


Article

Newly Built Construction Detection in SAR Images Using Deep Learning

Raveerat Jaturapitpornchai ^{1,*}, Masashi Matsuoka ¹ , Naruo Kanemoto ², Shigeki Kuzuoka ³,
Riho Ito ² and Ryosuke Nakamura ²

¹ Department of Architecture and Building Engineering, Tokyo Institute of Technology, Yokohama 226-8502, Japan; matsuoka.m.ab@m.titech.ac.jp

² National Institute of Advanced Industrial Science and Technology, Tokyo 135-0064, Japan; naruo.kanemoto@aist.go.jp (N.K.); pscrh.itou@aist.go.jp (R.I.); r.nakamura@aist.go.jp (R.N.)

³ Space Shift, Tokyo 105-0013, Japan; kuzuoka@spcsft.com

* Correspondence: jaturapitpornchai.aa@m.titech.ac.jp; Tel.: +81-45-924-5605

Received: 7 May 2019; Accepted: 17 June 2019; Published: 18 June 2019



Abstract: Remote sensing data can be utilized to help developing countries monitor the use of land. However, the problem of constant cloud coverage prevents us from taking full advantage of satellite optical images. Therefore, we instead opt to use data from synthetic-aperture radar (SAR), which can capture images of the Earth's surface regardless of the weather conditions. In this study, we use SAR data to identify newly built constructions. Most studies on change detection tend to detect all of the changes that have a similar temporal change characteristic occurring on two occasions, while we want to identify only the constructions and avoid detecting other changes such as the seasonal change of vegetation. To do so, we study various deep learning network techniques and have decided to propose the fully convolutional network with a skip connection. We train this network with pairs of SAR data acquired on two different occasions from Bangkok and the ground truth, which we manually create from optical images available from Google Earth for all of the SAR pairs. Experiments to assign the most suitable patch size, loss weighting, and epoch number to the network are discussed in this paper. The trained model can be used to generate a binary map that indicates the position of these newly built constructions precisely with the Bangkok dataset, as well as with the Hanoi and Xiamen datasets with acceptable results. The proposed model can even be used with SAR images of the same specific satellite from another orbit direction and still give promising results.

Keywords: satellite imagery; SAR; deep learning; U-net; urban change

1. Introduction

In developing countries, the high demand for the construction of new residential and business areas is common. Monitoring new construction is necessary in order to predict the expansion of cities in both political and economic terms. A commonly used approach for this type of observation is the use of remote sensing data from optical sensors, because such data allow the easy creation of maps. Despite the excellence of optical data, some developing countries are located in tropical areas where clouds cover parts of the area all year round. Unfortunately, optical sensors cannot capture Earth's surface below these clouds. Because synthetic-aperture radar (SAR) captures images using microwave signals that can penetrate clouds, the use of SAR data is a secondary option to handle the problem. However, the difficulty in the interpretation of SAR images makes it harder to identify locations of new constructions. In the literature, many methods have been proposed for SAR image change detection with threshold method and clustering methods [1–5]. Many of these publications have to generate a difference image from the pixel information of two SAR images, from which it is difficult to

identify one specific change, such as the appearance of new buildings, as any kind of change similar to the target change would be involved in the results. For instance, Y. Ban and O. Yousif [6] used the threshold-based method on a difference image in detecting urban change. Despite the good detection result, there is a possibility to detect falsely when the urban or non-urban area has unordinary intensity change behavior. In this research, our target is to monitor the increase in new construction in urban and suburban areas. To complete this objective, we used a deep learning technique to identify these newly built constructions from two SAR images directly without generating a difference image. The goal of deep learning is to simulate the experience-based learning mechanism of the human brain using training data and ground truth data in the same way that humans learn [7]. To date, deep learning has been highly effective, especially in the image processing field. One of the most successful deep learning networks that we considered using in this work is the U-net [8]. The U-net, proposed in 2016 for the purpose of medical image segmentation, was built on the basis of adding a skip connection to the fully convolutional network (FCN) [9] between the encoder part and decoder part. With the skip connection, the decoders can receive a low-level feature from the encoder and form the output without losing boundary information in the process. Because of its precisely predicted output at the boundary part of an image, it is now one of the most cited papers in the deep learning field. In our case, it is extremely important to preserve the boundary information because SAR data do not provide very clear information; this is because the observation mechanism of SAR is completely different from those of other sensors.

There are numerous publications involving high-precision remote sensing and deep learning [10–14], but they mostly involve optical imagery, whose data are clear and contain information that is similar to that in ordinary RGB images. Despite the frequent use of the RGB image and the good quality of its data resolution, other methods were explored for our study. The U-net has performed well in the extraction of buildings using very high resolution satellite imagery [15], which produces a very accurate result even at the building boundary area. Publications of studies that used deep learning with SAR images [16–19] also report excellent change detection results and prove that deep learning can be used with SAR images. S. Iino et al. [20] successfully used a convolutional neural network with an SAR image for land cover classification to find an urban distribution map for short-term change detection. However, their results included all of the changes that occurred on two occasions, regardless of the source of the changes, because they used only the information of the difference in intensities or the digital surface model. In a real-life application of change detection, we usually want to see only the changes of interest while ignoring all others; thus, instead of detecting all of the changes that occurred, we should only detect changes in a specific target, for example, detect changes in buildings while ignoring paddy field seasonal changes, as we are doing in this work. As most of the existing methods use only the differences in SAR intensities and the ground truth of all changes detected, they are not able to satisfy the objective of detecting only the changes related to buildings, even with the use of deep learning. To do so, a different approach to training a deep learning network to detect newly built constructions needs to be created. To this end, we aim to employ the U-net architecture to identify the location of newly built constructions in SAR images; this was implemented by training the network with SAR images from two different occasions and then guiding the network to determine which changes are from construction through the corresponding ground truth of building changes. While the results of difference image-based methods can contain such unwanted objects when similar intensity change behavior appears, our network can learn the change of the constructions and other areas by not just using the change of intensity, but also including visual features of constructions and non-constructions objects as well, which makes it able to tell the difference between the change from newly built construction and other kind of changes.

2. Dataset

2.1. SAR Data Description

The SAR data we used in this research are from ALOS-PALSAR in HH polarization with a resolution of 15 m/pixel. The images in the dataset were captured in ascending orbit mode at different times between 2008 and 2010. All SAR images were acquired in the right-looking direction with an off-nadir angle of 34.3° . The dataset includes three study areas: Bangkok, Thailand; Hanoi, Vietnam; and Xiamen, China. The images of the Bangkok area were taken at five different times: 1 January 2008, 27 November 2008, 12 January 2009, 21 November 2009, and 15 January 2010. The images of Hanoi and Xiamen were taken at two different times; the Hanoi images were taken on 2 February 2007 and 13 February 2011 and the Xiamen images were taken on 22 February 2007 and 2 November 2010. Although a variety of polarizations can be chosen, we selected HH polarization, as it is the most suitable for building detection because the double bounce effect of the building is clearest in HH polarization images. The unit of the backscatter (intensity) of the dataset is dB.

2.2. Ground Truth Preparation

We created ground truth data that correspond to our SAR data. The process of creating the ground truth was entirely manual and done by the authors. All of the ground truths were created by drawing polygons (red objects in Figure 1) directly onto the optical images (examples shown in Figure 1c,d) available in Google Earth software after comparing the images of the same location from two different times. The criteria used for selecting the date of the optical images corresponding to Time 1 and Time 2 of the SAR data is that the date must be as close as possible to the SAR data, while Time 1 of optical data must not exceed Time 1 of SAR data, and the Time 2 of optical data must not be before Time 2 of SAR data. Because the boundaries of our ground truths are large, the dates of the optical images we picked from Google Earth vary depending on the area within the ground truth boundary, the lack of optical information, and the cloud cover problem; for example, the dates for the optical data selected for Time 1 of the SAR pair 1 January 2008/12 January 2009 are 18 December 2004 and 10 February 2005; for Time 2, the dates 18 December 2009, 11 April 2010, and 15 April 2010 were selected. Please note that we only selected buildings with a size of more than 45×45 m (approximately 2025 m²) in the optical image. The number of polygons for each created ground truth is shown in Table 1.

Table 1. Acquisition information of dataset. SAR—synthetic-aperture radar.

Purpose	Location	Acquisition Date of SAR Images (Time 1–Time 2)	Number of Polygons in Ground Truth
Training	Bangkok, Thailand	1 January 2008–15 January 2010	164
		12 January 2009–15 January 2010	68
		1 January 2008–12 January 2009	38
Testing	Bangkok, Thailand	27 November 2008–15 January 2010	12
		12 January 2009–21 November 2009	16
	Hanoi, Vietnam	2 February 2007–13 February 2011	108
	Xiamen, China	22 January 2007–2 November 2010	68

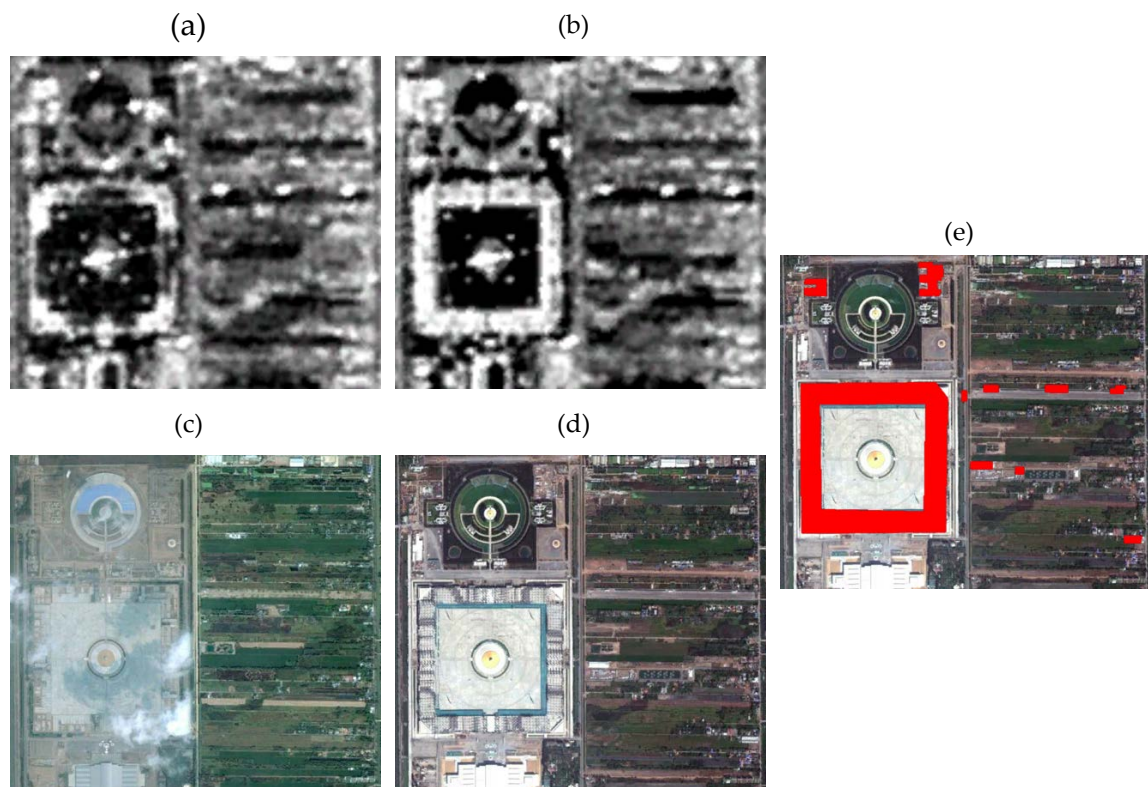


Figure 1. Examples of synthetic-aperture radar (SAR) dataset and ground truth data (red polygons) overlaid on an optical image from Google Earth of a temple construction in Bangkok: (a) SAR image from 12 January 2009, (b) SAR image from 21 November 2009, (c) optical image from 10 February 2005, (d) optical image from 18 December 2009, (e) created ground truth.

