NORTH SOUTH UNIVERSITY

Department of Electrical and Computer Engineering

Comparative Analysis of Deep Learning Methods for Detecting Car Crash

Fahim Faisal Deepto ID: 1811899042

Shakiful Alam Oyon ID: 1821512042

Faculty Advisor:

Dr. Shahnewaz Siqqique

Assistant Professor

CSE 498R, SUMMER 2022
Direct Research

Declaration

It is hereby acknowledged that:

- No illegitimate procedure has been practiced during the preparation of this document.
- This document does not contain any previously published material without proper citation.
- This document represents our own accomplishment while being Undergraduate Students in the North South University.

Sincerely,

Fahim Faisal Deepto
ID: 1811899042

Shakiful Alam Oyon
ID: 1821512042

Approval

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation.

Dr. Shahnewaz Siqqique
Assistant Professor
Department of Electrical and Computer Engineering
North South University
Dhaka, Bangladesh

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation.

Dr. Rajesh Palit
Professor & Chair
Department of Electrical and Computer Engineering
North South University
Dhaka, Bangladesh

Acknowledgement

We would like to begin with our gratitude towards North South University's Department of Electrical Engineering and Computer Science for providing us with the platform to showcase our design capabilities, troubleshooting ability and implementation of theoretical knowledge fed to us through the core courses designed in the program and ultimately leading to the completion of senior design project.

Our most sincere gratefulness is to our project supervisor Dr. Shahnewaz Siqqique Assistant Professor for his relentless support and motivation throughout the project term for which we shall remain indebted forever. The completion of this project would have been implausible with his support and supervision.

Last, but not the least, we would like to thank our family members, friends, fellow classmates and all other personal, to whom we might have caused anyinconvenience to, during the project term, for their understanding and support.

Contents

Abstract

In this day and age, accidents are rapidly becoming one of the most significant problems. Since more and more vehicle accidents take place, there is an increasing demand for extremely accurate systems that can identify oncoming crashes. We propose an automobile accident detection system that improves the performance of car crash detection by combining video data from different types of cameras, such as dash cams and CCTV cameras. This allows the system to better identify when a crash has occurred. Because different forms of data obtained from the same information source (for example, dashboard cameras or CCTV) may be viewed as numerous views of the same source, the proposed vehicle collision detection system makes use of an ensemble deep learning model that is based on multi-modal data (i.e., both videos). Because it's possible that one view will have information that another view won't, having so many different perspectives will help boost detection performance because they'll complement one another. The proposed automobile collision detection system is verified by comparing it to single classifiers that only utilize video data from dashboard cameras. The suggested system is built on multiple classifiers that use both video and audio data from the dashboard cameras. The authenticity of the findings of this study's results is supported by a number of recordings taken from dash cams and CCTV systems. The findings of the experiments demonstrate that the automotive collision detection system that was developed performs significantly better than single classifiers.

Introduction

Every year, automotive accidents are responsible for the deaths of around 1.3 million individuals. Even though these countries only account for around 60 percent of the world's automobiles, they are responsible for 93 percent of all deaths that occur on the world's roads. Road traffic accidents lead to huge financial losses for everybody involved, including the victims and their families as well as entire nations [1].

Accidents caused by vehicles are an even more widespread concern in Bangladesh. The number is climbing higher and higher with each passing year. The economy of Bangladesh is expanding and Bangladesh is a populous developing nation. On the other hand, the nation's infrastructure is not expanding at a rate that is fast enough to keep up with the rapid growth of the population. Even if there are more automobiles on the road, there simply aren't enough individuals who are able to drive them or enough roadways. Because of this, there have been an increased number of automobile accidents. There were 21 percent more incidents in 2021 than there were in 2020, which resulted in 6,284 fatalities and 7468 injuries, as opposed to 5431 fatalities and 7378 injuries in 2020, according to the annual report of the Bangladesh Road Safety Foundation (BRSF). In 2020, there were 7378 injuries and 5431 fatalities. According to the findings of the research, the cost to the country's human resources due to the fatalities and injuries was 9,631 crore (Tk) [2]. Accidents have turned Bangladesh's working population into a deathtrap. Sixty-seven percent of those who are killed or injured in car accidents are in economically productive age groups (15-64 years). It is anticipated that by the year 2030, accidents involving motor vehicles will rank as the seventh most common cause of death across the globe. As the number of traffic accidents rises, there has been a rise in demand for high-performance crash prediction systems.

We developed a model that is capable of predicting accidents before they take place and providing drivers with advance warning in order to successfully minimize the number of traffic accidents and the costs that are connected with them in order to solve this worldwide problem. My proposed approach, titled "The Most Effective Deep Learning Method for Detecting Car Crashes," would utilize an all-encompassing strategy to help reduce the number of accidents that occur on the roads. In order to offer drivers with early warnings, we will first find the best way for detecting the vehicle accident, and then we will develop a reliable accident prediction system. This will allow us to provide drivers with early warnings. In order to put this theory into reality, we will first compile video material from a variety of sources, then apply ensemble deep learning models to this data. In order to provide a reliable and precise prediction, we will take into account all of the possible modeling aspects that serve as early warning indicators of an impending catastrophe. We will be able to generate a prediction model that is more precise if we use this strategy [3].

