

Dashcam for Traffic Object Detection

Aura Syafa Aprillia Radim
*School of Electrical Engineering
Telkom University, Indonesia*

Bandung, Indonesia

syafaaura@student.telkomuniversity.ac.id

Muchammad 'Irfan Chanif Rusydi
*School of Electrical Engineering
Telkom University, Indonesia*

Bandung, Indonesia

Surya Michrandi Nasution
*School of Electrical Engineering
Telkom University, Indonesia*

Bandung, Indonesia

email address or ORCID

Casi Setianingsih
*School of Electrical Engineering
Telkom University, Indonesia*

Bandung, Indonesia

email address or ORCID

Abstract—Dashcam is a camera place on dashboard in vehicle. This tool serves to record all events in front of the vehicle. Security and safety have become a major concern in various sectors, including transportation and public roads. On the highway, traffic accidents caused by the driver's ignorance of objects around the vehicle are still a serious problem. In this study, the development of a simple dashcam built from an edge computer was carried out by combining the number of cameras. Image stitching is applied to combine images that have been collected by each camera. Next, object detection is carried out on the images that have been collected. The object detection system approach is carried out using YOLOv8 which is the latest variant of the YOLO series. This research is expected to be one step in the development of an Intelligent Transportation System that is in accordance with traffic conditions in Indonesia. The results obtained in testing using the system created exist using the configuration of 78,000 datasets, 3332 data validation with 8 epochs, batch size 32, linear learning rate and SGD optimization. Results are best in the morning and afternoon. The program can recognize predefined objects.

Index Terms—object detection, YOLOv8, dashcam

I. INTRODUCTION

Security and safety have become a major concern in various sectors, including transportation and public security. On the highway, traffic accidents caused by the driver's ignorance of objects around the vehicle are still a serious problems. Smart and effective object detection technology is becoming increasingly important for monitoring traffic [1].

Dashcam is a camera placed on the dashboard of a vehicle. This device usually serves to record all events in front of the vehicle. Dashcam is one of the devices that the demands is growing rapidly in the market. Currently, dashcam serves as a device to record events in front of the vehicle and then the recording is used as evidence in the event of an accident or proof of insurance claims. However, dashcam could also be used for other purposes, such as being used as one of the components of an Autonomous Driving Assistance System (ADAS).

In this paper, we propose a simple solution to prevent accidents amongs vehicles. Our solution is by implementing

object detection and single board computer (SBC) as processing unit. Object in front of the camera can be detected by Convolutional Neural Network. The content of this paper is organized as follows. Section 2 presents the literature review. Section 3 presents the system design. Section 4 presents our experimentaton in the process of developing the system. Section 5 presents the conclusion and future work.

II. LITERATURE REVIEW

A. Dashcam

Both China and South Korea governments obliges public transportation and commercial vehicles to install dashcam in order to assist the investigation of traffic accidents [2].

B. Object Detection

Object Detection is one of the important task in computer vision field, mainly dealing with detecting instances of visual object then categorize them into several classes [?]. With this kind of identification and localization, object detection can be used to count objects in a scene and determine and track their precise locations, all while accurately labeling them. Object detection has been widely used for face detection, vehicle detection, pedestrian counting, web images, security systems and driverless cars. Within the past twenty years, object detection have been going through a lot of changes and development. Although it is commonly divided into two periods: "traditional object detection" and "deep learning based". In 2012, Krizhevsky et al. [4] proposed a deep convolutional network trained on a subset of ImageNet.

This network, called AlexNet, was A year later, Girshick et al. proposed a new object detection framework called R-CNN [?] because it used region proposals combined with CNNs to detect objects in images. Since then, the field of object detection has been rapidly advancing, with new models, datasets, and techniques emerging in a rapid pace.

With the born of AlexNet, YOLO (You Only Look Once) model was introduced in 2015. Original base YOLO model can achieve 45 frame per second. While the sameller version,

Fast YOLO can achieve 155 frames per second. YOLO outperform DPM and R-CNN on Picasso Dataset and People-Art Dataset [6].

In the early period of making this paper, YOLOv7 was the latest version of YOLO. However, as January 2023, YOLOv8 was introduced by Ultralytics, the same software company that release YOLOv3 and YOLOv5. As of now, YOLOv8 is one of the latest State of The Art (SOTA) open source object detection model. Following its predecessor, YOLOv8 offers several model size. From the YOLOv8n as the smallest model, YOLOv8l as the medium model, and YOLOv8x as the largest model. We choose YOLOv8n as our model because it is the smallest model and can be run on edge computer. YOLOv8n have x parameters. With computation cost around 8 GFLOPs.

