Traffic Congestion Detection Using Fixed-Wing Unmanned Aerial Vehicle (UAV) Video Streaming Based on Deep Learning

Winahyu Utomo
Red Hat Cloud Consultant, RHCSA
Red Hat
California, USA
wutomo@redhat.com

Putu Wisnu Bhaskara
Dept. of Computer Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
wisnu16@mhs.te.its.ac.id

Arief Kurniawan
Dept. of Computer Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
arifku@ee.its.ac.id

Susi Juniastuti
Dept. of Computer Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
susi@te.its.ac.id

Eko Mulyanto Yuniarno
Dept. of Computer Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
ekomulyanto@ee.its.ac.id

***Abstract*—Population growth in the region has led to increased use of roads that causes traffic congestion. Traffic congestion also occurs during long weekend in outsides city. The roads in the area are usually smooth traffic flow, it becomes very congested. One method of the smart road monitoring system is a fixed camera sensor as input and artificial intelligent as the analysis. However, these methods require a great infrastructure on highways such as: power supply, protective CPU, good power supply and stable computer network connections. This will be difficult to fulfill if applied on roads outside the city. To overcome the problem, we propose a system of vehicles detection and road density classification using Fixed-Wing Unmanned Aerial Vehicle (UAV) video streaming. We chose Fixed-Wing UAV for its advantages: wide range and fast flight speed. The proposed system detects and classifies vehicles. Vehicles are detected and calcified using CNN's Deep Learning which uses the YOLO architecture. The level of traffic density is determined by the area of the road covered by vehicles to road area. We tested the proposed system using YouTube video and UAV video streaming. Both experimental scenarios have almost the same results. Precisions of recording video UAV and streaming video are: 90.75% and 90%, respectively.**

***Index Terms*—Congestion Detection, UAV, CNN, YOLO**

I. Introduction

,

Congestion is stagnating situation even stopped vehicles because of the large number of vehicles that exceed the capacity of the road. The condition causes: wasted time, increasing fuel usage, increasing pollution, affected the driver's mentality, and unproductive work. Traffic congestion caused by several reasons, there are: 1. Vehicle growth is not matched by the roads [1]. 2. Traffic congestion increases on intercity roads and tolls during long holidays. Some of the long holidays that cause congestion: religious holidays, school holidays and new year holidays. One effective way to avoid the traffic density is to use alternative roads.

To obtain alternative road information, road density information is required. Road density information is also used to control traffic [2]. The methods used to obtain information on road density: Vehicles number information using the fixed camera as a sensor. wifi and bluetooth signal information in a vehicle [3], Vehicles number information using the fixed camera as a sensor [4], [5]. However, the use of fixed cameras has limitations: they are static, having a small coverage, requires a complex infrastructure on the highway. The necessary infrastructure: power supply, protective CPU, voltage source and computer network connections. Due to the limited space to install cameras and the lack of computer network connections in suburban areas, it is necessary to have aerial data acquisition techniques in order to detect traffic congestion in large areas and temporary [6] [7].

UAV is a flying robot technology that is piloted using a remote controller or autonomous flying using an embedded computer system. The flying robot equipped with camera sensors and a wireless video streaming module for realtime monitoring [7]. UAV is divided into two types: Fixed Wing and Rotary Wing. Rotary Wing flying uses large propellers instead of wings, Fixed Wing fly using the wings as the giver of the main lift. In the study [4], the author proposed the detection and tracking of vehicles using rotary UAV types. Authors claimed methods provide monitoring solution in real time but the range of the rotary type UAV has limited flight ranges.

To overcome limited ranges, we propose a congestion detection that can provide realtime data with a wide coverage using UAV Fixed Wing. UAV sends video data to the control room to analyze the number of vehicles on a road. The CPU in the control room to calculate the number of vehicles and types in a segment of a road. Vehicle detection on the proposed method uses Convolution Neural Network (CNN) with You only Look Once (YOLO) architecture. The vehicles are classified into four based on the area of the vehicle detected by YOLO, namely: cars, motorbikes, trucks and buses. The density status is obtained from the calculation of road density with parameters: road width, length of road captured by the camera and the number of vehicles. The parameters of status density are: the width of roads, length of roads that captured camera and the number of vehicles.

