EARTH OBSERVATION AND GEOMATICS ENGINEERING
website: https://eoge.ut.ac.ir

Deep learning model improvement for building change detection by including local remote sensing training datasets

Saeid Abdolian[1], Ali Esmaeily[1], Mohammad Reza Saradjian[2*]
[1]Faculty of Civil and Surveying Engineering, Graduate University of Advanced Technology, Mahan, Kerman, Iran
[2]School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran

Article history:
Received: 2023-03-17, Accepted: 2023-06-22, Published: 2023-06-28

ABSTRACT

Deep learning networks that have been trained using a unique dataset normally do not produce convincing results when used on other datasets. In this study, it has been shown that the training of previously developed networks with the newly included local training datasets greatly increases accuracy. Accordingly, the aim of this study is to improve a DL model by including new local data along with an existing known dataset for building change detection. The STANet has not been previously challenged in industrial areas with the different building characteristics present in this new dataset. High-resolution satellite images have been used for building change detection in this study. The STANet has been implemented using a new dataset comprising an existing known dataset together with two local building datasets. The training and testing of the STANet are investigated using incremental training data in three stages. In the first stage, the STANet network that was trained on the already existing LEVIR-CD dataset was tested, and the changes obtained by the STANet network have achieved 89% accuracy, 55% precision, and 13% recall, which implies a not reasonable result. In the second stage, the network, which was trained by combining LEVIR-CD and a local dataset, was tested and 92% accuracy, 86% precision and 37% recall were obtained. In the third stage, the network, which was trained by combining the LEVIR-CD dataset, the local dataset from stage 2, and another more specific and complementary local dataset, was tested and reached a much better result than the previous two stages. As a result, the trained network with the corresponding data achieved a high 94% accuracy, 89% precision, and 57% recall. In conclusion, the significance of using the local corresponding training dataset on a DL model has been exhibited.

KEYWORDS

Deep learning improvement,
Local training datasets,
Change detection,
Newly constructed buildings

1. Introduction

Buildings as one of the important man-made objects have various applications. Change detection of buildings plays a crucial role in urban planning, natural resource management, risk assessment and land monitoring (Chandra & Vaidya, 2022). Information on building change aids city authorities in urban development, resource optimization, and management. Additionally, it supports environmental monitoring, city development assessment, and scientific investigations in various fields such as geography, environment, and urban planning. Therefore, building change detection is essential for effective decision-making and sustainable development in urban and environmental management (Darvishzadeh, 2000). Industrial parks, typically located on the outskirts of cities, serve as vital contributors to industrial and economic progress. They play an essential role in driving economic development, providing purpose-built spaces for large-scale enterprises. Government incentives are often extended to support the expansion of industrial parks due to their potential to stimulate economic growth (Liana-Eugenia & Nicoleta-Georgeta, 2013). Construction, demolitions, and alterations represent the main types of building changes, traditionally monitored through manual field surveys and visual interpretation which despite their accuracy are inefficient and lack real-time, practical, and global applicability; thus, automating building change detection is crucial to meet the increasing demand for dynamic monitoring and to achieve significant savings in labor, and time (Zhang *et al*., 2022).

* Corresponding author
E-mail addresses: saeid.abdolian.rs@gmail.com (S. Abdolian); aliesmaeily@kgut.ac.ir (A. Esmaeily); sarajian@ut.ac.ir (M.R. Saradjian)

The utilization of Google Earth imagery enables tracking of construction progress over time, facilitating trend analysis and change detection. Rapid and accurate identification of buildings change in industrial parks is essential for informed decision-making (Huang *et al*., 2022). Another important application of building change detection is to address the routine needs of property inspectors and government agencies and other entities tasked with construction inspection (Zhu *et al*., 2008). Building change detection involves the analysis of images capturing buildings at different time points to identify alterations (Lu *et al*., 2004). The utilization of Google Earth imagery enables tracking of construction progress over time, facilitating trend analysis and change detection (Huang *et al*., 2022).

Deep neural networks, which use multiple layers between the input and output layers, have been widely used to provide solutions for problems that were very complex previously. Lu *et al.* applied deep learning to data processing and presented that the depth of the artificial neural network can be very effective in improving the learning ability of the network (Lu *et al*., 2017). Deep learning models can automatically detect building footprints using high-resolution satellite images. It is essential to evaluate and compare the features of different deep learning models with different learning data. 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. It is shown in this study that the training of previously developed networks with the newly added local training dataset increases accuracy. This is because the development of networks alone using a specific dataset does not significantly improve the accuracy of the results when used in other areas. Accordingly, the aim of this study is to improve a deep learning model by including local datasets together with existing known datasets for building change detection.

In this research, two new datasets of industrial buildings in Iran have been produced. The new datasets have been combined with the LEVIR-CD (LEarning, VIsion and Remote sensing – Change Detection) dataset of buildings to improve STANet (Spatiotemporal Attention Network) deep learning model. In order to demonstrate the importance of using new datasets, three tests have been taken by the STANet network, and the importance of including local training datasets on the network has been shown in this study.

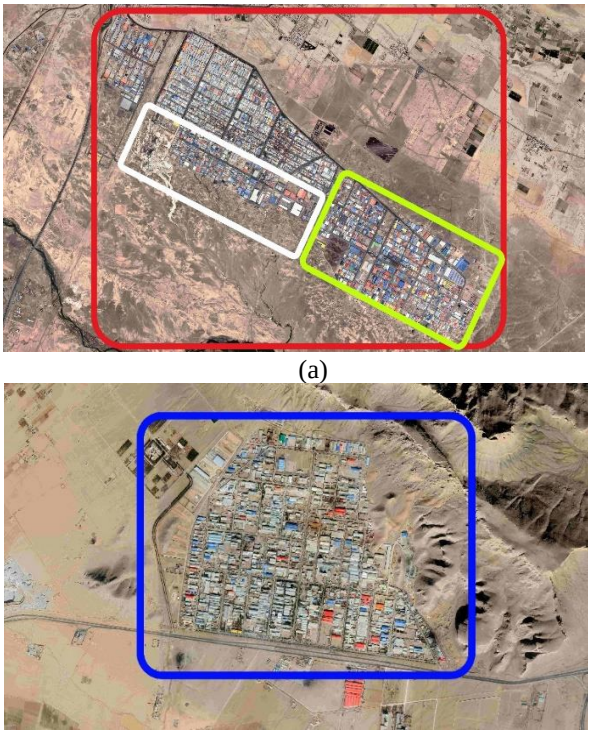
2. Material and Method

In this study, the material is data in the form of satellite 0.5m resolution images and training data as both LEVIR-CD and local data, and the method is the STANet as a deep learning model.

