Automatic Detector for Bikers with no Helmet using Deep Learning

Abstract— **The success of digital image pattern recognition and feature extraction using a Convolutional Neural Network (CNN) or Deep Learning was recently acknowledged over the years. Researchers have applied these techniques to many problems including traffic offense detection in video surveillance, especially for the motorcycle riders who are not wearing a helmet. Several models of CNN were used to solve these kinds of problem but mostly required the image pre-processing step for extracting the Region of Interest (ROI) area in the image before applying CNN to classify helmet. In this paper, we proposed to apply another interesting method of deep learning called Single Shot MultiBox Detector (SSD) into helmet detection problem. This method is the state-of-the-art that is able to use only one single CNN network to detect the bounding box area of motorcycle and rider and then classify that biker is wearing or not wearing a helmet at the same time. The results of the experiment were surprisingly good. The classification accuracy of bikers not wearing a helmet was extremely high and the detection of the ROI of biker and motorcycle in the image can be done at the same time as the classification process.**

Keywords—Helmet Detection and Classification; Traffic Offence; Convolutional Neural Network; Deep Learning; Single Shot MultiBox Detector; Pattern Recognition; Object Detection; Machine Learning

I. INTRODUCTION

The importance of automatic system in traffic control has been increased in the recent year. One goal is to improve the utilization of a traffic flow system, others are to reduce the cost of human labor and decrease the causes of an accident. In Thailand, one major reason for the accident is the motorcycle biker who drive without wearing a helmet [1]. According to the law, every motorcyclist needs to wear a helmet while riding the motorcycle. But many bikers ignored and use their vehicle without safety equipment. The policeman tried to control this problem manually but it is insufficient for the real situation. The ideal solution is to develop an electronic detection system that can be automated recognize this kind of problem without human cost.

Building an automatic system like this bring researchers into areas of Image Processing (IP), Computer Vision (CV), and Artificial Intelligence (AI). Because most data from the traffic control system usually came in a format of video surveillance data (image and video) that require the technique to analyze an image data such as image recognition, pattern matching, and image segmentation. To detect the bikers who don't wear a helmet, we need methods to detect the photo of motorcycle and driver from the image and then detect an area of the biker head before classify that this person is wearing a helmet or not. Several IP and AI approaches have been using to solve this problem, for example, Fourier transformation [2], Support Vector Machine (SVM) [3], Histogram of Oriented Gradients (HOG) [4], Artificial Neural Network [5] etc.

But the advancement in another technique called a Convolutional Neural Network (CNN) has proved to be the better method in the area of image recognition and computer vision. One historical method is "AlexNet" developed in 2010 [6]. The model of Alexnet is a Deep Layer Convolutional Neural Network consisted of 650,000 neurons and 60 million parameters with five convolutional layers and 1000-way softmax layer. This model was challenged in the ImageNet Large-Scale Visual Recognition Challenge 2010 (ILSVRC10) and won the competition proved that CNN method will likely be the best technique to solve the most image recognition problem in this era.

After the success of AlexNet, several CNN models have been introduced and tried to achieve better performance than the pioneer one. For example, VGG [7], GoogLeNet or Inception [8], and MobileNets [9]. But most of these models can use only to categorize or recognize one object from the image not for multiple objects. Another technique needs to include into these models for adding image segmentation feature by drawing the box on the area of a possible object in an image before categorization. This combination help CNN model to be able to detect multiple objects in one single frame of the image. The examples of this approach are Faster R-CNN [10], Single Shot Multi-Box Detector (SSD) [11], and YOLO [12].

Our previous work had applied the combination of the SSD method and the image classification model such as GoogLeNet and MobilesNet on the Thai License Plate Recognition problem and received a good result on that problem. The accuracy of detection and classification of a Thai character on the Thai License Plate was more than 90% for both models (GoogLeNet and MobileNets) [13]. It gave us the confidence to take a further step to apply this technique in a helmet detection and classification issue.

In this paper, we proposed to solve the biker and helmet detection problem from video surveillance data by using CNN models and the SSD method. Some of CNN models have used in this experiment (VGG, GoogLeNet, and MobileNets) to compare the result. From our initial experiment, we found that the combination of MobileNets model and SSD has achieved the best accuracy compared to GoogLeNet and VGG in helmet detection problem and MobileNets was the method which requires the smallest size of the overall network.

