International Cartographic Association
Association Cartographique Internationale

Demolished building detection from aerial imagery using deep learning

Shu Su [a, *], Takahiko Nawata [a]

[a] *Aero Asahi Corporation, shuu-so@aeroasahi.co.jp, takahiko-nawata@aeroasahi.co.jp*

* Corresponding author

Abstract: In this paper, we present a novel approach for demolished building detection using bi-temporal aerial images and building boundary polygon data. The building boundary polygon data can enable the proposed method to distinguish buildings from non-buildings. Moreover, it can enable the exclusion of non-building changes such as those caused by changes in tree cover, roads, and vegetation. The results of demolished building detection can be achieved by using the building-base. The proposed method classifies each building as demolished or undemolished. The architectures, which based on U-Net and VGG19, are implemented for realizing automatic demolished building detection. The result suggested that U-Net is a useful architecture for image classification problems as well as for semantic segmentation tasks. In order to verify the effectiveness of proposed method, the detection performance is evaluated using images of an entire city. The results suggest that the proposed method can accurately detect demolished buildings with a low mis-detection rate and low over-detection rate.

Keywords: building change detection, aerial imagery, deep learning, building boundary polygon data

1. Introduction

Over the last decade, with advances in computer vision techniques, building change detection has emerged as an active research area in the domain of photogrammetry and remote sensing. This increase in interest may be attributed to the wide range of applications of building change detection, including such map updating and disaster evaluation.

Traditionally, change detection of buildings was performed manually by comparing aerial images from different time periods. Owing to the tedious and time-consuming nature of this task, over the past decade, researchers have developed automatic detection techniques. However, accurate change detection of buildings still remains a challenging task because of the different characteristics of the two images that arise due to differences in camera, atmosphere, or shadows. Previous studies on building change detection have mainly focused on the spectral information and color variations in bi-temporal images (Bourdis et al., 2011). However, these methods yield too many errors, including mis-detection (i.e., a demolished building going undetected), over-detection (i.e., an undemolished building being detected as demolished, or a lot of changes in non-building area.), especially over-detection.

The development of digital surface models (DSMs) have proved to be effective in improving accuracy of building extraction and change detection (Rottensteiner et al., 2007; Tian et al., 2014). The variations in height represent a robust feature that enable the evaluation of building changes. However, DSMs cannot distinguish buildings changes from non-buildings changes such as those caused by trees and vegetation between two time periods. Moreover, it is difficult to determine the height of a building, if the building is partially obstructed by tree cover.

Recently, researchers have made efforts to apply machine learning techniques for change detection of buildings. In particular, convolution neural networks, which have been proven effective for identifying objects with their appearance variations, have attracted interest in buildings change detection as well as computer vision techniques (Daudt et al., 2018; Lim et al.,2018; Maltezos et al., 2018; Pang et al., 2018).

Given the fact that there are no reliable and stable approaches for realizing automated building change detection, additional research efforts are required to address the following challenges that exist in automated building change detection.

(1) Dense urban areas

Accurate detection of small sized buildings that area closely situated is challenging, particularly in dense urban areas.

(2) High noise

With increasing urbanization, there is significant noise due to the detection of non-building changes, such as cars, change in vegetation, and presence of temporary man-made structures.

(3) Object-base (building-base) detection.

Most detection methods are based on pixel level change. In these methods, changes in pixels is the base for detection. For example, in dense urban areas, if a building

https://doi.org/10.5194/ica-proc-2-122-2019 | © Authors 2019. CC BY 4.0 License.

and three neighboring buildings are demolished, these methods will detect changes over a large area, instead detecting four buildings as being demolished. Similarly, if a building and surrounding area undergoes changes over time, these methods can only detect change in the area but cannot differentiate building change from non-building changes. In summary, these pixel-based change detection methods cannot differentiate the boundary of buildings.

(4) City-level detection

Building change detection when applied to the landscape of an entire city is challenging given the large area comprising of tens of thousands of buildings, and presence of complex structures. Similar to building change detection, demolished building change detection is also an important application that should be automated.

In Japan, people who own buildings on January 1 are required to pay property tax of that year. Each year, the municipal government undertakes a survey of building change, such as newly-constructed or demolished buildings. Therefore, to prevent taxation on demolished buildings, there is a need to identify buildings that have been demolished.

Most municipal governments update building boundary polygon data (referred to as building polygon data) for urban planning on a yearly basis. However, before updating building polygon data, it is necessary to acquire information about demolished buildings. Therefore, building polygon data from the previous year can be utilized to improve the accuracy of demolished building detection at relatively low cost. It should be noted that methods for change detection of demolished buildings also experience the same difficulties as mentioned earlier.

