Kalpa Publications in Computing

Volume 22, 2025, Pages 247–258

Proceedings of The Sixth International Conference on Civil and Building Engineering Informatics

A Study on the Performance Comparison of Sediment Movement Detectors Using Unet-based Models and the Sliding Partitioning Method

Xiaosong Liu[1], Makoto Oda[2], Kei Kawamura[3]and Tsuyoshi Wakatsuki[4]

[1] Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Yamaguchi, Japan.

[2] Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Yamaguchi, Japan.

[3]D. Eng., Prof., Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Yamaguchi, Japan. Email: kay@yamaguchi-u.ac.jp

[4]Principal Investigator, the Storm, Flood and Landslide Research Division, National Research Institute for Earth Science and Disaster Resilience (NIED), Japan.

e115vgv@yamaguchi-u.ac.jp, e092vgv@yamaguchi-u.ac.jp, kay@yamaguchi-u.ac.jp,waka@bosai.go.jp

Abstract

Japan is covered by mountainous and hilly terrain and thus is prone to disasters such as landslides caused by earthquakes and heavy rainfall. Currently, the identification of sediment disaster areas is conducted through visual interpretation of aerial photographs taken after these disasters. To improve the efficiency of this process, research has been advancing in the automatic detection of sediment movement areas using deep learning techniques applied to aerial photographs. In this study, we compare the performance of sediment movement detectors using Unet and Unet++, in order to investigate the impact of changes in the loss function parameters during training on detection performance. Additionally, we evaluate the performance improvement when the sliding partitioning method is used for input images to the detectors. Results show that using sliding partitioning and adjusting the value of the loss function coefficient β can improve detection performance.

1 Introduction

Japan frequently experiences earthquakes, typhoons, and heavy rains, which increase the likelihood of sediment-related disasters such as debris flows and landslides. To analyze and develop countermeasures for such sediment disasters, sediment movement distribution maps are created by

visually interpreting sediment movement areas based on aerial photographs taken after disasters, allowing for a broad understanding of slope movement(Yang Erhong et al., 2007). However, creating these sediment movement distribution maps is time-consuming and labor-intensive. Due to these challenges, previous researches of this study into the automatic detection of sediment movement areas is currently being conducted, with a particular focus on applying various deep learning framework which have rapidly advanced in recent years(Liu et al., 2015; Tadataka ISHII et al., 2020). We have also been conducting research on the automatic detection of sediment movement areas from aerial photographs, including the construction of detectors using the learning models Unet and Unet++, and investigating the impact of adjusting loss function parameters during training on detection performance. Furthermore, they have proposed a method that uses slide-split images as input during inference. However, since these studies were conducted separately, this study aims to analyze the impact on detection performance when these methods are combined to better clarify the characteristics of the detectors.

2 Proposed Method in Comparison to Existing Research

2.1 Sediment Movement Detection Process

In existing detection processes, the system receives an image (aerial photograph) of the area where sediment movement is detected and sediment movement is calculated by the sediment movement detector. Our study, however, leverages Unet and its improved model Unet++ as detectors while adjusting the loss function parameters during training to investigate the effect on relative detection performance(Kei KAWAMURA et al., 2023). We then verified the impact on detection performance when using slide-split images for the input image to the detector by applying either "AND" or "OR" processing to the output images(Kei KAWAMURA et al., 2018).

2.2 Detector Model

(1) Features of Unet

Unet is a network developed for pixel-level region detection in images that employs a feature detector algorithm capable of learning the identifying characteristics present of the images. Using this feature detector, it can detect the image characteristics from novel images. The framworks of Unet is as shown in Figure 1(Ronneberger et al., 2015).

(2) Features of Unet++

Unet++ incorporates multiple decoders, forming four branch networks which are used to detect sediment movement. During training, errors from the four branch networks are calculated and averaged along with the errors from the encoder. This averaged value is then used as the error to train the neural networks that constitute the encoders. The framworks of Unet++ is as shown in Figure 2(Zhou et al., 2018).

