e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:06/June-2023 Impact Factor- 7.868 www.irjmets.com

OBJECT DETECTION AND IMAGE ANNOTATION USING DEEP LEARNING

S.S. Killikatt[*1], Avantika Patil[*2], Sarika Pharakate[*3], Supriya Koli[*4], Shreya Wankhade[*5], Maithili Patil[*6]

[*1]Assistant Professor, Department Of Computer Science Engineering, D Y Patil Engineering And Technology, Kasba Bawda, Kolhapur, India.

[2,3,4,5,6]Students, Department Of Computer Science Engineering, D Y Patil Engineering And Technology, Kasba Bawda, Kolhapur, India.

DOI : https://www.doi.org/10.56726/IRJMETS41723

ABSTRACT

Image annotation and object detection are fundamental tasks in computer vision that play a crucial different applications, including autonomous driving, surveillance systems, medical imaging. Image annotation involves the process of labeling specific objects or regions in an image with annotations or labels, providing information about their class, boundaries, or additional attributes. This process is typically carried out by human annotators who manually mark objects of interest using techniques like bounding boxes, segmentation masks, or key points. Object detection, on the other hand, is an automated process of identifying and localizing objects within an image. It involves training machine learning models, particularly Convolutional neural networks (CNNs), a type of deep learning model, to learn patterns and features that distinguish different object classes and their spatial locations. These models are trained on annotated datasets, where images are labeled with bounding boxes or other types of annotations. Once trained, the models can analyze new images and predict the presence, class, and location of objects accurately.In recent years, significant advancements in deep learning have propelled object detection techniques to new heights. Models such as Faster R-CNN, SSD, and YOLO have emerged as popular choices, combining convolutional layers for feature extraction and region proposal mechanisms to get cutting-edge performance. These models are capable of instant object detection, making them suitable for various real-world applications.The process of image annotation and object detection are intertwined, as accurate annotations are essential for training robust object detection models. Creating high-quality annotated datasets is a labor-intensive task that requires expertise and attention to detail. However, with the rapid development of annotation tools and the availability of pre-annotated datasets, the process has become more accessible.Image annotation and object detection have revolutionized computer vision applications by enabling automated object recognition, tracking, and understanding. These technologies continue to advance, contributing to the development of intelligent systems capable of perceiving and interacting with the visual world.

I. INTRODUCTION

Deep learning algorithms have revolutionised computer vision tasks in recent years., enabling us to develop sophisticated systems capable of detecting objects in images and annotating them accurately. Object detection and image annotation play vital roles in various domains ,including autonomous driving, surveillance systems, medical imaging, and content analysis.

This project aims to explore and implement a robust object detection and image annotation system using deep learning algorithms. By leveraging the power of regions with convolutional neural networks (CNNs) and other advanced techniques, we can automatically identify and locate objects of interest in images, as well as provide meaningful annotations describing those objects.

Object detection involves identifying multiple objects within an image, drawing bounding boxes around them , along with their corresponding class labels .It goes beyond simple image classification by providing precise localization information.

The goal of this research is to create object detection models that can locate things precisely and with high precision and recall.

The technique of giving semantic labels to objects or areas of interest in an image is known as image annotation. It involves associating meaningful textual information with specific objects, enabling better

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:06/June-2023 Impact Factor- 7.868 www.irjmets.com

understanding and interpretation of visual content. Our project will incorporate image annotation techniques, allowing us to automatically generate annotations for detected objects, enhancing the usefulness and comprehensibility of the system. In this project we will be giving images i.e , the Data Sets as input and our model will identify all the objects in the image, outline them and label them.

Object classification and identification are two examples. The user may select a region of a picture to emphasise and link to a label. We are promoting the use of machine learning, which is the function and structure of the brain known as artificial neural network, by employing a deep learning technique that learns data from images.

Need of the work :

Image annotation is required to make systems deliver accurate results, help modules identify elements to train computer vision and speech, recognition models. Any model or system that has a machine-driven decision making system at the support, data annotation is required to ensure the decisions are accurate and relevant.

