Object Detection using Artificial Intelligence: A Review

[1]Shibani Singh, [2]Mimi Ton Devi, [3]Kandarpa Kalita

[1]Student, [2]Student, [3]Assistant Professor
Department of Computer Science and Engineering,
Assam Downtown University, Assam, India

Abstract: **Image processing is the technique of manipulating an image to both decorate the great or extract applicable facts from it. It can be defined as the technical evaluation of an image through the usage of complicated algorithms. Here, image is used as the input, where the beneficial information returns as the output. Meanwhile, the artificial Intelligence industry is also turning to a huge boom curve. According to Forbes, it turned into believed that artificial intelligence and machine learning could have the capacity to create an additional $2.6T in price by 2020 in advertising and sales, and up to $2T in production and deliver chain planning.**

Index Terms: Artificial Intelligence, Object Detection, Convolutional Neural Network, R-CNN [1], Fast R-CNN [2], Faster R-CNN [3], YOLO [4].

1. INTRODUCTION

Artificial Intelligence [5] (AI) is intelligence demonstrated through machines, rather than natural intelligence displayed through animals including humans. Artificial intelligence is described as the study of "Intelligent agents", any system that perceives its environment and takes actions that maximize its chance of achieving goals.

Often what we refer to as AI is simply one component of AI, such as machine learning. AI usually uses a foundation of specialized hardware and software for writing and training machine learning algorithms. No one programming language is synonymous with AI, however few, which includes Python, R and Java are popular.

In general, AI systems work by ingesting massive quantities of classified training records, analyzing the data for correlations and patterns, and the usage of those patterns to make predictions about future states. AI programming basically focuses on three cognitive skills:

- Learning processes: This aspect of AI programming focuses on obtaining records and developing regulations for a way to show the records into actionable facts. The regulations that are known as algorithms offer computing gadgets with step-by-step instructions for a way to finish a particular task.
- Reasoning processes: This aspect of AI programming focuses on selecting the proper set of algorithms to attain a preferred outcome.
- Self-correction processes: This aspect of AI programming is designed to continually fine-tune algorithms and ensure that they offer the maximum correct outcomes possible.

IMPORTANCE OF ARTIFICIAL INTELLIGENCE

AI is crucial due to the fact that it may deliver businesses insights into their operations that they may not have been aware of previously. In a few cases, AI can carry out duties better as compared to humans. Particularly in relation to repetitive, detail-oriented duties such as analyzing large numbers of legal documents to ensure relevant fields are filled in properly, AI tools often completes tasks quickly and also the number of errors are less. Prior to the current wave of AI, it was often hard to imagine using computer software to connect riders to taxis, but today Uber has become one of the biggest companies in the world by doing just that. It utilizes highly experienced machine learning algorithms to predict when people are likely to need rides in certain areas, which helps proactively get drivers on the road before they are needed. As another example includes Google, that has become one of the largest players for a range of online services by using machine learning to understand how people use their services and then improve them.

ADVANTAGES AND DISADVANTAGES OF ARTIFICIAL INTELLIGENCE

Artificial neural networks and deep learning artificial intelligence technologies are evolving much faster and this is because AI processes large amounts of data quickly and is capable of making predictions more accurately as compared to humans. The advantages and disadvantages of Artificial Intelligence are mentioned below:

Advantages

- AI is good at detail-oriented jobs
- Delivers consistent outcomes
- AI-powered virtual agents are always available

Disadvantages

- It is very much expensive
- It requires deep technical expertise

TYPES OF ARTIFICIAL INTELLIGENCE [6]An assistant professor of integrative biology and computer science and engineering at Michigan State University named Arend Hintze, in 2016 article explained that artificial intelligence can be categorized into four categories on the basis of functionalities. These four types of artificial intelligence include the following:

1. Reactive Machines: These types of artificial intelligence are task specific and have no memory. The example of this type of AI is Deep Blue. It is the IBM chess program that beat Garry Kasparov in the 1990s. Deep Blue can make predictions and identify pieces on the chessboard, but as it has no memory, it cannot use past experience to inform future ones.
2. Limited memory: This type of AI systems has memory and as a result they can use the past experiences to inform about the future decisions. In self-driving cars, some of the decision-making functions are designed using this type of AI.
3. Theory of mind: The term 'Theory of mind' is a psychological term and when applied to AI, it means that the system would have the social intelligence to understand emotions. This AI will be able to judge human intentions and predict and conclude their behavior. It is a necessary skill for AI system for becoming an integral member of human teams.
4. Self-awareness: In this category of AI systems, it has a sense of self and it provides them consciousness.

APPLICATIONS OF ARTIFICIAL INTELLIGENCE [7]

Artificial Intelligence has made its way into a good kind of fields. Some of these fields includes:

- AI in HealthCare: The most important bets of this category are on improving patient outcomes and reducing costs. Various healthcare companies are applying machine learning to form better and faster diagnoses compared to humans. IBM is one in all best-known healthcare technologies. It can understand the natural language and responds to the questions asked of it. To predict, fight and understand pandemics like as COVID-19, an array of AI technologies is being used.

- AI in enterprises: Machine learning algorithms are getting used to seek out information on the way to serve their customers better and maintain proper Customer Relationship Management (CRM). Chatbots are designed to supply immediate service to the customers just in case needed.

- AI in education: AI can help students in learning better and faster because it provides high-quality learning materials and instructions. There are various students who need assistance outside classroom. With the help of AI chatbots, those students can get help in an exceedingly distinctive way of learning. These tools can help the students to determine their weak points and work upon them.

- AI in security: AI and machine learning are at the highest list in case security is concerned. Organizations use machine learning in security information and event management (SIEM) and related areas to detect defects and identify the activities associated with threats.

The paper represents a survey of object detection using AI. It is divided into multiple sections. The section two describes the term object detection, the various modes and kinds of object detection. It also states the importance of object detection. The third section discusses the literature review of object detection using AI. It highlights the varied works done previously associated to this topic. Section four describes the various algorithms used in object detection. Here, in this section we've discussed R-CNN, Fast R-CNN, Faster R-CNN and YOLO (You Only Look Once). The next section i.e., section five, provides a difference table about the various algorithms. The last section, i.e., section 6, concludes the overall article.

