REPUBLIC OF TÜRKİYE

ALTINBAŞ UNIVERSITY

Institute of Graduate Studies

Information Technologies

PREDICTING DANGER USING ARTIFICIAL INTELLIGENCE

Mustafa Ameen Sultan AL-TAMEEMI

Master's Thesis

Supervisor

Prof. Osman Nuri UÇAN

Istanbul, 2022

PREDICTING DANGER USING ARTIFICIAL INTELLIGENCE

Mustafa Ameen Sultan Al-TAMEEMI

Information Technologies

Master's Thesis

ALTINBAŞ UNIVERSITY

2022

The thesis titled PREDICTING DANGER USING ARTIFICIAL INTELLIGENCE prepared by MUSTAFA AMEEN SULTAN AL-TAMEEMI and submitted on 20/12/2022 has been **accepted unanimously** for the degree of Master of Science in Information Technologies.

Prof. Dr. Osman Nurı UÇAN

the Supervisor

Thesis Defense Committee Members:

Prof. Dr. Osman Nurı UÇAN	Department of Electrical and Electronics Engineering, Altınbaş University	_____
Asst. Prof. Dr. Abdullahi Abdu IBRAHIM	Department of Computer Engineering, Altınbaş University	_____
Prof. Dr. Hasan Hüesyin BALIK	Department of Computer Engineering, Yıldız Teknik University	_____

I hereby declare that this thesis meets all format and submission requirements of a Master's thesis.
Submission data of the thesis to Institute of Graduate Studies: ____/____/____

I hereby declare that all information/data presented in this graduation project has been obtained in full accordance with academic rules and ethical conduct. I also declare all unoriginal materials and conclusions have been cited in the text and all references mentioned in the Reference List have been cited in the text, and vice versa as required by the abovementioned rules and conduct.

Mustafa Ameen AL-TAMEEMI

Signature

DEDICATION

My God, the night is not good without your thanks, and the day is not good without your obedience...

Moments are not good except by remembrance of you...and there is no sweetness in the hereafter except by your forgiveness...and heaven is not sweet except by seeing you.

We thank God Almighty who enabled me to complete scientific research in good health

To the first teacher and leader of mankind, our Prophet Muhammad, may God bless him and grant him peace.

To the saviour of mankind, the waiter of the nation, and the supporter of truth in the dark, my master, the awaited imam

To whom was her prayer the secret of my success..... my beloved mother

To my dear sister, thank you

To my brother and Cindy, thank you

To whom I proudly bear his name... my dear father

To the one who helped me....Dr. Haider

To the one who gave me guidance and advice....my supervisor, Dr. Osman Nuri UCAN

To everyone who accompanied me with his prayers in my scientific careers my love

Thank you to Dr. Yasser, thank you to Dr. Muhammad, thank you to Dr. Wafa, thank you to Dr. Abdul Karim. Thank you to Dr Aqeel, Thank you to Dr Taha.

Thank you to the government and people of Turkey for everything they have given me personally. I wish prosperity and development for this beautiful country.

As they told us previously, its conclusion was caught while I was writing the last letters of this thesis. I am very proud of that great woman who accompanied me all the way and was a support for me and her supplication surrounded me from everywhere. The one that I describe as the tent peg, thanks to her, supported me as the tent supports after the awaited imam. I dedicate this thesis to her.

ABSTRACT

PREDICTING DANGER USING ARTIFICIAL INTELLIGENCE

AL-Tameemi, Mustafa Ameen

M.Sc., Information Technologies, Altınbaş University,

Supervisor: Prof. Dr. Osman Nuri Uçan

Date: 12/2022

Pages: (70)

This thesis discusses two types of dangerous: fire and criminals that threaten local areas in urban cities which could affect negatively on people life. The fire dangerous, usually, can be identified through surveillance cameras installed in city streets. However, identifying wanted criminals or dangerous persons can be detected through revealing faces. In this work, One Time Look (YOLO) v2 algorithm is invoked to develop a system of detecting warnings of such types of danger. Therefore, the firefighting and the criminal detection in real time through street surveillance cameras using computer vision-based AI would be applied using object detection and using YOLO algorithm. Matlab coding is used to implement the practical part of this research. Therefore, it is a subject of this thesis to conduct the use of surveillance cameras and computer vision to predict fire accident and recognize faces of wanted person in real time to alarm the nearest authority departments. This is a very challenging issue in many applications due to the limitations of traditional surveillance cameras use to recording accidents. However, in this work, the author applied the use such cameras real time predictions and detections. This is attempted by using the process of AI through applying YOLO algorithm based on training data for fire and face images in different scenarios. Therefore, the author conducted about 412 images and 345 images for the fire and faces, respectively, from different orientation and rotation axes as a trading data for the proposed algorithm. From this study, the obtained results show an excellent prediction response with fast detection process about 3 seconds with an excellent accuracy of treatment The author thought this work could be used in smart cities and high-security buildings.

Keywords: AI, Object Detections, Face Recognition, Computer Vision.

TABLE OF CONTENTS

Pages

ABSTRACT ..... **vi**
LIST OF TABLES ..... **ix**
LIST OF FIGURES ..... **x**
ABBREVIATIONS ..... **xii**
LIST OF SYMBOLS ..... **xiv**
1. INTRODUCTION **1**
1.1 INTRODUCTION 1
1.2 PROBLEM STATEMENT 4
2. LITERATURE SURVEY **5**
2.1 INTRODECTION 5
2.2 COMPUTER VISION 7
2.2.1 Computer Vision Work 8
2.2.2 Computer Vision's Historical Development 9
2.2.3 Computer Vision Software 10
2.2.4 Examples of Computer Vision 11
2.3 OBJECT DETECTION 12
2.3.1 Methods of Detecting Objects and their types 13
2.3.2 Object Detection Work 14
2.3.3 Applications and Use Cases of Object Detection 16
3. METHODOLOGY **18**
3.1 GENERAL INTRODUCTION 18
3.2 YOLO ALGORITHM 18
3.2.1 YOLO v1 19
3.2.2 YOLO v2 20
3.2.3 Comparison Between VOLO v1 and VOLO v2 20
3.2.4 YOLO and SSD Comparison 22
3.2.5 Versions and History of Yolo 23
3.2.6 How Does Yolo Work 24
3.2.7 Real-Time Capability 25

3.2.8 Real-time Object Detection Issues 25
3.2.9 Practical Applications of YOLO 27
3.3 MATALB 31
3.3.1 Matlab Features 31
4. RESULTS 32
4.1 CHAPTER INTRODUCTION 32
4.2 SYSTEM HARDWARE 32
4.3 SYSTEM SOFTWARE 32
4.4 HISTOGRAM 33
4.5 OUR RESUULTS 34
5. DISCUSSION AND CONCLUSIONS 56
5.1 CONCLUSION 56
5.2 FUTURE WORK 56
REFERENCES 58

LIST OF TABLES

Pages

TABLO 3.1: CHANGES BETWEEN YOLO AND YOLOV2 [32] 21

TABLO 4.1: NUMBER OF TRAINED DATA 36

TABLO 4.2: SUCCESS RATE FOR EACH ALGORITHM 55

LIST OF FIGURES

Pages

Figure 1.1: shows the annual motor vehicle fires on highways [29] 3

Figure 1.2: shows the number of deaths annually due to vehicle fire accidents on the highway [29] 3

Figure 2.1: Comparing human vision with computer vision 8

Figure 2.2: Computer vision technology to recognize objects. 13

Figure 2.3: A model that defines the detection of a single object 14

Figure 2.4: A more accurate model can suggest any number of locations that can include a bounding square [45]. 15

Figure 2.5: Illustrative diagram of Single Shot Detectors (SSDs) [19]. 16

Figure 3.1: Accuracy and Speed of Different Object Detection Models on VOC2007[32]. ……..22

Figure 3.2: A timeline showing the evolution of YOLO in recent years . ……..…24

Figure 3.3: Illustration of splitting an image into grid cells . …….……..…24

Figure 3.4: Architecture for the YOLO object detection service based on RSTP,[24]……..…25

Figure 3.5: Real-time processing issue with YOLO. ,[24] . …….……..…27

Figure 3.6: 2D kidney detection by YOLOv3. A-B: Normal kidneys with different CT scan [44] .. ……..…….……..…28

Figure 3.7: The mechanism of determining the maturity of agricultural crops using yolo [24] . …29

Figure 4.1: Selecting the part to be recognized (a,b) . ……..……..…35

Figure 4.2: shows the contents of the exported frame table 36

Figure 4.3: shows the fire detection mechanism..... 37

Figure 4.4: shows the number of successful and failed trials and the success rate of the algorithm. 37

Figure 4.5: Shows the failure and success experience while running the algorithm.. 38

Figure 4.6: shows the difference between a success and failure event according to the standard graph of a fire recognition event..... 40

Figure 4.7: shows the success and failure rate of Barack Obama's facial recognition experiment.. 41

Figure 4.8: shows the result of the experiment process, discovering Barack Obama’s face..... 42

Figure 4.9: shows the accuracy ratios for each pass-and-fail according to histogram. 45

Figure 4.10: shows the graph of success and failure rates for Narendra Modi's face recognition ..... .45

Figure 4.11: shows the results of an experiment with Narendra Modi..... 46

Figure 4.12: shows the accuracy rate of pass/fail Narendra Modi facial recognition according to the benchmark graph..... 49

Figure 4.13: shows the graph of success and failure rates for Sundar Pichai's face recognition.. 50

Figure 4.14: shows the results of an experiment with Sundar Pichai's 51

Figure 4.15: shows the accuracy rate of pass/fail Sundar Pichai facial recognition according to the benchmark graph..... 54

ABBREVIATIONS

AI	:	Artificial Intelligence
ML	:	Machine Learning
DL	:	Deep Learning
SSDs	:	Single Shot Selectors
YOLO	:	You Only Look Once
CNN	:	Convolutional Neural Network

LIST OF SYMBOLS

Dnq : The amount of time the nth frame's queue is waiting.

Ta : The interarrival time for the input frame.

Dns : The service time required in object detection of YOLO.

1. INTRODUCTION

1.1 INTRODUCTION

The possibility of any dangerous accident that threatens public life could occur at any time, especially with the recent population and traffic congestion increase [1]. The most common problems that threaten the public safety and life stability are fires and outlawing criminals [2]. These two types of danger are very important concepts in building infrastructures for modern smart cities to disturb the quality of public services and people life [3]. According to 2018 statistics, firefighting teams in the United States of America responded to about 212,500 car fires on public streets. According to these statistics, such accidents about 560 civilians were killed with 1500 injure. Also, massive losses in constructions with a cost of 1.9 billion dollars and direct material damage of properties were occurred [1]. However, among only 16% of the reported fire accidence can be treated by the auto-firefighting alarm systems mong 1.3 million times as announced by the US General Fire Department. Fire-cars, also, caused 15% of the total fire deaths [1]. Vehicle fires cause about 4.5% as many deaths from non-residential building fires, and 1.6% as many as apartment fires. Figure 1.1 displays annual highway vehicle fires, while, Figure 1.2 displays annual highway vehicle fire deaths.

Nowadays, video surveillance systems are being created and used practically everywhere with the aim of recording, supervising and evaluating damages. During criminal investigations, surveillance and security cameras are important evidence that helps identifying the situation of crimes. In fact, police officers and forensic professionals should have little problem using CCTV images to identify people at crime scenes and to match assembled facial images with gallery images of criminals [2]; but, this is not enough. Such system must provide a quick and effective identification of criminals by specifying locations and alerting the relevant authorities [4]. Over the past ten years, many researchers applied face recognition to enhance such systems efficacy and performance. Early face recognition research were discussed in [3, 4] to recognize individuals through captured images from surrounded areas to the accidents.

Rapid changes in human living have been brought about by the growth of "information technology" (IT). The usage of computers and mobile devices has increased due to recent advancements in electrical and communication technologies. One of the branches of this

technology is "artificial intelligence" (AI), which is a major branch through which many areas can be developed where AI can be used in many sectors such as transportation, finance, healthcare and banking, and it makes it possible for current computers to think extensively about processing video footage. It is automatic in real time and can also be used in image processing and helps us to carry out "object detection", detection and recognition through image and video input. In several business domains, computer vision is often used. The end-uses of "object detection", a crucial subfield of vision missions, are anticipated to include unmanned transportation, pedestrian detection, and video monitoring, etc. [1]. Companies, nowadays, chose ways to handle issues that were formerly dealt with traditional methods, such monitoring by a human operator or the use of additional mechanical sensors. No matter the application, motion detection and, if practical, moving are always the initial tasks of a video system analysis (segmentation). The circumstances for collection, the level of accuracy and the anticipated processing time are all affect how challenging of this operation. One of the most active fields in computer vision in recent years is "object recognition", and many academics are competing to develop the best "object detection" model [2]. As a consequence, several cutting-edge models are being created such as" RetinaNe, RCNN ,OpenCv, and YOLO".

In this research, the use of AI is invoked to predict danger; that takes two aspects to avoid danger in cities. The possibility of danger in life is expected to occur at any moment, so finding smart and innovative ways using modern technology to discover these dangers and control them quickly is a significant need. To avoid and treat it immediately, it will address two types of problems that threaten cities and frequent occurrences, which are fires and the detection of outlaws through the use of AI algorithms associated and programmed on cameras that can be placed in public streets, cities and external roads. Therefore, this study about the magnitude of the danger and the destruction caused by each part of the danger would be discussed in details. In 2001, only in Istanbul a fire brigade responded to 20,760 fires in 2021 [3]. As for facial recognition, there has been extensive research on the value of CCTV for crime prevention, but little in terms of its value as an investigative tool. Seeking to quantify how often CCTV is useful and how this is affected by conditions, an analysis of 251,195 British Transport Police recorded crimes that occurred on the British rail network between 2011 and 2015 [4]. The use of YOLO algorithm to monitor cities through special cameras, these cameras can detect fires early, in addition to recognizing faces and give information about people required by law if they pass through the streets of these cities at

once and we can send notifications to the relevant authorities in each part of the We will discuss later the mechanism used in this work.

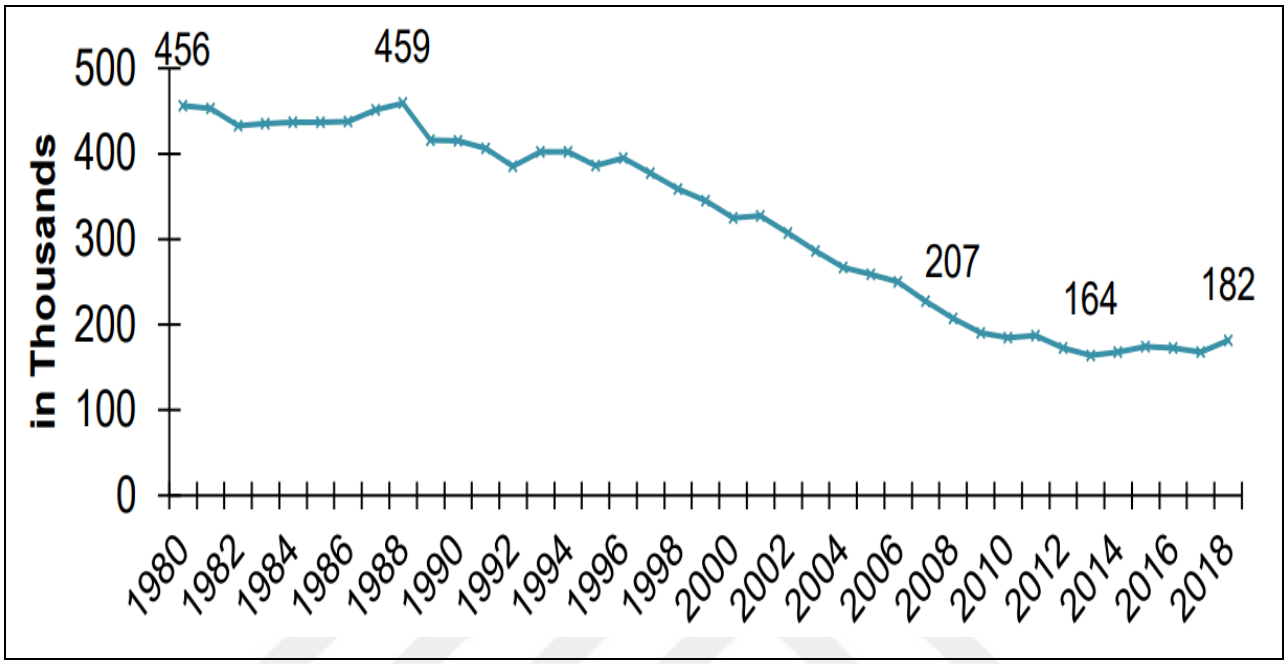


Figure 1.1 : shows the annual motor vehicle fires on highways [29].

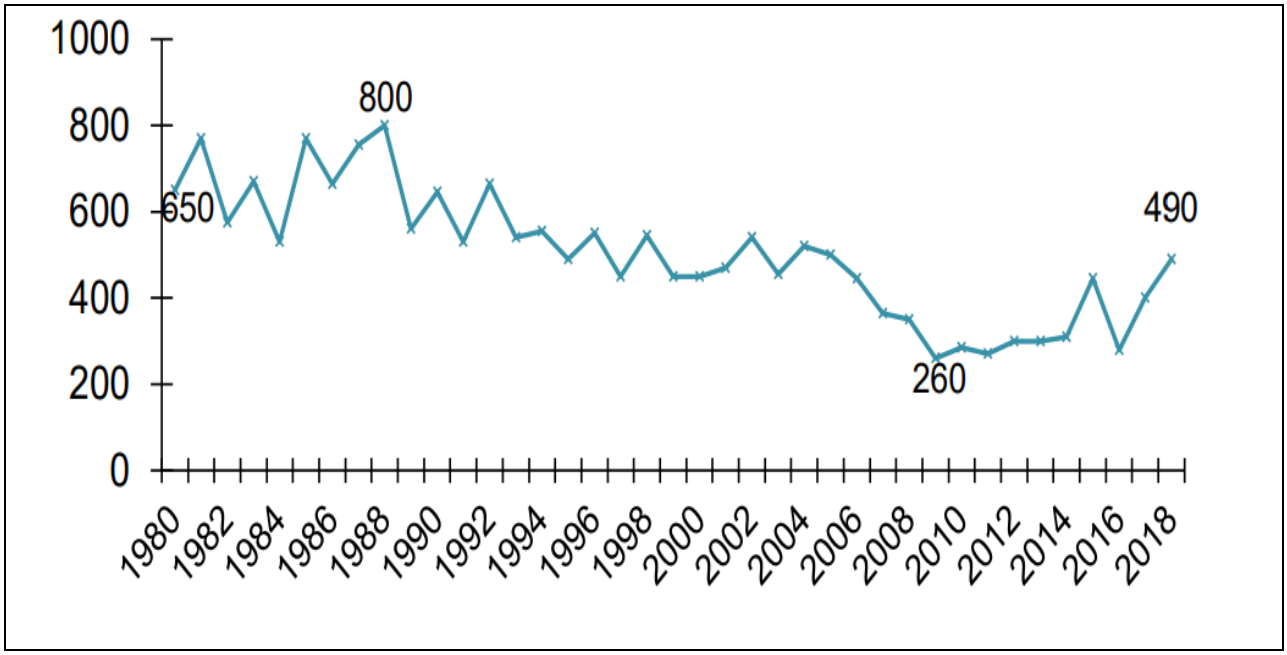


Figure 1.2: shows the number of deaths annually due to vehicle fire accidents on the highway [29].

1.2 PROBLEM STATEMENT

The ability of analyse image and video sequences to automatically recognize and detect objects is one of the most challenging problems in computer vision technology. Modern high-speed technologies with recent memory developments and powerful processing allowed modern computers to perform high object detection with low computational cost [6]. Over the past decades due to advances in the field of machine learning, specifically with the contributions of deep neural networks (DNN), many classification problems have been solved. Currently, computers can answer the question “Does this image contain this specific object?”. By comparing between the new versions with the previous detection processes, it is observed that the speed of objects detection and recognition is very slow with low detection quality. Therefore, to increase the speed of detection of a photo or video, a suitable model based YOLO algorithm was suggested in this work. The proposed algorithm shows an excellent faces recognition and fire detection at the same time with various sizes and distances in unclear places; where, previous studies suffered from.

