Journal of Information Systems Engineering and Management
2025, 10(30s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Human and Object Detection Deep Learning Model Using R-CNN

[1]Janarthanan Sekar, [2]Karthikeyan S, [3]Saravanakumar S
[1]*Department of CSE, Jain University (Deemed to be university),Bengaluru. Jana.mkce@gmail.com*
[2]*Department of CSE, Jain University (Deemed to be university,Bengaluru. link2karthikcse@gmail.com*
[3]*Department of CSE, Jain University (Deemed to be university),Bengaluru,sar112113118@gmail.com*

ARTICLE INFO

Received: 29 Dec 2024
Revised: 12 Feb 2025
Accepted: 27 Feb 2025

ABSTRACT

Human and object detection is deep learning model. Which identifies and detects human(people) and object from image. For implementing the human and object detection there many popular algorithms like YOLO (You Only Look Once), SSD (Single Shot Multi-Box Detector, CNN and R- CNN family. R-CNN family has R-CNN (Region Based Convolution Neural Network), FAST R-CNN, FASTER R- CNN. This time the R-CNN is very popular in market for more accuracy and efficient machine and deep learning object detection model. In this paper we have explain about algorithm and implementation of the human and object detection deep learning and machine learning model using Faster R-CNN and CV2. Faster R-CNN model take less time and create efficient model with more accuracy as comparison to Fast R-CNN and R-CNN [3]. For identifying object or human from image we have fast working approach that is Faster R-CNN with accuracy 97.85%, precision 89.80% and recall 91.30%.

Keywords: *FASTER R-CNN, Object-Human Detection, CNN*

I. INTRODUCTION

Technology, a word that is never going to get erased with the essence of time passing by. The reason that we will never forget technology to time is because it is the only thing that is making the life of every individual safe and relaxed. Advancements is technology is a constant change. We have evolved from Electronic Numerical Integrator and Computer (ENIAC), the very first computer to Windows, Mac and what not at present. This advancement of technology was necessary for finding solutions to many undefined problems. Technology advancement was the need of the hour because of which we are able to treat disease that were once incurable. Everything has its plus and minuses but the important part is it is used for the greater good. One such technology advancement is done in the surveillance department where police officials have installed cameras on red lights to check if someone if breaking the law and if they are then they capture the image and the image is sent to the department for forward procedures. Now due to human nature the department receives hundreds of photos in a day and it is practically impossible to attend to every picture to find out the credentials of the person who broke the law. Credentials related to finding the number plate getting a clear picture of it, getting a clear picture of the car and getting a clear picture of the person operating it. This is the case where computer vision field is used. For e.g., there is a pre-trained deep learning model called the "HaarCascade Classifier" which is basically used to detect the objects in an image. It will be able to distinguish between a car and a human in our case. Our project revolves around the image visualization techniques to figure out the different objects in an image and classify therm.

Human and object detection is a task of computer vision which identifies and locates the objects and human(people) within an image or video frame. Image classification has main role in human and object detection, which categorizes entire images into classes, object detection algorithms pinpoint the presence of distinct objects in an image and draw bounding boxes around them to precisely outline their locations. To analyze and forecast visual input, these algorithms frequently use machine learning and deep learning methods such as convolutional neural networks (CNNs). Bounding boxes, class labels, and confidence scores are predicted for each grid cell in an image or video frame by object identification algorithms. Bounding boxes are rectangles that tightly enclose objects, and confidence

scores reflect the algorithm's confidence in its predictions. This technology finds applications in various fields such as autonomous vehicles, surveillance systems, robotics, and augmented reality, facilitating tasks that require machines to perceive and interact with their visual environments accurately. In this tech world, the technology task efficiency is increasing day by day but convolution neural network algorithm takes more time to train a model in object detection. For improving the time complexity and efficiency of the train model in objection detection. R-CNN and Faster R-CNN take less time to train the model as compare to CNN. So Fast R-CNN and Faster R-CNN is better choice instead of normal CNN.

II. LITRATURE SURVEY

Zuopeng Justin Zhang, Ashwani Kumar and Hongbo Lyu

[1] had done research on "Object detection in real time based on improved single shot multi-box detector algorithm" they had worked to improve the single shot multi-box detector algorithm for increasing the classification accuracy of identifying object. That improvement had been done in multilayer convolution neural networks. There was different parameter to find accuracy in detecting the objects such as frames per second, loss function, mean per average precision and aspect ratio. Basically, they improved the SSD algorithm with high accuracy.

