

Article

Development of Deep Intelligence for Automatic River Detection (RivDet)

Sejeong Lee, Yejin Kong and Taesam Lee * 

Department of Civil Engineering, Gyeongsang National University, 501 Jinju-daero, Jinju 52828, Republic of Korea; dltpwjd312@gnu.ac.kr (S.L.); 2021080114@gnu.ac.kr (Y.K.)

* Correspondence: tae3lee@gnu.ac.kr; Tel.: +82-55-772-1797

Abstract: Recently, the impact of climate change has led to an increase in the scale and frequency of extreme rainfall and flash floods. Due to this, the occurrence of floods and various river disasters has increased, necessitating the acquisition of technologies to prevent river disasters. Owing to the nature of rivers, areas with poor accessibility exist, and obtaining information over a wide area can be time-consuming. Artificial intelligence technology, which has the potential to overcome these limits, has not been broadly adopted for river detection. Therefore, the current study conducted a performance analysis of artificial intelligence for automatic river path setting via the YOLOv8 model, which is widely applied in various fields. Through the augmentation feature in the Roboflow platform, many river images were employed to train and analyze the river spatial information of each applied image. The overall results revealed that the models with augmentation performed better than the basic models without augmentation. In particular, the flip and crop and shear model showed the highest performance with a score of 0.058. When applied to rivers, the Wosucheon stream showed the highest average confidence across all models, with a value of 0.842. Additionally, the max confidence for each river was extracted, and it was found that models including crop exhibited higher reliability. The results show that the augmentation models better generalize new data and can improve performance in real-world environments. Additionally, the RivDet artificial intelligence model for automatic river path configuration developed in the current study is expected to solve various problems, such as automatic flow rate estimation for river disaster prevention, setting early flood warnings, and calculating the range of flood inundation damage.



Academic Editor: Konstantinos X. Soulis

Received: 18 December 2024

Revised: 14 January 2025

Accepted: 17 January 2025

Published: 20 January 2025

Citation: Lee, S.; Kong, Y.; Lee, T. Development of Deep Intelligence for Automatic River Detection (RivDet). *Remote Sens.* **2025**, *17*, 346. <https://doi.org/10.3390/rs17020346>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: river; artificial intelligence; Roboflow; YOLOv8; augmentation

1. Introduction

Owing to the impact of recent climate change, the scale and frequency of extreme rainfall events and sudden downpours have increased, resulting in the development of floods [1–3]. As a result, extreme floods and various river disasters have intensified, so it is crucial to develop technologies to mitigate such extremes. Recent technical developments in the Fourth Industrial Revolution, such as unmanned aerial vehicles (UAVs), big data, and artificial intelligence, are being applied in various fields [4,5].

In particular, artificial intelligence has overcome various limits by implementing deep learning techniques on the basis of extensive amounts of data. Although there have been various attempts to apply artificial intelligence for river management, difficulties in applying technologies such as artificial intelligence exist due to the limited data available for rivers [6]. Additionally, river basins with steep slopes or dense vegetation are difficult

for people to access, and it is difficult to document a wide range of river survey information, as it takes a long time to obtain information from a broad area with conventional surveying approaches. Until now, river surveying has been conducted manually using GPS equipment. Various accidents have occurred during manual surveys, leading to injuries. Using UAVs could reduce the risk of human accidents. Additionally, by employing AI in UAV surveys, extensive river information and locations that are difficult to access can be obtained in a relatively short time using optimal routes, minimizing drone battery consumption and time wastage. Through this river detection artificial intelligence method (RivDet-AI), automatic flight can be adopted, and river surveying can be made much easier and faster. Therefore, we aimed to develop a RivDet artificial intelligence model in the current study to automatically detect river areas.

Several studies have investigated the application of machine learning (ML) and deep learning (DL) for identifying rivers and detecting floods, highlighting their potential to enhance flood modeling, hazard mapping, and real-time monitoring. A review of recent literature shows that DL techniques, in particular, have demonstrated significant improvements in flood modeling accuracy compared to traditional methods. In [7], hydrogeomorphic metrics were used for high-resolution fluvial landscape analysis to perform efficient feature extraction and interpretation. In [8], U-Net neural network models were utilized to extract hydrographic features and provide significant implications for hydrologic modeling. Furthermore, In [9], a 2D analysis was employed to unravel the spatial heterogeneity of inundation pattern domains, while in [10], transfer learning with convolutional neural networks was applied to delineate hydrological streamlines.

These methods are increasingly used to predict flood inundation and assess flood-prone areas by leveraging remote sensing data such as satellite imagery and unmanned aerial systems (UAS) to enhance flood detection accuracy. For instance, [11] investigated the use of transfer learning and water segmentation in river-level monitoring, demonstrating that these approaches enable automated and more accurate flood mapping and showed that incorporating remote sensing data from SAR (Synthetic Aperture Radar) and optical satellite imagery significantly enhances the detection of flood extent and water levels. Similarly, ref. [12] highlighted that when these techniques are integrated with hydrodynamic models, they offer promising tools for real-time flood forecasting by efficiently handling large datasets and complex flood dynamics.

In this study, the YOLO model, a type of deep learning algorithm within the realm of machine learning, is intended to be used. Research on detection technologies using the YOLO model has been conducted in various fields. In [13,14], they studied waste detection via UAV-based systems by modifying the loss function of YOLOv3 using Darknet-53 as the backbone network for stronger feature extraction. In [15], a monitoring system that detects waste in real time on beaches and at sea via UAVs was proposed by dividing images into multiple grid cells through a single neural network pass, predicting the likelihood of each cell containing an object. Additionally, a technology for detecting illegal logging sites in river basins via YOLOv5 was developed in [16], which includes techniques such as Mosaic Augmentation and Auto-Learning Bounding Box Anchors from its previous versions, and DeepLabv3+, a deep learning-based image segmentation model that classifies each pixel of an image.

A YOLO model capable of performing object detection in real time at high speed is a single unified model that performs image segmentation, object boundary box extraction, and object classification at once, and since it does not go through multiple steps like other models, its implementation is simple and efficient, providing a solution to these problems. Moreover, by integrating data augmentation techniques, we can improve the diversity of training data, resulting in a more robust and accurate river detection model. However, river

detection technology using the YOLO model has not been widely studied because of the limited aerial images of rivers. Although various machine learning (ML) and deep learning (DL) methods have been applied for object detection, river environments face limitations such as insufficient data for training, challenges in accessing complex river environments, and issues related to seasonal changes, variations in water volume, and vegetation growth. However, river detection technology using the YOLO model has not been widely studied because of the limited aerial images of rivers. The river environment presents various complex variables, such as seasonal changes, variations in water volume, and vegetation growth. Therefore, simply applying waste detection or illegal logging detection models to river detection is limited. To apply these systems to river detection, specialized data and models tailored to river environments are required.

Automated detection of UAV flight paths is particularly valuable for river surveying using aerial photogrammetry. If automated detection of river paths in UAV-based aerial surveys becomes feasible, it could prove highly beneficial for river engineers. However, this application has not been fully explored due to limited data and a lack of comprehensive studies in this specialized field. Therefore, an artificial intelligence model (RivDet) was developed in the current study via the YOLOv8 model, the latest standard version developed by Ultralytics in January 2023, which includes anchor-free detection and various upgraded features and is widely adopted in several fields [17,18]. Machine learning-based models are strongly influenced by the characteristics of the images used as input data for training, and securing high-quality images for training is vital. Aerial high-quality images from UAVs addressing river basins might not be enough to train the target AI model (i.e., RivDet). To overcome these limitations, the application of data augmentation techniques has been proposed as a method for generating new images. This approach aims to increase the diversity of training data, thereby facilitating the development of more robust artificial intelligence models.

2. Study Area

In the present study, susceptible locations where flood disaster prevention research has been conducted due to typhoons or guerrilla rainfall caused by climate change [19–21] across South Korea were selected as research sites. The river network in South Korea consist of 73 national rivers spanning 3602 km and 3842 local rivers covering 25,972 km (http://nationalatlas.ngii.go.kr/pages/page_1273.php), accessed on 14 August 2024. A total of 13 rivers were set as research sites, as described in Figure 1. The latitudes and longitudes of the nine rivers in Gyeongsangnam-do, as well as the two rivers in Gyeongsangbuk-do at the Andong River Experimental Center (Andong Kict), Poricheon Stream, and the two rivers in the Bophwacheon Stream and Chogangcheon Stream in Chungcheongbuk-do are presented. From these research sites, a total of 4177 UAV aerial survey photographs with 5472×3648 resolution and 1070×580 resolution were acquired using the Autel's evo2 Enterprise RTK drone and DJI's Phantom 4 Pro Version 2 drone from 2021.06 to 2024.04.