2. LITERATURE SURVEY

2.1 INTRODUCTION

Scale-Invariant Feature Transform (SIFT) technique is a computer detection approach to locate and identify geographical areas by specifying and recognizing the local visual properties and things may be more easily recognized. SIFT key-points relies on objects localization, background, and picture size, and rotation. This algorithm is very remarkable because it operates well under vibration, illumination, and little perspective changes. Such characters are highly recognizable and very simple to reach the precise target of identification with minimum mistakes as proposed by David G. Lowe in [5] for the first time and updated later in [6].

Histogram of directed gradients (HOG) a feature descriptor for computer-view and image analysis for object recognition of a histogram feature. The basic idea of such algorithm underlying the histogram process to localize objects existence and structure inside images which can represented by a sequence of intensity gradients or limits. In such algorithm, images are divided into small pixels are interconnected together to be called cells. A histogram technique of gradient axes is created for each pixel in each cell. The concatenation of these histograms serves as the description of the image. Therefore, HOG descriptor has a few key advantages over other descriptors. Except for object orientation, it is invariant to geometric and photometric alterations, since, it operates on local cells. Guangyuan Zhang et al. in [7] enhanced HOG characteristics to guarantee a successful human identification system and SVM classification for pedestrian identification. The results of the simulation showed how well the approach procedure worked. A typical method for detecting pedestrians used HOG and SVM.

In 2010, CHENG Guang-tao et al. [8] introduced a pedestrian segmenting approach based on vertical edges and symmetrical features. This approach is to find and segment the pedestrian from video pictures with special characteristics. Such an algorithm applies the use of HOG and SVM in a hybrid technique. According to their experimental findings, this algorithm performs a batter response in difficult situations with a fast detection rate. Similar to this, in 2011 YAO Xue-qin et al [9] introduced a pedestrian detection approach based on edge symmetry and HOG to address the slow detection speed issue in the old method. In this study, the symmetric difference was used to extract the input window vertical edges. Using HOG and SVM candidate pedestrians are quickly identified based on vertical edges symmetry. The results showed that doing so faster detection

process, while retaining a detection rate that was equivalent to the traditional strategy based on the HOG feature.

SURF is an effective local detector of object recognition and pedestrian tracking to be called speeded up robust features with minimum SIFT effects impacts. The proposed SURF standard edition publishers claim it to be more resistant to picture alterations than SIFT with quicker response. It was initially published in 2006 by Herbert Bay et al. [10]. The results demonstrate that SURF outperforms SIFT and even other earlier techniques in terms of both speed and accuracy. This is achieved by using combined pictures for signal convolutions, which capitalizes on the advantages of the top current detectors and descriptors. According to the comparison provided by Luo Juan et al. [11], SIFT is slow and not ideal for changes in illumination but is invariant to rotation, changes in scale, and affine transformations, while SURF is quick and performs as well as SIFT but is unstable when changing rotations and lighting. An effective, reliable local characteristic detector called ORB—Oriented FAST and rotated BRIEF—were utilized for object identification applications like pedestrian detection. In order to do this, the Binary Robust Elementary Features visual descriptor and the Quick keyboard detector are both used. It seeks to provide a quick and potent substitute for SIFT or SURF. The work was presented by Ethan Rublee et al. [12] in 2011. Experimental findings show that ORB functions with great accuracy and detection rates. Li Xiaohong et al. [13] introduced a unique ORB-based technique for object detection in dynamic situations in 2012.

Neural networks with convolutions (CNN) were proposed in 2014 by Ross Girshick et al. in [14]. They introduced a kind of scalable algorithm to raise the average accuracy by more than 30% in comparison to the performance of the earlier methods. When training data is labelled with sparse, supervised pre training for a subsidiary role and domain-specific refinement leads in a significant increase in efficiency. This employs high-capacity CNNs to discover and separate items geographical suggestions. The method is known as "R-CNN": CNN areas. A year later, Ross Girshick [15] proposed a rapid region-based technique for object identification ("Fast R-CNN") that effectively identified object suggestions using deep convolutional networks. This method was built on past research. The preparation, testing speed, and identification accuracy are all improved by Quick R-CNN. Later same year, Shaoqing Ren et al. [16] introduced the Regional Proposal Network (RPN), which is practically cost-free and shares full-scale convolutionary properties with the detection system. RPNs are fully prepared to offer excellent ideas for the regions that Fast R-

CNN uses to identify regions. By alternating optimization, RPN and Fast R-CNN may be set up to share convolutive features. It has been shown that this approach boosts the area's proposal's effectiveness and speed, increasing the accuracy of target identification overall. Joseph Redmon et al. presented YOLO, or first one-stage approach, in 2016 [17]. A single neural network may be used to quickly estimate bounding boxes and class probabilities on whole images. The design and construction of this architecture were excellent. False positives were less likely to happen in the background, despite the fact that YOLO generates more positional errors than other earlier standard tracking algorithms.

The detection technique for YOLO was created by Joseph Redmon at el. [18] in 2017. A novel multi-scale training algorithm, which outperforms Faster R-CNN and SSD while running considerably faster, was used in the upgraded YOLOv2 to make it simpler to cope with speed and accuracy. They provide a mechanism to detect and coach artefacts in their final suggestion. By using this method, they can concurrently train the Yolo9000 and forecast real-time detections for more than 9,000 different types of objects.

Jose Redmon once again issued several YOLO notices in 2018 [19]. The performance of YOLO has been enhanced by a number of architectural upgrades, cutting-edge algorithms, strong detection accuracy, and an extraordinarily quick detection speed.

2.2 COMPUTER VISION

Mankind gathered a lot of information and knowledge for centuries from their circumstances to be recorded in scripts. In addition, by collecting and analyzing information, they protect themselves from the external environment and express their responses to external stimulation. Among them, visual information is accepted faster, and their proportion is quite large. This is because visual information includes almost all the information necessary to interact with the surrounding environment, such as the shape, color, texture and movement of an object. Therefore, computers accept different information more than any other human sensory organ. Similar to how human vision works, computer vision also has certain advantages over human eyesight. The benefit of having a lifetime of context to learn how to discern between things, how far away they are, whether they are moving, and whether there is an issue with the picture is that human sight. Computer vision trains computers to do these tasks considerably more quickly by using (cameras, algorithms, and data) instead of the retina, optic nerves, and visual brain. As a machine can evaluate hundreds

of goods or processes per minute while spotting minute faults or problems, it may quickly outperform people in checking things or keeping an eye on production resources. It makes it possible for computers and other systems to take action or provide suggestions based on the information they are able to extract from digital photos, videos, and other visual inputs. A subfield of artificial intelligence is computer vision (AI). Machines can now sense, observe, and grasp the world just like humans can thanks to computer vision. Computer vision is used in the industrial, energy, utilities, and automotive industries, and the field is continually growing. It is projected to reach USD 48.6 billion by 2022. Figure 2.1 shows a comparison between human vision and computer vision.

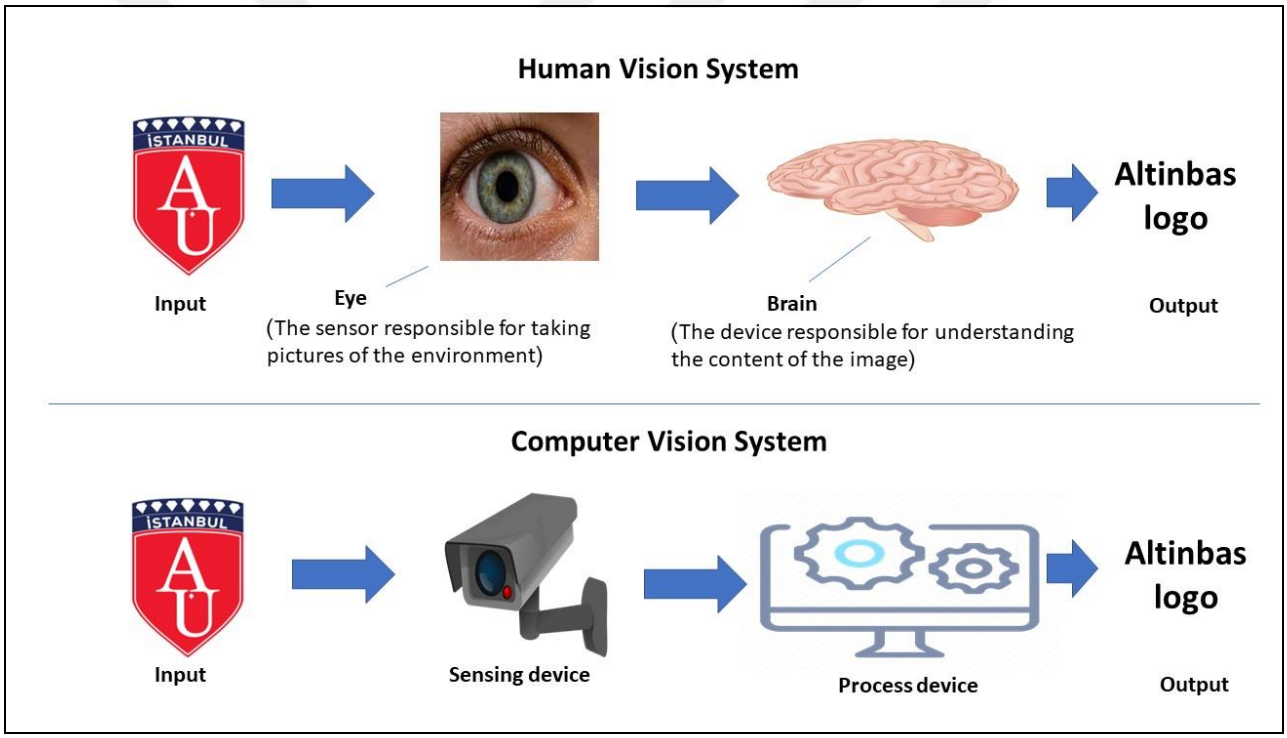


Figure 2.1: Comparing human vision with computer vision.

2.2.1 Computer Vision Work

Computer vision requires a lot of data. It runs data analysis repeatedly until it can discriminate between objects and recognize photos. To grasp the differences and recognize a tire, especially one without any flaws, a computer, for example, has to be fed a vast quantity of tire pictures and tire-related documents. Convolutional neural networks and deep learning, a kind of machine learning, are two fundamental technologies used in this (CNN). Using machine learning and algorithmic models, a computer may be trained how to recognize the context of visual data. If

enough data is fed through a certain model, computers "look" store and analyze the obtained data for training to distinguish between different pictures. Algorithms enable computers to learn on their own rather than requiring programming in order to recognize images. CNN algorithm facilitates the "seeing" capabilities of machine or deep learning models by breaking down images into pixels with labels or tags. By applying convolutions to the labels—a mathematical procedure on two functions to produce a third function—it generates predictions about what it is "seeing." The neural network performs convolutions and continuously assesses the precision of predictions up until they begin to come true. The next step is to recognize images in a manner comparable to how humans do. CNN algorithm first detects crisp shapes and basic forms before adding details as it repeatedly evaluates the predictions in a similar manner to human being sight. Recurrent neural networks (RNNs) are used in video applications in a way similar to such technology that helps computers to understand the links between the visuals in a succession of frames.

2.2.2 Computer Visions Historical Development

For almost 60 years, scientists and engineers have worked to create systems that would enable robots to see and comprehend visual information. Neurophysiologists began their first experiment in 1959 by exposing a cat to a range of images in an attempt to correlate a brain response. In terms of science, this meant that processing of images starts with simple shapes like straight edges. They noticed that it reacted initially to hard edges or lines. The first computer image scanning technology, which allowed computers to digitize and capture pictures, was created about the same time. The ability of computers to convert two-dimensional pictures into three-dimensional shapes in 1963 marked the achievement of another major milestone. The 1960s saw the emergence of AI as an area of research and the start of the AI effort to address the issue of human eyesight. Optical character recognition (OCR) technology, which could read text written in any font or typeface, was first introduced in 1974. Similar to this, intelligent character recognition (ICR) may use neural networks to understand handwritten text. Since then, (OCR and ICR) have made their way into a variety of popular applications, including the (processing of documents and invoices, the identification of license plates, mobile payments, machine translation) and more. Neuroscientist "David Marr" demonstrated that vision works hierarchically in 1982 and developed methods enabling computers to recognize edges, corners, curves, and other fundamental structures. In parallel, Kunihiko "Fukushima", a computer scientist, created a network of cells that could identify

patterns. Convolutional layers in a neural network were part of the network, which was known as the Neocognitron. The emphasis of research shifted to object identification in 2000 and the first real-time face recognition applications debuted in 2001. Through, in 2000, there was a standardization of the tagging and annotation of visual data sets. In 2010, the ImageNet data set became available. It contains millions of annotated photographs from a thousand distinct object classes and serves as the foundation for the current generation of CNNs and deep learning models. A University of Toronto team entered CNN in an image recognition competition in 2012. The AlexNet model drastically decreased the rate of error in picture recognition. Error rates have decreased to only a few percent since this discovery [42].

2.2.3 Computer Vision Software

There is much research being done on the subject of computer vision, but it is more than that. Applications in actual environments demonstrate how essential computer vision is to (activities in commerce, entertainment, transportation, healthcare, and day-to-day life). The growth of these applications is significantly influenced by the flood of visual data from smartphones, security systems, traffic cameras, and other visually instrumented devices. Although, it isn't being used right now, this information might be crucial to the operations of many different firms. The data serves as a testing ground for computer vision software and a point of entry for its incorporation into several human endeavors:

a. "IBM" used machine vision to create My Moments for the 2018 Masters golf tournament. After seeing several hours of Masters Film, IBM Watson was able to identify important shots looks and sounds. It gave viewers with customized highlight snippets of these important events after choosing them.

b. Users may use Google Translate to quickly translate signs that are written in other languages by pointing a smartphone camera at the sign.

c. In order to understand the visual data coming from a car's cameras and other sensors, computer vision is employed in the creation of self-driving automobiles. It is essential to recognize any other cars, traffic signs, lane markings, bicycles, pedestrians, and other visual objects on the road.

d. In order to deliver cutting-edge AI to the edge and assist automakers in identifying manufacturing flaws before a vehicle leaves the plant, IBM is using computer vision technologies with partners like Verizon.

2.2.4 Examples of Computer Vision

Numerous firms lack the resources needed to sustain computer vision research labs, create neural networks, and build deep learning models. Additionally, they may not have the computing power necessary to evaluate enormous visual data sets. Companies like IBM can help by offering software development services for computer vision. These services provide pre-built learning models that may be accessed through the cloud while putting less demand on computer resources. Users connect to the services using an application programming interface (API) and use them to create computer vision applications. A platform for machine vision that addresses concerns with development and computing resources has also been presented by IBM. Subject matter experts can label, train, and deploy deep learning vision models utilizing the features included into IBM Maximo Visual Inspection without any coding or deep learning expertise. The vision models may be used on edge devices, in the cloud, and local data centers. Even if acquiring the resources required to develop computer vision applications is becoming easier, it is essential to decide on their intended use before anything else. Projects and applications that are targeted, verified, and specified in terms of specific computer vision tasks make it easier to get started. Below are a few examples of well-known computer vision jobs:

Puppy, an apple, or persons face are examples of images that may be classified using image classification. More specifically, it can correctly guess which class a given picture belongs to. A social network corporation would wish to utilize it, for instance, to automatically recognize and sort out offensive photographs shared by users.

In order to identify a certain class of picture and then recognize and tabulate it is existence in an image or video, object detection may employ image classification. Examples include finding damaged equipment or spotting harm to a production line.

After an item is found, it is followed or tracked. This process is often performed using real-time video feeds or a collection of consecutively captured photos. For instance, autonomous vehicles

must monitor moving things like people, other vehicles, and road infrastructure in addition to classifying and detecting them in order to prevent crashes and follow traffic regulations. Instead of focusing on the pictures metadata tags, content-based image retrieval employs computer vision to explore, search, and retrieve images from massive data repositories. Automatic picture annotation may be used instead of human image labeling for this activity. These activities may be used to digital asset management systems to improve search and retrieval precision.

2.3 OBJECT DETECTION

Using a computer vision method called "object detection", certain objects in a photo or video can be recognized. "Object detection" may be used to count the objects in a scene, as well as, to locate and track them in real time while accurately identify and localize them. Imagine, for instance, a picture of a room containing a laptop, a cup, a bag and a chair as shown in Figure 2.2. By using "object detection", we can simultaneously categorize different items. It can be discovered and find examples of those objects in the image. Before, it is imperative that clarifies the distinctions between "object detection" and "image identification," since they may sometimes be confused.

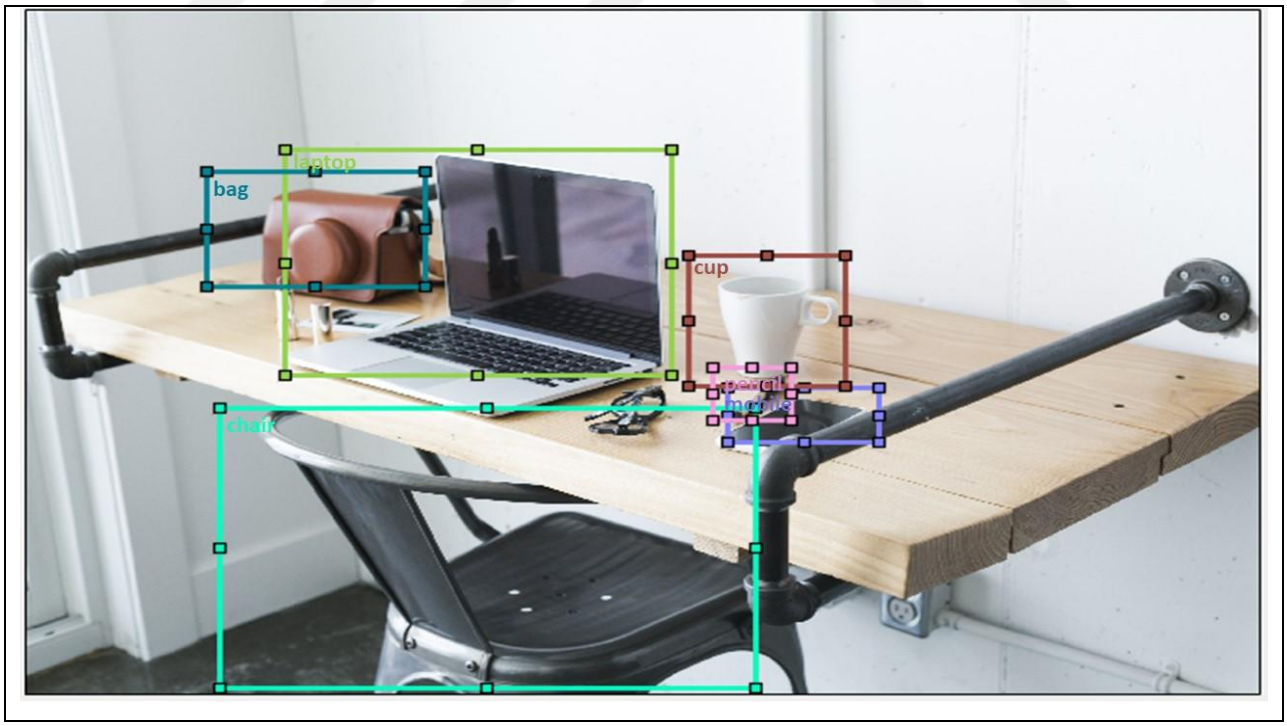


Figure 2.2: Computer vision technology to recognize objects.

2.3.1 Methods of Detecting Objects and their Types

In general, ML -based and DL – "Object recognition" methods; which are based on them may be differentiated. In more traditional ML-based techniques, clusters of pixels that might constitute an object are found by applying computer vision algorithms to examine different features of a picture, such as the color histogram or edges. These characteristics are then used as input into a regression model to predict the item location and labeling. Recognition of objects is essential. Object detection, which is integrally connected to other computer vision methods like image recognition and image segmentation, improves the comprehension and analysis of situations in photos and videos. But there are observable variations. While, picture segmentation develops a pixel-level comprehension of components scene, image recognition just generates a class name for each detected object. In contrast to these other occupations, "object detection" has the ability to locate individual items inside an image or video. We can count these items and afterwards trace them thanks to this.

