Semantic Segmentation in Red Relief Image Map by UX-Net

Tomoya Komiyama[1], Kazuhiro Hotta[1], Kazuo Oda[2], Satomi Kakuta[2] and Mikako Sano[2]
[1]*Meijo University, Shiogamaguchi, 468-0073, Nagoya, Japan*
[2]*Asia Air Survey co.,ltd, Kawasaki, 215-0004, Kanagawa, Japan*

Keywords: Semantic Segmentation, Red Relief Image Map, U-Net, UX-Net.

Abstract: This paper proposes a semantic segmentation method in Red Relief Image Map which a kind of aerial laser image. We modify the U-Net by adding the paths between convolutional layer and deconvolutional layer with different resolution. By using the feature maps obtained at different layers, the segmentation accuracy is improved. We compare the segmentation accuracy of the proposed UX-Net with the original U-net. Our proposed method improved class-average accuracy in comparison with the U-Net.

1 INTRODUCTION

Red Relief Image Map is a new topographical expression technique (Chiba Tatsuro et al., 2010). Figure 1 shows the example of Red Relief Image Map. Red Relief Image Map is created by Digital Elevation Model (DEM) data obtained from aerial laser survey and ground truth image is created by visual inspection with reference to DEM data. Red Relief Image Map expresses amount of inclination with red chroma and ridges, valleys, and the like with red brightness, and it is outstanding for reading performance. For example, it can understand roads and livers in the mountains and defective areas that we could not estimate the ground by trees. When there are topographic changes, the computer must understand the changes immediately from Red Relief Image Map. Therefore, in this paper, we carry out semantic segmentation of four classes (road, liver, defective areas by trees and others) in Red Relief Image Map.

Deep Learning gave high accuracy on various kinds of image recognition tasks such as object categorization (Huang et al., 2016), object detection (Ren et al., 2014) and object segmentation (Long et al., 2015). For object segmentation, the Encoder-Decoder Convolutional Neural Network (CNN) (Kendall et al., 2016) such as U-Net (Ronneberger et al., 2015) worked well. We modify the U-Net for improving the accuracy of semantic segmentation from Red Relief Image Map.

U-net used the path between encoder and decoder with the same resolution in order to compensate for the information eliminated by encoder. However, the information at different layer could be effective for semantic segmentation because each layer extracts different kinds of information. For example, shallower layer has fine information such as small object and correct position of objects. Deeper layer has the information related to classification. Thus, we add the path between encoder and decoder with different resolution to the U-net. By using the feature maps with different resolution, the segmentation accuracy is improved.

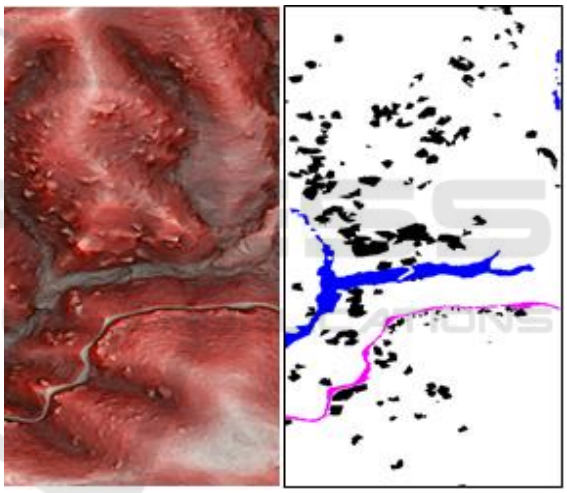

Figure 1: Example of Red Relief Image Map (left) and its ground truth image with 4 class labels (right). Black pixels are “defective areas by trees”, blue pixels are “road”, pink pixels are “river” and white pixels are “others”.

We evaluated our method on semantic segmentation problem using eleven Red Relief Image Maps. We segment four categories; trees,

Komiyama, T., Hotta, K., Oda, K., Kakuta, S. and Sano, M.
Semantic Segmentation in Red Relief Image Map by UX-Net.
DOI: 10.5220/0006716805970602
In *Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2018)*, pages 597-602
ISBN: 978-989-758-276-9

