SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Journal Publications & Informatics Engineering Research
Volume 4, Number 1, October 2019
DOI : https://doi.org/10.33395/sinkron.v4i1.10247
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Video Surveillance System with a Deep Learning Approach

Puji Lestari
Universitas Prima Indonesia
Medan, Indonesia
lpuji29@gmail.com

David Hamonangan D.Manik
Universitas Prima Indonesia
Medan, Indonesia
davidhamonangan05@yahoo.co.id

Nurseve Lina Br Sihotang
Universitas Prima Indonesia
Medan, Indonesia
nursevelina@gmail.com

Amir Mahmud Husein
Universitas Prima Indonesia
Medan, Indonesia
amirmahmud@unprimdn.ac.id

Abstract— The application of in-depth learning methods has been successfully applied in computer vision task with the ability to learn the features of differences in real world images by directly from the original image by passing layer after layer to get the high dimensions image, in this study we applied the YOLO method approach with network adaptation features based on Darknet-53 on a video dataset recorded by the activities of University of Indonesia Prima (UNPRI) students with are conditions of video with different objects as a surveillance system, based on the results of research into object classification produces an overall accuracy of 93%, but for the classification of objects bikes, buses, and cars have the lowest accuracy of 30% for bikes, 54% of cars and buses by 40% so it is necessary to develop methods to improve accuracy.

Keywords—YOLO; Object detection; Computer Vision, video surveillance;

I. INTRODUCTION

In this decade, research in the field of computer vision is increasingly being proposed by researchers in applications in areas such as signature identification (Harahap, Husein, & Dharma, 2017) - (Husein & Harahap, 2017), face (Husein & Harahap, 2017) - (Wijaya, Husein, Harahap, & Harahap, 2017), object detection, vehicle detection, face detection, pedestrians, video surveillance and people and people (Saqib, Khan, Sharma, & Blumenstein, 2018).

The application of in-depth learning methods has been successfully applied in computer vision tasks with the ability to learn the features of differences in real world images by extracting directly from the original image by passing layer after layer to obtain high dimensions of the image (Lan, Dang, Wang, & Wang, 2018), such as segmentation, classification, detection and recognition. Object detection and recognition aim to detect and classify objects that can be applied to various fields such as face detection, humans, pedestrians, vehicles (Huang, Pedoeem, & Chen, 2018), intelligent transportation, medical diagnosis and medical supervision (Fu, Liu, Ranga, Tyagi, & Berg, 2017) with various models such as R-CNN (Girshick, Donahue, Darrell, & Malik, 2018), Fast R-CNN (Ross, 2015), Faster R – NC (Ren, He, & Girshick, 2017), Single Shot MultiBox Detector (SSD) (Fu, Liu, Ranga, Tyagi, & Berg, 2017), Deconvolutional Single Shot Detector (DSSD) (Ren, Zhu, & Xiao, 2018), SPP-NET (He, Zhang, Ren, & Sun, 2015), You Only Look Once (YOLO) (Redmon, Divvala, Girshick, & Farhadi, 2016), YOLOv2 (Liu, et al., 2016),dan YOLOv3 (Redmon & Farhadi, 2018)

The R-CNN model generates independent area category proposals which are included in the second module to extract feature vectors of fixed length from each region. R-CNN produces high computation with several SVM classifications in training (Redmon, Divvala, Girshick, & Farhadi, 2016). Fast R-CNN is an improvement of the R-CNN model by training the network using multi-task loss in one single training stage, simplifying learning to improve runtime efficiency by combining the proposed RPN and Fast R-CNN into one network by sharing convolutional features, Faster R -NCN enables an integrated learning-based object detection system to run at a near-time frame rate (Li, Liu, Zhao, Zhang, & He, 2018).

YOLOv3 proposed by (Redmon & Farhadi, 2018), is one of the fastest and most accurate methods of using deep convolutional neural networks by

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Journal Publications & Informatics Engineering Research
Volume 4, Number 1, October 2019
DOI : https://doi.org/10.33395/sinkron.v4i1.10247
e-ISSN : 2541-2019
p-ISSN : 2541-044X

having a certain level of invariance to transformations, deformations, and geometric lighting so that it effectively overcomes the difficulties caused by changes in the appearance of objects . In addition, feature descriptions can be built adaptively under training data, showing higher flexibility and generalization capabilities so that it is accurately applied to vehicle detection Lane Detection, Detection of Traffic Congestion (Chakraborty, et al., 2018), Vehicle classification (Chauhan, Singh, Khemka, Prateek, & Sen, 2019), Pedestrians (Liu, Chen, Li, & Hu, 2018).