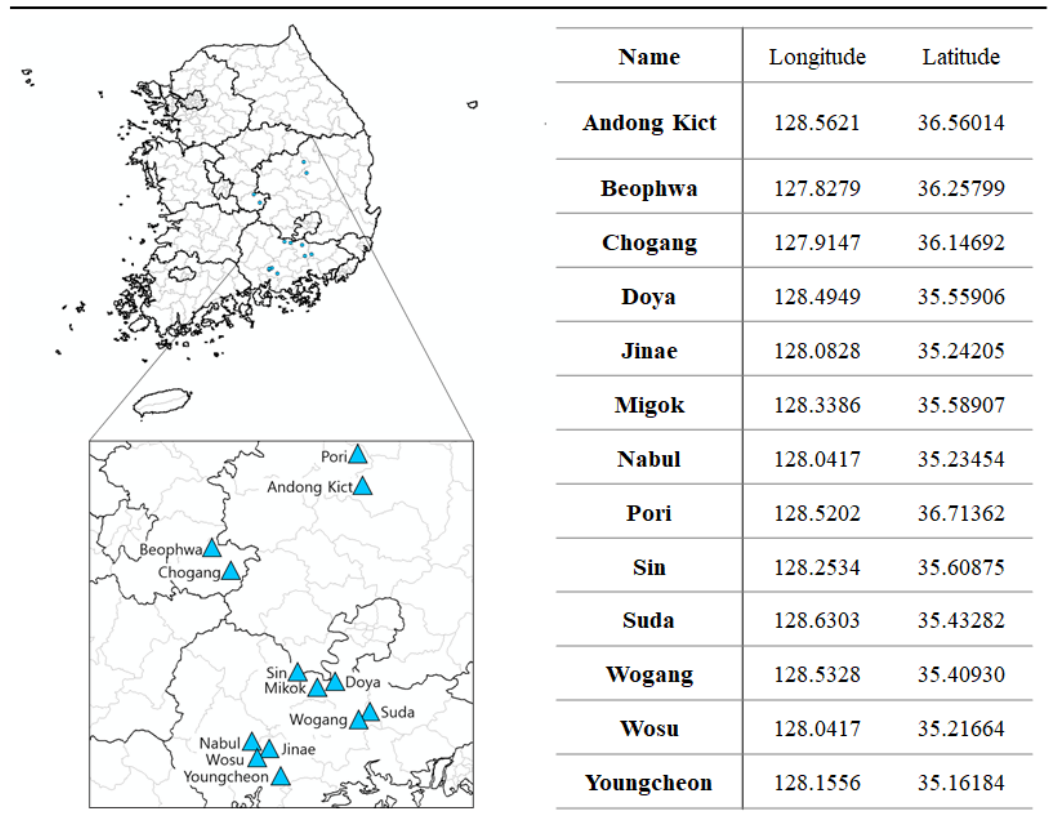


Figure 1. Map of the study area with each river name and its geological information, including latitude and longitude.

3. Development of River Detection Artificial Intelligence

In the present study, the YOLOv8 model on the Roboflow [22,23] platform was utilized for training the target UAV images for rivers and evaluating their performance. The procedure is illustrated in Figure 2. Its procedure is separated into three steps.

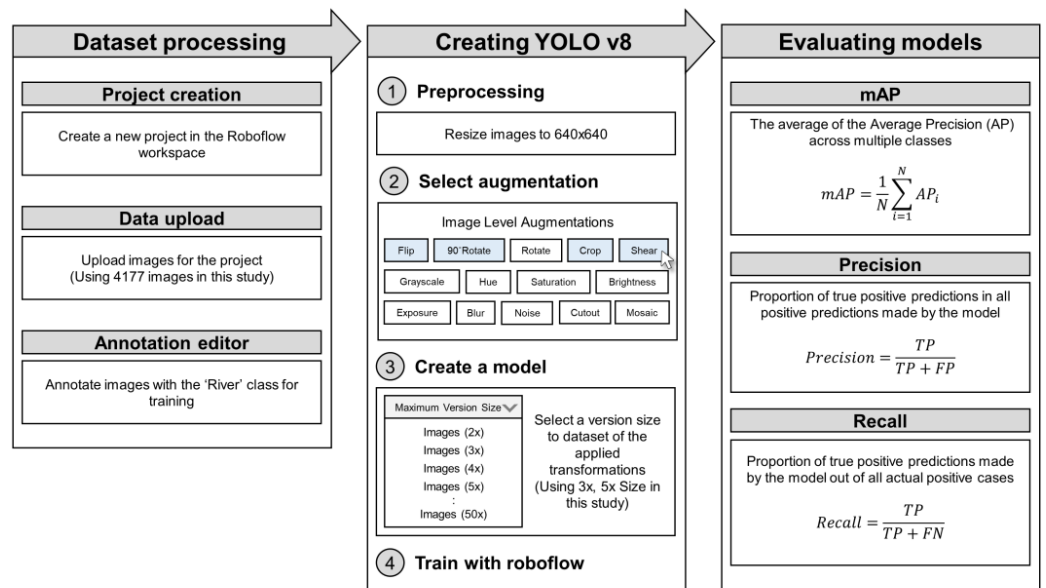


Figure 2. Process of model development via the YOLOv8 model on the Roboflow platform for river detection.

YOLOv8 is a neural network model for object detection, used for applications such as license plate recognition and airplane detection. It consists of three sections: the backbone, the neck, and the head. The backbone is a deep learning architecture that extracts features from the input image, while the neck combines the layer's features obtained from the backbone. Additionally, the head predicts the classes and bounding boxes of objects generated by the object detection model. The operation process of YOLOv8 is illustrated in Figure 3, based on materials provided by Ultralytics (<https://blog.roboflow.com/what-is-yolov8/>) (accessed on 2 January 2025).

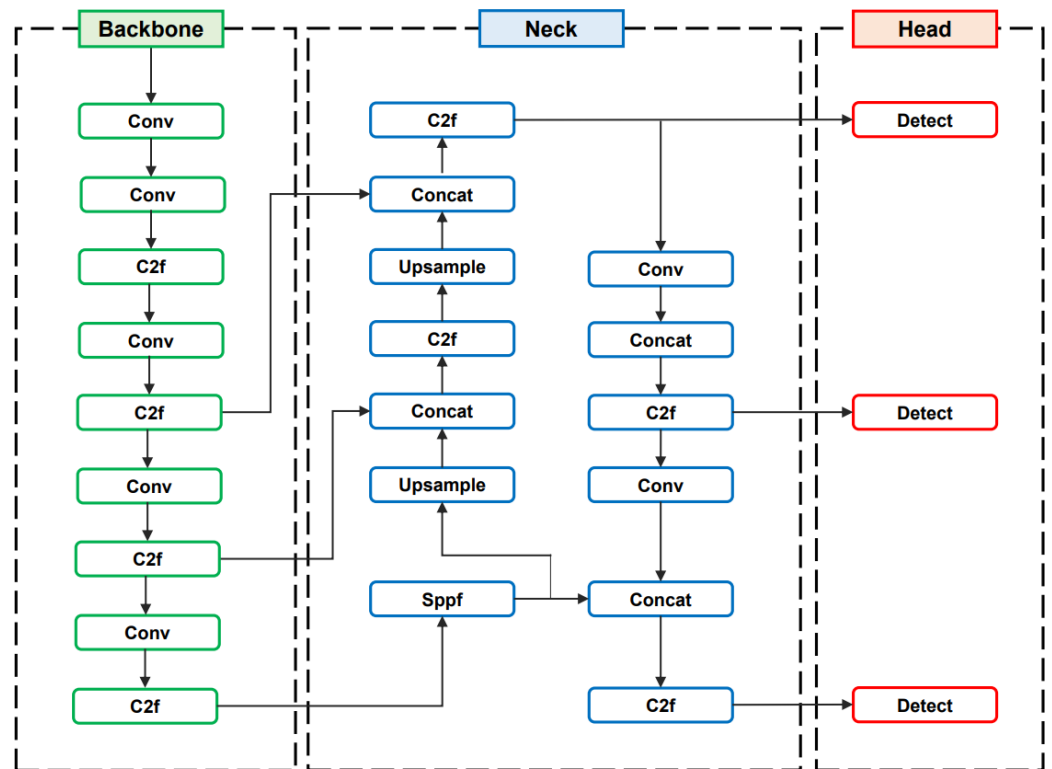


Figure 3. The operation process of the yolov8 model. Note that the process has been divided into three sections: backbone, neck, and head.

Conv (convolutional) layers are fundamental building blocks in neural networks. They apply a filter (or kernel) to the input image to create feature maps, capturing spatial hierarchies in the data. In YOLOv8, convolutional layers are used extensively in the backbone and head of the network. The c2f module is a variation in the CSP (cross stage partial) bottleneck. It consists of two convolutional layers and is designed to reduce computational complexity while maintaining performance. The c2f module helps in feature extraction and fusion. The concat operation combines multiple feature maps along the channel dimension. This is useful for merging information from different layers or branches of the network, allowing the model to learn more complex features. The upsample feature is the process of increasing the spatial resolution of feature maps. It is used in YOLOv8 to restore the original image size after downscaling during feature extraction. This helps in precise localization of objects in the final output. The detection module is responsible for the final object detection. It takes the processed feature maps and predicts bounding boxes, class probabilities, and other relevant information for each detected object (<https://abintimilsina.medium.com/yolov8-architecture-explained-a5e90a560ce5>) (accessed on 2 January 2025).

When an image is inputted into the network, it is first processed through the backbone. Features are extracted at multiple scales through convolutional layers and c2f modules.

Here, the sppf block is used after the last convolutional layer of the backbone to generate fixed feature representations of objects in various sizes without resizing the image or causing spatial information loss. The extracted features are then processed through the neck section, where concat and upsample operations combine the feature maps. The upsample feature maps are processed by the neck section, where the detection module generates the final output. The detection module includes the bounding boxes and class probabilities of the objects. Each block contains three detection blocks specialized in detecting different object sizes. The first detection block handles small objects from the c2f block. The second detection block handles medium-sized objects from the c2f block. The third detection block handles small objects from the c2f block again. This structure allows YOLOv8 to efficiently detect and predict objects.

3.1. Image Data Preparation

To conduct training on the Roboflow platform, first, a new project must be created in the workspace, and the images to be used must be subsequently uploaded. In the present study, a total of 4177 photographs of actual river sites were adopted, which can be categorized into artificial and natural rivers [24]. The 'River' class is then injected into the image as a bounding box via the Annotation Editor [25] in Figure 4. Here, the bounding box is a tool for displaying specific objects in an image. In this mode, crosshairs help to determine where to start drawing. A new annotation is created by clicking and dragging across an image; then, the class selector is employed to choose its label. Once this process is complete, dataset preparation for machine learning is complete.

