Comparison of algorithms in object detection

Chapter 1 :Abstract

In today's world, computer vision technology has become a very important direction in the field of Internet applications. As one of the basic problems of computer vision, object detection has become the basis of many vision tasks. Whether we need to realize the interaction between images and text or recognize fine categories, it provides reliable information. This article reviews comparison of algorithm in object detection. Object detection is one of the predominant and challenging problems in computer vision. Over the decade, with the expeditious evolution of deep learning, researchers have extensively experimented and contributed in the performance enhancement of object detection and related tasks such as object classification, localization, and segmentation using underlying deep models. Broadly, object detectors are classified into two categories viz. two stage and single stage object detectors. Two stage detectors mainly focus on selective region proposals strategy via complex architecture; however, single stage detectors focus on all the spatial region proposals for the possible detection of objects via relatively simpler architecture in one shot. Performance of any object detector is evaluated through detection accuracy and inference time. Generally, the detection accuracy of two stage detectors outperforms single stage object detectors. In this artical we will use different kind of algorithms of deep learning. We will compare those algorithms accuracy and performance. The algorithm is CNN (Convolutional Neural Network), YOLO (You Only Look Once), SSD (Single Shot MultiBox Detector) and D2Det (Dynamic-Driven Detector). CNN is a type of neural network designed for processing grid-like data, such as images and videos. It uses convolutional layers to automatically learn and extract features from input data. D2Det is an object detection algorithm that combines the advantages of two-stage detectors and deformable convolutional networks.It first generates region proposals using a dense prediction network. YOLO is a real-time object detection algorithm that directly predicts bounding boxes and class probabilities from an input image in a single forward pass.YOLO divides the input image into a grid and predicts bounding boxes and class probabilities for each grid cell.SSD is another real-time object detection algorithm that combines multiple bounding box aspect ratios and scales at different layers of a neural network.It predicts object classes and bounding box coordinates simultaneously, avoiding the need for region proposals.By reviewing the current research status of object detection networks, it provides suggestions for the further development trend and research of object detection.

Chapter 2: Introduction:

In this section, we will discuss about object detection, the techniques that we will use and problems of our proposed topic. We will discuss about different algorithms that are related to our topic and their parameters. Here we will alsoo find the impacts of our project.

Background:
Object detection is a computer vision task that involves identifying and locating objects within an image. The goal of object detection is to not only recognize what objects are present in the visual data but also to provide information about their precise positions by drawing bounding boxes around them.

Visual Understanding: Object detection enables machines to understand and interpret the visual world, making it possible to recognize and locate objects in images and videos. This is fundamental for tasks like image understanding, scene analysis, and video surveillance.

Automation: Object detection plays a key role in automating processes that require interaction with the physical world. For example, in autonomous vehicles, robots, and drones, object detection is used for identifying and avoiding obstacles, pedestrians, and other vehicles.

Security and Surveillance: Object detection is essential for security and surveillance applications. It can detect and track people, objects, or anomalies in real-time, helping to enhance security in public spaces, airports, and critical infrastructure.

Medical Imaging: Object detection is employed in medical imaging for identifying and localizing anatomical structures, tumors, and abnormalities in X-rays, MRI scans, and other medical images. These assists healthcare professionals in diagnosis and treatment planning.

Quality Control and Manufacturing: In industrial settings, object detection is used for quality control and defect detection in manufacturing processes. It ensures that products meet specific standards and reduces defects.

Technique:

CNN (Convolutional Neural Network): CNNs are a class of deep neural networks that are specifically designed for processing structured grid data. They have been widely used in image and video recognition, analysis, and processing. CNNs employ a variation of multilayer perceptron designed to require minimal preprocessing. They are capable of learning directly from raw data, and they automatically learn hierarchical representations. CNNs have significantly contributed to the advancement of compute.

YOLO (You Only Look Once): YOLO is a real-time object detection system that is one of the fastest and most popular algorithms used for this purpose. YOLO divides images into a grid and, for each grid cell, predicts bounding boxes and class probabilities directly. This results in a single network evaluation to predict bounding boxes and class probabilities for the entire image. YOLO

has several versions, including YOLOv1, YOLOv2, YOLOv3, and YOLOv4, each with various improvements in terms of speed and accurecy.

D2Det: D2Det (Deformable Two-Stage Detector) :is a more recent object detection framework that introduces dynamic receptive fields, which allows the network to automatically adjust the receptive field size according to the object scales. D2Det aims to handle objects of various scales in an image efficiently and accurately. By dynamically adjusting the receptive field, D2Det can effectively capture both small and large objects in the image, leading to improved detection performance.

