# OBJECT DETECTION VIA DEEP LEARNING AND NEURAL NETWORK

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF DEGREE
OF
MASTER OF TECHNOLOGY
IN
**COMPUTER SCIENCE & ENGINEERING**

Submitted by:

**ARCHI BAKAWLE**
**2K20/CSE/06**

Under the supervision of
**Dr. Vinod kumar**
( Professor)

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

MAY, 2022

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi - 110042

## CANDIDATE'S DECLARATION

I, **Archi Bakawle** , Roll No. 2K20/CSE/06 student of M. Tech (Computer Science and Engineering), hereby declare that the project Dissertation titled **"Object Detection via Deep Learning And Neural Network"** which is submitted by me to the Department of Computer Science & Engineering, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of and Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Archi

Place:Delhi

Archi Bakawle

Date:

2k20/CSE/06

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi - 110042

## CERTIFICATE

I hereby certify that the Project Dissertation titled **"Object Detection via Deep Learning And Neural Network"** which is submitted by Archi Bakawle, 2K20/CSE/06 Department of Computer Science & Engineering, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the students under my supervision. To the best of my knowledge, this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Vinod Kumar

Place: Delhi | Dr. Vinod Kumar

Date: | Professor

Head of CSE Department

# ACKNOWLEDGMENT

The success of this project requires the assistance and input of numerous people and the organization. I am grateful to everyone who helped in shaping the result of the project.

I express my sincere thanks to **Dr. Vinod Kumar**, my project guide, for providing me with the opportunity to undertake this project under his guidance. His constant support and encouragement have made me realize that it is the process of learning which weighs more than the end result. I am highly indebted to the panel faculties during all the progress evaluations for their guidance, constant supervision and for motivating me to complete my work. They helped me throughout with new ideas, provided information necessary and pushed me to complete the work.

I also thank all my fellow students and my family for their continued support.

Archi

ARCHI BAKAWLE

2K18/CSE/20

## ABSTRACT

DL approaches to Object Detection (OD) have attracted a lot of attention from researchers because of their implied strength in overcoming the drawbacks of traditional approaches that rely on handcrafted characteristics. DL algorithms have made major advances in object recognition during the previous few years. This paper discusses the most recent and effective DL framework for object recognition. Visual recognition systems, which include picture categorization, localization, and detection, are at the heart of all of these applications and have gathered a lot of research attention. These visual identification algorithms have achieved extraordinary performance due to considerable advancements in neural networks, particularly deep learning. OD is one of these sectors where computer vision has had a lot of success. The role of DL methods based on YoloWingNet for OD is proposed in this research.

In computer vision, classifying and detecting various items in an image is a crucial ability. Robust and efficient object identification is a critical method for engaging with one's surroundings. Humans utilize a technique known as a visual focus to swiftly determine which areas of an image require detailed processing and which can be avoided. However, identifying an object and its precise location in an image is a challenging problem for a machine. This paper studies features and methods of object detection and the algorithms related to object detection using deep learning.

# CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATIONS

1. NLP: Natural Language Processing
2. CNN: Convolutional Neural Network
3. OD: Object Detection
4. CV: Computer Vision
5. DL: Deep Learning
6. WED: Weighted Edit Distance
7. YOLO: You Look Only Once
8. RMSE: Root Mean Square Error
9. mAP: Mean Average Precision
10. CL: Convolutional Layer
11. PL: Pooling Layer
12. FM: Factorization Machine

# CHAPTER 1 : INTRODUCTION

## 1.1 GENERAL INTRODUCTION

Object identification and tracking in a variety of computer vision fields of research, as well as additional aspects in traffic detection, car navigation, and interpersonal interactions.It is connected to computer visualization and image processing. Face detection, face identification, and video object detection are just a few of the many applications in object detection. Monitoring the movement of the ball, following the ball during a match, and identifying a person in a video are among the possibilities. In some scenarios and camera angles, the object detection system detects the presence or absence of things. The many fields of object detection are categorized into specific and conceptual groups depending on the varied purposes. Detecting objects using different models, either explicitly or intuitively. Depending on the approach, the elements may be varied. The object selection is based on assumption and the item selection is based on similarity. The object detection method is a good fit for the processing.

Machines can now perform a significant part in executing routine daily tasks, owing to the tremendous advancements in computing intelligence. For humans, visible objects categorization and identification are frequent and automatic biological visual processing activities, but for machines, they are complex. It's simple to emulate due to the strong expressive power of object photos in the same class in a variety of viewing settings.The amount of graphical data in the digitized library is growing. Image processing algorithms that can autonomously identify interesting contexts are required to manage and evaluate this huge volume of image data. Among the most implications for visual detection, the task is the objects in the images. The foundation of a successful OD system is good image feature specifications.

According to the previous findings gained in the areas of image classification, OD, and natural language processing(NLP), DL technology became a buzzword worldwide. The development of DL is due to two factors: massive dataset distribution and efficient Graphics

Processing Units. Both of these needs have already been met in this era, as DL requires vast datasets and significant facilities for training. The goal of object detection DL approaches based on convolutional neural networks (CNNs).

## 1.2 DIFFERENT FEATURES OF OBJECT DETECTION

Identification, the selection of numerous characteristic aspects can limit the computer's work accessibility. When several algorithms are used to track distinct features, the composition of those features is calculated in various processes .

- **Color-** The characteristic of the computing device that is utilized to depict histogram appearances. The aspects of color visualizations for tracking are the ones with the most variety. The characteristics of color are tracking of a major problem that recognizes variations in illumination.

- **Histogram of gradients-** The HOG feature is the most widely utilized feature for detecting human bodies. The histogram feature's processes are based on the picture's grid connection unit. As a result, the optical deflections are influenced by geometric differences. Furthermore, the sampled location and local optimization keep the body and posture straight. These motions do not affect the detection process, which is the primary reason for the HOG characteristic in human detection.

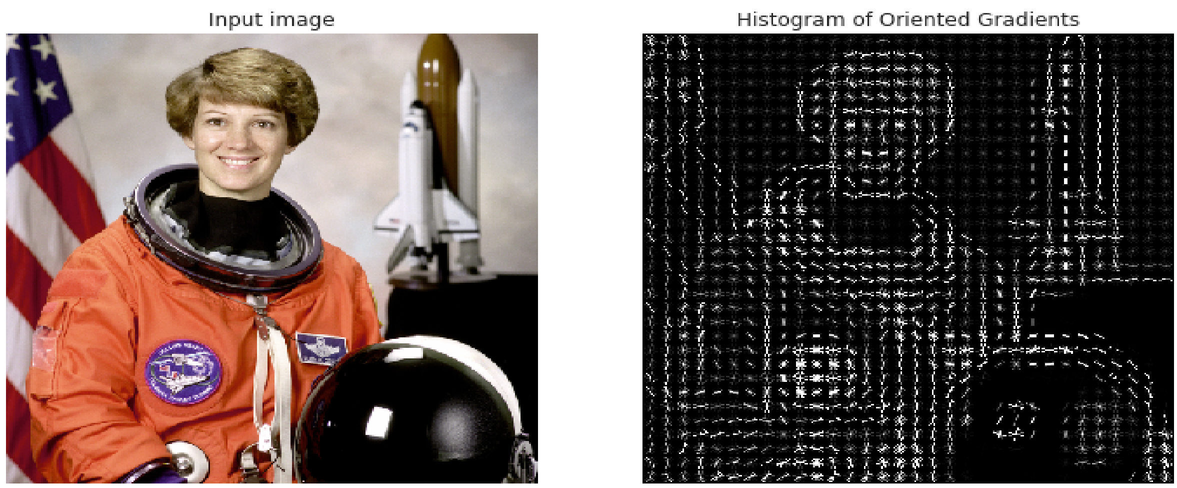



Figure 1.1 H.O.G.

- **Edges-** During  identification of  OD, the image intensities' bounds may shift. The OD has a distinct feature from the color features technique.