II. Related Work

A. Motorcycle and Helmet Detection

Detection of motorcycle and helmet in an image is one of the challenging problems in the area of image processing. The issues are the shape of the object (motorcycle) in the image, the detection of people riding on a motorcycle or it just an empty vehicle with no biker, the location of the biker head, and the detection of a helmet at the head location of the biker.

Several steps of image processing needed to apply on the video image before it can detect the location of the motorcyclist with no helmet. For example on the previous work of P. Wonghabut et al (See Fig. 1), They needed to use several pre-processing techniques such as HAAR or HOG to detect the location of motorcycle in image first (step a-b), before cut off an area of a bikers in the image and classify that it is a motorcycle or not (step c-e), after that they need to find an area of the head location and cropped that area before they are able to detect that the biker is wearing a helmet or not (step f-h) [14].

But the advancement in CNN and SSD techniques have promised us that they are able to do all these steps in only one single runtime that we are explaining further.

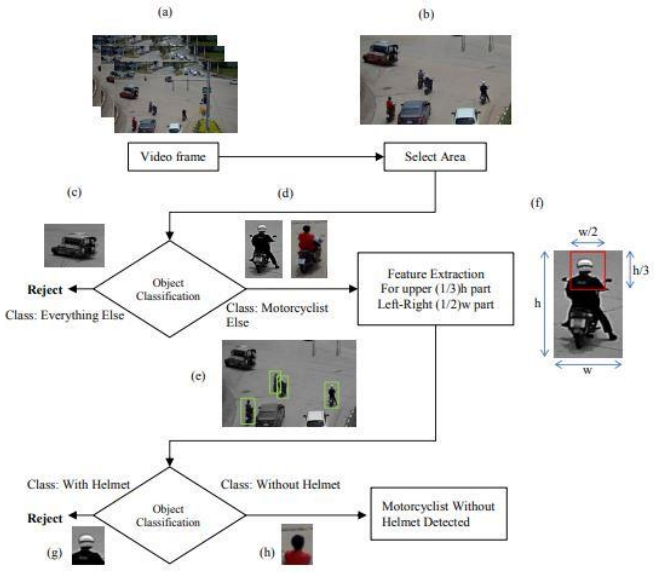

Figure 4. Diagram for detection of motorcyclist without helmet.
a) Input frame, b) Select Area A sample frame, c) Object classification as non motorcyclist, d) Object classification as Motorcyclist, e) Bounding box around Motorcyclist, f) Localized head of the Motorcyclist, g) Motorcyclist classified as 'with helmet' class, h) Motorcyclist classified as 'without helmet' class.

Fig. 1. Example of Previously Motorcycle and Helmet Detection Step [14].

B. VGG Net

The VGG network architecture was introduced by Simonyan and Zisserman in 2014 [7]. This model consists of 3×3 convolutional layers, max pooling, two fully-connected layers, and softmax classifier. Simonyan and Zisserman had shown that the performance of VGG network outperformed the other models who were the winner of ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) from the year 2012 and 2013.

The VGG model usually came in two different structures, VGG16 and VGG19. The number 16 and 19 means a number of weight layers in the network. VGG19 performed better than VGG16 because of a deeper layer but unfortunately increasing a lot of size for the whole network.

C. GoogLeNet (Inception V3)

GoogLeNet or Inception is the name of a CNN model created by Szegedy et al in 2014 [8]. The first version of Inception consisted of 22 layers neural network with the combination layers of convolution, max_pooling, fully connected, softmax, and a special layer called inception module. GoogleNet or Inception V1 was the winner on ILSVRC14 which was the competition on image recognition and the improved version came after year by year with a lot better performance. Now, version three or Inception V3, developed in 2016, has been the most used version for GoogLeNet models because its performance has surpassed its predecessor both in accuracy and time consuming [15].

D. MobileNets

MobileNets is another model of CNN which has proposed to decrease the size of the previous CNN model to make it available to use in a mobile platform. The idea is to replace the standard convolutional filters with two layers (Depthwise and Pointwise convolution) that build a smaller separable filter. The MobileNets network has achieved the good performance compare to another model and also come with the smallest size. Make it the good choice for the researcher to deal with the problem of large-scale data [9].

