Deep Learning and Image Processing for Disc Brake Pad Identification: A Case Study of Brake Pads Company

Nicha Khathinhorm[1] Rungrod Samankitesakul[2] Wasakorn Laesanklang[3]
Banpot Horbanluekit[4] Somkid Amornsamankul[5*]

[1,5*]School of Bioinnovation and Bio-based Product Intelligence, Faculty of Science, Mahidol University, Bangkok, Thailand
[2]Compact International (1994) Co., Ltd. (Head Office), Phetchaburi, Thailand
[3,5*]Department of Mathematics, Faculty of Science, Mahidol University, Bangkok, Thailand
[4]National Science and Technology Agency (NSTDA), Thailand Science Park, Pathum Thani, Thailand

*Corresponding Author. E-mail address: somkid.amo@mahidol.ac.th

Received: 21 March 2022; Revised: 25 April 2022; Accepted: 24 June 2022
Published online: 22 December 2022

Abstract

Disc brake pad identification is a difficult task that requires experiences of disc brake experts. However, disc brake pad retails must identify the part to deliver the right product to customers. In this research, the deep learning algorithms and object detection technologies to help identify disc brake pad in a case study of disc brake pads company are proposed. The goal is to implement disc brake identification system that finds the right brake pad model correctly in an instant time. We select two deep learning algorithms that are well known in object detection which are YOLOv5 and Faster R-CNN. The disc brake pad detection performance of the two algorithms is compared. There are four measurements: precision, detection speed, loss function (regression loss, classification loss), and training time. We use the two algorithms to detect and classify five disc brake pad models. The results show that YOLOv5 has better precision, detection speed, loss function, but Faster R-CNN requires less training time.

Keywords: Deep learning, Disc brake pad identification, Object detection, YOLOv5

I. INTRODUCTION

The global automotive aftermarket industry volume is increasing due to the rising demand of automobile [1]. Disc brake pad is the one of automotive aftermarket spare parts in which its demands are also increasing in the same trend with the demand of automotive market. Also, the modern vehicle models use disc brakes as the main vehicle braking system [2]. The increasing demand for automobiles affects to variety of spare parts feature (e.g., disc brake pad, car tyres, car accessories). Especially for disc brake pads, the variety of them is one of the concerns for the brake pad business due to the importance of identifying the disc brake pad that would be the information for installing disc brake pads in their cars correctly, which influences the brake system, brake performance, and safety of drivers.

A case study company in disc brake pad industry was tasked with identifying disc brake pad using human vision, which is well-known for disc brake pad drawing. But the variety of disc brake pad in terms of shape, size, and high volume caused the difficulty of identifying them with human vision precisely. Furthermore, there are a sizeable minority of experts who are well-known for their expertise in brake pad drawing. So, they could not identify a large number of disc brake pad fast enough. The constraint of human vision as above causes several problems for the brake pad business, such as slow response times for customer service and employee overload.

Nowadays, Object detection algorithms are getting better all the time. Object detection is a significant part of computer vision theory that is utilized in many practical image processing applications and is based on convolutional neural networks (CNNs) for extracting image features [3]. To identify each object, object detection is utilized to detect the single object or multiple objects and label with a class label in numerous photos [4]. There are two types of object detection model: one-stage and two-stage methods. The two-stage methods consists of region proposal followed by region classification, such as the R-CNN algorithm family, while one-stage object methods accomplishes both at the same time, such as SSD or YOLO series [5]. Furthermore, there are several libraries or application program interfaces (APIs) that can be used that are suitable for novice developers [6].

Due to the number of studies on disc brake identifying with object detection algorithm are rather tiny. The purpose challenge of this study is to recommend the best methods for identifying disc brake pads by comparing the efficiency of two methods, Faster R-CNN and YOLOv5 which were picked as the best representations of two-stage and one-stage object detection methods, respectively [7]. To achieve the fast and precision to identify disc brake pad, the performance evaluation consists of using the loss function, mean average precision (mAP), detection speed and training time.

II. LITERATURE REVIEW

Since the small number of research on disc brake identification using an object detection method. We investigated alternative object detection methods for other application to consider the optimal approach for detecting disc brake pads.

Sharma and Thakur [3] reviewed about object detection in deep learning field. Object identification is an important component of computer vision theory that is utilized in many practical image processing applications and is based on convolutional neural networks (CNNs) for extracting image features. In this topic, an object is characterized by its main characteristics, which include shape, size, color, texture, and other qualities.

