Seasonal-Aware Scale-Semantic Consistency Alignment Change Detection Network

Bing Shao[a,b], Hanchao Zhang[a,b,c,*], Mingzhu Li[a,b], Yunkun Zou[c], Ruiqian Zhang[a,b], Xiaogang Ning[a,b], Hao Wang[a,b]

[a] State Key Laboratory of Spatial Datum, Chinese Academy of Surveying and Mapping, Beijing, China - s2704800145@163.com; zhanghc@casm.ac.cn; 2157526732@qq.com; zhangrq@casm.ac.cn; ningxg@casm.ac.cn; wanghao@casm.ac.cn
[b]Liaoning Technical University Geomatics and Geographical Sciences, Fuxin, China
[c]Joint Laboratory of Spatial Intelligent Perception and Large Model Application - ht_wisdom_zyk@htwisdom.cn
*Corresponding author: zhanghc@casm.ac.cn

Keywords: Remote sensing imagery; Change detection; Deep learning; Multi-scale features.

Abstract

Change detection in remote sensing imagery is a crucial method for obtaining dynamic information about land cover. However, pseudo-changes caused by seasonal variations pose a significant challenge to detection accuracy. Seasonal variations, such as vegetation phenology and snow cover, introduce global appearance differences that are often mistaken for actual land cover changes. This phenomenon is particularly prominent in long-term monitoring tasks, where pseudo-changes dominate the detection results. Addressing the issues of global appearance differences and multi-scale feature fusion induced by seasonal changes, We propose a novel Seasonal-Aware Scale-Semantic Consistency Alignment Change Detection Network (SSCANet) for remote sensing image change detection. This approach incorporates a Seasonal-Aware Scale Alignment (ASA) module and a Seasonal-Aware Semantic Guided Fusion (SGF) module. By employing spatial scale transformation and semantic alignment, it reduces information mismatch in multi-scale feature fusion and enhances the perception of details in change regions. Experiments conducted on the GZ-CD and CDD datasets demonstrate that SSCANet achieves overall accuracy with F1 scores of 89.21% and 97.82%, with precision rates of 89.02% and 98.37%, respectively. These results represent significant improvements over other methods, demonstrating that SSCANet outperforms its counterparts in both overall accuracy and seasonal robustness. The findings confirm that this approach effectively suppresses seasonal false changes, enhancing the accuracy and reliability of change detection.

1. Introduction

Change detection in remote sensing imagery, as one of the core technologies in Earth observation (Zhang Jixian, 2021), identifies dynamic changes in land cover and land use attributes by analyzing remote sensing images acquired at different time points (Liu et al., 2023). It is widely applied in natural resource surveys (Dawei et al., 2021), land monitoring (Shunping et al., 2020), urban planning (Jinqi, 2023), disaster assessment (Jinqi et al., 2025; Haigang et al., 2021), and environmental management (Ziyi et al., 2022; Huimin Liu, 2024). With the rapid advancement of high-resolution remote sensing satellite technology (Libo, 2024), the spatial, temporal, and spectral resolutions of remote sensing imagery have significantly improved, providing a rich data foundation for detailed, large-scale monitoring of land surface changes (Xiyao et al., 2025). However, the massive volume of remote sensing data also presents new technical challenges: how to accurately and efficiently extract genuine change information from complex, multi-temporal imagery while suppressing interference caused by non-varying factors such as season, sensor type, and lighting conditions has become a critical issue in remote sensing information processing (Peng et al., 2019).

In change detection tasks, the input dual-temporal images are typically captured at different times. Influenced by factors such as seasonal shifts (Bellini et al., 2023), climate variations (Belova et al., 2023), sensor differences (Du et al., 2023), and imaging angles, they frequently exhibit phenomena like same objects with different spectral signatures and different objects with identical spectral signatures. This generates a large number of “pseudo-changes” (seasonal vegetation phases (Ngadi Scarpetta et al., 2023), shadow displacement, snow cover (Zhang, 2021)). Spectral and textural variations induced by seasonal shifts are particularly pronounced, severely disrupting the accurate extraction of genuine land cover changes. Such pseudo-changes not only complicate change detection but may also cause false positives and false negatives, undermining the reliability of monitoring outcomes. In applications like natural resource regulation, crop monitoring, and urban expansion analysis, the confusion between seasonal pseudo-changes and genuine changes directly impacts the scientific rigor and timeliness of decision-making.

Traditional change detection methods, such as image difference techniques (Bovolo and Bruzzone, 2007), ratio-based approaches (Healey et al., 2005), and principal component analysis(Jia et al., 2016), rely on manually designed features or shallow machine learning models. These methods struggle to handle complex scenes and nonlinear changes in high-resolution imagery. In recent years, deep learning technologies have achieved significant progress in remote sensing change detection due to their powerful feature learning and contextual modeling capabilities (Wang et al., 2024; Tian et al., 2023; Yang et al., 2023). Architectures like Fully Convolutional Networks, U-Nets, and Twin Networks have been widely adopted. Through end-to-end training, they automatically extract differential features, enhancing both the accuracy and automation of change detection. The FC-Siam series (Daudt et al., 2018) proposed by Daudt et al. employs a twin encoder structure to process dual-temporal images, achieving change detection via feature differencing or concatenation. While effective in simple scenarios, these methods struggle with multi-scale object changes. Additionally, U-Net variants like Otsu-UNet (Wang et al., 2025) enhance multi-scale feature fusion through nested skip connections, improving detail retention but remaining sensitive to seasonally induced spectral variations. Another approach, such as IFN, utilizes attention mechanisms to fuse multi-modal features, effectively suppressing some noise in remote sensing image processing, yet still suffers from false detections under complex seasonal transformations. Twin networks combined with Transformer architectures capture global dependencies through self-attention (Zhao et al., 2023;

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-593-2026

Liu et al., 2022), but suffer from high computational overhead and insufficient handling of local details. However, existing methods still face challenges such as inadequate multi-scale feature fusion, insufficient seasonal adaptability, and weak suppression of false variations, limiting their application in real-world scenarios. Therefore, the proposed SSCANet introduces a season-aware module to specifically address these limitations, enhancing the reliability and accuracy of change detection.