Therefore, in this paper, we present a novel approach for demolished building detection that uses bi-temporal aerial images and building polygon data. Demolished building detection based on building-base can be realized using building polygon data. Building polygon data not only makes it feasible to distinguish buildings from non-buildings, it also can be used to exclude non-building changes such as those caused by trees and vegetation. In this work, we used the U-Net and VGG19 networks. U-Net is regarded as a remarkable and one of the most successful and popular architectures for semantic segmentation (Ronneberger et al., 2015). In our proposed method, U-Net was used for the image classification task. On the other hand, VGG19 is a well-known architecture for image classification problems (Simonyan et al., 2015). Each building was classified into two classes—demolished or undemolished. The demolished building detection algorithm is based on building-base. The proposed method was tested on the entire area of a city to verify its performance.

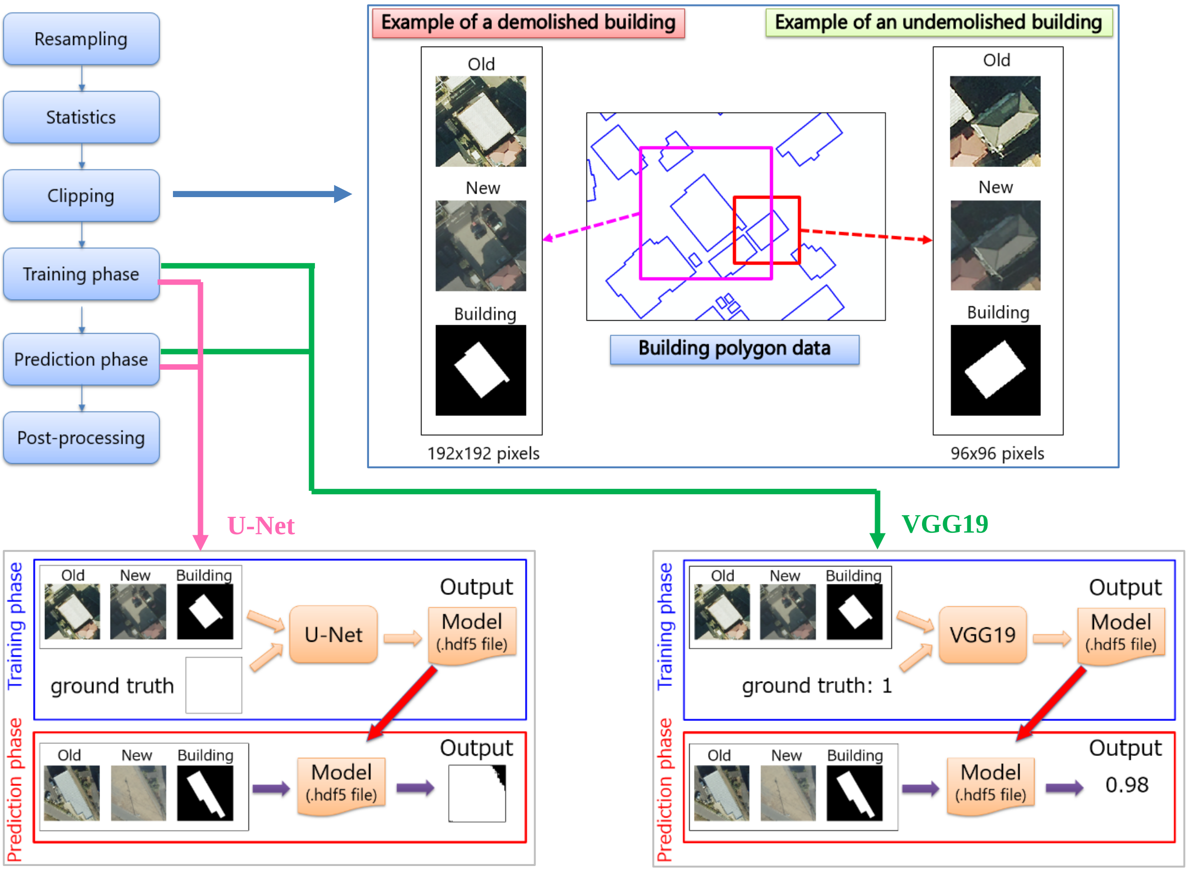

Figure 2. An overview of the proposed demolished building detection method

2. Related works

Figure 1 shows the histogram of the building area in a sample city. Buildings with area larger than 1280 m^2 only account for a small proportion of the total number of buildings. Buildings smaller with less than 5 m^2 area are mostly warehouses that are exempted from taxation. Therefore, this work only considers buildings between 5 m^2 and 1280 m^2 in terms of surface area.

