Vehicle Detection and Car Type Identification System using Deep Learning and Unmanned Aerial Vehicle

Chang Jin Seo

Abstract*: *Background/Objectives: UAV (unmanned aerial vehicle) based traffic measurement system has various advantages than the traditional traffic monitoring systems using fixed loop sensors. This paper proposes the designing and implementing method of vehicle detection and type identification using Deep Learning and UAV in UHD (ultra-high definition) 4K video images. Methods/Statistical analysis: The present study proposes the implementation method of detection and classification system that can be accomplished vehicle classification according to AUSTROADS's plan. The proposed system has designed two primary processes: detection and classification. This paper introduces the method that vehicle data training, detection, and classification method by applying a Darknet-53 for vehicles found in UHD images. Also, we propose the variable classification method due to parked and stopped cars for traffic flow monitoring. So, we considered the three conditions of driving, stopping, and parking. Findings: The results of the experiment show that the proposed approach resulted in errors that were twice as low as conventional methods that are using a fixed search area. Improvements/Applications: The proposed study can be applied to traffic flow monitoring system, ITS (intelligent transport system), vehicle detection and classification system.

Index Terms*: *Deep Learning, UAV, Car Type Identification, Object Detection, YOLOv3, ITS

I. INTRODUCTION

Vehicle detection and type-identification systems through UAVs with advantages in time, space and cost have recently gained attention for their active research areas in establishing traffic flow monitoring systems. In recent years, the performance improvements of UAVs have been enabled the recording of ultra-high-resolution aerial images at low cost, which can cover a wide area. Also, a variety of studies are being undertaken on measuring traffic volume using UAVs with various advantages. A Review of several recent studies shows that acquiring aerial images using satellites, airplanes, and helicopters are challenging to respond in real time to changes in weather and time and have a costly problem compared to UAVs [1,2]. In current traffic flow monitoring systems, traffic volume data are collected by fixed sensing systems, such as wireless sensors, ground loop sensors, and CCTV cameras. Such fixed sensor systems have some advantages to use for a long time once installed, but they have high maintenance costs [3]. Aerial traffic volume data can be recorded by aerial images from UAVs, airships, helicopters, aircraft, and satellites. Satellite and aircraft images are costly and hard to reflect weather and time changes on the fly [4,5]. However, airships and UAVs can provide UHD-4K aerial images with relatively low costs. The proposed method is consist of vehicle detection and type-identification method throughout deep learning based networks trained in UHD-4K aerial images. The composition of this paper is as follows. Section 2 describes faster R-CNN, deep learning, SSD, YOLOv3, and Kalman filter algorithms. Section 3, represents the vehicle detection and type identification using deep learning and traffic flow measurement method proposed in this article. In Section 4, the experiment results are presented for the proposed method. Finally, In Section 5, future work and conclusion are given.

II. RELATED WORK

A recent review of articles on vehicle detection and type-identification using UAV shows that a variety of approaches are underway to detect a variety of objects in aerial video.It is very important to know which object detection algorithm is selected in the system design of object detection because the performance of object detection algorithms greatly affects the performance of the object detection system [6]. In general, the detection speed and accuracy of object detection systems should be considered when designing object detection systems. The two performance indicators have closely correlated.Due to the complexity of detection algorithms, object detection accuracy is reduced if fast object detection is desired, an object detection speed is slowed if high object detection accuracy is desired.Therefore, the performance of the system depends on which object detection algorithm is selected, and the general object detection methods are divided into single frame analysis and multi-frame analysis techniques [7]. In this paper, we used single-frame features analysis for object detection for speed up. In recent, various object detection methods have been proposed to detect objects in images. The vehicle detection methods relying on spatial domain analysis has an advantage fast and easy detect vehicles, but their performance down when the cars feature information is changed such as rotate, move, or change in illumination, size, and color. The method based on frequency domain analysis are better strong to a rotation, move, and changes in the vehiclesillumination, size, and color, but it has a disadvantaged the detection speed is slow. Image analysis is the extraction of relevant information for object detection from an image analysis methods based on features such as texture, density, color,and frequency in an image. Naturally, the performance speed and accuracy of a vehicle detection algorithm depend on the complexity of image analysis algorithm. Sophisticated and robust object detection algorithms provide high accuracy, but their detection speed may not guarantee to make them suitable for the

*Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication*

real-time system. Simple object detection algorithms allow fast, real-time object detection but accuracy is down. The overall performance of the vehicle detection system that relies on quick, vehicle detection and classification is more influenced by the vehicle detector algorithms [8]. In recent years, vehicle detection and classification methods that use deep learning algorithms have emerged to provide fast vehicle detection without loss of precision. Various methods have been conducted to develop accurate and fast vehicle detection and identification systems based on deep learning algorithms [9]. Having a fair comparison among different vehicle detection algorithms performance is robust.There is no straight answer on which model is the best. For real applications, we make choices the detection algorithm to balance accuracy and speed. Besides the detector types, we need to aware of other decisions that impact the performance: (feature extractors, image resolutions, matching strategy, IoU threshold, the number of proposals or predictions, use of multi-scale images in training or testing, etc.). The present paper is designed in the object detection and classification process as the Darknet-53 algorithm. That is known as an excellent training algorithm as the fastest object training and detection method recently [10].

A. Deep Learning

Deep learning algorithms have been attracted a lot of attention in object detection and classification research groups recent. The basic concepts of deep learning algorithms are based on neural networks, developed in the 1980s. Early neural network algorithms have a disadvantage of some shortcomings such as overfitting, local minima, initial weights determination, and slow learning rates, and they could not recognize patterns of high complexity. But, in the mid-2000s, some researchers developed the learning methods that made better learning of many layers in neural networks with deep layers networks. Also, the development of GPGPU with substantial processing power has even much more decreased the calculation time for complex matrix operations recent, and deep learning algorithms have developed in a variety of object detection and classification algorithms. There are numerous deep learning algorithms, most of which are derived from early neural network algorithms. General deep learning networks consist of multiple layers. In general, deep learning networks are trained using the back-propagation that method used in neural networks to compute a gradient. Deep neural algorithms then train biases and weights using the gradient descent method. General deep learning algorithms have introduced concepts, such as dropout, mini batches, and pre-training for initial weights to overcome the problems in traditional neural network algorithms [11].

B. Faster R-CNN

R-CNN uses region proposals instead of sliding windows method for object detection.Selective Search used by R-CNN create potential bounding boxes and eliminate redundant detection through the ability to extract the Convolution Network, adjust the SVM scorebox, adjust linear model boundary box, and suppress non-maximum values.However, faster R-CNN speeds up the R-CNN framework through proposal regions using neural networks instead of Selective Search.Faster R-CNN which are often used for object detection [12]. Faster R-CNN improve speed and accuracy compared to R-CNN, but both are still not sufficient for real-time performance.

C. SSD(Single Shot MultiBox Detector)

SSD is a fast algorithm with high detection speed but slightly lower accuracy compared to the faster R-CNN. It uses a single CNN (convolution neural networks) for fast object detection to reframes object detection as a regression problem to object detection. And SSD is detected using multi-scale feature maps to make the detection results more accurate. For each feature map, a default set of boxes is used [13].

D. YOLO

YOLO (you only look once) is one of the faster object detection algorithm recent. It is a good algorithm when need rapid real-time detection without loss of much accuracy. YOLOv2 used a custom deep architecture darknet-19, 19-layer network supplemented with 11 more layers for object detection. With a 30-layer architecture, YOLOv2 often struggled with small object detections. It was attributed to a loss of fine-grained features as the layers downsampled the input. To overcome this, YOLOv2 used an identity mapping, concatenating feature maps from a previous layer to capture low-level features [14]. However, YOLOv2's architecture was still lacking some of the most critical elements that are now stapled in most of the state-of-the-art algorithms (no residual blocks, no skip connections and no upsampling). But YOLOv3 incorporated all of these. Darknet-53 used in YOLOV3 is an advanced model from Darknet-19 and consists of 53 convolutional layers. It still relies on successive 3×3 and 1×1 filters but added residual blocks. Darknet-53 is more than twice as efficient as ResNet-101 and ResNet-152 [10].

E. Kalman Filter

The Kalman filter is a linear minimum variance of error filter algorithm, it is recursive, and it can run in real time processing, using only the current input values and the previously valued state and its uncertainty matrix. Kalman filter algorithm is a particular case of Bayesian rule with linear, Gaussian assumptions and quadratic. The Kalman filter keeps tracking the estimated next state of the current vector state values over time. The future estimate state is updated using the state transition matrix and the Gaussian noise with known covariance [15]. A motionmodel is needed to construct a Kalman filter for vehicle tracking.