2. OBJECT DETECTION [8]

Object detection is a technique, which works to identify and locate objects within an image or video. With this kind of identification and localization, object detection can be used to count objects in a scene and also determine and track their accurate locations, all while accurately labelling them.

Most of the times, object detection is confused with image recognition. Image recognition assigns a label to an image. For example- an image of dog receives the label "dog". An image of two dogs, still receives the label "dog". On the other hand, object detection draws a box around each dog and marks the box "dog". As a result, object detection provides more information about an image when put next to image recognition.

- **MODES AND TYPES OF OBJECT DETECTION**

  Basically, object detection is categorized into two approaches- machine learning-based approaches and deep learning-based approaches.

  In ML-based approaches, various features of an image, such as the color histogram or edges, use computer vision techniques to look at them and identify groups of pixels that may belong to an image. These features are then fed into a regression model which predicts the location and label of the object.

  Deep learning-based approaches, on the other hand use convolution neural networks (CNNs) to perform end-to-end, unsupervised object detection. In this, features don't need to be defined and extracted separately.

- **IMPORTANCE OF OBJECT DETECTION**

  Object detection, which is inseparably linked to other similar computer vision techniques like image recognition and image segmentation, helps us to understand and analyze scenes in images or video.

  It is mostly used in computer vision tasks like image annotation, vehicle counting, activity recognition, face detection, face recognition.

3. LITERATURE REVIEW

In general, a lot many works have been done in this field of object detection using artificial intelligence. Object detection is a computer vision technology related to computer vision and image processing that deals with detecting instances of semantic

objects of a certain class (such as humans, buildings, or cars) in digital images and videos. Some of the related works which were done previously in the same area includes the following:

Joshi. R.C et. al [9], in their paper 'efficient multi-object detection and smart navigation using artificial intelligence for visually impaired people', proposed an artificial intelligence-based fully automatic assistive technology to recognize different object, and also provided auditory inputs to the user in real time, which gives better understanding to the visually impaired person about their surroundings. Also, they trained a deep-learning model with multiple images of objects that are highly applicable to the visually impaired people. In addition to computer vision-based techniques for object recognition, they integrated a distance-measuring sensor to make the device more comprehensive by recognizing obstacles which moving from one place to another.

Another paper was presented by Ashish Kumar [10] named 'Artificial Intelligence in Object Detection'. In this report, he presented major developments in this research field. His main research, which is based on face and motion detection is explained a little bit in this paper. Different architectures based on convolutional neural network is studied and different methodologies for object detection are presented and compared in this report. Convolution neural network (CNN or ConvNet) is a class of Artificial Neural Network (ANN) in deep learning, which is most commonly applied to analyze visual imagery. They are mostly used in applications such as image and video recognition, image classification, image segmentation, medical image analysis and many more.

Alexandrina. E.P et al [11], in their paper 'Image Processing using Artificial Neural Networks' presented the current status of Artificial Neural Networks used for image processing. They included issued resolved with artificial neural networks in civil engineering, geotechnical engineering, transportation, in the field of mechanics, military defence and industrial inspection, medical field and items solved with artificial neural networks in automation.

'Object Detection in 20 years: A survey' [12], was presented by Zhengxia Zou, Zhenwei Shi, Member, IEEE, Yuhong Guo, and Jieping Ye, Senior Member, IEEE. In this paper, they reviewed 400+ papers of object detection in the light of its technical evolution, spanning over a quarter-century's time (from the 1990s to 2019). They included a number of topics in their paper which includes milestone detectors in history, detection datasets, metrics, fundamental building blocks of the detection system, speed up techniques and the recent state of the art detection methods. They also reviewed some important detection applications such as face detection, text detection etc. Their also discussed the challenges currently me by the community and how these detectors can be further extended and improved.

Another paper was presented by Tu N. Nguyen et. al, titled 'Cyber security of smart grid: attacks and defense' [13]. They recommended a three-stage approach dubbed innovative identification method to increase the evaluation of projects for the vehicles of various scales. The researchers in their first step, used a picture cropping approach to split the system into multiple patches, so that they can prevent vehicle deformation in assessment scale and maintain more data in aerial photographs. In the second step, the primary image and two patches were combined into a batch and detected the vehicles with a Convolutional Neural Network (CNN).

In another research, which was done by P. Viola and M. Jones [14], around 18 years ago, achieved real-time detection of human faces for the first time without any constraints (e.g., skin color segmentation). The detector was ten to hundreds time faster than any other algorithms in its time under comparable detection accuracy. The detection algorithm was referred to "Viola-Joned (VJ) detector", in the memory of the significant contribution given by the authors.

Another researcher named R. Joseph et al. [15] in 2015, proposed a detector called YOLO (You Only Look Once). It is extremely fast. It follows a very different philosophy of applying a single neural network to the full image. This network further divides the image into regions and predicts bounding boxes and probabilities for each region simultaneously.

Anamika Dhillon and Gyanendra K Verma, in their paper named 'Convolutional neural network: a review of models, methodologies and applications to object detection' [16], provided a detailed review of various deep architectures and models and also highlighted the characteristics of particular models. At first, they described the functioning of Convolutional Neural Network (CNN) architectures and its components which was further followed by detailed description of the various CNN models which included LeNet model, AlexNet, ZFNet, GoogleNet, VGGNet, ResNet, ResNeXt, SENet, DenseNet, Xception and PNAS/ENAS. They mainly focused on the application of deep learning architectures to three major applications, namely, wild animal detection, small arm detection and human being detection.

In another paper, a group of researchers namely, S A Sanchez et. al, titled 'Comparison of performance metrics pretrained models for object detection using the TensorFlow framework [17], reviewed the state of the art, related to the performance of pre-trained models for the detection of objects in order to make a comparison of these algorithms in terms of reliability, accuracy, time processed and problems detected. The models they consulted were based on python programming language, the use of libraries based on TensorFlow, OpenCV and free image databases. For their research, they reviewed different pre-trained models for the object detection which included R-CNN, R-FCN, SSD and YOLO with different extractors of characteristics such as VGG16, ResNet, Inception, MobileNet.