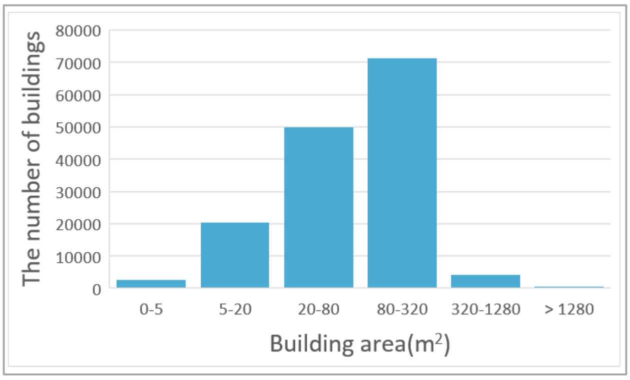

Figure 1. Histogram of building area of a sample city

An overview of the proposed demolished building detection method is illustrated in Figure 2. First, the bi-temporal aerial images are resampled into the same spatial resolution. Second, the average and standard deviation of the RGB values are obtained from the aerial image of each city for a given period. Third, the bi-temporal aerial images and building images are clipped with batch size. The training dataset is input to the network to train the detection model, and the test dataset is used for obtaining the model's prediction results. Finally, the model is applied to classify each building into two classes— demolished or undemolished.

2.1 Resampling

Bi-temporal aerial images that are acquired using different spatial resolutions are converted into the same spatial resolution.

2.2 Statistics

Figure 3 shows the histogram for Red (red line), Green (green line), and Blue (blue line) values of the bi-temporal images. The upper figure illustrates the RGB histogram of bi-temporal images of city AA acquired in 2017 (left) and 2018 (right), respectively. The bottom figure shows the RGB histogram of bi-temporal images of city BB that were acquired in 2017 (left) and 2018 (right), respectively. We observed significant differences in color variations in the bi-temporal aerial images, which may be attributed to the difference in cameras, atmosphere, and shadows. RGB histograms exhibit differences in the bi-temporal images of the same city; however, the difference is much smaller than that in observed between the bi-temporal images of two different cities. To eliminate the potential disturbances on the classification results due to these factors, the average and standard deviation from RGB values of each city in a given period were calculated. These values are used for color correction subsequently (section 2.3).

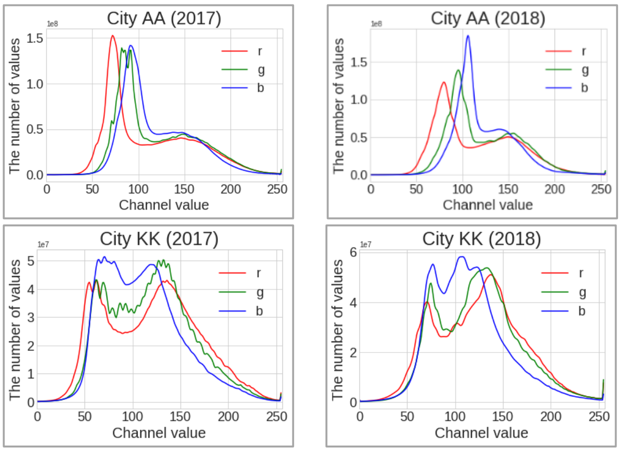

Figure 3. RGB histogram of bi-temporal aerial images

2.3 Clipping

To eliminate noise from adjacent buildings, the minimum bounding square for enclosing a building is regarded as a better choice for batch size. Figure 4 illustrates the process for obtaining the minimum bounding square using a building's polygon data (blue line). First, the minimum circle (yellow line) enclosing the building polygon is created; then, the minimum bounding square that encloses the circle (green line) is obtained.

