The Journal of The Institute of Internet, Broadcasting and Communication (IIBC)
Vol. 21, No. 1, pp.141-148, Feb. 28, 2021. pISSN 2289-0238, eISSN 2289-0246

https://doi.org/10.7236/JIIBC.2021.21.1.141
JIIBC 2021-1-19

YOLO v4 기반 혼잡도로에서의 움직이는 물체 검출 및 식별

Detection and Identification of Moving Objects at Busy Traffic Road based on YOLO v4

이추담*, 정석용*, 왕욱비*, 진락*, 손진구*, 송정영**

Qiutan Li*, Xilong Ding*, Xufei Wang*, Le Chen*, Jinku Son*, Jeong-Young Song**

요 약 일부 네거리나 혼잡도로에서 특정 시간대에 행인이 많고 도로가 막혀서 발생하는 교통사고가 적지 않다. 특히 인근에 학교교차로가 있어 바쁜 시간에 학생들의 교통안전을 지키는 것이 중요하다. 과거에는 교통 신호등을 설 계 했을 때 행인의 안전성을 고려하지 않고 자동차 인식과 교통 최적화에 대하여 연구 했다. 행인, 특히 학생들의 안전을 확보하는 전제에서 가능한 한 도로의 소통을 유지하는 것이 본 연구의 중점적인 연구 방향이다. 본 연구는 사람, 오토바이, 자전거, 자동차, 버스의 식별문제를 중점적으로 연구할 것이다. 조사와 비교를 통해 본 연구는 YOLO v4 네트워크로 목표물의 위치와 수량을 식별하는 것을 제시한다. YOLO v4는 작은 목표물의 식별 능력이 강하고 정밀도가 높으며 처리 속도가 빠르다는 특징을 가지고 있으며, 데이터 수집 대상을 설정하여 이미지 집합을 훈련하고 테스트 한다. 움직이는 영상에서 목표물의 정확도, 실수율과 누락율에 대한 통계를 사용하여, 본 연구에서 훈련된 네트워크는 움직이는 이미지 속의 사람, 오토바이, 자전거, 자동차와 버스를 정확하게 식별 할 수 있다.

Abstract In some intersections or busy traffic roads, there are more pedestrians in a specific period of time, and there are many traffic accidents caused by road congestion. Especially at the intersection where there are schools nearby, it is particularly important to protect the traffic safety of students in busy hours. In the past, when designing traffic lights, the safety of pedestrians was seldom taken into account, and the identification of motor vehicles and traffic optimization were mostly studied. How to keep the road smooth as far as possible under the premise of ensuring the safety of pedestrians, especially students, will be the key research direction of this paper. This paper will focus on person, motorcycle, bicycle, car and bus recognition research. Through investigation and comparison, this paper proposes to use YOLO v4 network to identify the location and quantity of objects. YOLO v4 has the characteristics of strong ability of small target recognition, high precision and fast processing speed, and sets the data acquisition object to train and test the image set. Using the statistics of the accuracy rate, error rate and omission rate of the target in the video, the network trained in this paper can accurately and effectively identify persons, motorcycles, bicycles, cars and buses in the moving images.

Key Words : person, motocycle, bicycle, bus, car, detection, YOLO v4

*준회원, 배재대학교 컴퓨터공학과
**정회원, 배재대학교 컴퓨터공학과(교신저자)
접수일자 2020년 11월 16일, 수정완료 2021년 1월 16일
게재확정일자 2021년 2월 5일

Received: 16 November, 2020 / Revised: 16 January, 2021 / Accepted: 5 February, 2021
**Corresponding Author: jysong@pcu.ac.kr
Dept. of Computer Engineering, Pai Chai University, Korea

Ⅰ. Introduction

In addition to relying on cars, Korean residents also rely on many motorcycles and bicycles. In recent years, there are more and more electric bicycles. The number of registered motorcycles with a displacement of more than 50cc in South Korea will reach about 2.2 million by 2020. In order to promote electric bicycles, from March 22, 2018, electric powered bicycles with a speed of less than 25 km/h and a vehicle weight of no more than 30 kg are allowed to pass on bicycle roads in South Korea.[1]

With the continuous development of urbanization, the number of cars, motorcycles and electric bicycles is growing, which brings convenience to people and brings huge traffic safety risks. After the outbreak of COVID-19, the South Korean government issued a social distance prevention measure. More and more people are going out less, and the number of mobile phone takeaway orders has increased sharply. Although the takeout industry is booming, the traffic accident rate of motorcycles and electric bicycles has increased significantly. Relevant statistics show that motorcycles, electric bicycles and other two wheeled vehicle traffic accidents occur frequently, and the fatality rate is high.