Literature Review

In the part on the review of the literature, the topic of detecting vehicle accidents using video and audio data is discussed.

A. System using dashboard cameras and multimodal data [4]

They created a car collision detection system by combining ensemble deep learning with multimodal data from dashboard cameras. It got the information from the YouTube website. They developed a system that can identify accidents by merging data from both audio and video sources. In order to achieve this goal, they developed an independent GRU and CNN that was centered on audio and video, which they then integrated in order to identify accidents. As a direct consequence of this, a cutting-edge model of categorization was developed [4]. This paper will help us train our model to spot an automobile collision. The detection modeling methodologies outlined in the paper will aid us in creating a more viable prediction system to help us achieve our objective.

B. Accident detection and messaging using GPS & GSM [5]

In the following study, we can observe that initially, hardware components are used to find the accident. It is equipped with three separate sensors that can detect an accident. Vibration sensors, speed sensors, and collision detection sensors are utilized in the sensing approach. The maximum amount of vibration that can occur while driving is programmed into the vibration sensor. When a collision happens, the car's vibration level is substantially altered. The vibration sensor detects this change in vibration, which contributes in the correct identification of the accident [5].

According to this paper, hardware components are used to identify accidents. However, we are looking for the most efficient deep learning technique in this instance. As a result, this study will not be included in our methodology.

C. Unsupervised detection [6]

In the next part, they provide an unsupervised method for detecting traffic accidents in first-person (dashboard-mounted camera) films. Our most significant innovation is the capacity to detect abnormalities by predicting the future positions of traffic participants and then analyzing the prediction accuracy and consistency criteria using three distinct approaches. Using AnAn Accident Detection (A3D), an unique dataset of heterogeneous traffic occurrences, and another publicly accessible dataset, we evaluate our methodology. According to experimental findings, our method outperforms the industry standard [6]. This article will assist us in training our model to recognize auto accidents. The detection modeling techniques presented in the article will assist us in developing a more promising prediction system that will aid us in reaching our objective.

D. Real time detection using YOLO and centroid tracking [7]

This paper presents a computer vision framework that can identify road traffic crashes (RTCs) using the installed surveillance/CCTV camera and notify them in real-time to emergency personnel with the specific location and time of the event. The framework consists of 5 components. The first module is vehicle recognition utilizing YOLO architecture; the second module is vehicle tracking with MOSSE tracker; and the third module is a novel method for identifying accidents based on collision estimation. The fourth module determines if an automotive collision has occurred

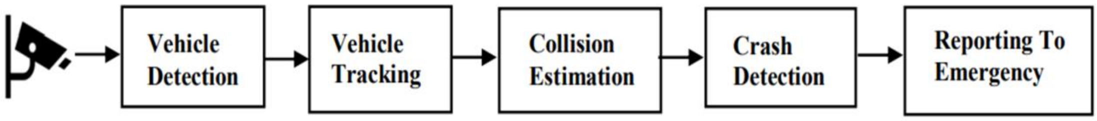


Figure: Proposed framework for accident detection [7]

prediction for every car [7]. CCTV footage is being used in the investigation of a car collision. This document will be used to establish the most efficient method for finding the car collision.

E. Crash detection using YOLO algorithm [8]

The proposed accident detection system can be trained on sample vehicle datasets using the YOLO (you only look once) regression-based algorithm and the vehicle detection process has been successfully executed by the trained model vehicle detector being tested on the test data set with live video feeds from the webcam. In terms of object prediction, the proposed system beats other object detection techniques, such as Faster-CNN and Fast CNN. Additionally, input can be modified to achieve improved outcomes. In addition, wireless communication devices are utilized to transmit messages to adjacent emergency vehicles [8]. In the subsequent investigation, it is shown that they employ dashboard cameras to identify and compare their model to other models, achieving superior results. As a consequence, we may discard the other model and apply this model to the other model's technique.

Proposed Methodology

Diagrammatically depicted in figure below is our approach to our analysis.

