FSFENet: An Efficient Frequency-Spatial Feature Extraction Network for Remote Sensing Change Detection

Hao Zhang[1], Tao Xu[1*], Ruke Shang[1], Jiayin Zhang[1], Yan Zhang[2], Yuqiang Wang[3]

[1]School of Information Science and Engineering, University of Jinan, Jinan, China
[2]Zouping Natural Resources and Planning Bureau, Binzhou, China
[3]Disaster Reduction Center of Shandong Province, Jinan, China
2511984431@qq.com; xutao@ujn.edu.cn; srkvictory@163.com; 2761852738@qq.com; xiao_yan9121@163.com; sdsjzzx@163.com
Corresponding author: Tao Xu

***Abstract*—Erroneous detection of pseudo-change targets is one of the main reasons for the performance degradation of change detection algorithms. Most research efforts improve the ability to identify pseudo-change targets by constructing complex network structures, which makes it difficult to balance detection accuracy and lightweight performance. To address this, this paper proposes a Frequency-Spatial Feature Extraction Network (FSFENet). First, a Frequency Filtering Module (FFM) is constructed, which uses Fast Fourier Transform and attention mechanisms for selective filtering, mitigating the impact of domain shift between dual-temporal image pairs and enhancing the effective representation of change targets. Second, a lightweight Dual-temporal Change Feature Extractor (DT-CFE) is designed to adaptively extract change features from dual-temporal features through channel-wise feature interaction, reducing the influence of pseudo changes. Finally, a Multi-level Feature Aggregation Module (MFAM) is designed to enhance the completeness of change regions by fusing low-level and high-level features. Experimental results on three public datasets—WHU-CD, SYSU-CD, and DSIFN-CD—show that with only 0.45M parameters and 2.34G FLOPs, FSFENet achieves F1 scores of 92.75%, 81.99%, and 67.77%, respectively, demonstrating the effectiveness of the proposed method. The code is available at https://github.com/zhang2209/FSFENet.**

***Index Terms*—Remote sensing change detection, frequency-domain analysis, lightweight network**

I. Introduction

Over the past years, deep learning-based change detection techniques have achieved remarkable progress [1], [2]. However, due to the complexity of real-world remote sensing scenarios, the development of change detection still faces several challenges. For instance, in remote sensing change detection tasks, factors such as illumination variations and irrelevant land cover changes often introduce pseudo changes in dual-temporal images. As shown in Fig. 1, areas highlighted by blue boxes in (1) and (2) exhibit some changes unrelated to the detection task, while areas within yellow boxes in (3) and (4) display apparent domain shift due to inconsistent illumination conditions during capture. Such pseudo changes are prone to cause erroneous detection, thereby compromising the reliability of detection results.

To address this issue, some researchers have proposed corresponding solutions. For example, Ma et al. [3] proposed STNet, which incorporates a temporal feature fusion (TFF) module that employs a cross-temporal gating mechanism to selectively enhance changes of interest while suppressing those of non-interest. Huang et al. [4] proposed SEIFNet, which incorporates a spatiotemporal difference enhancement module (ST-DEM) with a dual-branch structure to respectively capture global and local information from dual-temporal features, thereby enhancing the representation of change-related regions and suppressing irrelevant interference. Wei et al. [5] proposed STFF-GA network, which employs a guide aggregation (GA) module that utilizes deep feature guidance (DFG) mappings as prior information to guide the aggregation of multilevel semantic information, thereby correcting the positional information of change objects and filtering pseudo changes and other background noise. Although these methods have achieved promising results, their intricate network architectures introduce excessive parameters and computational cost, which limits their applicability in resource-constrained scenarios.

Therefore, this paper proposes a lightweight Frequency-Spatial Feature Extraction Network (FSFENet). Specifically, to address pseudo changes caused by domain shift between dual-temporal image pairs, we designed a Frequency Filtering Module (FFM). FFM utilizes the Fast Fourier Transform (FFT) and an attention mechanism to adaptively weight the low-frequency components in dual-temporal features, which contain global style information of the images. For pseudo changes caused by irrelevant land cover changes, we propose a Dual-temporal Change Feature Extractor (DT-CFE) and a Multi-level Feature Aggregation Module (MFAM). DT-CFE enhances the spatio-temporal correlation between dual-temporal features and models change feature through channel-wise interaction of pre-change, post-change, and their difference features. MFAM fuses shallow geometric features with deep semantic features and employs an attention mechanism to further enhance regions of interest while suppressing regions of non-interest, thereby generating more accurate change fea-

