Comparison of Plain and Dense Skip Connections on U-Net Architecture for Change Detection

Zamfirdaus Saberi(✉) and Noramiza Hashim

Multimedia University, Persiaran Multimedia, 63100 Cyberjaya, Selangor, Malaysia
mohdzamfirdaus99@gmail.com

Abstract. In recent years, identifying changes in multitemporal images in terms of land use and land cover is significant in a variety of applications including urban planning. CNN architectures are one of the most extensively utilised methods for change detection. The aim of this research is to investigate two types of skip connections that may be incorporated into CNN architecture to determine if they can improve the effectiveness of change detection during the CNN learning process. In this paper, we adopt the U-Net architecture to train the change detection model. We also modify the U-Net skip connection's path to include the dense skip connection and compare the modified U-Net with the original U-Net, which uses the plain skip connection. We also test the trained model with our collected local dataset in Cyberjaya to see how well it can anticipate changes in our location. The results of this study show that a U-Net with dense skip connections produces the best results and optimises change detection. It will help researchers understand how important the skip connection is to the model's performance.

Keywords: U-Net · Skip Connection · Change Detection · CNN

1 Introduction

The method of analyzing differences in the state of a matter or aspect through examining it over time is known as change detection [2]. In general, the concern is determining whether a change has occurred, or whether multiple changes have occurred, as well as determining the times of any such changes. Multi-temporal remote sensing data, such as aerial imaging and satellite imagery, can be incorporated to classify land use and land cover (LULC) changes at a specific location across time. Change detection, which necessitates finding differences in two images, is a common use of remote sensing data where the images were taken at two different times. The changes discovered will then produce a change map. A change map is made up of the changes that have been recognized. Figure 1 shows the general flowchart for change detection.

Change detection techniques have been applied in a variety of ways, depending on the application. In recent years, artificial intelligence.

algorithms, specifically convolutional neural networks (CNNs), have had a significant effect [3]. Additionally, CNN architectures for supervised change detection have

S.-C. Haw and K. Sonai Muthu (Eds.): CITIC 2022, 10, pp. 524–532, 2022.
https://doi.org/10.2991/978-94-6463-094-7_42

been suggested. Several authors have proposed convolutional neural networks (CNNs) for change detection in recent years. According to Liu et al. [4], U-Net is a convolutional network architecture that has been demonstrated to be effective at segmenting images semantically. It also works well for small number of training images [5]. There are several works that use U-Net in their change detection in multi-temporal satellite images which is Ahangarha et al. [6], and Jong et al. [7]. U-Net includes two paths: encoder and decoder. The encoder is effective at extracting features and spatial information, whereas the decoder is great at localising features.

However, with the advancement of deep neural networks, their architecture becomes more sophisticated, bringing a new issue as performance increases: When input or gradient information reaches the end of the network via multiple layers, it will vanish and "wash out," a phenomenon known as gradient vanishing [8]. According to the author, there have been some remarkable efforts to address this issue, including the use of data bypass or skip connections. The central concept is to establish a cross-layer link between the layers. The signal is passed from one layer to the next, both before and after the layer in order to optimise the flow of information between all of the network's layers and also to improve the stability of the model. According to [5], even U-Net incorporate skip connection, it has the difficulty to train through the layer due to the gap between from decoder to encoder.

According to Wang et al. [10], there are different types of skip connection scheme which includes plain and dense skip connections. The U-Net architecture use plain skip connection. Our work aims to investigate the effect of incorporating a different skip connection, which is dense, into the U-Net architecture. This is a promising method for training models that can extract rich spatial information and increase learning performance during model training.

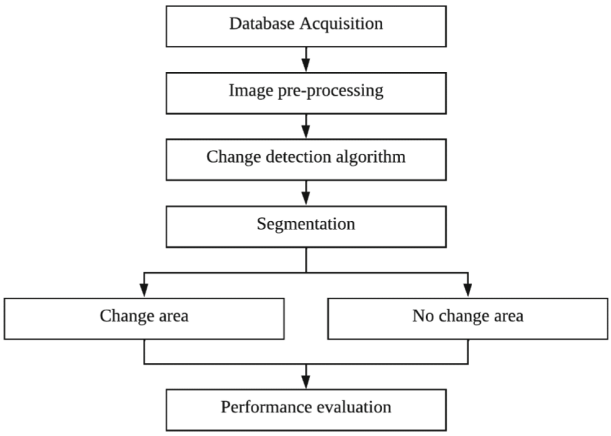


Fig. 1. General flow for change detection method [1]

