

# Vehicle-detection-based traffic density estimation at road intersections

Huu-Huy Ngo

**Abstract**—Traffic congestion is currently affecting people's lives, which is a big and very urgent problem that needs to be solved. Therefore, this study will present a model to estimate traffic density at intersections based on vehicle detection. This study will determine the traffic density at intersections and the times with the highest vehicle. The findings from this study will assist in proposing solutions to optimize traffic flow and reduce traffic congestion. This study created a labeled database (Veh5) which includes 2,500 images of 5 common types of road vehicles. First, the system extracts consecutive frames from the input video to perform processing on those separate frames. Next, the YOLOv8 model is used to detect objects on each frame. This model was trained on Veh5 dataset with high accuracy, reaching mAP50 and mAP50-95 values of 0.994 and 0.915 respectively. Additionally, this study also assesses traffic density at some intersections in Thai Nguyen province, Vietnam, through the C-ThaiNguyen app. Experimental results demonstrate the effectiveness of this model.

**Keywords**—Artificial intelligence, Vehicle detection, Traffic density, Video surveillance, YOLOv8.

## I. INTRODUCTION

Traffic congestion is currently a major problem affecting people's lives, especially in big cities. Traffic congestion can happen during peak hours, during holidays or due to unexpected events such as traffic accidents, bad weather, or lack of traffic infrastructure. Traffic congestion has been causing many negative consequences for people's lives. Traffic congestion makes the movement of people more difficult. Travel time is long, costly to transport and causes

discomfort for drivers and passengers. Traffic congestion also causes many traffic accidents and increases the risk of accidents. The situation of vehicles jostling, running in the opposite direction, illegally parking, running red lights puts drivers and pedestrians at risk.

In the 4.0 technology era, the development of artificial intelligence, especially the development of machine learning and deep learning, is playing a very important role in life and production. Object detection is one of the most popular and fascinating research topics in the field of computer vision because it can be widely applied in various applications, such as human-machine interaction, computer vision systems, and computer vision systems. surveillance and security systems, video surveillance and self-driving cars, intelligent transportation systems [1]. Therefore, the importance of accuracy when detecting object has encouraged many researchers to propose several methods based on deep learning in recent years, e.g. Faster R-CNN [2], SSD [3], RetinaNet [4], Mask R-CNN [5] and YOLOv8 [6].

Researchers are actively studying deep neural networks and have applied deep learning models extensively in the field of computer vision, especially in object detection. Compared with traditional object classification methods, deep learning offers many significant advantages. Specifically, this method can extract features directly from the data, which helps to ensure the accuracy of deep learning-based methods for problems in the field of computer vision [7]. Therefore, artificial intelligence technology has been applied in the transportation field to

