LETTER

Building Change Detection by Using Past Map Information and Optical Aerial Images

Motohiro TAKAGI[†a)], **Kazuya HAYASE**[††], **Masaki KITAHARA**[†], ***and*** **Jun SHIMAMURA**[†], ***Members***

SUMMARY This paper proposes a change detection method for buildings based on convolutional neural networks. The proposed method detects building changes from pairs of optical aerial images and past map information concerning buildings. Using high-resolution image pair and past map information seamlessly, the proposed method can capture the building areas more precisely compared to a conventional method. Our experimental results show that the proposed method outperforms the conventional change detection method that uses optical aerial images to detect building changes.
key words: *neural networks, change detection, map information*

1. Introduction

Due to the development of satellites and unmanned aerial vehicles, optical aerial images can easily be taken. Therefore, detecting changes in optical aerial images (change detection) is becoming more important for monitoring changes in environments and resources.

In change detection, building change detection is one of the key applications. Building change detection is useful for updating map information, assessing building damage caused by natural disasters. Usually, due to changes over time, it is difficult to detect changes in building accurately from images. Furthermore, building change areas tend to be relatively small, detection areas that do not contain building changes are often falsely detected in conventional methods.

In this paper, we propose a building change detection method using DNNs with map information. The proposed model consists of two DNNs: a segmentation network and a change detection network. The segmentation network detects building areas from optical aerial images. The change detection network detects changes from masked images by using building labels that are generated from optical aerial images and past map information. The proposed method integrates the segmentation task and the change detection task with past map information. Furthermore, the segmentation network and the change detection network are trained with Lovász hinge loss [1], [2]. Lovász hinge loss makes it possible to optimize the intersection-over-union (IoU) directly by DNNs. Using this loss significantly improves accuracy of change detection. The proposed method was evaluated on optical aerial images taken in Tokyo. To evaluate the proposed method, building label images and building change label images were made automatically from commercial quality map.

In this paper, we present a new method for building change detection that incorporates map data seamlessly. The proposed method can integrate the segmentation task into the change detection task. This paper is organized as follows. Section 2 discusses related works. Section 3 presents the proposed building change detection method using past map information seamlessly. Section 4 presents the results of the proposed method. Section 5 summarizes the study and discusses future work.

Manuscript received October 5, 2020.
Manuscript revised January 22, 2021.
Manuscript publicized March 23, 2021.
[†]The authors are with NTT Media Intelligence Laboratories, NTT Corporation, Yokosuka-shi, 239–0847 Japan.
[††]The author is with Nippon Telegraph and Telephone Corporation, Yokosuka-shi, 239–0847 Japan.
a) E-mail: motohiro.takagi.zc@hco.ntt.co.jp
DOI: 10.1587/transinf.2020EDL8129

2. Related Work

Recently, several change detection methods that use DNNs have been proposed [3]–[6]. In the studies of Daudt et al. [5], [6], a fully convolutional network has been used to detect changes between images. Daudt et al. [5] have used convolutional neural networks for detecting urban changes by using multispectral images. These networks were inspired by the Siamese network and early fusion network [7]. These networks take two 15×15×c as input, where c is the number of color channels. Fully convolutional neural network (FCNN) architectures that use a skip-connection structure [8] have also been proposed [6]. Their FCNN architectures can be trained end-to-end. These networks can learn changes from images directly. The model proposed in the present study was inspired by the network structure used in one of the state-of-the-art methods for change detection [6]. Their method uses optical images as input. However, it is difficult to detect changes by using images only. Therefore, map information is used to improve the performance of change detection.

3. Proposed Method

We propose a new CNN model that incorporates past map information seamlessly for estimating change areas in optical aerial images. In addition, to improve change detection accuracy, we apply the Lovász hinge loss [1], [2], which is a suitable loss for the change detection task.