- **Optical Flow-**Every pixel in the region is recognized by the displacement vector. The displacement vector is the one that specifies how each pixel in an image is transacted. In motion-based segmentation and tracking applications, optical flow is commonly employed as a characteristic. It's a complex field of displaced vectors that determines how each pixel in an area is translated. It's calculated with the brightness constraint, which specifies that pixels in successive frames have the same bright

## 1.3 METHODS OF OBJECT DETECTION

**Template Based-** Using the test image, the little sections of the picture can be recognized in this approach. This method is also known as template matching. The integrity of the multi-robot elements of the image is utilized as a quality assurance image. To identify the templates, the techniques are adopted using a search image.The spots with the highest scores are examined. Spatial filtration is the approach, and the filter mask is the template.

**Part Based-** The bundle of deformable configurations can be used to depict an item. Each portion of the models is individually constructed with a distorted configuration, which is indicated by the connections among pairs of sections. These models, which are useful for generic vision tasks, influence the visual aspect of subjective descriptors.

**Region-Based-** The input image is transformed into a directed graph using a set of rules specified by algorithms. The strategy entails traversing the graph after it has been preserved, which will reduce the graph's computing time. This approach is put to the test using a particular dataset.

**Contour Based-** The webcams are used to locate the robots and identify them by approaching items. Cameras are used to capture images to recognize items and place them in their ultimate positions. This procedure is divided into 2 stages. The 1st stage describes the polynomial shapes and is concerned with the position of the specific objects. Mostly every type of form has a strong relationship between the holes. Designs, on either hand, identify the type of object and compute the relative orientation.

**Appearance Based-** This technique uses a 3D recognition system to identify objects in the presence of interference and distortion. The characteristics are employed to give the photos and situations their appearance. The major classes that are used in the two-dimensional views of things are local and global solutions.

## 1.5 METHODOLOGIES

**Deep Learning methods**

The various deep learning approaches used in object detection are generally divided into three categories. Unsupervised feature learning is the first class, in which the theory and principles of deep learning are used solely for feature extraction. Dependent on the operations, these features will be provided to extremely simple machine learning algorithms for executing duties like categorization, detection, or tracking. When considerable quantities of labeled data are available, end-to-end learning is utilized to optimize the feature extractor and classifier components of the existing model.

**A. Neural Network (NN)-**

The neural network (NN) is made up of a sequence of interconnections that reflect brain activity. Each node has a weighted relationship with numerous different nodes in nearby layers. The data and weights from the individual nodes, as well as the basic capabilities to compute expected output, are used by the node. Neural systems exist in a variety of shapes and sizes. Considering the problem's complexities, the user must choose the NN architecture, which includes numerous hidden layers, the number of nodes, and their connection in certain hidden levels. It can learn and model nonlinear and dynamic interactions, which is critical because many input-output linkages in actual situations are nonlinear or complicated. It may infer hidden connections on insignificant information after learning from the fundamental information sources and their linkages, therefore impacting the model, to sum it up and anticipate hidden information[14].

**B. Convolutional Neural Network (CNN)-** It is a feedforward NN that was first utilized in computer vision in 1989. It is inspired by the visual cortex of animals. The three layers in CNN are input, feature selection, and classification[15]. Raw data is embedded in the input layer, a feature map is learned and produced by the convolution layer, and practical properties are extracted by the neural network. Finally, these attributes are given into a categorization layer that is fully integrated[16].

CNN has achieved flawless results in a variety of fields, such as NLP and SA, and hence for the added benefits. CNN uses a small number of variables that need little time to train. They've gained a lot of attraction in SA subsequently, can learn contextual qualities but are limited in long-term relationships. When constructed in a large number of samples, CNNs achieve adequate accuracy. However, this is not always the case, and the development of RNNs can address the problem of long-term dependencies[17].

**C. Restricted Boltzmann Machines (RBM)**

Ackley et al. originally proposed an RBM which is a little dissimilar from a stochastic Boltzmann machine. A random field with several layers of the visible stochastic layer or the hidden layer in a particular Markov region is generally believed to be RBM. The RBM is in the form of a bipartisan graph as shown in the figure below. This shows clearly that information flows bi-directionally during preparation or network use along with the same weight in both directions.

**D. Auto Encoders (AE)**

As a type of deep feedforward network, autoencoders may learn features from unlabeled data in an unsupervised manner. The output of an AE is set to be equal to its input, and the network is then trained to reduce the error between them. AE has been used as a pre-training approach to extract representative spatiotemporal characteristics in a variety of studies recently.

- To guide the AE and receive the learned data, use input data.

**E. Deep Belief Networks (DBNs)**

DBN was among the first non-convolutional networks to allow deep architecture training. The current deep learning revolution began with its release in 2006. A DBN consists of several RBM layers and one BP layer, and it can be used to solve classification and prediction difficulties.

**G. You Only Look Once (YOLO)-**

YOLO is a recent invention in the world of object detection. The RCNN sequence system has several drawbacks, including the fact that the entire network cannot perform end-to-end training, the intermediary training procedure necessitates a large amount of memory to retain some characteristics, and the computation performance is slow, among others. The YOLO algorithm proposes a novel concept that converts the object detection issue into a regression problem. It returns the target bounding box and its categorization categories in many regions of images given an input image.

YOLO detection speed can reach 155 frames per second. Unlike other object detection algorithms, YOLO takes a full picture as input, making appropriate use of the findings of the analysis during detection and having a low chance of predicting the inaccurate object information in the background. YOLO can learn highly generic traits and is more mobile. However, object detection accuracy isn't great, and it's simple to make positioning mistakes. Because a Grid cell can only anticipate two items at a time, it is ineffective at recognizing little ones[24].

# CHAPTER 2 : LITERATURE REVIEW

The advancement in computer technology hastens deep learning's progress and ushers in a new era of artificial intelligence. Deep learning has been steadily used to image identification and recognition in recent times, which has tremendously aided research in this subject. The efficiency of computer processing is rapidly increasing as the data era progresses. Object recognition models based on deep learning on a powerful computational platform have considerably increased detection capability and real-time performance, owing to the rapid advancements in computer hardware. Furthermore, these models have been widely utilized in diagnostic imaging, security prevention and control, image processing, and other domains. Automatic driving and other areas have significant research and application value. Recent OD methods have enormous storage requirements, high computer resource demands, and a large number of iterations, making them challenging to apply on platforms with poor processing capability and storage space. Further study is needed on lightweight networks with high real-time performance and accuracy, and lightweight target detection algorithms are still a research hotspot.

| **Author & Year** | **Title** | **Method Used** | **Evaluation Results** | **References** |
|---|---|---|---|---|
| Mohana & HV Ravish Aradhya | Object Detection and Tracking using Deep Learning and Artificial Intelligence for | CNN & YOLOv3 | Accuracy- 96% | [6] |

| | | | | |
|---|---|---|---|---|
| | Video Surveillance Applications | | | |
| Sheng Ding & Kun Zhao | Research on Daily Objects Detection Based on Deep Neural Network | Faster-RCNN-ResNet | Accuracy= 72.3 | [7] |
| Andrew G. Howard et al. [2017] | MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications | Faster-RCNN-VGG | mAP- 25.7% | [8] |
| Teena Sharma et al. | Deep Learning-Based Object Detection and Scene Perception under Bad Weather Conditions. | YOLOv5 | Accuracy- 72% | [9] |
| Faisal Saeed et al. [2021] | A robust approach for industrial small-object detection using an improved faster regional convolutional neural network | Enhanced Faster RCNN model with Bicubic Interpolation. | Accuracy- 90.57 | [10] |
| Vijayakumar Ponnusamy et al. [2021] | Deep Learning-Based X-Ray Baggage Hazardous Object Detection – An FPGA Implementation. | YOLO | Accuracy-98.9% | [11] |

| | | | | |
|---|---|---|---|---|
| Daulet Baimukashev et al. [2019] | Deep Learning-Based Object Recognition Using Physically-Realistic Synthetic Depth Scenes. | Faster-RCNN | Accuracy- 93.5% | [12] |
| Samet Akcay & Toby P. Breckon [2020] | An evaluation of region-based object detection strategies within x-ray baggage security imagery. | R-FCN with ResNet-101 | mAP- 96.3 | [13] |

Table 2.1 Related Work

**Meian Li [2021]-** Existing OD models have massive storage requirements, powerful computational resource demands, and a large number of components, making them challenging to apply on a platform with poor computer performance and limited capacity. This research offers a new network design of mobilenetv2-yolov5s that combines the compact network mobilenetv2 with yolo 5s to decrease the model complexity and increase the recognition rate. The upgraded yolo 5s provide a better identification effect than existing target detection techniques. On the MS coco dataset, the mobile etv2-yolo 5s network has an mAP value of 55.1. The system's detection speed is 31 frames per second, which is 25 frames per second faster than yolo 5s.

