Automatic Detection of Helmets on Motorcyclists Using Faster - RCNN

Aliyyah Nur Azhari*[1], Wahyono[2],
[1,2]Department of Computer Science and Electronics, FMIPA UGM, Yogyakarta, Indonesia
e-mail: aliyyah.n.a@mail.ugm.ac.id, wahyo@ugm.ac.id,

Abstract

Motorcycles have been a popular choice for a go-to daily means of transportation due to its lower price, making it affordable for high to low-class citizens. Helmets are required for every motorcycle owner so that the rider's head is protected from accidents. However, not many people follow the rules and tend to not wear helmets and plenty of them underestimate the usage of helmets. For this, it is necessary to implement a system that can detect which rider wears the helmet or not by applying deep learning techniques. This paper aims to implement one of the deep learning techniques, which is Faster R – CNN to detect the helmets and the motorcyclists. After training 400 images using different learning rates, the mean average precision (mAP) achieved the highest with 87% using the learning rate of 0.0001

Key words*— Helmets, Motorcycles, Objdect detection, Deep Learning, Faster R - CNN*

1. INTRODUCTION

In most Asian and developing countries, two wheeled vehicles or mostly known as motorcycles has been a popular choice for a go-to daily mean of transportation [1]. There are a lot of reasons why motorcycles are preferable than four wheeled vehicles, mainly because motorcycles costs less, making it affordable for middle to low class citizens. Motorcycles tend to be smaller in size, meaning that bike riders are able to drive through narrow roads and can find ways to pass in between cars in traffic jam, which can beat traffic and can reach the destination without having to worry about time consumption. In addition to that, for safety reasons one of the requirement equipment for every bike rider must have is wearing a helmet. Motorcyclists, and even for regular riders need to take careful precautions to avoid their body from accidents, and the most crucial part to be aware of is from the head.

In other aspects, apart from safety reasons indeed one must follow several strict rules as it is already provided from the country's laws. In Indonesia, such laws are already regulated in Law No. 22 of 2009 regarding Road Traffic and Transportation, which one of them states that every person who rides a motorcycle and the passenger must wear a helmet that meets the Indonesian national standards (UU No. 22 Year 2009 Article 106 Paragraph 8). For anyone who violates the mentioned article, a fine will be imposed for as much as Rp. 250.000,00 (UU No. 22 Year 2009 Article 291 Paragraph 1 and 2) [2].

The World Health Organization (WHO), traffic accidents that occurs are a major public health problem and are being the lead in the cause of death and injury globally. Due to its small size, motorcyclists are more at a risk of experiencing an accident since they often pass fast-moving four wheeled vehicles, and without wearing a protection especially helmet, this can lead to serious problems both physically and financially. As there is a considerable rise of the number of motorcycles in Asia, the higher the chances are that head injuries take place especially in low to middle income countries. Several low-income and middle-income countries are estimated to be responsible for up to 88% regarding to head injuries fatality, and the

Received June 1st,2012; Revised June 25th, 2012; Accepted July 10th, 2012

percentage will keep on increasing if the bike riders are still less aware about the importance of wearing helmets [3].

Having to manually monitor motorcyclists who wears and does not wear a helmet regularly is a very insufficient method and it is impossible for any human to do an activity like so, even by using the surveillance cameras, it is require to have a human assistance for technical operations. To ease this problem, the presence of Intelligent Transportation System has been conducted based on computer vision fields, and it is one of the many applications of object detection and classification in machine learning. The rise of motorcycle accidents has encouraged many scientists to develop a system that can automatically detect motorcyclists with and without helmets by using several types of approaches. The approaches that have been commonly used by scientists are CNN models, Histogram of Gradients (HOG), Support Vector Machine (SVM), and K – Nearest Neighbors (KNN). Despite having a reliable accuracy percentage, each of these approaches yet still face the same difficulties to accurately identify motorcyclists due to natural challenges, such as occlusions, quality of video, and lighting issues.

According to [4], they used one of CNN's well – known model, VGG16, which is already pre-trained on the ImageNet dataset. To segment the moving objects, the Canny edge detection algorithm is applied with the addition of morphological operators to remove unnecessary parts such as noise. To build the classifier model, only the convolutional layers of VGG16 were taken. Once the object is classified as a motorcyclist, the head portion is extracted taking the upper on-third portion of the original helmet. Using VGG16 shows an almost perfect percentage for accuracy, precision, and recall by 99.5% for the motorcycle vs. non-motorcycle classifier, while for the helmet vs. non-helmet is 99% also, and 98.92% on the recall.