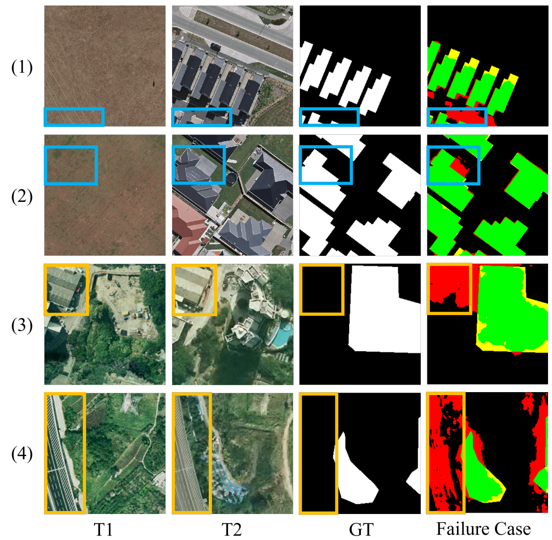

Fig. 1. Examples of some typical pseudo change situations. Red areas indicate erroneous detection, while yellow areas indicate missed detection.

ture representations. Moreover, all modules in FSFENet adopt a lightweight design, achieving a favorable balance between detection accuracy and computational efficiency.

The main contributions of this paper can be summarized as follows:

1) To address pseudo changes caused by domain shift between dual-temporal image pairs, we introduce the FFM, which utilizes Fast Fourier Transform and attention mechanisms for selective filtering, mitigating the impact of domain shift between dual-temporal image pairs and enhancing the effective representation of change targets.
2) To address pseudo changes caused by irrelevant land cover changes, we introduce the DT-CFE and MFAM. DT-CFE is designed to enhance the spatio-temporal correlation between dual-temporal features and models change feature through channel-wise feature interaction. MFAM is employed to enhance the completeness of change regions by fusing low-level and high-level features.
3) Comprehensive experiments are conducted on three public datasets—WHU-CD, SYSU-CD, and DSIFN-CD—and the results show that with only 0.45M parameters and 2.34G FLOPs, FSFENet achieves F1 scores of 92.75%, 81.99%, and 67.77%, respectively, demonstrating the effectiveness of the proposed method.

II. Related Work

A. Frequency Domain Analysis in Change Detection

Frequency domain analysis techniques were originally applied in electrical engineering and communication engineering. However, due to their ability to effectively separate and process the overall structure and detailed textures of images, some researchers have introduced them into change detection tasks. Zhang et al. [6] proposed a spatial-temporal wavelet attention aggregation network, which strengthens the representation of frequency domain information in change regions through a wavelet feature enhancement module to accurately capture complex building contours and subtle texture changes. Liu et al. [7] proposed a multidomain differential excavating network, which tackles the issues of style discrepancies caused by seasonal and illumination variations, as well as foreground-background feature confusion, in remote sensing image change detection through spatial-frequency collaborative analysis and a multi-neighborhood frequency gated attention mechanism. Chen et al. [8] proposed a fourier feature interaction and multi-scale perception model, which alleviates the style inconsistency in dual-temporal images caused by varying imaging conditions in change detection tasks through an adaptive frequency domain filtering module.

B. Feature Enhancement in Change Detection

Feature enhancement plays a vital role in improving the quality of feature representation. In change detection tasks, two common feature enhancement strategies are widely employed: attention mechanism-based enhancement and dual-temporal feature fusion-based enhancement.

In past few years, researchers have developed numerous attention mechanisms applicable to general computer vision tasks. For example, Hu et al. [9] proposed SENet, which first introduced channel attention into CNNs. By explicitly modeling inter-channel dependencies, it addressed the limitation of traditional CNNs in feature representation caused by treating all feature channels equally. Woo et al. [10] proposed the Convolutional Block Attention Module (CBAM), which integrated both channel and spatial attention mechanisms within a single module for the first time. This approach solved the dual problems of inadequate feature representation in traditional CNNs due to uniform processing of all feature channels and spatial locations, as well as the limitation of previous attention mechanisms that focused solely on channel dimension while neglecting spatial dimension.

In mainstream dual-branch change detection architectures, dual-temporal feature fusion techniques plays a crucial role in modeling spatio-temporal correlations between dual-temporal features. Early change detection methods typically adopted element-wise subtraction or channel-wise concatenation for dual-temporal feature fusion [11], [12]. However, these methods failed to adequately model spatio-temporal correlations between dual-temporal features. To address this limitation, researchers have proposed a series of more effective dual-temporal feature fusion techniques. For example, Ma et al. [3] designed a temporal feature fusion module to enhance changes of interest while suppressing changes of non-interest during feature extraction. Huang et al. [4] introduced a spatio-temporal difference enhancement module to capture both global and local information from dual-temporal feature, highlighting differences at identical spatial locations in different temporal phases.

III. METHOD

A. Overview

As shown in Fig. 2, the proposed FSFENet adopts a dual-branch U-shaped encoder-decoder architecture, where EfficientNet-b4 with pre-trained weights serves as the backbone network. The remaining structure primarily consists of three core modules: the Frequency Filtering Module (FFM), the Dual-temporal Change Feature Extractor (DT-CFE), and the Multi-level Feature Aggregation Module (MFAM).

