Video Surveillance System with a Deep Learning Approach

Puji Lestari  
Universitas Prima Indonesia  
Medan, Indonesia  
lpui29@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 -NC 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
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:

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.weight architecture for feature extraction. Following is the display of the dataset used for training the proposed method.
Table 3.1 Testing Video Sizes

<table>
<thead>
<tr>
<th>No</th>
<th>Video</th>
<th>Ukuran (MB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Original</td>
<td>Hasil</td>
</tr>
<tr>
<td>1</td>
<td>Video 1 Depan Kampus Skip</td>
<td>253</td>
</tr>
<tr>
<td>2</td>
<td>Video 2 Depan Kampus Skip</td>
<td>177</td>
</tr>
<tr>
<td>3</td>
<td>Video 3 (lobi Kampus Skip)</td>
<td>148</td>
</tr>
<tr>
<td>4</td>
<td>Video 4 (lobi Kampus Skip)</td>
<td>173</td>
</tr>
<tr>
<td>5</td>
<td>Video 5 (Parkiran Kampus Skip)</td>
<td>56</td>
</tr>
<tr>
<td>6</td>
<td>Video 6 (Parkiran Kampus Skip)</td>
<td>9</td>
</tr>
<tr>
<td>7</td>
<td>Video 7 Parkiran Kampus Katamso</td>
<td>246</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>No</th>
<th>Video</th>
<th>Klasifikasi</th>
<th>mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>bike</td>
<td>person</td>
</tr>
<tr>
<td>1</td>
<td>Video 1 Depan Kampus Skip</td>
<td>36%</td>
<td>97%</td>
</tr>
<tr>
<td>2</td>
<td>Video 2 Depan Kampus Skip</td>
<td>31%</td>
<td>97%</td>
</tr>
<tr>
<td>3</td>
<td>Video 3 (lobi Kampus Skip)</td>
<td>30%</td>
<td>97%</td>
</tr>
<tr>
<td>4</td>
<td>Video 4 (lobi Kampus Skip)</td>
<td>33%</td>
<td>98%</td>
</tr>
<tr>
<td>5</td>
<td>Video 5 (Parkiran Kampus Skip)</td>
<td>35%</td>
<td>99%</td>
</tr>
<tr>
<td>6</td>
<td>Video 6 (Parkiran Kampus Skip)</td>
<td>36%</td>
<td>95%</td>
</tr>
<tr>
<td>7</td>
<td>Video 7 Parkiran Kampus Katamso</td>
<td>37%</td>
<td>96%</td>
</tr>
<tr>
<td></td>
<td>Rata-rata hasil klasifikasi</td>
<td>34%</td>
<td>97%</td>
</tr>
</tbody>
</table>

In Table 3.2 is the result of object classification in the student activity video where the accuracy rate of person 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

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:
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


