6th Intercontinental Geoinformation Days

igd.mersin.edu.tr

Change detection of buildings using high spatial resolution images and deep learning

Saeid Abdolian[1], Ali Esmaeily[1], Mohammad Reza Saradjian*[2]

[1] Graduate University of Advanced Technology, Faculty of New Sciences and Technologies, Kerman, Iran
[2] University of Tehran, College of Engineering, School of Surveying and Geospatial Engineering, Tehran, Iran

Keywords
Change detection
Deep learning
Relevant training dataset
Buildings
Newly constructed areas

Abstract
The main aim of this study is to detect changes in buildings. The data required to achieve this goal are high spatial resolution images and the method used to achieve the goal is the use of a deep neural network and a new relevant dataset. In this research, the training and testing of a deep neural network is investigated to detect changes in buildings such that the accuracy is improved. The deep learning networks have not been previously used in industrial areas that have different characteristics such as building structures, building construction methods, materials and color variety of the roofs of the buildings. In this research, it is assumed that the development of networks alone does not improve the accuracy of the results considerably, but the training of previously developed networks with the relevant training dataset greatly increases the accuracy. Accordingly, a training dataset to detect the changes in the buildings of industrial areas have been produced to train an already developed deep learning network with the relevant data to detect the change in buildings. As a result, the trained network with relevant data was tested, and reasonable average accuracy and recall achieved.

1. Introduction

Deep neural networks, which use multiple layers between the input and output layers, have been widely used by providing solutions in solving problems that were very complex previously. Hinton and Salakhutdinov in 2006 applied deep learning to data processing and showed that the depth of the artificial neural network can be very effective in the learning ability of the network (Hinton and Salakhutdinov, 2006). Optimal features can be directly extracted from raw data by deep learning methods and the network can reach maximum accuracy in many applications such as building change detection.

2. Study area

The study area is two industrial parks named Mobarakeh Industrial Park in Mobarakeh City and Shamsabad Industrial Park in Rey City in Iran. Mobarakeh Industrial Park is located on the Isfahan-Shiraz Road and 28 km away from Mobarakeh city. It is located between 51.71° to 51.75° East longitude and 32.40° to 32.44° North latitude. Shamsabad Industrial Park is located in the south of Tehran province in Rey city. Shamsabad Industrial Park is the largest industrial park in Iran with an area of over 3 thousand hectares. The location of Shamsabad Industrial Park is at 45 km of Tehran-Qom highway, in the district of Hassanabad city. Shamsabad Industrial Park is located between 51.18° to 51.29° East longitude and 35.30 to 35.38° North latitude.

3. The Data

In order to collect training dataset, satellite images from the two industrial parks were downloaded through Google Earth Pro software (Figure 1). When downloading Google Earth images, the specification of the images is selected by image dimensions and scale. The images in this study were divided into three groups, each group containing two time series images. Two groups were used as network training and one group as network testing. Also, the test dataset is selected from another part of Shamsabad Industrial Park. Details of images including date, dimensions, and loading scale are provided in Table 1.

Due to the lack of internal training datasets, an external dataset selection has been used in the training process. The LEVIR-CD dataset is a new large-scale remote sensing binary structural change detection dataset that helps develop new deep learning-based algorithms for change detection in remote sensing images. The introduced dataset is a new benchmark for evaluating change detection algorithms, especially deep learning-based algorithms.

* Corresponding Author

*(sarajian@ut.ac.ir) ORCID ID 0000 – 0002 – 1734 – 5860
(aliesmaeily@hotmail.com) ORCID ID 0000 – 0002 – 5463 – 5613
(saeid.abdolian.rs@gmail.com) ORCID ID xxxx – xxxx – xxxx – xxxx

Cite this study

Abdolian, S., Esmaeily, A., & Saradjian, M. R. (2023). Change detection of buildings using high spatial resolution images and deep learning. Intercontinental Geoinformation Days (IGD), 6, 97-100, Baku, Azerbaijan

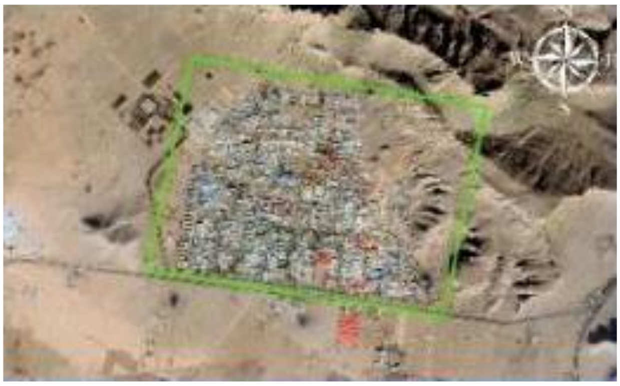

Figure 1. The study area: top) Shamsabad Industrial Park (green box, training area and yellow box, test area), bottom) Mobarakeh Industrial Park (green box, training area).