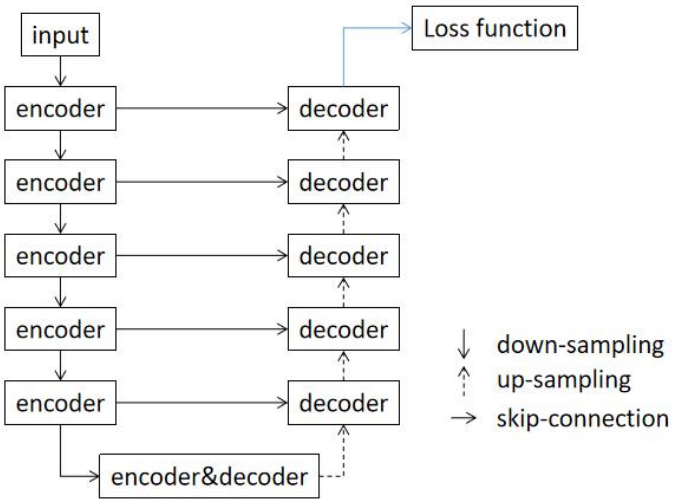


Figure 1: structure of Unet

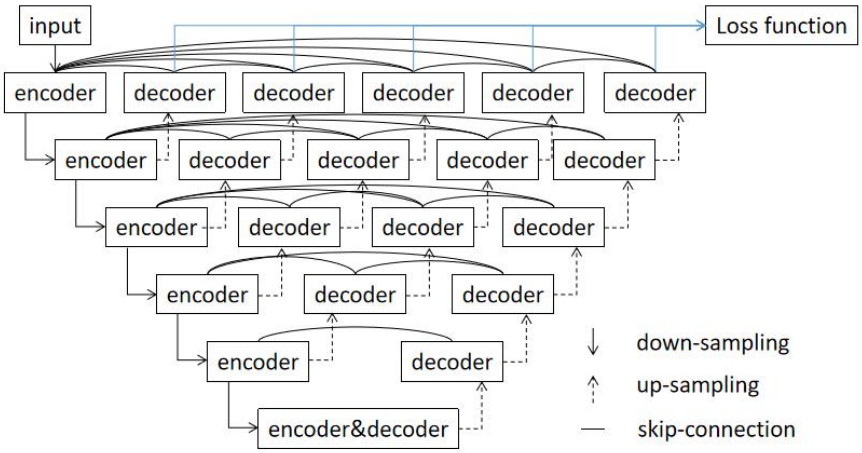


Figure 2: structure of Unet++

2.3 Processing of Input Images for the Detector

The size of the input images for the Unet and Unet++ detector models in this study is 224×224 pixels. Therefore, the training images or target region images for sediment detection (test images, etc.) must be divided into 224×224 pixel sections from the top-left to the bottom-right. In this study, any leftover regions smaller than 224×224 pixels that may occur on the right or bottom edges after dividing the aerial photographs were not considered as target areas for detection (see Figure 3 below). This division can be achieved using one of the following methods:

(1) Simple Division Method

The simple division method divides a large target image into predefined image sizes from the

top-left to the bottom-right without overlapping regions for sediment detection using aerial photographs.

(2) Slide-Split(overlaping) Method

The slide-split method divides the target image into predefined image sizes from the top-left to the bottom-right while sliding it horizontally and vertically by a certain width during the division process. However, the slide width is set smaller than the width and height of the input image in order to better analyze the division boundaries. In this experiment, the slide width was set to 112 pixels. The advantage of the slide-split method is that adjacent split images have overlapping regions, resulting in more accurate predictions.

2.4 Combining Output Images from the Detector

When combining slide-split images, two methods were compared for handling the overlapping regions: OR/AND processing. In OR processing, when any pixel in the overlapping region is judged as sediment in the split images, the result is output as sediment for that pixel. With AND processing, it is only output as sediment if the pixel is judged as sediment in all overlapping split images.

2.5 Training Process

For the training of the detector, we used Equation (1) as the loss function (equivalent to the Dice coefficient). Additionally, in this study, we analyzed the effect of changing the loss function coefficient β on detection performance.

$$F_\beta = (1+\beta^2)\frac{recall*precision}{\beta^2*recall+precision} \tag{1}$$

Here,

$$recall = \frac{TP}{TP+FN} \tag{2}$$

$$precision = \frac{TP}{TP+FP} \tag{3}$$

The symbols in the equation refer to recall (true positive rate), precision, true positives (TP), false negatives (FN), and false positives (FP)(Hand, D. & Christen, P., 2018).

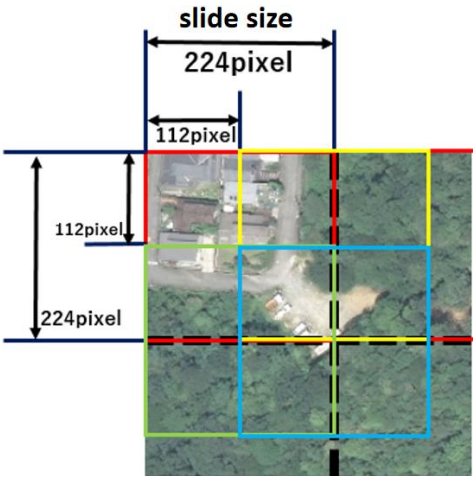


Figure 3: The simple segmentation method and the sliding segmentation method

3 Automatic Sediment Detection Methods and Performance Comparison Experiments

3.1 Experiment Flow

In this study, aerial photographs are input into the classifier using two image division methods: a simple division and the slide division method, which will be explained later. For the classifier, we used Unet and its improved model, Unet++, aiming to improve classifier performance. Additionally, by adjusting the loss function, we examined how to balance recall and precision better than in previous studies. We further performed performance comparisons using three result images: one using the simple division method combined output, and two others using the slide division method combined with "AND" and "OR" logic.