Problem Statement:

Object Detection and Image Annotation using Deep Learning techniques.

Objectives:

1. To take data and to preprocess them
2. To find the suitable deep learning model to detect object
3. Train the model for object detection
4. To train the model for object annotation
5. To test object detection and annotation on given image
6. To calculate the accuracy of work

Proposed work :

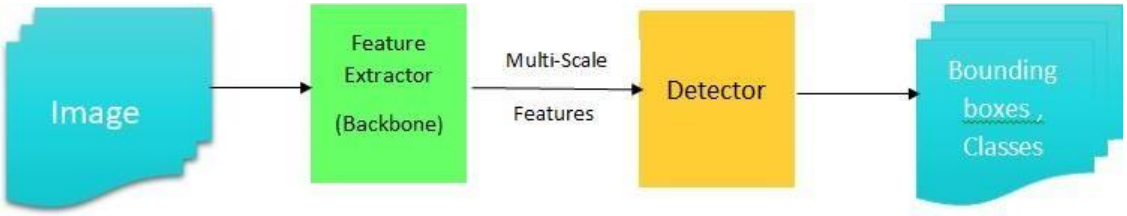

Fig.1: Flow of work

1. First we have to take the image i.e. the data as input and preprocess it.
2. Then we have to find a deep learning model which we will be using for detecting the objects from the given data.
3. We now will be training the chosen model for object detection.
4. Then we will have to train the model for object annotation also.
5. Further part will be testing the object detection and annotation on the given image.
6. And then we will be calculating the accuracy of the work.

REQUIREMENT ANALYSIS AND SPECIFICATION: INFORMATION GATHERING

Humans are able to recognise and locate items in an image. The human visual system is quick and precise and also capable of carrying out challenging tasks like object identification and obstacle detection with little conscious effort. We can now quickly train computers to detect and classify many items within a picture with high accuracy thanks to the availability of massive data sets, faster GPUs, and better algorithms. We must comprehend concepts like object localization, object detection, and loss functions for both, before examining the R-CNN object detection algorithm. To recognise items in digital photos, a group of related activities is referred to as object recognition. R-CNNs, also known as region-based convolutional neural networks, are a group of methods for tackling object localisation and recognition.

The goal of object detection is to identify every instance of a known class of items in a picture, such as people, cars, or faces. Even though there are often few instances of the object in the photograph, there are a vast array of locations and scales where they might appear that must be investigated in some way. Each image detection is

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:06/June-2023 Impact Factor- 7.868 www.irjmets.com

supplied together with some kind of pose data. This can be as straightforward as the object's location, scale, or the extent of the object as described by a bounding box.

II. LITERATURE REVIEW

Here is a discussion of earlier studies on object detection:

Using object detection, **Guo et al. (2012)** suggested a method for following objects in video frames. The results of the simulation show that this strategy was effective at identifying generic object types. For real-time object recognition, classification accuracy needs to be improved more.

Ben Ayed et al. (2015) published a big data analytics technique for text data identification based on a texture in video frames. Different fixed-size blocks are created from the video frames, and these blocks are then subjected to wavelet analysis. They also used a neural network to classify the text and non-text portions. This study should concentrate on filtering out text-like regions and extracting the regions to get rid of the noisy regions. in order to recognise pedestrians.

In **2015, Soundrapandiyan and Mouli** suggested a novel, adaptive method. They also employed image pixel intensities to differentiate between items in the foreground and background. The foreground edges were then sharpened using a high boost filter. Since the suggested methodology has a pedestrian recognition rate that is almost 90% greater than that of other single picture current methods, the results of the subject assessment and objective evaluation demonstrate the suggested approach's effectiveness. Future performance improvements were planned to bring the method into line with sequence image approaches in terms of higher detection rates and fewer false positives.

A modified frame difference approach was put out by **Ramya and Rajeswari (2016)**, which categorises pixels as foreground or background based on the correlation between the blocks of the current image and the background image. The blocks in the current image that strongly resemble the background image are referred to as the backdrop. Using the pixel-by-pixel comparison, the other block is classified as either the foreground or the background. The tests' findings demonstrated that this method improves the frame difference method, particularly in terms of promptly identifying correctness. This study must focus on extra data present in the blocks, such as shape and edge, in order to improve the detection accuracy.