Kumar and Srivastava [4] investigated the object detection system employing a single shot multi-box detector. To identify each object, object detection is utilized to detect the single object or multiple objects and label with a class label in numerous photos. They were training till the error rate was low and then testing the model with some example photos. The researcher discovered that precision could be checked using the loss function (LP), mean average precision (mAP), and frames per second (FPS).

Casado-Garcia and Heras [5] concluded that there are two types of object detection models. The first type is two-stage network, it consists of region proposal followed by region classification, such as the R-CNN algorithm family. The second type is one-stage network, it accomplishes both at the same time, such as SSD or YOLO. Furthermore, they found that the increasing cost of data collection means an increased cost too, to compensate for the lack of data, data augmentation techniques are employed.

Wang and Yan [7] researched a visual object detection which is the application in deep learning to classify types of tree leaf. In the research, YOLOv5 and Faster R-CNN were investigated as one-stage and two-stage detectors for detecting five different types of leaves. The results showed that YOLOv5 outperforms Faster R-CNN and flexibility for a variety of applications.

Turečková *et al.* [8] set a difficult situation of identifying a dog's face with YOLO network and deploy the system to mobile application for applying to open a pet entrance. At 0.5 IOU, YOLO tiny obtains an average precision of 92%. When compared to Faster RCNN-based, the average precision score of the YOLO tiny is 0.06 poorer, but it is much slightly speedier.

Ryu and Chung [9] investigated object detection to assess the surrounding situation rapidly and effectively in the driving environment. Recent research employed R-CNN to achieve precision, but they have constraints of real-time applying. This research proposed YOLO applying augmentation. The model has an 90.49% average precision. It was demonstrated that this method effectively identified objects and improved classification precision.

Yoon *et al.* [10] researched the classifying players and tracking ball movements in basketball game through video clips to help staff to design game strategies. The researchers utilized Yolo as the basic system, modifying it to identify player and ball motions from multiple perspectives. Yolo was chosen because of its ability to identify objects in real time. Furthermore, in terms of average precision and speed in frames per second, it outperformed other object identification algorithms such as Fast-RCNN and Faster-RCNN. The precision of detection is evaluated in this study by comparing it to human expert analysis.

Melek *et al.* [11] researched an object detection in shelf image to monitor the number of products on the shelves in retail. In terms of performance and speed, YOLOv2 was picked as the best deep learning algorithm. The total loss value is reduced as the number of iterations increases. Furthermore, the dataset is the most important component influencing the effectiveness of deep learning algorithm.

Bin Yan *et al.* [12] investigated a recognition system for recognizing existing apples on apple trees and classifying them as ungraspable apples or graspable apples. The study proposes YOLOv5 for real-time precise identification of apple picking robots in foggy and sunny conditions. The average identification time per image is 0.015 second and an average precision is 86.75%.

Ana Malta *et al.* [13] researched the recognized parts of an automobiles with YOLOv5 network. The dataset of car engines and the label of eight car parts were created by the researchers. The results in the test sets showed that YOLOv5 could recognize the car parts in real-time with high precision and recall above 96.8%. The results demonstrated that YOLOv5 is good and fast for detection problems.

Fangfei Shi *et al.* [14] proposed the object detection network, YOLOv5, to detect driver smoking. Since there is no public driver's smoking dataset, the researchers were created the dataset from video of various smoking with different angles, weather, and lighting conditions. The experiment results showed that decomposed YOLOv5 achieves the 93.5% precision and 4.9 second of detection time.

Md Jubayer *et al.* [15] researched an identifying mold on food surfaces such as fruit and bakery product based on YOLOv5 algorithm. In this research, a food images with mold dataset were produced. In compared to YOLOv3, YOLOv4, and YOLOv5, YOLOv5 model showed the highest precision, recall, and average precision with 205 epochs. Following that, as the number of epochs increased, the model's performance dropped.

Zixin Ning *et al.* [16] investigated MT-YOLOv5, a mobile terminal table detection model based on YOLOv5, that can identify the table location on the mobile terminal in real-time. The researchers created a dataset from a public source that included 1200 pictures for training and 439 images for testing. The results showed that YOLOv5 could determine the table location with a precision of 84% and it has a rapid reaction time of 24.2 frames per second, implying that real-time detection is possible.