3.1 Change Detection with Building Segmentation

Figure 1 shows our proposed framework. Inputs $\boldsymbol{x}_1$, $\boldsymbol{x}_2$,

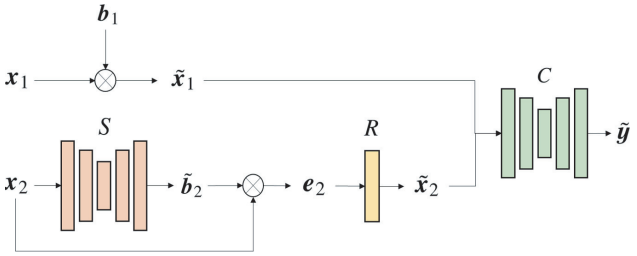


Fig. 1 Proposed method for detecting building changes.

and $\boldsymbol{b}_1$ are used to estimate the areas containing building changes. $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are optical aerial images containing areas with changes in buildings. $\boldsymbol{x}_1$ is a past aerial image, and $\boldsymbol{x}_2$ is a current aerial image. $\boldsymbol{b}_1$ is the binary building label of $\boldsymbol{x}_1$, and it is made from vector data of commercial maps. First, the building areas of image $\tilde{\boldsymbol{x}}_1$ (which is the masked image of $\boldsymbol{x}_1$) is extracted. $\boldsymbol{b}_1$ is used as the mask. Next, the building areas of $\boldsymbol{x}_2$ are estimated by using a segmentation network S. $\tilde{\boldsymbol{b}}_2$ expresses the estimated building area of $\boldsymbol{x}_2$, and $\boldsymbol{e}_2$ is the multiple value of $\boldsymbol{x}_2$ and $\tilde{\boldsymbol{b}}_2$.

$\tilde{\boldsymbol{x}}_1$ and $\tilde{\boldsymbol{x}}_2$ are used as the inputs of change detection network C, which estimates building change areas $\tilde{\boldsymbol{y}}$. The loss between $\tilde{\boldsymbol{y}}$ and change label image $\boldsymbol{y}$ is then calculated. A structure like U-Net [8] is used as the base model of the segmentation network S and the change detection network C.

3.2 Integration of Segmentation Network

Our proposed method uses the segmentation network and the change detection network, which have the same network structure except for the input. The segmentation network estimates the building areas of the current images.

$$\tilde{\boldsymbol{b}}_2 = S(\boldsymbol{x}_2). \tag{1}$$

$\tilde{\boldsymbol{b}}_2$ is the estimated building label image. To make an image pair with the images that were masked by the building areas from the past image and the current image, $\boldsymbol{x}_2$ is multiplied by the estimated building label image $\tilde{\boldsymbol{b}}_2$ in each channel.

$$\boldsymbol{e}_2 = \tilde{\boldsymbol{b}}_2 \boldsymbol{x}_2. \tag{2}$$

To simulate the mask operation, a threshold rectified linear unit (ReLU) R was applied as follows:

$$\tilde{\boldsymbol{x}}_2 = R(\boldsymbol{e}_2) \tag{3}$$

$$= \max(\boldsymbol{e}_2, T), \tag{4}$$

where T is the threshold. $\tilde{\boldsymbol{x}}_2$ is used as the input to be paired with the past image that was masked by using the building area. This operation can remove false-positive areas that the segmentation network predicts.

3.3 Model Optimization

An encoder-decoder network like U-Net is effective for detecting building changes [5]. However, it is difficult to capture the building area accurately with a building change label. Buildings are ideally labeled with building change labels according to the shape of the building. To capture the building change area accurately, we introduce the Lovász hinge loss [1], [2] to train the networks. Based on the Lovász extension, the loss is expressed as

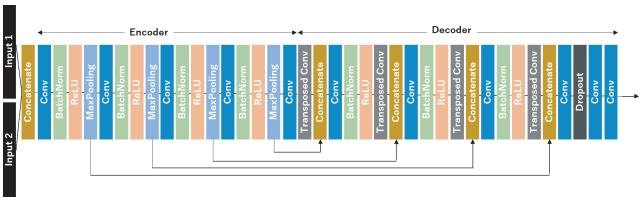


Fig. 2 Proposed network structure of the change detection network.

