Deep-learning-based detection of missing road lane markings using YOLOv5 algorithm

Nur Hanisah Azmi¹, Ali Sophian¹*, Ali Aryo Bawono²
¹Mechatronics Department, Kulliyyah of Engineering, International Islamic University Malaysia, Selangor, Malaysia
²Rapid Road Transport Department, TUMCREATE Limited, Singapore, Singapore

*Corresponding author: ali_sophian@iium.edu.my

Abstract. Road lane markings are critical for ensuring road user safety. To improve their safety, there are various types of road lane markings, such as single solid lines, double solid lines, dashed single line etc. Their colors generally are white and yellow. This road lane markings mainly used to provide guidance and information for road user to comply with the rule of the road. Unfortunately, these markings get worn out with time and may even disappear. In order to prevent this from happening, regular inspection and maintenance need to be conducted. Manual inspection is tedious, slow, and prone to human errors. With the recent technological advancement, especially in machine vision and artificial intelligence, automated or semi-automated missing road lane marking detection systems can potentially be developed. In this work, preliminary study of the implementation of one of the latest deep learning algorithms, i.e. YOLOv5, has been carried out in the detection and classification of missing road lane markings. This paper shows the preliminary results which look promising as the mean Average Precision (mAP@0.5) reaches 0.995.

1. Introduction
Traffic accidents have been identified as one of the most serious problems in Malaysia [1]. Road lane markings are one of the most important aspects of safe roadways as it is to delineate the roadway path and specific traffic lanes. They provide information including roadway alignment, vehicle positioning and other important driving-related tasks to road users. Uniformity of the lane marking is an important factor in minimizing confusion and uncertainty about their meaning among road users.

Unmarked or missing road lane markings are considered as areas of high potential hazards such as those new unmarked pavement roads in rural areas or those missing road lane markings that have been scraped away by either wind, water or other natural agents. These conditions can lead to fatal traffic accidents and missing road markings have been identified as one of causes of fatal road accidents in Malaysia [2].

Many road inspection activities, including survey on quality of road lane marking, are still done manually. The manual approaches are tedious and prone to human errors. Some automation that makes use advanced technology, such as computer vision and artificial intelligence (AI), can improve the performance and effectiveness of such inspection. Artificial intelligence is a replica of human intelligence programmed in a machine. While machine learning is a subset of artificial intelligence which is a technique used through trained data algorithms while deep learning is a subset of machine learning which is an approach of artificial neural networks. From the perspective of engineering, computer vision seeks to automate tasks that the human visual system can do [3].
There are several methods that have been reported in the research literature for the detection of the missing road lane markings, including those that use video cameras, mobile laser scanners and Light Detection and Ranging (LiDAR) [4]. Some development that uses the video cameras by using relatively low processing power computers, such raspberry pi, has also been discussed [5][6]. Many approaches typically use the same algorithm as Hough transform to generate lane marking by detecting straight edges according to the image features of the lane [7]. Another method suggested a road lane marking identification method based on the combination of fuzzy logic and random sample consensus (RANSAC) algorithm but with some drawback in the detection along a curved part of the road and the unstructured roads. The detection is more challenging due to factors such as bright light, shadows, water and so on. These factors degrade the detection performance [8][9].

A machine learning technique which is Support Vector Machine (SVM) was proposed for an accurate and efficient road marking detection system with the Principal Component Analysis on Histogram of Oriented Gradients (PCA-HOG) as its algorithm [10]. The paper comprises of three elements that are the vanishing point in the image is defined and is used to generate a bird’s-eye view of the road, the local median binarization approach is used to segment the image and it can avoid the interference of light and shadow and the feature vector is generated by using a combination of PCA-HOG and SVM to evaluate the feature vector as a particular road marking. Further studies on the SVM algorithm for the Digital Highway Data Vehicle (DHDV) system was used to acquire a laser image which can then be analysed for lane marking identification and reconstruction. Four phases were implemented in this research method, the first phase is the binarization of the laser images with a new threshold method [4]. This SVM tool is normally used in the advanced driver-assistance system (ADAS) to monitor and receive road information in real time.

