Seminar Report

Entitled

"YOLO : A SINGLE PASS TO DETECT & IDENTIFY OBJECTS "

Submitted in

Partial fulfilment for the award of the Degree of

Bachelor Of Technology

(Electrical Engineering)

Presented and Submitted By :

Pranav Dua

Under the guidance of

Dr H G Patel

B. Tech: 3rd Year - 5th Sem

Year: 2023-24

Sardar Vallabhbhai National Institute of Technology

Surat-395007, Gujarat, INDIA

Saradar Vallabhbhai National Institute Of Technology

Surat 395007 , Gujarat, INDIA

ELECTRICAL ENGINEERING DEPARTMENT

CERTIFICATE

This is to certify that the **seminar report** entitled **"YOLO : A Single Pass to Detect & Identify Objects "** is presented & submitted by Candidate **Pranav Dua** , bearing Roll No. **U21EE078 of** BTech. III , 5th Semester under the guidance of Dr. H G Patel in the partial fulfillment of the requirement for the award of B. Tech degree in **Electrical Engineering** for the academic year **2023-24**. He has successfully and satisfactorily completed his/her **Seminar Exam** in all respect. We, certify that the work is comprehensive, complete, and fit for evaluation.

Date:

Place: SURAT

Signature and date
Supervisor(s)

Signature and date
Examiners

Signature and date
Head of Department

Saradar Vallabhbhai National Institute Of Technology

Surat 395007 , Gujarat, INDIA

ELECTRICAL ENGINEERING DEPARTMENT

DECLARATION

I hereby declare that the **seminar report** entitled **" YOLO : A Single Pass to Detect & Identify Objects"** being presented & submitted by Candidate **Pranav Dua** , bearing Roll No. **U21EE078 of** BTech. III , 5th Semester in the partial fulfillment of the requirement for the award of B. Tech degree in **Electrical Engineering** for the academic year **2023-24** is an authentic record of my own work done under the guidance of Dr. H G Patel

Dr. H G Patel

Seminar Guide Signature and date

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to all those who have been instrumental in the successful completion of this seminar report. It is with their support and encouragement that this endeavor has come to fruition.

First and foremost, I would like to extend my sincerest appreciation to Dr. HG Patel sir, my seminar guide, for their invaluable guidance, unwavering support, and patience throughout this project. Their expertise and mentorship were crucial in shaping the content and direction of this report.

I would like to acknowledge my fellow students and friends for their intellectual discussions, feedback, and moral support, which greatly enriched my understanding of the subject matter and kept me motivated during the course of this project.

I am deeply thankful to my family for their unwavering support, love, and understanding. Their encouragement has been a constant source of strength throughout my academic journey.

This report would not have been possible without the collective efforts and contributions of these individuals and organizations. I am truly grateful for their assistance in making this seminar report a reality.

PRANAV DUA

SARDAR VALLABHBHAI NATIONAL INSTITUTE OF TECHNOLOGY SURAT

20 / 10 / 2023

ABSTRACT

When we look at images or videos, we can easily locate and identify the objects of our interest within moments. Passing on this intelligence to computers is nothing but object detection - locating the object and identifying it. Object Detection has found its application in a wide variety of domains such as video surveillance, image retrieval systems, autonomous driving vehicles and many more. Various algorithms can be used for object detection but we will be focusing on the Yolo algorithm. YOLO stands for "You Only Look Once". The YOLO model is very accurate and allows us to detect the objects present in the frame. YOLO follows a completely different approach. Instead of selecting some regions, it applies a neural network to the entire image to predict bounding boxes and their probabilities. YOLO is a single deep convolutional neural network that splits the input image into a set of grid cells, so unlike image classification or face detection, each grid cell in YOLO algorithm will have an associated vector in the output that tells us if an object exists in that grid cell, the class of that object, the predicted bounding box for that object. The model here is progressive so it learns more over time, increasing its prediction accuracy over time. The way the model works is that it makes many predictions in one frame and decides to use the most accurate prediction, thus discarding the other. The predictions are made randomly, so if the model feels like there is an object in the frame which is of a very small pixel it will take that also into consideration. To make it more precise and clearer, the model simply creates bounding boxes around everything in the frame, it would make predictions for each box and pick the one with the most confidence score. All this is done in a small-time frame, thus showing why this specific model is the best to use in a real time situation.

CONTENTS

CHAPTER 1: INTRODUCTION

1.1 Object Detection

Object detection is a computer equipment which is related to computer visualization and image processing that deals with detecting examples of semantic objects of a certain class (such as humans, buildings, or cars) in digital images and videos. The wide area of applications in object detection is face detection, face recognition and video object detection. Some of the applications are, tracking motion of the ball, tracking ball during the match, tracking person in a video. Normally, Object detection has uses in many areas of computer vision, including image fetching and video surveillance .The object detection system recognizes the presence or the absence of the objects in certain scenes and the cameras viewpoints. The various domains of the object detection based on the different objectives and classified on specific and conceptual categories. The object detection based on the various models either explicitly or implicitly. The components may vary based on the different approaches. The selection of the object based on the hypothesis and the selection based on the matching. The object detection is an appropriate technique for the processing. The object detection is the searching of the objects where the images are found in real world applications[1].

1.1 Object Detection

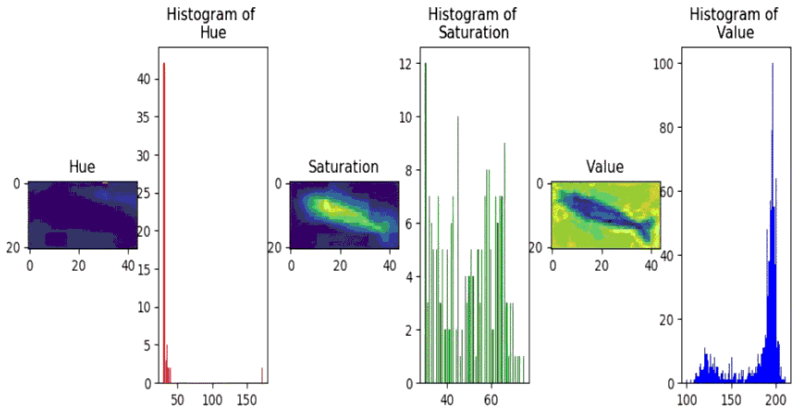
1.2 Features of the Object Detection

In the object detection, tracking and the selection of the various characteristics features that can reduce the work accessibility of the computer. When the tracking is done using various algorithms the combination of the different features determined in various steps:-

- **Color** - The feature of the computer system that is used for the histogram appearance representations. The widest features of the color representations are the features of the

color representations for the tracking. The features of the color are tracking of serious problem which recognize the illumination variation.