X Yang, C Zheng, Y Feng, C Tang IEE Explore: 2017 4th International..., 2017

An important application of deep learning technology is object identification, which is distinguished by its high feature learning and feature extraction capabilities.In comparison to the conventional object detecting techniques. The paper begins by introducing the traditional methods for object detection and then explains how they relate to and differ from the deep learning approaches for object recognition.

MS Naik, PG Raikar, and SR Prasad, 2019. Most computer vision systems and robot vision systems depend heavily on the capacity to detect objects. Recent research in this field has advanced significantly in many areas. A number of applications exist for object recognition and tracking, and this paper discusses some of them. Here, we talk about the numerous domains in which object detection systems are currently and in the future used.

Life Cycle Model

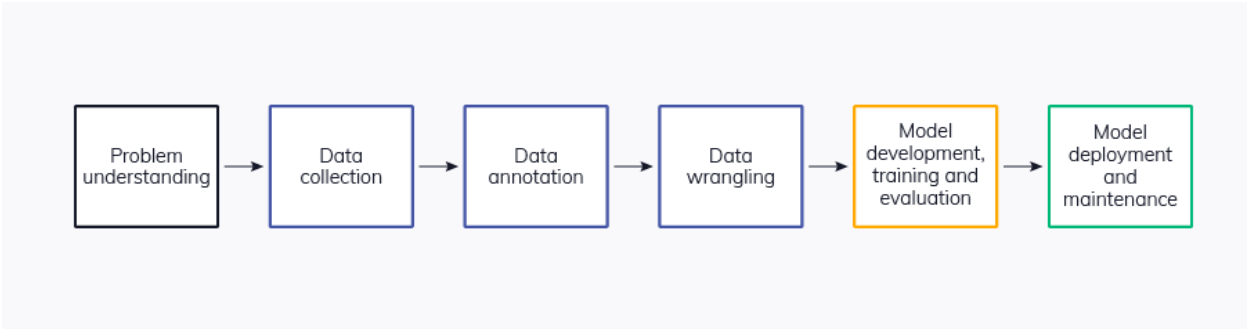

Fig.2: Life Cycle Model

Step 1: Recognise the issue

Every project begins with a challenge that needs to be overcome. A precise problem definition should ideally be mathematically explained. With the use of numbers, you can not only determine where you are beginning

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:06/June-2023 Impact Factor- 7.868 www.irjmets.com

from but also monitor the effects of future changes.

Step 2: Gathering data

To gather as much pertinent data as you can is your aim. If we are discussing tabular data, this typically means obtaining data across a broad timeframe. Keep in mind that your future model will be more accurate the more samples you have.

Step 3: Preparing the data

It (also known as data wrangling) is one of the most time-consuming but crucial processes because it has a direct impact on the calibre of the data that will be uploaded to the internet.

Step 4: Annotating the data

You will require a label for every sample in your dataset if your work falls under the category of supervised learning. Data annotation or data labelling is the practise of giving labels to data samples.

Step 5: modelling

A unique validation dataset is used for evaluation during training. It monitors the generalisation performance of our model while avoiding bias and overfitting.

Step 6: Deploying the model

Models deployed require monitoring. To ensure that the deployed model continues to perform at the level that the business demands, you must monitor its performance.

III. SRS

Project Perspective:

SRS(Software Requirements Specification) document outline for Deep learning project object detection and image annotation:

Introduction:

Finding one or more useful targets from still images or video data is the primary goal of object detection. It completely incorporates a range of methodologies, including machine learning, pattern recognition, and image processing.

Scope:

The majority of computer and robot vision systems depend heavily on the capacity to detect objects. Although there has been significant development in recent years, object detection is still necessary for robots that will explore places that humans have never been, including the deep sections of the ocean or other planets. The detection systems will need to learn new item classes as they are encountered.