2.2.1 Object Detection Work

Models for "object recognition" based on deep learning typically include two parts. When, an encoder takes an image as input, it processes via a variety of layers and blocks technology how to extract statistical traits that are needed to find and identify objects. The bounding boundaries and labels for each item are decided by a decoder using the encoder outputs. The most straightforward decoder is a pure regression. By attaching to the encoder's output, the regression is able to directly anticipate the location and size of each bounding box. The output of the model is X and Y coordinates of the item and area in the picture. Despite being simple, this paradigm has certain drawbacks. You must let us know in advance how many boxes there will be. Figure 2.3 shows this. One Cat will be left unidentified if there are two in your picture but your model can only recognize one. However, if you know in advance how many elements you need to predict in each picture, pure regression-based models could be a good option.

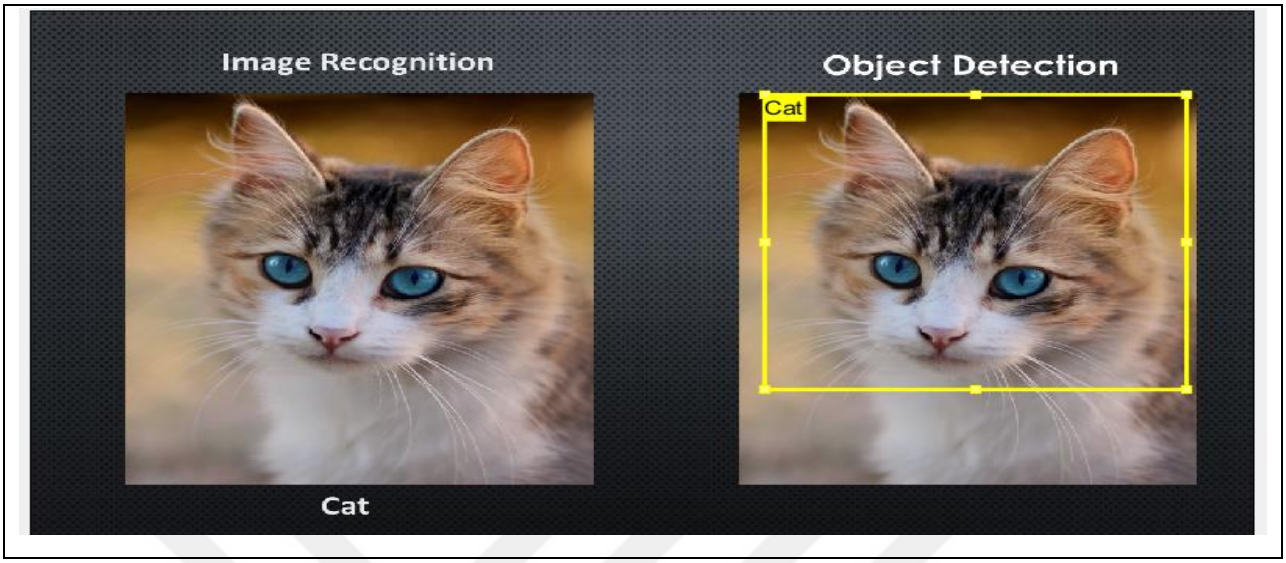


Figure 2.3: A model that defines the detection of a single object.

An improvement on the regression approach is a region proposal network. This decoder model recommends regions of an image where it believes an object could be. The pixels at these places are subsequently given a label using a classification subnetwork (or reject the proposal). Following that, the pixels that include these regions are put via a categorization network. The benefit of this method offers more accuracy, flexible model that can recommend any number of sites that might contain a bounding box. But the greater accuracy comes at the price of the decreased computational efficiency as seen in Figure 2.4.

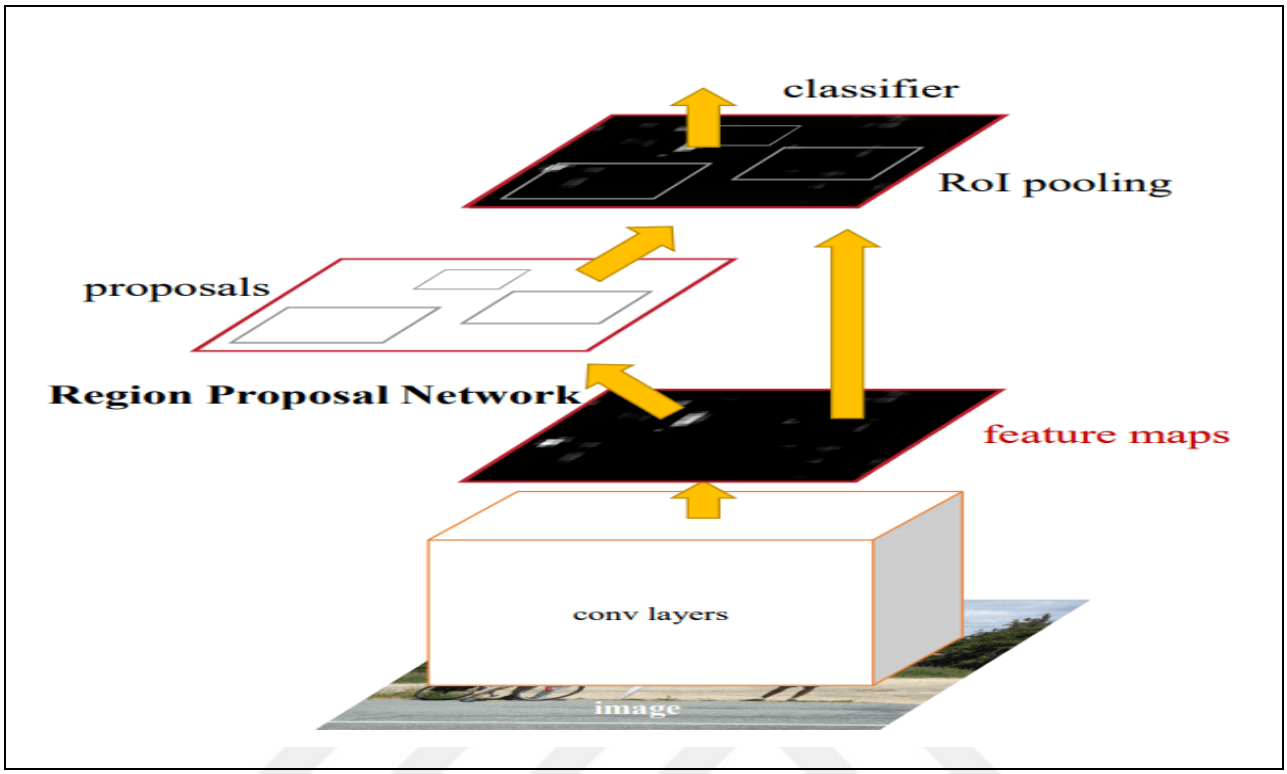


Figure 2.4: A more accurate model can suggest any number of locations that can include a bounding square [45].

The Single Shot Detector, or SSDs, is one of the models for real-time object identification that is much quicker than previous algorithms. By using default boxes and multi-scale capabilities, it may achieve an astounding five to tenfold boost in performance when compared to RCNNs. The action mechanism is shown in Figure 2.5. SSDs depend on a set of specified regions rather than a subnetwork to suggest areas. The input picture is covered with a grid of anchor points, and regions are placed at each anchor point as boxes in a variety of sizes and shapes. The model gives modifications to the location and size of the box to better fit the object, as well as a forecast of whether or not an item exists within the region for each box at each anchor point. Since, there are many boxes at each anchor point and anchor points might be close together, SSDs can result in a lot of overlapping detections. Post-processing of SSD data is required to exclude the bulk of these predictions and choose the best one. The post-processing technique with the most widespread use is non-maximum suppression.

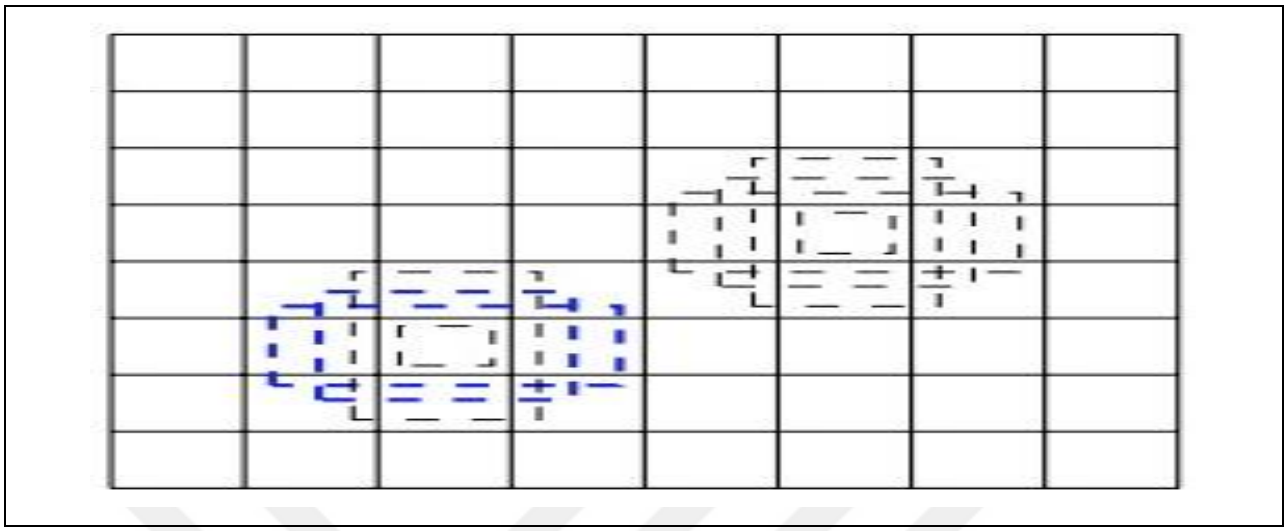


Figure 2.5: Illustrative diagram of Single Shot Detectors (SSDs) [19].

A last word about accuracy; the position and label of each item are produced by object detectors, but how can be determined well from the model performance. The most used measure for determining the position of an item is intersection-over-union (IOU). We calculate the intersections area and divide it by the union area given two bounding boxes. This number is in the range of 0 (no contact) to 1 (perfectly overlapping). A straightforward "% accurate" may be used for labels.

2.3.3 Applications and Use Cases of Object Detection

In this part, an overview of real-world "object detection" application cases will be provided. It will go further and consider how this computer vision technology may affect a variety of industries.

a. Video Monitoring

Modern "object recognition" algorithms easily lend themselves to automated video surveillance systems because they can precisely identify and track many occurrences of a particular item in a scene. For instance, "object detection" models can track a large number of people concurrently and in real-time as they move around a scene or across video frames. This type of exact monitoring might provide valuable insights about safety, worker productivity and safety, retail foot traffic, and more.

b. Crowd Statistic

Crowd counts are another effective use of "object detection" . In densely populated areas like theme parks, retail centers, and city squares, "object detection" may help businesses and municipalities more effectively analyze different sorts of traffic, whether on foot, in cars, or in other ways. With the ability to track people as they travel through different locations, businesses may be able to optimize everything from shop hours and shift scheduling to logistics networks and inventory management. Similar to this, item recognition might help cities plan events, distribute resources, etc.

c. Anomaly Detection

The easiest way to describe the "object detection" use case for anomaly detection is using examples from a particular business. A tailored object recognition model in agriculture, for instance, may accurately identify and localize likely plant disease outbreaks, allowing farmers to recognize risks to their crop yields that would otherwise not be apparent to the human eye. In the medical field, "object detection" may be utilized to treat illnesses with unique lesions. Skin care and acne therapy are two examples; the "object detection" model may identify acne within seconds. These potential use cases are very important and convincing because they make use of and transmit knowledge and information that are often only accessible to agricultural professionals or physicians, respectively.

d. Autonomous Vehicles

The successful operation of autonomous vehicle systems depends on real-time automobile detection models. For these systems to move around the environment safely and effectively, they need to be able to recognize, find, and track nearby items. Object recognition still the primary strategy that underpins current efforts to make self-driving cars a reality, even if tasks like picture segmentation may be (and often are) applied to autonomous vehicles.

3. METHODOLOGY

3.1 INTRODUCTION

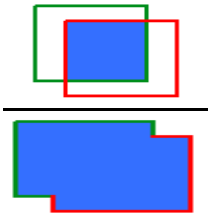
In this chapter, an argument about two important parts of this thesis is discussed: The first part about the algorithm, Yolo algorithm (you only look once), that will be applied for detection. In that discussion, a brief explanation about Yolo algorithm including: the history of development, the mechanism of the process and the most famous applications. The second part, a brief overview of the Matlab coding is explored in detailed.

3.2 YOLO ALGORITHM

The algorithm known as You Only Look Once (YOLO) is popular and has gone viral [20]. YOLO (You Only Look Once) is a cutting-edge and reliable object identification method that works in conjunction with SSD (Single Shot Detector) and Faster R-CNN [21], [22]. YOLO is renowned for its ability to identify objects. The first YOLO version was introduced in 2015 by Redmon et al. [23]. Before YOLO, object detection algorithms would employ classifiers to conduct detection, but YOLO suggests using an end-to-end neural network that predicts bounding boxes and class probabilities, simultaneously. YOLO provides state-of-the-art results, outperforming existing real-time object identification algorithms by a significant margin by taking a fundamentally new approach to object recognition, YOLO runs at up to 45 FPS, making it a far faster algorithm than other competitors [26]. Scholars have recently produced a number of YOLO following versions known as (YOLO V2, V3, V4, and V5) in [22 - 25]. There are a few limited-revision variations, such (YOLO-LITE) [11-12]. The compact size and quick computation speed of the model form the basis of the YOLO target identification technique. YOLO has an easy-to-understand framework. The neural network may immediately output the location and category of the bounding box. The speed of YOLO is quick since it simply requires uploading an image to the network to get the final detection result, allowing it to also accomplish video time detection. YOLO employs the picture file directly for detection, which has the advantage of encoding the global information and lowering the mistake rate of mistaking the backdrop for the item. Due to its capacity to acquire highly generalized characteristics that may be applied to several sectors, YOLO has significant generalization ability. The target detection issue is changed into a regression issue, although detection precision still has to be increased. The test of objects that are close to each other and that

are in the form of groups, the results of the YOLO algorithm are weak. The poor performance is due to the expectation that there are only two bins in the network, and they only belong to a new class of objects within the same class, which results in an anomalous aspect ratio, as well as other factors such as the limited ability to generalize. The positioning error is the primary factor enhancing the detection efficiency because of the loss function. It is necessary to improve handling of both big and tiny things. The key to successful implementation is finding a loss function that strikes the right balance between these three factors [21]. In order to increase the effectiveness of the detection, YOLO employs a number of lower sampling layers and does not utilize all of the target properties that were learnt by the network. 24 convolution layers precede two fully coupled layers in the original YOLO design. Although, YOLO predicts numerous bounding boxes for each grid cell, the highest Intersection Over Union (IOU) with the ground truth is chosen, which is referred to as non-maxima suppression [13]. IOU is a popular measure used to evaluate the accuracy of localization and identify localization errors in object identification models. The starting point for computing IOU using the predictions and the ground truth is the intersection region between the bounding boxes for a certain prediction and the ground truth bounding boxes of the same area. The union, or the combined area covered by the two bounding boxes, is then determined. By dividing the intersection by the union, it can determine from the overlapping proportion to the whole area. It can get a good notion of how well the bounding box reflects the original prediction using this ratio [26].

$$IOU = \frac{\text{area of overlap}}{\text{area of union}} = \frac{\text{}}{\text{}} \qquad (3.1)$$



3.2.1 YOLO v1

The input image is divided using a (SXS) grid, and if the centres of two adjacent items match, object identification is carried out in that grid cell [17]. The bounding box, confidence level, and class probability for each grid are predicted by a grid cell. This prediction is calculated using a tensor named (SXS(B×5+C)), where B is the number of bounding boxes and C is the number of conditional classes of cells. The accuracy and confidence level of the model are given as a

numerical value known as the confidence score, which is used to decide whether an item is included in Equation 3.1.

$$CS = Pr(Obj) \times IUO_{predt}^{ruth} \quad (3.2)$$

If the item is absent from the cell, the confidence score is reduced to 0, and if it is present, it has the IOU value between the ground truth and the prediction box. Intersection over Union is referred to as IOU. The grid's bounding boxes each have the following values: x, y, w, h, and confidence. Following the key class probability is each grid cell. The conditional class probability is multiplied by the bounding box's confidence score at the time of execution to get a class-specific confidence score for each bounding box (Equation 3.2).

$$\begin{aligned} &Class\ Specif\ CS \\ &= Conditional\ Class\ Prob * CS \\ &= Pr(Classi\ /Obj) *\ Pr\ (Obj) *\ IUO_{predt}^{ruth} \\ &= Pr(Classi) *\ IUO_{predt}^{ruth} \end{aligned} \quad (3.3)$$

3.2.2 YOLO v2

The YOLO v2 technique, which is meant to employ a lot of classification data, allows object detectors and classification data to be trained simultaneously with the use of a combined training algorithm. Compared to YOLO v1, a batch normalizing layer was introduced to improve accuracy and speed. In order to stabilize early learning, a convolution anchor size was further set. Additionally, by making adjustments to the classification network, performance was improved even at high-resolution input, and the output resolution was improved while the network was shrunk by carrying out the prediction of the bounding box in the anchor box rather than the fully connected layer [18].

3.2.3 Comparison Between VOLO v1 and VOLO v2

a. The performance of versions 1 and 2 is compared in Table 3.1, and design enhancements were employed to show the performance gain at mA. This indicates that:
b. Adding batch normalization to all convolutional layers has an influence on mAP of around 2%.

c. In version 2, the resolution classifiers was applied with less epochs and trained with 224 224 images, and the mAP increased by 4% as a result.
d. d. Anchor boxes forecast 1,000 layers of boxes per image, while v1 forecast 100. v1 had 81% recall at 69.5% mAP, but v2 recovered 88% at 69.2%.
e. v2 adds a pass-through layer and correlates high- and low-resolution characteristics [20]. With a resolution of 26 26, this layer is similar to the previous one but with a percentage increase. Version 2 (v2) showed that k-means clustering on several bounding boxes for prior identification improves the effect.
f. It showed about 5% gain in mAP in v2 and overcome the constrained chi-prediction by utilizing multidimensional clustering to directly predict the center position of the bounding box.
g. Repeated multi-scale training on the input image size and detection prediction for different image resolutions in the same network allowed the performance on small-size images in v2 to be faster and more accurate.
h. As discussed in [43], Figure 3.1 shows VOC2007's item recognition techniques' effectiveness and accuracy. YOLO v2 performed well in processing time and mAP.

Table 3.1: changes between YOLO and YOLOv2 [32] .

	YOLO								**YOLO2**
Batch norm?		✓	✓	✓	✓	✓	✓	✓	✓
hi-res classifier?			✓	✓	✓	✓	✓	✓	✓
convolutional?				✓	✓	✓	✓	✓	✓
Anchor boxes?				✓	✓				
New network?					✓	✓	✓	✓	✓
Dimension priors?						✓	✓	✓	✓
Location prediction ?						✓	✓	✓	✓
Passthrough ?							✓	✓	✓
Multi-scale?								✓	✓
Hi-res detector?									✓
VOC2007 mAP	63.4	65.8	69.5	69.2	69.6	74.4	75.4	76.8	78.6

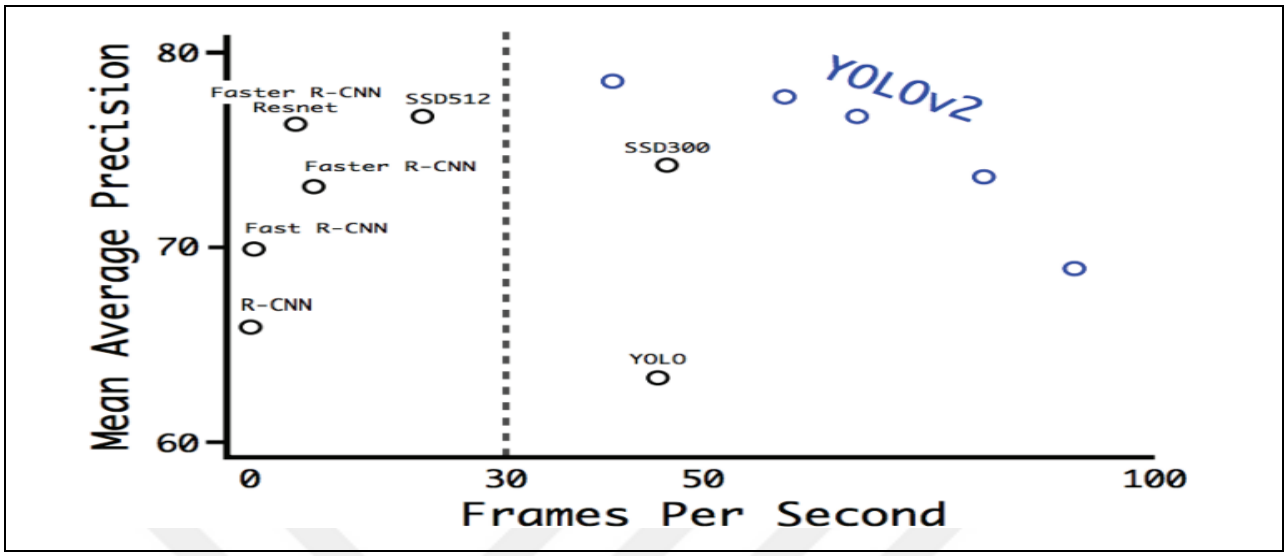


Figure 3.1: Accuracy and Speed of Different Object Detection Models on VOC2007 [32].

