Transactions on Environment, Energy and Earth Sciences
ISSN: 2960-1924, eISSN: 2960-2300 | Volume 5
EMCEME 2025

A Novel Transformer-Enhanced Framework for Multi-scale Feature Fusion in Remote Sensing Change Detection

Sizhe Fan [1, *], La Zhu [2]

[1] School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science & Technology, Nanjing, China, 210044

[2] School of Geographical Sciences, Nanjing University of Information Science & Technology, Nanjing, China, 210044

* Corresponding Author Email: 18921449219@163.com

Abstract. With the rapid advancement of remote sensing technology, high-resolution satellite imagery has become increasingly accessible for monitoring urban expansion and surface changes. However, change detection in remote sensing images faces several significant challenges, including inconsistent image registration, inadequate handling of multi-scale features, class imbalance in training samples, and the occurrence of pseudo-changes that appear as modifications but represent no actual surface alterations. To address these limitations, this paper presents MA-CDNeXt, a novel deep learning model that enhances the Transformer architecture for improved change detection in high-resolution remote sensing imagery. The proposed model incorporates three key innovations: (1) an enhanced feature encoder that better captures spatial-temporal information, (2) integration of MAFGNet's multi-scale attention fusion mechanism within CDNeXt's spatial-temporal module to effectively process features at different scales, and (3) an optimized feature fusion decoder equipped with a dual-branch graph convolution module for more accurate change localization. To mitigate the class imbalance problem commonly encountered in change detection datasets, we employ a combined loss function incorporating cross-entropy loss, Dice loss, and focal loss. Experimental evaluation on the LEVIR-CD+ dataset demonstrates the effectiveness of our approach, achieving an overall accuracy of 84.09% and a recall rate of 78.39%. The results indicate that MA-CDNeXt successfully addresses the key challenges in remote sensing image change detection and provides a robust solution for monitoring surface changes in high-resolution imagery.

Keywords: Remote sensing image change detection; Multi-scale feature fusion; Enhanced Transformer; Attention mechanism; Graph convolutional network.

1. Introduction

In recent years, the development of remote sensing technology has been particularly rapid, and high-resolution remote sensing images have become the main tool for monitoring land changes, assessing ecological environments, and observing urban expansion [1] [2]. With continuous advancement of satellite technology [3], we can now easily obtain large amounts of remote sensing data from different periods, bringing unprecedented opportunities for dynamic Earth observation [4]. This technological evolution has made change detection a critical component in understanding surface transformations and environmental monitoring.

Traditionally, change detection methods face significant challenges when dealing with high-resolution remote sensing images [5]. These challenges encompass several critical aspects: image registration issues caused by measurement platform position uncertainties and varying atmospheric conditions[6] [7], difficulties in handling multi-scale characteristics of buildings and roads[8] [9], class imbalance problems where unchanged regions vastly outnumber changed regions[10], and pseudo-changes caused by illumination variations and seasonal differences that create false change signals [11]. These limitations have historically constrained the effectiveness of conventional approaches such as Principal Component Analysis (PCA), Change Vector Analysis (CVA), and direct pixel comparison methods.

The emergence of deep learning has revolutionized change detection capabilities, gradually replacing traditional methods due to its ability to extract multi-level features and identify complex correspondences between multi-temporal images [12] [13] [14] [15]. Convolutional Neural Networks (CNNs) have demonstrated superior performance in change detection tasks, yet they suffer from inherent limitations in handling long-range dependencies due to their localized convolutional operations, particularly when global spatial-temporal understanding is required [16]. While Transformer models can capture global dependencies through self-attention mechanisms, their computational complexity becomes prohibitive when processing high-resolution images. [17] Recent advances have explored hybrid approaches, such as Swin Transformer implementations and methods combining local and global information processing, showing promising results in addressing these fundamental challenges [18].

Analyzing change detection dynamics enables researchers and practitioners to develop more robust monitoring systems for environmental assessment, urban planning, and disaster management. The advantage of advanced change detection lies in providing comprehensive frameworks for understanding temporal surface variations, enabling more strategic responses to environmental changes and supporting evidence-based decision-making processes. However, accurately defining optimal feature representations and applying them effectively in diverse scenarios presents ongoing challenges.

Moreover, the balance between computational efficiency and detection accuracy remains a critical concern. Current methods struggle with the trade-off between model complexity and practical applicability, while the generalizability across different geographical regions and temporal scales continues to be debated. The dynamic nature of environmental changes and the variety of imaging conditions make comprehensive change detection analysis particularly challenging.