Table 1. Images used as a training set

Location	Image date	Image dimensions
Shamsabad	2005/03/09	2086
Industrial Park	2017/03/13	×
		4195
Shamsabad	2002/05/05	2071
Industrial Park	2017/03/31	×
		4739

The LEVIR-CD set consists of 637 pairs of very high resolution (0.5 m/pixel) image segments with a size of 1024 × 1024 pixels. These two-time images with a period of 5 to 14 years have significant land use changes, especially the growth of constructions. LEVIR-CD covers different types of buildings such as villas, high-rise buildings, garages and warehouses. In this collection, an attempt has been made to focus on building-related changes, including change from soil/vegetation/built-up land or building under construction to new construction areas, and also building decline. These two-time images are labeled by remote sensing image interpretation experts using binary labels (i.e., 1 for change and 0 for no-change). LEVIR-CD contains 31,333 examples of change building. Using this dataset, it was tried to create a separate dataset for large warehouses and create a new dataset. This new dataset, which contains 522 images in dimensions of 256 × 256, was added to internal dataset to increase the volume of training data.

4. Method

Change detection in Shamsabad Industrial Park was done by STANet (Spatial-Temporal Attention-based neural network). This network consists of convolutional deep learning algorithm (CNN) (Shen et al. 2019), advanced ResNet *(He et al. 2016)* model and combination of several modules in order to improve the training process. In other networks, two-time images are normally coded separately, while in this network, a self-attention change detection mechanism is designed. This module makes full use of the spatio-temporal relationship to obtain features. By doing this, the examination of pixels that have become challenging in terms of brightness and misregistration have also been improved to some extent.

The STANet (Zhou et al. 2022) network consists of three general parts: 1) Basic feature extractor by ResNet architecture with FCN (Sun et al. 2020) algorithm based on transfer learning approach, 2) Extraction of more distinct and accurate features by PAM (Pyramid spatial-temporal Attention Module) (Chen and Shi 2020), and 3) Check changes by metric module.

5. Implementation

In order to implement and analyze the results, the training dataset for the detection of construction changes produced from Shamsabad and Mobarakeh industrial parks were used to train the STANet network. After the training of STANet network, a part of Shamsabad Industrial Park was tested by this network. First, data preparation, then network training and hyperparameter estimation were done. Then the network was tested and the test results were presented along with their performance analysis. At the end, the test output was validated and a more detailed change representation was provided. Data preparation consists of three parts, geometric corrections, change label generation, and image patches generation.

First, geometrical corrections were made. Then, the change label was generated to train the network and validate the experiment. The actual land change image, which is a binary image, is a change map or label, where the white color represents the areas of change and the black color represents the areas of no change. By visually comparing the time series images, the pixels of newly built buildings were changed to white and other areas of the image were changed to black. In this research, in order to train the STANet network, a change map was produced from the two industrial parks. After training the STANet network, it is necessary to test the network. For this purpose, a part of Shamsabad Industrial Park was tested by this network. Then, in order to validate the test from this region, the change label of this region was used and the validation criteria were applied.

In order to train and test the STANet network, input images with specific dimensions should be input to the network. These dimensions may be set to 32, 64, 128, 256, 512 and 1024 pixels. In dimensions smaller than 32 pixels, the amount of information for analysis is too little. Dimensions larger than 1024 pixels slow down the processing speed. The smaller the size, the less RAM is used for processing which increases the processing speed and vice versa. The dimensions of the images downloaded from Google Earth are larger than the dimensions defined in the grid input because they were taken from the entire area. Therefore, the original images were divided into smaller patches. The dimensions of 256 × 256 were appropriate in terms of the amount of RAM used for processing speed and the amount of information for network analysis.

The STANet network was trained in two stages, and tested in three stages according to Figure (2). First, the STANet network which was trained by the LEVIR-CD dataset, was tested for the internal region to determine the capability of the network as well as the capability of the LEVIR-CD training dataset. In the following, first dataset of Mobarakeh Industrial Park, which contains 117 images of 256 × 256 pixels was combined with a selection of warehouse changes of the LEVIR-CD dataset. Then, the STANet network was trained using the combination of this dataset, and a part of Shamsabad Industrial Park was tested. Then, the second training dataset of Shamsabad Industrial Park which contains 128 images with dimensions of 256 × 256 pixels, were added to this dataset. By combining these three datasets, a more appropriate dataset was obtained. In the last step, the STANet network was trained again using the combination of this newly formed dataset. The same part of Shamsabad Industrial Park was tested again. As a result, the importance of adding internal training data was revealed.

