Object Detection For Autonomous Vehicles

Abhinav Bharali-191IT201
Information Technology
National Institute of Technology Karnataka
Surathkal, India 575025
abhinavbharali.191it201@nitk.edu.in

Pratham Nayak-191IT241
Information Technology
National Institute of Technology Karnataka
Surathkal, India 575025
pratham.191it241@nitk.edu.in

Alimurtaza Mustafa Merchant-191IT205
Information Technology
National Institute of Technology Karnataka
Surathkal, India 575025
alimurtaza.191it205@nitk.edu.in

Naveen Shenoy-191IT134
Information Technology
National Institute of Technology Karnataka
Surathkal, India 575025
naveenshenoy.191it134@nitk.edu.in

***Abstract*—Vehicle detection plays an essential role in many applications, such as highway traffic surveillance, control, accident prevention, etc. Vehicle detection processes on the road are used for vehicle tracking, measuring the vehicle's average speed, traffic analysis, vehicle categorizing, and implementing different environmental changes. Autonomous vehicles will enable safer and accurate transportation, avoiding deaths and accidents all over the world. In this review, we present a method for detecting various objects during autonomous driving, such as pedestrians, cars, trucks, traffic lights, etc. To implement this task of object detection in a video, we will use Convolutional Neural Networks(CNNs). The object detection system's job is to detect and classify everyday objects of the real world present either in a digital image or video (real-time or pre-recorded). The object can belong to any class, such as pedestrians, cars, trucks, traffic lights, etc.**

***Index Terms*—Autonomous vehicles, Object detection, Convolution Neural Networks**

I. Introduction

In recent years, autonomous vehicles have gained popularity in research for both industry and academia. A vehicle must navigate through the environment in which it is driving to be genuinely autonomous. It must identify and keep track of objects (moving and stationary) in its surroundings. This information about its surroundings can be extracted using cameras and inertial sensors. It is a fundamental problem because millions of people die in road accidents every year. This project aims to provide autonomous vehicles with all the object's information on their radar, reducing mishaps on the road and allowing for safe and convenient travel to the passengers.

With advancements in deep learning and incremental improvements in computing power, object detection using images outperforms other methods to detect and classify objects. Deep Learning brings greater accuracy in various tasks, namely object detection and image classification when compared to Computer Vision. We can take advantage of the vast amount of video data available today to train the neural networks. DL outweighs the benefits of computer vision, which requires lots of expert analysis and fine-tuning since it is programmed, unlike neural networks, which can be trained accordingly to fulfill the requirement of specific models. DL also provides better performance because CNN models can be re-trained using a custom dataset for any use case, contrary to CV algorithms, which tend to be more domain-specific. The thesis aims to develop an algorithm to detect, classify, and track objects commonly visible on the road. The deep learning model outperforms the conventional machine learning algorithms, provided that it has sufficient computational power and a large amount of data to train on.

II. Literature Survey

The project uses the paper "Rich feature hierarchies for accurate object detection and semantic segmentation" as the base for implementing the idea of detecting objects for autonomous vehicles.

The model described in this paper consists of three modules.
1. Category-independent region proposals are generated in the first module, which defines the set of candidate detections available to the detector.
2. The second module extracts a fixed-length feature vector from each region using an extensive convolutional neural network.
3. The third module is a group of class-specific linear SVMs used to present each module's design decisions, describe their test-time usage, detail how their parameters are learned, and show detection results on PASCAL VOC 2010-12 and ILSVRC2013.

