FUSION OF MULTISPECTRAL IMAGERY AND DIGITAL SURFACE MODEL FOR BUILDING CHANGE DETECTION

Pei-Cheng Chen[a], Tee-Ann Teo[b]

[a] Ph.D. Student, Dept. of Civil Engineering, National Yang Ming Chiao Tung University, Hsinchu, Taiwan 300093.
[b] Professor, Dept. of Civil Engineering, National Yang Ming Chiao Tung University, Hsinchu, Taiwan 300093.

ABSTRACT
Urbanization often involves construction and deconstruction. In map updating processing, most procedures focus on the changed areas instead of reproducing the entire map. The automatic building change detection technology efficiently locates the change areas and can be utilized in map updating. The airborne multispectral images and digital surface model (DSM) provide useful features to detect building regions. This study proposes a UNetFormer fusion approach to detect building change areas. We also analyze the impact of different input features. In the evaluation, we use pixel-wise evaluation and object-based evaluation. Combining all the data, the best combination reached 99% of F1-Score in the overall pixel-wise assessment and 93% in object-based evaluation. The result demonstrated the integration of spectral, height, and old map building area information can detect the building area effectively.

Keywords: Building Change Detection, Multispectral Images, DSM, Data Fusion.

1. INTRODUCTION

In rapidly evolving urban environments, change occurs at an accelerated pace, characterized by the replacement of aging structures with new constructions. Consequently, topographic maps quickly become outdated, necessitating frequent updates. When addressing the task of map updating, it becomes imperative to identify areas needing high-priority updates and those requiring immediate attention. Recent advancements in remote sensing and deep learning have introduced several innovative techniques that can significantly enhance the efficiency of the mapping process.

Previous research on building change detection can be broadly classified into three categories by data source: spectral imagery (Tan et al., 2016), digital surface models (DSM) (Lyu et al., 2020), and data fusion (Su et al., 2020). Spectral imagery encompasses color and texture features. Applying a Mask R-CNN-based model to aerial images has demonstrated the capability to detect building changes (Ji et al., 2019). However, aerial or satellite imagery with projective projection often encounters challenges such as relief displacement and shadows. The DSM approach directly conveys building height information through the difference between surface and ground heights. Consequently, areas of building change can be identified by comparing bitemporal DSMs (Warth et al., 2019).

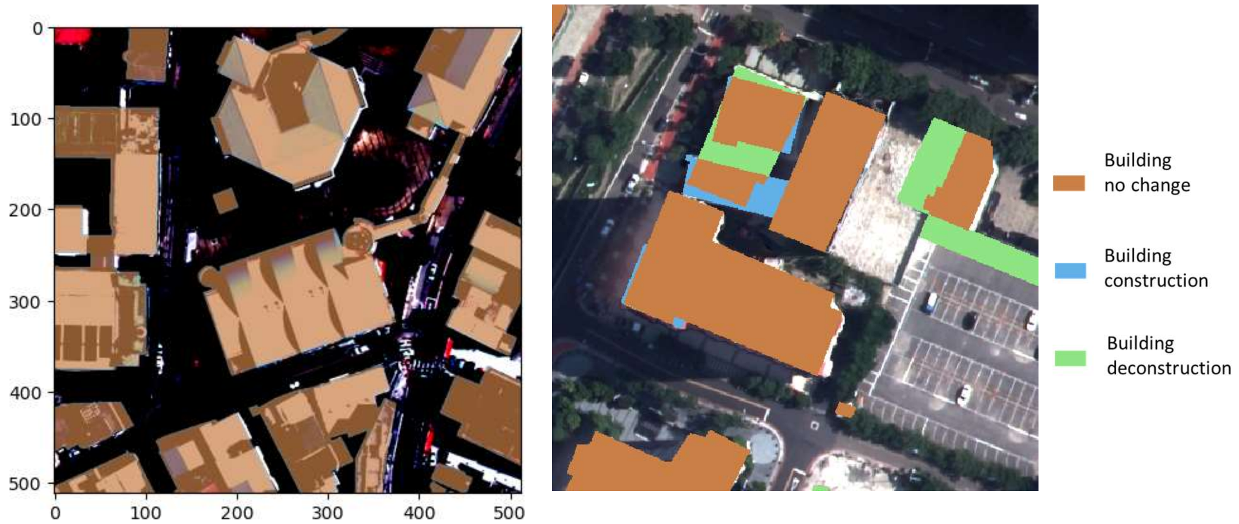
The fusion method combines spectral and height information, complementing each other. Furthermore, the inclusion of an old building map provides insight into previous building states, enhancing the accuracy of change detection (Su et al., 2020). Su et al. (2020) employed an UNet-based encoder-decoder network and integrated three distinct data sources: new and old ortho images, height variations, and old map images. Comparative analysis of various input combinations revealed that utilizing all available input data yielded the most favorable outcomes.

