

Building Change Detection Based on a Dual-Stream Network with Adaptive Feature Fusion and Cross-Attention Guidance Using SAR and Optical Remote Sensing Imagery

Weiwei Fang^{1,2,3}, Taotao Zheng^{1,2,3}, Qing Zhao^{1,2,3,*}
¹Key Laboratory of Geographical Information Science, Ministry of Education, East China Normal University, Shanghai 200241, China; ²School of Geographic Sciences, East China Normal University, Shanghai 200241, China; ³Key Laboratory of Spatial-Temporal Big Data Analysis and Application of Natural Resources in Megacities, Ministry of Natural Resources, Shanghai 200241, China

Abstract
Urban building change detection has a wide range of applications, including urban planning, disaster assessment, and digital map updating. With the continuous diversification of remote sensing satellite platforms (e.g., optical and SAR), the integration of multimodal remote sensing imagery for detecting urban changes has become an important technological approach. SAR satellites can acquire imagery under all weather conditions and regardless of illumination, as they are not affected by light source or atmospheric conditions; however, they are prone to speckle noise, and their backscatter intensity can be disturbed by building layout, orientation, and surface material. In contrast, optical satellites can capture abundant building texture and spectral information, although their acquisition may be compromised by cloud and haze interference. By synergistically utilizing multimodal remote sensing imagery, the complementary advantages of each modality can be fully exploited to improve the precision of urban building change detection. This study proposes a change detection method based on an adaptive-feature-fusion and dual-cross-attention-guided dual-stream network (ADDNet). In the proposed network, the dual-stream architecture is employed to extract features in a manner tailored to the distinct characteristics of SAR and optical data. The cross-attention mechanism effectively fuses multi-scale features from both the encoder and decoder, reinforcing the complementary relationship between low-level details (such as edge textures) and high-level semantic context (such as global structure). Furthermore, an adaptive feature fusion module dynamically generates spatial weight maps to automatically regulate the contribution ratios of SAR and optical features, thereby achieving efficient integration of the complementary information.

Keywords: building change detection; synthetic aperture radar (SAR); optical; deep learning; multimodal remote sensing;

1. Introduction

Deep learning is a very successful method in the field of machine learning. It can learn complex and deep features through large networks and network models can be optimized quickly through backpropagation. There are already some studies of change detection based on deep learning. Pang et al. [1] proposed CD-TransUNet, a building change detection model combining U-Net and Transformer layers. Du et al. [2] integrated U-Net++ with Transformer to introduce the TransUNet++SAR model. Xia et al. [3] presented an end-to-end change detection network model called EGCTNet. Chen et al. [4] proposed ADF-Net, an attention-guided dual-branch change detection model. In this study, we employ ADDNet to extract and fuse deep features from SAR and optical images. During the downsampling stage, the AnisoResConv and EdgeGuidedConv modules are specifically designed to extract modality-specific features from SAR and optical data, respectively. Subsequently, an AS Module is utilized at the bottleneck layer to capture comprehensive local-global features. The multi-scale encoder outputs are processed through a Dual Cross-Attention (DCA) module before being merged with decoder outputs via skip connections, thereby effectively preserving shallow-level features. Finally, the derived SAR and optical features are adaptively fused through an Adaptive Feature Fusion (AFF) Module to generate the final change detection results.

2. Methodology

ADDNet was applied to detect building changes in Shanghai by employing a dual-stream network architecture. In this architecture, each stream is based on an encoder–decoder structure, with the encoder designed with specialized down-sampling convolution modules that are customized according to the data type (SAR or optical) for precise feature extraction. The encoder then performs up-sampling to restore the feature maps to the original resolution. Concurrently, the encoder outputs are refined using a cross-attention mechanism and merged with the decoder outputs via skip connections to better preserve low-level features. Finally, an adaptive feature fusion module is introduced to seamlessly integrate the complementary information from SAR and optical data.