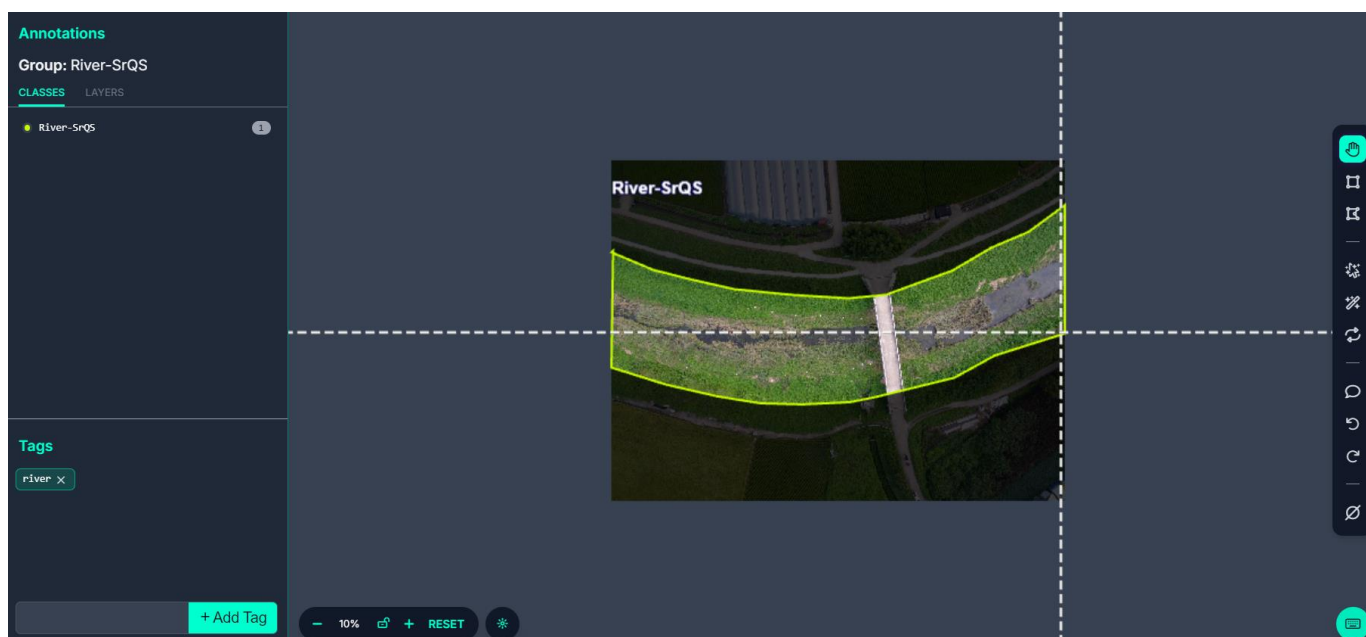


Figure 4. Example of inserting the 'River' class in Roboflow.

3.2. AI Model (RivDet) Creation and Training

A new river detection AI (RivDet) model can be created on the basis of the prepared image dataset. The dataset is divided into train set, valid set, and test set. The dataset is classified into three categories using Roboflow's random sampling feature. By default, Roboflow allocates 70% of the data to the train set, 20% to the validation set, and 10% to the test set. In this process, the images require adjustment and normalization before being inputted into the YOLOv8 model [26]. This involves applying a 640×640 transformation of all images in the dataset to reduce training time and improve performance [27]. It is

essential to unify and downsize all the UAV images into one resolution for training the AI model. Images with different resolutions should be unified to combine all the photographs for training an AI model, and the training speed can be accelerated by downsizing the images with respect to the river feature.

Furthermore, training a RivDet AI model requires a large amount of data to obtain satisfactory results, although more than 4000 images have been adopted in the current study. The augmentation features of the Roboflow platform can be highly adoptable, and its usage has improved model performance [14,28–31]. Therefore, a data augmentation procedure was applied in the current study, and we analyzed its relative performance thoroughly. Roboflow offers numerous image-level augmentation operations, including flip, rotate, crop, shear, grayscale, hue, saturation, brightness, exposure, blur, noise, cutout, and mosaic operations.

Horizontal or vertical flipping inverts the river image while allowing the model to recognize objects in the mirror image of the data, as shown in Figure 5a. This is particularly useful in cases where the object may appear in the opposite direction [28]. River images from UAV aerial surveying often cannot be obtained from both sides. Therefore, this flip feature can be useful when rivers are perceived in the opposite direction. The 90° rotate operation turns the image by 90°, enabling the model to recognize objects from various angles (see Figure 5b). This is important for aerial or satellite images where the river direction can vary at right angles [32].

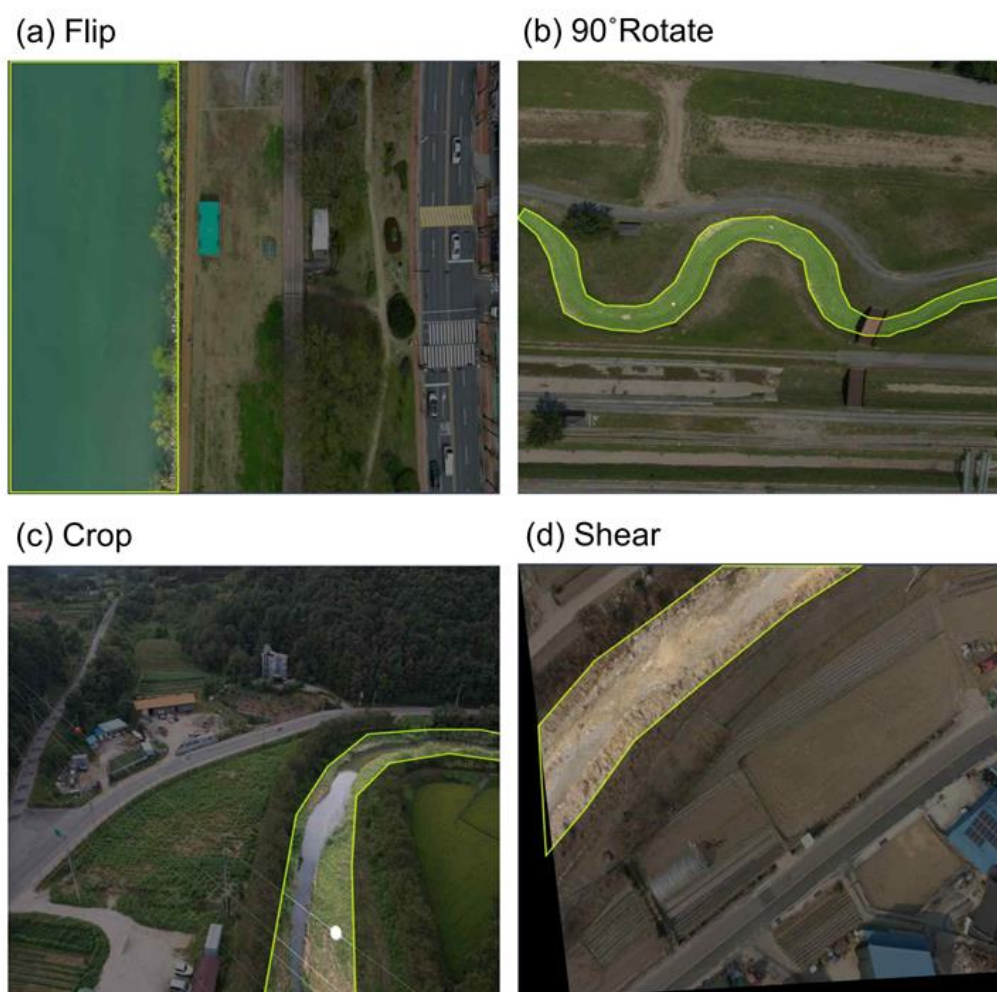


Figure 5. Example of augmentation for a single case. Note that the left plot represents the basic model, whereas the right plot represents the model with the corresponding augmentation applied: (a) flip, (b) 90° rotate, (c) crop, and (d) shear.

The crop operation removes arbitrary parts of the river image, focusing the model on different parts of the object and allowing it to detect objects that are partially visible or obscured, as shown in Figure 5c. Additionally, the cutout operation randomly removes some parts of the river image, enabling the model to recognize objects despite missing information [33]. The shear operation tilts the river image to simulate cases where an object appears distorted owing to changes in perspective, as shown in Figure 5d. Through this approach, the model is trained to recognize distorted rivers as well [34].

A comprehensive set of transformations was provided to mimic various scenarios in which objects appear with different directions, sizes, and perspectives in real-world settings. To this end, a group of augmentation operations—flip, 90° rotate, crop, and shear—was utilized in the current study. The other combinations were further tested, but no better results were obtained. Therefore, we presented those four augmentation features in the current study.

The selected combination of augmentations further allows a multiplicative increase in the amount of input data. This process involves reviewing the selected items and choosing a version size to create a moment-in-time snapshot of the dataset via the applied transformations. Increases from 2× to 50× are possible. Although larger increases increase the training duration, they might result in better model performance. In the present study, an increase of up to 5× was applied. To increase the diversity of the input model data, combinations of four augmentations were applied at both a maximum increase of 5 and an intermediate increase of 3×.

3.3. Model Evaluation

The performance of the object detection model was assessed by summing the differences between the mean average precision (mAP), precision, and recall values of the model without augmentation and those of the model with augmentation. The mAP is the average of the average precision (AP) for multiple classes and is used to comprehensively evaluate the performance of an object detection model. The AP is calculated for each class; then, the average of these APs is determined. The AP is computed as the area under the curve (AUC) of the precision-recall graph, and the precision values must be averaged across all recall values. The mAP is calculated as

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (1)$$

Precision is the ratio of actual true instances among those predicted as true by the model, indicating the accuracy of the predictions. The precision is calculated as

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Here; true positive (TP) refers to cases where the model accurately detects an object; false-positive (FP) refers to cases where the model detects an object that is not present in the image; false-negative (FN) refers to cases where the model fails to detect an object that is present; and true negative (TN) refers to cases where the model correctly identifies that no object is present in the image. Recall is the proportion of actual true instances that the model predicted as true, indicating how well the model identifies actual positive cases. The recall is calculated as

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

4. Result

The mean average precision (mAP) graph is represented in Figure 6, as a result of the basic model. A higher mAP value indicates better model performance. Therefore, during the learning process, the mAP graph should show an increasing trend. If the mAP value decreases or fluctuates significantly, it may indicate overfitting by the model. Box loss represents the prediction error of the bounding box. The lower this value is, the more accurately the model can predict the location of the bounding box. Class loss (CL) measures how well the class is predicted by the model (i.e., the type of object detected) and matches the actual class. The object loss (OL) measures how well the model determines whether an object is present. All loss values should show a decreasing trend. If the loss value increases or fluctuates significantly, it may mean that the model is overfitting. The result indicates that the employed images are sufficient to train the RivDet AI model with no overfitting.