- **Histogram of gradients** - The HOG feature is the most popular feature used for the detection of the human body. The operations of the histogram feature based on the local grid unit of the image. So the geometric variations influence the optical deformations. Moreover, the sampling orientation and local optimization maintain the upright posture and body movements. These movements do not influence the detection phase which is the main reason of HOG feature in detection of humans.
- **Edges** - The boundaries of the image intensities may change during the identification of the object detection. The feature of the object detection is different from the color features technique.
- **Optical Flow** - The feature based on the motion segmentation and the applications of the tracking. The displacement vector recognize about the every pixel of the region. The displacement vector is that which determines the transactions of each pixel of each image. Optical flow is usually used as a feature in motion-based segmentation and tracking applications. It is a dense field of the displacement vectors which defines the translation of each pixel in a region[2]. It is computed using the brightness constraint, which assumes brightness constancy of consistent pixels in consecutive frames. With the development of technology, there are many popular techniques for computing dense optical flow, such as Horn-Schunck Algorithm.



1.2 Features of object detection

1.3 Challenges of Object Detection

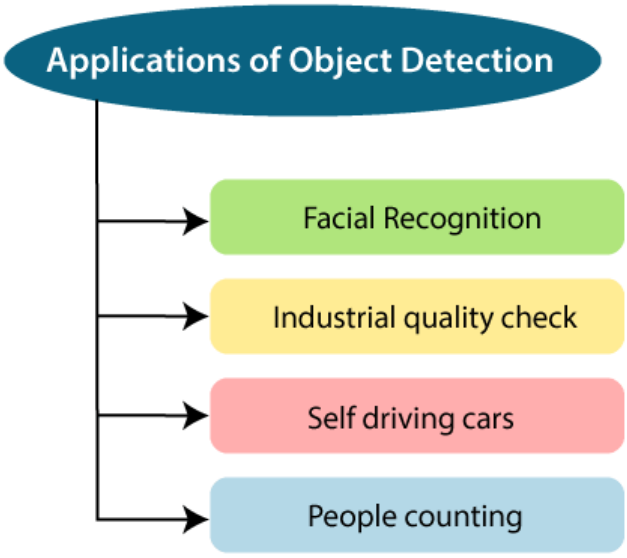
- **Positioning** -In this process, the position of the image can be changes at any time. In the template matching the system will handle the images uniformly in the system.
- **Lighting** - In this lighting conditions may change during the course of the system. The changes in the weather may affects the lighting of an image. In such case, the lighting condition may vary with the time. The shadow of the image affects the image lighting system. The detection of object from an image can be done during any condition of the lighting.
- **Rotation** -The images may be rotated t where the system may be capable of handling such type of the difficulty. For instance, character may appear in any form, but the orientations of an image are not affected by the detection of the character.
- **Mirroring** - The images which are mirrored of any object can be detected by the object detection system
- **Occlusion condition** - When an object are not visible then then image and that condition is referred as occlusion.
- **Scaling method-** The object detection system are not affected by the change in the size of the object. The challenges may occur due to the object detection. The scaling method is the process of the recognition of the scaling of the images in the object detection[3].

1.3 Challenges of Object Detection

1.4 Applications in Object Detection

The various applications of the object detection include various categories are:-

- **Biometric Detection**- The security and the authentication are determined through the physical and the behaviour traits. The biometric features is the identification of the individual that is based on the biological features such as finger prints, hand geometry, retina and iris patterns, DNA. The object detection method can determine through the template-based matching.
- **Surveillance systems** - The videos and object are tracked through the surveillance systems. The object detection is the tracking of the suspected persons or the individuals.
- **Inspection of the industries** - The different parts of the machines of the industry are recognised using the object detection method and the manufactured products or the damage to the products through the monitoring procedure.
- **Content-based image retrieval (CBIR)** - The retrieval of the image based on the content of the image is Content-based image retrieval. In this supervised learning system, automated keyword, annotation of the images and content based image[4].
- **Autonomous Robotics**- The autonomous robots method of the research is the main issue in the recent world of the research. The human robotic systems are the most popular system technique. The vision of the relied system based on the computational behaviour. The termination data is taken as the final data that recognize the functional methods. The computation errors can be decreased through the features of the object obtained from the object recognition methods.
- **Analysis of the medical analysis** -Tumour detection in MRI images, skin cancer detection can be some examples of medical imaging for object recognition.
- **Document Recognition** -In this the documents are scanned through the detection methods.
- **Interaction of computer Systems** - In this method the storage of the human gestures in the real time environment and the interaction with the humans. In the interaction system the type of applications are mobile phone , interactive games.
- **Intelligent vehicle systems** - The traffic signs are detected using traffic signs detections and recognitions methods. In this detection phase, the scanning of the scene image determined by the region of the interest (ROI).The sign candidates in the ROI recognised through the waveform features. The speeds up features determine the templates of the data set and the template images are recognized using the maximum number of matches.



1.4 Application of Object Detection

1.5 Methods of Object Detection

The various object detection methods are described as

- **Template Based Object Detection** - In this method, the small parts of the image can be recognised using the template image. This technique is also called as the template matching. The quality of the mobile robot parts of the image used as the quality control image and also detects the edges of the image. The relation between the template image and the real image are detected through the geometrical parameters. The data image for the template matching use different iterations for the geometrical parameters. The geometrical parameters with the search images S(x,y) where (x, y) represent the coordinates of each pixel in the search image. The methods are implemented by using search image to find the templates. The origin of the template move over every point of the search image and compute the amount of products between the coefficients over the whole area spanned by the template. The positions are considered with highest score positions. The method is spatial filtering and the template is filter mask.
- **Part Based Object Detection** - The collection of the deformable configuration for the representation of an object. Each part of the models is separately arranged with the deformed configuration and that are represented by the connections between the pairs of the parts. Such models determine the visual appearance of the qualitative descriptions and that are suitable for generic recognition problems[5].
- **Region Based object detection** - The transformation of the input image into directed graph through various rules determined by an algorithm. The graph characteristics

symbolize the global shape information of the object inside the input image, and are extracted during the graph construction. The technique represents the traversing of the post preserving of the graph and that will improve the computational time of the graph. Such algorithm is tested through a specific database and that demonstrates the two problems which are object class recognition and similar image retrieval.