3.2.4 YOLO and SSD Comparison

Unlike YOLO, SSD does not split the picture into random-sized grids. Forecasts are made for each location on the feature map about the offset of present anchor boxes (default boxes). Each box has a set size, aspect ratio, and placement in relation to the relevant cell. The whole feature map is convolutionally covered by all of the anchor boxes. The anchors in YOLO 5 and SSD are somewhat different from one another. Because YOLO employs anchors that might be as small as a single grid cell or as large as the whole picture, it bases all of its predictions on a single grid. The SSD's anchors place too much emphasis on target shape dimensions ratios and various practical perspectives. Unlike the anchors of SSD, which are created using a simple method, the anchors of YOLO are calculated using k-means clustering on the training data. Although YOLO calculates the confidence score to show confidence in the anticipated results, SSD may not use it. SSD could realize the task via a unique backdrop class. A poor YOLO confidence score matches the anticipated SSD background class outcomes. Both illustrate the impossibility of the detector ever discovering a target.

3.2.5 Versions and History of Yolo

YOLO developed significantly by Joseph Redmon at first using a phrase on Darknet. Here are some factors that helped YOLO first iteration triumph against R-CNN and DPM:-

YOLO processes 45 frames per second in real time Less background false positive greater detecting precision (although lower accuracy on localization). Since the algorithm's first publication in 2016, it has continued to be developed. Joseph Redmon is the author of YOLOv2 and YOLOv3. New writers who anchored their own objectives in each subsequent YOLO release appeared after YOLOv3. YOLOv2 was released in 2017 and was given an honorable mention at CVPR 2017 due to considerable resolution and anchor box improvements. YOLOv3, in 2018, version included links to the backbone network layers and an extra objectivity score for the bounding box prediction. Because predictions could be made at three distinct levels of granularity, it also offered better performance on small objects. YOLOv4 was published in 2020 paper at the first time by Alexey Bochkovski who revealed brand-new advancements such as better feature aggregation and mind activation. YOLOv5 was developed by Glenn Jocher who focused on the architecture itself in 2020 and releases it with more advancement.

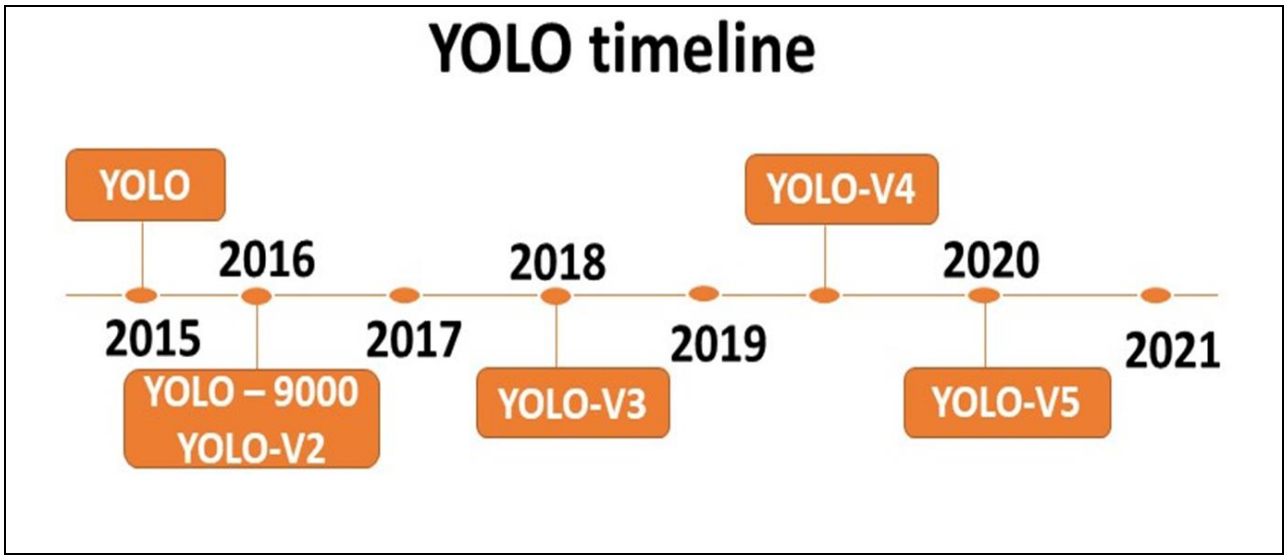


Figure 3.2: A timeline showing the evolution of YOLO in recent years.

3.2.6 How does YOLO Work?

The operation of YOLO-based models is built on three key strategies, none of which seize to gain control of the area:

a) Residual blocks: At this step, the model separates the input picture into equal-sized grids, each of which is in charge of identifying an item or a portion of an object that appears within the grid.

b) Bounding box regression: Each cell has a bounding box with properties such as weight, height, class, and centre that highlights the objects within. With a bounding box regression that represents the likelihood of an item appearing in the bounding box, YOLO forecasts these.

c) Bounding box overlap is referred to as IOU. The predicted bounding boxes and their confidence ratings are the responsibility of each grid cell. By dividing the overlapped area by the union area, the IOU is determined. If the anticipated and ground-truth bounding boxes are identical, the IOU is equal to 1. In this method, it is simpler to get rid of bounding boxes that deviate too far from the actual box.

After partitioning the picture into grid cells, each cell predicts bounding boxes for each item with specific likelihood scores and class probabilities. If you had three separate items, such as a kid, a tree, dog ,bicycle , car and a ball, Figure 3.3 illustrates this all of your predictions would still be made simultaneously. In order for the final detection to result in distinctive bounding boxes that precisely contain objects, IOU guarantees that the predictions are in accordance with the reality.

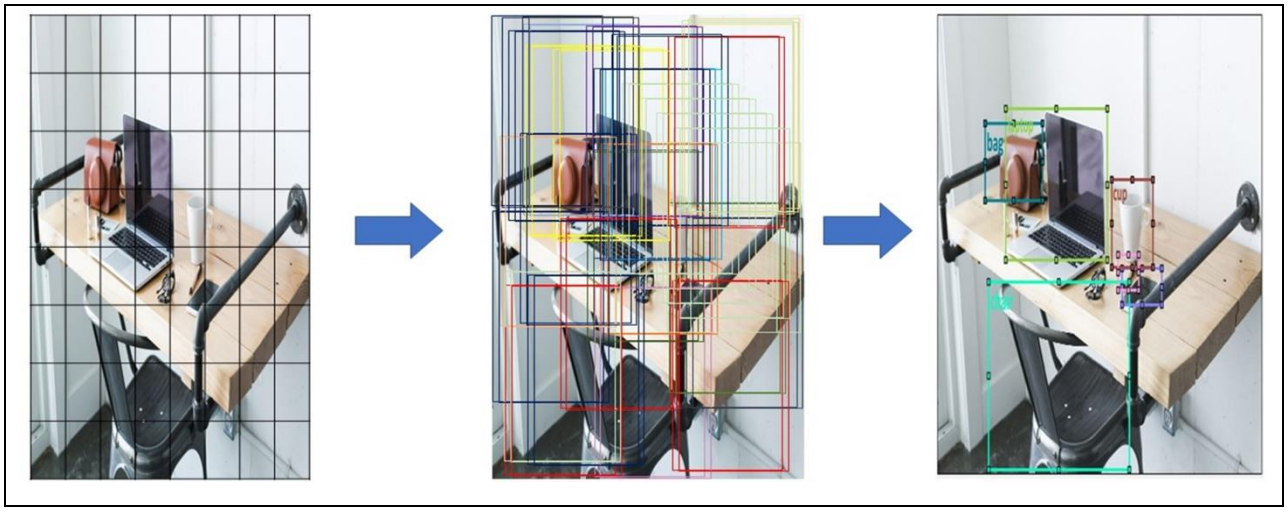


Figure 3.3: Illustration of splitting an image into grid cells.

3.2.7 Real-Time Capability

The majority of network cameras are built on RTSP [31], which can ensure the real-time video frame rate when broadcast over the network. In order to analyze real-time video data from Darknet YOLO permits the usage of IP cameras using the RTSP protocol as input devices source through option setting during execution. The overall design of RTSP-based YOLO object identification is shown in Figure 3.4. RTSP is a queue for buffering network-transmitted frames as assumed little network latency that transmits noticeable detection. If the pace of object detection service (Ts) in the event that YOLO is quicker than the frame interval (Ta) from RTSP, each frame may be analyzed at the highest rate for object recognition in the system to be capable for. However, this processing speed varies significantly depending on the hardware to be used [24]. In general, a system with a powerful GPU with an AGX Xavier or above specification is needed to meet the real-time detection service rate, which is the same as an input rate [24].

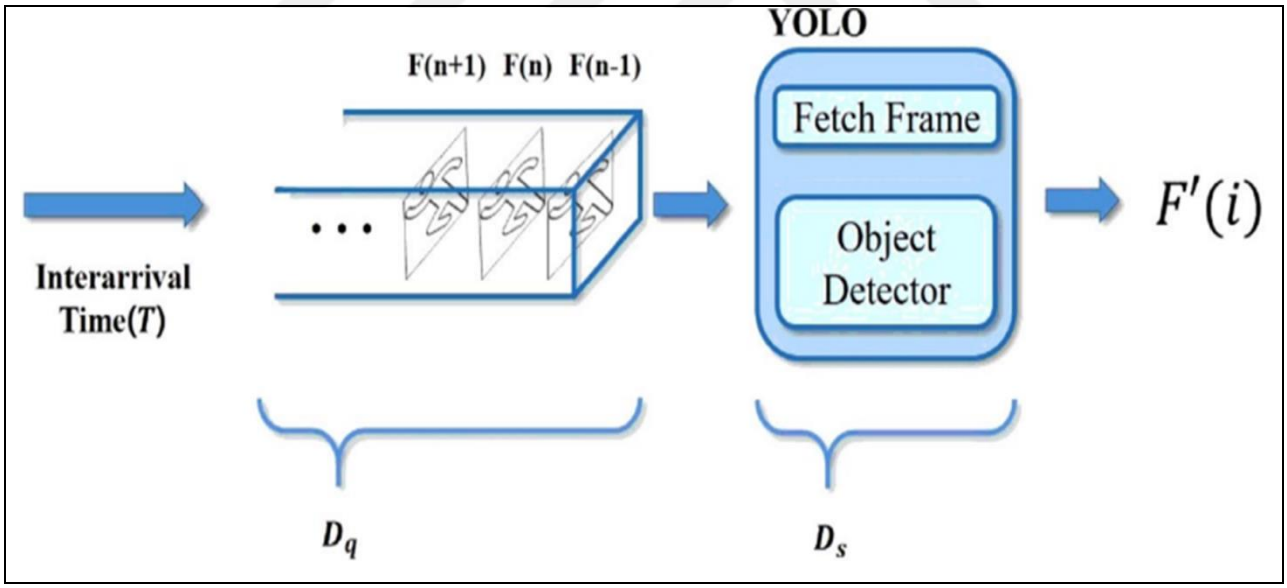


Figure 3.4: Architecture for the YOLO object detection service based on RSTP [24].

3.2.8 Real-Time Object Detection Issues

An application requires a frame processing speed of several hundred milliseconds for real-time object detection. That is real-time object detection rates of around 3-5FPS or 10FPS are adequate. The present YOLO, on the other hand, may have a serious problem with real-time processing if

the object detection service rate is slower than the frame rate provided by the camera. Real-time frames from network cameras come at a nearly constant rate, as seen in Figure 3.5. Ta arriving frames are maintained in RTSP queue, while the object detection service is active when the object detection service time Ts exceeds Ta. The frames are then one by one taken out of the queue and processed when the object detection service is done. The hardware capabilities of the system that runs YOLO for object detection service realizes time variation. Even with the same hardware; the execution time may vary other applications execution time simultaneously. As shown in Fig. 3.5, the waiting time for further incoming frames grow with D(n) growing. If the object detection service time (Ts) is longer than the input period, the total service delay of each input frame, D(n), may be equivalent to the following (Ta).

$$D(n) = \begin{cases} Dq(n-1) + Ds(n-1) + Ta + Dns, (Dq(n-1) + Ds(n-1) - Ta > 0) \\ Dns, (Dq(n-1) + Ds(n-1) - Ta \leq 0) \end{cases} \quad (3.4)$$

where

a. D^n_q is the amount of time the nth frame's queue is waiting.

b. T_a "is the interarrival time for the input frame".

c. D^n_s is the service time required in object detection of YOLO.

As the delay lengthens over time, the overall service completion time for each frame increases until it eventually exceeds the application deadline. As a result of this cumulative, a delay getting longer over time for the object detection is achieved.

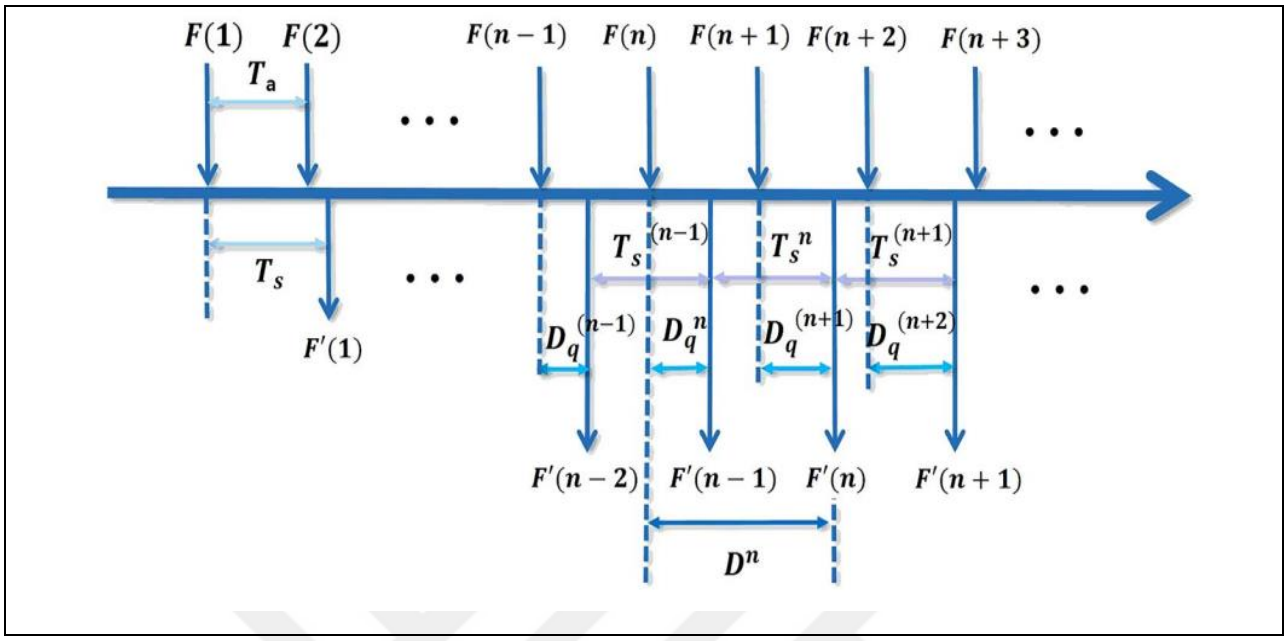


Figure 3.5: Real-time processing issue with YOLO [24].

Such application could process frames from previous periods (t-n), rather than in real-time (t). When using traditional YOLO, a network camera's real-time processing is mostly based on how well its underlying hardware performs. Real-time processing may become difficult if object detection processing is delayed in particular since it will inevitably affect following frames. In the instance of YOLO, an extra component should be included to enable real-time processing despite good recognition performance as its primary objective was to maximize the processing FPS by input files.

3.2.9 Practical Applications Of YOLO

Many useful areas, including healthcare, sports, Security and agriculture, have used object detection. Let's use concrete examples to clarify each one:

a. Healthcare

Planning operations or identifying disorders are only two examples of how organ detection is important in medicine. Before using further image processing techniques like segmentation [28], [21], it is sometimes desirable to add bounding boxes to the organs. Adaptive radiotherapy [30]or laparoscopic surgery [31] may both benefit from real-time organ tracking. When compared to the

Faster R-CNN, SSD and YOLO have the benefit of being real-time models [22]. YOLO starts to show up in the medical industry. Recently, the use of YOLO [31] based models was investigated for the localisation of healthy active organs in 3D PET scans [28], for the identification of lung nodules for the prevention of lung cancer [33], and for the automated detection of nasal cavities in CT images [21]. Taking into the account its durability, quickness, and accuracy on other medical photos. Due to biological differences between patients, localizing organs in real time during surgery might be difficult. Using YOLOv3 for kidney recognition in CT was able to more easily locate kidneys in CT images as shown in Figure 3.6.

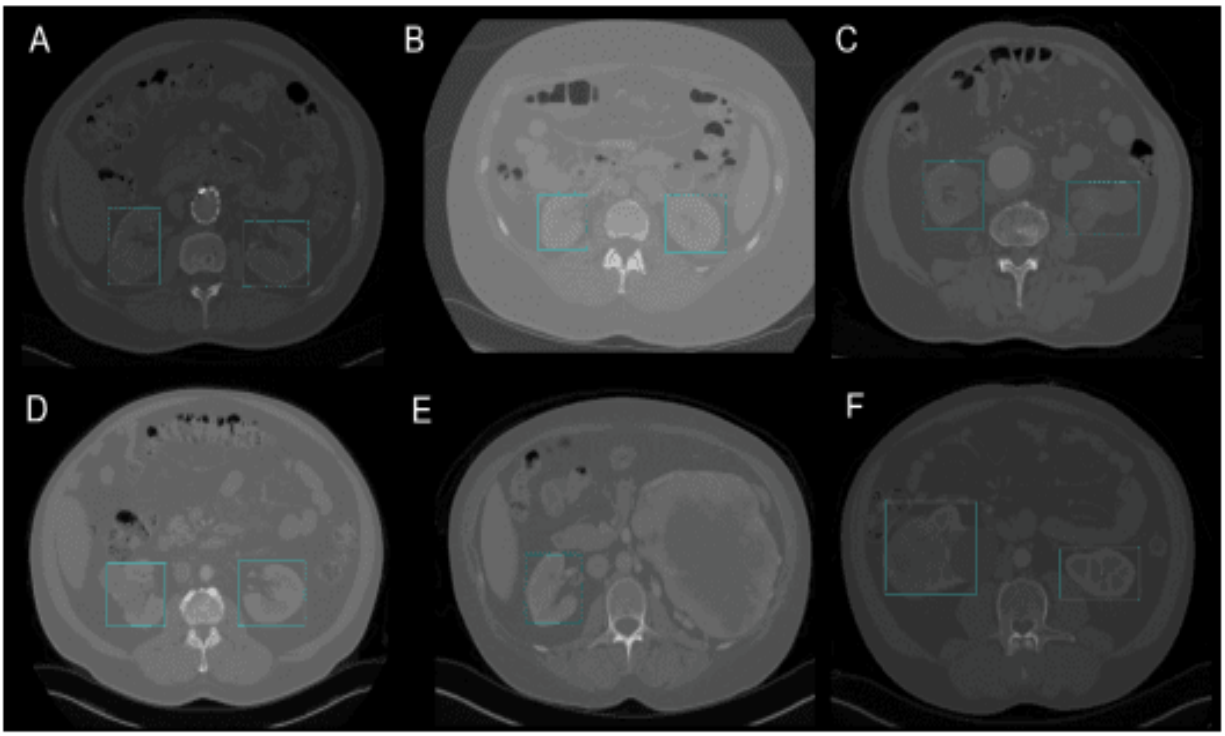


Figure 3.6: 2D kidney detection by YOLOv3. A-B: Normal kidneys with different CT scan [44].

b. Agriculture

The use of artificial intelligence in agriculture has drawn attention from all around the globe, especially in the creation of harvesting robots. This fruit-harvesting robot was developed to take the role of labour-intensive, costly, and sometimes mistake-prone hand fruit picking. The first crucial stage for harvesting robots is the autonomous detection of fruits or other agricultural goods.