SSD (Single Shot Multibox Detector): SSD is another popular object detection algorithm known for its single-shot detection capabilities. It simultaneously predicts multiple bounding boxes and class probabilities for these boxes at various scales in an image. By using a set of default bounding boxes with different aspect ratios, SSD efficiently detects objects of different scales and aspect ratios. SSD is known for its balance between speed and accuracy, making it suitable for real-time applications.

Anchor Boxes: Anchor boxes, also known as default bounding boxes, are used to predict object locations and sizes at different scales and aspect ratios. They provide a prior on what types of objects to expect in different regions of the image.

Backbone Networks: The choice of backbone network architecture can significantly impact the performance of object detectors. Common choices include architectures like ResNet, VGG, Inception, and EfficientNet.

YOLOv3 is the most accurate but slowest object detection system while SSD is the fastest one with the lowest accuracy. YOLOv2 has a lower accuracy than YOLOv3 but it is faster. For object detection in recorded images and videos, YOLOv3 is the best one since it detects the objects with the highest accuracy.Since there is a trade-off between accuracy and speed in all these systems, the most appropriate system for each application depends on the application requirements.

Objectives:

The main objectives are to

1)Use four algorithms named CNN, YOLOv, SSD, D2Det in object detection

2)Discuss about those algorithms in detailed way.

3) Evaluate and compare selected algorithms on the basis of the performance, speed, accuracy.

Scope of the study:

Real-time Object Detection:

Researching techniques to achieve real-time or near-real-time object detection for applications like autonomous vehicles and robotics.

Small Object Detection:

Addressing the challenges of detecting small objects in images, which are common in applications like microscopy, satellite imagery, and medical imaging.

Multi-Object Detection:

Developing methods for detecting multiple objects within a single image or video frame.
Extending object detection to crowded scenes.

Impact of study in national:

Developing methods for detecting multiple objects within a single image or video frame.
Extending object detection toObject detection plays a vital role in national security and public safety. It is used for identifying and tracking objects, individuals, and potential threats in surveillance systems, airports, and public spaces. crowded scenes.In agriculture, object detection helps optimize crop management, detect crop diseases, and automate farming tasks. It can lead to increased agricultural productivity and food security.Police and law enforcement agencies use object detection for criminal investigations, missing person searches, and traffic monitoring. It aids in solving crimes and maintaining law and order.

Impact of study in international:

Object detection technologies are used in international security and defense to protect borders, airports, and critical infrastructure. It aids in detecting and preventing threats from terrorism and other security concerns. Object detection technologies support global food production and food security initiatives by optimizing agricultural practices and monitoring crop health.

Chapter 2- Literature Review:

In this section we will review some papers related to our proposed topic.

1)From first paper author gave the information that 2 types of object detection methods are found. One stage method like YOLO, SSD and two stage method for instance R-CNN, fast R-CNN, faster R-CNN. CNN is more effective than traditional handmade image extraction method from image. But CNN is more time consuming than YOLO method. Limitation of YOLO algorithm is it can't detect overlapping small objects.SSD means single shot multibox detector has significant difference from YOLO algorithm. SSD use direct detection of convolution and can overcome problems of YOLO algorithm.

Paper link: https://www.scirp.org/journal/paperinformation.aspx?paperid=115011

2) The YOLOv5 model produces results with an exceptionally good visualization function. This study shows that the TSR in the YOLOv5 experiment is remarkably accurate. The definitions of "road bump," "cross walk," "give way," and "no entry" are given in detail in this document. The "No U-turn" method produced the lowest precision of 0.94. Almost eight classes have values that are all over 90.00%, demonstrating YOLOv5's exceptional TSR performance in our dataset.

Paper link: https://link.springer.com/article/10.1007/s11042-022-12163-0

3) In this paper, Objects that may be a risk to the safety of an autonomous vehicle when driving on a road are classified as hurdles, cars, and passengers. The advantage of YOLOv4, a typical one-stage detector technique, is its quick detection speed.

Paper link: https://www.sciencedirect.com/science/article/pii/S2405959521001818

4) In this paper for object detection experimentation, there are three types of photos utilized process: 409 for training, 46 for validation, and 51 for testing. Important metrics that show how accurately object detection algorithms recognize objects are AP and mAP. The accuracy of the YOLOv4 model is 93.97%, while YOLO-GD obtains 97.38%, with the larger the value of AP or mAP.