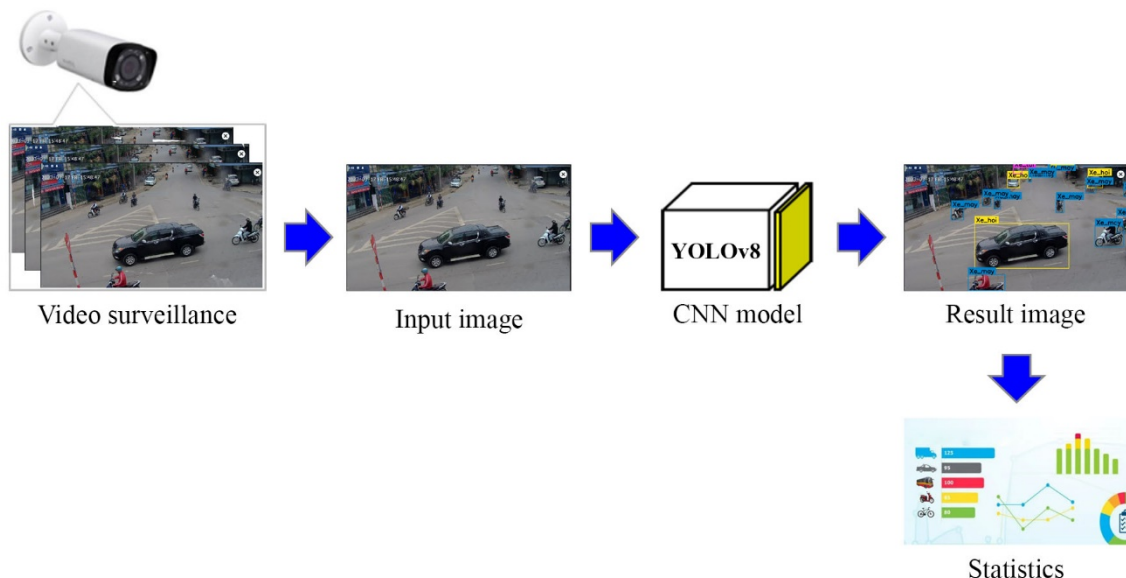


Figure 1. System overview

reduce traffic congestion. This is a promising topic and attracts a lot of research [8- 13]. Biswas et al. [8] presented a method to estimate traffic density using Single Shot Detection (SSD) and MobileNet-SSD models. In particular, SSDs are capable of handling different shapes, sizes, and viewing angles of objects. MobileNet-SSD is a cross-trained model from SSD to MobileNet architecture, thus faster than SSD. Yeshwanth et al. [9] presented a method to estimate traffic density based on convolutional neural network. First, they determined the Region-of-interest (ROI) of the vehicular road. They then estimated the traffic density in the region of interest using a combination of background subtraction and deep learning methods. Prasad et al. [11] has proposed an efficient method to estimate the traffic density on the intersection using image processing techniques and machine learning methods in real time. They combine a Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and Support Vector Machine (SVM) for estimating traffic density. Zhu et al. [12] presented a solution for estimating urban traffic density, using deep learning techniques to process ultra-high resolution traffic videos captured from Airplanes. Unmanned Aerial Vehicle (UAV). This urban traffic analytics solution includes deep neural network-based vehicle detection and localization, and vehicle tracking and counting over time.

Therefore, this study proposes a system to determine traffic density at intersections based on vehicle detection, using deep learning method. This system will determine the

traffic density at intersections and the times with the highest vehicle traffic. The findings from this study will assist in proposing solutions to optimize traffic flow and reduce traffic congestion. The proposed system is capable of detecting 5 common types of road vehicles in Vietnam and from there giving information about the number of vehicles and the density of vehicles at the intersections.

The remainder of this paper is organized as follows: Section II introduces the model for determining traffic density at road intersections. Details of the process of data collection and labeling are presented in Section III. Section IV presents experimental results. Finally, Section V summarizes conclusions and future research directions.

II. THE PROPOSED TRAFFIC DENSITY ESTIMATION MODEL

A. System Structure

The overview of the proposed system is shown in Figure 1. The input data are surveillance videos at road intersections. This study will detect 5 common types of road vehicles in Vietnam, including: bicycle, motorcycle, car, bus and truck. First, the system extracts consecutive frames from the input video to perform processing on those separate frames. The sampling rate is the number of frames per second (fps) and is custom set. Next, a convolutional neural network (CNN) model is used to detect objects on each frame. The YOLOv8 model is used for system implementation and development. This model is one of the most advanced and highly accurate