Gudipudi Tripura, Cherku Bhaskara Gupta, Govardhana Siri, Swarna Kuchibhotla Teja, Hrishitha Veginati, Srithar

S, [2] have done research on "Improved Human Detection and Classification using Supervised Machine Learning Algorithms" their work suggests an exploration of human detection using deep learning. They had performed YOLO, SSD, and Faster R-CNN algorithm to create deep learning models for human detection from image. They had also mentioned the advantage and disadvantage of these algorithm.

Jaspreet Singh, Shivam Tiwari and Gurpreet Singh [3] had done survey on "Real Time Object Detection using Neural Networks: A Comprehensive Survey" they had explored the deep learning concept for object detection using multiple algorithm R-CNN, FAST R-CNN and Faster R-CNN and their application. But they had more focused on Convolutional neural network for building object detection deep learning model. They had also explored the main component of CNN and deep learning model.

P S V S Sridhar, Mannem Ponika, Kavuri Verma and Kopalli Jahnavi [4] had done research on "Developing a YOLO based Object Detection Application using OpenCV" they had built an application which used YOLO algorithm and OpenCV for detecting object from image or video. They had faced main problem with CNN was that this consumed more time to trained the deep learning object detection model. They also mentioned the time of 47sec per picture and to analyze two thousand regions per picture it took more time. Basically, they had aim to create faster model than CNN model which can solve slow training speed of model.

Md. Ariful Islam, Rohani Amrin, Abdullah Al Zubaer, Md. Romzan Ali and Md. Mehedi Hasan Naim [5] had done research on "Object Detection from Image using Convolution Nural Network based on Deep Learning". They had given two explorer the model using convolution neural network algorithm for object detection for images into two parts. First part explained theoretical concept of the importance of convolutional neural network to improve deep learning and computer vision. Second part is experimental concept that is implementation of convolutional neural network easily for object detection.

These studies collectively contribute to the field of object detection using various algorithms, with a focus on improving accuracy, exploring different models, addressing training speed issues, and combining theoretical concepts with practical implementations. We have researched on CNN advance family that is R-CNN and explained how faster R-CNN model is better than other model for Human and Object Detection.

III. PROPOSED METHEDOLOGY

In this research the faster Regional Convolutional Network (FASTER R-CNN) algorithm is used for human and object detection in python environment with various packages and libraries. When it comes to projects related to machine learning and deep learning then what other than python programming language will fit best in the list of tools and technologies used for developing a project in the artificial intelligence domain. Python is a very strong language when it comes to do things related to machines because of its ease of learning and readability. The problems related to machines and artificial intelligence already seem impossible to solve and already having a problem statement so hard plus having a language that is complex to be used will be like taking all the time in the world to develop

something and with this speed the world where we stand now would not have been able to develop so much.

Detectron2 is an advanced open-source object detection framework created by Facebook AI Research (FAIR). Serving as a flexible and modular successor to the original Detectron, this framework is designed to support experimentation with various components of object detection models. A key feature is its seamless integration with the PyTorch deep learning framework, leveraging PyTorch's dynamic computation graph for enhanced flexibility. Detectron2 provides an extensive array of pre- trained models for diverse computer vision tasks, including instance segmentation, object detection, and keypoint detection, covering a wide range of applications. Noteworthy for supporting cutting-edge models like variations of Faster R-CNN and Mask R-CNN, Detectron2 allows users to customize and extend these models to meet specific requirements.

Python is full of rich libraries and frameworks that help us in data visualization from making graphs to calculating hard equations. Everything is available in the libraries and framework of python and the development is still going on. The different libraries that we have used in our code are NumPy, Matplotlib and OpenCV. Let's us see what these libraries are basically used for

A. Libraries

NumPy – It is a tool for engineers, researchers and data scientists that provides a solid foundation for various data related work. It stands for Numerical Python and it supports for large data related to multi-dimensional array and also matrices. There are many more benefits of using NumPy such as Element-Wise Operations, Array Slicing and Indexing.

Matplotlib – It is a very popular data visualization library in python. It is used in wide range of fields for creating quality, static, animated etc. visualizations. It is also used in many fields from data analysis to scientific research and much more. It is used for creating large variety of charts and plots. It is used in conjunction with other libraries such as NumPy for data manipulation.