A manipulator is often directed to select the fruits based on the results of the detection accuracy tests. However, it is exceedingly challenging to create a computer vision system for fruit and vegetable recognition that is as clever as humans. This is due to a variety of causes, including occlusion, uneven lighting, non-structural fields, and other uncontrollable elements [34]. For the harvesting robots, substantial efforts have been undertaken throughout the years to identify fruit. Wei et al. in [35] utilized segmentation based on colour to separate fruits from their background, whereas Yin et al. in [36] used L×a×b color space to extract ripe tomatoes. However, it might be challenging to choose the optimum colour model in practical settings [35] since fruit recognition relies heavily on colour space effectiveness for the extraction of colour information. The effects of light and leaves on the performance were noticed in Kelman et al.'s [37] examination of the shape approach for localizing ripe apples. Using an ideal threshold on fusion picture characteristics, Zhao et al.'s [38] separated ripe tomatoes from background with 93% accuracy. Due to the use of just features, the lighting had an impact on the researchers' findings Figure 3.7 shows the mechanism of detection of ripe tomatoes.

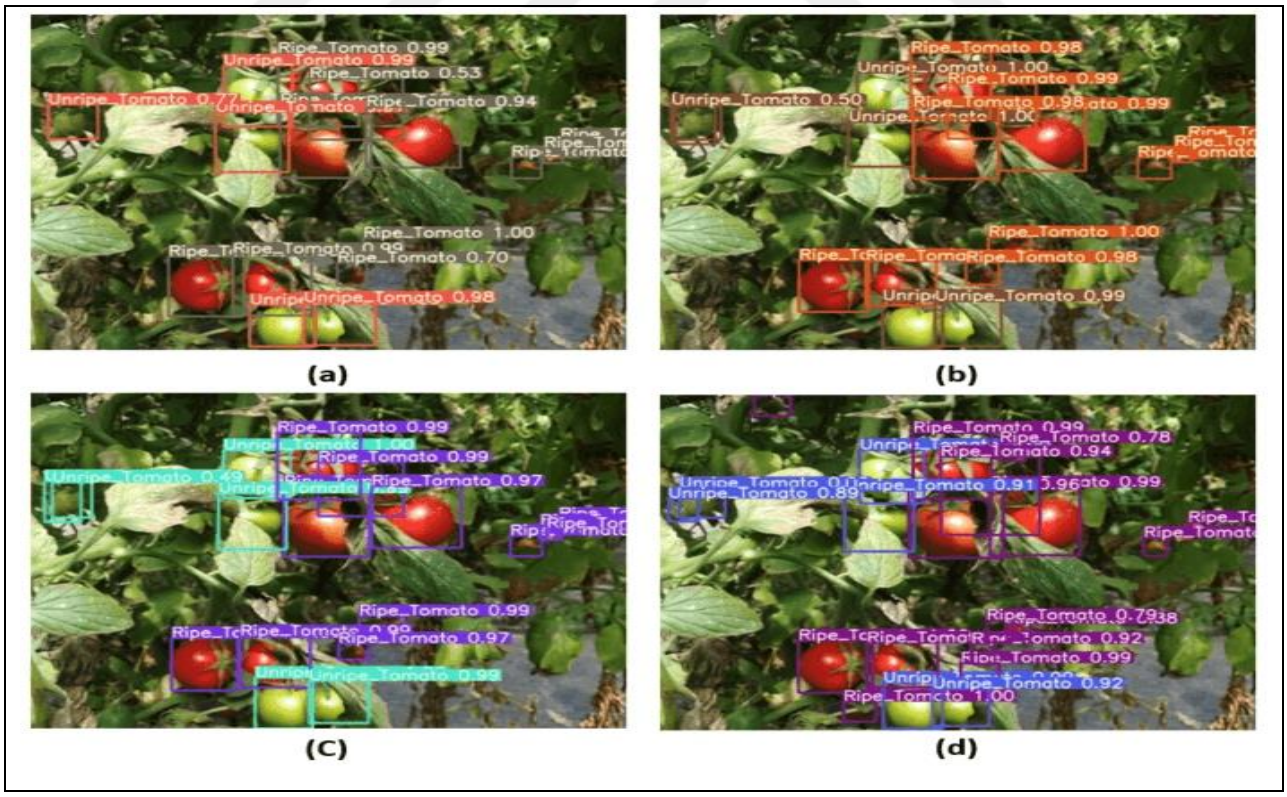


Figure 3.7: The mechanism of determining the maturity of agricultural crops using yolo [24].

c. Security Surveillance

Although, security monitoring is the primary use for object detection, it is not the only one. YOLOv3 has been used to quantify social distance violations between people during the COVID19 epidemic.

d. Self-driving Automobiles

A nation growth and wealth are symbolized by its transportation system. However, it also leads to a number of major issues including traffic congestion and accidents. In the first quarter of 2021, there were 3,206 traffic incidents throughout the country, which resulted in 1672 fatalities. Drunkenness, exhaustion, and improper vehicle control are subjective reasons that affect drivers of vehicles. Countries are building smart systems, including infrastructure and cars, based on the foundation of cutting-edge networks like 5G and 6G networks, in order to eliminate negative effects and increase the efficiency of transportation. Lane recognition and obstacle detection are two of the smart vehicles' most essential functions. They have a direct impact on driving habits. A smart car that communicates its exact location on a road surface based on a lane allows a driver to effectively steer the vehicle. Driving efficiency and safety are significantly increased by obstacles like position and distance from other smart cars or animals on the road, as well as identification of items like signs or traffic signals. Radio, sound, and light sensors like RADAR and LiDAR are used for these tasks. The authors [39] suggest lane recognition and warning method that makes use of a very precise map, an inertial sensor, and a global positioning system (GPS). The LiDAR sensor and camera are combined using a technique by the authors [40]. A LiDAR sensor that provides real-time automobile identification based on distance and light ray intensity data is presented by the authors [41]. Real-time object detection is ingrained in autonomous vehicle systems from the beginning. Because autonomous cars must accurately recognize the right lanes, all nearby objects, and people to maximize road safety, this integration is essential. When compared to straightforward picture segmentation methods, YOLO is a superior contender due to its real-time feature [17].

3.3 MATLAB

A high-level programming language is Matlab. The environment is interactive and is also used to create algorithms and analyse data. It is a crucial component of developing applications and models and offers the user a set of tools and mathematical operations that aid in coming up with quick answers using spreadsheets or even conventional programming languages. Most notably Java (JAVA, C++, C), its use is rising among programmers in the fields of control systems, computational biology, and other areas. Because it was developed by Mathworks, MATLAB is also a matrix or algorithm created expressly for the aim of constructing a digital computing environment with many models. Communication with programs written in other languages can be developed such Python and Fortran Java, as well as user interfaces.

3.3.1 MATLAB FEATURES

MATLAB differs from other programming languages in a variety of ways, the most significant of which are as follows:

a. Its users may get the answers in a familiar mathematical manner, making it simple to use.

b. Providing tools and means that constitute solutions to the problems facing applications and their development.

c. An effective and standard educational tool for several areas, including principles of engineering, mathematics, science, and others.

d. A true model for the development and advancement of software.

e. Best choice for use in writing programs that need a medium range of commands and editing to solve problems.

f. The general performance of the language is abbreviated to manipulate and change numbers.

4. RESULTS

4.1 CHAPTER INTRODUCTION

In the previous chapters, the computer vision and discovering things are discussed along with YOLO algorithm. In this chapter, four main aspects of extracting results and training the proposed algorithm are discussed as follow: First, the computer that is used to train the proposed algorithm with all specifications. Next, the talk about the used code for such algorithm is discussed. Then, specifying the criteria by which the quality of images can be evaluated. Finally, the obtained results of this thesis are discussed.

4.2 SYSTEM HARDWARE

The system relies on a personal computer to process the proposed algorithm. The image data is recorded through a surveillance camera or it can be analyzed from videos or personal photos. The used computer specifications are given as:

a. Operating System: Windows 10 Home 64-bit.
b. System Model: Alienware 17R4.
c. BIOS:1.0.8.
d. Processor : Intel(R) Core(™) i7-6700HQ CPU @ 2.60Hz(8 CPUs), ~2.6GHz.
e. Memory :16384MB RAM.
f. Graphics card: GTX 106 - RAM 8 G.

4.3 SYSTEM SOFTWARE

The proposed system algorithm is implemented using MATLAB R2021a program. The standard programming language, C, may be used with MATLAB to design the proposed algorithms, visualize data, analyze data, and conduct numerical computations. The proposed algorithms for face recognition are created using the image acquisition toolbox, image processing toolbox, and the neural network toolbox. Image acquisition from a frame grabber or imaging equipment that MATLAB supports is made possible via the image acquisition toolbox. This toolbox supports the following acquisition parameters: frame grabber acquisition resolution, trigger specification, color space, number of acquisitions at trigger, area of interest during acquisition, etc. This toolkit serves

as a link between the MATLAB environment and the frame grabber. Numerous reference algorithms, graphical tools, analyses, etc. are offered by image processing toolbox. Fast algorithm development is made possible by reference algorithms. Code creation is made simpler by ready-to-use utilities such as filters, transformations, and upgrades. This toolkit is used in the face recognition and face detection portions. Designing, implementing, visualizing, and simulating neural networks are all provided by Neural Network Toolbox. Tools for clustering, data fitting, and pattern recognition are offered. Both supervised and unsupervised learning networks—such as FeedForward, RadialBasis, Time Delay, etc.—as well as self-organizing maps and competitive layers—are supported.

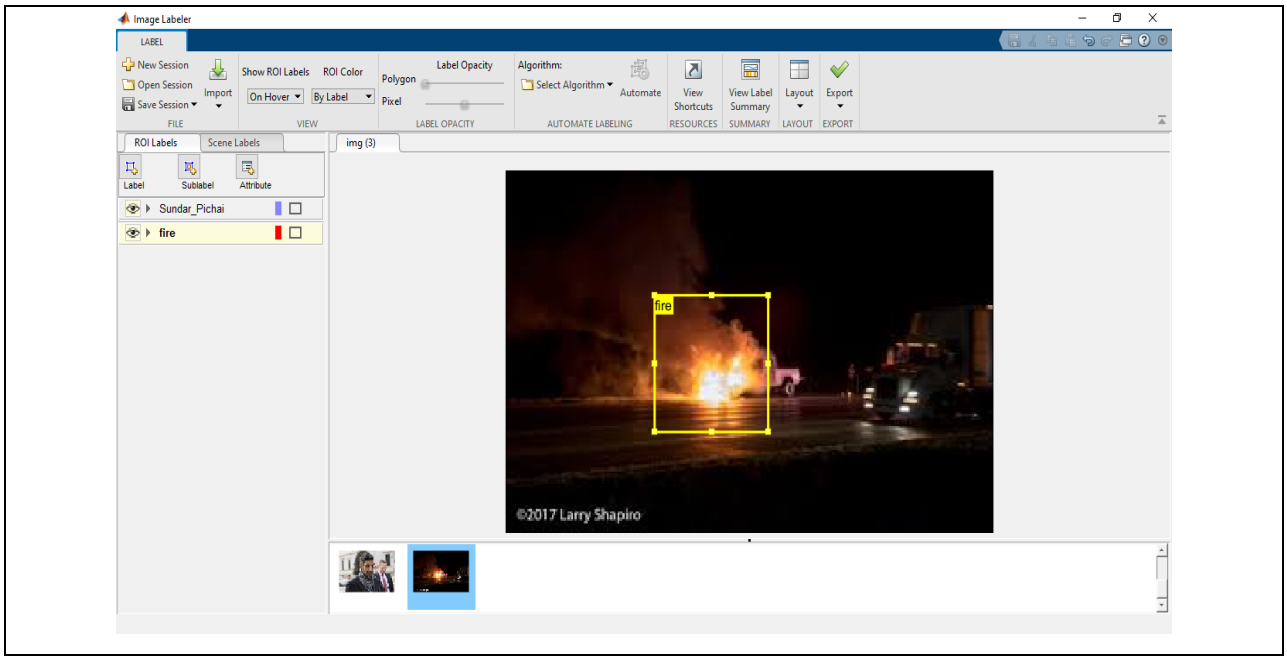
4.4 HISTOGRAM

With digital photography entering the world of photography, it has become easier to know the nature of the contrast in an image after it has been taken or even before it has been taken. All digital cameras give the ability to display a curve histogram after the picture is taken, and some modern digital cameras have what is known as (Live Histogram) through which (i.e. the camera); in which the contrast before taking the picture can be determined. It is a feature that gives the photographer an opportunity to change the orientation of the image being taken so that the contrast is optimal. Also, the ability to review graphs is an important factor to be considered in all image processing programs. Histogram is a curve that represents the final form of color contrast in a scene that is calculated on the basis of grayscale from black to white (from shadow areas to areas of high light). Suppose you want to calculate the sum of the different monetary denominations; which is clear first to classify the denominations in preparation for the collection process and make them in the form of sums that each group represents a particular monetary denomination. For example, stacking similar coins on top of each other and put these denominations one next to the other could provide a curve representing the height of each denomination with the number of coins in that denomination. This is exactly what an electronic digital image calculator applies to plot a curve graph. A digital image is made of millions of light points to be called pixels. The calculator classifies these pixels into categories and each category represents an amount of light. The number (256 assuming an eight-bit color depth) was chosen for the number of light classes starting at color zero. It is assigned to the black color, then the first and second color category, down to the fifty-five color category, which is the white color category. So the histogram is the curve that shows the

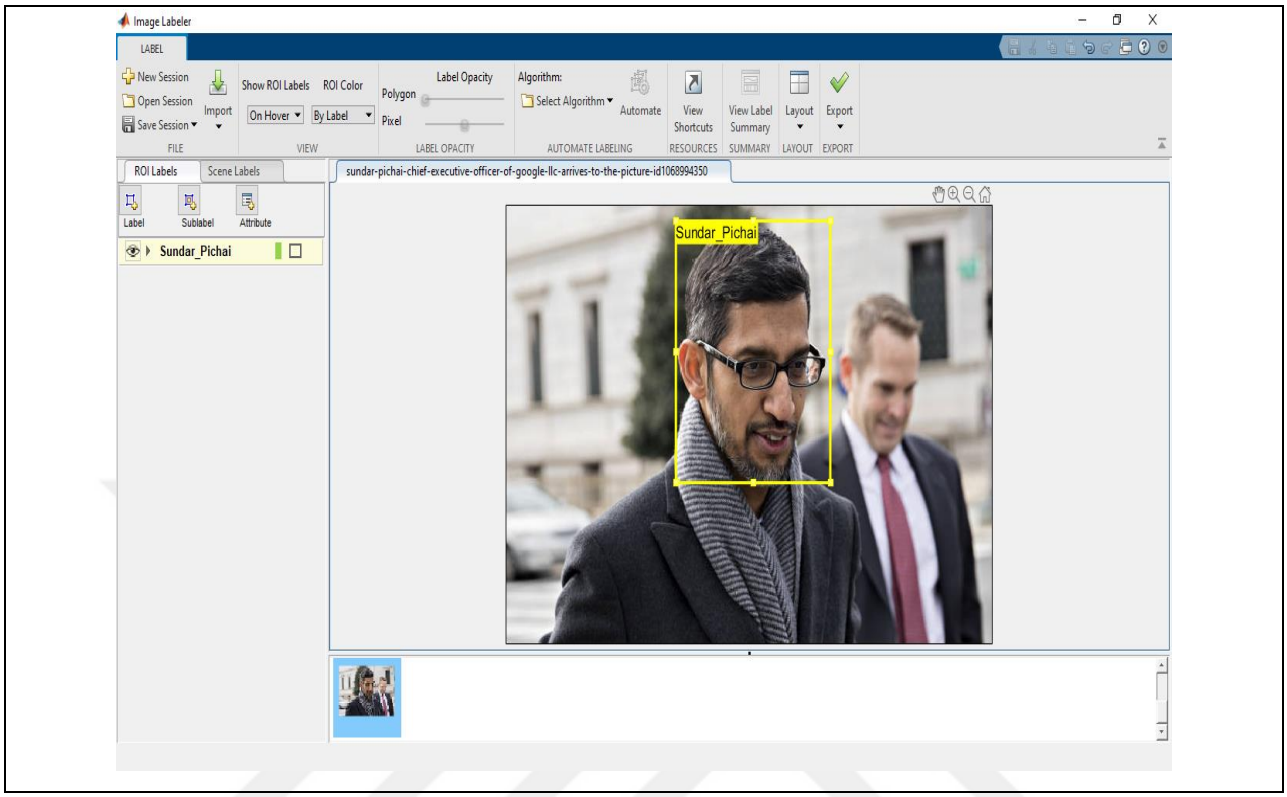
nature of contrast in an image, and it's best when it covers all areas of light, from deep shadows to high light areas. I can say that Curve Graph is an idea inspired by the world famous photographer Ansel Adam's System of Areas. The difference of the histogram from the system of areas is that the histogram is a method that can be used to find out the color contrast of an image or scene when using digital cameras or after converting the image to an electronic calculator, that is, after digitizing it. The histogram will be the tool with which can be determined accurately and performance in this thesis.

4.5 OUR RESULTS

We developed the object detection work through two parts, in the first part, the process of object detection is developed through fire detection, and in the second part, the detection of people wanted by the state through facial recognition. This system was built using a second generation algorithm for object detection ("You only look once"). We will explain how this? The system was built and designed in two phases. The first stage is data collection and training. Images similar to the part to be detected are entered into Matlab software. Then the Image Labeler tool, one of the tools in Matlab, is used. With this tool, you can select the part to be recognized in the image and give it a name or a symbol. As shown in Figure 4.1.



(a)



(b)

Figure 4.1 : Selecting the part to be recognized: (a) Determine the shape of the fire in the image and (b) Determine the shape of the person to be recognized in the photo.

After completing the process of defining the frame for the part to be discovered, the program creates a network within this frame as explained in the previous chapters. It is done exporting outputs to a table that contains (X, Y, W, H) as shown in Figure 4.2 .

