

The implementation of intelligent systems in automating vehicle detection on the road

Susanto, Dimas Dwi Budiarjo, Aria Hendrawan, Prind Triajeng Pungkasanti

Department of Information Technology and Communication, Universitas Semarang, Indonesia

Article Info

Article history:

Received Feb 20, 2020

Revised May 25, 2021

Accepted Jun 6, 2021

Keywords:

Artificial intelligence

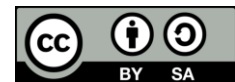
Convolutional neural network

Object detection

ABSTRACT

Indonesia is a country with a high population, especially in big cities. The road always crowded with various types of vehicles. Sometimes the growth of vehicles is not matched by road construction. During peak hours, too many vehicles can cause traffic jams on the road. The road is needed to be widened to accommodate the number of vehicles that pass each day. In order for road widening to be precise at locations that frequently occur in traffic jams, data on the number and classification of vehicles passing is required. Therefore, a system that can calculate and recognize the type of vehicle that passes is needed. The development of various studies on artificial intelligence especially about object detection can classify and calculate the type of vehicle. In this study, the authors used the you only look once (YOLO) object detection system using a convolution neural network (CNN) method to classify and count vehicles that pass automatically. The author uses a dataset of 600 images with 4 classes which are car, truck, bus, and motorbikes that pass through the road. The results showed that the YOLO object detection system can recognize objects consistently with accuracy more than 80% on CCTV video that installed on the road.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Aria Hendrawan

Department of Information Technology and Communication

Universitas Semarang

Soekarno Hatta Street, Tlogosari, Semarang 50196, Indonesia

Email: ariahendrawan@usm.ac.id

1. INTRODUCTION

In Indonesia, especially in big cities, vehicle growth rates are very high, this is because people prefer to use private vehicles rather than public vehicles. But the high rate of vehicle growth is not matched by adequate road construction, so that too many vehicles and inadequate roads cause congestion which is detrimental to society. Therefore, road construction or widening is needed in order to reduce congestion [1]. Construction or widening of the highway to reduce congestion requires careful planning and must be in accordance with needs. To find out the location of congestion points, it requires data on the number of vehicles passing through that location. More vehicles passing by means that road construction or widening is needed at the location to reduce congestion. On every highway there are various types of passing vehicles such as cars, trucks, buses, and motorbikes. Data of each type of vehicle that passes can be used for consideration of development plans or road widening. To count the number of vehicles that pass each day can automatically use the application by implementing a smart system [2].

As research on artificial intelligence develops, one of them is object detection, which can help recognize objects in an image. Object detection is one area in computer vision. Computer vision is the study of how computers see and analyze objects in an image. Object detection is useful for detecting or recognizing

objects in an image based on the shape, color or from the dataset created. There are various ways to implement object detection applications, including using the convolutional neural network (CNN) method and *you only look once* (YOLO) detection system. In research conducted by Redmon *et al.* [3], the YOLO detection system was proven to recognize an object more quickly in the image so that it is very suitable when applied to real-time object detection on video [4]-[7].

In real-time object detection the speed of object detection is very important because it is different from an image, in a video can process 24 frames per second (fps) or more. If the object detection process takes too long then the resulting video will not be good, there will be a delay of each frame so that the video becomes broken. By applying object detection using YOLO to an application can help the calculation and classification of each vehicle that passes on the highway in real-time via CCTV video. Passing vehicles will be counted and classified automatically based on their detection results and accuracy level [5].

Based on the description above the application of an intelligent system that is object detection using YOLO can be a solution to classify and count every vehicle that passes on the highway in real-time every day. Vehicle calculation and classification data obtained can be used for consideration of determining road construction or widening [8].

2. RESEARCH METHOD

We actualize this model as a convolutional neural system and assess it on the identification dataset. The underlying convolutional layers of the system extricate highlight from the picture while the completely associated layers foresee the yield probabilities and directions. We likewise train a quick form of YOLO intended to push the limits of quick item location. Quick YOLO utilizes a neural system with less convolutional layers (9 rather than 24) and fewer channels in those layers. Other than the size of the system, all preparation and testing parameters are the equivalent among YOLO and Fast YOLO. The last yield of our system is the $7 \times 7 \times 30$ tensor [3], [9].