To address these limitations, this study develops MA-CDNeXt, a novel framework that enhances the Transformer architecture through improved multi-scale feature fusion capabilities for high-resolution remote sensing image change detection. The model integrates an enhanced feature encoder utilizing improved ConvNeXt modules, incorporates MAFGNet's multi-scale attention fusion mechanisms into CDNeXt's temporal-spatial interaction module, and implements an optimized feature fusion decoder with dual graph convolution modules. To handle class imbalance issues, the study employs a composite loss function combining cross-entropy loss, Dice loss, and focal loss. The research systematically evaluates model performance using the LEVIR-CD+ dataset, demonstrating significant improvements in detection accuracy and computational efficiency compared to existing methods.

2. Model Construction and Algorithm Design

In recent years, remote sensing technology has developed particularly rapidly, and high-resolution remote sensing images have become particularly useful in many ways, such as being able to observe changes on the ground, assess the ecological environment, and analyze urban expansion. However, traditional change detection methods still have many problems in practical applications, especially when dealing with complex scenes. The first problem is that images taken at different times often deviate during the registration process, mainly because the location of the measurement platform is not fixed, and the weather conditions and imaging angles are different at the time of shooting, so it is easy to cause image deformation. Combined with the structure and multi-scale features, a new change detection method was designed, which well integrated the characteristics of the CEXt network framework. In addition, some new ideas for target detection by MaRSNet were also referenced, and finally a good change detection system was built.

The method described in this paper mainly uses four parts to make up, including encoders and decoders, as well as interactive parsers and detectors, and the specific structure can be seen in Figure1.

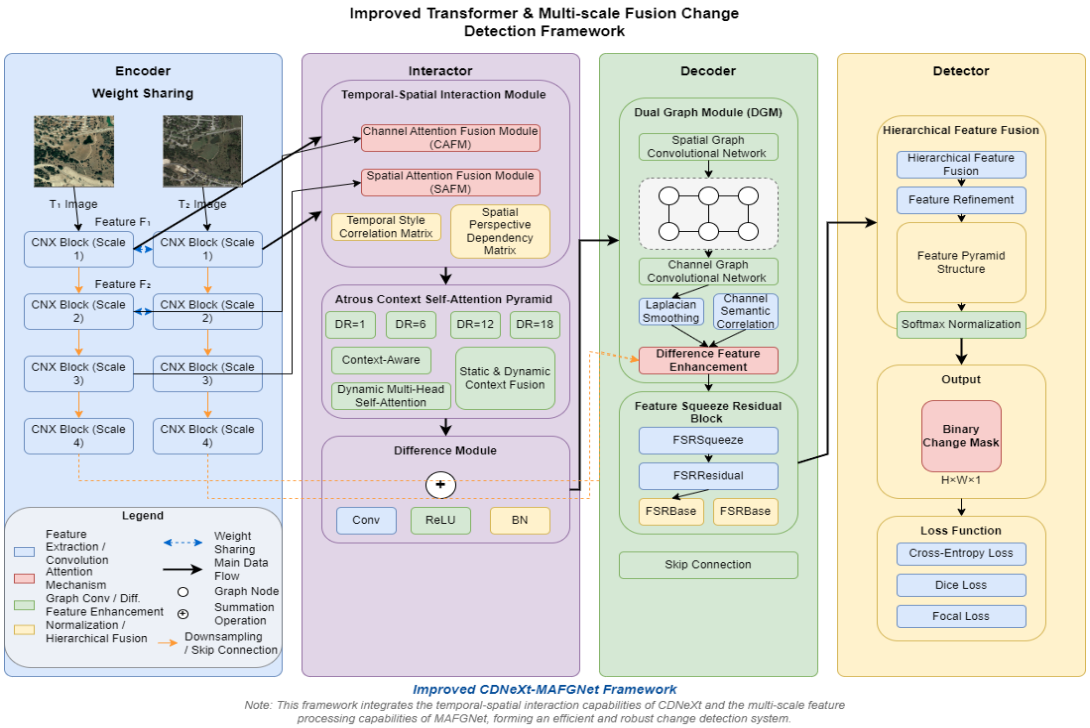

Figure 1. Improved Transformer & Multi-scale Fusion Change Detection Framework.