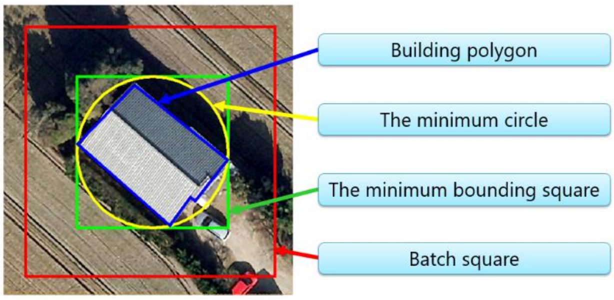

Figure 4. The minimum bounding square of a building

However, feeding the dataset to a network with tens of thousands of different batch sizes is unrealistic and time-consuming. Therefore, based on building attributes (individual house, mansion, factory, warehouse, etc.), buildings were divided into four areal ranges 5–20 m^2, 20–80 m^2, 80–320 m^2, and 320–1280 m^2. To obtain the optimal batch size for each areal range, the histogram of the minimum side length of the bounding square for each area range were plotted, as shown in Figure 5; the side of the square is measured in pixels. The red dashed lines in Figure 5 indicate the batch size used in this study. It should be noted that we did not utilize the maximal square side length in each areal range, because it is too big for most of the buildings in that areal range. However, very narrow buildings will be partially excluded from the batch.

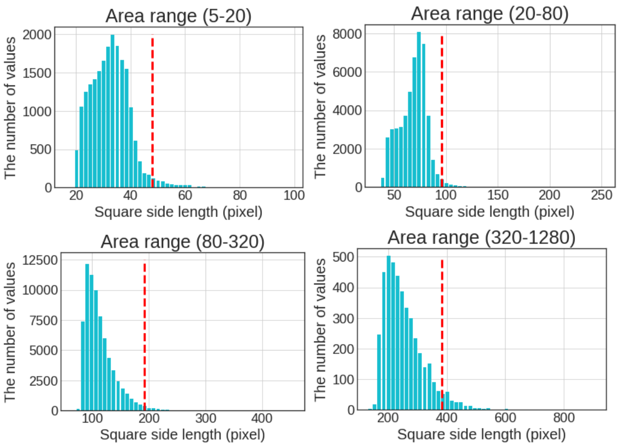

Figure 5. Histogram of square side length

Using the four steps shown in Figure 2, the dataset (training or test) for each building polygon was obtained. The procedure is described as follows:

(1) create a square per batch size. There are four batch sizes: 48x48, 96x96, 192x192, and 384x384 that bound the building area based on ranges listed in Table 1.

(2) clip the old image and the new image using the above square.

(3) correct the color of the old and new images by using the average and standard deviation of the RGB values.

(4) transform the building polygon data to a black and white image. The area inside the building boundary is white, while all the adjacent buildings, surrounding region, and non-building area is blacked out. In other words, excluding the building boundary area, all other objects/areas are masked.

In this work, the geospatial data abstraction library (GDAL) was used to generate the dataset (training or test). GDAL is a free and open source software for reading and writing raster and vector geospatial data formats. It can be installed and run in Ubuntu operating system. Therefore, the processes in this work can be processed in the Ubuntu 16.04 LTS.

Building area range(m^2)	Batch size(pixels)
5–20	48x48
20–80	96x96
80–320	192x192
320–1280	384x384

Table 1. Batch size different building area ranges

2.4 Training phase

We assigned a ground truth value for each training dataset created. For U-Net, when a building is demolished, a white image is labelled as the ground truth; Conversely, a black image is labelled as the ground truth. In the VGG19 dataset, the ground truth is 1 if the building was demolished, and 0 otherwise.

Both the U-Net and VGG19 datasets were trained for four batch sizes (48x48, 96x96, 192x192 and 384x384 pixels). We trained each size for 100 epochs. The model of the 100th epoch model was used. The model was output into a hdf5 file.

2.5 Prediction phase

For each test dataset, the output of the U-Net network is a white and black image, as illustrated in Figure 6. A higher proportion of white pixels indicates the possibility of the building in the center of image being demolished. While the output value of the VGG19 network lies between 0 and 1, a higher value indicates that a building has a higher probability of being demolished.

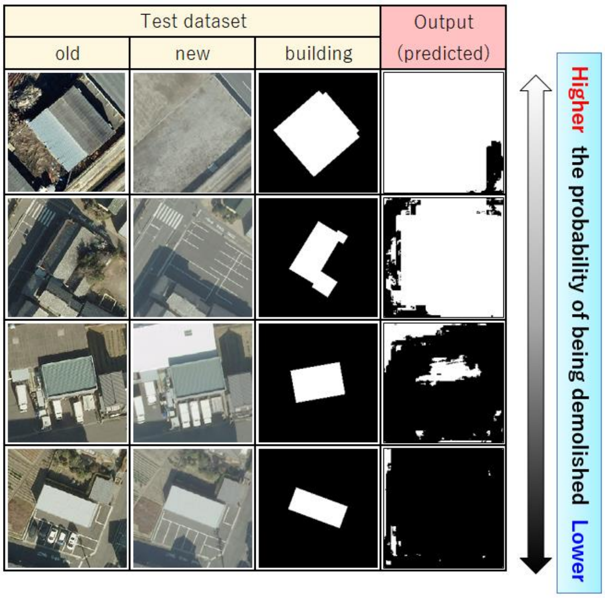

Figure 6. Sample outputs obtained using U-Net

2.6 Post-processing

Finally, the threshold of the output of U-Net is obtained as a value between 0 and 1. In this study, each building was classified as either a demolished building or an undemolished building.