Therefore, this paper focuses on addressing the challenges of global appearance variations and multi-scale feature fusion caused by seasonal changes, while enhancing the model's robustness against seasonal disturbances and sensitivity to multi-scale feature variations. To this end, we propose a novel Seasonal-Aware Scale-Semantic Consistency Alignment Change Detection Network (SSCANet). Its core concept involves explicitly modeling the alignment between seasonally invariant features and multi-scale semantic information to enhance change detection accuracy and robustness. The main core innovations of this paper include:

The Seasonal-Aware Scale Alignment (ASA) Module explicitly targets the spatial misalignment in multi-scale features that arises during the fusion of feature pyramids in the decoder. By adaptively upsampling and recalibrating adjacent-level features based on spatial attention, ASA ensures that features from different scales are precisely aligned before fusion, addressing the challenge of inaccurate boundary delineation.

The Seasonal-Aware Semantic Guided Fusion (SGF) Module is designed to tackle the semantic inconsistency and noise interference from seasonal variations. It employs dual-attention (channel and spatial) mechanisms to suppress irrelevant seasonal noise while semantically highlighting potential change regions. This operation enhances the model's ability to discriminate between real changes and seasonal pseudo-changes.

The synergistic operation of these two modules enables SSCANet to achieve more precise and reliable change detection by explicitly modeling the alignment between seasonally invariant features and multi-scale semantic information. This approach provides robust technical support for high-frequency, large-scale dynamic monitoring in real-world scenarios where seasonal variations are inevitable.

The remainder of this paper is structured as follows: Section 2 describes the overall architecture of SSCANet and its two innovative modules. Section 3 details experimental evaluations on the CDD and GZ-CD datasets, which incorporate seasonal variations. Finally, Section 4 provides concluding remarks and further discussion.

2. Seasonal Perception Scale-Semantic Consistency Alignment Change Detection Network

2.1 Methodological Framework

The overall architecture of SSCA Net is based on a dual-branch encoder and triple-branch decoder framework. The core idea is to reduce spatial and semantic differences between multi-scale features through scale alignment and semantically guided fusion, thereby enabling more precise detection of change regions while accommodating sister variations. The main components of the architecture include the encoder, decoder, and fusion decoder.

The encoder employs a dual-encoder architecture with weight sharing, processing dual-temporal images separately to output multi-scale features. Each encoder consists of multiple SE-CBR modules (incorporating convolution, batch normalization, ReLU activation, and SE attention mechanisms), extracting both shallow detail features and deep semantic features(The SE (Squeeze-and-Excitation) module, enhances feature representation by modeling channel-wise dependencies through squeeze and excitation operations). The decoder comprises dual-temporal decoders and a fusion decoder. The dual-temporal decoder fuses encoder features through skip connections to generate change feature maps for each temporal phase. The fusion decoder integrates dual-temporal features and performs multi-scale feature optimization via ASA and SGF modules. The fusion decoder is a specialized component within the overall decoder architecture, specifically designed to integrate multi-temporal features through scale alignment and semantic fusion operations.The initial input is the global semantic features from the deepest encoder output, followed by gradual resolution recovery through upsampling and feature fusion. ASA and SGF modules are applied at different layers of the fusion decoder to achieve scale alignment and semantic enhancement.

The overall workflow is as follows: Dual-branch encoders process front and rear temporal phase images respectively, extracting multi-scale features (including shallow details and deep semantic information), followed by preliminary feature fusion via the decoder; The core innovation stage introduces a Seasonal Awareness Scale Alignment (ASA) module to resolve spatial mismatches in multi-scale features, ensuring consistent feature dimensions. Subsequently, the Seasonal Awareness Guided Fusion (SGF) module employs a channel-spatial attention mechanism to dynamically weight key information, suppressing pseudo-variations like seasonal changes. Finally, through upsampling and convolution operations, it outputs a high-precision binary change detection map, effectively enhancing accuracy and robustness in complex scenes.The overall architecture of SSCANet is shown in Figure 1.