Paper link: https://www.mdpi.com/2075-1702/10/5/294

5) This is mainly a review paper. The three stage object detectors—RCNN, Fast-RCNN, and Faster-RCNN—as well as their significant applications were studied in this paper. This research reviewed in detail single stage object detectors, in particular YOLOs objects, their architectural developments, and their loss function.

Paper link: https://link.springer.com/article/10.1007/s11042-022-13644-y

6) This is a review paper that compare the performance of various object detectors on PASCAL VOC 2012 and Microsoft COCO datasets. Those models are compared on average precision (AP) and processed frames per second (FPS) at inference time. This paper intentionally compare the performances of detectors on similarly size input image, where possible, to provide a reasonable account.

Paper link: https://www.sciencedirect.com/science/article/pii/S1051200422001312#fg0060

7) This study describes the YOLOv5, a deep learning-based bug detector. For the purpose of training, validation, and testing, this model created a new twenty-three classes IP-23 dataset. The model that produced the most success, YOLOv5x, was determined to be notable. The YOLOv5x model, which was designed for this project and trained with specific parameters, produced an average precision value of 98.3%, recall value of 97.8%, precision value of 94.5%, and F1 score of 96% in terms of detection rate.

Paper link: https://www.mdpi.com/2076-3417/12/19/10167

8) An improved YOLO technique for target detection in high-resolution zoom sensing photos is presented in this study. It uses the SLIC super pixel segmentation technique to distinguish between light and dark areas, hence addressing issues like image blur and distortion. In order to deal with the unique properties of the image, the suggested technique modifies the vertical grid number of the YOLO network structure. Experiments done with multiple datasets show that performs regular YOLO and other popular algorithms in terms of accuracy and real-time performance.

Paper link: https://www.frontiersin.org/articles/10.3389/fbioe.2022.905583/full

9)This paper describes the use of machine learning and DWT for human face recognition. This paper employs four distinct algorithms: the principal component analysis (PCA) error vector, the PCA eigen vector, the CNN eigen vector, and the Linear Discriminant Analysis (LDA) eigen vector. The four results are then combined using the fuzzy system and the entropy of detection probability. The paper's combined approach produces a recognition rate of 93.34% in the best scenario and 89.56% in the worst.

Paper link: https://www.sciencedirect.com/science/article/pii/S1319157819309395

10) From this paper author showed that Normal machine learning in object detection does not perform well but combination of YOLO and SSD can accurately detect object and CNN based YOLO enhances processing time. Also we find that YOLOV5 is faster than YOLOv3 that's why YOLOv5 is used over YOLOv3.

Paper link: https://www.mdpi.com/20799292/11/4/563/htm?ref=blog.roboflow.com

11) The paper named Hyperspectral Anomaly Detection Using Deep learning , author named helped us by providing many important information about anomaly detection from image using deep learning. CNN one of the deep learning method has better fault tolerance, adaptability and strong self learning ability. CNN can extract features from image automatically. End to end cube CNN is better than pixel based CNN but it can be affected by noise and inference.

Paper link: https://www.mdpi.com/2072-4292/14/9/1973

12) Multiscale object recognition in Synthetic Aperture Radar (SAR) pictures is growing as an important area of research in SAR image interpretation. In this paper author proposed a approach named SARFNet. SARNet has the highest detection accuracy over other state-of- the-art methods. This paper also gave information adding saliency information in SSD algorithm guides SSD to understand Salient feature of the SAR target.

Paper link: https://www.mdpi.com/2072-4292/14/4/973

13) This paper is about object Instance Segmentation. To overcome the problems of object instance segmentation detection-based Method and single stage method are used. Detection based method has highest accuracy and single stage methods have faster speed. Future work of this paper is to segment object instance accurately in bad weather condition like rain, snow etc and multiscale object segmentation.

Paper link: https://www.spiedigitallibrary.org/journals/journal-of-electronic-imaging/volume-31/issue-4/041205/Review-of-object-instance-segmentation-based-on-deep-learning/10.1117/1.JEI.31.4.041205.full?SSO=1

14) The latest developments in computer vision are covered in this excerpt, with a focus on deep learning methods. It highlights the importance of tasks like object detection, object recognition, and image classification while highlighting the critical role that convolutional neural networks (CNNs) play in these processes. The evolution of object recognition from single to multi-object recognition is highlighted in the text, along with the use of deep learning in this field. It also discusses how deep learning-based methods—like RCNN—help achieve greater accuracy across a range of datasets.