TABLE I: Summary of Literature Survey

Paper Name	Authors	Methdology	Merits	Limitations
Rich feature hierarchies for accurate object detection and semantic segmentation	Ross Girshick, Jeff Donahue,Trevor Darrell, Jitendra Malik.	High-capacity CNNs are applied to bottom-up region proposals to localize and segment objects. An excellent performance boost is obtained by implementing supervised pre-training and domain-specific fine-tuning	A simple, scalable, and straight forward object detection algorithm. Works well even when data is scarce	This algorithm is slow as compared to later computer vision algorithms like YOLO, etc
MobileNets: Efficient Convolutional Neural Networks for mobile vision applications	Andrew G. Howard, Menglong Zhu,Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam	Light-weight deep neural networks are built using depthwise separable convolutions.	Small, low-latency, and low power models.	Fewer parameters may lead to slower performance than other models like ResNet-50, which has better accuracy than MobileNet
MobileNetV2: Inverted Residuals and Linear Bottlenecks	Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, Liang-Chieh Chen	Splits convolution operator into depthwise convolution(light-weight filtering) and pointwise convolution (builds new features through a linear combination of input channels)	Memory-efficient model Simple network architecture	Fewer parameters may lead to slower performance than other models like ResNet-50, which has better accuracy than MobileNet
object Detection with Deep Learning: A Review	Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, and Xindong Wu	To provide a review of various deep learning -based object detection frameworks.	Brings greater accuracy in various tasks such as image classification, object detection when compared to Computer Vision.	DL applications require huge manually annotated data sets that are hard to obtain. Annotation is time-consuming, expensive, and often ambiguous.

III. PROBLEM STATEMENT

Object detection for autonomous vehicles using deep learning models like R-CNNs and Mobile Netv2 Architecture.

A. OBJECTIVES

- To be able to identify the presence of obstacles for autonomous vehicles.
- Be able to distinguish the obstacles into two categories:
  - Automobiles/Vehicles
  - Pedestrian/Non-Automobiles
- Be able to locate the position of these obstacles accurately.
- To understand and efficiently fulfill the above objectives by implementing R-CNN and MobileNetv2.

IV. METHODOLOGY

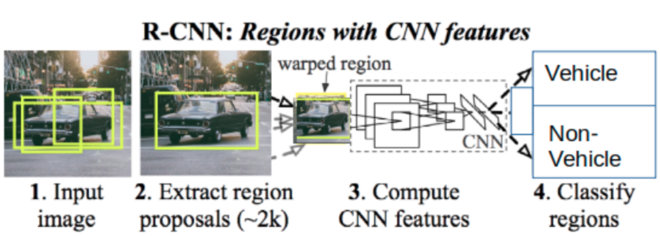

Fig. 1: R-CNN

A. *Extracting regions proposals (2000) using selective search*

First, around 2000 region proposals were extracted from the images. This region proposal algorithm was preferred over the traditional sliding window because of its computational efficiency.
In this method, a superpixel clustering algorithm is applied to the input image to over-segment the image. The segments are then merged based on similarities in color, texture, size, shape, and a meta-similarity that linearly combines all the similarities mentioned above. This is a greedy algorithm that helps us combine similar regions into larger ones. This allows us to get proposals in the image where an object could be potentially present and generate final candidate region proposals.

B. *Using MobileNetV2 as the base model for extracting features from the input image*

MobileNetV2, with weights trained on the ImageNet dataset, was used as our base model. There are two types of blocks in MobileNetV2. The first one is a residual block with a stride of 1, and the second one is a block with a stride of 2 for downsizing.
Layers in each type of block are:-

- The first layer is a 1×1 convolution layer with ReLU6.
- Depthwise Convolution layer
- 1x1 convolution layer without any non-linearity

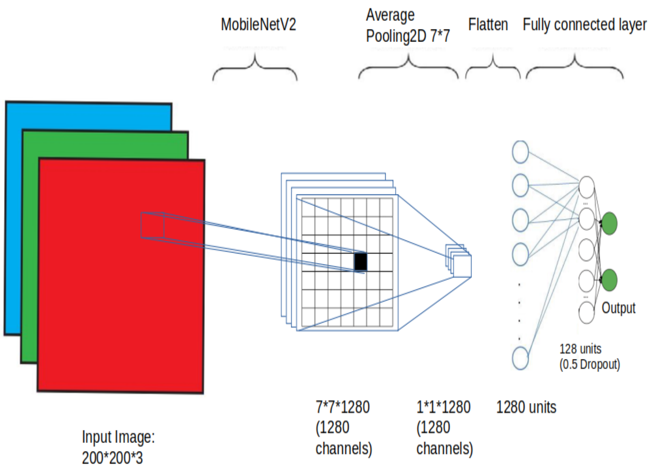

Fig. 2: Model Architecture

C. *Using neural networks at the top layer to classify the ROI using the features extracted by the base model.*

First 2D average pooling was performed on the features extracted by the MobileNetV2 model. The output was then flattened and fully connected to a dense layer with 128 neurons having ReLU activation, followed by a fully connected layer of 2 neurons having softmax activation, which act as the classifier. A dropout of 0.5 probability was used in the fully connected layer for regularisation.