2.1 Data

Satellite images of change in industrial buildings necessary for the case study in this research have been downloaded with the help of Google Earth Pro Windows version with the highest possible quality. Google Earth image sources with very high resolution are provided by Maxar Technologies, which uses several satellites to collect Google Earth images. The satellite images of this technology include IKONOS, QUICKBIRD, WorldView-1/2/3/4/Legin, GeoEye-1 and Pleiades (Internet communication). This technology does not mention the type of satellite or the type of sensor in its images for each region, specifically in Google Earth. In order to prepare a continuous image for the globe, images from such different satellites and at different times are mosaicked together. Hence, the name of the sensor is not mentioned as the source of the images.

The images used as the case study area are from two industrial parks in Iran. They are Shamsabad industrial park in Rey City and Mobarakeh industrial park in Mobarakeh City (Figure 1).



(a)

(b)

Figure 1. The study area: a) Shamsabad industrial park (green box, training area and white box, test area), b) Mobarakeh industrial park (blue box, training area).

Shamsabad Industrial Park is located in the South of Tehran Province, in Rey City. It is located 45 km off the Tehran-Qom highway in the district of Hassanabad City. Its geographical location is between 51.18° and 51.29° East longitude and 35.30° to 35.38° North latitude. Mobarakeh Industrial Park is located on the Isfahan-Shiraz Road, 28 km away from Mobarakeh City. Its geographical location is between 51.71° to 51.75° East longitude and 32.40° to 32.44° North latitude. When downloading satellite images, the specifications of the images were provided as image dates and dimensions. Then, these local images were divided into three groups, each group containing bi-temporal images.

As the first network training dataset, a bi-temporal images of Mobarakeh industrial park selected. They originally had a size of 3570×2451 and were taken on 2010/01/09 and 2021/11/11. As the second network training dataset, another bi-temporal image of Shamsabad industrial park was selected. They originally had the size of 4195×2086 and were taken on 2005/03/09 and 2017/03/13. As the test dataset, a bi-temporal images of Shamsabad industrial park selected. They originally had

the size of 4739×2071 and were taken on 2002/05/05 and 2017/03/31.

In order to train and test the model for building change detection, there exist a variety of datasets. The LEVIR-CD dataset contains 637 image pairs of 0.5m resolution of building change. The images in the dataset were collected manually from Google Earth (Chen & Shi, 2020). Shen *et al*. (2021) published a new building change detection dataset called S2Looking and made a comparison with LEVIR-CD+ as the world's most up-to-date and reliable dataset. The S2Looking dataset contains 5000 binary pairs of rural images (Shen *et al*., 2021). Another dataset called the WHU building dataset consists of an aerial and a satellite imagery dataset of building change samples due to the Earthquake. The aerial dataset consists of more than 220,000 independent buildings extracted from aerial images with 0.075 m spatial resolution. The satellite imagery dataset consists of two subsets of cities from various remote sensing resources, including QuickBird, Worldview series, IKONOS, ZY-3, etc., with 2.7 m ground resolution. Another dataset called OSCD which is used in urban areas, contains 24 pairs of multispectral satellite images with a spatial resolution of 10 meters (Daudt *et al*., 2018). Also in such studies, the SZTAKI dataset has been used. It includes 12 pairs of aerial optical images with a 1.5-meter resolution. It has been extracted from a combination of building changes and afforestation (Benedek & Szirányi, 2008; Benedek & Szirányi, 2009).

As the most frequently used, reliable and suitable resolution dataset, the LEVIR-CD dataset has been selected for this study. It is a known binary building change detection dataset and helps develop a new deep learning-based algorithm for change detection. It is a benchmark for evaluating change detection algorithms, especially for deep learning-based algorithms.

In order to train the deep learning model, a selection of LEVIR-CD dataset has been used. As mentioned, the LEVIR-CD dataset consists of 637 pairs of 0.5 m/pixel resolution image segments with a size of 1024×1024 pixels. These bi-temporal images with a period of 5- to 14-year time intervals show significant land use changes, especially in the increase of building construction. LEVIR-CD covers different types of buildings, such as villas, high-rise buildings, garages and warehouses. In the LEVIR-CD collection, an attempt has been made to focus on building-related changes, including changes from soil/vegetation to built-up land or buildings under construction to new construction areas. These bi-temporal images are labeled 1 for change and 0 for no-change by remote sensing image interpretation experts. LEVIR-CD contains 31,333 examples of changed buildings. Buildings in industrial areas have different characteristics, such as structures, construction methods, materials and color variety of the roofs. From this dataset, another new dataset for large warehouses has been created and used in this study. This new dataset, which contains 522 images in dimensions of 256×256, was later used along with two local datasets to increase the amount of training data.

2.2 Method

In similar studies, comparisons have been made between different networks in the field of change detection. By comparing the methods of FC-EF, FC-Siam-Conc, FC-Siam-Diff, DTCDSCN, STANet, CDNet and so on, the accuracy obtained by the STANet method has achieved high accuracy (Su *et al*., 2022; Wu *et al*., 2024). Therefore, the STANet network has been selected as a suitable network for research in the regions of Iran in this study. The STANet network consists of four main parts: Feature Extractor, Spatial-Temporal Attention Module (consisting of BAM (Basic Attention Module) and PAM (Pyramid spatial-temporal Attention Module) modules), Metric Module and Loss Layer Design. Fully Convolutional Network (FCN) (Long *et al*., 2015) is a kind of CNN without fully connected layers that is widely used for dense classification tasks. Remote sensing image change detection requires pixel-wise prediction and benefits from the dense features of FCN-based methods (Chen *et al*., 2017; Badrinarayanan *et al*., 2017). Our work uses ResNet (He *et al*., 2016) for constructing the feature extractor. As illustrated in Figure ٢, we have used STANet as an FCN-like feature extractor, which is based on ResNet-18 (He *et al*., 2016).