Deep Convolutional Neural Network (DCNN) is a category of deep neural networks which is commonly used for visual image analysis that outperforms the traditional methods. A study presented an approach for robust road lane marking identification quality using a deep convolutional neural network known as Lane Marking Detector (LMD) [11]. A dilated convolution is chosen for enhancing performance with lower complexity together with shallower and thinner architecture for low computational cost. In contrast with the DCNN, a discovery was made on a modified faster region based conventional neural network (Faster RCNN) and proposed a fast, deep convolutional neural network that surpasses the Faster RCNN but still restricted by the scale of its convolution-based map performance capability though doing good in both accuracy and efficiency in object detection [12]. A lot of network architecture was discovered later in recent years, but this deep convolutional neural network method was frequently used to outperform the traditional approaches on many applications.

A revolutionary approach to object detection is presented as YOLO which was first published in 2016 and turned out to be one of the major real-time object detection evolution known as YOLOv1 [13]. The next YOLO generation was introduced as YOLO 9000 or YOLOv2 that can detect over 9000 object classes [14]. The same developers again successfully launched YOLOv3 as the most popular and reliable version at that time [15]. YOLOv4 was then introduced with a high margin and surprising accuracy rate compared to the EfficientDet Family, which surpasses YOLOv3 [16]. The latest version in YOLO family is the PyTorch based YOLOv5 was launched by Glenn Jocher in June 2020 which outperforms previous versions in terms of COCO AP. High accuracy object detection is not the main aim of YOLOv5 as it was developed primarily for ease of use in real-time, high-quality, and persuasive object detection performance.

In the next sections, the methodology and preliminary results of our work in the development of automated detection of road lane markings of various types based on YOLOv5 will be presented. The conclusion and further works are to follow at the end of the paper.

2. Methodology

In this project, the focus is primarily on machine vision systems that can be used to identify a missing road lane marking for inspection purposes. Therefore, unlike advance driving assistance (ADAS), the deep learning model developed or used in this project does not have to work real-time. Although real-
time capability is not required for this project, the detection speed is still crucial as the amount of data from tens of kms of roads will be huge. In addition to the detection of the missing line, the system should also detect and register road lane markings and classify them correctly.

The detection of the lane markings is developed by using YOLOv5 model which is one of the latest object detection deep learning models that have been developed. The use of this family is also deemed attractive as the community support is widely available and growing. YOLOv5’s architecture consists of the backbone, the neck, and the head sections. The backbone is a convolutional neural network that receives the input image, extracts the image features, and generates a fixed-length representation despite the possible variation in the input image size. The neck consists of network layers that combine the image features to be fed to the next stage. The head makes the prediction of the categories of detected object.

Out of the YOLOv5 family, YOLOv5s is selected to be used in this project due to its superior speed while its detection performance is comparable to other YOLOv5 algorithms.

2.1. The Development of the Road Lane Marking Detection Algorithm

The stages of the development of the road lane marking detection algorithm based on YOLOv5s is shown in the flowchart in Figure 1. Started by building the dataset and the annotation of the images, the dataset was then used to train and test the deep learning model. Some performance metrics, which will be described later, are used to evaluate the performance of the model during training and testing. The training is done through many iterations (epochs) in order to improve the performance of the detection model. Improving metrics indicate that the YOLOv5 model would be potentially used in this application. On the other hand, if the metrics are stagnant, then it is likely that YOLOv5 is not suitable for this application.

![Figure 1. YOLOv5s-based model development for road lane marking detection.](image)
2.2. Dataset and Object Classes
In this work, the dataset collection for this missing road lane marking detection project was built by using a recorded video of roads in Kuala Lumpur by using a dashboard camera. A total amount of 298 image frames were extracted from the video for building the dataset. Each image has 1280x1280 pixels. Figures 2 shows some typical samples of the images in the dataset.

![Figure 2](image_url)

**Figure 2.** Typical images grabbed from the video in the developed dataset collection.

2.3. Data Annotation
Annotation or labelling was performed to build the training image dataset by using a labelling tool called labelImg, which is recommended as it supports the annotation format that is compatible with the YOLOv5 model used in this missing road lane marking detection project. Figure 3 shows an annotated image to be used for training the deep learning model. In this particular image, the green box indicates a missing marking, while the yellow and purple are for single solid line and single broken line, respectively. The missing marking was purposely labelled inside the single broken white line to indicate the type of missing road lane marking.