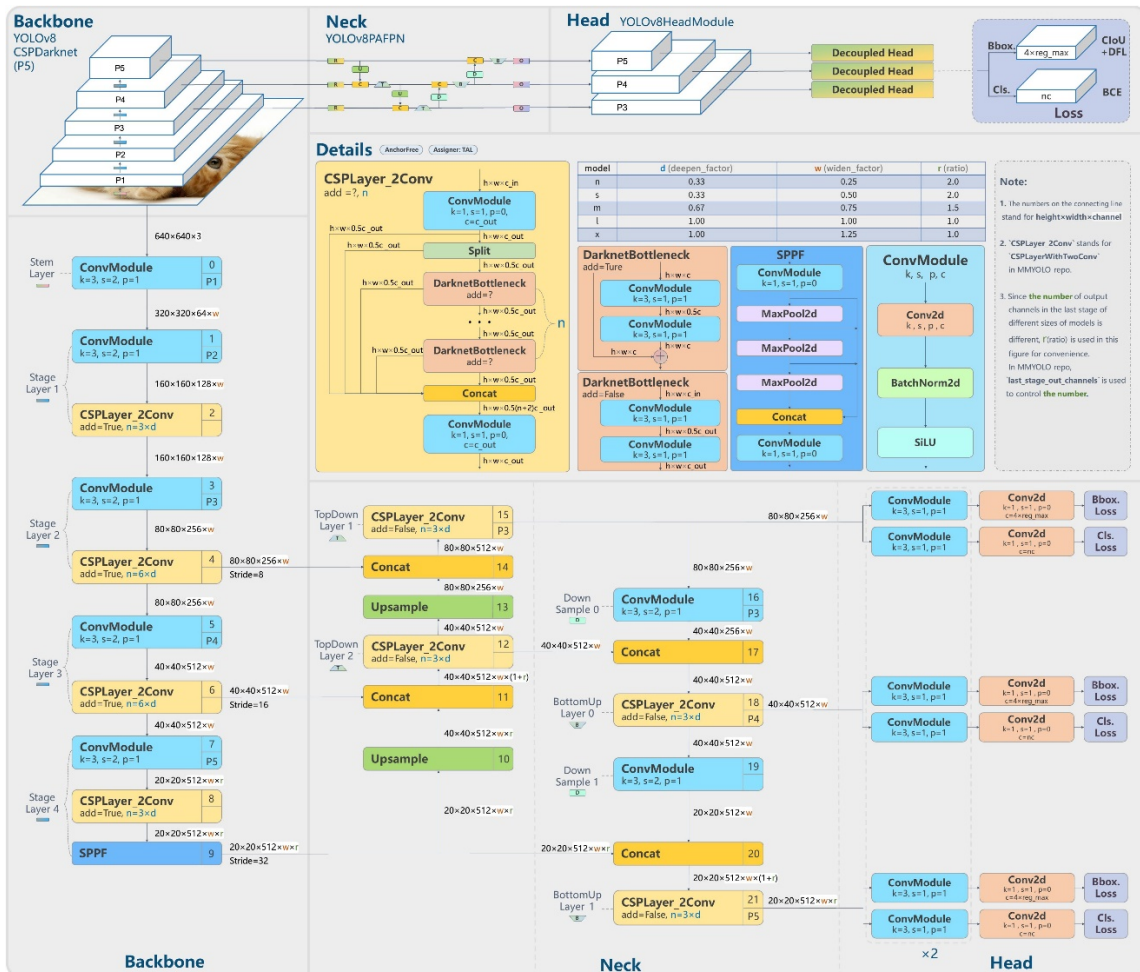


Figure 2. The structure of YOLOv8 model [15]

object detection models. The newly trained YOLOv8 model can detect the above 5 types of road vehicles. Finally, the resulting image is an image that identifies the location of vehicles by bounding boxes. Consequently, the system can calculate the number of each type of vehicle and the total number of traffic at the intersection.

### B. Model YOLOv8

In this study, the YOLOv8 model [6] is used to build the system. The YOLOv8 model was developed by the research group Ultralytics – founded by Glenn Jocher [14]. The YOLOv8 model is a state-of-the-art model that builds on the success of previous YOLO versions. At the same time, this model introduces new features and enhancements to increase performance and flexibility even further. YOLOv8 is designed to be fast, accurate, and easy to use, making it an excellent choice for a variety of object detection, image segmentation, and image classification tasks [6]. The structure of the YOLOv8 model is shown in Figure 2 [15]. The YOLOv8 model has several different versions, such as: YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l and YOLOv8x.

In the system building process, the first phase is to train the model. In this phase, we collect and label data to create a database for training and testing the model. This phase will be detailed in Section III. Next, the YOLOv8n model is selected and trained for system development. Phase 2 is training, testing and application development. This phase will be detailed in Section IV.

## III. DATASET

### A. Data Labeling Tool (Make Sense)

In the process of creating the database, first, the data collected in raw data is the image of 5 types of vehicles that need to be detected, as mentioned above. These images are then labeled with data to provide information so that the development model can learn from that data. In this study, the Make Sense tool [16] was used to label the images.

The Make Sense tool is a free online tool for image labeling. This tool runs on the web platform, so we don't

need to install it, we just need to visit the website of this tool [16] to use it directly. Therefore, this tool can run on different operating systems. So Make Sense tool is a very good choice for small computer vision deep learning projects, making data set preparation much easier and faster.

Figure 3 shows the working interface on the Make Sense tool. This tool allows to annotate images in the style of polygon, rectangle, line and point. In this study, the problem to be solved is that object detection should annotate images in a rectangular fashion. After labeling images, this tool can save image annotations in many different formats, such as: YOLO, VOC XML, CSV, VGG JSON, COCO JSON. Depending on the type of image annotation and the program code, we can save these image captions in the appropriate file format standard.

### B. Dataset Collection

*Dataset of 5 types of road vehicles (Named Veh5):* This database includes 2,500 images of 5 common types of road vehicles in Vietnam, including: bicycle, motorcycle, car, bus and truck. In which, the number of images in each object class is equal to 500 images. This dataset is divided into two parts, the training dataset and the testing dataset, with the proportions of 70% and 30% respectively. This dataset is aggregated from a number of different data sources, including: Poribohon-BD dataset [17], Car dataset [18] and Bus dataset [19]. Detailed information about feature classes in the database is shown in Table 1.

