Verification of the effect of UAV shooting altitude on detection accuracy in asphalt crack detection using deep learning

Eita Uotani[1*] and Mitsuharu Tokunaga[2]

[1]Graduate Student, Department of Civil and Environmental Engineering,

Kanazawa Institute of Technology, Japan

[2]Professor, Department of Civil and Environmental Engineering,

Kanazawa Institute of Technology, Japan

c6500506@st.kanazawa-it.ac.jp

Abstract *Accurate identification of early deterioration in asphalt pavement and the implementation of appropriate preventive maintenance are critical issues in reducing life cycle costs (LCC). However, at present, regular inspections are conducted in only about 80% of prefectures and approximately 20% of municipalities, and standardized methods for data collection and management remain underdeveloped. This study aims to improve the efficiency and quality of pavement inspections by combining unmanned aerial vehicles (UAVs) with artificial intelligence (AI). A machine learning model was developed to automatically detect cracks from UAV-captured images of asphalt pavement. The study evaluated how detection accuracy is affected by imaging conditions such as flight altitude, camera angle, and lighting, as well as by model architecture and training methods. Analysis under multiple conditions revealed a tendency for detection accuracy to decline as flight altitude increases. However, the introduction of model optimization techniques and data augmentation was found to effectively suppress this decline. As a result, high-accuracy crack detection was demonstrated to be feasible even from relatively high altitudes, enabling more extensive and efficient pavement inspection. This approach is expected to contribute to the advancement of road management practices.*

Keywords: Deep Learning, YOLO, Asphalt Crack, UAV altitude

1 Introduction

A Maintenance and management of asphalt pavement in our country play an extremely important role in ensuring safe and smooth road traffic. In particular, many arterial and residential roads with high traffic volumes are constructed with asphalt pavement. If deterioration or damage to these pavements is left unaddressed, it not only compromises the safety of vehicles and pedestrians but may also lead to sudden repairs, resulting in traffic restrictions, increased construction costs, and potential social and economic losses. Therefore, accurately identifying early deterioration of asphalt pavement and implementing appropriate preventive maintenance are critical issues for extending the lifespan of road infrastructure and reducing life cycle costs (LCC). However, it is difficult to say that the inspection system for pavements is sufficiently developed nationwide. Especially at the municipal level, issues such as labor shortages and the aging of technical personnel make continuous maintenance and management challenging. According to a report by the National Institute for Land and Infrastructure Management (2024), currently, pavement inspections are conducted in only about 80% of prefectures and about 20% of municipalities, and unified data acquisition methods and management standards have not yet been established [1]. In response to this situation, next-generation pavement inspection methods utilizing UAVs (Unmanned Aerial Vehicles) and AI (Artificial Intelligence) have attracted attention in recent years. UAVs can quickly and extensively perform inspection tasks that were previously conducted manually, and by adjusting the flight altitude and camera angle, they offer flexible adaptation to various pavement conditions. Furthermore, advances in AI-based image analysis technology have made it possible to automatically detect and classify cracks and damaged areas from captured road surface images, thereby assisting technical personnel in their assessments and streamlining inspection work. There are also reports that the introduction of AI has improved inspection efficiency by two to three times [2], and its usefulness is highly regarded. On the other hand, several challenges remain regarding the introduction of such technologies. In particular, it is known that the accuracy of crack detection by AI is greatly influenced by shooting conditions (such as flight altitude, camera angle, and illumination), as well as differences in model architecture and training methods. As the shooting altitude increases, image resolution decreases and the visibility of cracks tends to deteriorate, so ingenuity is required to maintain accuracy even in high-altitude UAV photography. In addition, to maximize model performance, factors such as annotation accuracy, the quality and quantity of training data, and data augmentation techniques are also important elements.

