the epochs. However, classification loss is slightly increasing in YOLOv7 after 10 epochs, which means the model is overfeeding the data.

- **mAP@ 0.5:** IoU > 0.5 often indicates a success; otherwise, it indicates a failure. One can figure it out for each class. True Positive TP(c): Class c was proposed, and a class c object truly existed. False Positive FP(c): A class c proposal was submitted, but no class c objects exist.tio

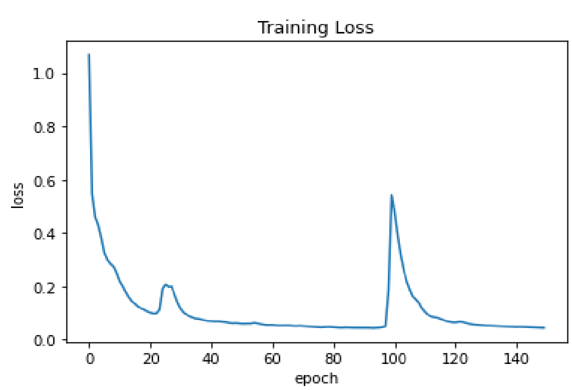
$$mAP = \frac{1}{|classes|} \sum_{c \in classes} \frac{\#TP(c)}{\#TP(c) + \#FP(c)}$$

❖ **Intersection over Union (IoU):**

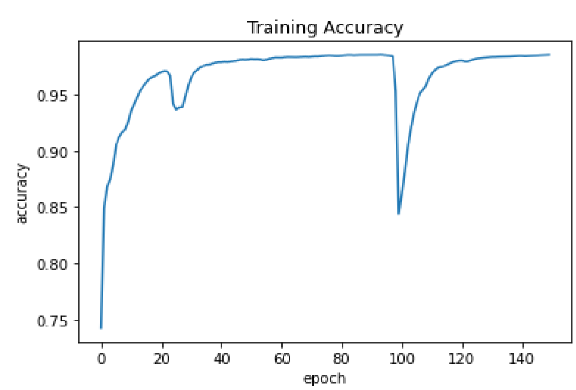
$$IOU = \frac{Area\ of\ Intersection\ of\ two\ boxes}{Area\ of\ Union\ of\ two\ boxes}$$

❖ **mAP 0.5:0.95:0.05 :** Average precision where IoU is 0.5 to 0.95 in common difference 0.05 often indicates a success; otherwise, it indicates a failure. One can figure it out for each class. True Positive TP(c): Class c was proposed, and a class c object truly existed. False Positive FP(c): A class c proposal was submitted, but no class c objects exist.

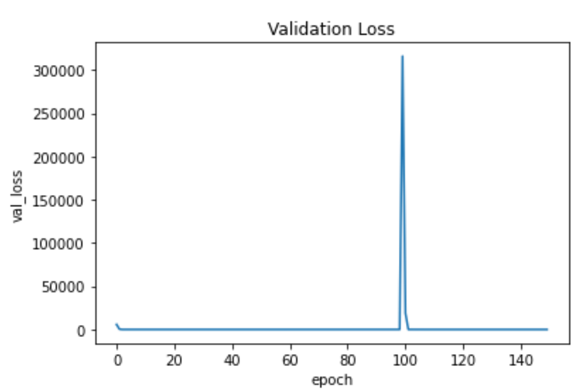
4.5.b)Object Segmentation Models: One such technique employed in these intelligence systems is object segmentation, and every day new papers and algorithms are being created. The image segmentation problem can be better understood with the aid of this article. In my thesis work, I trained a few image segmentation models. Those are DeepLabv3+,UNet, LinkNet, and PSPNet.



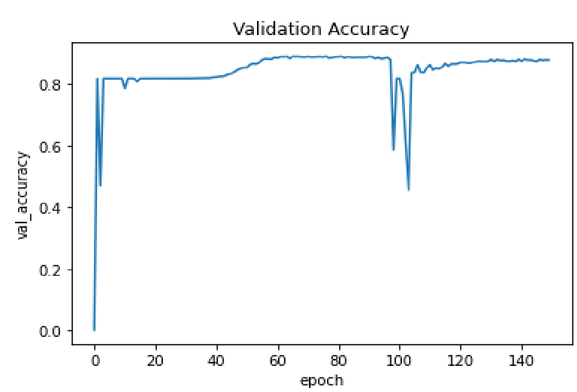
(a)



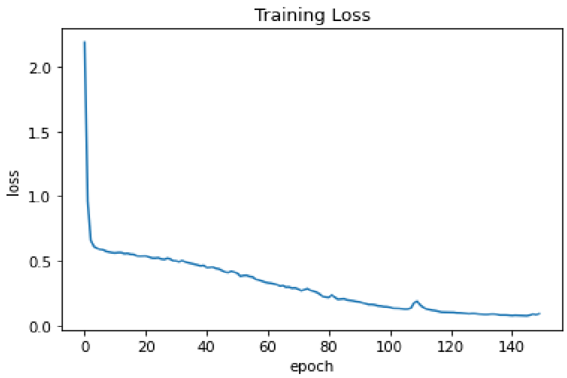
(b)



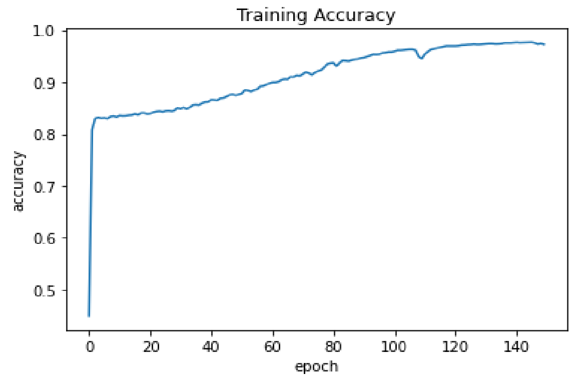
(c)



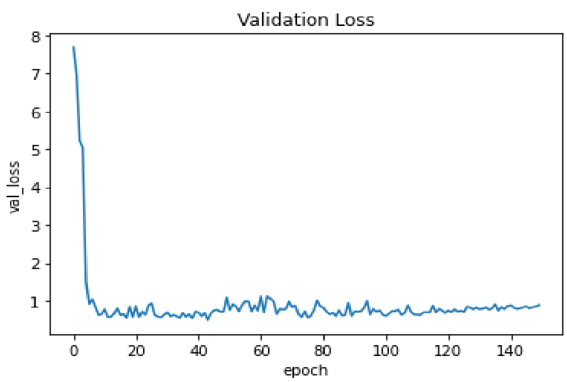
(d)


Training Loss
loss
epoch

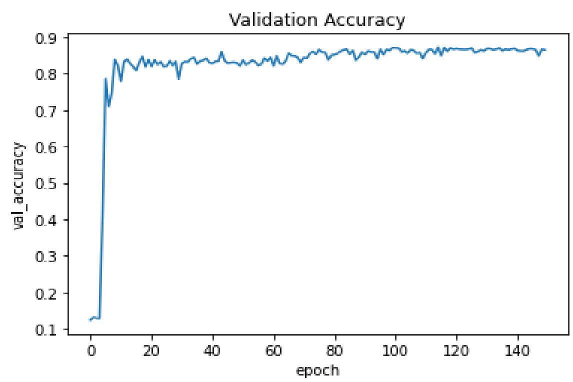
(e)


Training Accuracy
accuracy
epoch

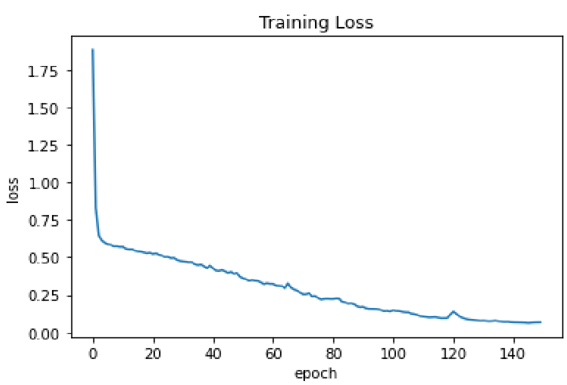
(f)


Validation Loss
val_loss
epoch

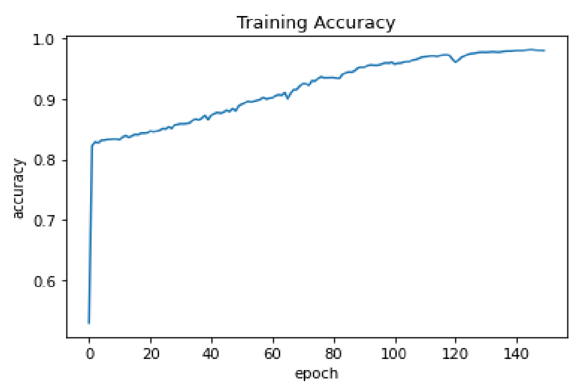
(g)


Validation Accuracy
val_accuracy
epoch

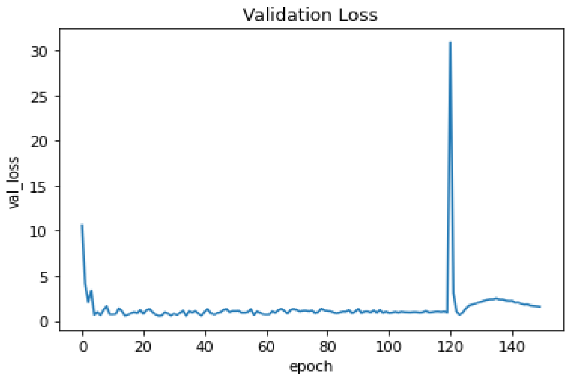
(h)

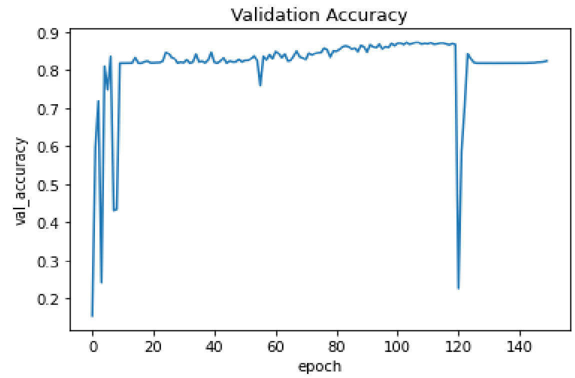

Training Loss
loss
epoch

(i)


Training Accuracy
accuracy
epoch

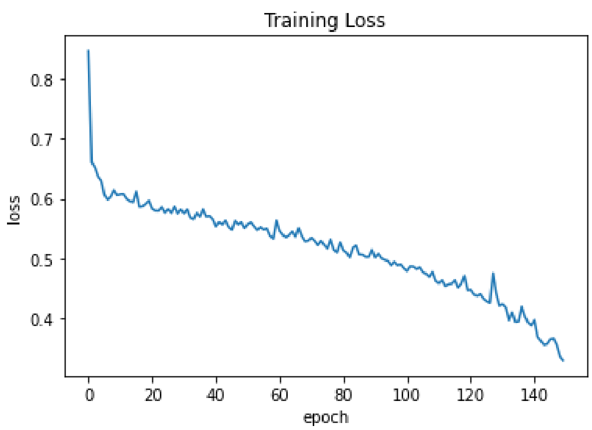
(j)