S N David Chua, S. F. Lim and T K Chang, in their paper, titled 'development of a Child detection System with Artificial Intelligence using Object Detection Method' [18], proposed a technique for child detection with transfer learning. They established a real-time child detection system that consisted of a camera as an input medium, a classifier as a detector to detect the presence of a child and a triggering system in audio and visual forms. As a starting point of the training process, they trained the modern convolutional object detector, SSD Mobilenet v1 with Microsoft Common Objects in context (MS COCO) dataset. Then the model was assessed and retained to possess the ability to distinguish human into an adult or a child. They measured the accuracy of the model by counting the pixels labelled correctly per class. Based on the mean Average Precision (mAP), their detection system provided an overall precision of 0.969 and the experimental results obtained showed a precision of 0.883, giving an error of less than ten percent.

Another paper was written by Ruixin Yang and Yingyan Yu on the topic 'Artificial convolutional neural network in object detection and semantic segmentation for medical imaging analysis' [19]. In today's era of digital medicine, a large number of medical imaging are produced every day and as a result the demand for intelligent equipment's are increasing for adjuvant diagnosis to help medical doctors with different disciplines. In their review paper, the introduced the progression of object detection and semantic segmentation in medical imaging study. They also discussed on how to define the location and boundary of diseases accurately.

Xin Qiu and Chun Yuan, in their paper 'Improving Object Detection with Convolutional Neural Network via Iterative Mechanism' [20], proposed to use the iterative mechanism to improve the object detection performance of the CNN algorithms. The main contribution of their work included two aspects: firstly, they trained an iterative version of Faster R-CNN to show the application of the iterative mechanism in improving the localization accuracy; secondly, they presented a prototype CNN model that iteratively searched for objects on a very simple database to generate proposals. The thoughtful experiments they did on object detection benchmark datasets showed that the two iterative models they proposed, consistently improved the performance of the baseline methods e.g., in PASCAL VOC2007 test set, their iterative version of Faster R-CNN had 0.7115 mAP, which was about 1.5 points higher than the baseline Faster R-CNN (0.6959 mAP).

Yunchao Bai, Libo Zhang, Tianxing Wang and Xianzhong Zhou [21], in their research, proposed a skeleton-based object detection method to recognize the dynamic gesture. They utilized the Kinect depth camera to capture the skeleton of human beings in the dynamic gesture motions. Then they marked and trained all the gestures of the same dynamic gestures. After this, a modified Single Shot MultiBox Detector (SSD) network was adopted to locate the arm in the skeleton images. Then, the dynamic gesture was recognized based on the arm skeleton movements in the motion. In order to balance the precision and recognition time in the identification period, a corresponding comprehensive index was constructed to find out suitable proportion of arm skeleton images for detection by the trained neural network. They found that, it could reduce the redundancy and improve the detection efficiency. It would also decrease the impact of noise and enhance the recognition accuracy. At the end, the effectiveness of the proposed method was validated by an experimental analysis.

Aman Dureja in his research, 'A Review: Image classification and Object Detection with Deep Learning' [22], included a thorough analysis of the different deep learning architectures and frameworks illustrating the model specifications. The main goal of his review paper was to bring some prominent models and techniques back into the light and provide their results on different popular datasets.

Farhana Sultana, Abu Sufian and Paramartha Dutta, in their article 'A review of object detection models based on convolutional neural network' [23], reviewed some popular state-of-the-art object detection models based on convolutional neural networks. They distinguished and categorized the detection models according to two different approaches which included one-stage approach and two-stage approach. Then after, they explored various gradual developments in one-stage object detectors from YOLO to RefineDet and in two-stage object detection models from R-CNN to latest mask R-CNN. They also focused on the training details of each model and also made a comparison among those models.

Eric Crawford and Joelle Pineau in their work 'Spatially Invariant Unsupervised Object Detection with Convolutional Neural Networks' [24], developed a neural network architecture that effectively addressed the large-image, many object setting. They combined the ideas from Attend, Infer, Repeat (AIR), which performs unsupervised object detection but did not scale well, with recent developments in supervised object detection. The researchers replaced AIR's core recurrent network with a convolutional network and made use of an object-specification scheme that described the location of objects with respect to local grid cells instead of the image as a whole. After this, they demonstrated a number of features of their architecture and unlike AIR, their architecture was able to discover and detect objects in large, many-object scenes. It also had a significant ability to generalize to images that are larger and contain more objects than images encountered during training. It was also able to discover and detect objects with enough accuracy to facilitate non-trivial downstream processing.

Meghna Raj Saxena, Akarsh Pathak, Aditya Pratap Singh and Ishika Shukla in their paper 'Real time object detection using machine learning and OpenCV' [25], addressed the HAAR-CASCADE classifier. Their main focus was on the case study of a face detection and object detection such as watch detection, pen detection. Their attempt was to create their own HAAR classifier using OpenCV. They also demonstrated some of the fundamental techniques implemented in Python OpenCV and MATLAB which can be used in human detection and tracking in video.

Kuo-Ching Hung, Meng-Chun Lin and Sheng-Fuu Lin, in their article 'A Novel Power Saving Reversing Camera System with Artificial Intelligence Object Detection' [26], implemented a reversing camera system with AI to increase the information of the reversing image. The system they implemented consisted of an image processing chip (IPC) with a wide-angle image distortion correction and an image buffer controller, a low-power KL520 chip and an optimized artificial intelligence model, MobileNetV2-YOLOV3-Optimized (MNYLO). The results of their experiment showed three advantages of their system. The first advantage was that through the image distortion correction of IPC, they could restore the distorted reversing image. The second was that by using a public dataset and collected images of different weathers for artificial intelligence model training, their system did not need to use image algorithms that eliminated bad weathers such as rain, fog and snow to restore polluted images. Objects were still detected by their system in mages contaminated by weather. The last advantage was that when compared with the AI model Tiny_YOLOV3, not only the parameters of their MNYLO reduced by 72.3%, but also the amount of calculation reduced by 86.4% and also the object detection rate was maintained and avoided sharp drops.

