Enhanced Object Detection for Driving Assistance

Nagesh V[1], Manaswini Pisipati[2], Deepthi S[2], Kavya Tummala[2], Venkata Akhil[2]

[1]Associate Professor, Sreyas institute of engineering and Technology, JNTUH, India.

[2]Student, Sreyas institute of engineering and Technology, JNTUH, India.

Corresponding Author: nageshmathew@gmail.com

Abstract: As a significant technology of intelligent transportation systems, the intelligent vehicle is the carrier of comprehensive integration of many technologies. Although vision-based autonomous driving has shown excellent prospects, there is still a problem of how to analyze the complicated traffic situation by the collected data. In this study, a vision-based system was developed to detect and identify various objects and predict the intention of pedestrians in the traffic scene. The main contributions of this research are an optimized model was presented to detect 10 kinds of objects based on the structure of Faster RCNN a fine-tuned Part Affinity Fields approach was proposed to estimate the pose of pedestrians; Explainable Artificial Intelligence (XAI) technology is added to explain and assist the estimation results in the risk assessment phase; an elaborate self-driving dataset that includes several different subsets for each corresponding task was introduced; and an end-to-end system containing multiple models with high accuracy was developed. Experimental results proved that the total parameters of optimized Faster RCNN reduced by 74%, which satisfies the real-time capability. In addition, the detection precision of the optimized Faster RCNN achieved an improvement of 2.6% compared to the state-of-the-art.

Key Words: —*Object Detection, Driving Assistance, Intelligent Vehicle.*

I. Introduction

Rapid urbanization has highlighted a series of problems, especially in the aspect of transportation, which severely limits travel and has certain security risks. Even though some progress has been made in the existing object detection technologies in self-driving, there still exist potential risk factors of collision as motor cars are surrounded by many objects in daily life, including some uncontrollable moving objects (pedestrians and vehicles) and static objects (traffic lights and signs). Therefore, it is necessary to promptly detect various static objects and accurately estimate the intention of moving objects.

In the object detection tasks, the main deep learning methods are divided as one-stage detection algorithms and two-stages detection algorithms. YOLO and SSD are one-stage detection methods that directly convert the detection problem to a unified regression problem.

Manuscript revised July 06, 2022; accepted July 07, 2022. Date of publication July 09, 2022.

This paper available online at www.ijprse.com

ISSN (Online): 2582-7898; SJIF: 5.59

Due to the characteristics of the structure, the one-stage methods are faster than two-stage methods. Faster R-CNN is a typical two-stage network that generates a series of candidate bounding boxes and then classifies each object by using the Convolutional Neural Network (CNN). From the aspects of detection and localization precision, the two-stage methods perform better than most of the one-stage methods. In this study, the proposed model with multiple tasks is based on the one-stage methods to reduce the time used for the object detection phase.

1.1 Motivation

Our object detection system, called Faster R-CNN, is composed of two modules. The first module is a deep fully convolutional network that proposes regions, and the second module is the Fast R-CNN detector that uses the proposed regions. The entire system is a single, unified network for object detection using the recently popular terminology of neural networks with 'attention' mechanisms; the RPN module tells the Fast R-CNN module where to look.

1.2 Objective

Deep convolution of CNN features is applied at smaller feature output scales, which is further fused with features at larger feature output scales, to provide richer context for object detection at individual feature output scales. Such enhancement can effectively address the large object scale variation challenge.

1.3 Scope

In this paper, we show that an algorithmic change computing proposal with a deep convolutional neural network leads to an elegant and effective solution where proposal computation is nearly cost-free given the detection network's computation

1.4 Outline

In this project, we proposed an effective incremental training method based on learning automata for deep neural networks. The main thought is to train a deep model with dynamic connections which can be either ``activated'' or ``deactivated'' on different datasets of the incremental training stages. Our proposed method can relieve the destruction of old features while learning new features for the newly added training samples, which can lead to better training performance on the incremeIn this project, we proposed an effective incremental training method based on learning automata for deep neural networks. The main thought is to train a deep model with dynamic connections which can be either ``activated'' or ``deactivated'' on different datasets of the incremental training stages. Our proposed method can relieve the destruction of old features while learning new features for the newly added training samples, which can lead to better training performance on the incremental learning stage.

