Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

Number plate recognition on vehicle using YOLO - Darknet

To cite this article: Budi Setiyono *et al* 2021 *J. Phys.: Conf. Ser.* **1821** 012049

Number plate recognition on vehicle using YOLO - Darknet

Budi Setiyono[1], Dyah Ayu Amini[2] and Dwi Ratna Sulistyaningrum[3]

[1,2,3)]Department of Mathematics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

Email : budi@matematika.its.ac.id

Abstract. Character recognition is one of the steps in the number plate recognition system. Character recognition is done to get text character data. The method used is YOLOv3 (You Only Look Once), and Darknet-53 is used as a feature extractor. In this study, the data used were number plate images derived from the extraction and cropping of motorized vehicle videos that had been taken using cellphones and cameras. Testing is done with two different models, namely the model obtained with additional preprocessing data and the model obtained without any preprocessing data. Data preprocessing is done to improve the quality of the number plate image. Testing is done on an uninterrupted number plate image dataset and a number plate image dataset with interference and a reduction of color intensity (brightness) in the image. For the model obtained from the data without preprocessing, the highest number plate recognition accuracy obtained is 80%, and the character recognition accuracy is 97.1%. Meanwhile, for the model obtained from preprocessing data, the highest number plate recognition accuracy obtained was 88%, and the character recognition accuracy was 98.2%.

1. Introduction

In the Smart Transportation System, one of the most important things is to identify the types of vehicles passing on the road. Video acquisition using CCTV that is already installed on the highway, is the easiest way to implement.[1][2]. A number plate is one type of motor vehicle identification. Number plates are also called vehicle registration plates. The form is a piece of metal or plastic plate attached to a motorized vehicle as official identification. Usually, a pair of number plates must be installed at the front and rear of the vehicle. Various methods of identifying vehicle license plates have been carried out. In general, these algorithms are developed from 3 steps, namely searching the number plate area, segmenting the characters from the license plate, and recognizing each character [4]. The vehicle number plate recognition system is an application that replaces the function of human vision in recognizing motorized vehicle license plates.

One of the critical stages in number plate recognition is the character recognition stage. In this character recognition stage, output in the form of text characters will be generated. The purpose of character recognition is to support the increasingly rapid development of technology in digital form. So that if there is a physical data that is desired to be a digital form, this character recognition system can be used [5].

2. Related work

The YOLO method was first introduced by Redmon et al. (2016) under the title "You Only Look Once: Unified, Real-Time Detection". In his research, besides its simple architecture, it is said that YOLO is very fast in identifying objects, and the average accuracy obtained is relatively high, with a percentage reaching 88% in ImageNet 2012 Validation [3].

In 2018 Rayson Laroca et al. conducted a research that applies the YOLO method to detect and recognize number plates. This research is entitled "A Robust Real-Time Automatic License Plate

Recognition Based on the YOLO Detector" [7]. This study discussed the problem of detecting and recognizing number plates with video as input data. There are several stages carried out in this study. The first stage is detecting vehicles and number plates using the Fast-YOLO and YOLOv2 methods, where Fast-YOLO requires faster detection time compared to YOLOv2. Meanwhile, for the accuracy obtained, Fast-YOLO has an accuracy of 97.3%, and YOLOv2 has an accuracy of 99%. The next step is character segmentation using the CNN method, followed by character recognition, which is also carried out using the CNN method.

Hendry and Rung-Ching Chen, in 2019, conducted a research using the YOLO method. The research that was conducted entitled "Automatic License Plate Recognition via sliding window darknet YOLO Deep Learning" [6]. This research discusses seven problems of number plate detection and recognition using the YOLO (You Only Look Once) method applied to license plate images. In this study, modifications were made to the YOLO network architecture and formed tinyYOLO with 13 convolutional layers. The accuracy obtained at the time of number plate detection reached 98.22%, while for number plate recognition, the accuracy obtained only reached 78%.

3. Proposed method

The primary process in this research is training and testing, as shown in Figure 1. researchers conducted training on 145 vehicle license plate images, which were obtained from the video. Furthermore, the researcher annotated by making bounding boxes and labeling each character. The number of characters is 801. The number of class labels is 36, namely from letters (A-Z) and numbers (0-9). The number of classes is thirty-six because it corresponds to 26 uppercase letters in the alphabet and ten numbers.