- **Contour Based Object Detection** -The image data base determines the various types of the objects that are stored by single prototype images. The cameras are used where the robots are located and identified by the passing objects. In order to recognize the objects and put them to the final position the image are taken by using cameras. This process based on the two phases. The first phase is the location of the individual objects and describes the polynomial shapes. Almost every kind of the shapes includes the strong connections of the holes. The other phase is the prototypes and detects the type of the objects and calculates the relative orientation. The alternative approach is the segmentation of the image and combining the polygon method that depends on the initial detection of the image. The precision of the triangular approximation is the statistical approach.
- **Appearance Based Detection** -In this method the object detection system based on dealing with the 3D recognition system of the objects in the presence of the occlusion and clutters. The appearances are used for the appearances of the images and the scenes. The two dimensional views of the objects based the main classes which are local and global approaches.
- **Background Subtraction**- In this method the foreground objects are subtracted from the background organizations in the photo frames. The technique in the background subtraction in non-recursive methods are Frame differencing, Median Filter, Linear predictive filter The main objective of this method is that the video frames are estimated using the background model and that statistical properties of frames consumes high memory data storage. The various methods maintain the background estimation based on the input frames and approximation of the median filter, Kalman filter. The recursive method can be declared as the process of the repetitions of the objects in similar way. The single background model can be updated using the new video frames that have the minimum memory storage that is compared to the non-recursive and the computational method[6].

1.6 Machine Learning

Machine learning is a form of artificial intelligence (AI) that allows computers to learn and develop on their own without being explicitly programmed. Machine learning is concerned with the development of computer programs that can access data and learn on their own.

The learning process begins with observations or data, such as examples, direct experience, or teaching, so that we may seek for patterns in data and make better judgments in the future based on the examples we offer. The basic objective is for computers to learn on their own, without human involvement, and to change their activities accordingly.

Methods of Machine Learning

There are three types of machine learning algorithms:-

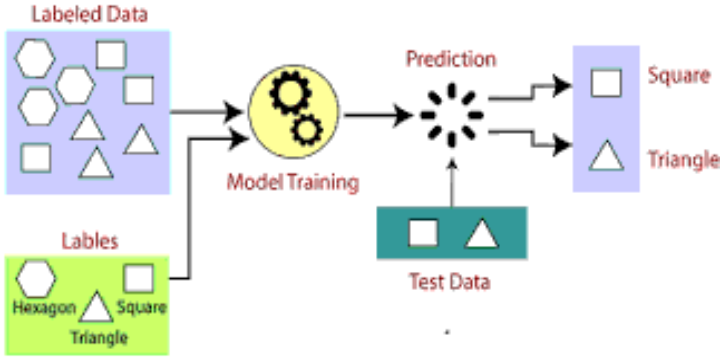
- **Supervised machine learning algorithms** may utilize labeled examples to apply what they've learnt in the past to fresh data and anticipate future events. Based on a study of a given training dataset, the learning algorithm generates an inferred function to make predictions about output values. After a sufficient amount of training, the system may offer objectives for any new input. The learning algorithm may also compare its output to the desired output and find faults, allowing the model to be appropriately adjusted.
- **Unsupervised machine learning algorithms** are utilized, on the contrary conjunction, when the data being trained is neither categorized nor labeled. Unsupervised learning studies how systems may infer a function to describe a hidden structure from unlabeled data. The system cannot choose the correct output, but it can explore the data and infer hidden structures from unlabeled data using datasets.
- **Reinforcement machine learning algorithms** are a form of learning algorithm that interacts with its environment by producing actions and recognizing faults or rewards. The key features of reinforcement learning are trial and error search and delayed reward. This strategy allows machines and software agents to automatically select the optimum course of action in a given scenario in order to maximize their efficiency. Simple reward feedback is necessary for the agent to learn which action is optimal; this is known as the reinforcement signal.

HOW MACHINE LEARNING WORKS

The learning system of a machine learning algorithm is broken down into three main parts-

- **A Prediction or Classification Process**: Machine learning algorithms are used to make predictions or classifications in general. Your algorithm will generate an estimate about a pattern in the data based on some input data, which can be labelled or unlabelled.
- **An Error Function**: An error function is used to assess the model's prediction. If there are known examples, an error function can be used to compare the model's accuracy.

- **Optimisation process** -Weights are adjusted to reduce the discrepancy between the known example and the model estimate if the model can fit better to the data points in the training set. This evaluate and optimise process will be repeated by the algorithm, which will update weights on its own until a certain level of accuracy is reached.



1.6 How machine learning works

1.7 Applications of Machine Learning

Image and Video Analysis:

Object Detection and Recognition: Machine learning enables the identification and localization of objects within images or video frames. This technology is used in various fields, including self-driving cars, surveillance systems, and medical imaging.

Facial Recognition: ML algorithms can recognize and verify individuals based on facial features, which is used in security systems, unlocking smartphones, and even identifying missing persons.

Content Recommendation: Video streaming platforms use ML to analyze user preferences and viewing history to provide personalized content recommendations.

Natural Language Processing (NLP):

Sentiment Analysis: NLP algorithms can determine the sentiment (positive, negative, neutral) of text data. This is widely used for social media monitoring, brand reputation management, and customer feedback analysis.

Machine Translation: Machine learning enables the automatic translation of text from one language to another, powering tools like Google Translate and making international communication more accessible.

Chatbots and Virtual Assistants: ML-driven chatbots and virtual assistants provide automated customer support, answer queries, and perform tasks, improving user experience.

Healthcare:

Disease Diagnosis and Medical Image Analysis: ML models are trained to analyze medical images, such as X-rays, MRIs, and CT scans, to aid in diagnosing diseases and conditions like cancer or fractures.

Drug Discovery and Development: ML helps identify potential drug candidates, predict their effectiveness, and streamline the drug development process.

Patient Risk Prediction and Personalized Treatment: ML models predict patient risks, assist in early diagnosis, and recommend personalized treatment plans based on medical data.

Finance:

Credit Scoring and Risk Assessment: Machine learning assesses an individual's creditworthiness by analyzing their financial history, helping banks and lenders make more informed lending decisions.

Algorithmic Trading: ML algorithms analyze market data to make high-frequency trading decisions, optimizing portfolio management and risk reduction.

Fraud Detection and Prevention: ML helps identify fraudulent transactions by analyzing patterns and anomalies in financial data, reducing financial fraud.

Autonomous Vehicles:

Self-Driving Cars: Machine learning plays a crucial role in self-driving vehicles, allowing them to perceive their surroundings, make driving decisions, and navigate safely.

Advanced Driver-Assistance Systems (ADAS): ADAS features such as lane-keeping assistance, adaptive cruise control, and automatic emergency braking utilize ML to enhance vehicle safety.

Industrial Automation:

Predictive Maintenance: ML models analyze sensor data to predict when machinery or equipment is likely to fail, enabling proactive maintenance and minimizing downtime.

Quality Control and Defect Detection: ML-powered computer vision systems inspect products on assembly lines, identifying defects and ensuring quality.