The rest of this paper is organized as follows: Section 2, gives proposed methods, deep learning data set for training

978-1-7281-8283-4/20/$31.00 ©2020 IEEE

process and several assumptions to get the level of traffic congestion. In Section 3, shows the experiment results and discussion. Finally, Section 4, we conclude the proposed methods based on our experiment.

II. Methodology

The proposed traffic congestion detection using fixed wing UAV video streaming is illustrated in the figure 1, each block of the method will be explained in the subsection. Fixed Wing UAV transmits video of highway traffic to the control room. Vehicles are detected and classified into: motorbikes, cars, trucks and buses. In addition to detecting vehicles, the control room also analyzes the status of the road density based on the number of vehicles in each frame.

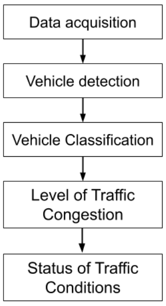

Fig. 1. Block diagram of the proposed system.

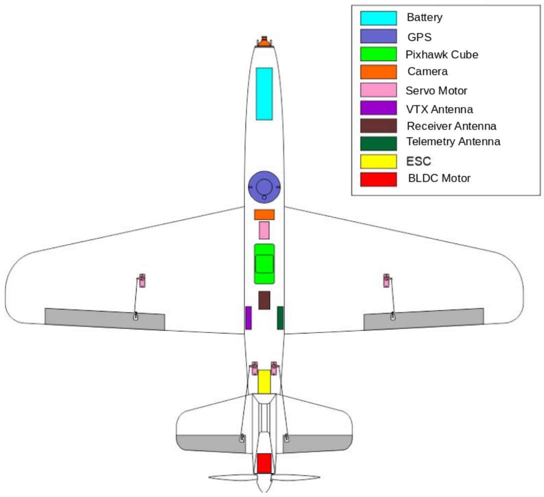

Fig. 2. Design of Fixed-Wing UAV.

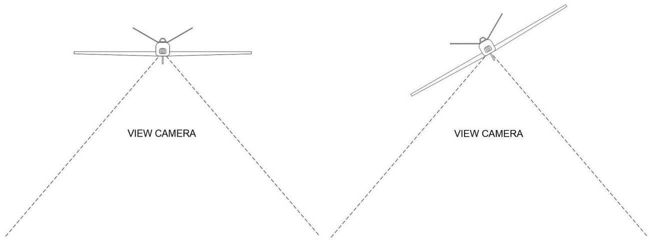

Fig. 3. The UAV camera is equipped with a stabilizer.

A. Data Acquisition

We propose Fixed Wing UAV, the design of the UAV and the layout of the UAV components as shown in figure 2. The motor and propeller drive are designed at the rear which aims to reduce vibration of the camera. The size of the UAV is 114.6 cm long, 130 cm wide and 10 cm high. The size accommodates all the required hardware, such as cameras, driving components, and autonomous systems. The shape of the tail is a Y-Tail which aims to increase the efficiency of the turn. The drive system of the UAV consists of: Propeller, Motor BLDC, Electronic Speed Controller (ESC), LIPO battery, and servo as driving control surface. The autonomous system uses the following modules: GPS GNSS, Microcontroller Pixhawk Cube based 32bit STM32F427 CortexM4F core with FPU. The GPS GNSS module has an update rate of 10 Hz, a precision level of 1 meter, a fix position time of 26 seconds on cold start, and is equipped with an internal compass. The proposed UAV flies according to the flight routes that have been planned using the Ardu Pilot program [8].