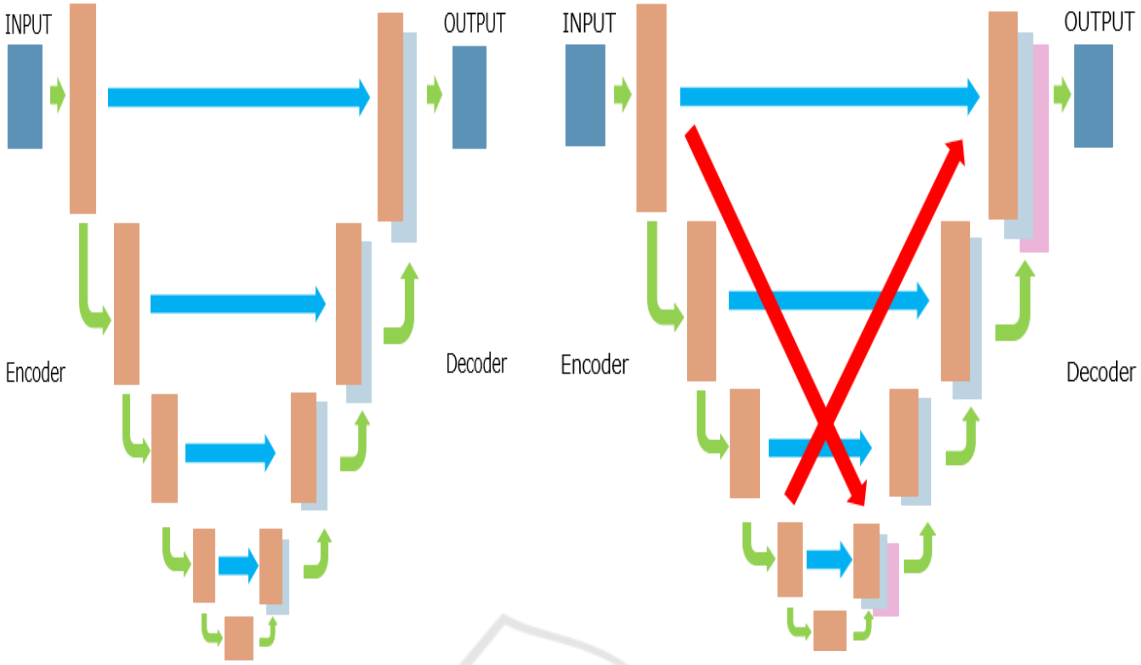

Figure 2: Structure of two networks. (a) Structure of U-Net (left). (b) Structure of UX-Net (right).

road, river and others in experiments. Our proposed method improved the accuracy in comparison with the U-Net.

This paper is organized as follows. Section 2 describes the details of the proposed method. Section 3 shows the experimental results. Comparison with the original U-net is also shown. Finally, we describe conclusion and future works in Section 4.

2 PROPOSED METHOD

In general, the number of training data for the U-net depends on the number of pixels in training images. Thus, we do not need to use a large number of training images. In this paper, we have only 11 Red Relief Image Map with ground truth. Therefore, we use the U-net as the baseline and modify it.

We explain the original U-Net in section 2.1. The proposed method is explained in section 2.2.

2.1 U-Net

U-Net is a kind of encoder-decoder CNN and is effective for semantic segmentation. In recent years, it is also used for image generation task such as pix2pix (Isola et al., 2017) which improved Deep Convolutional Generative Adversarial Networks (Radford, et al., 2016). Encoder-Decoder CNN carries out convolution at encoder part and deconvolution at decoder part in order to make the segmentation result.

U-Net improved the segmentation accuracy by using the feature map at the encoder parts in decoder parts with the same resolution as shown in Figure 2 (a). The paths from encoder part to decoder part compensate for the small objects and edges eliminated at encoder parts.

2.2 UX-Net

A structure of the proposed network is shown in Figure 2 (b). In addition to the original path of the U-net, we give the path from the shallow layer at encoder part to the beginning of decoder part in order to use the fine information at the shallow layer in the decoder part with small resolution. Since the beginning of decoder part does not have fine information such as small objects, edges and correct position of object, the feature at shallow layer should be useful. Furthermore, we also add the path from deep layer at encoder part to the final layer at decoder part. Since the feature map at the deep layer of encoder part has the information about object categories, the information should be useful to make a final segmentation result. New adding paths are like “X” shape. Thus, we call the proposed network “UX-Net”.

Table 1: Accuracy of the proposed method and U-Net.

	Pixel-wise accuracy	Defective areas by trees	River	Road	Others	Class-average accuracy
U-Net	98.08%	50.25%	82.62%	93.65%	99.46%	81.49%
UX-Net1	97.76%	74.23%	89.87%	96.67%	98.45%	89.81%
UX-Net2	97.70%	77.32%	90.34%	97.05%	98.31%	90.76%

However, the size of feature maps of shallow layer at encoder part and that of beginning layer at decoder part is different. Thus, we use pooling to be the same size. Similarly, since the size of deep layer at encoder part and that of final layer at decoder part is different, we use unpooling to be the same size.