2 Research Objectives

This study aims to develop a machine learning model for the automatic detection of cracks in asphalt pavement using images captured by an Unmanned Aerial Vehicle (UAV). The primary objective is to clarify the impact of imaging conditions, such as flight altitude and camera angle, on detection accuracy. Furthermore, the study seeks to compare the changes in detection performance when various data augmentation techniques are applied, in addition to examining different model architectures and training methods. Through this, the goal is to propose imaging and training approaches that can consistently achieve high detection accuracy. In addition, by collecting and analyzing data under diverse conditions that simulate real-world road environments, the study emphasizes the development of a model with high field applicability. This approach aims to enhance the efficiency and frequency of pavement inspection, which has traditionally relied heavily on manual labor, thereby reducing maintenance costs and alleviating the workload of inspection personnel. The outcomes of this research are expected to contribute not only to the realization of wide-area and efficient pavement inspection, enabling high-quality infrastructure management with limited personnel and budget, but also to the advancement of smart infrastructure management technologies and potential applications in other related fields.

3. Experiment

3.1 Experimental Preparation

Data Acquisition for Model Training

In this study, image data were collected from a municipal road located near the Amaike Nature School Golf Practice Range in Hakusan City, Ishikawa Prefecture, Japan. This location was selected because both longitudinal and transverse cracks were present on the surface layer of the asphalt pavement, and the traffic volume was relatively low, allowing for safe and repeated image acquisition. Furthermore, the pavement width was typical for municipal roads, suggesting that the results of this study could be applied to other roads of similar scale. To ensure consistent weather conditions and clear visibility of the pavement surface, data collection was conducted on sunny days. The UAV flight altitude was set between 5 and 10 meters, balancing the need to avoid obstacles such as surrounding vegetation and structures while maintaining a high-resolution capture of pavement cracks. During image acquisition, factors affecting the visibility of cracks were carefully controlled. Specifically, to minimize strong shadows and glare caused by sunlight, images were captured between 10:00 AM and 2:00 PM. Additionally, image collection was avoided during and immediately after rainfall, as wet surfaces obscure cracks. To accurately record the shape and width of the cracks, the UAV was controlled so that the camera remained nearly perpendicular to the pavement surface, thereby minimizing image distortion caused by oblique viewing angles. By standardizing these conditions, variations due to flight altitude and external environmental factors were minimized, resulting in a dataset well-suited for training deep learning models.

The specifications of the UAV used for image acquisition are shown below:

Model: Mavic 3 Pro

Camera: Hasselblad 4/3 CMOS, equivalent to 24 mm

Resolution: 3840 × 2160 (4K UHD)

Figure 1：Mavic3pro

Figure 2:Drone camera

Pre-processing of Captured Images

In this study, training data were created for deep learning models using asphalt pavement images captured by UAVs. For annotation, Label Studio, an open-source annotation tool, was used. Label Studio is a flexible, web-based platform that enables multiple users to work simultaneously and track progress. It also supports various output formats, such as YOLO and COCO, which makes it suitable for efficiently generating datasets for common deep learning models. In the annotation process, cracks in each image were enclosed with rectangular Bounding Boxes. This annotation format is suitable for object detection models such as YOLO and was considered appropriate for accurately learning to detect cracks. The target cracks were limited to those with clear contrast against the pavement surface, where both width and length could be visually identified. Specifically, cracks that were easy to distinguish from pavement texture and minor surface changes were targeted. The following cases were excluded:

・Cracks that were extremely small in width and could not be clearly identified in the image.

・Patterns that were difficult to visually distinguish from shadows, fallen leaves, dirt, and other artifacts.

By excluding these cases, we aimed to reduce noise in the training data and minimize false positives. These criteria ensured consistency in the annotation process and improved the accuracy of the deep learning model. The annotation work proceeded as follows. First, UAV-captured images were uploaded to Label Studio, where workers manually set the Bounding Boxes while checking the locations of the cracks. Even if a crack had a curved shape, the Bounding Box was standardized as a rectangle, and one continuous crack was enclosed by a single Bounding Box in principle. When multiple cracks existed in close proximity, each was annotated separately with its own Bounding Box.