Three experiments in total were conducted using aerial photographs taken in Hiroshima after the 2018 sediment disaster and targets interpreted by workers. The first experiment focused on changing the value of the loss function's β parameter and assessing the impact of this change on the F-score using Unet and Unet++ (referred to as Experiment 1). The second experiment compared the evaluation values of the conventional simple division method and the newly introduced slide division method for the test images (referred to as Experiment 2). The third experiment examined the effect on recall, precision, and F-score when using the β value obtained from Experiment 1 in conjunction with the division methods from Experiment 2 (referred to as Experiment 3).

3.2 Experiment Method

This experiment used orthophotos of Hiroshima Sakacho Area (taken on July 9 and 11) after the 2018 heavy rainfall in Western Japan as its primary source data. To create the training data, we selected twelve pairs of regions (Training 1–12) from the aerial photos taken after the sediment disaster where workers had interpreted the sediment movement areas, which were subsequently divided into 244×244 pixel (px) sections. One aerial photograph covers an area of 12,560×6,087px, approximately 7.4×3.6 km. The actual size of each divided image is about 82m×82m. The targets are of the same size as the aerial photographs. In total, 9,264 pairs of training data were created, of which 80% (7,408 pairs) were used for training while the remaining 20% (1,856 pairs) were used for validation.

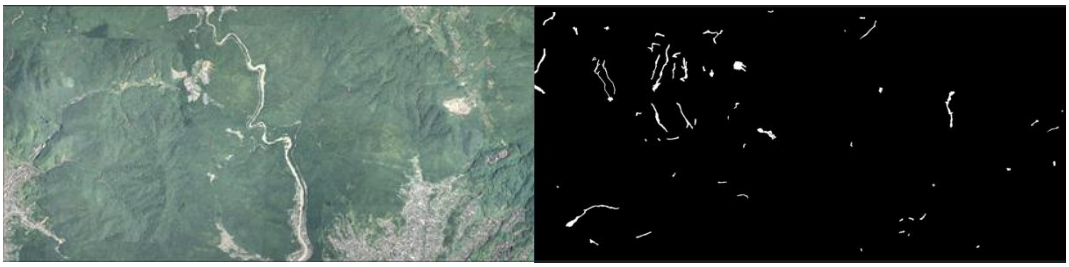
Following this, the created data was input into the network for training, in which one training cycle on the entire data set was considered a single epoch. At the end of each epoch, the weights of the trained network were saved, and prediction accuracy was calculated using both the training and validation datasets based on the error function. The epoch with the highest prediction accuracy for the validation data set was then identified, and its weights were used for the classifier. In this study, the Dice coefficient was used as a measure of prediction accuracy. The Dice-score, which measures the similarity between two sets, is determined using the following formula: if the two sets are identical, the value is 1, and if they are completely different, the value is 0.

$$(\text{Dice} - \text{score}) = \frac{2\times|X\cap Y|}{|X|+|Y|} \tag{4}$$

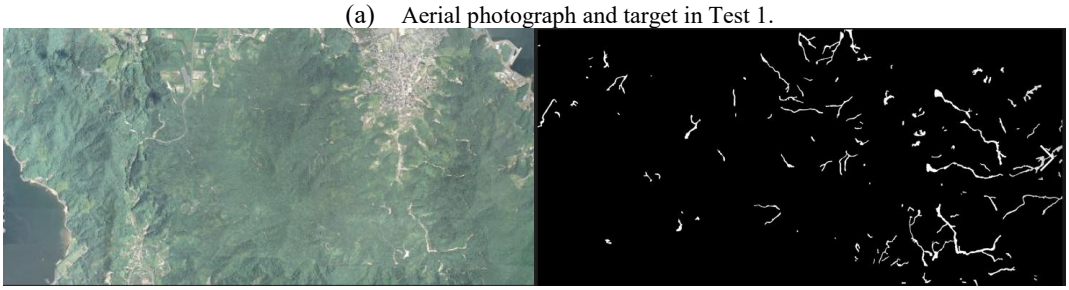
Here, X is the predicted data, and Y is the target data. To evaluate the generalization ability of the sediment classifier, we used four test regions (Test 1 to Test 4) as shown in Figure 4. The size of each image is the same as the training data, 12,560×6,087 px, covering an area of approximately 3.6 km×7.4 km. The top of the image corresponds to the north, the bottom to the south, the right side to the east, and the left side to the west. These images are from regions other than the twelve pairs used