E. SSD (Single Shot Multibox Detector)

Single Shot Multibox Detector or SSD is the name of the deep neural network model that design to use only one single network to do both tasks of image recognition (image segmentation and image classification) in the same time [11]. An idea of SSD is to find the proper bounding box in each image that should be considered to be an object first, and then used that area of the bounding box to classify the type of object. SSD is the model that can combine its part with the structure of other networks such as GoogLeNet and MobileNets into one single network and makes it faster and easier to solve image recognition problem both in term of the position of an object in an image and the accuracy of object classification. Moreover, the best part is it can detect multiple objects in the image using only one runtime (See Fig. 2).

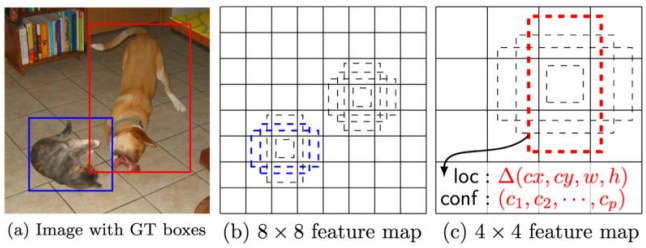

Fig. 2. Example of SSD multi-object detector [11].

III. Research Methodology

In this paper, we focus on applying four CNN models (VGG16, VGG19, GoogLeNet (Inception_v3), and MobileNets) to do an image classification experiment on our

input datasets about motorcyclist with a helmet and no helmet. And we apply one deep learning method for image detection (SSD) to do a biker with no helmet detection from the video image. This research has been carried out based on the four-step process: video and image gathering, image classification experiment, image detection experiment, and result interpretation.

A. *Video and Image Gathering*

Our input datasets were collected from the video surveillance system of Loei Rajabhat University in Loei province, Thailand. A camera we chose to start an experiment is the camera at the front gate of the university. We collected 50 videos of a vehicle passing through the gate, each video is 5 minutes long then the total of all videos length is 250 minutes (See Fig. 3).

Fig. 3. Examples of the input video from a front gate camera.

After that, we manually classify the image of a biker wearing a helmet and no helmet from the video data. Then we crop an area of a motorcycle with a biker and helmet into one image dataset call "Biker_with_helmet" and the area of a motorcycle with a biker who wears no helmet into another dataset call "Biker_with_no_helmet" (See Fig. 4). The total input image we have in "Biker_with_helmet" dataset is 336 images and for "Biker_with_no_helmet" we have 157 images. The total of them is 493 images.

Fig. 4. Examples of the cropped image from video dataset. (Left) Biker_with_no_helmet, (Right) Biker_with_helmet.

B. *Image classification experiment*

After gathering 493 images for our training dataset, we split our images into two groups, one for training data and another for test data to use in classification experiment. This experiment we test them with four CNN models for image classification (VGG16, VGG19, Inception V3, and MobileNets). For the evaluation, we used 10-fold cross-validation experiment which we set a number of test data for 10% of the total image. The training networks are trained using Python TensorFlow library, then we calculate the accuracy and choose two good models to use in image detection step.

C. *Image Detection Experiment*

In this step, we use all 50 videos that we collected to do image detection experiment using SSD technique combine with two CNN models we chose from the previous step. All videos will be tested and calculated the accuracy of the biker with helmet and no helmet detection in the video. We also count a number of undetected motorcyclists to be an error.

D. *Result Interpretation*

The last step, we compare the performance from two previous steps and make the conclusion. The accuracy of the experiments will show the performance of each technique in terms of image classification and image detection.

IV. Data Analysis and Results

For the image classification experiment. We calculate the accuracy for each model (VGG16, VGG19, Inception V3, and MobileNets). The overall results are shown in Table I.

TABLE I. Accuracy of Biker with helmet and No Helmet Classification

Network Model	Accuracy	Model Size (KB)
VGG16	78.09 %	434,580
VGG19	79.11 %	451,258
Inception V3	84.58 %	85,447
MobileNets	85.19 %	16,754

The result from Table I shows that MobileNets is the best CNN model to recognize biker with helmet and no helmet image sets. The accuracy of MobileNets is the highest (85.19%) follow by Inception V3 (84.58%). VGG16 is the lowest models compare to these four models with VGG19 slightly better than the VGG16 but both of VGG models generate the huge size of the network (> 400 MB) compare to Inception V3 and MobileNets. Then we decide to choose Inception V3 and MobileNets for the next experiment on video image detection with the SSD method.

