ACRS 2022
MONGOLIA

DETECTION OF RESIDENTIAL AND WATER AREAS IN HISTORICAL TOPOGRAPHIC MAPS OF JAPAN USING U-NET

Ami Ogita[1], Takayuki Shinohara[2], Haoyi Xiu[1], Haruki Oshio[1], Masashi Matsuoka[1]
[1] *Tokyo Institute of Technology*, *4259 Nagatsuta, Midori, Yokohama, Kanagawa 226-8502*, Japan,
Email: *ogita.a.aa@m.titech.ac.jp, xiu.h.aa@m.titech.ac.jp, oshio.h.aa@m.titech.ac.jp, matsuoka.m.ab@m.titech.ac.jp*
[2] *Pasco Corporation*, *4-9-6 Aobadai, Meguro, Tokyo153-0043, Japan*,
Email: *taarkh6651@pasco.co.jp*

KEY WORDS: *disaster mitigation, land use detection, deep learning, semantic segmentation*

ABSTRACT: Historical topographic maps are valuable materials that provide information concerning past land uses, which are useful for hazard assessment. In particular, past residential and water areas are considered important as they are known to be related to the current disaster damage such as liquefaction and flood. Therefore, it is desirable to detect past land uses automatically from historical topographic maps and utilize them for hazard assessment purposes for large areas. To enable automatic detection, we first digitized and geo-referenced 1,300+ maps (from the 1890s to 1930s). Next, we created the training data manually to train and evaluate the performance of the deep learning model. We use U-Net, a standard network for semantic segmentation, for our experiments. Furthermore, to increase the number of training data without additional annotations, we adopt a data augmentation method that effectively improves detection accuracy. As a result, our trained model achieves an F1-score of 0.87 for residential areas and 0.97 for water areas on unseen test data.

1. INTRODUCTION

Historical topographic maps were created and published by the Geospatial Information Authority of Japan in the past, and they are valuable materials that provide information concerning past land uses and topography. Therefore, the spatial and temporal geographic information in them is utilized in many research fields such as disaster *mitigation*, urban planning, environmental engineering, and social sciences.

Especially, in the field of disaster *mitigation*, past land uses are known to be related to current disaster damages and ground conditions. Among past land uses, residential and water areas (examples are shown in Figure 1) are known to be highly relevant to current ground disaster damages such as floods and liquefaction. The specific relationship between past residential and water areas and current disaster damages is shown in Figures 2 and 3. As shown in Figure 2, the flood that occurred in 2018 did not affect most of the past residential areas, indicating that those areas are still useful today to identify low disaster risk areas. Furthermore, the comparison of the current and historical topographic maps reveals that liquefaction damages are highly relevant to past water areas since they are more likely to have a high groundwater level and loose sandy soil (Figure 3).

To utilize the above observations for hazard assessment purposes, it is necessary to analyze the relationship between past land uses and current disaster damages on GIS. However, in Japan, most the historical topographic maps have not been digitized and georeferenced. Furthermore, since the number of historical topographic maps is huge, it is required to automatically detect land uses from them to conduct the analysis.

Therefore, in this study, we digitized and georeferenced 1,337 historical topographic maps to create a well-prepared dataset for subsequent analysis. Moreover, we propose a method to automate the detection of residential and water areas from digitized and geo-referenced historical topographic maps using deep learning.