OpenCV – CV2 is another frequent abbreviation for it. An open-source software library for machine learning and

computer vision is called OpenCV. The primary applications for this library are vision-related activities, such as those involving images and videos. It provides a wide range of tools and functions for working with images and videos. The functions that it provides are Image Processing, Video Analysis, Camera Calibration and many more functions. It is widely used in processes like facial recognition, surveillance etc.

PyTorch- The AI Research lab at Facebook (FAIR) created the open-source machine learning framework PyTorch. It offers a dynamic and adaptable framework for creating and refining deep learning models. What sets PyTorch apart is its dynamic computation graph, allowing developers to modify and iterate on models dynamically during runtime, making it particularly popular among researchers and practitioners for prototyping and experimentation. PyTorch supports both CPU and GPU acceleration, enabling efficient computation on a wide range of hardware. Its user-friendly interface and seamless integration with Python have contributed to its widespread adoption, making it one of the most popular choices for deep learning tasks. Because of its extensive ecosystem of tools and libraries, PyTorch is a great choice for a wide range of machine learning applications, including reinforcement learning, computer vision, and natural language processing.

TensorFlow- The Google Brain team has built an open- source machine learning framework called TensorFlow. It is extensively employed in several deep learning and machine learning applications. Researchers and developers may design and implement machine learning models more effectively with TensorFlow's extensive and adaptable ecosystem of tools, libraries, and community resources. One of its key features is its ability to create static or dynamic computation graphs, allowing for both high-level model construction and fine-grained control over network architecture. TensorFlow supports efficient numerical computations, making it suitable for training and deploying large-scale machine learning models. It has a versatile architecture that enables deployment on various platforms, including CPUs, GPUs, and specialized hardware accelerators like Google's Tensor Processing Units (TPUs). TensorFlow's popularity is attributed to its scalability, flexibility, and extensive support for deep learning techniques, making it a preferred choice for machine learning projects across academia and industry.

A group of object identification models known as the RCNN (Region-based Convolutional Neural Network) family

locate and identify items in a picture using a region- based method. These models have developed throughout time to address issues with accuracy and speed, and they are essential in computer vision jobs. The following are some important RCNN family members:

i. A convolutional neural network based on regions, or R-CNN:

• First presented in 2014 by Ross Girshick et al.

• The process is divided into two phases: object detection and region proposal generating.

• Selective Search is employed while proposing regions.

• CNN features are taken out of every suggested region and categorised using Support Vector Machines.

ii. Fast R-CNN: Developed by Ross Girshick in 2015, it addressed the speed problems with the original R-CNN.

• Uses a single common convolutional network to extract features, removing the requirement to independently compute CNN features for every area proposal.

iii. Faster R-CNN: Presented in 2015 by Shaoqing Ren and colleagues.

• Enhanced the neural network's region proposal procedure by including it.

• To facilitate the model's ability to propose areas directly from feature maps, the Region Proposal Network (RPN) is presented.

iv. Mask R-CNN: - Debuting in 2017, Kaiming He et al.

• Adds a second branch to Faster R-CNN to predict segmentation masks for every instance concurrently with bounding box recognition.

• Allows for instance segmentation, which makes pixel-by-pixel distinctions between object instances.

B. Algorithm used

Faster R-CNN is an advanced object detection algorithm designed for efficient and accurate detection of objects within images. Unlike its predecessors, Faster R-CNN integrates a Region Proposal Network (RPN) directly into the model, making the process more streamlined and faster.

Region Proposal Network (RPN): The key innovation of Faster R-CNN is the RPN, which generates potential object bounding box proposals. Instead of relying on external algorithms to propose regions, Faster R-CNN has an internal neural network, the RPN, which predicts potential object locations. These proposals are areas in the image where objects might be located.

Feature Extraction: Convolutional neural networks, or CNNs, are used by Faster R-CNN to extract features from the input picture. After processing the complete image, this CNN creates a feature map that encodes different aspects of the objects in the image.

Region of Interest (RoI) Pooling: Once the RPN generates proposals, RoI pooling is applied. RoI pooling allows these variable-sized regions to be transformed into a fixed-sized feature map. This step ensures that the subsequent layers of the network can process regions of interest in a consistent manner, regardless of their original sizes.