D. *Removing Redundant proposals using NMS and IoU*

- **Intersection over Union (IoU) :**
  It is an evaluation metric that is used to measure an object detector's accuracy on a particular dataset. IoU is used to measure the performance of HOG + Linear SVM object detectors and Convolutional Neural Network detectors (R-CNN, Faster R-CNN, YOLO, etc.).

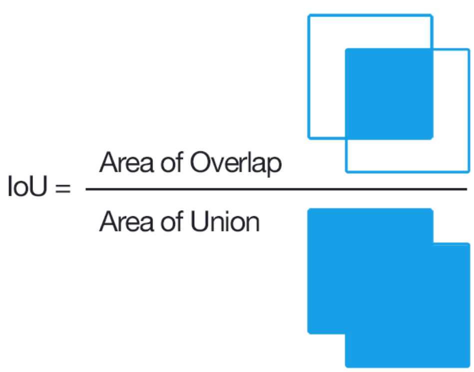

Fig. 3: IoU

But the actual algorithm used to generate the predictions does not matter. IoU can be used to evaluate any algorithm that provides bounding boxes as its prediction. To apply IoU to evaluate an object detector, we need-

  – The ground-truth bounding boxes (i.e., the hand-labeled).
  – The predicted bounding boxes from our model.

It is basically the ratio of the area of overlap to the area of union of 2 bounding boxes. Higher IoU indicates greater similarity between the 2 bounding boxes used for calculation.

- **Non-max Suppression (NMS) :**

It is a technique used for removing redundant proposals based on the IoU of the predicted bounding boxes.

The following steps are followed to perform NMS on a list of bounding boxes

1.First, all the bounding boxes are sorted in decreasing order of their confidence scores.
2.The proposal with the current highest confidence score is selected and compared with all other proposals.
3.The proposals having having an IoU greater than a specific threshold when evaluated with the selected box, are removed from consideration.
4.Step 2 and 3 is repeated for all the proposals in decreasing order of their confidence scores until the list becomes empty.

Fig 4 and 5 show the working NMS.