$$loss(\tilde{\boldsymbol{y}}) = \overline{\Delta_J}(\boldsymbol{m}(\tilde{\boldsymbol{y}})). \tag{5}$$

This is the Lovász hinge loss applied to the Jaccard loss. $\overline{\Delta_J}$ is the Lovász extension of the Jaccard set function [1], [2]. $\tilde{\boldsymbol{y}}$ is the output scores of network C. The i-th element of misprediction $\boldsymbol{m}$ is defined below:

$$m_i = \max(1 - \tilde{y}_i y_i, 0). \tag{6}$$

$\tilde{y}_i$ is the i-th element of the output scores $\boldsymbol{C}$. y_i is the i-th element of the labels $\boldsymbol{y}$. Equation (5) is used to optimize the change detection network. The loss of building area detection can be expressed as

$$loss(\tilde{\boldsymbol{b}}_2) = \overline{\Delta_J}(\boldsymbol{m}(\tilde{\boldsymbol{b}}_2)). \tag{7}$$

$\tilde{\boldsymbol{b}}_2$ is the output scores of network S.

As for the order of the learning procedure, first, the segmentation network is trained by using Eq. (7) to segment the buildings from the current images. Next, segmentation network S and change detection network C are trained end-to-end by using Eq. (5).

3.4 Network Structure

The structure of the change detection network is shown in Fig. 2. The encoder of the network consists of a repeated structure with 3×3 convolutions, batch normalization, ReLU, and 2×2 max pooling operation. The convolutional filters are initialized by the normalization method proposed by He et al. [9]. The decoder of the network consists of a repeated structure with 3×3 convolutions, batch normalization, ReLU, and transposed convolution with stride 2. The decoder concatenates the output of the max pooling of the encoder with the output of the transposed convolution. Dropout [10] is used to improve the generalization. After dropout is used, two convolutional filters are applied, and the network outputs the estimated change labels. The output of the network is not activated by a sigmoid function. The segmentation network has a similar structure to the change detection network. The only difference between the structures is their input. The segmentation network uses a single optical aerial image as input.

4. Experiments

4.1 Experimental Setting

Optical aerial images taken in Tokyo were used as evaluation data. Building changes in regard to the optical aerial images taken in 2016 and the optical aerial images taken in 2017 were extracted from the vector data (shapefile) of the GEOSPACE electronic map [11], which contains seasonal, shadow, and brightness changes. Change labels for the buildings were created by taking the difference between the vector data of the buildings created from the GEOSPACE electronic map. There are two change cases of the change labels: one is when the building demolished, and the other is when the building is newly built. Hence, we estimate the change areas, building areas to non-building areas and non-building areas to building areas. The areas where buildings become different types of buildings are not targeted. The original size of the optical aerial images is 8000×6000 with a ground resolution of 25 cm/pixel. The images were divided into 128×128. The divided images containing a building change area were extracted from the optical aerial images taken in Tokyo. The evaluation data was generated from 100 optical aerial images. 70 images were used as training data and validation data, and 30 images were used as test data. 20% of the images in the training and validation data were used for validating. 22,656 divided images were used as training and validation data, and 7,813 divided images were used as test data. AdaBound [12] was used for optimizing model parameters, and the initial learning rate was set to 0.001. The batch size was 16, and the maximum number of training epochs was 200. The threshold of ReLU T was set to 0.05. This parameter was decided experimentally.