Paper link: https://www.sciencedirect.com/science/article/abs/pii/S1051200422004298

15) In the passage, the significance of object detection in computer vision is discussed, and the "DeepMultiBox" detector is presented as a possible fix for the computing difficulties associated with the exhaustive search method. This detector uses a single Deep Neural Network (DNN) in a class-agnostic manner to produce a finite number of bounding boxes as object possibilities. It highlights the novel use of regression to define object detection, which enables the net to generate a confidence score for every projected box. Compared to conventional techniques that score features inside specified boxes, this unique approach is different. The paper's unique loss function, which makes it easier to train bounding box predictors inside the network, is one of its main contributions. Through the resolution of an assignment issue involving ground truth boxes and predictions, as well as the updating of matched box locations, confidences, and updating matched box coordinates, confidences, and underlying features, the model is tailored towards precise localization

Paper link:
https://openaccess.thecvf.com/content_cvpr_2014/html/Erhan_Scalable_Object_Detection_2014_CVPR_paper.html

Chapter 3:Methodology:

One kind of Deep Learning neural network architecture that is frequently utilized in computer vision is the convolutional neural network (CNN). The branch of artificial intelligence known as "Computer Vision" gives computers the ability to comprehend and analyze images and other visual input.

1. Convolutional Layer

The convolutional layer, which produces the majority of the network's computations, is the fundamental layer used to build convolutional neural networks. Keep in mind that the number of parameters does not equal the amount of calculation. Compared to a fully connected network of the same size, the convolution operation can effectively reduce the training complexity of the network model as well as the network connection and parameter weights. Standard convolution, transposed convolution, hole convolution, and depth separable convolution are examples of common convolution operations.

2. Activation Layer

Artificial neural networks can be filled with an Activation Function to aid in the network's ability to recognize complex patterns in data. Rectified Linear Units (ReLU), Randomized LeakyReLUs (RReLU), Exponential Linear Units (ELU), and others are examples of common activation functions. Among the most important unsaturated activation functions is the linear rectification function (ReLU). As shown in its mathematical expression is as follows:

f(x)=max (0, x)

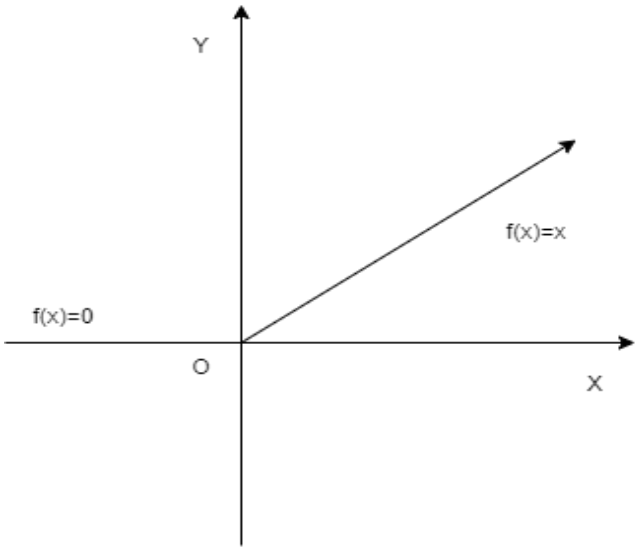

Fig: ReLU function image.

3. Pooling Layer

These days, convolutional neural networks frequently use it as one of their constituent parts. In order to reduce overfitting, the amount of data and parameters are compressed by placing the pooling layer between successive convolutional layers. The pooling layer's primary job is to compress images if that's the input type.The pooling layer's primary job when the input is an image is to compress it.

By performing collective statistical operations on the special diagnosis at various positions in the local area of the image, the pooling layer can effectively reduce the size of the matrix. This reduces the parameters in the final fully connected layer, speeds up calculation speed, and lessens the excessive sensitivity of the convolutional layer to the image position. The common operations of the pooling layer include the following: max-pooling, average pooling, Spatial Pyramid Pooling etc.