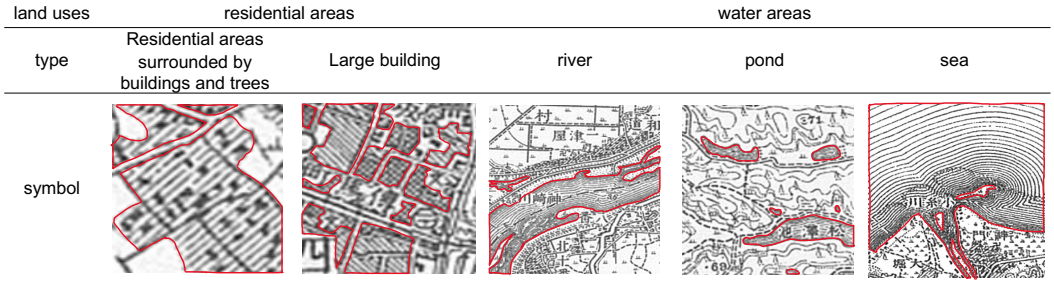
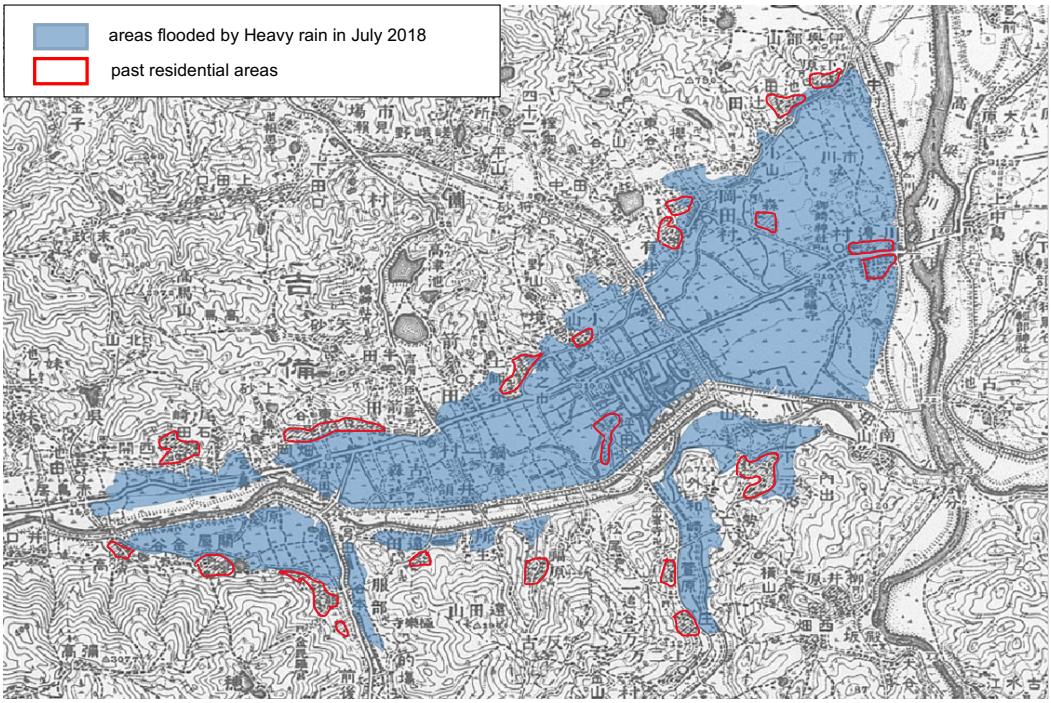
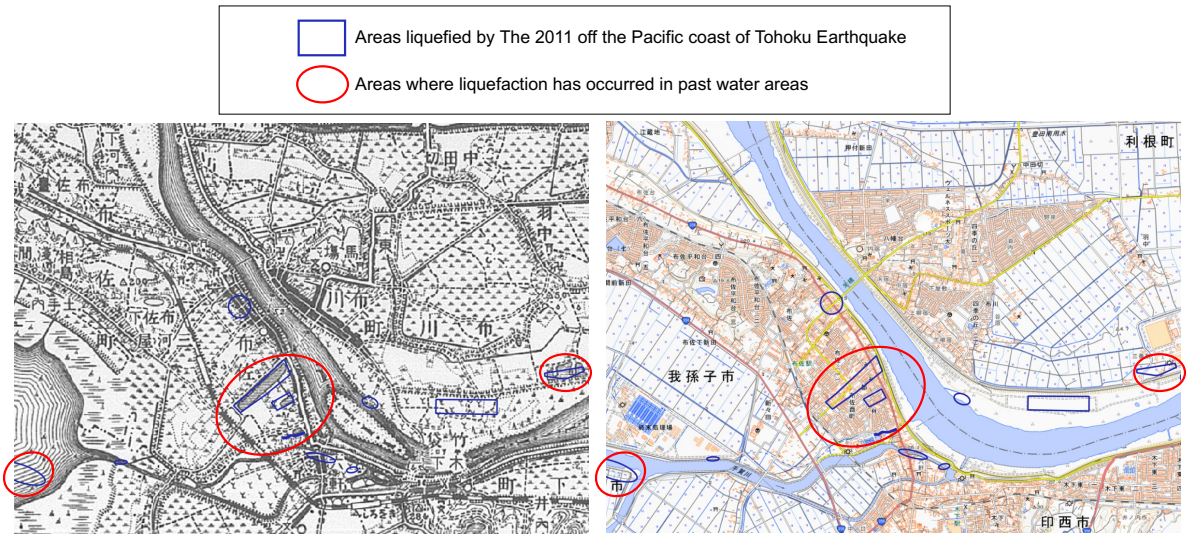


Figure 1 Examples of residential and water areas (The red lines indicate the symbols for each type of land use.)