2.3. Training Data Preparation

Before any further action, we first reduced the speckle noise in the entire dataset using the Lee filter [21] with a filter size of 3×3 to prevent potential errors due to noisy values from occurring during the training process. We then normalized the intensity value of the data to a range of $[-1, 1]$ to facilitate network training by avoiding inconsistent SAR intensities. To enable identification of the positions of new constructions that were built between two different times, we selected data from the dataset acquired on different dates with the same data acquisition conditions and geolocations. We then matched the selected data to form a pair of Time 1 and Time 2 SAR images. The images from Time 1 and Time 2 and their corresponding ground truth were then stacked and prepared for cutting into small patches for training the network. To cut the SAR images taken at two different times and the corresponding ground truth to use in network training (as Time 1, Time 2, and the ground truth) for loss calculation, we used a sliding window with a sliding step of 50 pixels along the images to cut them into patches. Fifty was deemed the most suitable number of pixels for the sliding step because it results in a patch that is cut without skipping buildings, but is also not too repetitive. Only the patches containing at least one polygon according to the corresponding ground truth were selected for use in the training process. As a result, we had 2028 pairs after discarding patches that contained only negative pixels (please note that 10 percent of the patches from 2028 pairs were randomly selected for the validation of the model at the end of each training epoch). The patches with only negative pixels were removed because we want the network to learn from positive samples so that it can locate the construction of a building; also, we are likely to maintain a balance between positive and negative data during training as patches containing positive pixels also contain negative pixels. The patches cut for training the network were 256×256 pixels, which is a size that is suitable for detecting a building, as shown in Figure 2, as it has the appropriate proportion of positive and negative pixels.

The details are discussed in Section 4.2, in which we compare the accuracies resulting from using a patch size of 128×128 and 256×256 . Time 1 and Time 2 of each paired dataset are shown in Table 1. Another thing to note is that the areas used for testing purposes in Bangkok, Hanoi, and Xiamen were manually selected at 400×400 pixels, which differs from the training data and was chosen for the ease of inspection.

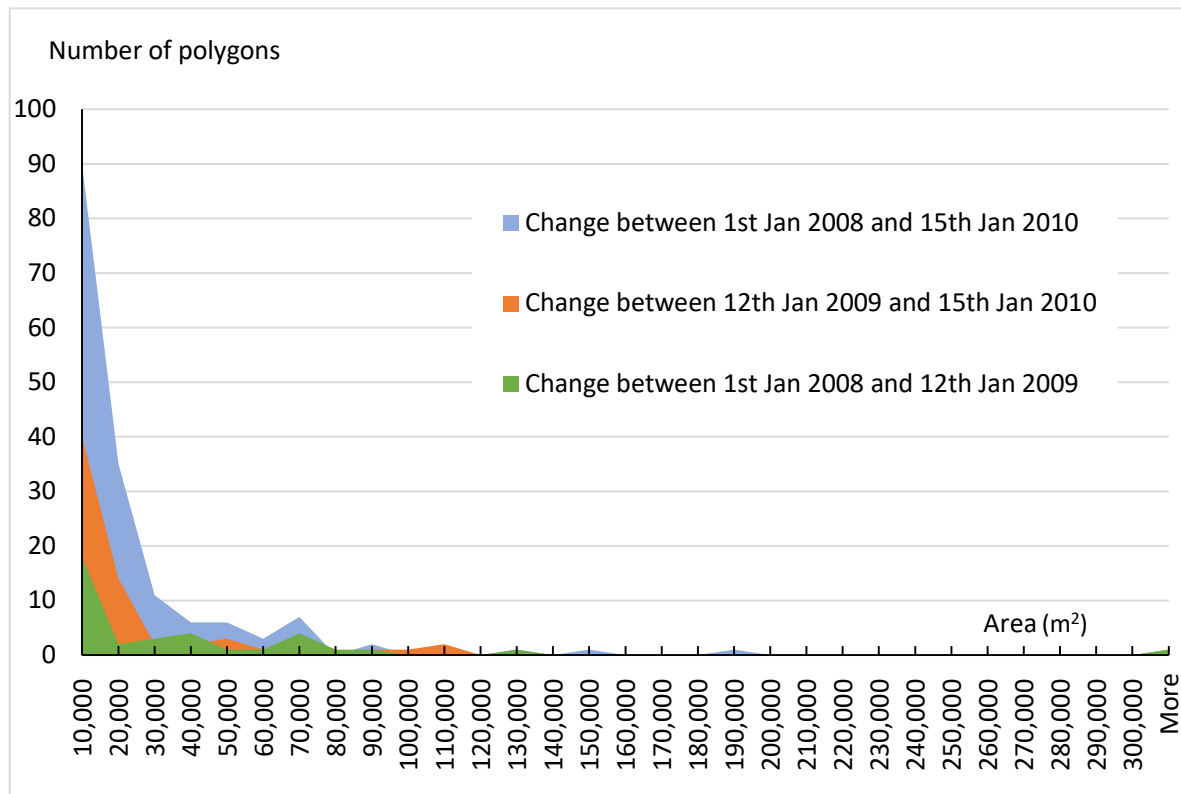


Figure 2. The histogram of change area in training data.

3. Network Description

Because the U-net is an FCN model, in order to express the reasons for its selection among all other models, the background of some other FCNs needs to be explained first.

The FCN is an architecture built only upon locally connected layers, such as the convolution, pooling, and upsampling layers. The network is usually divided into an encoder part and a decoder part. The encoder is responsible for gathering the information or features of objects in an input image, while the decoder is for recovering spatial information. One of the best examples of FCN architecture is SegNet [22], which was proposed for the semantic segmentation of an RGB image. The architecture consists of the same number of encoders and decoders, and each encoder applies convolution, batch normalization, ReLU, and max pooling to downsample the result. The decoder carries out almost the same procedure as the encoder, but without a ReLU step and with upsampling instead of downsampling. The output of the last decoder is then subjected to the Softmax function to generate the segmentation prediction result.

The architecture of the U-net is very similar to that of SegNet, but with an additional skip connection between each corresponding encoder and decoder. The skip connection makes a huge difference. Without a skip connection, the output prediction result lacks sharpness around the boundary areas, which is especially crucial for the SAR images in our case. Although comparing our result from the U-net with that from SegNet would be informative, it is impossible to generate the output using SegNet because features are too blurry to be identified. The result of using SegNet indicates that the

skip connection is very important when dealing with images without significant sharpness, as is the case for our dataset.

Besides SegNet, any other FCN that involves a deconvolution layer [23] as an upsampling layer cannot generate a suitable result, as the checkerboard phenomenon will occur [24]. Therefore, it would be difficult to compare the results of these FCNs with those of the proposed network.

Our network is shown in Figure 3. Each encoder block consists of a convolution–BatchNorm–ReLU layer. The values of the number of channels, spatial filter size, and stride size of the convolution filters in each step are shown in Figure 4. As our modules are in the form of convolution–BatchNorm–ReLU [25], it is noted that the first layer in the encoder does not apply BatchNorm. As we followed the method applied by Isola et al. [26], in the encoder, all ReLU functions are leaky with a slope of 0.2, while the ReLU functions in the decoder are not leaky. The dropout rate is 0.5. Our skip connections in the U-net architecture were placed to concatenate activations between each layer i in the encoder and layer $n - i$ in the decoder, where n is the total number of layers. The concatenation leads to a change in the number of channels in the decoder. At the last layer in the decoder, a convolution function is applied to map the output, followed by a sigmoid function.

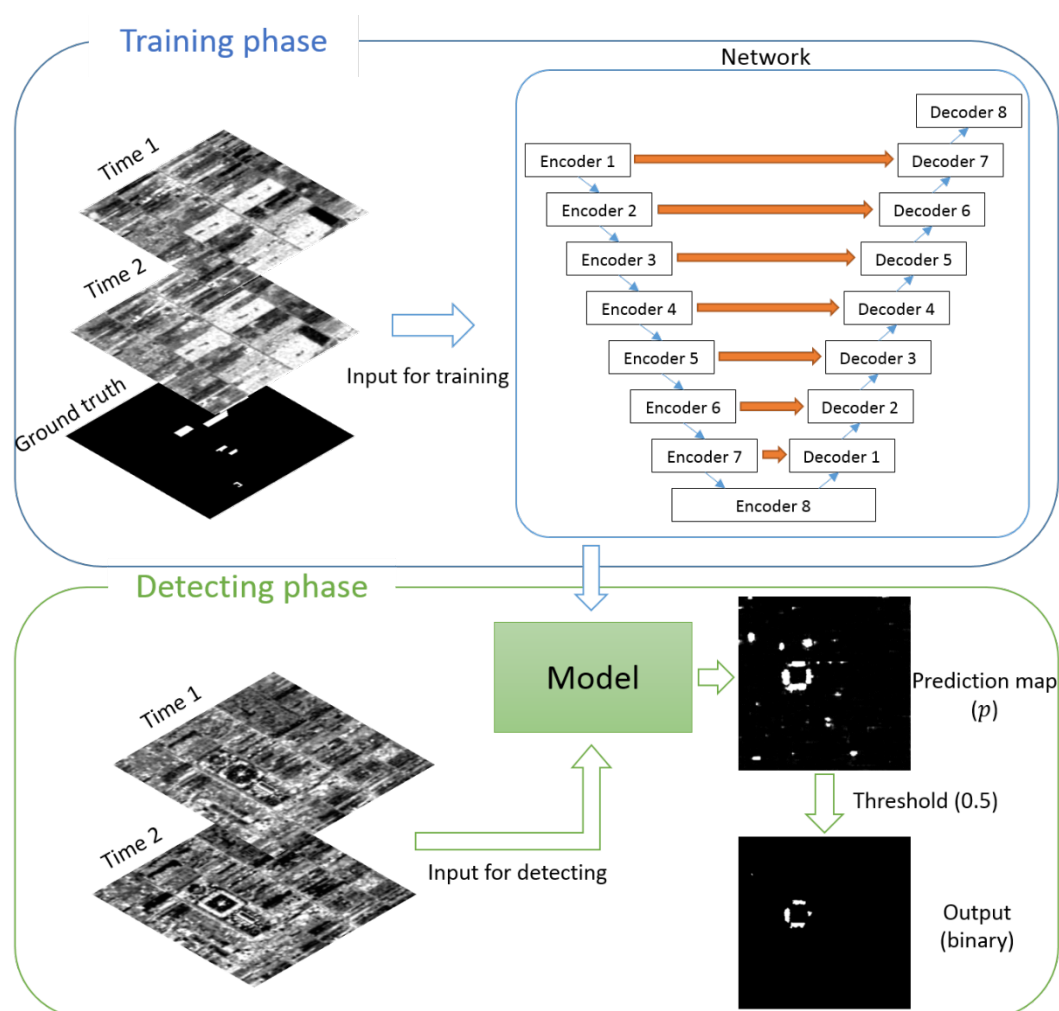


Figure 3. Process of detecting newly built constructions.

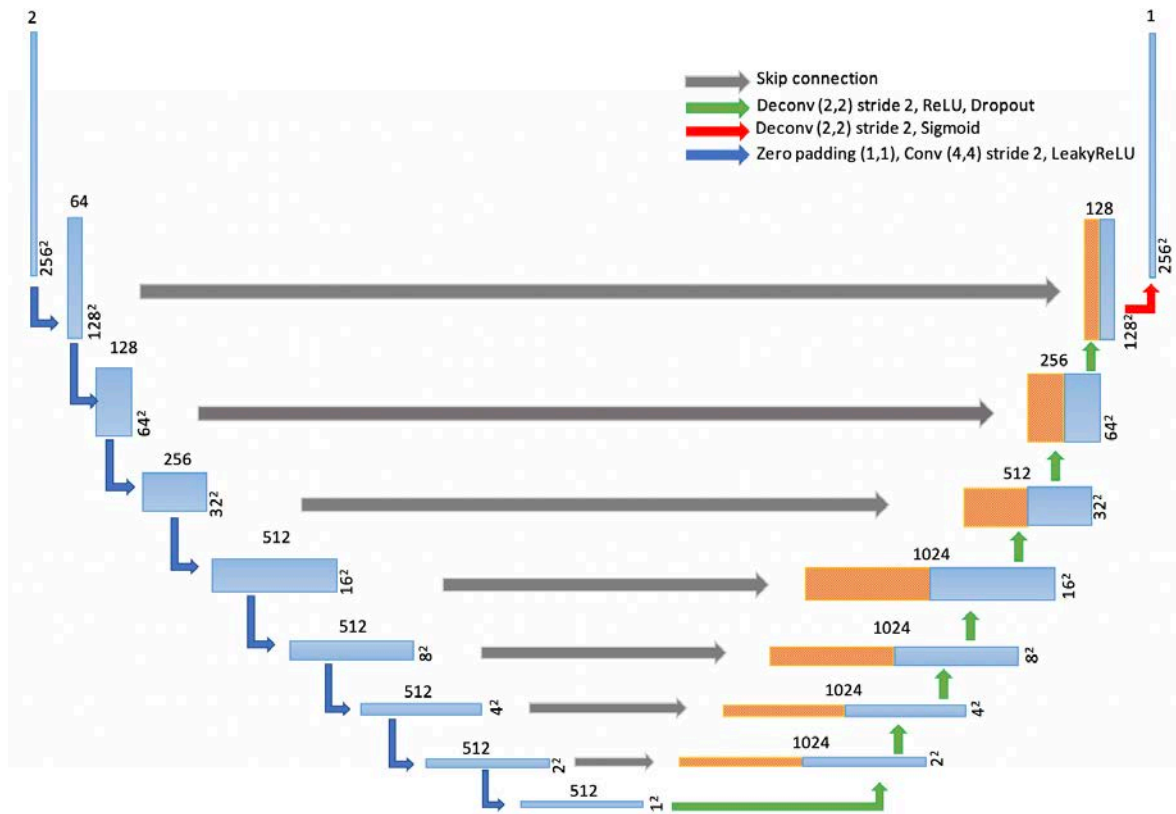


Figure 4. Detail of network architecture.