According to Seoul news, South Korea's Traffic Safety Association and Ministry of land and transportation said on September 15, 2020 that the number of people killed in two wheeled vehicle accidents has gradually become younger, and the number of young people aged 10-20 years is gradually increasing. In the first half of 2020, 9880 two wheeled vehicle accidents occurred, with a year-on-year increase of 2.8%.[2]

The detection of persons, motorcycles bicycles, cars and buses is not only an important part of driving environment information recognition in unmanned driving and active safety, but also widely used in the design of traffic signal command system. Person, motorcycle and electric bicycle detection is essentially target detection. If the above test results can be used in the traffic control system under the principle of "pedestrian priority", more traffic protection will be provided for pedestrians.

According to the detection requirements of high recognition accuracy, good real-time performance and good detection effect for small targets, the target detection algorithms based on CNN (Convolutional Neural Network) are compared and selected, and finally the YOLO v4 (You Only Live Once) algorithm is selected as the detection algorithm in this paper.[3]

Ⅱ. Relational Research

In the ImageNet large-scale image classification competition held in 2012, the team using the deep neural network won the championship, and then the person detection using the deep neural network method received high attention. The regional convolution network (RCNN) target detection method proposed in 2014 uses the selective search method to obtain 2000 candidate frames, and then extracts the convolution features of the candidate frames, and then classifies the extracted features with support vector machine. Although the accuracy of RCNN target detection network has been improved, the calculation speed is greatly reduced.

The spatial pooling pyramid network (SPP) proposed in 2015, which can make RCNN output any size, can extract the convolution features of different candidate frames, so that the computing speed is greatly improved. Fast RCNN can classify candidate frames and combine the tasks before regression, can greatly shorten the training time, and the calculation speed is 50 times faster than before, and the calculation accuracy is also greatly improved. Then, the

Faster RCNN method is faster than Fast RCNN by extracting candidate frames directly from convolution feature map, and then using Fast RCNN to classify and regress.[4]

Through online search, it is found that there are few papers on target detection of motorcycles and electric bicycles. At present, the target detection based on deep learning develops rapidly, and the detection effect is getting better and better, so the algorithm based on deep learning is used to identify electric persons and motorcycles.

Ⅲ. Selection of target detection algorithm

The first task of this paper is to select a target detection algorithm with high accuracy and good real-time performance. By comparing and selecting the target detection algorithms based on CNN, the YOLO v4 target detection algorithm with high accuracy, good real-time performance and good detection effect on small targets is finally selected.

CNN is one of the important network structures in deep learning technology. It has been used in image recognition, speech recognition and other occasions. It has achieved very good results in computer vision applications. At the same time, it has made great achievements in the competition datasets such as ImageNet and coco. Compared with the image processing algorithm, the advantage of CNN is that compared with the traditional full connection method of neural network, CNN avoids the problem of excessive parameters. CNN can optimize the calculation speed and save space by local connection, weight sharing and other methods.

At present, the mainstream target detection algorithm is mainly based on CNN deep learning model.

YOLO transforms the target detection task into regression problem, which greatly speeds up the detection speed. YOLO series has been updated to the fourth generation, namely YOLO v4. YOLO v4 is the most balanced target detection network in terms of speed and accuracy so far. Through the integration of a variety of advanced methods, all the short boards of YOLO series (including those not good at detecting small objects in YOLO V1) are made up to achieve amazing results and speed of group selection. The algorithm not only has a good effect for the object, but also has good compatibility for other objects, such as art works.[5]

Ⅳ. Deep learning environment building and YOLO v4 training

After selecting the target detection algorithm, the next main task is to configure the software, build a deep learning platform, and then configure the parameters of YOLO v4 network according to the identification content of this paper, and finally train the YOLO v4 network.