A map of Kurashiki, Okayama Pref. in 1903

Figure 2 Past residential areas were almost unaffected by the 2018 flood, indicating that they are still effective in identifying areas of low disaster risk today.



A map of Abiko, Chiba Pref. in 1903 A map of Abiko, Chiba Pref. in 2014

Figure 3 Relationship between past water areas and liquefaction damages (The source of the image on the right is the Geospatial Information Authority of Japan.)

2. MATERIALS AND METHODS

2.1 Materials

2.1.1 Geo-referencing of Maps

We first scanned and digitized 1,337 historical topographic maps, which have a different geographical coordinate system from the modern maps. We then converted the geographical coordinate system from “Tokyo Datum” to “Japanese Geodetic Datum 2000” and georeferenced them by aligning the triangulation station with the current ones.

2.1.2 Dataset

We chose 7 map sheets for residential areas and 6 map sheets for water areas of scale 1:50,000 (Table 1) and divided them and manually created ground truth images (positive for residential or water areas, negative for non-residential or non-water areas) vertically into two equal parts and then split 6:2:2 from left to right for training, validation, and test data (Figure 4(d), (e)).

In general, a large amount of training data is preferred to train deep learning models. In this study, however, the amount of training data was limited due to the time-consuming processes of manually creating the ground truth images, as mentioned above. Therefore, we augmented the training data by cropping 512 x 512-pixel patches with shifting horizontally and vertically, and further rotating or horizontally flipping the patches (Figure 4(g), (h)).

From Table 2 that shows the composition of each data set and the balance between the number of positive and negative pixels, it can be seen that the data used in these experiments is imbalanced, with fewer positive pixels than negative ones. Therefore, to reduce the negative pixels, we used only patches containing more than 1000 positive pixels as training data.

Table 1 List of map sheets used in each dataset

map sheets for residential areas							
covering area	Nagoya (Aichi)	Fukuoka (Fukuoka)	Kumagaya (Saitama)	North east of Osaka	North west of Tokyo	Kadena (Okinawa)	Kokura (Fukuoka)
measurement year	1891	1900	1907	1908	1909	1921	1923

map sheets for water areas						
covering area	Atsuta (Aichi)	Kashima (Ibaraki)	Ryugasaki (Ibaraki)	Sawara (Chiba)	North east of Osaka	Kishiwada (Osaka)
measurement year	1891	1903	1903	1906	1908	1909