Jiali Cui *et al.* [17] suggested a biometric identification approach based on a Faster Region based Convolutional Neural Network (Faster R-CNN) for detecting eyes. The approach is composed of three parts: convolutional layers, a region proposal network, and a detection network. Faster R-CNN is faster than other R-CNN series on a professional GPU, achieving 8-9 frames per second. Furthermore, there are comparison of test time and recall with YOLO (tiny), they found that the speed of YOLO (tiny) is better than Faster R-CNN, while the recall on the dataset with Faster R-CNN method may approach 95-96%, while YOLO (tiny) method achieved 47.4-63.5%, suggesting that the proposed technique based on Faster R-CNN has excellent precision for detecting eyes.

Wan and Goudos [18] investigated multiple fruit classes consisting of apples, mangoes, and orange detection using robotic vision with the development of Faster R-CNN. This study is a first in that it uses 4,000 images from the real world to construct a multi-labeled. In the experiment, they compared the Fast R-CNN, Faster-RCNN, YOLO series: YOLO, YOLOv2, and YOLOv3, and their own improved Faster R-CNN. The test results show that their own improved Faster R-CNN outperforms standard detectors in terms of detecting precision, 91% mAP, followed by YOLOv3, YOLOv2, Faster R-CNN, Fast R-CNN, and YOLO, respectively. While YOLOv3 has the fastest processing time, it is followed by YOLOv3, YOLO, Faster R-CNN, YOLO, Faster R-CNN, and Fast R-CNN, in that order.

Li [19] studied about object detection performance based on R-CNN series. This paper compared multiple pre-training models, to evaluate the efficiency of the R-CNN series. It was based on three different datasets in common public data, including PASCAL VOC2007, COCO, and ILSVRC. The results showed that Faster R-CNN outperformed R-CNN and Fast R-CNN in terms of precision, as measured by mAP, and detection speed, as measured by the second metric.

III. RESEARCH METHODOLOGY

We implement two deep learning algorithms for identifying disc brake pads, consisting of YOLOv5 and Faster R-CNN which are dominant in precision and detection speed [18], [19]. This section explains convolutional neural network (CNN) and the two algorithms: YOLOv5, and Faster R-CNN. Furthermore, this section also describes data preparation, deep learning environment and performance evaluation methods.

A. The Study of Convolutional Neural Network (CNN)

Convolutional neural network (CNN) [20] is a type of deep learning that has gained popularity in a variety of computer vision application. CNN is a mathematical structure consisting of three layers: convolution, pooling, and fully connected layers. The convolution and pooling layers perform feature extraction of object while the final output is extracted by the fully connected layers. Digital images consist of values of pixel in two-dimensional grid. The grid cells of image pixel are called kernel which represent the feature extraction for highly making performance of image processing.

B. The Study of YOLOv5 Principle

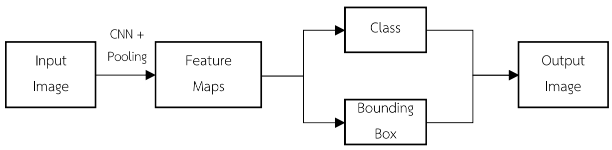

Figure 1: The work methodology of YOLOv5 algorithm

YOLO (You Only Look Once) [21] is the one of CNN, which can process image features and classify objects at the same time. There have been five versions of YOLO algorithms. Each version has been upgraded and has different architecture. For latest version of YOLO is YOLOv5 which has a strength of user-friendly framework, speed, and precision.

The work methodology of YOLOv5 [22] is to apply CNN to extract image features and use pooling to decrease size of feature maps. Then, it creates bounding boxes along with compute the confidence score of each bounding box to represent the object in the images. Finally, YOLOv5 generates the bounding box with highest confidence score and determines the class of objects as shown in Figure 1.

C. The Study of Faster R-CNN Principle

Faster R-CNN, or Faster Region-Based Convolutional Neural Network, [23] has two main steps for detecting and classifying object. The initial step is to propose regions using a CNN followed by the step of region proposal network (RPN) using to consider and categorize the object class. The adding of RPN is the key success that make Faster R-CNN be the best represent of CNN family.

The work methodology of Faster R-CNN [24] starts with the feature map creation using CNN to extract the features from input images. Then the region proposal network (RPN) will construct multiple bounding boxes on feature map known as regions of interest (ROI) that are highly probable to contain any objects. Finally, the ROI pooling layer is used to classify the class and generate the bounding box on output images as shown in Figure 2.