As the fusion approach produces better building change detection results, this study aims to adopt the UNetformer (Wang et al., 2022), a UNet-like transformer network, for building change detection tasks. Moreover, we evaluate the efficacy of incorporating DSM and old building maps into the process.

2. MATERIAL

2.1 Research area and data pre-processing

The research area was located in Hsinchu, Taiwan, covering Taiwan eMap 23 frames about 161 km^2. We split 22 frames as train/validation and 1 as an independent test. The multiview aerial images were captured in 2016 and 2018 with spatial resolution 0.25m/pixel. The DSM was generated from these images and the true ortho images at via phogrammetry processing. To compose our building change dataset, every image was clipped into 512x512 pixels patches (Figure 1a) with a 50% lateral and horizontal overlap rate. The entire train/validation dataset 1048575 patches were split into 80% train and 20% validation. The old building map was rasterized from 2016 eMap building layers at the same spatial resolution. Ground truth building label was also rasterized from 2016 and 2018 eMap building layers, then computed building change label (Figure 1b)



(a) True orthoimage with building area overlay (b) Building change label example

Figure 1. True orthoimage with building area overlay and label image example

2.2 Data augmentation in model training

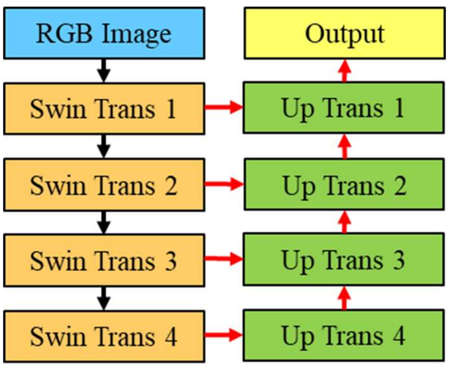
In order to gain more dataset for traning, we designed a data augmentation workflow during model training (Table 1). The augmentation parameters of orthoimage are color and brightness random shift, while the DSM only applies height value augmentation. All the patches were applied the random flip augmentation in both horizontal and vertical direction.

Table 1. Training data augmentation

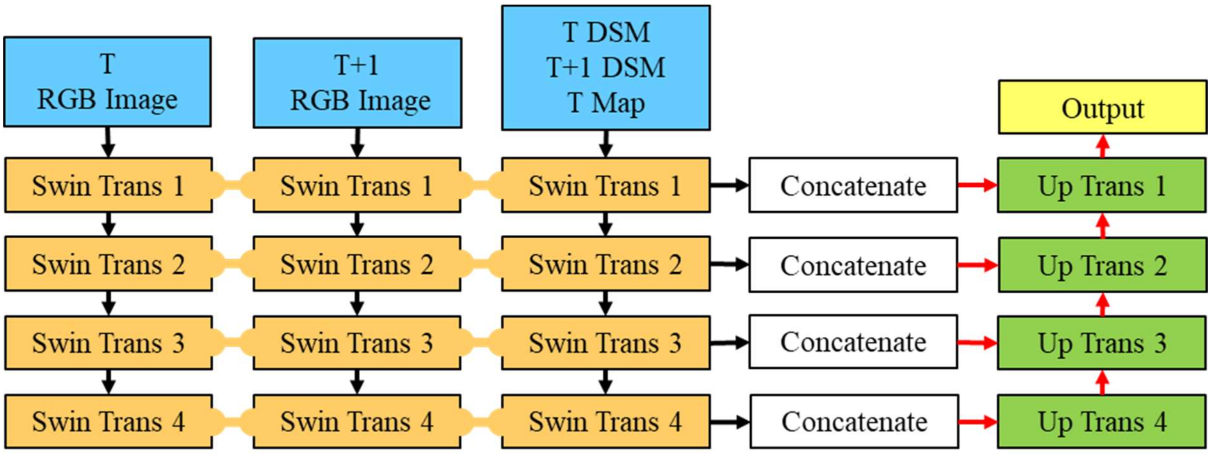
	Random augmentation	Random flip
RGB image	Color / Brightness	Horizontal / Vertical
DSM image	Brightness (height)	Horizontal / Vertical
Old building image	None	Horizontal / Vertical
Label image	None	Horizontal / Vertical

3. METHODOLOGY

The original UNetFormer structure (FTUNetFormer) was designed for a single RGB input image (Figure 2a). We modified the structure by duplicating Swin Transformer (Liu, et al. 2021) encoding layers and made it a three-branch weight-sharing siamese structure (Figure 2b). The transformer siamese structure integrates two images on change detection task (Bandara & Patel, 2022). Our network was designed to fuse different data types and different information, including RGB color, DSM height value, and old building map.