We reframe object identification as a solitary relapse issue, directly from picture pixels to jumping box arranges and class probabilities. Utilizing our framework, you only look once (YOLO) at a picture to foresee what articles are available and where they are. YOLO prepares on full pictures and straightforwardly enhances location execution. This bound-together model has a few advantages over conventional strategies for object identification [10]-[14].

We outline location as a relapse issue we need not bother with an unpredictable pipeline. We essentially run our neural system on another picture at test time to foresee identifications. Our base system runs at 45 casings for each second with no group handling and a quick form runs at in excess of 150 fps. This implies we can process spilling video continuously with under 25 milliseconds of dormancy. Besides, YOLO accomplishes more than double the mean normal accuracy of other constant frameworks.

Second, YOLO reasons all around about the picture when making expectations. Not at all like sliding window and locale proposition based procedures, YOLO sees the whole picture during preparing and test time so it encodes relevant data about classes just as their appearance. Quick CNN, a top location technique. Botches foundation fixes in a picture for objects since it cannot see the bigger setting. YOLO makes not exactly a large portion of the number of foundation mistakes contrasted with Fast CNN.

Third, YOLO learns generalizable portrayals of articles. At the point when prepared on normal pictures and tried on the work of art, YOLO outflanks top identification strategies like CNN by a wide edge. Since YOLO is exceptionally generalizable it is more averse to separate when applied to new areas or startling information.

In the implementation of this smart system the author uses 4 classes (vehicle types) to detect the type of vehicle, namely cars, trucks, buses, and motorcycles where the training process uses pretrained weight from Yolo and then runs in GPU mode with 200 epochs. During the training process will produce file weights that will be used to predict objects [15]-[17].

3. RESULTS AND DISCUSSION

That has been collected as many as 600 images are labeled in accordance with the name of the object in it, labeling the image using the label IMG software, then the results of labeling the image are saved with Pascal VOC format. To run the program, select the source video to be used, then draw the calculation area marked with a green line, then select the weights from the file generated from the training process and the configuration file. If the configuration is correct press the start button, the program will run the video and start classifying and counting the number of vehicles. Vehicles that are detected and pass through the calculation area will be marked with a red bounding box [18].

In Table 1, the results of the study using a CNN, with a video size of 640×480, were able to detect the presence of a car vehicle object by 88% for its success rate, while in a video with a size of 640×360, car vehicle objects were able to be detected by 83%. Likewise, for detecting truck vehicle objects, CNN works quite well with a 640×480 video where the results obtained are 100% successful in detecting a truck vehicle object compared to a 640×360 video which is only able to detect the success of the truck object by 90% [19], [20]. This is because the quality of the objects obtained must be of good quality to be detected well enough [21].

Table 1. Result automating vehicle detection

Video	Duration	Size	Actual	Detection results	Accuracy
CCTV Perintis Kemerdekaan Road, Banyumanik, Semarang	1 minute 9 seconds	640×480	Cars: 18	Cars: 16	Cars: 88%
			Trucks: 8	Trucks: 8	Trucks: 100%
			Bus: 0	Bus: 0	Bus: 0%
			Motor: 23	Motor: 22	Motor: 95%
Sample CCTV toll road	6 minutes 24 seconds	640×360	Cars: 159	Car: 132	Car: 83%
			Trucks: 20	Truck: 18	Truck: 90%
			Bus: 2	Bus: 2	Bus: 100%
			Motor: 3	Motor: 3	Motor: 100%

Figure 1 is a desktop application that is used to get automatic counting of vehicles passing on a road that has been installed a CCTV camera. Figure 2 shows the catch of a vehicle object that can be detected from an application that has been made using a CNN algorithm.