**P.Devaki [2019]-** The goal of this paper is to use OD techniques to help a blind person. It enables visually sighted persons to walk freely by allowing them to know about the items around them. On a Raspberry PI 3 with OpenCV libraries, a model has been implemented, and promising results have been obtained. A complete review of object detection utilizing (RCNN) based learning programs for practical applications is discussed in this work. This study examines the various approaches for detecting objects and follows through the detection phase, along with a DNN for SSD constructed using the Caffe model.

**Rohini Goel et al. [2019]-** This article addresses some of the most current and effective DL frameworks for object identification. It is provided an up-to-date analysis of recently evolved DNN -based object recognition algorithms. The many benchmark datasets which are utilized to evaluate performance are also covered. The visual recognition approach's applicability for various sorts of objects (such as faces, buildings, plants, and so on) is also discussed. They end by discussing the advantages and disadvantages of existing approaches, as well as their future potential in this field[4].

**Mohana and HV Ravish Aradhya [2019]-** On the KITTI and COCO datasets, the CNN model is created for single OD on the urban vehicles dataset and YOLOv3 for numerous object recognition.The peculiarity of other networks like DarkNet is emphasized in this study. On a dataset of urban vehicles, effective detection and recognition can be seen. Real-time, exact, accurate detections are provided by the techniques, which are ideal for real-time traffic applications.

# CHAPTER 3 : TERMINOLOGIES

# ,TOOLS AND TECHNOLOGY USED

### 3.1 PYTHON

It is a high-quality programming language with excellent comprehension, interactivity, and artifacts. This makes extensive use of English keywords, whereas others make use of punctuation. In comparison to other languages, it contains fewer syntactic structures.

· Interpreted: The interpreter runs Python at runtime. Until you run it, you do not have to develop your software. PERL & PHP are close.

· Interactive: Also, in a Python prompt, you can directly communicate with an interpreter for writing programs.

· Object-oriented: Python promotes object-oriented application encapsulation type or technique in modules.

· Beginner's Software: The software Python provides beginning programmers with a broad variety of queries for easy WWW application word processing for sports.

More precisely, It allows the user to concentrate more quickly to solve domain problems instead of struggling with the complexities of how a machine works in comparison to other programming languages like C, Fortran, or Java. by the following characteristics Python achieves this objective:

· Python is a language of high level, which means it sums up the technical details of computers. For example, Python does not make the users too much worry about computer storage or correct variables definitions and makes sure what the programmer is trying to convey. To get closer to English prosthetics or math, a high-level language can also be conveyed. The simple, "small ceremonies" character of Python is suitable for literary programming.

· Python is a common language that is capable, instead of being specialized in a certain area like statistical analysis, of being used for each problem. Python for artificial and statistical analyzes, for example, Python can be used, as the UCAR scientist described earlier, Python can be used for several heterogeneous workflows.

· The meaning of Python was that code analysis can happen directly instead of having to undergo a time-consuming compilation and execution cycle, which thereby speeds up the processes of thinking and experimenting. IPython is an interactive version of Fernando Perez's Python language as well. These environments are excellent for quick code prototypes or simple experiments with new ideas.

· Python has the main library but many third-party implementations, which provides an extensive range of popular codebases and models of problem-solving.

· The programmers can quickly find solutions and sample code for problems with the help of Google and Stackoverflow.

· Python has several users.

Python is a fast, modern programming language for computers. It contains some commonalities to the Fortran language, one of the earliest, although far stronger than the Fortran language. Without defining variables (i.e. specifying the forms implicitly) Python permits you to use them, relying on indentation as a control structure. You are not obligated to create Python classes (as opposed to Java), but you can do so as required.

Python is free software. It has been developed by Guido van Rossum. You can get Python without spending as much time as in "free beer". But in other important ways, Python is also free, for instance, free to copy it as often as you want and free to study and change the source code. A global movement was launched by Richard Stallman in 1983, behind the idea of free software.

For mathematical calculations, Python is a good option, since we can write code easily, quickly check as well as its syntax resembles how mathematical ideas are expressed in the mathematical literature. You will learn a key tool for many software developers by learning Python. Real-world Python implementations:

· GUI-Based Desktop Applications

**3.1.1 Python Environment Variables**

· **PYTHONPATH:** The Python source library directory or Python source code directories will be included. Sometimes, PYTHONPATH is predetermined by Python.

· **PYTHONCASEOK:** In Windows, the first case insensitive match is to be found in an import declaration. Set a value for this variable to be allowed.

· **PYTHONHOME:** It is a quest path for an alternate module. Used to make swapping module libraries simple it is incorporated within PYTHONSTARTUP or PYTHONPATH directories.

### 3.2 JUPYTER

The Jupyter Notebook is a dominant instrument for interactively creating or viewing data science projects. In a single document, a notebook combines code or its output with an image, text, mathematics, or further media. Instinctual systems encourage reiterative but quick expansion, making notebooks ever more likely to be a candidate at the heart of modern data science, technology & innovation. Worst of all, the project Jupyter open source is completely free.

The Jupyter project is an inheritor to the preceding IPython Notebook, released as a reference for the first time in 2010. But you can also use several different programming languages in Jupyter Notebooks, Python is the most widely used case in the above article. You can also use pip only when you are an experienced Python user or prefer manual package management.

#### 3.2.1 Running Jupyter

On Windows, you could run Jupyter via Anaconda ads to your start menus to create a new tab that looks like some kind of screenshot in your default browser.


Figure 3.1 Jupyter Notebook

This is a dashboard for your Jupyter Journals. This was made for you. Think of this as the starter to search your journals, edit them, or build them. In the top right of the page, click the "New" drop-down button or pick "Python 3").

You'll see Untitled. ipynb and you'll see a green word which means your notebook is going to run when you switch back to the dashboard.

### 3.2.2 The Notebook Interface

Since you have an open notebook before you, I hope that the interface doesn't look completely alien. In reality, Jupyter is a professional word processor. Why are you not taking a look? See menus, to display the list of commands, which are the small button that has either symbol, in particular, in a few minutes (or Ctrl + Shift + P).

You will note 2 very popular words, probably novel to you: cells or kernels are both important to understand Jupyter or to what it renders as much as a word processor. These ideas are fortunately not problematic to understand.

· "Program processor" is a kernel that performs a code of notebook text.

· The cell is a container that shows the transcript of a notebook or the code that the notebook kernel executes.

### 3.2.3 Cells

A little earlier we'll go back to kernels, but first, let's get cells under control. The body of a notebook consists of cells. Box with green contour is an empty cell in a screenshot of a notebook in the above section. Two major types of cells are covered by us:

· A code cell comprises code to run in the kernel or its output is shown in the following.

· A Markdown cell produces Markdown-formatted text and displays its output on the spot.