1. Dataset construction

The main research content of this paper includes the recognition of person, motorcycle, bicycle, car and bus, so it is necessary to establish a dataset for YOLO v4 training. In this paper, the main ideas of establishing dataset are as follows: identify the object, collect the dataset, label the dataset, and construct the PASCAL VOC2012 dataset available for YOLO v4.

(1) Target detection object

The target detection objects in this paper are persons, motorcycles, bicycles, cars and buses. Restricted by the conditions, the person dataset in this paper is composed of images at some intersections, mainly including persons, motorcycles, bicycles, cars and buses. There may be some persons, motorcycles, bicycles, cars and

buses models that can not be collected. Therefore, in order to enrich the dataset, some photos are downloaded from the Internet.[6]

(2) Data collection

At present, the main popular datasets are VOC (The PASCAL Visual Object Classes Challenge), ImageNet and COCO (Microsoft Common Objects in Context), which have reached a high degree in quantity and diversity. A very good dataset not only reflects its quantity and photo quality, but also its diversity. The official weight file of YOLO v4 target detection algorithm is obtained by training VOC and coco datasets. After comparison, it is decided to make VOC2012 dataset for YOLO v4.[7,8]

In this paper, we first consider the quantity and quality of the images in the dataset, as well as the richness of the data.

Therefore, when selecting pictures, in addition to the pictures provided by VOC2012, the images of UA-DETAC (a challenging real-world multi-object detection and multi-object tracking benchmark) dataset are also added, as shown in Fig 1. The UA-DETRAC benchmark dataset consists of 100 challenging videos captured from real-world traffic scenes (over 140,000 frames with rich annotations, including illumination, vehicle type, occlusion, truncation ratio, and vehicle bounding boxes) for multi-object detection and tracking. After sorting out, the testing dataset collected about 5100 images, and the training dataset collected about 11000 images.

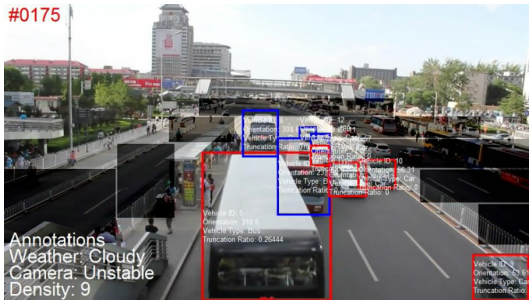


Fig. 1. UA-DETAC dataset
그림 1. UA-DETAC 집합

(3) Annotation of datasets

LabelImg is a visual image calibration tool, which is used to mark the position and name of objects in the image in depth learning. At present, the popular target detection networks such as Faster RCNN, YOLO, SSD can use this tool to label the target in the image.

After the completion of the image collection, it is necessary to convert the image into a format that can be used for the network training of YOLO v4. First install Python 3.7.9 on computer, and then install labelImg on python. In the tagging process, only a single recognition object is annotated, such as person, motorcycle, bicycle, car, bus, as shown in Fig 2.

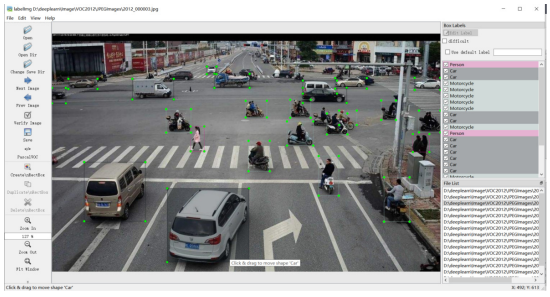
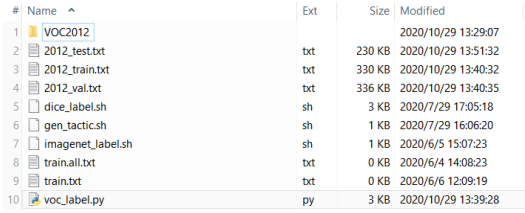


Fig. 2. labelImg annotation
그림 2. labelImg 주석

(4) Making VOC2012 dataset of YOLO v4

Pascal VOC provides a set of standardized dataset for image recognition and classification. We will not use the whole dataset, but prepare our own dataset according to its format. The VOC2012 dataset contains 11530 images for training and verification, including 27450 calibrations in ROI and 20 classifications.

