EfficientDet Model for Accurate Detection and Classification of Road Damages in Indonesia

Bernard Hugo *1, Alfi Yusrotis Zakiyyah2, Kelvin Asclepius Minor 3,
1, 2, 3 Mathematics Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia
Email: 1 bernard.hugo@binus.ac.id, 2 alfi.zakiyyah@binus.edu, 3 kelvin.minor@,ac.id
*Corresponding Author

***Abstract*—This electronic Road infrastructure is an important aspect that must be maintained to ensure public road safety. Traditional road damage detection methods are labor-intensive, costly, and inefficient, highlighting the need for an automated solution. By utilizing the EfficientDet model, we tried to assess the model's performance in detecting and classifying diverse types of road damage in Indonesia. After that, we leverage a built-in data augmentation technique to improve the model's succession. We achieved the best result with the validation F1 Score of 59.7%, bypassing the performance of the previous work. The moderate performance of the model is caused by the complexity in learning road damages features and challenges in generalizing on unseen data.**

Keywords— EfficientDet, road damage, road damage detection and classification, road infrastructure, Indonesia.

I. INTRODUCTION

Road infrastructure requires high-quality maintenance to ensure public road safety. Cracks or damage in road surfaces greatly affect the safety of the traffic and road drivers [1]. Until now, Indonesia has experienced a lot of road damage accidents which have resulted in several problems: long traveling time for drivers, traffic jams, and traffic accidents [2]. Severe road damages made road drivers have to be more alert to damaged roads, which can lead to accidents: vehicle damage and collisions between vehicles [3]. The causes of these road damages are overweight vehicles, climate factors, poor road drainage, soil conditions that are unstable, and poor material pavement construction. Appropriate repairment techniques and high economic cost are required to overcome these road damages issues [2]. That is why detection and classification of road damages is crucial to be performed to prevent further damage or cracks from happening [1].

Traditional detection techniques are dominant for detecting road damage. However, these techniques are labor-intensive, costly, and high risk to give information about road damages [4]. As damaged road surfaces have gotten worse over time, there is a need for an effective and efficient solution to maintain a high-quality road condition [5]. Recent advancements in artificial intelligence and deep learning have offered new approaches to automate and facilitate the detection and classification of road damage. These approaches are beneficial to speed up the process and offer low cost and fast time in performing the task [6]. To address the road damage issues, we adopt an EfficientDet model and data augmentation techniques from the previous paper made by Naddaf-Sh et al. [7]. The goal of this research is to assess the EfficientDet model's performance in detecting and classifying damages on road surfaces in various regions in Indonesia.

In this research, we involve collecting dataset containing road images from various regions in Indonesia, leveraging the EfficientDet-D0 model, and implementing data augmentation techniques to enhance the model's quality. For the model, we leverage it from the previous paper made by Naddaf-Sh et al. [7] to be used in our research task. This research involves independent variables: road damage coordinates, types of road damages, and hyperparameter tuning to optimize the model's performance. The dependent variable in this research is the F1-Score to evaluate the effectiveness of the model in detecting and classifying road damages accurately.

Previous studies have explored deep learning methods or approaches for road damage detection and classification. CNN, Faster R-CNN, and YOLO are deep learning models or methods that are commonly used for the task. These studies with various methods have shown results that contribute to real-time road damage detection and classification.

This research explores how well the EfficientDet-D0 model can detect and classify diverse types of road damage in Indonesia. These types consist of pothole, alligator crack, longitudinal crack, and lateral crack. In addition, this research examines the role of data augmentation techniques in improving the model's quality. By adapting the model to detect and classify various road damages, this research seeks to demonstrate its effectiveness for detection and classification of road damages in Indonesia. After that, the research aims to determine how well the data augmentation techniques contribute to achieving more consistent output.

We apply a pre-trained EfficientDet-D0 model and assessing its performance in detecting and classifying road damages in Indonesia. First, we search for the dataset from Kaggle with a dataset containing road images from various regions in Indonesia and each of them has a road damage annotation coordinate as a focus point on specific cracks or damage [8]. We map the original dataset to the COCO detection dataset before we load it to the model. After that, we perform hyperparameter tuning with the Adam optimizer.