III. PROPOSED METHOD

The proposed system isdesigned to detect and identify vehicles according to the AUSTROADS's classification used Darknet-53 for vehicle images training and Kalman filter for traffic volume measurement [16]. In the implementation of the proposed system, some classes of the classification AUSTROADS were merged and processed due to visually indistinguishable viewing angles in aviation images. [Table 1] shows the differences between the AUSTROADS classification system and the proposed classification.Due to the viewing angle and

Table 1: The AUSTROADS and the Proposed Classification

AUSTROADS Vehicle Classification System		Class	Proposed Vehicle Classification
Short Vehicle less 5.50m	Motorcycles	1	Class 1
	Sedan, Light Van, etc.	1	Class 2
Medium Vehicle 5.50m to 14.50m	Short Towing Trailer, Caravan, Boat, etc.	2	Class 3
	Two, Three, Four Axle and Heavy Truck	3,4,5	Class 4
	Bus	3	Class 5
Long Vehicle over 14.50m	Heavy and Long Vehicle	6-12	Class 3

altitude of the UAV camera, it is difficult to distinguish between different vehicle types with just the axle configuration. The main difference between the merged classes is the number of axles.

A. Implementation of the Proposed System

The proposed method is designed the training, detection and classification process using the Darknet-53 and the Kalman filter algorithm to set a variable detection area. And, we propose the variableclassification method due to parked and non-moved vehicles. So, in this paper, we consider the three conditions of driving, stopping, and parking at the measurement of traffic volume. The proposed system is designed as the following conditions. To train and detect for the vehicle, we use Darknet-53 [17]. This method has fast training and detection speed and has excellent detection speed and accuracy compared to SSD and Faster R-CNN [10]. And we propose the classification method which has the variable classification area to measure traffic flow on the road due to a parked vehicle. For the implementation of the proposed study, we use at the SDX-4195 Deep Learning Server with which are installed in OpenCV, CUDA, Xeon E5-2650 4-CPU, and GTX-1080TI 4-GPU.

B. Data Training and Vehicle Detection

For the implementation of the proposed system, the deep learning network is trained various types of car images which are short cars (sedan, wagon, light van, etc.), medium cars (short towing, short trailer, caravan, bus, etc.) and long vehicles using Darknet-53 algorithm with real vehicle images. The vehicle images were taken in an almost orthogonal direction driving cars about the intersection at an altitude of 100-120m using the DJI Phantom 3 Professional. The test video is composed of UHD 3840×2160 images with a refresh rate of 30 FPS. The performance of a vehicle detector is dependent mainly on the performance of its deep neural network. We use Darknet-53, a variant of the framework for YOLOv2, to implement a deep learning network and to detect cars on aerial images. [Figure 1] shows the number of training cars in each class and [Figure 2] is an example of aerial vehicle images.

C. Traffic Flow Measurement using Kalman Filter

It is needed to calculate the position information which is the vehicle's position in two consecutive frames for traffic flow measurement when vehicles are detected in the aerial images. This information is also needed to find the same cars in consecutive frames. The distance of the vehicle positions in the two consecutive frames is not significant because the aerial video is recorded by 30 FPS (frames per second). So, In general approaches, the vehicle nearest to the previous frame of the target vehicle within the detection range is determined as the target vehicle in the current frame. But, this method can rise to track error when the detection range is disappeared with cars driving closely and when cars direction moves other direction. The proposed plan minimize the current frame errors by using a Kalman prediction value extracted in the previous frames. We define a vehicle movement model to the traffic flow measurement in successive frames like the following steps.