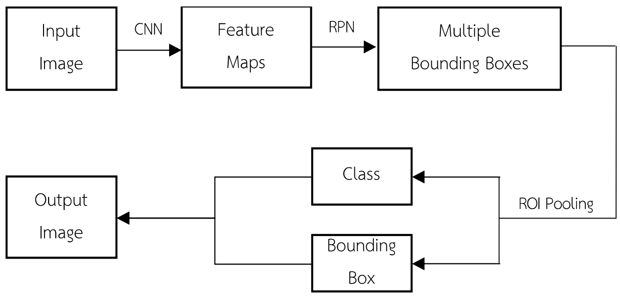

Figure 2: The work methodology of Faster R-CNN algorithm

D. Dataset Preparation

This step consists of three parts: collect images data, data annotation, and data preprocessing as follow:

1) Collect Images Data

All disc brake pad images were collected from a case study company. There are two sides of disc brake pad model consist of friction material side and backing plate side as shown in Figure 3. Images of both sides of each disc brake pad were collected with different arrangements. In addition, we collected disc brake pad images in visible shot sizes and clear surroundings, including a white background and sufficient light as shown in Figure 4. For this test dataset, we took image samples from five brake pad models: Model A, Model B, Model C, Model D and Model E.

Figure 3: Friction material side and backing plate side of disc brake pad

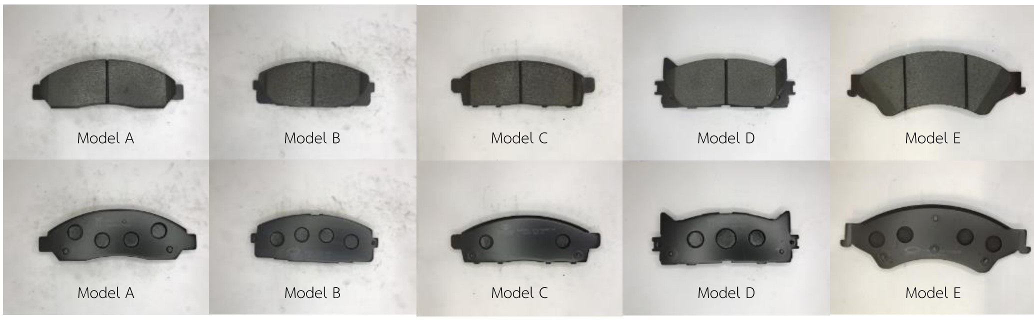

Figure 4: The image of disc brake pads collection: Model A, Model B, Model C, Model D and Model E

2) Data Annotation

The annotation process used Roboflow to label the class name of disc brake pads: Model A, Model B, Model C, Model D and Model E. Each class is represented of one to three car model names which use in automotive industry.

3) Data Preprocessing

We applied data augmentation to increase the number of instance in our dataset. This is essential part for preparing training dataset to improve the training model of deep learning algorithm. The data augmentation was created by Roboflow. It consists of flipping, rotating, shearing, grayscaling, brightening, exposuring, and blurring techniques. After data preprocessing was finished, the split of the training and validation dataset is 80% and 20% respectively as shown in table 1, which is suitable for multiple classes [25]. The number of samples for Model A, B, C, D and E are 651, 510, 665, 449 and, 458 respectively. Otherwise, the test dataset were collected from the single disc brake pad and multiple disc brake pads on a white background, as shown in Figure 5, to examine the results of two object detection methods.

Table 1: The image of test dataset collection

Algorithm	Train dataset (80%)	Validation dataset (20%)
YOLOv5	1,368	342
Faster R-CNN	1,368	342

Figure 5: The image of test dataset collection

E. Environment Setting and Training

After preparation of data, we trained both algorithms using Google Colab Pro. The epoch and batch size of train setting were defined by trial and error for receiving the best precision of both algorithms.

F. Performance Evaluation

The performance evaluation for comparing two algorithms consists of five parameters: confusion matrix, loss function, precision, detection speed, and training speed. All parameters are computed by algorithms at the end of training.

1) Loss function

The loss function [7] determines how close an approximated value is to the true value. It mainly includes classification loss and regression loss. The classification loss optimizes the classification performance while regression loss optimizes the bounding box that locates objects. So, in this research, we selected both loss functions to evaluate algorithms.