Received: 24- 12- 2024
Revised: 01-03- 2025
Published: 2- 5- 2025

When all is finished, we train and evaluate the model with a certain batch size and a certain epoch. Below are computational resources and softwares used in this research.

- CPU: Intel Core Ultra 7
- RAM: 32 GB
- Python: 3.7.6
- GPU: Intel Arc Graphics
- PyTorch: 1.8.0

II. LITERATURE REVIEW

A. *EfficientDet*

EfficientDet [9] is a state-of-the-art CNN model for real-time object detection to achieve a balance between accuracy and computational efficiency. The model uses a compound scaling method that scales the resolution, depth, and width of the backbone, feature network, and box or class prediction networks simultaneously and uniformly. The model used a combination of EfficientNet [10], BiFPN, and compound scaling approach as the backbone that made the EfficientDet achieve a better accuracy with fewer parameters. The EfficientDet model has also achieved better computational cost efficiency and high accuracy than previous object detection and semantic segmentation models across various resource constraints. The EfficientDet model comes in several variants ranging from D0 to D7. Each of these variants is optimized for a specific level of accuracy and computational cost. EfficientDet-D0 is the smallest and fastest variant with low computational resources and EfficientDet-D7 is the largest variant and requires more computational resources.

B. *Japan Road Association Standard*

Japan Road Association (JRA) standard is a standard provided by Japan Road Association that categorizes road damages into several types. The types of road damage were categorized using the following codes [11]:

- Longitudinal crack (D00)
- Lateral crack (D10)
- Alligator crack (D20)
- Pothole (D40)

C. *Evaluation Metrics*

Intersection over Union (IoU) is a metric that is used to compare similarities between two shapes that change dynamically. IoU captures and translates the shape characteristics of the objects that are being compared, which are width, height, and bounding box location of each object into region properties. Then, it calculates a normalized measure that shows the object's area [12]. The IoU is calculated with formula:

$$IoU = \frac{area(B_{Bp} \cap B_{gt})}{area(B_{Bp} \cup B_{gt})} \quad (1)$$

where $area(B_{Bp} \cap B_{gt})$ represents the area of intersection between the predicted bounding box and the ground truth bounding box. Then, $area(B_{Bp} \cup B_{gt})$ represents the area of union of the predicted bounding box and the ground truth bounding box [7].

F1 Score is an evaluation metric that is used to evaluate the performance of classification models. F1 Score is calculated using the formula [7]:

$$F1\ Score = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (2)$$

where

$$precision = \frac{tp}{tp + fp} \quad (3)$$

and

$$recall = \frac{tp}{tp + fn} \quad (4)$$

where *tp* stands for true positive, *fp* stands for false positive, and *fn* stands for false negative.

D. *Loss Metrics*

1) *Focal Loss*

Focal loss [13] is a loss function that is used to address the class imbalance problem in object detection tasks. This loss function is used to reduce the relative loss of well-classified examples and help to focus more on hard or misclassified examples. First, the focal loss is defined from the balanced cross-entropy loss. The balanced cross-entropy loss is defined as:

$$CE(p_t) = -\alpha_t \log(p_t) \quad (5)$$

where α_t is a weighting factor for the class and p_t is the probability of the predicted true class. p_t is defined as:

$$p_t = \begin{cases} p & y = 1 \\ 1 - p & y = 0 \end{cases} \quad (6)$$

where y is the ground truth class (1 for positive class and 0 for negative class) and p is the model predicted probability for positive class.
The focal loss is defined as:

$$FL(p_t) = -(1 - p_t)^{\gamma} \log(p_t) \quad (7)$$

If the dataset has imbalance classes, the balanced version of focal loss is defined as:

$$FL(p_t) = -\alpha_t (1 - p_t)^{\gamma} \log(p_t) \quad (8)$$

where γ is the focusing parameter that adjusts the number of instances in the dataset that the model can classify accurately with high confidence.