The workflow of FSFENet is as follows: First, a pair of dual-temporal remote sensing images is fed into the EfficientNet-b4 backbone network to extract multi-level features. Then, decoding is performed using the original input and the outputs from the last three layers of the backbone. Specifically, the dual-temporal features at each level are first passed to the FFM, which leverages Fast Fourier Transform (FFT) and attention mechanisms to adaptively weight the low-frequency components containing global style information, thereby mitigating the impact of domain shift between dual-temporal image pairs. Subsequently, the DT-CFE enhances the spatio-temporal correlation between dual-temporal features and models the change features. Finally, the MFAM progressively aggregates and refines the multi-level change features, and the final output is passed to a classifier to obtain the prediction result.

B. Frequency Filtering Module (FFM)

Due to variations in imaging conditions, images of the same region captured at different times often exhibit significant style differences. This can cause substantial variations in the representations of unchanged areas, thereby interfering with the model's judgment. To address this issue, we propose the FFM, whose detailed structure is illustrated in Fig. 2. For the dual-temporal features F_1^i and F_2^i, $i \in \{0,1,2,3\}$, we first transform them from the spatial domain to the frequency domain using 2D-FFT [13]. Since the features after FFT are in complex form, for computational convenience, the real and imaginary parts of the features are stacked along the channel dimension to obtain $F_{FFT,1}^i$ and $F_{FFT,2}^i$. This process can be expressed as:

$$\begin{aligned} F_{FFT,1}^i &= FFT\left(F_1^i\right) \\ F_{FFT,2}^i &= FFT\left(F_2^i\right) \end{aligned} \tag{1}$$

Subsequently, an element-wise subtraction between $F_{FFT,1}^i$ and $F_{FFT,2}^i$ is performed to obtain a coarse representation of feature changes in the frequency domain, which can be expressed as:

$$D_{FFT}^i = F_{FFT,1}^i - F_{FFT,2}^i \tag{2}$$

To leverage the original contextual information for generating more accurate attention maps, we integrate D_{FFT}^i with $F_{FFT,1}^i$ and $F_{FFT,2}^i$ respectively through channel-wise concatenation followed by convolution. Here, depthwise separable convolution is employed instead of standard convolution to maintain the lightweight design of the module. This process can be expressed as:

$$\begin{aligned} \hat{F}_{FFT,1}^i &= \delta(BN(DSC(Cat(D_{FFT}^i, F_{FFT,1}^i)))) \\ \hat{F}_{FFT,2}^i &= \delta(BN(DSC(Cat(D_{FFT}^i, F_{FFT,2}^i)))) \end{aligned} \tag{3}$$

where δ denotes the PReLU activation function, BN represents batch normalization, DSC indicates depthwise separable convolution, and Cat refers to the feature concatenation operation along the channel dimension.

Subsequently, to mitigate the impact of domain shift between dual-temporal image pairs on change detection, we apply an attention mechanism to adaptively weight the frequency-domain features $F_{FFT,1}^i$ and $F_{FFT,2}^i$. This process can be expressed as:

$$\begin{aligned} \widetilde{F}_{FFT,1}^i &= F_{FFT,1}^i \otimes \sigma(\mathrm{Conv}_{1\times1}(\hat{F}_{FFT,1}^i)) \\ \widetilde{F}_{FFT,2}^i &= F_{FFT,2}^i \otimes \sigma(\mathrm{Conv}_{1\times1}(\hat{F}_{FFT,2}^i)) \end{aligned} \tag{4}$$

where $\otimes$ denotes element-wise multiplication, σ represents the Sigmoid activation function, and $\mathrm{Conv}_{1\times1}$ indicates 1×1 convolution.

Finally, we apply the Inverse Fast Fourier Transform (IFFT) to convert the weighted frequency-domain features $\widetilde{F}_{FFT,1}^i$ and $\widetilde{F}_{FFT,2}^i$ back to the spatial domain, obtaining the final output of the FFM. This process can be expressed as:

$$\begin{aligned} F_{FFM,1}^i &= IFFT\left(\widetilde{F}_{FFT,1}^i\right) \\ F_{FFM,2}^i &= IFFT\left(\widetilde{F}_{FFT,2}^i\right) \end{aligned} \tag{5}$$

C. Dual-temporal Change Feature Extractor (DT-CFE)

Effectively extracting change features from dual-temporal features is crucial for improving the model's accuracy in identifying change regions. Existing change detection methods predominantly employ dual-temporal feature fusion strategies. Early change detection approaches often used element-wise subtraction or channel-wise concatenation to fuse dual-temporal features. While these two methods are straightforward, they both have certain limitations: element-wise subtraction focuses only on pixel-level feature differences, overlooking long-range spatial correlations, which can easily lead to misinterpreting pseudo changes as real changes; channel-wise concatenation does not explicitly model feature differences and may introduce excessive redundant information, requiring more parameters and computational effort for subsequent processing. Nowadays, although many more effective dual-temporal fusion strategies have been proposed, they often come with complex computational mechanisms, increasing the operational burden. Therefore, to balance performance and computational cost, we propose a simple yet effective DT-CFE, whose structure is shown in Fig. 2. It enhances the spatio-temporal correlation between dual-temporal features and models change features through channel-wise interactions of the pre-change, post-change, and their difference features, serving as an extension to element-wise subtraction and channel-wise concatenation methods.