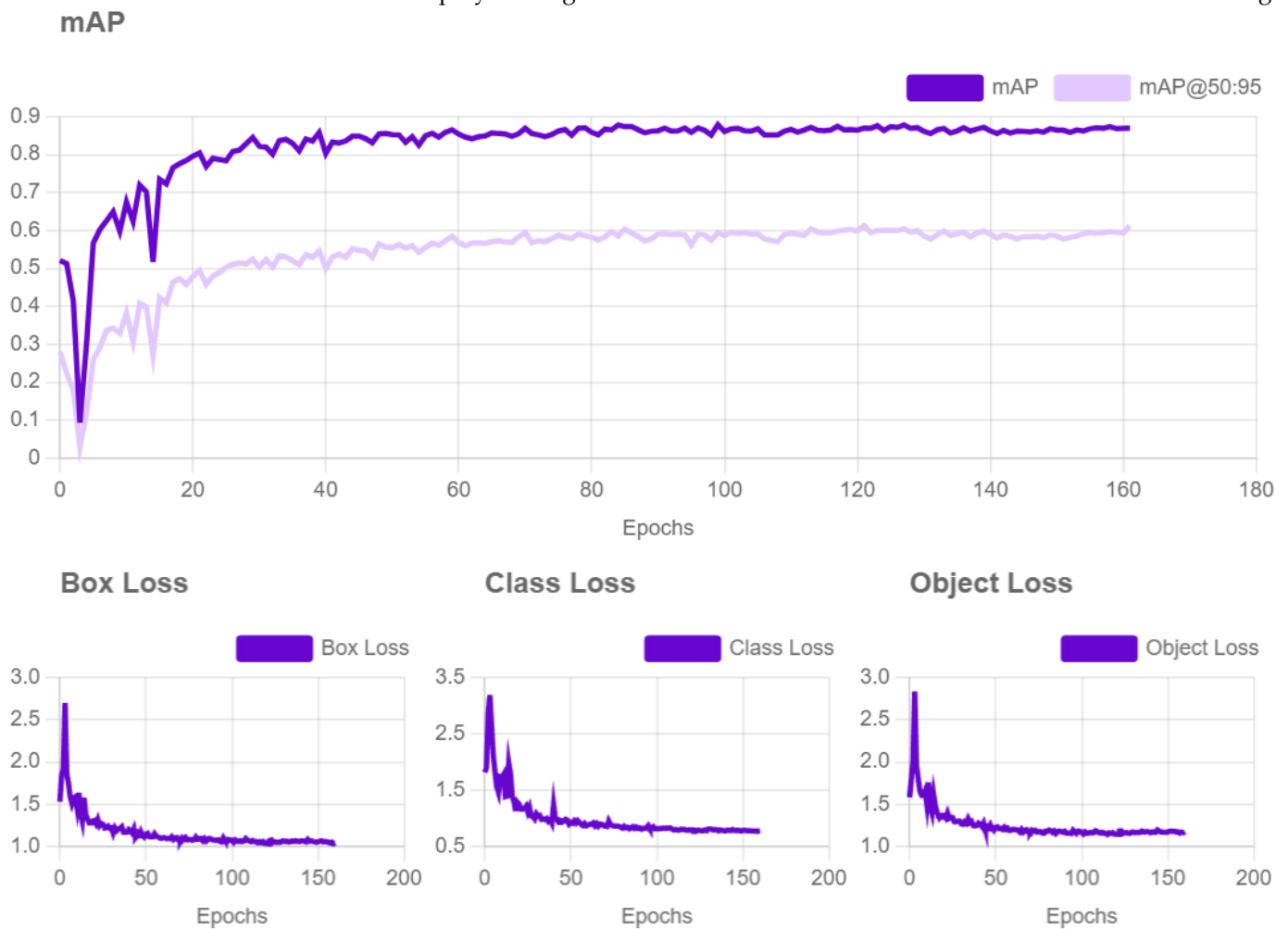


Figure 6. mAP calculation during the training procedure through each epoch for the basic model.

In the present study, we further analyzed two significant features during RivDet training. First, the existence of river levees was investigated since these hydraulic structures can significantly affect the overall identification of river areas. Although this approach might help to detect rivers when they actually exist and are included in an image, this feature can reduce the detection of river areas when they do not exist in testing data. Therefore, a comparative study with and without levees in training data was performed.

Furthermore, the augmentation feature was thoroughly investigated since it is a highly useful technique for improving model performance and covering limited UAV river images. The results of these two features are discussed in the following two subsections.

4.1. Comparative Analysis with River Levees

As discussed, a total of 4177 measured river models were analyzed via the Roboflow platform with further data augmentation. In the process of determining feature indicators as bounding boxes in images, the model that included river levees as feature indicators was compared and analyzed with the model that did not include river levees. In the analysis of the models that included river levees, 3195 out of the actual 4177 measured images that contained river levees were used.

The prediction results of the bounding box for the river model including levees were a mAP of 79.5%, a precision of 79.1%, and a recall of 76.8%, whereas the prediction results of the bounding box for the river model without levees were a mAP of 87.20%, a precision of 81.60%, and a recall of 82.00%, as presented in Table 1.

Table 1. Performance of the RivDet AI model with or without river levees.

	mAP	Precision	Recall
With Levees	0.795	0.791	0.768
Without Levees	0.872	0.816	0.820

Overall, the predicted values of the bounding box without river levees were higher than those with river levees. This suggests that levees in rivers can increase structural complexity within the image, making it difficult for the model to identify objects accurately. Furthermore, structures such as river levees can disrupt the flow within the image and cause localized erosion, altering patterns related to water flow, which in turn can affect the ability of the model to identify the boundaries of rivers [35]. On the basis of these results, this study selected the model without river levees as the basic model and proceeded to predict bounding boxes.

4.2. Augmentation

Given the nature of deep learning-based image classification models, which learn from the diverse features of images to classify targets, it is expected that performance can be enhanced through the training of various image characteristics. To this end, model development was conducted by utilizing the augmentation feature and comparing it with the basic model. Augmentation can improve the model ability by further generalizing the images through a variety of transformations [36].

Usually, UAV photogrammetry is carried out during the day or in sunny weather, and image data taken mainly during the day provide relatively high accuracy [37]. Most of the images used in this study were also taken during clear weather. To analyze only the performance of simple shape transformations except for the color or brightness control function, which is useful for images taken at night or in cloudy weather, four out of fourteen augmentation methods were applied, and models for each augmentation method were developed accordingly. Each model was classified on the basis of the application of one augmentation function: flip, 90° rotate, crop, and shear functions. Also, their combinations for two augmentation functions were tested: flip and 90° rotate (see Figure 7a), flip and crop (see Figure 7b), flip and shear (Figure 7c), 90° rotate and crop (see Figure 7d), 90° rotate and shear (see Figure 7e), and crop and shear (see Figure 7f).

Additionally, the models with three augmentation features were categorized as flip and 90° rotate and crop, as shown in Figure 8a; flip and 90° rotate and shear, as shown in Figure 8b; flip and crop and shear, as shown in Figure 8c; and 90° rotate and crop and shear, as shown in Figure 8d, as well as a model that applies four augmentation features, as shown in Figure 8e.



Figure 7. Example of the augmentation for the cases with two augmentations: for each panel, the left plot represents the original photo, whereas the right plot represents the model with the corresponding augmentation applied: (a) flip and 90° rotate, (b) flip and crop, (c) flip and shear, (d) 90° rotate and crop, (e) 90° rotate and shear, and (f) crop and shear.

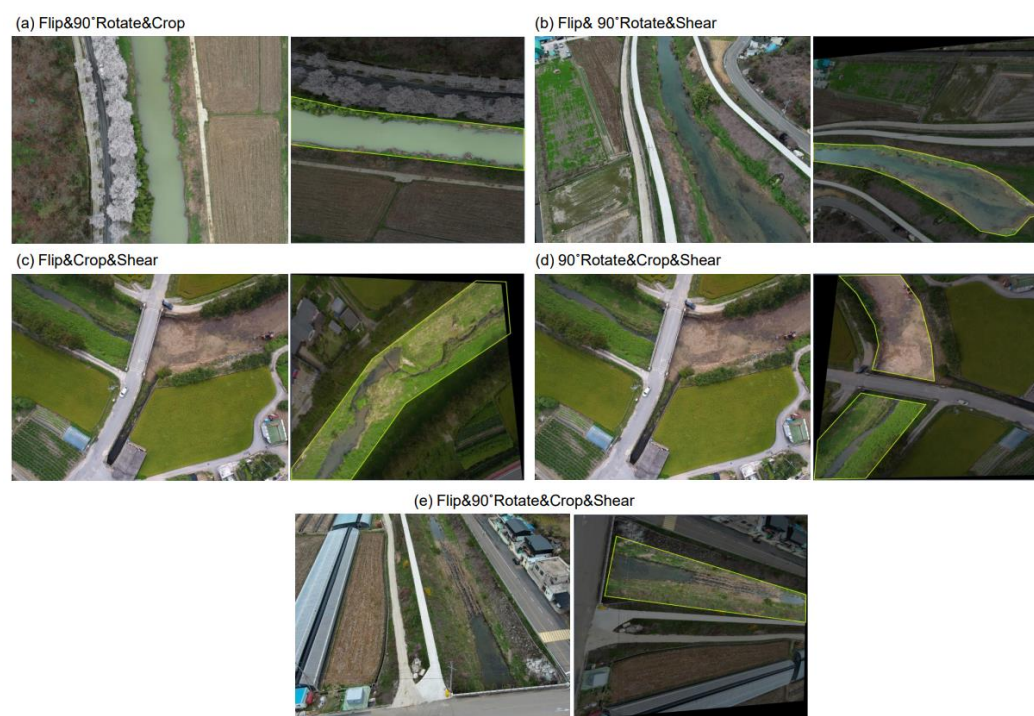


Figure 8. Example of the augmentation for the cases with three (a–d) and four (e) augmentations. Note that in each panel, the left plot represents the original photo, whereas the right plot represents the model with the corresponding augmentation.