2) *Huber Loss*

Huber loss [14] is a commonly used loss function for regression tasks including bounding box regression. Huber Loss is a loss function that is used to calculate bounding box loss. In the context of object detection, bounding boxes are defined by their center coordinates and dimensions. Huber loss is used to calculate differences between the predicted bounding box coordinates and the ground truth coordinates. Huber loss is defined as:

$$L_\delta(a) = \begin{cases} \frac{1}{2}a^2 & if|a| \le \delta \\ \delta|a| - \frac{1}{2}\delta^2 & if|a| > \delta \end{cases} \quad (9)$$

where a is the difference between the ground truth bounding boxes and the predicted bounding boxes, and δ represents the threshold parameter [15].
The final loss is defined as:

$$Loss = FL(p_t) + \text{weight} * L_\delta(a) \quad (10)$$

E. Related Works

A research paper made by Sadra Naddaf-Sh, M-Mahdi Naddaf-Sh, Amir R. Kashani, and Hassan Zargarzadeh [7] aims to develop scalable and efficient models that can detect road damages automatically with the goal of improving road safety, reducing maintenance costs, and supporting infrastructure management. Data augmentation techniques are applied by using selected augmentation policy (V1) that can improve the performance of smaller models and improve the detection capability. However, this research has a weakness where the model produces false positives and false because of the bounding box problem. Because the effectiveness of the EfficientDet model is dependent on the bounding box. The results of their research were an achieved F1 Score of 56% on the RDD-2020 dataset and an average inference time from 178-10 images per second.

Research about the introduction of an enhanced lightweight deep learning network which is an optimized version of EfficientDet called E-EfficientDet from D0 until D2 variants has been done by Hui Luo, Chenbiao Li, Mingquan Wu, and Lianming Cai [16]. The interesting thing about this research is the result of a lightweight model E-EfficientDet-D0 has the size of 32.31 MB and reached the detection speed of 27.08 FPS. Then, an E-EfficientDet-D2 from the result shows that it can reach a detection accuracy of 57.51% with the size of 61.78 MB. Which made it suitable for real-time applications and feasible for deployment on mobile devices, drones, and smartphones. Overall, the result of the research shows that lightweight networks E-EfficientDet-D0 and E-EfficientDet-D2 offered a high accuracy in road damage detection with small model size, and they can be suitable for real-time application scenarios.

Mandal et al [17] conducted a study to perform analysis of deep learning frameworks for pavement distress classification. They used, evaluated, and compared three state-of-the-art deep learning algorithms namely, YOLO, CenterNet, and EfficientDet. It aims to address the critical need for timely pavement surfaces maintenance and rehabilitation and to limit further degradations and maintain high quality pavement surfaces. Among these algorithms, YOLO is the best model with achieved F1 Scores of 0.5814 and 0.5751 on two different test datasets.

A study made by Vung Pham, Du Nguyen, and Christopher Donan [18] used YOLOv7 with coordinate attention technique and fine-tuning technique to classify road damages. The dataset used in this research are RDD-2020 dataset and USA road images data that were collected using Google Street View. As a result, it achieved an F1 Scores of 81.7% in the U.S. Road data and 74.1% on overall test images. This study shows the adaptability of an advanced version of the YOLO model for detecting and classifying various road damage types in real-world applications.

A study made by Arman et al. [19] focused on using Faster R-CNN and R-CNN to develop an automated solution for detecting and classifying road damage. These algorithms were tested on street images from Dhaka City and compared to discover which one can perform better. As a result, Faster R-CNN achieved a high accuracy rate of 98.88% with a low loss of 0.01 and it outperformed R-CNN. The study shows the effectiveness of Faster R-CNN in handling road damage detection with high precision in complex environments.

A study conducted by Rahul Vishwakarma and Ravigopal Vennelakanti [20] focused on comparing two-stage Faster R-CNN with Resnet and FPN backbones with a one-stage YOLOv5 with CSPNet backbone on the RDD-2020 dataset. It achieved a mean F1 Scores of 0.542 or 54.2% on Test2 and 0.536 or 53.6% on Test1 by using a multi-stage Faster R-CNN model with Resnet-50 and Resnet-101. It demonstrated that Faster R-CNN with Resnet-101 gives better evaluation in Test1 and Faster R-CNN with Resnet-50 gives better evaluation in Test2.