A code cell is always the first cell in a new notebook. Let us try it with the classic example of a hello world. Tap on the run button into a cell in the toolbar above or press Ctrl + Enter. This should be an outcome

```
print('Hello World!')
Hello World!
```

Figure 3.2 Different cells

After a cell is going, its output is shown below, as well as a label to the left has been modified from In[] to In[1]. The code cell output is included in the text, so you can see that in this object. The differentiation between code or markdown cells is always possible because coding cells are that label on the right & Markdown cells do not. "in" of the label is brief for "data," whereas when label no. is on a kernel, the first cell is running. Run cell again & label switches to In[2], the cell being 2 and most important in the kernel now. When we study kernels more closely, you'll find out why this is so beneficial. To establish your first new code cell and use the code below

This cell does not generate a single output but runs for 3 seconds. Remember how Jupyter means that the cell now runs when its label is modified [*]. The output of a cell type is dependent on all text information that is printed explicitly during cell running and on the value of the last line in a cell, whether it's a lonely attribute, function call, or rather. For instance,

### 3.2.4 Data Type Conversion

There are multiple built-in functions for transferring between data types. These functions restore a new object that represents value converted. The few are listed below.

| S.no. | Functions & Descriptions |
|---|---|
| 1. | int(x [,base]); Conversion of x to an integer, base stipulates that x is string. |
| 2. | float(x); Transforms x into a set of floating points. |
| 3. | complex(real [,imag]); Produces an interesting no. |
| 4. | str(x); Conversion of object x to a string file. |
| 5. | repr(x); Turns object x to a string expression. |
| 6. | eval(str); Returns an object to evaluate a string. |
| 7. | tuple(s); Turns s into tuple. |
| 8. | list(s); Turns s into list. |
| 9. | set(s); Turns s into set |
| 10. | dict(d); Make a dictionary. d shall be a tuple (key, value) sequence. |

Table 3.2- Functions & Descriptions

# CHAPTER 4 : PROPOSED WORK

## 4.1 PROPOSED METHOD

### A. Dataset Collection

PASCAL VOC is a set of image databases for object class detection that has been standardized. It offers a standardized toolset for interacting with data sets and descriptions. It allows for the assessment and comparison of various methodologies, as well as the assessment of object class recognition performance. Collect the dataset (there are almost 24,000 photos in this dataset).

### B. Convolutional Neural Network

Convolution is assimilation that illustrates how one functionality intersects with another by multiplying two units together. Pooling layers (PL) are used to construct simplified FM, that reduces the network's spatial complexities. FM is formed as a consequence of repeating this technique for the desired set of parameters. Ultimately, this FM is merged with fully linked layers to provide an image detection output with sentiment polarity for expected classifiers.

### Convolutional Layer (CL)

Filters and FM are included in the CL. Filters are processors that work on a certain layer. These filters are not interchangeable. They accept pixel values as input and produce an FM as a result. One filter layer's outcome is an FM.

### Pooling layer (PL)

To decrease dimensions, a PL is used. To generalize characteristics learned from prior FM, pooling layers are added after one or two CL. This reduces the risk of overloading throughout the training process.

### Fully connected layer (FCL)

After collecting and aggregating features from the CL and pooling subsequently, the FCL is utilized to allocate the feature to class probabilities. Linear activation functions or softmax activation functions are used in these layers.

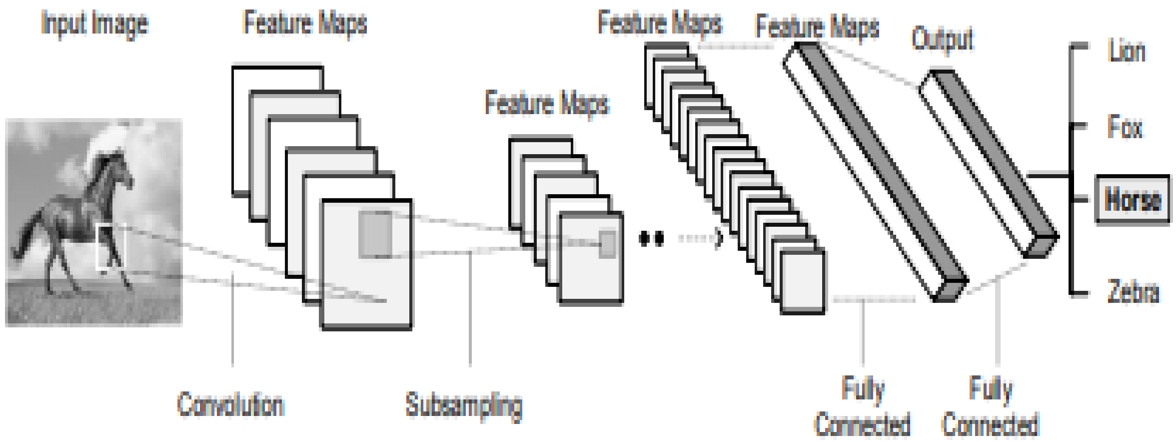



Fig4.1 – CNN architecture

### C. You Only Look Once (YOLO)

YOLO is a single-stage object detector that was inspired by Faster RCNN. Its primary use is in the identification of real-time images. On a Titan X GPU, the YOLO system works at 45 frames/sec, whereas Faster RCNN runs at 0.5 and 5 frames/sec, respectively.

The image is segmented into (g*g) grids. Object categorization is based on features derived from each grid cell. B bounded boxes (BB) are predicted for each grid cell, and C class possibilities for C object categories are calculated for every box. For each BB, two components are used: first, the probability (P) of the BB is used to determine if it corresponds to any item, and then the IoU among the underlying data and the BB is used to determine how correctly the BB includes that object. The YOLO network has 24 CL and 2 FCL. Object localization is a problem for YOLO, which limits its detection performance.

For OD, the YOLO uses DL and CNN. Every image just needs to be "seen" once. As a result, YOLO is among the quickest detection techniques available. It can identify things in real-time at 30 frames/sec. The image is partitioned into an S*S grid for detection (left image). Each cell will estimate the number of feasible bounding boxes as well as the possibility of each one. S*S*N boxes are computed as a result of this.

## 4.2 SUMMARY OF PROPOSED ALGORITHM

1. Defining all of YoloWingNet's layers, which are based on yolov4 and include some customizable layers, new layers, and batch normalization.

2. Defining the WeightReader class for reading the model's weights.

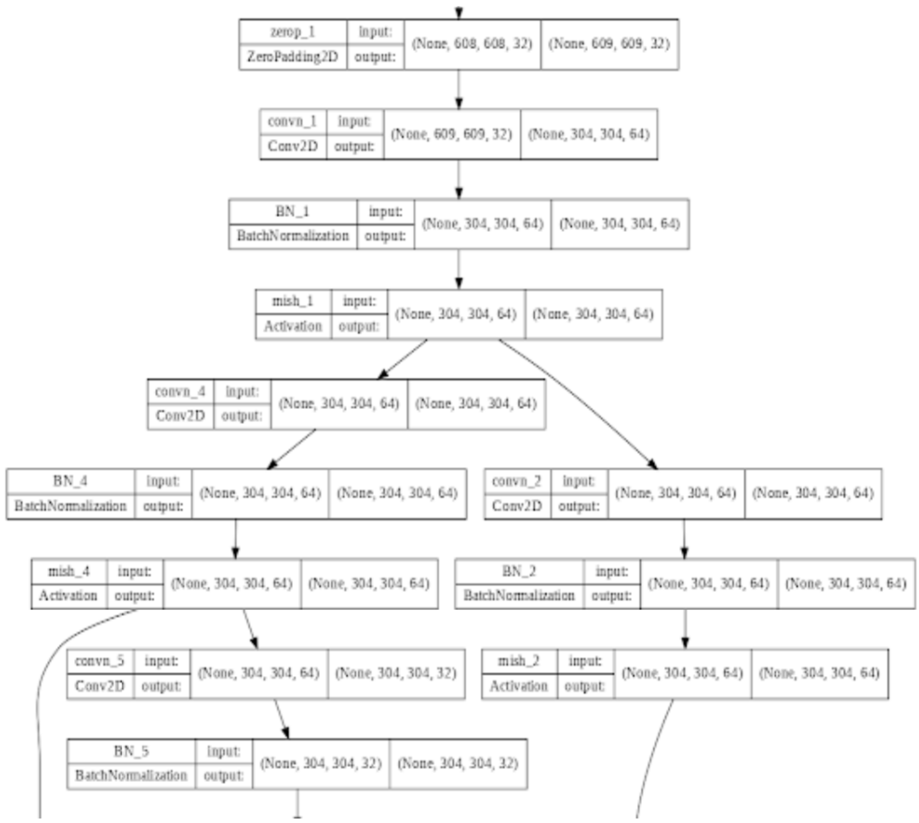
3. Adding more functions to build bounding boxes, find, load images, and so on.