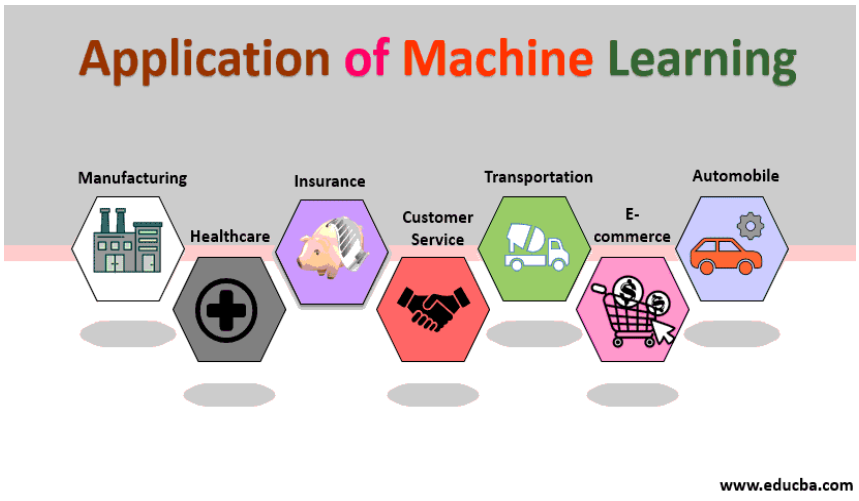
Supply Chain Optimization: ML optimizes supply chain operations, including demand forecasting, inventory management, and logistics, leading to cost savings and efficiency improvements.

Agriculture:

Crop Monitoring and Yield Prediction: ML models analyze satellite and drone imagery to monitor crop health, predict yields, and guide farming decisions.

Pest and Disease Detection: Image recognition and ML detect signs of pests or diseases in crops, allowing farmers to take timely actions.

Precision Agriculture: ML-driven tools optimize the use of resources like water and fertilizers, reducing waste and environmental impact.



1.7 Application of Machine Learning

1.8 Challenges to Machine Learning

1**. Data Quality and Quantity**: Machine learning models require large volumes of high-quality data for training. Obtaining and preparing such data can be resource-intensive and challenging, particularly in domains with limited or noisy data.

2. **Data Privacy and Security:** The use of personal and sensitive data for training machine learning models raises concerns about privacy and security. Ensuring compliance with data protection regulations and safeguarding against data breaches is crucial.

3. **Bias and Fairness:** Machine learning models can inherit biases present in the training data, leading to biased predictions and discrimination. Ensuring fairness and mitigating bias in machine learning algorithms is a critical challenge.

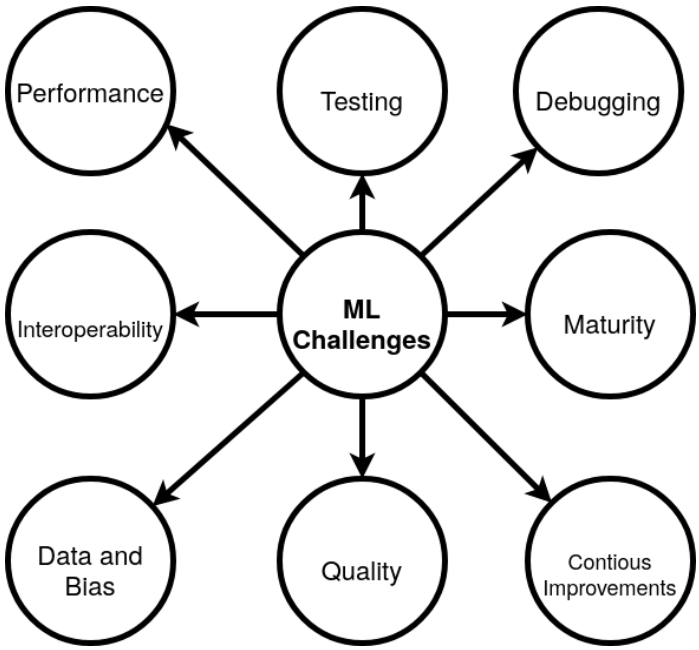
4. **Interpretability:** Many machine learning models, such as deep neural networks, are often seen as "black boxes" due to their complexity. Understanding and interpreting model decisions is important, especially in critical applications like healthcare and finance.

5. **Generalization and Overfitting:** Achieving a balance between a model's ability to generalize from the training data and its tendency to overfit (fitting noise in the data) is a common challenge in machine learning.

6. **Algorithm Selection:** Selecting the appropriate machine learning algorithm for a specific task is not always straightforward. The choice depends on the nature of the problem, the available data, and other factors, and it can significantly impact model performance.

7. **Scalability:** Training large machine learning models can be computationally intensive and may require specialized hardware. Scaling up models for big data can be a technical and cost challenge.

8. **Model Robustness:** Machine learning models can be vulnerable to adversarial attacks, where slight modifications to input data can lead to incorrect or malicious outputs. Ensuring model robustness is crucial in security-critical applications.



1.8 Machine Learning Challenges

CHAPTER 2: YOLO (YOU LOOK ONLY ONCE)

2.1 Background

The YOLO (You Only Look Once) algorithm has emerged as a transformative force in the domain of computer vision, specifically in the context of object detection. Its significance lies in its unparalleled efficiency and speed, enabling the rapid and precise detection and classification of objects within images or video frames, all within a single pass through a neural network. Unlike conventional object detection methods, which often involve intricate multi-step procedures, YOLO simplifies the process by simultaneously performing object

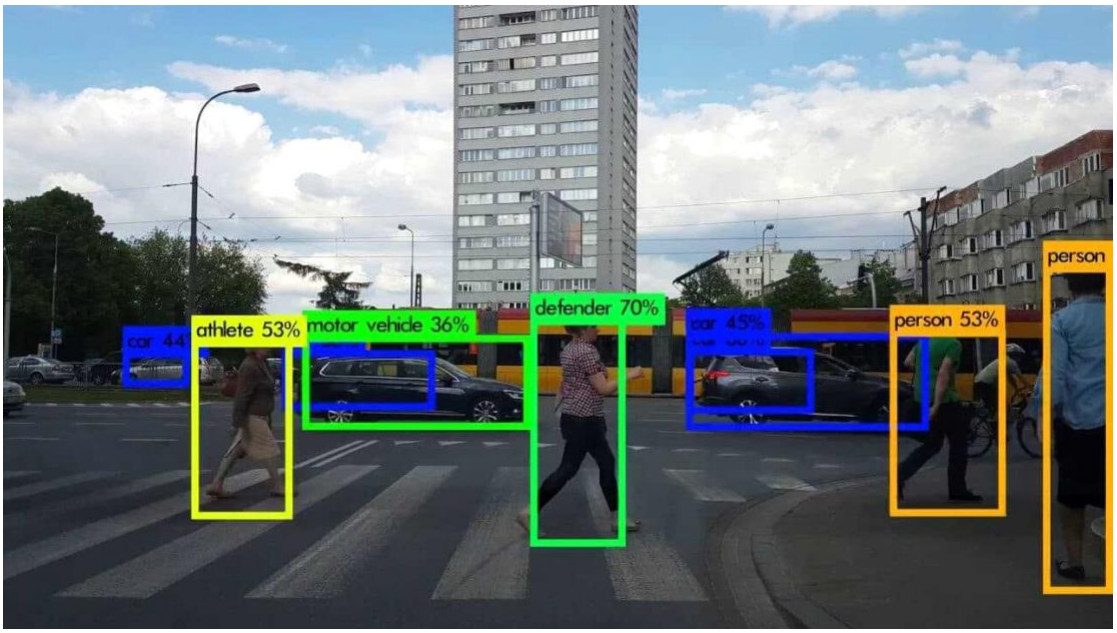
localization, through bounding boxes, and categorization, placing each object into predefined classes or categories.