Recent research made by Pham et al. [21] has explored and developed an effective automated solution for detecting and classifying road damage by applying Faster R-CNN and Detectron2. Detectron2 is applied to the Faster R-CNN model to enhance efficiency and speed up the model development process. The Faster R-CNN model with X101-FPN and Detectron2 base model shows less efficiency and generalization by achieving an F1 Score result of around 51%. Another study made by Kortmann et al. [22] created a solution by using Faster R-CNN to address the urgent need for effective and automated detection of road damage, enhance the infrastructure management, and support the development of autonomous driving technologies. In addition, a regional expert network is used to enhance the model's accuracy in detecting specific road damage. The achieved F1-score result of the model is 48.7% across all regions for identifying road damage types.

Wan et al. [23] conducted a study that focused on developing a lightweight and efficient model for road damage detection called YOLO-LRDD to balance detection accuracy and computational efficiency. By incorporating YOLO with BiFPN and optimizing the sample imbalance processing method, the model can improve the feature extraction in terms of the detection precision, and it can improve the sensitivity of small objects recognition. The resulting YOLO-LRDD model has a smaller number of parameters and less computation resources and achieved an accuracy score of 59.2%.

A research paper made by Anand et al. [24] proposed a Convolutional Neural Networks (CNN)-based method for an automated road damage detection aiming to improve the efficiency and accuracy in detecting road damages. The objective of this research was to design and develop a system with a model that can detect and classify road damages accurately. The model is trained on a dataset consisting of 6.000 road images containing various road damages consist of cracks, potholes, and patches, with 4.500 for training and 1.500 for testing. As a result, the developed system demonstrated a strong performance with a promising accuracy score of 94% along with a precision of 92% and a recall of 96%. In summary, the research highlighted a practical application of the developed system and its potential to improve road safety, prevent any accidents, and reduce maintenance costs.

III. METHODOLOGY

The objective of this research lies in the theories related to object detection and deep learning. This research uses the state-of-the-art CNN model called EfficientDet-D0 for real-time object detection. The model uses BiFPN (Bidirectional Feature Pyramid Network) as a backbone for easy and fast multi-scale feature fusion [9]. For classifying road damage types, this research uses a JRA (Japan Road Association) standard. The standard helps the model to classify each road damage type using codes: D00 for a longitudinal crack, D10 for a transverse crack, D20 for an alligator crack, and D40 for a pothole [11]. Overall, the theories aim to contribute to an experiment for real-time road damage detection and classification in Indonesia by utilizing the EfficientDet-D0 model.

From Fig. 1, the flow of research starts with a literature review where we research papers, collect them, and select the most relevant paper for our research topic. Following this, the model selection is conducted to select the most relevant model for our research. After we have done literature review, we done the experiment. The experiment we perform is detailed in Fig. 2, consist of preparing dataset, preparing model, augmenting data, tuning hyperparameter, training the model, and evaluating the model. Next, we analyze the result produced by the model and determine whether the model is good or bad in detecting and classifying road damages, consist of creating required graphs and comparing with the previous paper. Finally, we create a conclusion that concludes the overall methodology and the result score and mention future works that can improve the model performance and results.

From Fig. 2, the main method of our research is the EfficientDet-D0 model and the data augmentation technique from the previous paper [7]. First, we prepare the dataset that we collected from Kaggle to be used in our research by load the images into a new directory and convert the images coordinates data from raw txt files into a structured format, which is required to perform model training and evaluation. The dataset contains 6643 files where one text file represents the road damage classes, 3321 files are road images, and the rest of them are text files containing annotation coordinates representing damage(s) of the road within each image. Next, we propose the EfficientDet-D0 model as the main model. After that, we utilized data augmentation techniques from the previous paper made by Naddaf-Sh et al. [7] to improve the model's quality. Next, we perform hyperparameter tuning by tuning various parameter values: batch size, learning rate, number of epochs, and optimizer to optimize the model's performance.