Software Requirements:

- Language: Python
- IDE: Google Colab, Jupyter Notebook.
- Tool: Anaconda, LabelImg

Hardware Requirements:

- Operating System: Windows 10
- Processor: Inteli5 and above
- RAM: 4 GB and above

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:06/June-2023 Impact Factor- 7.868 www.irjmets.com

IV. DESIGN

System Architecture:

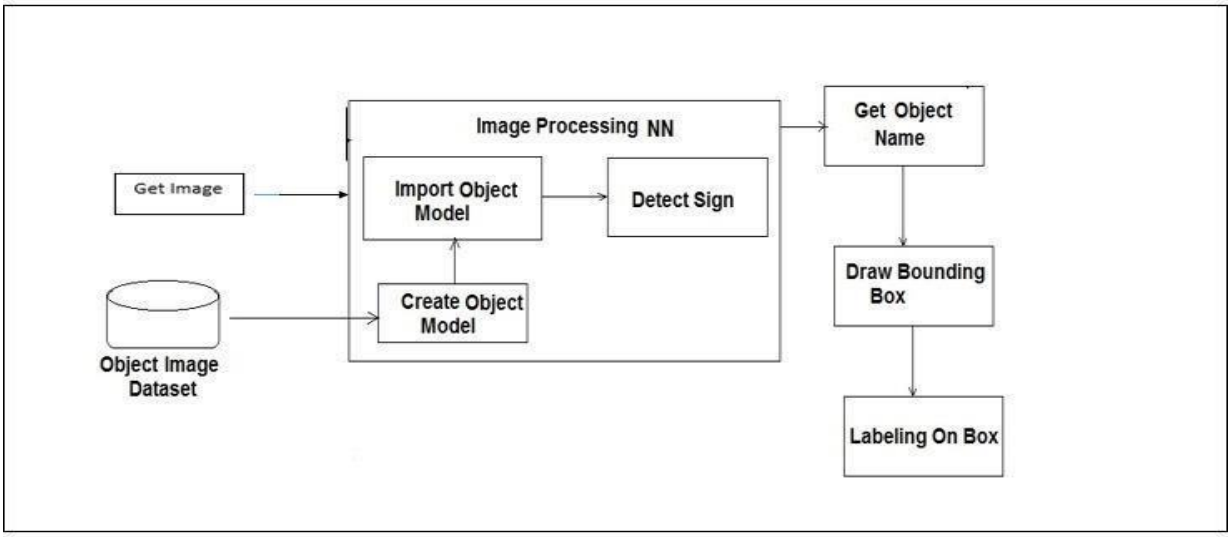

Fig.3: System Architecture

Description:

The System Architecture consists of three parts Input , Processing and Output.

1. For Input , we have to take an image from image dataset for object detection.
2. In the processing part the following is done :

Creating an object model for image detection and annotation typically involves the following steps:

a. Data Collection : Gather a diverse and representative dataset of images that contain the objects you want to detect and annotate. The dataset should include a wide variety of object instances, backgrounds, orientations etc.

b. Data Annotation : Annotate the objects of interest in the images with bounding boxes or masks. This annotation process involves marking the object's location and shape in each image.

c. Training Data Preparation : Split the annotated dataset into training and validation sets.

d. Model Selection : Choose an appropriate object detection model architecture for your task.

e. Model Training : Initialize the selected object detection model with the pretrained weights.

f. Model Evaluation : Evaluate the trained model's performance using the annotated validation dataset.

g. Deployment : Integrate the object detection model into your desired application or system.

h. Import object model : To import an object model for image detection and annotation

i. You need to follow these general steps :

A. Choose a Deep Learning Framework

B. Install the Required Libraries:

- Model Loading
- Visualization and Integration

It is common practise to use a collection of annotated sign images to train a deep learning model, such as a convolutional neural network (CNN), in order to detect signs. Images of signs are included in the annotated dataset together with the relevant bounding box coordinates and class labels. The CNN gains knowledge of the visual patterns and characteristics of signs, enabling it to spot them in fresh, previously undiscovered images.