The YOLO algorithm's distinguishing feature is its real-time processing capability, which has made it an essential tool for a range of applications, including autonomous driving, where it aids in the recognition and avoidance of obstacles, pedestrians, and traffic signs, contributing to enhanced safety and control. It is also invaluable in surveillance systems, providing real-time object tracking and security enhancements, and it has enabled augmented reality applications to offer seamless and responsive experiences for users[7].

Over the years, YOLO has evolved through various versions, including YOLOv1, YOLOv2 (YOLO9000), YOLOv3, and YOLOv4, with each version refining and enhancing the algorithm's performance and capabilities. These iterations have broadened the scope of applications, from retail, where YOLO is applied to inventory management and customer behavior monitoring, to agriculture, where it is utilized for crop monitoring and pest detection. Moreover, YOLO has made a significant impact in the field of medical imaging, aiding in the identification and localization of anomalies in radiological and diagnostic images, leading to more accurate and timely diagnoses.

In this report, we delve deep into the inner workings of the YOLO algorithm, offering a comprehensive understanding of how it operates and its architectural components. We also explore its extensive array of applications across multiple industries and the transformative influence it has had on the field of computer vision. As we continue to embrace the digital age, YOLO stands as a testament to the power of innovation and its ability to reshape our interactions with visual data in a fast-paced, dynamic world.

Given below is the example of car detections using YOLO on highway with bounding boxes.



2.1 Object Detection

There are two factors to bear in mind while developing an object detector: First and foremost, maintain a high degree of accuracy on a constant basis, since if this area is not given the attention it needs, the ultimate result might be disastrous. Second, it is necessary to be able to swiftly recognize the things in the photos. There are two types of deep learning object identification algorithms: multi stage detectors and single shot detectors. Before attempting to establish the values of bounding boxes, the multi-stage detectors recognize region suggestions.

Single shot detectors execute both jobs simultaneously, yielding quicker results. Both types of algorithms have limitations, which are being addressed by the researchers. In essence, the multistage algorithms are being evaluated to see if they can provide quicker results, whilst the single shot detector methods are evaluated to see if they can create greater levels of accuracy. In this article, we employed bounding boxes to train the YOLO algorithm to recognize objects on photos using our proprietary classes. Thus, we showed how the YOLO might be valuable for autonomous driving on the road since it can offer a very precise result. In the actual world, many YOLO models have been deployed.

2.2 Literature Survey

Rodrigo Verschae and Javier Ruiz-del-Solar obtained an overview of past research on object detection, outline the current main research directions, and discuss open problems possible future directions. Baohua Qiang et al., proposed an object detection algorithm by jointing semantic segmentation (SSOD) for images. Author constructed a feature extraction network that integrates the hourglass structure network with the attention mechanism layer to extract and fuse multi-scale features to generate high level features with rich semantic information. Second, the semantic segmentation task is used as an auxiliary task to allow the algorithm to perform multi-task learning[8].

Finally, multi-scale features are used to predict the location and category of the object. algorithm substantially enhances object detection performance and consistently outperforms other three comparison algorithms, and the detection speed can reach real-time, which can be used for real-time detection. Zhong-Qiu Zhao et al., provided a review on deeplearning based object detection frameworks with Convolutional Neural Network(CNN) focused on typical generic object detection architectures along with some modifications and useful tricks to improve detection performance.

As distinct specific detection tasks exhibit different characteristics, survey Experimental analyses several specific tasks, including salient object detection, face detection and pedestrian detection. Christian Szegedy Alexander and Toshev Dumitru Erhan addressed the problem of object detection using DNNs, Deep Neural Networks classified and also precisely localizing objects of various classes.

2.3 YOLO Upgradations

- **YOLOv2**

YOLOv2 was intended to address two major concerns with YOLO: the identification of tiny objects in groups and the precision of localization.By implementing batch normalization, YOLOv2 boosts the network's mean Average Precision.

Anchor boxes, as described by YOLOv2, were a significantly more substantial contribution to the YOLO algorithm. YOLO forecasts just one item per grid cell, as we all know. While this simplifies the created model, it introduces complications when a single cell includes several objects, because YOLO can only give a single class to a cell.

This constraint is solved by YOLOv2, which allows several bounding boxes to be predicted from a single cell. To do this, the network predicts 5 bounding boxes for each cell[[9]. .

- **YOLO9000**

YOLO9000 was described as a strategy to identify more classes than COCO as an object detection dataset might have made possible using a similar network design to YOLOv2.In comparison to ImageNet, which has 22.000 classes, the object identification dataset (COCO) on which these models were trained has just 80.

By merging labels from ImageNet and COCO, YOLO9000 finds many more classes, combining classification and detection tasks into a single detection operation. Because certain COCO classes are supersets of ImageNet classes, YOLO9000 employs a WordNet-inspired hierarchical classification-based method in which classes and their subclasses are represented as trees.

- **YOLOV3**

While YOLOv2 is a lightning-fast network, alternative choices with better accuracies have arisen, such as Single Shot Detectors. Despite being substantially slower, they exceed YOLOv2 and YOLO9000 in terms of accuracy. YOLOv3 was presented as an improvement to YOLO by using newer CNNs that employ residual networks and skip connections.

While the DarkNet-19 is used as the model architecture in YOLOv2, the DarkNet-53, a 106-layer neural network with residual blocks and up sampling networks, is used as the model backbone in YOLOv3. Because of YOLOv3's architectural innovation, it can forecast at three distinct scales, with feature maps collected for these predictions at layers 82, 94, and 106.

YOLOv3 compensates for the shortcomings of YOLOv2 and YOLO by detecting features at three different scales. This helps it detect smaller objects. The fine-grained features that have been extracted are preserved thanks to the architecture that allows the concatenation of the up sampled layer outputs with the features from previous layers. This makes detecting smaller objects easier.