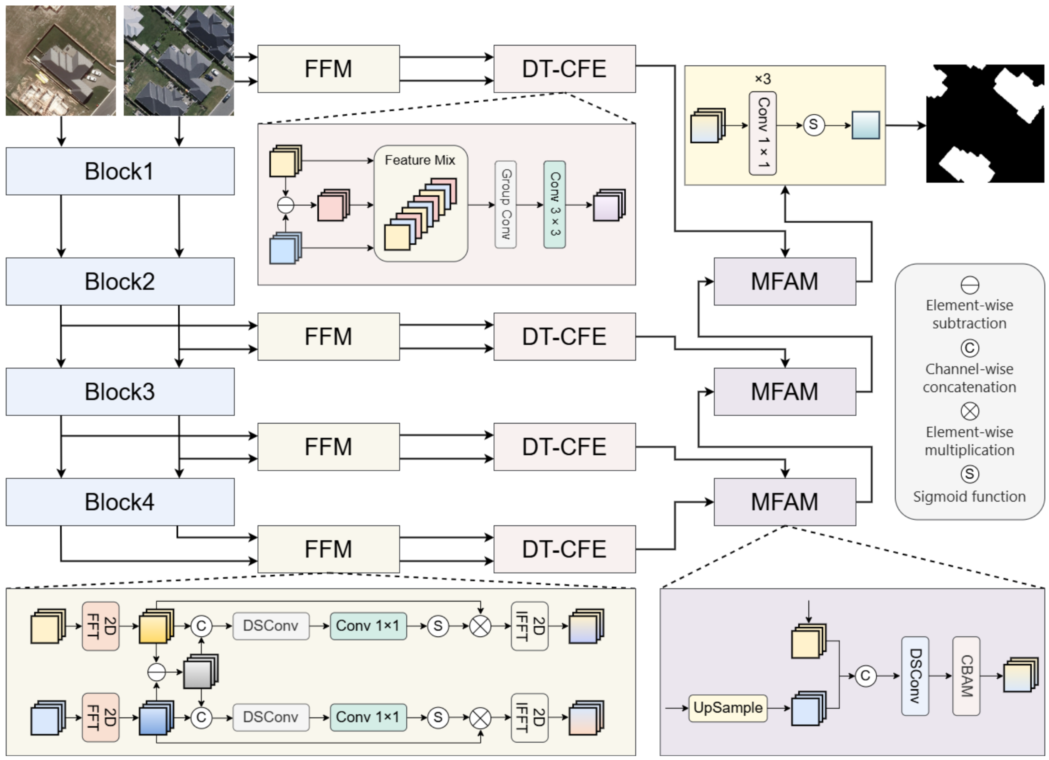

Fig. 2. Overall architecture of the FSFENet.

For the dual-temporal features $F^i_{FFM,1}$ and $F^i_{FFM,2}$ after frequency domain filtering, their differential feature is first computed through element-wise subtraction. This process can be expressed as:

$$D^i_{FFM} = F^i_{FFM,1} - F^i_{FFM,2} \tag{6}$$

Then, following the computational approach outlined in Equation (7), a channel-wise feature mixing operation is performed on the dual-temporal features and their difference features to obtain F^i_{mix}.

$$\begin{matrix} channel\ c \\ of\ F^i_{mix} \end{matrix} = \begin{cases} channel\ k\ of\ F^i_{FFM,1} & c = 3k-2 \\ channel\ k\ of\ F^i_{FFM,2} & c = 3k-1 \\ channel\ k\ of\ D^i_{FFM} & c = 3k \end{cases} \tag{7}$$

where $k \in \{1, 2, ..., C_i\}$, C_i denotes the number of channels in $F^i_{FFM,1}$ and $F^i_{FFM,2}$.

Subsequently, group convolution is employed to fuse $F^i_{FFM,1}$, $F^i_{FFM,2}$, and D^i_{FFM}, enhancing the spatio-temporal correlation between dual-temporal features. This process can be expressed as:

$$\begin{aligned} \hat{F}^i_{mix} = \delta(BN(\mathrm{GroupConv}_{3\times3}(F^i_{mix}, \mathrm{ch}_{in} = 3C_i, \\ \mathrm{ch}_{out} = C_i, groups = C_i))) \end{aligned} \tag{8}$$

where δ denotes the PReLU activation function, BN represents batch normalization, and $Group\mathrm{Conv}_{3\times3}$ indicates group convolution with a 3×3 kernel size.

Finally, we apply a 3×3 convolution to $\hat{F}^i_{mix}$ to obtain the final change features. This process can be expressed as:

$$F^i_{DT-CFE} = \delta(BN(\mathrm{Conv}_{3\times3}(\hat{F}^i_{mix}))) \tag{9}$$

It is noteworthy that neither the element-wise subtraction nor the feature mixing operations introduce additional parameters. Furthermore, the feature fusion operation employs group convolution, ensuring that DT-CFE achieves effective dual-temporal feature fusion without excessive computational overhead, thereby enhancing the model's capability to represent change features.