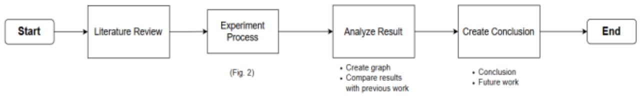

Fig. 1. The flow of research.

After hyperparameter tuning, we train the model using tuned hyperparameters and augmented image data. We begin by loading the prepared dataset images into the model and we train it using the tuned parameter values. Training EfficientDet-D0 is a supervised process. This is different from YOLO where it does not require a ground truth as its baseline. Next, we evaluate the trained model to test the model's performance on unseen data. We use F1 Score as a metric to provide a balanced measure of the model's performance by combining precision and recall into a single metric.

Here are the methodology steps of this research:

A. *Prepare Dataset*

First, we perform the dataset preparation which involves loading the images from the raw dataset into a new directory containing all images data. Next, we convert the raw txt file from the dataset containing the coordinate matrix of the damaged area within the image into a structured JSON file that is suitable for model training and evaluation. We split these txt files into two subsets: training set and validation set along with converting them into structured JSON files. We use a validation split ratio of 5 where it indicates that 20% of the total images were appended to the validation set. Then, the remaining images were appended to the training set. As a result, the training set contains 2978 image annotations data, and the validation set contains 1225 image annotations data.

Each image's metadata includes its filename, image height, width, and its unique identifier is encapsulated in the images field. While the annotations field encapsulates the

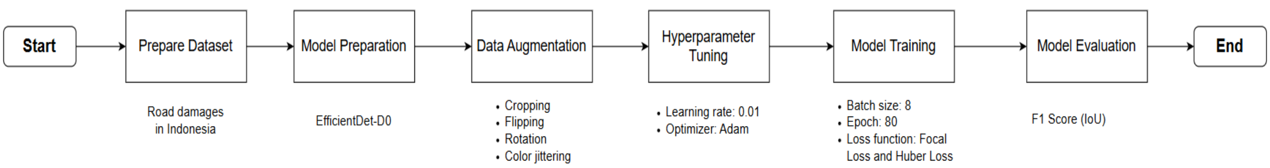

Fig. 2. The proposed method.

bounding box information that specifies the damaged regions of the corresponding image. The categories of road damages were labeled as D00, D10, D20, and D40 with respective IDs ranging from 1 to 4. Inside the training and validation JSON data, the IDs representing the road damage types in the dataset were standardized to align with the following provisions:

1: Longitudinal crack
2: Lateral crack
3: Alligator crack
4: Pothole

B. Model Preparation

Based on the specification table of the EfficientDet model shown in [25], we use the EfficientDet-D0 model due to its faster inference time, small parameters, and small model size. This model is useful to be utilized in devices that have limited computational resources, and it requires less memory. The model's faster inference time can accelerate the process of image processing during model training and evaluation. We tried to use D4 variant to achieve better accuracy, but we could not continue due to lack of computational resources.

C. Data Augmentation

We use data augmentation approaches from the previous research paper made by Naddaf-Sh et al. [7] to improve the quality of the model. First, the image augmentation is performed by utilizing a random augmentation policy through techniques: cropping, flipping (vertical and horizontal), rotation, and color jittering. Next, every image is resized to the target size which is 224. These techniques allow the model to learn from images of varying scales with the aim of improving the model's robustness to objects that may appear in different sizes within the same class. Bounding box augmentation is also applied by transforming coordinates of corresponding bounding boxes in each image annotation to keep it aligned with the augmented image and ensuring each object annotation remains accurate after the transformation.

D. Hyperparameter Tuning

The hyperparameter tuning process involves adjusting key parameters to optimize the EfficientDet-D0 model's performance. We tuned the following values: learning rate of 0.01 and Adam optimizer. These parameters will then be used for model training and evaluation.