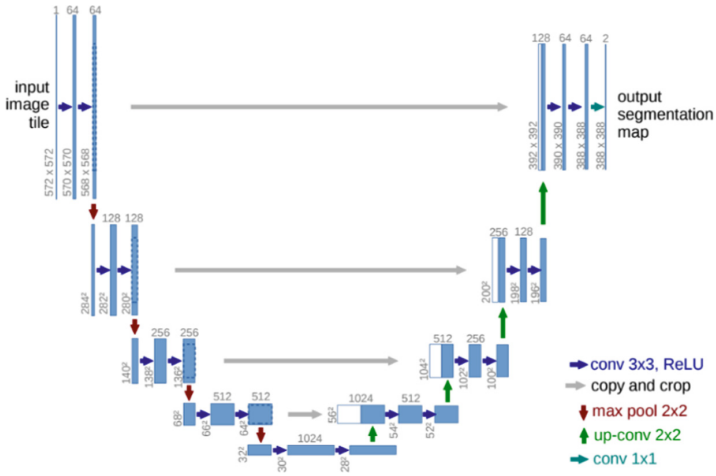


Fig. 2. U-Net architecture (example for 32 × 32 pixels in lowest resolution) [9]

2 Methodology

In this section, we propose the network architecture where we will modify the U-Net [8] by constructing and incorporating the dense skip connection into the U-Net that will be nested. It is different from U-Net, which has a denser architecture with numerous skips' connection pathways. For this project, two types of skip connection will be compared, namely plain and dense.

2.1 Network Architecture

Figure 2 shows the U-Net architecture where it uses the plain skip connection. This architecture has two paths: encoder and decoder. The encoder is the first path that is used to retrieve the image's background information [6]. The encoder reduces the spatial dimensions of each layer while increasing the number of channels used to extract the features. The encoder is composed of stacked max-pooling layers and a simple convolution. The Max-pooling layer is used to reduce the dimensionality of the convolutional layer's feature map [11]. There are ten 3 × 3 convolutional layers followed by a ReLU activation function and four max-pooling layers of 2 × 2 make up the encoding part. We increase the number of feature channels while halving the spatial dimensions with each downsampling step.

The next path is the decoder, which is used to localise features through transpose convolution. The spatial dimensions are increased while the channel number is decreased. Each layer in the decoder is composed of an upsampling of the feature map followed by 22 transpose convolutions, resulting in a channel number of half. The feature maps will next be concatenated with their respective counterparts in the encoding section by using plain skip connection, followed by 3 × 3 convolutional neural networks (each followed by a ReLU).

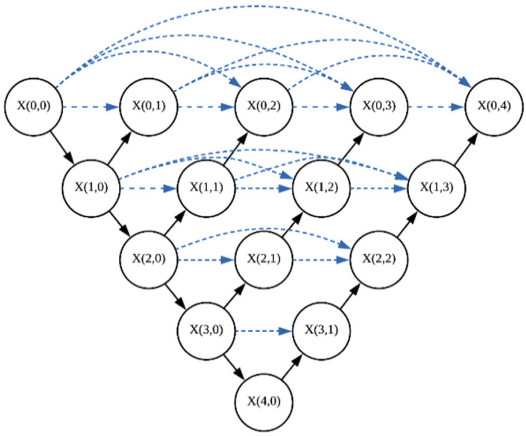


Fig. 3. Nested U-Net with dense skip connection

From the U-Net, a nested U-Net with a dense skip connection is formed, as seen in Fig. 3. The original U-Net which use plain skip connection method is modified where more skip connection paths are added into the architecture and we add more nodes into the architecture so that it can allow us to design a dense connection.

2.2 Loss Function

For the loss function for this model, cross-entropy is used. The use of cross-entropy is suggested by [12] as it solves the problem of the imbalance of binary class samples in the dataset. It is used when modifying model weights during training. The goal is to reduce the loss as much as possible; the less the loss, the better the model. The cross-entropy loss of a perfect model is zero. The formula is defined as:

$$L_{CE} = -\sum_{i=1}^{n} t_i \log(p_i), \text{ for n classes,} \tag{1}$$

where t_i is the truth table and p_i is the Softmax probability of ith class.

3 Experimental Setup

Two models will be trained for the experiment, notably the nested U-Net with plain and dense skip connections. Separately, these two models will be trained using the same parameters. The outcome of the model training will be examined in more detail in the following section.

Table 1. Details of data received from MYSA

Location name	Cyberjaya
Coordinates	upper left latitude: 2.9363 upper left longitude: 101.6265 lower left latitude: 2.8989 lower right longitude: 101.6601
Years	2017, 2018, 2021
Satellite	IKONOS
Resolution	1m/pixel
Format	JPEG 2000 Core Image File (.JP2)
Size	512 x 512 pixel
Format	JPEG 2000 Core Image File (.JP2)

3.1 Dataset

To train the model, the LEVIR-CD dataset will be used. The dataset is selected because it contains a pair of multi-temporal images along with their respective ground truth. There are a total of 637 pairs of high-quality (0.5 m/pixel) Google Earth image patch pairs with a 1024 $\times$ 1024 pixel resolution. These bitemporal images has changes in land use, notably in terms of development growth. LEVIR-CD is applicable to a wide variety of construction types, including villas, high-rise apartments, teeny-tiny garages, and massive structures.