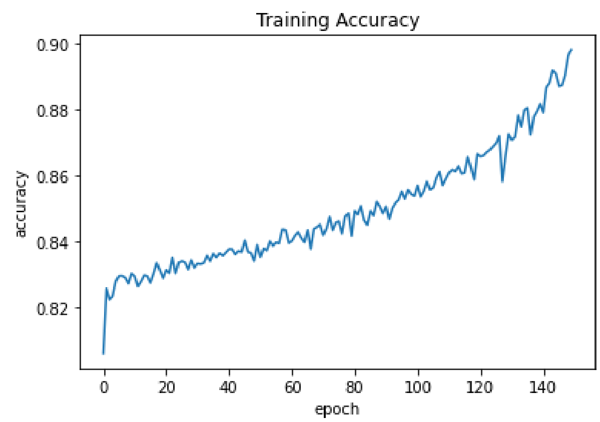

Validation Loss
val_loss
epoch


Validation Accuracy
val_accuracy
epoch

(k)

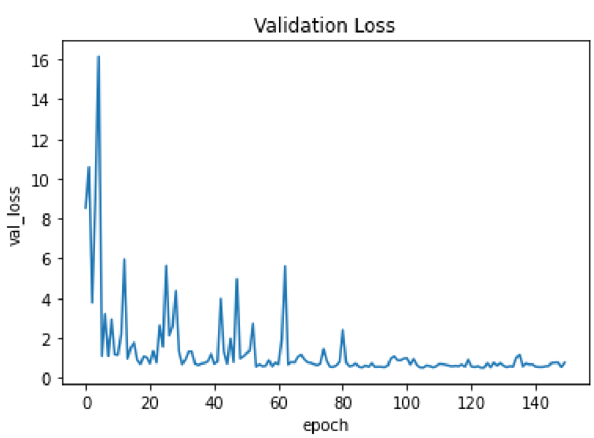
(bl)

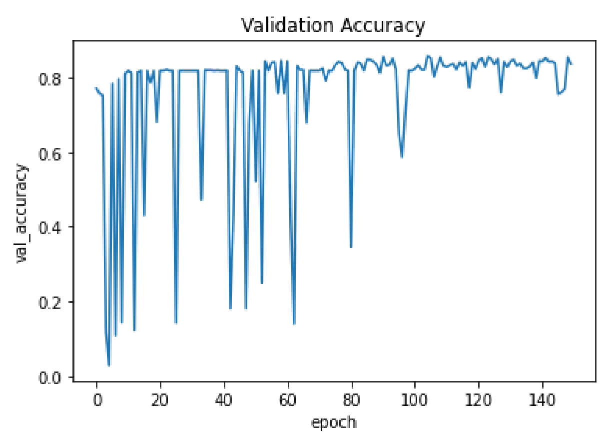

Training Loss
loss
epoch


Training Accuracy
accuracy
epoch

(m)

(n)


Validation Loss
val_loss
epoch


Validation Accuracy
val_accuracy
epoch

(o)

(p)

The graphs are: (a) Training Loss of DeepLabv3+ (b) Training Accuracy of DeepLabv3+ (c) Validation Loss of DeepLabv3+ (d) Validation Accuracy of DeepLabv3+ (e) Training Loss of UNet (f) Training Accuracy of UNet (g) Validation Loss of UNet (h) Validation Accuracy of UNet (i) Training Loss of LinkNet (j) Training Accuracy of LinkNet (k) Validation Loss of LinkNet (l) Validation Accuracy of LinkNet (m) Training Loss of PSPNet (n) Training Accuracy of PSPNet (o) Validation Loss of PSPNet (p) Validation Accuracy of PSPNet

- **Training Loss:** An indicator of how well a deep learning model fits the training data is the training loss. In other words, it evaluates the model's error on the training set. Keep in mind that the training set is a subset of the dataset that was initially used to train the model. The total of errors for each sample in the training set is used to computationally determine the training loss.

  It's also crucial to remember that each batch ends with a measurement of the training loss. Typically, a training loss curve is plotted to show this. The object detection model tries to decrease the training loss over epochs. In my dataset, training loss is decreasing gradually with increasing epochs for DeepLabv3+,UNet, LinkNet, and PSPNet. But for DeepLabv3+ and PSPNet model loss, loss is increased like impulse nature at 120 and 95 epochs, respectively.

- **Validation Loss:** On the other hand, a deep learning model's performance on the validation set is evaluated using a statistic called validation loss. The dataset's validation set is a section set aside to check the model's efficacy. Similar to the training loss, the validation loss is determined by adding the errors for each sample in the validation set.The validation loss is also measured after every epoch. This helps us determine whether the model needs to be adjusted or tuned further. We typically plot a learning curve for the validation loss to do this. In my dataset, the validation loss is approximately equal to the training loss .

- **Training Accuracy:** An indicator of the model's performance across all classes is accuracy. When all classes are equally important, it is helpful. The number of accurate forecasts divided by the total number of predictions is used to compute it. Test accuracy means that the trained model correctly detects independent images that were not used in training, whereas training accuracy means that the same images are used for both training and testing. Generally, accuracy is increasing with increasing epochs, but for DeepLabv3+ and PSPNet models, accuracy is decreasing then loss increases.

- **Validation Accuracy:**The accuracy calculated on the data set that was not used for training but was instead used (during the training phase) to validate the generalization ability of your model or for "early stopping," in other words, is referred to as the test (or testing) accuracy.

CHAPTER 5

EXPERIMENT RESULT

In my thesis work I had done two different types of object detection work one is localization and other one is segmentation. For benchmarking of localization datasets, I have trained five different state-of-the-art models, i.e. YOLOv3, YOLOv5, YOLOR, MT-YOLOv6 and YOLOv7, and to compare the proposed segmentation model on the developed dataset, I used DeepLabv3+ and the UNet model. Through those models, I have got two different types of results, which are qualitative results and quantitative results.

5.1) Qualitative Result: My annotated dataset ran through the models and I got those types of results which are shown below:

Annotated Sample

Predicted Sample

Table 5.1

Annotated and Predicted Data Sample of YOLOv3 Model

Annotated Sample

Predicted Sample

Table 5.2
Annotated and Predicted Data Sample of YOLOv5 Model

Annotated Sample

Predicted Sample

Table 5.3
Annotated and Predicted Data Sample of YOLOR Model

Annotated Sample

Predicted Sample

Table 5.4
Annotated and Predicted Data Sample of MT-YOLOv6 Mod

Annotated Sample

Predicted Sample

Table 5.5
Annotated and Predicted Data Sample of YOLOv7 Model

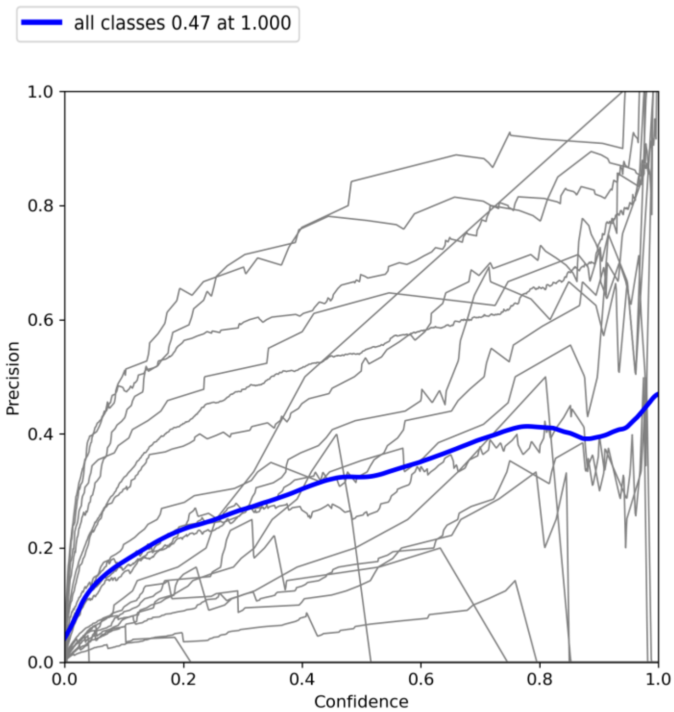
5.2) Quantitative Result:

PRESSION TABLE:

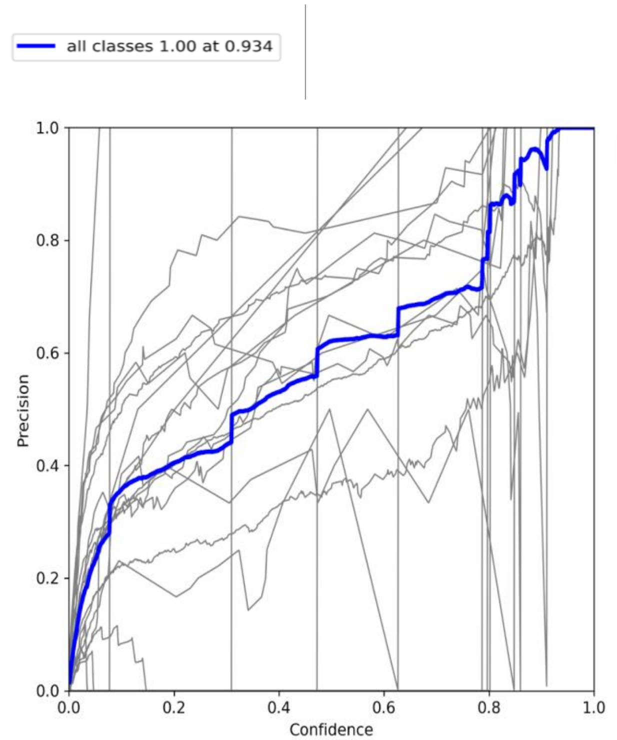
	YOLOv3	YOLOv5	YOLOR	YOLOv7
Plastic carry	0.324	0.321	0.277	0.493
Paper	0.163	0.237	0.242	0.38
Paper cup	0.476	0.52	0.434	0.584
Plastic	0.157	0.395	0.418	0.546
Plastic container	0.303	0.361	0.347	0.44
Paper cardboard	0.174	0.383	0.345	0.552
Cigarette packet	0.427	0.385	0.345	0.654
Plastic bottle	0.508	0.716	0.581	0.804
Newspaper	0.0968	0.326	0.0552	0.24
Thermocol	0.078	0.389	0.235	0.367
Match Box	0.387	0.557	0.376	0.499
Metal	0.0622	0.326	0.14	0.308
Cloth	0.0755	0	0.273	0.167
Plastic sack	0.0622	0	0	0.0866
Medicine wrapper	0.113	0.417	0.0958	0.35
Glass bottle	0.207	0.571	0.356	0.62
Glass	0.0312	0	0	0
Nylon	0	0	0.0552	0
Cotton	0.0885	1	0.323	0.617
Aluminium_foil	0	0.365	0.143	0.365
Rubber	0	1	0.138	0

Table 5.6 Precision Table for each classes of YOLOv3,YOLOv5,YOLOR and YOLOv7

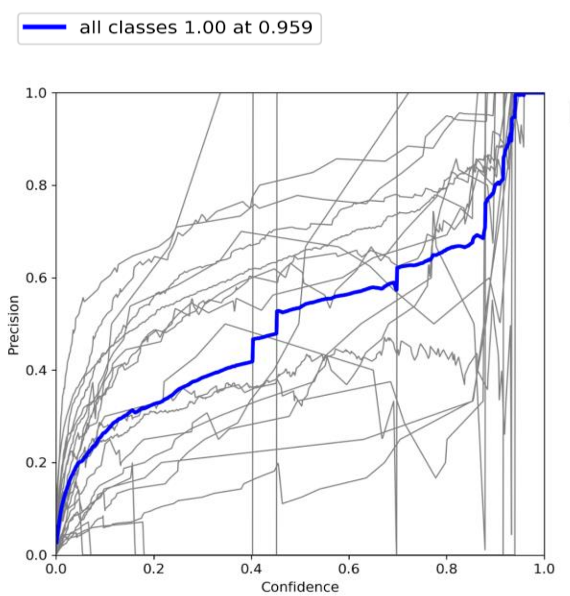
PRECISION VS. CONFIDENCE:



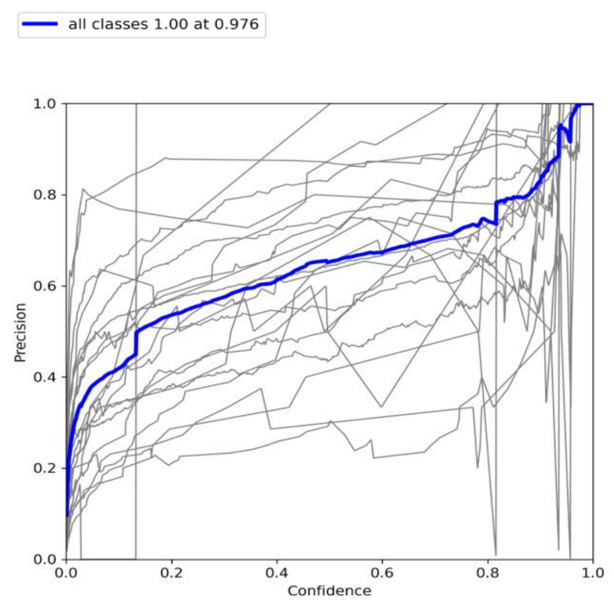
YOLOv3



YOLOv5



YOLOR



YOLOv7

Fig 5.1 Precision vs. Confidence graph for YOLOv3,YOLOv5,YOLOR and YOLOv7 models

PRECISION: The precision shows how accurate the model is in identifying positive samples. The accuracy is calculated as the ratio of positive samples that were correctly classified to all samples that were classified as positive (either correctly or incorrectly). The precision gauges how well the model categorizes a sample as positive.

$$Precision = True\ positive/(True\ positive\ +\ False\ positive)$$

The denominator rises and the precision becomes low when the model makes many wrong positive classifications or few correct positive classifications. However, the precision is high when:

1) Numerous accurate positive classifications are made by the model (maximize True Positive).

2) Less inaccurate positive classifications and thus minimizes false positives).

RECALL TABLE:

	YOLOv3	YOLOv5	YOLOR	YOLOv7
Plastic carry	0.66	0.207	0.3	0.2
Paper	0.316	0.342	0.461	0.168
Paper cup	0.666	0.667	0.75	0.704
Plastic	0.25	0.72	0.733	0.629
Plastic container	0.558	0.538	0.769	0.673
Paper cardboard	0.291	0.273	0.345	0.2
Cigarette packet	0.587	0.543	0.708	0.658
Plastic bottle	0.69	0.586	0.793	0.655
Newspaper	0.222	0.111	0.111	0.167
Thermocol	0.625	0.312	0.375	0.312
Match Box	0.733	0.733	0.8	0.667
Metal	0.333	0.486	0.5	0.5

Cloth	0.143	0	0.214	0.0596
Plastic sack	0.5	0	0	0.25
Medicine wrapper	0.0833	0.0833	0.168	0.167
Glass bottle	0.778	0.778	0.778	0.667
Glass	0.188	0	0	0
Nylon	0	0	0.0769	0
Cotton	0.3	0.137	0.2	0.1
Aluminium_foil	0	0.5	0.5	0.5
Rubber	0	0	1	1

Table 5.7 Recall Table for each classes of YOLOv3,YOLOv5,YOLOR and YOLOv7

RECALL VS. CONFIDENCE:

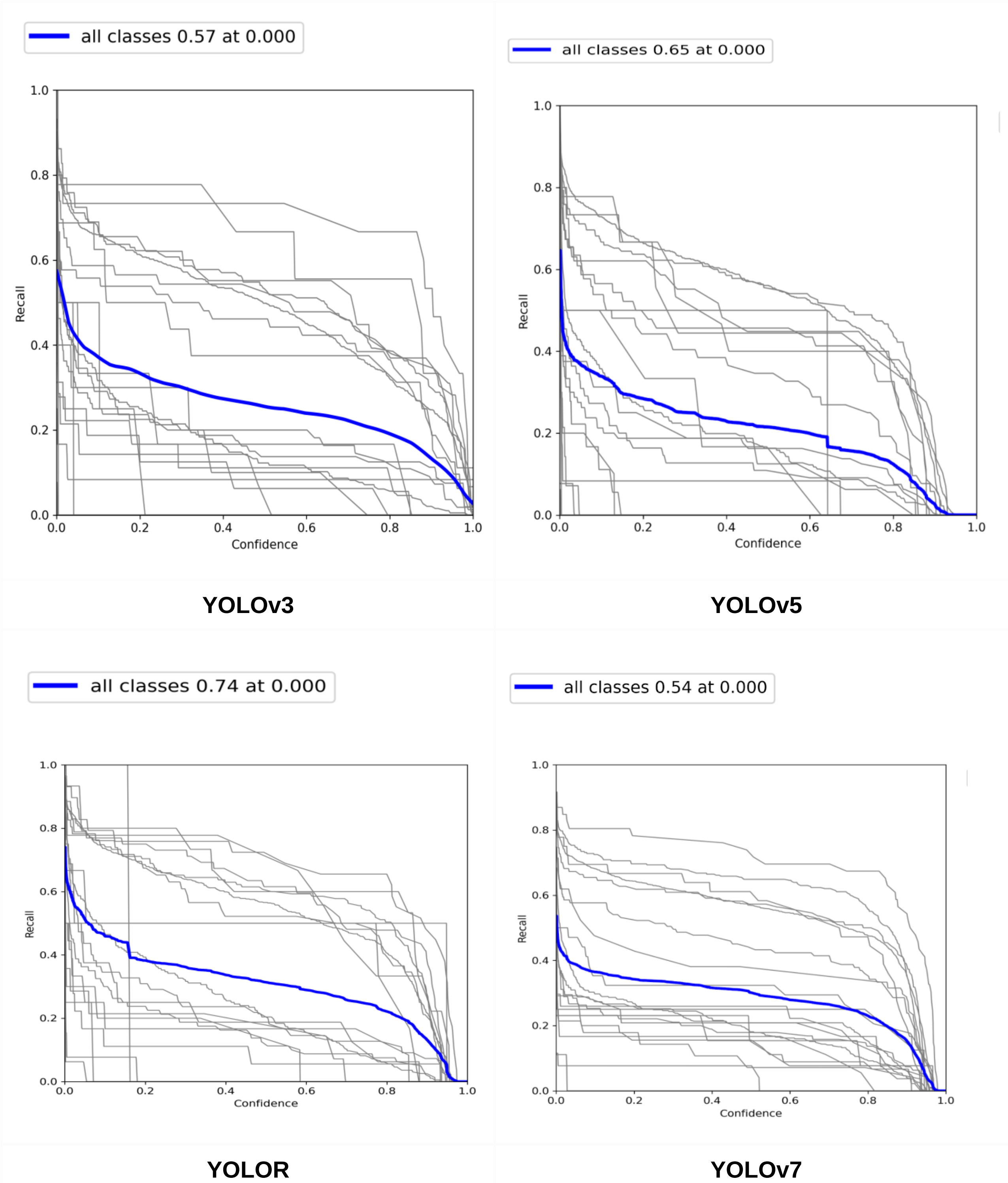


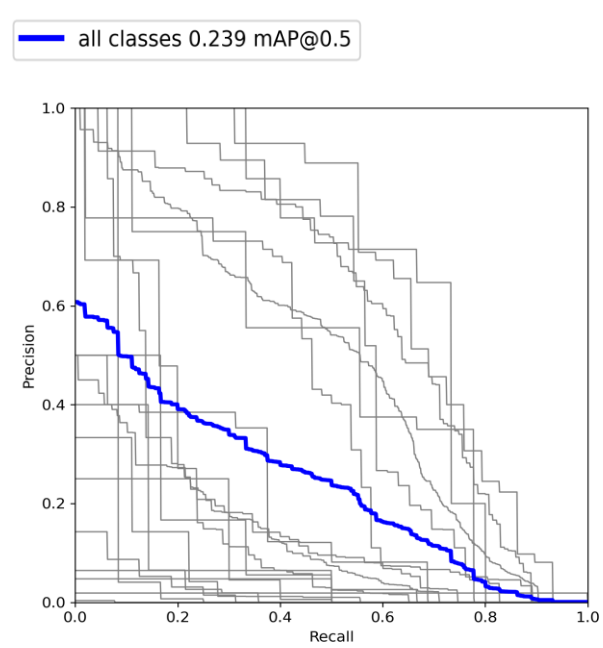
Fig 5.2 Recall vs. Confidence graph for YOLOv3,YOLOv5,YOLOR and YOLOv7 models

RECALL: The recall is determined as the proportion of positive samples that were correctly identified as positive to all positive samples. The recall gauges how well the model can identify positive samples. The more positive samples that are identified, the larger the recall.

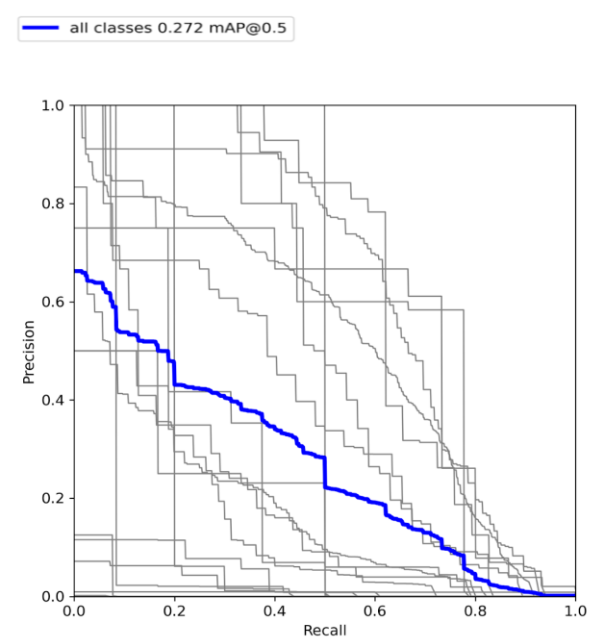
$$Recall = True\ positive/(True\ positive + False\ negative)$$

Only the classification of the positive samples is important to the recall. This is unrelated to the manner in which the negative samples are categorized, such as for precision. Even if all of the negative samples were mistakenly labeled as positive, the recall would be 100% when the model labels all of the positive samples as positive. Let's examine a few instances.