D. Multi-level Feature Aggregation Module (MFAM)

In change detection tasks, low-level features contain detailed information such as object edges and textures, while high-level features encompass abstract semantic information like object categories. To integrate shallow geometric features with deep semantic features for generating more accurate and detailed change feature representations, skip connections are introduced in the decoder stage, and MFAM is employed to achieve the aggregation of features from different levels. The structure of MFAM is illustrated in Fig. 2.

For the low-level feature F^i_{DT-CFE}, $i \in \{0, 1, 2\}$ and high-level feature F^{i+1}_{MFAM} (denoting F^3_{DT-CFE} as F^3_{MFAM}), a bilinear interpolation upsampling operation is first applied to F^{i+1}_{MFAM} to achieve size alignment with F^i_{DT-CFE}. This process can be expressed as:

$$\hat{F}^{i+1}_{MFAM} = UpSample(F^{i+1}_{MFAM}) \tag{10}$$

Subsequently, channel-wise concatenation followed by depth-wise separable convolution is employed to fuse $\hat{F}^{i+1}_{MFAM}$ and

F^i_{DT-CFE}. This process can be expressed as:

$$Mix = \delta(BN(DSConv(Cat(\hat{F}^{i+1}_{MFAM}, F^i_{DT-CFE})))) \tag{11}$$

where δ denotes the PReLU activation function, BN represents batch normalization, DSC indicates depthwise separable convolution, and Cat refers to the feature concatenation operation along the channel dimension.

Finally, a CBAM block [10] is employed to further enhance the information relevant to the change detection task while suppressing irrelevant interference. This process can be expressed as:

$$F^i_{MFAM} = CBAM(Mix) \tag{12}$$

E. Loss Function

The change detection task can essentially be regarded as a pixel-level binary classification task. Therefore, we employ the cross-entropy loss function as the loss function for model training, which can be expressed as:

$$\mathcal{L}(G,P) = -\frac{1}{H_0 \times W_0} \sum_{i,j} G_{i,j} \log P_{i,j} \tag{13}$$

where G represents the ground truth map, P denotes the model's predicted map, and H_0 and W_0 indicate the height and width of the dual-temporal images, respectively.

IV. Experiments

A. Datasets and Implementation Details

To evaluate the proposed model, we use three public remote sensing datasets. The first is the WHU-CD dataset [14], containing a pair of aerial images (size 32,507 × 15,354 pixels) documenting building changes in New Zealand after the 2011 earthquake and the 2016 reconstruction. The second is the SYSU-CD dataset [15], comprising 800 pairs of aerial images (size 1024×1024 pixels) from Hong Kong, recording various surface changes between 2014 and 2017, including reclamation projects, vegetation changes, and road modifications. The third is the DSIFN-CD dataset [16], consisting of 6 pairs of high-resolution satellite images from six major Chinese cities, covering changes in multiple land cover types such as roads, buildings, farmland, and water bodies. In the experiments, all dataset samples are cropped into 256×256 pixel non-overlapping images and randomly divided into training, validation, and test sets.

The proposed method is implemented based on the PyTorch framework, with model training and testing conducted on an NVIDIA GeForce RTX 3090 GPU (24GB VRAM). All parameter weights are randomly initialized. The AdamW optimizer is employed during training, with an initial learning rate set to 0.003 and a weight decay coefficient of 0.009. A cosine annealing strategy is adopted to dynamically adjust the learning rate. The training process is configured for 200 epochs with a batch size of 8. The model's performance was evaluated using precision, recall, F1 score, IOU and overall accuracy.

B. Comparison Experiments

To better demonstrate the effectiveness of our method, we compare it with several existing change detection approaches: FC-EF [11], FC-Siam-diff [11], FC-Siam-conc [11], IFNet [16], SNUNet [17], L-UNet [18], P2V-CD [19], TinyCD [20], SEIFNet [4], and LCD-Net [21]. To ensure fairness in experimental results, all models were retrained and tested under identical hardware conditions using publicly available GitHub code with default hyperparameters.

a) Quantitative and Qualitative Comparisons: Quantitative comparison results are presented in Table I. Bold and underlined scores indicate the best and second-best performances in each column, respectively. Specifically, on the WHU-CD dataset, the proposed FSFENet achieves the highest performance in F1 and OA metrics, and the second-best in Rec and IoU metrics. Compared to the second-highest scores, FSFENet improves F1 by 0.29% and OA by 0.02%. On the SYSU-CD dataset, FSFENet achieves the best performance in Rec, F1, and IoU, and ranks second in OA. It outperforms the second-best methods by 0.36% in Rec, 0.42% in F1, and 0.59% in IoU. On the DSIFN-CD dataset, FSFENet obtains the highest scores in F1 and IoU, with improvements of 0.86% and 0.97%, respectively, over the second-best results. Qualitative comparison results are shown in Fig. 3. Evidently, compared to other existing methods, the proposed approach demonstrates significantly fewer erroneous detections caused by domain shift and irrelevant land cover changes. Both quantitative and qualitative results confirm that our method achieves competitive performance compared to existing approaches.

b) Computational Efficiency Comparison: Computational efficiency is also a key criterion for evaluating a model. It is typically measured by two metrics: the number of parameters (Params) and floating-point operations (FLOPs). The computational efficiency comparison between FSFENet and other existing methods is shown in Table I. It is evident that among the selected comparative methods, FSFENet's computational efficiency is second only to TinyCD. Nevertheless, it achieves competitive change detection performance, demonstrating its strong capability in balancing detection accuracy and computational efficiency.