Architecture of CNN:

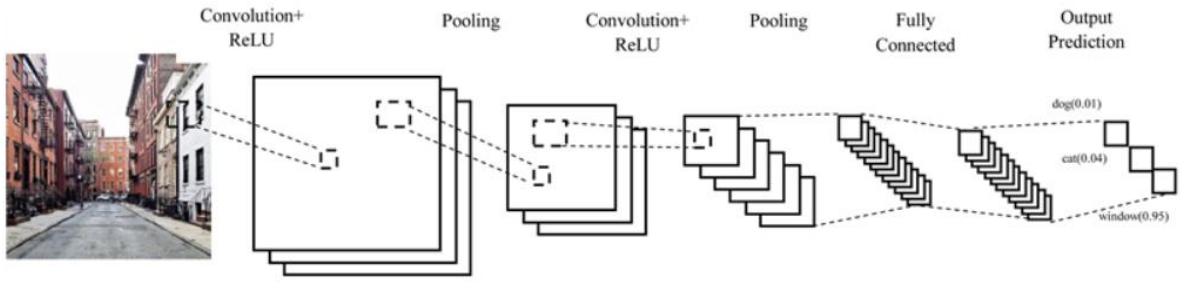

Fig: Architecture of CNN:

Working procedure of CNN

Step:

1. import the necessary libraries
2. set the parameter
3. define the kernel
4. Load the image and plot it.
5. Reformat the image
6. Apply convolution layer operation and plot the output image.
7. Apply activation layer operation and plot the output image.
8. Apply pooling layer operation and plot the output image.

Flowchart:

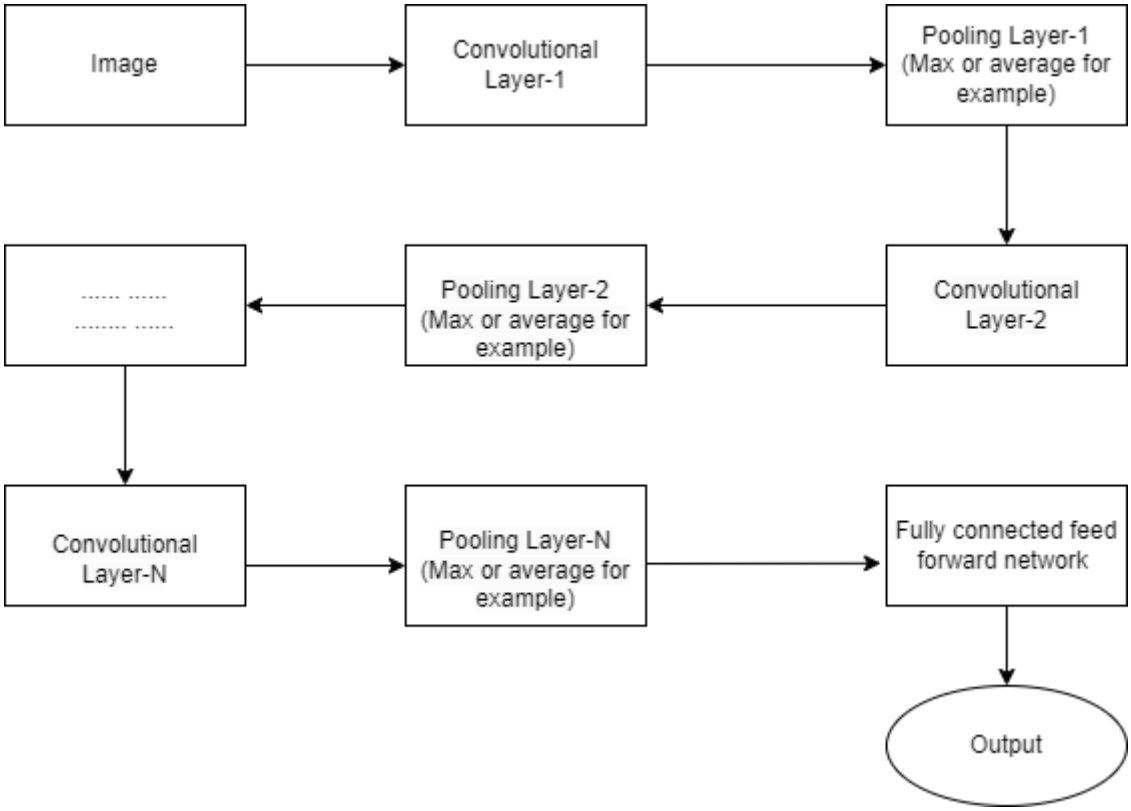

Fig:Flow chart of CNN network

Here are some applications of CNNs:

Image Classification:

CNNs are frequently used in image classification applications, where the network is trained to classify images into various groups. This program is widely used in many different domains, from diagnosing medical images to recognizing objects in photos.