3. In the output section, the item name, a bounding box around the image, and the label on the needed image with the bounding box and accuracy will all be provided.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:06/June-2023 Impact Factor- 7.868 www.irjmets.com

Problem Modules:

- **Module 1 : Data Collection**

The first step in an object detection is to collect a dataset of images that contain the objects you want to detect. We have two models, one is the pre- trained model and another model is created by us. We collected data for the pre-trained model from Kaggle dataset viz.coco.names . And for the model created by us we have collected various types of objects (data) from Google. While creating the dataset for the model created by us we have used Labeling tool. Labeling tool is used to annotate the objects by drawing bounding boxes around object and assign corresponding labels.

- **Module 2 : Training a Dataset**

The usual training dataset comprises of bounding box annotations and captioned images. A model is trained using the dataset to find and identify items in images.In the case of a pre-trained model, the model has already been trained, hence further training is not necessary. The annotated dataset is fed into the chosen model during training, and model parameters are changed to reduce prediction error. This can be a laborious operation that may call for specialised gear, such GPUs.

- **Module 3 : Object detection**

A trained deep learning model that is capable of reliably recognising and localising objects in real time will be the project's ultimate deliverable. Three elements often make up the final product.:

1. Bounding box – The input images are annotated with bounding box to highlight detected object. The bounding boxes are usually drawn around object with different colors for each class .
2. Class label – The predicted class labels are often displayed along with the bounding boxes

to provide clear identification of detected object .

3. Confidence score – The confidence score associated with each detection can be visualized , such as displaying them as numerical values to represent accuracy.

V. IMPLEMENTATION AND CODING: TECHNOLOGY USED

- **OpenCV**

A set of programming tools called OpenCV is primarily focused on real-time computer vision. With support for deep learning, OpenCV (Open Source Computer Vision Library) is a robust and frequently used open-source computer vision library. Even while OpenCV is well known for its powerful computer vision capabilities, its integration with deep learning frameworks expands its capability and elevates it to the status of a useful tool for deep learning tasks. The capacity of OpenCV to handle real-time deep learning jobs is one of its key features. Real-time inference on live video streams or camera inputs is made possible by integrating deep learning models with OpenCV. When deep learning models must handle video frames for applications like real-time item detection or facial recognition, this is especially helpful.

pip install opencv-python-command

- **R-CNN**

A two-stage detection algorithm is R-CNN. A subset of regions in an image that potentially contain an item are found in the first stage. The object is categorised in each region in the second stage. A deep learning architecture called RCNN (Region-based Convolutional Neural Network) was created for object detection in photos. By introducing a region-based strategy that combines the strength of convolutional neural networks (CNNs) with precise localization and classification skills, it revolutionised the field of object detection.The RCNN model is divided into several phases. With the aid of a comparable algorithm or selective search, it first creates a list of region proposals. These suggestions represent possible areas in the image where things might be visible. Then, to extract high-level information, each region suggestion is independently distorted and fed into a pre-trained CNN.

- **Tensorflow**

A complete open source machine learning platform is called TensorFlow. The class concentrates on using a specific TensorFlow API to create and train machine learning models, despite the fact that TensorFlow is a robust system for managing all parts of a machine learning system. A computer vision method called

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:06/June-2023 Impact Factor- 7.868 www.irjmets.com

TensorFlow object detection finds, tracks, and detects an item in a still image. Google created TensorFlow, an effective and popular open-source deep learning framework. It gives users access to a complete environment for creating and using deep neural networks. TensorFlow is a preferred option for deep learning researchers, developers, and practitioners because to its flexibility and adaptability. TensorFlow's capacity to automatically differentiate is one of its main advantages. It effectively determines gradients for a parameter

- **Tflite-model-maker**

When deploying a TensorFlow neural-network model for on-device machine learning applications, the TensorFlow Lite Model Maker module streamlines the process of converting and adapting the model to specific input data. TensorFlow's high-level deep learning library, Tflite-model-maker, makes it easier to train specialised machine learning models for a variety of applications, including image classification, object identification, and text categorization. It seeks to make it possible for developers, including those without considerable machine learning knowledge, to easily and effectively design their own models. The user-friendly interface of tflite-model-maker streamlines the model training procedure. The intricacies of model architecture selection, data preprocessing, training, and assessment are handled by a collection of simple-to-use APIs and abstractions. This enables programmers to concentrate on the current work.