Figure 3:Example of training data

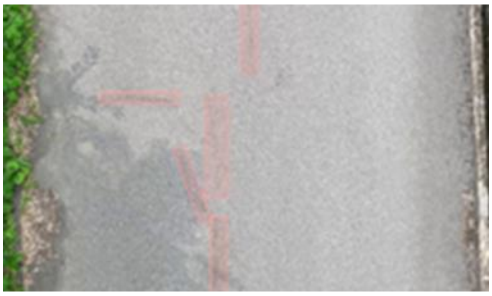

Figure 4:Enlarged view

Finally, the annotation data were exported in YOLO format, which is directly compatible with YOLO-based object detection models. This format generates a corresponding text file for each image, containing the coordinate information and class labels of the Bounding Boxes.

Training Parameter Settings

In this study, for the object detection task using YOLO, the detection target was limited to cracks only, and the number of classes was set to one. Based on the results of preliminary experiments, it was observed that around 50 epochs were insufficient for adequate loss reduction, while the accuracy tended to stabilize around 100 epochs. Therefore, the number of epochs was set to 100. The input image size was fixed at 640 pixels to ensure sufficient resolution for recognizing fine crack details while minimizing the load on GPU memory. The batch size was set to 16, considering a balance between training stability, processing speed, and GPU memory usage. The directory was structured by allocating 200 images for training (train) and 48 images for validation (val), resulting in an approximately 8:2 ratio of annotated data. Using this dataset and training configuration as the foundation, this study examines in detail the impact of flight altitude and the presence or absence of data augmentation on model accuracy.

Object Detection Algorithm

In this study, YOLOv8 (You Only Look Once, version 8) was adopted to automatically detect cracks in asphalt pavement. YOLOv8 is widely recognized in the field of real-time object detection, with a significantly redesigned architecture compared to previous YOLO versions. It incorporates advanced deep learning optimization techniques to achieve both high accuracy and fast processing speed. This is particularly advantageous for pavement inspection, where large volumes of road images must be processed efficiently. YOLOv8 has significantly improved its ability to detect small objects and fine structures compared to previous models, making it effective for identifying elongated and irregular cracks commonly found in asphalt pavement. Zhang et al. (2025) reported detecting microcracks in pavement images using YOLOv8-DGS, a derivative model of YOLOv8. Their model achieved an mAP50 of 92.4%, an F1 score of 90.8%, and an inference speed of approximately 85 FPS, representing more than a 20% performance improvement compared to YOLOv5 and related models [3]. Studies have also shown that a lightweight, high-performance variant called YOLOv8-ES can maintain high detection accuracy even with low-resolution images by incorporating dynamic convolution and the Spatial Guided Attention Module (SGAM) [4]. This suggests that the model may be suitable for inspection under conditions such as high-altitude UAV imaging and variable lighting environments. Moreover, YOLOv8 provides a Python API and an open-source official implementation by Ultralytics, making it straightforward to implement and scalable for practical applications. Transfer learning can be easily applied to custom datasets, allowing stable models to be built even with a small dataset, thereby enabling effective use of the limited image data resources available to local governments [5]. In summary, YOLOv8 is highly accurate, fast, lightweight, and scalable, making it a practical and flexible solution for pavement inspection applications. In this study, efficient and high-precision crack detection on UAV-acquired asphalt pavement images was achieved by leveraging these capabilities.

3.2 Experimental Environment

Operating System: Windows 11

CPU: Intel Core i7-10700 @ 2.90 GHz

GPU: NVIDIA GeForce RTX 2070 Super

Python Version: 3.12.9

YOLO Framework: Ultralytics YOLOv8 (version 8.3.163)

3.3 Experiment Comparing the Effects of Data Augmentation (Experiment 1)

The purpose of Experiment 1 was to evaluate the effect of data augmentation on model performance. Two models were trained under identical learning conditions: one with data augmentation applied and the other without. The detection accuracy of the two models was then compared.