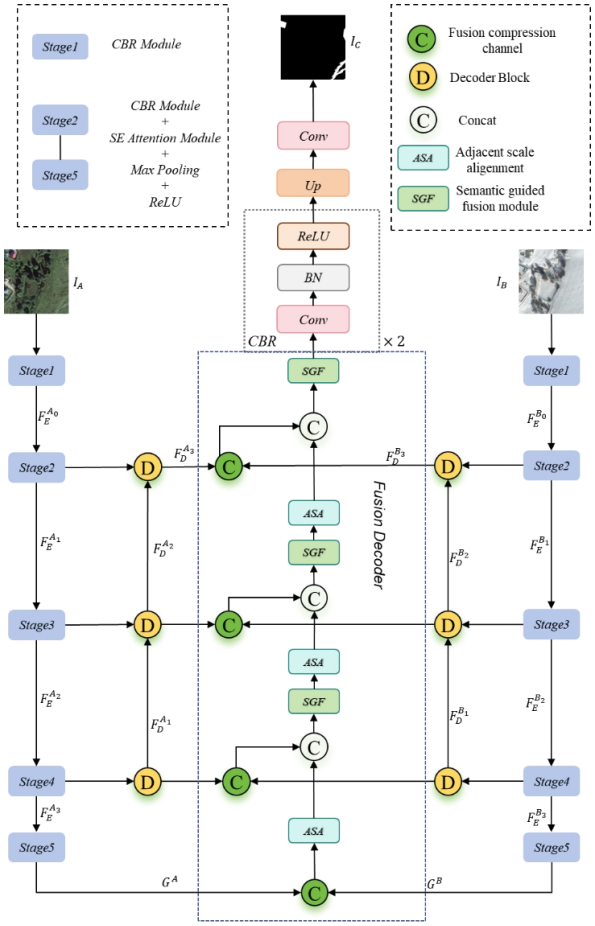

Figure 1. The overall structure of SSCANet

2.2 Seasonal-Aware Scale Alignment module

The ASA module addresses spatial scale mismatches between adjacent features in the fusion decoder, which originate from hierarchical downsampling in the encoder (resolutions reduced by factors of 2, 4, 8, and 16). Its inputs are multi-scale features from the fusion decoder, where dual-temporal decoder features are first fused and channel-compressed via 1×1 convolution. Specifically, ASA takes features F_D^i and F_D^{i+1} as inputs, upsampling the lower-resolution F_D^i to align spatially with F_D^{i+1}. Spatial attention weights are then generated to enhance consistency. The detailed structure is illustrated in Figure 2.The specific workflow is as follows:

It selects the dual-temporal decoded features F_D^i from different layers as input and processes them through a 1×1 convolutional layer. This operation serves two primary purposes: reducing the number of feature map channels to half to decrease computational complexity and parameter count in subsequent steps, and learning and integrating information across the channel dimension without altering spatial dimensions to enhance feature expressiveness. Batch normalization is then applied to stabilize the training process, while the SiLU activation function introduces nonlinearity to help the model better capture data features. Subsequently, the features processed through the above steps undergo linear upsampling, doubling their resolution to align spatially with adjacent high-resolution feature maps. This initially mitigates the scale disparity between feature maps of different resolutions.

$$SA_{up} = up(CBS_{1\times1}(F_D^i)) \quad (1)$$

$CBS_{1\times1}$ denotes a 1×1 convolution using the BN and SiLU activation functions, while upindicates upsampling.

The up sampled features are processed through two convolutional layers and two activation functions to generate spatially aligned weights, thereby enhancing the spatial consistency of the feature maps.

$$SA^{i+1} = Sigmoid\left(conv_{1\times1}\left(ReLU\left(conv_{1\times1}(SA_{up})\right)\right)\right) \quad (2)$$

$conv_{1\times1}$ denotes processing using a 1×1 convolution layer, $ReLU$ represents processing through the ReLU activation function, and $Sigmoid$ denotes processing through the Sigmoid activation function.

The ASA module effectively reduces spatial scale differences between adjacent scale features, making adjacent scale features more consistent in spatial dimensions and feature representation. While the ASA module does not explicitly incorporate seasonal parameters, it addresses seasonal variations indirectly by aligning features that may contain seasonal-invariant characteristics. This alignment helps mitigate the effects of seasonal appearance changes on feature consistency.

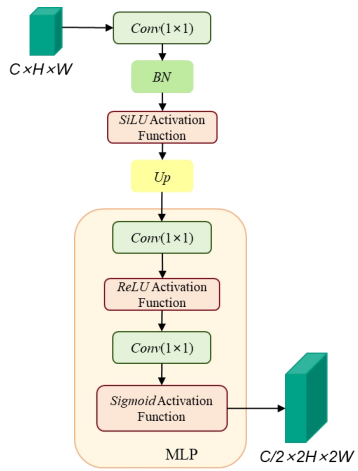

Figure 2. Illustration of the architecture of ASA

2.3 Seasonal-Aware Semantic Guided Fusion module

The SGF module builds upon the ASA module to further address semantic conflicts and redundancy in feature fusion. It employs a channel-space dual attention mechanism to dynamically focus on critical change information, enhancing the semantic consistency of multi-scale features. This improves perception capabilities in areas with subtle changes, such as small objects and edges. The architecture of the SGF module is illustrated in Figure 3. The specific workflow is as follows:

The module takes the feature SA^{i+1} processed by the ASA module and the corresponding dual-time decoding feature F_D^{i+1} as input after concatenation. By concatenating these two features, it generates the integrated feature S_F^{i+1} that combines different information.

$$S_F^{i+1} = cat\left(SA^{i+1}, F_D^{i+1}\right) \quad (3)$$

The cat operation denotes channel-wise concatenation of feature maps
In the channel attention section, the concatenated features undergo global average pooling, compressing them into a scalar to obtain the global information S_C^{avg}. Subsequently, a convolutional layer reduces the original channels to one-quarter of their size. After nonlinear activation via ReLU, the compressed features S_C^{zip} are obtained. A convolutional layer restores the features to their original number. Subsequently, the Sigmoid function normalizes the features, generating the importance weights S_C^{ext} for each channel.

$$S_C^{ext} = Sigmoid\left(conv_{1\times1}\left(S_C^{zip}\right)\right) \tag{4}$$

$conv_{1\times1}$ denotes the operation through a 1×1 convolution layer, while $Sigmoid$ represents processing through the Sigmoid activation function.