in the training data. Furthermore, the test data was not used during the training phase. The four test datasets were created by dividing these four test images using the method described earlier. The characteristics of each test region are as follows: Test 1 contains a large river in the center of the image; Test 2 features an urban area in the upper right corner; Test 3 has urban areas in the lower right and upper left corners; Test 4 contains many sediment movement areas in a mountainous region.

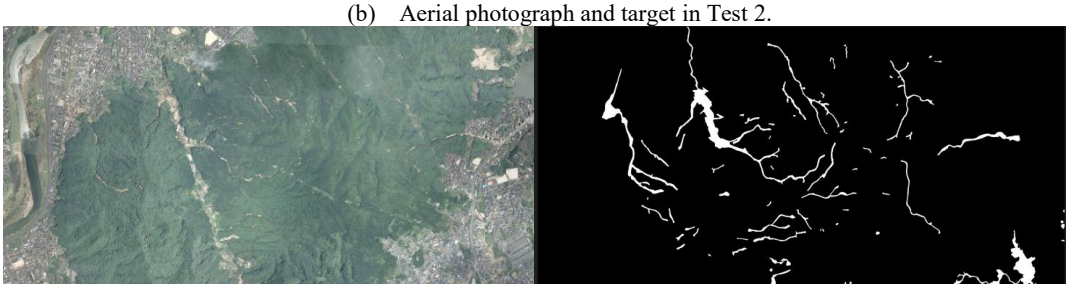
The characteristics of the four test regions are as follows: Test 1 has a large river in the center of the image, Test 2 features an urban area in the upper right corner, Test 3 has urban areas in the lower right and upper left corners, and Test 4 is a mountainous area with many sediment movement sections.



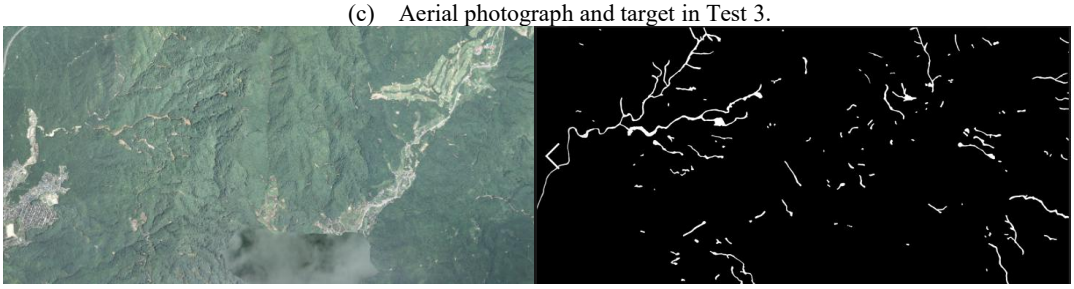
(a) Aerial photograph and target in Test 1.



(b) Aerial photograph and target in Test 2.

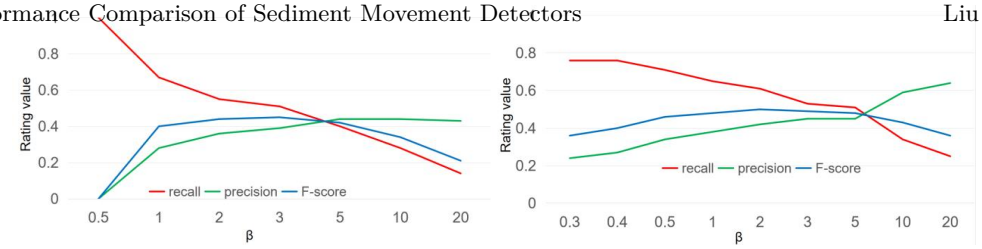


(c) Aerial photograph and target in Test 3.



(d) Aerial photograph and target in Test 4.

Figure 4: Aerial photograph and target in the test data.



3.3 Evaluation Metrics

To evaluate the detection results against the targets, each pixel is classified into one of four categories: True Positive (TP), False Positive (FP), False Negative (FN), or True Negative (TN), as shown in the table. Then, recall, precision, and the weighted F-score are calculated using the formulas below and are used for evaluation.