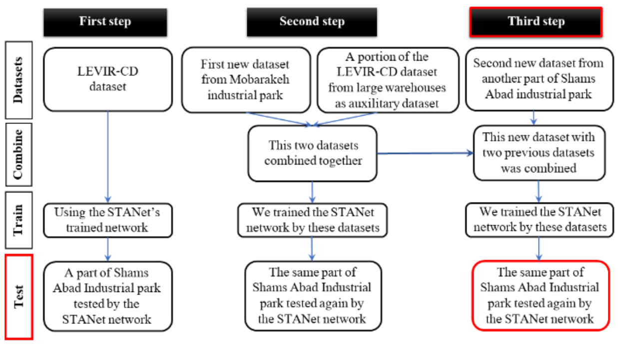

Figure 2. Research stages in which, in the first stage, only the network trained using the external region was tested for the internal region and in the second and third stages, the network was trained and tested with the addition of new datasets of the internal regions.

5.1. Tools and processing environment

The programming code in this research was written in Python language. Python as an object-oriented and high-level programming language has many applications in various fields of data science, machine learning, deep learning, data mining, etc. The code was implemented in the Google Colab environment with free GPU (graphic processing units) which makes it possible to run heavy computer programs at a very high speed. The STANet network was implemented by two main libraries in deep learning. Pytorch library was used for vector calculations through GPU and deep neural network construction based on autodiff system. The second library was Torchvision library which was used to detect the edges of buildings in the image.

5.1.1. Network training

Network training was done in two stages. First, the setting of the algorithm and hyperparameters for the purpose of training and then the training steps were performed.

Choosing the right optimization algorithm for the deep learning model is very important and has a great impact on the time to reach the desired result. In this network, the Adam's optimization algorithm was used. This algorithm is a method that adjusts the learning rate during the training process. Adam's optimization algorithm is a generalized version of the stochastic gradient descent (SGD) algorithm, which has recently been widely used for deep learning applications in the field of computer vision. In the following, the hyperparameters set in the training process are explained.

In order to increase the amount of data and enrich (i.e. augmentation) the training dataset, the network creates a set of data different from the existing data by rotating and cropping the images. In this way, the amount of data increases and leads to an increase in accuracy. In the learning phase, the rotation rate was applied as a random rotation (-15° to 15°).

The whole training process was set to 200 epochs. The entire dataset was trained in each epoch and the training accuracy was validated by a series of training data. In each epoch compared to the previous epoch, the training process got better by resetting the new learning parameters.

The learning rate controls how much the model changes in response to the estimation error each time the model weights are updated. Determining learning rate is considered as a hyperparameter. If the learning rate is chosen low, the algorithm may get stuck in local minima. If a large training rate is selected, the network may reach an unstable state. The learning rate in this network was set to 0.001 for the first 100 epochs and then it was set to zero in the next 100 epochs with a linear decrease.

In stage 1 training, the first dataset was selected for training from an industrial park that has examples of change and no-change in proportion. The network was trained for both the changed buildings and the buildings that had a color difference in the two images as the no-change. Due to the lack of training data produced inside the country, the LEVIR-CD dataset was also used in the training process and a selection of large warehouses in this dataset was extracted and integrated with the internal training dataset. The learning process of the network in the first stage in 200 epochs is presented in four diagrams in Figure 3.

Stages 2 and 3 of training are combined here. In order to increase the internal training dataset, a second internal training dataset was also produced. This dataset was a part of Shamsabad Industrial Park. In the second and third stages of training, network training was performed by combining two internal training datasets (i.e. first and second datasets) and a selection of large buildings and warehouses (i.e. LEVIR-CD dataset). The training process of the network in these stages in 200 epochs is presented in four diagrams in Figure 4.

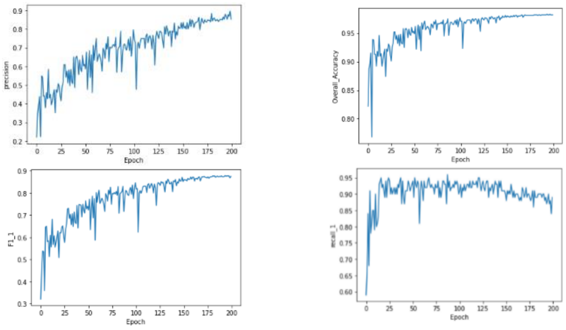

Figure 3. Four evaluation criteria of the training process of stage 1 in 200 epochs: the top-left graph shows the accuracy criterion, the top-right graph is the precision criterion, the bottom-left graph is the F1 criterion, and the bottom-right graph is the recall criterion.