C. COCO Dataset

The COCO dataset is a large-scale object detection, segmentation, and keypoint dataset. In total, The Microsoft Common Objects in COntext contains 91 common object categories with 82 of them having more than 5,000 labeled instances [7]. The first release of COCO Dataset was in 2014. In 2014, COCO Dataset has 83,000 image in train split and 41,000 image in validation split. After taking community feedback, COCO Dataset increase total number of images and changes its split proportion. In 2017 release, train2017 split have 118,000 image files and 5,000 images in val2017 split. Given its size, COCO Dataset widely used by state of the Art Object Detection Model to train and evaluate the model performance.

Despite its popularity, COCO Dataset has some drawbacks. COCO Dataset contains 80 classes. However, COCO Dataset have imbalance class distribution. For example, the number of person class is 64115, while the number of hair drier class is only 100.

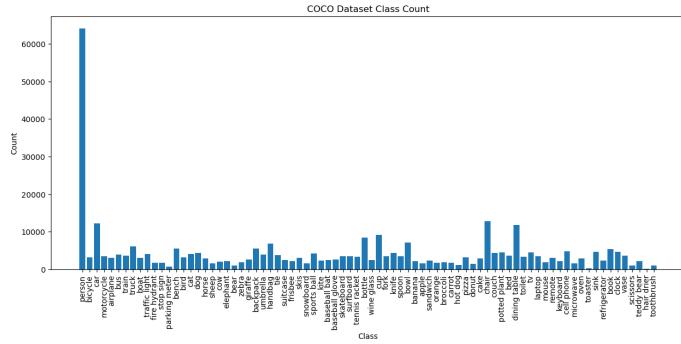


Fig. 1. COCO Dataset Class Count

D. Filtering Dataset

As mention earlier, COCO dataset contains 80 classes. However, we only need traffic related object classes. Therefore, we filtered the dataset to only contains 12 classes. The classes are: Car Truck Bus Motorcycle Bicycle Traffic Light Stop Sign Train Hydrant Cat Dog The result is 78,663 images in training split with 12 classes. This amount of data was reduced from

122,125 images in original dataset. By reducing the dataset, our model expected to take less time to train.

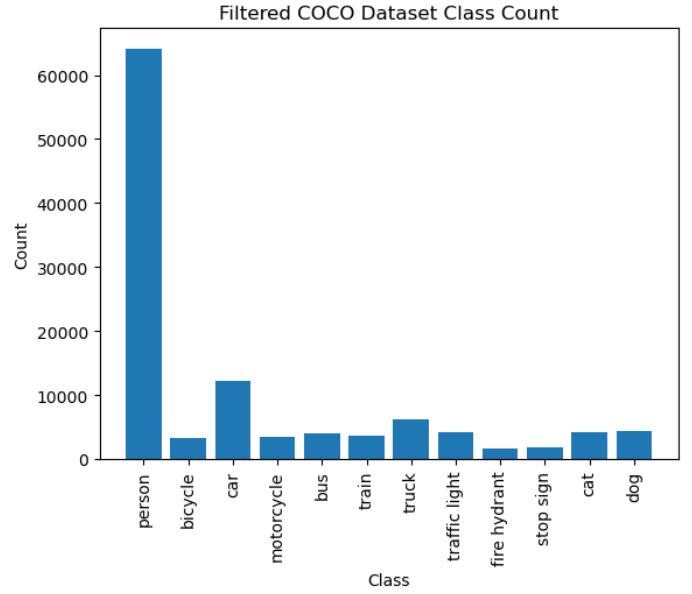


Fig. 2. Filtered COCO Dataset Class Count

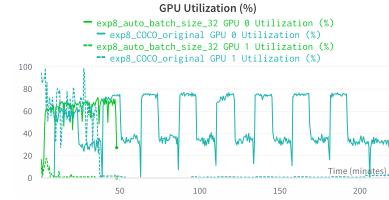


Fig. 3. Comparison of GPU Utilization between training the model using filtered dataset and original dataset

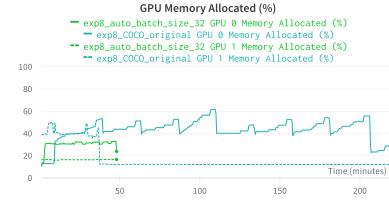


Fig. 4. Comparison of GPU Memory Utilization between training the model using filtered dataset and original dataset

In addition, as shown in Fig. 3 and Fig. 4, the filtered dataset has lower GPU Utilization and Memory Utilization. This means that the filtered dataset is more efficient to train. Thus allowing us to push model training further. Metrics wise, the model trained using filtered dataset showed better result.