As mentioned, the STANet network consists of a convolutional deep learning (CNN) algorithm (Ji *et al*., 2019), an advanced ResNet (He *et al*., 2016), and a combination of several modules. In other networks, bi-temporal images are normally coded separately, while in this network, a self-attention change detection mechanism is used. 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 has also been improved to some extent. The STANet performs in 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 (Chen & Shi, 2020); and 3) Check changes by metric module as shown Figure 2. The STANet is more stable to overfitting because it uses ResNet in its architecture.

After extracting the basic features, more distinct and detailed features are extracted from those features with the help of the spatial-temporal attention of the PAM module. The PAM module consists of a combination of two mechanisms: the BAM spatial-temporal attention module and the pyramid structure. The BAM module uses a global spatial-temporal relation to obtain more distinct features. This module is easily integrated with a deep Siamese fully convolutional network for change detection.

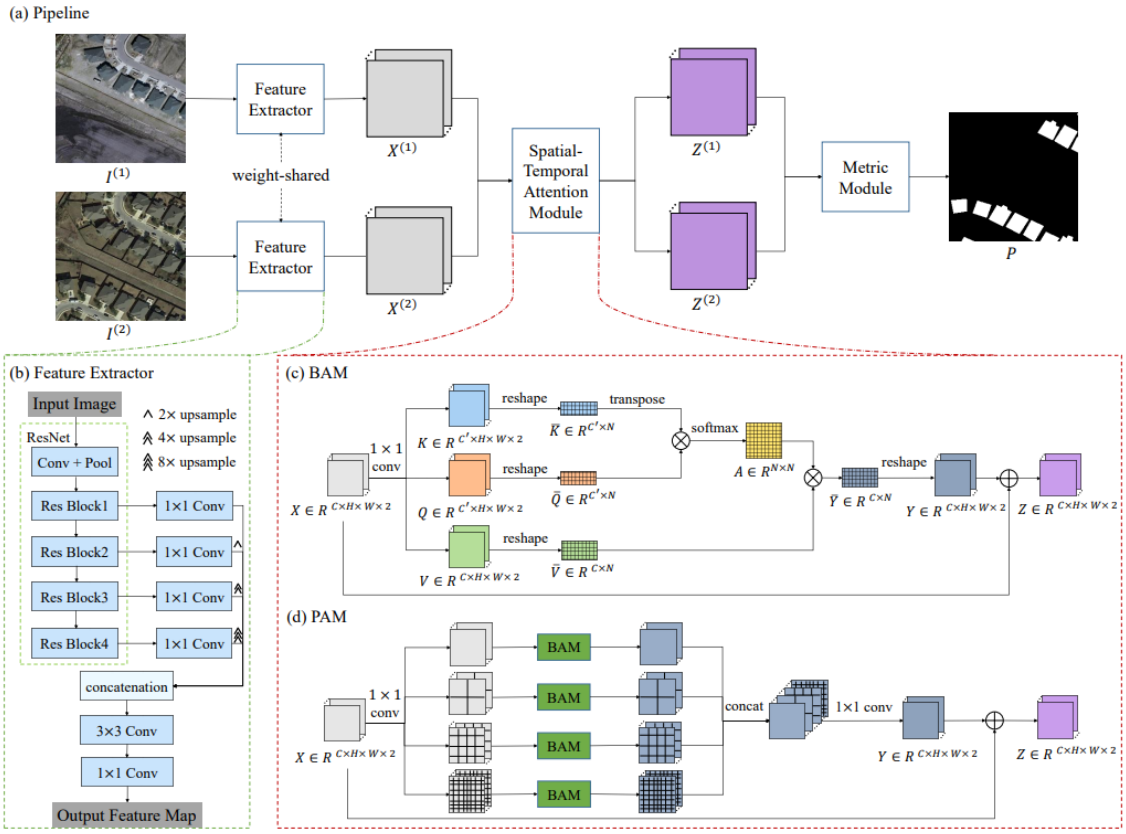


Figure 2. The STANet model: a) STANet with two self-attention BAM and PAM modules, b) Feature extractor, c) Basic spatial–temporal attention module (BAM), d) Pyramid spatio–temporal attention module (PAM) (Chen & Shi, 2020; Zhou et al., 2022).

Attention in deep learning can be interpreted as a vector of importance weights. The Attention module is designed to improve network results. The attention mechanism, which originates from the human visual system, models the dependencies between input and output sequences and has been used in various tasks such as neural network machine translation, image description, and scene parsing. The attention module makes it easier to extract convolutional neural network features. This module is used to learn and focus more on non-useful background information using convolutional algorithms. The attention mechanism enables the network to learn where or what to pay attention to by explicitly establishing the interdependencies of weighted spatial regions (Liu *et al*., 2020).

The attentional feature map can make full use of the interdependencies within channels and locations and is combined with a shortcut connection operation with the input feature map, which improves the feature detection ability without losing information (Zhang *et al*., 2020). The attention module consists of a simple 2D convolutional layer, MLP, with a Softmax function at the end to create a mask of the input feature map. This module takes a feature map of size W×H×C as input and outputs a 1D attention map of W×H×1 or a 3D attention map of W×H×C. W and H are the width and length of the feature map, and C is the number of channels in the feature map. This attention map is then multiplied element-by-element with the input feature map to produce a more accurate and significant output. In general, attention mechanisms are applied in two modes for spatial and channel dimensions. The two attention mechanisms (spatial and channel attention map production) are performed sequentially or in parallel. Figure 3 shows the general structure of a Convolutional Block Attention Module.