C. Ablation Experiments

FFM and DT-CFE are two core components of FSFENet. To further validate the effectiveness of these two modules, we conducted ablation studies for FFM and DT-CFE on two datasets under the same experimental conditions. In the baseline model, DT-CFE is replaced with channel-wise concatenation followed by 3×3 convolution. The quantitative results of the ablation studies are presented in Table II.

a) Ablation on FFM: As shown in Table II, after introducing FFM into the baseline model, the F1 scores on the WHU-CD and DSIFN-CD datasets increase by 0.08% and 1.59%, respectively, while IoU increases by 0.13% and 1.8%. Removing FFM from the complete FSFENet leads to a decrease of 0.4% and 1.35% in F1 scores, and 0.7% and 1.52% in IoU on the WHU-CD and DSIFN-CD datasets, respectively.

TABLE I
COMPARISON OF THE RESULTS WITH OTHER CHANGE DETECTION METHODS ON THE WHU-CD,SYSU-CD AND DSIFN-CD DATASETS

Method	Params (M)	FLOPs (G)	WHU-CD					SYSU-CD					DSIFN-CD				
			Pre	Rec	F1	IoU	OA	Pre	Rec	F1	IoU	OA	Pre	Rec	F1	IoU	OA
FC-EF [11]	1.35	3.58	91.39	86.57	88.92	80.05	99.14	81.04	77.00	78.97	65.25	90.33	64.98	65.05	65.02	48.17	88.11
FC-Siam-diff [11]	1.35	4.73	90.86	79.09	84.57	73.26	98.85	81.51	75.45	78.36	64.42	90.17	58.66	71.65	64.50	47.61	86.60
FC-Siam-conc [11]	1.54	5.33	77.28	86.34	81.56	68.86	98.45	84.73	76.57	80.44	67.28	91.22	55.23	**80.54**	65.53	48.73	85.60
IFNet [16]	35.73	82.26	**96.46**	84.36	90.00	81.82	99.26	82.31	79.94	81.01	68.08	91.18	62.99	65.62	64.28	47.36	87.61
SNUNet [17]	10.20	44.38	86.24	91.51	88.79	79.85	99.08	81.95	78.74	80.31	67.10	90.90	59.23	71.57	64.81	47.94	86.80
L-UNet [18]	8.45	17.33	86.80	86.91	86.86	**98.96**	76.77	83.38	75.47	79.27	65.66	90.69	**65.43**	68.47	66.91	50.28	**88.49**
P2V-CD [19]	5.42	32.96	93.78	86.77	90.14	82.04	99.25	82.91	74.66	78.56	64.70	90.39	63.47	69.76	66.46	49.77	88.04
TinyCD [20]	0.29	1.54	91.12	90.05	90.58	82.79	99.25	79.33	79.90	79.61	66.14	90.35	54.02	77.01	63.50	46.52	84.95
SEIFNet [4]	27.91	8.37	89.37	87.59	88.47	79.33	99.09	**86.40**	76.90	81.37	68.60	**91.69**	61.30	70.16	65.43	48.62	87.40
LCD-Net [21]	2.56	4.45	92.67	**92.26**	92.46	85.99	99.40	84.55	78.80	81.57	68.88	91.60	60.53	67.72	63.93	46.98	87.01
FSFENet	0.45	2.34	93.49	92.03	**92.75**	86.49	**99.42**	83.79	**80.26**	**81.99**	**69.47**	91.68	60.91	76.37	**67.77**	**51.25**	87.65