![Figure 3](image_url)

**Figure 3.** An example of data annotation of different types of road lane marking. In this work, the tool called labelImg was used for annotating the images.

Table 1 shows the statistics of the annotated objects for each class. As can be seen, there are seven classes, which are single broken white line, single broken yellow line, single solid white line, single solid yellow line, double solid white line, double solid yellow line, and the missing line. The missing
line is a part of the road where a road lane line marking is supposed to be present or line marking that is visibly worn out. The data will later be used for identification which parts of the road that need lane marking repainting.

| No. | Class                     | Number of Objects |
|-----|---------------------------|-------------------|
| 1   | SingleBrokenWhite         | 523               |
| 2   | SingleSolidWhite          | 201               |
| 3   | SingleSolidYellow         | 140               |
| 4   | SingleBrokenYellow        | 40                |
| 5   | DoubleSolidWhite          | 55                |
| 6   | DoubleSolidYellow         | 58                |
| 7   | MISSING                   | 269               |

Custom object detection data were generated by uploading the 298-image dataset together with its annotations to an online tool called Roboflow. The dataset was split into 208 images for training, 60 (20%) for validation, and 30 (10%) for test. Then, a dataset health check was performed to make sure the class balance is relatively evenly distributed. Data augmentation is used in this study in order to increase the amount of training data. Subsequently, the dataset were exported to the Google Colab in the YOLOv5 PyTorch format for training.

Figure 4 and Figure 5 show some of both the original ground truth and augmented ground truth data that are used in the training. The use of the augmented data improves the quality of the training of the model.

**Figure 4.** Samples of typical ground truth data used in the training.  
**Figure 5.** Augmented training images
2.4. Model Training

The YOLOv5 repository was cloned and its dependencies were installed to set up the programming environment to be ready to run object detection training and inference commands. The minimum environment setup that is required to run the YOLOv5 lane detection training and inference command are PyTorch version 1.5, Python version 3.7 and CUDA version 10.2. Other libraries used are Numpy 1.17, Torch 1.5, Torchvision 0.6.0 and pyAML5.3.1.

The training was run with 1000 epochs. After each cycle or epoch of the training, generally metrics showed that the model became better and better. The main metrics used are precision (P), recall (R) and mAP@0.5, which is the mean Average Precision that takes the detection as positive if the intersection over union (IoU) is larger or equal to 0.5. Precision P and recall R are given by

\[ P = \frac{TP}{TP+FP} \]  
\[ R = \frac{TP}{TP+FN} \]

, where TP = true positive, FP = false positive and FN = false negative. Recall tells us how many true positive predictions out of all actual positives, while precision indicates how many are true positives predictions out of all the positive predictions.

From the training the weights of the model that gave the best performance metrics, especially the mAP@0.5 were chosen and saved for the actual use of the model in detection of the lines and their classification.

3. Results

3.1. Training Output

The performance evaluation of the model during the training after each epoch can be seen in the plots shown in Figure 6. The plots of the recall, precision and mAP metrics show that the training has managed to develop the model to become an effective road lane marking detection model. Each of these metrics has been described briefly in the previous section. The mAP@0.5:0.95, shown in Figure 6(d), represents the average mAP at different IoU thresholds ranging from 0.5 to 0.95 in increments of 0.05, i.e. the thresholds are 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, and 0.95. The growing mAP@0.5:0.95 plot also shows that the model is getting better in the detection and classification of the road lane markings or the missing ones as the training iteration is performed repeatedly.
The recall is specially very high at more than 0.95, while the precision is lower at just above 0.8. This means that the model can detect and recognize the types of most of the lines in the images, including if they are missing. However, the model also produces a rather high number of false positives. Nevertheless, the overall performance is still very good as shown by the high mAP, which is higher than 0.995.