Object Detection:

CNNs are useful for locating and identifying objects in pictures. CNNs are used in popular object detection frameworks such as Faster R-CNN and YOLO (You Only Look Once). Robotics, autonomous cars, and surveillance systems are a few examples of applications.

Facial Recognition:

Facial recognition systems use CNNs to recognize and verify people. They can be used for social media tagging, access control, and security systems.

D2DET:

A deformable two-stage detector is a type of object detection model used in deep learning for computer vision tasks. It combines two key concepts: the two-stage architecture and deformable convolutional networks.

1.Two-stage architecture: Object detection models typically follow a two-stage approach. In the first stage, they generate region proposals, which are potential bounding boxes that may contain objects. In the second stage, these proposed regions are classified and refined to produce the final object detections.

2.Deformable Convolutional Networks (DCN): Deformable Convolutional Networks are a type of convolutional layer that can adaptively adjust their receptive fields based on the features within an image. This allows them to capture more accurateand flexible information about object shapes, especially when objects are occluded, deformed, or occur at various scales.

Working procedure of D2Det algorithm:

Certainly, here's a step-by-step procedure for the working of a deformable two-stage detector algorithm:

Step 1: Input Image

Begin with an input image that you want to perform object detection on.

Step 2: Feature Extraction

Pass the input image through a convolutional neural network (CNN) to extract feature maps.These feature maps capture visual information from the image.

Step 3: Region Proposal Network (RPN)

The RPN operates on the feature maps and generates a set of anchor boxes (potential object bounding boxes) at various scales and aspect ratios. Each anchor box is associated with a score indicating the likelihood of containing an object.Apply deformable convolutions in the RPN to adaptively adjust the receptive fields and capture more relevant information.

Step 4: Anchor Scoring

Assign objectness scores to each anchor box based on how likely they are to contain an object. This is done using a classification layer.

Step 5: Anchor Refinement

Predict adjustments (bounding box regressions) for the anchor boxes to better align them with the true object locations. This is done using a regression layer.

Step 6: Anchor Selection

Filter the anchor boxes based on their objectness scores, selecting the top-ranked anchors as proposals for further processing. These proposals represent regions likely to contain objects.

Step 7: Region-of-Interest (RoI) Pooling

Extract fixed-size feature vectors for each selected proposal by applying RoI pooling or a similar technique on the feature maps. This step ensures that all proposals have consistent input sizes for further processing, regardless of their original sizes.

Step 8: Deformable Convolutional Networks (DCN)

Pass the RoI features through deformable convolutional layers, which adaptively adjust their receptive fields. Deformable convolutions help capture information from different position within the RoI, which is particularly beneficial for deformable or occluded objects.

Step 9: Object Classification

Use a classifier network to assign a class label to each RoI, determining the type of object contained in the proposal.

Step 10: Bounding Box Regression

Apply a regressor network to refine the coordinates of the bounding boxes associated with the RoIs. This step helps adjust the bounding boxes to better fit the precise object locations.

Step 11: Post-processing

Perform non-maximum suppression (NMS) to remove redundant or highly overlapping bounding boxes.This ensures that only the most confident detections are retained and eliminates duplicates.

Step 12: Output

Provide the final set of detected objects, along with their class labels and refined bounding box coordinates, as the output of the deformable two-stage detector.

The deformable two-stage detector algorithm combines the strengths of deformable convolutional layers with the two-stage object detection approach to achieve improved accuracy in detecting and localizing objects, especially in cases where objects exhibit deformations, occlusions, or variations in scale.

SSD:

SSD operates as a one-shot detector. It predicts the boundary boxes and the classes directly from feature maps in a single pass and does not have a gave region proposal network. SSD adds offsets to default boundary boxes and small convolutional filters to predict object classes in order to increase accuracy.

The SSD object detection composes of 2 parts:

Extract feature maps, and Apply convolution filters to detect objects.

SSD extracts feature maps using VGG16. Next, it makes use of the Conv4_3 layer to detect objects. As an example, Four object predictions are made for each cell, also known as location. To improve accuracy, SSD can be trained from beginning to end. SSD has better coverage on location, scale, and aspect ratios and makes more predictions.

Working procedure of SSD:

A well-liked object detection model for real-time object detection in photos and videos via deep learning is the Single Shot MultiBox Detector (SSD). SSDs are made to anticipate a set of bounding boxes and the class labels that correspond with them in order to identify objects in an image.