E. Performance Metrics

For training and validation purposes, we focus on mAP (mean Average Precision), Precision, and recall. mAP is the

average of AP (Average Precision) over all classes. AP is the area under the precision-recall curve. We compare mAP averaged for IoU (Intersection over Union) thresholds from .50 to .95 with .05 increments (MS COCO standard metric, abbreviated as mAP50-95) and mAP50 (PASCAL VOC metric, abbreviated as mAP50) [7]. Precision and Recall in this experiment is a relative measure, because it depends on the threshold value. Precision calculated by dividing True Positive by the sum of True Positive and False Positive. Precision is calculated using (1).

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

Recall calculated by dividing True Positive by the sum of True Positive and False Negative.

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \quad (2)$$

Threshold value determines whether the prediction is Positive or Negative. For example, if the threshold is 0.5, then the prediction is positive if the IoU is greater than 0.5. Otherwise, the prediction is negative.

Both Precision and Recall are relative to the threshold value and usually used in the form of Precision-Recall Curve. We can calculate the area under the curve to get the Average Precision To get the value of mAP, we need to calculate Average Precision for each class. Then, we calculate the mean from all Average Precision we calculate earlier.

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (3)$$

η is the number of classes. This is the main reason of filtering Dataset would be beneficial in our training process. As we use filtered dataset, the value of η in equation 3 is 12 instead of 80. Figure 3 shows the comparison of mAP50 and mAP50-95 between training the model using filtered dataset and original dataset. We can see that the filtered dataset has better mAP50 and mAP50-95 than original dataset.

III. SYSTEM DESIGN

A. Related Work

Nasution et.al [9] proposed a system that can detect vehicles and also street lane. Image feed captured using smartphone, then processed using by ImageAI library with RetinaNet for COCO as object detection model

Nasution and Dirgantara (2023) proposed pedestrian detections for Autonomous Driving Assistance System (ADAS) using YOLOv5 and Raspberry Pi 4 as the edge computer. The pedestrian detection system is only capable to deliver 0,9 Frame Per Second.

B. System Overview

In this section, we will discuss about the system overview. The system consist of 3 main hardware components. The first component is the camera. The camera is used to capture the image. The second component is the edge computer. The edge

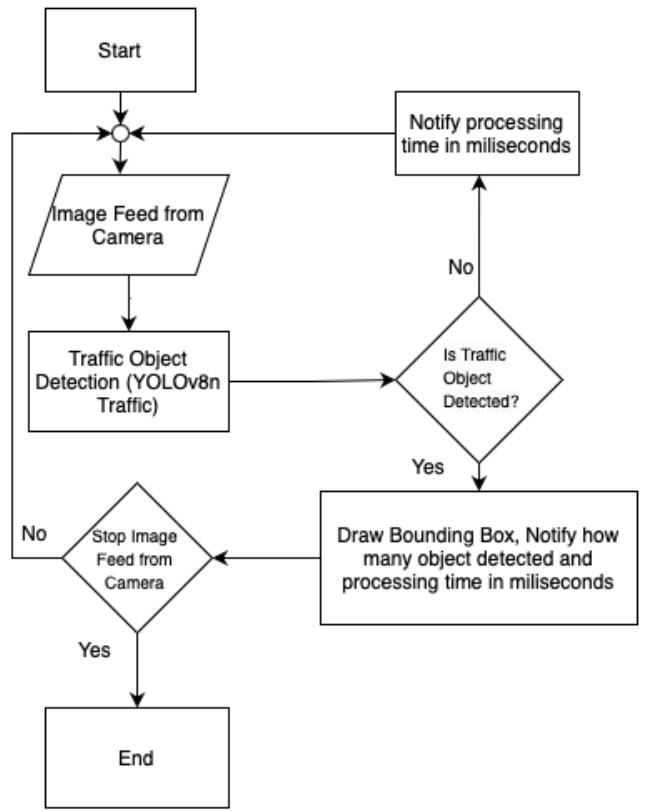


Fig. 5. Proposed Flowchart of the system

computer is used to process the image. The third component is the display. The display is used to show the result of the image processing.

C. Experimental Setup

In this section, we will discuss about the experimental setup. We are using Pytorch as our deep learning framework in Python 3. MNode training was conducted on Geforce RTX 3090, 24 GB VVideo RAM, 32GB RAM, Intel Core i3-12100 4 Core 8 Thread, running on Windows 10 Model inference and testing was performed on Macbook Air M1, and Jetson Nano 4GB. The camera used in this experiment is Lifecam Studio by Microsoft. Jetson Nano that we use running on Ubuntu 20.04.2 LTS on top of Jetpack 4.6. We use Jetson Nano 4GB as our edge computer because it is the most affordable edge computer in Nvidia Jetson Series.