The camera used is Runcam 2, which has HD resolution 1080 60fps. The camera is equipped with a stabilizer uses 1-axis servo controlled by Pixhawk Cube. So that the captured of the UAV camera facing down becomes stable when the UAV moves turn, as shown in figure 3. The UAV flies at an altitude of 40 meters, a speed of 10 m/s or 36 Km/h. The first time flying, the aircraft collects a top-view vehicle image for training data set that consists: cars, motorbikes, buses, and trucks as depicted figure 4.

B. Detection and Classification

We use the YOLO architecture based on deep learning CNN for detection and classification [9]. Yolo divides the image input into an 13 $\times$ 13 grid which consists of seven convolution layers, six max-pooling layers to predict the target of vehicle, there are: motorcycles, cars, buses and trucks. If the centre of the object is located in a grid cell, the grid cell is responsible for detecting the object. Each grid cell predicts bounding box.Each bounding box contains five predictions: x, y, w, h and confidence score. The coordinates (x, y) represent the coordinates of the box relative to the grid cell, (w, h) represent length and width of the bounding box and confidence score shows the level of accuracy of the detected object in the bounding box. We use 6000 data sets of vehicle objects. The composition of the data sets is: 5400 as training data and 600 as validation data. The image of the vehicle is labeled into 4 classes: Car, Motorcycle, Bus and Truck as depicted figure 4. The training process requires 2000 iterations and uses the Adam optimizer.

C. Traffic Density

In this study, Traffic density (D) is the number of covered area by the vehicles on the road divided the area of the road that capture by the camera as follow equation 1. Length of vehicle bonding box is represented by L (meter), Width of

Fig. 4. Top View of the vehicles.

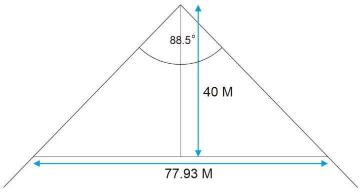

Fig. 5. Road Length Calculation.

vehicle bonding box is represented by W (meter), A_r is the variable of the Area of road in a single frame video. $L(i,j)$, $W(i,j)$ are the length and Width of the i^{th} object vehicle and j is the type of vehicles. There are four types of vehicles are: motorcycles, cars, trucks, buses that have values 0,1,2,3, respectively. The values of $L(i,j)$ and $W(i,j)$ are described by the equation 2 and 3.

$$D = \sum_{i=0}^{n} \frac{L(i,j) \times W(i,j)}{A_r} \tag{1}$$

$$L(i,j) = \begin{cases} 2 \text{ m}; \forall i, \text{ if } j = 0 \\ 4 \text{ m}; \forall i, \text{ if } j = 1 \\ 8.5 \text{ m}; \forall i, \text{ if } j = 2 \\ 12.5 \text{ m}; \forall i, \text{ if } j = 3 \end{cases} \tag{2}$$

$$W(i,j) = \begin{cases} 0.6 \text{ m}; \forall i, \text{ if } j = 0 \\ 1.75 \text{ m}; \forall i, \text{ if } j = 1 \\ 2.5 \text{ m}; \forall i, \text{ if } j = 2 \\ 2.5 \text{ m}; \forall i, \text{ if } j = 3 \end{cases} \tag{3}$$

$$Nmax_j = \left\lfloor \frac{A_r}{(L_j + (2 \times s_l)) \times (W_j + +(2 \times s_w))} \right\rfloor \tag{4}$$

The detected road area can be calculated the length multiplied by the width of the road caught on camera. We assume the UAV is flying at an altitude of 40 m and a camera angle of 88.5 is used. The road length calculation that the camera can capture is 77.93 meters according to figure 5. We assume that the road width is 7 meters and that each queue of vehicles has a minimum distance of one meter. The values of s_l and s_w used in the figure 6 are 0.5 meters where s_l, s_w is the value of half the distance between the vehicles to the front, rear and side. Based on the assumption of the width, length of a vehicle in equations 2, 3 and the distance between vehicles, the maximum number of one type vehicles on the road $Nmax_j$ that is captured by the camera is as follows equation 4. Figure 7 shows the simulation of maximum vehicles in a queue of one type of vehicle captured on camera. The maximum number of one type of vehicle $Nmax_j$ with a distance of 1.0 meters has according to table I.