Although we have datasets for three cities, we chose to train our network with the Bangkok dataset. From the images at Time 1 and Time 2, each 256×256 -pixel patch is concatenated and fed to the network along with the ground truth. The U-net returns the change detection result, which is used to calculate the loss for comparison with the corresponding ground truth. In the loss calculations, the loss function L in our method is the cross-entropy, which normally can be calculated as

$$L = - \sum_{c=1}^M y_{o,c} \log(p_{o,c}), \quad (1)$$

where M is the number of classes, y is the binary indicator (0 or 1) that represents whether class label c is the correct classification for observation o , and p is the predicted probability that observation o is of class c . However, in our case, M equals 2 because it is a binary classification (changed or unchanged). Thus, L from (1) can be derived as

$$L = -(y \log(p) + (1 - y) \log(1 - p)). \quad (2)$$

However, because there are far fewer positive pixels than there are negative pixels, we considered applying class weight balancing to the loss function in order to prevent the network from excessive activations for negative parts and never for positive parts. The weighted loss function has proven its efficiency in handling imbalance class dataset [27], which is also applicable to our case. As a result, the calculation of the loss function becomes

$$L = -(y \log(p)(\omega_p) + (1 - y) \log(1 - p)), \quad (3)$$

where

$$\omega_p = \frac{\text{percentage of negative pixels in training set}}{\text{percentage of positive pixels in training set}}. \quad (4)$$

This weight ω_p makes the network focus equally on how changes happened in positive areas and in negative areas. Although the negative area should be given a higher priority in training because, in most cases, the majority of the area is negative, we want the model to be applicable to any situation regardless of the ratio of positive to negative area, so we decided to use weights that result in the network learning both classes to an equal extent.

The value of ω_p can vary depending on the dataset used to train the network. In our case, the value is 181.5, which is the result of the rate of white pixels (new construction areas) = 0.548% and the rate of black pixels (non-changed areas) = 99.452%. We did not use the ratio from ground truths corresponding to the whole SAR image because it contains too many black pixels in patches that were discarded (i.e., negative patches) and thus excluded from the network training process; thus, the ground truth ratio would not match the ratio received by the network from the training set.

As the selection of parameters can affect the model efficiency, we determined the best parameters by observing training loss and testing loss during the network training. As a result, the number of epochs we used in this work is 10 with the batch size of 16. The model was trained with an Adam optimizer at a learning rate of 0.001. We found that using an epoch of around 10 lowers the training loss while keeping the testing loss stable; this was determined by observing that the testing loss of the network started to fluctuate around the 10th epoch and the loss increased afterward. In addition, using more than 10 epochs may cause overfitting and is also time-consuming, as there is no significant difference between using 10 epochs and more than 10 epochs.

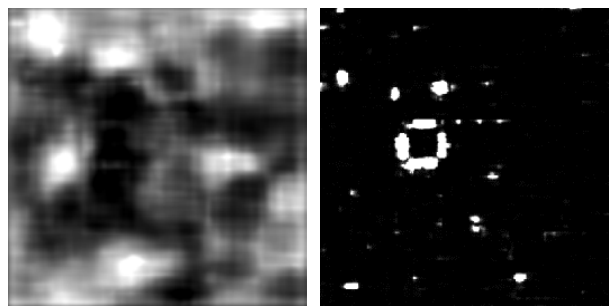
4. Experimental Results

In this section, the experiments for evaluating the impact of weighting loss compared with that of not weighting are described. The experiment for evaluating different patch sizes is also described in this section.

To evaluate the results, we used the Bangkok testing dataset from the two date pairs shown in Table 1. The accuracy in this research is calculated in the form of overall accuracy, precision, recall, F measure, F1 measure, Kappa, intersect over union (IOU), false negative (FN) rate, and false positive (FP) rate. The false negative rate is obtained by the number of pixels that are in the ground truth, but not in our predicted result multiplied by 100 and then divided by the total number of positive pixels in the ground truth; the false positive rate is the number of pixels that are not in the ground truth, but are in our predicted result multiplied by 100 and then divided by the total number of negative pixels in the ground truth. The calculation of each validation method, excluding the false negative and false positive rates, is shown in Table 2. The TP in the Table 2 stands for true positive while TN stands for true negative. Please note that the β value of our F measure is 0.3. The result of the proposed network is compared with the results using fuzzy c-means (FCM) clustering [28] and Otsu thresholding [29]. Some other experiments that we conducted are also shown in this section. The FCNs are not used in the comparison in this section because they cannot generate a decent detection result, as shown in Figure 5. Please note that in Figure 5, the range of prediction value of SegNet is $[0.39 \times 10^{-2}, 1.01 \times 10^{-2}]$, while that of our proposed network (based on the U-net) is $[8.39 \times 10^{-6}, 0.99]$. As the prediction range of SegNet is very small, it is difficult to generate the binary output map as the proper threshold value cannot be obtained.

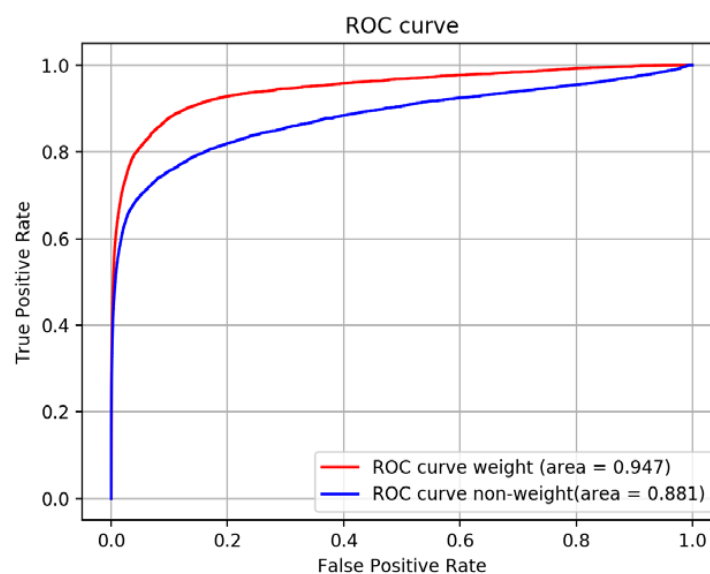
Table 2. The calculation of each validation method. IOU—intersect over union.

Validation Method	Calculation
Overall accuracy	$Overall\ accuracy = \frac{TP+TN}{TP+TN+FP+FN}$
Precision	$Precision = \frac{TP}{TP+FP}$
Recall	$Recall = \frac{TP}{TP+FN}$
F measure	$F_{\beta} = (1 + \beta^2) \cdot \frac{precision \cdot recall}{(\beta^2 \cdot precision) + recall}$
F1 measure	$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$
Kappa	$Kappa = \frac{Observed\ agreement - chance\ agreement}{1 - chance\ agreement}$
IOU	$IoU = \frac{target \cap prediction}{target \cup prediction}$

**Figure 5.** Comparison between prediction map of using SegNet (left) and proposed network (right).

4.1. Comparison between Weighted and Non-Weighted Loss Function

While the loss function of the network used in this research is weighted, an experiment to compare the weighted and non-weighted loss function was conducted to show the benefit of weight balancing. The advantage of weight balancing is illustrated in Figure 6, in which the area under the curve (AUC) of the receiver operating characteristic curve (ROC curve) of the weighted loss is higher than that of the non-weighted loss. The ROC curve is a graph of the true positive rate plotted against the false positive rate, and the closer the AUC is to 1, the higher the model's efficiency in separating classes.

**Figure 6.** The receiver operating characteristic (ROC) curve of the weighted loss model compared with that of the non-weighted loss model.

As we previously stated with regard to the unbalanced class weight, we noticed that our model can generate results with a decent accuracy without involving a positive weight in the loss function, but ultimately, using the weighted loss still generates a better result, as shown in Table 3.

Table 3. The accuracy of non-weighted loss compared with that of weighted loss for the Bangkok testing site.

Validation Method	Non-Weighted Loss	Weighted Loss ($\omega_p = 181.5$)
False negative	84.5628	55.9093
False positive	0.0090	0.2238
Overall accuracy	99.14%	99.22%
Precision	0.9458	0.6669
Recall	0.1544	0.4409
F measure	0.6645	0.6398
F1 measure	0.2654	0.5308
Kappa	0.2633	0.5270
IOU	0.1530	0.3613

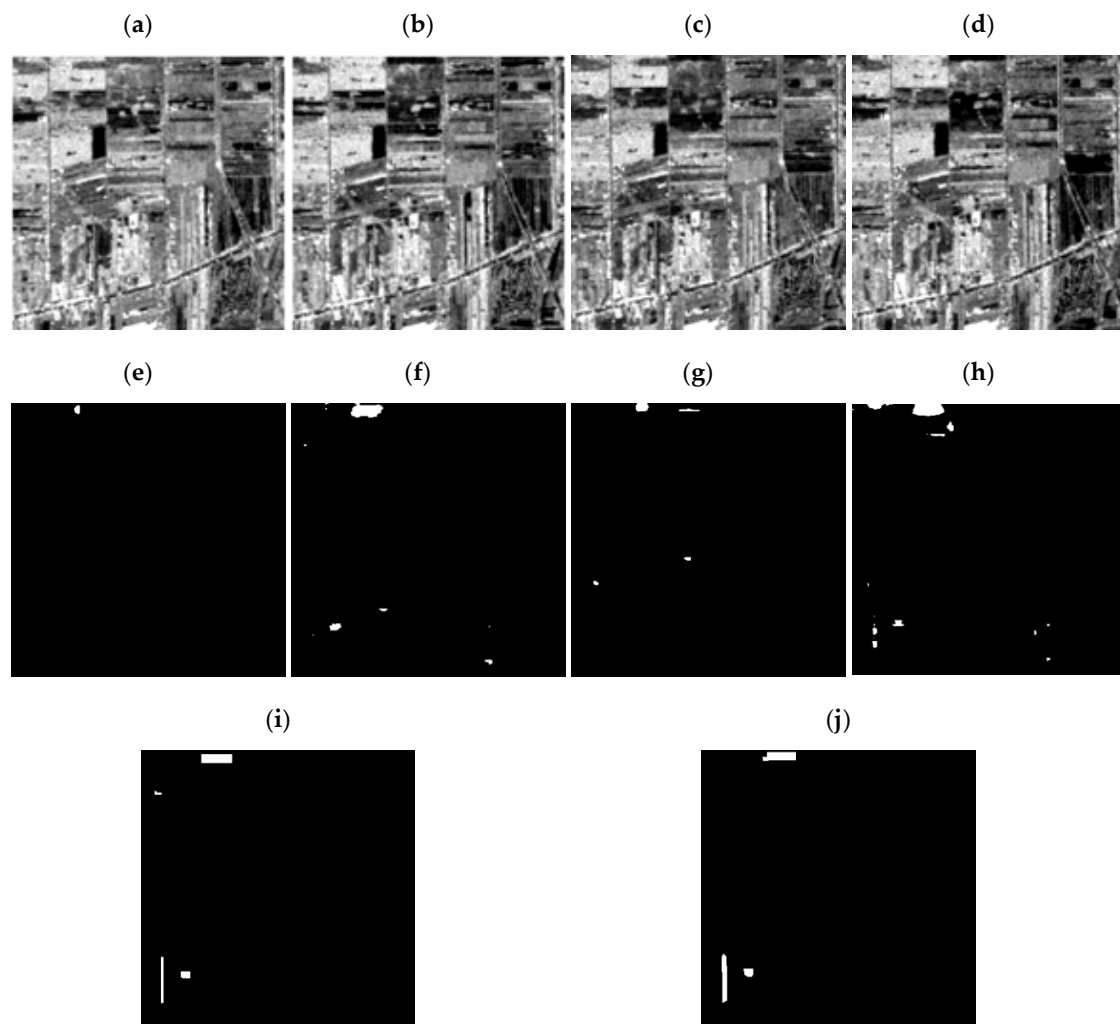
4.2. Comparison between Different Patch Sizes

As we stated in Section 2.3, we aimed to use a patch size that yields the most suitable ratio between black pixels and white pixels that enables the network to learn the positions of building constructions from positive samples; in this research, we used a patch size of 256×256 pixels. Before we selected this size, we conducted an experiment to test which patch size— 128×128 or 256×256 pixels—results in better accuracies. The results are shown in Figures 7 and 8, in which the threshold applied to the prediction maps is 0.5, the ω_p of the 128×128 patch size is 80, and the ω_p of the 256×256 patch size is 181.5. $\omega_p = 80$ is obtained when the proportion of white pixels = 1.23% and the proportion of black pixels = 98.77%, while $\omega_p = 181.5$ is obtained when the proportion of white pixels = 0.55% and the proportion of black pixels = 99.45%. The process of applying ω_p to our network is described in Section 3.