VI. TESTING

Using a different dataset that wasn't utilised during training, a trained model is evaluated and validated during testing. Testing is necessary to evaluate the model's performance and generalisation skills on untested data as well as to analyse the model's accuracy, precision, recall, and other pertinent metrics. Testing is used to assess the efficacy, generalizability, and performance of trained models on unobserved data. The best model is chosen, problems like overfitting are identified and dealt with, and feedback is given for iterative development. Testing guarantees dependability, fosters trust, and facilitates continual model monitoring for long-term performance.

An object detection model's accuracy is influenced by the calibre and quantity of training samples, the input imagery, the model parameters, and the accuracy threshold.

Precision—The ratio of genuine positives to all positive predictions is known as precision. The precision would be 90% if the model detected 100 trees and 90 of them were correct.

Precision = (True Positive)/(True Positive + False Positive)

Recall—The proportion of genuine (relevant) objects to all true positives is known as recall. For instance, the recall is 75% if the model accurately identifies 75 trees in an image when there are actually 100 trees there.

Recall = (True Positive)/(True Positive + False Negative)

VII. RESULTS

Fig.4: GUI

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:06/June-2023 Impact Factor- 7.868 www.irjmets.com

Fig.5: Selecting an image

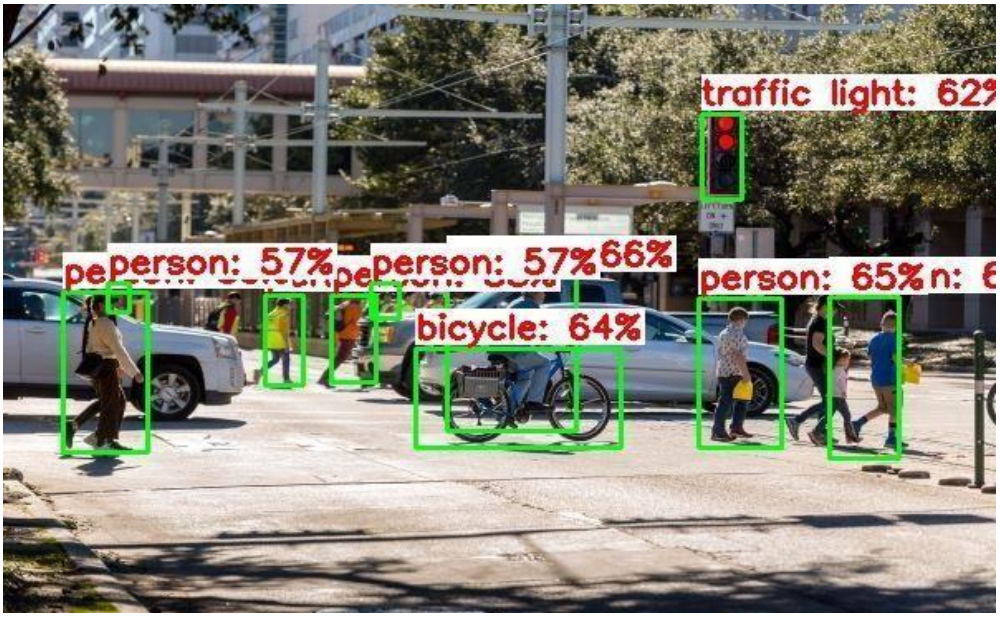

Fig.6: Output 1

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:06/June-2023 Impact Factor- 7.868 www.irjmets.com