(a) Original UNetFormer Structure



(b) Modified UNerFormer Structure

Figure 2. UNetFormer structure and our modification

3.1 Model training

The computer, equipped with i7 9700KF, 96GB RAM, RTX Titan 24GB, trained the model for 50 epochs about 14 hours. The best model checkpoint was monitored by validation mIoU index. We designed three different combinations, (1)RGB + DSM, (2) RGB + Map, and (3)RGB + DSM + Map (Table 2), to evaluate the performance of different input dataset.

Table 2. Training input combination

	RGB image	DSM image	Building Map
RGB_DSM	√	√	
RGB_Map	√		√
RGB_DSM_Map	√	√	√

3.2 Model prediction

Although the training process used a 512x512 input size, the prediction process could be larger, such as 1024x1024, because Swin transformer was designed to be a multi-scale feature map. Each prediction patch has pixel-wise classification confidence value. We aggregated these values in place and merged the entire prediction image. A single eMap frame prediction took about 3 minutes on GTX 1080Ti.

4. EXPERIMENT RESULTS AND DISCUSSION

4.1 Pixel-wised evaluation

In order to evaluate the difference between the predicted result and the ground truth label, we compared each value in two images at the same pixel. True positive (TP) occurs when the prediction matches the ground truth label. False positive (FP) happens when there is no ground truth label, but the prediction has a value or the prediction value does not matches to ground truth label. False negative (FN) occurs when there is no prediction value, but the ground truth has a label. Precision = TP/(TP+FP), Recall = TP/(TP+FN), F1_Score = 2 x Precision x Recall / (Precision + Recall). Pixel-wise evaluation showed that DSM improved building construction detection (Table 3), while building map improved deconstruction and no change. RGB_DSM_Map successfully fused all information and got better results.

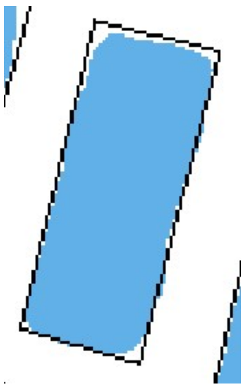
Table 3. Three combination pixel-wised evaluation

RGB_DSM				
	Deconstruction	No Change	Construction	Overall
Precision	76.72%	99.09%	80.86%	99.07%
Recall	42.20%	88.63%	82.83%	98.90%
F1_Score	54.44%	93.57%	81.84%	98.98%
RGB_Map				
	Deconstruction	No Change	Construction	Overall
Precision	46.91%	99.10%	99.47%	99.13%

Recall	98.30%	99.60%	64.83%	99.67%
F1_Score	63.51%	99.35%	78.49%	99.40%
RGB_DSM_Map				
	Deconstruction	No Change	Construction	Overall
Precision	74.08%	99.45%	99.09%	99.74%
Recall	96.93%	98.94%	80.90%	99.74%
F1_Score	83.97%	99.14%	89.08%	99.74%

4.2 Object-based evaluation

Several small change units on the label were too small to distinguish. We established two thresholds to solve the issue. A change unit area must be larger than 9 m^2 and at least IoU > 0.1. An example is shown in Figure 3, where the black line area represents a label change unit, and the blue pixels are prediction results. Equation 1 shows the calculation of IoU. RGB_DSM_Map was the best combination in change unit evaluation (Table 4). Adding an old building Map could improve prediction results more than adding DSM.



label: Construction prediction: Construction (correct)

Figure 3. change unit example

$$\text{IoU} = \frac{TP}{TP+FP+FN} = \frac{3724}{3724+0+363} = 0.91 \quad (1)$$

Table 4. Three combination change unit evaluation

RGB_DSM				
	Deconstruction	No Change	Construction	Overall
Precision	68.52%	91.80%	73.30%	88.79%
Recall	37.76%	90.80%	42.02%	78.85%
F1_Score	48.68%	91.30%	53.42%	83.53%
RGB_Map				
	Deconstruction	No Change	Construction	Overall
Precision	62.65%	98.40%	60.98%	91.27%
Recall	71.23%	99.34%	58.14%	91.27%
F1_Score	66.67%	98.87%	59.52%	91.27%
RGB_DSM_Map				
	Deconstruction	No Change	Construction	Overall