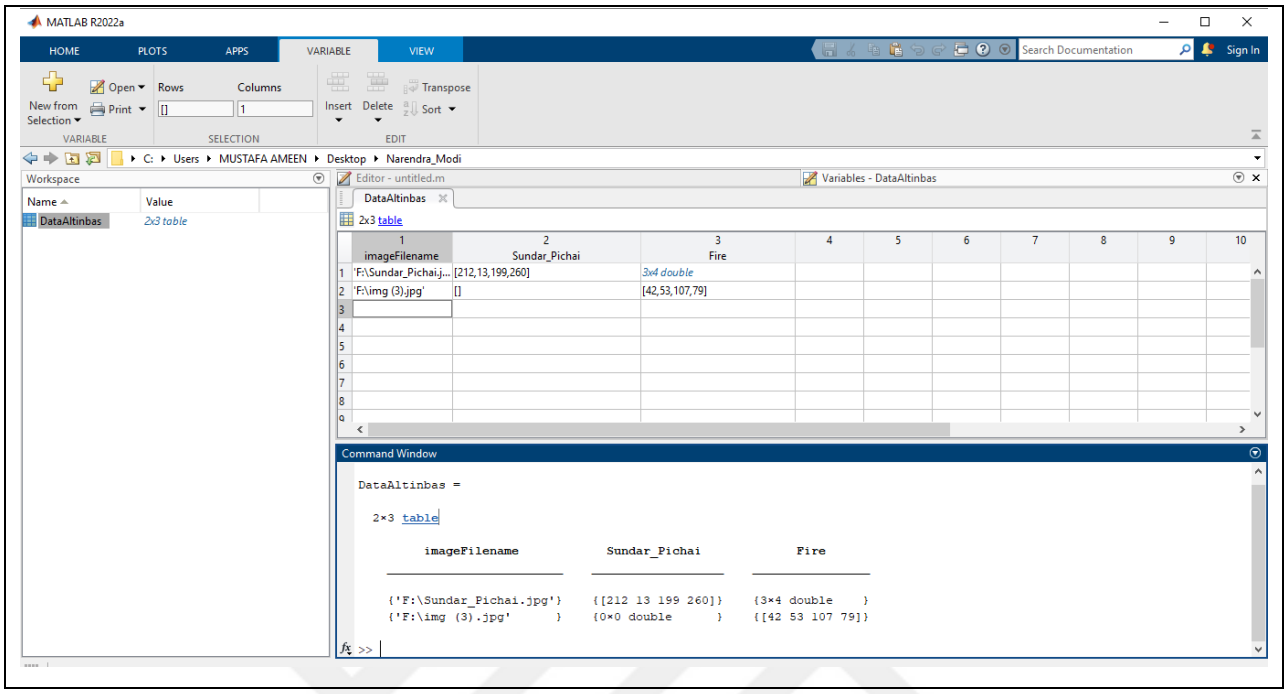


Figure 4.2: shows the contents of the exported frame table.

In table, the pixels of the selected frame are listed. We trained two parts, the first part is fire images, and the second part is the face recognition structure. It is applied in the proposed algorithm to be trained based three famous personalities (Barack_Obama, Sundar_Pichai and Narendra_Modi) to reach 757 images as shown in Table 4.1.

Table 4.1: Number of trained data.

Name	Number of trained images
Fire Detection	412
Barack_Obama	163
Sundar_Pichai	55
Narendra_Modi	127

After completing the training process, the algorithm is now ready to determine what is required. The algorithm has enough knowledge to identify the fire and the faces of the wanted people. Now, the explanation of fire recognition experiment images is shown in Figure 4.3.

Figure 4.3: shows the fire detection mechanism.

After the fire training process, it is tested the algorithm and performed 90 tests that passed 88 tests, two of which failed. Figure 4.4 shows the mechanism of success and failure rate.

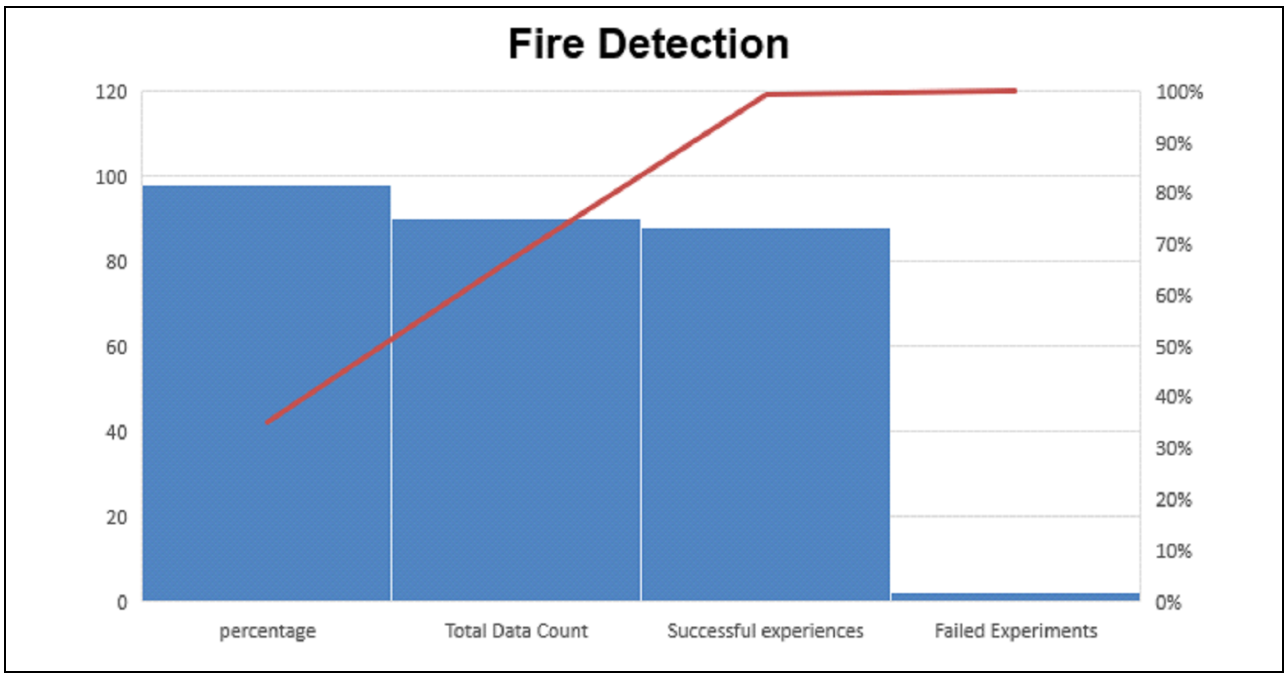


Figure 4.4: shows the number of successful and failed trials and the success rate of the algorithm.

Figure 4.5 shows one of the successful and unsuccessful exercises in the algorithm.

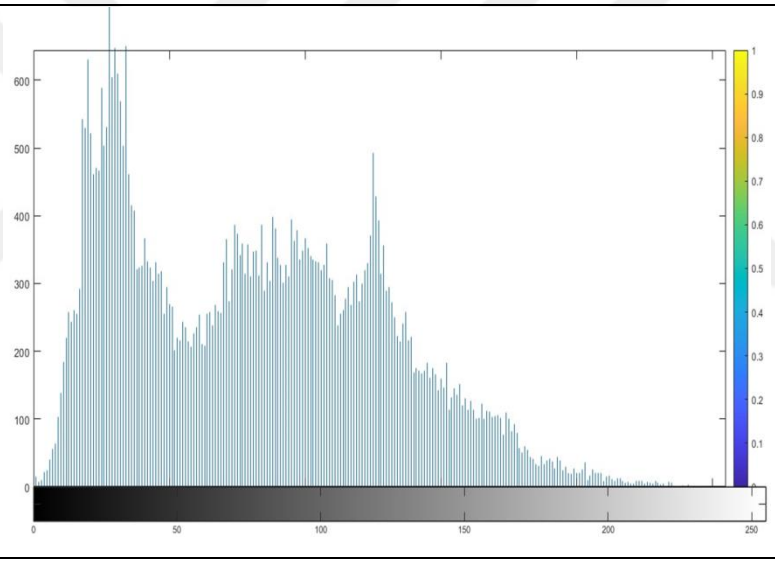
(a)

(b)

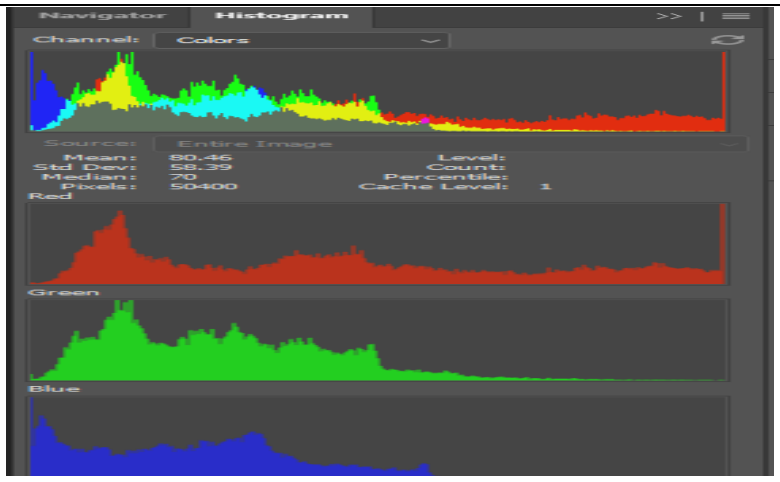
Figure 4.5: Shows the failure and success experience while running the algorithm: (a) The event that succeeded during training and (b) The event that succeeded during training.

After conducting an experimental test to one event fails and susses the cause of the failure by comparing the success event with the failure event according to histogram as shown Figure 4.6.

a. The photo of the event that has proven successful is in grey.

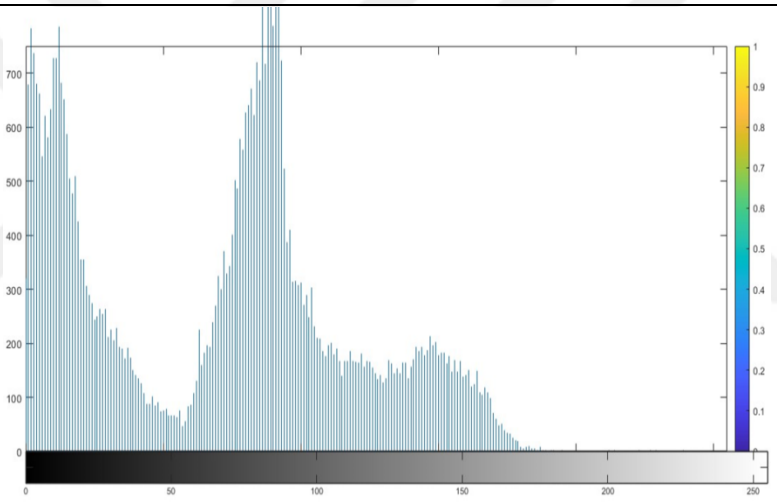


b. Split histogram pixels according to a standard histogram of a successful event.

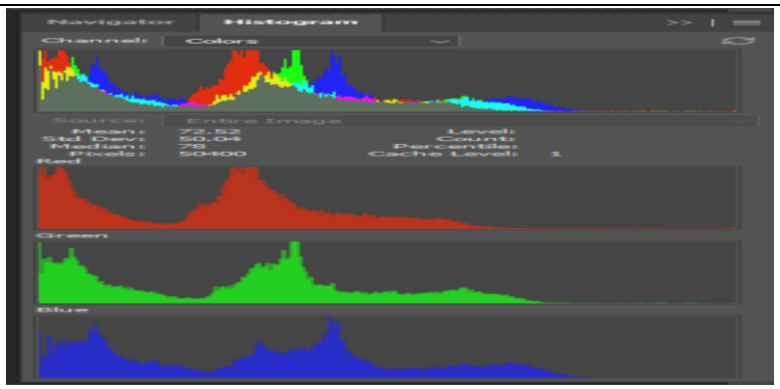


c. Colour segmentation according to a standard histogram of the successful event.

d. The image of the failed event is greyed out.



e. Split histogram pixels according to a standard failure event histogram.



f. Colour segmentation according to a standard failed event histogram.

Figure 4.6: shows the difference between a success and failure event according to the standard graph of a fire recognition event.

Now, the second part is applied by determining the wanted face among three famous personalities as seen below:

a. Barack Obama picture is conducted after 34 experiments to identify Barack Obama's face after 28 successes and 6 fails as shown in Figure 4.7.

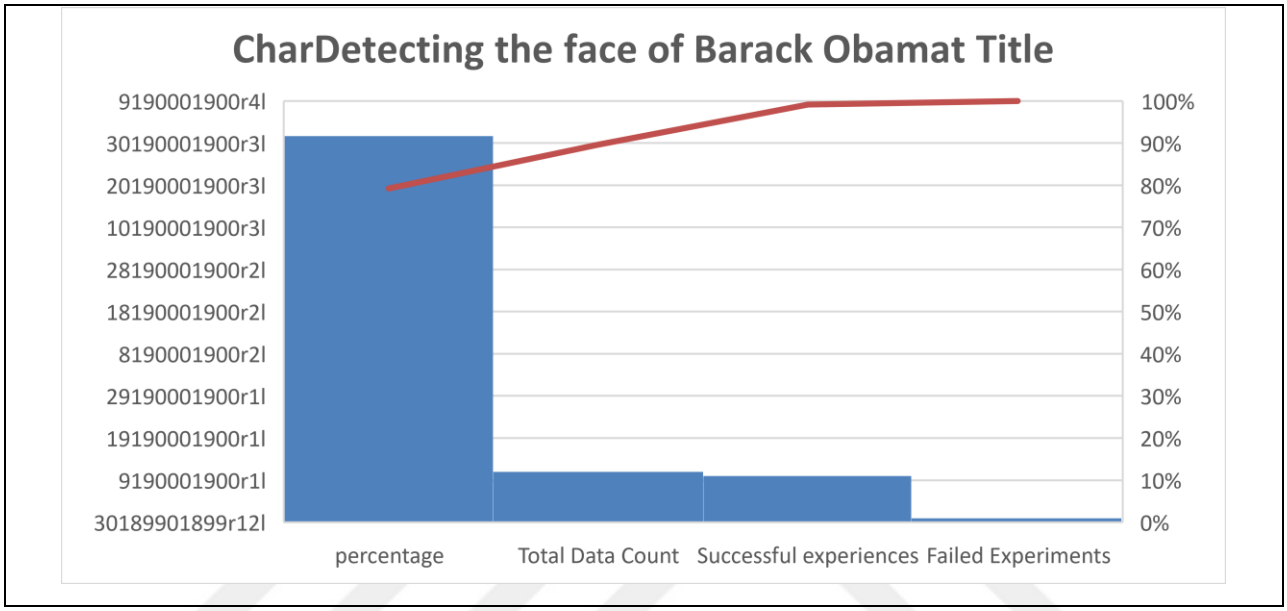


Figure 4.7: shows the success and failure rate of Barack Obama's facial recognition experiment.

After determining the success and failure rates; the practical results of testing the performance of the proposed algorithm is discussed in this part as shown in Figure 4.8.

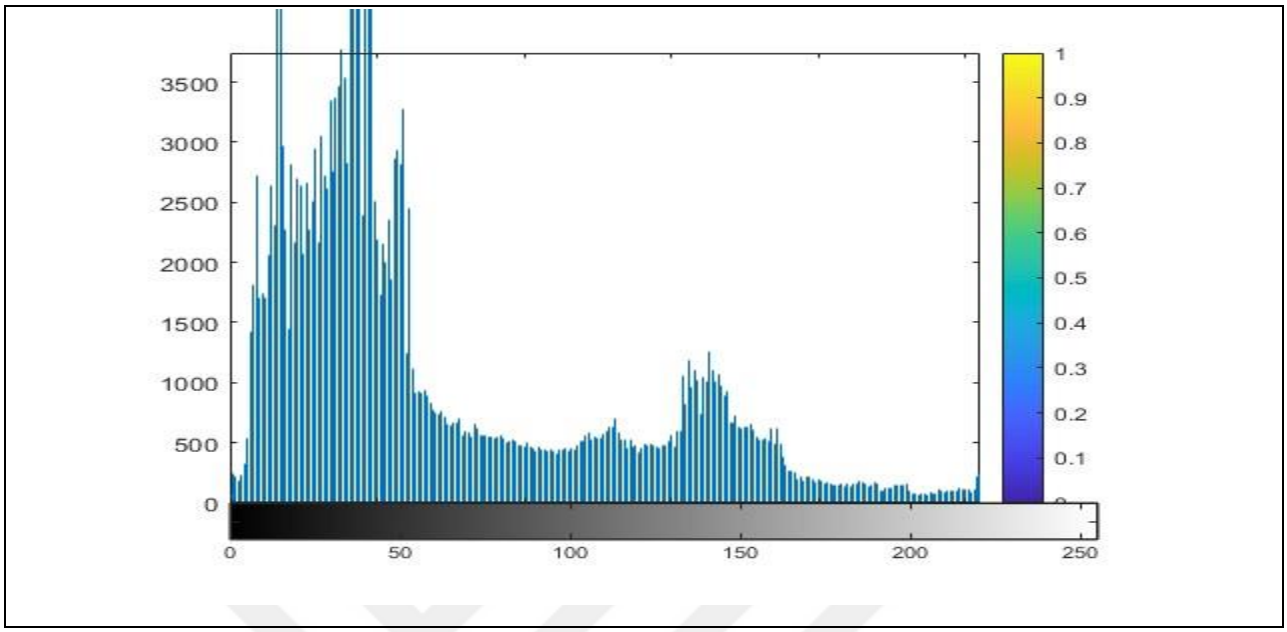
a. Experiments that succeeded and the face was recognized.

b. Experiments that failed and the face was not recognized.

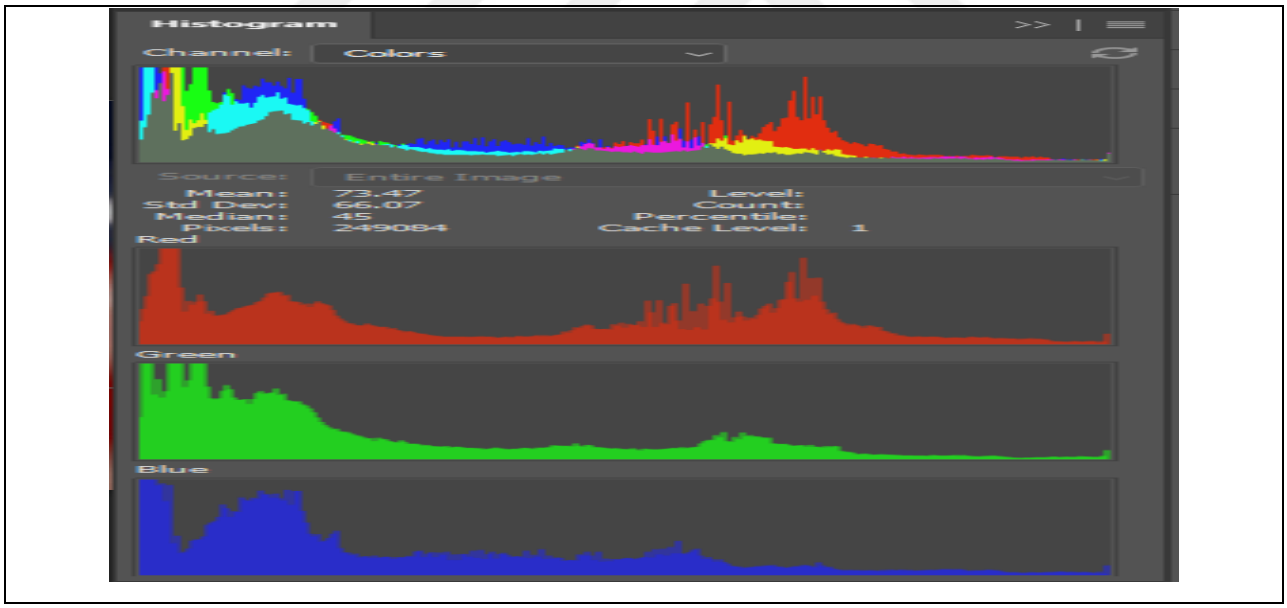
Figure 4.8: shows the result of the experiment process, discovering Barack Obama's face.

After reaching the final conclusions, it is necessary to find out why some events failed during the experiment as seen in Figure 4.9.

a. Grayscale image of success experience.

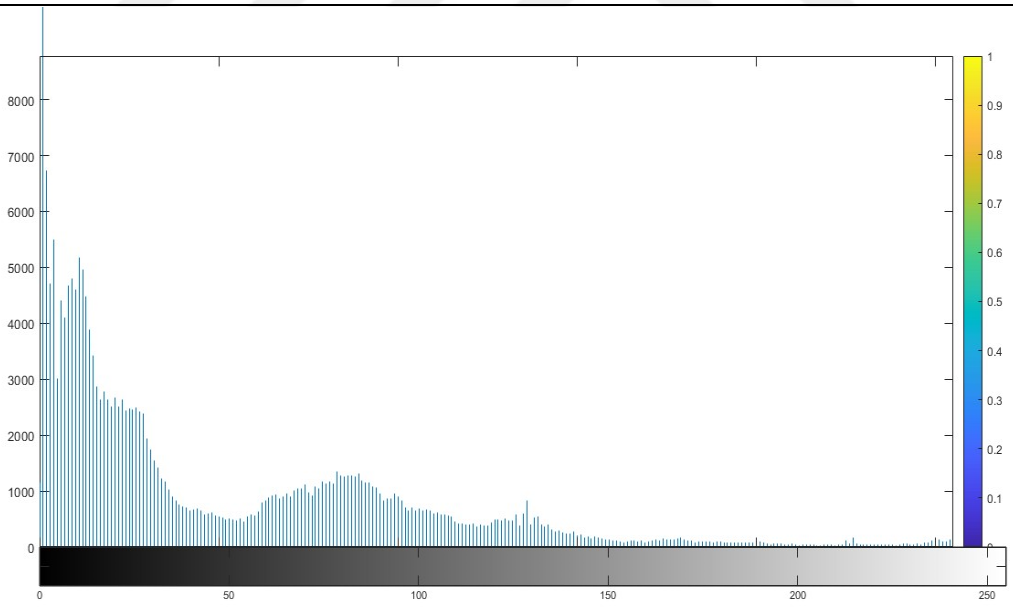


b. Split histogram pixels according to a standard histogram of the success event .