Step 1: The state vector of the vehicle at time k is expressed as next in Equation 1 and 2.

$$x(k) = [x(k) \quad y(k) \quad \Delta x \quad \Delta y]^T \quad (1)$$

$$x(k+1) = \Phi(k)x(k) + w(k) \quad (2)$$

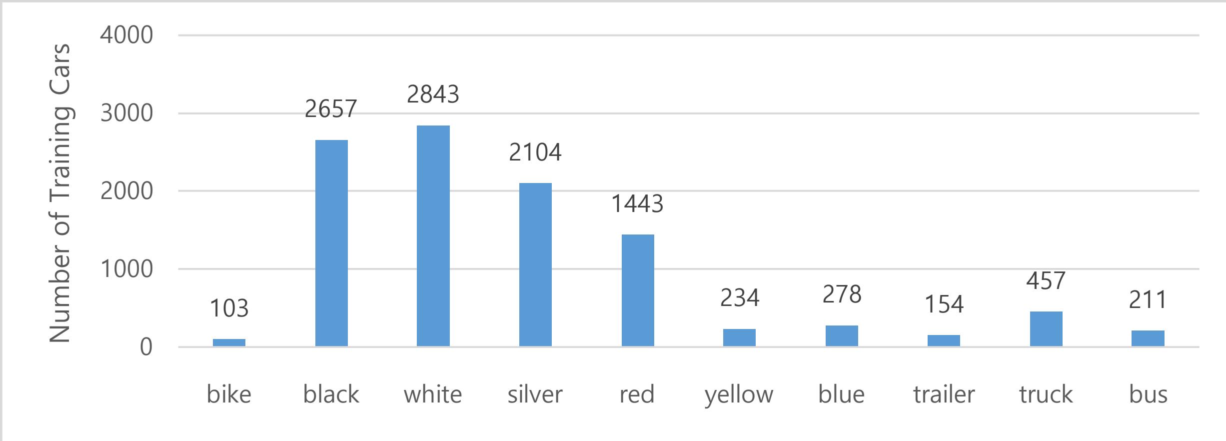
Step 2:$w(k)$ is the motion prediction noise and $v(k)$ is themeasurement noise.

Step 3: The values $v(k)$ between the previous-frame estimated values $\hat{z}(k|k-1)$ and the current-frame values $z(k)$ is computed using Equation 3 and 4.