4.2.1. One and Two Augmentations

The number of training data samples used in the study, learned through the augmentation feature of the Roboflow platform, is based on the basic model with all adopted images.

Augmentation of each method can be applied to increase the number of images by multiple times [38]. In this study, three times and five times the number of original images are used. The maximum value provided by the augmentation function of the starter plan from the Roboflow platform is five times the reinforcement, and the minimum value is three times the reinforcement excluding two times the basic provision. Therefore, learning was conducted three times or five times.

For models applying a one augmentation function, the counts are as follows. The flip operation was applied three times with 8027 images and five times with 10,211 images; the 90° rotate operation was applied three times with 9283 images and five times with 12,156 images; the crop operation was applied three times with 10,029 images and five times with 15,881 images; and the shear operation was applied three times with 10,027 images and five times with 15,863 images.

For models applying two augmentation functions, the counts are as follows: the flip and 90° rotate operations were applied three times with 9866 images and five times with 14,822 images, and the flip and crop, 90° rotate and crop, and crop and shear operations were applied three times with 10,029 images and five times with 15,881 images. Additionally, the flip and shear and 90° rotate and shear models were applied three times with 10,027 images and five times with 15,873 images. The learning results for each model were displayed via a heatmap, as shown in Figures 9 and 10. Heatmaps utilize color to allow for a quick visual assessment of data distributions and patterns, and they intuitively present the interrelationships between models that influence each other [39].

Figures 9 and 10 present the interrelated data results for mAP, precision, and recall. Figure 9 shows the results of three times the input data using the augmentation feature of the Roboflow platform, whereas Figure 10 shows the results of five times the input data. The x-axis and y-axis represent the augmentation functions used in the current study, and the values in each table indicate the correlation data between the axes. The range of mAP values is 0.870×0.893 , the range of precision values is 0.798×0.838 , and the range of recall values is 0.790×0.848 .

According to the $3 \times$ mAP results in Figure 9, the 90° rotate and shear model scored the highest, whereas the flip model scored the highest according to the $3 \times$ precision results, and the flip and 90° rotate and 90° rotate and crop models scored the highest according to the $3 \times$ recall results. Similarly, as shown in Figure 10, the flip and shear model scored the highest according to the $5 \times$ mAP results, the crop and shear model scored the highest according to the $5 \times$ precision results, and the shear model scored the highest according to the $5 \times$ recall results. These results indicate that certain augmentation methods may be more advantageous for specific metrics. For example, while the flip augmentation positively impacts precision, the combination of 90° rotate and shear may increase the mAP. However, one cannot rely solely on a single metric when selecting the optimal model [40]. Since each metric evaluates different aspects of the model, the most effective augmentation method may vary according to the metric. This implies that multiple metrics must be considered comprehensively when optimizing a model. For example, a model with a high mAP might actually have low precision or recall, which could affect performance in practical applications. Therefore, when evaluating and selecting a model, a balanced approach that considers various metrics, such as mAP, precision, and recall, is essential. Additionally, it is necessary to identify the most critical metric for a specific task or application and adjust the model accordingly.

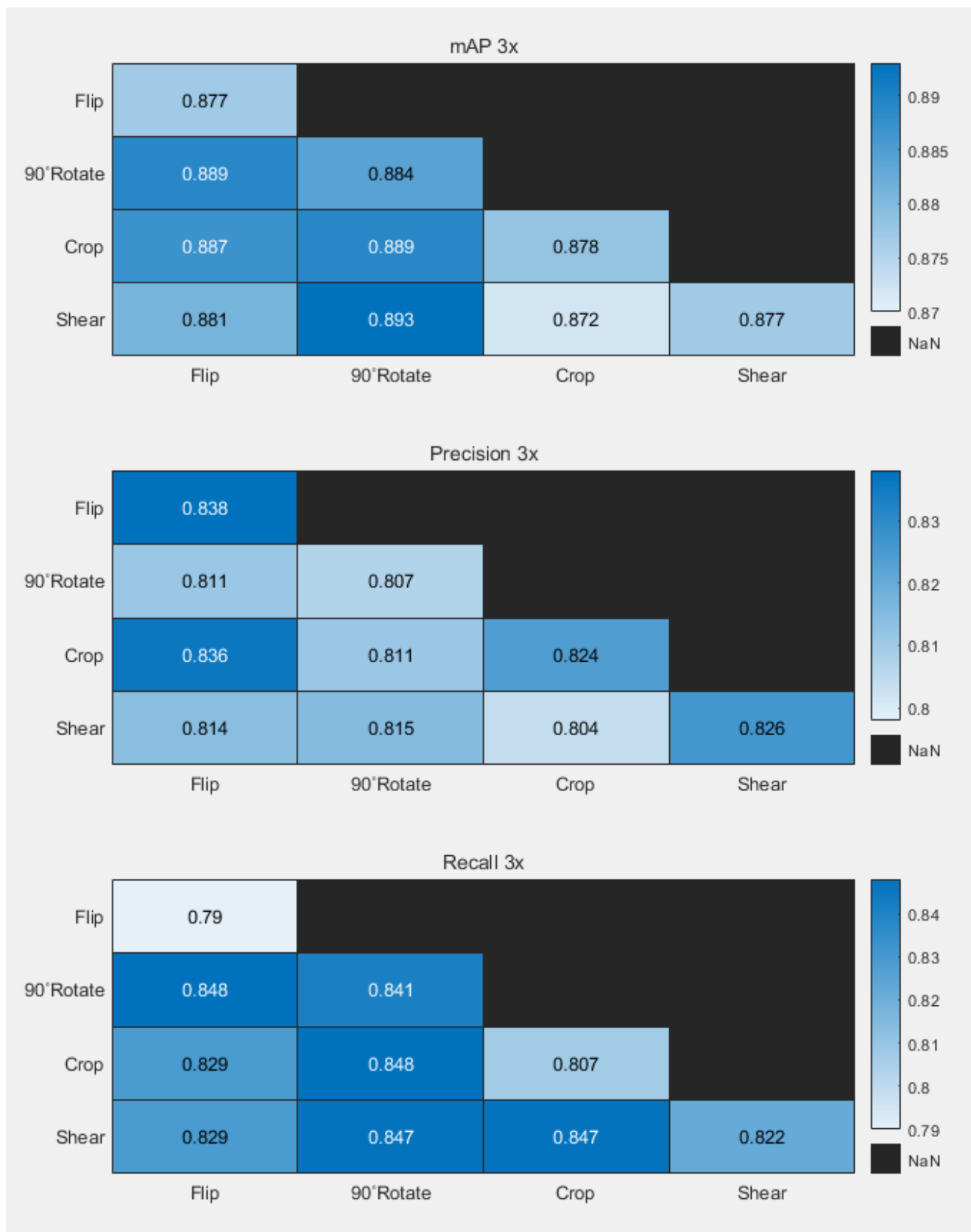


Figure 9. Heatmap of each augmentation effect on the model performance metrics: the corresponding value between each augmentation model displays the mAP, precision, and recall values for the models with three times augmentations of flip, 90° rotate, crop, or shear.