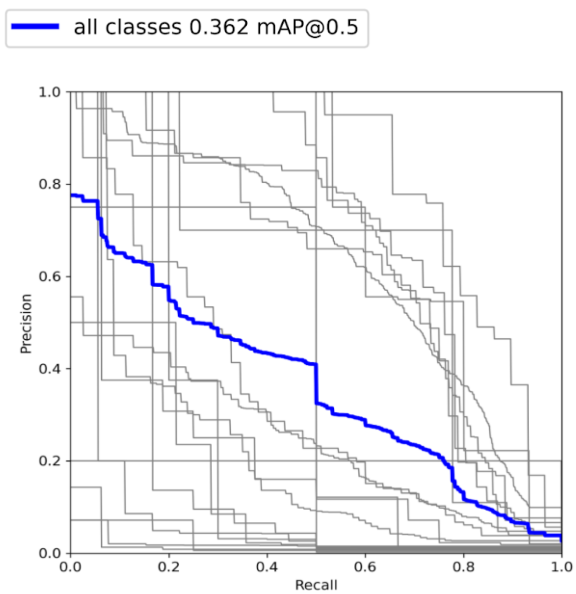
PRECISION VS. RECALL



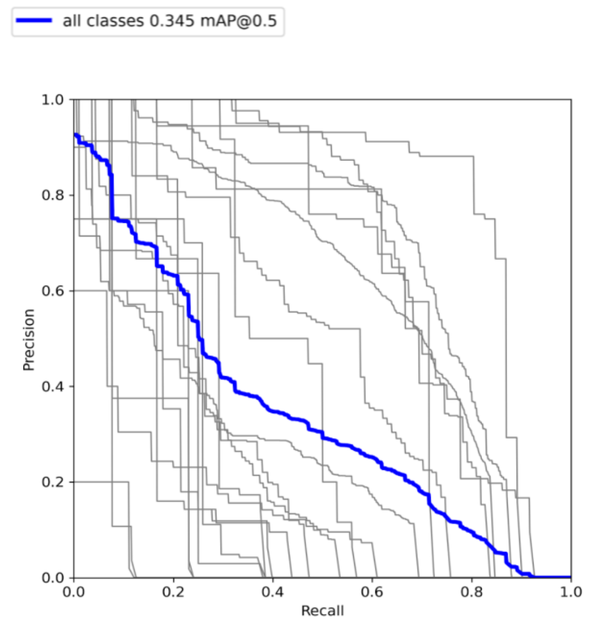
YOLOv3



YOLOv5



YOLOR



YOLOv7

Fig 5.3 Precision vs. Recall graph for YOLOv3,YOLOv5,YOLOR and YOLOv7 models

On a Precision vs. Recall (PR) graph, plot the precision and recall values. Precision and recall are always in a trade-off because of the monotonically declining PR graph. Each other's increases will be affected by each other's. Due to specific exceptions and/or a lack of data, the PR graph does not always show a monotonic decline.

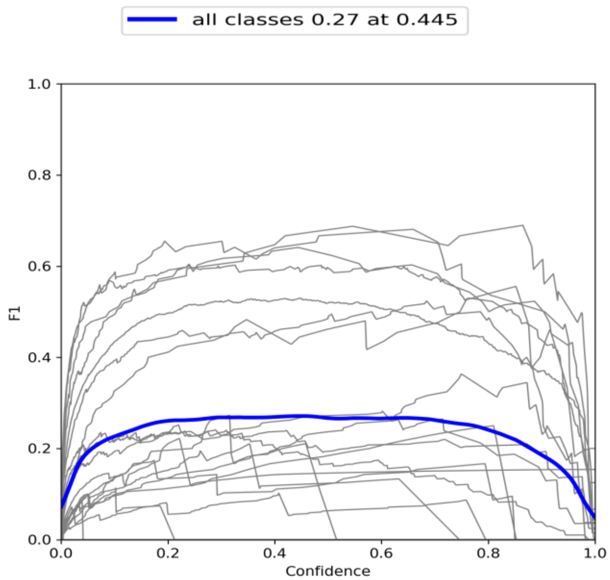
The detection of objects has a threshold. Raising the threshold lowers the possibility of over-detecting items while raising the possibility of missing detections. For instance, if the threshold is set to 1, no item will be found, the precision is set to 1, and the recall is set to 0. In contrast, an infinite number of objects will be discovered, precision will be 0.0, and recall will be 1.0 if the threshold is set to 0.0. In contrast, an infinite number of objects will be discovered, precision will be 0.0, and recall will be 1.0 if the threshold is set to 0.

F1 Score TABLE:

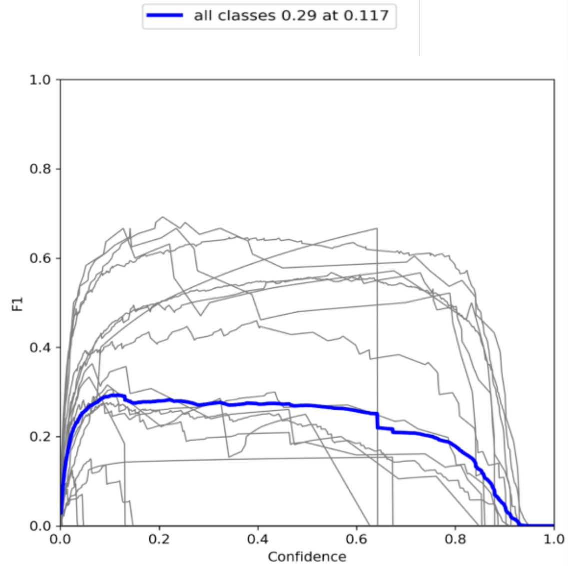
	YOLOv3	YOLOv5	YOLOR	YOLOv7
Plastic carry	0.435	0.252	0.288	0.285
Paper	0.215	0.281	0.318	0.233
Paper cup	0.555	0.585	0.550	0.639
Plastic	0.193	0.510	0.532	0.584
Plastic container	0.393	0.432	0.478	0.532
Paper cardboard	0.218	0.319	0.345	0.293
Cigarette packet	0.494	0.450	0.464	0.656
Plastic bottle	0.585	0.645	0.671	0.072
Newspaper	0.135	0.166	0.074	0.197
Thermocol	0.139	0.346	0.289	0.338
Match Box	0.506	0.633	0.512	0.571
Metal	0.105	0.390	0.219	0.381
Cloth	0.0988	0	0.241	0.087
Plastic sack	0.111	0	0	0.128
Medicine wrapper	0.096	0.139	0.122	0.226
Glass bottle	0.327	0.659		0.642
Glass	0.0535	0	0	0
Nylon	0	0	0.0643	0
Cotton	0.137	0.240	0.248	0.172
Aluminium_foil	0	0.422	0.223	0.422
Rubber	0	0	0.243	0.999

Table 5.8 F1 Score Table for each classes of YOLOv3,YOLOv5,YOLOR and YOLOv7

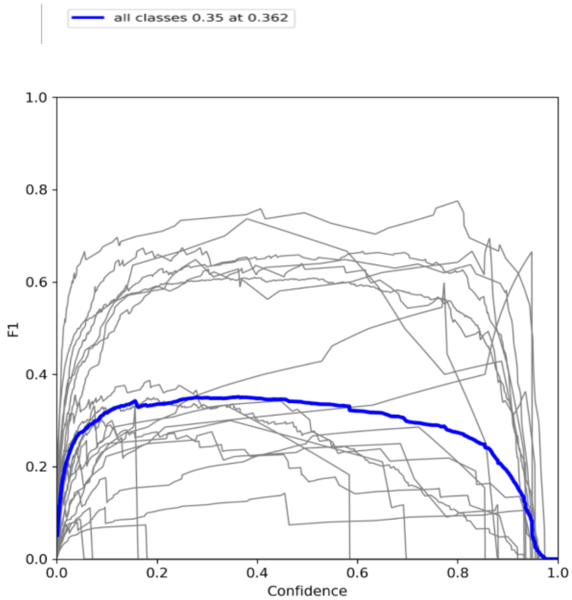
F1 Score VS. CONFIDENCE:



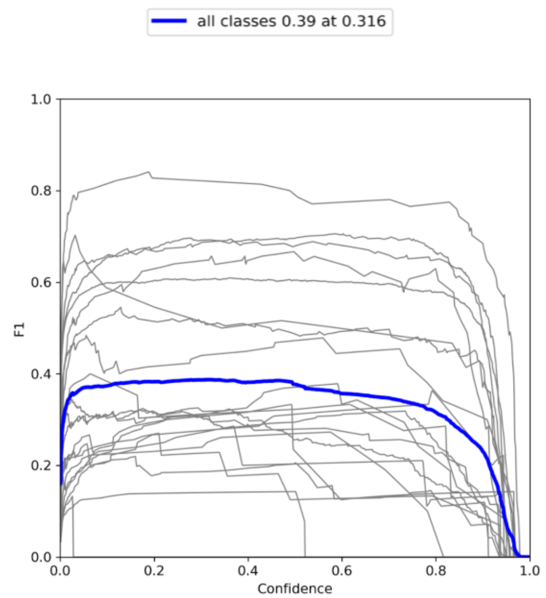
YOLOv3



YOLOv5



YOLOR



YOLOv7

Fig 5.4 F1 Score vs. Confidence graph for YOLOv3,YOLOv5,YOLOR and YOLOv7 models

The F-measure, often known as the F-score, is a measurement of a test's accuracy used in statistical analyses of binary categorization. It is derived from the test's precision and recall, where precision is the proportion of "true positive" results to "all positive results," including those incorrectly identified as positive, and recall is the proportion of "true positive" results to "all samples that should have been identified as positive."

In diagnostic binary classification, recall is also referred to as sensitivity, while precision is also referred to as positive predictive value.
The harmonic mean of precision and recall is the F1 score. The more general F(beta) score applies extra weights and gives preference to either precision or recall.

An F-score can have a maximum value of 1.0, which denotes perfect precision and recall, and a minimum value of 0, which occurs when either precision or recall is zero.

mAP 0.5 TABLE:

	YOLOv3	YOLOv5	YOLOR	YOLOv7
Plastic carry	0.477	0.229	0.202	0.213
Paper	0.127	0.189	0.242	0.201
Paper cup	0.59	0.632	0.647	0.682
Plastic	0.17	0.535	0.623	0.579
Plastic container	0.403	0.404	0.622	0.547
Paper cardboard	0.199	0.237	0.333	0.305
Cigarette packet	0.56	0.534	0.57	0.689
Plastic bottle	0.682	0.665	0.785	0.772
Newspaper	0.0593	0.0422	0.0865	0.107
Thermocol	0.181	0.186		0.248
Match Box	0.642	0.545	0.723	0.625
Metal	0.124	0.193	0.391	0.452
Cloth	0.0658	0.0219	0.123	0.0452
Plastic sack	0.032	0.0113	0.0212	0.0491
Medicine wrapper	0.0889	0.0937	0.198	0.129
Glass bottle	0.482	0.619	0.639	0.673
Glass	0.0206	0.000124	0.0174	0.00728
Nylon	0.000274	0.00939	0.0141	0.0195
Cotton	0.0789	0.205	0.24	0.196
Aluminium_foil	0.0236	0.495	0.496	0.495
Rubber	0.0188	0	0.199	0.995

Table 5.9 mAP@ 0.5 Table for each classes of YOLOv3,YOLOv5,YOLOR and YOLOv7

mAP 0.5:0.95 TABLE:

	YOLOv3	YOLOv5	YOLOR	YOLOv7
Plastic carry	0.477	0.144	0.135	0.153
Paper	0.127	0.136	0.181	0.154
Paper cup	0.59	0.469	0.499	0.502
Plastic	0.17	0.358	0.445	0.409
Plastic container	0.403	0.267	0.507	0.443
Paper cardboard	0.199	0.166	0.277	0.246
Cigarette packet	0.56	0.408	0.57	0.566
Plastic bottle	0.682	0.51	0.658	0.644
Newspaper	0.0593	0.0313	0.0841	0.103
Thermocol	0.181	0.145	0.166	0.191
Match Box	0.642	0.347	0.524	0.434
Metal	0.124	0.0901	0.23	0.319
Cloth	0.0658	0.0153	0.084	0.0419
Plastic sack	0.032	0.00575	0.201	0.0295
Medicine wrapper	0.0889	0.0827	0.162	0.114
Glass bottle	0.482	0.459	0.555	0.589
Glass	0.0206	8.67e-05	0.013	0.00582
Nylon	0.000274	0.00282	0.00639	0.0063
Cotton	0.0789	0.062	0.201	0.182
Aluminium_foil	0.0236	0.446	0.447	0.495
Rubber	0.0188	0	0.179	0.896

Table 5.10 mAP@0.5:0.95 Table for each classes of YOLOv3,YOLOv5,YOLOR and YOLOv7

TABLE FOR ALL:

	YOLOv3	YOLOv5	YOLOR	MT-YOLOv6	YOLOv7
Precision	0.178	0.382	0.254	0.157	0.432
Recall	0.377	0.334	0.459	0.271	0.394
F1 Score	0.228	0.356	0.327	0.198	0.412
mAP@0.5	0.239	0.278	0.358	0.21	0.382
mAP@0.5:0.95	0.239	0.197	0.283	0.156	0.311

Table 5.11 Precision, Recall,F1 Score, mAP@0.5 and mAP@0.95 Table for each classes of YOLOv3,YOLOv5,YOLOR and YOLOv7

CONFUSION MATRIX: A table called a confusion matrix is used to describe how well a classification system performs. The output of a classification algorithm is shown and summarized in a confusion matrix.