*C-ThaiNguyen app:* After the model is trained, testing and evaluation will be implemented. This study will estimate the traffic density at the intersections allowed at C-ThaiNguyen app [20]. Figure 4 shows the main interface on C-ThaiNguyen app. In particular, the "camera online" function allows direct observation of the traffic situation at 15 intersections in Thai Nguyen province, Vietnam, including: "Cau Gia Bay; Cho Thai; Cong Tam Quan; Duong tron Dong Quang; Duong tron Gang Thiep; Duong tron Tan Long; Duong tron Trung Tam; Luong Ngoc Quyen-LTV; Minh Cau-HVT; Minh Cau Hoang Ngan; Nga ba Mo Bach; Nga ba Nong Lam-DTM; Quang Trung-Viet Bac; Quang Trung Z115; Quang trung Vo Nguyen Giap" (*Names of*



Figure 3. Make Sense tool

**Table 1.** Details of experimental datasets

No.	Object Class ID	Object Class Name	Sources	Number of Original Images	Number of Used Images
1	Xe_dap	Bicycle	Poribohon-BD dataset: Folder “Bicycle” [17]	707	500
2	Xe_may	Motorcycle	Poribohon-BD dataset: Folder “Bike” [17]	864	500
3	Xe_hoi	Car	Car dataset: Folder “cars_train” [18]	8,144	500
4	Xe_buyt	Bus	Poribohon-BD dataset: Folder “Bus” [17]	452	250
			Bus dataset: Folder “train” [19]	436	250
					500 (Sum)
5	Xe_tai	Truck	Poribohon-BD dataset: Folder “Truck” [17]	730	500

**Figure 4.** The interface of C-ThaiNguyen app

intersection points are listed on the C-ThaiNguyen app). During the conduct of this study, videos were recorded for processing, analysis and statistical systems.

#### IV. EXPERIMENTAL RESULTS

##### A. Model Training Evaluation

**Hardware device:** The proposed system is implemented on a computer with 3.10 GHz core i5 CPU, 64-bit Windows operating system, 16 GB RAM, GPU (NVIDIA GeForce RTX 3050 Laptop GPU). In addition, the online camera system is used from the C-ThaiNguyen app.

After collecting the database, we will proceed to train the YOLOv8n network model. This model training task is very important because it directly affects the quality of the proposed system. In this section, the mean average precision (mAP) is used to evaluate the model quality. This metric is

one of the most popular metrics used to evaluate the performance of models in object detection, and is defined in Equation (1). Where,  $AP_i$  is the average accuracy at class  $i$ . Figure 5 shows the accuracy of the model during training. Accuracy is improved after each epoch. At the end of the 20th epoch, the accuracy of the model reached a high level, with mAP50 and mAP50-95 of 0.994 and 0.915 respectively.

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

##### B. Model Testing Evaluation

As mentioned above, the dataset Veh5 includes 2,500 images of 5 common types of road vehicles in Vietnam. This dataset is divided into two parts, the training dataset and the testing dataset, with the proportions of 70% and 30% respectively. Therefore, the number in the testing