Zhihui Li et al. in their paper, 'Zero-Shot object detection with textual descriptions' [27], addressed the challenging problem of zero-shot object detection with neural language description, which aimed to simultaneously detect and recognize novel concept instances with textual descriptions. A novel deep learning framework to jointly learn visual units, visual-unit attention and word-level attention, which were combined to achieve word-proposal affinity by an element-wise multiplication were proposed by them.

Abidha Pandey, Manish Puri and Aparna Varde, in their work 'Object detection with neural models, deep learning and common sense to aid smart mobility' [28], mainly focused on object detection which could potentially enhance autonomous driving and other types of automation in transportation system. They also provided expanded analysis of recent object detection techniques including neural models, deep learning and related advances. They highlighted a novel object detection system called YOLO (You Only Look Once) and conducted its performance evaluation on real-time data. Then they pointed out the challenges in this field and explored the use of Commonsense Knowledge (CSK) in object detection with neural models and deep learning. They also explained how their work will potentially enhance autonomous vehicles and transportation systems.

Junwei Han et al., in their research paper 'Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning' [29], proposed a novel and effective geospatial object detection framework by combining the weakly supervised learning (WSL) and high-level feature learning. At first, they adopted a deep Boltzmann machine to derive the spatial and structural information encoded in the low-level and middle-level features to effectively describe objects in optical Remote Sensing Images (RSI). Then, a novel weakly supervised learning (WSL) approach was presented to object detection where the training sets required only binary labels that indicated whether an image contained the targeted images or not. They also developed an evaluation criterion which could detect model drift and cease the iterative learning.

Jisoo Jeong et al. [30], in their paper work proposed a Consistency-based Semi-supervised learning method for object detection (CSD). CSD is a way of using consistency constraints as tool for enhancing detection performance by making full use of available unlabeled data. They also proposed Background Elimination (BE) to avoid the negative effect of the predominant backgrounds on the detection performance.

Yang Yang, Guang Shu and Mubarak Shah [31], in their work proposed an approach to boost the performance of generic object detectors on videos by learning video specific features using a deep neural network. The main aim of their proposed approach was that an object appearing in different frames of a video clip should share similar features, which can be learned to build better detectors. Moreover, their method didn't require any extra annotations or utilized temporal correspondence. At first, they started with the high-confidence detections from a generic detector, then they learned the new video specific features and refined the detection scores. They also proposed a new feature learning method using a deep neural network based on auto encoders in order to learn discriminative and compact features.

Sabit UI Hussain et al., proposed a work in 2012 'Machine learning method for visual object detection' [32]. Their main aim was to provide computers with artificial visual systems having human-like image understanding capabilities so that they can achieve their goals. Their fundamental task was the interpretation and labeling of scene content.

Marc 'Aurelio et al. [33], in their research presented an unsupervised method for learning a hierarchy of sparse feature detectors that are invariant to small shifts and distortions. Their resulting feature extractor consisted of multiple convolution filters, which was followed by a feature-pooling layer.

Jahanzaib Latif et al., in their research 'Medical imaging using machine learning and deep learning algorithms' [34], surveyed image classification, object detection, pattern recognition, reasoning etc. concepts in medical imaging. Their main aim was to focus on machine learning and deep learning techniques being used in medical images.

S Pu et al., presented research on 'Unsupervised object detection with scene-adaptive' [35]. They proposed a novel scene-adaptive evolution unsupervised video object detection algorithm. This algorithm, which they proposed could decrease the impact of scene changes through the concept of object groups.

L. Liu et al., in their article 'Deep Learning for Generic Object Detection' [36], provided a comprehensive survey of the recent achievements in this field brought about by deep learning techniques. More than 300 research contributions are included in their survey, covering many aspects of generic object detection: detection frameworks, object feature representation, object proposal generation, context modeling, training strategies, and evaluation metrics.

S. kihyuk et al., in their article 'A Semi-Supervised Learning Framework for object detection' [37], proposed STAC (Self-Training and Augmentation driven Consistency regularization). It's a simple but effective SSL framework for visual object detection together with a data augmentation strategy. They proposed experimental protocols to evaluate the performance of semi-supervised object detection MS-COCO and also showed the efficacy of STAC on both MS-COCO and VOC07.

Jeong. J. et. al., in their paper 'Consistency-based Semi-Supervised learning for Object detection' [38], proposed a consistency based semi-supervised learning method for object detection (CSD). They also proposed Background Elimination (BE) process for avoiding the negative effect of the predominant backgrounds on the detection performance.

Han. J et. al. [39], in their research, 'Object Detection in Optical Remote Sensing Images based on weakly Supervised Learning and High-Level Feature Learning', proposed a geospatial object detection framework by combining weakly supervised Learning (WSL) and high-level feature learning. At first, they adopted a deep Boltzmann machine to infer the spatial and structural information encoded in the low-level and middle-level features which effectively described objects in optical RSIs. Next, a novel WSL approach was presented to object detection where the training sets required only binary labels which indicated whether an image contain the targeted object or not. They also developed a novel evaluation criterion to detect model drift and cease the iterative learning.Luo. A. et. al., in their research paper 'Webly-Supervised learning for salient object detection' [40], proposed an approach that could utilize large amounts of web data for learning a deep salient object detection model. They introduced a novel quality evaluation method that could help in picking out images with high-quality masks for training. They also presented a self-training approach to boost the performance of their network by selecting more hard web images for training.

Ali. S. et. al., in their paper 'A Supervised learning framework for generic object detection in images' [41], presented a novel framework for object class detection that combined both the feature of reduction and selection abilities of kernel PCA and AdaBoost respectively. They successfully tested the proposed method on wide range of object classes (cars, air-planes, pedestrians, motorcycles, etc.) using standard datasets and also achieved detection rates of above 95% with minimal false alarm rates in most object categories.

Rhee. P.K. et. al., in their article 'Active and semi-supervised learning for object detection with imperfect data' [42], addressed the combination of the active learning (AL) and semi-supervised learning (SSL), called ASSL, to leverage the strong points of both the learning paradigms so that the performance of object detection is improved. Their proposed method demonstrated outstanding performance when compared with state-of-art methods on the challenging Caltech pedestrian detection dataset. It reduced the miss rate to 12.2%, which was significantly smaller than current state-of-art.