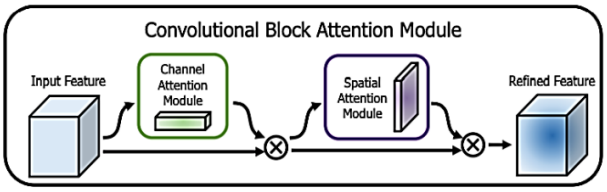


Figure 3. Spatial and channel attention modules can be applied sequentially or in parallel (Woo *et al*., 2018).

Pyramidal attention module means that each pixel in the image space is involved in the mechanism of attention to itself in sub-regions at different scales. It can be imagined that these sub-regions are arranged from small to large, just like the structure of a pyramid. The pyramid module introduced in this network has four parallel branches. Each branch of the feature tensor equally divides into $s \times s$ sub-regions, where s in the subsets (1, 2, 4, 8) defines four pyramidal scales (Chen & Shi, 2020).

In addition to the LEVIR-CD dataset, and in order to improve the training process, the training datasets of buildings change generated from Shamsabad and Mobarakeh industrial parks were used to train the STANet network. This step has been taken to implement the model and analyze the results. First, data preparation, including geometric corrections, change label generation, and image patch generation, has been done.

The ground-truth land change image was generated using

the two industrial parks images. It is a binary image as change map or label where white color represents the areas of change and black color represents the areas of no-change. This step has been done by visually comparing the bi-temporal images and changing the pixels of newly built buildings to white and leaving the other areas as black. In order to generate change labels, time series images were used and by searching region by region in the images, changes have been detected. In order to validate the changed areas that were visually complicated, they were input into the Google Earth software and the time of images of that area were checked. This review increased the accuracy in detecting changes that were difficult to detect in terms of pixel color and shadows. Then, color of changed buildings, were converted to white and areas of no building change were converted to black. Then, the training of STANet network performed and hyperparameter were estimated.

In order to train the STANet network, input images with specific dimensions should be input to the network. Many dimensions were examined. It was shown that in dimensions smaller than 32×32 pixels, the amount of information for analysis is too little, and the dimensions larger than 1024×1024 pixels slow down the processing speed. The smaller the size, the less RAM is used for processing, which increases the processing speed. As shown in "Table 1", the dimensions of the images are larger than the dimensions defined for the input. Therefore, the 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.

2.2.1 Network Training

The network training was performed by setting the algorithm and tuning the hyperparameters. Choosing the right optimization algorithm for the deep learning model is very important and has a great impact on the time it takes to reach the desired result. The way neural network optimization methods work can be considered as a function of a chain of mountains and valleys, where each point in this function represents different weights of the network. The network error is less if the valley is deeper. This concept is called the loss perspective. Searching from a lossy perspective to find the deepest valley in the global minimum can be defined as gradient minimization optimization. The direction of movement to decrease with the greatest slope indicates the direction of the gradient, and the amount of the decrease indicates the amount of the gradient. After determining the movement path, the size of the movement step should be determined. It is called the learning rate. In the network learning process, various optimization methods are used to adjust the weights of the network and minimize its errors. Choosing the right optimization algorithm for the deep learning model is very important and has a great impact on the time it takes to reach the desired result. Meanwhile, methods based on gradient reduction, such as Adam (Adaptive Moment Estimation) have been used more than others. In this network, the Adam's optimization algorithm has been 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 (Figure 4), which has recently been widely used for deep learning applications in the field of computer vision (Bashetty *et al*., 2022).

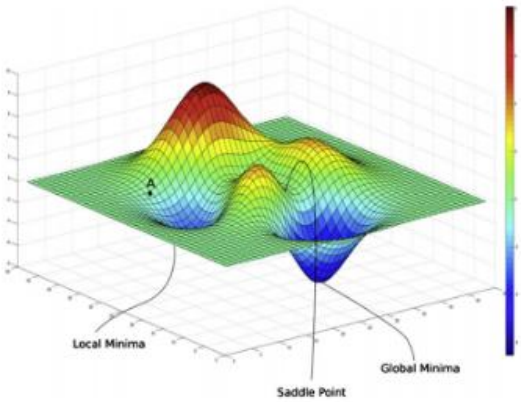


Figure 4. Stochastic Gradient Descent (SGD) (Bashetty *et al*., 2022).

In order to perform the hyperparameters setting in the training process, the following steps have been done. First, the augmentation which is increasing the amount of the training dataset has been done. When the amount of data increases, it leads to an increase in accuracy. The STANet created a set of extra data out of the existing data by rotating and cropping the images. The rotation was applied as a random rotation from -15° to 15°. Then, the whole training process was set to 200 epochs. The entire dataset was trained at each epoch. Results from each epoch were compared to the previous epoch, and the training process improved by resetting the new learning parameters.

Learning rate determination is considered as a hyperparameter. The learning rate controls how much the model changes in response to the estimation error each time the model weights are updated. If the learning rate is chosen too low, the algorithm may get stuck in local minima. If a too 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.

When data is sent in batches, it makes the model more comprehensive than a model that receives all data at once. The optimal batch size by means of trial and error was set to 6 due to the limitation of processor power. The STANet network was trained and tested in three stages, as shown in Figure 5.