The framework proposed in this paper mainly uses two remote sensing images at different times as input data, which contain information on the three dimensions of channel count, height, and width. First, a multi-scale feature encoder is used to extract different levels of feature representations, and then establish temporal correlations through an improved space-time interaction module. Next, these processed features are fed into the feature fusion decoder for integration, and finally a specially designed area detector to generate a divalued change detection diagram.

$$M = \mathcal{D}(\mathcal{U}(\mathcal{T}(\mathcal{E}(I_A, I_B)))) \tag{1}$$

In this framework, D represents the coding module, and it is also responsible for interaction and decoding, and finally completes the task of detecting the module, and the result is a binary form of a change mask.

2.1. Multi-Scale Feature Encoder

The paper uses an improved ConvNeXt block as a core part of the multiscale feature encoder, which extracts features well on each branch of different sizes. This module is the basic component of the entire system, and its main role is to extract and process the original image features on the same scale. The specific details of the entire calculation process can refer to Equation (2), and the specific structural design of this module is detailed in Equation (3).

$$F^i = \mathcal{F}_{\text{CNX}}(F^{i-1}) \tag{2}$$

$$\mathcal{F}_{\text{CNX}}(x) = \text{Conv}_{1\times1}(\text{GELU}(\text{Conv}_{1\times1}(\text{LN}(\text{Conv}_{7\times7}(x))))) \oplus x \tag{3}$$

Here is how the convolutional operation is done, what is the normalization of the layer, and how the activation function of the Gauss error linear unit is used, and finally also mentioned how to add between the elements. Specifically, the convolutional operation is to calculate the data according to a certain rule, the layer normalization is to make the data more stable, Gauss error linear unit is a special activation function, the elements add up to the value of the corresponding position of the two things.

This paper wanted to make the features of different scales perform better, so we added a scale adaptive normalization mechanism based on the standard ConvNeXt Block. This new mechanism can make some minor adjustments to the output characteristics of the branches at each scale, and how to adjust can refer to equation (4).

$$\hat{F}^i = \frac{F^i - \mu^i}{\sqrt{(\sigma^i)^2 + \epsilon}} \odot \gamma^i + \beta^i \tag{4}$$

This section is about the average and standard deviation of the scale feature, and a small constant is added to ensure that the value does not go wrong, and the scale and bias parameters that can be adjusted by itself. With this scalable normalization method, the model can learn to adjust the strategy according to the characteristics of different sizes, so that the processed features will be more consistent and differentiated, especially suitable for dealing with objects with large changes in size in remote sensing images.

2.2. Enhanced Temporospatial Interaction Module

2.2.1. Spatial Attention Fusion Module (SAFM)

The paper uses two equations, (5) and (6), to calculate spatial attentional weights, by analyzing the statistical characteristics of the channel to produce results.

$$\begin{aligned} S_{sp} = \mathrm{Concat}(\mathrm{Mean}(F'^l, dim=1), \mathrm{Max}(F'^l, dim=1), \\ \mathrm{Mean}(F'^{l+1}, dim=1), \mathrm{Max}(F'^{l+1}, dim=1)) \end{aligned} \tag{5}$$

$$\theta = \mathrm{Sigmoid}(\mathrm{Conv}_{3\times3}(S_{sp})) \tag{6}$$

The spatial fusion process is expressed in equation (7):

$$F_{SAFM} = F'^l_A \odot \theta + F'^{l+1}_A \odot (1 - \theta) \tag{7}$$

2.2.2. Temporal Correlation Modeling

The enhanced time interaction and aggregation modules proposed in this paper are mainly used to deal with the relationship between time and space. This module is based on the original improvement, it can better analyze the characteristic diagrams of the two time points before and after the changes, and find their links by designing correlation matrices. These connections are not only changes in time, but also involve spatial reactions and interactions.

Initially, features are projected into a latent space as defined in equations (8) and (9):

$$F_{SAFM} = F'^l_A \odot \theta + F'^{l+1}_A \odot (1 - \theta) \tag{8}$$

$$Q_i = \mathrm{Conv}^{(i)}_{1\times1}(\mathrm{BN}^{(i)}(F_i)), \quad K_i = Q_i^T \tag{9}$$

In neural network training, batch normalization operations are usually represented by Where, a symbol that often appears in the paper. Here, subscripting operations specifically designed to generate query representations are specifically indicated, so that readers can easily distinguish between different computational steps.