The directory structure of the VOC2012 dataset is shown in Fig 3, which contains five folders. Because there is no image segmentation in this paper, the SegmentationClass and SegmentationObject folders are not used.



#	Name	Ext	Size	Modified
1	VOC2012			2020/10/29 13:29:07
2	2012_test.txt	txt	230 KB	2020/10/29 13:51:32
3	2012_train.txt	txt	330 KB	2020/10/29 13:40:32
4	2012_val.txt	txt	336 KB	2020/10/29 13:40:35
5	dice_label.sh	sh	3 KB	2020/7/29 17:05:18
6	gen_tactic.sh	sh	1 KB	2020/7/29 16:06:20
7	imagenet_label.sh	sh	1 KB	2020/6/5 15:07:23
8	train.all.txt	txt	0 KB	2020/6/4 14:08:23
9	train.txt	txt	0 KB	2020/6/6 12:09:19
10	voc_label.py	py	3 KB	2020/10/29 13:39:28

Fig. 3. VOC2012 dataset structure
그림 3. VOC2012 집합 구조

This paper first establishes a folder directory with the same file name as the VOC2012 dataset, and then puts all the pictures into the JPEGImages folder, and puts all the XLM files marked by ourselves under the Annotations file. Then, the corresponding program is applied to generate 4 ImageSets / Main / files, and the production of this VOC2012 dataset is completed. YOLO v4 can't use this dataset directly, it needs to run VOC_label.py. After running, three more TXT files will be generated in the main directory, namely 2012_test.txt, 2012_train.text, 2012_val.text as shown in Fig 3, the dataset suitable for the training of YOLO v4 target detection algorithm is completed.

2. Training environment construction and training

In this paper, on the basis of mastering the construction of deep learning model, the model is trained and the parameters of the model are optimized. In a deep learning model, a large number of parameters need to be calculated and optimized, which will inevitably cost a lot of computing power and time. Good computing hardware can greatly reduce the time cost of training, especially when we need to repeatedly adjust the parameters of the model, the increase of time cost will be disastrous. Therefore, this section will build a simple deep learning workstation according to its own training model.

(1) Hardware and software configuration

This section will build a deep learning workstation to select the configuration in the deep learning workstation. The software and hardware parameters to be selected in this paper are shown in Table 1.

Table 1. Hardware and software parameters
표 1. 하드웨어 및 소프트웨어 매개 변수

Hardware and software	Detailed parameters	Hardware and software	Detailed parameters
System	Windows10 64bit	cuDNN	7.5.1
Graphics card	GTX1080Ti	OPENCV	3.4.6
CPU	i7-7700k	Python	3.7.9
CUDA	10.0.130	Darknet	Darknet53

(2) YOLO v4 configuration

Because the persons, motorcycles, bicycles, cars and buses that need to be identified in this paper, they are different from the 81 objects of official identification, it is necessary to change the file of YOLO v4 before training.

(3) Selection of training parameters

The training parameter settings mainly include the maximum training times, learning rate, input format of pictures, etc. the specific parameter meanings are shown in Table 2:

Table 2 Description of training parameters
표 2. 훈련 매개 변수 설명

Training parameters	Function	Explain
batch	sample size	The parameters are updated once for each batch of samples.
subdivisions	The number of samples sent into the trainer at one time	The bigger the batch, the better the training effect; The greater the subdivisions can reduce the pressure of the graphics card.
widths	Input the width of the image	Set it to a multiple of 32 pixels
heights	Input the height of the image	Set it to a multiple of 32 pixels
channels	The number of channels to enter the image	
momentum	Momentum parameters	Momentum parameters in the optimization method in DeepLearning1
decay	The weight decays the regular term	
saturation	saturation	Generate more training samples by adjusting saturation
exposure	exposure	Generate more training samples by adjusting exposure
learning_rate	Determines the speed of weight updating	Setting it too large causes the result to exceed the optimal value; setting it too small causes the decline to slow

On the basis of the above analysis, the main parameters are set in this paper. The relevant parameters are as follows, and other parameters are the default values, as shown in Table 3.