Data Augmentation Applied

Data augmentation is a technique that applies various transformations to training images, enabling the model to learn efficiently under diverse imaging conditions.

In the task of crack detection on asphalt pavements, both shooting conditions (e.g., illumination, shadows, camera angles) and pavement surface conditions (e.g., dryness, wetness, dirt, and crack patterns) can vary significantly. When only a limited number of training images are available, the model may overfit to specific conditions, resulting in decreased generalization performance. To address this issue, this study introduced data augmentation to artificially generate images simulating a wide range of imaging environments. This approach allowed the construction of a realistic and diverse training dataset, improving both the detection accuracy and robustness of the model. The data augmentation methods used in this study were those implemented by default in YOLOv8 [6], as described below.

Geometric Transformations

・Horizontal Flip (Fliplr)

Flips the image horizontally to increase the diversity of crack orientations and shape patterns, while compensating for asymmetry in the pavement and bias in structural layouts.

・Vertical Flip (Flipud)

Flips the image vertically to reproduce subtle variations in UAV orientation or inversion during flight.

・Rotation (Degrees)

Rotates the image to simulate slight tilts occurring during UAV imaging, thereby enhancing the model's robustness to changes in shooting angles.

・Translation (Translate)

Shifts the image horizontally and vertically to replicate UAV positional variations.

・Scaling (Scale)

Enlarges or reduces the image within a defined range to simulate UAV altitude changes and corresponding variations in the apparent crack size.

・Shear Transformation (Shear)

Warps the image into a trapezoidal shape to simulate perspective distortion caused by oblique imaging.

・Perspective Transformation (Perspective)

Applies slight perspective changes to simulate variations in camera viewpoint.

The transformation intensity was kept low to prevent inconsistencies between the transformed images and their corresponding bounding boxes.

Color Space Augmentation

・HSV Adjustment (hsv_h, hsv_s, hsv_v)

Modifies the hue, saturation, and value of the image to simulate different weather and lighting conditions, such as sunny, cloudy, or dusk environments.

This enhances the model's ability to accurately detect cracks under varying illumination and to handle contrast changes caused by shadows or surface dirt.

Advanced Image Composition

・Mosaic

Randomly combines four images into a single composite training image.

This increases the diversity of crack counts, spatial arrangements, and background variations, thereby improving the model's generalization capability.

・MixUp

Blends two images into a single input image to reproduce complex overlaps between cracks and the pavement background.

This helps the model to learn crack boundaries more precisely, reducing both false positives and missed detections.

Data Augmentation Intensity

The intensity of data augmentation was configured at four levels: low, mid, high, and max.

The specific parameter settings for each intensity level are presented below.

Table 1: Strength of each data extension

	low	mid	high	max
degree	45	90	135	180
fliplr	0.25	0.5	0.75	1
flipud	0.25	0.5	0.75	1
hsv_h	0.25	0.5	0.75	1
hsv_s	0.25	0.5	0.75	1
hsv_v	0.25	0.5	0.75	1
mixup	0.25	0.5	0.75	1
mosaic	0.25	0.5	0.75	1
perspective	0.00025	0.0005	0.00075	0.001
scale	0.25	0.5	0.75	1
shear	5	10	15	20
translate	0.25	0.5	0.75	1

Evaluation Metrics

In Experiment 1, the performance of the crack detection model for asphalt pavement using UAV imagery was quantitatively evaluated using mean Average Precision at IoU 0.5 for bounding boxes (mAP50(B)) and Recall as evaluation metrics. These metrics are widely used in object detection tasks and enable a quantitative assessment of how accurately the model can detect cracks. The mAP50(B) provides a comprehensive measure of the balance between Precision and Recall, while Recall directly indicates the proportion of actual cracks that are correctly detected. High values for both metrics indicate a high-precision model that achieves both low false positives and low false negatives. In this study, evaluating these two metrics together allowed for a comprehensive assessment of model performance from both perspectives: false detection and missed detection. In particular, mAP50(B) is widely used for comparing the overall accuracy of object detection models, as it offers a more integrated performance evaluation than individual metrics alone [7].