Object Classification and Bounding Box Refinement: The network's branches get the fixed-sized ROI characteristics separately. Determining what sort of item (if any) is present in the ROI is done by one branch's object categorization process. In order to improve the bounding box's accuracy of fit and fit the object more precisely, the other branch carries out bounding box regression.

Loss Function and Training: A multi-task loss function is used in the training of a faster R-CNN. This function combines classification loss (to classify objects) and regression loss (to refine the bounding boxes). The network learns to minimize this combined loss during the training process, which refines the RPN and the object detection parts simultaneously.

Inference: During the inference phase, the trained Faster R- CNN model can detect objects in new images. The RPN proposes regions, which are then refined and classified to obtain the final object detection results.

Faster R-CNN's integration of the RPN within the network architecture has significantly sped up the object detection process. By eliminating the need for external algorithms to propose regions, Faster R-CNN achieves impressive accuracy and efficiency, making it a cornerstone in the field of object detection.

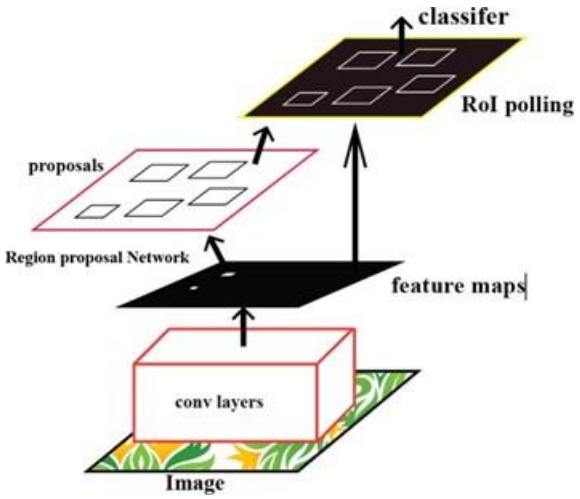

Fig. 1. FASTER R-CNN Model Digram

Here is how the Faster R-CNN works-

- Region suggestions are generated by the RPN.
- A fixed-length feature vector is retrieved from each area utilising the ROI Pooling layer for each region proposed in the picture.
- Next, a Fast R-CNN classification is performed on the retrieved feature vectors.
- Along with their bounding boxes, the class scores of the objects that were discovered are returned.

IV. RESULT AND ANALYSIS

For object detection Region Convolutional Neural Network family is most used algorithm for getting efficient model with high accuracy. There are three member of this family you know very well. Region Convolution Neural Network, FAST R-CNN and FASTER R-CNN.

Problem pertaining to R-CNN

- Every picture must categorise 200 region suggestions. For this reason, training the network takes a long period.
- The GPU takes 49 seconds to identify the items in a picture.
- A large amount of disk space is also needed in order to store the feature map for the region proposal.

Problem pertaining to R-CNN

- An approach for selective search region proposal generation takes

Faster R-CNN model has 97.85% Accuracy mentioned in Fig no 2

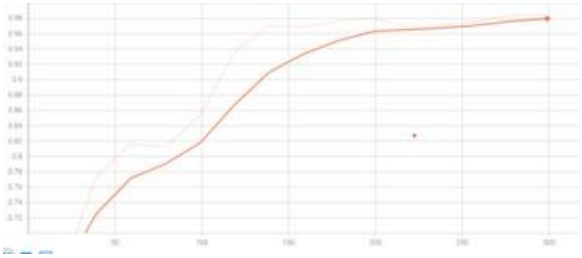

Fig. 2. Faster R-CNN accuracy graph Mask R-CNN model gave has 97.28% Accuracy mentioned in Fig

