Enhanced Change Detection Method in Historical Districts: A Lightweight Visual Transformer Integration Model with Context-Aware Local Feature Augmentation

Lujin Hu[1] , Senchuan Di[1]

[1]School of Geomatics and Urban Spatial Information, Beijing University of Civil Engineering and Architecture, Beijing 102616,China, hulujin@bucea.edu.cn,2108570023123@stu.bucea.edu.cn

Keywords: Change detection, Transformer, Siamese network, Historical Districts.

Abstract

Due to rapid urbanization and the continuous increase in building stock, significant challenges arise for historic district preservation. To overcome the persistent challenge of insufficient small-scale unauthorized structure detection in dense historic districts—a critical limitation of existing deep learning-based change detection frameworks—this paper introduces a Siamese network integrated with a lightweight visual transformer, effectively resolving subtle change omission in complex scenarios. The model utilizes context-aware local enhancement to capture high-frequency local information, significantly improving its accuracy in identifying changed regions. Within the change detection network, a CNN feature extractor first performs downsampling on the input image pair to preliminarily extract feature information. Subsequently, a semantic extraction module extracts and enhances semantic information from the feature maps. Finally, a prediction module calculates the differences between the features of the two images and generates the change prediction results. The reasrech comprehensively validated the model on the public LEVIR-CD dataset. Experimental results demonstrate significant improvements in performance metrics compared to other models. The findings indicate that the improved model also performs excellently on this dataset, verifying its effectiveness and robustness, and showcasing its ability to substantially reduce both omissions and false detections. This study offers a solution for high-accuracy remote sensing change detection by improving deep learning-based models.

1. Introduction

1.1 Research Background

Historic districts refer to urban or rural areas that preserve a significant number of historical remnants (such as buildings, street layouts, spatial patterns, etc.), possess a specific historical character, embody local cultural characteristics, and hold high historical, cultural, artistic, social, or scientific value. Historic districts, rich in architectural and spatial landmarks that embody diverse periods, bear witness to a multifaceted social past. However, with the continuous advancement of urbanization in China and the ever-increasing number of buildings, the protection of historic districts faces severe challenges (Zhang et al., 2016). Traditional manual inspection methods struggle to efficiently monitor subtle changes in the urban fabric and architectural features of these districts. Unauthorized modifications, damage, or improper renovations can lead to irreversible loss of cultural heritage.

To address the preservation needs of historic districts, the utilization of Remote Sensing Change Detection techniques is necessitated. Remote sensing change detection technology offers a new approach for large-scale dynamic monitoring of historic districts. Change detection is a technique for identifying differences in the state of observed objects or phenomena by examining them at different times (Lyu et al., 2016). It holds significant application value in numerous fields, such as land cover change monitoring (Lyu et al., 2018), urban expansion and change studies (Liu et al., 2020), and natural disaster assessment (Li et al., 2019). With the recent continuous progress in theories and technologies within the Artificial Intelligence (AI) field, deep learning methods have gradually been applied to change detection, enhancing its development prospects and potential (Peng et al., 2019). Deep learning techniques possess exceptional feature extraction and representation capabilities, enabling them to fully exploit the deep feature information within remote sensing imagery. Compared to traditional change detection methods constrained by manually designed features, deep learning algorithms can better represent complex surface conditions in images, thereby yielding more accurate change detection results.

However, with the development of high-spatial-resolution imagery, current deep learning methods exhibit certain limitations when confronted with rich spatial feature information, diverse scale characteristics of ground objects, and the massive volume of remote sensing data (Peng er al., 2020). Consequently, neural network-based algorithms still suffer from issues such as insufficient capability to detect pseudo-changes, inadequate multi-scale feature extraction, missed detection of small objects, incomplete detection of changed regions, insufficient extraction of semantic information, and poor representation of image difference information. To address these challenges, this paper proposes a Siamese network integrated with a lightweight vision Transformer (Lightweight ViT).