Table 3. Setting of training parameters
표 3. 훈련 매개 변수 설정

Training parameters	Parameter values	Training parameters	Parameter values
batch	8	max_batches	40000
subdivisions	32 pictures	saturation	1.5
widths	416 pixels	exposure	1.5
heights	416 pixels	learning_rate	0.001
channels	3	Burn_in	1000
momentum	0.9	decay	0.0005

V. Verification and analysis of training results

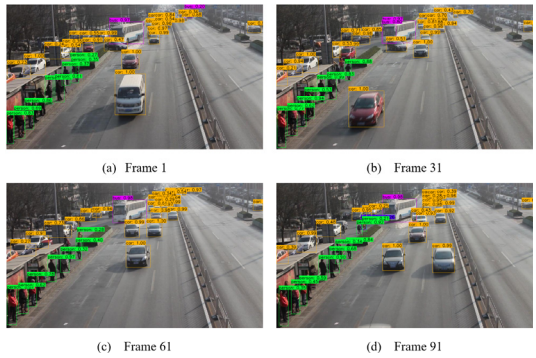
1. Training results

After four days of training, the average loss of training results is less than 0.060730 avg on the deep learning experimental platform built in this paper, and the training results have been satisfactory. Due to too many data results, this paper only shows the final training results, and the recognition results are shown in Fig 4.

Fig. 4. Training results
그림 4. 훈련 결과

The targets of this training test are person, motorcycle, bicycle, car and bus, because the detection of fewer targets, so the speed is relatively fast. In the figure, "40000" refers to the number of iterations of the current training; in the figure, "0.043076" is the overall loss; in the figure, "0.043057 avg" is the average loss, and the lower the value, the better. Generally speaking, once this value is lower than 0.060730 avg, the training can be terminated. In the figure, "0.000100 rate" represents the current learning rate, and its initial value and adjustment strategy are defined in the. cfg file; "10.501258" seconds in the figure represents the total time spent in the current batch training; and "9267 images" in the figure represents the total number of pictures participated in the training so far.

2. Training results analysis



(a) Frame 1 (b) Frame 31

(c) Frame 61 (d) Frame 91

Fig. 5. Identification results
그림 5. 식별 결과

In order to show the training results, some of the pictures in UA-DETRAC dataset is selected. The MVI_20011 folder contains 664 pictures, the resolution is 960X540, and the first four frames are captured by capturing one picture every 30 frames. The recognition results are shown in Fig 5.

On frame 1, 35 boxes were identified, 32 were identified, 0 were misidentified and 6 were not identified. The specific results are shown in Table 4.

From the detection results of frame 1, there are no motorcycles or bicycles in the picture, and the detection accuracy of cars and buses is very high. Due to the serious occlusion, some vehicles far away are not detected. Person detection accuracy is relatively high.

Table 4. Image detection results of frame 1
표 4. 프레임의 영상 검출 결과 1

	Detection Accuracy rate	Detection Error rate	Omission rate
Person	75. 00%	No	25. 00%
Motorcycle	No	No	No
Bicycle	No	No	No
Car	88. 89%	No	11. 11%
Bus	100. 00%	No	0. 00%

In this video, 22325 boxes were identified , which can accurately judge the area where travelers and electric vehicles are located. At the same time, the processing speed reaches about

30fps, which is not different from the real-time (30fps). It meets the requirements of real-time and achieves the research purpose of this paper.

VI. Discussion

The main task of target detection is to find all the objects interested in the image. The objects of interest in this paper are pedestrians, motorcycles, bicycles, cars, buses, etc.

In this paper, based on the network of Yolo V4, the VOC dataset suitable for YOLO v4 is constructed, and the ideal recognition effect is achieved. In the selection of data sets, in addition to some of the pictures in the VOC data set, we also use some pictures from the ua-detac data set. If we can collect some clearer pictures, we will get better training results.