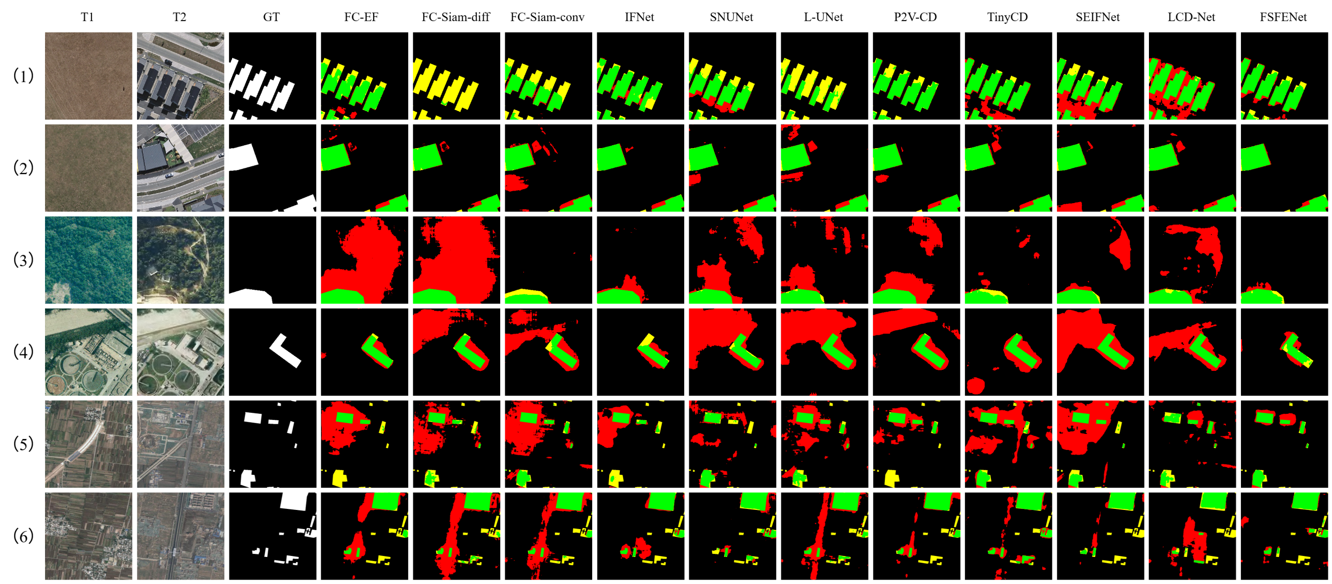

Fig. 3. Qualitative results of the proposed method compared with other methods on the WHU-CD, SYSU-CD, and DSIFN-CD dataset. Different colors are used for a better view, i.e., TP (green), FP (red), TN (black), and FN (yellow).

TABLE II
ABLATION STUDY RESULTS OF FFM AND DT-CFE IN FSFENET ON TWO DIFFERENT DATASETS.

FFM	DT-CFE	WHU-CD		DSIFN-CD	
		F1	IoU	F1	IoU
×	×	91.77	84.80	65.86	49.09
✓	×	91.85	84.93	67.45	50.89
×	✓	92.35	85.79	66.42	49.73
✓	✓	**92.75**	**86.49**	**67.77**	**51.25**

Additionally, Fig. 4 visually demonstrates the impact of FFM on model predictions. Clearly, the inclusion of FFM reduces erroneous detections in pseudo changes caused by domain shift between dual-temporal image pairs.

b) Ablation on DT-CFE: As shown in Table II, introducing DT-CFE into the baseline model improves the F1 scores on the WHU-CD and DSIFN-CD datasets by 0.58% and 0.56%, respectively, and the IoU scores by 0.99% and 0.64%, respectively. Removing DT-CFE from the complete FSFENet reduces the F1 scores on the WHU-CD and DSIFN-CD datasets by 0.9% and 0.32%, respectively, and the IoU scores by 1.56% and 0.36%, respectively. Furthermore, Fig. 5 visually illustrates the impact of DT-CFE on model predictions. Evidently, the inclusion of DT-CFE reduces erroneous detections of pseudo changes caused by irrelevant land cover changes.

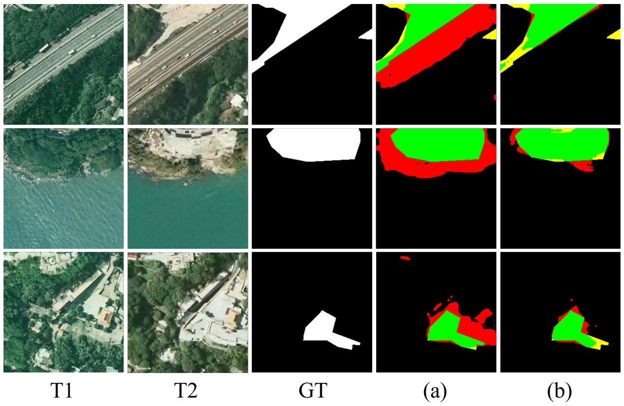

Fig. 4. Visualization results of ablation study on FFM. (a) and (b) show the change prediction results of FSFENet without FFM and FSFENet, respectively. Red areas indicate erroneous detection, while yellow areas indicate missed detection.

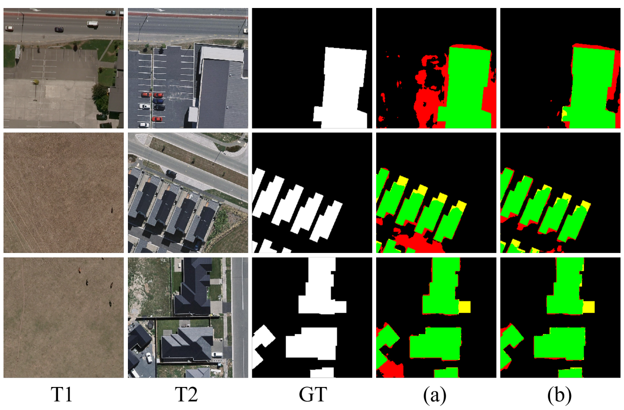

Fig. 5. Visualization results of ablation study on DT-CFE. (a) and (b) show the change prediction results of FSFENet without DT-CFE and FSFENet, respectively. Red areas indicate erroneous detection, while yellow areas indicate missed detection.