2) Precision

In part of the precision measurement, we selected mean average precision (mAP) [26] for estimating the precision of object detection. The value of mAP starts from 0 to 1 which means the greater value means the better precision. We select the mAP at IOU threshold at 0.5 to measure the algorithm precision in all classes and in each class. Since the good precision is the prioritization factor for classification problems. We also used a confusion matrix [27] to evaluate the quality of classification and show the cross of prediction and true classification. The confusion matrix is obtained from the k-nearest neighbors clustering where the one axis is defined as the target class or true class labels and another axis is the output class or class predictions.

3) Detection speed

The detection speed is measured by second per image to investigate speed performance and evaluate detection speed for practical use.

4) Training speed

The training time is measured as well, and it is computed in hours to compare the amount of time spent training between two algorithms.

IV. RESULTS AND DISCUSSIONS

A. Comparison of Disc Brake Pad Detection Results

The test datasets are separated in two types: single disc brake pad and multiple disc brake pads which were tested on YOLOv5 and Faster R-CNN algorithm with over 50 percentage confidence setting.

The disc brake pad detection on both YOLOv5 and Faster R-CNN consist of bounding box creation and class name identification, the experiment results in single disc brake pad detection as shown in Figure 6 showed that YOLOv5 could perform as good as Faster R-CNN, they could create the bounding box on object position precisely and identify the name of disc brake pad correctly in every class name with high confidence. While the experiment results in multiple disc brake pads detection as shown in Figure 7, showed that both YOLOv5 and Faster R-CNN create the bounding box on object position worse than single disc brake pad image because they created overlapped bounding box in some images. Since YOLOv5 created bounding box on empty background while Faster R-CNN created multiple bounding boxes on one object. But in term of class name identification, YOLOv5 could identify the name of disc brake pad correctly except the one which create the mistaken bounding box on empty background while Faster R-CNN was confused the name of disc brake pad in some images such as the wrong identify of Model A (50%) as Model C (96%) in Figure 7.

B. Comparison of Two Algorithms Performance

The confusion matrix as shown in Table 2 indicated that the performance of the identified disc brake pad of YOLOv5 is better than that of Faster R-CNN due to the higher precision of predicted name, while Faster R-CNN has high precision as well, but it was still confused in few class names such as the predicted name of Model A as Model B, C, and E.

In Figure 8 showed the mAP of disc brake pad identification in all classes and each class. In all classes, YOLOv5 has a 99.394% mAP while Faster R-CNN has 99.215% mAP. On the other hand, the mAP of YOLOv5 in each class is range from 99% to 99.5% while the mAP of Faster R-CNN in each class is range from 88.586% to 93.333%. The result of mAP showed that YOLOv5 had better average precision percentage than Faster R-CNN in each class which correspond to average precision of all classes. The investigation of detection speed after algorithms training in Figure 9 showed that YOLOv5 outperformed with 0.008 second/image while Faster R-CNN could detect 0.174 second/image. The result showed that YOLOv5 could detect disc brake pad image faster than Faster R-CNN about 22 times. On the other hand, in Figure 10 showed that the training time of Faster R-CNN is 1.3 hour while the training of YOLOv5 is 8.138 hour. The results showed that Faster R-CNN required less training time than YOLOv5 about 6 times.

For the loss function following Figure 11 and Figure 12 showed that YOLOv5 has 1.585% and 0.415% in regression loss and classification loss, respectively, which had better performance than Faster R-CNN.

Table 2: The confusion matrix of YOLOv5 and Faster R-CNN.

YOLOv5					
Predict / True	Model A	Model B	Model C	Model D	Model E
Model A	1.00				
Model B		1.00			
Model C			0.99	0.01	
Model D				1.00	
Model E					1.00

Faster R-CNN					
Predict / True	Model A	Model B	Model C	Model D	Model E
Model A	0.95	0.02	0.02		0.01
Model B		0.98	0.01	0.01	
Model C			1.00		
Model D				0.99	0.01
Model E	0.03				0.97