Additionally, we will validate our trained model using satellite images obtained in the local area of Cyberjaya. We obtain the data from Malaysian Space Agency (MYSA). This is to determine the model's robustness against new data. Table 1 summarizes the data received from MYSA.

4 Result and Discussion

The results of model training and testing on the LEVIR-CD dataset are shown in Table 2. From observation, nested U-Net with dense skip connections had a slightly higher accuracy than other models, with a value of 0.98212. The accuracy of a prediction is obtained by dividing the total number of predictions by the total number of true predictions (true positives+true negatives). However, when working with class-imbalanced data sets, accuracy alone is insufficient as a criterion for evaluation. Therefore, additional measures are used to evaluate the trained models' performance.

U-Net with a dense skip connection achieved the greatest F1 score of 0.56148, which reflects the balance average of recall and precision values, implying that it delivers great value for both. For recall, it illustrates the model sensitivity, which is used to quantify the number of positive class predictions made from all the positive classes in the dataset. Meanwhile precision shows the model's capacity to quantify the number of positive class predictions that are genuinely positive class predictions. U-Net with a dense skip

Table 2. Result of model training and testing

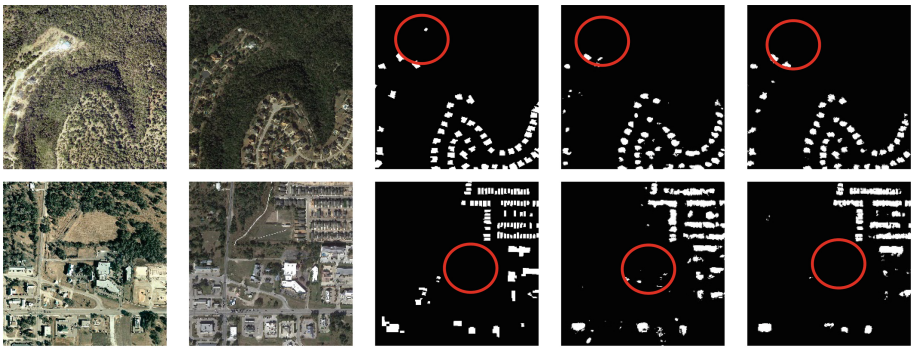
Skip connection scheme	Plain	Dense
Total parameters	1,941,537	9,057,633
Accuracy	0.97058	0.98212
F1 Score	0.56148	0.65008
Recall	0.58483	0.59995
Precision	0.56948	0.74835
Intersection over Union (IoU)	0.61699	0.70622
Average processing time (s)	0.07100	0.13400

connection achieves the maximum score of 0.70622 for Intersection over Union (IoU), indicating that the model is capable of predicting changes as comparable as the ground truth. On the other hand, the average processing time for a model with a plain skip connection is faster at 0.07100 s than a dense skip connection at 0.13400 s. This is due to the total parameters for dense skip connection being large, with a total parameter of 9,057,633.

Figure 4 illustrates the prediction results of model testing using trained models on LEVIR-CD. As a result of the observation, it is clear that U-Net with dense skip connections in Fig. 4g has a robust prediction change map. It is because of the model's capability to detect changes in buildings similar to those observed in the ground truth. It can be seen from the prediction result where it can predict true positive of the changes. On the other hand, U-Net with a plain skip connection can also predict in general, but it is less accurate at detecting areas of change than others. This is because the plain skip connection has areas of change that cannot be recognized adequately, causing false negatives to be reported.

The trained model will be evaluated using the dataset gathered locally in Cyberjaya. This will determine how effectively it can anticipate a change in a building's structure when presented with new and random data. The best model will be chosen and applied for this experiment, which will feature U-Net with a dense skip connection. 6 pairs of images from the local dataset will be used in a quantitative investigation in which the ground truth for 6 pairs of images where the building has changed will be manually annotated.

Table 3 shows the results of trained model testing on the local Cyberjaya dataset. If viewed quantitatively, it shows that the dense skip connection method gives better results where the performance of the model gets a high value. The accuracy and F1 score of U-Net using dense skip connection are higher than those using plain skip connection. It shows the effect of a good dense skip connection where it has more skip connection paths when compared to a plain skip connection. So, it allows the model to learn and predict better when using a dense skip connection. The IoU for dense is also high, with a value of 0.3864, which indicates the ability of the model to predict new and random data.