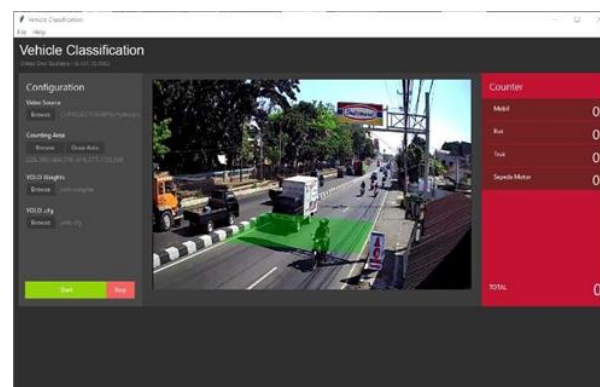


Figure 1. Calculation area

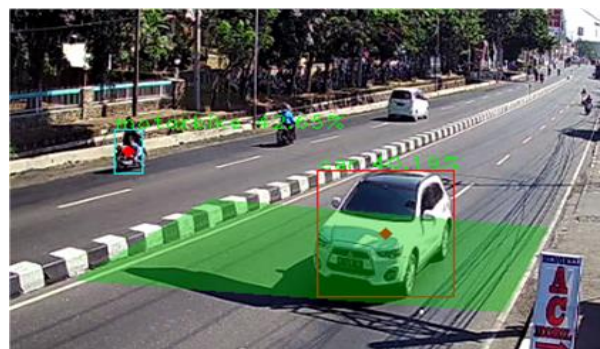


Figure 2. Bounding the box

YOLO commits far less foundation errors than Fast CNN. By utilizing YOLO to dispense with foundation detections from fast CNN we get a critical lift in execution. For each bouncing box that CNN predicts we verify whether YOLO predicts a comparable box. In the event that it does, we give that forecast a lift dependent on the likelihood anticipated by YOLO [3], [22]-[25].

4. CONCLUSION

We present YOLO and CNN, a bound-together model for object detection. Our model is easy to build and can be prepared straightforwardly on full pictures. Not at all like classifier-based methodologies, YOLO is prepared on a misfortune work that straightforwardly relates to identification. Execution and the whole model are prepared mutually. Quick YOLO is the quickest broadly useful article detector in the writing and YOLO pushes the best in class continuously object recognition. Range of values confidence generated is 40% to 90% depending on the point of view and clarity of the object. The accuracy value when testing gets a value between 80% to 95%.