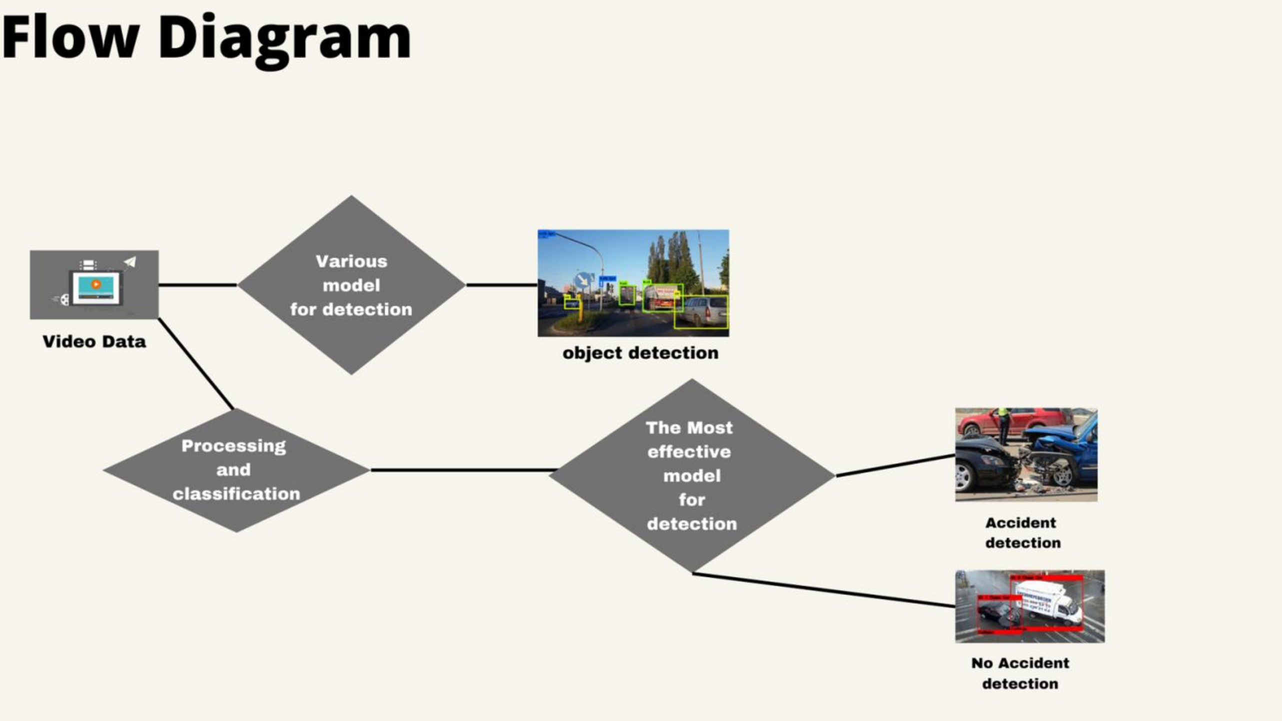

Flow Diagram
Video Data
Various model for detection
object detection
Processing and classification
The Most effective model for detection
Accident detection
No Accident detection

Figure: Work Flow

A. Obtaining Data and Preprocessing

In the first step of this process, we will gather video examples of accidents from a variety of online sources. A frame-by-frame analysis will be performed on the videos. Therefore, each frame will be processed in the same manner as a picture. The information that was gleaned from the films will be subjected to stringent processing before being made available for training. In addition, because the data may contain a large number of picture classes, we will only categorize those photographs that are essential to the completion of our task.

This will be done so that we could tell why our system is making a prediction and get proper clarity. By doing all these, our repository would have been successfully created that is ready for training and testing.

B. Training the Model and Creating Prediction System

Following the completion of the picture processing, the next step will be to train the model. However, the Deep Learning algorithms can only comprehend numerical values. Therefore, we will utilize YOLO to apply the Bounding Box properties to the photos. This will categorize all of the items contained inside our picture collection, bringing us one step closer to having a finished dataset. After that, a model will be created from the dataset by applying the YOLO technique to do training. A relatively insignificant fraction of the dataset will be retained for use in testing the model at a later date. However, this will just provide object detection and the location of the items inside the photos. Following this, we will proceed to add constraints to our model based on the data that we gathered from the various accident scenarios. Our model will gain an understanding of the conditions under which it may accurately anticipate the occurrence of an accident if we proceed in this manner. Because of this, our model will be able to make accidents foreseeable and, as a result, preventable based on the data that has been collected. In the end, the entire model is going to be validated using the test data from our dataset. If the accuracy isn't good enough, we're going to have to restart the training process and make the required adjustments until we have a result that's good enough.

YOLO (You Only Look Once)

Real-time detection requires an algorithm that is quicker than 24 frames per second (fps), which is the standard frame rate for videos. A typical movie has 24 fps. The YOLO system is a real-time object detection system that is state-of-the-art and is based on the architecture of the darknet. The algorithm ensures faster results at the expense of some accuracy; however, this trade-off is acceptable. Training on whole images allows YOLO to increase its detection performance. In comparison to more traditional methods of object detection, this unified model offers several benefits. In the first place, YOLO has a blazingly quick speed. YOLOv3 does image processing at a rate of 30 frames per second on a Pascal Titan X and has a mAP (mean Average Precision) of 57.9 percent on the COCO test-dev(2); a faster version analyzes pictures at a pace of 150 frames per second (with less mAP) [9]. Because it is 100 times quicker than fast RCNN and 1000 times quicker than RCNN, YOLOv3 is suitable for real-time object detection [10]. This is because it maintains the same degree of accuracy despite its increased speed. The concept of "you only live once" originates from the fact that the whole prognosis is based on a single visual analysis. Because we consider detection to be a matter of regression, we do not consider a sophisticated process to be necessary. Executing our neural network on a fresh picture during the testing phase allows us to more accurately anticipate detections [9]. Because we consider detection to be a matter of regression, we do not consider a sophisticated process to be necessary. In order to make detection predictions, all we need to do is run our neural network on a fresh picture during the testing phase. In addition, when generating its predictions, YOLO takes into consideration the image in its whole. Unlike sliding window and area proposal-based techniques, YOLO observes the whole image during training and testing; hence, it implicitly stores contextual information about classes in addition to their appearance [9]. Therefore, not only do we find out whether or not the object is present, but also the location of the object inside the picture itself. The YOLO architecture is quite similar to the one seen in figure below [10].