1.2 Current Related Work

The development of deep learning and big data technologies has led to significant advancements in remote sensing image change detection in fields such as computer vision. By utilizing deep learning-based methods, change features can be directly learned from dual-phase, multi-phase, or time series remote sensing images, and change maps can be generated via image segmentation. These features demonstrate strong robustness, and compared to traditional methods, deep learning approaches not only eliminate the impact caused by dependence on change difference images but also handle remote sensing data acquired from different sensors, exhibiting strong versatility.

Currently, various remote sensing image change detection models have been gradually proposed based on convolutional neural networks, stacked autoencoders, deep belief networks, deep neural networks, and recurrent neural networks. For example, Zhang et al. (2016) analyzed the outstanding feature extraction capabilities of denoising autoencoder models and applied them to change detection in SAR and optical images, achieving remarkable results. Recurrent neural network models are typically used to handle time series data, and since multi-phase remote sensing images are inherently related to temporal changes, some researchers have applied recurrent neural networks to change detection in remote sensing images. Lyu et al. (2016) successfully applied the Long Short-Term Memory (LSTM) model in the field of land cover change detection in remote sensing images for the first time, demonstrating change detection over a 20-year time span.

Fully Convolutional Networks (FCN) are widely used in image classification and change detection. They utilize deconvolution to extract change maps from high-dimensional features, enabling FCN to perform change detection in an end-to-end manner. Liu et al. (2020) improved FCN using the idea of depthwise separable convolution, making the network more lightweight and enhancing its performance. Li et al. (2019) incorporated unsupervised modules into FCN, achieving unsupervised change detection. In 2015, a variant network based on the FCN structure, named U-Net, was published. Due to its superior performance in semantic segmentation tasks, many scholars have also utilized this model for change detection tasks. For instance, Peng et al. (2019) proposed an end-to-end change detection method based on U-Net++, which uses early fused dual-phase remote sensing images as input to the convolutional neural network. Peng et al. (2020) also incorporated differential enhancement dense attention blocks into the U-Net++ structure for change detection in dual-phase optical remote sensing images.

Siamese neural networks can perform comparisons between images using multiple inputs, with shared weights between the two subnetworks, reducing the number of network parameters and ensuring features are extracted in the same manner. This approach is widely used in remote sensing image change detection, with many researchers opting for Siamese neural network models in their studies. Hughes et al. (2018) designed a pseudo-Siamese convolutional neural network with two different subnetwork structures, applying it to change detection in optical and SAR images. Daudt et al. (2018) took U-Net as a baseline and designed three different Siamese fully convolutional network structures for change detection based on different feature fusion or difference extraction methods. Zhang et al. (2020) proposed a dual-branch structure that extracts highly expressive deep features and uses a deep supervision recognition network to classify deep feature differences for change detection in remote sensing images. Chen et al. (2020) introduced a dual-attention Siamese network structure that effectively captures long-term dependencies through a dual attention mechanism, resulting in more discriminative feature representations. Additionally, they employed a weighted bilateral contrastive loss to adjust the influence of invariant pixels in public datasets, reducing the weight of invariant features and achieving certain results.

Furthermore, other scholars have proposed different processes and model structures for change detection. For example, Gong et al. (2017) applied Generative Adversarial Networks (GAN) for change detection. Wu et al. (2021) performed image segmentation based on Graph Convolutional Networks (GCN), then constructed graphs from image patches for change recognition. Chen et al. (2022) developed a change detection network based on a Transformer approach that combines an encoder and decoder with semantic features.

2. Research Methodology

Given the prevalent insufficient extraction of semantic information in current building change detection models—which leads to loss of detail features (particularly multi-scale details) and consequent incomplete boundary delineation—we propose a context-aware enhanced Siamese network for change detection. The architecture comprises: (1) a CNN-based feature extraction module, (2) asemantic token extraction module, (3) a semantic enhancement module, (4) context-aggregating encoder-decoder blocks, and (5) a prediction head. This design strategically reinforces semantic representation to refine image features, thereby enhancing detection precision and ultimately improving the model's capacity to extract precise building change information from remote sensing data.