TABLE I
THE MAXIMUM NUMBER ONE TYPE OF VEHICLE

j	Vehicle	Length(m)	Width(m)	Number of Vehicles
0	Motorcycle	2	0.6	156
1	Car	4	1.75	30
2	Truck	8.5	2.5	16
3	Bus	12.5	2.5	10

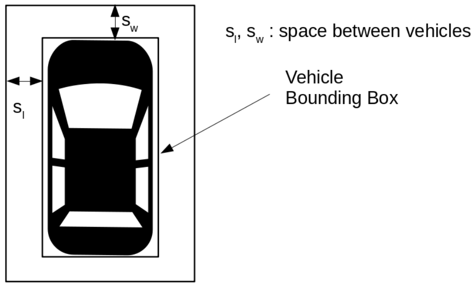

Fig. 6. Minimum space between vehicles when queuing occurred.

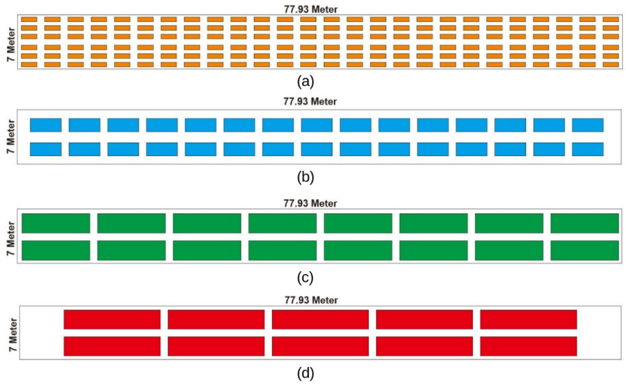

Fig. 7. Simulation of Maximum Number of (a) motorcycles, (b) cars, (c) trucks, (d) buses in a road.

We define the status of the density based on the level density D of roads captured by the AUV camera, namely: smooth, congested, and jammed. The status is based on the threshold limit based on equation 5. A road section is considered smooth if the traffic density D is less equal to T1, congested if the density D is between T1 and T2, Conversely, if the density is greater T2 then the status of road is jammed. The threshold value T1 is 25 % of the average number of vehicles captured by the camera on the road and T2 is 50 %. The value of T1, T2 are 0.12 and 0.36, respectively.

$$Status = \begin{cases} \text{Smooth; if } D \leqslant T_1 \\ \text{Congested; if } T_1 < D \leqslant T2 \\ \text{Jammed; if } D > T_2 \end{cases} \tag{5}$$

III. Experiments and Analysis

This study proven using two kinds of experiments, there are: vehicle detection testing and status of road. Both of these experiments using two scenarios, namely: aerial video from YouTube (scenario 1) and real time detection using UAV video streaming (scenario 2) using the streaming proposed UAV.

A. *Vehicle Detection Test*

We investigate performance of our proposed vehicle detection using a predictive positive P according to the equation 6. Where TP stands for True Positive is the total number detecting of vehicle using our proposed system. Meanwhile, FP stands for False Positive is the total number of not vehicles but identified as the vehicles.

$$P = \frac{TP}{TP + FP} \times 100\% \tag{6}$$

Vehicle detection test uses live streaming of Fixed Wing UAV as shown in figure 8. In one frame of video, 2 cars and 2 vehicles were detected, so the figure has a predictive positive of 100%. Experiments using live streaming UAV, our system only detects motorcycles and cars. Because the UAV flying around campus of ITS Surabaya. At that time, there were no buses and trucks. As a result, the precision for detecting the motorbike is 100% and the precision for detecting the car is 80%. The overall results of vehicle detection testing using scenario 1 and scenario 2 are shown in table II.

Fig. 8. Vehicle Detection Test.