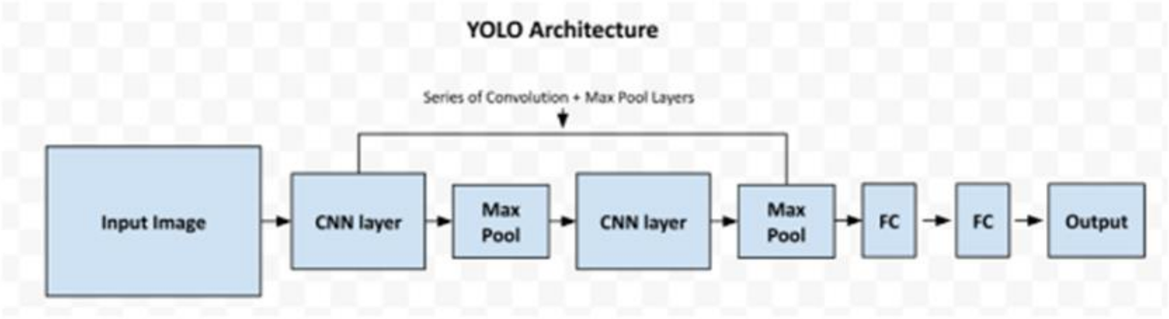


Figure: How YOLO operates [10]

Let's have a look at how the YOLO filter works with a few different examples.

The bounding box detection method is utilized by YOLO in order to locate an item within an image.

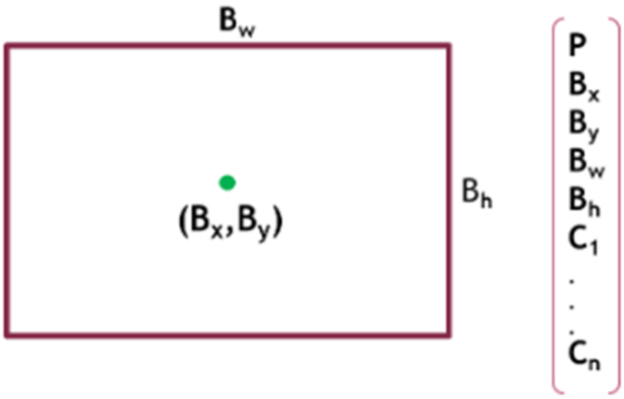


Figure: Bounding box of YOLO

Each bounding box has a vector, such to the one seen in the picture above, that encodes many properties of the item identified within the bounding box.

P= Probability of a class being in the box (value can be 0 to 1).

(Bx , By)= Center of the bounding box

Bw= Width of the bounding box

Bh = Height of the box

C1 ….. Cn = Classes of image(1 if the image class is present, 0 if not)

If there is just one object in a picture, as in Figure 4, the bounding box will detect the item and the probability variable in the matrix will equal 1 and the class detection variable will equal 1. We can determine whether or not the item is in the present based on these variables. But using YOLO, we may obtain more information, as we know the object's bounding box's center as well as its height and breadth. These values will provide the object's location inside the picture [10].

Now, when many items are present in a picture, there will be multiple classes (C1,...Cn). This is handled by the YOLO algorithm step by step. Consider two items of classes 1 (C1) = Car and 2 (C2) = Human. Initially, it separates the picture into grids, as seen in the figure below. Each of these grids will be responsible for detecting any objects within their boundaries.

Now consider a grid in which no object classes are declared. Then the probability variable in the vector will be P=0, and class type is unnecessary as there is no object. In contrast, a grid containing a vehicle will yield a vector with the values P=1 and C1=1, as it contains an object of the car type. From the remaining values of the vector, we will simultaneously determine its location in the picture.