The YOLO algorithm classifies and places objects in one step to get object positions and categories directly at the output layer, while YOLOv2 proposes a hangar box and trains the bounding box to find better box dimensions automatically using the K-Means method while YOLOv3 uses the network adaptation feature based on Darknet-53, and softmax loss in YOLO v2 is replaced by logistical loss so that it has the ability to detect small objects. In this paper we propose the YOLOv3 framework as a surveillance system to detect objects in video datasets recorded by the activities of University of Indonesia Prima (UNPRI) students with rare conditions of videos with different objects.

II. RESEARCH METHODS

A. Types of research

In this study using the type of verification research in which the dataset used is the result of recording the activities of Prima Indonesia University students recorded using video, then applied to the YOLOv3 framework for testing in the supervision system in real time. This type of research is useful for testing how far the goals that have been outlined are achieved or in line with expectations as well as standard theories. Verification research aims to test existing theories with the aim of creating new knowledge.

B. Time and Place of Research

The research took place in the laboratory of Prima Indonesia University with a period of 8 (eight) months starting from November 2018 until August 2019.

C. Working Procedures

The work procedure in the study is shown in the figure below:

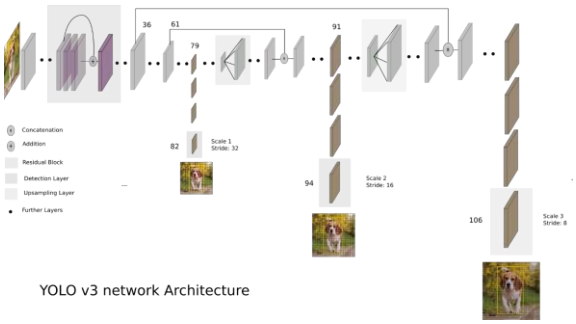

Figure 2.1 Work Procedure

III. RESULTS AND DISCUSSION

A. RESULTS

The results of tests conducted on the study consisted of several stages which included the stages of object recognition training in student activity videos recorded in various places with the number of video datasets used in the training totaling 15 videos with a size of 4.08 GB, then the object classification consisted of persons, car, bus, bike and others, then applied using the YOLO version 3 approach with the Yolo3.wigth architecture for feature extraction. Following is the display of the dataset used for training the proposed method.

Figure 3.1 Training Dataset

The results of the training on the unprivil student's active video will be evaluated to improve the test results by cutting the original video into several videos into 27 videos with a size of 3.44 GB, where the video distribution results will be evaluated so that the results of the focus video contain object data, this is useful to facilitate conduct a video evaluation of method testing. In table 3.1 The results of testing the introduction of objects in the video of student activities.

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Journal Publications & Informatics Engineering Research
Volume 4, Number 1, October 2019
DOI : https://doi.org/10.33395/sinkron.v4i1.10247
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 3.1 Testing Video Sizes

No	Video	Ukuran (MB)	
		Original	Hasil
1	Video 1 Depan Kampus Skip	253	469
2	Video 2 Depan Kampus Skip	177	343
3	Video 3 (lobi Kampus Skip)	148	450
4	Video 4 (lobi Kampus Skip)	173	250
5	Video 5 (Parkiran Kampus Skip)	56	176
6	Video 6 (Parkiran Kampus Skip)	9	16
7	Video 7 Parkiran Kampus Katamso	246	873

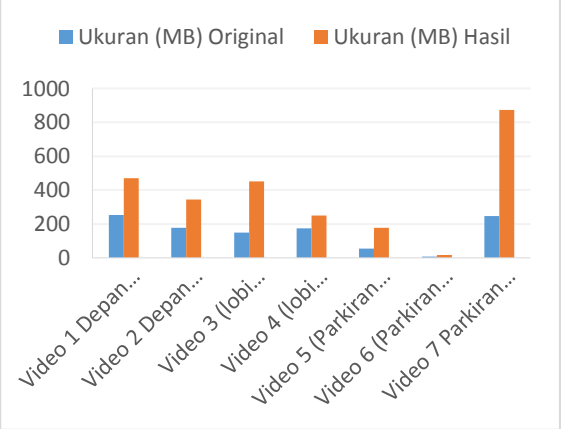

Figure 3.2 Size Comparison Chart

In Table 3.1 is the result of comparing the original video size with the object classification test results video, where the Yolo method applied re-making video by displaying object data known as person, car, bus, bike and other classifications as shown in Figure 3.3 and Table 3.2 Object Classification Results