4. Using the annotation file to read and extract annotations.

5. Defining the image training and validation functions.

6. Creating a custom loss function, which aids in adjusting the predicted annotation's x and y axes, class likelihood, class abjectness, coefficients, and returning the sum of the losses. In several locations.

## 4.3 MODEL ARCHITECTURE



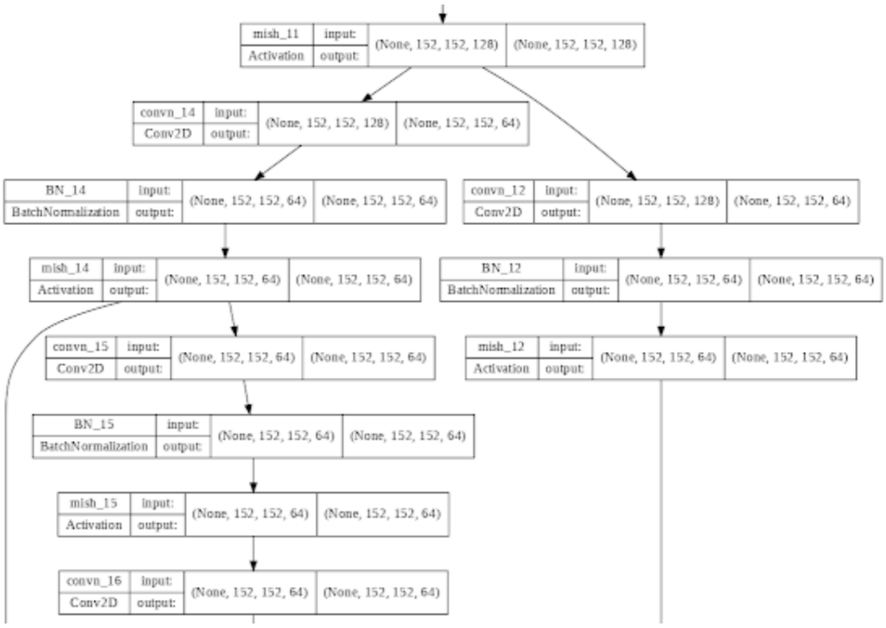



Figure 4.2 Model Architecture

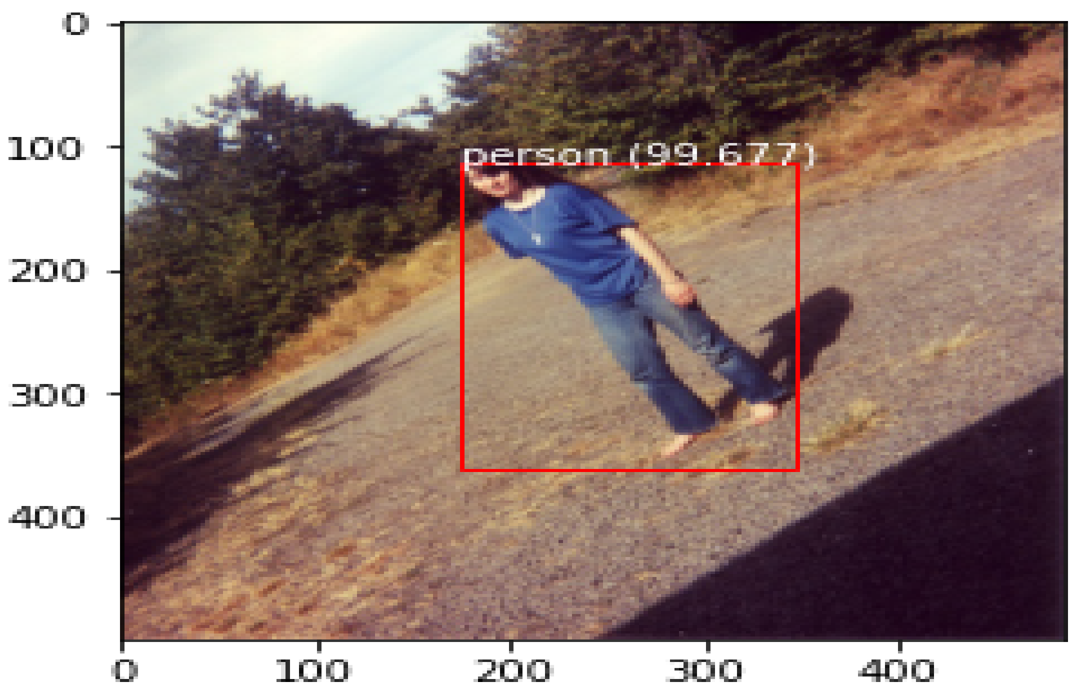



Figure 4.3

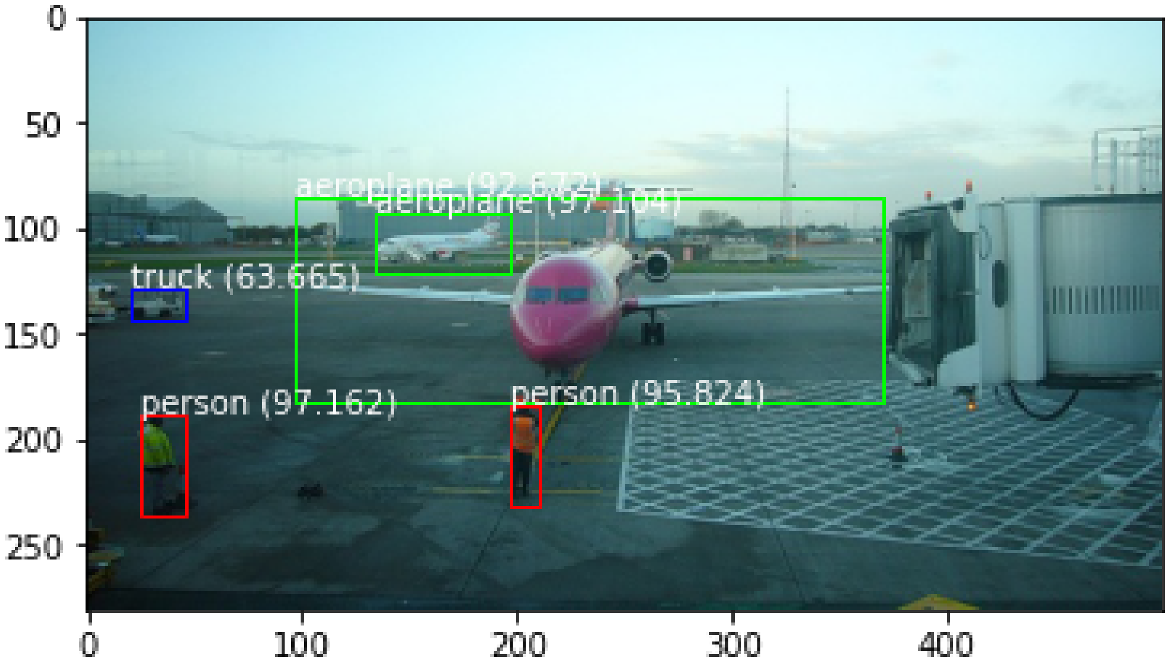

aeroplane (92.672)
aeroplane (97.104)
truck (63.665)
person (97.162)
person (95.824)
0
50
100
150
200
250
0
100
200
300
400