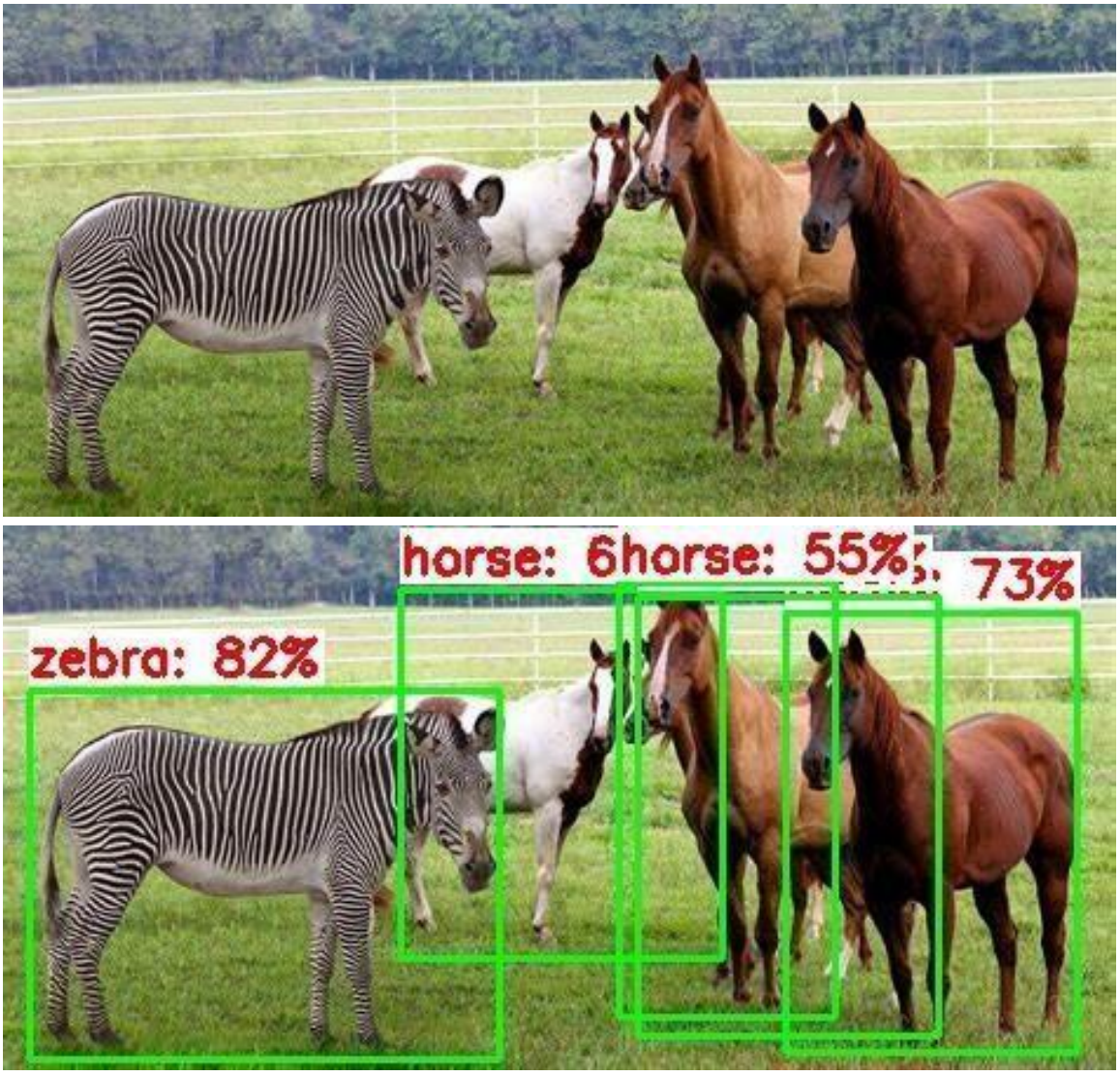

Fig.7: Output 2

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:06/June-2023 Impact Factor- 7.868 www.irjmets.com