The model then calculates the relationship between the temporal and spatial characteristics of the two tenses, specifically the two-part relationship. One part is about the correlation of the time style, and the other is the dependence of spatial perception, which can be calculated in a way that refers to the specific formulas given in equations (10) and (11).

$$S_1 = \text{Softmax}(Q_A \otimes K_B^T) \tag{10}$$

$$S_2 = \text{Softmax}(Q_B \otimes K_A^T) \tag{11}$$

We use a correlation matrix to represent these relationships, which can simultaneously reflect the laws of change in time and spatial interconnections. By analyzing these correlation matrices, we can make some improvements to the original eigenvalues to better reflect the interaction of time and space. Specifically, the two formulas given in the paper are calculated and processed.

$$Z_A = S_1^T \otimes V_B \otimes S_2 \tag{12}$$

$$Z_B = S_1 \otimes V_A \otimes S_2^T \tag{13}$$

At the end of the step, these structurally enhanced feature plots are put together and the most started feature diagrams, plus the remaining connected parts, so that the final output of this module can be obtained. The whole process can be represented by equations (14) and (15), which are a good illustration of how the feature diagrams are combined.

$$\hat{F}_A = \text{BN}(\text{Conv}_{1\times1}(Z_A)) \oplus F_A \tag{14}$$

$$\hat{F}_B = \text{BN}(\text{Conv}_{1\times1}(Z_B)) \oplus F_B \tag{15}$$

2.3. Feature Fusion Decoder

In the change detection task, the main task of the decoding stage is to resize the deep semantic features extracted by the encoder to the same size as the original image, and to find out exactly where the changes have changed. In order to make the two steps of feature fusion and change recognition fast and good, this article has designed a new feature fusion decoder. The decoder contains a special dual-graphic convolutional module, the DGM module, which can be improved on the dual-graphic attention structure of the MAFGNet network, and is now used in the CDEXt feature fusion decoder.The long-term dependence on the two dimensions of the channel can better integrate the characteristics of different scales together, and the recognition effect of the changing area is also stronger.

2.3.1. Difference Feature Enhancement

Difference features are computed through element-wise subtraction:

$$F_{diff} = \text{Diff}(\hat{F}_A, \hat{F}_B) = |\hat{F}_A - \hat{F}_B| \tag{16}$$

These difference features are then enhanced:

$$F_{enhanced} = \text{CGCN}(\text{SGCN}(F_{diff})) + F_{diff} \tag{17}$$

2.3.2. Feature Squeeze Residual Block (FSR)

FSR is actually the most basic component of the decoder, and its specific construction method can be seen in the formula given by Equation (18).

$$\text{FSR}(x) = \text{FSRResidual}(\text{FSRSqueeze}(x)) \tag{18}$$

Where FSRSqueeze performs channel compression:

$$x_{squeeze} = \text{GELU}(\text{BN}(\text{Conv}_{1\times1}(x))) \tag{19}$$

And FSRResidual performs feature refinement:

$$\text{FSRResidual}(x) = \text{Conv}_{1\times1}(\text{Conv}_{3\times3}(x)) \oplus x \tag{20}$$

2.4. Change Region Detector

In the system we designed, the changing area detector uses a hierarchical feature fusion method that integrates multi-scale semantic information from different encoder stages. The decoder generates feature diagrams of different resolutions, and these feature maps contain different semantic abstraction levels. The lighter color feature diagram contains a lot of detail information, which can see small changes in space, and the deeper feature map captures the more advanced semantic content.The method.

$$F_{fused} = \text{Concat}(F'_1, F'_2, F'_3, F'_4) \tag{21}$$

This section is about adjusting the feature diagrams of different resolutions to the same size. After the feature fusion steps, this article uses a module called FSR to further deal with these features, the specific structure of which can refer to the design given by Equation (22). The full name of the FSR module is the feature extrusion residue block, which mainly plays the role of optimizing the feature, so that the fused features can better express information.

$$F_{final} = \text{FSR}(\text{FSR}(F_{fused})) \tag{22}$$

The final change mask is derived through softmax normalization:

$$P = \text{Softmax}(\text{Conv}_{1\times1}(F_{final})) \tag{23}$$

$$M = \arg\max(P) \tag{24}$$

3. Experimental Analysis and Results Discussion

3.1. Overall Performance Metrics

Table 1 shows the specific performance of our MA-CDNeXt model on the LEVIR-CD+ test set. From the experimental results, the model has achieved good results on several important indicators, including accuracy, recall rate, F1 score, and IoU value. These data are a good proof that the model structure we have proposed for remote sensing image change detection is indeed effective and has advantages over other methods. This shows that the network architecture we designed to do a good job in handling such tasks.

Table 1. Overall performance metrics of the MA-CDNeXt model on the LEVIR-CD+ dataset.