Network Used	Unet	Unet++
Image Size	224×224(pixel)	
Loss Function	Fβ(β=0.3,0.4,0.5,1,2,3,10,20)	
Number of Training Data	7408 pairs	
Number of validation Data	1856pairs	
Number of Training Iterations	100 times	300times
Batch Size	16	
Optimization Method	Adam(learning rate=1.0×10-4)	

Table 1: experimental conditions.

$$\text{recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{5}$$

$$\text{precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{6}$$

$$\text{F}_\beta - \text{score} = (1+\beta^2)\frac{\text{recall}*\text{precision}}{\beta^2*\text{recall}+\text{precision}} \tag{7}$$

Here, "recall" refers to the percentage of sediment areas in the target that were correctly identified as sediment by deep learning. Precision refers to the percentage of all areas identified as sediment by deep learning that were actual sediment areas. The F-score is a performance metric calculated based on recall and precision. The best value for each metric is 1, and higher values indicate better accuracy. Table 1 shows the experimental conditions.

4 RESULTS

4.1 Experiment 1

Figure 5 shows the changes in the loss function values for Unet and Unet++, respectively, as the β values were varied, as listed in the tables. Figures 6 and 7 show the difference images for β = 0.5, 1, 3, and 20. From Figure 5, it can be observed that recall decreases with an increase in β, while precision increases with an increase in β. Thus, it is possible to adjust recall and precision by modifying the β values. Since recall and precision have a trade-off relationship, the F-value does not change significantly with varying β. However, there is a slight change in the maximum F-value observed at β=2 for this test image. Figures 6 and 7 show that both Unet and Unet++ generated false positives for areas with similar hues to the sediment, such as sandy areas.

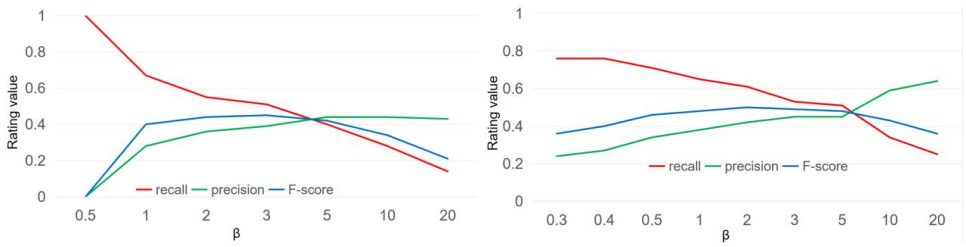


Figure 5: Comparison of evaluations when modifying the Dice coefficient β for Unet (left) and Unet++ (right).

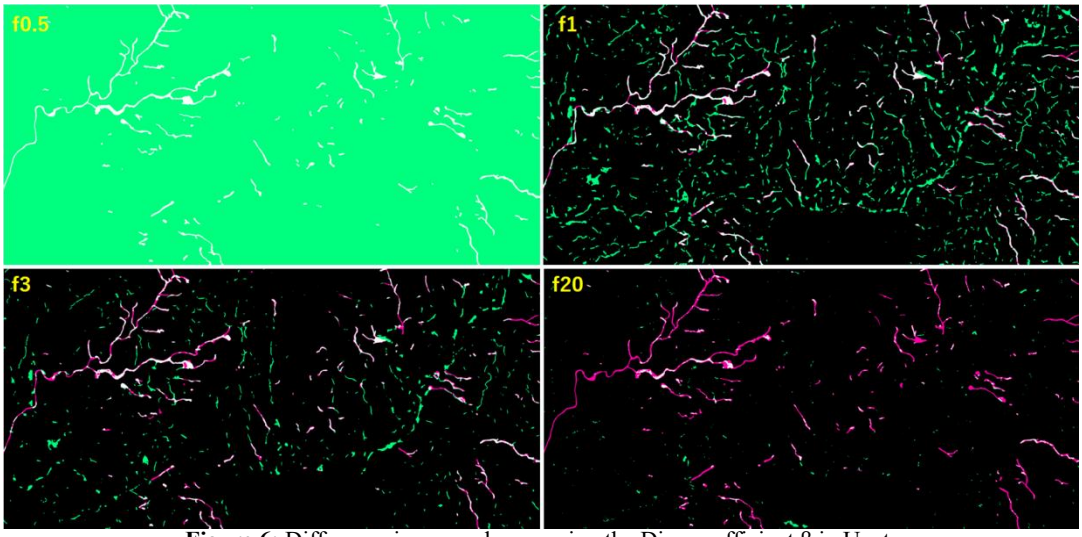


Figure 6: Difference images when varying the Dice coefficient β in Unet.