Finally, the channel attention weights are multiplied by the original input features to obtain the channel attention-adjusted features S_C^{i+1}, emphasizing the feature information from important channels while suppressing that from less significant channels.

In the spatial attention component, the input features undergo both max pooling and average pooling operations. These features are then concatenated along the channel dimension to yield the feature S_S^{pool}, which incorporates spatial positional information.

$$S_S^{pool} = cat\left(maxpool\left(S_F^{i+1}\right), avgpool\left(S_F^{i+1}\right)\right) \tag{5}$$

$maxpool$ and $avgpool$ represent max pooling and average pooling operations, respectively, while cat denotes a concatenation operation.

After concatenating features across channel dimensions, the S_S^{pool} application applies a 5×5 convolutional kernel to convolve the pooled concatenated features, extracting multi-scale spatial context information. Subsequently, the Sigmoid function generates spatial attention weights A_S. Finally, the obtained spatial attention weights A_S are multiplied by the original input features S_F^{i+1} to yield spatially attention-adjusted features. This operation highlights feature information from critical spatial locations, enabling the model to focus on regions of change.

$$S_S^{i+1} = S_F^{i+1} \times Sigmoid\left(conv_{5\times5}\left(S_S^{pool}\right)\right) \tag{6}$$

$conv_{5\times5}$ denotes the application of a 5×5 convolution layer operation, while $Sigmoid$ represents the Sigmoid activation function operation.

The features adjusted by channel attention and those adjusted by spatial attention are combined to achieve channel-spatial collaborative enhancement. The combined features undergo further refinement through: - Deep separable convolution, batch normalization, and activation operations to reduce parameters while preserving local details; - Dropout for random masking of neurons to prevent overfitting; - A 1×1 convolution layer to adjust the number of feature channels, ultimately producing the semantic feature S^{i+1}.

$$S^{i+1} = conv_{1\times1}\left(dropout\left(DWConv\left(S_C^{i+1}, S_S^{i+1}\right)\right)\right) \tag{7}$$

$conv_{1\times1}$ denotes the operation of a 1×1 convolution layer, $DWConv$ represents depthwise separable convolution, and $dropout$ is a regularization operation.

The SGF module employs a channel-space dual attention mechanism to perform in-depth feature analysis, effectively reducing semantic conflicts and redundant information while enhancing the semantic expressiveness of features. Through a series of processing steps and outputs, including separable convolutions, it efficiently guides the model to focus on critical semantic information, preventing it from being distracted by irrelevant details. The SGF module is not applied to Stage 5 of the fusion decoder because Stage 5 corresponds to the highest resolution feature map of the model, which mainly focuses on the fine-grained boundary refinement of change regions. The semantic information at this stage is relatively shallow, and the dual-attention mechanism of the SGF module has limited optimization effect on shallow features; in addition, omitting SGF at Stage 5 can reduce the computational overhead of the model while ensuring detection accuracy.

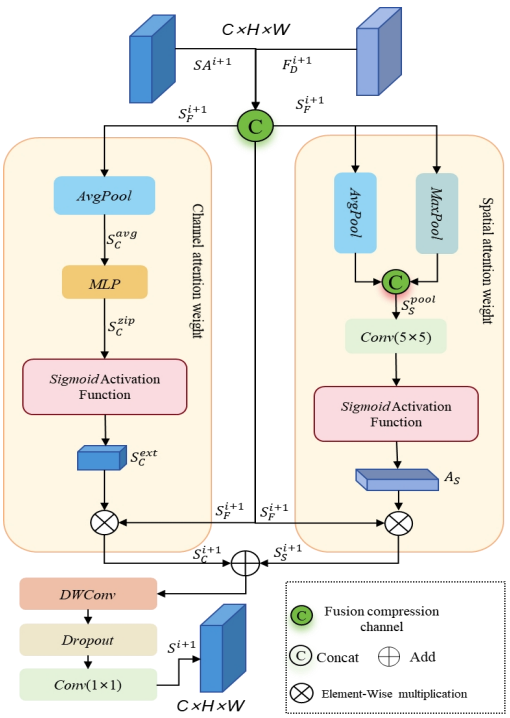

Figure 3. Illustration of the architecture of SGF

3. Experimental Analysis

3.1 Dataset

The CDD dataset encompasses a broader range of elemental changes, including alterations in man-made features like buildings, roads, and vehicles, as well as natural seasonal variations such as vegetation growth and snow cover. The diversity of this dataset aids in evaluating a model's generalization capability under complex changing conditions. The GZ-CD dataset(Peng et al., 2021) features a resolution of 0.55 meters and exhibits pronounced seasonal variations, such as vegetation growth cycles and differences in light conditions, providing an ideal scenario for verifying seasonal robustness. During experimentation, each original image was cropped into

multiple non-overlapping 256×256 image patches. The CDD dataset contains image pairs exhibiting pronounced seasonal variations such as summer-winter cycles, designed to validate the model's robustness against extreme seasonal differences. The GZ-CD dataset primarily features spring-autumn pairs, focusing on building change detection with milder seasonal variations. This contrast enables a comprehensive evaluation of SCANet's generalization capabilities under varying seasonal intensities.

To visually illustrate challenges posed by seasonal and environmental variations, Figure 4 presents representative pseudo-change cases from two datasets. In the CDD dataset, typical pseudo-changes include: (a) vegetation phenological differences between summer and winter scenes—winter deciduous trees may be misclassified as land cover removal; (b) snow cover changes obscuring ground features, creating apparent “changes” where none exist. In the GZ-CD dataset, false change primarily manifests as: (c) differences in vegetation growth cycles between spring and autumn; (d) shifting building shadows due to changing solar angles. These visual examples demonstrate that false change can closely mimic genuine land cover alterations, underscoring the necessity of robust change detection methods like SCANet that incorporate seasonal awareness.