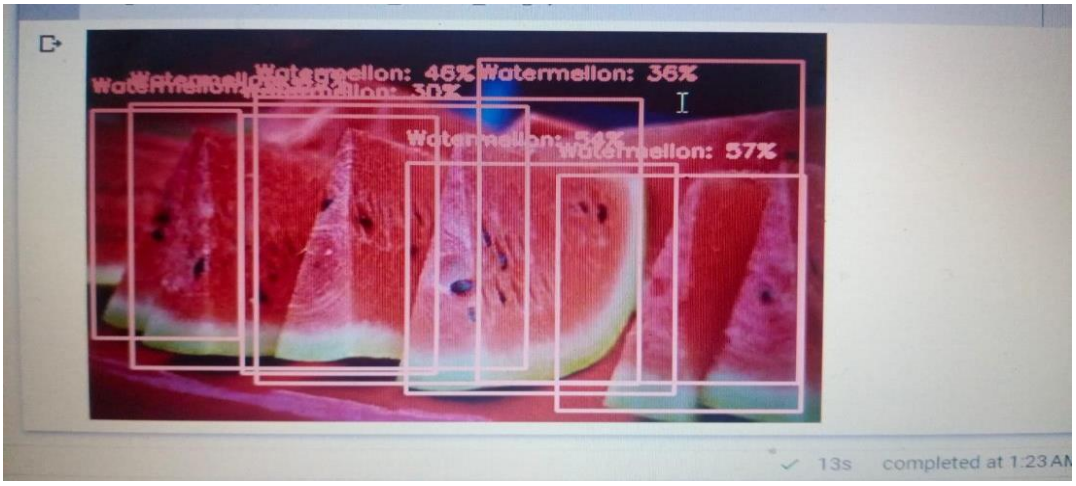

Fig.8: Output 3

VIII. CONCLUSION

Deep learning methods for object detection and image annotation have completely changed the field of computer vision and image understanding. Convolutional neural networks (CNNs), in particular, have shown astounding ability in precisely recognising and localising objects within images. Deep learning approaches for object detection and picture annotation have revolutionised computer vision applications by offering precise and effective methods for identifying things and comprehending visual content. The future of computer vision and its effects on many businesses and areas look very bright thanks to the ongoing developments in this area. It's crucial to note that there are still certain difficulties with deep learning-based object detection. Further research is needed in the areas of huge labelled dataset requirements, computational resources, and model optimisation for real-time applications. Globally, the

IX. FUTURE WORK

Vehicle plate recognition, autonomous vehicles, object tracking, face recognition, medical imaging, object counting, object extraction from an image or video, and human detection are the main uses of object detection that will be advantageous in the future.

X. REFERENCES

[1] https://labelbox.com/guides/image-annotation/

[2] https://www.v7labs.com/blog/image-annotation-guide

[3] https://research.aimultiple.com/image-annotation/

[4] Padigel, Adwitiya, Tushar Chintanwar, Shruti Landge, Pooja Khobragade, Tanu Awachat, and Manoj Lade. "Real Time Object Detection Using Deep Learning."

[5] Zhao, Zhong-Qiu, Peng Zheng, Shou-tao Xu, and Xindong Wu. "Object detection with deep learning: A review." IEEE transactions on neural networks and learning systems 30, no. 11 (2019): 3212-3232.

[6] Padigel, Adwitiya, Tushar Chintanwar, Shruti Landge, Pooja Khobragade, Tanu Awachat, and Manoj Lade. "Real Time Object Detection Using Deep Learning."

[7] Zhang, Shuai, Chong Wang, Shing-Chow Chan, Xiguang Wei, and Check-Hei Ho. "New object detection, tracking, and recognition approaches for video surveillance over camera network." IEEE sensors journal 15, no. 5 (2014): 2679-2691.

[8] Wang, Xiaogang. "Deep learning in object recognition, detection, and segmentation." Foundations and Trends® in Signal Processing 8, no. 4 (2016): 217- 382.

[9] Padigel, Adwitiya, Tushar Chintanwar, Shruti Landge, Pooja Khobragade, Tanu Awachat, and Manoj Lade. "Real Time Object Detection Using Deep Learning."

[10] Li, Ce, Yachao Zhang, and Yanyun Qu. "Object detection based on deep learning of small samples." In 2018 Tenth International Conference on Advanced Computational Intelligence (ICACI), pp. 449-454. IEEE, 2018.