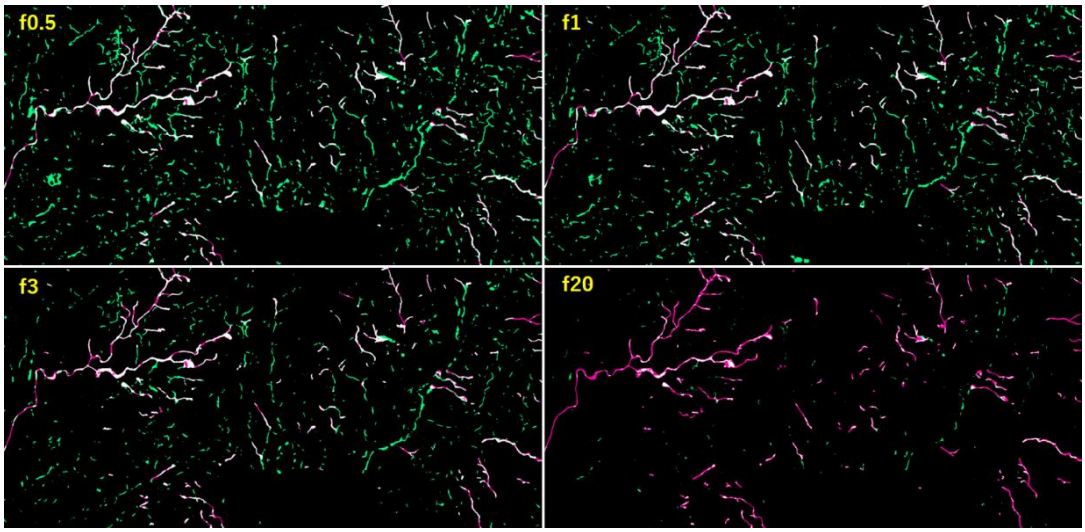


Figure 7: Difference images when varying the Dice coefficient β in Unet++.

4.2 Experiment 2

Table 2 presents the difference in F-scores between the results obtained by applying the sliding segmentation method and the simple segmentation method. Unet++ was trained using the loss function coefficient patterns β from Table 1, and the sliding segmentation method was applied to Test 4. The F-scores (values in parentheses in Table 2) were calculated after applying either "and" or "or" processing to the output from the detector. The difference values were then derived by subtracting the F-scores of the simple segmentation method from those of the sliding segmentation method.
For β = 1, 2, and 3, the F-score difference values (AND) are positive, indicating that the sliding segmentation method outperforms the simple segmentation method. Similarly, for β = 10, the F-score difference values (OR) are positive, further demonstrating the reliability of the sliding segmentation method over the simple segmentation method. These results suggest that the sliding segmentation method can enhance the performance of the detector as shown in the image comparison in Figure 8.

β	1	2	3	5	10
AND combination's F-score difference value (F-score)	0.05 (0.53)	0.04 (0.55)	0.03 (0.54)	-0.02 (0.52)	-0.04 (0.43)
OR combination's F-score difference value (F-score)	-0.05 (0.43)	-0.05 (0.46)	-0.03 (0.49)	-0.02 (0.52)	0.04 (0.51)

Table 2: The difference in F-score between the OR/AND combination and the simple combinations.

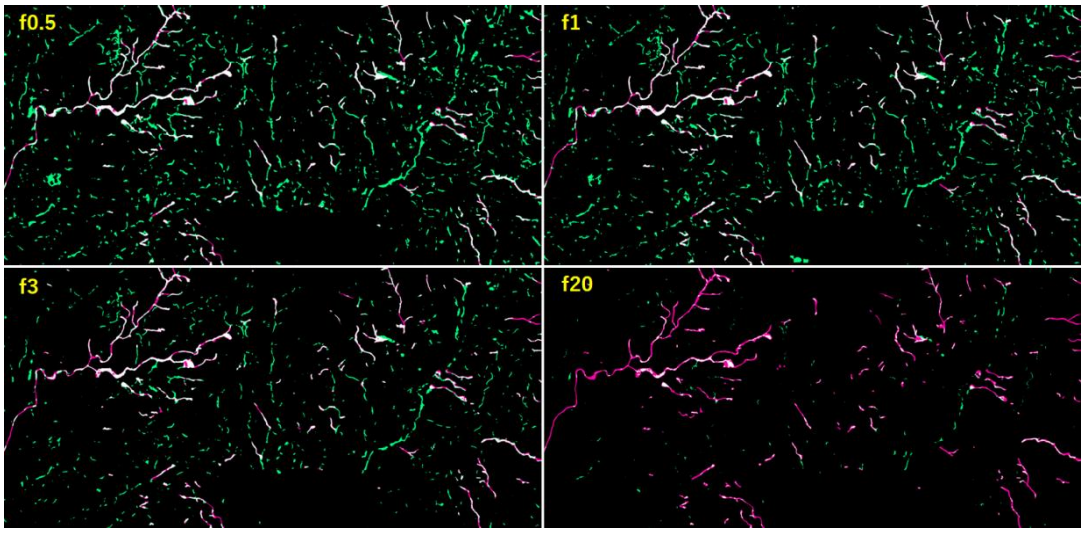


Figure 8: Difference images using different combinations.