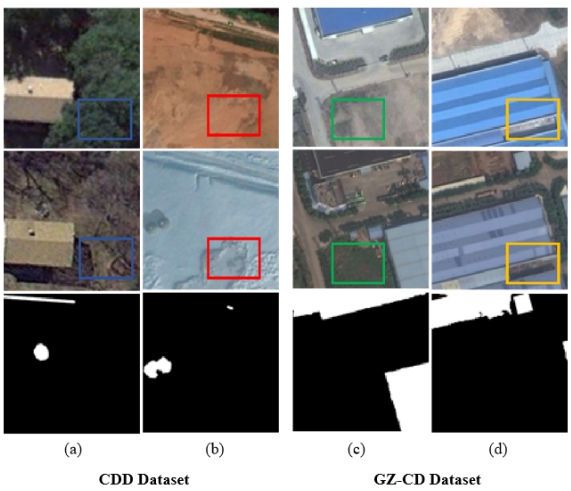

Figure 4. Display of pseudo changes in CDD and GZ-CD dataset

3.2 Comparative experiment

To comprehensively and thoroughly evaluate the effectiveness of the SSCANet model under identical datasets and experimental conditions, we conduct a comparative analysis between SSCANet and both classical and state-of-the-art change detection methods. Training was conducted on an NVIDIA GeForce RTX 4090 computer with an initial learning rate of 0.000001, a preset learning rate of 0.0002, a weight decay coefficient of 0.003, and a batch size of 8. This analysis assesses the performance of various models in change detection tasks, primarily including IFN (Zhang et al., 2020), Interactive Fusion Network, representing the application of attention mechanisms in change detection. L-UNet (Papadomanolaki et al., 2021), Lightweight U-Net architecture, representing the direction of efficient network design.DMINet (Feng et al., 2023), Dense Multi-scale Interaction, representing multi-scale feature fusion strategies. USSFCNet (Lei et al., 2023), Combines spatio-temporal feature fusion, representing cross-temporal modeling approaches.

Method	CDD Dataset			
	F1	OA	Precision	Recall
IFN	95.70	99.00	97.15	94.27
L-UNet	90.60	97.88	94.61	86.91
DMINet	96.55	99.23	96.54	96.57
USSFCNet	95.01	98.82	95.00	95.01
SSCANet	**97.82**	**99.46**	**98.37**	**97.27**

Table 1. Experimental results of different methods on the Change Detection Dataset⋆

Method	GZ-CD Dataset			
	F1	OA	Precision	Recall
IFN	86.16	93.81	88.30	84.12
L-UNet	85.06	93.20	85.60	84.52
DMINet	85.06	93.20	85.60	84.52
USSFCNet	85.07	93.37	87.95	82.37
SSCANet	**89.21**	**95.04**	**89.02**	**89.40**

Table 2. Experimental results of different methods on the GZ-CD Dataset⋆

*All values in the table are expressed as percentages (%) and all tables apply.

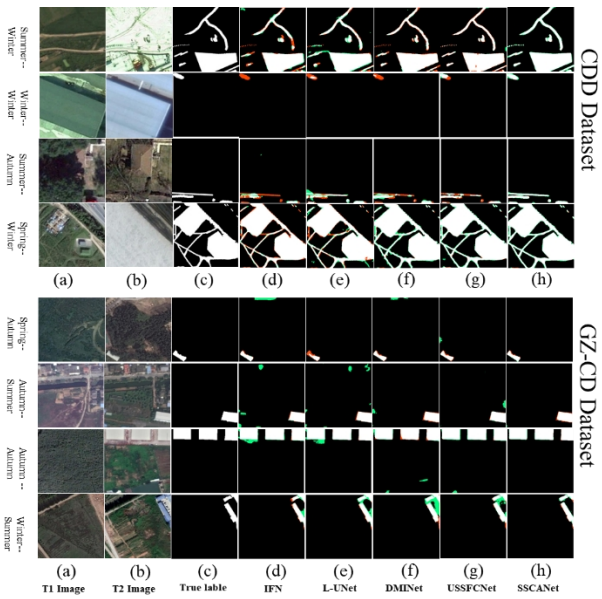

Figure 5. Experimental comparison results on GuangZhou Dataset and Change Detection Dataset# .In the Figure4,(a)T1Image, (b) T2Image, (c)True label, (d)IFN, (e)L-Uet, (f)DMINet, (g)USSFCNet, (h)SSCANet

#Red indicates missed detection areas, green indicates false detection areas, and white indicates areas with positive change detection.

Comparative experiments demonstrate that SSCANet effectively addresses global appearance variations caused by seasonal changes and multi-scale feature fusion challenges.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-593-2026 | © Author(s) 2026. CC BY 4.0 License.

Specifically, we compared SSCANet with IFN, L-UNet, DMINet, and USSFCNet on the CDD and GZ-CD datasets. As shown in Tables 1 and 2, on both the CDD and GZ-CD datasets, SSCANet achieved F1 scores of 97.82% and 89.21%, respectively, with precision rates reaching 97.82% and 89.21%. Compared to other methods, this represents maximum improvements of 2.14% and 3.05%, demonstrating SSCANet's superior efficiency. All F1 scores reported in this paper are calculated at the pixel level, which is the most commonly used evaluation metric for binary change detection in remote sensing imagery and is consistent with the evaluation standards of comparative SOTA methods. The visualization in Figure 5 indicates that red areas represent missed detections, while green areas denote false positives. IFN exhibits more false positives and noticeable missed detections in elongated regions and boundaries. L-UNet and DMINet tend to produce false positives when features are dispersed, whereas USSFCNet shows more missed detections in irregular building configurations. Consequently, SCANet significantly enhances detail perception in variable regions, demonstrating outstanding performance.