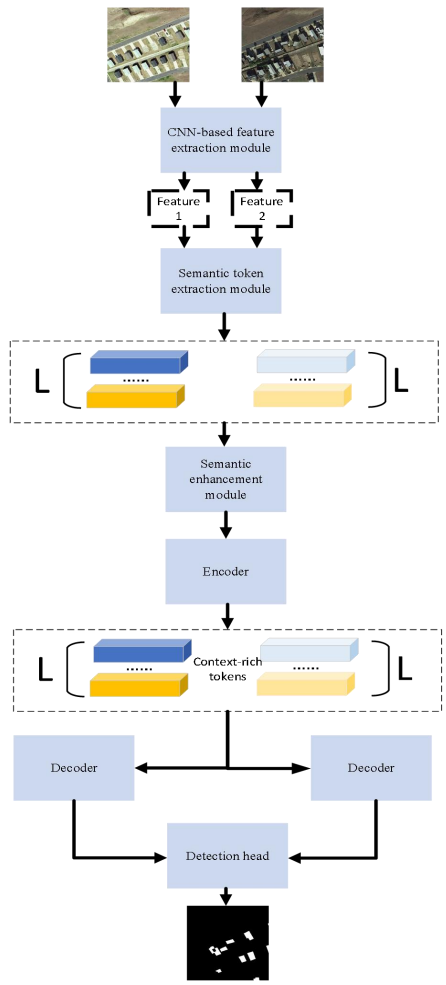

Figure 1. Operational Workflow of the Model.

2.1 CNN-based feature extraction module

The CNN module can utilize multiple convolutional layers to achieve hierarchical feature representation and leverage a larger receptive field to capture global contextual information. When the change scenarios are relatively concentrated and the areas of change are densely distributed, this approach can yield superior detection results. This module is capable of simultaneously

processing dual-timeinput images to extract spatial-spectral features ranging from shallow to intermediate levels. Its structure is fundamentally similar to that of the VGG network, but it incorporates fewer convolutional kernels and exhibits lower computational complexity, akin to ResNet18 (K et al., 2016). Its structure is illustrated in Figure 2.

The model consists of 34 layers and is divided into five stages, incorporating downsampling operations. Most of these layers are composed of 3×3 convolutional kernels with appropriate padding, except for the initial layer, which utilizes a 7×7 kernel. The network concludes with a global average pooling layer followed by a fully connected layer with 1000 output nodes. The pooling layers are employed to progressively reduce the spatial dimensions of the feature maps, thereby decreasing the number of parameters and computational cost. The fully connected layer further consolidates the features for classification. When processing an input RGB image of 224×224 pixels, the model traverses these five stages, undergoing gradual spatial reduction until the feature maps reach a 1×1 dimension. This results in a 1D feature vector, which is subsequently processed by the fully connected layer for feature classification and class probability output. It can be expressed mathematically as:

$$F_i = softmax\left(Avgpool\left(Conv(X_i)\right)\right) \tag{1}$$

Where X_i is the input image at the i-th time point, and F_i is the corresponding feature of the image output from the CNN feature extraction module at the i-th time point.

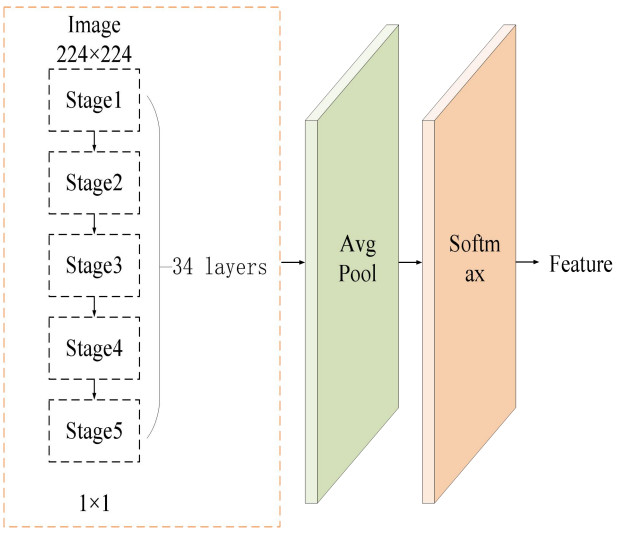

Figure 2. The structure of the CNN feature extraction module.