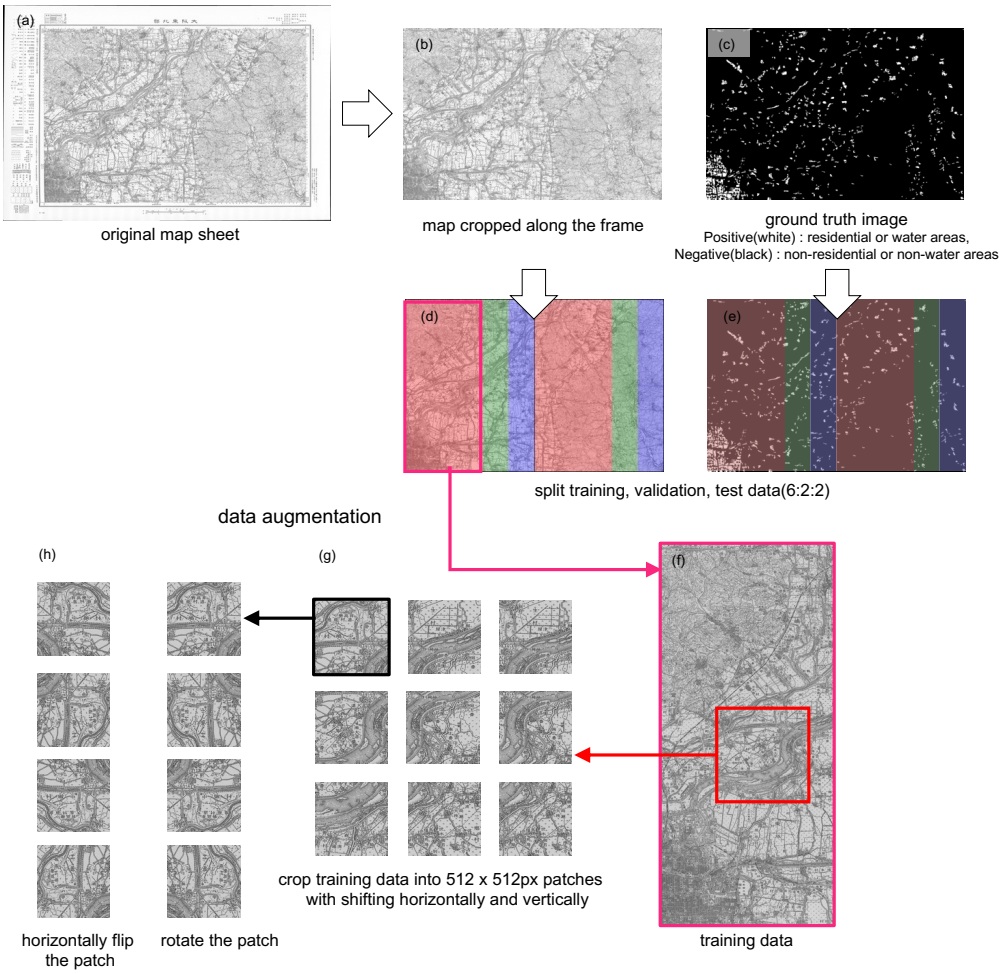


Figure 4 How to create datasets

Table 2 The composition and the balance of positive and negative pixels of each dataset

Residential areas	the number of patches	the balance of positive and negative data	
		positive	negative
train	4416	8.18%	91.82%
valid	368	6.78%	93.22%
test	368	6.91%	93.09%
total	5152	6.36%	93.64%

Water areas	the number of patches	the balance of positive and negative data	
		positive	negative
train	6040	20.59%	79.41%
valid	312	21.07%	78.93%
test	312	21.65%	78.35%
total	6664	18.77%	81.23%

2.2 Methods

2.1.1 U-Net Architecture

In this study, we used U-Net (Ronneberger et al., 2015), a standard network for semantic segmentation (Cordts et al., 2016). Semantic segmentation is a task that performs classification for each pixel in an image. We used partially modified the original U-Net architecture for our experiments, changing the activation function from ReLU to ELU to create a more accurate model and applying a Dropout to each layer to prevent overfitting (Figure 5). Finally, we attempted to classify each pixel of the historical topographic maps into two classes, residential or not, using one trained model for residential areas, and water areas or not, using each of the three trained models for water areas.