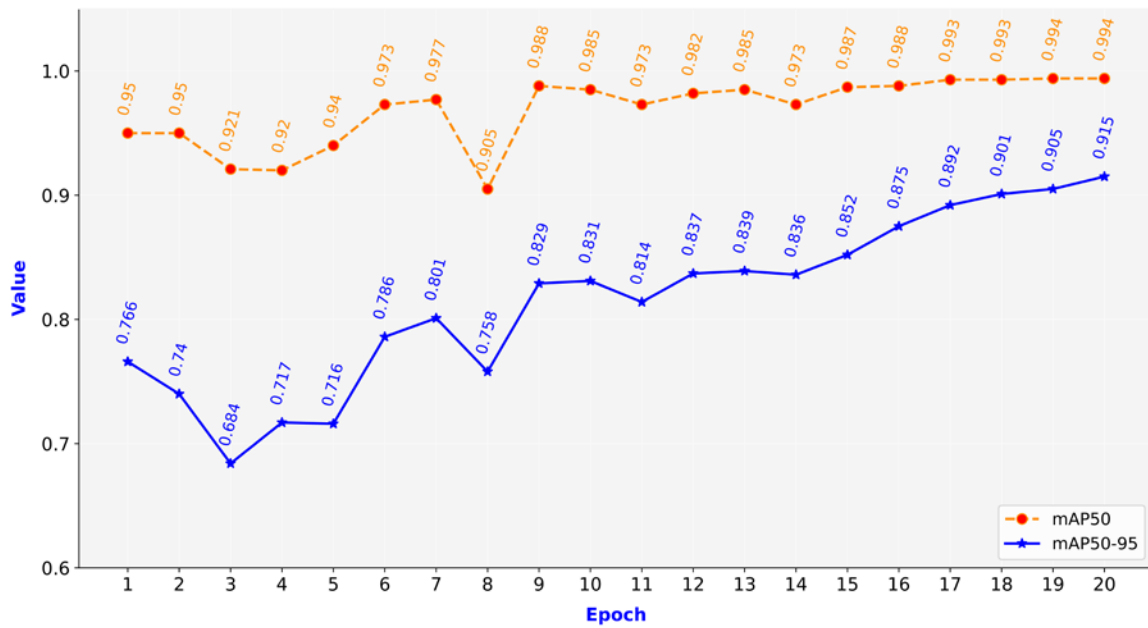


Figure 5. Results of model training

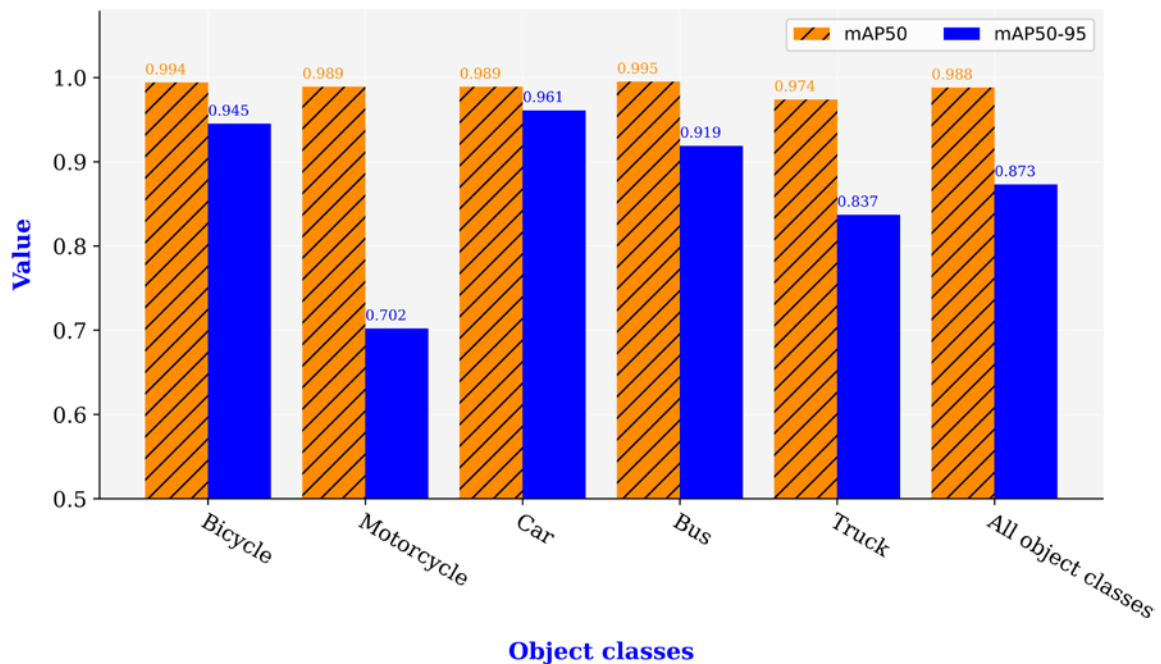


Figure 6. Results of model testing

dataset is 750 images. In this section, we perform model evaluation on the testing dataset. Figure 6 shows the model test results. The mAP50 value in the object classes is very high and approximately equal. Thus, the mAP50 value of the average of all object classes is also very good at 0.988. Meanwhile, the mAP50-95 value in the object classes is different and also high. The car class with the highest mAP50-95 value was 0.961. In contrast, the motorcycle class had the lowest mAP50-95 value of 0.702. However, this value is still acceptable. Therefore, the mAP50-95 value of the average of all classes achieved is 0.873. Figure 7 illustrates some of the object detection results of model testing. These findings are all very accurate.

### C. Comparison of Traffic Density at several Road Intersections

In this section, we evaluate and compare traffic density results at several intersections is allowed at C-ThaiNguyen app. In particular, the "camera online" function allows to directly observe the traffic situation at 15 intersection points in Thai Nguyen province, Vietnam. However, during the data collection process, 12 intersections have been identified with video signal, while there are 3 intersections without video signal, including: "Luong Ngoc Quyen-LTV; Minh Cau-HVT; Quang trung Vo Nguyen Giap" (Names of intersection points are listed on the C-ThaiNguyen app). Therefore, in the experimental program, this study recorded videos of 12 intersections for 1 hour (From 15:00 to 16:00 on March 17, 2023), and then performed the extraction of