2.2 Semantic Token Extraction Module

The Semantic Tokenizer extracts semantic information and obtains semantic tokens from the feature maps of dual-temporal remote sensing images extracted by the CNN module. It segments the images based on semantic information, and these semantic tokens are aggregated to form a set of token sets. Let F_1 and F_2 denote the feature maps of the dual-temporal input images, with height, width, and channel dimensions represented as H×W×C. The Semantic Token Extraction Module will produce two sets of tokens, referred to as T_1 and T_2.

Assuming that F_i (for i=1,2) represents a pixel in the feature map, applying pointwise convolution to each such pixel in the feature maps will yield L semantic groups, where each group corresponds to a distinct semantic concept. The computation process can be expressed as follows:

$$Output = T_i = Extraction(F_i) = \sum_{j=1}^{L} w_j \cdot F_i \tag{2}$$

where w_j represents the weights for each semantic group.

2.3 Semantic enhancement module

The Semantic Enhancement Module is implemented by the Cloblock module, where the Cloblock module is divided into a local branch and a global branch (Fan et al., 2023). It processes the token sequences from the Semantic Tokenizer and utilizes a self-attention mechanism to establish long-range dependencies between the tokens. By stacking multiple Transformer encoder layers, this module learns complex interactions between objects and scenes, capturing contextual relationships at the level of semantics throughout the entire image. This capability is essential for accurately interpreting the changing semantic context.

The Cloblock module consists of a local branch and a global branch, the structure of which is shown in Figure 3:

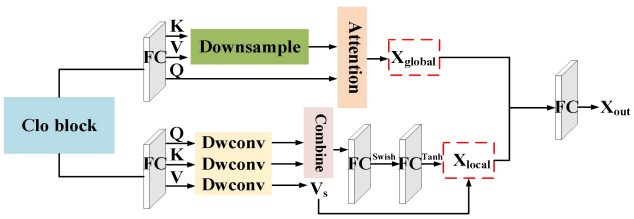

Figure 3. The structure of the two branches of the Cloblock module.

In the diagram, FC represents the fully connected layer and DWConv denotes Depthwise Convolution. In global branch, first, the key (K) and value (V) vectors undergo downsampling, followed by the standard attention mechanism applied to the query (Q), key (K), and value (V) vectors to extract low-frequency global information. Specifically, the query vector Q_i, key vector K_i, and value vector V_i are generated by applying linear transformations to the tokens. Subsequently, K_i and V_i undergo downsampling before being subjected to the attention operation alongside Q_i to obtain the output. The process can be represented by the following formula:

$$A_{global} = Attention\left(Q_i, Pool(K_i), Pool(V_i)\right) \tag{3}$$

In the local branch, during this process, the query (Q) and key (K) undergo a depthwise convolution, followed by the computation of the Hadamard product between Q and K. A series of transformations are then applied to obtain the context-aware weights A_1. After passing through two fully connected layers (FC) as well as the Swish and Tanh activation functions, higher-quality context-aware weights A_2 are obtained. Finally, these weights are jointly computed with the value (V), which has also undergone depthwise convolution, to produce the output A_{local}. The process can be represented by the following formula:

$$V_s = DWconv(V) \tag{4}$$

$$Q_s = DWconv(Q) \tag{5}$$

$$K_s = DWconv(K) \tag{6}$$

$$A_1 = FC\left(Swish\left(FC(Q_s \odot K_s)\right)\right) \tag{7}$$

$$A_2 = Tanh\left(\frac{A_1}{\sqrt{d}}\right) \tag{8}$$

$$A_{local} = A_2 \odot V_s \tag{9}$$

Here, d is the number of token's channels, the symbol " $\odot$ " represents the Hadamard product. Thus, both the local branch and the global branch have been obtained, and their fusion results in the final output:

$$A_{out} = FC\left(Concat\left(A_{local}, A_{global}\right)\right) \tag{10}$$

"Concat" means to concatenate the two outputs along the channel dimension.