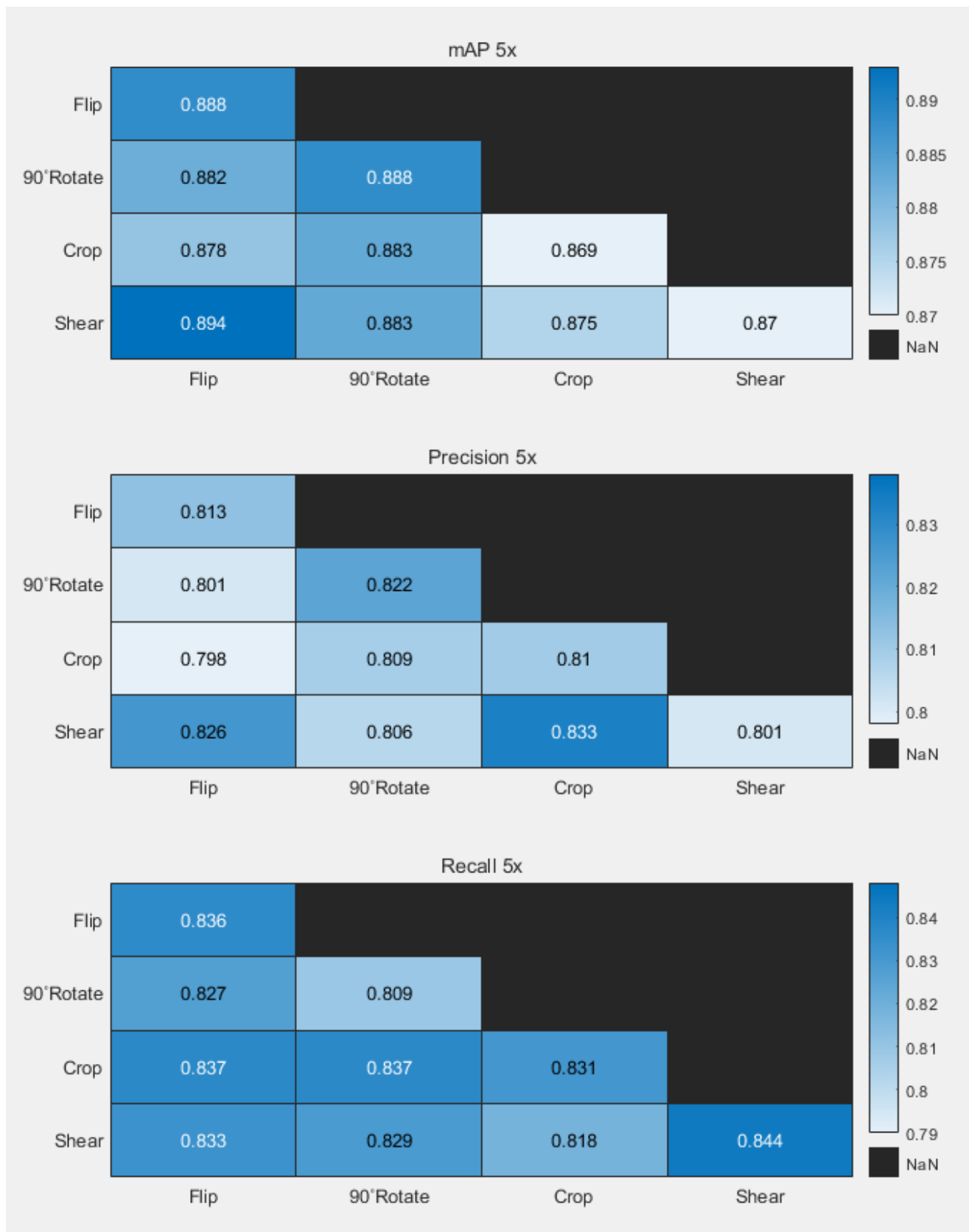


Figure 10. Heatmap of each augmentation effect on the model performance metrics. The corresponding value between each augmentation model displays the mAP, precision, and recall values for the models with five times the amount in data and augmentations of flip, 90° rotate, crop, and shear.

To facilitate this, the performance of each model was analyzed by using the sum of the differences between the metrics of the basic model without augmentation and the models with augmentation applied, which is displayed in Table 2. For ease of analysis of the augmentation models, one augmentation model was categorized as S1 to S8, and two augmentation models were categorized as D1 to D12. Model S1 is the flip model with a three-fold increase, Model S2 is the 90° rotate model with a three-fold increase, and Model D12 is the crop and shear model with the five-fold increase. The differences between the mAP values of the basic and augmentation models are denoted D-mAP, the differences in precision are denoted D-Precision, and the differences in recall are denoted D-Recall.

Table 2. Differences in performance between the original model and the model with augmentation (one or two augmentations). Note that D indicates the difference between the measurements from the model with augmentation and the original model without any augmentation.

No.	Augmentation	D-mAP	D-Precision	D-Recall	Sum
S1	Flip	0.005	0.022	−0.030	−0.003
S2	90° Rotate	0.012	−0.009	0.021	0.024
S3	Crop	0.006	0.008	−0.013	0.001
S4	Shear	0.005	0.010	0.002	0.017
D1	3× Flip and 90° Rotate	0.017	−0.005	0.028	0.040
D2		0.015	0.020	0.009	0.044
D3		0.009	−0.002	0.009	0.016
D4		0.017	−0.005	0.028	0.040
D5		0.021	−0.001	0.027	0.047 [†]
D6		0.000	−0.012	0.027	0.015
S5	Flip	0.016	−0.003	0.016	0.029
S6	90° Rotate	0.016	0.006	−0.011	0.011
S7	Crop	−0.003	−0.006	0.011	0.002
S8	Shear	−0.002	−0.015	0.024	0.007
D7	5× Flip and 90° Rotate	0.010	−0.015	0.007	0.002
D8		0.006	−0.018	0.017	0.005
D9		0.022	0.010	0.013	0.045
D10		0.011	−0.007	0.017	0.021
D11		0.011	−0.010	0.009	0.010
D12		0.003	0.017	−0.002	0.018

[†] The numbers in bold and underlined indicate the best-performing model.

The comparative analysis of the basic model and the augmentation models shows that Model D5, which applies 90° rotate and shear operations, has the highest performance among the models with three-fold increases, whereas Model D9, which applies flip and shear operations, excels compared with the models with five-fold increases. A common feature of Model D5 and Model D9 is the use of shear transformation. Shear transformation is used to simulate geometric distortion in images by skewing objects. This is thought to aid the model in better handling various visual transformations that can occur in real-world scenarios.

In particular, Model D5 has the highest performance among all augmentation models, with a score of 0.047. The augmentations applied to Model D5, i.e., 90° rotate and shear operations, can have complementary effects. For example, the 90° rotate transformation aids the model in recognizing objects rotated along the horizontal or vertical axis, whereas the shear transformation enhances the recognition of asymmetric tilt changes. This suggests that Model D5 is not overly dependent on training data and can generalize better with regard to new data, as evidenced by its relatively good ability to recognize new images affected by different tilt and rotation angles.

Conversely, Model S1 has the lowest performance, with a score of -0.003 . This is attributed to the application of fewer training data compared with Model S5 and the use of only the flip transformation, which is a relatively simple transformation that flips images horizontally or vertically. The use of a single transformation may limit the ability of the model to learn various visual transformations, resulting in lower performance.

4.2.2. Three and Four Augmentations

The number of training samples learned through the three and four augmentations were 10,029 with three times augmentation and 15,881 with five times augmentation for each model. A performance analysis was conducted on the basis of the augmentation combinations, with the values of mAP, precision, and recall being calculated. The models based on three augmentations are denoted T1 to T8, whereas the models with four augmentations are represented as Q1 and Q2 in Table 3. The models with three-fold increases are classified as T1 to T4 and Q1. The models with five-fold increases are represented as T5 to T8 and Q2. According to the mAP results in Table 3, Model T6, the model with a five-fold increase and a combination of flip, 90° rotate and shear operations, achieves the highest performance, with a mAP of 0.890, and the highest precision, with a score of 0.842. Furthermore, on the basis of the recall results, Model T1, the model with a three-fold increase and a combination of flip, 90° rotate and crop operations, achieves the highest performance. These findings support the earlier statement that the flip operation has a positive effect on precision and that the combination of 90° rotate and shear operations can increase mAP.

Table 3. mAP, precision, and recall calculated through a combination of three and all augmentation functions. Note that Roboflow augmentation was used to increase the amount of input data by three or five times.

No.	Augmentation		mAP	Precision	Recall
		Basic	0.872	0.816	0.820
T1	3×	Flip and 90° Rotate and Crop	0.882	0.788	0.857
T2		Flip and 90° Rotate and Shear	0.878	0.820	0.825
T3		Flip and Crop and Shear	0.888	0.830	0.848
T4		90° Rotate and Crop and Shear	0.885	0.813	0.833
T5	5×	Flip and 90° Rotate and Crop	0.886	0.818	0.825
T6		Flip and 90° Rotate and Shear	0.890	0.842	0.795
T7		Flip and Crop and Shear	0.885	0.788	0.856
T8		90° Rotate and Crop and Shear	0.884	0.839	0.788
Q1	3×	Flip and 90° Rotate and Crop	0.884	0.802	0.842
Q2	5×	and Shear	0.887	0.835	0.816

Despite the variety of augmentation combinations, reliance on a single metric for selecting the optimal model is insufficient. Therefore, the performance of each model was analyzed via the sum of the differences between the metrics of the basic model and the models with three or four augmentations applied. The analyzed results are presented in Table 4. The comparative analysis of the basic model and the augmentation models indicates that Model T3 has the highest performance among the models with three-fold increases, whereas Models T5 and T7 have the highest performance among the models with five-fold increases. The models with the three-fold and five-fold increases applying flip, crop and shear operations demonstrated the best performance.

Table 4. Differences in performance between the original model and the model with augmentations (three and four augmentations). Note that D indicates the difference in the measurements from model with augmentations and the original model without any augmentation.

No.	Augmentation	D-mAP	D-Precision	D-Recall	Sum
T1	Flip and 90° Rotate and Crop	0.010	−0.028	0.037	0.019
T2	Flip and 90° Rotate and Shear	0.006	0.004	0.005	0.015
T3	Flip and Crop and Shear	0.016	0.014	0.028	0.058[†]
T4	90° Rotate and Crop and Shear	0.013	−0.003	0.013	0.023
T5	Flip and 90° Rotate and Crop	0.014	0.002	0.005	0.021
T6	Flip and 90° Rotate and Shear	0.018	0.026	−0.025	0.019
T7	Flip and Crop and Shear	0.013	−0.028	0.036	0.021
T8	90° Rotate and Crop and Shear	0.012	0.023	−0.032	0.003
Q1	3× Flip and 90° Rotate and Crop	0.012	−0.014	0.022	0.020
Q2	5× and Shear	0.015	0.019	−0.004	0.030

[†] The numbers in bold and underlined indicate the best-performing model.

This suggests that the models are helpful in detecting rivers in images where the boundaries are tilted in various directions or partially obscured by surrounding environments. This finding reflects the ability of the models to adapt to a variety of scenarios and changes similar to real river environments. The interactions between augmentation functions can sometimes act in unexpected ways, potentially having a negative impact on performance. Moreover, excessive augmentation may interfere with the ability of the model to learn important features and lead to overfitting. Therefore, the selection and combination of augmentation functions should be performed with care, and an experimental approach with meticulous analysis is necessary to optimize model performance.

Overall, the models with augmentation performed better than the basic model without augmentation. This is due to the augmentation features allowing the model to experience a greater variety of transformations and scenarios, enabling it to learn about the diverse situations that may occur in real-world settings.

4.3. Results for Each River

Based on the simulation results from Sections 4.1 and 4.2, The confidence was extracted by applying them to the 11 study rivers as shown in Figure 11. The confidence score is calculated by multiplying the objectness score and class probability. The objectness score is the probability that an object exists within an estimated bounding box while class probability is the likelihood that the detected object belongs to a specific class. The confidence score represents the predictive confidence of the class predicted by the YOLO model, with higher values indicating a better model. Parts of the data were excluded at Andong river experiment center, Beophwacheon stream (D2~D5, D6~D11, and Q1~Q2) due to data limitation to calculate the confidence score.