Figure 7 shows the Precision-Recall Curve that shows the tradeoff between the precision and recall values when the threshold is varied. The resulting curve indicates that the model can reach a high performance, which is characterized by having both high precision and high recall values. The red box in the plot shows that it is possible to have both precision and recall that are higher than 0.9.
3.2. Inference Testing on Image

The inference testing with the best weights was done by using images that had not been used in the training. Figure 8 shows some examples of the results. In these particular images, it can be clearly seen that the model has successfully detected the lines and the missing lines. It can also be seen that the missing marking is located inside of other lane marking types to indicate the type of missing marking.

This result is very promising in order for us to develop it further to design an automatic identification system of parts of the inspected road that need to have marking repainting. When this information is complemented with geolocation information, the system will be very useful and informative for the road inspection and maintenance operators.

Figure 8. Inference testing results show that the model successfully classifies the missing line and the lines of different types.

4. Discussion

In this preliminary work, dataset was built by using a video of roads in Kuala Lumpur. A detection and classification model based on YOLOv5 was built and trained by using the dataset. From the results of the experiment that was done, the model evaluation performance of the custom YOLOv5 detector was very promising with mAP@0.5 reached higher than 0.99. This preliminary work will be continued by
incorporating further algorithms in order to allow the quantification of missing lane marks along a given route and GPS data will be used for the positioning to the missing lines. It is expected that the final solution will help the implementation of automated and more informative road inspection systems.

5. Conclusion
Road lane markings serve as an important navigation aid for road users during day and night which plays an important role in ensuring road in any country around the world. There is a demand for automatically detecting the missing or significantly faded road lane markings for improved road inspection and maintenance for road users’ safety. The development of deep learning techniques has provided a great opportunity to make this happen. This paper has presented some preliminary results in the study of the use of YOLOv5, which is a very latest deep learning model, in the detection and classification of the missing road lane markings. The results are very promising, with a mean Average Precision (mAP@0.5) of 0.995, paving the way for the implementation of automated inspection of road lane markings by using an RGB camera. Further work will include optimization of the accuracy of detection and classification by improving the training dataset, among others.

Acknowledgement
The authors would like to thank the owner of KL Stuff channel on YouTube who has given the permission to use their video, titled “KL Drive | City Centre, 2020”.

References

[1] Abu Mansor S N, Ahmad Saman M S, Tengku Razman, T M S, and Masnel H 2019 IOP Conf. Ser. Mater. Sci. Eng. 512 12023
[2] Abdul Manan, M M, Várhelyi A, Çelik A K, and Hashim H H 2018 IATSS Res. 42
[3] Aziz M V G, Hindersah H, and Prihatmanto A S 2017 4th International Conference on Electric Vehicular Technology (ICEVT) 149–153
[4] Li L, Luo W, & Wang K C P 2018 Sensors (Switzerland) 18
[5] Mirunalini K, David VK 2019 Int Journal of Research and Analytical Reviews 6
[6] Blaga B C Z, Deac M A, Al-Doori R W Y, Negru M, & Danescu R 2018 Proc. - 2018 IEEE 14th Int. Conf. on Intelligent Computer Communication and Processing, ICCP 2018, 229–236
[7] Hernández D C, Kurnianggoro L, Filonenko A, & Jo K H. 2016 Sensors (Switzerland), 16
[8] Xu Y, Shan X, Chen B Y, Chi C, Lu Z F, & Wang Y Q 2017 7th Int. Conf. on Power Electronics Systems and Applications
[9] Du X., & Tan K K 2016 IEEE Transactions on Image Processing 25 2075–2088
[10] Ding L, Zhang H, Xiao J, Li B, Lu S, Klette R, Norouzifard M, & Xu F 2020 Multimedia Tools and Applications 79 17193–17210
[11] Chen, P R, Lo S Y, Hang H M, Chan S W, & Lin J J 2019 International Conference on Digital Signal Processing
[12] Tian Y, Gelernter J, Wang X, Chen W, Gao J, Zhang Y, & Li X. 2018 Neurocomputing 280 46–55
[13] Redmon J, Divvala S, Girshick R, & Farhadi A. 2016 Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 779–788
[14] Redmon J, & Farhadi A. 2017 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017 6517–6525
[15] Redmon J, & Farhadi A 2018 arXiv 1804.02767
[16] Bochkovskiy A, Wang, C Y & Liao H Y M 2020 arXiv 2004.10934