2.4 Context-aggregating encoder-decoder blocks

Following semantic enhancement, the subsequent step involves transforming the tokens into pixel-level features. This task is achieved by modeling contextual dependencies and refining image features through an Encoder-Decoder architecture operating on tokens at each timestep. The encoder receives the semantically enriched dual-temporal token sequences and derives contextually enriched tokens via multi-head self-attention mechanisms. These tokens are subsequently refined by the decoder based on the relationship between each pixel and the final token set, thereby revealing information pertaining to objective changes within the image data. The specific process can be expressed by the following formulae:

$$T_{final}^{i*} = \text{Transformer\_Encoder}(T_{final}^{i}) \tag{11}$$

$$F^{i*} = \text{Transformer\_Decoder}(F^{i}, T_{final}^{i*}) \tag{12}$$

Here, i takes the values of 1 and 2, representing the two temporal points.

2.5 Detection head

The primary objective of change detection is to identify and extract information pertaining to surface changes by comparing remote sensing imagery acquired at different temporal epochs. For building targets, such changes can be characterized as either their emergence or disappearance. This phenomenon can be effectively represented in the image domain by computing the absolute difference between the features extracted from the two time points. Consequently, the structure of the detection head is intentionally kept relatively simple. It utilizes the semantically enriched features encoded by the Transformer and employs a very shallow fully convolutional network (FCN) to perform change discrimination. Specifically, given the two upsampled feature maps F_1 and F_2 output from the preceding stage, the absolute difference between these dual-temporal feature sets is computed. This resulting difference map is subsequently passed through a classifier and finally processed by a softmax function to generate the predicted change probability map P, formulated as follows:

$$P = \text{softmax}\left(g\left(|F^{1*} - F^{2*}|\right)\right) \tag{13}$$

3. Experiment

3.1 Dataset

This study employs the LEVIR-CD dataset for model training and evaluation. LEVIR-CD is a publicly available, large-scale, high-resolution benchmark dataset for remote sensing image change detection, constructed by a research team from Wuhan University (Chen et al., 2020). The dataset comprises 637 pairs of bitemporal remote sensing image patches (with temporal intervals ranging approximately from 5 to 14 years). Each pair consists of two 1024×1024 pixel images of the same geographical area at 0.5-meter per pixel resolution, accompanied by corresponding pixel-level binary change masks. Primarily covering urban areas in Texas, USA (e.g., Austin, League City), the dataset focuses on building changes (construction and demolition), with approximately 31,333 independent changed building instances annotated. It provides an official split into training set (445 pairs), validation set (64 pairs), and test set (128 pairs), ensuring fair and reproducible evaluation. The scenes encompass diverse building types and complex conditions (e.g., illumination variations, shadows, vegetation occlusion), establishing LEVIR-CD as a challenging standard benchmark widely adopted for building change detection algorithms within the research community.

To validate the generalization capability of the model, this study conducted experiments on a private dataset. The proposed model was applied to detect remote sensing images of the Shichahai historical districts in Beijing, China, at two time points: 2013 and 2023. Each image has a resolution of [256×256] pixels and encompasses the hutong neighborhoods, water bodies, and traditional architectural clusters within the Shichahai historical districts, documenting the processes of urban renewal and landscape evolution under the constraints of cultural heritage.

3.2 Parameter Settings

The proposed method was evaluated on the LEVIR-CD benchmark. All models were trained using the Adam optimizer with an initial learning rate of 5×10^{-4} and cosine annealing scheduler. Training configurations included: batch size = 8, maximum epochs = 1,000, and fixed random seed (11) for reproducibility. Data augmentation followed the standard protocol with random cropping and flipping. Data Preparation: High-resolution (1024×1024) images from LEVIR-CD were processed into 256×256 patches using a sliding window approach (50% overlap) to balance contextual preservation and computational demands. The resulting 14,328 patches underwent an 8:2 training-validation split, with rigorous spatial stratification ensuring that there was no geographic overlap between the training and validation sets.