$$\hat{z}(k|k-1) = H(k)\hat{x}(k|k-1) \quad (3)$$

$$v(k) = z(k) - \hat{z}(k|k-1) \quad (4)$$

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication



The proposed method uses the measurement error that is

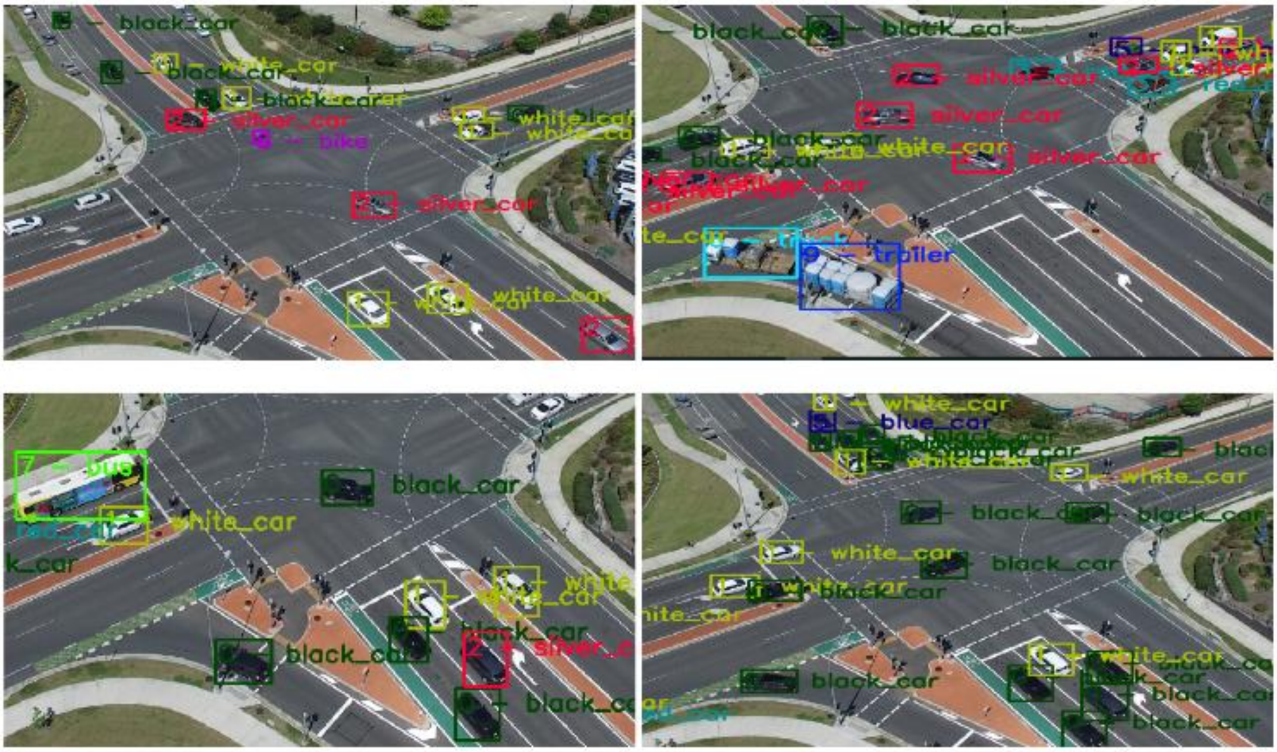


Figure 2. Vehicle Images for Network Training [17]

called "innovation" value to set a search area for vehicle detection in the next frame. The proposed plan can be possible to set the variable detection area when the vehicle movement speed of the target vehicle is fast or slow. When the vehicle moves fast, a wide detection area is set. And if the vehicle moves slowly, a narrower detection area is given, enabling more fine-grained searches. In this paper, using the parameters of the Kalman filter, the condition of the vehicle is divided into three cases.

In the experiment, when the vehicles' Kalman filter values have changed the detection and classification are performed only for vehicles driving and stopping to measure traffic volume, and in case of the Kalman filter values have not changed there were not applied to the classification.

1. Driving: when the vehicles' Kalman filtervalues are changed and thecase of vehicle moves.

2. Stopping: In the case of vehicle speed gradual reduced and stop.

3. Parking: when the vehicles' Kalman filter values are not changes and the case of vehicle not movement.

IV. THE RESULT OF EXPERIMENTAL

The aerial images use in the experiment are images of vehicles on the road recorded under 120m altitude using Phantom3 Professional UAV of DJI's. The size of the experimental images is 3940×2160, 30 FPS (frames per second) in UHD resolution. The experimental result [Table 2] is showed that the proposed method resulted in better classification performance than the process of fixed search. As can be seen in [Table 2], the proposed method is about two times more accurate than the fixed search ranges. [Figure 3] shows the vehicle detection and classification result in aerial images. In this experiment, we only use vehicles that are stopping and driving for traffic measurement and target tracking except for parked cars.In the experiment result, two problems have occurred in the following cases. The first problem is the "combination"; this case

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

occurred when vehicles, which are spatially separated at the beginning, overlapped on the aerial image. And the second problem is the "dissociation"; this case is when the combined cars, a group of vehicles which were being tracked together, diverged. In the proposed method, it flags which get separated from the combined vehicle with the 'dissociated vehicle' label and keeps classified of them individually. These problems occurred purely due to detection errors.

V. CONCLUSION

UAV based traffic monitoring system is an essential system to build ITS. The proposed study proposed a detection and classification system using UAV that could carry out vehicle classification by AUSTROADS' plan and be compared with non-invasive detection systems. And the proposed method was designed and implemented the deep neural network and Kalman filter for the real-time vehicle detection and classification system. In the experiments, the proposed method successfully detected andclassified vehicle which moves at 23 FPS maximum and 17-20 FPS on average with UHD-4k video. The proposed method with variable search ranges made minor errors than the method with fixed search ranges. Even though the experiment dataset used in the test was not enough big and diverse to make a clear statement, the proposed method performed better precious than other methods. In the future, the proposed method will be studied with an expanded set of aerial images, and the system to detect and classify various types of objects (e.g., buildings, animals, pedestrian, traffic lanes, traffic lights) will also be