Method	Avg Loss	Accuracy	Precision	Recall	F1 Score	IoU
MA-CDNeXt	0.0324	0.9873	0.8409	0.7839	0.8114	0.6826

This paper compares our improved models with several common change detection methods, including FC-EF and FC-Diff. On the LEVIR-CD+ dataset, we get specific comparison results, which are detailed in Table 2. With this comparison, it is more clear how our improved model behaves.

Table 2. Performance comparison on LEVIR-CD+ dataset.

Method	P(%)	R(%)	F1(%)	IoU(%)
FC-EF	70.60	73.64	72.09	56.36
FC-Diff	79.94	72.55	76.06	61.37
MA-CDNeXt	84.09	78.39	81.14	68.26

From the experimental data in Tables 1 and 2, the MA-CDNeXt model presented in this paper performed particularly well on the LEVIR-CD+ dataset. The model has achieved good results on many important indicators, such as accuracy, accuracy, recall rate, F1 score, and IoU value, and the average loss is also very low. Compared with the old methods of FC-EF and FC-Diff, MA-CDNeXt has made great progress in all aspects. The recall rate was 78.39%, which is also better than FC-EF's 73.64% and FC-Diff's 72.55%; the F1 score increased to 81.14%, much higher than FC-EF's 56.36% and FC-Diff 61.37%. These data show that our model is indeed much stronger than the traditional method.

3.2. Visualization Result Analysis

This paper uses Figure 1 to show the detection effect of the MA-CDNeXt model on the LEVIR-CD+ test set, which will look clearer. There are four columns on the graph, the first column is the T1 period image before the change, the second column is the T2 period image after the change, the third column is the real-world label of the actual change situation, and the last column is the change map predicted by our MA-CDNeXt model.

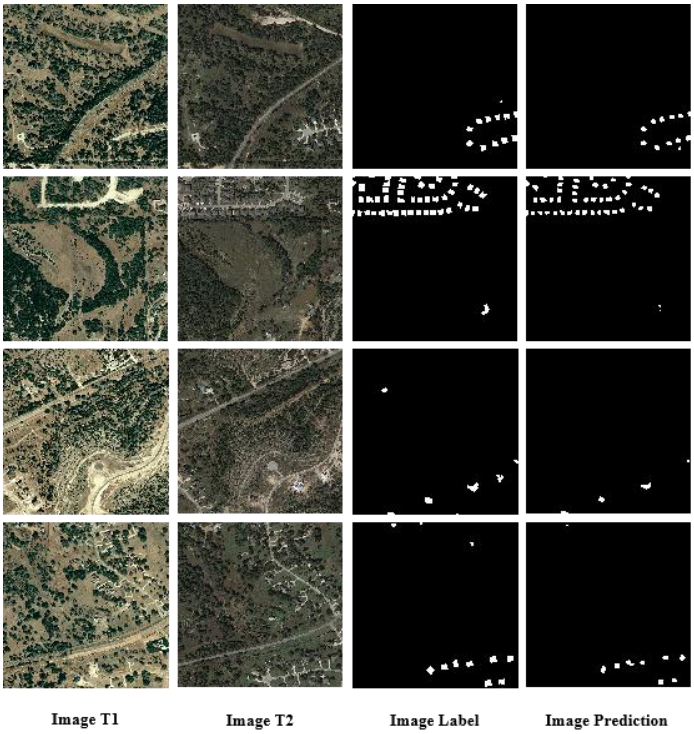

Figure 1. Example change detection results of MA-CDNeXt model across representative scenarios.

Each row is lined with a complete sequence of changes: on the left is an image of T1 before change, followed by a modified T2 image, followed by a ground-based truth map of actual changes, and finally a map of changes we have predicted using the MA-CDNeXt model. A closer look at these results shows that our MA-CDNeXt model is very stable in identifying changes in different buildings, even in very complex contexts.

The first and third examples are about changes in the roadside environment, although these changes are subtle, but the model can still be well-discovered, which shows that even in the complex areas of both natural landscapes and artificial buildings, it can accurately identify the changes. The second

example is the change of rectangular plots, and the results predicted by the model are basically the same as the actual situation, which proves that it can be particularly flexible when it can be seen that the different structural characteristics of the building. Changes in shape.

From the prediction results, the model can be more accurately divided into the boundaries of the changing area on many different samples. These predictions rarely appear to be over-divided or insufficiently divided, indicating that the model's judgment of the spatial position is quite accurate, and the processing of boundary details is also very good. The model also well overcomes the interference caused by background noise, and the detection accuracy remains at a high level even in those more complex areas.