II. Literature Survey

2.1 Object Detection

Recently, multi-objects detection has been a prevailing topic that is attracting a lot of researchers in the field of autonomous driving. A one-stage method You Only Look Once (YOLO) was first presented to address object detection as a regression problem. As one of the state-of-the-art works, YOLO can achieve robust and fast performance in object detection, but the spatial constraint of the model limits the predicted amount of objects [1]. Another one-stage method is named the single shot Multibox detection (SSD) [20]. For a 300 × 300 input size, the SSD model can achieve at 59 FPS, and 74.3% mean Average Precision (mAP) on the PASCAL VOC dataset, which is greatly superior to the real-time YOLO [1]. Besides, a unified network was performed for object detection. Experiments show the processing speed of the method is slower than the YOLO [1], but it achieved better performance in mAP due to the improved gripping process [2]. Compared with most of the one-stage methods, the two-stage methods can obtain more accurate detection results, but the detection speed is slower. For example, Faster R-CNN is a kind of two-stage method, which optimizes the overall accuracy by introducing a region proposal network [3]. In previous work, the performance of the autonomous driving dataset BDD100K has not been reported. Thus, the latest method Faster RCNN algorithm [18] is superior to the other state-of-art detectors and was optimized and tested on the BDD100K dataset in this study.

2.2 Intention Recognition

Since deep learning has significantly enhanced object detection performance, some extensions have been proposed to estimate the postures of pedestrians and vehicles. Based on the appearances of pedestrians, a CNN model was proposed to classify the pedestrians' head pose and body orientation, and the method is available for still images and image sequences [6]. In another work, a neural network using appearance features was provided to predict the location and key points in pose estimation [28]. In contrast to CNN-appearance-based methods, the dynamical model with Gaussian processes was presented to predict paths and poses of pedestrians by analyzing fitted skeletons [7]. The proposed skeleton-based intention recognition was compared with the appearance-based model to evaluate the effectiveness, and the results proved that the former achieved better performance [5]. However, classification accuracy on the skeleton features obtained 88% by using Random Forest algorithm, which is not satisfactory in the self-driving system

2.3 Risk Assessment

For risk assessment, the recognition of traffic lights and the moving trend for vehicles are important factors to avoid traffic accidents. A recurrent network with effective performance was used to predict the intention of drivers at different types of intersections. In another work, an end-to-end method combining CNN and Long Short-Term Memory (LSTM) was applied to recognize the direction of vehicles based on vehicle tail lights. As the key signal on the road, the detected traffic lights were recognized by a CNN model after filtering the traffic light candidates by the importance map in a real time system. A multi-task learning approach combining object detection and distance estimation was presented to probe the characteristics of dangerous objects with different distances, which achieved a better performance of 2.27% than the SSD method on the KITTI dataset. Even though there are some existing methods for safe driving, the method that can jointly process object detection, intention recognition, and risk assessment is not considered in the previous work.

There are two main categories considered in self-driving, including still objects and dynamic objects. Here, a vision-based model with multiple tasks was proposed to detect various objects and assess the posture of pedestrians and vehicles (dynamic objects) in this study. Besides, the traffic lights (still objects) are recognized to indicate whether the automatic driving system should continue to drive.

III. Existing System

Despite the fast growth of CNN in object detection over datasets with a large number of object classes, real time visual object detection in driving environments is still very challenging. It is observed that the object detection performance of the popular CNN detectors is not very good over the benchmark datasets.

- In the existing multi-scale CNN models, feature maps from feature output scales are processed separately to predict the existence of objects at fixed scales.

In most existing CNN detectors, non-maximal suppression (NMS) method is used for suppression of overlapping object proposals. With such a process

- there is very little chance for proper detection of occluded objects. But in driving environments occluded objects are normal and are potential driving hazards.
- In the existing CNN detectors, default anchor boxes with certain sizes are used to generate object proposals. In the driving environment the interested objects have strong features in shape, for example, the width of a car should not exceed lane width.