Mitash. C. et. al., in their research 'A self-supervised learning system for object detection using physics simulation multi-view pose estimation' [43], proposed an autonomous process for training a Convolutional Neural Networks (CNN) for object detection. Their focus was on detecting the objects which were placed in cluttered, tight environments, such as a shelf with multiple objects. The models were placed in physically realistic poses with respect to their environment to generate a labeled synthetic dataset.

Sharma. K.U. et. al., in their research 'A review and an approach for object detection in images' [44], presented a review of the various techniques that are used to detect an object, localize an object, categories an object, extract features and many more, in images and videos. They also presented an idea about the possible solutions for the multi-class object detection.

Xie. E. et. al., in their article 'DetCo: Unsupervised contrastive learning for object detection' [45], presented a simple yet effective self-supervised approach for object detection, called Detco. Unsupervised pre-training methods were designed for object detection but they were insufficient in image classification. However, DetCo transferred well on downstream instance-level dense prediction tasks and also maintained competitive image-level classification accuracy.

Nair. V. et. al., in their paper 'An Unsupervised, online learning framework for moving object detection' [46], presented a framework which learns the classifier online with automatically labeled data for specific case of detecting moving objects from video. Their framework was demonstrated on a person detection task for an office corridor scene. In this method, they used background subtraction to automatically label the training examples. The frameworks ran by itself on the scene video stream to train an accurate detector after the initial manual effort of implementing the labelling method.

Dai. Z. et. al., in their research paper 'UP-DETR: Unsupervised Pre-training for Object Detection with Transformers' [47], proposed a pretext task named random query patch detection to Unsupervisedly Pre-train DETR (UP-DETR) for object detection. They randomly cropped patches from the given image and fed them as queries to the decoder. Their model was pre-trained to detect the query patches from the original image.

Arruda. V.F. et. al., in their article 'Cross-Domain Car Detection using Unsupervised Image-to-Image Translation: From Day to Night' [48], presented a method for training a car detection system. Their proposed method achieved significant and consistent improvements, including the increasing by more than 10% of the detection performance as compared to the training with only the available annotated data i.e., day images.

Amin. P. et. al., presented a paper titled 'Object Detection using Machine Learning Technique' [49]. Their main aim was to build a system that could detect objects from image or a stream of images given to the system. Their system could also classify the object to the classes they belong to. For the object detection, they used python programming and a machine learning algorithm named YOLO (You Only Look Once) using convolutional neural network.

Szegedy. C. et. al., in their research titled 'Deep Nural Networks for Object Detection' [50], addressed the problem of object detection using DNNs. They defined multi-scale inference procedure that was able to produce high-resolution object detection at a low cost by a few network applications. They showed that the simple formulation of detection as DNN-base object mask regression could yield strong results when applied using a multi-scale course-to-fine procedure.

Galvez. R.L. et. al., in their research 'Object Detection Using Convolutional Neural Networks' [51], used convolutional neural network to detect objects in the environment. They compared two state-of the art models for object detection which included Single Shot Multi-Box Detector (SSD) with MobileNetV1 and a Faster Region-based Convolutional Neural Network (Faster-RCNN) with InceptionV2. The result they obtained showed that one model was ideal for real-time application because of speed and the other one was ideal for more accurate object detection.

Mittal. N. et. al., in their article 'Object Detection and Classification using YOLO' [52], demonstrated YOLO. Their paper concentrated on Convolutional Neural Networks (CNN), the fundamental structure of CNN, object location dependent on YOLO and also the library utilized in executing their task. They executed YOLO with the assistance of the open-source OpenCV library utilizing CNN.

4. OBJECT DETECTION ALGORITHMS [53]

Computer vision is one of the interdisciplinary fields that is gaining huge recognition in the recent years and also has expanded a lot. Object detection is one of the integral parts of computer vision. The difference between classification algorithms and object detection algorithms is that in object detection algorithms, a bounding box is drawn around the object of interest to locate it within the image. But there is a problem with this as the number of occurrences of the objects of interest is not fixed. Therefore, algorithms like R-CNN, Fast R-CNN, Faster R-CNN and YOLO have been developed to overcome this problem.

- **R-CNN:** To overcome the problem of selecting a huge number of regions, Ross Girshick et al. proposed a method. In this method, he used selective search to extract just 2000 regions from the image and that was called as region proposals. As a result, now we can just work with 2000 regions instead of trying to classify a huge number of regions.

Problems with R-CNN

1. A huge amount of time is taken to train the network as we will have to classify 2000 region proposals per image.
2. As it takes around 47 seconds for each image testing so, cannot be implemented in real time.

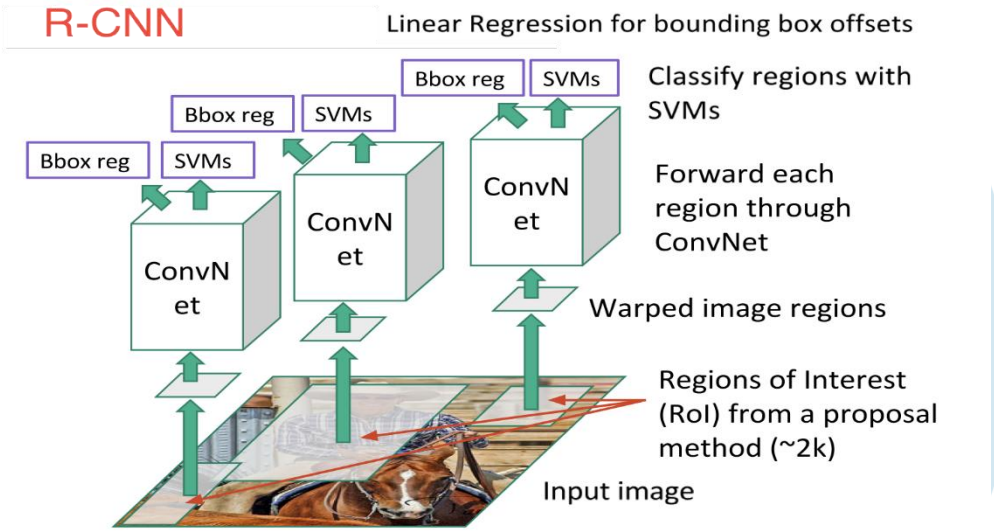

Figure 1: Object Detection with R-CNN.