The results in Table 4 indicate that using a 256×256 patch size leads to better accuracies. As a result, we decided to use a patch size of 256×256 in all experiments. A 256×256 patch size results in better accuracies because a 128×128 patch size is too small, causing the loss of features of some parts, such as paddy fields, and leading to the network's inability to fully learn the change pattern of these areas. Even though the negative part is not the focus of this study, it is indispensable for network training, which can be more recognizable in a 256×256 patch size. Moreover, because the sliding step in patches cutting is smaller in a 128×128 patch size, the cut patches would have too many repetitive patterns of both positive and negative features, which can cause the model to be overfitted at an early stage.

Table 4. The accuracy of a 128×128 patch size compared with a 256×256 patch size on the Bangkok testing site.

Validation Method	128×128 Patch Size	256×256 Patch Size
False negative	94.3625	55.8006
False positive	0.0680	0.4033
Overall accuracy	98.98%	99.04%
Precision	0.4572	0.5269
Recall	0.0564	0.4420
F measure	0.2881	0.5187
F1 measure	0.1004	0.4807
Kappa	0.0984	0.4759
IOU	0.0528	0.3164



SAR pair: 27 November 2008/15 January 2010

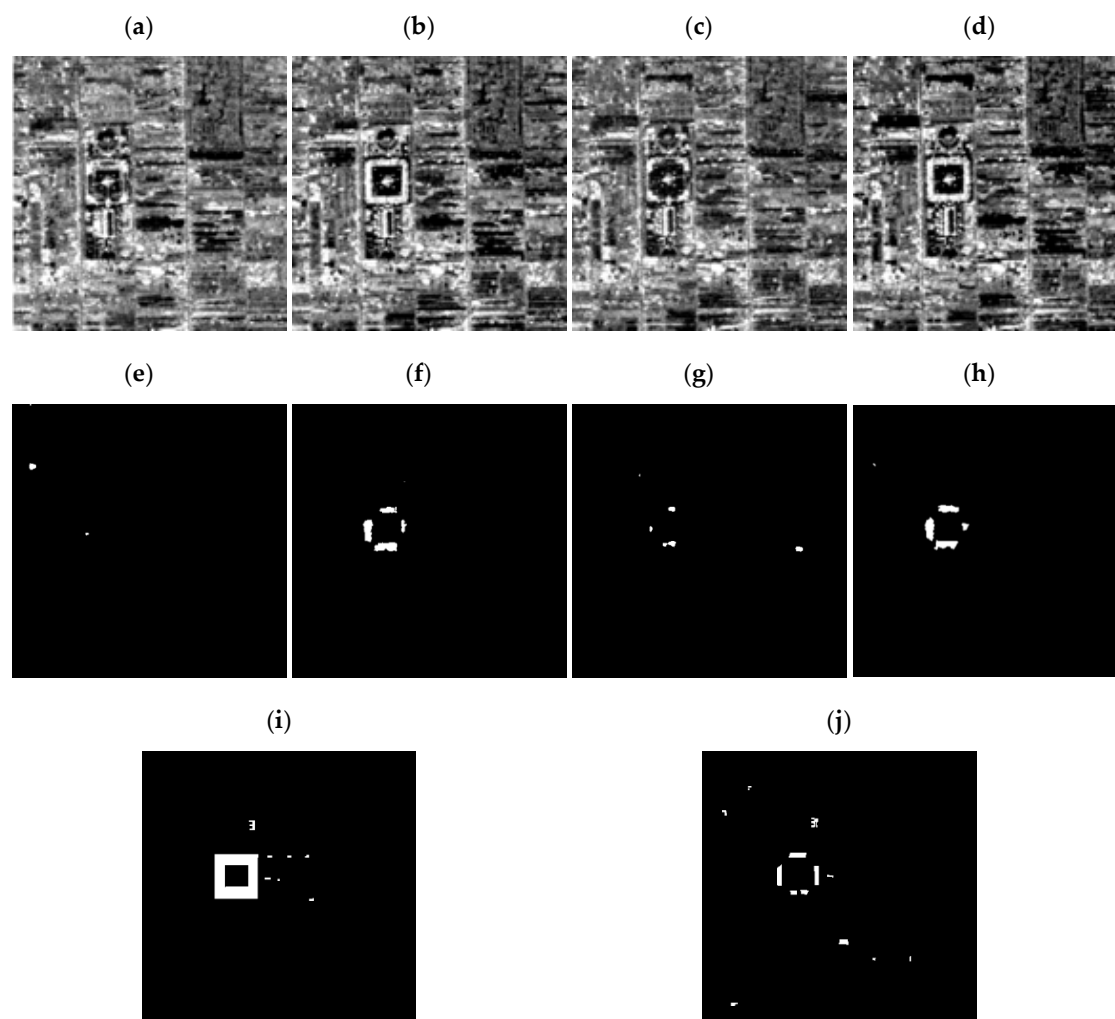
SAR pair: 12 January 2009/21 November 2009

Figure 7. Comparison between each patch size in the first area of the Bangkok site. The resolution of each image is $6 \text{ km} \times 6 \text{ km}$ (for SAR pairs 27 November 2008/15 January 2010 and 12 January 2009/21 November 2009, respectively: (a,c) Time 1 SAR image, (b,d) Time 2 SAR image, (e,g) result of 128×128 patch size, (f,h) result of 256×256 patch size, (i,j) ground truth).

4.3. Result of Bangkok Testing Site

The Bangkok test site includes two date pairs: 27 November 2008/15 January 2010 and 12 January 2009/21 November 2009. The results of the model shown in this section are the binary maps obtained from prediction maps with a threshold of 0.5. The model used in this experiment was trained with a weight of $\omega_p = 181.5$. The results of our model are shown for two different areas in Figures 9 and 10.

From the results of the first test area, in which paddy fields account for the majority of the area, the model can predict the construction of buildings while avoiding the changes in paddy fields caused by seasonal effects. On the other hand, while both FCM and Otsu can capture most of the building changes, they fail to ignore the changes in other parts; this is especially the case for Otsu, which is very sensitive to intensity changes, resulting in about half of the image being detected as a building change.

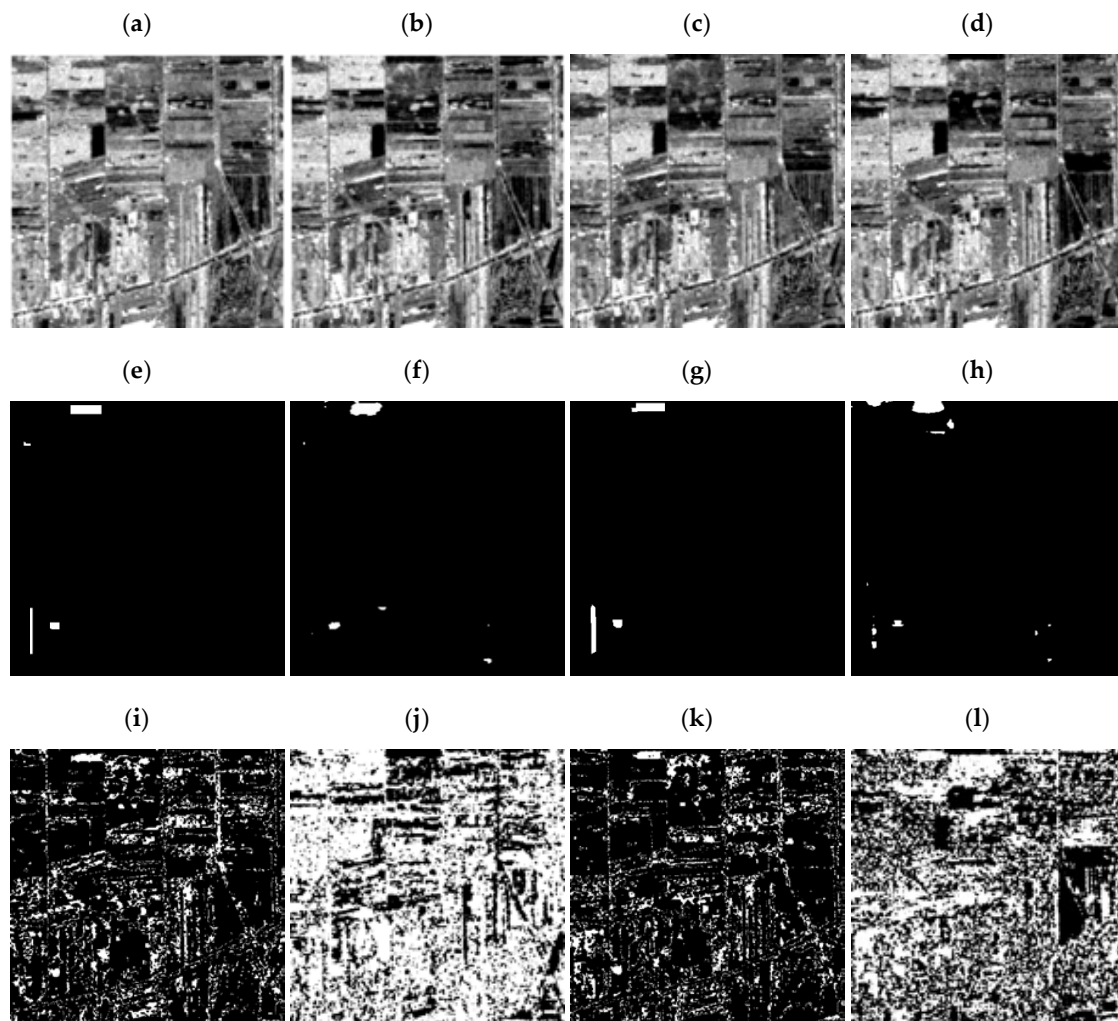


SAR pair: 27 November 2008/15 January 2010

SAR pair: 12 January 2009/21 November 2009

Figure 8. Comparison between each patch size in the second area of the Bangkok site. The resolution of each image is $6 \text{ km} \times 6 \text{ km}$ (for SAR pairs 27 November 2008/15 January 2010 and 12 January 2009/21 November 2009, respectively: (a,c) Time 1 SAR image, (b,d) Time 2 SAR image, (e,g) result of 128×128 patch size, (f,h) result of 256×256 patch size, (i,j) ground truth).

Similar to the first area, the changes in the second area in paddy fields are ignored, while the construction of the temple (the big square object in Figure 10e) and surrounding constructions are detected. The construction of the temple starts with the appearance of four corners of the square as construction preparation tools in Figure 8a, followed by the development of the construction site in Figure 8c, and then the complete construction in Figure 10b,d. The results from the FCM and Otsu methods are similar to those for the first area—they fail to detect only the building changes.