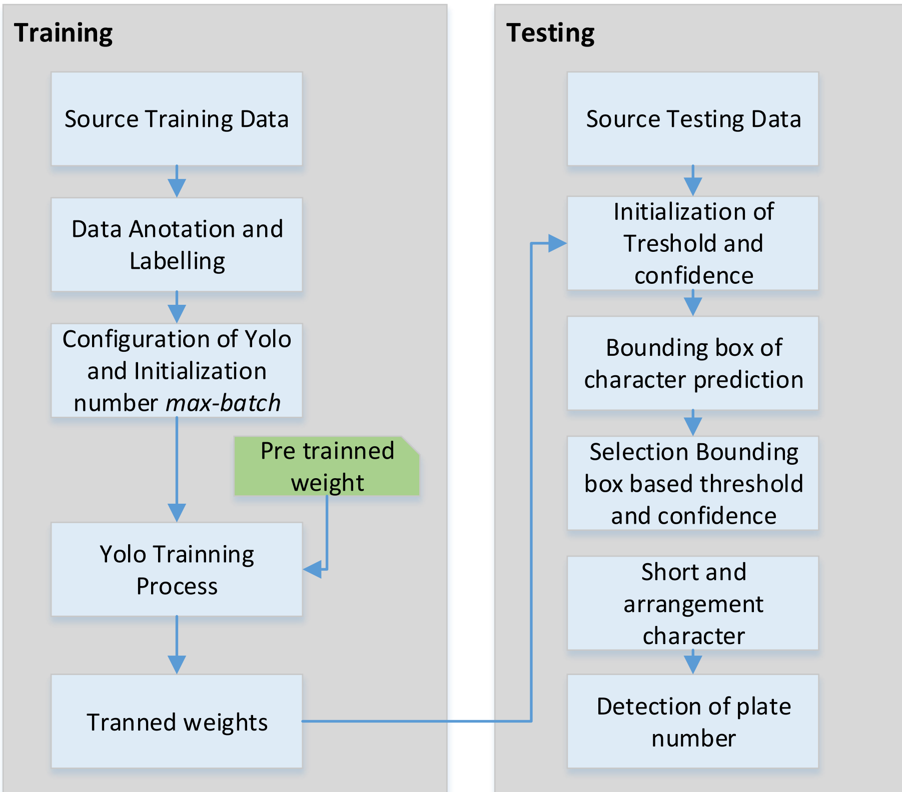

Figure 1. Proposed Method

Meanwhile, testing will use 25 vehicle license plate data consisting of data sets without noise and noise. 2 types of noise are given in the data set, namely the addition and reduction of color intensity noise in the image. This noise is given because there are frequent changes in lighting conditions in the field, usually in the form of brightness. In the following sections, we will describe the processes that will be carried out at each stage in detail.

3.1. Data Design

Recognition of objects based on digital images requires training data to obtain weights that will be used to recognize objects. Besides, data is also needed for the testing process to test the accuracy of the method. The data that has been collected in this study amounted to 170 data. The data consisted of number plate images from video extraction and cropping. Then it will be divided by a ratio of 85:15, which is 85% for the training process of 145 data and 15% for the testing process, which is 25 data.

Video of a speeding vehicle is taken manually using a cellphone and camera. The video was taken during the day with sunny conditions. The video was taken in a one-way lane with the vehicle approaching the camera. From the video obtained, it is necessary to extract and crop it to separate the background's number plate image. The number of plate images obtained is, on average, 80 × 20 pixels in size. Figure 2 is a number plate image obtained after extracting and cropping the video. It can be seen that the number plate size obtained is quite small and a little blurry.

Figure 2. Number plate from video extraction.

Some data are needed to be used for the training process. The number of training data used in this study was 145 number plate images detected and video cropping. The total number of characters on the number plate used for this training process is 801. Training data uses original data without preprocessing and training with data carried out by the preprocessing process first. Preprocessing that is done is applying filters to improve the image quality of the number of plates. To find out the distribution of the training data is used, a graph is displayed in Figure 3.