3.3 Ablation Studies

To validate the proposed ASA module and SGF module, we incorporated each module into the baseline and conducted comparative analysis. The SSCANet baseline comprises two weight-shared encoders, a weight-shared dual decoder, and a fusion decoder. SSCANet further optimizes semantic consistency in the multi-scale feature fusion process by incorporating the ASA module and SGF module..The experimental results are shown below.

Method	CDD Dataset			
	F1	OA	Precision	Recall
Baseline	97.47	99.38	97.89	97.04
BL+ASA	97.61	99.41	98.04	97.19
BL+SGF	97.69	99.43	98.14	97.25
SSCANet	**97.82**	**99.46**	**98.37**	**97.27**

Table 3. Accuracy evaluation results of SSCANet ablation experiments on the Change Detection Dataset⋆

Method	GZ-CD Dataset			
	F1	OA	Precision	Recall
Baseline	88.90	94.93	89.29	88.51
BL+ASA	89.09	95.03	89.63	88.56
BL+SGF	89.11	94.97	88.39	89.84
SSCANet	**89.21**	**95.04**	**89.02**	**89.40**

Table 4. Accuracy evaluation results of SSCANet ablation experiments on the GuangZhou Dataset⋆

*All values in the table are expressed as percentages (%) and all tables apply.

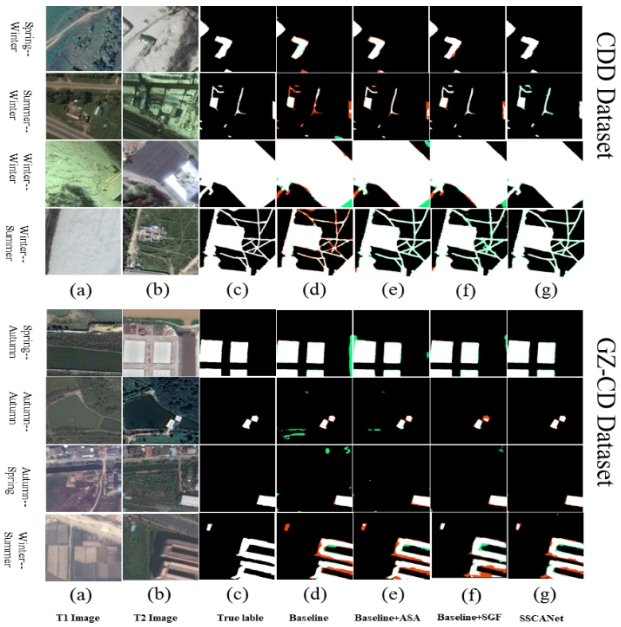

Figure 6. Accuracy evaluation results of SSCANet ablation experiments on GuangZhou Dataset and Change Detection Dataset#. In the Figure 6, (a)T1 Image, (b) T2 Image, (c)True label, (d)Baseline, (e) Baseline+ASA, (f) Baseline+SGF, (g)SSCANet

Ablation studies validated the contributions of each module in SSCANet, demonstrating that the combined use of ASA and SGF effectively addresses global appearance variations caused by seasonal changes and multi-scale alignment issues. As shown in the visualizations in Figure 6, the introduction of the ASA module significantly reduces missed detections, particularly for irregular objects. The SGF module effectively mitigates false positives, enabling precise detection of elongated objects and scattered small features. Results in Tables 3 and 4 demonstrate that both ASA and SGF modules substantially enhance model performance. Compared to the baseline model, the F1 score improved by 0.14% and 0.22% on the CDD dataset, and by 0.19% and 0.21% on the GZ-CD dataset, respectively, after incorporating the ASA and SGF modules. The modest improvement validates the effectiveness of the ASA and SGF modules in optimizing multi-scale feature fusion. It also demonstrates the synergistic effect of both modules in suppressing seasonal pseudo-variations and enhancing multi-scale feature fusion. This indicates that even on mature baselines, performance gains can still be achieved through refined feature alignment and semantic consistency constraints. This further confirms the model's capability to effectively address appearance variations caused by seasonal changes and resolve multi-scale alignment issues.

4. Conclusions

In this paper, we introduce a novel method, SSCANet, which primarily addresses the challenges of global appearance variations caused by seasonal changes and multi-scale feature fusion misalignment. This approach aligns adjacent features spatially via spatial scale transformation and semantic weight generation, thereby mitigating fusion errors. Simultaneously, it employs a channel-spatial dual attention mechanism to dynamically weight features, suppressing background interference while enhancing semantic consistency in change

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-593-2026 | © Author(s) 2026. CC BY 4.0 License.

regions and improving change detection accuracy. Experimental results on the CDD and GZ-CD datasets demonstrate that our method outperforms state-of-the-art algorithms under seasonal variation conditions. Although the absolute improvements in quantitative metrics may appear modest, their significance is substantial when considered against the inherent complexity of the task. These achievements were attained amid intense competitive benchmarks, demonstrating that SSCANet has achieved fine-grained optimization at the performance frontier — where further improvements are increasingly challenging. Quantitative and visual analyses confirm that our approach significantly reduces the likelihood of false negatives and false positives, delivering superior change detection performance. Specifically, the ASA module effectively reduces missed detections of irregular and small objects, while the SGF module mitigates false positives caused by spectral similarities.