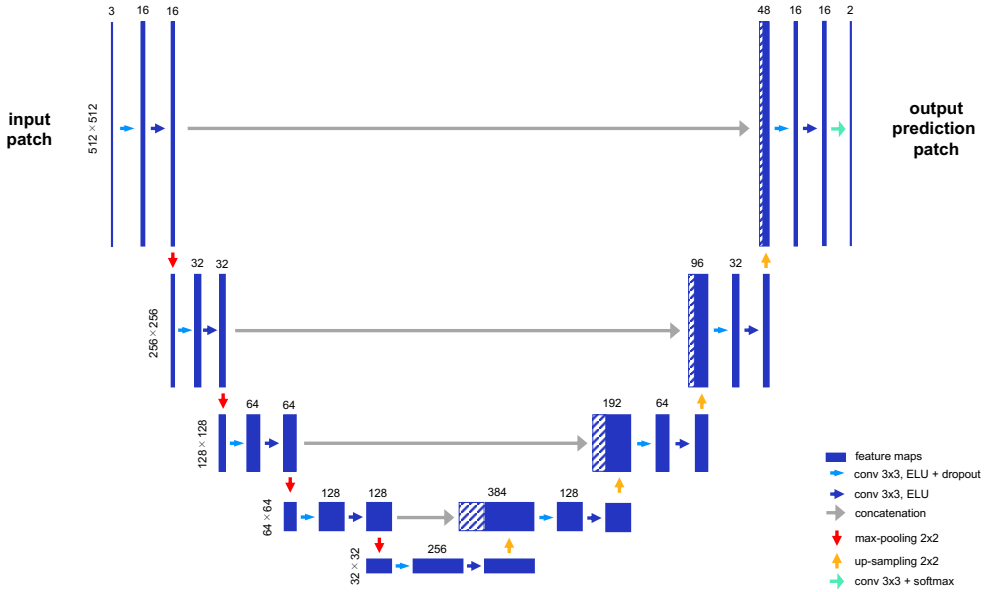


Figure 5 U-Net architecture used in our experiments

2.2.2 Training Setup and Hyperparameters

We implemented all of our networks using PyTorch (Paszke et al., 2019). For all processes of training, we used cross-entropy loss as the loss function and the Adam optimizer with an initial learning rate of 0.001 and ReduceLROnPlateau as the learning-rate schedule. The learning rate decayed by 10% if the loss value does not decrease for 3 epochs. The batch size was set to 8. All networks were trained using NVIDIA Tesla P100 GPUs in TSUBAME3.0. All computations in this article were carried out by using the TSUBAME3.0 supercomputer at the Tokyo Institute of Technology.

3. EXPERIMENT RESULTS

The following experiments demonstrate the effectiveness of the trained model on unseen test data.

3.1 Results of Residential Area Detection

For the training of the residential area detection model, the number of epochs was set to 100 because the loss continued to decrease steadily up to 100 epochs. After the training, we set the cutoff point to the value at which Precision and

Recall (definitions are shown in Table 3) were balanced and high in the PR curve (a graph with Recall on the horizontal axis and Precision on the vertical axis) when validation data was input to the model (Figure 6). Accordingly, the cutoff point was 0.6, and the performance of the model on unseen test data is shown in Table 4 and Figure7. The blue circles in Figure 7 show relatively small residential areas (3,600~4900 m^2) were either omitted or falsely detected. We speculate that this is caused by the lack of tiny residential areas in the training data. Overall, however, Table 4 and Figure 7 verify that our model detects residential areas accurately.

Table 3 Confusion matrix and evaluation indices

		Predicted	
		Negative	Positive
Actual	Negative	TN (True Negative)	FP (False Positive)
	Positive	FN (False Negative)	TP (True Positive)

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1score = \frac{2Precision \cdot Recall}{(Precision + Recall)}$$

$$IoU = \frac{TP}{TP + FN + FP}$$

Figure 6 PR curve (The orange dot is the cutoff point.)