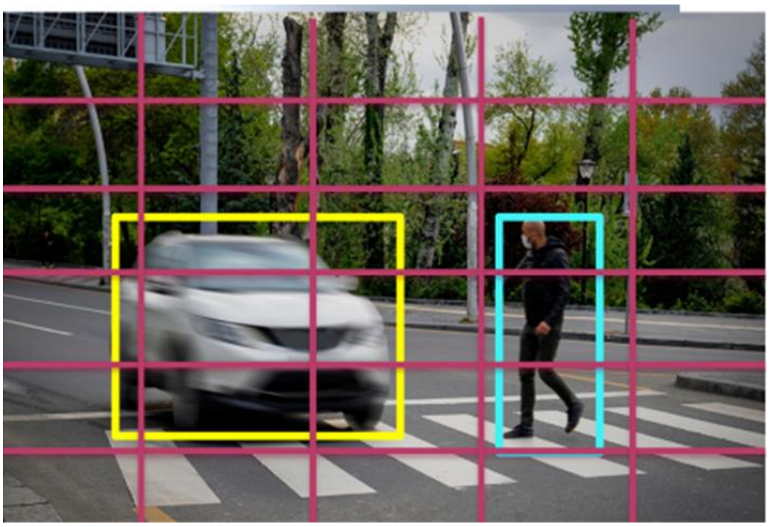


Figure: Multiple objects in an image

However, if YOLO forecasts two bounding boxes for an object-

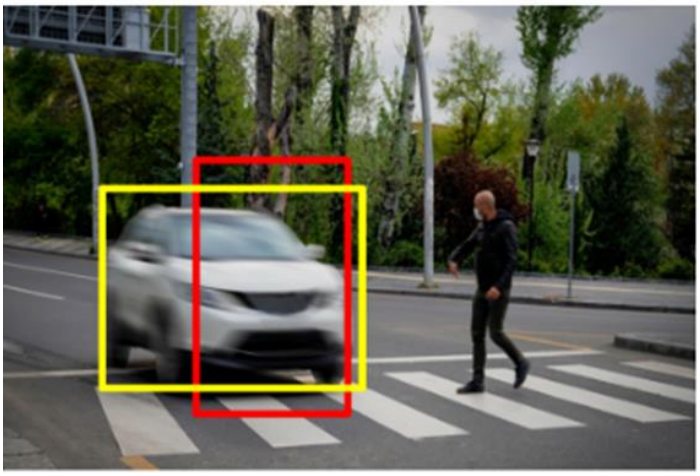


Figure: Two bounding boxes for a single object

The difficulty is addressed by YOLO by the utilization of the notion of Intersection over Union. This indicates that the algorithm will calculate the area that is created when these two bounding boxes intersect with one another and then divide that result by the area that is created when these two boxes are combined in order to determine both the prediction and the location of the object within the image.

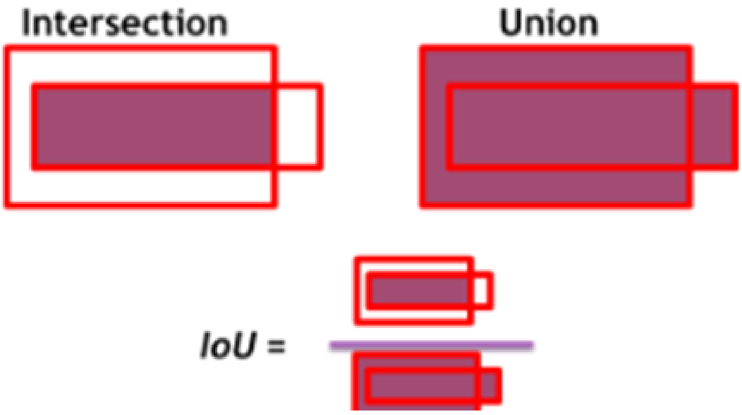


Figure: Intersection over union

OpenCv

OpenCV is a free and open-source library that focuses on finding solutions to issues related to computer vision. Assuming that your computers already have Python 3 installed, the pip package manager is the most user-friendly approach to add OpenCV support to Python. You may accomplish this by entering the following command line into the command prompt on your computer.

Paul Viola and Michael Jones devised the Viola-Jones method, which is used for object detection. The algorithm described above is based on machine learning. Training a cascade function with a large number of negative and positive labeled pictures is the initial stage. After training the classifier, "HAAR Features" are retrieved from the training pictures. HAAR characteristics consist mostly of rectangular areas of pixels with contrasting brightness and darkness [11]. The value of each feature is determined as the difference between the sum of pixel intensities in the bright region and those in the dark region. All conceivable picture sizes and locations are considered to determine these characteristics. A picture may have a large number of irrelevant characteristics and a small number of identifying features. The classifier is trained with a dataset that has been pre-labeled in order to extract the most helpful features and obtain the fewest mistakes by weighting each feature appropriately. An individual characteristic is known as a weak feature. The final classifier is the weighted sum of the weak characteristics. Only a small portion of the picture has the item to be detected; the majority of the image comprises the backdrop. Cascaded classifiers are employed to boost the detection rate. In this method, if a portion of an image has even a single negative characteristic, the algorithm disregards that part and continues on to the next region. In the picture, the sole region containing all the identifying characteristics is highlighted as the needed object.

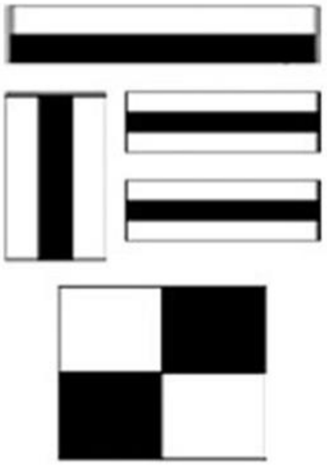


***Figure: Common HAAR features** [11]*

DenseNet

Each layer in DenseNet receives extra inputs from all of the layers that came before it and then passes on its own feature maps to all of the levels that come after it. There is a use for concatenation. Each layer is imparted with some of the "collective wisdom" of all the layers that came before it [12].

Because each layer is given feature maps from all of the layers that came before it, the network may be made more thin and compact; in other words, there can be fewer channels. The growth rate, shown by the symbol k, is the total number of extra channels added to each layer. As a result, both its computational and memory efficiencies are significantly improved [13]. The notion of concatenation is shown in the accompanying image, which depicts forward propagation.