We use batch normalization (Ioffe and Szegedy, 2015) at each layer though original U-net did not use it. Class balancing (Badrinarayanan et al., 2016) is also used to improve the segmentation accuracy of objects with small area.

3 EXPERIMENTS

We show experimental results on semantic segmentation in Red Relief Image Map. At first, we explain the dataset that we use in the following experiments in section 3.1. Comparison methods are explained in section 3.2. Experimental results are shown in section 3.3.

3.1 Dataset

In this paper, we use eleven Red Relief Image Maps. Five images are used for training images and remaining six images are used for test. Since some quantity of training images are necessary for training deep learning, we crop a local region of 256 x 256 pixels with overlapped ratio 0.7 from Red Relief Image Map of 1,500 x 2,000 pixels. In addition, we rotate those cropped regions at the interval of 90 degrees to enlarge the number of training images. As a result, the number of training images is 7,344. Test regions of 256 x 256 pixels are cropped without overlap from the original six images. The total number of test regions is 185.

3.2 Comparison Methods

We compare our method with some networks including the original U-net. The first method is the U-Net. The second method is our proposed method. When we concatenate the feature maps of different resolution, the size of each feature map is changed by pooling and convolution or unpooling and deconvolution. We call this method “UX-Net1”.

The third method is also our method but we do not use convolution and deconvolution when we change the size of feature map. Only pooling and unpooling are used to change the size of feature maps. We call this network “UX-Net2”.

3.3 Experimental Results

We show the experimental results of all methods. As evaluation measure, we use the pixel-wise accuracy and class average accuracy. Pixel-wise accuracy is the accuracy in all pixels. This is influenced by objects of large area such as background. Class-average accuracy is the average accuracy of each class. This is influenced by objects of small area such as defective areas by trees, road and river. In this paper, class average accuracy is more important than pixel-wise accuracy because we want to segment defective areas by trees, road and river well.

We show the segmentation results of all methods in Figure 3 and 4. The first row shows input image and ground truth label. The second rows show the result by U-Net and UX-Net1. The bottom row shows the result by UX-Net2.

We show the pixel-wise accuracy and the class-average accuracy of each method in Table 1. The best result at each class is shown in red.

We found that our proposed UX-Net has higher accuracy for defective areas by trees, road and river than the original U-Net. The pixel-wise accuracy of the proposed method is worse than the U-net because the pixel-wise accuracy is influenced by the background which is not the main target.

Note that our proposed method can improve the accuracy of defective areas by trees that are hard to segment by the U-net. This is because we use the “X-path” that the fine information obtained at shallow layer is used in deep layer and semantic information obtained at deep layer is used to general the final segmentation result. When we compare UX-Net1 with UX-Net2, UX-Net2 gave better result than UX-Net1. The main difference is how to change the feature map. Experimental results show that only pooling and unpooling is effective to change the size. When we use pooling and