Author [5] uses the combination of four CNN models (VGG16, VGG19, GoogLeNet, and MobileNets) for classification in which these four models shows different accuracy percentage. It is shown that GoogLeNet and MobileNets shows better accuracy than the other two, having 84.58% and 85.19% respectively. For the detection method, the authors uses the two models from the previous step with the addition of Single Shot Multi-Box Detection (SSD) technique to the input video from a surveillance system. Training consist of collecting 493 images, and from these images, the two models successfully detect all 493 bikers from the video datasets and leave only zero undetected.

As CNN and YOLO has been conducted by previous researches, these authors [6] applied Faster R-CNN in which the system compromises two modules such as region proposal network (RPN) and fast RCNN. The RPN proposes at least 300 regions to detect the object boxes with a confidence score and to speed it up, region of interest (ROI) pooling layer is used. The motorcycle accuracy is quantitatively evaluated by mAP (mean Average Precision) in percentage, meaning that the average of a maximum precision at diverse recall altitudes. Subsequently, the helmet detection again uses Faster RCNN with VGG16 pre – trained model and softmax classifier. The authors conducted four different places for the datasets which contains side-view and front-view of the motorcycles and is surrounded by different natural phenomena, including shadows and occlusions.

Faster R-CNN is a state-of-the-art object detection architecture that is comprised of two modules; a region proposal network (RPN) and Fast R-CNN [7]. The RPN is a fully convolutional layer that predicts if there is an object and creates a bounding box of the detected objects, which takes from the output of a pre – trained CNN model, and the most commonly used is VGG-16 [8]. Faster R-CNN has a faster testing speed rate and a higher mean average precision compare to the other R-CNN models such as R-CNN and Fast R-CNN, meaning that creating a system of detecting motorcyclists wearing a helmet by using this state-of-the-art is efficient, as it requires less computational time.

2. METHODS

2.1 Data Preparation

This research uses 2 different datasets: The first is a custom dataset which contains a recorded public road somewhere in Jakarta [9], and the second is a CCTV footage of a public road as well provided by IITH India [10]. After achieving the datasets, the videos are separated into frames and are manually chosen that must includes the main objects, which is motorcycles, and the riders that wears the helmet or not. The total frames that were going to be experimenting is 500 images, and it is split 80% for training set and 20% testing set. The datasets are labeled according to its corresponding class: “Motorcycle”, “Helmet”, and “No_Helmet”. The Figure 1 shows the workflow for this research.

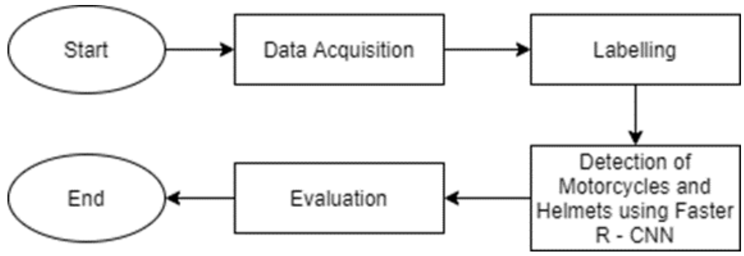

Figure 1 Flowchart of the main research

2. 2 Faster – RCNN Architecture

Faster R – CNN [7] will be applied to detect the motorcycles and helmets which compromises of two main models, the region proposal networks (RPN) and fast R – CNN. After achieving the frames from splitting the videos, these frames are passed towards the convolutional layers to achieve the feature maps. The convolutional layers uses VGG – 16’s architecture which contains 16 deep layers; first 13 layers are convolutional layers whereas the last three are fully connected layers. However, for this system the last fully connected layers are not applied since the feature maps are needed for the RPN layer to generate proposals.