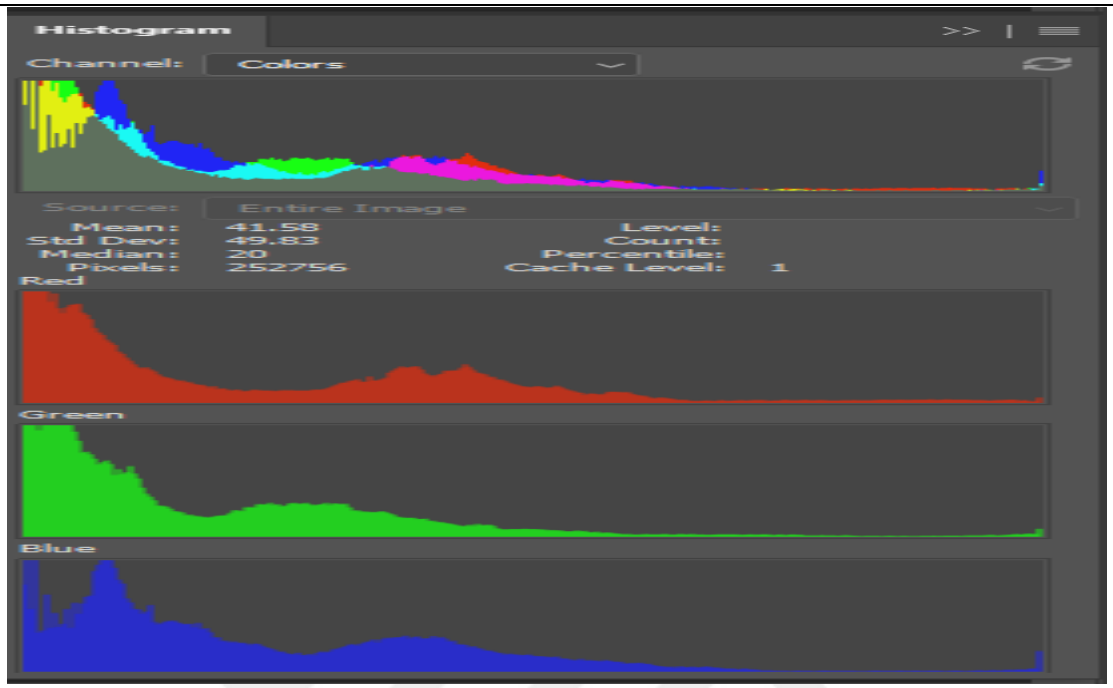


c. Colour segmentation according to a standard histogram of the success event.

d. Grayscale image of a failed experiment.



e. Split histogram pixels according to a standard failure event histogram.



f. Colour segmentation according to a standard failed event histogram.

Figure 4.9: shows the accuracy ratios for each pass-and-fail according to histogram.

b. Narendra Modi Face Recognition is invoked to include 26 images for face recognition experiments. 25 events are of them succeeded and only one failed. Figure 4.10 shows through the graph the success and failure rate of this experiment.

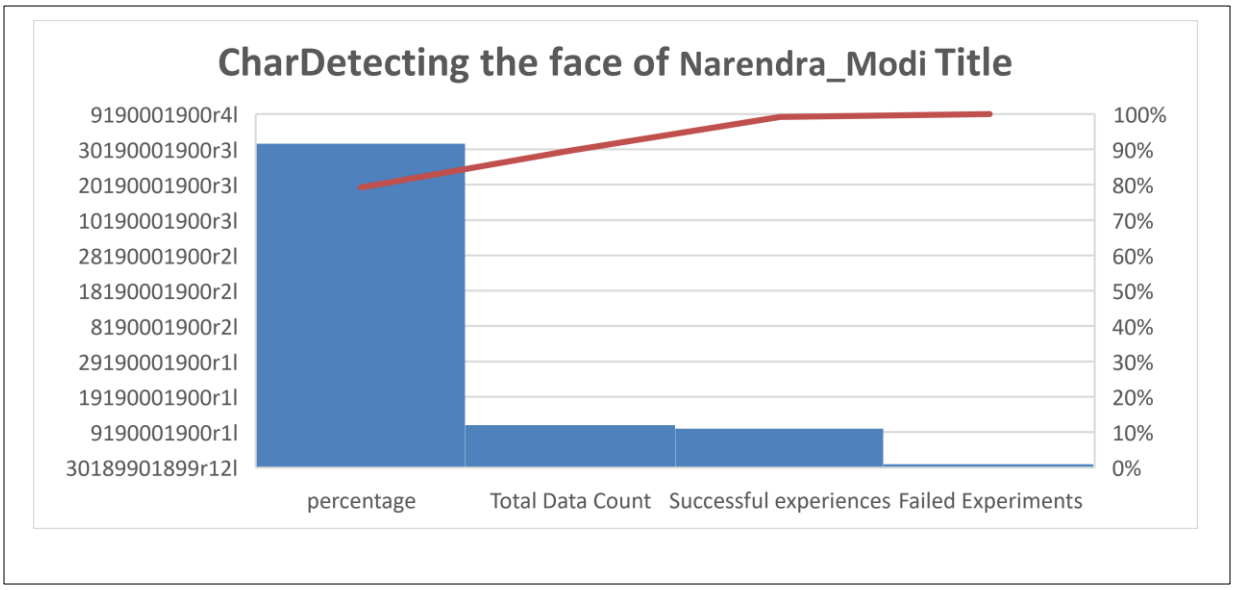


Figure 4.10 shows the graph of success and failure rates for Narendra Modi's face recognition.

After looking at the success and failure rates of Narendra Modi's facial recognition experiment, Figure 4.11 now shows us the results of the experiment.

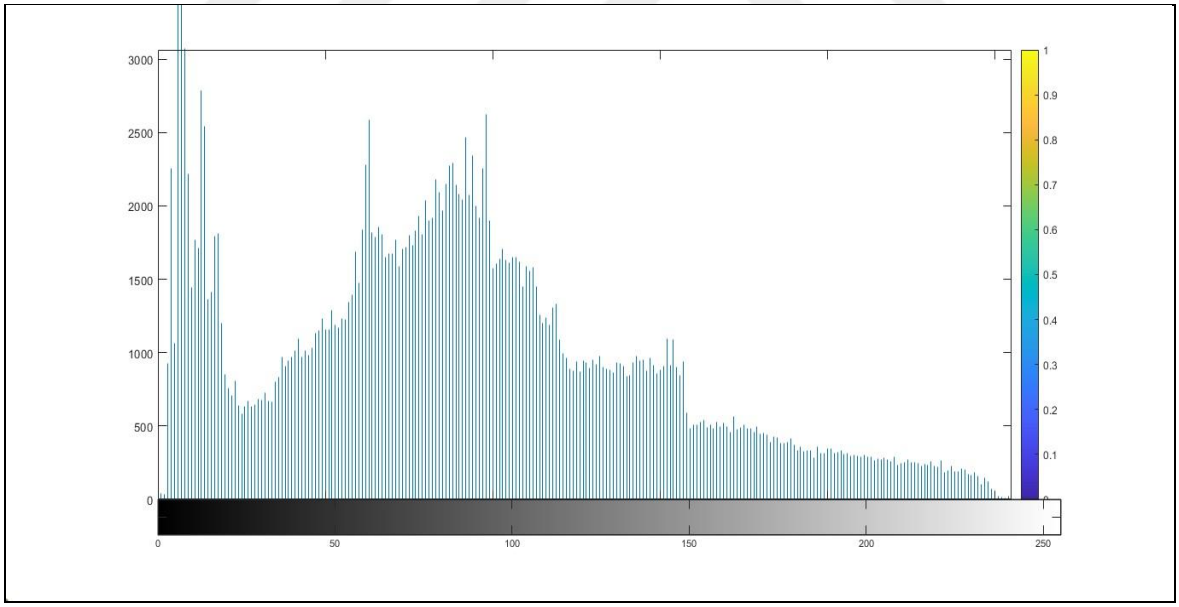
a. Successful Events.

b. Events that failed

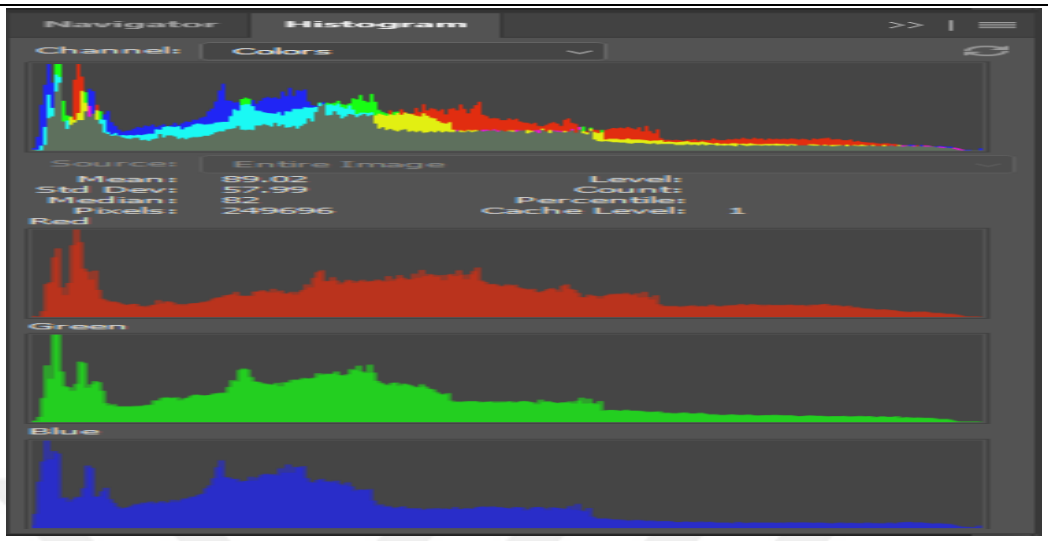
Figure 4.11: shows the results of an experiment with Narendra Modi.

After showing the results, it should know how good the failure and success results are according to the criterion of histogram Figure 4.12 shows.

a. Grayscale image of success experience.

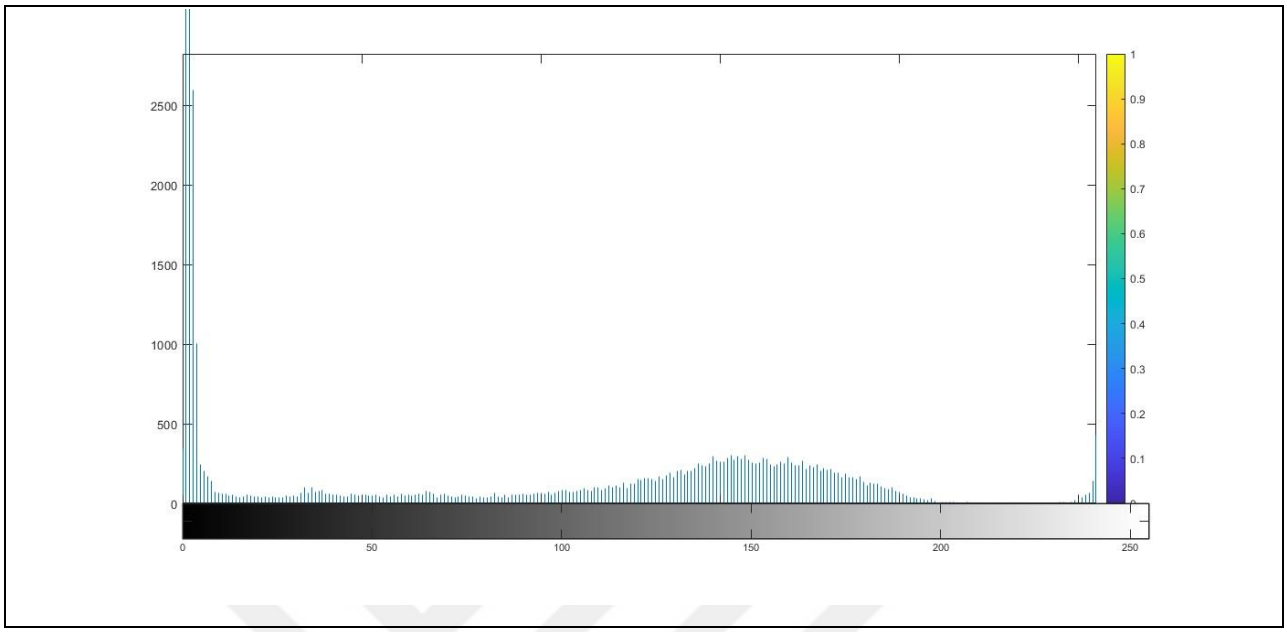


b. Split histogram pixels according to a standard histogram of the success event.

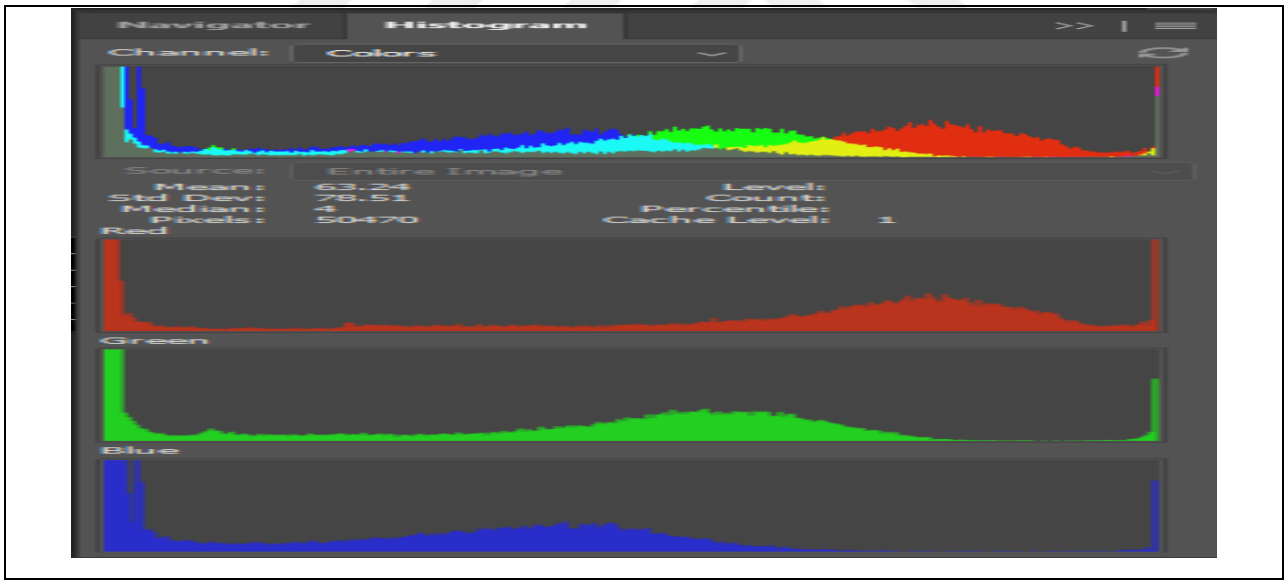


c. Colour segmentation according to a standard histogram of the success event.

d. Grayscale image of a failed experiment.



e. Split histogram pixels according to a standard failure event histogram.



f. Colour segmentation according to a standard failed event histogram

Figure 4.12: shows the accuracy rate of pass/fail Narendra Modi facial recognition according to the benchmark graph.

c. Facial recognition Sundar Pichai image is ran with 12 facial recognition experiments in Sundar Pichai, of which 11 succeeded and only one failed. Figure 4.13 shows through the graph the success and failure rate of this experiment.

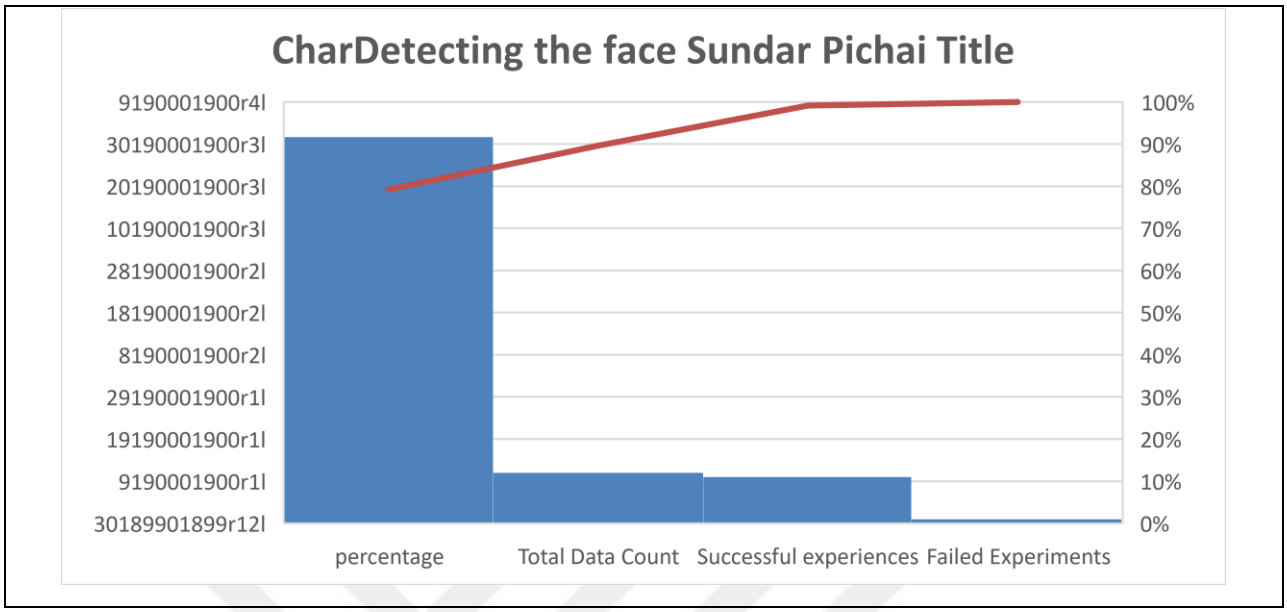


Figure 4.13: shows the graph of success and failure rates for Sundar Pichai's face recognition.