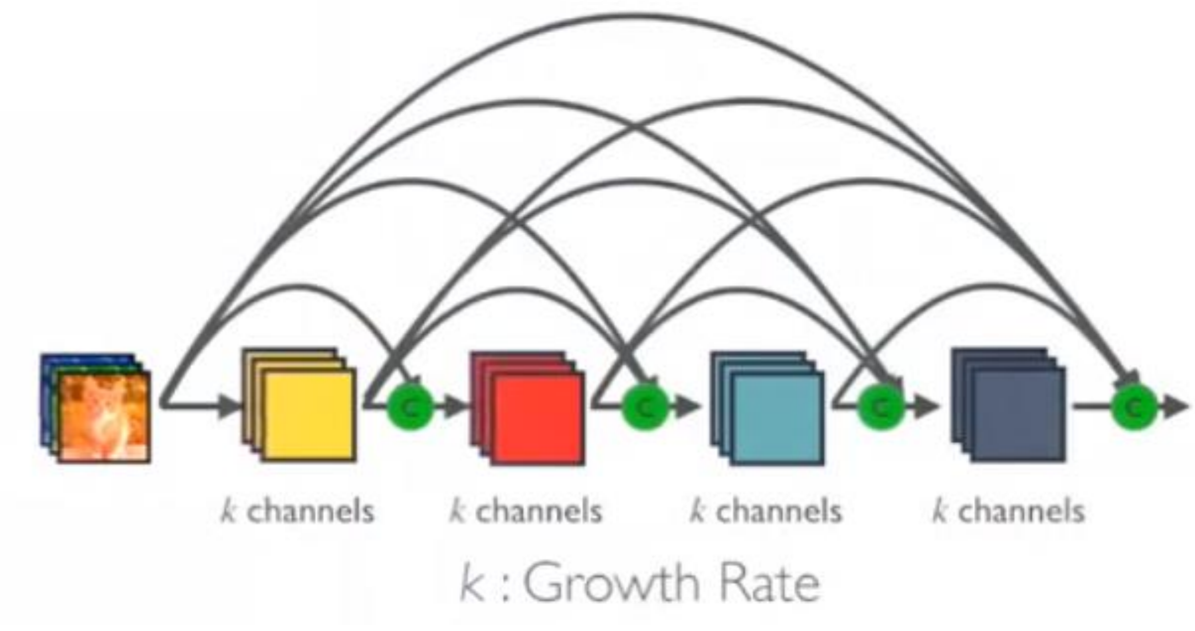


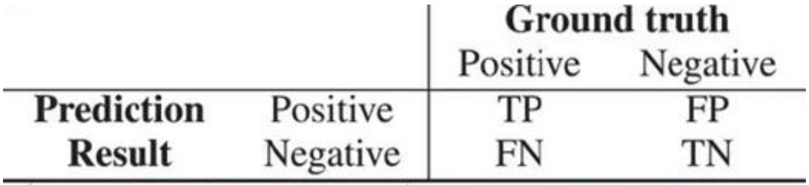
Figure: Dense block with growth rate of K [12]

Result Analysis

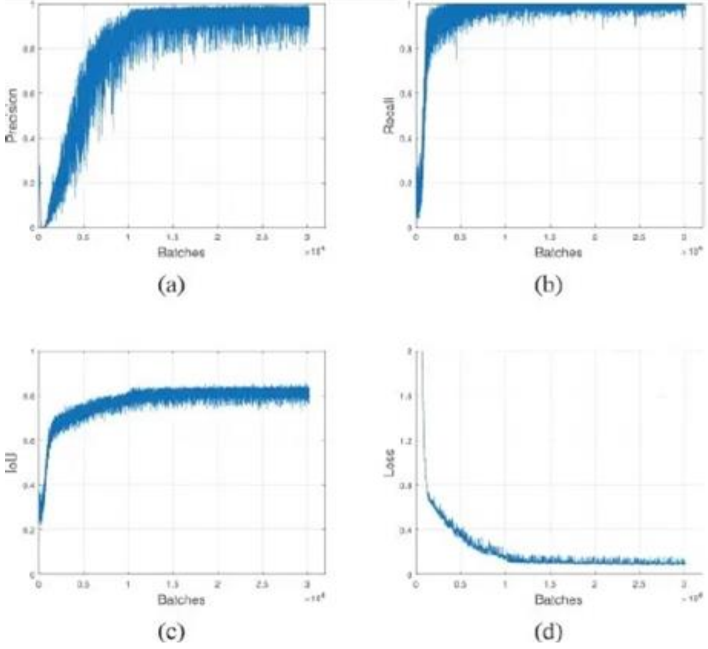
In the following section of our technique, we will talk about the many kinds of models that may be used to quickly identify objects while maintaining the necessary level of accuracy. We test each model with the identical kind of video data in order to determine which one is the best at recognizing whether a car has been involved in an accident. in order for us to accurately determine the real model for the detecting process. Here, we need to keep a few things in mind, the most important one being that we are going to employ the most efficient model for auto accident detection in our future plan, which is entitled "You only live once." Within the scope of that research, we will make use of this model to determine an appropriate prediction model for the driver of a vehicle or bike. Here, initially, we compile a wide variety of video data that is comparable to the situation in our nation. First, we employ a manual filtering process to separate the CCTV footage from the rest of the movie. According to the findings of the Background study, people have achieved a satisfactory level of accuracy in the detection process by making use of CCTV footage. However, CCTV footage cannot be included in the prediction model for the foreseeable future. It is possible to utilize it as part of an emergency call project in order to save someone who has been injured in an accident.