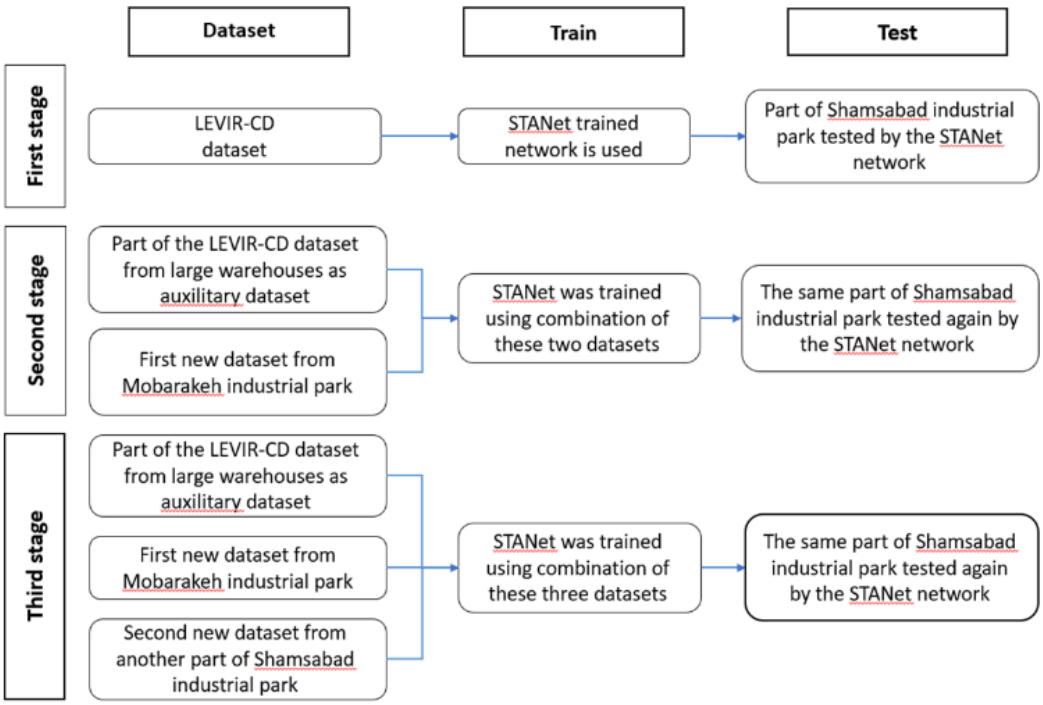


Figure 5. The stages of the performance: In the first stage, the network was trained using LEVIR-CD data and tested on Shamsabad test data. In the second and third stages, it was trained with including two local datasets and tested.

Stage 1: The STANet network was first trained by the LEVIR-CD dataset, and then was tested using the local dataset of Shamsabad to determine the capability of the network as well as the capability of the LEVIR-CD training dataset.

Stage 2: The local training dataset of Mobarakeh industrial park (Figure 6) were combined with a selection of warehouses extracted from the LEVIR-CD dataset. The training dataset of Mobarakeh industrial park contains 117 images of 256×256 pixels. The STANet network was trained using this dataset. 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. The learning process of the network in the first stage in 200 epochs were performed using evaluation metrics of accuracy, precision, F1-score, and recall. It was then tested on a part of Shamsabad industrial park.

Stage 3: The training dataset of Shamsabad Industrial Park (Figure 7), which contains 128 images with dimensions of 256×256 pixels, was added to the dataset used in the stage 2. By combining these three datasets, a more appropriate dataset was generated. The STANet network was trained again using the combination of this newly formed dataset. The STANet in stage 3 was tested on the same part of Shamsabad industrial park as was used in stage 2. As a result, the importance of adding local training data was revealed.

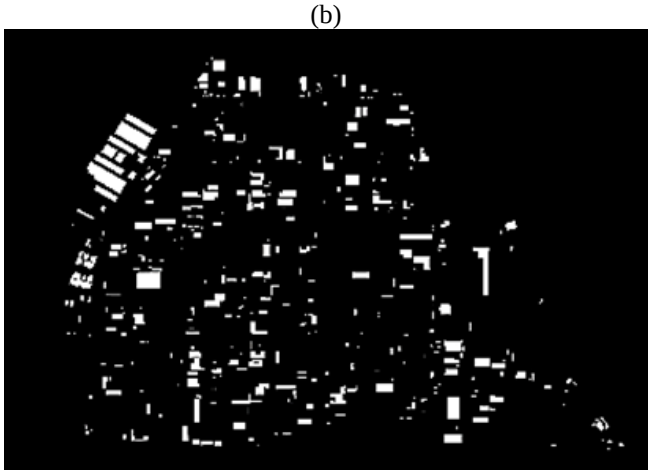
In order to increase the volume of the training dataset, as mentioned, a second local training dataset was produced. This dataset was part of the Shamsabad industrial park. In the third stage, the network training was performed by combining two local training datasets and a selection of large buildings and warehouses from the LEVIR-CD dataset. The training process of the network in this stage in 200 epochs was performed using four evaluation metrics: accuracy, precision, F1-score, and recall.

2.2.2 Network Testing

The test phase was also performed in three stages. To carry out the test phase, after training the network, ready weights were generated so that during the test, only by introducing these weights could the test be 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, a binary image of the change was produced at each stage. However, the initial output image is in the form of small pieces and is very difficult to visually check. In the post-processing stage, the pieces of the images as well as the generated change detection images are combined to form a whole image with the same size as the original image. Figure 8 shows the results of the detected changes in this area.

(a)

(b)



(c)

Figure 6. Training dataset of Mobarakeh industrial park, a) before change, b) after change, and c) the ground-truth change label map in white.

The evaluation of the three test stages by STANet network in Shamsabad industrial park has been done with five metrics of Precision, Recall, F1-Score, Accuracy and κ. The formulas for the metrics are:

$$Precision = \frac{TP}{TP+FP} \qquad Recall = \frac{TP}{TP+FN}$$

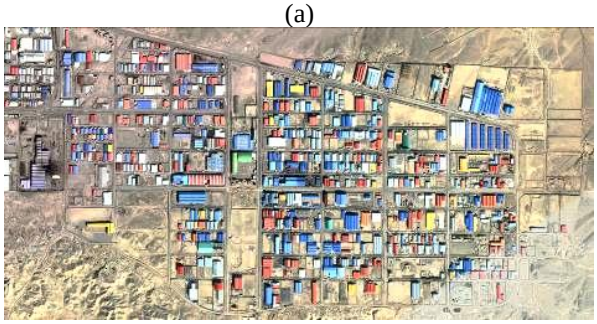
$$F1-Score = \frac{2TP}{2TP+FP+FN}$$

$$Accuracy = P0 = \frac{TP+TN}{TP+TN+FP+FN}$$

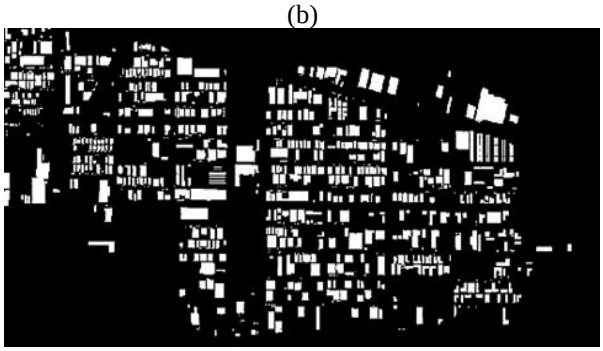
$$PE = \frac{(TP+FN)(TP+FP)}{(TP+TN+FP+FN)^2} + \frac{(TN+FP)(TN+FN)}{(TP+TN+FP+FN)^2}$$

$$\kappa = \frac{P0-PE}{1-PE}$$

(a)