Table 1 shows the results obtained by applying each method. The proposed method was compared to the fully convolutional neural network (FCNN) proposed by Daudt et al. [6]. Three cases were evaluated. In the first case, the change detection network with the Lovász hinge loss (CDNet) learns the parameters from images only. The input and output are the same as those used by FCNN [6]. The network structure is the same as that of the proposed change detection network C. In the second case, the segmentation network and the change detection network learned independently (Independent). The segmentation network learned the parameters and estimated current building areas. The masked images of the current images were made by using the estimated building labels, and those of the past images were made by using the past building labels obtained from the map information. The estimated building labels were obtained by thresholding the output values of the segmentation network manually. A threshold is the same as the threshold of the ReLU (0.05). The change detection network used the masked images as input and learned the parameters independently. In the third case, the proposed method (end-to-end) connected the segmentation network and the change detection network seamlessly and learned the parameters in an end-to-end manner.

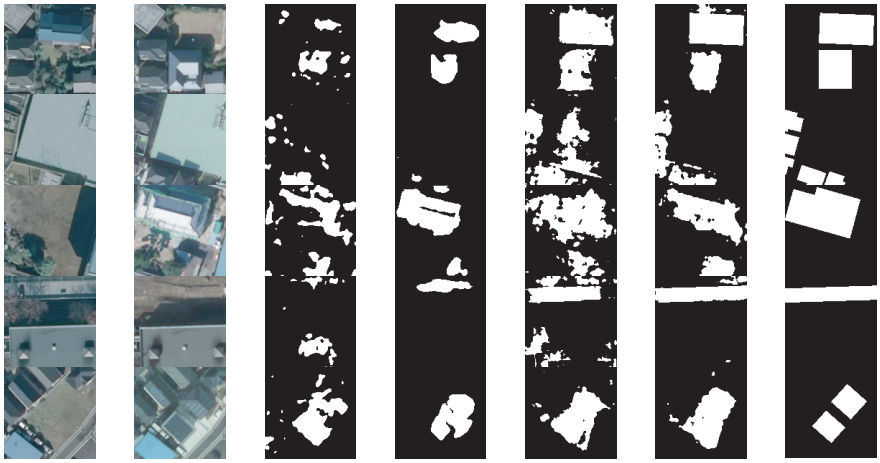
Table 1 Evaluation of building change detection.

Method	Precision	Recall	F-score	mIoU
FCNN [6]	0.332	0.523	0.391	0.258
CDNet	0.517	0.534	0.503	0.358
Independent	0.710	0.549	0.591	0.455
End-to-end	**0.760**	**0.647**	**0.675**	**0.541**

As shown in the table, the proposed method outperforms the conventional method in terms of each metric. The mean IoU (mIoU) of CDNet is 0.358. CDNet has a higher mean IoU than that of FCNN (i.e., 0.258). This result means the Lovász hinge loss is effective in regard to improving the accuracy of change detection. The mIoU of the independent model is 0.455. The independent model has a higher mIoU than that of CDNet. The independent model used the segmentation network with mIoU, precision, recall of 0.708, 0.775, and 0.841, respectively, for building detection. This result shows that masking the building areas in advance is effective, and the estimated building label estimated by the segmentation network is effective as the mask of the current images. The end-to-end model has higher values for all metrics. Back propagating the loss of the change detection network to the segmentation network is effective for improving mIoU. In the end-to-end model, the segmentation network was optimized to detect change areas. Hence, the performance of the segmentation network is an important factor for change detection. We used the segmentation network that has a low mIoU (0.356) for evaluating the difference of the segmentation network in the end-to-end model. In our experiment, the mIoU of the end-to-end model was 0.539. If the segmentation network has low mIoU, the end-to-end model can obtain similar mIoU when the segmentation network that has high mIoU is used. A more accurate segmentation model can improve change-detection results intuitively. Therefore, by replacing the segmentation network with the more accurate segmentation model, there is a possibility that the model will achieve higher mIoU.