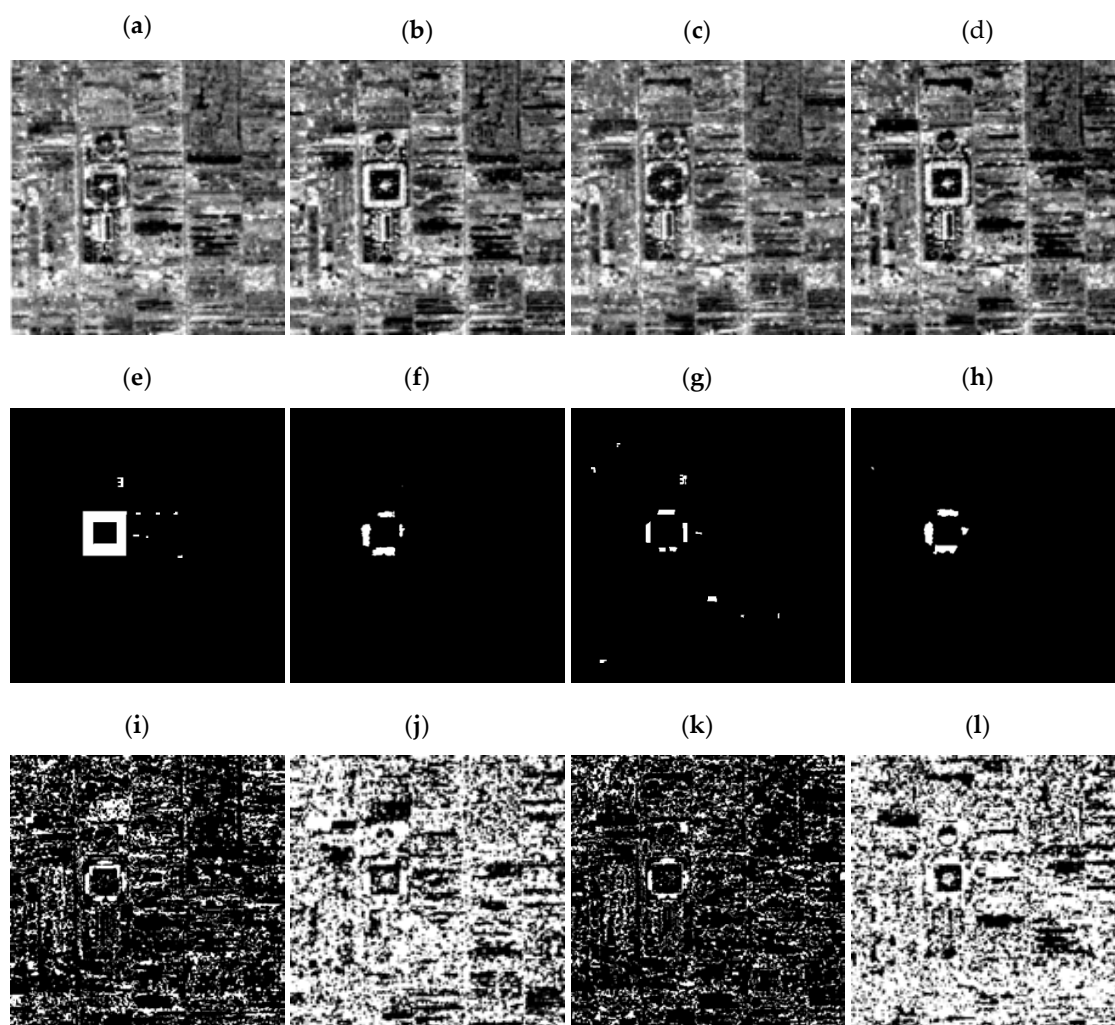


SAR pair: 27 November 2008/15 January 2010 SAR pair: 12 January 2009/21 November 2009

Figure 9. Results of the Bangkok site in the first area. The resolution of each image is $6 \text{ km} \times 6 \text{ km}$ (for SAR pairs 27 November 2008/15 January 2010 and 12 January 2009/21 November 2009, respectively: (a,c) Time 1 SAR image, (b,d) Time 2 SAR image, (e,g) ground truth, (f,h) proposed result, (i,k) result of fuzzy c-means (FCM), (j,l) result of Otsu).

The accuracy of each method for the Bangkok test site is shown in Table 5.

Despite the very high overall accuracy, our model has quite a high false negative rate, which means that it detects buildings as being smaller or in the wrong shape compared with those in the ground truth. However, the low false positive rate means that our model has a very low chance of detecting other types of changes as a building change, and this is the target of our research. Other accuracies are not very high, but they are all at an acceptable level, especially when compared with the FCM and Otsu methods.



SAR pair: 27 November 2008/15 January 2010 SAR pair: 12 January 2009/21 November 2009

Figure 10. Results of the Bangkok site in the second area. The resolution of each image is $6 \text{ km} \times 6 \text{ km}$ (for SAR pairs 27 November 2008/15 January 2010 and 12 January 2009/21 November 2009, respectively: (a,c) Time 1 SAR image, (b,d) Time 2 SAR image, (e,g) ground truth, (f,h) proposed result, (i,k) result of FCM, (j,l) result of Otsu).

Table 5. Accuracy of each model in the Bangkok area. FCM—fuzzy c-means.

Validation Method	Proposed Network	FCM	Otsu's Threshold
False negative	55.8006	51.4676	21.8357
False positive	0.4033	14.8646	58.3693
Overall accuracy	99.04%	84.77%	42.00%
Precision	0.5269	0.0321	0.0134
Recall	0.4420	0.4853	0.7816
F measure	0.5187	0.0348	0.0146
F1 measure	0.4807	0.0602	0.0264
Kappa	0.4759	0.0422	0.0068
IOU	0.3164	0.0311	0.0134

5. Applicability to Other Datasets and Discussion

Although our model can generate the prediction output that identifies building constructions in the Bangkok area, the training set for the model is also from the Bangkok area, so its applicability to

other areas can be questioned. To prove that our model can be used globally, we tested it with the Hanoi and Xiamen areas, which are completely different from the training area, to see if it can detect constructions as effectively as it does in the Bangkok area.

5.1. Hanoi Testing Site

The first testing data we selected are from the area of Hanoi. Hanoi is not only considered to be a developing city comparable to Bangkok, but it also has a similar environment. The results of the Hanoi area from the proposed model trained with the Bangkok dataset are shown in Figure 11. While Bangkok and Xiamen each have two test areas, we validated the Hanoi test site with one area because of the lack of available data.

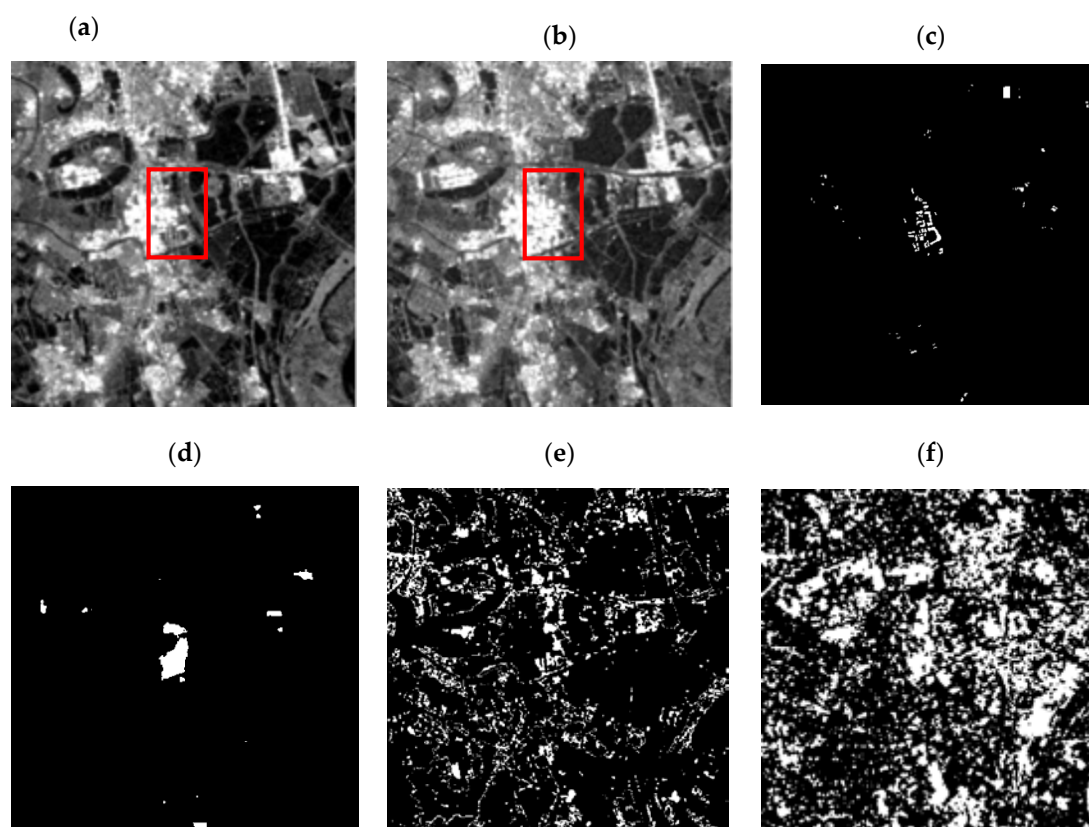


Figure 11. Result of the Hanoi site. The resolution of each image is $6\text{ km} \times 6\text{ km}$. (a) Time 1 SAR data, (b) Time 2 SAR data, (c) ground truth, (d) result of proposed model, (e) result of FCM, (f) result of Otsu thresholding.

The changes occurring between the two times of the SAR image pair comprise mainly the construction of a small building. The building changes detected by the proposed model are larger than in those in the ground truth, especially at the buildings around the center of the image. This can be explained by the large group of intensity changes in the SAR images, as indicated by the red rectangles in Figure 11a,b. Also, because the sizes of the constructions in the Bangkok area used to train the network are larger than those in the Hanoi area, the model tends to perform better in the detection of constructions with sizes that are similar to those in the training data. As a result, the predicted sizes of building changes in the Hanoi area are larger than those in the ground truth. While the detected size and shape resulting from the FCM and Otsu methods are closer to the ground truth, as shown before, they cannot distinguish between building changes and other kinds of changes.

The accuracy is shown in Table 6.

Table 6. Accuracy of the model applied to the Hanoi area.

Validation Method	Proposed Network	FCM	Otsu's Threshold
False negative	58.3236	55.1804	29.2308
False positive	0.9218	10.7332	30.7600
Overall accuracy	98.77%	89.03%	69.25%
Precision	0.1962	0.0220	0.0084
Recall	0.4168	0.4482	0.7077
F measure	0.2051	0.0239	0.0091
F1 measure	0.2668	0.0420	0.0165
Kappa	0.2614	0.0321	0.0094
IOU	0.1539	0.0215	0.0083

Even though the model has lower accuracies in Hanoi compared with those in Bangkok, the accuracy of the proposed network is higher than that of the other methods in almost every aspect.

5.2. Xiamen Testing Site

Similar to the Hanoi case, we validated our model with the Xiamen test site, which has been developing rapidly throughout the last decade. The two areas were tested, and the results of our model are as follows.

In the first test site, a faded line from the center to the bottom, as seen in Figure 12b, is not detected by our model; we assume that this line is noise in the SAR image because it does not appear in any optical images close to these dates, while both FCM and Otsu failed to ignore this line. On the other hand, our model can predict most of the constructions built, even those on the artificial island created around the center of the images between Time 1 and Time 2.