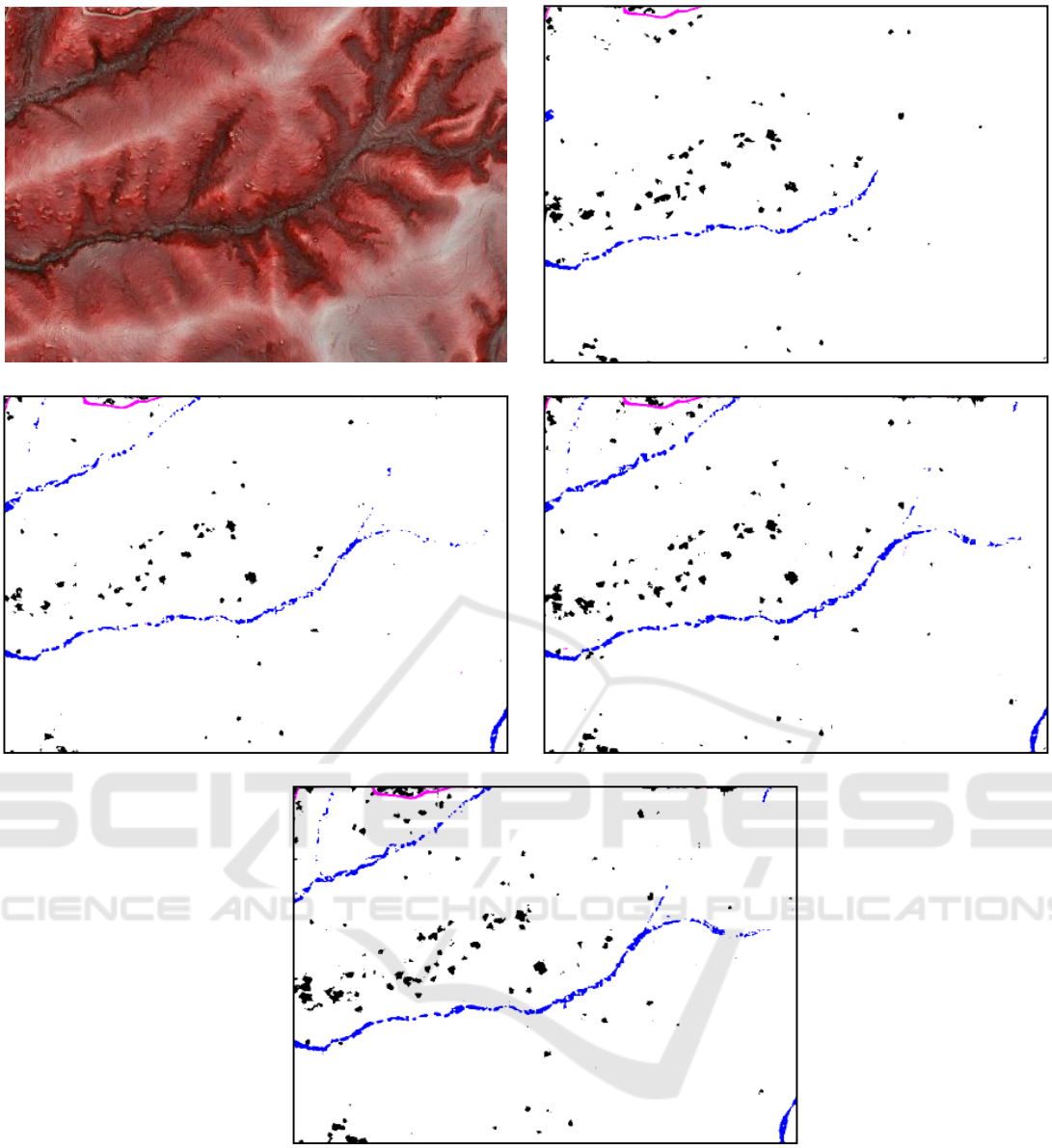

Figure 3: Segmentation results from Red Relief Image Maps. The first row shows input image and ground truth label. The second rows show the result by U-Net and UX-Net1. The bottom row shows the result by UX-Net2.

convolution, the feature map obtained by shallow layer is changed by convolution, and fine information is lost. Similarly, the semantic information may be lost by unpooling and deconvolution. These are the reason why UX-Net2 is better.

4 CONCLUSION

In this paper, we carried out semantic segmentation from Red Relief Image Map which is a kind of aerial laser image. We add "X-path" to the original U-net. X-path means that fine information is used in deep layer and semantic information is used to generate final segmentation result. Experimental results demonstrated the effectiveness of our proposed UX-Net. In particular, the accuracy of defective areas by trees, road and river is much improved in comparison with the original U-Net.

However, our proposed method has over-detection of defective areas by trees. Therefore, we want to improve the accuracy by using not only information at shallow encoder part and deep