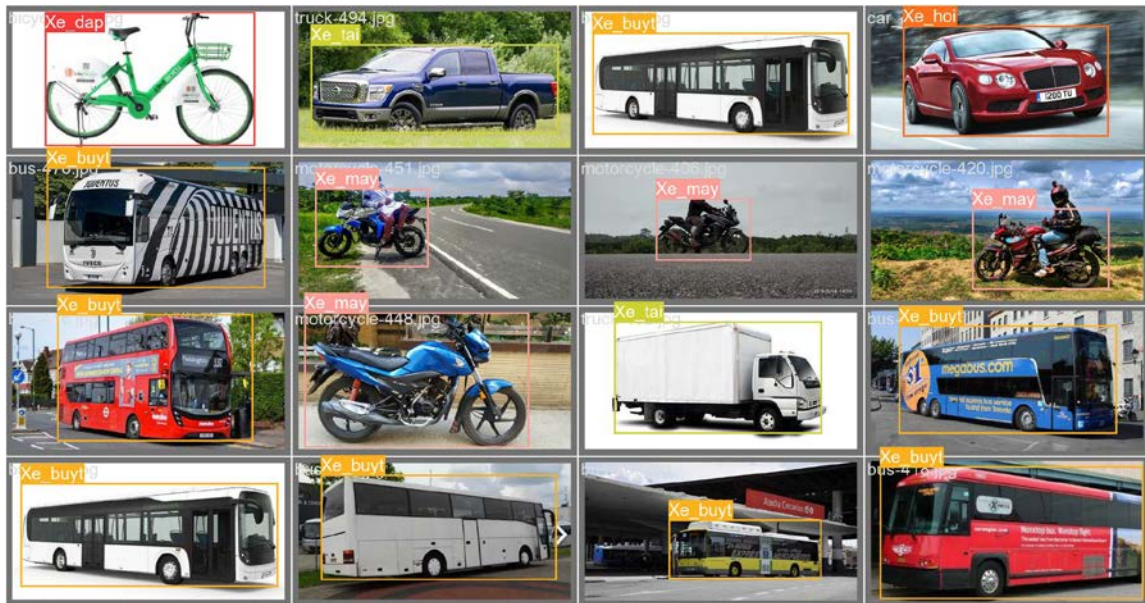


Figure 7. Results of object detection during model testing

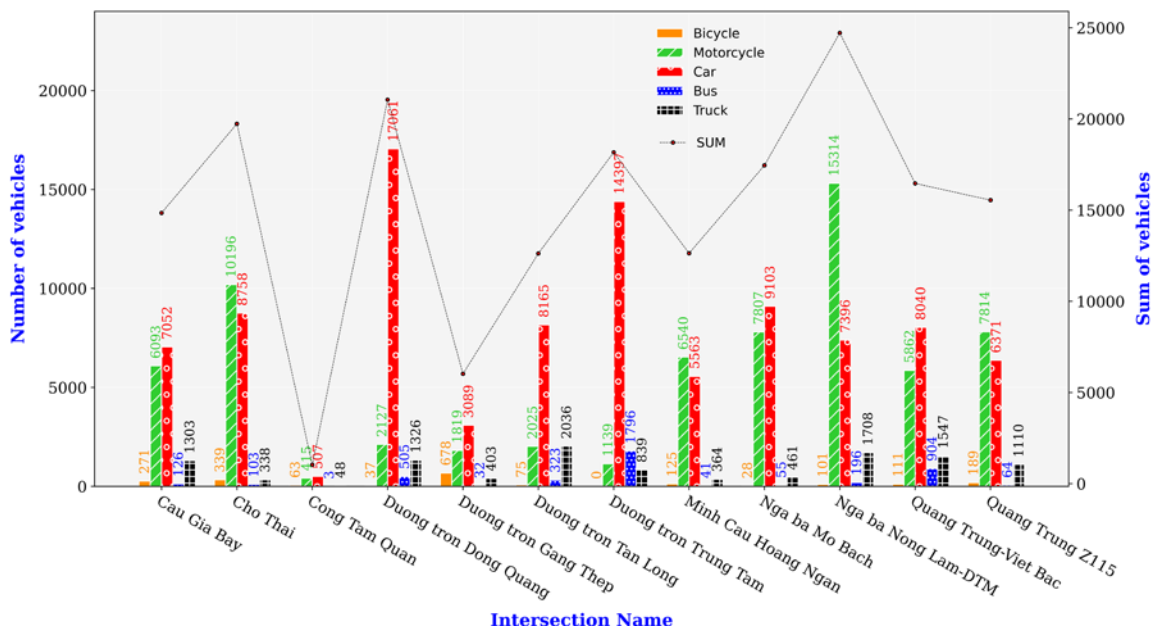


Figure 8. Traffic density at several intersections

frames for analysis, with a sampling rate of 4 s/frame.

Figure 8 shows traffic density results at intersections. From this figure, we can see that there are some intersections with a large number of vehicles such as “Nga ba Nong Lam-DTM, Duong tron Dong Quang and Cho Thai”. Meanwhile, some other intersections have a very small number of vehicles and low traffic density such as “Cong Tam Quan and Duong tron Gang Thiep”. During the system testing period, cars and motorcycles were the two most popular vehicles, especially at “Duong tron Dong Quang and Duong tron Trung tam”, with the number of cars being 17,061 and 14,397 respectively. Meanwhile, bicycle is the least vehicle out of 5 vehicles evaluated. In addition, Figure 9 also illustrates the results of vehicle detection at intersections with high accuracy.