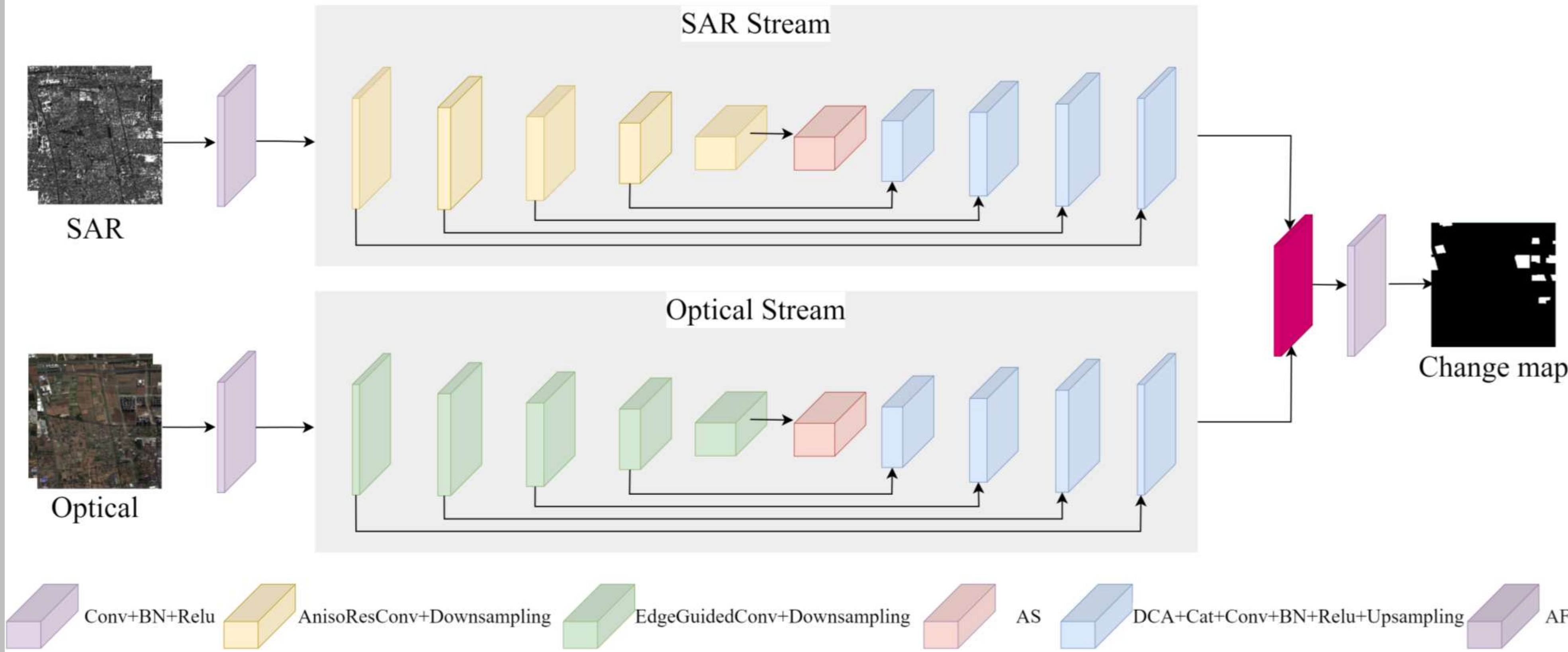


Figure 1. The structure of the ADDNet

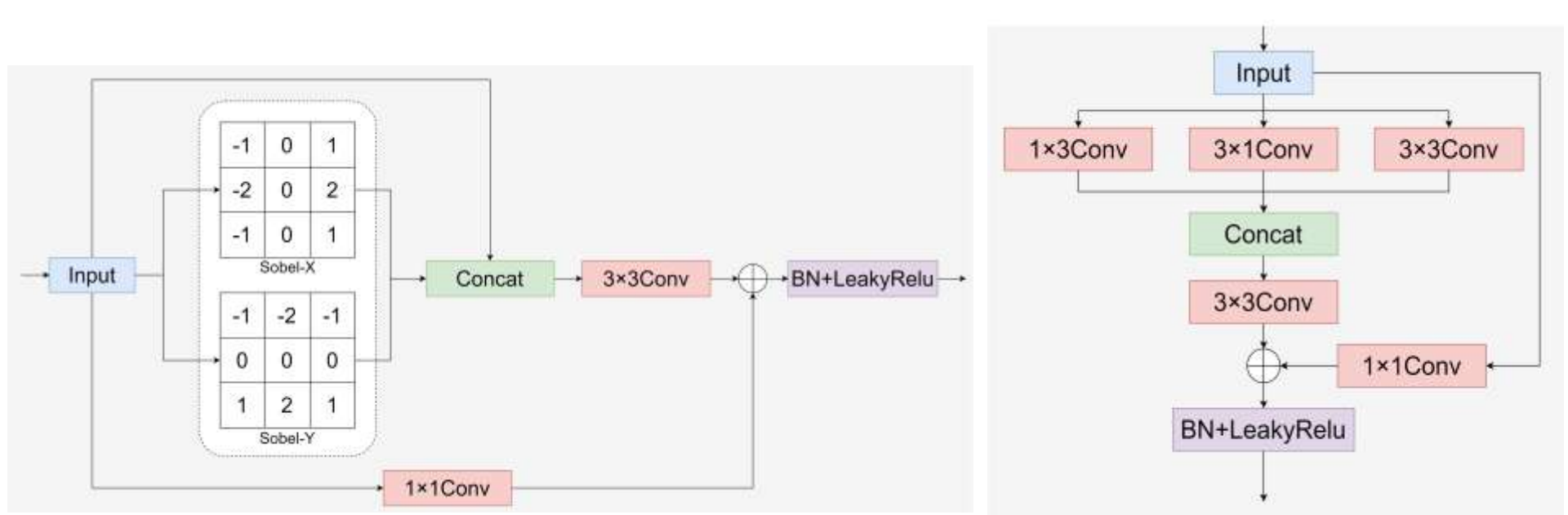


Figure 2. The structure of EdgeGuidedConv and AnisoResConv

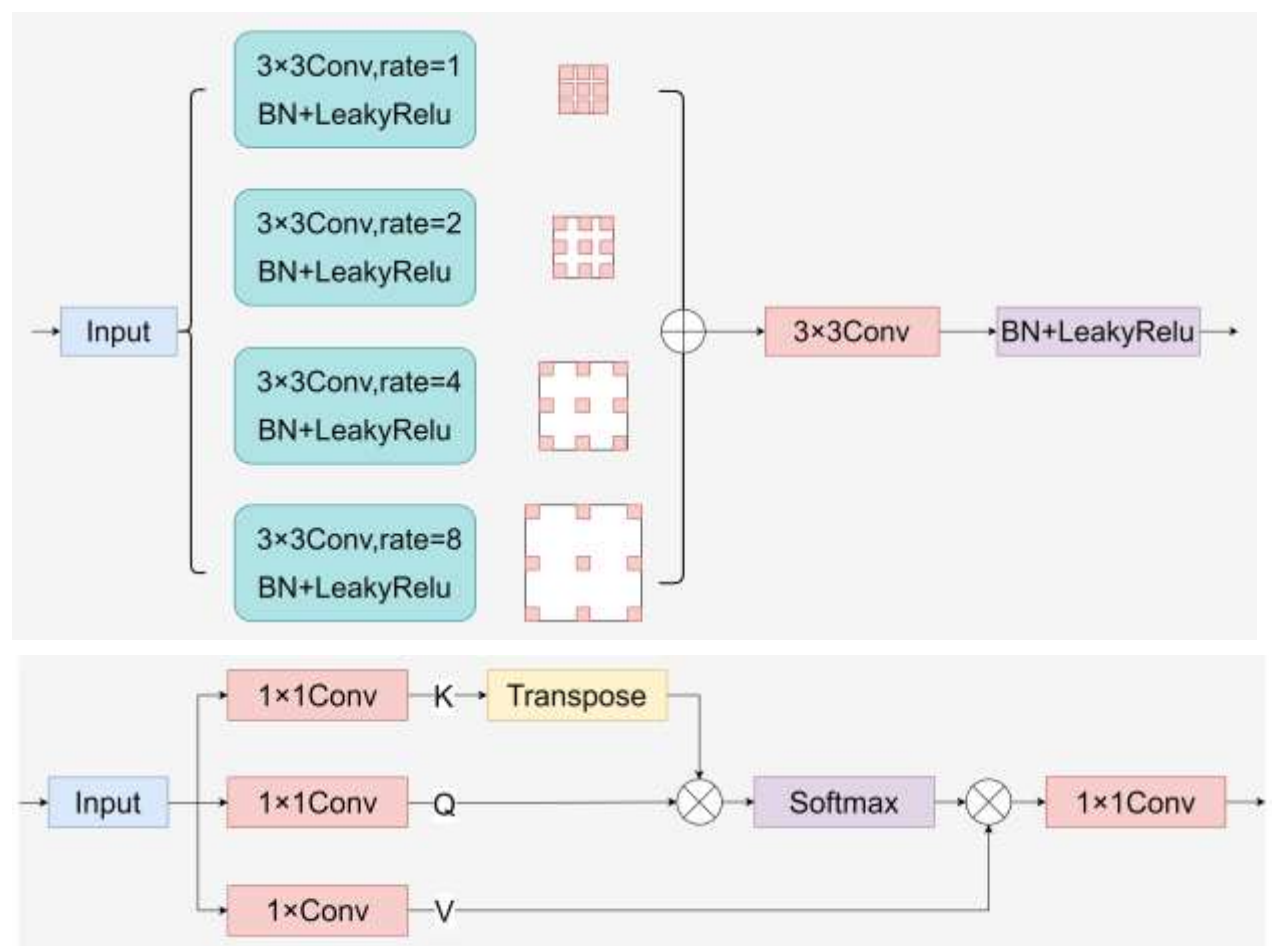


Figure 3. The structure of ASPP and Self-attention