Figure 6: Single disc brake pads detection results

Figure 7: Multiple disc brake pads detection results

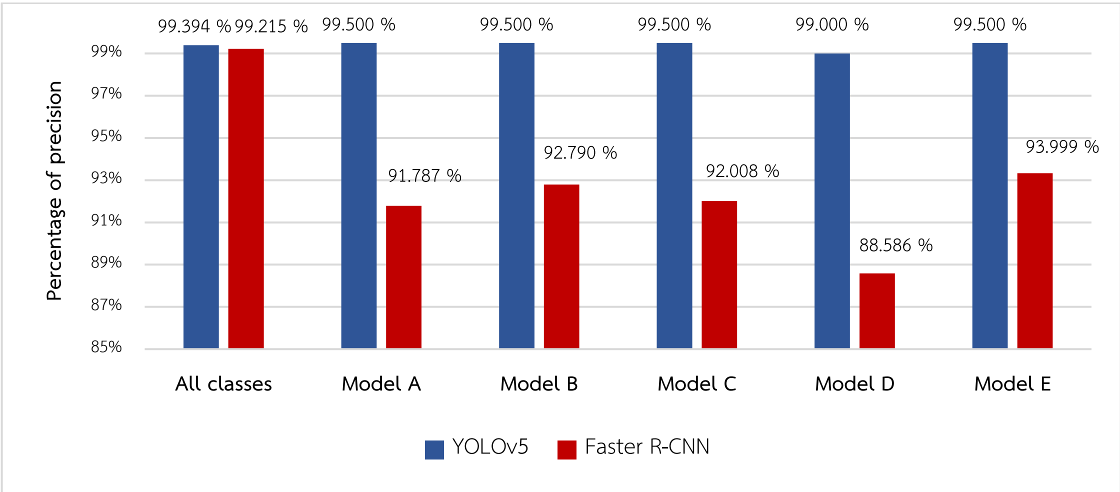

Figure 8: Precision comparison (mAP@0.5) of YOLOv5 and Faster R-CNN

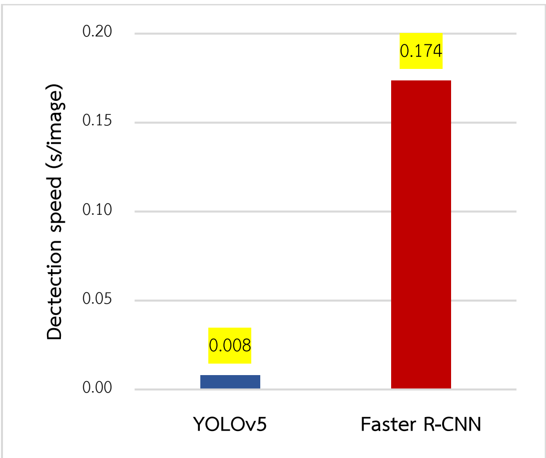

Figure 9: Detection speed (s/image) comparison of YOLOv5 and Faster R-CNN

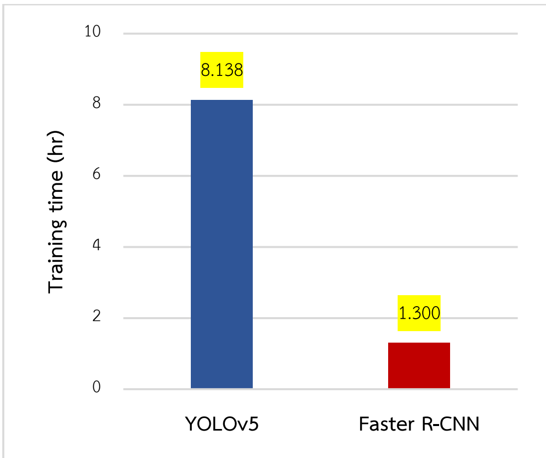

Figure 10: Training time (hr) comparison of YOLOv5 and Faster R-CNN.