3.3. Ablation Experiments

The main reason for this paper is to see whether several key parts of the MA-CDNeXt model are useful, including multi-scale feature extraction MSF, improved temporal and spatial attention mechanism MSAF is TIAM, and attention-guided multi-scale selection of ACSP. We first use a simple model as a comparison, this model only CDNeXt's original attention mechanism, there is no MSF module, but retains the basic differences in feature enhancement DFE module.

3.3.1. Quantitative Analysis

Table 3-3 gives specific performance data for several different model configurations, which are expressed in percentage form, and each number is accurate to four digits after the decimal point. The purpose is to make the data look more accurate and it is convenient for us to compare the differences between different models.

Table 3. Ablation experiment results for different MA-CDNeXt model components.

Methods	Accuracy	Precision	Recall	F1 Score	IoU
Baseline CDNeXt	0.9833	0.8057	0.6858	0.7409	0.5884
Baseline CDNeXt + MSF	0.9832	0.8108	0.6736	0.7359	0.5821
Baseline CDNeXt + MSF + MSAF	0.9826	0.7831	0.6920	0.7348	0.5807
MA-CDNeXt	0.9873	0.8409	0.7839	0.8114	0.6826

Based on the experimental data presented in Table 3, we conduct the following analysis:
MSF Module Effectiveness Analysis:

This paper compares the difference in the effectiveness of the two methods: Baseline CDNeXt and Baseline CDNeXt plus the MSF module. After adding the multi-scale feature extraction module to the experiment, the accuracy of the model increased from 0.8057 to 0.8108, which is a small but still visible. However, the recall rate is slightly reduced, and this change makes the two important indicators of F1 score and IoU also get a little worse.

MSAF (Enhanced TIAM) Module Effectiveness Analysis:
This paper attempts to add the improved time and space attention module MSAF on the basis of the original CDNeXt plus MSF model, and found that this change has brought some interesting changes. From the experimental results, MSAF does allow the model to find more areas of change, and the recall rate data is better than before, but the accuracy rate has decreased a little. If the two important indicators of F1 score and IoU are still relatively different, the new model may not change much, and even slightly worse than the previous.Adjustment, there should be more room for cooperation between them has not been fully exerted.

ACSP Module Comprehensive Effect Analysis (Complete MA-CDNeXt Model):

The paper concludes with a combination of the multi-scale selection module guided by the focus and the previous base model configuration to complete the MA-CDNeXt architecture we designed. With the addition of this module, all indicators have been greatly improved, with accuracy from 0.7831 to 0.8409, and the recall rate increased from 0.6920 to 0.7839. This result, F1 scores 0.8114 and IoU have 0.6826. The AF module works well together, bringing together information at different scales and fine time attention, and finally making the detection performance better.

When the two modules, MSF and MSA, used alone or in part, were mixed and less stable. But when they were completely combined with the ACSP modules, the results were good and well matched. It was this complete combination of the parts that allowed our MA-CDNeXt model to perform better on all important indicators than the previous benchmark model and the intermediate version, which proved that the MA-CDNeXt architecture we designed was indeed justified.

3.3.2. Visualization Comparative Analysis

This paper wants everyone to see clearly what impact each component has on the test results, so we specially selected some typical test samples, and then compare the predictions of the models under different configurations together, and can see Figure 3-2.