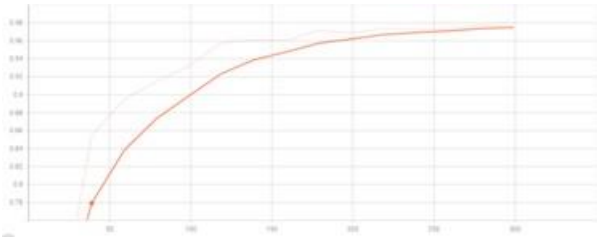

Fig. 3. Mask R-CNN accuracy graph Faster R-CNN Precision and Recall

- Precision- 89.80%
- Recall- 91.30%

Mask R-CNN Precision and Recall

- Precision-95.80%
- Recall-97.70%

There are outputs of Faster R-CNN with Mask R-CNN

Fig. 4. People and object detected by model with Image segmentation

Fig. 5. People detected by model with region

Fig. 6. People and object detected by model with region and image segmentation

Fig. 7. People detected by model with region

Fig. 8. People and object detected by model with Image segmantation

Fig. 9. People and object detected by model with region and image segmentation

Fig. 10. People and object detected by model with Image segmantation

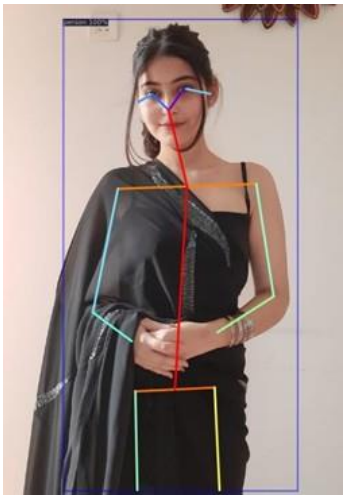

Fig. 11. People detected by model with region

V. CONCLUSION AND FUTURE

Human-Object detection is working on real life problem which detects people and object from Image using Faster r-cnn with best accuracy and taking less time to identify region of human and object from image for detection. Currently Faster r-cnn is faster than other algorithm available in market. It is used in education sector for education and research, surveillance domain like for detecting all activity for specific region, security domain like face

recognition system, industry sector for detecting quality of product or object and traffic control room etc. In this digital world there are many places where Human- Object detection is used. AI robots also use the human- object detection to identify the object from their vision. In the domain there are many other researchers working for make faster and more efficient algorithm than r-cnn family. Which can detect the object in mini second from video and Image.

ACKNOWLEDGMENT

Without the invaluable advice and help of our guide, Dr-

S. Janarthanan, we would not have been able to finish this paper. we learned a lot about working as a team, leadership qualities and some important technical aspects of our project. We would like to thank our guide too for their suggestions.

REFERENCES

[1] Kumar, A., Zhang, Z.J. & Lyu, H. Object detection in real time based on improved single shot multi-box detector algorithm. J Wireless Com Network 2020, 204 (2020). https://doi.org/10.1186/s13638-020- 01826-x

[2] C. B. Gupta, G. Tripura, G. S. Teja, H. Veginati, S. S and S. Kuchibhotla, "Improved Human Detection and Classification using Supervised Machine Learning Algorithms," 2023 8th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2023, pp. 1045-1050, doi: 10.1109/ICCES57224.2023.10192660.

[3] G. Singh, S. Tiwari and J. Singh, "Real Time Object Detection using Neural Networks: A Comprehensive Survey," 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), Coimbatore, India, 2023, pp. 1281-1286, doi: 10.1109/ICAIS56108.2023.10073826.

[4] M. Ponika, K. Jahnavi, P. S. V. S. Sridhar and K. Veena, "Developing a YOLO based Object Detection Application using OpenCV," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2023, pp. 662-668, doi: 10.1109/ICCMC56507.2023.10084075.

[5] Md. Mehedi Hasan Naim, Rohani Amrin, Md. Romzan Ali, Abdullah Al Zubaer and Md. Ariful Islam, “Object Detection from Image using Convolution Nural Network based on Deep Learning” International Journal of Engineering Research & Technology (IJERT) http://www.ijert.org ISSN: 2278-0181 IJERTV9IS090014(This work is licensed under a Creative Commons Attribution 4.0 International License.)Published by :www.ijert.org Vol. 9 Issue 09, September- 2020

[6] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In CVPR, 2014.

[7] S. Ren, K. He, R. Girshick, X. Zhang, and J. Sun. Object detection networks on convolutional feature maps. arXiv:1504.06066, 2015.

[8] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional neural networks. In ECCV, 2014.

[9] N. Chavali, H. Agrawal, A. Mahendru, and D. Batra. Object-Proposal Evaluation Protocol is ’Gameable’.arXiv: 1505.05836, 2015.

[10] J. Dai, K. He, and J. Sun. Convolutional feature masking for joint object and stuff segmentation. In CVPR, 2015.

[11] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov. Scalable object detection using deep neural networks. In CVPR, 2014.