3. Cross-validation

Cross-validation is a statistical technique that is used to estimate the performance of a model. Several cross-validation techniques such as, k fold cross-validation, leave one out cross-validation, stratified cross-validation, and time series cross-validation are commonly used. In this study, we used the k fold cross-validation, which divides the dataset into k groups. One group is kept for testing, and the model is trained on the other k-1 groups. Then, the process is repeated k times.

3.1 Sample dataset

Nine urban cities from eight prefectures in Japan were used for the cross-validation. The eight prefectures are located in northeast Japan, eastern Japan, and western Japan. For each city, a sample dataset was selected, as shown in Table 2.

No.	City	Sample dataset	
		Demolished	Undemolished
1	A	27	364
2	B	1258	653
3	C	719	395
4	D	56	516
5	E	58	1007
6	F	1016	2530
7	G	257	1168
8	H	937	5287
9	I	90	1837
Total		4418	13757

Table 2. Sample dataset for cross-validation

It is preferred to split the dataset sample into k groups of equal sizes. The number of demolished buildings in sample cities B, F, and H are greater than those in others cities. Therefore, we split the data into three groups (k = 3), as illustrated in Table 3.

Group	City	Sample dataset	
		Demolished	Undemolished
1	A, B, I	1375	2854
2	C, H	1656	5682
3	D, E, F, G	1387	5221

Table 3. Three groups for cross-validation

The networks were trained and evaluated using the test dataset. Figure 7 shows the image of cross-validation.

a) Trained on group 1 and group 2; Tested on group 3

b) Trained on group 2 and group 3; Tested on group 1

c) Trained on group 1 and group 3; Tested on group 2

The training dataset is randomly divided into actual training dataset (95%) and validation dataset (5%).

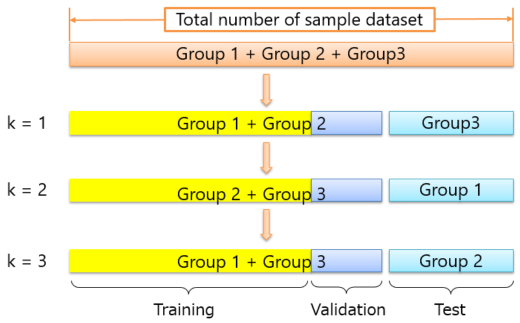

Figure 7. The image of cross-validation

3.2 Metrics

As this paper, we are interested in both mis-detection as well as over-detection of demolished building. The performance of the model is evaluated in terms of both mis-detection and over-detection. We evaluated the performance using accuracy, recall, precision, and F-score at the building-base, which is given by the following equations.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$F - score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (4)$$

Here, TP = True Positive
TN = True Negative
FP = False Positive
FN = False Negative

The confusion matrix is defined as shown in Table 4.

In this work, we define accuracy as the ratio of buildings correctly classified relative to the total number of buildings detected.

The recall rate determines the ratio of correctly predicted buildings relative to all demolished buildings. A low recall value indicates a high mis-detection rate.

Precision is defined as the ratio of the number of actually demolished buildings to the number of buildings being classified as demolished; a lower precision indicates higher over-detection rate.

F-score is the overall measure of accuracy that combines precision and recall. In other words, a good F-score indicates both low mis-detection and low over-detection rates. The ideal value for accuracy, precision, recall, and F-score is 1 (100% success), while the worst case is 0 (complete failure).