SVM classifier is what we utilize in this place for the video pre-processing. After we were through with the preprocessing, we trained our data on a number of different models to see which models had a high level of accuracy when it comes to detecting vehicle accidents.

The YOLO-CA training outcomes, including the changes in accuracy, recall, IoU, and loss that occurred throughout each batch iteration procedure. During the YOLO-CA training procedure, we consider the prediction result to be a true result if it has an IoU that is more than 0.5 and the correct classification. All other predictions are considered to be false results.



		Ground truth	
		Positive	Negative
Prediction Result	Positive	TP	FP
	Negative	FN	TN



(a) (b)

(c) (d)

Figure: YOLO CA training outcomes- (a)Precision (b)Recall (c IoU (d)Loss [8]

The outcomes of the predictions may be broken down into the following categories:

1. TP stands for "Truth Positive."
2. FP: Shorthand for "false positive."
3. FN: known as a “false negative”
4. TN: Shorthand for "True Negative."

The formula for calculating precision is as follows-

Precision = TPTP + FP

and the formula for calculating recall is as follows-

Recall = TPTP + FN

The precision of YOLO-CA is improving progressively along with the increase in the number of iterations, and it has converged to around 90 percent [8]. In addition, recollection ultimately reaches a point where it is more than 95 percent [8].

Experiments in comparative research are carried out in order to evaluate seven different detection models:

- Models with only one stage, including SSD, our suggested YOLO-CA, the standard YOLO-v3, and YOLO-v3 with no MSFF (Multi-Scale FeatureFusion).
- Two-stage models, including fast R-CNN, faster R-CNN, and fast R-CNN with forward propagation of noise.

The indices have been chosen for comparison among the seven models in order to illustrate the comparative validity of YOLO-CA as well as affirm its strength in terms of the comprehensive performance on the accuracy and real-time.

- Average Precision (AP) is defined as the average value of precision under varying recall, which may be altered by modifying the confidence threshold. The AP index measures the accuracy of detection models. Calculating the average precision is possible.

  $$AP =\sum \text{precision}(r), \text{ where r is recall.}$$

- Average Intersect over Union (Average IoU) is used to measure the object finding performance of a computer vision system detection models. The Average IoU is the mean of the IoUs between each predicted bounding box and the ground truth.
- Frames Per Second (FPS) Inference time is defined as the average time cost of finding a frame among test set. FPS is the inverse of inference time, which is defined as the average number of frames per second that may be identified.