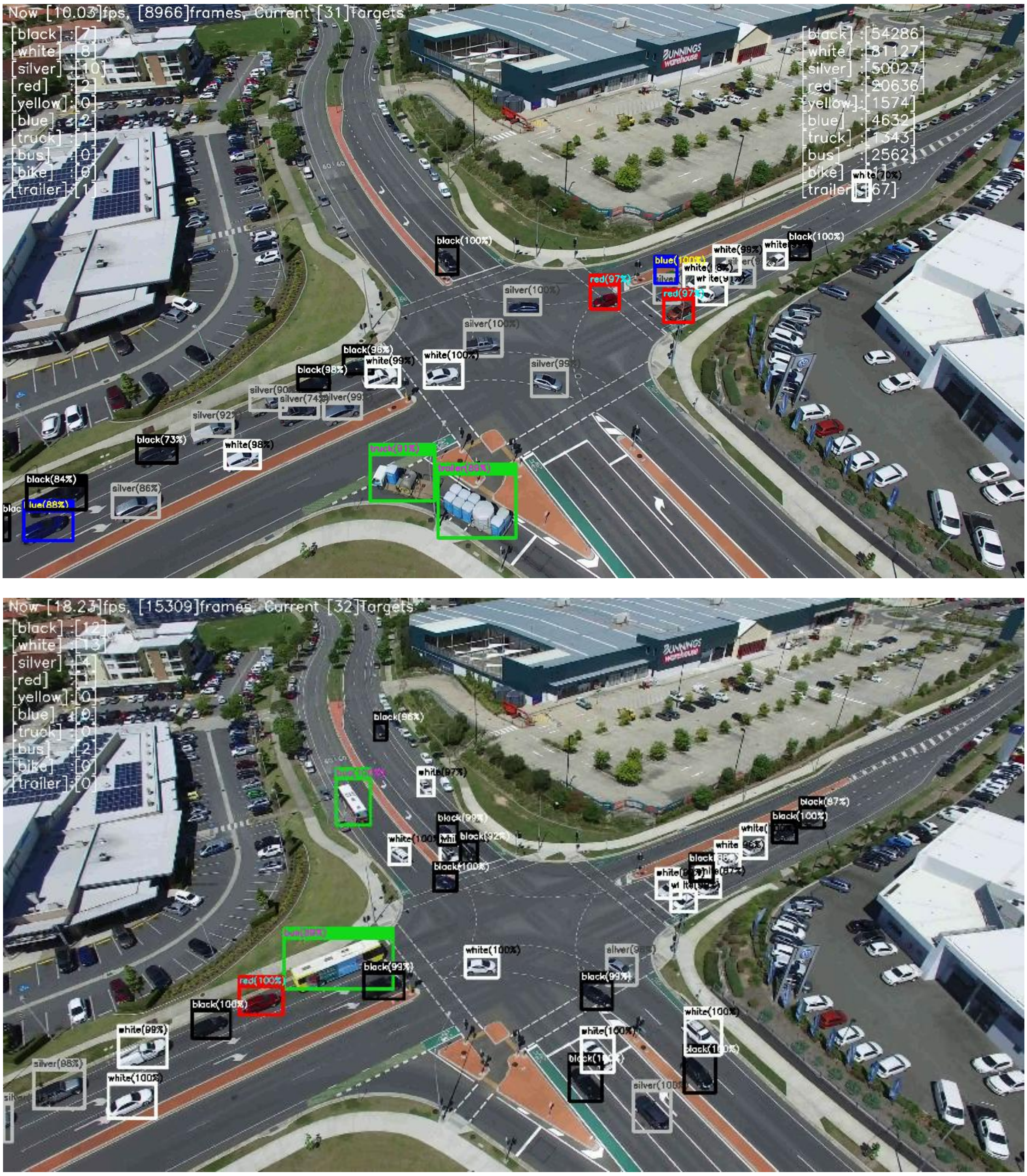


Figure 3. The result of vehicles classification

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Table 2: The Experiment Results of the Proposed Method

Vehicle type		# of cars	# of proposed method errors	# of conventional method errors
Short Vehicle	Motorcycles	243	17	43
	Black Cars	15,425	519	1,373
	White Cars	17,527	625	1,745
	Silver Cars	10,342	372	939
	Red Cars	4,183	152	418
	Yellow Cars	253	14	32
	Blue Cars	814	29	93
Medium Vehicle	Bus	828	35	92
	Heavy Truck	327	21	39
LongVehicle	Long Trailer	68	4	15

developed.

ACKNOWLEDGMENT

This research was funded by a 2019 research Grant from Sangmyung University.