#### D. Analysis of Traffic Density at different Times

In this section, we analyze the results of traffic density over

time at the intersection point "Quang Trung Z115". Video is still collected through the "camera online" function of C-ThaiNguyen app. The analysis time was from 07:00 to 21:00 on March 30, 2023, with a sampling rate of 4 s/frame. Figure 10 shows traffic density results over time at the intersection “Quang Trung Z115”. It is easy to see that the period with the greatest number of vehicles (i.e. the highest traffic density) is from 17:00 to 18:00, for a total number of vehicles of 16,652. Next is the period from 16:00 to 17:00 and from 07:00 to 08:00, with a total number of vehicles of 15,092 and 12,033 respectively. These are all peak hours when people go to work and get out of work. Meanwhile, in the evening, the number of traffic is significantly reduced, i.e. the traffic density is less. The time from 19:00 to 20:00 and from 20:00 to 21:00 only has a total number of vehicles of 7,782 and 7,810 respectively.



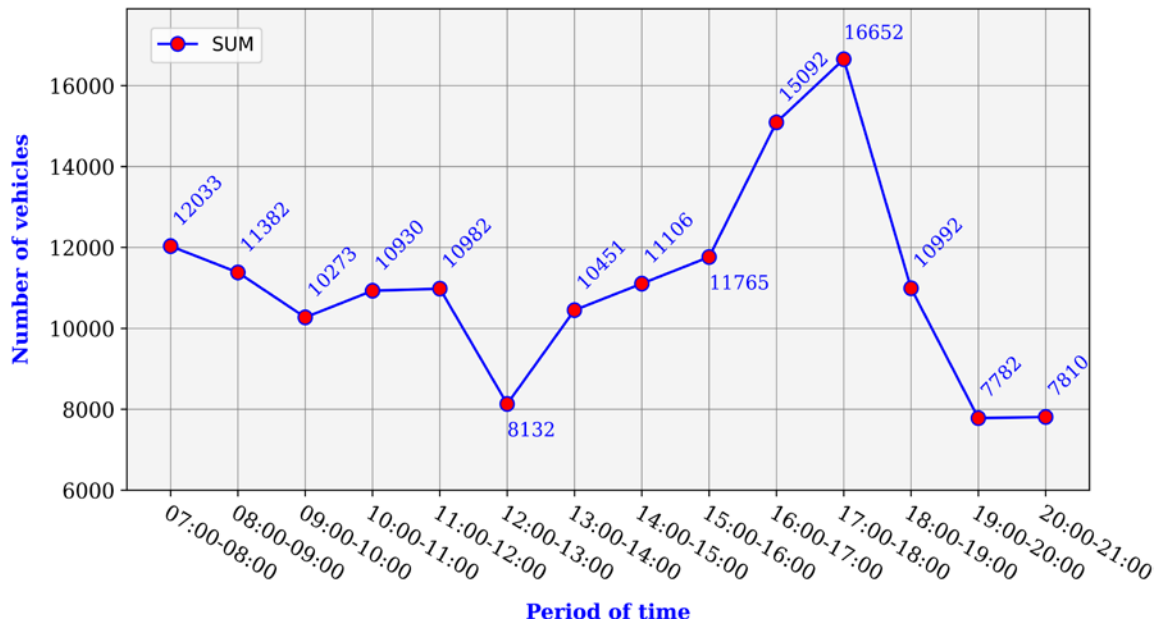
Figure 9. Results of vehicle detection at several intersections

## V. CONCLUSIONS

This study presented a model to determine traffic density at intersections based on vehicle detection, using deep learning with YOLOv8 model. This study created a dataset (Veh5) of 2,500 images of 5 common types of road vehicles and labeled these data. The YOLOv8n model structure was used to develop the system. This model was trained on Veh5 dataset with high accuracy, reaching mAP50 and mAP50-95 values of 0.994 and 0.915 respectively. Besides, this study also analyzed, compared and evaluated the traffic density at several intersections in Thai Nguyen province, Vietnam, through the C-ThaiNguyen app. In the future, we plan to continue to improve the quality of the proposed model and integrate the system to detect more other vehicles. On the other hand, we intend to use the object tracking method to improve the traffic density estimation.