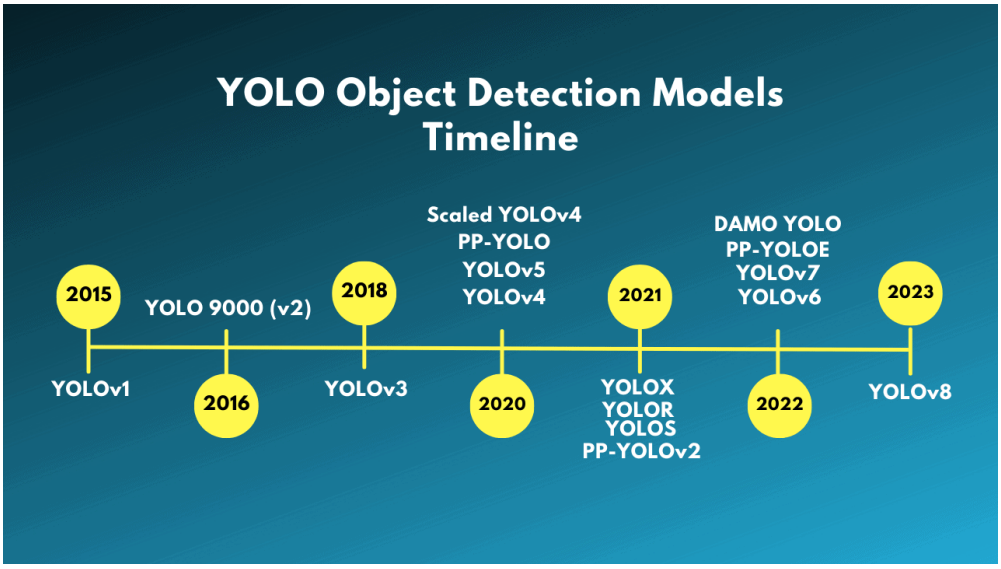
YOLOv3 only predicts three bounding boxes per cell (compared to five in YOLOv2), but it does so at three different scales, resulting in a total of nine anchor boxes.

- **YOLOV4**

YOLOv4 presents Weighted Residual Connections, Cross Mini Batch Normalization, Cross Stage Partial Connections, Self-Adversarial Training, and Mish Activation as methodological modifications to the contemporary techniques of regularisation and data augmentation. The authors also include a YOLOv4 Tiny version, which has quicker object identification and a higher frame rate but lower prediction accuracy.

- **YOLOV5**

YOLOv5 is an open-source project that provides a series of object identification models and algorithms based on the pre-trained YOLO model using the COCO dataset. It is maintained by Ultralytics and reflects the company's open-source research on the future of computer vision.



2.3 YOLO Models

2.4 Designing of YOLO

The YOLO (You Only Look Once) architecture is a convolutional neural network designed for real-time object detection. It processes an entire image in a single pass through the network and outputs bounding boxes and class predictions for the objects it detects. Here is an overview of the YOLO architecture:

1. Input Image:

The YOLO architecture takes an input image of fixed dimensions (e.g., 416x416 pixels), which is divided into a grid.

2. Backbone Network:

YOLOv4 and YOLOv5 typically use CSPDarknet53 as the backbone network, which extracts features from the input image. This backbone consists of convolutional layers and other operations to build feature maps of different resolutions.

3. Neck:

The neck of the network is responsible for merging and fusing features from multiple layers with different spatial resolutions. This helps the network handle objects of varying sizes.YOLOv4 introduced the PANet (Path Aggregation Network) to improve feature fusion.

4. Detection Head:

The detecting head is an important component of the YOLO design, since it predicts bounding boxes and class probabilities. It is made up of convolutional layers and prediction layers.The detection head forecasts object bounding boxes, which comprise coordinates (x, y, width, and height) relative to grid cells, objectness scores (showing the presence of an object within the box), and class probabilities for each specified category.Anchor boxes are used by YOLO to assist anticipate bounding box locations and size. Because YOLO predicts numerous bounding boxes for each grid cell, it may identify many objects in close proximity[10].

5. Grid Cells:

The image is divided into a grid of cells, and each cell is responsible for predicting objects within its boundaries.

6. Non-Maximum Suppression (NMS):

After the detection head's predictions, post-processing steps are applied, including non-maximum suppression, to remove redundant or overlapping bounding boxes.

The YOLO architecture's key innovation is the ability to perform object detection in a single forward pass through the network, offering real-time processing for applications like autonomous vehicles, surveillance systems, and more. It simplifies object detection by simultaneously handling localization and classification tasks.

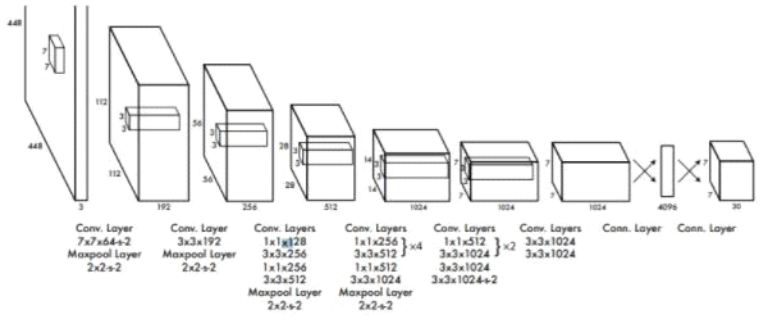


Figure 2.4(a) YOLO Architecture

YOLO has total 24 convolutional layers with 2 fully connected layers at the end.

What is the significance of the YOLO algorithm?

The YOLO algorithm is significant for the following reasons:

Speed: Because it can predict objects in real-time, this algorithm improves detection speed.

High precision: YOLO is a predictive technique that yields precise results with minimal background noise.

Learning abilities: The algorithm has exceptional learning abilities, allowing it to learn object representations and apply them to object detection.

2.5 YOLO Functionality

YOLO algorithm works using the following three techniques:

- Residual blocks
- Bounding box regression
- Intersection Over Union (IOU)

Residual blocks

First, the image is divided into various grids. Each grid has a dimension of S x S. The following image shows how an input image is divided into grids.

Figure 2.5(a) Residual Block

In the image above, there are many grid cells of equal dimension. Every grid cell will detect objects that appear within them. For example, if an object center appears within a certain grid cell, then this cell will be responsible for detecting it.

Bounding box regression

A bounding box is an outline that highlights an object in an image.

Every bounding box in the image consists of the following attributes:

- Width (bw)
- Height (bh)

Class (for example, person, car, traffic light, etc.)- This is represented by the letter c.

Bounding box center (bx,by)

The following image shows an example of a bounding box. The bounding box has been represented by a yellow outline.

