Vision-Based Motorcycle Crash Detection and Reporting Using Deep Learning

Sanjit Gawade[1], Nadine Dias[2]

[1]Student, Information Technology Department, Goa College of Engineering, India
[2]Assistant Professor, Information Technology Department, Goa College of Engineering, India

---

Abstract - *In recent years, two-wheelers have gained popularity amongst daily commuters, youngsters, and urban residents as two-wheelers are easy to operate in congested traffic conditions and are fuel-efficient. With the increase in the number of vehicles on road, there has been a significant increase in road accidents. When we compare safety features offered on vehicles, motorcycles are equipped with the least features compared to cars and other vehicles. This project endeavors to provide a means to explore a vision-only approach to detecting traffic anomalies. Considering the real-time operating aspect of the system, the YOLOv4 algorithm is considered for object detection and classification. The algorithm is trained using a custom dataset with 398 images of road anomalies involving a motorcycle. The Model performed exceptionally well and achieved mAP@50 of 74% and precision of 60%.*

***Key Words*: Deep Learning, Computer Vision, Yolov4, Object detection, Real-time, CNN.**

1. INTRODUCTION

One of the most compelling types of artificial intelligence is computer vision. Computer vision is the field of computer science that focuses on understanding complex human vision systems and enabling a computer to identify and process information in images and video streams like humans do. The application of computer vision is expanding in fields like healthcare, surveillance, action and activity recognition, military, agriculture, and manufacturing.

Computer vision algorithms analyze certain criteria in images and videos and then apply interpretation to predictive or decision-making tasks. Computer vision models are designed to translate visual data based on feature and contextual data learned during training.

Convolutional Neural Network is the foundation of modern computer vision algorithms. CV algorithms are based on CNN, which provides a dramatic performance improvement compared to traditional image processing algorithms. CNNs are neural networks with a multi-layered architecture that is used to gradually reduce data and calculation to the most relevant set.

After the initial survey of automated accident detection systems, it was identified that existing systems are mainly based on GPS and GSM technology, and ultrasonic and impact sensors inside cars to detect and report the occurrence of an accident. Researchers are working on vision-based accident detection, it was observed that those systems were trained using cars and heavy vehicle data. Such systems are based on finding acceleration anomaly, trajectory anomaly, and change in angle anomaly [1], this approach is effective to determine car accidents in normal traffic flow and good visibility conditions and requires post-processing of data.

Our proposed system emphasizes a vision-only approach to detect anomalies on road. The main focus is to detect motorcycle accidents, as there are fewer safety features on motorcycles, and injuries attained may prevent a wounded person from contacting emergency medical services and receiving medical attention. This research endeavors to provide a means to explore a vision-only approach to detecting traffic anomalies. The main focus is to detect incidents involving motorcycles, as there are fewer safety features on motorcycles, and injuries attained may prevent a wounded person from contacting emergency medical services. Considering the real-time operating aspect of the system, 3 algorithms were finalized Faster-RCNN, Yolov4, and Yolov4-Tiny. After training the models, it was found that the YOLOv4 algorithm Outperforms Faster R-CNN in terms of speed and Yolov4-Tiny in terms of accuracy. The algorithms were trained using a custom dataset with 398 images of road anomalies involving a motorcycle. The Yolov4 Model performed exceptionally well and achieved mAP@50 of 74% and precision of 60%.

2. Related work

The evolution of object detection started in the early 2000s. They followed the low and mid-level vision and followed the method of recognition by components. This method enabled object detection as a measurement of similarity between the object components, shapes, and contours and the feature that were taken into consideration were shape context, edgelessness, and distance transformation.[doi2000]. object detection genre was not making any progress as the performance of hand-crafted features became saturated. However in 2012, with the advancement in convolutional neural networks and deep convolutional networks, they were successful at learning robust and high-level feature representations of an image. The deadlocks of object detection were broken in 2014 by the proposal of the Regions with CNN features (RCNN) for object detection. In this deep learning era, object detection is grouped into two genres: two-stage detection and one-stage detection.

Traffic accidents have a short duration in driving videos, and the backgrounds of driving videos are dynamic and complex. These make traffic accident detection quite challenging. To effectively and efficiently detect accidents from the driving videos [2], in this work researchers have proposed an accident detection approach based on Spatio-temporal feature encoding with a multilayer neural network. A multilayer neural network is used to encode the temporal features of video for clustering the video frames. From the frame cluster obtained by encoding they detect the border frames as the potential accident frames.

In this work[5], "A review of object detection based on deep learning" by Xiao, Y., Tian, Z., Yu, J. Authors have compared traditional feature-based methods with the deep learning object detection methods for both low-level and high-level image features. This work includes a detailed study of backbone networks, loss function, and training strategies for the deep convolutional neural network.

In this work[4], Fast car accident detection in videos, by V. Machaca Arceda and E. Laura Riveros. The authors have proposed a three-stage proposal to detect fast car accidents. First using convolutional neural networks object detection is performed, You Only Look Once (YOLO) algorithm is used; the second stage is

a tracker to focus on each car; then the final stage for each car they have used the Violent Flow descriptor with a Support Vector Machine (SVM) to detect the car crashes.