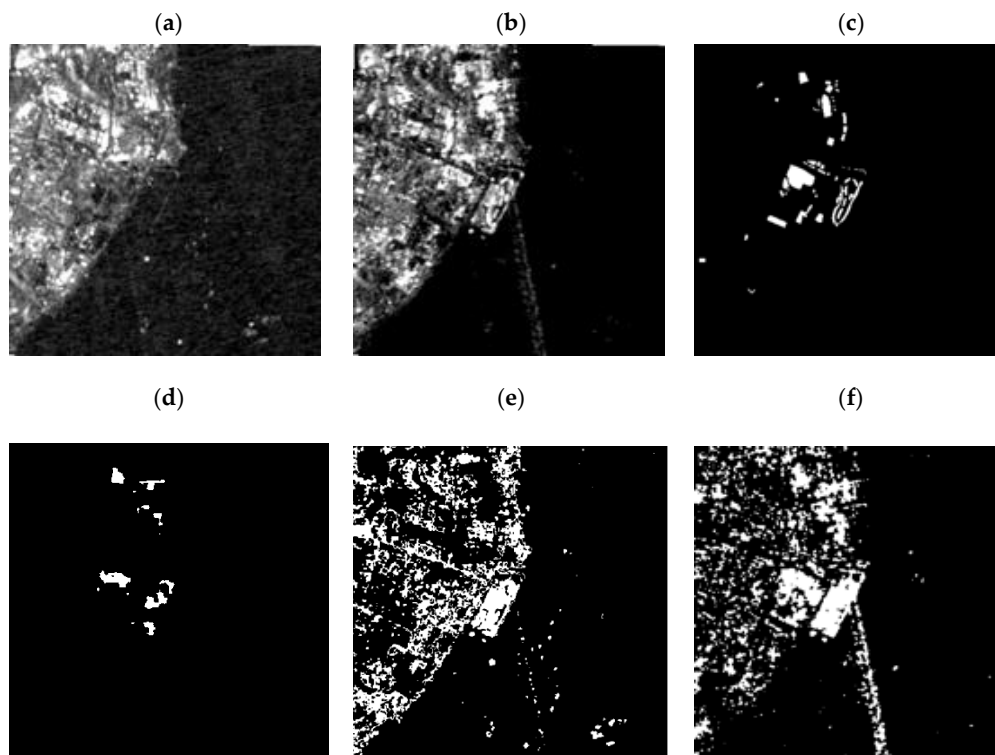


Figure 12. Result of the first Xiamen test site. The resolution of each image is $6 \text{ km} \times 6 \text{ km}$. (a) Time 1 SAR data, (b) Time 2 SAR data, (c) ground truth, (d) result of proposed model, (e) result of FCM, (f) result of Otsu thresholding.

At the second test site, three bridges are correctly excluded from our model's prediction, as this is not the objective of this research. While these bridges are in the middle of construction, please note that bridges in Figure 13a have a higher intensity than those in Figure 13b because of the nearly complete condition of the bridge surfaces at the time the image in Figure 13b was acquired; these surfaces cause more reflectance of the SAR signal.

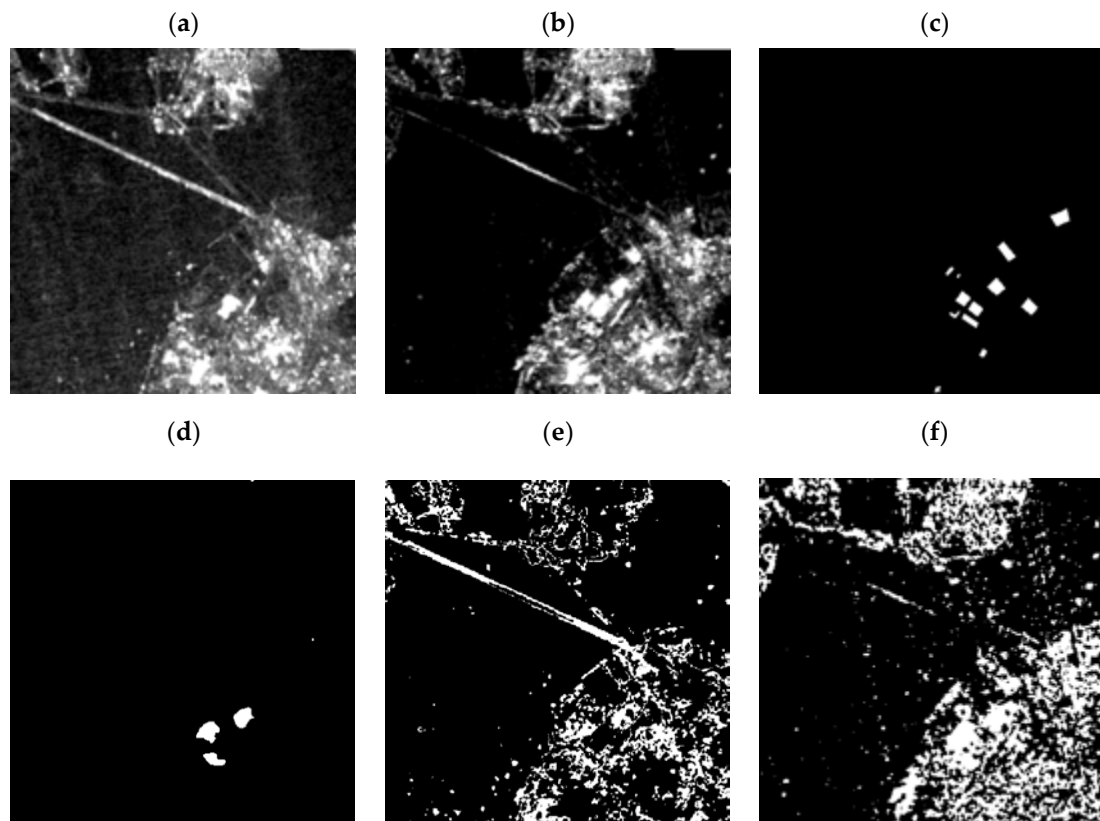


Figure 13. Result of the second Xiamen test site. The resolution of each image is $6 \text{ km} \times 6 \text{ km}$. (a) Time 1 SAR data, (b) Time 2 SAR data, (c) ground truth, (d) result of proposed model, (e) result of FCM, (f) result of Otsu thresholding.

The accuracy is shown in Table 7.

Table 7. Accuracy of the model in the Xiamen area.

Validation Method	Proposed Network	FCM	Otsu's Threshold
False negative	77.5769	63.0076	30.4775
False positive	0.5083	12.1638	15.7484
Overall accuracy	98.4121%	87.88%	84.05%
Precision	0.3852	0.0414	0.0590
Recall	0.2242	0.3699	0.6952
F measure	0.3636	0.0447	0.0638
F1 measure	0.2834	0.0745	0.1088
Kappa	0.2756	0.0506	0.0852
IOU	0.1651	0.0387	0.0575

Compared with the accuracies for the Hanoi area, the accuracies for Xiamen are improved, even though they are not as good as those in the Bangkok area. The main reason is that Xiamen is an island area surrounded by water. Because the model was trained with the Bangkok dataset, in which there is no water, the accuracy of the Xiamen result is slightly lower than that of the Bangkok result.

5.3. Other Experiments

We also applied our model to an area that is outside of the ground truth boundary for the Bangkok image pair acquired on 27 November 2008 and 15 January 2010. The proposed model detects the appearance of new construction that is actually happening, as shown in Figure 14. As it is difficult to see the corresponding area between optical images and SAR images because of the low resolution of SAR images, we add red rectangles in order to allow the readers see the boundary of buildings clearer. Please note that area in Figure 14 is cropped from the tested image of 400×400 pixels ($6 \text{ km} \times 6 \text{ km}$ resolution).

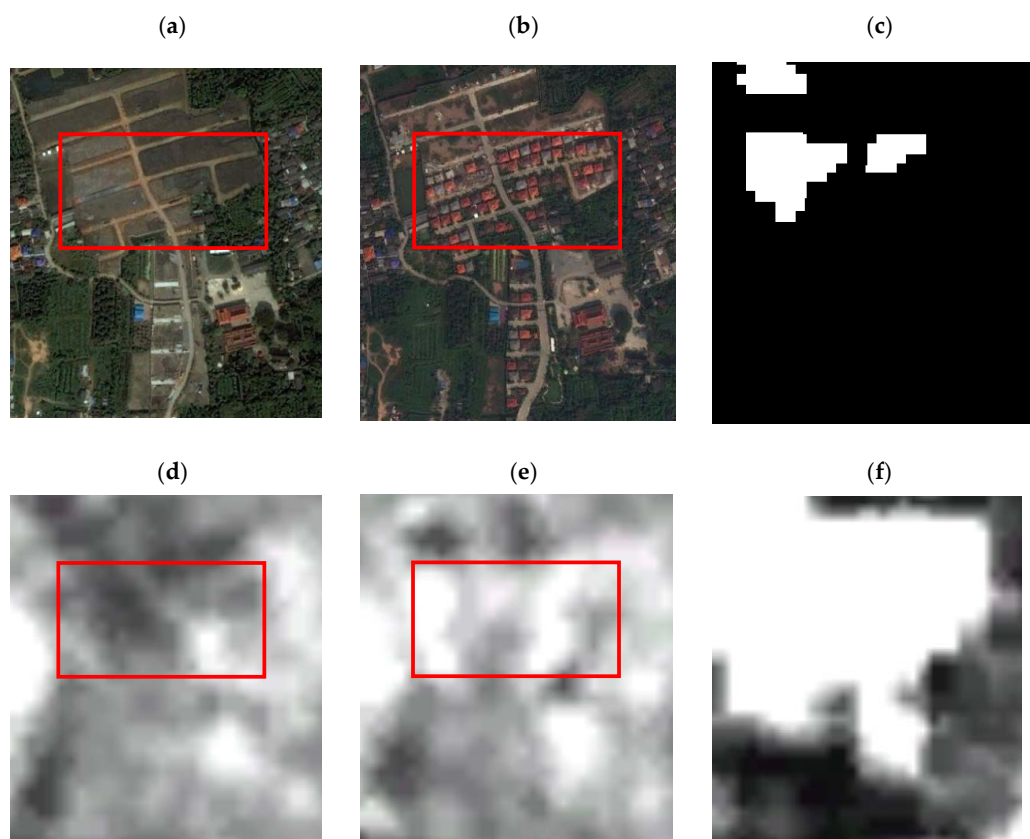


Figure 14. The result of the proposed model with an area outside the ground truth boundaries. The resolution of each image is $0.5 \text{ km} \times 0.45 \text{ km}$. (a) Time 1 optical data, (b) Time 2 optical data, (c) proposed result, (d) Time 1 SAR data (zoom), (e) Time 2 SAR data (zoom), (f) prediction map.

In addition to testing other networks and testing areas, we further tested our model, which was trained with ascending SAR data, with descending SAR data. The result in Figure 15 shows that the proposed model can also be used with SAR data from another orbit. Please note that the size of the descending SAR images used in this experiment is 300×300 pixels, and the ground truth is the same for the 12 January 2009/15 January 2010 SAR pair.

The accuracy shown is in Table 8.

Most of the accuracies resulting from using the model with descending SAR images are even slightly higher than those with ascending SAR images of Bangkok (Table 5). This is because there are more significant differences between a pair of descending images compared with ascending images, as the model tends to detect construction more easily when there is a significant change in intensity values. As can be seen in Figure 16, the highest number of pixels of the descending images at Time 1 and Time 2 is different, resulting in intensity values of approximately -8 and -8.5 , respectively; on the other hand, in the ascending images, the intensity values are the same at approximately -8.5 , although

the number of pixels is different. Please note that the size of each image used to create the histogram is 400×400 pixels.

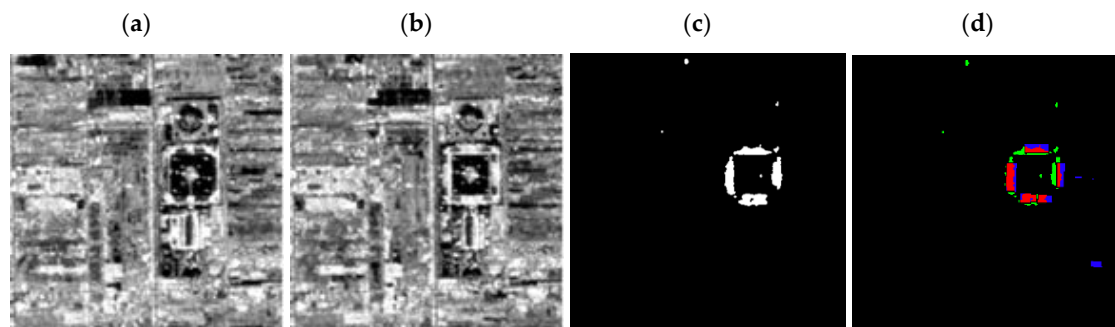


Figure 15. The result of the proposed method with descending SAR image data. The resolution of each image is $3.5 \text{ km} \times 3.5 \text{ km}$. (a) SAR image from 18 September 2008; (b) SAR image from 9 August 2010; (c) the proposed result; (d) comparison of the proposed result and the ground truth: (red) true positive area, (green) false positive area, (blue) false negative area.