E. Model Training

The training data is loaded into the model for training. The model is trained with a batch size of 8 for 80 epochs using optimized hyperparameter values. Each epoch consists of 373 iterations, the total number of iterations for training is 29840. We monitored the average loss values throughout the training process to assess the model's learning performance on the training data. To calculate losses, we use a combination of two loss functions, namely Focal loss and Huber loss. For the weight defined in eq. 10, we use the implicitly defined weight by 50, used by [7]. Therefore, this loss combination provides a balanced learning or training performance between the classification task and the localization task.

F. Model Evaluation

We evaluate the model on the validation data as soon as the training process is done at each epoch. We evaluate how well the model performs in detecting and classifying road damages based on the produced F1 Score. The F1 Score used in our research is a combination of F1 Score classification and IoU. The classification F1 Score evaluates the balance between precision and recall. While IoU measures the overlap between ground truth and predicted bounding boxes. This combination provides a metric to assess both classification score performance and detection performance. In practice, the F1 Score of the model is calculated by calculating the ground truth and detected class labels to calculate metrics: true positives, false positives, and false negatives. True positives are obtained by calculating the minimum overlap (IoU) between the ground truth and the number of detections for each class. While false positives and false negatives are obtained by calculating differences between detection and ground truth counts to determine over-predictions and under-predictions. These metrics will be used to calculate precision, recall, and F1 Score of the model.

IV. Results and Discussion

From Fig. 3, we can see that the model shows a limited learning ability of the features within images. Though the loss value reduces over iterations, the average loss at the last iteration still shows a high value. The average loss at the first iteration shows a value of 7.0037 or 700.37%. As the number of iterations increases, the average loss value decreases. Until it reaches a stable value of 1.1735 or 117.35% in the last iteration. The reduction of average loss value shows that the model is still limited in learning patterns of road damage within each image data.

From Fig. 4, there is an improvement of F1 Score throughout 73 epochs. The F1 Score at the first epoch shows a value of 12.59%. It increases along with the increasing epoch, and it reaches the highest F1 Score of 59.7% at the epoch of 72. After that, the model shows a slight fluctuation in the next epochs. Based on this result, the EfficientDet-D0 model performs moderately to detect and classify road damages effectively with the best achieved F1 Score of 59.7%.

Overall, it demonstrates that the model's performance is still limited in terms of learning features and generalization to unseen data. The reduction of average loss value from 7.0037 to 1.1742 indicates that the model is still struggling to capture features within the dataset and the error is still high. In terms of how well the model performs, the highest F1 Score of 59.7% indicates that the prediction performance of the model is moderate. Therefore, the EfficientDet-D0 model still has limitations in detecting and classifying road damages effectively, and it still faces challenges to reduce its accuracy error.

The results shows that the EfficientDet-D0 model performs moderately in detecting and classifying road damages in Indonesia, as the best F1 Score is 59.7%. We found that the model struggling to fit the dataset at first, but gradually got the moment at the later stage. The quality of the, which may caused by the complex features of road damage and noises in the dataset. These issues made the

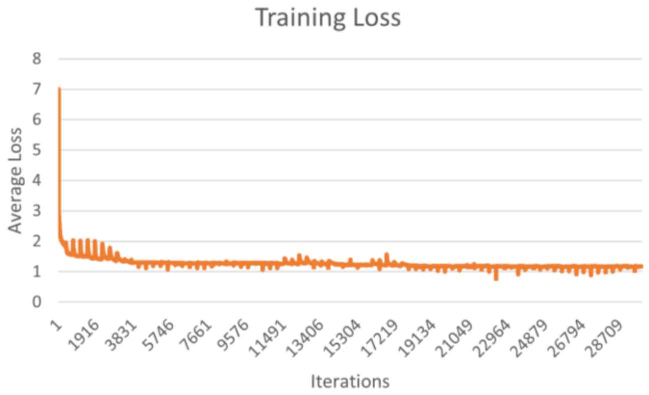

Fig. 3. Model training loss.