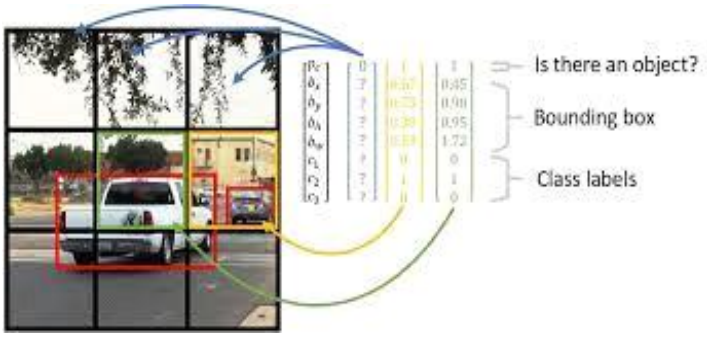


Figure 2.5(b) Bounding Box

YOLO uses a single bounding box regression to predict the height, width, center, and class of objects. In the image above, represents the probability of an object appearing in the bounding box.

Intersection over union (IOU)

Box overlapping is described by the object detection phenomena known as intersection over union (IOU). IOU is used by YOLO to create an output box that properly encircles the items.

Each grid cell forecasts the bounding boxes and their confidence ratings. The IOU is 1 if the expected and actual bounding boxes match. Bounding boxes that are less or larger than the real box are eliminated using this approach. The IOU is demonstrated in the graphic below. Formally, we require the following in order to apply Intersection over Union to assess a (arbitrary) object detector:

- The ground-truth bounding boxes (i.e., the hand labelled bounding boxes from the testing set that specify where in the image our object is).

- The predicted bounding boxes from our model.

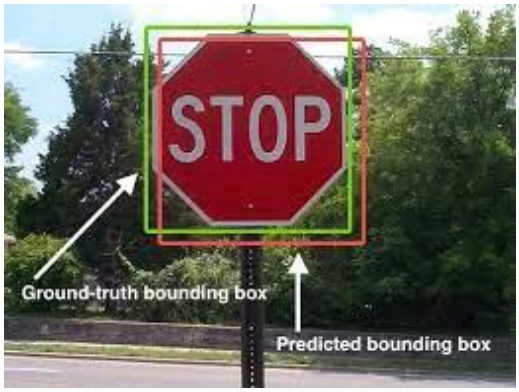


Figure 2.5(c) IOU Method

In the image above, there are two bounding boxes, one in green and the other one in red. The red box is the predicted box while the green box is the real box. YOLO ensures that the two bounding boxes are equal.

The following image shows how the three techniques are applied to produce the final detection results.

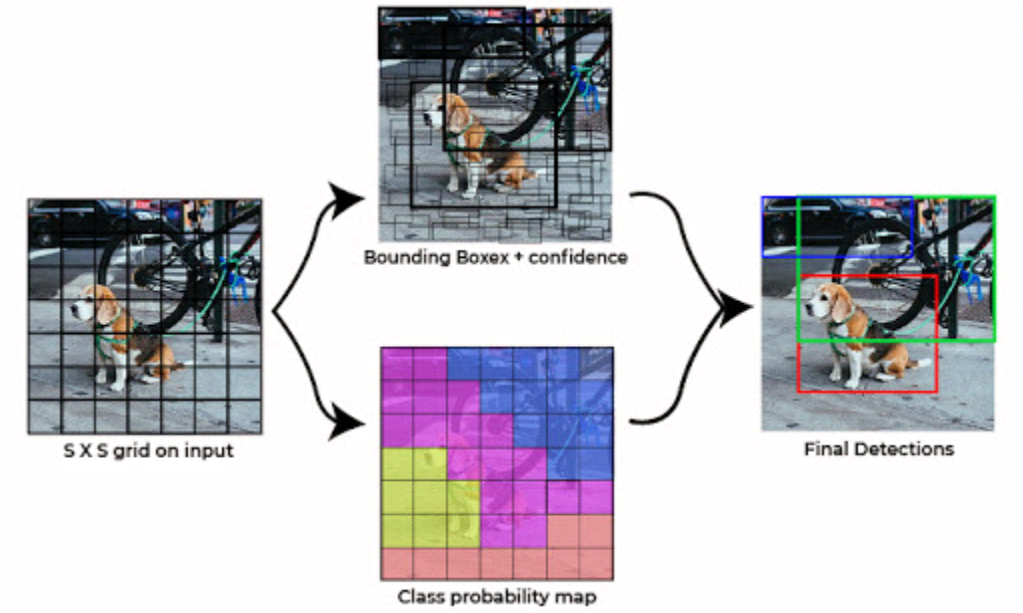


Figure 2.5(d) Three techniques of YOLO

The image is first subdivided into grid cells. B bounding boxes are forecasted in each grid cell, along with their confidence scores. To determine the class of each object, the cells predict the class probabilities.

We can see at least three types of objects, for example: a car, a dog, and a bicycle. A single convolutional neural network is used to make all of the predictions at the same time.

When intersection over union is utilized, the predicted bounding boxes are equal to the actual boxes of the objects. This phenomenon eliminates any extra bounding boxes that don't match the dimensions of the objects (such as height and breadth).The final detection will be made up of special bounding boxes that precisely suit the objects. For instance, the automobile is surrounded by the pink bounding box, whereas the bicycle is surrounded by the yellow bounding box. The dog has been highlighted using the blue bounding box[12].

The models used for object detection today are split into two parts: the head of the detection model, which determines the coordinates of the bounding boxes and the class of the item in contention, and the backbone of the model, which is trained on the IMAGENET dataset. The probable backbones in competition if the detector is operated on the GPU platform are RESNET, DENSENET, RESNEXT, or the VGG NET. The probable backbones in conflict if the detector is executed on the CPU platform are RESNET, DENSENET, RESNEXT, or the VGG NET. The backbones of Mobile Net, Squeeze Net, and Shuffle Net will compete, however, if the detector is operating on a CPU platform.

The way a YOLO algorithm works is that the image being considered is divided into grid cells, and the algorithm only goes through the image once, allowing the objects present in the image to be detected using bounding boxes circumventing its perimeter. The following image depicts a model representation of the image. We used eight different objects to train our model in our

project, so it will be a 13-dimensional one-hot encoded vector. The first component represents the object's existence, followed by the bounding box and finally the object classes.

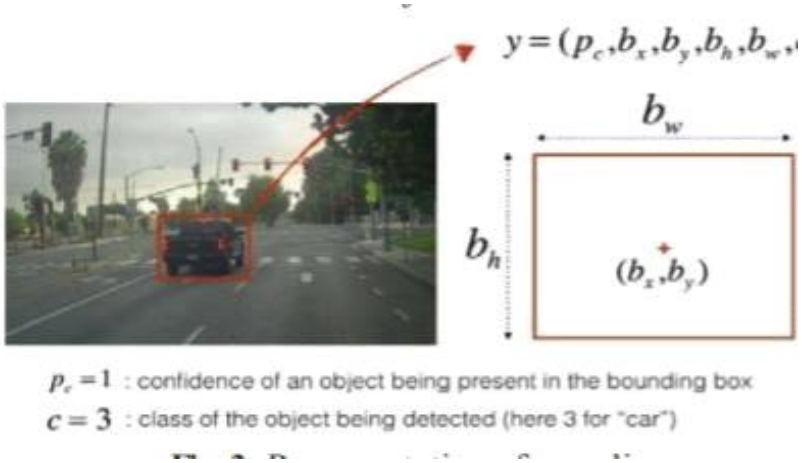


Figure 2.5(e) coordinates of object

Overlapping of boxes results from the formation of several anchor boxes. As a result, we are unable to identify the automobile or object inside. Therefore, we do non max suppression. In this case, we choose the box with the highest likelihood by using a threshold function to remove the picture. The non max suppression is the name of this phenomenon. The picture below demonstrates a non-max suppression that prevents repeated instances of image overlapping.