D. Testing Dataset

As mention earlier, we are using COCO dataset. However, we filtered the dataset to only contain 12 classes. The classes are: Car, Truck, Bus, Motorcycle, Bicycle, Traffic Light, Stop Sign, Train, Hydrant, Cat, Dog. The result is 78000 images with 12 classes. For inference and real world testing, we gather our own data by recording traffic condition in Bandung, Indonesia. The data was recorded using camera mentioned earlier.

E. Training

Before we begin training process of our YOLOv8n-Traffic model with high number of epoch, we want to find the optimal hyperparameter configuration for our use case. We conducted various training configuration with 8 epoch to save time. Although each run for one configuration is only 8 epoch, it's still takes about 1 to 2 hours. However, we can get the result faster than training with 100 epoch for each configuration.

We tried different batch size, learning rate, and optimizer. We use batch size 16 and 32.

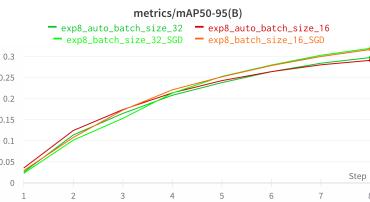


Fig. 6. Comparison of mAP50-95 between training the model using batch size 16 and 32

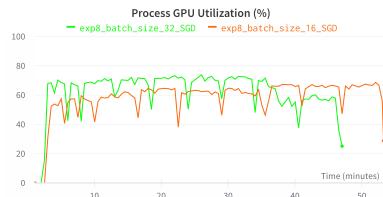


Fig. 7. Comparison of GPU Utilization between training the model using batch size 16 and 32

Fig.6 shows the result of training with different batch size. We can see that batch size 32 is better than batch size 16. Especially in terms of training time, batch size 32 is faster than batch size 16. Therefore, we use batch size 32 for our training. We also tried using SGD and Adam optimizer. YOLOv8 implement optimizer switching depends on the number of iterations.

F. Real World Testing

In this section, we will discuss about the real world testing. We bring our system to real world environment scenario, placed our whole system into the dashboard of vehicles.

We use USB car charger to power our system. We drove around the city of Bandung to test our system.

IV. RESULTS AND DISCUSSION

In this section, we will discuss about the result of our experiment. We will discuss about the result of training and real world testing.

A. Training Result

In this section, we will discuss about the result of our training. In order to compare the performance of YOLOv8n model, we also train YOLOv7-tiny model.



Fig. 8. Front View



Fig. 9. Side View

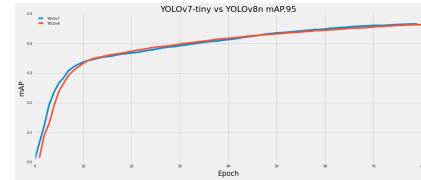


Fig. 10. Comparison of mAP50-95 between YOLOv8n and YOLOv7-tiny

B. Real World Testing Result

Our test result shows that our system can detect traffic object in real world environment. As mentioned earlier, we use car charger to power the system. While the system is running, we also use the other port on the same car cahrgers to charge our smartphone. This results in different power available for our system. As by doing this we can simulate the real user scenario. In user scenario, the user might use the other port on the car charger to charge their smartphone. As many car only have one port for car charger and no other USB power inside the car. As a result, our system can run in 3 different power consumption. 5 Watt, 7 Watt, and 10 Watt.

$$\text{FramePerSecond} = \frac{1000}{\text{latency}} \quad (4)$$

We also found that Motorcycle along with its rider is the most difficult traffic object to detect. While some random red or green lights can be detected as traffic light. On the other hand, our system can detect Car 90% of the time.