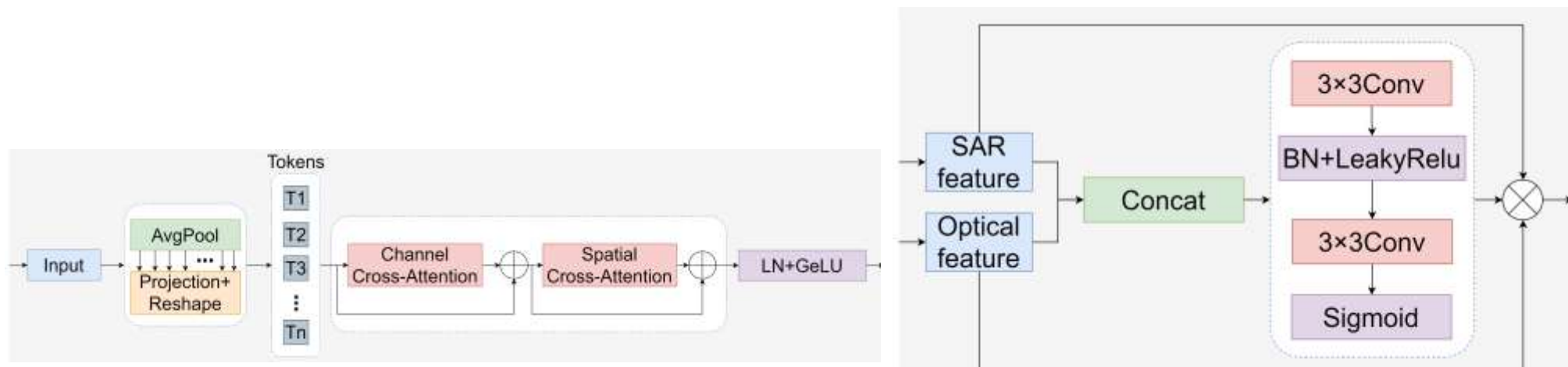


Figure 4. The structure of DCA and AFF

3. Experiments and results

A. Experimental Data

This study acquired Sentinel-1 (SAR) and Sentinel-2 (optical) satellite image pairs from Shanghai, covering pre-change and post-change phases (November 2021 to December 