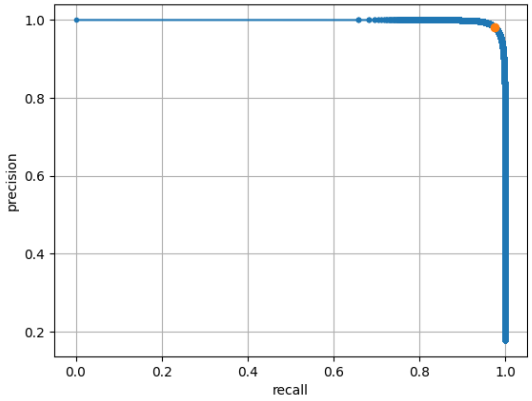


Table 4 The performance of the model on test data

Precision	0.857
Recall	0.896
F1-score	0.876
IoU	0.780

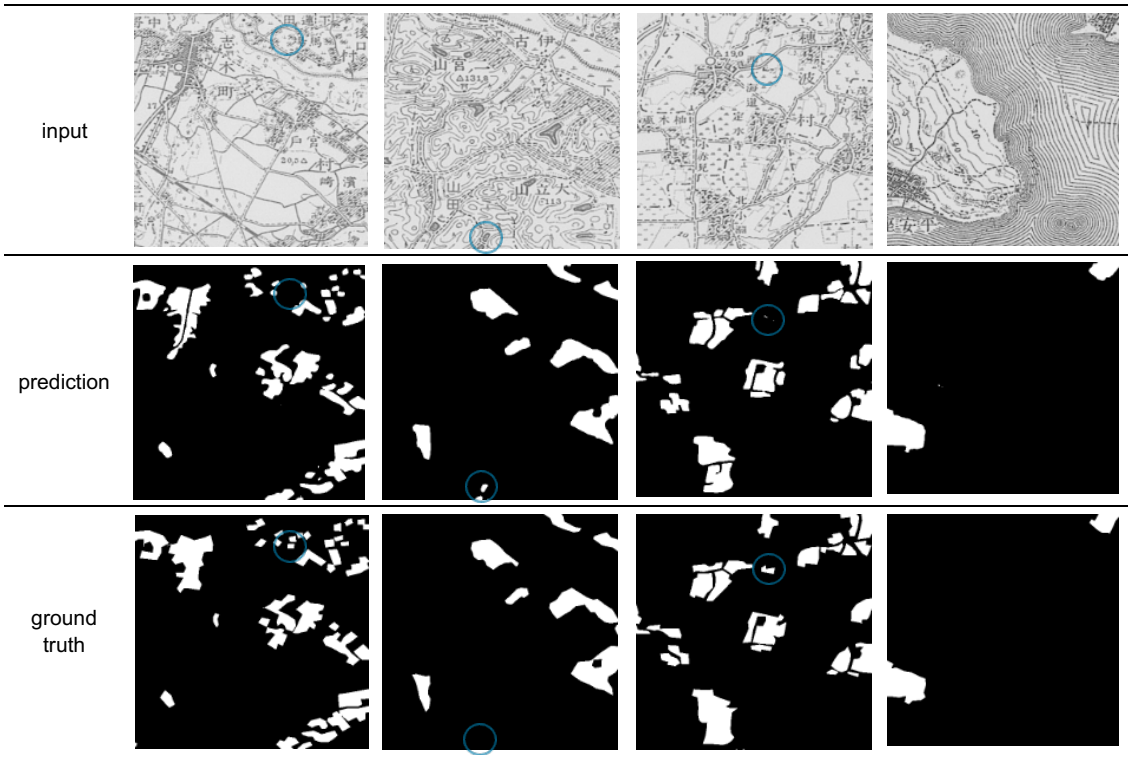


Figure 7 The performance of the model on test data (The blue circles indicate false positives/negatives.)

3.2 Results of Water Area Detection

When the model for detection of water areas was trained under the same conditions as for residential areas (only the number of epochs was changed to 200 because the loss continued to decrease steadily up to 200 epochs) and the cutoff point was set, the performance of the model (hereinafter called model I), was poor, and there were cases where contour lines were mistakenly detected as water areas (Figure 8 (c), (d)). Since water areas have similar patterns to contour lines, we needed Model I to learn to distinguish between the two. However, some patches containing contour lines (i.e., mountainous areas) had no water areas at all, and these were not included in the training data under the condition of " training only patches that contained water areas at least 1000 pixels". Therefore, we conjecture that model trained under the condition failed to learn the contour lines well.

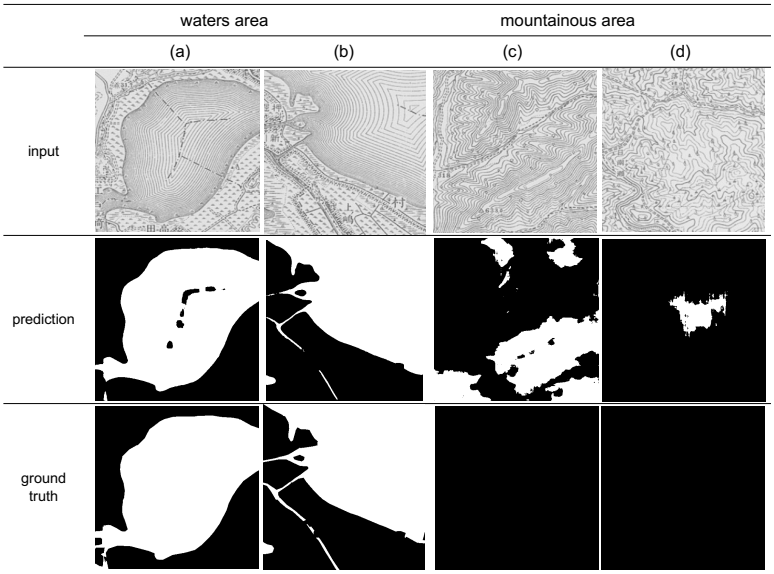


Figure 8 The performance of the model I on test data

Thus, to validate the above issue, we additionally trained model I, using training data that also included patches with less than 1000 pixels of water areas. In addition, to make a model more distinguishable between water areas and contour lines, we added map sheets containing many mountainous areas to the original six map sheets (Table 5). The composition of the dataset including the added data and the number of positive and negative pixels are shown in Table 6. Regarding the additional trained model, referred to as model II (the training methods for model II are schematically illustrated in Figure 9), the number of epochs was set to 100. The batch size was set to 32 due to GPU's memory usage and to speed up the training. The cutoff point was 0.67 when determined in the same way as for residential areas, and the performance of model II on unseen test data is shown in Table 7 and Figure 10.