Table 3.2 Object Classification Results

No	Video	Klasifikasi				mAP
		bike	person	car	bus	
1	Video 1 Depan Kampus Skip	36%	97%	50%	45%	94%
2	Video 2 Depan Kampus Skip	31%	97%	52%	43%	94%
3	Video 3 (lobi Kampus Skip)	30%	97%	52%	32%	93%
4	Video 4 (lobi Kampus Skip)	33%	98%	60%	34%	94%
5	Video 5 (Parkiran Kampus Skip)	35%	99%	60%	34%	92%
6	Video 6 (Parkiran Kampus Skip)	36%	95%	52%	45%	92%
7	Video 7 Parkiran Kampus Katamso	37%	96%	55%	48%	91%
Rata-rata hasil klasifikasi		**34%**	**97%**	**54%**	**40%**	**93%**

In table 3.2 is the result of object classification in the student activity video where the accuracy rate of persin object recognition is 97% with the highest value while the lowest value in the bus object classification is 40%, it shows that the proposed Yolo method still needs to be developed to identify bus objects, car. The overall result of the proposed method MAP is 93% for all objects identified in the video. The results in graphical form can be seen in Figure 3.3

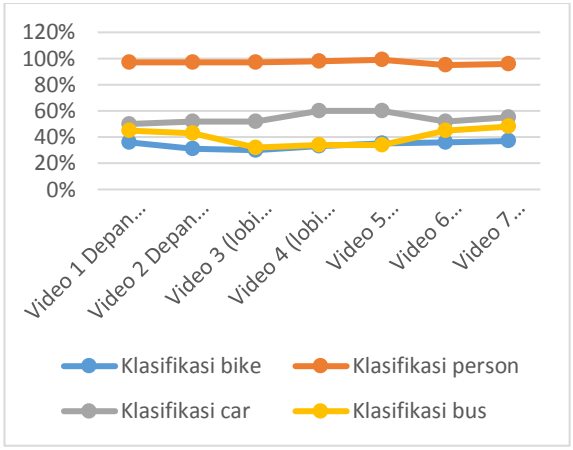

Figure 3.3 Graph of Object Classification Results

B. DISCUSSION

Based on the results of testing on the video of student activities recorded using the Yolo model deep learning method approach version 3, the results of the classification recognition of objects with criteria for person, car, bus, bike and others with an overall accuracy of 93%, but for bicycle, bus and object objects car has the lowest accuracy of 30% for bikes, 54% cars and 40% buses, some analysis is done because the size of the bike is too small, while the car and bus due to the same size, this can be used as part of developing the method to the next stage .

IV. CONCLUSIONS AND SUGGESTIONS

A. CONCLUSIONS

Based on the results of tests conducted using a deep learning approach with the Yolo model to identify objects in UNPRI student activity videos, several conclusions can be drawn .

1. The application of YOLO model version 3 which is used to classify objects derived from video recordings of student activities produces an overall 93% accuracy.
2. The classification of bicycle, bus, and car objects has the lowest accuracy of 30% for bicycles, 54% of cars and buses at 40% so it is necessary to develop methods to improve accuracy.
3. Accuracy results greatly affect the level of accuracy where videos with good quality with higher resolution levels will produce better classification accuracy results compared to low resolution videos.

B. SUGGESTIONS

Some suggestions proposed for further research development are:

Journal Publications & Informatics Engineering Research
Volume 4, Number 1, October 2019
DOI : https://doi.org/10.33395/sinkron.v4i1.10247
e-ISSN : 2541-2019
p-ISSN : 2541-044X

1. The Yolo method has a low level of accuracy to recognize bike, car and bus objects, so that it can be applied using other methods.
2. Video quality has a significant effect on the accuracy of object recognition, so other testing needs to be done with a focus on improving video quality before object recognition testing is performed.