Here's an overview of how an SSD works in deep learning:

1. Input Image:

A neural network is used by the SSD model to extract features from an input image of any size. The backbone network is usually a convolutional neural network (CNN) that has been pre-trained, like VGG16 or ResNet. CNN uses several scales to extract features from the picture.

2. Feature Maps:

Various-sized feature maps are produced as the image is processed by CNN. Information is captured at different spatial resolutions by these feature maps. In general, higher-level features are captured by the deeper layers of the CNN, whereas lower-level features are captured by the shallower layers.

3. Multi-scale Feature Fusion:

SSD uses convolutional layers with different kernel sizes and moves to obtain feature maps at different scales. These layers are in charge of combining features from the backbone network's various layers. To identify objects of various sizes, feature maps at various scales are utilized.

4. Default Anchor Boxes:

A set of default anchor boxes, also called default bounding boxes, with various aspect ratios and scales are defined for every location in the feature maps. To begin object detection, these anchor boxes are used. In order to better fit the real objects in the picture, the SSD model predicts offsets, or deltas, for these anchor boxes.

5. Object Detection Head:

After feature extraction and anchor box definition, the SSD network splits into two parallel subnetworks:

- Localization Head: This part predicts the offsets (deltas) for each anchor box to adjust their positions and sizes to match the ground-truth objects in the image.

- Classification Head: This part predicts the class probabilities for each anchor box, indicating the likelihood that an object of a particular class is present in that box.

6. Non-Maximum Suppression (NMS):

Following the receipt of predictions from the localization and classification heads, low-confidence and redundant detections are filtered out using a post-processing technique known as non-maximum suppression. NMS makes sure that as final detections, only the most certain and non-overlapping bounding boxes are kept.

7. Output:

An array of bounding boxes with the corresponding class labels and confidence scores is the SSD model's final result. The objects that were found in the input image are represented by these bounding boxes.

Because SSD is effective at predicting class labels and bounding boxes at multiple scales in a single forward pass, it can detect objects in real time. This makes it a well-liked option for many applications, such as image recognition, autonomous driving, and surveillance. By modifying the quantity and dimensions of the default anchor boxes and fine-tuning the model on particular datasets, the SSD architecture's flexibility allows it to be tailored for a variety of object detection applications.

YOLOv:

YOLO is an algorithm that provides real-time object detection using neural networks. The accuracy and speed of this algorithm makes it popular.

The abbreviation YOLO stands for "You Only Look Once." This algorithm (in real-time) finds and recognizes different objects in an image. YOLO employs object detection as a regression problem, resulting in the class probabilities of the identified images. Convolutional neural networks (CNN) are used by the YOLO algorithm to detect objects in real-time. As the name implies, the algorithm can detect objects with just one forward propagation via a neural network. This indicates that a single algorithm run is used to predict the entire image. Multiple class probabilities and bounding boxes are simultaneously predicted by the CNN.

There are several variations of the YOLO algorithm. Tiny YOLO and YOLOv3 are a couple of the popular ones.

Working procedure of yolo algorithms:

YOLO algorithm works using the following three techniques:

1. Residual blocks
2. Bounding box regression
3. Intersection Over Union (IOU)

Residual blocks

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

Bounding box regression

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

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

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

- Bounding box center (bx,by)

Intersection over union (IOU)

In object detection, the phenomenon known as intersection over union (IOU) characterizes how boxes overlap. YOLO creates an output box that exactly covers the objects by using IOU. The task of predicting the bounding boxes and their confidence scores falls on each grid cell. If the expected and actual bounding boxes match, the IOU is equal to 1. Bounding boxes that are not equal to the actual box are removed by this mechanism.

Combination of the three techniques

The image is first split up into grid cells. B bounding boxes are predicted by each grid cell, along with their confidence scores. To determine each object's class, the cells make predictions about the class probabilities.

A car, a dog, and a bicycle are just a few examples of the at least three classes of objects that are visible. A single convolutional neural network is used to make all of the predictions at the same time.

The predicted bounding boxes and the actual boxes of the objects are guaranteed to be equal by intersection over union. The effect removes irrelevant bounding boxes that don't match the object's dimensions (height and width). The final detection will be made up of distinct bounding boxes that precisely match the objects.

For instance, the yellow bounding box surrounds the bicycle and the pink bounding box covers the car. The blue bounding box has been used to highlight the dog.

YOLO algorithm can be applied in the following fields:

Autonomous vehicles can utilize the YOLO algorithm to identify nearby objects, including people, cars, and parking signals. In forests, this algorithm is used to identify different kinds of animals. Journalists and wildlife rangers use this kind of detection to recognize animals in photos.