3. Methodology

Building a network model to perform accident detection on real-time video we propose to train 3 algorithms to analyze their accuracy and inference speed

Our approach consists of

A. Building a custom dataset.

B. Training the candidate algorithms on the custom dataset.

C. Perform analysis

D. Used best-performing model weights for accident detection.

E. Build a Darknet on a local system to run inference on live videos.

   Using best-performing model weights, build a darknet to detect motorcycle accidents, and once an accident is detected generate an alert.

4. Dataset

There are only a few high-resolution datasets publically available for accident detection. While conducting the literature survey, it was observed that the datasets were densely populated with accidents involving four-wheelers, which fall under the category of large objects. To serve our purpose of detecting motorcycle crashes a new data set was composed of videos sourced from the web, containing CCTV footage involving a motorcycle accident. Using a python script, the video sequence data is broken down into individual frames and stored in jpg format. 166 unique images containing incidents involving a motorcycle, from 905 extracted frames containing are selected to train the network.

Image annotation is the process of labeling an image, which provides the location and class labels of the objects present in the image to the computer vision model. Roboflow annotation tool is used to annotate the images, there are six distinct class labels used while annotating the images. The classes are Accident, Bike, Person, Truck, Car, and Bus. Preprocessing is applied to training, validation, and testing set to assure learning and inference occur on the same image properties. The images have been resized from 1920x1080 to 416x416 to maintain the SxS aspect ratio used by the YOLO algorithm.

Dataset Augmentation is performed on the dataset to increase the diversity of the data. To train the algorithm for low-light scenes a grayscale filter is applied and the brightness level is adjusted to mimic night conditions. A gaussian blur filter is also applied to represent adverse weather conditions

Fig -1: Dataset After performing augmentation

4.1 Test/Train Split

After performing augmentation the total number of samples in the dataset is 398 images. Out of which 348 are used in the training set for the algorithm, 25 samples are used in the validation set, and 25 samples are kept for testing purposes.

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper.

5. Implementation

The research was conducted in 4 phases, at each phase different algorithm was trained on the same dataset and evaluated for Accuracy, mean Average Precision, recall, and Speed.

1 **Phase 1**

In phase 1 Faster R-CNN algorithm was trained for 2000 iterations. We achieved Average precision of 34% with a recall rate of 29.57 and mAP@50 66.52

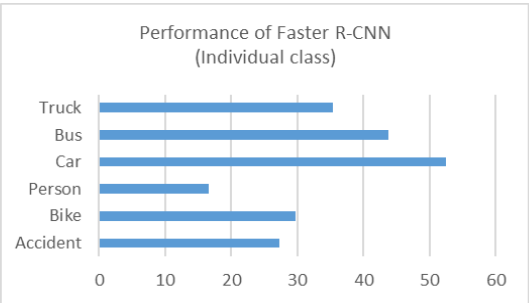

Fig -2: Performance of Faster R-CNN for individual class

2 **Phase 2**

In Phase 2 Yolov4 algorithm was trained for 2000 iterations with the default configuration. We achieved Average precision of 50% with a recall rate of 37 and mAP@50 of 40.02.

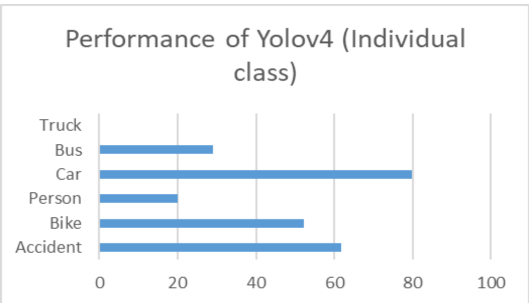

Fig -2: Performance of Yolov4 for individual class

3 **Phase 3**

In Phase 1 Yolov4-Tiny algorithm was trained for 10000 iterations with the default configuration. We achieved Average precision of 69% with a recall rate of 11% and mAP@50 of 32.23%

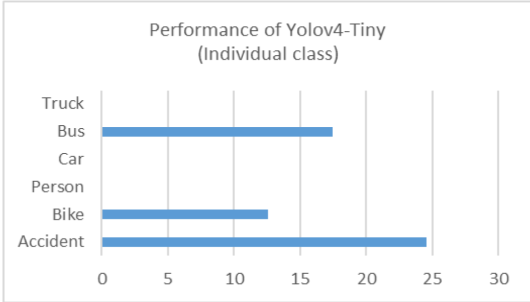

Fig -3: Performance of Yolov4-Tiny for individual class

4 **Phase 4**

In Phase 4 Yolov4 algorithm was trained for 2000 iterations with a reduced batch size of 12, instead of 64 in the default configuration. No pre-trained weights were used to train the backbone CNN. We achieved Average precision of 60% with a recall rate of 67 and mAP@50 of 74.3