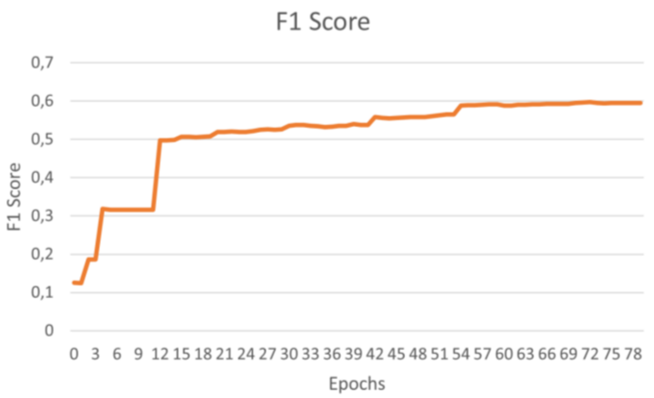

Fig. 4. F1 Score results.

model facing challenges in learning features and the model still has limitations to achieve a better result.

Our result shows that the model performs better in road images from Indonesia and outperforms the research conducted by Naddaf-Sh et al.. The result in our research is attributed to the diversity of the dataset that we used. This diversity contributed to the model's ability to adapt and perform better within local images data and to achieve a higher score.

V. CONCLUSIONS

We assessed how well the EfficientDet-D0 model performs in detecting and classifying road damages in Indonesia. The research focused on evaluating the model's ability to identify four types of road damages: pothole, alligator crack, longitudinal crack, and lateral crack. The model achieved a moderate validation F1 Score of 59.7%, indicating that the model faces challenges that could impact its ability to generalize on unseen data. Followed by high average loss values during model training, the model is still limited in learning features within the image dataset. This moderate performance is due to challenges when dealing with the complexity of object features within image data. Data augmentation techniques from the paper by Naddaf-Sh et al. shows contribution to slightly improve the model quality robustness effectively. Overall, it shows that the research has limitations that hinder the ability of the model to achieve higher performance or accuracy.

When compared to the performance reported in the previous research paper by Naddaf-Sh et al., our research shows a better model performance in detection and classification of road damages in Indonesia and outperforms them. This performance score disparity, caused by the diversity in quality of the dataset used in the research where our dataset contains a better sight of the object within several images. Some images have the same settings and conditions, but the difference lies in the distance at which each picture was taken. So, it could affect the F1 Score in model evaluation where the model is still limited to be able generalize on unseen road damages.

To achieve better performance, we can perform improvements by utilizing advanced hyperparameter and data augmentation techniques to enhance the model's generalization ability. Because the used data augmentation techniques are still limited to improve the model's generalization on unseen data and training performance. Another limitation of this research is the dataset size which is relatively small compared to the needs in real scenario. Therefore, adding more data to the dataset can be beneficial to increase the dataset quality and enhance the model's ability to learn and generalize. While, EfficientDet-D0 model is better in terms of its small size and low resource requirements, it is limited to achieve better accuracy. Applying advanced EfficientDet variant models (D1 or higher) and incorporating feature refinement techniques can contribute to improve feature learning ability and achieve better accuracy.

REFERENCES

[1] Y. Jiang, "Road Damage Detection and Classification Using Deep Neural Networks," Discover Applied Sciences, vol. 6, no. 421, p. 1, 2024.

[2] E. Suhartono, M. A, R. BP, M. M. D. R. Kusumastuti, D. B. Setiawan and N. S. P, "Estimated Cost of Repairing Road Pavement Damage Assessment: A Case Study of Indonesia," IJASRE, vol. 5, no. 6, p. 1-2, 2019.

[3] H. Kusumah, M. R. Nurholik, C. P. Riani and I. R. N. Rahman, "Deep Learning for Pothole Detection on Indonesian Roadways," Journal Sensi, vol. 09, no. 02, pp. 175-176, 2023.

[4] G. Guo and Z. Zhang, "Road Damage Detection Algorithm for Improved YOLOv5," scientific reports, vol. 12, no. 15523, p. 1, 2022.