The feature maps that are generated from the VGG – 16 model is then proceeded to the RPN layer to achieve region proposals. To define the anchors, The RPN proposes at approximately 300 regions using fully convolutional layers to detect the objects, the aspect ratios of RPN’s for the anchors in Faster R – CNN are 1:1, 1:2, and 2:1. The region proposals are proceeded to the RoI pooling to pool the feature maps into a fixed size since fully connected layers are required for the output layer to perform classification and bounding boxes. RoI pooling takes the regions from the feature maps of VGG16 and performs max pooling to achieve a fixed size, thus the output layer will have a size of $7 \times 7 \times 512 \times N$, having N the number of anchors. The RPN layer consist of 3 fully connected layer; the first has 4096 neurons and uses ReLU as the activation function, and the second and the third represents the cls (classification) and reg (regression).

For the output layer, it consists of the softmax function that classifies both motorcycles and helmets, and a linear regressor to adjust the position of the bounding boxes. If the RPN proposals overlaps, non-maximum suppression (NMS) is applied with a threshold of 0.7 on the proposal regions on the *cls* scores. The Figure 2 shows the flowchart of the proposed detection system.

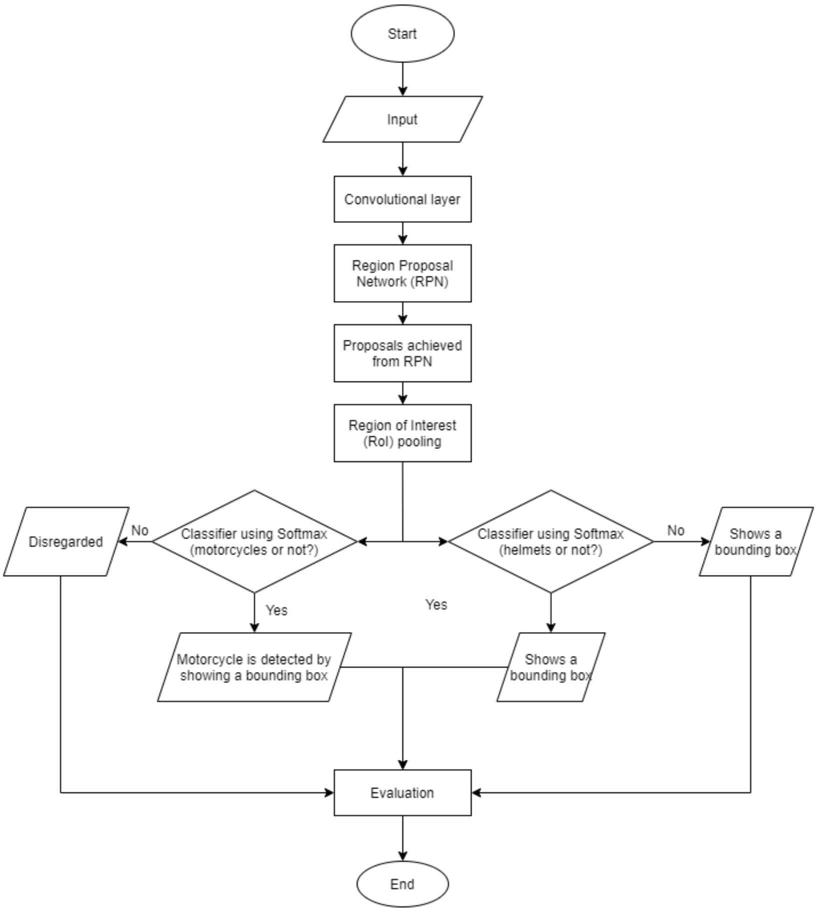

Figure 2 Flowchart of the proposed detection system

2. 3 Training and Testing

During the training process, some anchors will overlap the image's boundaries, or the cross boundary anchors occur. To avoid this, the Non Max Supression (NMS) is used with a threshold of 0.7 so that unwanted bounding bozes are removed [11]. Training is done by using the Adam optimizer with the learning rate of 0.0001 and 0.00001, with 700 epochs. The results are evaluated by comparing 300, 500, and 700 epochs respectively, measuring its accuracy, loss, and the mean average precision (mAP).

3. RESULTS AND DISCUSSION

After tuning the optimal hyperparameters from the Faster – RCNN model to detect the motorcycle rider who wears the helmet or not, the results from the training and testing are to be analyzed.

3. 1 Learning rate 0.0001 (1e-4)

3. 1.1 Results after training 300 epochs

Training after 300 epochs can be seen in the Table 1. Here, it is seen that the IITH dataset achieves higher results than the custom dataset, although the IITH dataset achieved a higher accuracy, it shows a relatively low mAP at 69.4%.