https://stackoverflow.com/questions/43402760/object-detection-with-r-cnn

- **Fast R-CNN**: The same author of R-CNN, i.e., Ross Girshick, solved some of the R-CNN related drawbacks to build a faster version of object detection algorithm and it was named as Fast R-CNN. The reason behind Fast R-CNN being faster compared to R-CNN is that we don't need to feed 2000 region proposals to the convolutional neural network every time. Instead, it is done only one time per image and the map is generated from it. But, at the time of testing, when we look at the performance of Fast R-CNN, including region proposals, it slows down the algorithm when compared to not using region proposals.

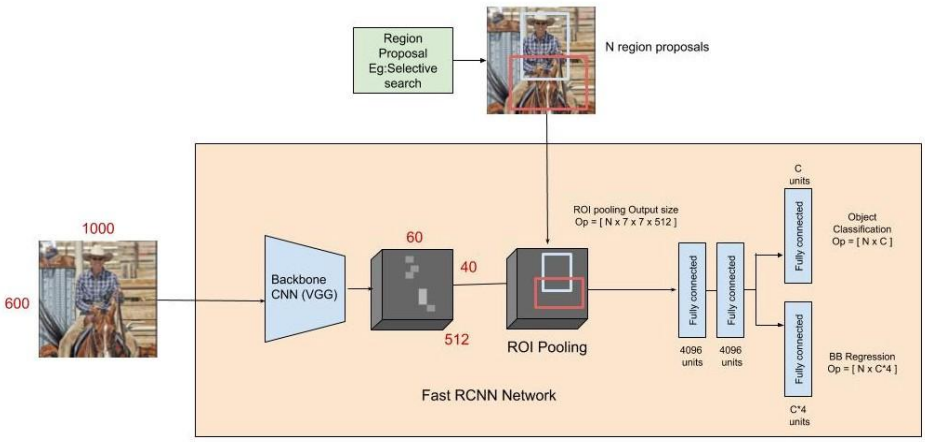

Figure 2: Fast R-CNN for Object Detection.

https://towardsdatascience.com/fast-r-cnn-for-object-detection-a-technical-summary-_a0ff94faa022

- **Faster R-CNN**:

Both R-CNN and Fast R-CNN uses selective searching to find out the region proposals. But selective search is a time-consuming and slow process that affects the performance of the network. Therefore, an object detection algorithm called as Faster R-CNN was proposed by Shaoqing Ren et al. to eliminate the selective search algorithm and lets the network learn the region proposals.In Faster R-CNN, the image is provided as an input to a convolutional network that provides the convolutional feature map. Instead of using selective search algorithm on the feature map which identifies the region proposals, it uses a separate network that predicts the region proposals. Then the reshaping of the predicted region proposals is done using RoI pooling layer which is further used to classify the image within the proposed region and predict the offset values for the bounding boxes. It is also applicable for real-time object detection.

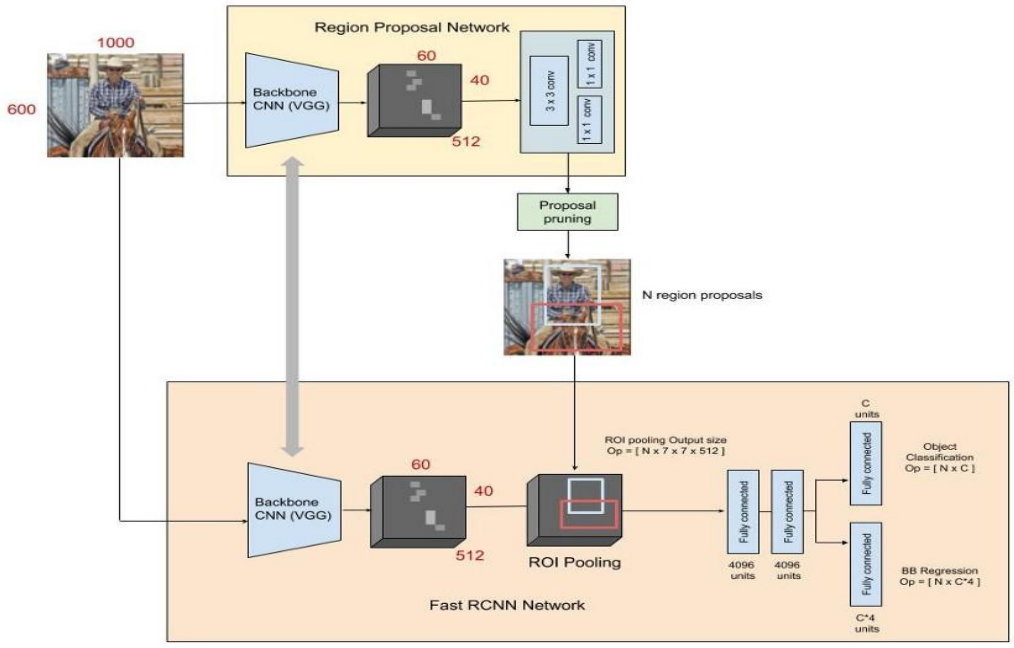

Figure 3: Faster R-CNN for Object Detection

https://appliedsingularity.com/2021/06/08/object-detection-part-5-faster-r-cnn/

- **YOLO (You Only Look Once)**:

YOLO is an object detection algorithm but it is very much different from the region-based algorithms which are R-CNN, Fast R-CNN and Faster R-CNN. In YOLO, a single convolutional network is used for predicting the bounding boxes and the class probabilities of these boxes. YOLO is faster when being compared to the other object detection algorithms. But it has a limitation that it struggles with small objects within the image and this is due to the spatial constraints of the algorithm.