Crucially, the practical value of this methodological framework has been validated through its successful application in real-world natural resource monitoring scenarios. SSCANet was deployed to detect cultivated land changes in Dongguang County for the period 2017 – 2018. In this highly challenging and imbalanced scenario (change-to-non-change ratio approximately 1:1140), SSCANet yielded high-precision results and significantly outperformed comparative methods. The framework reliably distinguishes genuine land conversions (farmland to construction land) from seasonal pseudo-changes, providing a high-precision, automated technical support system for the large-scale, high-frequency dynamic monitoring demands of land resource management departments. By reducing manual verification workloads and facilitating more timely, evidence-based policy-making, it converts technical metrics into practical operational benefits. Consequently, the proposed algorithm holds considerable potential for addressing the challenges of global appearance variations and multi-scale feature fusion misalignment caused by seasonal changes.

Despite its advantages, SSCANet has certain limitations: (1) It relies on high-quality manual annotations, with large-scale applications being highly dependent on annotation quality; (2) It is unable to identify the specific type of detected changed features. Therefore, to address limitations arising from label quality, our future work will incorporate semi-supervised and unsupervised learning methods, which will enable the model to learn the intrinsic structural patterns of change within the data and thus reduce its reliance on manual annotations. Concurrently, we will investigate semantic change detection approaches that integrate feature classification information with semantic segmentation techniques to achieve precise identification of distinct change types. This advancement will allow change detection technology to deliver greater practical value across a broader range of real-world applications.

5. Acknowledgements

.
This Research was Supported by Open Project Funds for the Joint Laboratory of Spatial Intelligent Perception and Large Model Application [Grant No. SIPLMA-2025-ZD-02]; the National Natural Science Foundation of China [grant number 42401500]; the Fundamental Research Funds for Chinese Academy of Surveying and Mapping [grant number AR2503].

References

Bellini, E., Moriondo, M., Dibari, C., Leolini, L., Staglianò, N., Stendardi, L., Filippa, G., Galvagno, M., and Argenti, G.: Impacts of Climate Change on European Grassland Phenology: A 20-Year Analysis of MODIS Satellite Data, 15, 218, 2023.

Belova, N., Ermolov, A., Novikova, A., Ogorodov, S., and Stanilovskaya, Y.: Dynamics of Low-Lying Sandy Coast of the Gydan Peninsula, Kara Sea, Russia, Based on Multi-Temporal Remote Sensing Data, 15, 48, 2023.

Bovolo, F. and Bruzzone, L.: A Theoretical Framework for Unsupervised Change Detection Based on Change Vector Analysis in the Polar Domain, IEEE Transactions on Geoscience and Remote Sensing, 45, 218-236, 10.1109/TGRS.2006.885408, 2007.

Daudt, R. C., Saux, B. L., and Boulch, A.: Fully Convolutional Siamese Networks for Change Detection, 2018 25th IEEE International Conference on Image Processing (ICIP), 7-10 Oct. 2018, 4063-4067, 10.1109/ICIP.2018.8451652,

Dawei, Z., Yonghong, J., Jianrong, B., Linyu, L., and Lian, L.: Prediction Method and Application of Dust from Land Creation in Lanzhou Northern Mountain Area, Geomatics and Information Science of Wuhan University, 46, 1106-1113, 10.13203/j.whugis20190254, 2021.

Du, Y., Zhong, R., Li, Q., and Zhang, F.: TransUNet++SAR: Change Detection with Deep Learning about Architectural Ensemble in SAR Images, 15, 6, 2023.

Feng, Y., Jiang, J., Xu, H., and Zheng, J.: Change Detection on Remote Sensing Images Using Dual-Branch Multilevel Intertemporal Network, IEEE Transactions on Geoscience and Remote Sensing, 61, 1-15, 10.1109/TGRS.2023.3241257, 2023.

Haigang, S., Bofei, Z., Chuan, X., Mingting, Z., Zhuotong, D., and Junyi, L.: Rapid Extraction of Flood Disaster Emergency Information with Multi-modal Sequence Remote Sensing Images, Geomatics and Information Science of Wuhan University, 46, 1441-1449, 10.13203/j.whugis20210465, 2021.

Healey, S. P., Cohen, W. B., Zhiqiang, Y., and Krankina, O. N.: Comparison of Tasseled Cap-based Landsat data structures for use in forest disturbance detection, Remote Sensing of Environment, 97, 301-310, 2005.

Huimin LIU, C. Z., Kaiqi CHEN, Min DENG, Chong PENG: Deep learning-based spatio-temporal prediction and uncertainty assessment of urban PM2.5 distribution, 53, 750-760, 10.11947/j.AGCS.2024.20230071, 2024.

Jia, L., Li, M., Zhang, P., Wu, Y., An, L., and Song, W.: Remote-Sensing Image Change Detection With Fusion of Multiple Wavelet Kernels, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9, 3405-3418, 10.1109/JSTARS.2015.2508043, 2016.

Jinqi, G.: Research on building change detection from high-resolution remote sensing images in complex urban scenes, 52, 1233-1233, 10.11947/j.AGCS.2023.20210728, 2023.

Jinqi, Z., Zhangjie, C., Yufen, N., Shuangcheng, Z., Pengfei, Y., and Xiaying, W.: Building Damages Detection of the 2023 Ms 6.2 Jishishan (Gansu, China) Earthquake Using Multi-temporal Dual Polarization ALOS-2 /PALSAR-2 Data, Geomatics and Information Science of Wuhan University, 50, 284-296, 10.13203/j.whugis20240118, 2025.