The result indicated that Wosucheon stream had the highest confidence score of 0.842, while Sincheon stream had the lowest as 0.697. This shows that the models recognized Wosucheon stream the best, while Sincheon stream had the lowest recognition confidence. Wosucheon stream used the most images for training, accounting for about 12% of the 4177 drone photographs used in this study. Additionally, Sincheon stream showed a tendency for its boundaries between the levees and river to be unclear due to the influence of plant debris.

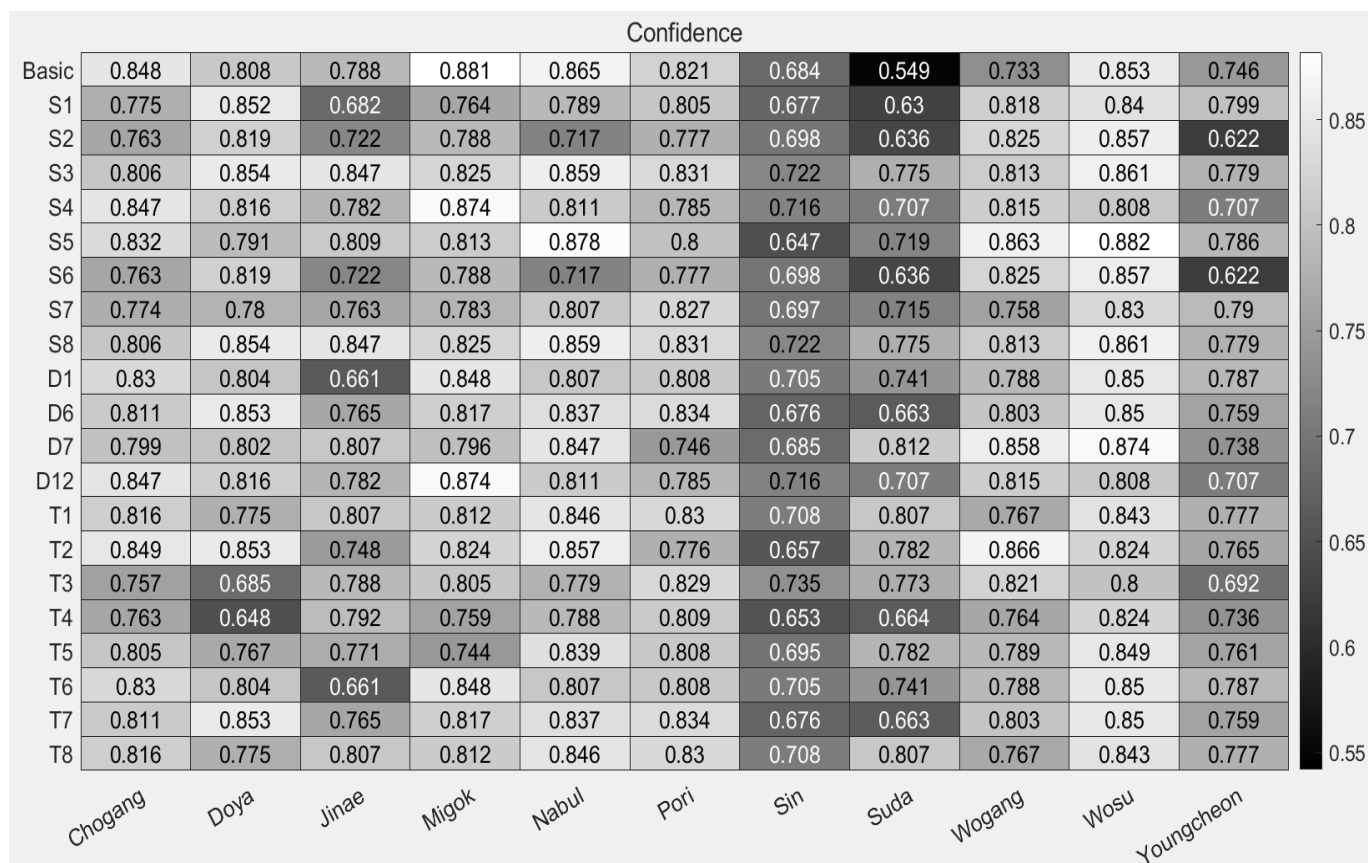


Figure 11. Confidence score for 11 rivers according to its augmentation models. The x-axis represents the driver and the y-axis represents the model. The types of augmentation of the model are shown in Tables 2 and 4.

Based on the results in Figure 11, the models with the highest confidence for each river were identified and are presented in Table 5. The maximum confidence ranged from 0.88 (at Migok with the basic model) to 0.74 (at Sin with T8). This suggests that data augmentation does not always enhance model performance, highlighting the need for careful testing of augmentation methods before conducting further analysis.

Table 5. Max confidence score among 20 augmentation models at each river.

River	No.	Augmentation	Confidence
Chogang	D6	Crop Shear 3×	0.849
Doya	T8	90° Rotate and Crop and Shear 5×	0.867
Jinae	S3	Crop 3×	0.847
Migok	Basic	-	0.881
Nabul	S5	Flip 5×	0.878
Pori	T4	90° Rotate and Crop and Shear 3×	0.841
Sin	T8	90° Rotate and Crop and Shear 5×	0.741
Suda	S7	Crop 5×	0.812
Wogang	D12	Crop and Shear 5×	0.866
Wosu	T3	Flip and Crop and Shear 3×	0.885
Youngcheon	T7	Flip and Crop and Shear 5×	0.865

Except for the Basic Model and Model D5, all other models incorporated crop augmentation. Crop augmentation enhances model performance by selectively removing arbitrary portions of the image, reducing data redundancy, and ensuring that essential features of the image are retained. As a result, this technique enables more accurate model

predictions. This capability of crop augmentation significantly contributes to achieving higher maximum confidence scores for most rivers.

According to the results in Table 5, models utilizing 90° rotate, crop, and shear achieved the highest confidence, closely followed by those incorporating flip, crop, and shear. Among models with one or two augmentation techniques, 90° rotate and shear demonstrated the best performance, while flip, crop, and shear outperformed others in models with three or four augmentation techniques. These findings suggest that model performance significantly influences the confidence levels when applied to river data. Furthermore, data augmentation techniques, particularly the crop method, effectively enhance the model's prediction confidence.

5. Conclusions

Given the recent impacts of climate change, which have led to floods and various river-related disasters, securing technology to address these issues has become crucial. The current state of artificial intelligence-based detection technology for rivers can be highly adaptable for river-related management, and aerial images from UAVs can be advantageous in training AI to detect river sections. Therefore, this study developed an AI model named RivDet for automatic river path detection, which was performed by applying augmentation through the YOLOv8 model on the Roboflow platform.

Two main issues were addressed when training RivDet AI: the existence of river levees and the augmentation features. A river levee is crucial for protecting against and mitigating floods, and it defines the river section. However, its clear boundary might not be helpful if no bank exists. The results indicate that training river images without separating levees can be more beneficial for better performance.

In addition, data augmentation was further investigated to determine whether this augmentation can be helpful in training the RivDet AI model for detection, and which feature is more reliable. The results showed that among the models with one or two augmentations applied, the combined augmentation method using 90° rotate and shear operations exhibited the highest performance, whereas the method using the flip augmentation alone showed the lowest performance. This suggests that training the model using 90° rotate and shear transformations can improve performance reasonably in river detection tasks by making the model more robust to various directional and geometric transformations. Moreover, among the models with three or four augmentations applied, the model combining the flip, crop and shear operations demonstrated the highest performance. This finding indicates that training under conditions similar to those of the actual river environment is advantageous for adapting to various scenarios and changes in field conditions and that excessive augmentation can hinder the model from learning important features.

Based on the results, confidence was calculated by applying it to the rivers. When all models, except those with limitations, were applied, it was found that the Wosucheon stream had a higher average confidence. This suggests that training the models with a large number of images from the specific river can yield higher confidence. Additionally, the max confidence, representing the highest reliability for each river, was extracted based on the confidence results. Therefore, Chogangcheon stream is best represented by Model D6, Doyacheon stream by Model T8, Jinaecheon stream by Model S3, Migokcheon stream by Basic, Nabalcheon stream by Model S5, Poricheon stream by Model T4, Sincheon stream by Model T8, Sudacheon stream by Model S7, Wogangcheon stream by Model D12, Wosucheon stream by Mode T3, and Youngcheongang stream by Model T7. Except for two rivers, the recognition accuracy of the augmentation models using the crop function was found to be

higher in all other rivers. This suggests that including crop in the simulations can result in higher confidence in detecting rivers.

However, the effects of data augmentation can vary depending on factors such as the characteristics of the dataset, the structure of the model, and the training process, which indicates that the results may not always reflect general cases. These outcomes are limited to specific datasets and experimental settings, and different results may emerge under other conditions. Furthermore, these augmentation methods do not guarantee optimal results in all cases, and it is important to find the best augmentation combination through experimentation with various datasets and models. Therefore, it is believed that appropriately applying augmentation functions according to the river field environment can yield the highest performance.

The training models with augmentations applied are expected to help the model generalize better with regard to new data and improve performance in real-world environments. Additionally, the RivDet artificial intelligence model for the automatic river path setting developed in this study is expected to solve various problems, such as finding the best river path, automatically calculating the flow rate, establishing an early flood warning setup, and estimating the flood inundation range for river disaster prevention.

Author Contributions: S.L. carried out the research as well as writing the original draft. T.L. conceptualized the research and supervised the overall research while Y.K. edited this manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) (2023R1A2C1003850).

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author/s.

Conflicts of Interest: The author declares that they have no conflict of interest.