		Actual	
		Demolished	Undemolished
Predicted	Demolished	TP	FP
	Undemolished	FN	TN

Table 4. The confusion matrix

	Building area range (m^2)	Training dataset		Test dataset		Accuracy		Recall		Precision		F-score	
		D^*	Un^*	D^*	Un^*	VGG19	U-Net	VGG19	U-Net	VGG19	U-Net	VGG19	U-Net
k = 1	5–20	999	1981	318	1636	0.904	0.944	0.918	0.745	0.644	0.897	0.910	0.942
	20–80	1157	1951	527	434	0.955	0.955	0.935	0.937	0.982	0.980	0.955	0.955
	80–320	814	4310	491	2391	0.982	0.965	0.936	0.810	0.960	0.982	0.982	0.963
	320–1280	61	294	51	760	0.321	0.944	0.764	0.215	0.067	0.68	0.426	0.930
k = 2	5–20	851	2492	466	1125	0.899	0.908	0.841	0.806	0.820	0.870	0.899	0.907
	20–80	1221	1536	463	849	0.968	0.968	0.978	0.967	0.935	0.945	0.968	0.968
	80–320	884	5889	421	812	0.964	0.957	0.980	0.942	0.919	0.934	0.964	0.957
	320–1280	87	986	25	68	0.913	0.946	0.880	0.840	0.814	0.954	0.914	0.945
k = 3	5–20	784	2761	533	856	0.901	0.897	0.881	0.842	0.863	0.883	0.901	0.896
	20–80	990	1283	694	1102	0.940	0.939	0.963	0.972	0.890	0.883	0.940	0.940
	80–320	912	3203	393	3498	0.985	0.985	0.964	0.936	0.895	0.922	0.985	0.985
	320–1280	76	828	36	226	0.919	0.652	0.833	0.861	0.66	0.264	0.923	0.706

Table 5. The results of cross-validation (D^* = Demolished, Un^* = Undemolished)

3.3 Discussion

The performance of U-Net was evaluated by comparison with the results of VGG19, as shown in Table 5. The value of all metrics is calculated in terms of building-base. In Table 5, a cell is highlighted in yellow if the prediction rate of U-Net is better than VGG19; similarly, if the rate of U-Net is equal to that of VGG19, the cell is highlighted in gray. The description of the results is as follows:

(1) U-Net and VGG19 both have high F-scores (almost greater than 90%). This implies that both networks show high accuracy in the demolished building detection task.

(2) U-Net yields a better precision rate, while VGG19 yields a better recall rate.

(3) The F-score for U-Net is higher than that for VGG19; in other words, U-Net exhibits lower mis-detection and over-detection rates for demolished building.

(4) VGG19 yields a low F-score (0.426, text in red) for k = 1, and for area ranges between 320-1280 m^2. This can be attributed to the small size of the sample dataset. In summary, the proposed method can detect demolished buildings with an optimal mis-detection and over-detection rate. Moreover, U-Net shows good performance despite the small size of the training data.

4. City-level Experiments

To verify the effectiveness of demolished building change detection method at the city level, we used city F with area greater than 50 km^2, including 60699 buildings with areas ranging between 5–1280 m^2 as test. As shown in Table 6, three experiments were performed using the same test dataset. Each experiment was implemented using U-Net and VGG19.

Building area range (m^2)	Test dataset
5–20	8871
20–80	24955
80–320	24857
320–1280	2016
Total	60699

Table 6. Test dataset for city-level experiments

4.1 Experiment 1

The training datasets were acquired from three cities, with over 250,000 buildings. Because of the skewness in the number of demolished buildings vis-a-vis undemolished buildings, we augmented the number of demolished buildings by randomly rotating demolished buildings. Then, we selected the same number of random undemolished buildings. The training dataset is described in Table 7.

Building area range (m^2)	Training dataset	
	Demolished	Undemolished
5–20	3888	3888
20–80	12060	12060
80–320	5058	5058
320–1280	378	378

Table 7. Training dataset for experiment 1

29th International Cartographic Conference (ICC 2019), 15–20 July 2019, Tokyo, Japan. This contribution underwent single-blind peer review based on submitted abstracts. https://doi.org/10.5194/ica-proc-2-122-2019 | © Authors 2019. CC BY 4.0 License.

4.2 Experiment 2

The model developed in experiment 1 was retrained by using the training dataset shown in Table 8. The training dataset was obtained from the same city of test dataset; however, the aerial images were captured during a different time period.

Building area range (m^2)	Training dataset	
	Demolished	Undemolished
5–20	150	323
20–80	196	753
80–320	137	433
320–1280	2	0

Table 8. Training dataset for retraining the model

4.3 Experiment 3

The model developed in experiment 2 was retrained by using the training dataset in Table 2; however, city F was excluded.

4.4 Results

The test dataset contains a total of 60699 buildings, which includes 1009 demolished buildings and 59690 undemolished buildings. Table 9 shows the rate of mis-detection and over-detection for each experiment. The performance of each model was measured using the same threshold. In experiment 1, the over-detection ratio is very high. However, the results suggest that retraining is very effective in experiment 2, as the over-detection ratio is reduced significantly. Moreover, the over-detection ratio was also decreased effectively in experiment 3.