2024). Sentinel-1 SLC data were processed with multi-looking, radiometric correction, and geocoding to extract intensity information, while Sentinel-2 multispectral data were processed with atmospheric correction and radiometric calibration. All datasets were resampled to a uniform spatial resolution and spatially aligned through geocoding. SAR, optical label images are shown in Figure 5.

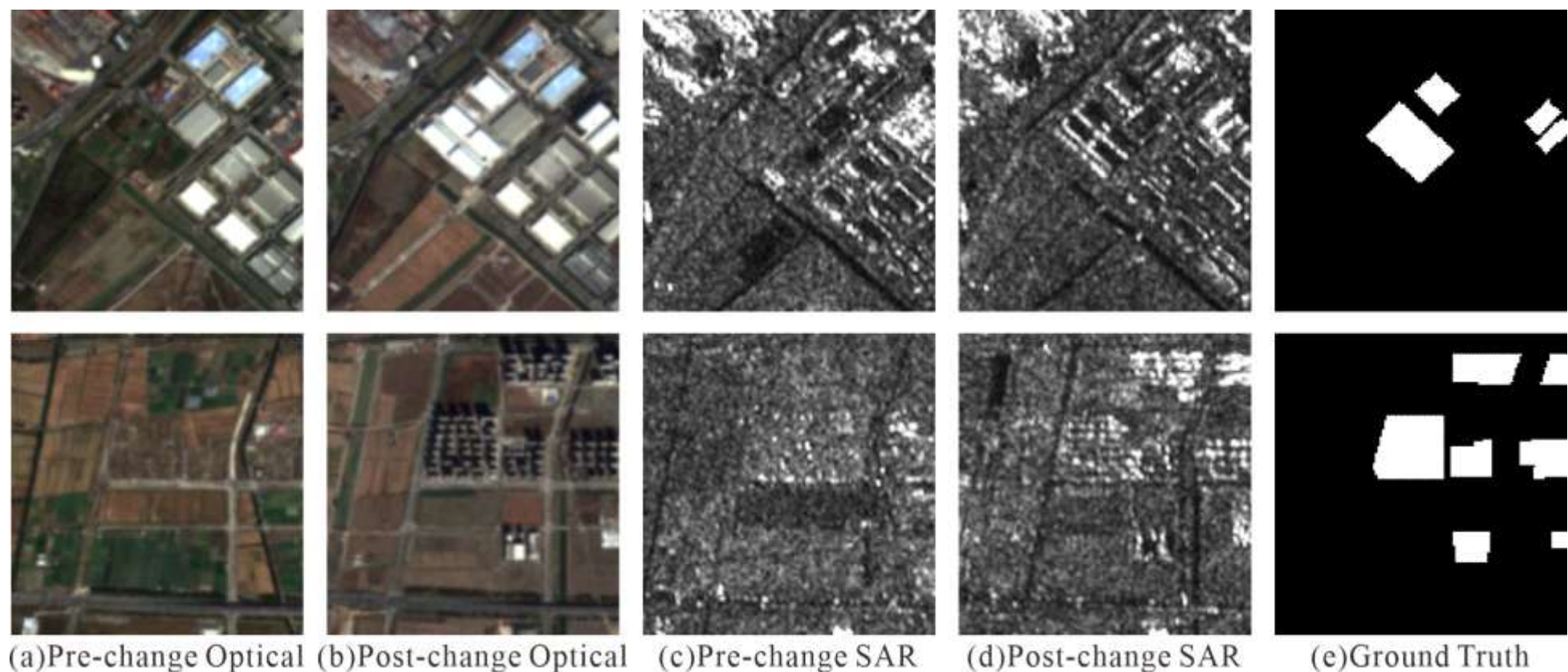


Figure 5. Dataset . (c)Pre-change optical image. (d)Post-change optical image. (c)Pre-change SAR image. (d)Post-change SAR image. (e) Ground-truth image