A confusion matrix for the Inception_v3 and MobileNets for image detection with SSD are presented in Table II and III. From the result, we found that MobileNets also performs better than Inception V3 in terms of image detection. MobileNets detected 117 bikers with helmet and 304 bikers with no helmet correctly from overall 493 images. But Inception V3 can detect only 115 bikers with helmet and 301 bikers with no helmet. However, both models successfully detect all 493 bikers in the video datasets and leave only 0 biker that is undetected. The error of our image detection experiment found only on the misclassification issue but we have a 100% image detection accuracy.

TABLE II. ACCURACY OF BIKER WITH HELMET AND NO HELMET DETECTION ON VIDEO IMAGE (INCEPTION V3+SSD)

Inception V3	**Detected Class**		
Actual Class	**With Helmet**	**No Helmet**	**Undetected**
With Helmet	115	41	0
No Helmet	36	301	0

TABLE III. ACCURACY OF BIKER WITH HELMET AND NO HELMET DETECTION ON VIDEO IMAGE (MOBILENETS+SSD)

MobileNets	**Detected Class**		
Actual Class	**With Helmet**	**No Helmet**	**Undetected**
With Helmet	117	39	0
No Helmet	33	304	0

V. CONCLUSION

In this paper, we presented our experiment on applying some deep learning techniques to solve the issues of motorcyclists wearing a helmet and no helmet detection and classification. We used four Convolutional Neural Networks (CNN) in these experiments (VGG16, VGG19, GoogLeNet or Inception V3, and MobileNets) for image classification step and we also combine these models with the SSD technique to do an image detection step.

The results of our experiment were looking good. In the classification step, we found that MobileNets (85.19%) and Inception V3 (84.58%) achieved the better accuracy than VGG16 (78.09%) and VGG19 (79.11%). For the detection step, MobileNets is the winner which detected 421 (85.40%) correct motorcyclists class from the total number of 493 video images. Follow by Inception V3 which detected only 416 (84.38%) correct images. And the best part of these is the SSD technique can detect these images by using only one single runtime and require no other image pre-processing algorithms.

These results show us that a Deep Learning or CNN techniques are the good algorithms that we can apply on the problem of image detection and classification about bikers wearing a helmet or no helmet problem. In future work, we will expand our experiment by adding more CNN models or other Deep Learning techniques to compare with MobileNets and Inception V3 in term of image classification. And we also plan for adding other techniques such as Faster R-CNN or YOLO to compare with the SSD for the image detection experiment.

REFERENCES

[1] World Health Organization, "GHO by category road traffic deaths data by country", 2013, http://apps.who.int/gho/data/node.main.A997.

[2] R. Gonzales, R. Woods, "Digital Image Processing", Addison-Wesley Publishing Company, 1992, pp 81 - 125.

[3] Vapnik, V.N., "Statistical Learning Theory", Wiley, September 1998.

[4] Dalal, N. and Triggs, B., "Histograms of oriented gradients for human detection". In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. Vol. 1, pp. 886-893.

[5] Bishop, C.M., 1995. Neural networks for pattern recognition. Oxford university press.

[6] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[7] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition". arXiv preprint arXiv:1409.1556, 2014.

[8] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., 2015. Going deeper with convolutions (2014). arXiv preprint arXiv:1409.4842, 7.

[9] Howard, A.G., M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam. "Mobilenets: Efficient convolutional neural networks for mobile vision applications." arXiv preprint arXiv:1704.04861 (2017).

[10] Ren, S., He, K., Girshick, R. and Sun, J., "Faster r-cnn: Towards real-time object detection with region proposal networks". In Advances in neural information processing systems, pp. 91-99, 2015.

[11] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y. and Berg, A.C., 2016, October. "Ssd: Single shot multibox detector". In European conference on computer vision (pp. 21-37). Springer, Cham.

[12] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A., 2016. "You only look once: Unified, real-time object detection". In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

[13] Puarungroj W., and Boonsirisumpun N., "Thai License Plate Recognition based on Deep Learning", Procedia Computer Science, Vol. 135, pp. 214-221, 2018. DOI://10.1016/j.procs.2018.08.168

[14] Wonghabut, P., J. Kumphong, T. Satiennam, R. Ung-arunyawee, and W. Leelapatra. "Automatic helmet-wearing detection for law enforcement using CCTV cameras." In IOP Conference Series: Earth and Environmental Science, vol. 143, no. 1, p. 012063. IOP Publishing, 2018.

[15] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2818-2826).