The confusion matrix is a matrix that displays how a classification model performed on a set of test data. It is sometimes referred to as the error matrix.

Four fundamental properties (numbers) make up the confusion matrix, which is used to provide the classifier's measurement parameters. These are the four numbers:

1)True positive (TP): Observation is predicted positive and is actually positive.

2)False positive (FP): Observation is predicted positive and is actually negative.

3)True negative (TN): Observation is predicted negative and is actually negative.

4)False negative (FN): Observation is predicted negative and is actually positive

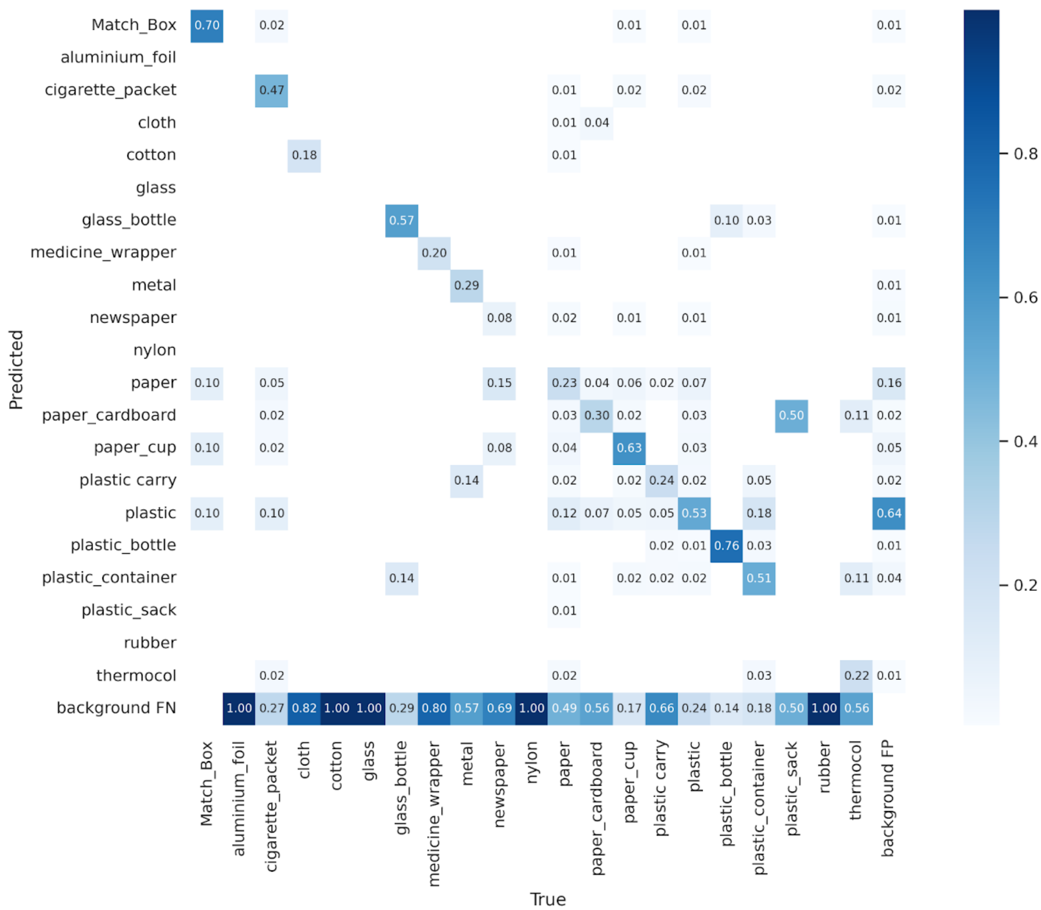


Fig 5.5 Confusion Matrix of YOLOv5 model

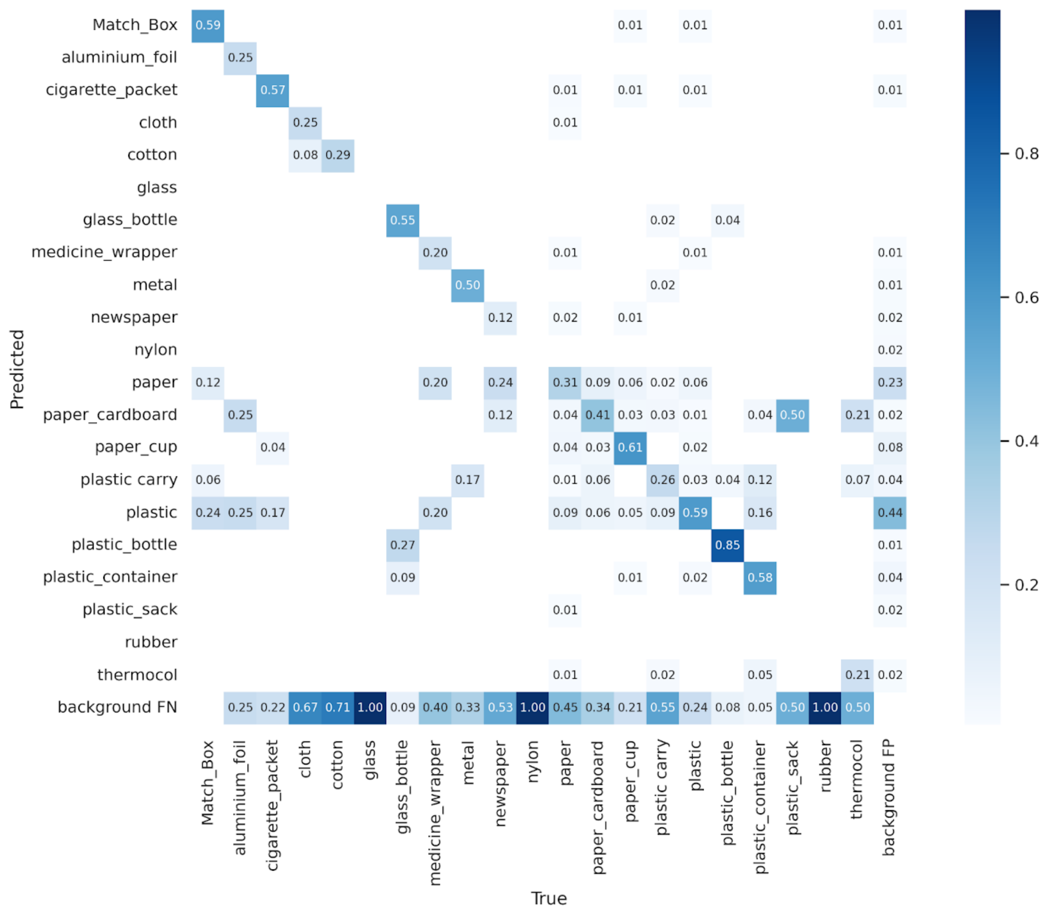


Fig 5.6 Confusion Matrix of YOLOR model

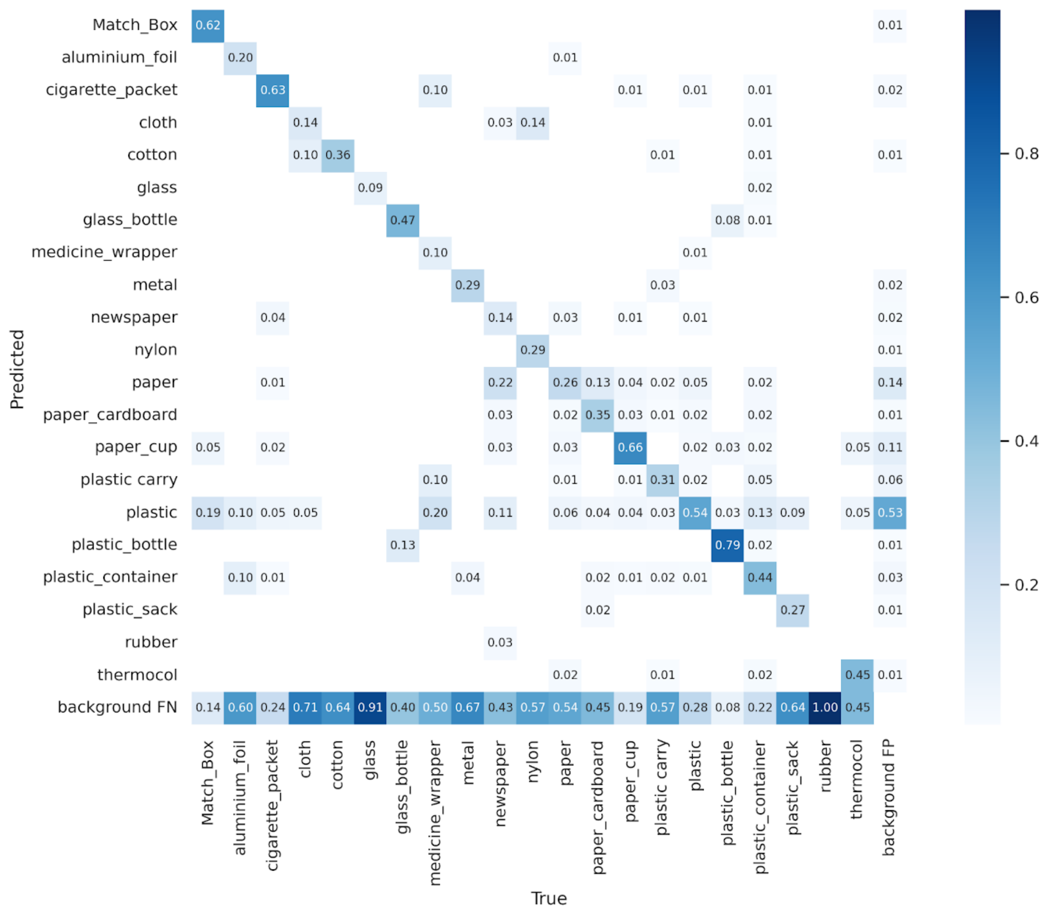
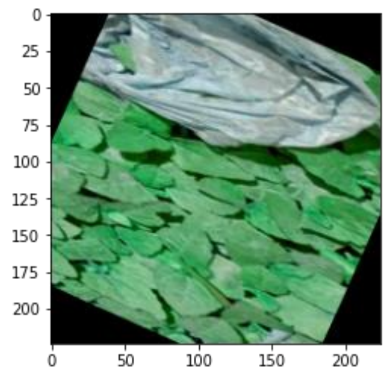