The implementation is executed using Python 3.9 and the PyTorch framework on an NVIDIA GeForce RTX 4060 GPU (8GB VRAM) with CUDA acceleration. A balanced weighted cross-entropy loss function is utilized. To ensure reproducibility, all experiments fix the random seed to 11 and are conducted in a single-GPU environment.

3.3 Result

In this study, our model was compared with several other models using the publicly available LEVIR-CD dataset. We selected the following change detection models for comparative experiments: FC-Siam-Di (Daudt et al., 2018), FC-Siam-Conc (Daudt et al., 2018), DTCTSCN (Liu et al., 2020), BIT (Chen et al., 2021), and SNUNet (Fang et al., 2021). In the experiments, Precision (Pre), Recall (Rec), F1-score (F1), Intersection over Union (IoU), and Overall Accuracy (OA) were selected as accuracy metrics to quantitatively assess the detection precision of building changes. The quantitative analysis of the model's accuracy is shown in the table below:

models	Pre/%	Rec/%	F1/%	IoU/%	OA/%
FC-Siam-Di	89.53	83.31	86.31	75.92	98.67
FC-Siam-Conc	91.99	76.77	83.69	71.96	98.49
DTCTSCN	88.53	86.83	87.67	78.05	98.77
BIT	89.24	89.37	89.31	80.68	98.92
SNUNet	89.18	87.17	88.16	78.83	98.82
Ours	91.68	90.18	90.92	82.16	98.97

Table 1. Quantitative Comparison Results of LEVIR-CD Dataset

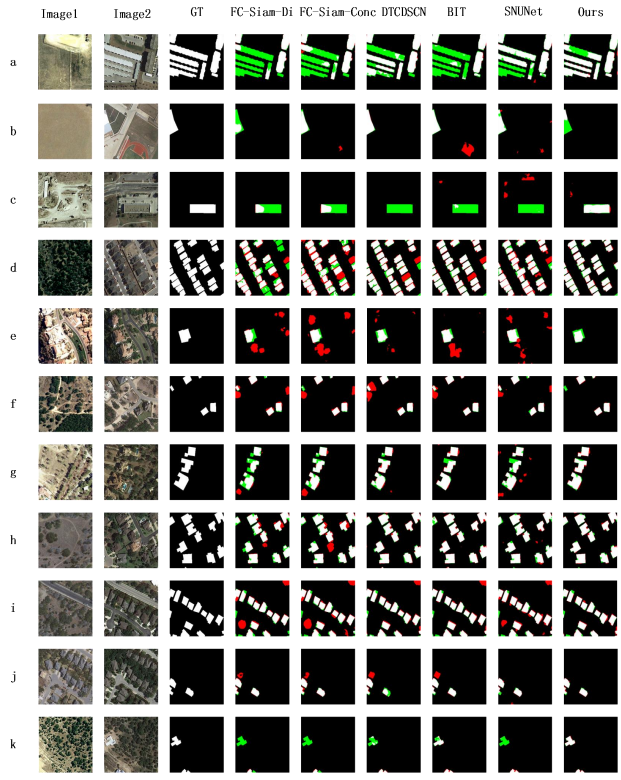

Figure 4. Visual Results on LEVIR-CD Dataset.

Compared to FC-Siam-Di, FC-Siam-Conc, DTCTSCN, BIT, and SNUNet, the Recall metric improved by 6.87%, 13.41%, 3.35%, 0.81%, and 3.01%, respectively. The F1 metric saw increases of 4.61%, 7.23%, 3.25%, 1.61%, and 2.76%, respectively. Overall, our model demonstrates superior performance in terms of Recall, F1 score, Intersection over Union (IoU), and overall accuracy (OA) compared to the other models, with a significant improvement in accuracy. We conducted a qualitative analysis of our model's performance on the test set, comparing it with other models, and the visual representation of the results is shown in Figure 4.