Lei, T., Geng, X., Ning, H., Lv, Z., Gong, M., Jin, Y., and Nandi, A. K.: Ultralightweight Spatial–Spectral Feature Cooperation Network for Change Detection in Remote Sensing Images, IEEE Transactions on Geoscience and Remote Sensing, 61, 1-14, 10.1109/TGRS.2023.3261273, 2023.

Libo, W.: Vision Transformer Based Building Segmentation Methods for High-Resolution Remote Sensing Images, Geomatics and Information Science of Wuhan University, 49, 2355-2355, 2024.

Liu, M., Chai, Z., Deng, H., and Liu, R.: A CNN-Transformer Network With Multiscale Context Aggregation for Fine-Grained Cropland Change Detection, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15, 4297-4306, 10.1109/JSTARS.2022.3177235, 2022.

Liu, S., Du, K., Zheng, Y., Chen, J., Du, P., and Tong, X.: Remote sensing change detection technology in the Era of artificial intelligence : Inheritance , development and challenges, National Remote Sensing Bulletin, 27, 1975-1987, 10.11834/jrs.20222199, 2023.

Ngadi Scarpetta, Y., Lebourgeois, V., Laques, A.-E., Dieye, M., Bourgoin, J., and Bégué, A.: BFASTm-L2, an unsupervised LULCC detection based on seasonal change detection – An application to large-scale land acquisitions in Senegal, International Journal of Applied Earth Observation and Geoinformation, 121, 103379, 2023.

Papadomanolaki, M., Vakalopoulou, M., and Karantzalos, K.: A Deep Multitask Learning Framework Coupling Semantic Segmentation and Fully Convolutional LSTM Networks for Urban Change Detection, IEEE Transactions on Geoscience and Remote Sensing, 59, 7651-7668, 10.1109/TGRS.2021.3055584, 2021.

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

Peng, D., Bruzzone, L., Zhang, Y., Guan, H., Ding, H., and Huang, X.: SemiCDNet: A Semisupervised Convolutional Neural Network for Change Detection in High Resolution Remote-Sensing Images, IEEE Transactions on Geoscience and Remote Sensing, 59, 5891-5906, 10.1109/TGRS.2020.3011913, 2021.

Shunping, J., Siqi, T., and Chi, Z.: Urban Land Cover Classification and Change Detection Using Fully Atrous Convolutional Neural Network, Geomatics and Information Science of Wuhan University, 45, 233-241, 10.13203/j.whugis20180481, 2020.

Tian, S., Tan, X., Ma, A., Zheng, Z., Zhang, L., and Zhong, Y.: Temporal-agnostic change region proposal for semantic change detection, ISPRS Journal of Photogrammetry and Remote Sensing, 204, 306-320, 2023.

Wang, M., Zhu, B., Zhang, J., Fan, J., and Ye, Y.: A Lightweight Change Detection Network Based on Feature Interleaved Fusion and Bistage Decoding, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17, 2557-2569, 10.1109/JSTARS.2023.3344635, 2024.

Wang, P., Huang, J., Chen, S., Gao, S., Lin, J., and Tao, Y.: A novel Otsu-UNet coupling approach to extract cage aquaculture in changing bay environments, International Journal of Applied Earth Observation and Geoinformation, 142, 104738, https://doi.org/10.1016/j.jag.2025.104738, 2025.

Xiyao, L., Jiayi, L., Jianxun, D., Limin, D., Xiaoyu, Y., Min, P., and Xin, H.: Spatial-Spectral Fusion Lightweight Network Based on Improved Residual Structure for Large-Scale Crop Classification, Geomatics and Information Science of Wuhan University, 50, 937-948, 10.13203/j.whugis20230008, 2025.

Yang, Z., Wu, Y., Li, M., Hu, X., and Li, Z.: Unsupervised change detection in PolSAR images using siamese encoder–decoder framework based on graph-context attention network, International Journal of Applied Earth Observation and Geoinformation, 124, 103511, 2023.

Zhang, C., Yue, P., Tapete, D., Jiang, L., Shangguan, B., Huang, L., and Liu, G.: 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, https://doi.org/10.1016/j.isprsjprs.2020.06.003, 2020.

ZHANG Jixian, L. H., GU Haiyan, ZHANG He, YANG Yi, TAN Xiangrui, LI Miao, SHEN Jing: Study on man-machine collaborative intelligent extraction for natural resource features, 50, 1023-1032, 10.11947/j.AGCS.2021.20210102, 2021.

Zhang, Q.: Characterization of a seasonally snow-covered evergreen forest ecosystem, International Journal of Applied Earth Observation and Geoinformation, 103, 102464, https://doi.org/10.1016/j.jag.2021.102464, 2021.

Zhao, S., Zhang, X., Xiao, P., and He, G.: Exchanging Dual-Encoder–Decoder: A New Strategy for Change Detection With Semantic Guidance and Spatial Localization, IEEE Transactions on Geoscience and Remote Sensing, 61, 1-16, 10.1109/TGRS.2023.3327780, 2023.

Ziyi, Z., Yanfang, L., Yang, Z., Yaolin, L., Yanchi, L., and Qiran, R.: Spatial Non-Stationary Response of the Ecosystem Services Synergy and Tradeoff to Influencing Factors: A Case Study of Ecological Function Area in Fujian Province %J Geomatics and Information Science of Wuhan University, 47, 111-125, 10.13203/j.whugis20200700, 2022.