5. RESULTS AND DISCUSSIONS

ALGORITHMS	FEATURES	LIMITATIONS
R-CNN	Composed of 2 steps: • First step is selective search for region identification • Extraction of CNN features from each region independently for classification	• Training is expensive and slow • The process involves 3 separate models without much shared computation • Cannot be implemented in real time because it takes around 47 seconds to run each test image

FAST R-CNN	Each image is passed one time to the CNN • Feature maps are used to detect objects • Uses a single R-CNN model • Much faster compared to R-CNN in both training and testing time	• Selective search is slow and as a result the computational time is high • This process is incredibly expensive as region proposals are generated separately employing different model
FASTER R-CNN	• It uses a unified model composed of RPN (region proposal network) and fast R-CNN with shared convolutional feature layers.	• Object proposals with RPN are time consuming • The performance of the current system gets affected by the performance of the previous system.
YOLO	• Process frames at the speed of 45 fps (larger network) to 150 fps (smaller network) which is better compared to real-time. • The network is ready to generalize the image better.	• It produces comparatively low recall and more localization error compared to Faster R-CNN. • It struggles to detect close objects as each grid can propose only 2 bounding boxes. • Small object detection is difficult.

6. CONCLUSION

We can easily detect and identify the various objects present in an image. Our visual system is much faster and accurate and we can also perform difficult tasks like identifying multiple objects and detect the problems and obstacles with little conscious thought. But, with today's growing technology and availability of large amount of data, faster GPUs and better algorithms, we can even train machines and computers to do the same with high accuracy.

In our paper, we have done a survey on the topic object detection using artificial intelligence. We have explained the terms like object detection, artificial intelligence and also the various algorithms which are used for object detection. Object detection provides a faster and accurate means to predict the location of an object in an image. It is useful in various applications such as video surveillance or image retrieval system. Some of the popular algorithms used to perform object detection includes Region-Based Convolutional neural Networks (R-CNN), Fast R-CNN, Faster R-CNN and YOLO. Object detection has grown rapidly in recent years and is one of the most fundamental and challenging problems in computer vision.

REFERENCES

1. https://www.geeksforgeeks.org/r-cnn-region-based-cnns/
2. https://blog.paperspace.com/faster-r-cnn-explained-object-detection/
3. https://www.geeksforgeeks.org/faster-r-cnn-ml/
4. https://www.section.io/engineering-education/introduction-to-yolo-algorithm-for-object-detection/
5. https://en.m.wikipedia.org/wiki/Artificial_intelligence
6. https://chethankumargn.medium.com/artificial-intelligence-definition-types-examples-technologies-962ea75c7b9b
7. https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence
8. https://www.fritz.ai/objectdetection/#:~:text=Object%20detection%20is%20a%20computer,all%20while%20accurately%20labeling%20them.
9. Joshi RC, Yadav S, Dutta MK, Travieso Gonzalez CM. Efficient Multi-object Detection and Smart Navigation Using Artificial Intelligence for Visually Impaired People. Entropy (Basel). 2020;22(9):941. Published 2020 Aug 27. doi: 10.3390/e22090941
10. Kumar, Ashish. (2020). Artificial Intelligence In Object Detection - Report.
11. Pandelea, Alexandrina-Elena & Budescu, Mihai & Covatariu, Gabriela. (2015). Image Processing Using Artificial Neural Networks. Bulletin of the Polytechnic Institute of Jassy, CONSTRUCTIONS. ARCHITECTURE Section. LXI(LXV). 9-21.
12. Zou, Z., Shi, Z., Guo, Y., & Ye, J. (2019). Object Detection in 20 Years: A Survey. *ArXiv, abs/1905.05055*.
13. T. N. Nguyen, B. -H. Liu, N. P. Nguyen and J. -T. Chou, "Cyber Security of Smart Grid: Attacks and Defenses," *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1-6, doi: 10.1109/ICC40277.2020.9148850.