TABLE II
PRECISION DETECTION USING SCENARIO 1 AND SCENARIO 2.

Scenario	Vehicle	Precision	Average
YouTube Video	Motorcycle	74%	90.75%
	Car	97%	
	Truck	92%	
	Bus	100%	
UAV Streaming	Motorcycle	100%	90%
	Car	80%	

In the proposed vehicle detection test, the average precision has almost the same results between scenario 1 and scenario 2. Scenario 1 has an average precision of 90.75%, while in scenario 2, the real-time detection results have an average precision of 90%.

B. Road Density Test

Our proposed system for determining density status in real time is based on the number of vehicles on the road in one video frame. Road Density level and status test uses scenario 1 as shown in figure 9. Results of road density test in scenario 1 and scenario 2 shown in the table III. The road status in the table is obtained from calculating the density level using equations 1, 2, 3, and 5.

Fig. 9. Road Density Level and Status Test using YouTube Video.

TABLE III
EXPERIMENTS OF DENSITY LEVEL.

Video Source	Vehicle	Total	Density level	Status
Youtube Video #1	Motorcycle	0	23.32%	Congested
	Car	11		
	Truck	1		
	Bus	1		
UAV Streaming #1	Motorcycle	1	2.71%	Smooth
	Car	2		
	Truck	0		
	Bus	0		
UAV Streaming #2	Motorcycle	1	3.96%	Smooth
	Car	3		
	Truck	0		
	Bus	0		

IV. CONCLUSION

This study proposed a Traffic Congestion Detection Using Fixed-Wing UAV base on YOLO. The method is suitable when implemented for traffic monitoring on the roads outside city. Since Fixed-Wing UAV that has a flying far and fast. The traffic density is detected via video streaming UAV based on YOLO deep learning. YOLO is deep learning algorithm based on CNN. The output of the detection and classification is used to update the status of road density. We use two experimental scenarios to detect and classify, there are: public video recording (YouTube video) and real time Fixed-Wing UAV video streaming. Based on the test results, the system can detect vehicles and road density in both scenarios. There are four types of vehicles are detected and classified, namely: cars, motorcycles, buses, and trucks. Some vehicles that move horizontally against the camera capture which is not detected by the system, this is because the data set used only contains vertical vehicle images. The experiment is to use public video recording and streaming UAV video have nearly same result precision, there are 90.75% and 90%, respectively. Based on road density tests, the level of road density can be used as a basis for determining the road density status.

ACKNOWLEDGMENT

This research was supported by Computer Engineering Department, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember (ITS).

REFERENCES

[1] S. Maji, “Traffic congestion and possible solutions a case study of asansol,” 2017, pp. 42–46.
[2] N. V. Hung, L. C. Tran, N. H. Dung, T. M. Hoang, and N. T. Dzung, “A traffic monitoring system for a mixed traffic flow via road estimation and analysis.” IEEE, 2016, pp. 375–378.

[3] T. Tsubota and T. Yoshii, “An analysis of the detection probability of mac address from a moving bluetooth device,” 2017.
[4] L. Wang, F. L. Chen, and H. M. Yin, “Detecting and tracking vehicles in traffic by unmanned aerial vehicles,” 2016, pp. 294–308.
[5] S. B. Prakoso, “Pengembangan sistem penghitung jumlah kendaraan memanfaatkan citra aerial yang diambil dengan pesawat tanpa awak.” Institut Teknologi Sepuluh Nopember, 2018.
[6] A. Alkaabi, “Applications of unmanned aerial vehicle (uav) technology for research and education in uae,” 2017, pp. 4–11.
[7] R. Yanushevsky, “Guidance of unmanned aerial vehicle,” 2011, pp. 1–5.
[8] M. Oborne, “Mission planner,” https://ardupilot.org/planner, 2019.
[9] R. G. A. F. Joseph Redmon, Santosh Divvala, “You only look once: Unied,real-time object detection,” 2016.