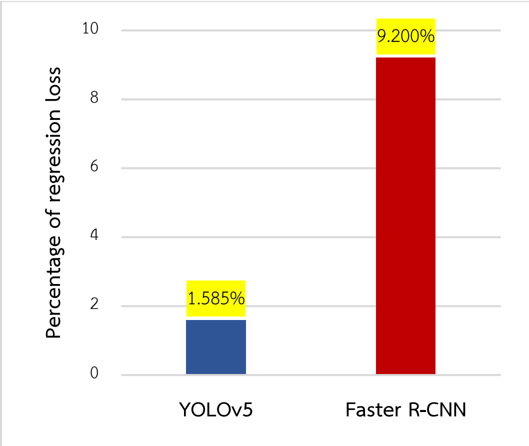

Figure 11: Regression loss comparison of YOLOv5 and Faster R-CNN

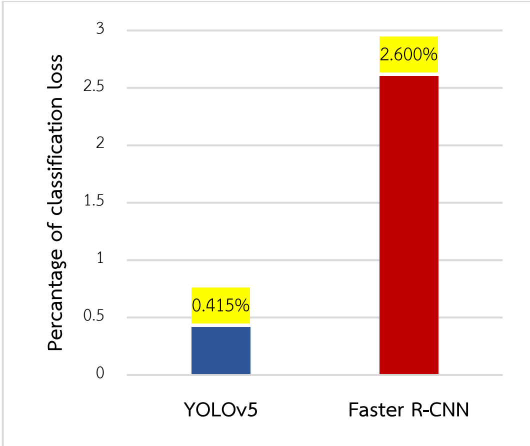

Figure 12: Classification loss comparison of YOLOv5 and Faster R-CNN

V. CONCLUSION

The experiment of algorithm performance comparison between YOLOv5 and Faster R-CNN for recommend the proper algorithm to identify disc brake pad is started from collect five models disc brake pad images in a case study company and pre-processing all images before training both algorithms. The results of training and testing determined that YOLOv5 has better precision, detection speed and loss function. Even though Faster R-CNN required less training time, YOLOv5 still significantly outperformed Faster R-CNN in disc brake pad identification.

By the way, the applying algorithm to identify disc brake pad in practical use should be concern about the quantity of models in disc brake industry and various background. So, the increasing of disc brake pad models and sufficient images background for real work coverage are essential tasks to develop the algorithms for using in all kinds of disc brake pad identification applications. Furthermore, the establishment of cross validation, which helps developers select the proper training data and the consider of balance of sample for each class, are a necessary task as well.

REFERENCES

[1] J. Laborda and M. J. Moral, “Automotive aftermarket forecast in a changing world: The stakeholders’ perceptions boost!,” *Sustainability*, vol. 12, no. 18, p. 7817, 2020.

[2] A. Borawski, “Conventional and unconventional materials used in the production of brake pads – review,” *Sci. Eng. Compos. Mater.*, vol. 27, no. 1, pp. 374–396, 2020, doi: 10.1515/secm-2020-0041.

[3] K. U. Sharma and N. V. Thakur, “A review and an approach for object detection in images,” *Int. J. Comput. Vision Robot.*, vol. 7, no. 1/2, pp. 196–237, 2017, doi: 10.1504/ijcvr.2017.081234.

[4] A. Kumar and S. Srivastava, “Object detection system based on convolution neural networks using single shot multi-box detector,” *Procedia Comput. Sci.*, vol. 171, pp. 2610–2617, 2020.

[5] Á. Casado-Garcia and J. Heras, “Ensemble methods for object detection,” in *Proc. 24th Eur. Conf. Artif. Intell.*, Santiago de Compostela, Spain, Jun. 2020, pp. 2688–2695.

[6] M. Noman, V. Stankovic, and A. Tawfik, “Object detection techniques: Overview and performance comparison,” in *Proc. IEEE 19th Int. Symp. Signal Process. and Infor. Technol. (ISSPIT)*, Ajman, United Arab Emirates, Dec. 2019, pp. 489–493.

[7] L. Wang and W. Q. Yan, “Tree leaves detection based on deep learning,” in *Geometry and Vision* (Communications in Computer and Information Science vol. 1386), M. Nguyen, W. Q. Yan, and H. Ho, Eds., Cham, Switzerland: Springer, 2021, pp. 26–38.

[8] A. Turečková, T. Holík, and Z. Komínková Oplatková, “Dog face detection using YOLO network,” *Mendel*, vol. 26 no. 2, pp. 17–22, Dec. 2020.

[9] S.-E. Ryu and K.-Y. Chung, “Detection model of occluded object based on YOLO using hard-example mining and augmentation policy optimization,” *Appl. Sci.*, vol. 11, no. 15, p. 7093, Jul. 2021, doi: 10.3390/app11157093.

[10] Y. Yoon *et al.*, “Analyzing basketball movements and pass relationships using realtime object tracking techniques based on deep learning,” *IEEE Access*, vol. 7, pp. 56564–56576, 2019, doi: 10.1109/ACCESS.2019.2913953.