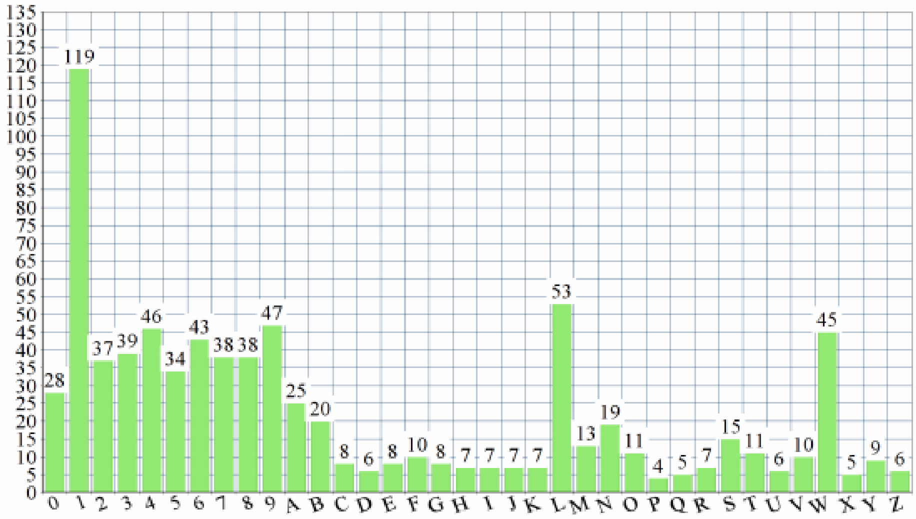

Figure 3. Distribution of training data

A total of 25 new data (other data that were not used for the training process) were used for the testing process. The data consists of a dataset without interruption and a dataset with disruption. There are two types of disturbances given to the dataset: the disturbance with the addition of color intensity and a reduction in the intensity of the image's color. Distraction is given on several levels. This process is done as a precaution when the captured data experiences lighting disturbances such as too dark or too bright. The undisturbed dataset is a dataset containing 25 image data, and the entire data is undisturbed. An image with an increase and decrease in color intensity noise is shown in Figure 4 and Figure 5.

Figure 4. Image without preprocessing with interference with the addition and reduction of color intensity

.

However, it should be noted that the test data used to test each model is different. The test data used is also without prior preprocessing for the model obtained from the training results using data without preprocessing. The model obtained from the training results using data by preprocessing; first, the test data used is also carried out by the preprocessing process. This is done to determine the accuracy of the model in detecting training data patterns.

Journal of Physics: Conference Series **1821** (2021) 012049 doi:10.1088/1742-6596/1821/1/012049

Figure 5. Image with preprocessing with the interference of increase and decrease of color intensity

3.2. Data Preprocessing

The data obtained from the video's extraction and cropping does not have a good quality, such as the image size are too small, and there is a blur effect on the image. So, it is necessary to preprocessing data to improve image quality. The preprocessing stage of this data consists of several steps, including resizing, namely changing the size of the data so that it is not too small, sharpening the image, removing noise, and reducing color intensity.

3.3. Data Annotation

The data annotation process begins by drawing a bounding box on each object in the image, which in this case, is a character, then saving the bounding box description in a file. The contents stored in the file are the values of c, (x, y), (w, h), respectively, the object class, the coordinates of the center point of the bounding box, and the bounding dimensions box. The description of the bounding box that has been made will be compared with the dimensions of the original size image; this aims to maintain the bounding box information in different image dimensions according to the proportions.

There are 36 object classes in the annotation, namely letters and numbers (A-Z and 0-9). Thirty-six classes were chosen because the system is used to recognize the number of plates. The data annotation process was carried out using Labelimg 1.7.0 software. This software was chosen because in labeling the stored data, it is under the YOLO format. An illustration of the data annotation is shown in Figure 6.

Figure 6. Annotate the data with Labelimg 1.7.0

3.4. Character Recognition using YOLO

3.4.1. You Only Look Once (YOLO)

YOLO is a Real Object Detection which has recently been very popular to develop. Most previous

Journal of Physics: Conference Series **1821** (2021) 012049 doi:10.1088/1742-6596/1821/1/012049

detection systems used a classifier or localizer to perform detection by applying the model to the image at multiple locations and scales and assigning a value to the image as a material for detection. YOLO uses a very different approach from the previous method, namely applying a single neural network to the entire image. This network will divide the image into regions then predict the bounding and probability boxes; for each bounding area box, the probability is weighted to classify as an object or not [3].

3.4.2. Network Architecture

YOLOv3 has 106 Convolutional layers that underlie the architecture followed by Leaky ReLU activation and batch normalize and shortcuts, routes, and upsample. Darknet-53 is used to perform feature extraction from image input. The network architecture of YOLOv3 is illustrated in Figure 7.