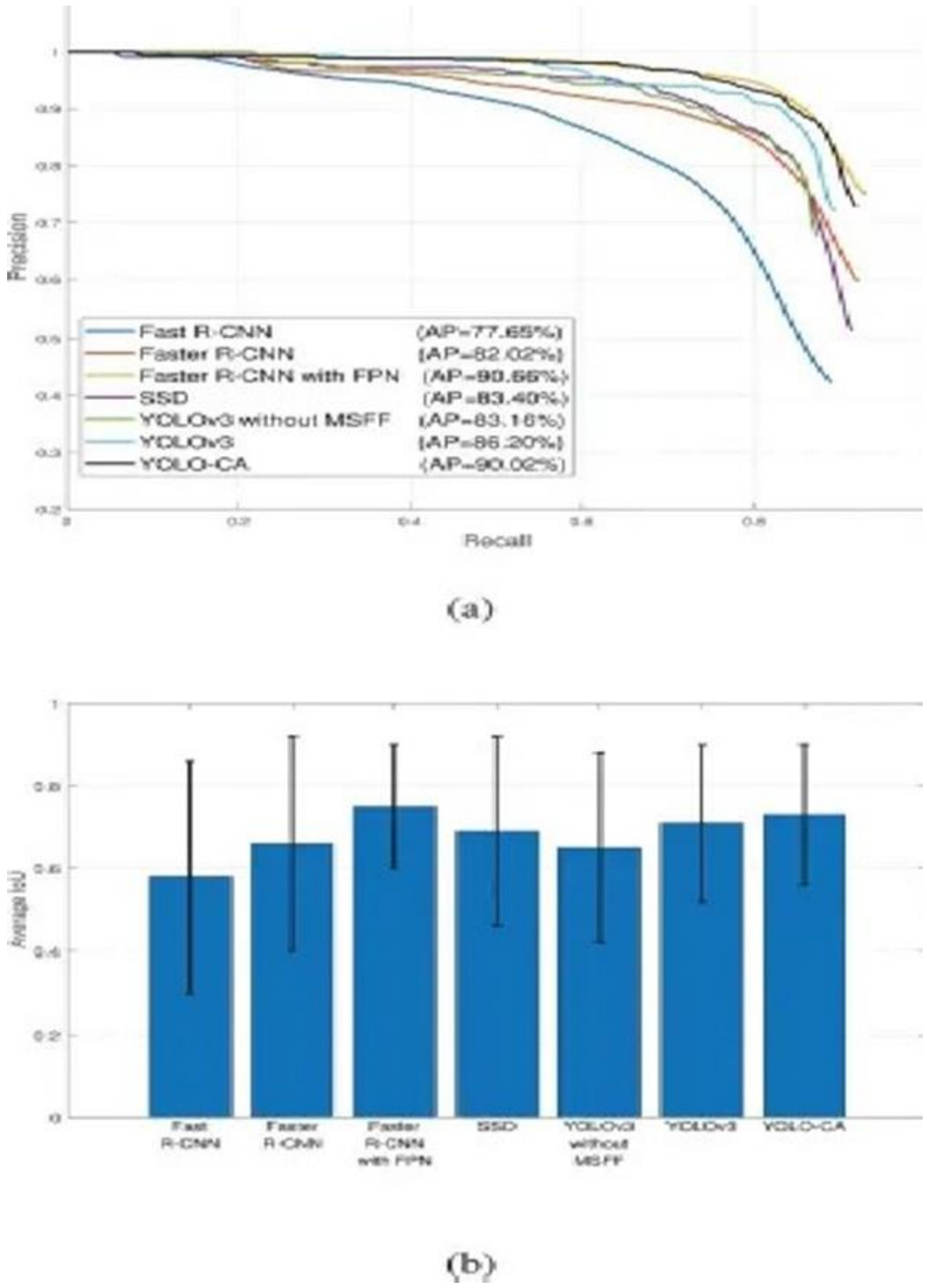


Figure: AP and IoU results from various models. (a)Precision-Recall curve (b)IoU average *[8]*

Conclusion

In this investigation, the proposed accident detection system is able to be trained by employing a regression-based algorithm known as the YOLO (you only look once) algorithm on the sample vehicle datasets. Additionally, the vehicle detection process has been successfully carried out by the trained model vehicle detector being tested on the test data set with the live video feeds from the webcam. The suggested system detects objects more accurately than previous object identification algorithms, such as Faster-CNN and Fast CNN, and detects them more quickly than other object detection methods. Additionally, the input can be improved, which will lead to superior outcomes. We now know that YOLO operates quickly in the event that it detects a car collision. Now, for the future implantation of this system, we are going to use this YOLO in conjunction with DenseNet to determine the most appropriate prediction system for the driver of a car or bike. Because of this, the rate of accidents in our nation as well as those in other countries will decrease, and this will be beneficial in a variety of different ways.

Bibliography

[1] WHO. (2022) World Health Organization. [Online]. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

[2] GRSF. Global Road Safety Facility. [Online]. https://www.roadsafetyfacility.org/country/bangladesh

[3] Atlas Magazine, "Road Safety in 2018," *Atlas Magazine: Insurance news around the world*, 2020.

[4] Jae Gyeong Choi and Chan Woo Kong and Gyeongho Kim and Sunghoon Lim, "Car crash detection using ensemble deep learning and multimodal data from dashboard cameras," *Expert Systems with Applications*, vol. 183, p. 115400, https://doi.org/10.1016/j.eswa.2021.115400.

[5] Mantra and Parmar, Angel and Parmar, Kelvin and Patel, Dhruvil and Darji, Mittal Sanathra, "Car Accident Detection and Notification: An Analytical Survey," 2019.

[6] Yu and Xu, Mingze and Wang, Yuchen and Crandall, David J. and Atkins, Ella M. Yao, "Unsupervised Traffic Accident Detection in First-Person Videos," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019.

[7] Zillur and Ami, Amit and Ullah, Muhammad Rahman, "A Real-Time Wrong-Way Vehicle Detection Based on YOLO and Centroid Tracking," , 2020, doi: 10.1109/TENSYMP50017.2020.9230463.

[8] Prof. L. K. Wani, Md Maaz Momin, Sharwari Bhosale, Abhishek Yadav, and Manas Nil, "Vehicle Crash Detection using YOLO Algorithm," *International Journal of Computer Science and Mobile Computing*, 2022.

[9] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi, "You Only Look Once:Unified, Real-Time Object Detection," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[10] Manish Chablani. (2017) Towards Data Science. [Online]. https://towardsdatascience.com/yolo-you-only-look-once-real-time-object-detection-explained-492dc9230006

[11] Vipul Kumar, "How to Detect Objects in Real-Time Using OpenCV and Python," *Towards Data Science*, 2020.

[12] Sik-Ho Tsang. (2018) Towards Data Science. [Online]. https://towardsdatascience.com/review-densenet-image-classification-

b6631a8ef803

[13] Gao Huang and Zhuang Liu and Kilian Q. Weinberge, "Densely Connected Convolutional Networks," *CoRR*, 2016, URL: http://arxiv.org/abs/1608.06993.

[14] Mate and Ivasic-Kos, Marina and Pobar, Miran Krišto, "Thermal Object Detection in Difficult Weather Conditions Using YOLO," *IEEE Access*, vol. 8, pp. 125459-125476, 2020, doi: 10.1109/ACCESS.2020.3007481.

[15] Kh. Azizul Hakim, Md. Mehedi Hasan, and Surma Akter, "Automatic Vehicle Accident Detection and Messaging System Using GSM and GPS Module," DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING, CITY UNIVERSITY, DHAKA, BANGLADESH, 2020.