IV. Proposed System

In this paper, we show that an algorithmic change computing proposal with a deep convolutional neural network leads to an elegant and effective solution where proposal computation is nearly cost-free given the detection network's computation. To this end, we introduce novel Region Proposal Networks (RPNs) that share convolutional layers with state-of-the-art object detection networks. By sharing convolutions at test-time, the marginal cost for computing proposals is small (e.g., 10ms per image). Our observation is that the convolutional feature maps used by region-based detectors, like Fast R-CNN, can also be used for generating region proposals. On top of these convolutional features, we construct an RPN by adding a few additional convolutional layers that simultaneously regress region bounds and objectness scores at each location on a regular grid. The RPN is thus a kind of fully convolutional network (FCN) and can be trained end-to-end specifically for the task for generating detection proposals.

V. Implementation

5.1 Module description

NumPy: NumPy is a module for Python. NumPy enriches the programming language. Python with powerful data structures, implementing multidimensional arrays and matrices.

CV2(OpenCV): OpenCV is an open-source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products.

Matplotlib: images, sorting contours, detecting edges, and much easier with OpenCV.

ArgParse: The argparse module makes it easy to write user-friendly command-line interfaces.

OS: OS module provides allows you to interface with the underlying operating system that Python is running on – be that Windows, Mac or Linux.

5.2 Implementation stages:

- Object Detection
- Background Subtraction
- Object tracking

5.3 Functional Requirements

a. Software Requirements

Programming Language: python

Operating system : windows 7/10

UML Design : Staruml

Tools : PIP

IDE : pycharm, jupyter

b. Hardware Requirements

Processor : intel core i3 and above

Speed : 1.1 GHZ

RAM : 4GB andHigher

Hard Disk : 500 GB(Minimum)

5.4 System Architecture

The relationship between various components would be displayed using a system architecture diagram. They are typically developed for systems that contain both hardware and software, and these are represented in the diagram to show how they interact. It can, however, also be developed for web applications.

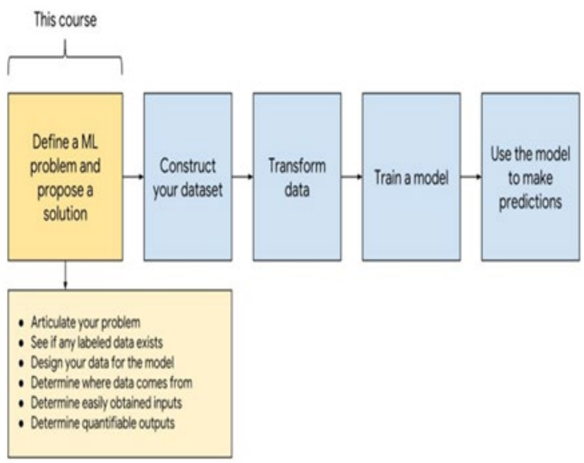


Fig.1. system architecture

5.5 System Design

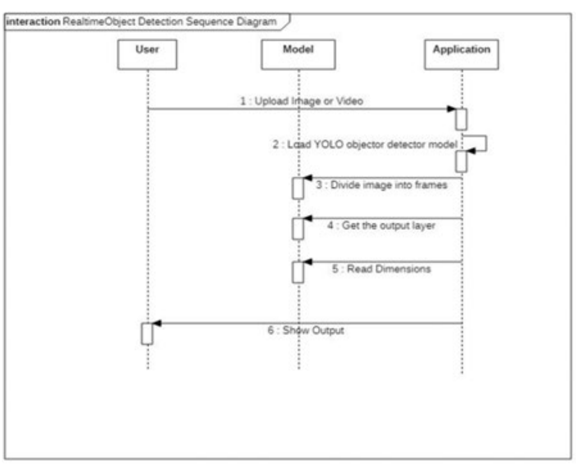


Fig.2. Class diagram for healthcare chatbot

The class diagram is the main building block of object-oriented modeling. It is used for general conceptual modeling of the structure of the application, and for detailed modeling translating the models into programming code. Class diagrams can also be used for data modeling.