(b)



(c)

Figure 7. Training dataset of Shamsabad industrial park, a) before change, b) after change, and c) the ground-truth change label map in white.

In the first stage test, STANet network that was already trained by LEVIR-CD was tested using a part of the local Shamsabad dataset to assess both the capabilities of the network and the LEVIR-CD dataset. With the help of the shared weights, changes have been detected in the Shamsabad industrial park. The detection of the obtained changes did not lead to an acceptable result (*i.e.* 89% accuracy, 55% precision, 21% F1-Score, and 13% recall) due to the very large difference between LEVIR-CD data and the local Shamsabad industrial park dataset (Table 1).

In the second stage test, the network, which was trained by a combination of the Mobarakeh and LEVIR-CD training datasets, was tested on the local Shamsabad industrial park and obtained a better result than the previous test (*i.e.* 92% accuracy, 86% precision, 52% F1-Score, and 37% recall). However, it has not yet reached acceptable accuracy due to the training that has been done for limited examples of building change detection.

In the third stage test, the network, which was trained by the two previous training datasets (*i.e.* LEVIR-CD, and local Mobarakeh) and combined with Shamsabad training dataset, was tested on Shamsabad industrial park and reached a much better result than the previous two tests (*i.e.* 94% accuracy, 89% precision, 69% F1-Score and 57% recall).

Table 1. Evaluation metrics of the three test stages by the STANet network.

STANet test on Shamsabad industrial park	F1-score	Recall	Precision	Accuracy	Kappa
Trained with LEVIR-CD	21%	13%	55%	89%	0.18
Trained with LEVIR-CD and Mobarakeh	52%	37%	86%	92%	0.49
Trained with LEVIR-CD, Mobarakeh and Shamsabad	69%	57%	89%	94%	0.66

The implementation showed that with the increase of the local training dataset, the results continued to move towards greater accuracy. Figure 9 shows the change detection result obtained from Shamsabad industrial park by STANet.

Figures (a, b), (c, d) and (e, f) in Figure 9 show the changes for the three stages, respectively.

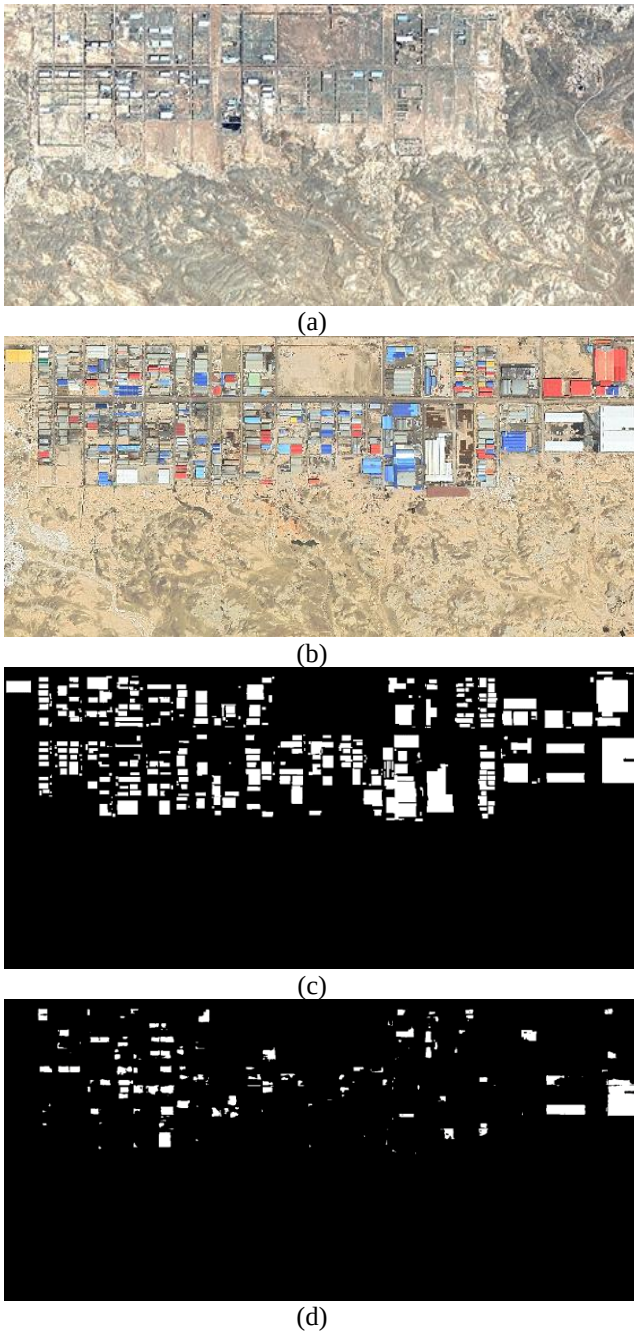
Figures (a, c, and e) in Figure 9 show the ground-truth of change (white) and no-change (black).

Figures (b, d, and f) in Figure 9 show the amount of correctness of the detected change (white), correctly detected no-change (blue), mistakenly detected no-change (red).