Table 1 Results of both dataset trained after 300 epochs

	Custom Dataset	**IITH Dataset**
Accuracy	81.70%	83.30%
Total Loss	1.29	0.86
Mean Average Precision	81.9%	69.4%

3. 1.2 Results after training 500 epochs

Training after 500 epochs can be seen in the Table 2. The accuracy for both of the dataset increased by comparing from the previous table, yet the total loss for the custom dataset slightly decreased by 0.35 while the IITH dataset decreased by 0.29, which is an indication that the model is performing better. The mAP for the custom dataset however dropped by 3%, contrariwise for the IITH dataset increased by 3.7%.

Table 2 Results of both dataset trained after 500 epochs

	Custom Dataset	**IITH Dataset**
Accuracy	85%	87.50%
Total Loss	0.94	0.57
Mean Average Precision	78.9%	73.1%

3. 1.3 Results after training 700 epochs

Training after 700 epochs can be seen in the Table 3. Again, the accuracy for both dataset has increased, but the total loss increased for the IITH dataset by 0.26. The mAP for both of the datasets decreased, and the detection turns to be overfitting since the mAP seems to be continuously decreasing. Overall, training with 700 epochs achieved a relatively high accuracy.

Table 3 Results of both dataset trained after 700 epochs

	Custom Dataset	**IITH Dataset**
Accuracy	88.30%	90.0%
Total Loss	0.68	0.83
Mean Average Precision	76.5%	64.7%

3. 2 Learning rate 0.00001 (1e-5)

3. 2.1 Results after training 300 epochs

After training 300 epochs, by comparing the results from using the learning rate of 1e-4, the accuracy tend to be higher, as shown in Table 4. The loss using learning rate of 1e-5 is lower compare to 1e-4, which indicates that the model's prediction performs better. Although, the mAP achieved a higher result for the custom dataset.

Table 4 Results of both dataset trained after 300 epochs

	Custom Dataset	**IITH Dataset**
Accuracy	87.50%	93.30%
Total Loss	0.61	0.48
Mean Average Precision	88%	63.9%

3. 2.2 Results after training 500 epochs

After training 500 epochs, the accuracy increased and is also higher compared to using the learning rate of 1e-4, as shown in Table 5. There is a major jump of the mAP result for the IITH dataset, increasing 15.5%.

Table 5 Results of both dataset trained after 500 epochs

	Custom Dataset	**IITH Dataset**
Accuracy	91.70%	95%
Total Loss	0.44	0.35
Mean Average Precision	91%	79.4%

3. 2.3 Results after training 700 epochs

Finally at 700 epochs, there has a slight difference from using the learning rate of 1e-5. Previously, training using 1e-4 tend to show that the accuracy increased as the epochs increases, but here, it does not follow that path. This may be because stopping the training session at a certain epoch may affect the weights from using the Google Colaboratory environment. The accuracy got lower for both datasets, and so does the mAP, yet the mAP for the IITH dataset does not change, nor stays the same at 79.4%, as shown in Table 6.

Table 6 Results of both dataset trained after 700 epochs

	Custom Dataset	**IITH Dataset**
Accuracy	88.30%	93.30%
Total Loss	0.59	0.33
Mean Average Precision	87%	79.4%

3. 3 Prediction results

The Figures 3, 4, and 5 are the predictions made from the Faster – RCNN model. The model successfully predicts the motorcycles, helmets, and non - helmets in its right location, but there are some predictions that did not predicted the objects. To cover that, though originally planned to train with 300 epochs only, the epochs were increased so that the false positives, false negatives, and true negatives were reduced.