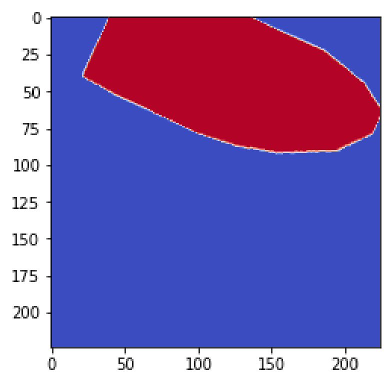
Fig 5.7 Confusion Matrix of YOLOv7 model

Model Name	Training Accuracy	Validation Accuracy
DeepLabv3+	98.12%	89.39%
UNet	97.28%	88.99%
LinkNet	98.00%	86.46%
PSPNet	89.80%	83.80%

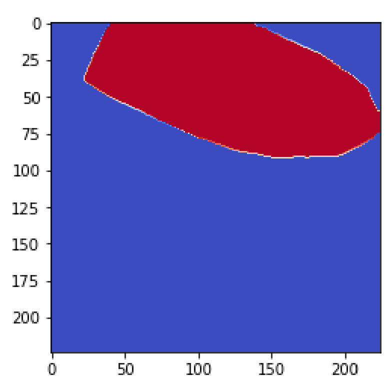
Table 5.12 Training Accuracy and Validation Accuracy for DeepLabv3+, UNet, LinkNet and PSPNet



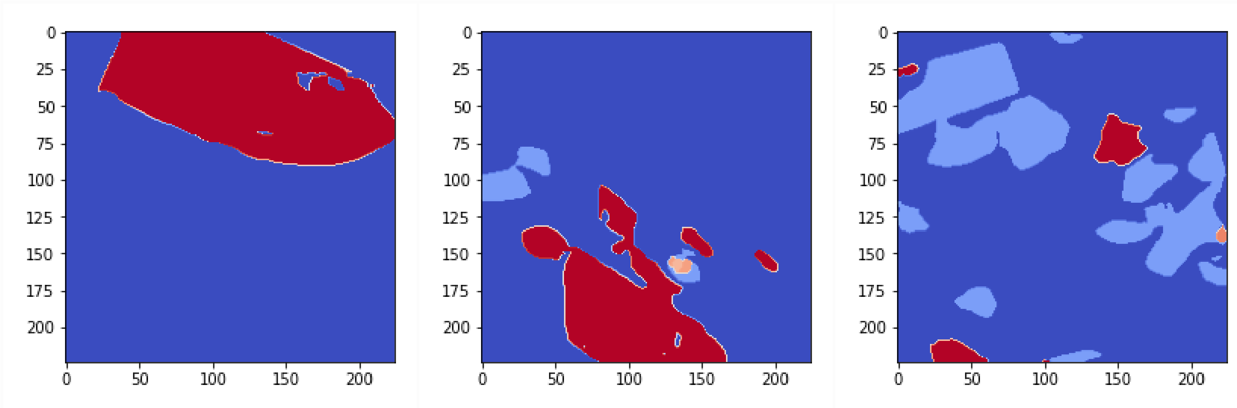
Images



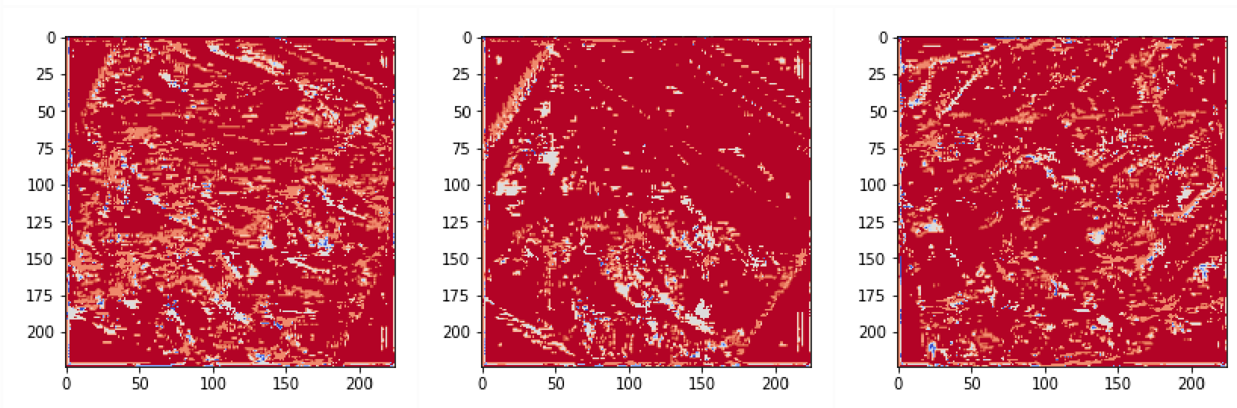
Ground Truth



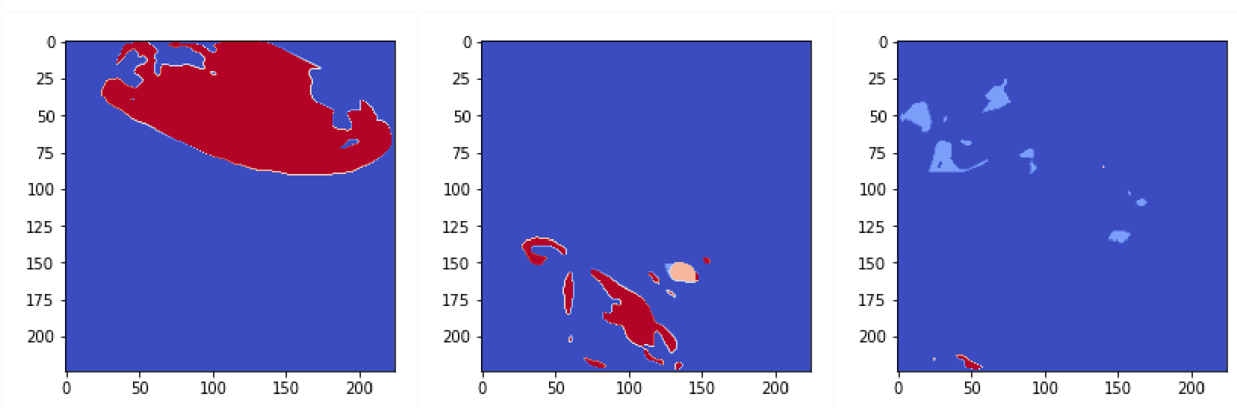
Prediction result of DeepLabv3+



Prediction result of UNet



Prediction result of LinkNet



Prediction result of PSPNet

CHAPTER 6

CONCLUSION

This study proposed an automated waste detection framework using deep learning algorithms and image processing techniques to reduce the effects caused by improper trash disposal. The framework was built on a huge collection of images, training methods, and predicative patterns for object detection and classification. In this thesis, I have shown how to use the YOLO algorithm to classify waste items into 21 classes based on the number of objects in a single image. In the majority of the papers studied for this work, I found classifications were conducted on a single object in an image with three or four categorizations using YOLO or other machine learning methods. My methodology offers an improvement in waste material classification by proposing novel procedures to detect waste items while retaining a higher level of precision. Not only can many waste products be detected from images, but they can also be recognized and categorized from any image, videoclip, or live webcam broadcast. The methods utilized in this study will contribute to a reduction in contamination levels caused by disintegration and mixing of waste with soil and water , with a long-term emphasis on the improvement of the global waste management system. Therefore, it can be said that this endeavor represents a huge benefit to society.

The dataset created for this research contains images of regional waste products that are slightly diverse, and is the cause of the model's inaccurate predictions on a few local garbage images. This should be taken into account for future work and employ a similar methodology but improve the models whose accuracy came up to 98.02% for segmentation dataset and 35.8% for localization dataset of the actual result. Images of the trash and other waste items in the training dataset that are unclean and appear dirty must be attached. This will aid the model's ability to predict local waste materials, which primarily consist of filthy domestic goods. In doing so, it may be possible to obtain improved categorization with a greater degree of accuracy. So that the framework can easily be trained to predict many objects in a single image without making any errors in the detection process. The dataset should also comprise images containing various waste products. The inclusion of other types of bulky trash categories in the dataset should also be considered for future research on this topic. This framework will become more developed and unquestionably contribute to the enhancement of an effective waste management procedure. Future work may also involve analyzing and contrasting various models, like DeepLabv3+, UNet, LinkNet, YOLOs , and others. Applying each model's various classification algorithms to the framework independently can be done for this analysis, which can then be used to determine which model is optimal for obtaining faster and more accurate item identification and prediction. In order to safeguard the environment and reduce pollution, future work will need to integrate this technology into robotic arms so that it is very simple for the user to segregate waste items and dispose of them in the appropriate disposal bin.

BIBLIOGRAPHY

[1] Majchrowska, S., Mikołajczyk, A., Ferlin, M., Klawikowska, Z., Plantykow, M.A., Kwasigroch, A. and Majek, K., 2022. Deep learning-based waste detection in natural and urban environments. Waste Management, 138, pp.274-284
[2] Gupta, A.K., Seal, A., Prasad, M. and Khanna, P., 2020. Salient object detection techniques in computer vision—A survey. Entropy, 22(10), p.1174.
[3] Zhu, L., Kolesov, I., Gao, Y., Kikinis, R. and Tannenbaum, A., 2014, September. An effective interactive medical image segmentation method using fast growcut. In MICCAI workshop on interactive medical image computing.
[4] Mottaghi, R., Chen, X., Liu, X., Cho, N.G., Lee, S.W., Fidler, S., Urtasun, R. and Yuille, A., 2014. The role of context for object detection and semantic segmentation in the wild. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 891-898).
[5] Yang, R. and Yu, Y., 2021. Artificial convolutional neural network in object detection and semantic segmentation for medical imaging analysis. Frontiers in oncology, 11, p.638182.
[6] Zhou, Q., Wang, Y., Fan, Y., Wu, X., Zhang, S., Kang, B. and Latecki, L.J., 2020. AGLNet: Towards real-time semantic segmentation of self-driving images via attention-guided lightweight network. applied soft computing, 96, p.106682.
[7] Girshick, R., Donahue, J., Darrell, T. and Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).
[8] Akgün, A., Eronat, A.H. and Türk, N., 2004, July. Comparing different satellite image classification methods: an application in Ayvalik District, Western Turkey. In The 4th International Congress for Photogrammetry and Remote Sensing, Istanbul, Turkey.
[9] Albawi, S., Mohammed, T.A. and Al-Zawi, S., 2017, August. Understanding of a convolutional neural network. In 2017 international conference on engineering and technology (ICET) (pp. 1-6). Ieee.
[10] Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural networks, 61, pp.85-117.
[11] Cheng, H.D., Jiang, X.H., Sun, Y. and Wang, J., 2001. Color image segmentation: advances and prospects. Pattern recognition, 34(12), pp.2259-2281.
[12] Liu, Z., Li, X., Luo, P., Loy, C.C. and Tang, X., 2015. Semantic image segmentation via deep parsing network. In Proceedings of the IEEE international conference on computer vision (pp. 1377-1385).
[13] Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C. and Torr, P.H., 2015. Conditional random fields as recurrent neural networks. In Proceedings of the IEEE international conference on computer vision (pp. 1529-1537).