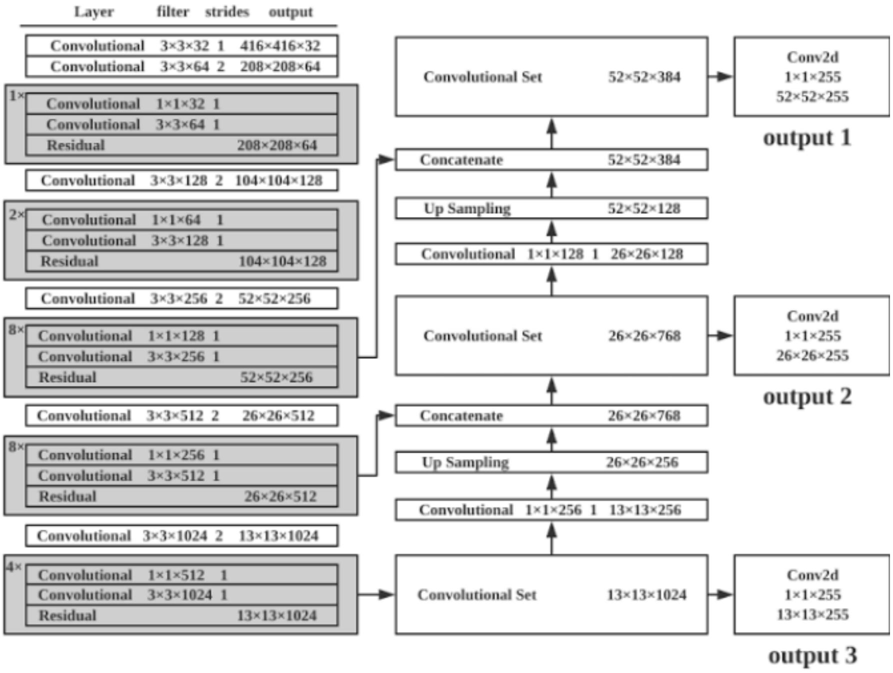

Figure 7. YOLOv3 network architecture.

3.4.3. Detection on Three Scales

Based on the YOLOv3 architectural design in Figure 6, it is known that the Yolov3 network performs detection by predicting at three different scales. The first scale occurs in the 82nd layer with a stride of 32. Since the input image dimensions are 416 × 416, the scale size for this layer is (13 × 13). The second scale occurs in the 94th layer with a stride of 16 so the scale size for this layer is (26 × 26). Meanwhile, the third scale occurs at the 106th layer with a stride 8. Then the scale size for this layer is (52 × 52)

.

3.4.4. Anchor Box

At YOLOv3, each scale will predict the object class using regression classification. The size of the anchor boxes used is obtained from data clustering using the k-means method. Each scale (13 × 13, 26 × 26, 52 × 52) on the pixels has three anchor boxes so that the total anchor boxes used are 9 with different sizes according to the clustering results of the data annotations. Each anchor box has 5 + C attributes, where the value of 5 represents five attributes, including Objectivity Score, Coordinate x midpoint of the bounding box, Y coordinate of bounding box midpoint, boundary box width, height bounding box, and object class score.

3.4.5. Non-Maxima Suppression (NMS) and Loss Function

Non-maxima Suppression (NMS) is used to solve overlapping bounding boxes in recognizing the same object. NMSs use an essential function called Intersection over Union, or IoU. The illustration of IoU will be shown in Figure 8. Loss Functions on YOLOv3 include Loss Function Localization.

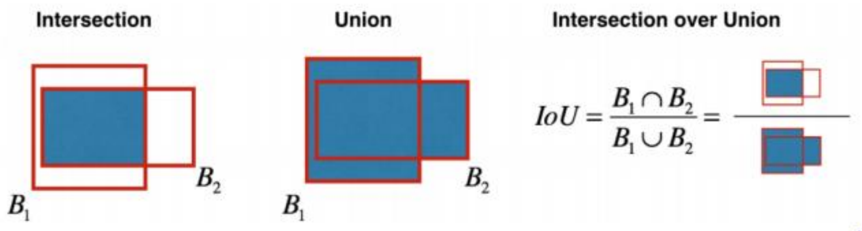

Figure 8. Intersection over Union.

3.5. Number Plate Recognition

The output of the applied YOLOv3 method is a bounding box for each character recognized by the class label and confidence value. The bounding box provides coordinate and class information of recognized characters on a number plate image.