[11] C. G. Melek, E. B. Sonmez, and S. Albayrak, “Object detection in shelf images with YOLO,” in *Proc. IEEE EUROCON 2019 - 18th Int. Conf. Smart Technol.*, Novi Sad, Serbia, Jul. 2019, pp. 1–5.

[12] B. Yan, P. Fan, X. Lei, Z. Liu, and F. Yang, “A real-time apple targets detection method for picking robot based on improved YOLOv5,” *Remote Sens.*, vol. 13, no. 9, p. 1619, Apr. 2021, doi: 10.3390/rs13091619.

[13] A. Malta, M. Mendes, and T. Farinha, “Augmented reality maintenance assistant using YOLOv5,” *Appl. Sci.*, vol. 11, no. 11, p. 4758, May 2021, doi: 10.3390/app11114758.

[14] F. Shi, H. Zhou, C. Ye, and J. Mai, “Faster detection method of driver smoking based on decomposed YOLOv5,” *J. Phys.: Conf. Ser.*, vol. 1993, 2021, Art. no. 012035, doi: 10.1088/1742-6596/1993/1/012035.

[15] F. Jubayer *et al.*, “Mold detection on food surfaces using YOLOv5,” *Curr. Res. Food Sci.,* vol. 4, pp. 724–728, 2021.

[16] Z. Ning, X. Wu, J. Yang, and Y. Yang, “MT-YOLOv5: Mobile terminal table detection model based on YOLOv5,” *J. Phys.: Conf. Ser.*, vol. 1978, 2021, Art no. 012010, doi: 10.1088/1742-6596/1978/1/012010

[17] J. Cui, F. Chen, D. Shi, and L. Liu, “Eye detection with faster R-CNN,” in *Proc. Int. Conf. Adv. Comput. Technol., Inf. Sci. and Commun.*, Xiamen, China, Mar. 2019, pp. 111–116.

[18] S. Wan and S. Goudos, “Faster R-CNN for multi-class fruit detection using a robotic vision system,” *Comput. Netw.*, vol. 168, Feb. 2020, Art. no. 107036.

[19] W. Li, “Analysis of object detection performance based on Faster R-CNN,” *J. Phys.: Conf. Ser.*, vol. 1827, 2021, Art no. 012085, doi: 10.1088/1742-6596/1827/1/012085.

[20] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, “Convolutional neural networks: an overview and application in radiology,” *Insights Imaging*, vol. 9, pp. 611–629, Jun. 2018.

[21] D. Thuan, “Evolution of YOLO algorithm and YOLOV5: the state-of-the-art object detection algorithm,” B.S. thesis, Dept. Inf. Technol., Oulu Univ. Appl. Sci., Oulu, Finland, 2021. [Online]. Available: https://www.theseus.fi/handle/10024/452552

[22] P. Punyato and N. Sidahao, “Implementation of a low-cost real-time people counting system on raspberry pi based on applied tiny YOLO V3,” (in Thai), *Eng. Trans.: A Res. Publication Mahanakorn Univ. Technol.*, vol. 22, no. 2, pp. 72–78, 2019.

[23] J. Du, “Understanding of object detection based on CNN family and YOLO,” *J. Phys.: Conf. Ser.*, vol., 1004, 2018, Art. no. 012029, doi: 10.1088/1742-6596/1004/1/012029.

[24] X. Yao, X. Wang, S. H. Wang, and Y. D. Zhang, “A comprehensive survey on convolutional neural network in medical image analysis,” *Multimed. Tools Appl.*, 2020, doi: 10.1007/s11042-020-09634-7.

[25] A. Rácz, D. Bajusz, and K. Héberger, “Effect of dataset size and train/test split ratios in QSAR/QSPR multiclass classification,” *Molecules*, vol. 26, no. 4, p. 1111, Feb. 2021, doi: 10.3390/molecules26041111.

[26] P. Henderson and V. Ferrari, “End-to-end training of object class detectors for mean average precision,” in *Proc. 13th Asian Conf. Comput. Vis.*, Taipei, Taiwan, Nov. 2016, pp. 198–213.

[27] S. Visa, B. Ramsay, A. Ralescu, and E. van der Knaap, “Confusion matrix-based feature selection,” in *Proc. 22nd Midwest Artif. Intell. and Cogn. Sci. Conf.*, Cincinnati, OH, USA, Apr. 2011, pp. 120–127.