In the figure:
GT (Ground Truth): Manually annotated changed regions (reference data).
Image1 & Image2: Dual-temporal remote sensing images.
White pixels: Correctly detected building changes (True Positives).
Black pixels: Correctly identified unchanged areas (True Negatives).
Green pixels: Omitted changes (False Negatives; actual changes undetected).
Red pixels: False alarms (False Positives; erroneous change detection).

Red and green pixels quantitatively represent commission errors (false positives) and omission errors (false negatives), respectively. The spatial extent of these colored areas directly correlates with prediction inaccuracies—larger areas indicate greater errors. Fig. 4 showcases results across diverse challenging scenarios: severe vegetation occlusion (e.g., h, j), densely built-up areas (e.g., d), and small-target changes (e.g., e, k). Crucially, our model demonstrates significantly smaller red/green areas than comparative methods in all cases. This empirical evidence confirms a marked reduction in both commission and omission error rates, substantiating the superior performance of our approach.

3.4 Model Generalization Capability Verification in Historical Districts

The Shichahai historical districts Change Detection Dataset is a remote sensing image dataset constructed for the detailed analysis of urban changes within the historical and cultural protection zone of Shichahai in Beijing. The detection results of the proposed model for the historical districts are illustrated in Figure 5.

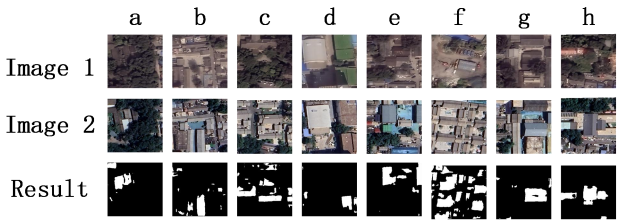

Figure 5. The prediction results of the model on the Shichahai Change Detection Dataset.

As illustrated in the figure, the proposed model exhibits robust performance in practical applications, excelling in detecting historical building changes within urban districts. This further demonstrates the model's strong generalization capability.

4. Conclusion

This study proposes a novel lightweight change detection model that aims to efficiently and accurately identify building changes in high-resolution remote sensing imagery, particularly in historical districts, by combining a twin network with a lightweight visual Transformer. The architecture employs a context-aware local enhancement module that captures high-frequency local details through a two-branch structure in

conjunction with depthwise convolution. Extensive validation in the LEVIR-CD benchmark has demonstrated its leading performance (F1: 90.92%, IoU: 82.16%, OA: 98.97%), significantly outperforming existing methods in complex scenarios involving vegetation occlusion, dense urban environments, and small target changes. Importantly, when applied to the Shichahai historical district dataset, the model exhibits exceptional generalization capabilities, confirming its robustness in real-world cultural heritage monitoring scenarios. By effectively balancing precision and recall through the dual-branch structure and semantic labeling enhancement mechanism, this work provides a solution with both computational efficiency and practical value for large-scale urban dynamic analysis and heritage preservation, advancing the technological development in the field of high-resolution remote sensing building change detection.

Although our proposed model achieves improved accuracy in identifying changes within historical districts, the study still presents some limitations. Specifically, our model exhibits a relatively high computational cost (parameter count). Future research will focus on reducing model complexity and computational burden while maintaining or even enhancing accuracy. Furthermore, extending the model's capability for fine-grained identification of 3D structural changes and their underlying causes – such as renovation, restoration, and demolition – is also essential. This enhancement will provide stronger decision-making support for smart city renewal and the targeted preservation of cultural heritage.

References

Chen, H., Qi, Z., Shi, Z., 2022. Remote Sensing Image Change Detection With Transformers. *IEEE Transactions on Geoscience and Remote Sensing* 60, 1-14, doi:10.1109/TGRS.2021.3095166.

Chen, H., Shi, Z., 2020. A Spatial-Temporal Attention-Based Method and a New Dataset for Remote Sensing Image Change Detection. *Remote Sensing* 12, 1662,doi:10.3390/rs12101662.

Chen, J., Yuan, Z., Peng, J., Chen, L., Huang, H., Zhu, J., Liu, Y., Li, H., 2021. DASNet: Dual Attentive Fully Convolutional Siamese Networks for Change Detection in High-Resolution Satellite Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14, 1194-1206, doi:10.1109/JSTARS.2020.3037893.