3.4 Results of Experiment 1

Regarding metrics/mAP50(B)

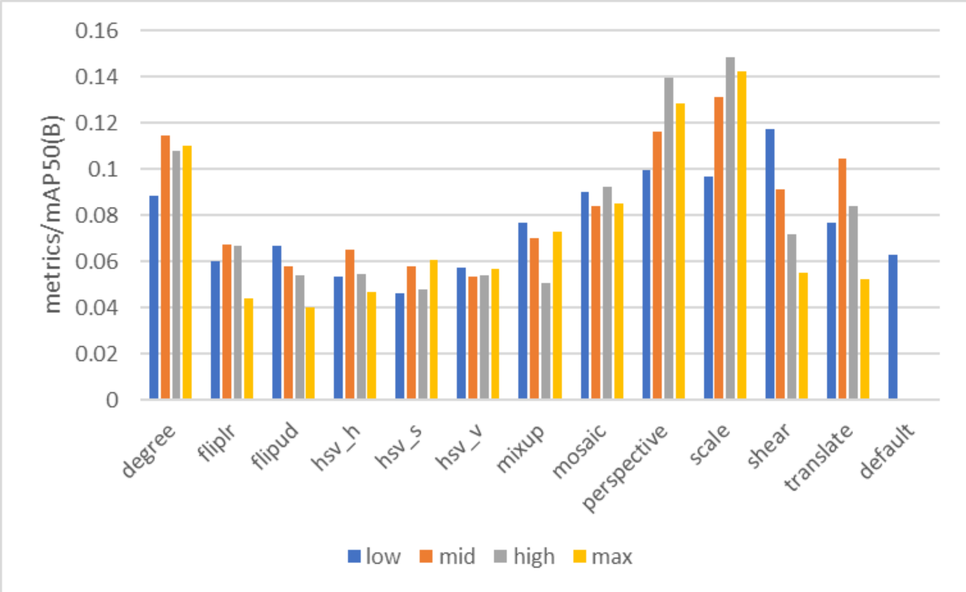

Figure 5: metrics/mAP50(B)

A comparison of mAP50(B) values was conducted for each data augmentation technique. As a result, scale and perspective demonstrated higher detection accuracy than the other methods, with the high-level scale setting achieving the best performance, reaching an mAP50(B) of 0. 148.In contrast, fliplr, flipud, hsv_h, hsv_s, and hsv_v consistently showed low accuracy, with all mAP50(B) values remaining below 0.07, indicating limited effectiveness. The degree transformation exhibited moderate accuracy, with a tendency for performance to improve as the intensity increased. On the other hand, for shear and translate, it was observed that excessively high intensity levels led to a decline in accuracy. These results indicate that, for crack detection using UAV-acquired pavement surface images, scale and perspective are the most effective augmentation techniques for improving model performance. Additionally, shear and translate can be beneficial at moderate intensity levels, but overly aggressive settings may have an adverse effect. Furthermore, color space augmentations based on hue, saturation, and brightness adjustments were found to have only limited impact on model performance.

Regarding metrics/recall(B)