## REFERENCES

- [1] H. H. Ngo, F. C. Lin, Y. T. Sehn, M. Tu and C. R. Dow, "A Room Monitoring System Using Deep Learning and Perspective Correction Techniques," *Applied Sciences*, vol. 10, no. 13, pp. 1–22, Jul. 2020.
- [2] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection With Region Proposal Networks," in *Proceedings of Advances in Neural Information Processing Systems* 28, Montreal, Canada, pp. 91–99, Dec. 2015.
- [3] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu and A. C. Berg, "SSD: Single Shot MultiBox Detector," *arXiv:1512.02325*, pp. 1–17, Dec. 2016.
- [4] T. Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, "Focal Loss for Dense Object Detection," in *Proceedings of IEEE International Conference on Computer Vision*, Venice, Italy, pp. 2980–2988, Oct. 2017.
- [5] K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN," *arXiv:1703.06870*, pp. 1–12, Jan. 2018.
- [6] Ultralytics YOLOv8, Ultralytics GitHub, <https://github.com/ultralytics/ultralytics>. [20-Feb-2023].
- [7] C.-R. Dow, H.-H. Ngo, L.-H. Lee, P.-Y. Lai, K.-C. Wang and V.-T. Bui, "A Crosswalk Pedestrian Recognition System by Using Deep Learning and Zebra Crossing Recognition," *Practice and Experience*, vol. 50, no. 5, pp. 630–644, Aug. 2019.
- [8] D. Biswas, H. Su, C. Wang, A. Stevanovic and W. Wang, "An Automatic Traffic Density Estimation Using Single Shot Detection (SSD) and Mobilenet-SSD," *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 110, pp. 176–184, Apr. 2019.
- [9] C. Yeshwanth, P. S. A. Sooraj, V. Sudhakaran and V. Raveendran, "Estimation of Intersection Traffic Density on Decentralized Architectures with Deep Networks," in *Proceedings of International Smart Cities Conference (ISC2)*, Wuxi, China, pp. 1–6, Sep. 2017.
- [10] M. A. Aljamal, H. M. Abdelghaffar and H. A. Rakha, "Estimation of Traffic Stream Density Using Connected Vehicle Data: Linear



**Figure 10.** Traffic density over time at the intersection “Quang Trung Z115”

and Nonlinear Filtering Approaches,” *Sensors*, vol. 20, no. 15, pp. 1–15, Jul. 2020.

- [11] D. Prasad, K. Kapadni, A. Gadpal, M. Visave and K. Sultanpure, “HOG, LBP and SVM based Traffic Density Estimation at Intersection,” in *Proceedings of IEEE Pune Section International Conference (PuneCon)*, Pune, India, pp. 1–5, Dec. 2019.
- [12] J. Zhu, K. Sun, S. Jia, Q. Li, X. Hou, W. Lin, B. Liu and G. Qiu, “Urban Traffic Density Estimation Based on Ultrahigh-Resolution UAV Video and Deep Neural Network,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 12, pp. 4968–4981, Dec. 2018.
- [13] J. S. Chang, J. Hwang and M. Choi, “Vehicle Detection Approach Adjusting Road Curves to Estimate Local Traffic Density under Real Driving Conditions,” *Transportation Research Record*, vol. 2677, no. 3, pp. 1382–1396, Sep. 2022.
- [14] Ultralytics: Revolutionizing the World of Vision AI, Ultralytics, <https://ultralytics.com/about>, [20-Feb-2023].
- [15] YOLOv8 Structure, MMYOLO GitHub, <https://github.com/open-mmlab/mmyolo>, [20-Feb-2023].
- [16] Make Sense, <https://www.makesense.ai/>, [14-Feb-2023].
- [17] Poribohon-BD, Mendeley Data, <https://data.mendeley.com/datasets/pwyyg8zmk5/>, [14-Feb-2023].
- [18] Cars Dataset, Mendeley Data, [http://ai.stanford.edu/~jkrause/cars/car\\_dataset.html](http://ai.stanford.edu/~jkrause/cars/car_dataset.html), [14-Feb-2023].
- [19] Bus Dataset, Images.cv, <https://images.cv/dataset/bus-image-classification-dataset>, [14-Feb-2023].
- [20] C-ThaiNguyen, Google Play, <https://play.google.com/store/apps/details?id=com.paht.thainguyen.smart.app&hl=vi&gl=US>, [14-Feb-2023].

**Huu-Huy Ngo** received his B.S. and M.S. degrees from Thai Nguyen University of Information and Communication Technology, Vietnam, in 2010 and 2012, respectively, and Ph.D. degrees in Information Engineering and Computer Science from Feng Chia University, Taiwan, in 2021. Currently, he is a lecturer at the Thai Nguyen University of Information and Communication Technology, Vietnam. His research interests include computer vision, deep learning, embedded system, neural networks, and object detection. Email: nhhuy@ictu.edu.vn.