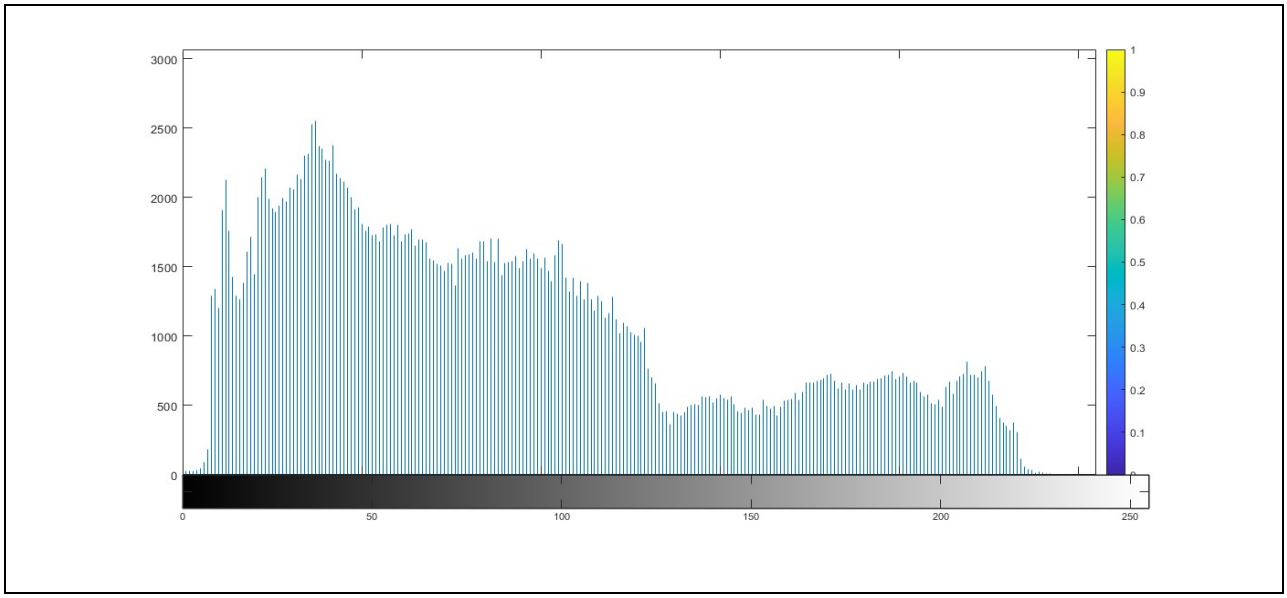
a. Successful Events.

b. Events that failed

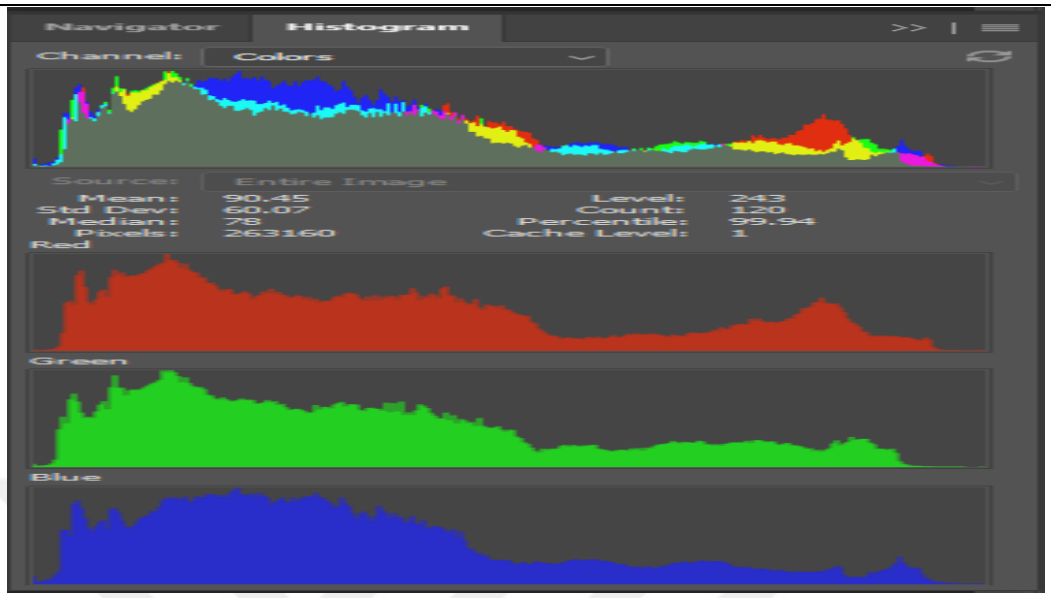
Figure 4.14: shows the results of an experiment with Sundar Pichai's.

After viewing the results that ran to determine the face of 12 Sundar Pichai, it should know how good the failure and success results are according to the standard Chart Figure 4.15 shows this.

a. Grayscale image of success experience.

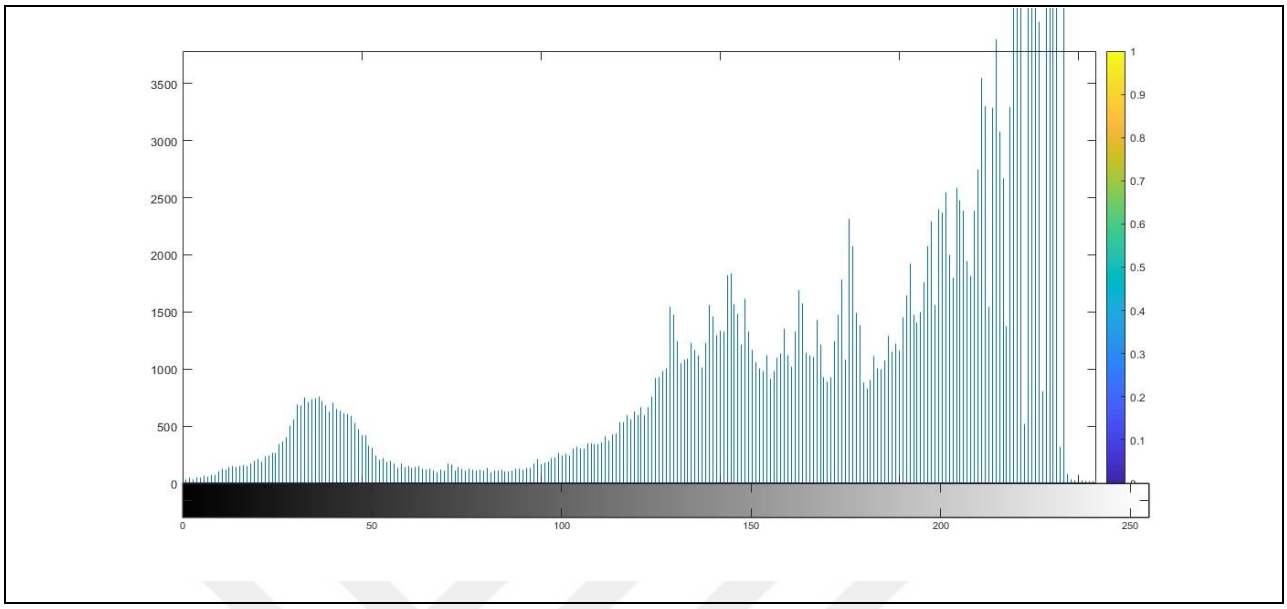


b. Split histogram pixels according to a standard histogram of the success event.

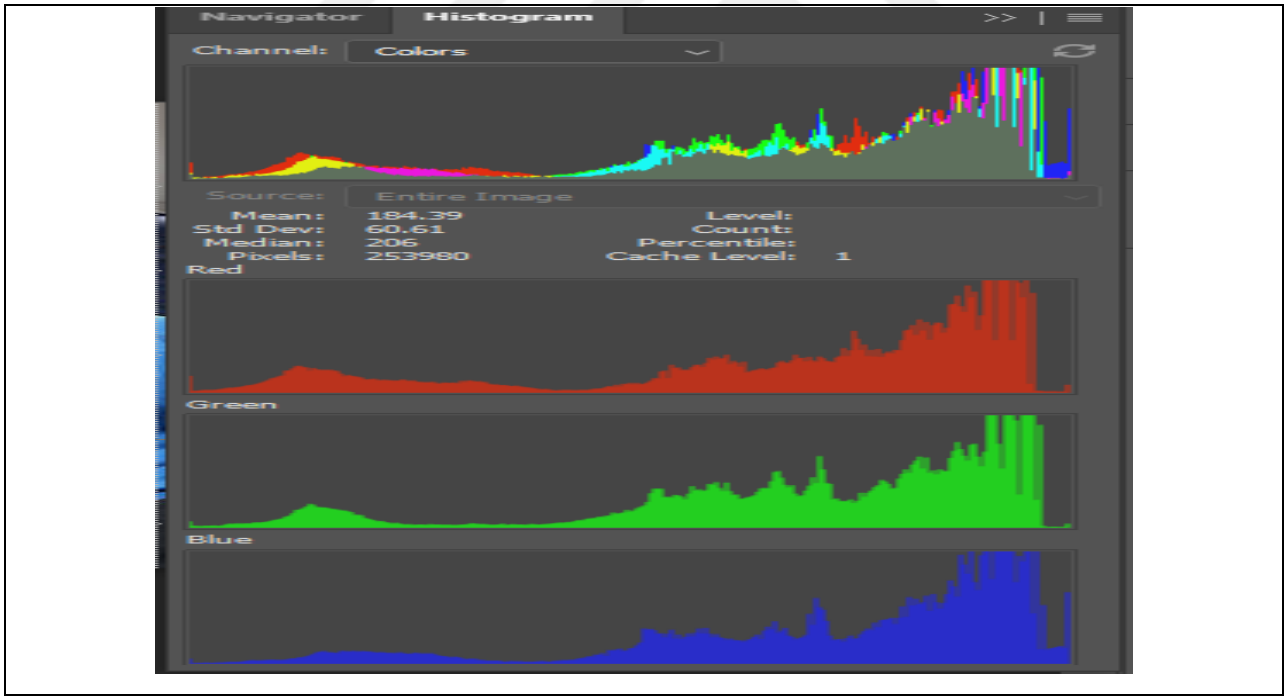


c. Color segmentation according to a standard histogram of the success event.

d. Grayscale image of a failed experiment.



e. Split histogram pixels according to a standard failure event histogram.



f. Color segmentation according to a standard failed event histogram.

Figure 4.15: shows the accuracy rate of pass/fail Sundar Pichai facial recognition according to the benchmark graph.

Matlab code was used to identify the pixels in each event, and Photoshop code was used to identify the colours in each event according to the graph standard. After displaying the results of fire events and discovering the faces of three famous personalities. Nevertheless, displaying the results of failure and success for each of case is displayed. Then, an analysis was conducted for each of them to prove the quality of performance of each event. It is discovered that the failed events do not correspond to the event pixels correctly, or they are in a blurry form as shown in Figure (6,9,12,15-e), and it was also clear to us that the colours were poor and the light intensity was weak in the events which failed as shown in Figure (6, 9, 12, 15 f). It is concluded from the proposed algorithm recognizable events because the quality of the bad events and not because of the algorithm. As the algorithm is able to identify the rest of the events, although the quality of these events is not perfect, that the proposed algorithm is able to identify and discover these events. The outcomes of the proposed algorithm are found very responsive to 3 seconds. The algorithm is also able to identify faces of different sizes and dimensions with different accuracy. Ii is also identified indirect faces as shown in the results can be detected. Finally, Table 4.2 shows the success rates of each of the four detections.

Table 4.2: Success rate for each algorithm.

Name	Success rates
Fire Detection	97.78%
Barack_Obama	82.35%
Sundar_Pichai	91.15%
Narendra_Modi	96.15%

5. DISCUSSION AND CONCLUSIONS

5.1 CONCLUSION

Due to the wide use of surveillance cameras in public streets and cities for recording accident and their ability to be compatible with computer vision algorithms in real time processing, this thesis is conducted the use of such cameras to develop a high detective system with high prediction technique for fire and faces. Therefore, the use of YOLO algorithm based on AI networks is invoked to develop such systems after calling about 757 images for faces and fire from different rotational and rotational axes. This work is applied by generating a Matlab code to develop the computer vision through a traditional digital camera. It is found from the proposed algorithm and excellent response for fire detection and face prediction within about 3 seconds. Therefore, this work is considered an excellent enhancement in the accuracy of computer vision technology with minimum processing time. This work is found to be a novel candidate through what was previously discussed and reviewed for the previous studies and algorithms. Nevertheless, it is concluded that the success rate of the proposed algorithm increases with the increasing the rate of knowledge based on training data to give a knowledge of all aspects of detections.

5.2 FUTURE WORK

After the success of the algorithm, whether in detecting the fire or identifying the wanted persons, the algorithm proved its efficiency in making a decision at a rate of only three seconds with the possibility of knowledge and discovery from remote places in addition to the possibility of knowing faces of different sizes. The future work will be proposed as following:

a-Since the algorithm was able to identify two different things, one over the other, and showed the same efficiency; it is possible to train the algorithm to recognize floods and all disasters that threaten cities.

b- Linking imaging and surveillance devices to state systems that specialize in each field, such as sending a notification to police departments in case the state requests identification of a person in a particular location.

c- Training this system to make decisions on its own, for example, to run sprinklers on the streets when there is a fire.

d-Training the algorithm to understand suspicious events, train itself on its own, and understand the danger before it happens through the various data that pave the way for this danger, for example, when a car emits high smoke, the algorithm can understand that this car will burn before the flame appears.

REFERENCES

[1] M. J. Karter, “Fire loss in the United States in 1993.,” *NFPA J.*, vol. 88, no. 5, pp. 57–60, 62, 1994.

[2] S. Afra and R. Alhajj, “Early warning system: From face recognition by surveillance cameras to social media analysis to detecting suspicious people,” *Phys. A Stat. Mech. its Appl.*, vol. 540, p. 123151, 2019,

[3] G. M. Zafaruddin and H. S. Fadewar, “Face recognition using eigenfaces,” *Adv. Intell. Syst. Comput.*, vol. 810, pp. 855–864, 2018,

[4] T. Ahonen, A. Hadid, and M. Pietikäinen, “Face description with local binary patterns: Application to face recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, 2006,

[5] D. G. Lowe, “Object recognition from local scale-invariant features,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2, pp. 1150–1157, 1999,

[6] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004,

[7] G. Zhang, F. Gao, C. Liu, W. Liu, and H. Yuan, “A pedestrian detection method based on SVM classifier and optimized Histograms of Oriented Gradients feature,” *Proc. - 2010 6th Int. Conf. Nat. Comput. ICNC 2010*, vol. 6, no. Icnc, pp. 3257–3260, 2010,

[8] C, Guang-tao, X,CHEN, and Z,GUO. "Pedestrian detection method of vision based on HOG features.", *Transducer and Microsystem Technologies* vol, 30, pp, 479-488 Jun2011: 7 .

[9] Y. Xue-qin, LI. Xiao-hua, and Z, Ji-liu. “Pedestrian detection method based on edge symmetry and HOG”, *Computer Engineering*, ,vol, 38.5: pp, 179-182, Jun 2012

[10] H. Bay, T. Tuytelaars, and L. Van Gool, “Computer Vision – ECCV 2006 SURF: Speeded Up Robust Features,” *Comput. Vis. – ECCV 2006*, vol. 3951, pp. 404-417–417, 2006, [Online]. Available: http://www.springerlink.com/content/e580h2k58434p02k/

[11] L. Juan and G. Luo, “A Comparison of SIFT, PCA-SIFT and SURF Luo,” *Urban Stud.*, vol. 38, no. 1, pp. 207–223, 2007,

[12] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “ORB: An efficient alternative to SIFT or SURF,” *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 2564–2571, 2011

[13] Y, Li. X, Xie . C, Jia, and Z, Guofu. “Rapid moving object detection algorithm based on ORB features ”,2013, Journal of Electronic Measurement and Instrument 5 (2013).

[14] R. Girshick, J. Donahue, T. Darrell, J. Malik, and U. C. Berkeley, “Rich feature hierarchies for accurate object detection and semantic segmentation,” pp. 580–587, 2014,

[15] Girshick, Ross. "Fast r-cnn." Proceedings of the IEEE international conference on computer vision, Santiago, Chile, 2015, pp, 2380-7504

[16] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017,

[17] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once : Unified , Real-Time Object Detection,” 2016,

[18] J. Redmon and A. Farhadi, “YOLO9000: Better, faster, stronger,” *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 6517–6525, 2017,

[19] D, Anguelov. et al, “Single Shot MultiBox Detector”. 2016 ,Lecture Notes in Computer Science vol, 9905, pp. 21–37, Sep 2016.

[20] F. Sultana, A. Sufian, and P. Dutta, “A Review of Object Detection Models based on Convolutional Neural Network,” 2020.

[21] C. O. Laura, P. Hofmann, and G. Systeme, “AUTOMATIC DETECTION OF THE NASAL CAVITIES AND PARANASAL SINUSES USING DEEP NEURAL NETWORKS Stefan Wesarg Technische Universit ¨ Aachen University of Applied Sciences , Aachen , Germany,” no. Isbi, pp. 1154–1157, 2019.

[22] Z. Zhao, P. Zheng, S. Xu, and X. Wu, “Object Detection With Deep Learning : A Review,” vol. 30, no. 11, pp. 3212–3232,Dec, 2019.

[23] W. Zhiqiang and L. Jun, “A Review of Object Detection Based on Convolutional Neural Network,” pp. 11104–11109,Jan 2017.

[24] M. O. Lawal, “Tomato detection based on modified YOLOv3 framework,” pp. 1–12, 2021.

[25] J. Lu *et al.*, “A Vehicle Detection Method for Aerial Image Based on YOLO,” pp. 98–107, 2018,

[26] R. K., Rohit Kundu YOLO versus. California, Riverside other detectors is discussed in https://www.v7labs.com/blog/yolo-object-detection#h2 in Jan, 2023.

[27] J. Lee, “YOLO with adaptive frame control for real - time object detection applications,” *Multimed. Tools Appl.*, pp. 36375–36396, 2022,

[28] S, Afshari . A, BenTaieb , and Ghassan Hamarneh. “Automatic localization of normal active organs in 3D PET scans”. 2018 Computerized Medical Imaging and Graphics, 70:111–118, Dec 2018.

[29] M, Marty Ahrens, National Fire Protection Association. US, fires vehicle news is discussed in https://www.nfpa.org/News-and-Research in Dec, 2020.

[30] M. J. Menten *et al.*, “The impact of 2D cine MR imaging parameters on automated tumor and organ localization for MR-guided real-time adaptive radiotherapy,” *Phys. Med. Biol.*, vol. 63, no. 23, 2018, doi: 10.1088/1361-6560/aae74d.

[31] Y. Lee and Y. Kim, “Comparison of CNN and YOLO for Object Detection,” vol. 19, no. 1, pp. 85–92, Jul, 2020.

[32] S. R. S, J. George, and S. Skaria, “Using YOLO based deep learning network for real time detection and localization of lung nodules from low dose CT scans,” no. February 2018, 2023, doi: 10.1117/12.2293699.

[33] Y. Zhao, L. Gong, Y. Huang, and C. Liu, “A review of key techniques of vision-based control for harvesting robot,” *Comput. Electron. Agric.*, vol. 127, pp. 311–323, 2016, doi: 10.1016/j.compag.2016.06.022.

[34] X. Wei, K. Jia, J. Lan, Y. Li, Y. Zeng, and C. Wang, “Optik Automatic method of fruit object extraction under complex agricultural background for vision system of fruit picking robot,” *Opt. - Int. J. Light Electron Opt.*, vol. 125, no. 19, pp. 5684–5689, 2014, doi: 10.1016/j.ijleo.2014.07.001.

[35] H. Yin, Y. Chai, S. X. Yang, and G. S. Mittal, “Ripe Tomato Extraction For A Harvesting Robotic System,” no. October, pp. 2984–2989, 2009.

[36] E. E. Kelman and R. Linker, “ScienceDirect Vision-based localisation of mature apples in tree images using convexity,” *Biosyst. Eng.*, vol. 118, no.2000, pp.174–185, 2014,

[37] U. Feature and I. Fusion, “Robust Tomato Recognition for Robotic Harvesting Using Feature Images Fusion,” 2016, doi: 10.3390/s16020173.

[38] J. M. Clanton, D. M. Bevly, A. S. Hodel, and S. Member, “A Low-Cost Solution for an Integrated Multisensor Lane Departure Warning System,” vol. 10, no. 1, pp. 47–59, 2009.

[39] Q. Li, L. Chen, M. Li, S. Shaw, and A. Nüchter, “A Sensor-Fusion Drivable-Region and Lane-Detection System for Autonomous Vehicle Navigation in Challenging Road Scenarios,” vol. 63, no. 2, pp. 540–555, 2014.

[40] J. Stanisz and K. Lis, “Hardware-software implementation of car detection system based on LiDAR sensor data - a demo,” no. October, 2019.

[41] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017, doi: 10.1145/3065386.

[42] University of Oxford, .“Pascal Visual Object Classes Homepage”.2020 ,[Online]. Available: http://host.robots.ox.ac.uk/ pascal/VOC/index.html.

[43] A. Lemay, “Kidney Recognition in CT Using YOLOv3, Polytechnique Montréal, ” no. Figure 1, pp. 1–5, Oct, 2019, [Online]. Available: http://arxiv.org/abs/1910.01268

[44] M. Maity, S. Banerjee, and S. Sinha Chaudhuri, “Faster R-CNN and YOLO based Vehicle detection: A Survey,” *Proc. - 5th Int. Conf. Comput. Methodol. Commun. ICCMC 2021*, no. Iccmc, pp. 1442–1447, 2021, doi: 10.1109/ICCMC51019.2021.9418274.