Figure 4.4

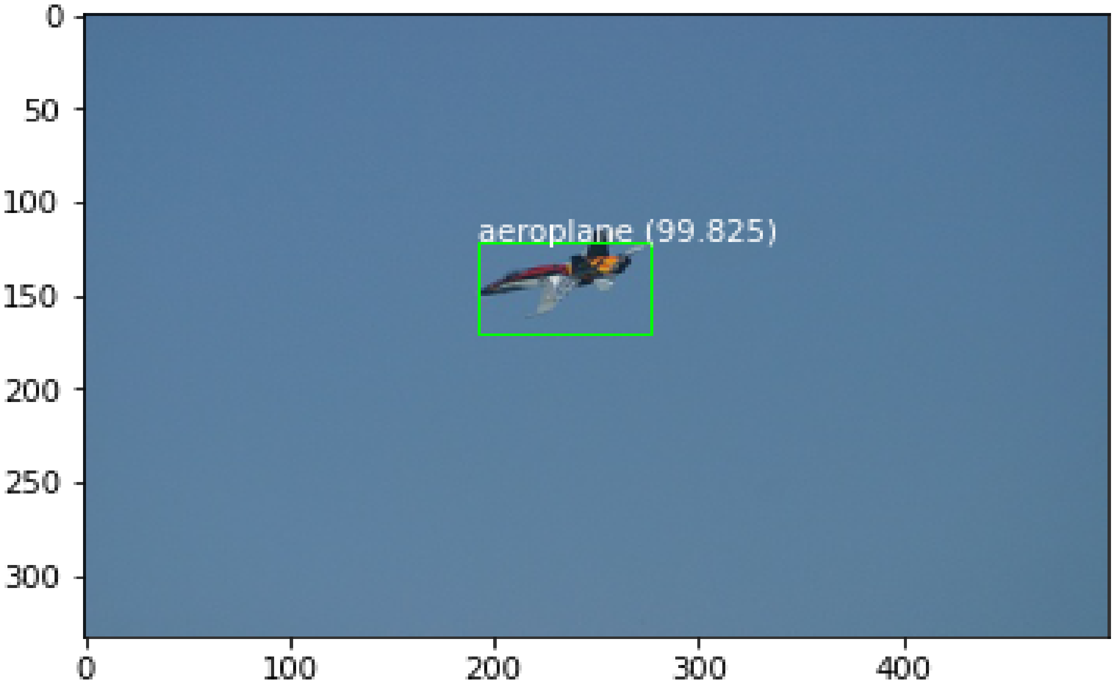



Figure 4.4

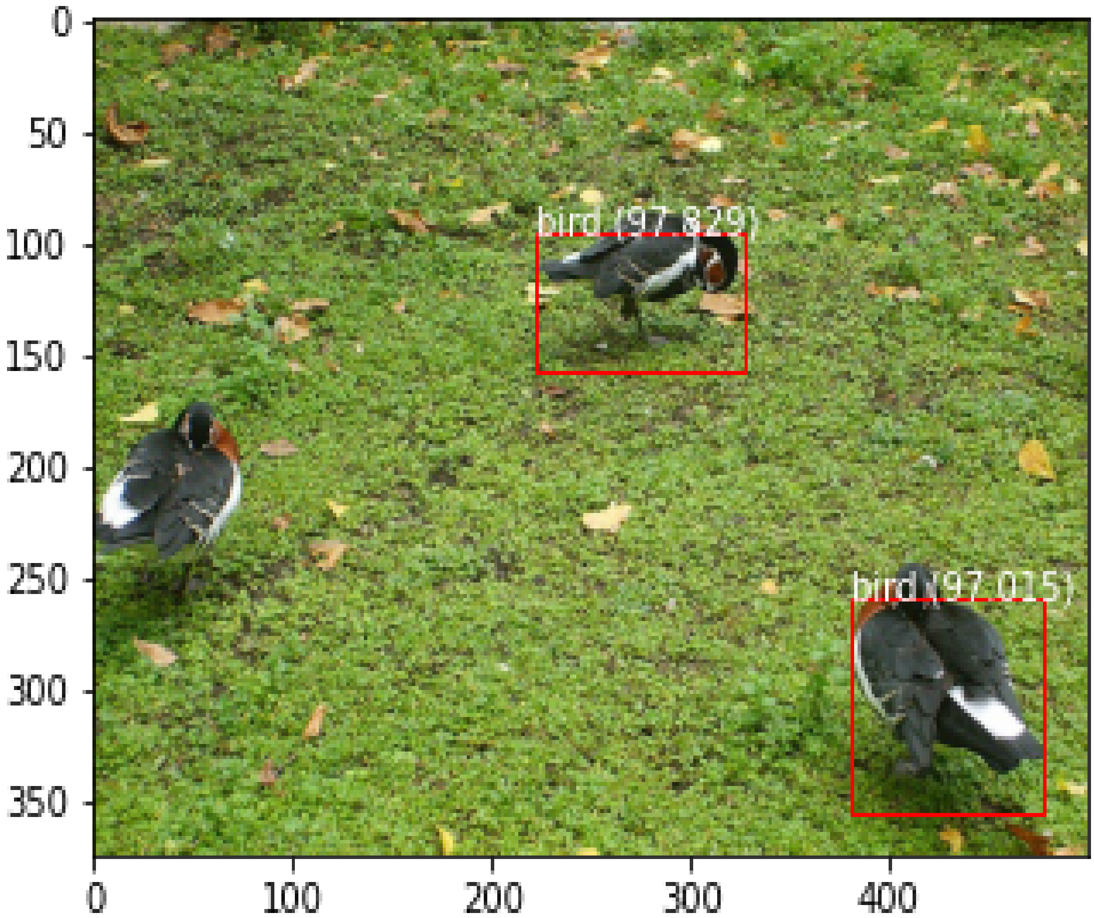



Figure 4.5

# CHAPTER 5 : EXPERIMENTS AND RESULTS

## 5.1. Mean Average Precision (mAP)

Each class's AP is computed separately. This signifies that the number of AP values is equal to the no. of classes (loosely). The mean Average Precision (mAP) is calculated by averaging these AP values. (mAP) is the average of all AP values across all classes.

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k \qquad \text{-equation 5.1}$$

where $AP_K$= the AP of the class k.

n= the no. of classes.

## 5.2 Root Mean Square Error (RMSE)

The sq. root of the mean of the square of all errors in the RMSE. The RMSE error measure is widely used, and it is regarded as a good overall error meter for numerical forecasts.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(S_i - O_i)^2} \qquad \text{- equation 5.2}$$

where Oi is the no. of observations, Si denotes the projected value of a variable, and n denotes the total number of observations accessible for evaluation. Because RMSE is scale-dependent, it should only be used to evaluate prediction errors of various models or model setups for a single variable, not between variables.

| | mAP | BN_loss | Root MSE |
|---|---|---|---|
| **Base** | 0.78 | - | - |
| **Proposed** | 0.86 | 3.5 | 1.05 |

Table 5.1- Evaluated Outcomes.

