**Comparison of algorithms in object detection**

**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.

**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.

**What we need:**

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. This 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.

**Use of best Algorithms:**

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.

Report outline:

Chapter 1- Introduction

This section of the report describes the studies explanation,goals,scope and possible impacts.

Chapter 2- Literature Riview

This section of the report describes the studies explanation,goals,scope and possible impacts.

Chapter 2- Literature Riview

This section of the report describes the studies explanation,goals,scope and possible impacts

Chapter 2- Literature Riview

This section of the report describes the studies explanation,goals,scope and possible impacts

Chapter 2- Literature Riview

This section of the report describes the studies explanation,goals,scope and possible impacts

**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>

**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.