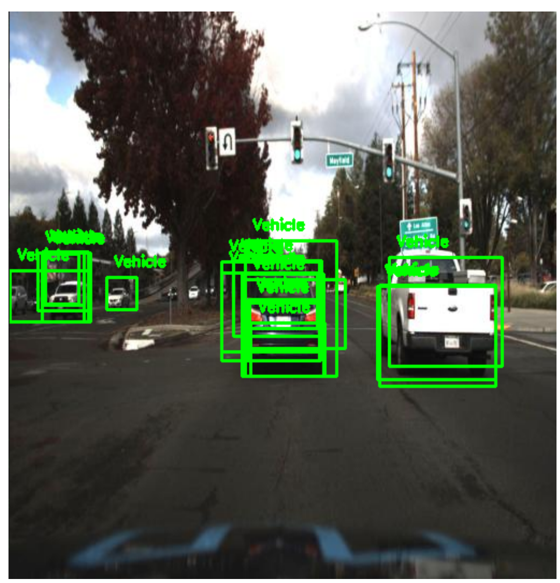

Fig. 4: Before NMS

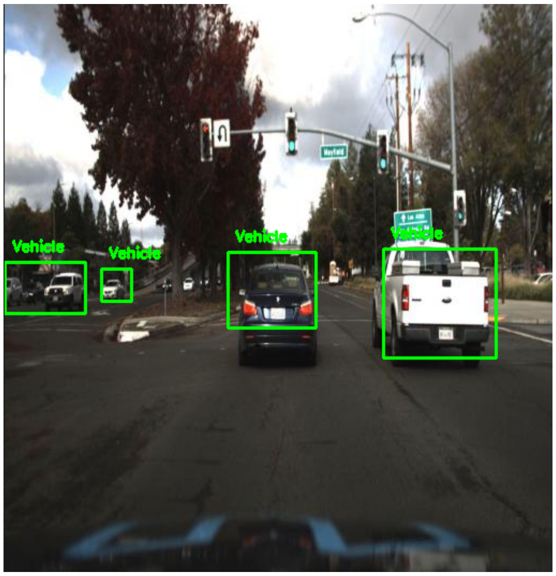

Fig. 5: After NMS

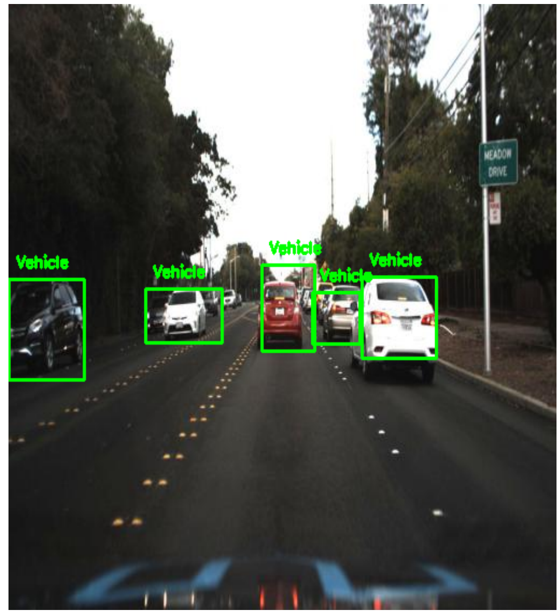

Fig. 6: Result

V. RESULTS AND ANALYSIS

Although R-CNN serves as a simple, scalable and straightforward algorithm, it is a bit slower compared to some other state of the art computer vision algorithms like YOLO and SSD, however its ability to give accurate results on scarce data overshadows its speed. However, the computational speed can be increased using Fast RCNN or Faster-RCNN.

The CNNs usually used for object detection InceptionNet, ResNet and MobileNet. InceptionNet focuses on computational cost. On the other hand, ResNet focuses on computational accuracy. Usually, in computer vision, deeper neural networks are built to improve accuracy. However, deeper networks come with the tradeoff of size and speed.

In real applications such as an autonomous vehicle, the object detection task must be done on the computationally limited platform. Hence we have used MobileNet which is specifically developed to solve this problem. This is because it is a small, low latency model which can achieve accuracy compared to ResNet with minimal computational complexity and cost compared to InceptionNet.

For the classification part of the detection algorithm performance of the model was quite impressive.In the classification part,the model accuracy was 98% with low loss. The model was trained on 12000 images to get the best results. For the object detection part, for analysing the model, we mainly used three categories: true positives,false positives and false negatives. The testing was done on 100 images to get a proper estimation of the model.

Performance of the model

- We found that the percentage of true positives was an impressive 86%.
- As for the percentages of false positive ,it was at 14% which is quite good given that our model is not much computationally expensive.
- The false negatives for the test data-set was found to be 32% mainly because of the presence of minimally visible vehicles marked in the dataset.

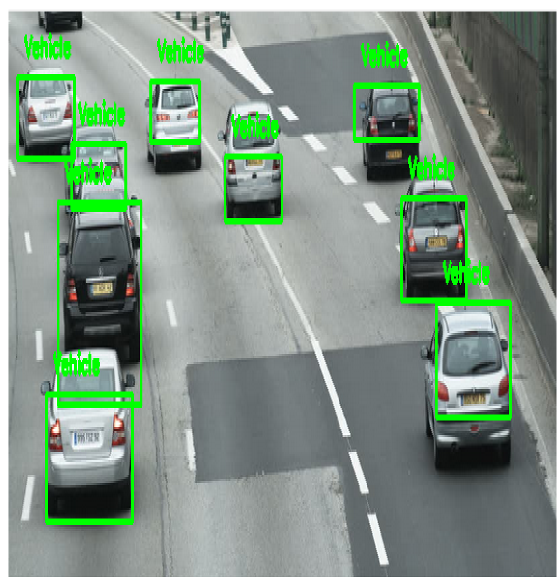

Fig. 7: Result

Fig. 8: Result

VI. Conclusion

The project detects vehicles on the roads to assist in autonomous driving. This project has helped us understand the fundamental concepts of image processing, convolution neural networks(CNNs), transfer learning and other algorithms needed to develop ML systems that can automate our tasks.