REFERENCES

1. Adrian Carrio, Carlos Sampedro, Alejandro Rodriguez-Ramos& Pascu al Campoy, A Review of Deep Learning Methods and Applications for Unmanned Aerial Vehicles. Journal of Sensors, Article ID 3296874, 2017, https://doi.org/10.1155/2017/3296874
2. Yaw Okyere, Adu-Gyamfi, Sampson Kwasi Asare, Anuj Sharma, Tienaah Titus, Automated Vehicle Recognition with Deep Convolutional Neural Networks. Transportation Research Record: Journal of the Transportation Research Board, 2017 2645(1):113-122. https://doi.org/10.3141/2645-13
3. HeikkiHuttunen, Fatemeh ShokrollahiYancheshmeh, Ke Chen, Car Type Recognition with Deep Neural Networks, IEEE Intelligent Vehicles Symposium (IV), 2016 :1115-1120. DOI : 10.1109/IVS.2016.7535529
4. Yongzheng Xu, Guizhen Yu, Yunpeng Wang, Xinkai Wu, and Yalong Ma, Car Detection from Low-Altitude UAV Imagery with the Faster R-CNN. Journal of Advanced Transportation, Article ID 2823617, 2017, https://doi.org/10.1155/2017/2823617
5. Tao Qu, Quanyuan Zhang, Shilei Sun, Vehicle detection from high-resolution aerial images using spatial pyramid pooling-based deep convolutional neural networks. Multimedia Tools and Applications, October 2017 76(20):21651–21663. https://doi.org/10.1007/s11042-016-4043-5
6. C. Sheppard, M. Rahnemoonfar, Real-time scene understanding for UAV imagery based on deep convolutional neural networks. 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, 2017: 2243-2246.doi: 10.1109/IGARSS.2017.8127435
7. Herrera, G J, Dechant,Jason A, Green E K, Klein, Ethan A, Technology Trends in Small Unmanned Aircraft Systems (sUAS) and Counter-UAS: A Five Year Outlook. Institute for Defense Analyses Alexandria, 2017. https://apps.dtic.mil/dtic/tr/fulltext/u2/1047506.pdf
8. C. Kyrkou, G. Plastiras, T. Theocharides, S. I. Venieris, C. Bouganis, DroNet: Efficient convolutional neural network detector for real-time UAV applications. 2018 Design, Automation & Test in Europe Conference & Exhibition (DATE), Dresden, 2018:967-972.doi: 10.23919/DATE.2018.8342149
9. T. Moranduzzo, F. Melgani,Detecting Cars in UAV Images with a Catalog-Based Approach. IEEE Transactions on Geoscience and Remote Sensing, 2014 52(10):6356-6367. DOI : 10.1109/TGRS.2013.2296351
10. J. Redmon, A. Farhadi, YOLOv3: An incremental improvement. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018, arXiv:1804.02767v1
11. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. Nature, 2015 (521):436-444. DOI: https://doi.org/10.1038/nature14539
12. Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, June 2017 39(6):1137–1149. arXiv:1506.01497v3
13. Liu W. et al. SSD: Single Shot MultiBox Detector. Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science,
14. Springer, Cham, 2016 9905:21-37. https://doi.org/10.1007/978-3-319-46448-0_2
15. Joseph Redmon, Ali Farhadi, YOLO9000: Better, Faster, Stronger. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017:7263-7271. arXiv:1612.08242v1
16. S. Y. Chen, Kalman Filter for Robot Vision: A Survey. IEEE Transactions on Industrial Electronics, Nov 2012 59(11):4409 – 4420. DOI: 10.1109/TIE.2011.2162714
17. AUSTROADS Vehicle Classification System. [Internet]. AUSTROADS Australia; [cited 2019 April 8]. Available from: https://austroads.com.au/publications/pavement/agpt04k/austroads-vehicle-classification (website)
18. Alexey. (2018). GitHub. Yolo-v3 Windows and Linux version. [Internet]. [cited 2019 April 8], Available from: https://github.com/AlexeyAB/darknet (website)

AUTHORS PROFILE

ChangJinSeo has received his Ph. D. in Multimedia Engineering from Pusan National University, South Korea in 2003. He started his career as Sensor Technology Research Center (STRC) in 1999 and thereafter he moved to Sungduk University in 2000 as a Professor. From 2013 to currently, he is working as a Professor at Sangmyung University, Dept. of Information Security Engineering, South Korea and involved actively in teaching and research mainly in the area of Deep Learning, Object Detection and Target Tracking.