14. Viola, P., Jones, M.J. Robust Real-Time Face Detection. *International Journal of Computer Vision* **57,** 137–154 (2004).
15. Redmon, Joseph & Divvala, Santosh & Girshick, Ross & Farhadi, Ali. (2016). You Only Look Once: Unified, Real-Time Object Detection. 779-788. 10.1109/CVPR.2016.91
16. Dhillon, Anamika & Verma, Gyanendra. (2019). Convolutional neural network: a review of models, methodologies and applications to object detection. Progress in Artificial Intelligence. 9. 10.1007/s13748-019-00203-0.
17. Sánchez Hernández, Sergio & Romero, H & Morales, A. (2020). A review: Comparison of performance metrics of pretrained models for object detection using the TensorFlow framework. IOP Conference Series: Materials Science and Engineering. 844. 012024. 10.1088/1757-899X/844/1/012024.
18. Chua, S.N. & Lim, Soh Fong & Lai, S. & Chang, T.. (2019). Development of a Child Detection System with Artificial Intelligence Using Object Detection Method. Journal of Electrical Engineering & Technology. 14. 10.1007/s42835-019-00255-1.
19. Yang, Ruixin & Yu, Yingyan. (2021). Artificial Convolutional Neural Network in Object Detection and Semantic Segmentation for Medical Imaging Analysis. Frontiers in Oncology. 11. 638182. 10.3389/fonc.2021.638182.
20. Qiu, Xin & Yuan, Chun. (2017). Improving Object Detection with Convolutional Neural Network via Iterative Mechanism. 141-150. 10.1007/978-3-319-70090-8_15.
21. Y. Bai, L. Zhang, T. Wang and X. Zhou, "A Skeleton Object Detection-Based Dynamic Gesture Recognition Method," *2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC)*, 2019, pp. 212-217, doi: 10.1109/ICNSC.2019.8743166.
22. Aditi, & Dureja, Aman. (2021). A Review: Image Classification and Object Detection with Deep Learning. 10.1007/978-981-33-4604-8_6.
23. Sultana, Farhana & Sufian, A. & Dutta, Paramartha. (2019). A Review of Object Detection Models based on Convolutional Neural Network.
24. Crawford, Eric & Pineau, Joelle. (2019). Spatially Invariant Unsupervised Object Detection with Convolutional Neural Networks. Proceedings of the AAAI Conference on Artificial Intelligence. 33. 3412-3420. 10.1609/aaai.v33i01.33013412.
25. https://www.ripublication.com/irph/ijisaspl2019/ijisav11n1spl_04.pdf
26. Hung, Kuo-Ching & Lin, Meng-Chun & Lin, Sheng-Fuu. (2022). A Novel Power-Saving Reversing Camera System with Artificial Intelligence Object Detection. Electronics. 11. 282. 10.3390/electronics11020282.
27. Li, Zhihui & Yao, Lina & Zhang, Xiaoqin & Wang, Xianzhi & Kanhere, Salil & Zhang, Huaxzhang. (2019). Zero-Shot Object Detection with Textual Descriptions.
28. A. Pandey, M. Puri and A. Varde, "Object Detection with Neural Models, Deep Learning and Common Sense to Aid Smart Mobility," *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*, 2018, pp. 859-863, doi: 10.1109/ICTAI.2018.00134.
29. J. Han, D. Zhang, G. Cheng, L. Guo and J. Ren, "Object Detection in Optical Remote Sensing Images Based on Weakly Supervised Learning and High-Level Feature Learning," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3325-3337, June 2015, doi: 10.1109/TGRS.2014.2374218.
30. Jeong, J., Lee, S., Kim, J., & Kwak, N. (2019). Consistency-based Semi-supervised Learning for Object detection. *NeurIPS*.
31. Y. Yang, G. Shu and M. Shah, "Semi-supervised Learning of Feature Hierarchies for Object Detection in a Video," *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1650-1657, doi: 10.1109/CVPR.2013.216.
32. Hussain, S.U. (2012). Machine Learning Methods for Visual Object Detection.
33. M. Ranzato, F. J. Huang, Y. Boureau and Y. LeCun, "Unsupervised Learning of Invariant Feature Hierarchies with Applications to Object Recognition," *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1-8, doi: 10.1109/CVPR.2007.383157.
34. J. Latif, C. Xiao, A. Imran and S. Tu, "Medical Imaging using Machine Learning and Deep Learning Algorithms: A Review," *2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, 2019, pp. 1-5, doi: 10.1109/ICOMET.2019.8673502
35. Pu, Shiliang & Zhao, Wei & Chen, Weijie & Yang, Shicai & Xie, Di & Pan, Yunhe. (2021). Unsupervised object detection with scene-adaptive concept learning基于场景自适应概念学习的无监督目标检测. Frontiers of Information Technology & Electronic Engineering. 22. 638-651. 10.1631/FITEE.2000567.
36. Liu, Li & Ouyang, Wanli & Wang, Xiaogang & Fieguth, Paul & Chen, Jie & Liu, Xinwang & Pietikäinen, Matti. (2018). Deep Learning for Generic Object Detection: A Survey.
37. Sohn, Kihyuk & Zhang, Zizhao & Li, Chun-Liang & Zhang, Han & Lee, Chen-Yu & Pfister, Tomas. (2020). A Simple Semi-Supervised Learning Framework for Object Detection.
38. Jeong, J., Lee, S., Kim, J., & Kwak, N. (2019). Consistency-based Semi-supervised Learning for Object detection. *NeurIPS*.

39. Han, J., Zhang, D., Cheng, G., Guo, L., & Ren, J. (2015). Object Detection in Optical Remote Sensing Images Based on Weakly Supervised Learning and High-Level Feature Learning. *IEEE Transactions on Geoscience and Remote Sensing*, *53*, 3325-3337.
40. Luo, Ao & Li, Xin & Yang, Fan & Jiao, Zhicheng & Cheng, Hong. (2020). Webly-Supervised Learning for Salient Object Detection. Pattern Recognition. 103. 107308. 10.1016/j.patcog.2020.107308.
41. Saad Ali and Mubarak Shah, "A supervised learning framework for generic object detection in images," *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, 2005, pp. 1347-1354 Vol. 2, doi: 10.1109/ICCV.2005.22.
42. Rhee, P.K., Erdenee, E., Kyun, S.D., Ahmed, M.U., & Jin, S. (2017). Active and semi-supervised learning for object detection with imperfect data. *Cognitive Systems Research*, *45*, 109-123.
43. Mitash, Chaitanya & Bekris, Kostas & Boularias, Abdeslam. (2017). A Self-supervised Learning System for Object Detection using Physics Simulation and Multi-view Pose Estimation.
44. Sharma, K.U., & Thakur, N.V. (2017). A review and an approach for object detection in images. *Int. J. Comput. Vis. Robotics*, *7*, 196-237.
45. Xie, E., Ding, J., Wang, W., Zhan, X., Xu, H., Li, Z., & Luo, P. (2021). DetCo: Unsupervised Contrastive Learning for Object Detection. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 8372-8381.
46. V. Nair and J. J. Clark, "An unsupervised, online learning framework for moving object detection," *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, 2004, pp. II-II, doi: 10.1109/CVPR.2004.1315181.
47. Dai, Zhigang & Cai, Bolun & Lin, Yugeng & Chen, Junying. (2020). UP-DETR: Unsupervised Pre-training for Object Detection with Transformers.
48. Arruda, V.F., Paixão, T.M., Berriel, R., Souza, A.F., Badue, C.S., Sebe, N., & Oliveira-Santos, T. (2019). Cross-Domain Car Detection Using Unsupervised Image-to-Image Translation: From Day to Night. *2019 International Joint Conference on Neural Networks (IJCNN)*, 1-8.
49. https://www.irjet.net/archives/V6/i5/IRJET-V6I51170.pdf
50. Szegedy, Christian & Toshev, Alexander & Erhan, Dumitru. (2013). Deep Neural Networks for Object Detection. 1-9
51. Galvez, R.L., Bandala, A.A., Dadios, E.P., Vicerra, R.R., & Maningo, J.M. (2018). Object Detection Using Convolutional Neural Networks. *TENCON 2018 - 2018 IEEE Region 10 Conference*, 2023-2027.
52. https://ijsret.com/wp-content/uploads/2019/03/IJSRET_V5_issue2_219.pdf
53. https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e