Improvements:

1) Detection of pedestrians and traffic signals to make autonomous driving safer

2) Using the detected objects to make intelligent autonomous driving decisions.

Individual Contribution

All the team members contributed equally in the learning phase and report compilation, however the various implementational sub tasks were divided amongst the team members based on their areas of interest as follows

Pratham Nayak

- Design of Project mode
- Implementation of R-CNN and MobileNetV2 base model on extracted region proposals
- Testing overall performance of the model and madeim-provements
- Project Report

Alimurtaza Merchant

- Implementation of R-CNN and MobileNetV2 base model on extracted region proposals
- Dataset Collection
- Project Report

Abhinav Bharali

- Using neural networks at the top layer to classify the ROI using the features extracted by the base model
- Dataset Collection
- Project Report

Naveen Shenoy

- Removing redundant proposals using NMS and IOU
- Analysis of results and efficient working of the model
- Project Report

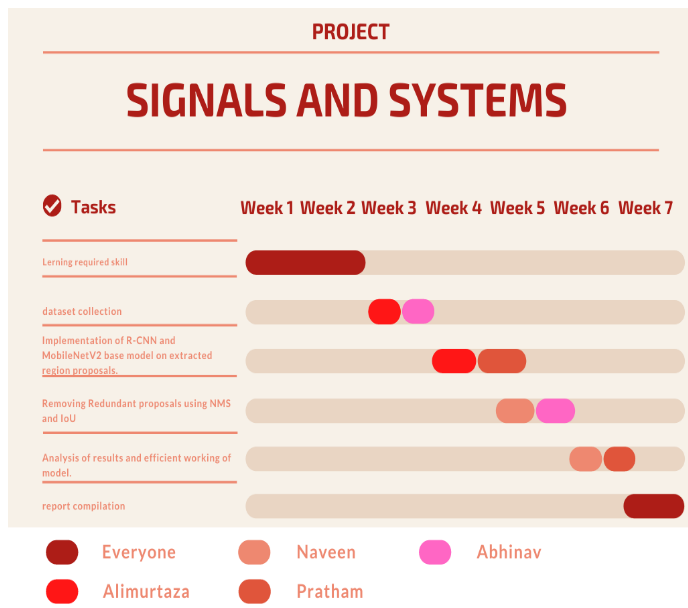

Fig. 9: Gantt Chart

References

[1] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, June 2014.

[2] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam." MobileNets: Efficient Convolutional Neural Networks for mobile vision applications" arXiv preprint arXiv:1704.04861, 2017.

[3] Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, and Xindong Wu." Object Detection with Deep Learning: A Review" . arXiv preprint arXiv:1807.05511

[4] Mark Sandler, Andrew Howard, Menglong Zhu, Liang-Chieh Chen, Andrey Zimoghinov." MobileNetV2: Inverted Residuals and Linear Bottlenecks". Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.

[5] OpenCV, 2015. Open Source Computer Vision Library,

[6] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Rafal Jozefowicz, Yangqing Jia, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Mike Schuster, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[7] M. V. Rajasekhar and A. K. Jaswal, ”Autonomous vehicles: The future of automobiles,” 2015 IEEE International Transportation Electrification Conference (ITEC), Chennai, 2015, pp. 1-6, doi: 10.1109/ITEC-India.2015.7386874.

[8] Bimbraw, Keshav. (2015). Autonomous Cars: Past, Present and Future - A Review of the Developments in the Last Century, the Present Scenario and the Expected Future of Autonomous Vehicle Technology. ICINCO 2015 - 12th International Conference on Informatics in Control, Automation and Robotics, Proceedings. 1. 191-198. 10.5220/0005540501910198