(a) Image 1 (b) Image 2 (c) Ground truth (d) Plain (g) Dense

Fig. 4. Prediction result of model testing on LEVIR-CD using U-Net with different skip connections

Table 3. Result of model testing on the local dataset

Skip connection scheme	Plain	Dense
Accuracy	0.7214	0.8128
F1 Score	0.8614	0.9257
Recall	0.8628	0.9771
Precision	0.8614	0.8828
Intersection over Union (IoU)	0.3364	0.3864

5 Conclusion

Our proposed U-Net with nested dense skip connection provides an improvement to the result change detection where a dense skip connection path is added into the U-Net in order to allow the architecture to learn more robustly. The application of dense skip connection has increased the effectiveness of change detection in multi-temporal satellite images and can produce a more precise change map. It also confirmed the statement by Chen and Qi [8] that using data bypass or skip connection can improve the stability of the model. Modification of U-Net in this study, which was originally proposed by Olaf et al. [8], can provide insight to another researcher to use skip connection to improve the performance of the deep learning model, especially in the field of change detection.

Acknowledgments. The authors would like to acknowledge Multimedia University Internal Research Fund (IRFund) for the financial support of the project.

Authors' Contributions. Methodology: Zamfirdaus Saberi and Noramiza Hashim; Formal analysis: Zamfirdaus Saberi; writing—original draft paper: Zamfirdaus Saberi; Revision: Noramiza Hashim.

References

1. Y. Chen, G. Liu, and H. Chen, "Multi-temporal remote sensing image registration based on multi-layer feature fusion of deep residual network," *ICIIBMS 2019 - 4th Int. Conf. Intell. Informatics Biomed. Sci.*, pp. 363–367, 2019, doi: https://doi.org/10.1109/ICIIBMS46890.2019.8991506.
2. W. Shi, M. Zhang, R. Zhang, S. Chen, and Z. Zhan, "Change detection based on artificial intelligence: State-of-the-art and challenges," *Remote Sens.*, vol. 12, no. 10, 2020, doi: https://doi.org/10.3390/rs12101688.
3. R. Caye Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, "Multitask learning for large-scale semantic change detection," *Comput. Vis. Image Underst.*, vol. 187, no. October 2018, p. 102783, 2019, doi: https://doi.org/10.1016/j.cviu.2019.07.003.
4. J. Liu *et al.*, "Convolutional Neural Network-Based Transfer Learning for Optical Aerial Images Change Detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 1, pp. 127–131, 2020, doi: https://doi.org/10.1109/LGRS.2019.2916601.
5. T. Ando and K. Hotta, "Cell image segmentation by Feature Random Enhancement Module," Jan. 2021, Accessed: Jun. 13, 2022. [Online]. Available: http://arxiv.org/abs/2101.07983.
6. M. Ahangarha, R. Shah-Hosseini, and M. Saadatseresht, "Deep Learning-Based Change Detection Method for Environmental Change Monitoring Using Sentinel-2 Datasets," *Environ. Sci. Proc.*, vol. 5, no. 1, p. 15, 2020, doi: https://doi.org/10.3390/iecg2020-08544.
7. K. L. De Jong and A. Sergeevna Bosman, "Unsupervised Change Detection in Satellite Images Using Convolutional Neural Networks," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2019-July, Dec. 2018, doi: https://doi.org/10.48550/arxiv.1812.05815.
8. C. Chen and F. Qi, "Single Image Super-Resolution Using Deep CNN with Dense Skip Connections and Inception-ResNet," *Proc. - 9th Int. Conf. Inf. Technol. Med. Educ. ITME 2018*, pp. 999–1003, Dec. 2018, doi: https://doi.org/10.1109/ITME.2018.00222.
9. T. B. Olaf Ronneberger, Philipp Fischer, "U-Net: Convolutional Networks for Biomedical Image Segmentation," 2015.
10. Z. Wang *et al.*, "Multi-memory convolutional neural network for video super-resolution," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2530–2544, 2019, doi: https://doi.org/10.1109/TIP.2018.2887017.
11. V. Christlein, L. Spranger, M. Seuret, A. Nicolaou, P. Kral, and A. Maier, "Deep generalized max pooling," *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR*, pp. 1090–1096, Sep. 2019, doi: https://doi.org/10.1109/ICDAR.2019.00177.
12. Y. Li, W. Shi, G. Liu, L. Jiao, Z. Ma, and L. Wei, "SAR Image Object Detection Based on Improved Cross-Entropy Loss Function with the Attention of Hard Samples," pp. 4771–4774, Oct. 2021, doi: https://doi.org/10.1109/IGARSS47720.2021.9554061.