Table 5 List of map sheets added to the dataset of water areas

covering area	Tsushima (Aichi)	Okayama (Okayama)	Asahikawa (Hokkaido)	Yamaguchi (Yamaguchi)	Ashiya (Fukuoka)	Fukuoka (Fukuoka)
measurement year	1891	1897	1898	1899	1900	1900

covering area	Tateoka (Yamagata)	Kanazawa (Ishikawa)	North west of Kyoto	South east of Kyoto	Turugi (Ishikawa)	Kokawa (Wakayama)
measurement year	1901	1909	1909	1909	1909	1910

Table 6 The composition and the balance of positive and negative dataset of water areas for model II,Ⅲ

Waters	the number of patches	the balance of positive and negative data	
		positive	negative
train	17136	14.05%	85.95%
valid	590	14.94%	85.06%
test	590	14.67%	85.33%
total	18316	14.44%	85.56%

Table 7 The performance of the model Ⅰ,II, Ⅲ on test data

	model Ⅰ	model II	model Ⅲ
Precision	0.828	**0.974**	0.927
Recall	0.938	**0.975**	0.916
F1-score	0.880	**0.974**	0.922
IoU	0.786	**0.950**	0.854

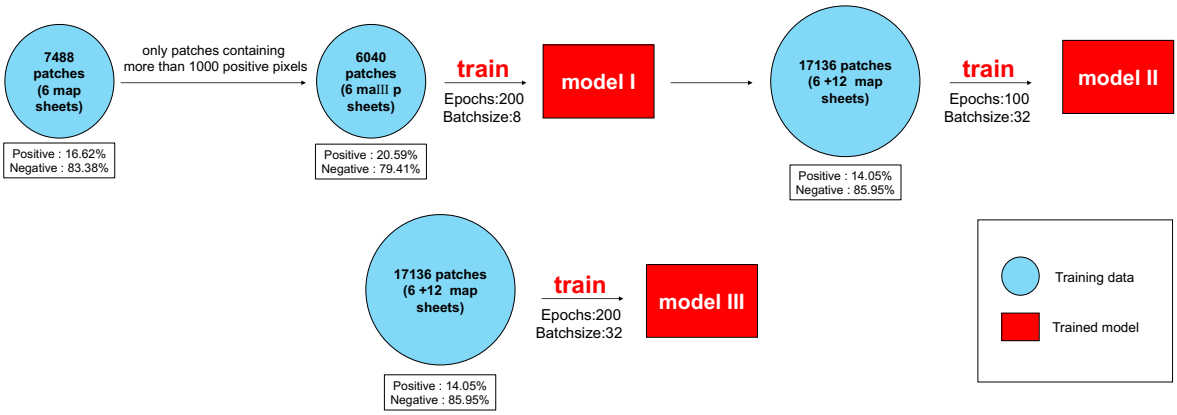


Figure 9 How to train each model

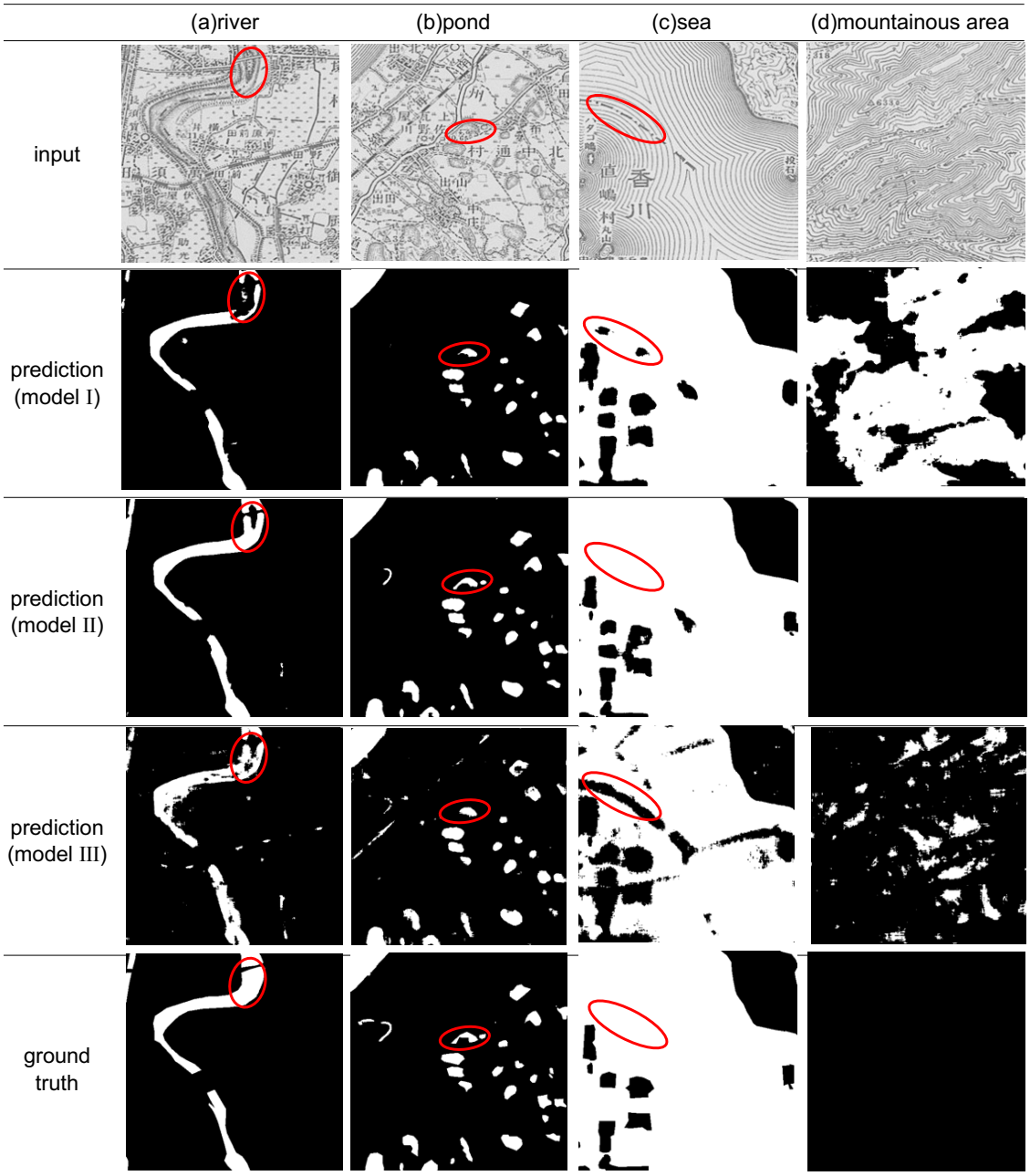


Figure 10 The performance of the model I &II on test data
(The red circles indicate areas of improvement in model II.)

Table 7 and Figure 10 also show the performance of the model trained from the beginning with the data in Table 6 (hereafter referred to as Model III) with 200 epochs (the number of epochs was determined as described because the losses continued to decrease steadily as in the previous model up to 200 epochs) and 32 batch sizes. In conclusion, model II has improved performance over models I and III, with fewer false positives/negatives. In other words, the performance of a model was improved by initially learning more balanced data and then having the model