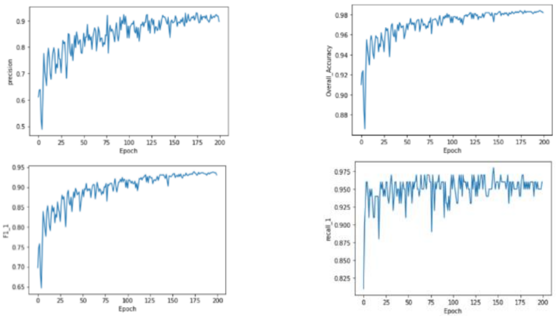

Figure 4. Four evaluation criteria of the training process of stages 2 and 3 in 200 epochs: the top-left graph shows the accuracy criterion, the top-right graph is the precision criterion, the bottom-left graph is the F1 criterion, and the bottom-right graph is the recall criterion.

5.1.2. Network testing

To carry out the test phase, after training the network, ready weights were generated so that during the test, only by introducing these weights, the test processed at a very high speed. The algorithm processed the data after installing the necessary libraries with the prepared weights obtained in the training stage. As a result, the binary image of the changes was produced as a prediction. The test phase was performed in three stages. In the third stage test, the network was trained by combining the two previous training datasets and the second internal training dataset (Shamsabad Industrial Park). The test was again performed on Shamsabad Industrial Park and reached a much better result than the previous two tests. The results showed that with the increase of the internal training dataset, the results continued to increase with great accuracy. Figure 5 shows the total result of change detection obtained from Shamsabad Industrial Park by STANet network. The first, second and third rows show the first, second and third stages of predicted changes, respectively. The left column shows the predicted change and no-change and the right column shows the correctness of the predicted change and no-change. In the left column, the white color is the change and the black color is no-change. In the right column, the white color is the obtained change, the blue color is the unpredicted change, and the red color is no-change that were classified as change mistakenly.

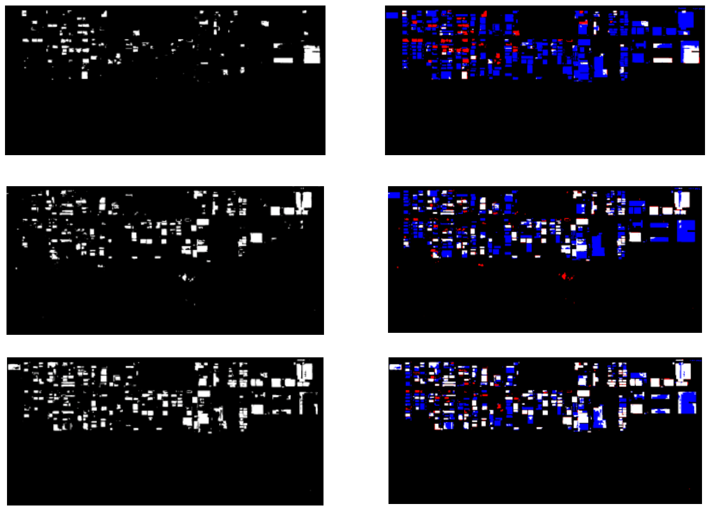

Figure 5. The total result of change detection obtained from Shamsabad Industrial Park by STANet network.

6. Conclusion

As mentioned, the result of prediction in the first stage was not accurate. In the second stage of the test, the average accuracy was 52% and the recall was 37% due to the training of the network by combining two datasets of internal and external training. The first training dataset were selected in such a way that the ratio of change and no-change was the same so that the network pays enough attention to both modes during learning. In this phase of testing, the network error was reduced. Next, the network was trained with more internal data. As a result, in the last stage of the test, the average accuracy was 69% and the recall was 57%. In addition, it was shown in this study that development of networks alone does not improve the accuracy of the results considerably, but the training of previously developed networks with the relevant training dataset greatly increases the accuracy.

References

Chen, H., & Shi, Z. (2020). A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. Remote Sensing, 12(10), 1662.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. science, 313(5786), 504-507.

Zhou, S., Dong, Z., & Wang, G. (2022). Machine-learning-based change detection of newly constructed areas from GF-2 imagery in Nanjing, China. Remote Sensing, 14(12), 2874.

Sun, Y., Zhang, X., Huang, J., Wang, H., & Xin, Q. (2020). Fine-grained building change detection from very high-spatial-resolution remote sensing images based on deep multitask learning. IEEE Geoscience and Remote Sensing Letters, 19, 1-5.