When the speed is high, the training network for high-speed motorcycle recognition effect is poor, after recognition rate less than 83%, the main reason is that the data set of motorcycles in the data set is not rich enough, adding the corresponding data can achieve more than 96% of the recognition effect. The target recognition should also include position recognition and movement trend recognition. This paper only does the pedestrian position recognition, but not the movement trend recognition.

In the future, based on the principle of "pedestrian priority", the design of traffic control system will continue.

VII. Conclusion

With the continuous development of target detection algorithm based on deep learning, the recognition accuracy and recognition speed have been greatly improved. Therefore, it is feasible to use the target detection algorithm based on deep learning to recognize pedestrians, motorcycles and vehicles.

In this paper, firstly, the fixed image of electric motorcycle and pedestrian are identified. The recognition results show that the trained YOLO v4 network has high recognition accuracy, which can be applied in the field of unmanned driving and traffic target recognition.

References

[1] Ensanian A. Discovering the Motorcycle: The History. The Culture. The Machines[M]. Hillcrest Publishing Group, 2016:35-50

[2] ZHANG W. Ataraxia. Emotional care & driverless car service[J]. 2018.

[3] Yi Z, Yongliang S. An improved tiny-yolov3 pedestrian detection algorithm[J]. Optik, 2019, 183: 17-23.
DOI : 10.1016/j.ijleo.2019.02.038

[4] Silver D, Huang A, Maddison C J, et al. Mastering the game of Go with deep neural networks and tree search[J]. nature, 2016, 529(7587): 484-489.
DOI: https://doi.org/10.1038/nature16961

[5] Liu S, Huang D, Wang Y. Pay Attention to Them: Deep Reinforcement Learning-Based Cascade Object Detection[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019.

[6] Espinosa J E, Velastín S A, Branch J W. Detection of Motorcycles in Urban Traffic Using Video Analysis: A Review[J]. IEEE Transactions on Intelligent Transportation Systems, 2020.

[7] Behailu H. A Real-time Obstracle Detection And Interframe Information Storage System From Motion Pictures For Assisting Blind People[D], 2020.

[8] Y. S. Im, E. Y. Kang, "MPEG-2 Video Watermarking in Quantized DCT Domain," The Journal of The Institute of Internet, Broadcasting and Communication(JIIBC), Vol. 11, No. 1, pp. 81-86, 2011.

저 자 소 개

Qiutan Li(준회원)

• Qiutan Li received the master degree in Computer Engineering from Shandong University of Science and Technology. He is currently studying for the computer engineer degree of Pai Chai University. His research interests include Computer Network, Artificial Intelligence.

Xilong Ding(준회원)

• Xilong Ding received the master's degree in Computer Application Technology form Ocean University of China in 2007. PhD student at Pai Chai University. His research interests include Machine learning and Computer vision technology.

Xufei Wang(준회원)

• Xufei Wang received the master degree in mechanical engineering from Xinjiang University. He is an associate professor at Shanxi University of Technology. He is pursuing a doctorate in computer engineering at Pai Chai University in Korea. His research interests include Self-driving, Machine learning.

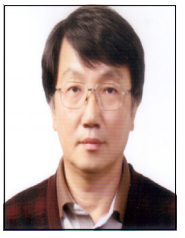
Le Chen(준회원)

• Le Chen received the master degree from Nanjing University of Technology in 2018. He is currently studying for the computer engineer degree of Pai Chai University. His research interests include Software Engineer, Artificial Intelligence..

Jinku Son(준회원)

• Jinku Son received master degree in computer engineering from Waseda University. Japan in 2011. His research interests include AI, Big data, BPM.

Jeong-Young Song(정회원)



• 1984.2: B.S. Degree, Computer Engineering, Hannam Univ. S. Korea
• 1992.3: M.S. Degree, Electrical Information and System, Waseda Univ., Japan.
• 1995.3: Ph.D. Degree, Electrical Information and System, Waseda Univ., Japan.
• 1995.3~1997.2: Computer Science, CheongUn Univ., Korea.
• 1997.3~Present: Computer Engineering, Professor, PaiChai Univ., S. Korea.
• 2011.9~2012.8: Invited Scholarship Professor, Department of EE(Electrical Engineering), ISU(Idaho State University), USA.
• Research Interests : Pattern Processing(Image, Speech, Character), Machine Learning. etc..