Daudt, R.C., Saux, B.L., Boulch, A., 2018. Fully Convolutional Siamese Networks for Change Detection, 2018 25th IEEE International Conference on Image Processing (ICIP), pp. 4063-4067.

Fan, Q., Huang, H., Guan, J., He, R., 2023. Rethinking Local Perception in Lightweight Vision Transformer, p. arXiv:2303.17803.

Fang, S., Li, K., Shao, J., Li, Z., 2022. SNUNet-CD: A Densely Connected Siamese Network for Change Detection of VHR Images. *IEEE Geoscience and Remote Sensing Letters* 19, 1-5, doi:10.1109/LGRS.2021.3056416.

Gong, M., Niu, X., Zhang, P., Li, Z., 2017. Generative Adversarial Networks for Change Detection in Multispectral Imagery. *IEEE Geoscience and Remote Sensing Letters* 14, 2310-2314, doi:10.1109/LGRS.2017.2762694.

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), pp. 770-778.

Hughes, L.H., Schmitt, M., Mou, L., Wang, Y., Zhu, X.X., 2018. Identifying Corresponding Patches in SAR and Optical Images With a Pseudo-Siamese CNN. *IEEE Geoscience and Remote Sensing Letters* 15, 784-788, doi:10.1109/LGRS.2018.2799232.

Li, X., Yuan, Z., Wang, Q., 2019. Unsupervised Deep Noise Modeling for Hyperspectral Image Change Detection. *Remote Sensing* 11, 258.

Liu, R., Jiang, D., Zhang, L., Zhang, Z., 2020. Deep Depthwise Separable Convolutional Network for Change Detection in Optical Aerial Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13, 1109-1118, doi:10.1109/JSTARS.2020.2974276.

Liu, Y., Pang, C., Zhan, Z., Zhang, X., Yang, X., 2021. Building Change Detection for Remote Sensing Images Using a Dual-Task Constrained Deep Siamese Convolutional Network Model. *IEEE Geoscience and Remote Sensing Letters* 18, 811-815, doi:10.1109/LGRS.2020.2988032.

Lyu, H., Lu, H., Mou, L., 2016. Learning a Transferable Change Rule from a Recurrent Neural Network for Land Cover Change Detection. *Remote Sensing* 8, 506.

Lyu, H., Lu, H., Mou, L., Li, W., Wright, J., Li, X., Li, X., Zhu, X.X., Wang, J., Yu, L., Gong, P., 2018. Long-Term Annual Mapping of Four Cities on Different Continents by Applying a Deep Information Learning Method to Landsat Data. *Remote Sensing* 10, 471.

Peng, D., Zhang, Y., Guan, H., 2019. End-to-End Change Detection for High Resolution Satellite Images Using Improved UNet++. *Remote Sensing* 11, 1382.

Peng, X., Zhong, R., Li, Z., Li, Q., 2021. Optical Remote Sensing Image Change Detection Based on Attention Mechanism and Image Difference. *IEEE Transactions on Geoscience and Remote Sensing* 59, 7296-7307, doi:10.1109/TGRS.2020.3033009.

Wu, J., Li, B., Qin, Y., Ni, W., Zhang, H., Fu, R., Sun, Y., 2021. A multiscale graph convolutional network for change detection in homogeneous and heterogeneous remote sensing images. *International Journal of Applied Earth Observation and Geoinformation* 105, 102615, doi:10.1016/j.jag.2021.102615.

Zhang, C., Yue, P., Tapete, D., Jiang, L., Shangguan, B., Huang, L., Liu, G., 2020. A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing* 166, 183-200, doi:10.1016/j.isprsjprs.2020.06.003.

Zhang, Y., Zhang, E., Chen, W., 2016. Deep neural network for halftone image classification based on sparse auto-encoder. *Engineering Applications of Artificial Intelligence* 50, 245-255, doi:10.1016/j.engappai.2016.01.032.