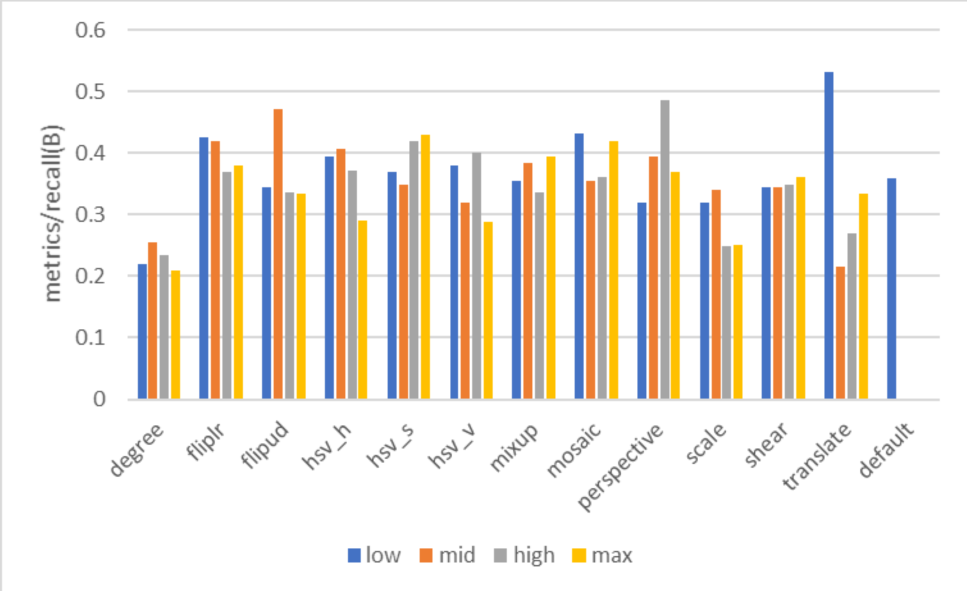

Figure 6: Metrics/Recall (B)

Figure 6 shows that overall recall generally distributed within the range of 0.2 to 0.55, with clear differences observed depending on the method and intensity setting. Notably, fliplr and mosaic consistently demonstrated high recall, maintaining values around 0.4. Since fliplr performs transformations close to real-world shooting conditions, it is effective in preventing detection omissions. Mosaic is also thought to contribute to improved detection performance by generating complex scenes. On the other hand, recall for degree (rotation) at all intensities fell below the default, indicating limited improvement. Additionally, scale and translate showed a tendency for recall to decrease as intensity increased. Excessive scaling may increase detection failures by diverging from actual object sizes. Particularly for translate, while it showed high recall at low intensity, recall dropped sharply at mid intensity and above. These results indicate that scale and translate are effective when used within moderate ranges, but excessive transformations negatively impact the model. Furthermore, the HSV-based adjustments (h, s, v) showed moderate effectiveness with recall values around 0.3 to 0.45. Specifically, hsv_s and hsv_v demonstrated slight performance improvements from mid to high levels, indicating that adjusting saturation and brightness yields consistent benefits. In summary, for crack detection using pavement images captured by UAVs, data augmentation centered on fliplr and mosaic is effective when prioritizing recall. Combining this with mild scale or HSV adjustments can further improve performance. Conversely, overly intense transformations like translate or shear are likely to increase detection failures and require careful configuration.

3.5 Experiment Comparing the Effects of Flight Altitude（Experiment 2)

We deployed the trained model from Experiment 1 in validation mode and performed inference on multiple image datasets acquired by a UAV at different heights (5m and 10m), then compared the results. This aimed to clarify the impact of varying capture heights on the model's inference accuracy. For evaluation, we used the model that achieved the highest mAP50 score among the trained models created with four intensity settings for each data augmentation. We validated images captured at 5m and 10m in validation mode and compared the differences in detection performance.

3.6 Results of Experiment 2

Using the trained model that demonstrated a high mAP50 score, we evaluated images captured at 5m and 10m in validation mode to compare detection performance. The comparison results are shown in Figures 7 and 8.