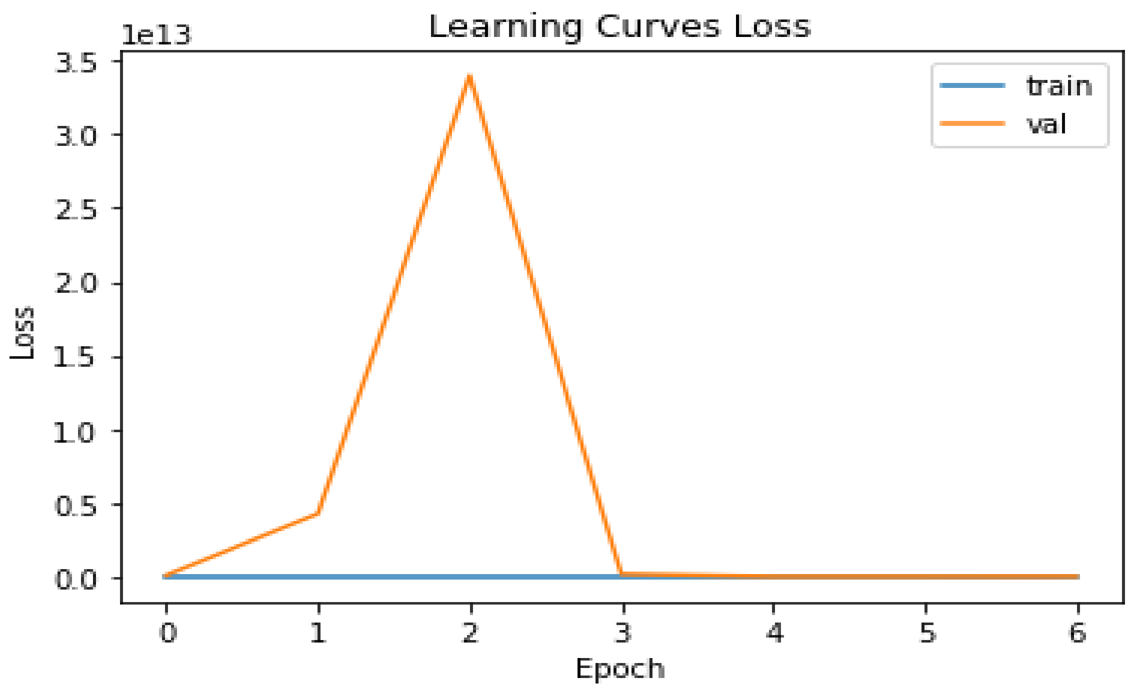



Figure 5.1– Loss curve graph of the proposed approach.

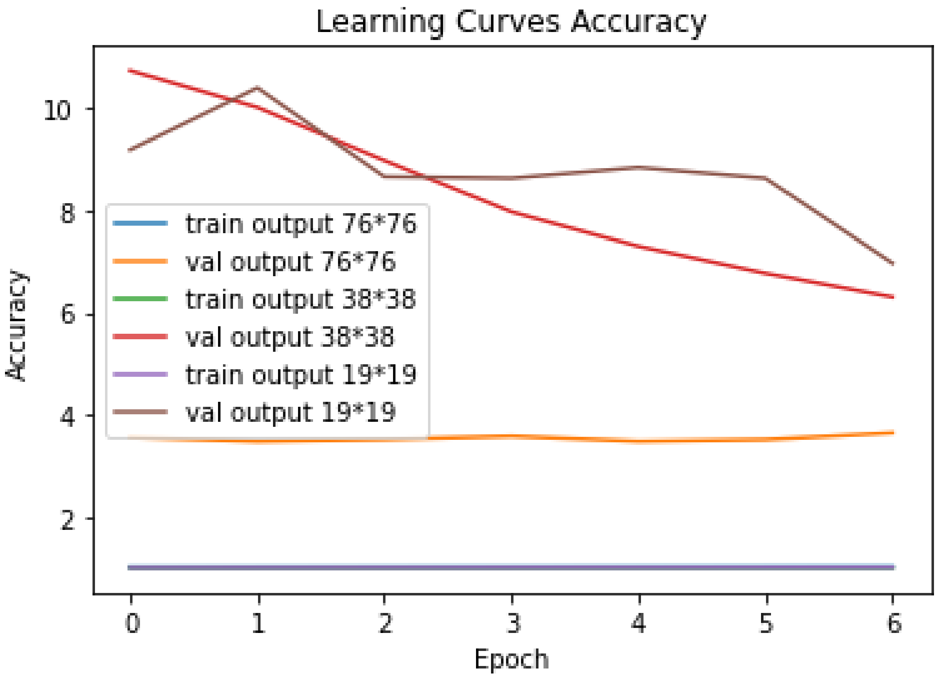



Figure 5.2– Learning Curve Accuracy

# CHAPTER 6 : CHALLENGES AND CONCLUSION

## 6.1 Challenges of Object Detection

The image's location in position can be changed at any moment. The program will manage the images consistently in the template matching process.
During the span of a lighting system, circumstances may alter. The lighting of a photograph can be affected by weather changes. In this circumstance, the lighting situation may change over time. The picture lighting system is affected by the image's shadow. Object detection from a picture can be performed in any lighting environment.
The photos may be rotated in rotation, and the machine may be capable of overcoming such a challenge. For example, a character can take any shape, yet the identification of the character does not affect the image's alignment.
The object detection system is capable of detecting mirrored images of any object.
The situation of occlusion occurs when an object is not viewable.
The object detecting system is unaffected by changes in the object's dimensions. The issues could arise as a result of object detection. In object detection, the scaling phase is the method of recognizing the scale of photographs.

## 6.2 Conclusion

Object detection is a critical and difficult subject in computer vision that has piqued public interest. Target detection has undergone significant modifications as a result of the evolution of deep learning technologies. This paper provides a comprehensive overview of object detection techniques, as well as the dataset and assessment criteria utilized in target detection. The evolution of object detection technology is also discussed, as well as the classic and new domains. Future object recognition research could concentrate on the following aspects: Lightweight OD, Video OD, Weak OD, Small OD.

Object detection is the initial step in the development of self-driving automobiles and robots. In this paper, we decoded the role of CNN-based DL systems for object detection. There are further datasets for object localization and identification that have been presented in international competitions. The pointers to the domains where object detection can be used were discussed. Object identification algorithms based on deep learning that are state-of-the-art have been evaluated and compared.

# REFERENCES


[1] R. Zhang, S. Li, G. Ji, X. Zhao, J. Li, and M. Pan, “Survey on Deep Learning-Based Marine Object Detection,” *J. Adv. Transp.*, vol. 2021, 2021, doi: 10.1155/2021/5808206.

[2] Z. Q. Zhao, P. Zheng, S. T. Xu, and X. Wu, “Object Detection with Deep Learning: A Review,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 30, no. 11, pp. 3212–3232, 2019, doi: 10.1109/TNNLS.2018.2876865.

[3] B. Kaur and S. Singh, “Object Detection using Deep Learning: A Review,” *ACM Int. Conf. Proceeding Ser.*, pp. 328–334, 2021, doi: 10.1145/3484824.3484889.

[4] Sunil and Gagandeep, “Study of Object Detection Methods and Applications on Digital Images,” *IJSDR1905088 Int. J. Sci. Dev. Res.*, vol. 4, no. 5, pp. 491–497, 2019, [Online]. Available: www.ijsdr.org.

[5] K. L. Masita, A. N. Hasan, and T. Shongwe, “Deep learning in object detection: A review,” *2020 Int. Conf. Artif. Intell. Big Data, Comput. Data Commun. Syst. icABCD 2020 - Proc.*, no. August, 2020, doi: 10.1109/icABCD49160.2020.9183866.

[6] Mohana and H. V. Ravish Aradhya, “Object detection and tracking using deep learning and artificial intelligence for video surveillance applications,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 12, pp. 517–530, 2019.

[7] S. Ding and K. Zhao, “Research on Daily Objects Detection Based on Deep Neural Network,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 322, no. 6, 2018, doi: 10.1088/1757-899X/322/6/062024

[8] A. G. Howard *et al.*, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” 2017, [Online]. Available: http://arxiv.org/abs/1704.04861.

[9] T. Sharma, B. Debaque, N. Duclos, A. Chehri, B. Kinder, and P. Fortier, “Deep Learning-Based Object Detection and Scene Perception under Bad Weather

Conditions," *Electron.*, vol. 11, no. 4, pp. 1–11, 2022, doi: 10.3390/electronics11040563.

[10] F. Saeed, M. J. Ahmed, M. J. Gul, K. J. Hong, A. Paul, and M. S. Kavitha, "A robust approach for industrial small-object detection using an improved faster regional convolutional neural network," *Sci. Rep.*, vol. 11, no. 1, pp. 1–13, 2021, doi: 10.1038/s41598-021-02805-y.

[11] V. Ponnusamy, D. R. Marur, D. Dhanaskodi, and T. Palaniappan, "Deep Learning-Based X-Ray Baggage Hazardous Object Detection – An FPGA Implementation," *Rev. d'Intelligence Artif.*, vol. 35, no. 5, pp. 431–435, 2021, doi: 10.18280/ria.350510.