3. Discussion

The STANet network was first trained on the LEVIR-CD dataset only. It was then tested using the local dataset of Shamsabad. As mentioned in the previous section, the detection of the changes obtained in the first stage has 89% accuracy, 55% precision, 21% F1-Score, and 13% recall, as shown in Table 1. This stage of testing has less detection and more error than the next two stages. The network at this stage is tuned with training data unrelated to the test data. The buildings in industrial parks are dense and side by side. The roof color of these buildings is very diverse. Also, the color of the roof of the unchanged buildings is shown differently in the before and after change images. Hence, these regions are very different from the training data regions of the LEVIR-CD dataset. As a result, an inappropriate prediction has been obtained in the first stage.

In the second stage test, the network, which was already trained by a combination of the LEVIR-CD and local Mobarakeh training datasets, was tested on the local Shamsabad industrial park and obtained a better result than the previous test, with 92% accuracy, 86% precision, 52% F1-Score, and 37% recall obtained. This increase in accuracy is obtained due to network training by combining two datasets of local training and LEVIR-CD. The images before and after the changes to the first training data set have been selected in such a way that the ratio of changes to no- changes is the same, so that the network pays enough attention to both during learning. At this stage of the test, the network error



(a) (b) (c) (d)

Figure 8. The test area of Shamsabad Industrial Park, a) image before change, b) image after change, c) ground-truth label of changes, and d) result of change detection obtained by STANet for the stage 1 test.

has decreased and the correct detection of changes has increased.

Then, the network is trained with more local data. As a result, in the third stage of the test, the 94% accuracy, 89% precision, 69% F1-Score and 57% recall obtained. There are many building changes in the last training dataset. The color variety of the roofs of the buildings has also increased considerably. For this reason, in this stage of the test, buildings with more color diversity have been detected better, as a result, the accuracy has increased in this stage. Figure 10 shows the results of step by step increasing of the performance in the form of graphs where each bar represents the number of pixels.

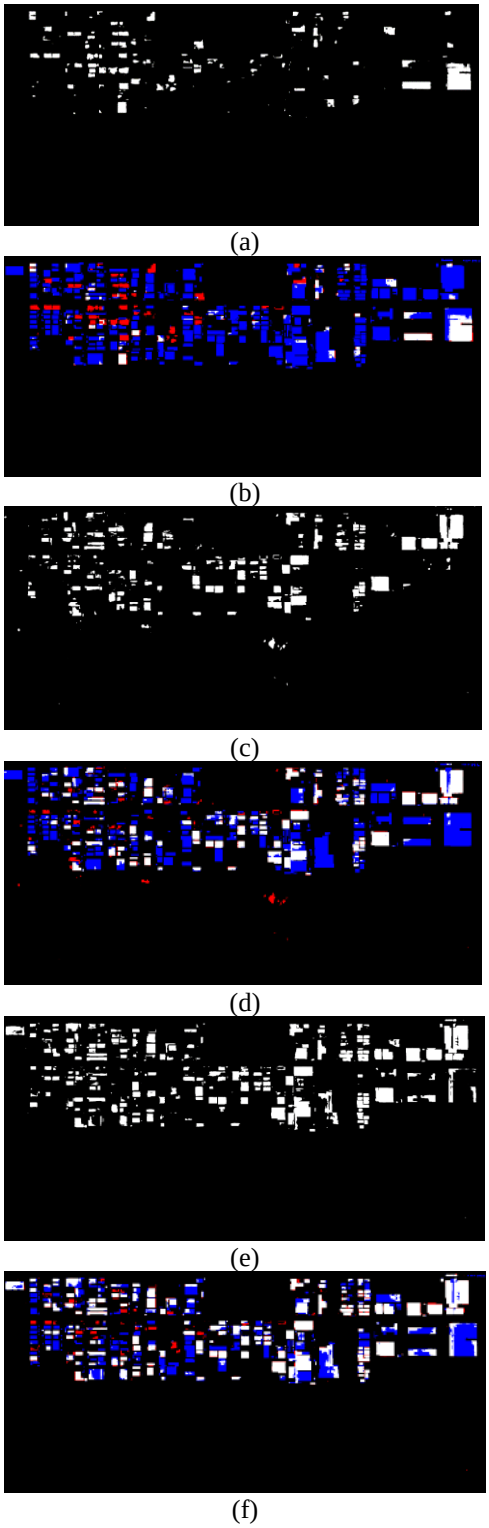


Figure 9. Result of change detection of Shamsabad industrial park. (a, b), (c, d), (e, f) the changes for the three stages, respectively. (a, c, e) ground-truth of change (white) and no-change (black). (b, d, f) the amount of correctness of the detected change (white), correctly detected no-change (blue), mistakenly detected no-change (red).

3.1 Tools and processing environment

The programming code in this research was written in the 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 a free GPU (graphic processing unit), which makes it possible to run heavy 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 the autodiff system. The second library was the Torchvision library, which was used to detect the edges of buildings in the image.

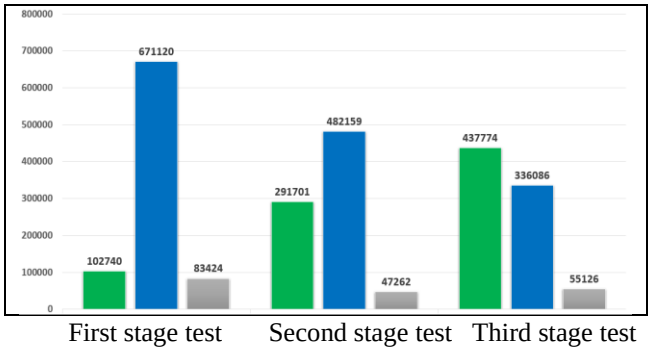


Figure 10. Graph of the change detection results.