Precision	84.53%	99.34%	72.46%	94.88%
Recall	69.61%	99.49%	59.17%	91.18%
F1_Score	76.34%	99.42%	65.14%	92.99%

4.3 Case Discussion

4.3.1 Omission cases in RGB_DSM_Map model

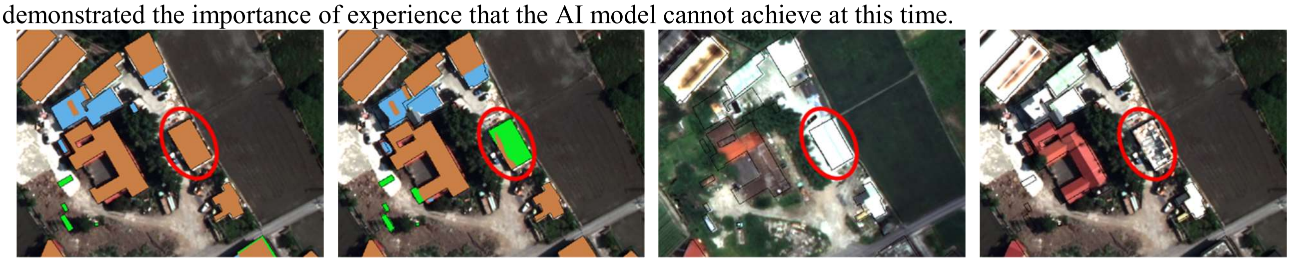
The canopy has existed in past and current images. However, the past build label didn't account for it (Figure 4). After that, we reviewed the training dataset and found the building area definition might have been changed. Several building canopy areas were not labelled on the past building map but on the current building map. The building area definition change could led to this omission errors.

(a) Change label (b) Prediction (c) past image (d) now image

Figure 4. An omission case at building canopy

4.3.2 Commission cases in RGB_DSM_Map model

A part of the building was under renewal, but the dataset tent to classify it into construction areas. It might be the feature of this dataset because the same condition happened to other buildings in the training dataset. Some under-construction areas were labeled as building, while others were not. We were not involved in the eMap production and didn't know how to determine the completeness required for labelling an area as building. This commission error case demonstrated the importance of experience that the AI model cannot achieve at this time.



(a) Change label (b) Prediction (c) past image (d) now image

Figure 5. A commission case at an under-renwal building area

5. CONCLUSION

The input data experiment revealed the best combination of spectral image, height elevation, and building area. In this study, we modified UNetformer, which could effectively fuse different information by sharing weight siamese structure. The best model achieved pixel-wised overall F1-Score 99%, and the best change unit evaluation overall F1-Score of 93%. The test dataset had 115 deconstruction units, 1387 no-change units, and 353 construction units. Because no change in building category was the main component, adding a building map directly introduced existing information that significantly improved prediction results. Additionally, DSM height information enhanced building construction detection, as the height difference could obviously represent ground object change.

ACKNOWLEDGMENTS

This research was partially supported by the Ministry of the Interior, Taiwan (112PL024A).

REFERENCES.

1. Bandara, W. G. C., & Patel, V. M., 2022. A transformer-based siamese network for change detection. In *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium* (pp. 207-210). IEEE.
2. 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.

3. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S. & Guo, B., 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 10012-10022).
4. Lyu, X., Hao, M., & Shi, W., 2020. Building change detection using a shape context similarity model for LiDAR data. *ISPRS International Journal of Geo-Information*, *9*(11), 678.
5. Su, S., Nawata, T., & Fuse, T., 2020. Building change detection from bitemporal aerial images using deep learning. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2, 565-571.
6. Tan, K., Jin, X., Plaza, A., Wang, X., Xiao, L., & Du, P., 2016. Automatic change detection in high-resolution remote sensing images by using a multiple classifier system and spectral–spatial features. *IEEE Journal of Selected topics in applied Earth observations and Remote Sensing*, *9*(8), 3439-3451.
7. Wang, L., Li, R., Zhang, C., Fang, S., Duan, C., Meng, X., & Atkinson, P. M., 2022. UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *190*, 196-214.
8. Warth, G., Braun, A., Bödinger, C., Hochschild, V., & Bachofer, F., 2019. DSM-based identification of changes in highly dynamic urban agglomerations. *European Journal of Remote Sensing*, *52*(1), 322-334.