[12] D. Baimukashev *et al.*, "Deep Learning Based Object Recognition Using Physically-Realistic Synthetic Depth Scenes," *Mach. Learn. Knowl. Extra.*, vol. 1, no. 3, pp. 883–903, 2019, doi: 10.3390/make1030051.

[13] S. Akcay and T. P. Breckon, "An evaluation of region-based object detection strategies within X-ray baggage security imagery," *Proc. - Int. Conf. Image Process. ICIP*, vol. 2017-September, pp. 1337–1341, 2018, doi: 10.1109/ICIP.2017.8296499.

[14] D. Li, L. Deng, and Z. Cai, "Research on image classification method based on convolutional neural network," *Neural Comput. Appl.*, vol. 7, 2020, doi: 10.1007/s00521-020-04930-7.

[15] V. Aggarwal and G. Kaur, "A review: deep learning technique for image classification," *Accent. Trans. Image Process. Comput. Vis.*, vol. 4, no. 11, pp. 21–25, 2018, doi: 10.19101/tipcv.2018.411003.

[16] N. Handa, Y. Kaushik, N. Sharma, M. Dixit, and M. Garg, "Image Classification Using Convolutional Neural Networks," *Commun. Comput. Inf. Sci.*, vol. 1393, no. 3, pp. 510–517, 2021, doi: 10.1007/978-981-16-3660-8_48.

[17] [17] L. Alzubaidi *et al.*, *Review of deep learning: concepts, CNN architectures, challenges, applications, future directions*, vol. 8, no. 1. Springer International Publishing, 2021.

[18] [18] S. t S. and E. And, "Elman's recurrent neural network applications to

condition monitoring in nuclear power plant and rotating machinery," *Research Gate*, 2003. https://www.researchgate.net/publication/245214071_Elman's_recurrent_neural_network_applications_to_condition_monitoring_in_nuclear_power_plant_and_rotating_machinery.

[19] M. Shuaib, "a Survey on Deep Learning Techniques," *Strad Res.*, vol. 7, no. 8, pp. 1–4, 2020, DOI: 10.37896/sr7.8/037.

[20] C. et Al., "Gated Recurrent Unit," *papers with code*, 2014. https://paperswithcode.com/method/gru.

[21] Y. Bengio, "Learning deep architectures for AI," *Found. Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–27, 2009, doi: 10.1561/2200000006.

[22] A. Yazdan, "Stacked Auto Encoders." https://www.researchgate.net/figure/A-stacked-autoencoder-based-deep-learning-model_fig1_310671661.

[23] A. Bilogur, "Notes on deep belief networks," *Kaggle*, 2019. https://www.kaggle.com/residentmario/notes-on-deep-belief-networks.

[24] J. Wang, T. Zhang, Y. Cheng, and N. Al-Nabhan, "Deep learning for object detection: A survey," *Comput. Syst. Sci. Eng.*, vol. 38, no. 2, pp. 165–182, 2021, doi: 10.32604/CSSE.2021.017016.

[25] S. K. Pal, A. Pramanik, J. Maiti, and P. Mitra, *Deep learning in multi-object detection and tracking: state of the art*, vol. 51, no. 9. 2021.

[26] J. Deng, X. Xuan, W. Wang, Z. Li, H. Yao, and Z. Wang, "A review of research on object detection based on deep learning," *J. Phys. Conf. Ser.*, vol. 1684, no. 1, 2020, doi: 10.1088/1742-6596/1684/1/012028.

[27] A. R. Pathak, M. Pandey, and S. Rautaray, "Application of Deep Learning for Object Detection," *Procedia Comput. Sci.*, vol. 132, no. Iccids, pp. 1706–1717, 2018, doi: 10.1016/j.procs.2018.05.144.

[28] M. A. Hossain and M. S. Alam Sajib, "Classification of Image using Convolutional Neural Network (CNN)," *Glob. J. Comput. Sci. Technol.*, vol. 19, no. 2, pp. 13–18, 2019, doi: 10.34257/gjcstdvol19is2pg13.

[29] P. Devaki, S. Shivavarsha, G. B. Kowsalya, M. Manjupavithraa, and E. A. Vima, “Real-Time Object Detection using Deep Learning and Open CV,” *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 12S, pp. 411–414, 2019, doi: 10.35940/ijitee.l1103.10812s19.

[30] M. Li, H. Zhu, H. Chen, L. Xue, and T. Gao, “Research on Object Detection Algorithm Based on Deep Learning,” *J. Phys. Conf. Ser.*, vol. 1995, no. 1, 2021, doi: 10.1088/1742-6596/1995/1/012046.

[31] R. Dhamija, “Analysis of Clustering Methods with CBIR- A Survey,” vol. 9, no. 4, 2020.

[32] A. Hossain and M. Ali, “Recognition of Handwritten Digit using Convolutional Neural Network (CNN),” 2019

[33] M. Tan, R. Pang, and Q. V. Le, “EfficientDet,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 10778–10787, 2020.

[34] D. Wang, J. G. Wang, and K. Xu, “Deep learning for object detection, classification and tracking in industry applications,” *Sensors*, vol. 21, no. 21, pp. 3–5, 2021, doi: 10.3390/s21217349.

[35] S. Beery, G. Wu, V. Rathod, R. Votel, and J. Huang, “Context R-CNN: Long term temporal context for per-camera object detection,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 13072–13082, 2020, doi: 10.1109/CVPR42600.2020.01309.

[36] R. Goel, A. Sharma, and R. Kapoor, “Object recognition using deep Learning,” *J. Comput. Theor. Nanosci.*, vol. 16, no. 9, pp. 4044–4052, 2019, doi: 10.1166/jctn.2019.8291.

# LIST OF PUBLICATIONS

**[1]** Vinod Kumar,Archi Bakawle, "Object Recognition via Deep Learning and YoloWingNet Deep Neural Network ". Accepted at the **4th International Conference on Advances in Computing , Communication Control and Networking (ICAC3N-22)**
Indexed by Scopus, IEEE,

**Abstract-** DL approaches to Object Detection (OD) have attracted a lot of attention from researchers because of their implied strength in overcoming the drawbacks of traditional approaches that rely on handcrafted characteristics. DL algorithms have made major advances in object recognition during the previous few years. This paper discusses the most recent and effective DL framework for object recognition. Visual recognition systems, which include picture categorization, localization, and detection, are at the heart of all of these applications and have gathered a lot of research attention. These visual identification algorithms have achieved extraordinary performance due to considerable advancements in neural networks, particularly deep learning. OD is one of these sectors where computer vision has had a lot of success. The role of DL methods based on YoloWingNet for OD is proposed in this research.

[2] Vinod Kumar, Archi Bakawle, " Deep Learning-Based Object Detection : A Review". Accepted and presented at the **4th International Conference on Advances in Computing , Communication Control and Networking (ICAC3N-22)**.

Indexed by
Scopus,IEEE.

**Abstract-** In computer vision, classifying and detecting various items in an image is a crucial ability. Robust and efficient object identification is a critical method for engaging with one's surroundings. Humans utilize a technique known as a visual focus to swiftly determine which areas of an image require detailed processing and which can be avoided. However, identifying an object and its precise location in an image is a challenging problem for a machine. This paper studies features and methods of object detection and the algorithms related to object detection using deep learning.

Similarity Report ID: oid:27535:17894799

turnitin

PAPER NAME

final thesis archi.docx

WORD COUNT

7319 Words

CHARACTER COUNT

39808 Characters

PAGE COUNT

42 Pages

FILE SIZE

1.8MB

SUBMISSION DATE

May 30, 2022 2:14 PM GMT+5:30

REPORT DATE

May 30, 2022 2:15 PM GMT+5:3

## 13% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 10% Internet database
- 1% Publications database
- Crossref Posted Content database database
- 11% Submitted Works

## Excluded from Similarity Report

- Crossref database
- Bibliographic material