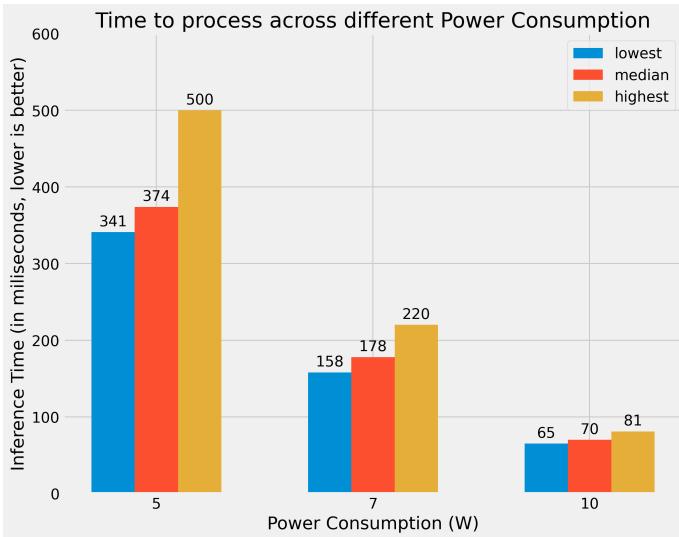


Fig. 11. Inference Time Comparison across different Power Consumption, Power Consumption error range is within 1 Watt

In terms of inference time, we gather the data by running the system in 3 different power consumption. 5 Watt, 7 Watt, and 10 Watt. We record the inference time data in miliseconds. Many research use FPS (Frame Per Second) as their inference process metric. We can convert inference time into FPS by using Eq (4)

Based on Eq (4) and Fig.11, we can conclude that FPS and inference time is inversely proportional. As the inference time increase, the FPS decrease. If we want to get smoother video output, we need to increase the FPS. Obviously, we need to reduce the inference time to increase the FPS.

V. CONCLUSION

Dashcam has became a popular device for drivers to record the road condition. However, the dashcam video is not only used for recording the road condition, but also used for other purposes, such as the insurance claim. In this paper, we proposed a dashcam system to detect traffic object. The object detection system is based on the YOLOv8n model. We also proposed a dataset for dashcam traffic object detection. Our dataset created by filtering MS COCO Dataset. The dataset contains 78000 images with 12 traffic objects class. By filtering dataset, we can reduce the training time and increase the model performance. We also propose a dashcam system that can be used for real world testing. The system is built using Jetson Nano 4GB as the edge computer.

Based on the Real World Testing, we can conclude that our system can detect traffic object in real world environment. Inference Time shows different result across different power consumption. Our system power draw is dependendt on the electrical power available in the vehicle. We managed to get the result of inference time in 3 different power consumption, 5,7 and 10 Watt. 10 Watt is the recommended power consumption for Jetson Nano 4GB. According to Nvidia, Jetson Nano 4GB can run in with MAXN mode. With this power

consumption, our system can do the inference in under 90 miliseconds for image size 640x480. While in lower power consumption, 7 Watt, our system can do the inference in around 150 - 220 miliseconds for image size 640x480. In 5 Watt power consumption, our system can do the inference in around 340 - 500 miliseconds for image size 640x480. If we convert processing time into Frame Per Second, we managed to get 2 to 15 FPS.

Although using jetson nano was sufficient to run the system. Better performance can be achieved by using more powerful edge computer. Such as using newer hardware from Nvidia Jetson Series. Further optimization can be done by making use of full potential of GPU in Nvidia Jetson Series. Rewriting or running inference in C++ might improve the performance of the system. Further development towards Autonomous Driving Assistance System (ADAS) can be done by adding more camera and sensor.

ACKNOWLEDGMENT

This research is supported by Telkom University, Indonesia. We would like to personally thank M Rayhan Aryana for his help in this research. Especially in the process of gathering the dataset and testing the system in real world environment. We also would like to thank QEngineering for open sourcing their Jetson Nano Ubuntu 20.04 image. With their Operating System image that was preinstalled with OpenCV and Pytorch, compiled with CUDA support, we can save a lot of time in setting up the environment.

REFERENCES

Please number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [?]. Refer simply to the reference number, as in [4]—do not use “Ref. [4]” or “reference [4]” except at the beginning of a sentence: “Reference [4] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [?]. Papers that have been accepted for publication should be cited as “in press” [?]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [?].

REFERENCES

- [1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.
- [2] Kim, J., Park, S., & Lee, U. (2020). Dashcam Witness: Video Sharing Motives and Privacy Concerns Across Different Nations. IEEE Access, 8, 110425–110437. doi:10.1109/access.2020.3002079
- [3] <https://arxiv.org/pdf/1905.05055.pdf>

- [4] Krizhevsky, A., Sutskever, I., Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84–90. <https://doi.org/10.1145/3065386>
- [5] <https://arxiv.org/abs/1311.2524>
- [6] <https://arxiv.org/abs/1506.02640>
- [7] <https://arxiv.org/pdf/1405.0312>
- [8] <https://link.springer.com/content/pdf/10.1007/s11042-021-11480-0.pdf>
- [9] <https://online-journals.org/index.php/i-jim/article/view/11530/6449>

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove the template text from your paper may result in your paper not being published.