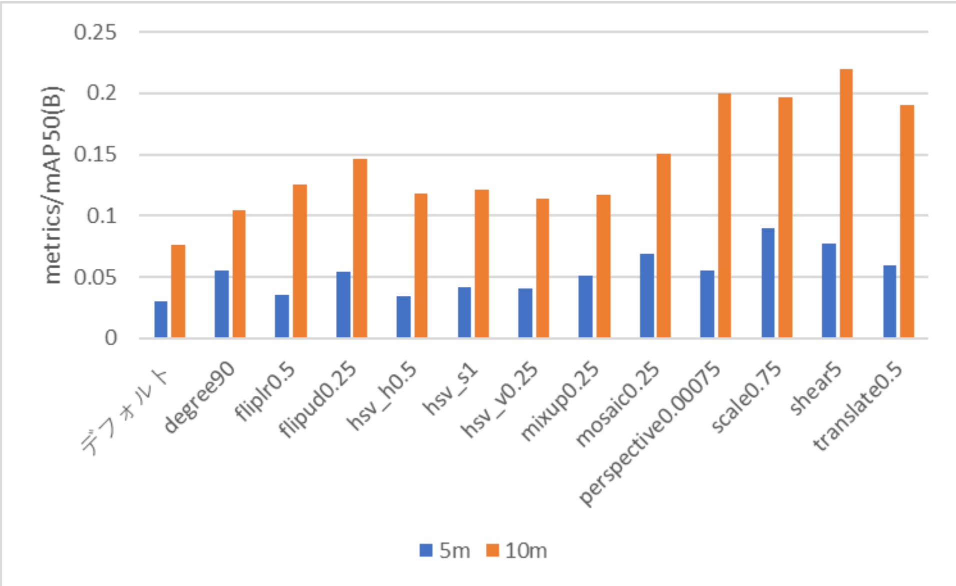

Figure 7: Comparison of mAP50 at 5m and 10m in the val directory

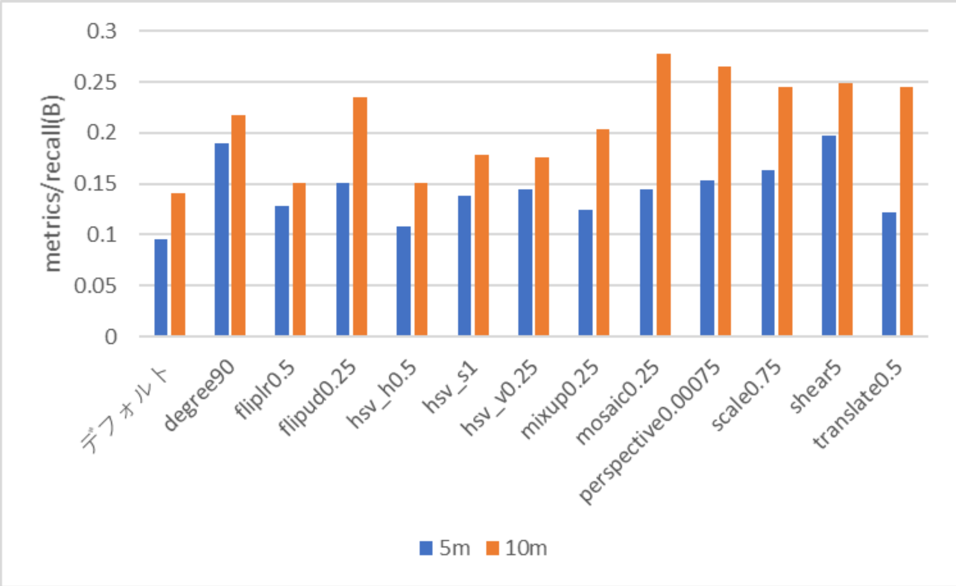

Figure 8: Recall comparison at 5m and 10m in the val directory

At 5m, mAP50(B) was generally low overall, and accuracy improvements through data augmentation were limited. However, slight improvements were observed with some geometric transformations like scale and shear. Scale yielded the highest accuracy in particular. At 10m, improvements through geometric transformations such as scale, shear, and translate were significant, with shear showing the highest accuracy. Furthermore, HSV-based color transformation parameters showed moderate improvement effects, but techniques like flip and perspective yielded only minor accuracy gains, indicating limited effectiveness.