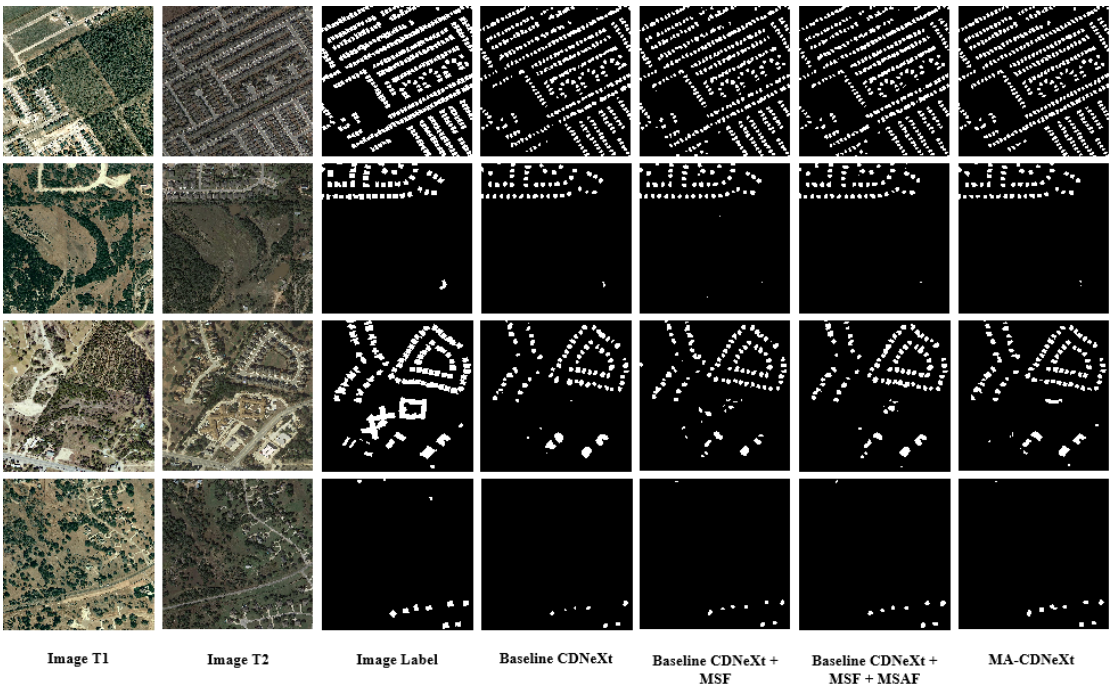

Figure 2. Comparison of detection results on selected test samples across different model configurations.

The visualization from Figure 2 clearly shows how well the performance of the different models varies greatly from one example to another example. Let's look at the first example, which shows the high-density building area, where the buildings are neatly arranged in rectangular. Although the basic CDNeXt model can roughly identify the distribution of the building, it is not good enough to do well in the processing of the edges of the building, the boundaries look a little blurred, and sometimes the edges of the buildings that should not be connected are mistakenly connected.The structure is also more complete. When the complete MA-CDNeXt model was used, the results were almost identical to the real situation, accurately identifying the spatial distribution of dense buildings, and accurately depicting the shape of each building, which matches the reference label very highly.

The second example mentioned in this paper is the more varied road area, which has both a cluster of buildings with scattered small changes below. With the original baseline model, it can be very good to find the large buildings above, but it is not accurate to find the small changes below. Later, with the addition of MSF and MSF, the two methods of building, which really make the above complex detection more accurate, but it is still uncertain.It's all a lot better.

This example shows us a particularly irregular cluster of buildings with several different shapes, such as triangles and rectangular ones, and some other geometrical buildings mixed together. Because the structure is particularly messy, the detection algorithm is very difficult to process. The model used in the beginning can only approximate the main building structure, but it is not very good to find the complex shapes and details, especially the few individual buildings in the lower right corner, which have not been found.The structure of the entire irregular complex has been restored, and the boundaries of the buildings of various shapes have been drawn more accurately, and even the small buildings that are separate in the lower right corner have been found.

The fourth example shows the scene where the change is less obvious, the building changes are relatively small, and this subtle change is easy to miss. Although the baseline model finds some changes, there are still many places that have not been found. Add the MSF module after the detection effect is a little better, but it is still less sensitive to those particularly small targets.

From these visual comparisons, it is more clear that our quantitative analysis results are reliable. By adding optimized components step by step, the effect of building change detection is improved in many ways. For example, in ordinary building-dense areas, structural recognition is more accurate; for buildings with irregular shapes, the edge detection effect is better; those small scattered changes can also be found; and in complex environments, the details are retained more.When it changes, its advantages are particularly obvious.

4. Conclusion

This research addresses critical challenges in high-resolution remote sensing image change detection by developing MA-CDNeXt, a novel framework that integrates enhanced multi-scale feature processing, improved temporal-spatial interaction mechanisms, and optimized decoding strategies. The model combines an improved ConvNeXt-based Siamese network with multi-scale feature extraction, incorporates MAFGNet's channel and spatial attention fusion modules (CAFM and SAFM) into CDNeXt's temporal interaction framework, and implements an Atrous Context Self-attention Pyramid (ACSP) structure for enhanced contextual understanding. Additionally, the dual graph convolutional module (DGM) with differential feature enhancement improves change region discrimination, while a composite loss function addresses class imbalance issues. Experimental validation on the LEVIR-CD+ dataset demonstrates significant performance improvements, achieving 84.09% accuracy, 78.39% recall, 81.14% F1-score, and 68.26% IoU, representing approximately 7 percentage points improvement in F1-score and nearly 10 percentage points in IoU compared to baseline CDNeXt. Despite these achievements, the research acknowledges limitations including computational complexity, dataset specificity, dependency on labeled data, and hyperparameter sensitivity. Future work should focus on model lightweighting and efficiency optimization, enhancing generalization through domain adaptation, exploring weakly-supervised learning paradigms, integrating multi-modal data fusion capabilities, conducting larger-scale benchmarking, and advancing model interpretability research to develop more robust and practical change detection systems for diverse remote sensing applications.