4.3 Experiment 3

Figure 9 shows the transitions in Unet's performance using AND and OR combinations for different β values. Figures 10 and 11 show the difference images for β = 0.5, 1, 2, and 3. From these, we can see that recall and precision follow the same trends for simple division, AND combination, and OR combination. As in Experiment 1, recall and precision have a trade-off relationship, so the F-value does not change significantly with varying β. In Experiment 2, as with the AND combination,

recall for the slide combination was lower than that for simple division. However, precision improved because only the detected areas were output. The recall for the OR combination improved since it outputs areas detected in any of the images, outperforming simple division, although precision decreased compared to simple division. When increasing β with the AND combination, the precision further improved compared to the AND combination alone. These results suggest that classifier performance is influenced by changes in β, even when using the slide combination.

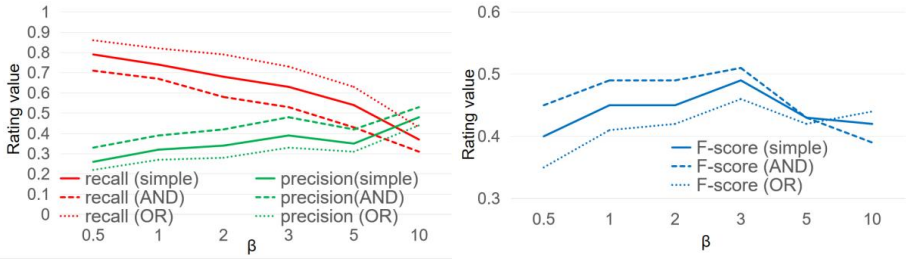


Figure 9: Comparison of evaluations when modifying the Dice coefficient β for Unet (left) and Unet++ (right).

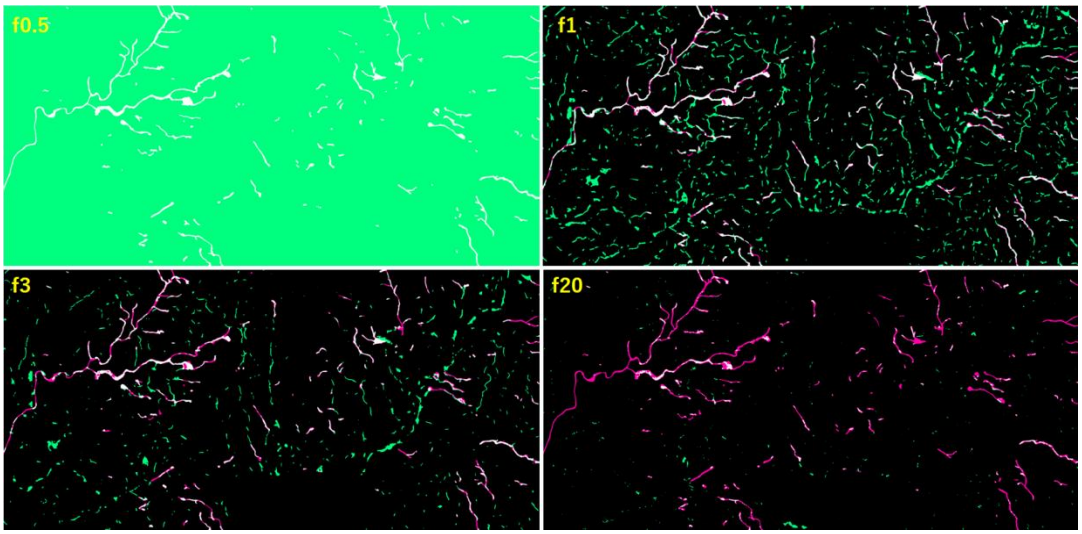


Figure 10: Difference images using AND fusion with varying Dice coefficient β.