[5] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, A. Mraz, T. Kashiyama and Y. Sekimoto, "Deep Learning-based Road Damage Detection and Classification for Multiple Countries," Automation in Construction, vol. 132, p. 1, 2021.

[6] V. Pham, L. D. T. Ngoc and D.-L. Bui, "Optimizing YOLO Architectures for Optimal Road Damage Detection and Classification: A Comparative Study from YOLOv7 to YOLOv10," IEEE, p. 1, 2024.

[7] S. Naddaf-Sh, M.-M. Naddaf-Sh, A. R. Kashani and H. Zargarzadeh, "An Efficient and Scalable Deep Learning Approach for Road Damage Detection," IEEE, pp. 1-2, 5, 2020.

[8] Basily, A. (2020, August). Road Damage, Version 1. Retrieved January 11, 2025, from https://www.kaggle.com/datasets/alvarobasily/road-damage/data.

[9] M. Tan, R. Pang and Q. V. Le, "EfficientDet: Scalable and Efficient Object Detection," IEEE, pp. 1-8, 2020.

[10] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," Proceedings of the 36th International Conference on Machine Learning, pp. 6105-6114, 2019.
[11] JRA. Maintenance and repair guide book of the pavement 2013. Japan Road Association, 1st. edition, 2013.
[12] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid and S. Savarese, "Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression," IEEE, pp. 658-659, 2019.
[13] T.-Y. Lin and P. G. R. G. K. H. P. Doll´ar, "Focal Loss for Dense Object Detection," IEEE, pp. 2980-2982, 2017.
[14] G. P. Meyer, "An Alternative Probabilistic Interpretation of the Huber Loss," IEEE, pp. 5257-5259, 2021.
[15] M. u. I. Arif, M. Jameel and L. Schmidt-Thieme, "Directly Optimizing IoU for Bounding Box Localization," in Pattern Recognition, Auckland, Springer, 2019, p. 545.
[16] H. Luo, C. Li, M. Wu and L. Cai, "An Enhanced Lightweight Network for Road Damage Detection Based on Deep Learning," Electronics, pp. 1-20, 2023.
[17] V. Mandal, A. R. Mussah and Y. Adu-Gyamfi, "Deep Learning Frameworks for Pavement Distress Classification: A Comparative Analysis," IEEE, pp. 1-7, 2020.
[18] V. Pham, D. Nguyen and C. Donan, "Road Damages Detection and Classification with YOLOv7," IEEE, pp. 1-8, 2022.
[19] M. S. Arman, M. M. Hasan, F. Sadia, A. K. Shaki, K. Sarker and F. A. Himu, "Detection and Classification of Road Damage Using R-CNN and Faster R-CNN: A Deep Learning Approach," LNICST 325, pp. 730-741, 2020.
[20] R. Vishwakarma and R. Vennelakanti, "CNN Model & Tuning for Global Road Damage Detection," IEEE, pp. 1-7, 2021.
[21] V. Pham, C. Pham and T. Dang, "Road Damage Detection and Classification with Detectron2 and Faster R-CNN," IEEE, pp. 1-10, 2020.
[22] F. Kortmann, K. Talits, P. Fassmeyer, A. Warnecke, N. Meier, J. Heger, P. Drews and B. Funk, "Detecting Various Road Damage Types in Global Countries Utilizing Faster R-CNN," IEEE, pp. 1-9, 2020.
[23] F. Wan, C. Sun, H. He, G. Lei, L. Xu and T. Xiao, "YOLO LRDD: a lightweight method for road damage detection based on improved YOLOv5s," EURASIP, vol. 2022, no. 98, pp. 1-18, 2022.
[24] A. Anand, A. Shinde, H. Borgave, and S. S. Thigale, "Road Damage Detection using CNN in Machine Learning," International Journal of Engineering Research & Technology (IJERT), vol. 12, no. 04, pp. 505-509, 2023.
[25] M.-M. Naddaf-Sh, "roadDamageDetection2020," GitHub, 26 October 2020. [Online]. Available: https://github.com/mahdi65/roadDamageDetection2020/blob/main/README.md. [Accessed 11 January 2025].