REFERENCES

Chakraborty, P., Adu-Gyamfi, Y. O., Poddar, S., Ahsani, V., Sharma, A., & Sarkar, S. (2018). Traffic Congestion Detection from Camera Images using Deep Convolution Neural Networks. *Journal of the Transportation Research Board*, 222-231. doi:https://doi.org/10.1177/0361198118777631

Chauhan, M. S., Singh, A., Khemka, M., Prateek, A., & Sen, R. (2019). Embedded CNN based vehicle classification and counting in non-laned road traffic. *ArXIV*, 1-10.

Fu, C.-Y., Liu, W., Ranga, A., Tyagi, A., & Berg, A. C. (2017, January 23). *Archive*. Retrieved from Archive Cornell University: https://arxiv.org/abs/1701.06659

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2018). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *2014 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1-10). Columbus, OH, USA: IEEE.

Harahap, M., Husein, A. M., & Dharma, A. (2017). Identifikasi Tanda Tangan Dengan Kohonen Som Berbasis Principal Component Analysis. *Seminar Nasional APTIKOM (SEMNASTIKOM) 3*, (pp. 333-337). Medan.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (pp. 10-15). -: IEEE.

Huang, R., Pedoeem, J., & Chen, C. (2018). YOLO-LITE: A Real-Time Object Detection Algorithm Optimized for Non-GPU Computers. *IEEE International Conference on Big Data (Big Data)* (pp. 2503–2510). Seattle, WA, USA, USA: IEEE. doi:https://doi.org/10.1109/BigData.2018.8621865

Husein, A. M., & Harahap, M. (2017). Penerapan Metode Distance Transform pada Kernel Discriminant Analysis untuk Pengenalan Pola Tulisan Tangan Angka Berbasis Principal Component Analysis. *SinkrOn*, 31-36.

Husein, A. M., & Harahap, M. (2017). Pengenalan Multi Wajah Berdasarkan Klasifikasi Kohonen SOM Dioptimalkan dengan Algoritma Discriminant Analysis PCA. *Query: Journal of Information Systems*, 33-39.

Lan, W., Dang, J., Wang, Y., & Wang, S. (2018). Pedestrian Detection Based on YOLO Network Model. *IEEE International Conference on Mechatronics and Automation (ICMA)* (pp. 123-126). Changchun, China: IEEE. doi:https://doi.org/10.1109/ICMA.2018.8484698

Li, X., Liu, Y., Zhao, Z., Zhang, Y., & He, L. (2018). A Deep Learning Approach of Vehicle Multitarget Detection from Traffic Video. *Journal of Advanced Transportation*, 1-11. doi:https://doi.org/10.1155/2018/7075814

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). SSD: Single Shot MultiBox Detector. *European Conference on Computer Vision* (pp. 21-37). Amsterdam, The Netherlands: Springer.

Liu, Z., Chen, Z., Li, Z., & Hu, W. (2018). An Efficient Pedestrian Detection Method Based on YOLOv2. *Mathematical Problems in Engineering*, 1-10.

Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. New York, Ithaca, New York.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 112-128). Las Vegas, NV, USA: IEEE. doi:https://doi.org/10.1109/CVPR.2016.91

Journal Publications & Informatics Engineering Research
Volume 4, Number 1, October 2019
DOI : https://doi.org/10.33395/sinkron.v4i1.10247
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Ren, S., He, K., & Girshick, R. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (pp. 1137–1149). IEEE. doi:https://doi.org/10.1109/TPAMI.2016.2577031

Ren, Y., Zhu, C., & Xiao, S. (2018). Deformable Faster R-CNN with Aggregating Multi-Layer Features for Partially Occluded Object Detection in Optical Remote Sensing Images. *Remote Sensing*, 56-66.

Ross, G. (2015). Fast R-CNN. *IEEE International Conference on Computer Vision (ICCV)* (pp. 1-10). Santiago, Chile: IEEE.

Saqib, M., Khan, S. D., Sharma, N., & Blumenstein, M. (2018). Person Head Detection in Multiple Scales Using Deep Convolutional Neural Networks. *International Joint Conference on Neural Networks (IJCNN)* (pp. 10-15). Rio de Janeiro, Brazil: IEEE.

Wijaya, B. A., Husein, A. M., Harahap, M., & Harahap, M. K. (2017). Implementation Distance Transform Method in Kernel Discriminant Analysis for Face Recognition Using Kohonen SOM. *International Journal of Engineering Research & Technology (IJERT)*, 28-31.