B. Result and Accuracy

This comparative study demonstrates that the ADDNet model with SAR and optical imagery inputs achieves optimal performance in building change detection (BCD) tasks. As shown in Figure 6, the dual-input ADDNet exhibits enhanced capability in identifying building change within test set results, significantly reducing false positives and missed detections. Quantitative analysis (Table 1) reveals that compared to the U-Net baseline, the proposed model improves F1-score, IoU, and Accuracy by 5.60%, 7.98%, and 0.23%, respectively. When compared to its SAR-only variant, the dual-source ADDNet further enhances these metrics by 8.21%, 11.41%, and 0.33%, while outperforming the optical-only variant with 7.74%, 10.44%, and 0.28% improvements respectively. These results confirm the significant performance gains enabled by multisource data fusion in deep learning models.

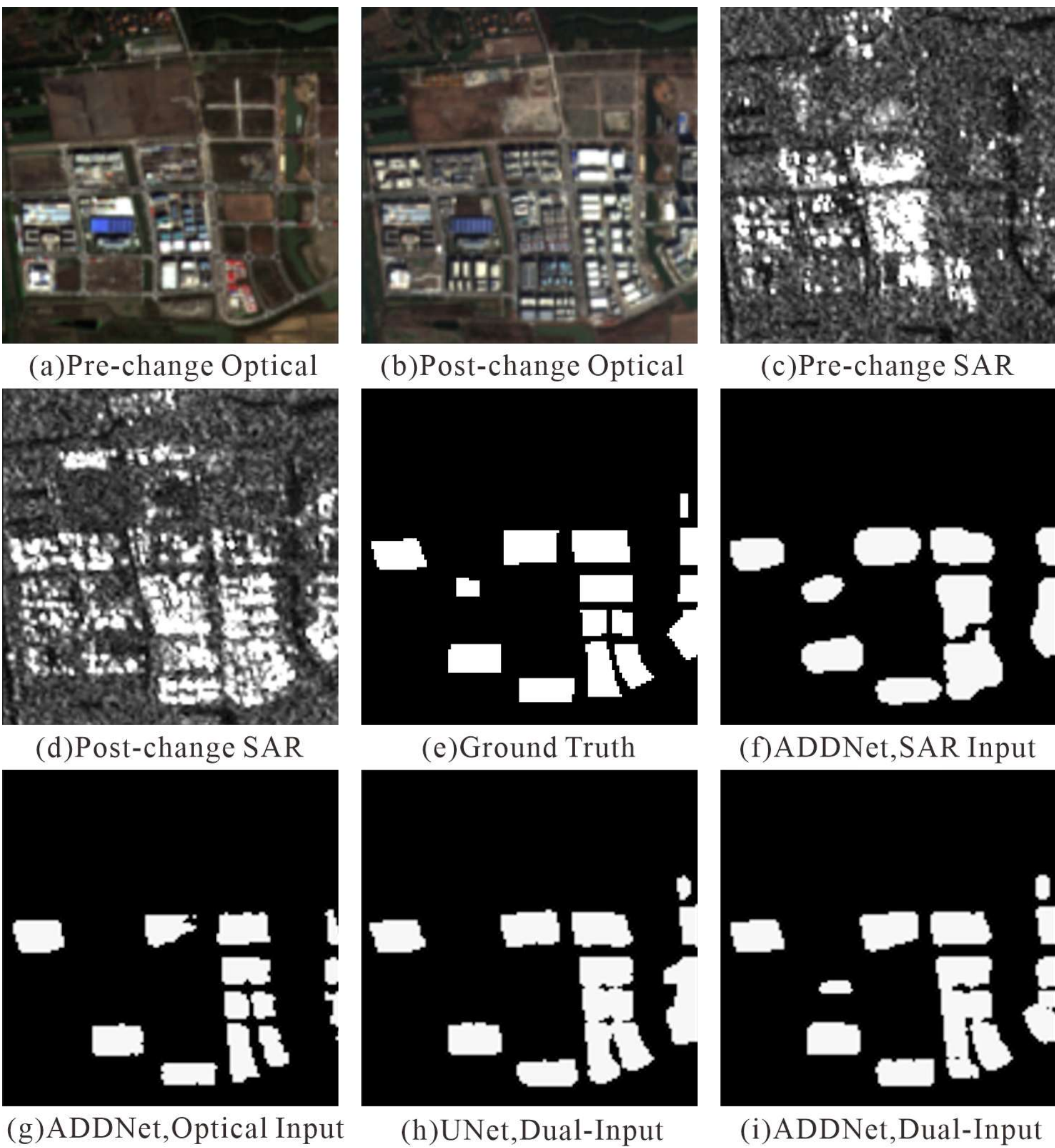


Figure 6. The visual results of building change detection. (a) Pre-change optical image. (b) Post-change optical image. (c) Pre-change SAR image. (d) Post-change SAR image. (e) Ground-truth image. (f) ADDNet, SAR as input data. (g) ADDNet, optical as input data. (h) U-Net, SAR and optical as input data. (i) ADDNet, SAR and optical as input data.

Method	Input	Precision	Recall	F1	IoU	Accuracy
U-Net	Optical, SAR	0.8434	0.8282	0.8258	0.7187	0.9904
ADDNet	Optical	0.9374	0.7296	0.8081	0.6941	0.9899
ADDNet	SAR	0.8287	0.7925	0.8034	0.6844	0.9894
ADDNet	Optical, SAR	0.8920	0.8843	0.8855	0.7985	0.9927

Table 1. Comparisons of building change detection methods. SAR and optical represent the different sources of input data

4. Conclusion