[14] Voon, Z.C. and Chaw, J.K., 2020. Crack Segmentation Using DeepLab (Doctoral dissertation, Tunku Abdul Rahman University College).
[15] Huang, Q., Xia, C., Wu, C., Li, S., Wang, Y., Song, Y. and Kuo, C.C.J., 2017. Semantic segmentation with reverse attention. arXiv preprint arXiv:1707.06426.
[16] Dai, J., He, K. and Sun, J., 2015. Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. In Proceedings of the IEEE international conference on computer vision (pp. 1635-1643).
[17] Shetty, S., 2016. Application of convolutional neural network for image classification on Pascal VOC challenge 2012 dataset. arXiv preprint arXiv:1607.03785.
[18] Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z. and Lu, H., 2019. Dual attention network for scene segmentation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 3146-3154).
[19] Stehling, R.O., Nascimento, M.A. and Falcao, A.X., 2002, November. A compact and efficient image retrieval approach based on border/interior pixel classification. In Proceedings of the eleventh international conference on Information and knowledge management (pp. 102-109).
[20]BenAbdelkader, C. and Griffin, P., 2005, September. A local region-based approach to gender classi. cation from face images. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops (pp. 52-52). IEEE.
[21] Backes, A.R., Casanova, D. and Bruno, O.M., 2009. A complex network-based approach for boundary shape analysis. Pattern Recognition, 42(1), pp.54-67.
[22] Liu, M., Jiang, X. and Kot, A.C., 2005. Fingerprint reference-point detection. EURASIP Journal on Advances in Signal Processing, 2005(4), pp.1-12.
[23] Shrivakshan, G.T. and Chandrasekar, C., 2012. A comparison of various edge detection techniques used in image processing. International Journal of Computer Science Issues (IJCSI), 9(5), p.269.
[24] Aggarwal, N. and Karl, W.C., 2006. Line detection in images through regularized Hough transform. IEEE transactions on image processing, 15(3), pp.582-591.
[25] Hoang, N.D., Nguyen, Q.L. and Tien Bui, D., 2018. Image processing–based classification of asphalt pavement cracks using support vector machine optimized by artificial bee colony. Journal of Computing in Civil Engineering, 32(5), p.04018037.
[26] White, H.E., Ignatiou, A., Clare, D.K. and Orlova, E.V., 2017. Structural study of heterogeneous biological samples by cryoelectron microscopy and image processing. BioMed research international, 2017.
[27] Voloshynovskiy, S., Herrigel, A., Baumgaertner, N. and Pun, T., 1999, September. A stochastic approach to content adaptive digital image watermarking. In International workshop on information hiding (pp. 211-236). Springer, Berlin, Heidelberg.
[28] Likas, A., Vlassis, N. and Verbeek, J.J., 2003. The global k-means clustering algorithm. Pattern recognition, 36(2), pp.451-461.
[29] Agatonovic-Kustrin, S. and Beresford, R., 2000. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. Journal of pharmaceutical and biomedical analysis, 22(5), pp.717-727.
[30] Nyma, A., Kang, M., Kwon, Y.K., Kim, C.H. and Kim, J.M., 2012. A hybrid technique for medical image segmentation. Journal of Biomedicine and Biotechnology, 2012.

[31] Cheriet, M., Said, J.N. and Suen, C.Y., 1998. A recursive thresholding technique for image segmentation. IEEE transactions on image processing, 7(6), pp.918-921.
[32] Chaubey, A.K., 2016. Comparison of the local and global thresholding methods in image segmentation. World Journal of Research and Review, 2(1), pp.1-4.
[33]Noh, J.S. and Rhee, K.H., 2005, July. Palmprint identification algorithm using Hu invariant moments and Otsu binarization. In Fourth Annual ACIS International Conference on Computer and Information Science (ICIS'05) (pp. 94-99). IEEE.
[34] Kasar, T., Kumar, J. and Ramakrishnan, A.G., 2007, September. Font and background color independent text binarization. In Second international workshop on camera-based document analysis and recognition (pp. 3-9).
[35] Christov, I.I., 2004. Real time electrocardiogram QRS detection using combined adaptive threshold. Biomedical engineering online, 3(1), pp.1-9.
[36] Kaganami, H.G. and Beiji, Z., 2009, September. Region-based segmentation versus edge detection. In 2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (pp. 1217-1221). IEEE.
[37] Farag, A.A., 1992. Edge-based image segmentation. Remote sensing reviews, 6(1), pp.95-121.
[38] Haralick, R.M., 1987. Digital step edges from zero crossing of second directional derivatives. In Readings in Computer Vision (pp. 216-226). Morgan Kaufmann.
[39] Gould, S., Gao, T. and Koller, D., 2009. Region-based segmentation and object detection. Advances in neural information processing systems, 22.
[40] Chang, Y.L. and Li, X., 1994. Adaptive image region-growing. IEEE transactions on image processing, 3(6), pp.868-872.
[41] Aneja, K., Laguzet, F., Lacassagne, L. and Merigot, A., 2009, November. Video-rate image segmentation by means of region splitting and merging. In 2009 IEEE International Conference on Signal and Image Processing Applications (pp. 437-442). IEEE.
[42] Shafarenko, L., Petrou, M. and Kittler, J., 1997. Automatic watershed segmentation of randomly textured color images. IEEE transactions on Image Processing, 6(11), pp.1530-1544.
[43] Bieniek, A. and Moga, A., 2000. An efficient watershed algorithm based on connected components. Pattern recognition, 33(6), pp.907-916.
[44] Carranza, E.J.M. and Hale, M., 1997. A catchment basin approach to the analysis of reconnaissance geochemical-geological data from Albay Province, Philippines. Journal of Geochemical Exploration, 60(2), pp.157-171.
[45] Yang, Z., Chung, F.L. and Shitong, W., 2009. Robust fuzzy clustering-based image segmentation. Applied soft computing, 9(1), pp.80-84.
[46] Bezdek, J.C., Ehrlich, R. and Full, W., 1984. FCM: The fuzzy c-means clustering algorithm. Computers & geosciences, 10(2-3), pp.191-203.
[47] Hartigan, J.A. and Wong, M.A., 1979. Algorithm AS 136: A k-means clustering algorithm. Journal of the royal statistical society. series c (applied statistics), 28(1), pp.100-108.
[48] Liu, X., Song, L., Liu, S. and Zhang, Y., 2021. A review of deep-learning-based medical image segmentation methods. Sustainability, 13(3), p.1224.
[49] He, K., Gkioxari, G., Dollár, P. and Girshick, R., 2017. Mask r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).

[50] Fan, Q., Zhuo, W., Tang, C.K. and Tai, Y.W., 2020. Few-shot object detection with attention-RPN and multi-relation detector. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4013-4022).
[51] De Brabandere, B., Neven, D. and Van Gool, L., 2017. Semantic instance segmentation with a discriminative loss function. arXiv preprint arXiv:1708.02551.
[52] Beiker, S.A., 2012. Legal aspects of autonomous driving. Santa Clara L. Rev., 52, p.1145.
[53] Bolya, D., Zhou, C., Xiao, F. and Lee, Y.J., 2019. Yolact: Real-time instance segmentation. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 9157-9166).
[54] Balaguer, C. and Abderrahim, M. eds., 2008. Robotics and automation in construction. BoD–Books on Demand.
[55] Kaymak, Ç. and Uçar, A., 2019. A brief survey and an application of semantic image segmentation for autonomous driving. Handbook of Deep Learning Applications, pp.161-200.
[56] Agarap, A.F., 2018. Deep learning using rectified linear units (relu). arXiv preprint arXiv:1803.08375.
[57] Demir, A., Yilmaz, F. and Kose, O., 2019, October. Early detection of skin cancer using deep learning architectures: resnet-101 and inception-v3. In 2019 medical technologies congress (TIPTEKNO) (pp. 1-4). IEEE.
[58] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M. and Berg, A.C., 2015. Imagenet large scale visual recognition challenge. International journal of computer vision, 115(3), pp.211-252.
[59] Al-Qizwini, M., Barjasteh, I., Al-Qassab, H. and Radha, H., 2017, June. Deep learning algorithm for autonomous driving using googlenet. In 2017 IEEE Intelligent Vehicles Symposium (IV) (pp. 89-96). IEEE.
[60] Recht, B., Roelofs, R., Schmidt, L. and Shankar, V., 2019, May. Do imagenet classifiers generalize to imagenet?. In International Conference on Machine Learning (pp. 5389-5400). PMLR.
[61] Mallya, A., Davis, D. and Lazebnik, S., 2018. Piggyback: Adapting a single network to multiple tasks by learning to mask weights. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 67-82).
[62] Caicedo, J.C. and Lazebnik, S., 2015. Active object localization with deep reinforcement learning. In Proceedings of the IEEE international conference on computer vision (pp. 2488-2496).
[63] Wang, X., Shrivastava, A. and Gupta, A., 2017. A-fast-rcnn: Hard positive generation via adversary for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2606-2615).
[64] Yang, M.; Thung, G. Classification of trash for recyclability status. CS229 Project Report 2016, 2016.
[65] Lowe, D.G. Distinctive image features from scale-invariant keypoints. International journal of computer vision 2004, 60, 91–110.
[66] Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. Communications of the ACM 2017, 60, 84–90.
[4] M. Yang and G. Thung. Trashnet. https://github.com/garythung/trashnet, 2016.

[67] Cenk Bircanoglu and Nafiz Arıca. A comparison of activation functions in artificial neural networks. In ˘ 2018 26th Signal Processing and Communications Applications Conference (SIU), pages 1–4, 05 2018.
[68] Yinghao Chu, Chen Huang, Xiaodan Xie, Bohai Tan, Shyam Kamal, , and Xiaogang Xiong. Multilayer hybrid deep-learning method for waste classification and recycling. Computational Intelligence and Neuroscience, 2018, nov 2018.
[69] Ibrahim, Mostafa S.; Badr, Amr A.; Abdallah, Mostafa R.; Eissa, Ibrahim F. (2012). Bounding Box Object Localization Based On Image Superpixelization. Procedia Computer Science, 13(), 108–119. doi:10.1016/j.procs.2012.09.119
[70] M. Fulton, J. Hong, M. J. Islam, and J. Sattar. Robotic detection of marine litter using deep visual detection models. In 2019 International Conference on Robotics and Automation (ICRA), pages 5752–5758, 2019.
[71] Jungseok Hong, Michael Fulton, and Junaed Sattar. Trashcan: A semantically-segmented dataset towards visual detection of marine debris, 2020.
[72] Pedro F Proença and Pedro Simões. Taco: Trash annotations in context for litter detection. arXiv preprint arXiv:2003.06975, 2020.
[73] B. D. Carolis, F. Ladogana, and N. Macchiarulo. Yolo trashnet: Garbage detection in video streams. In 2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS), pages 1–7, 2020.
[74] Tao Wang, Yuanzheng Cai, Lingyu Liang, and Dongyi Ye. A multi-level approach to waste object segmentation. Sensors, 20(14), 2020.
[75] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In ECCV, 2020.
[76] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020
[77] 4. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521 (7553):436-444
[78] Liu L, Ouyang W, Wang X, Fieguth P, Chen J, Liu X, Pietikäinen M (2020) Deep Learning for Generic Object Detection: A Survey. International Journal of Computer Vision 128 (2):261-318. doi:10.1007/s11263-019-01247-4
[79] Ren S, He K, Girshick R, Sun J (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (6):1137-1149. doi:10.1109/TPAMI.2016.2577031
[80] He K, Gkioxari G, Dollar P, Girshick R (2018) Mask R-CNN. IEEE transactions on pattern analysis and machine intelligence. doi:10.1109/tpami.2018.2844175
[81] He K, Gkioxari G, Dollar P, Girshick R (2018) Mask R-CNN. IEEE transactions on pattern analysis
and machine intelligence. doi:10.1109/tpami.2018.284417
[82] Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A Learning deep features for discriminative localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016. pp 2921-2929
[83] R. Girshick and J. P. N. Fotheringham-Smythe and G. Gamow. Fast R-CNN. In International Conference on Computer Vision (ICCV), 2015.