After obtaining the bounding box with the class label and the recognized character's confidence value, the next step is to read the class label based on the bounding box coordinates from the top left to the bottom right. This is done because the desired result in this study is the recognition of number plates, but the results obtained are still in the form of character labels whose order is still random based on the recognized characters first.

4. Result and Discussion

4.1. Preprocessing Data

The quality of the data obtained from the video's extraction and cropping is not too good quality, such as the image size is too small, and there is a slight blur effect on the image. Because it has been applied several filters to improve image quality, the results of the application of the filter used can be seen in Figure 9. From this figure, it can be seen that the results of preprocessing data look quite good. The initially small image and has a blurry effect becomes clearer so that character recognition is easier to do.

Figure 9. Results of preprocessing data.

4.2. Max Batch

During the training data process, the value of the maximum batch used is determined. The maximum batch is the maximum number of iterations performed to obtain the testing process's weight. In this study, the maximum batch values used were 1000, 5000, 10000, 15000, 18000, 20000, and 25000. During the training data process, the maximum batch value greatly affected the computation time required. The greater the maximum batch value, the higher the computation time required. The display in the graph can be seen in Figure 10.

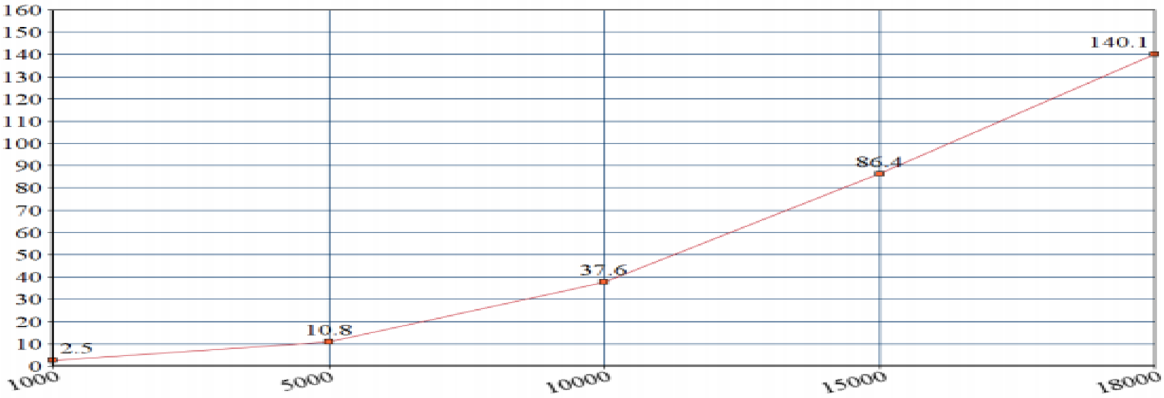

Figure 10. Graph of the effect of the maximum batch on computation time

In addition to the difference in computation time, the accuracy obtained from the difference in the maximum batch used is also different when the weight of the training results is used to test the uninterrupted image test data totaling 25 images. The display in the graph can be seen in Figure 11.

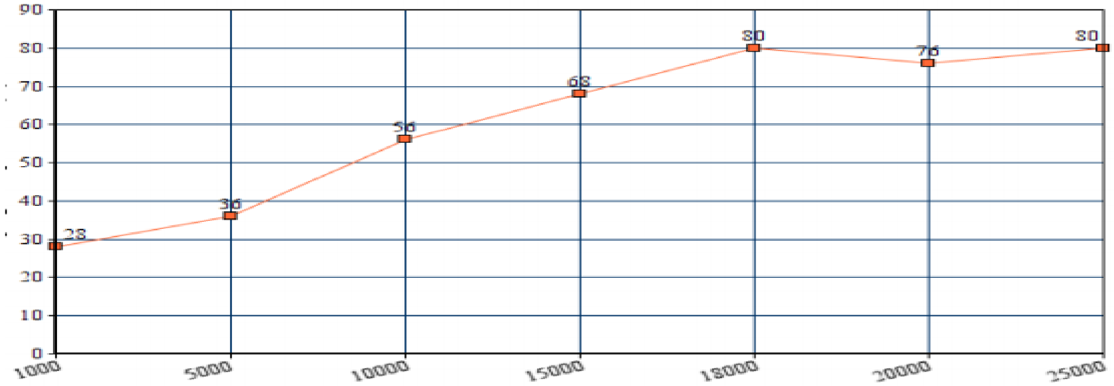

Figure 11. Graph of the effect of the maximum batch on accuracy

4.3. System Interface

To do experiments with ease, we create a GUI interface using tkinter Python to put all processes into one system. The GUI interface for the system is given in Figure 12.