V. Conclusion

In this paper, to address the challenge of balancing accurate change region identification and lightweight performance in current change detection algorithms, we propose a FSFENet, which incorporates a FFM to adaptively weight the low-frequency components in dual-temporal features, thereby mitigating the impact of domain shift between dual-temporal image pairs. And a DT-CFE is employed to adaptively extract change features from dual-temporal features through channel-wise feature interaction, reducing the influence of pseudo changes. Furthermore, a MFAM is used to enhance the completeness of change regions by fusing low-level and high-level features. All modules in FSFENet adopt a lightweight design, achieving a favorable balance between detection accuracy and lightweight performance.

References

[1] A. Shafique, G. Cao, Z. Khan, M. Asad, and M. Aslam, "Deep learning-based change detection in remote sensing images: A review," *Remote Sensing*, vol. 14, no. 4, p. 871, 2022.

[2] T. Bai, L. Wang, D. Yin, K. Sun, Y. Chen, W. Li, and D. Li, "Deep learning for change detection in remote sensing: a review," *Geo-spatial Information Science*, vol. 26, no. 3, pp. 262–288, 2023.

[3] X. Ma, J. Yang, T. Hong, M. Ma, Z. Zhao, T. Feng, and W. Zhang, "Stnet: Spatial and temporal feature fusion network for change detection in remote sensing images," in *2023 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2023, pp. 2195–2200.

[4] Y. Huang, X. Li, Z. Du, and H. Shen, "Spatiotemporal enhancement and interlevel fusion network for remote sensing images change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–14, 2024.

[5] H. Wei, N. Wang, Y. Liu, P. Ma, D. Pang, X. Sui, and Q. Chen, "Spatio-temporal feature fusion and guide aggregation network for remote sensing change detection," *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[6] X. Zhang, K. Dong, D. Cheng, Z. Hua, and J. Li, "Stwanet: Spatio-temporal wavelet attention aggregation network for remote sensing change detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025.

[7] J. Liu, S. Li, R. Dian, Z. Song, and X. Kang, "Mdenet: Multidomain differential excavating network for remote sensing image change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–11, 2024.

[8] Y. Chen, S. Feng, C. Zhao, N. Su, W. Li, R. Tao, and J. Ren, "High-resolution remote sensing image change detection based on fourier feature interaction and multi-scale perception," *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[9] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[10] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[11] R. C. Daudt, B. Le Saux, and A. Boulch, "Fully convolutional siamese networks for change detection," in *2018 25th IEEE international conference on image processing (ICIP)*. IEEE, 2018, pp. 4063–4067.

[12] H. Chen, Z. Qi, and Z. Shi, "Remote sensing image change detection with transformers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2021.

[13] P. Duhamel and M. Vetterli, "Fast fourier transforms: a tutorial review and a state of the art," *Signal processing*, vol. 19, no. 4, pp. 259–299, 1990.

[14] S. Ji, S. Wei, and M. Lu, "Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set," *IEEE Transactions on geoscience and remote sensing*, vol. 57, no. 1, pp. 574–586, 2018.

[15] Q. Shi, M. Liu, S. Li, X. Liu, F. Wang, and L. Zhang, "A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection," *IEEE transactions on geoscience and remote sensing*, vol. 60, pp. 1–16, 2021.

[16] C. Zhang, P. Yue, D. Tapete, L. Jiang, B. Shangguan, L. Huang, and G. Liu, "A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 166, pp. 183–200, 2020.

[17] S. Fang, K. Li, J. Shao, and Z. Li, "Snunet-cd: A densely connected siamese network for change detection of vhr images," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021.

[18] M. Papadomanolaki, M. Vakalopoulou, and K. Karantzalos, "A deep multitask learning framework coupling semantic segmentation and fully convolutional lstm networks for urban change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 9, pp. 7651–7668, 2021.

[19] M. Lin, G. Yang, and H. Zhang, "Transition is a process: Pair-to-video change detection networks for very high resolution remote sensing images," *IEEE Transactions on Image Processing*, vol. 32, pp. 57–71, 2022.

[20] A. Codegoni, G. Lombardi, and A. Ferrari, "Tinycd: A (not so) deep learning model for change detection," *Neural Computing and Applications*, vol. 35, no. 11, pp. 8471–8486, 2023.

[21] W. Liu, J. Li, H. Wang, R. Tan, Y. Fu, and Q. Tian, "Lcd-net: A lightweight remote sensing change detection network combining feature fusion and gating mechanism," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025.