[84]] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[85] Jifeng Dai, Yi Li, Kaiming He, and Jian Sun. R-FCN: object detection via region-based fully convolutional networks. In Advances in Neural Information Processing Systems (NIPS), 2016.
[86] Wen, X., Zhou, X., and Hu, H. 2008. The new process in integrated e-waste management in China. In 2008 IEEE International Symposium on Electronics and the Environment (pp. 1-6).
[87] Yuan, F., and Guanying, X. 2009. The Research of Hazardous Waste Management System in Guangdong Province Based on WebGIS. In 2009 International Conference on Environmental Science and Information Application Technology (pp. 317-319).
[88] Viola, P. and Jones, M. (2001) Rapid Object Detection Using a Boosted Cascade of Simple Features. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, 8-14 December 2001, I-511-I-518.
[89] Hoff, S., Meyer, M., and Sachs, J. 1998. Analysis of the General Packet Radio Service (GPRS) of GSM as access to the Internet. In undefined (pp. 415 - 419 vol.1).
[90] O'Carroll P. Informatics Training for CDC Public Health Advisors: Introduction and Background, http://faculty.washington.edu/~ocarroll/infrmatc/home.htm, posted July 1997.
[91] Hafiz, A.M. and Bhat, G.M., 2020. A survey on instance segmentation: state of the art. International journal of multimedia information retrieval, 9(3), pp.171-189.
[92] Liang, S. and Gu, Y., 2021. A deep convolutional neural network to simultaneously localize and recognize waste types in images. Waste Management, 126, pp.247-257.
[93] Wu, T., Huang, Q., Liu, Z., Wang, Y. and Lin, D., 2020, August. Distribution-balanced loss for multi-label classification in long-tailed datasets. In European Conference on Computer Vision (pp. 162-178). Springer, Cham.
[94] Hsu, D.J., Kakade, S.M., Langford, J. and Zhang, T., 2009. Multi-label prediction via compressed sensing. Advances in neural information processing systems, 22.
[95] Nowakowski, P. and Pamuła, T., 2020. Application of deep learning object classifier to improve e-waste collection planning. Waste Management, 109, pp.1-9.
[96] Mao, W.L., Chen, W.C., Wang, C.T. and Lin, Y.H., 2021. Recycling waste classification using optimized convolutional neural network. Resources, Conservation and Recycling, 164, p.105132.
[97] L. Breiman. Random forests. Machine learning, 45(1):5–32, 2001
[98] R.M. Bell and Y. Koren. Lessons from the netflix prize challenge. ACM SIGKDD Explorations Newsletter, 9(2):75–79, 2007.
[99] A. Berg, J. Deng, and L. Fei-Fei. Large scale visual recognition challenge 2010. www.imagenet.org/challenges. 2010
[100] Joseph Redmon, et al. "YOLOv3: An Incremental Improvement". CoRR abs/1804.02767. (2018).
[101]Glenn Jocher. ultralytics/yolov5:v3.1- Bug Fixes and Performance Improvements. https://github.com/ultralytics/yolov5, October 2020.
[102] Wang, C.Y., Yeh, I.H., & Liao, H.Y. (2021). You Only Learn One Representation: Unified Network for Multiple Tasks. arXiv preprint arXiv:2105.04206.
[103] https://github.com/meituan/YOLOv6
[104]J.R.R. Uijlings, K.E.A. van de Sande, T. Gevers, A.W.M. Smeulders Selective search for object recognition Int. J. Comput. Vis., 104 (2) (2013), pp. 154-171

[105] Ronneberger, O., Fischer, P. and Brox, T., 2015, October. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.
[106] Q. Tong, M. Ning, W. Si, X. Liao, and J. Qin, "3D deeplysupervised U-net based whole heart segmentation," in International Workshop on Statistical Atlases and Computational Models of the Heart, 2017, pp. 224–232.
[107]. Ciresan, D.C., Gambardella, L.M., Giusti, A., Schmidhuber, J.: Deep neural networks segment neuronal membranes in electron microscopy images. In: NIPS, pp. 2852–2860 (2012)
[106]. Hariharan, B., Arbel´aez, P., Girshick, R., Malik, J.: Hypercolumns for object segmentation and fine-grained localization (2014), arXiv:1411.5752 [cs.CV]
[108] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NIPS, pp. 1106–1114 (2012)
[109]. Cardona, A., et al.: An integrated micro- and macroarchitectural analysis of the drosophila brain by computer-assisted serial section electron microscopy. PLoS Biol. 8(10), e1000502 (2010)
[110] Dosovitskiy, A., Springenberg, J.T., Riedmiller, M., Brox, T.: Discriminative unsupervised feature learning with convolutional neural networks. In: NIPS (2014)
[111] Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T. and Ronneberger, O., 2016, October. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In International conference on medical image computing and computer-assisted intervention (pp. 424-432). Springer, Cham.
[112] Lienkamp, S., Ganner, A., Boehlke, C., Schmidt, T., Arnold, S.J., Sch¨afer, T., Romaker, D., Schuler, J., Hoff, S., Powelske, C., Eifler, A., Kr¨onig, C., Bullerkotte, A., Nitschke, R., Kuehn, E.W., Kim, E., Burkhardt, H., Brox, T., Ronneberger, O., Gloy, J.,Walz, G.: Inversin relays frizzled-8 signals to promote proximal pronephros development. PNAS 107(47), 20388–20393 (2010)
[113] Hariharan, B., Arbel´aez, P., Girshick, R., Malik, J.: Hypercolumns for object segmentation and fine-grained localization. In: Proceeding CVPR, pp. 447–456 (2015)
[114] Milletari, F., Ahmadi, S., Kroll, C., Plate, A., Rozanski, V.E., Maiostre, J., Levin, J., Dietrich, O., Ertl-Wagner, B., B¨otzel, K., Navab, N.: Hough-CNN: deep learning for segmentation of deep brain regions in MRI and ultrasound. CoRR abs/1601.07014 (2016)
[115] Kleesiek, J., Urban, G., Hubert, A., Schwarz, D., Maier-Hein, K., Bendszus, M., Biller, A.: Deep mri brain extraction: a 3D convolutional neural network for skull stripping. NeuroImage (2016)
[116] Tran, D., Bourdev, L.D., Fergus, R., Torresani, L., Paluri, M.: Deep end2end voxel2voxel prediction. CoRR abs/1511.06681 (2015)
[117] LeCun, Y. and Bengio, Y., 1995. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, *3361*(10), p.1995.
[118] Konovalenko, I., Maruschak, P., Brezinová, J., Viňáš, J. and Brezina, J., 2020. Steel surface defect classification using deep residual neural network. *Metals*, *10*(6), p.846.
[119] Glouche, Y. and Couderc, P., 2013, June. A smart waste management with self-describing objects. In *The second international conference on smart systems, devices and technologies (SMART'13)*.
[120] Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q., 2017. Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).

[121] Majchrowska, S., Mikołajczyk, A., Ferlin, M., Klawikowska, Z., Plantykow, M.A., Kwasigroch, A. and Majek, K., 2022. Deep learning-based waste detection in natural and urban environments. Waste Management, 138, pp.274-284.
[122] Mulim, W., Revikasha, M.F. and Hanafiah, N., 2021, October. Waste Classification Using EfficientNet-B0. In 2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI) (Vol. 1, pp. 253-257). IEEE.
[123] Majchrowska, S., Mikołajczyk, A., Ferlin, M., Klawikowska, Z., Plantykow, M.A., Kwasigroch, A. and Majek, K., 2022. Deep learning-based waste detection in natural and urban environments. Waste Management, 138, pp.274-284.
[124] Tan, M. and Le, Q., 2021, July. Efficientnetv2: Smaller models and faster training. In International Conference on Machine Learning (pp. 10096-10106). PMLR.
[125] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. Lecun. Overfeat: Integrated recognition, localization and detection using convolutional networks. In ICLR, 2014.
[126] M. Najibi, B. Singh, and L. S. Davis. Fa-rpn: Floating region proposals for face detection. In CVPR, 2019.
[127] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. In CVPR, 2016.
[128] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg. SSD: Single shot multibox detector. In ECCV, 2016.
[129] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar. Focal loss for dense object detection. In ICCV, 2017.
[130] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In NIPS, 2015.
[131] M. Everingham, L. Gool, C. K. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. IJCV, 88(2):303–338, 2010.
[132] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. DollÃ ˛ar, and C. L. Zitnick. Microsoft coco: Common objects in context. In ECCV, 2014.
[133] C. C. Loy, D. Lin, W. Ouyang, Y. Xiong, S. Yang, Q. Huang, D. Zhou, W. Xia, Q. Li, P. Luo, et al. WIDER face and pedestrian challenge 2018: Methods and results. arXiv:1902.06854, 2019.
[134] Z. Cai and N. Vasconcelos. Cascade r-cnn: Delving into high quality object detection. In CVPR, 2018.
[135] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask R-CNN. In ICCV, 2017.
[136] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie. Feature pyramid networks for object detection. In CVPR, 2017.
[137] Elharrouss, O., Al-Maadeed, S., Subramanian, N., Ottakath, N., Almaadeed, N. and Himeur, Y., 2021. Panoptic segmentation: a review. arXiv preprint arXiv:2111.10250.
[138] Ghosh, S., Das, N., Das, I. and Maulik, U., 2019. Understanding deep learning techniques for image segmentation. ACM Computing Surveys (CSUR), 52(4), pp.1-35.
.[139] Sohn, K., Shang, W., Yu, X. and Chandraker, M., 2018, September. Unsupervised domain adaptation for distance metric learning. In International Conference on Learning Representations.
[140]European household waste collection (Eurostat). 2018. Waste statistics [Data file]. Retrieved from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Waste_statistics.

[141] Kaza, Silpa, Lisa C. Yao, Perinaz Bhada-Tata, and Frank Van Woerden. 2018. What a Waste 2.0. World Bank Publications 30317. The World Bank. https://ideas.repec.org/b/wbk/wbpubs/30317.html.
s. IOT ’18. New York, NY, USA: Association for Computing Machinery. doi: 10.1145/3277593.3277627.
[145] Big data analytics t[142] Plastic waste in the marine environment: a review of sources, occurrence and effects
Sci. Total Environ., 566-567 (2016), pp. 333-349
[143] White, Gary, Cabrera, Christian, Palade, Andrei, Li, Fan, Clarke, Siobhan, 2020. WasteNet: Waste Classification at the Edge for Smart Bins.
[144] Tavakoli, Arash, Ashrafi, Amir, Heydarian, Arsalan, Behl, Madhur, 2018. The internet of wasted things (Iowt). In: IOT ’18: Proceedings of the 8th International Conference on the Internet of Thingo identify illegal construction waste dumping: a Hong Kong study
Resour. Conservat. Recycl., 141 (2019), pp. 264-272, 10.1016/j.resconrec.2018.10.039
[146] M. Kraft, M. Piechocki, B. Ptak, K. Walas
Autonomous, onboard vision-based trash and litter detection in low altitude aerial images collected by an unmanned aerial vehicle
Remote Sens., 13 (5) (2021), 10.3390/rs13050965