4 Discussion

In the image verification at 5m and 10m altitudes using a UAV in Experiment 2, the results showed lower mAP50 and recall values for the 5m images. While lower altitudes should theoretically capture finer crack details more precisely, leading to improved inference accuracy, multiple factors likely combined to cause the model performance on 5m images to unexpectedly decline. This result suggests that simply lowering the altitude does not necessarily guarantee higher accuracy in UAV-based crack detection. First, the impact of fine-scale noise emerges as a significant factor. Low-altitude images capture minute details such as road surface irregularities, small stones, dirt, and intricate patterns beyond the cracks themselves. Such information can be misinterpreted as cracks by the learning model, leading to false detections. Therefore, at low altitudes, the risk increases that the model will learn erroneous features. Conversely, as the shooting altitude increases, the ground pixel size becomes larger. This makes the impact of noise caused by fine shadows and light reflections—which are more problematic at low altitudes—relatively smaller, leading to a stable improvement in detection accuracy. Next, differences in crack distribution are also thought to affect the verification results. At low altitudes, the area of paved road surface within the image field of view is narrower, tending to result in fewer cracks captured in a single image.

Consequently, the learning model may not fully grasp the evaluation target, potentially leading to a decrease in metrics like mAP and recall. Conversely, at 10m altitude, a wider area can be captured, making it more likely for multiple cracks to appear in a single image. This allows the model to evaluate the target more evenly. Consequently, metrics tended to be more stable and generally higher in the 10m images. Furthermore, regarding the effects of data augmentation, from the mAP perspective, scale and translate were confirmed to have a high effect on improving accuracy when set to appropriate intensities. These methods are thought to contribute to improved inference accuracy by providing the learning model with variations in crack size and position, enabling it to capture more generalizable features. On the other hand, from the recall perspective, while the overall improvement from each data augmentation method was not substantial, a certain effect was observed. However, regarding degree (rotation), recall values decreased compared to the default setting at all intensity levels, suggesting it does not contribute to improving inference accuracy. Therefore, it is advisable to avoid using degree or set it with caution. In summary, for pavement crack detection using UAV images, moderate use of scale and translate can be expected to improve accuracy, while degree carries a risk of performance degradation and requires caution. Furthermore, by considering the characteristics of images and the impact of noise at different flight altitudes, and by optimizing data augmentation and learning settings, it is possible to maximize the performance of the learning model. This enables stable crack detection with high accuracy while efficiently capturing a wide area, allowing for flexible selection of drone flight altitudes. Therefore, this method is expected to contribute to streamlining road inspection operations and reducing operational costs, while also being useful for optimizing practical flight plans and imaging conditions.

References

1 National Institute for Land and Infrastructure Management (2024). FY 2023 Report on the Status of Infrastructure Inspection Implementation [令和5年度 インフラ点検実施状況調査報告書, in Japanese]. Retrieved September 24, 2025, from https://www.nilim.go.jp/lab/bcg/siryou/2024report/ar2024hp.pdf

2 Cabinet Office (2020). Manual for Functional Diagnosis Surveys Utilizing UAVs (Draft) [UAVを活用した機能診断調査マニュアル（案）, in Japanese]. Retrieved September 24, 2025,from

https://www5.cao.go.jp/Keizai-shimon/kaigi/special/reform/ab1/20220411/shiryou194.pdf

3 Zhang, H., et al. (2025). YOLOv8-DGS: Real-time pavement crack detection based on improved deep learning. PLOS ONE. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0324512

4 Zeng, Y., et al. (2025). Efficient and accurate road crack detection technology based on YOLOv8-ES. ResearchGate.

https://www.researchgate.net/publication/388850505

5 Labellerr (2024). Understanding YOLOv8: Architecture, features and applications. https://www.labellerr.com/blog/understanding-yolov8-architecture-applications-features/

6 Ultralytics Inc. (2025). YOLO Data Augmentation Guide [YOLO データ拡張ガイド, in Japanese]. Retrieved July 23, 2025, from

https://docs.ultralytics.com/ja/guides/yolo-data-augmentation/

7 Ultralytics Inc. (2025). YOLO Performance Metrics [YOLO パフォーマンス指標の詳細, in Japanese]. Retrieved July 25, 2025, from

https://docs.ultralytics.com/ja/guides/yolo-data-augmentation/