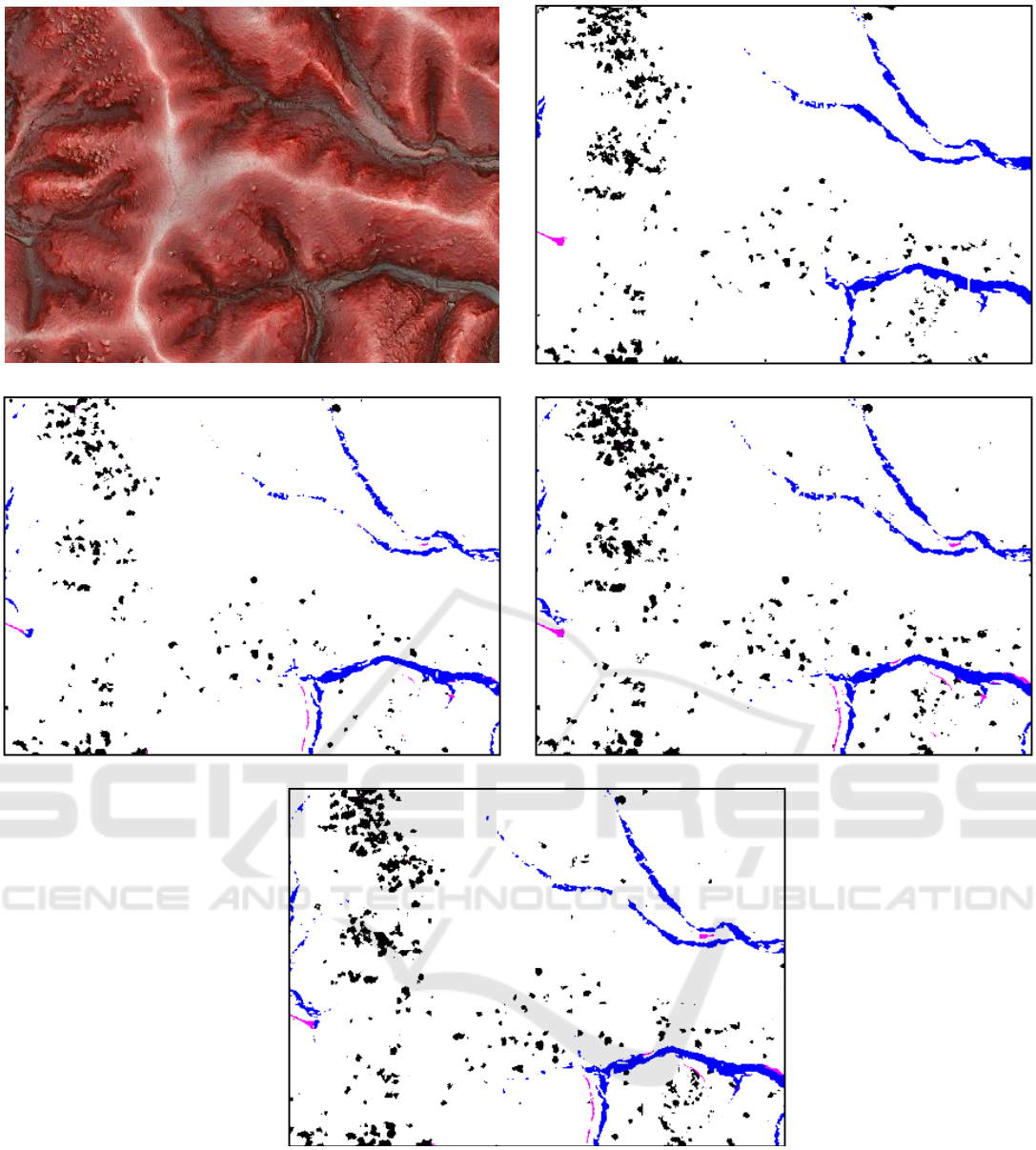

Figure 4: Segmentation results from Red Relief Image Maps. The first row shows input image and ground truth label. The second rows show the result by U-Net and UX-Net1. The bottom row shows the result by UX-Net2.

encoder part but also effectively information at various feature maps. Moreover, we adopt a loss function for considering objects which are hard to detect, and we would like to improve the class average accuracy further. These are subjects for future works.

REFERENCES

Chiba, T., Suzuki, Y., Arai, K., Tomita, Y., Koizumi, S., Nakashima, K., Ogawa K., 2010. The measurement of magma discharge volume of the "Jogan" eruption in Aokigahara on Fuji volcano, based on the micro topography by LiDAR and result of the drilling. *Journal of the Japan Society of Erosion Control Engineering*.

Huang, S., Xu, Z., Tao, D., Zhang, Y., 2016. Part-Stacked CNN for Fine-Grained Visual Categorization. *Computer Vision and Pattern Recognition*.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully Convolutional Networks for Semantic Segmentation. *Computer Vision and Pattern Recognition*.

Ren, S., He, K., Girshick, R., Sun, J., 2014. Faster R-CNN: Towards Real-Time Object Detection with

Region Proposal Networks. *Computer Vision and Pattern Recognition*.

Badrinarayanan, V., Kendall A., Cipolla R., 2016. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer Assisted Intervention*.

Isola, P., Zhu, J., Zhou, T., Efros A. A., 2017. Image-to-Image Translation with Conditional *Adversarial Networks. Computer* Vision and Pattern Recognition.

Radford, A., Metz, L., Chintala, S., 2016. Unsupervised Representation Learning With Deep Convolutional Generative Adversarial Network. *International Conference on Learning Representations*.

Ioffe, S., Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training by ReducingInternal Covariate Shift. *arXiv preprint* arXiv:1502.03167.

Badrinarayanan, V., Kendall, A., and Cipolla, R., 2016. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.