Experiment	Model	Mis-detection FN/(TP+FN)	Over-detection FP/(FP+TN)
1	VGG19	2.55%	24.02%
	U-Net	0.55%	57.25%
2	VGG19	3.37%	13.54%
	U-Net	2.09%	9.8%
3	VGG19	3.82%	8.33%
	U-Net	2.64%	4.56%

Table 9. Results of the three experiments

Because of skewness in the number demolished buildings (1009) relative to the undemolished building (59690) in the test dataset, we used the receiver operating characteristics (ROC) curve and the area under the curve (AUC) to evaluate the results. The ROC curve is obtained by plotting the false positive rates on the x-axis and the true positive rates on the y-axis for different threshold values. AUC is defined as the area under the ROC curve. A higher AUC indicates better performance.

Figure 8 shows the ROC curve and AUC of U-Net and VGG19. It can be seen that U-Net yields higher AUC than VGG19 in all three experiments. This result indicates that using the same training dataset, U-Net has higher correction rate than VGG19. In addition, more representative training samples can optimize the accuracy of the model for both U-Net and VGG19. For experiment 2, the datasets of same city show remarkable improvements in the AUC value. Therefore, the results suggest that the same city taken at different periods can improve the accuracy of the classification, not only with U-Net, but also with VGG19.

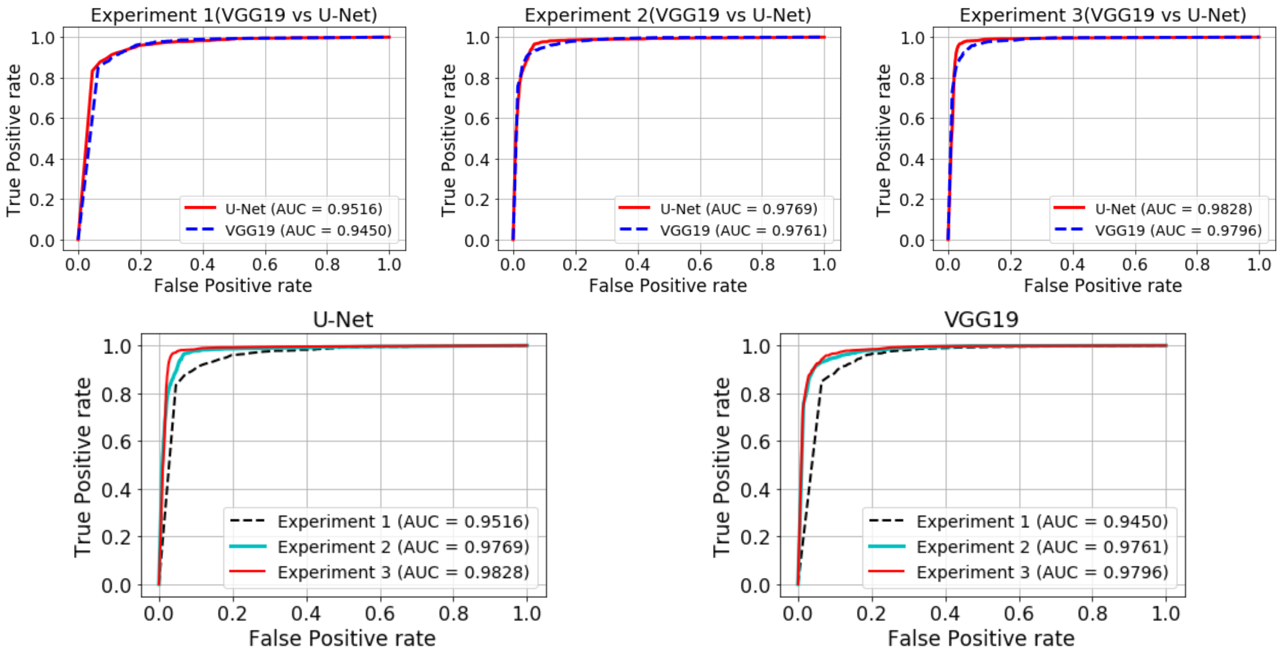

Figure 8. The results of ROC curve and AUC

It should be noted that while it is challenging to extract special types of buildings, such as those where the roofs are characterized by different shadows, color intensity, and partially covered by trees. Figure 9 illustrates three undemolished examples of U-Net in experiment 3, listed with the sample of different shadows, color intensity, and partially covered by trees. The probability of a building being classified as demolished is low (0.17, 0.15, and 0.04). The proposed method can successfully classify such special buildings. The results implied that the performance of the demolished building detection method proposed in this work is satisfactory.