Figure 2.5(e) Non Max suppression

2.6 YOLO Implementation

The YOLO (You Only Look Once) algorithm has a wide range of applications across various domains, primarily in real-time object detection. Here are some of the key applications of YOLO:

1. **Autonomous Vehicles**: YOLO is used in self-driving cars for real-time detection of objects like other vehicles, pedestrians, traffic signs, and obstacles. It's essential for navigation and collision avoidance.

2. **Surveillance and Security:** YOLO is employed in surveillance systems to detect and track people, vehicles, and suspicious activities in real time, enhancing security and situational awareness.

3. **Face Detection and Recognition:** YOLO is employed for real-time face detection and recognition, used in security systems, access control, and social media tagging.

4. **Traffic Management**: In smart traffic management systems, YOLO helps monitor and manage traffic flow, detect accidents, and ensure road safety.

5. **Environmental Monitoring**: YOLO is applied to detect and track wildlife in environmental conservation efforts, such as tracking animal populations and preventing poaching.

6. **Industrial Automation:** YOLO is used for quality control in manufacturing, including defect detection and ensuring product quality on production lines.

7. **Robotics:** YOLO assists robots in perceiving and interacting with their environments, making them more adaptive and safer for collaborative tasks[11].

2.6 Application of YOLO

These applications demonstrate the versatility and impact of the YOLO algorithm in enabling real-time object detection across various industries and technologies, enhancing safety, efficiency, and automation.

CONCLUSION

Machine learning in object detection is an important technique in dealing with the occlusion , positioning ,scale transformation and lighting. The machine learning method has shown the impressive performance on the various vision tasks such as image classification, object detection and object classification. Particularly, the machine learning technique improves the performance based on image classification that discriminate the sub level features. The object detection system recognises the presence or the absence of the objects in certain scenes and the cameras viewpoints. The various domains of the object detection based on the different objectives and classified on specific and conceptual categories

The YOLO (You Only Look Once) algorithm has proven to be a groundbreaking and highly effective approach to object detection in computer vision. Throughout the course of this seminar report, we have explored the fundamental principles, architecture, and practical applications of YOLO, shedding light on its numerous advantages and some inherent challenges.

One of the key strengths of YOLO lies in its real-time object detection capabilities, which have revolutionized a wide range of industries, from autonomous driving and surveillance to healthcare and agriculture. By allowing us to detect and localize objects within images and video streams with remarkable speed and accuracy, YOLO has opened up new possibilities for automation, safety, and efficiency in various domains.

We have also discussed the trade-offs associated with YOLO, including the challenge of detecting small objects, which can sometimes be a limitation in certain applications. Nonetheless, ongoing research and development in the field of object detection continue to address these challenges, making YOLO an evolving and promising technology.

As we move forward, it is clear that YOLO and similar algorithms will play an increasingly vital role in our rapidly advancing digital world. Their impact on autonomous systems, augmented reality, and various AI-driven applications cannot be overstated. Moreover, the YOLO algorithm serves as a testament to the power of deep learning and convolutional neural networks in solving complex computer vision tasks.

In conclusion, YOLO is a significant milestone in the field of computer vision, and its impact on our daily lives and industries is undeniable. It is essential for researchers, engineers, and developers to continue exploring the frontiers of object detection, seeking innovative solutions to the challenges that lie ahead, and harnessing the full potential of YOLO for the betterment of society.

This seminar report has provided a glimpse into the exciting world of YOLO, and it is my hope that it inspires further research and applications in this field.

REFERENCES

1. C.Stauffer and W.Grimson.Adaptive background mixture models for real-time tracking. Proceedings of IEEE Conf.Computer Vision and pattern Recognition,vol.2,1999, 246-252.

2. R.T Collins A System for video surveillance and monitoring: VSAM final report.CMU-TR-0012, Technical Report, Carnegie Mellon University, 2000.

3. A.J. Lipton, H. Fuijiyoshi, R.S Patil. Moving target classification and tracking from real-time video. Proceedings of the IEEE Workshop on Application of Computer Vision, 1998, 8-14.

4. Y. Kuno, T. Wantance , Y. shimosakoda , S. Nakagawa. Automated detection of human for visual surveillance system. Proceedings of the International Conference on Pattern Recognition, 1996. 865-869.

5. R. Culter, L.S Davis. Robust real-time periodic motion detection, analysis, and application. IEEE Trans. Pattern Anal.Mech. Intell. 22 (8) (2000) 781-796.

6. Local application of optic flow to analysis rigid versus non-rigid motion. http://www.eecs.lehigh.edu/FRAME/ Lipton/iccvframe.html.

7. A.Selinger, L. Wixson. Classifying moving objects as rigid or non-rigid without correspondences. Proceedings of the DAPRA Image Understanding Workshop, Vol. 1, 1998, 341-358.

8. M. Oren, et al. Pedestrian detection using wavelet templates. Proceedings of the IEEE CS Conference on Computer Vision and Pattern Recognition. 1997,193-199.

9. Rodrigo Verschae, Javier Ruiz-del-Solar, "Object Detection: Current and Future Directions Perspective", Article in Frontiers in Robotics and AI , December 2015.

10. Baohua Qiang, Ruidong Chen, Mingliang Zhou, Yuanchao Pang, Yijie Zhai,Minghao Yang,"Convolutional Neural Networks-Based Object Detection Algorithm by Jointing Semantic", Segmentation for Images, Sensors 2020.

11. Kamate, S., &Yilmazer, N. (2015). Application of object detection and tracking techniques for unmanned aerial vehicles. Procedia Computer Science, vol61,no.3, pp. 436-441.

12. Kurian, M. Z., & MV, C. M. (2011). Various Object Recognition Techniques for Computer Vision. Journal of Analysis and Computation, vol7, no 1,pp. 39-47.

13. https://www.mygreatlearning.com/blog/yolo-object-detection-using-opencv/

14. https://shorturl.at/kAHJM

15. https://t.ly/JUf5w

16. https://rb.gy/0c5au

17. https://rb.gy/i140j

18. https://rb.gy/mvonb

19. https://rb.gy/0xhmx
20. https://rb.gy/0kbji