Table 8. Accuracy of the model with descending SAR data.

Validation Method	Descending SAR Image
False negative	41.9199
False positive	0.4677
Overall accuracy	98.51%
Precision	0.6209
Recall	0.5808
F measure	0.6174
F1 measure	0.6002
Kappa	0.5663
IOU	0.4287

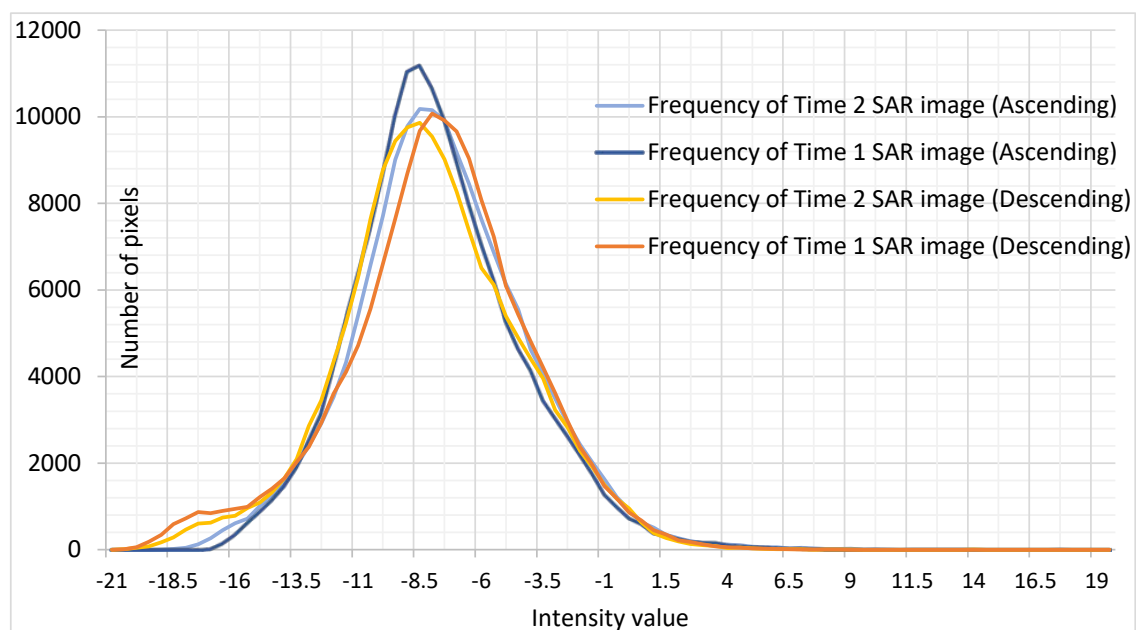


Figure 16. Histogram of Time 1 and Time 2 ascending and descending SAR images (the dates of acquiring the Time 1 and Time 2 ascending image pair are 10 January 2008/15 January 2010, and the dates of acquiring the Time 1 and Time 2 descending image pair are 18 September 2008/9 August 2010).

5.4. Model Accuracy Discussion

Because of the variety and complexity of the shapes of buildings in satellite images, it is difficult for the proposed method to give the exact shape of a detected building change, which is reflected by the low recall of our results. On the other hand, our proposed method can give the position of almost all building changes that occurred between two different dates, and these changes can be confirmed by visual inspection. The model can identify almost all changes according to the ground truth. The false positive rate is low because the score was obtained by a pixel-based validation method. As can be seen in Figure 17, almost all of the blue areas are attached to the red areas, which means that our model can predict newly built construction positions that are very close to the ground truth.

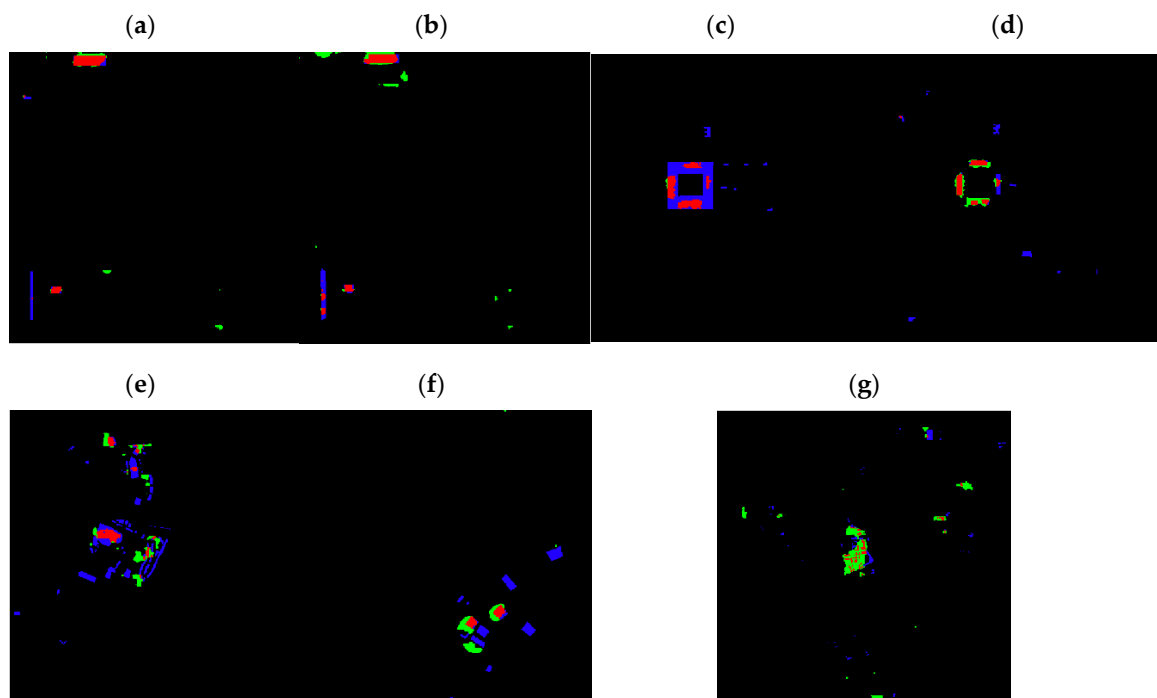


Figure 17. Comparison of the proposed model's result and the ground truth of (a) the first test area of Bangkok SAR pair 27 November 2008/15 January 2010, (b) the first test area of Bangkok SAR pair 12 January 2009/21 November 2009, (c) the second test area of Bangkok SAR pair 27 November 2008/15 January 2010, (d) the second test area of Bangkok SAR pair 12 January 2009/21 November 2009, (e) the first test area of Xiamen SAR pair, (f) the second test area of Xiamen SAR pair, (g) the test area of Hanoi SAR pair ((red) true positive area, (green) false positive area, (blue) false negative area) (the resolution of each image is $6\text{ km} \times 6\text{ km}$).

The area in Figure 18 is a cropped version of the first test area of Bangkok to discuss whether the model can or cannot detect newly built constructions from SAR images. The model can precisely detect the construction if the change in SAR intensity is significant, as shown by the blue rectangles in Figure 18d,e. Please note that the southern part of the blue rectangle has high intensity in the Time 1 SAR image, but there is no house in the optical image because of a time gap between the available optical image and our SAR dataset. On the other hand, in some cases, the model is not able to detect constructions if the difference in SAR intensity is too small, as in the case of the row of houses in the red rectangles in Figure 18d,e. However, it is nearly impossible for any algorithm or even manual inspection to detect changes if the difference in intensity is very low. The reason for the low intensity of the houses in the red rectangular area compared with the intensity of those in the blue rectangular area is the difference in the orientation of the houses. As they are constructed in different orientations, it is possible that they reflect the SAR signal differently. The high intensity in the blue rectangle is

possibly the result of the double bounce on the houses' walls or a strong single bounce on the houses' roofs, but these phenomena do not happen with the houses in the red rectangle because the orientation of the houses is different, and the latter may end up reflecting the SAR signal at their roof edge. It is also worth mentioning that our model does not detect any changes that are not construction changes. In Figure 19a,b, the change due to cutting forests is not detected by our model even though there is a significant change in SAR intensity, as shown in Figure 19c,d. The change in paddy fields caused by seasonal effects is also not detected by our model, as shown in Figure 19e–h. Despite its efficiency in distinguishing constructions from paddy fields or open spaces, it is possible that the model would fail to detect an area such as land with snow cover, for which the intensity change differs from that of the constructions and paddy fields or open spaces that we used to train the model. Constructions with very different shapes from those we used for training also have a small chance of not being detected by our model. We believe that the result of the model is strong enough for the goals of this study and can be developed further to use as the ground truth of building changes in any other work.

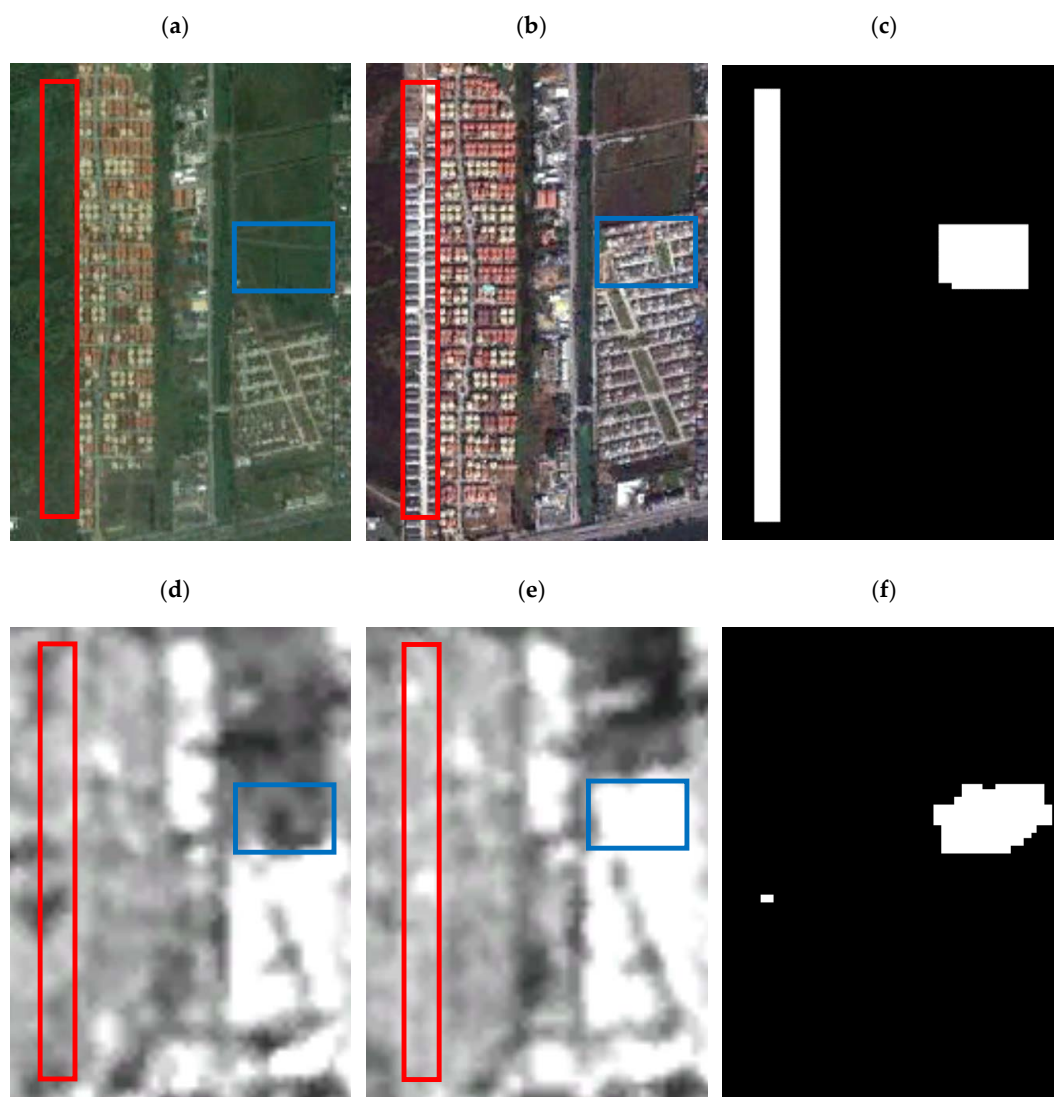


Figure 18. Example of successful detection and failed detection: (a) optical image from 22 August 2008, (b) optical image from 15 April 2010, (c) ground truth, (d) SAR image from 27 November 2008, (e) SAR image from 15 January 2010, (f) result of the model (the resolution of each image is $1 \text{ km} \times 0.75 \text{ km}$).