Figure 12. GUI Interface

4.4. Accuracy

The calculation of accuracy is divided into two, namely calculating accuracy based on the number of recognized number plates and the calculation of accuracy based on the number of recognized characters. We have been carried out with data without preprocessing; the accuracy obtained is shown in Table 1, and based on the experiments that have been carried out with data without preprocessing, the accuracy obtained is shown in Table 2.

Table 1. Accuracy of the dataset without preprocessing.

Dataset without preprocessing	Accuracy of character recognition (%)	Accuracy of plate recognition (%)
Without distraction	97.1	80
Brightness (+25)	96.0	72
Brightness (+50)	94.2	60
Brightness (+75)	94.2	60
Brightness (-25)	96.0	78
Brightness (-50)	90.6	48
Brightness (-75)	85.3	36

Table 2. Accuracy of a dataset with preprocessing.

Dataset with preprocessing	Accuracy of character recognition (%)	Accuracy of plate recognition (%)
Without distraction	98.2	88
Brightness (+25)	97.6	88
Brightness (+50)	95.9	80
Brightness (+75)	94.1	68
Brightness (-25)	98.2	88
Brightness (-50)	97.7	84
Brightness (-75)	97.0	84

5. Conclusion

In this research, the YOLO-Darknet method has been successfully developed to identify motor vehicle license plates. The steps taken are data preparation, preprocessing, data annotation, training, and testing.

Without preprocessing, the undisturbed dataset yields the highest number plate recognition accuracy value of 80%, with a character recognition accuracy of 97.1%. For models obtained from data without preprocessing, a dataset with disturbed good effects or increased color intensity resulted in the highest number plate recognition accuracy value of 72%, with a character recognition accuracy of 96%. For models obtained from data without preprocessing, a dataset with disturbing dark effects or reduced color intensity resulted in the highest number plate recognition accuracy value of 76% with a character recognition accuracy of 96%. For the model obtained from the data by preprocessing, the undisturbed dataset resulted in the highest number plate recognition accuracy value of 88%, with a character recognition accuracy of 98.2%.

While for the model obtained from the data by preprocessing, a dataset with disturbed good effects or increased color intensity resulted in the highest number plate recognition accuracy value of 88%, with a character recognition accuracy of 97.6%. For the model obtained from the data by preprocessing, a dataset with disturbing dark effects or reduced color intensity produces the highest number plate recognition accuracy value of 88% with a character recognition accuracy of 98.2%

6. References

[1] B. Setiyono, et al,. 2020 *The Rain Noise Reduction using Guided Filter to Improve performance of Vehicle Counting*, International Journal of Innovative Computing, Information and Control Volume 16, Number 4, 1349-4198

[2] B. Setiyono, et al., 2019 *Multi Vehicle Speed Detection Using Euclidean Distance Based on Video Processing*, International Journal of Computing, 18(4), 431-442

[3] Redmon Joseph, Santosh Divvala, Ross Girshick, Ali Farhadi. 2016 *You Only Look Once : Unified, Real-Time Object Detection*, The University of Washington, Allen Institute for AI, Facebook AI Research.

[4] C. Anagnostopoulos, I. Anagnostopoulos, E. Kayafas, et. al., 2008 *License plate recognition from still images and video sequences*, IEEE Trans. Intell. Transp. Syst., vol. 9, no. 3.

[5] Cheriet, M, et al,. 2007 *Character Recognition Systems*, John Willey and sons inc. publication.

[6] Hendry, Rung-Ching Chen. 2019 *Automatic License Plate Recognition via sliding-window-darknet-YOLO Deep Learning* Image and Vision Computing 87, p 47-56.

[7] Rayson Laroca, Evair Severo, et al. 2018 *A Robust Real-Time Automatic License Plate*

Recognition Based on the YOLO Detector, International Joint Conference on Neural Networks (IJCNN)

Acknowledgment.
This work is partially supported by PDUPT with the title (Sistem Robus terhadap *Noise* pada *Tracking* dan *Counting* Kendaraan Bergerak Berbasis Pengolahan Citra Digital sebagai Pendukung Manajemen Transportasi Cerdas, contract number 1227/PKS/ITS/2020 Tanggal 20 Maret 2020), from the Ministry of Research, Technology and Higher Education of the Republic of Indonesia. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.