Proposed System Model

In our proposed topic we used four algorithms. Here is a flowchart of our proposed model

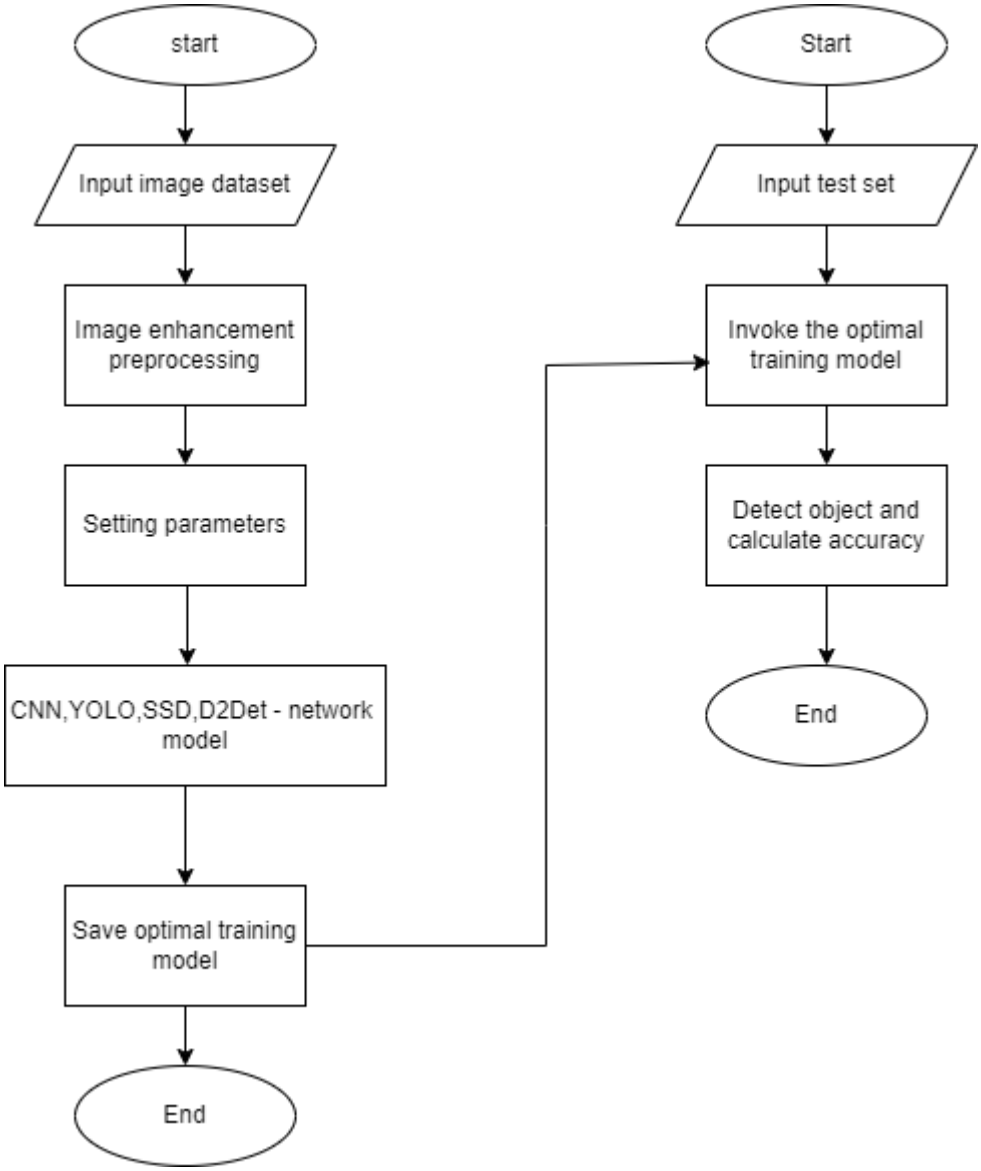

Fig:Proposed model of our project

Conclusion:

Comparing CNN (Convolutional Neural Networks), YOLO (You Only Look Once), D2Det (Deformable Two-Stage Detector), and SSD (Single Shot MultiBox Detector) algorithms for object detection in deep learning involves considering their strengths, weaknesses, and use cases. Each algorithm has its unique characteristics and is suited for specific scenarios.In future we will merge this algorithms and will try to make a model.We will try to use it for real image.We will also use it in disease detection. We will make a device a which will recognize leaf name.