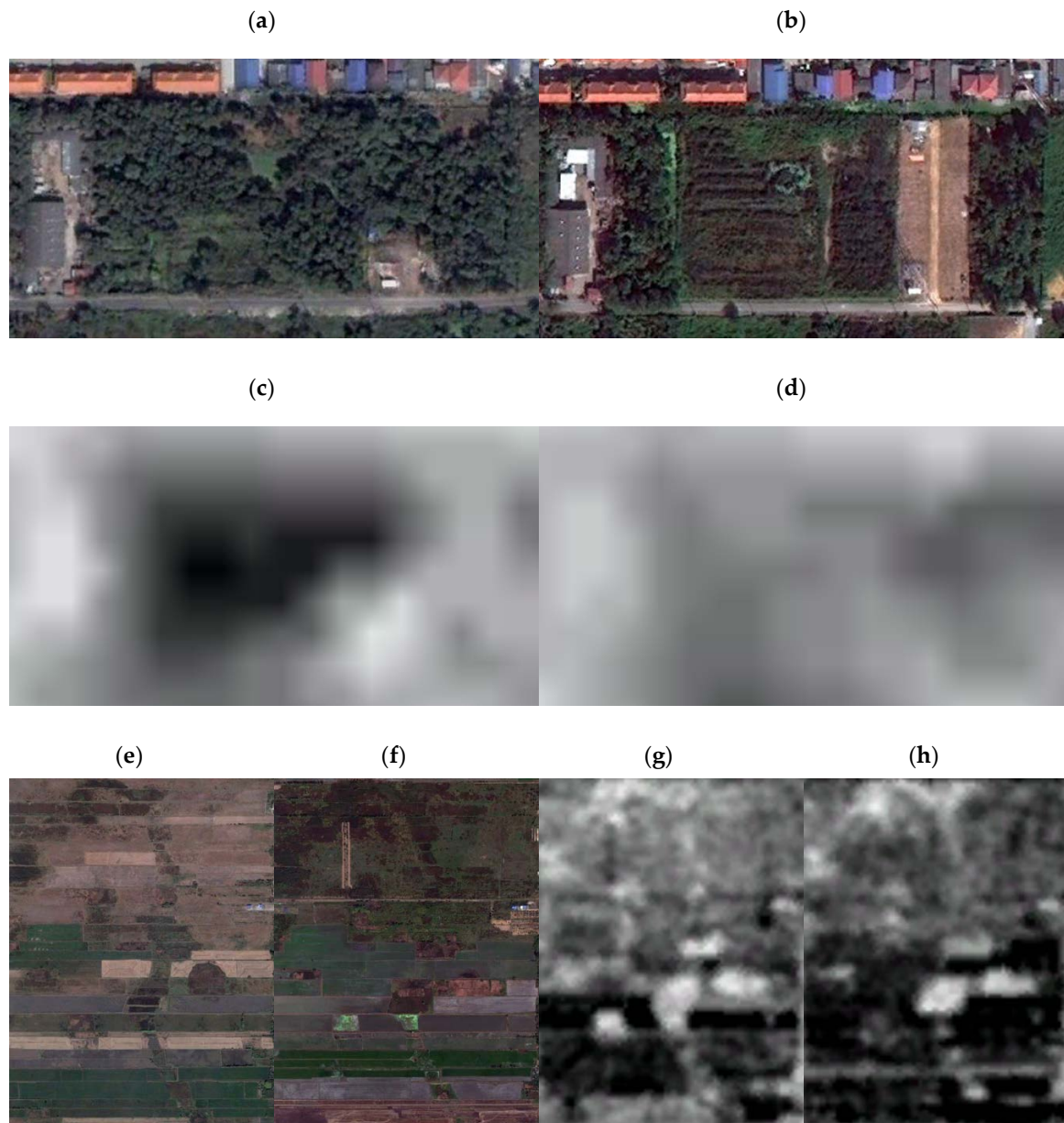


Figure 19. Examples of changes ignored by the model: (a) optical image of forest area from 18 December 2004, (b) optical image of forest area from 15 April 2010, (c) SAR image of forest area from 27 November 2008, (d) SAR image of forest area from 15 January 2010, (e) optical image of paddy field area from 18 December 2004, (f) optical image of paddy field area from 15 April 2010, (g) SAR image of paddy field area from 27 November 2008, (h) SAR image of paddy field area from 15 January 2010 (the resolution of each image is $0.14 \text{ km} \times 0.265 \text{ km}$ for (a–d) and $1.5 \text{ km} \times 1.2 \text{ km}$ for (e–h)).

6. Conclusions

In this research, we propose a U-net-based network to detect the new construction of buildings in developing areas between two SAR images taken at different times. Because the proposed model is based on the U-net, which includes a skip connection, it can generate good results without losing boundary information, while other FCNs cannot, even with our dataset from ALOS-PALSAR at a resolution of 15 m/pixel. The dataset was preprocessed to reduce noise with a 3×3 Lee filter, and the intensity was normalized to $[-1, 1]$. Then, images were cut into patches before their use in the network training process. The ground truth used for network training and testing was created manually by drawing polygons on optical images obtained from Google Earth. Because of the unbalanced class

weights in the training dataset, we weighted the loss function with the ratio between the positive class percentage and negative class percentage. The suitable weight, patch size, and epoch number used for network training were obtained after conducting several experiments. Our U-net-based model satisfies our objective, which is to identify the position of the newly built constructions. By comparing the results with the ground truth, we validated the proposed model with conventional methods, and it achieves a higher accuracy with any testing area. In addition to its effectiveness using the ascending SAR data of the Bangkok area, which we used as training data, the model can also return the position of changes in Hanoi and Xiamen, and is successful when used with descending SAR data. As the current ground truth data do not contain the constructions smaller than 2025 m² and the training data of Bangkok area do not contain mountains or water, which can lead to failure when using the model with such areas, we will try to further generalize our method by investigating this topic in the future.

Author Contributions: R.J. proposed the method, conducted experiments, and wrote the manuscript. M.M. improved the structure of the manuscript and provided information of SAR data. N.K. provided opinion and information on testing areas. S.K. revised the manuscript and provided information of SAR data analysis. R.I. improved the method and supported the implementation of the algorithm used in the method. R.N. provided dataset and corresponding information, and supplied the server for conducting experiments.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bazi, Y.; Bruzzone, L.; Melgani, F. Automatic Identification of the Number and Values of Decision Thresholds in the Log-Ratio Image for Change Detection in SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 349–353. [\[CrossRef\]](#)
2. Mu, C.; Li, C.; Liu, Y.; Sun, M.; Jiao, L.; Qu, R. Change detection in SAR images based on the salient map guidance and an accelerated genetic algorithm. In Proceedings of the 2017 IEEE Congress on Evolutionary Computation (CEC), San Sebastian, Spain, 5–8 June 2017; pp. 1150–1157. [\[CrossRef\]](#)
3. Liu, M.; Zhang, H.; Wang, C.; Wu, F. Change Detection of Multilook Polarimetric SAR Images Using Heterogeneous Clutter Models. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 7483–7494.
4. Hu, H.; Ban, Y. Unsupervised Change Detection in Multitemporal SAR Images Over Large Urban Areas. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 3248–3261. [\[CrossRef\]](#)
5. Gong, M.; Zhou, Z.; Ma, J. Change Detection in Synthetic Aperture Radar Images based on Image Fusion and Fuzzy Clustering. *IEEE Trans. Image Process.* **2012**, *21*, 2141–2151. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Ban, Y.; Yousif, O. Multitemporal Spaceborne SAR Data for Urban Change Detection in China. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 1087–1094. [\[CrossRef\]](#)
7. Arel, I.; Rose, D.C.; Karnowski, T.P. Deep Machine Learning—A New Frontier in Artificial Intelligence Research [Research Frontier]. *IEEE Comput. Intell. Mag.* **2010**, *5*, 13–18. [\[CrossRef\]](#)
8. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
9. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Pacifici, F.; Del Frate, F. Automatic Change Detection in Very High Resolution Images With Pulse-Coupled Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 58–62. [\[CrossRef\]](#)
11. Xu, Z.; Wang, R.; Zhang, H.; Li, N.; Zhang, L. Building extraction from high-resolution SAR imagery based on deep neural networks. *Remote Sens. Lett.* **2017**, *8*, 888–896. [\[CrossRef\]](#)
12. De Jong, K.L.; Bosman, A.S. Unsupervised Change Detection in Satellite Images Using Convolutional Neural Networks. *arXiv* **2018**, arXiv:1812.05815.
13. Bai, Y.; Mas, E.; Koshimura, S. Towards Operational Satellite-Based Damage-Mapping Using U-Net Convolutional Network: A Case Study of 2011 Tohoku Earthquake-Tsunami. *Remote Sens.* **2018**, *10*, 1626. [\[CrossRef\]](#)

14. El Amin, A.M.; Liu, Q.; Wang, Y. Convolutional neural network features based change detection in satellite images. *Proc. SPIE* **2016**. [[CrossRef](#)]
15. Xu, Y.; Wu, L.; Xie, Z.; Chen, Z. Building Extraction in Very High Resolution Remote Sensing Imagery Using Deep Learning and Guided Filters. *Remote Sens.* **2018**, *10*, 144. [[CrossRef](#)]
16. Gong, M.; Yang, H.; Zhang, P. Feature learning and change feature classification based on deep learning for ternary change detection in SAR images. *ISPRS J. Photogramm. Remote Sens.* **2017**, *129*, 212–225. [[CrossRef](#)]
17. Gong, M.; Zhao, J.; Liu, J.; Miao, Q.; Jiao, L. Change Detection in Synthetic Aperture Radar Images Based on Deep Neural Networks. *IEEE Trans. Neural. Netw. Learn. Syst.* **2016**, *27*, 125–138. [[CrossRef](#)] [[PubMed](#)]
18. Ajadi, O.A.; Meyer, F.J.; Webley, P.W. Change Detection in Synthetic Aperture Radar Images Using a Multiscale-Driven Approach. *Remote Sens.* **2016**, *8*, 482. [[CrossRef](#)]
19. Bazi, Y.; Bruzzone, L.; Melgani, F. An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 874–887. [[CrossRef](#)]
20. Iino, S.; Ito, R.; Doi, K.; Imaizumi, T.; Hikosaka, S. CNN-based generation of high-accuracy urban distribution maps utilising SAR satellite imagery for short-term change monitoring. *Int. J. Image Data Fusion* **2018**, *9*, 302–318. [[CrossRef](#)]
21. Lee, J.S. Speckle analysis and smoothing of synthetic aperture radar images. *Comput. Graph. Image Process.* **1981**, *17*, 24–32. [[CrossRef](#)]
22. Badrinarayanan, V.; Alex, K.; Roberto, C. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *arXiv* **2015**, arXiv:1511.00561. [[CrossRef](#)]
23. Zeiler, M.D.; Taylor, G.W.; Fergus, R. Adaptive deconvolutional networks for mid and high level feature learning. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Barcelona, Spain, 6–13 November 2011; pp. 2018–2025.
24. Aitken, A.P.; Ledig, C.; Theis, L.; Caballero, J.; Wang, Z.; Shi, W. Checkerboard artifact free sub-pixel convolution: A note on sub-pixel convolution, resize convolution and convolution resize. *arXiv* **2017**, arXiv:1707.02937, .
25. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv* **2015**, arXiv:1502.03167.
26. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. *arXiv* **2016**, arXiv:1611.07004.
27. Haixiang, G.; Yijing, L.; Shang, J.; Mingyun, G.; Yuanyue, H.; Bing, G. Learning from class-imbalanced data: Review of methods and applications. *Expert Syst. Appl.* **2017**, *73*, 220–239. [[CrossRef](#)]
28. Bezdek, J.C.; Ehrlich, R.; Full, W. FCM: The fuzzy c-means clustering algorithm. *Comput. Geosci.* **1984**, *10*, 191–203. [[CrossRef](#)]
29. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man* **1979**, *9*, 62–66. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).