4. Conclusion

In this study, it has been shown that when a deep learning model is implemented on an unsuitable dataset, it may produce low-quality results. Training of previously developed networks with newly added local and suitable training datasets increases accuracy. The STANet deep learning network has not been previously used in industrial areas where there are different building structures. The STANet has been implemented using a new dataset comprising an existing dataset and two local building datasets. At the first stage, the STANet network, which has been trained using the LEVIR-CD dataset, was tested on the Shamsabad industrial park. At this stage, the lowest accuracy was achieved with 89% accuracy, 55% precision, 21% F1-Score, and 13% recall. At the second stage, the first training dataset was selected in such a way that the ratio of change to no-change was the same so that the network paid enough attention to both modes during learning. This dataset was combined with a local dataset. The evaluation metrics rates increased to 92% accuracy, 86% precision, 52% F1-Score, and 37% recall due to the training of the network by combining the two datasets. At the third stage, the network was trained with more local data. As a result, in the final stage of the test, the metrics increased to 94% accuracy, 89% precision, 69% F1-Score, and 57% recall. It has been shown in this study that implementing networks on inappropriate datasets does not produce convincing accuracy, but the training of previously developed networks with relevant and local training datasets greatly increases the accuracy. Therefore, the importance of using local corresponding training datasets on networks has been shown.

The link to access to the change detection dataset in industrial buildings in Iran created in this study: https://b2n.ir/b33897 or

https://drive.google.com/drive/folders/1iZyISdBb47tN5WOiS5_5EEEKhR8o3kry?usp=sharing

References

Badrinarayanan, V., Kendall, A. & Cipolla, R. (2017). Segnet: (2017), A deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell., 39, pp. 2481–2495.

Bashetty, S., Raja, K., Adepu, S. & Jain, A. (2022). Optimizers in Deep Learning: A Comparative Study and Analysis, International Journal for Research in Applied Science & Engineering Technology (IJRASET), Vol. 10, Issue XII, Dec 2022, pp. 1032-1039.

Benedek, C., & Szirányi, T. (2008). A mixed Markov model for change detection in aerial photos with large time differences. In 2008 19th International Conference on Pattern Recognition, IEEE, pp. 1-4.

Benedek, C., & Szirányi, T. (2009). Change detection in optical aerial images by a multilayer conditional mixed Markov model. IEEE Transactions on Geoscience and Remote Sensing, 47(10), pp. 3416-3430.

Chandra, N., & Vaidya, H. (2022). Building detection methods from remotely sensed images. Current Science, pp. 1252-1267.

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.

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K. & Yuille, A.L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans. Pattern Anal. Mach. Intell., 40, pp. 834–848.

Darvishzadeh, R. (2000). Change detection for urban spatial databases using remote sensing and GIS. In Proceedings of ISPRS, Vol. 320.

Daudt, R. C., Le Saux, B., Boulch, A., & Gousseau, Y. (2018). Urban change detection for multispectral earth observation using convolutional neural networks. In IGARSS -2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 2115-2118.

He, K., Zhang, X., Ren, S. & Sun, J., (2016). Deep Residual Learning for Image Recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, pp. 770-778.

Huang, L., Xiang, S., & Zheng, J. (2022). Fine-Scale Monitoring of Industrial Land and Its Intra-Structure Using Remote Sensing Images and POIs in the Hangzhou Bay Urban Agglomeration, China. Int. Journal of Environmental Research and Public Health, 20(1), 226.

Internet communication, https://gis.stackexchange.com/questions/11395/spatial-resolution-of-google-earth-imagery.

Ji, S., Shen, Y., Lu, M., & Zhang, Y., (2019). Building instance change detection from large-scale aerial images using convolutional neural networks and simulated samples, Remote Sensing, 11 (11), 1343.

Liana-Eugenia, M., and Nicoleta-Georgeta, B. (2013). The Role of Industrial Parks in Economic Development, The Annals of the University of Oradea, 123.

Liu, Y., Pang, C., Zhan, Z., Zhang, X. & X. Yang, (2020). Building change detection for remote sensing images using a dual-task constrained deep Siamese convolutional network model, IEEE Geoscience and Remote Sensing Letters, 18(5), 811-815.

Long, J.; Shelhamer, E.; Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015, pp. 3431–3440.

Lu, D., Mausel, P., Brondizio, E., & Moran, E. (2004). Change detection techniques. International journal of remote sensing, 25(12), pp. 2365-2401.

Lu, Z., H. Pu, F. Wang, Z. Hu & L. Wang, (2017). The Expressive Power of Neural Networks: A View from the Width, Proceedings of 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, pp. 1-9.

Shen, L., Lu, Y., Chen, H., Wei, H., Xie, D., Yue, J., & Jiang, B. (2021). S2Looking: A satellite side-looking dataset for building change detection. Remote Sensing, 13(24), 5094.

Su, E., Siqi Cai, Longhan Xie, Haizhou Li, Tanja Schultz, (2022). STAnet: A Spatiotemporal Attention Network for Decoding Auditory Spatial Attention From EEG, IEEE Trans Biomed Eng. 69(7), pp. 2233-2242.

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 Geosci. Remote Sens. Lett., 19, pp. 1-5.

Woo, S., Park, J., Lee J.Y. & Kweon, I.S. (2018). CBAM: Convolutional Block Attention Module, arXiv, v2, 2018, pp. 1-17.

Wu, C., Yang, L., Guo, C. & Wu, X. (2024). High-Resolution Remote Sensing Image Change Detection Based on Cross-Mixing Attention Network, Electronics, 13, 630, pp. 1-15.

Zhang, J., Pan, B., Zhang, Y., Liu, Z., & Zheng, X. (2022). Building change detection in remote sensing images based on dual multi-scale attention. Remote Sensing, 14(21), 5405.

Zhang, J., Zhou, Q., Wu, J., Wang, Y., Wang, H., Li, Y., Chai, Y. & Liu, Y. (2020). A Cloud Detection Method Using Convolutional Neural Network Based on Gabor Transform and Attention Mechanism with Dark Channel Subnet for Remote Sensing Image, Remote Sensing J. (MDPI), 12, 3261, pp. 1-19.

Zhou, S., Dong, Z. & Wang, G., (2022). Machine- Learning-Based Change Detection of Newly Constructed Areas from GF-2 Imagery in Nanjing, China, *Remote Sens.* 14, 2874.

Zhu, L., Shimamura, H., Tachibana, K., Li, Y., & Gong, P. (2008). Building change detection based on object extraction in dense urban areas. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences., Commission VI, WG VII/5.