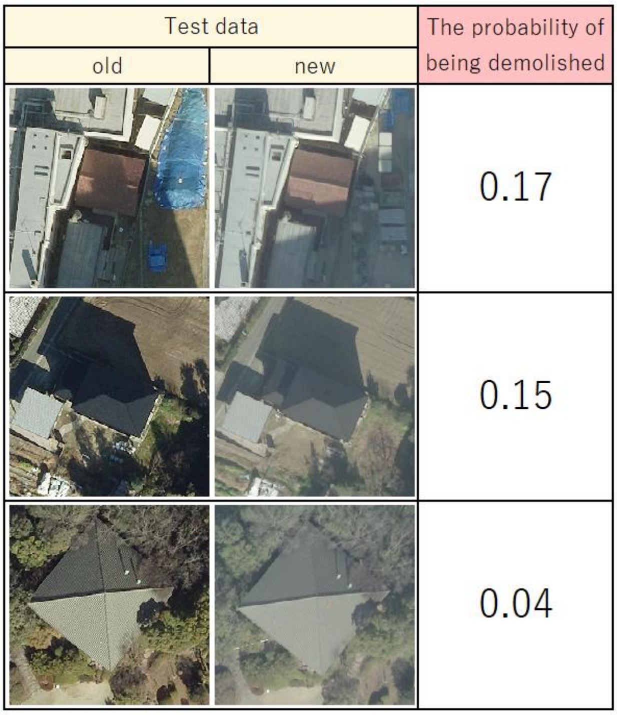

Figure 9. Examples of undemolished buildings

5. Conclusions

In this paper, we proposed a novel demolished building detection method that used bi-temporal aerial images and building polygon data based on the U-Net and VGG19 architectures. The main contributions of our work can be summarized as follows:

(1) Non-buildings changes from real building changes are excluded by using building polygon data. The results also suggest a significant reduction in noise due to non-building changes, such as trees, ground, and roads.

(2) The architectures, which are based on U-Net and VGG19, were implemented for demolished building detection. In terms of classification performance, the results of U-Net were better than that of VGG19, which suggested that U-Net is a useful architecture for image classification problems as well as for semantic segmentation tasks.

(3) In this work, the demolished building detection is based on building-base rather than pixel level, wherein each building in the building polygon data was classified as demolished or undemolished.

(4) The proposed approach accurately identifies demolished buildings, as well as addresses the problems associated with false-detection. The proposed method achieved demolished building detection with low mis-detection (2.64%) and over-detection (4.56%) rates, obtained from an entire urban city with more than 60,000 buildings.

6. References

Bourdis, N., Marraud, D. and Sahbi, H. (2011). Constrained optical flow for aerial image change detection. IGARSS, July 24-29, 2011, Vancouver, Canada

Daudt, R., Saux, B., and Boulch, A. (2018) Fully Convolutional Siamese Networks for Change Detection. ICIP 2018, October 7-10, 2018, Athens, Greece.

Lim, K., Jin, D. and Kim, C. (2018). Change Detection in High Resolution Satellite Images Using an Ensemble of Convolutional Neural Networks. APSIPA ASC 2018, November 12-15, 2018, Honolulu, Hawaii.

Maltezos, E., Ioannidis, C., Doulamis, A. and Doulamis, N. (2018). Building Change Detection using Semantic Segmentation on Analogue Aerial Photos. FIG Congress 2018, May 6-11, 2018, Istanbul, Turkey.

Pang, S., Hu, X., Cai, Z., Gong, J. and Zhang, M. (2018). Building Change Detection from Bi-Temporal Dense-Matching Point Clouds and Aerial Images. Sensors (Basel), May 24, 2018.

Ronneberger, O., Fischer, P. and Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, LNCS, Vol.9351, pp.234-241.

Rottensteiner, F. (2007). Building change detection from Digital Surface Models and multi-spectral images. Photogrammetric Image Analysis, September 19-21, 2007, Munich, Germany.

Simonyan, K. and Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. International Conference on Learning Representations (ICLR) 2015, May 7-9, 2015, San Diego,USA.http://www.robots.ox.ac.uk/~vgg/research/very_deep, 2019.

Tian, J., Cui, S., and Reinartz, P. (2014). Building change detection based on satellite stereo imagery and digital surface models. IEEE Transactions on Geoscience and Remote Sensing 52(1), pp. 406-417.

https://www.gdal.org/