learn again all of the data, including the unbalanced data, rather than learning all of the imbalanced data from the beginning. We consider the improved performance of model II is due not only to the increased amount of training data but also to the fact that once the model learned the characteristics of the water areas on the balanced data, it became more robust to imbalanced data compared to the model that learned only imbalanced data (model III).

4. CONCLUSIONS

We trained models using U-Net, one of the networks of deep learning, and attempted to automatically detect residential and water areas from historical topographic maps to analyze the relationship between past land uses and current disaster damages on GIS. As a result, our trained models achieved an F1-score of 0.876 for residential areas and 0.974 for water areas on unseen test data. We also confirmed that when learning imbalanced data, we can expect to improve the performance of models by first training them with more balanced data and then training them with imbalanced data.

In the future, we will analyze and quantify the relationship between past land uses and current disaster damages to utilize residential and water areas detected from historical topographic maps for hazard assessment purposes.

5. REFERENCES

Cordts, M. et al., 2016. The cityscapes dataset for semantic urban scene understanding. Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 3213–3223.

Paszke, A. et al., 2019. PyTorch: An imperative style, high-performance deep learning library. Proc. Adv. Neural Inf. Process. Syst., pp. 8026–8037.

Ronneberger, O. et al., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention (MICCAI) Vol. 9351, pp234-241.