References

[1] REN W, WANG Z, XIA M, et al. MFINet: Multi-scale feature interaction network for change detection of high-resolution remote sensing images[J]. Remote Sensing, 2024, 16(7): 1269.

[2] WANG H, YE Z, XU C, et al. TTMGNet: Tree topology mamba-guided network collaborative hierarchical incremental aggregation for change detection[J]. Remote Sensing, 2024, 16(21): 4068.

[3] SODNOM E, DAMDINSUREN A. Improvement of Mongolian Height System Using a Satellite Technology[J]. Open Journal of Applied Sciences, 2020, 10(4): 154-168.

[4] CAI L, YIN J, YAN X, et al. The environmental analysis and site selection of mussel and large yellow croaker aquaculture areas based on high resolution remote sensing[J]. Acta Oceanologica Sinica, 2024, 43(3): 66-86.

[5] LIU J, GU H, LIU F, et al. CE-CDNet: A Transformer-Based Channel Optimization Approach for Change Detection in Remote Sensing[J]. Computers, Materials & Continua, 2025, 83(4): 803-822.

[6] YAN W, CAO L, YAN P, et al. Remote sensing image change detection based on swin transformer and cross-attention mechanism[J]. Earth Science Informatics, 2025, 18(1): 106.

[7] RAO M S, RAO G R, KRISHNA G M, et al. Revolutionizing Groundwater Suitability with AI-Driven Spatial Decision Support—A Remote Sensing and GIS Approach for Visakhapatnam District, Andhra Pradesh, India[J]. Journal of Geographic Information System, 2025, 17(1): 23-44.

[8] SARR A. Multi-Scale Characteristics of Precipitation and Temperature over West Africa Using SMHI-RCA Driven by GCMs under RCP8.5[J]. American Journal of Climate Change, 2017, 6(3): 455-486.

[9] LI Y M, ZHAN R F, DING Y H. Multi-Scale Influencing Factors and Prediction of Interannual Variability in Rapid Intensification Magnitude of Northwest Pacific Tropical Cyclones[J]. Journal of Tropical Meteorology, 2025, 31(1): 75-86.

[10] LI Z, CAO S, DENG J, et al. STADE-CDNet: Spatial-temporal attention with difference enhancement-based network for remote sensing image change detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62(1): 1-15.

[11] WANG Z, GU G, XIA M, et al. Bitemporal attention sharing network for remote sensing image change detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62(1): 1-15.

[12] ACHOUR S, ELMEZOUARI M C, TALEB N, et al. A PCA-PD fusion method for change detection in remote sensing multi temporal images[J]. International Journal of Remote Sensing, 2019, 40(15): 2-15.

[13] DALMIYA C P, SANTHI N, SATHYABAMA B. A novel feature descriptor for automatic change detection in remote sensing images[J]. Egyptian Journal of Remote Sensing and Space Science, 2018, 21(1): 1-10.

[14] FAN C L, CHUNG Y J. Integrating Image Processing Technology and Deep Learning to Identify Crops in UAV Orthoimages[J]. Computers, Materials & Continua, 2025, 82(2): 1925-1945.

[15] SAIDI S, IDBRAIM S, KARMOUDE Y, et al. Deep-learning for change detection using multi-modal fusion of remote sensing images: A review[J]. Remote Sensing, 2022, 14(10): 2343.

[16] NOMAN M, FIAZ M, CHOLAKKAL H, et al. ELGC-Net: Efficient local-global context aggregation for remote sensing change detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62(1): 1-10.

[17] ZOU C, LIANG W, LIU L, et al. Hyperspectral image change detection based on an improved multi-scale and spectral-wise transformer[J]. International Journal of Remote Sensing, 2024, 45(6): 1903-1924.

[18] XU L, LI F, CHANG S. Unifying Convolution and Transformer Decoder for Textile Fiber Identification[J]. Journal of Donghua University (English Edition), 2023, 40(4): 357-363.