References

1. Salman, A.M.; Li, Y. Flood risk assessment, future trend modeling, and risk communication: A review of ongoing research. *Nat. Hazards Rev.* **2018**, *19*, 04018011. [[CrossRef](#)]
2. Zhang, J.; Xu, W.; Liao, X.; Zong, S.; Liu, B. Global mortality risk assessment from river flooding under climate change. *Environ. Res. Lett.* **2021**, *16*, 064036. [[CrossRef](#)]
3. Liu, W.; Feng, Q.; Engel, B.A.; Yu, T.; Zhang, X.; Qian, Y. A probabilistic assessment of urban flood risk and impacts of future climate change. *J. Hydrol.* **2023**, *618*, 129267. [[CrossRef](#)]
4. Jain, A.; Ramaprasad, R.; Narang, P.; Mandal, M.; Chamola, V.; Yu, F.R.; Guizan, M. AI-Enabled Object Detection in UAVs: Challenges, Design Choices, and Research Directions. *IEEE Netw.* **2021**, *35*, 129–135. [[CrossRef](#)]
5. Varatharasan, V.; Rao, A.S.S.; Toutounji, E.; Hong, J.-H.; Shin, H.-S. Target detection, tracking and avoidance system for low-cost UAVs using AI-based approaches. In Proceedings of the 2019 Workshop on Research, Education and Development of Unmanned Aerial Systems (RED UAS), Cranfield, UK, 25–27 November 2019; pp. 142–147.
6. Eum, T.S.; Shin, E.T.; Song, C.G. Analysis of present status and characteristics of elementary technologies for smart river management. *J. Korean Soc. Disaster Secur.* **2022**, *15*, 13–21.
7. Bernard, T.G.; Davy, P.; Lague, D. Hydro-Geomorphic Metrics for High Resolution Fluvial Landscape Analysis. *J. Geophys. Res. Earth Surf.* **2022**, *127*, e2021JF006535. [[CrossRef](#)]
8. Stanislawski, L.V.; Shavers, E.J.; Wang, S.; Jiang, Z.; Usery, E.L.; Moak, E.; Duffy, A.; Schott, J. Extensibility of U-Net Neural Network Model for Hydrographic Feature Extraction and Implications for Hydrologic Modeling. *Remote Sens.* **2021**, *13*, 2368. [[CrossRef](#)]
9. Costabile, P.; Costanzo, C.; Lombardo, M.; Shavers, E.; Stanislawski, L.V. Unravelling spatial heterogeneity of inundation pattern domains for 2D analysis of fluvial landscapes and drainage networks. *J. Hydrol.* **2024**, *632*, 130728. [[CrossRef](#)]
10. Jaroenchai, N.; Wang, S.; Stanislawski, L.V.; Shavers, E.; Jiang, Z.; Sagan, V.; Usery, E.L. Transfer learning with convolutional neural networks for hydrological streamline delineation. *Environ. Model. Softw.* **2024**, *181*, 106165. [[CrossRef](#)]

11. Patil, S.; Sawant, S.; Joshi, A. Flood detection using remote sensing and deep learning approaches. In Proceedings of the 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 6–8 July 2023; pp. 1–6.
12. Vandaele, R.; Dance, S.L.; Ojha, V. Deep learning for automated river-level monitoring through river-camera images: An approach based on water segmentation and transfer learning. *Hydrol. Earth Syst. Sci.* **2021**, *25*, 4435–4453. [[CrossRef](#)]
13. Niu, G.; Li, J.; Guo, S.; Pun, M.-O.; Hou, L.; Yang, L. SuperDock A Deep Learning-Based Automated Floating Trash Monitoring System. In Proceedings of the 2019 IEEE International Conference on Robotics and Biomimetics (ROBIO), Dali, China, 6–8 December 2019; pp. 1035–1040.
14. Rizk, H.; Shokry, A.; Youssef, M. Effectiveness of data augmentation in cellular-based localization using deep learning. In Proceedings of the 2019 IEEE Wireless Communications and Networking Conference (WCNC), Marrakesh, Morocco, 15–18 April 2019; pp. 1–6.
15. Liao, Y.-H.; Juang, J.-G. Real-Time UAV Trash Monitoring System. *Appl. Sci.* **2022**, *12*, 1838. [[CrossRef](#)]
16. Lee, K.; Wang, B.; Lee, S. Analysis of YOLOv5 and DeepLabv3+ Algorithms for Detecting Illegal Cultivation on Public Land: A Case Study of a Riverside in Korea. *Int. J. Environ. Res. Public Health* **2023**, *20*, 1770. [[CrossRef](#)] [[PubMed](#)]
17. Wang, G.; Chen, Y.; An, P.; Hong, H.; Hu, J.; Huang, T. UAV-YOLOv8: A Small-Object-Detection Model Based on Improved YOLOv8 for UAV Aerial Photography Scenarios. *Sensors* **2023**, *23*, 7190. [[CrossRef](#)] [[PubMed](#)]
18. Sultana, F.; Sufian, A.; Dutta, P. A review of object detection models based on convolutional neural network. In *Intelligent Computing: Image Processing Based Applications*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 1–16.
19. Kim, T.; Jung, J.; Ha, T.; Kong, Y.; Lee, a. Determining the Optimum Altitude Parameter for River Surveying Using UAVs: Case Study on Jinae-cheon Stream. *J. Korean Soc. Hazard Mitig.* **2022**, *22*, 187–193. [[CrossRef](#)]
20. Kong, Y.; Kim, T.; Lee, T. UAV-Based Floodwater-Level Establishment for FEWS for Abrupt River Section Change in Imsan. *J. Korean Soc. Hazard Mitig.* **2022**, *22*, 377–384. [[CrossRef](#)]
21. Kim, T.; Park, J.; Hwang, S.; Lee, T. Whole Watershed-based Estimation of FEWS Installation Site Using UAV Photogrammetry. *J. Korean Soc. Hazard Mitig.* **2023**, *23*, 233–241. [[CrossRef](#)]
22. Alexandrova, S.; Tatlock, Z.; Cakmak, M. RoboFlow: A flow-based visual programming language for mobile manipulation tasks. In Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, USA, 26–30 May 2015; pp. 5537–5544.
23. Ciaglia, F.; Zuppichini, F.S.; Guerrie, P.; McQuade, M.; Solawetz, J. Roboflow 100: A rich, multi-domain object detection benchmark. *arXiv* **2022**, arXiv:2211.13523.
24. Kim, H.-J.; Shin, B.-K.; Kim, W. A study on hydromorphology and vegetation features depending on typology of natural streams in Korea. *Korean J. Environ. Ecol.* **2014**, *28*, 215–234. [[CrossRef](#)]
25. Pei, J.; Ananthasubramaniam, A.; Wang, X.; Zhou, N.; Sargent, J.; Dedeloudis, A.; Jurgens, D. Potato: The portable text annotation tool. *arXiv* **2022**, arXiv:2212.08620.
26. Terven, J.; Cordova-Esparza, D. A comprehensive review of YOLO: From YOLOv1 to YOLOv8 and beyond. *arXiv* **2023**, arXiv:2304.00501.
27. Talebi, H.; Milanfar, P. Learning to resize images for computer vision tasks. In Proceedings of the IEEE/CVF International Conference On Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 497–506.
28. Perez, L.; Wang, J. The effectiveness of data augmentation in image classification using deep learning. *arXiv* **2017**, arXiv:1712.04621.
29. Mikołajczyk, A.; Grochowski, M. Data augmentation for improving deep learning in image classification problem. In Proceedings of the 2018 International Interdisciplinary PhD Workshop (IIPhDW), Swinoujście, Poland, 9–12 May 2018; pp. 117–122.
30. Gu, S.; Pednekar, M.; Slater, R. Improve image classification using data augmentation and neural networks. *SMU Data Sci. Rev.* **2019**, *2*, 1.
31. Asperti, A.; Mastronardo, C. The effectiveness of data augmentation for detection of gastrointestinal diseases from endoscopic images. *arXiv* **2017**, arXiv:1712.03689.
32. Khalifa, N.E.; Loey, M.; Mirjalili, S. A comprehensive survey of recent trends in deep learning for digital images augmentation. *Artif. Intell. Rev.* **2022**, *55*, 2351–2377. [[CrossRef](#)] [[PubMed](#)]
33. Takahashi, R.; Matsubara, T.; Uehara, K. Data augmentation using random image cropping and patching for deep CNNs. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *30*, 2917–2931. [[CrossRef](#)]
34. Goceri, E. Medical image data augmentation: Techniques, comparisons and interpretations. *Artif. Intell. Rev.* **2023**, *56*, 12561–12605. [[CrossRef](#)]
35. Yu, H.; Kawaike, K.; Yamanoi, K.; Koshiha, T. 3d Simulation of the Effect of Spur Dikes Spacing on Bed Deformation, Flow and Performance Evaluation in Meandering Channels. *J. JSCE* **2024**, *12*, 23–16178. [[CrossRef](#)]
36. Nagaraju, M.; Chawla, P.; Kumar, N. Performance improvement of Deep Learning Models using image augmentation techniques. *Multimed. Tools Appl.* **2022**, *81*, 9177–9200. [[CrossRef](#)]

37. Burdziakowski, P.; Bobkowska, K. UAV photogrammetry under poor lighting conditions—Accuracy considerations. *Sensors* **2021**, *21*, 3531. [[CrossRef](#)]
38. Mumuni, A.; Mumuni, F. Data augmentation: A comprehensive survey of modern approaches. *Array* **2022**, *16*, 100258. [[CrossRef](#)]
39. Słomska-Przech, K.; Panecki, T.; Pokojski, W. Heat maps: Perfect maps for quick reading? comparing usability of heat maps with different levels of generalization. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 562. [[CrossRef](#)]
40. Sekeroglu, B.; Ever, Y.K.; Dimililer, K.; Al-Turjman, F. Comparative Evaluation and Comprehensive Analysis of Machine Learning Models for Regression Problems. *Data Intell.* **2022**, *4*, 620–652. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.