REFERENCES

- [1] P. Newman, "Reducing automobile dependence," *Environ. Urban.*, vol. 8, no. 1, pp. 67-92, 1996, doi: 10.1177%2F095624789600800112.
- [2] A. Bull, *Traffic Congestion - The Problem and How to Deal with it?*, ECLAC, 2004.
- [3] G.-S. Peng, "Performance and Accuracy Analysis in Object Detection," p. 32, 2019.
- [4] Z. Q. Zhao, P. Zheng, S. T. Xu, dan X. Wu, "Object Detection with Deep Learning: A Review," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 30, no. 11, pp. 3212-3232, 2019, doi: 10.1109/TNNLS.2018.2876865.
- [5] R. Yamashita, M. Nishio, R. K. G. Do, dan K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights Imaging*, vol. 9, no. 4, pp. 611-629, 2018, doi: 10.1007/s13244-018-0639-9.
- [6] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol. Cybern.*, vol. 36, no. 4, pp. 193-202, 1980, doi: 10.1007/978-3-642-46466-9_18.
- [7] J. E. Espinosa Oviedo, "Detection and Tracking of Motorcycles in Urban Environments by using Video Sequences with High Level of Occlusion," Ph.D. thesis, Universidad Nacional de Colombia, 2019, doi: 10.13140/RG.2.2.29372.18563.
- [8] F. Shao, X. Wang, F. Meng, J. Zhu, D. Wang, dan J. Dai, "Improved Faster R-CNN Traffic Sign Detection Based on a Second Region of Interest and Highly Possible Regions Proposal Network," *Sensors (Basel)*, vol. 19, no. 10, 2019. <https://doi.org/10.3390/s19102288>.
- [9] J. Redmon, S. Divvala, R. Girshick, dan A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.
- [10] J. D. Owens, M. Houston, D. Luebke, S. Green, J. E. Stone, and J. C. Phillips, "GPU computing," in *Proceedings of the IEEE*, vol. 96, no. 5, 2008, doi: 10.1109/JPROC.2008.917757.
- [11] L. Z *et al.*, "Low-Grade Glioma Segmentation Based on CNN with Fully Connected CRF," *Journal of Healthcare Engineering*, vol. 2017, 2017, doi: 10.1155/2017/9283480.
- [12] F. Shao, X. Wang, F. Meng, J. Zhu, D. Wang, and J. Dai, "Improved Faster R-CNN Traffic Sign Detection Based on a Second Region of Interest and Highly Possible Regions Proposal Network," *Sensors (Basel, Switzerland)*, vol. 19, no. 10, 2019, doi: 10.3390/s19102288.
- [13] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611-629, 2018, doi: 10.1007/s13244-018-0639-9.
- [14] J. Xu *et al.*, "Short Text Clustering via Convolutional Neural Networks," in *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, pp. 62-69, 2015, doi: 10.3115/v1/w15-1509.
- [15] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," in *52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference*, vol. 1, pp. 655-665, 2014, doi: 10.3115/v1/p14-1062.
- [16] K. Cao and A. K. Jain, "Latent orientation field estimation via convolutional neural network," in *Proceedings of 2015 International Conference on Biometrics, ICB 2015*, no. i, pp. 349-356, 2015, doi: 10.1109/ICB.2015.7139060.
- [17] C. J. Huang and P. H. Kuo, "Multiple-Input Deep Convolutional Neural Network Model for Short-Term Photovoltaic Power Forecasting," *IEEE Access*, vol. 7, pp. 74822-74834, 2019, doi: 10.1109/ACCESS.2019.2921238.
- [18] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, dan A. Zisserman, "The Pascal Visual Object Classes (VOC) Challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303-338, 2010, doi: 10.1007/s11263-009-0275-4.
- [19] A. Li *et al.*, "Evaluating Modern GPU Interconnect: PCIe, NVLink, NV-SLI, NVSwitch and GPUDirect," *IEEE Transactions on Parallel and Distributed Systems*, vol. 31, no. 1, 2020, doi: 10.1109/TPDS.2019.2928289.
- [20] Y. Yuan, F. Shi, J. T. Kirby, and F. Yu, "FUNWAVE-GPU: Multiple-GPU Acceleration of a Boussinesq-Type Wave Model," *Journal of Advances in Modeling Earth Systems*, vol. 12, no. 5, 2020, doi: 10.1029/2019MS001957.
- [21] W. Chen, H. Huang, S. Peng, C. Zhou, and C. Zhang, "YOLO-face: a real-time face detector," *Visual Computer*, vol. 37, no. 4, 2021, doi: 10.1007/s00371-020-01831-7.
- [22] P. Ren, L. Wang, W. Fang, S. Song, and S. Djahel, "A novel squeeze YOLO-based real-time people counting approach," *International Journal of Bio-Inspired Computation*, vol. 16, no. 2, 2020, doi: 10.1504/ijbic.2020.109674.
- [23] S. Chaudhari, N. Malkan, A. Momin, and M. Bonde, "Yolo Real Time Object Detection," *International Journal of Computer Trends and Technology*, vol. 68, no. 6, 2020, doi: 10.14445/22312803/ijctt-v68i6p112.

- [24] J. Li, J. Gu, Z. Huang, and J. Wen, "Application research of improved YOLO V3 algorithm in PCB electronic component detection," *Applied Sciences (Switzerland)*, vol. 9, no. 18, 2019, doi: 10.3390/app9183750.
- [25] D. Xu and Y. Wu, "Improved YOLO-V3 with DenseNet for multi-scale remote sensing target detection," *Sensors (Switzerland)*, vol. 20, no. 15, 2020, doi: 10.3390/s20154276.

BIOGRAPHIES OF AUTHORS



Susanto is the Dean of the Faculty of Information and Communication Technology, University of Semarang, Indonesia. Susanto is a researcher in the field of website networking, information system and desktop applications.



Dimas Dwi Budiarmo is a student majoring in information technology with an information systems study program from the Faculty of Information and Communication Technology, University of Semarang, Indonesia.



Aria Hendrawan is a lecturer in informatics engineering at the Faculty of Information and Communication Technology, University of Semarang, Indonesia. Aria Hendrawan is a researcher in the field artificial intelligence, open source system, and data science.



Prind Triajeng Pungkasanti is a lecturer at the Information Systems Faculty of Information and Communication Technology, University of Semarang, Indonesia. Prind is a researcher in the field information system and expert system.