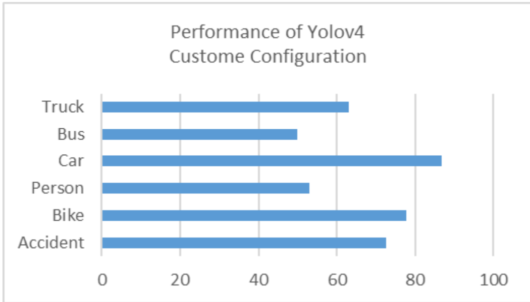

Fig -4: Performance of Yolov4 with a custom configuration for individual class

5.1 Accident Detection on a video stream

With the help of cuDNN Darknet was built on a local machine and weights from Yolov4 with a custom configuration were used to detect motorcycle anomalies. The model was successfully able to detect accidents from the video stream and upon detecting accidents a snapshot was captured of the accident and sent over to emergency response contact via email.

Fig -5: Accident Detection on a video stream

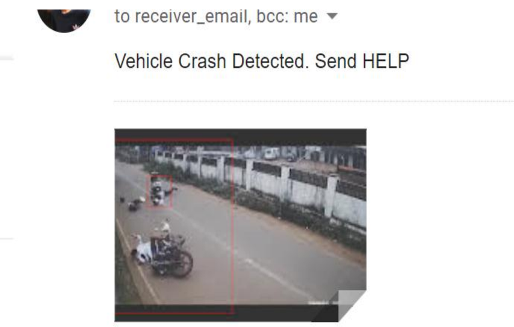

Fig -6: E-mail Alert

5.2 Analysis

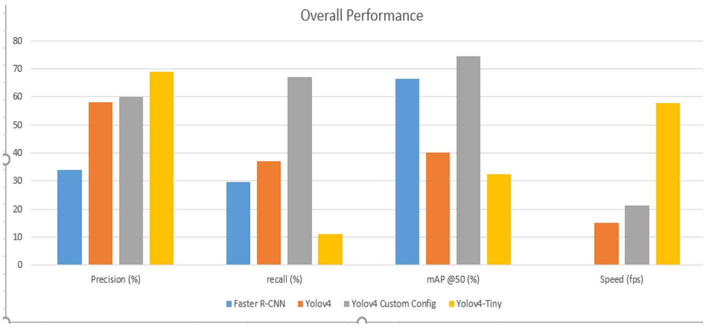

Fig -7: Performance of Analysis of Algorithms

After conducting the experiments it was observed that the Yolov4 algorithm with custom configuration performed the best. By reducing the batch size, the time it takes to train the model was drastically reduced by 67% and mAP@0.5 of 74.3% was achieved. The model achieved an average fps of 21.2 Frames per second.

The Faster R-CNN model can successfully identify accidents on images but failed to detect objects on a video stream.

Yolov4-Tiny is very fast, running inferences at an average of 58 fps, and the accuracy of detecting the object was low.

5. Conclusion

Vision-only approach to detection was successfully implemented and tested. Single-stage detectors like Yolov4 are suitable for real-time computer vision applications as they maintain the balance between accuracy and speed. Two-stage detectors are very slow for real-time application use, if high accuracy is the final goal, two-stage detectors are a suitable choice due to the dense network they can generate better feature maps and identify objects with high precision.

REFERENCES

[1] Z. Zhou, X. Dong, Z. Li, K. Yu, C. Ding, and Y. Yang, "Spatio-Temporal Feature Encoding for Traffic Accident Detection in VANET Environment," in IEEE Transactions on Intelligent Transportation Systems, DOI: 10.1109/TITS.2022.3147826.

[2] Vipul Gaurav, Sanyam Kumar Singh, and Avikant Srivastava, “Accident Detection, Severity Prediction, Identification of Accident Prone Areas in India and Feasibility Study using Improved Image Segmentation, Machine Learning and Sensors ”, International journal of computer science and network security 17, no. 6 (2017): 22-28

[3] Iman M. Almomani, Nour Y. Alkhalil, Enas M. Ahmad and Rania M. Jodeh, “Ubiquitous GPS Vehicle Tracking and Management System”, 2011 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), DOI: 10.1109/AEECT.2011.6132526R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[4] Arceda, Vicente & Riveros, Elian. (2018). Fast car Crash Detection in Video. 632-637. 10.1109/CLEI.2018.00081.

[5] Xiao, Y., Tian, Z., Yu, J. *et al.* A review of object detection based on deep learning. *Multimed Tools Appl* **79,** 23729–23791 (2020). https://doi.org/10.1007/s11042-020-08976-6

[6] Xiao, Y., Tian, Z., Yu, J. et al. A review of object detection based on deep learning. Multimed Tools Appl 79, 23729–23791 (2020). https://doi.org/10.1007/s11042-020-08976-6.

[7] Sowmya, V., and R. Radha. "Heavy-Vehicle Detection Based on YOLOv4 featuring Data Augmentation and Transfer-Learning Techniques." Journal of Physics: Conference Series. Vol. 1911. No. 1. IOP Publishing, 2021.