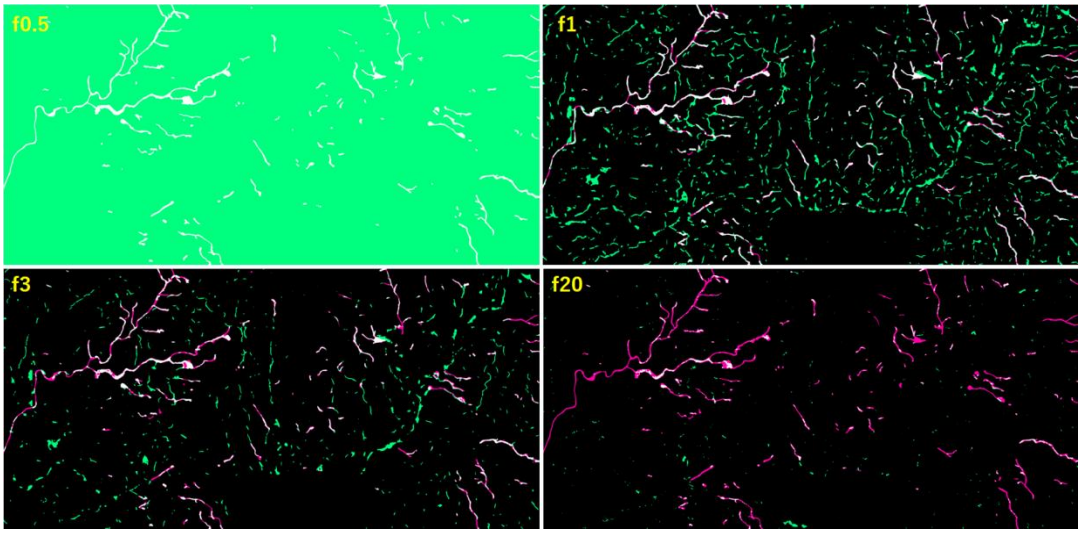


Figure 10: Difference images using AND fusion with varying Dice coefficient β.

5 Conclusion

This study aimed to suppress false-positive detections that occurred during the analysis of sediment areas using the Unet classifier via the introduction of the Unet++ classifier, adjusting the loss function coefficient β, and adopting the sliding partitioning method. Experimental results suggested that the detection performance of sediment areas improved with the Unet++ classifier, as the F-score increased compared to the Unet classifier. By changing the loss function coefficient β, it is possible to adjust the balance between recall and precision without significantly affecting the F-score. Moreover, the AND combination method in the sliding partitioning technique was able to enhance precision while suppressing the decline in recall. As a result, the F-score improved, leading to an overall enhancement in classifier performance.

When developing deep learning software, it is necessary to design a classifier tailored to the preferences of the user performing the interpretation tasks. In this study, the Unet++ classifier, using sliding partitioning, showed an improvement in precision and an increase in the F-score compared to the Unet classifier. Additionally, by adjusting the value of the loss function coefficient β, it is possible to emphasize either recall or precision without changing the F-score.

It remains essential to further refine the combination methods in the sliding partitioning technique to enhance precision without reducing recall. Although we compared detection performance using the Unet++ model, there are other deep learning methods such as pix2pix that are utilized for image segmentation. In the future, it will be necessary to conduct performance comparisons with models other than the Unet system and explore ways to further improve detection performance.

References

Erhong, Y., Zhihua, W., & Shaohui, J. (2007). The application of remote sensing digital technique in the investigate of the tiantai landslide in Xuanhan County. Journal of Jilin University (Earth Science Edition), 37(3), 557-563.

Liu, Y., & Wu, L. (2016). Geological disaster recognition on optical remote sensing images using deep learning. Procedia Computer Science, 91, 566-575.

Tadataka ISHII, Michiyuki HIROKANE, Kazumasa KURAMOTO, & Naoki NSHIHARA. (2020). CONSIDERING SOCIAL FACTOR IN SETTING OF SEDIMENT DISASTER WARNING AREA USING DEEP LEARNING. Japanese Journal of JSCE F6 , 76(2), I_193-I_199.

Kei KAWAMURA, Yushi NAKAMURA, Tsuyoshi WAKATSUKI, & Toshikazu SAMURA. (2018). A STUDY ON THE AUTOMATIC DETECTION OF LANDSLIDE FROM AERIAL PHOTOGRAPHS USING DEEP LEARNING. Japanese Journal of JSCE F3, 74(2), I_132-I_143.

Kei KAWAMURA, Shunsuke SUGAHARA, Syoki RYU, & Tsuyoshi WAKATSUKI. (2023). A STUDY ON THE AUTOMATIC DETECTION OF LANDSLIDE FROM AERIALPHOTOGRAPHS USING UNet AND UNet+. Japanese Journal of JSCE, 79(22), 22-22026.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18 (pp. 234-241).

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4 (pp. 3-11).

Hand, D., & Christen, P. (2018). A note on using the F-measure for evaluating record linkage algorithms. Statistics and Computing, 28, 539-547.