Figure 3 Prediction from custom dataset (No helmet)

Figure 4 Prediction from custom dataset (Helmet)

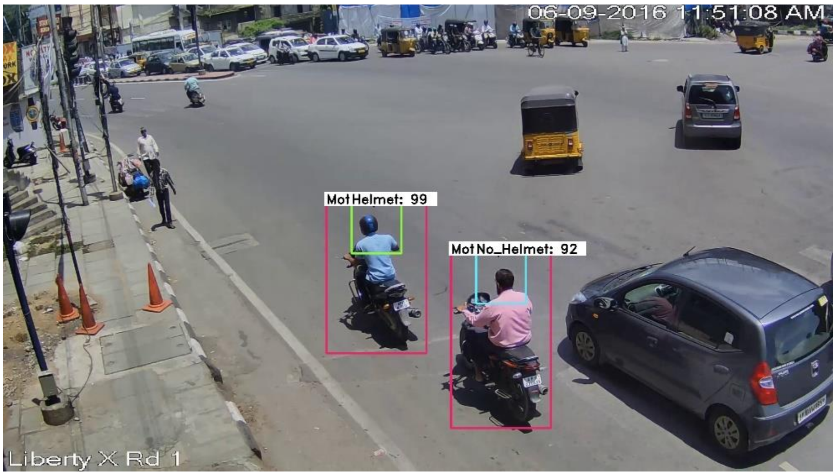

Figure 5 Prediction from IITH dataset

4. CONCLUSIONS

To conclude, the Faster R – CNN model shows a decent performance on detecting motorcycles, helmets, and non-helmets with the VGG – 16 network as the feature extraction. Using the learning rate of 0.0001, the custom dataset and the IITH dataset achieved the highest accuracy after being trained using 700 epochs by 88.30% and 90% respectively. Meanwhile, using 0.00001 for both dataset achieved the highest after being trained by 500 epochs at 91.70% and 90%. The mAP achieved at learning rate 0.0001 for the custom and IITH dataset are 76.5% and 64.7%, while 0.00001 are 87% and 79.4% respectively after being trained at 700 epochs. It shows that using the learning rate of 0.00001 has a better result than that of 0.0001, although changing the hyperparameters such as the anchor scales and anchor ratios can execute different values.

REFERENCES

[1] M. Zargar, A. Khaji, and M. Karbakhsh, “Pattern of motorcycle-related injuries in Tehran, 1999 to 2000: a study in 6 hospitals,” *Eastern Mediterranean Health Journal = La Revue De Sante De La Mediterranee Orientale = Al-Majallah Al-Sihhiyah Li-Sharq Al-Mutawassit*, vol. 12, no. 1–2, pp. 81–87, Jan. 2006.

[2] Constitution Law of Republic of Indonesia (UU No. 22 Year 2009) Article 291 Paragraph 1 and 2.

[3] World Health Organization, (2006). Helmets: A Road Safety Manual for Decision-makers and Practitioners. Available at: https://apps.who.int/iris/bitstream/handle/10665/43261/9241562994_eng.pdf

[4] Y. Kulkarni, S. Bodkhe, A. Kamthe, and A. Patil, “Automatic number plate recognition for motorcyclists riding without helmet,” *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, Mar. 2018, doi: 10.1109/icctct.2018.8551001.

[5] N. Boonsirisumpun, W. Puarungroj, and P. Wairotchanaphuttha, “Automatic Detector for Bikers with no Helmet using Deep Learning,” *2018 22nd International Computer Science and Engineering Conference (ICSEC)*, Nov. 2018, doi: 10.1109/icsec.2018.8712778.

[6] B. Yogameena, K. Menaka, and S. Saravana Perumaal, “Deep learning-based helmet wear analysis of a motorcycle rider for intelligent surveillance system,” *IET Intelligent Transport Systems*, vol. 13, no. 7, pp. 1190–1198, Jul. 2019, doi: 10.1049/iet-its.2018.5241.

[7] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, Jun. 2015, doi: 10.1109/tpami.2016.2577031.

[8] K. Simonyan, A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition”, arXiv:1409.1556v6, 10 Apr 2015, Available Online: https://arxiv.org/abs/1409.1556v6

[9] A.N. Azhari, “Automatic Detection of Helmets on Motorcyclists Using Faster Region Based Convolutional Neural Networks (Faster R – CNN)”, Undergraduate Thesis, Universitas Gadjah Mada, 2021.

[10] C. Vishnu; Dinesh Singh; C. Krishna Mohan; S. Babu, “Detection of motorcyclists without helmet in videos using convolutional neural network”, *2017 International Joint Conference on Neural Networks (IJCNN)*, 14-19 May 2017, doi: 10.1109/IJCNN.2017.7966233

[11] J. Hosang, R. Benenson, B. Schiele, “Learning non-maximum suppression”, arXiv:1705.02950v2, 9 May 2017 Available: https://arxiv.org/abs/1705.02950.