This study proposes an urban building change detection method integrating SAR and optical data. Unlike conventional SAR-optical change detection approaches, we develop a heterogeneous dual-stream architecture where the SAR branch employs denoising convolution while the optical branch enhances texture detail preservation. The framework introduces a dual cross-attention module that dynamically establishes spatial-channel correlations during the encoder phase to improve collaborative representation of shallow details and deep semantics. An adaptive feature fusion module is designed to generate spatial weighting maps based on local context, optimizing the contribution ratios of dual-stream features. Experimental results demonstrate the method's superior reliability in urban building change detection, achieving Precision (89.20%), Recall (88.43%), F1-score (88.55%), IoU(79.85%), and Accuracy (99.27%), outperforming existing advanced techniques.

KEY REFERENCES

- [1] Pang L, Sun J, Chi Y, Yang Y, Zhang F, Zhang L. CD-TransUNet: A Hybrid Transformer Network for the Change Detection of Urban Buildings Using L-Band SAR Images[J]. Sustainability, 2022, 14(16): 9847.
- [2] Du Y, Zhong R, Li Q, Zhang F. MDPI AG, 2022. TransUNet++SAR: Change Detection with Deep Learning about Architectural Ensemble in SAR Images[J]. Remote Sensing, 2022, 15(1): 6.
- [3] Xia L, Chen J, Luo J, Zhang J, Yang D, Shen Z. Building Change Detection Based on an Edge-Guided Convolutional Neural Network Combined with a Transformer[J]. Remote Sensing, 2022, 14(18): 4524.
- [4] Chen P, Lin J, Zhao Q, Zhou L, Yang T, Huang X, Wu J. MDPI AG, 2024. ADF-Net: An Attention-Guided Dual-Branch Fusion Network for Building Change Detection near the Shanghai Metro Line Using Sequences of TerraSAR-X Images[J]. Remote Sensing, 2024, 16(6): 1070.