5.6 Test Cases

Table.1. Test Cases with their respective results

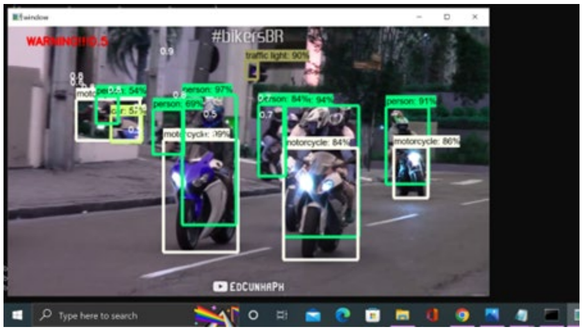
Tested	Test name	Inputs	Expected output	Actual Output	status
1	giving test image or video	image or video	input taken	successfully taken	success

Tested	Test name	Inputs	Expected output	Actual Output	status
2	loading model	invoking YOLO model	model loaded successfully	Model Loaded	success

Tested	Test name	Inputs	Expected output	Actual Output	status
3	object detection	frames divided by model	object detected	successfully detected	success

Tested	Test name	Inputs	Expected output	Actual Output	status
4	displaying output	detected object in image	image with object name in box	successfully displayed	success

5.7 Results



VI. Conclusion

In this study, a vision-based object detection and recognition framework was proposed for autonomous driving. The proposed framework contains one object detection task and three recognition tasks. Various objects are detected by using an optimized Faster RCNN algorithm model with fewer parameters, which can achieve faster processing speed and higher detection accuracy than the original. For detected objects, vehicles, pedestrians, and traffic lights are extremely important objects in the self-driving topic. Thus, there are three recognition tasks for the corresponding objects. By comparing with different CNN models, the most suitable model with the highest accuracy is selected for each recognition task. Besides, the RISE algorithm is used to explain the classification results by making the corresponding saliency maps for each image.

Future Enhancement:

- In the future, more attention should be paid to improving the overall speed of the proposed framework.
- To improve the performance of the system, a separate pipeline that can efficiently process single frame based and multi frame-based recognition can be applied in the following study.
- Considering that the distance between the autonomous vehicles with other objects is important, distance prediction also should be integrated into this framework.

References

[1]. Wei Liu and Alexander C. Berg, "SSD: Single Shot MultiBox Detector", Google Inc., Dec 2016.

[2]. Andrew G. Howard, and Hartwig Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications", Google Inc., 17 Apr 2017.

[3]. Justin Lai, Sydney Maples, "Ammunition Detection: Developing a Real Time Gun Detection Classifier", Stanford University, Feb 2017.

[4]. ShreyamshA mate, "UAV: Application of Object Detection and Tracking Techniques for Unmanned Aerial Vehicles", Texas A&M University, 2015.

[5]. Adrian Rosebrock, "Object detection with deep learning and OpenCV", pyimagesearch.

[6]. Mohana and H. V. R. Aradhya, "Elegant and efficient algorithms for real time object detection, counting and classification for video surveillance applications from single fixed camera," 2016 International Conference on Circuits, Controls, Communications and Computing (I4C), Bangalore, 2016, pp. 1-7.

[7]. Akshay Mangawhai, Mohana, Mohammed Leesan, H. V. Ravish Aradhya, "Object Tracking Algorithms for video surveillance applications" International conference on communication and signal processing (ICCSP), India, 2018, pp. 0676-0680.

[8]. Apoorva Raghunandan, Mohana, Pakala Raghav and H. V. Ravish Aradhya, "Object Detection Algorithms for video surveillance applications" International conference on communication and signal processing (ICCSP), India, 2018, pp. 0570-0575.

[9]. Manjunath Jogin, Mohana, "Feature extraction using Convolution Neural Networks (CNN) and Deep Learning" 2018 IEEE International Conference on Recent Trends in Electronics Information Communication Technology, (RTEICT) 2018, India.

[10]. Arka Prava Jana, Abhiraj Biswas, Mohana, "YOLO based Detection and aClassification of Objects in video records" 2018 IEEE International Conference on Recent Trends In Electronics Information Communication Technology, (RTEICT) 2018, India.