4.2 Performance of Building Change Detection

Examples of the estimated change labels used in each method are shown in Fig. 3. As shown in the figure, the conventional method cannot capture the building areas precisely. On the other hand, CDNet can capture the building areas. CDNet tends to estimate a more integrated area than FCNN. This is considered to be due to the direct optimization of the Lovász hinge loss that represents IoU. However, CDNet gives many false-negative areas, and false detection occurs due to various differences such as color changes. The independent model and the end-to-end model give less false-negative areas compared to CDNet and FCNN. In addition, the end-to-end model can capture the building areas more precisely compared to the change areas estimated by the independent model. The end-to-end model used the segmentation network that is the same network as the segmentation network that was used in the independent model. The



(a) (b) (c) (d) (e) (f) (g)

Fig. 3 Estimated change label images: (a) past images, (b) current images, (c) change label images estimated by FCNN [6], (d) change label images estimated by CDNet, (e) change label images estimated by the independent model, (f) change label images estimated by the proposed end-to-end model, and (g) correct change label images.

main difference between the independent model and proposed end-to-end model is thresholding. The end-to-end model applies the threshold ReLU to $\boldsymbol{e}_2$ instead of thresholding manually. By training the segmentation network in an end-to-end manner, the output values of the segmentation network were optimized to detect change areas, and the estimated building areas were also optimized for the threshold ReLU. Hence, the end-to-end model can remove false-positive and false-negative areas properly and improve performance compared to the independent model. Therefore, it is concluded that the end-to-end model is significantly better compared to FCNN by using end-to-end learning with the past map information and Lovász hinge loss. This conclusion shows that the end-to-end learning can obtain a shape representing the estimated building change areas that is closer to the shape of the actual buildings.

5. Conclusion

This paper proposes a building change detection method that incorporates map data seamlessly. The proposed method integrates the segmentation task and the building change detection task with past map information. The proposed method improves the accuracy of detecting building change compared to the method that uses images only as input. For future work, we will apply the proposed network to detect changes in various classes of objects such as roads and rivers.

Acknowledgments

The aerial images and map information used in this paper were provided by NTT GEOSPACE CORPORATION (currently, it is NTT InfraNet).

References

[1] J. Yu and M.B. Blaschko, "Learning submodular losses with the Lovász hinge," Proc. Int. Conf. Machine Learning (ICML), pp.1623–1631, July 2015.
[2] M. Berman, A.T. Rannen, and M.B. Blaschko, "The Lovász-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks," Proc. IEEE Comput. Vis. Pattern Recognit. (CVPR), June 2018.
[3] J. Liu, M. Gong, K. Qin, and P. Zhang, "A deep convolutional coupling network for change detection based on heterogeneous optical and radar images," IEEE Trans. Neural Networks and Learning Systems, vol.29, no.3, pp.545–559, March 2018.
[4] Y. Zhan, K. Fu, M. Yan, X. Sun, H. Wang, and X. Qiu, "Change detection based on deep siamese convolutional network for optical aerial images," IEEE Geosci. Remote Sens. Lett., vol.14, no.10, pp.1845–1849, Oct. 2017.
[5] R.C. Daudt, B.L. Saux, A. Boulch, and Y. Gousseau, "Urban change detection for multispectral earth observation using convolutional neural networks," Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS), July 2018.
[6] R.C. Daudt, B.L. Saux, and A. Boulch, "Fully convolutional Siamese networks for change detection," Proc. IEEE Int. Conf. Image Process. (ICIP), pp.4063–4067, Oct. 2018.
[7] S. Zagoruyko and N. Komodakis, "Learning to compare image patches via convolutional neural networks," Proc. IEEE Comput. Vis. Pattern Recognit. (CVPR), June 2015.
[8] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp.234–241, Oct. 2015.
[9] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," Proc. IEEE Int. Conf. Comput. Vis. (ICCV), pp.1026–1034, Dec. 2015.
[10] N. Srivastava, G.E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," J. Mach. Learn. Res., vol.15, no.1, pp.1929–1958, Jan. 2014.
[11] "GEOSPACE electrical map," https://www.ntt-geospace.co.jp/geospace/denshi.html.
[12] L. Luo, Y. Xiong, Y. Liu, and X. Sun, "Adaptive gradient methods with dynamic bound of learning rate," Proc. Int. Conf. Learning Representations (ICLR), April 2019.