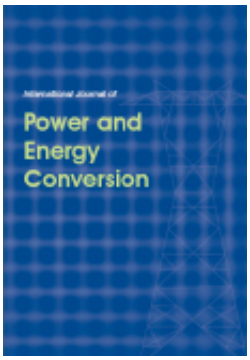
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research




**International Journal of Power and Energy Conversion**

ISSN online: 1757-1162 - ISSN print: 1757-1154
https://www.inderscience.com/ijpec

# Change detection framework for power facilities in disaster scenarios

Jiangyi Wu, He Tang, Guangdu Cen, Ke Wang

**DOI:** 10.1504/IJPEC.2026.10078900

**Article History:**

| | |
|---|---|
| Received: | 26 August 2025 |
| Last revised: | 14 April 2026 |
| Accepted: | 01 May 2026 |
| Published online: | 16 June 2026 |

# Change detection framework for power facilities in disaster scenarios

Jiangyi Wu and He Tang

Foshan Power Supply Bureau,
Guangdong Power Grid Co., Ltd.,
Foshan 528000, China
Email: wujiangyi@fs.gd.csg.cn
Email: tanghe@fs.gd.csg.cn

Guangdu Cen* and Ke Wang

China Southern Power Grid Technology Co., Ltd.,
Guangzhou 510000, China
and
Guangdong Engineering Technology Research
Center of Special Robots for Special Industries,
No. 1, Fenjiang South Road, Chancheng District,
Foshan City, Guangdong Province, 510000, China
Email: ciju4477@21cn.com
Email: wangkewoft@163.com
*Corresponding author

**Abstract:** In order to regularly monitor geological surface changes around the power transmission line corridors and to swiftly and effectively respond to the negative impacts of natural disasters (specifically landslides) on electrical facilities, thereby enhancing the disaster resistance and recovery speed of the power system, while ensuring the stability of the tower foundations, a change detection model oriented towards disaster scenarios is proposed. Firstly, a twin-network-based change detection framework is designed, which utilises pre-processed remote sensing images before and after the disaster to extract multi-scale visual features using a basic visual model and a designed change detection module, respectively. Secondly, to filter effective feature information and integrate the features from both modules, an attention mechanism-based alignment module is proposed to fuse general features and domain-specific features, guiding the domain-specific features with general features for efficient feature extraction, thus ensuring the robustness and accuracy of the proposed model. Finally, experiments are conducted on real disaster datasets, and the proposed method is compared with excellent algorithms proposed in recent years to verify the accuracy of the method.

**Keywords:** change detection; CD; deep learning; disaster monitoring; remote sensing; RS; vision foundation models; VFMs; Siamese neural network.

**Reference** to this paper should be made as follows: Wu, J., Tang, H., Cen, G. and Wang, K. (2026) 'Change detection framework for power facilities in disaster scenarios', *Int. J. Power and Energy Conversion*, Vol. 17, No. 5, pp.1–20.

**Biographical notes:** Jiangyi Wu holds a Master's degree. He is an Engineer engaged in the research of key technologies for digital grids and intelligent operation and maintenance.

He Tang holds a Master's degree. He is an Engineer engaged in the research of key technologies for digital grids and intelligent operation and maintenance.

Guangdu Cen holds a Master's degree. He is an Engineer engaged in the research of key technologies for digital grids and intelligent operation and maintenance.

Ke Wang holds a Master's degree. He is an Engineer engaged in the research of key technologies for digital grids and intelligent operation and maintenance.

## 1 Introduction

With the continuous advancement of technology and societal development, the demand for monitoring and analysing the Earth's surface has become increasingly prominent. In this process, remote sensing change detection has emerged as a key technology; this technology offers robust methodologies for acquiring a more profound comprehension of the dynamic transformations occurring across the Earth's surface. Change detection can be defined as the analytical process of discerning alterations in the state of an object or phenomenon through the examination of multi-temporal data (Paranjape et al., 2025). The significance of change detection extends beyond academic inquiry to inform critical decision-making in numerous real-world applications. It provides the foundational data for strategic activities, including land use planning (Li et al., 2017), urban development (Jasim, 2025), and environmental monitoring (Bikis et al., 2025). The versatility of deep learning paradigms has been extensively proven across diverse domains (Dulal and Dulal, 2026; Kumari et al., 2026; Haque et al., 2026), driving its rapid adoption in complex remote sensing tasks.

To address this, recent research has increasingly pivoted towards transformer-based architectures and attention mechanisms. Several studies (Gu et al., 2020; Tian et al., 2020; Liang et al., 2022; Li et al., 2023) have demonstrated that self-attention can effectively capture global context and dynamic changes in remote sensing imagery without being constrained by local perception. Furthermore, to tackle the issue of feature misalignment in bi-temporal images, researchers have introduced various alignment strategies, such as feature interaction modules (Fang et al., 2024; Zhao et al., 2023; Zhang et al., 2024) and pseudo-transition video generation (Lin et al., 2023). Others have focused on enhancing boundary precision through dilated convolutions (Ji et al., 2020) or deformable attention (Lei et al., 2024). These innovations have significantly driven the development of change detection techniques.

Particularly in the power sector, the geographic layout of large-scale transmission line corridors makes power infrastructure vulnerable to natural disasters. Transmission towers are frequently situated in remote, mountainous areas with complex geological conditions. In these scenarios, the primary threat to the power grid often comes from the instability of the surrounding environment rather than damage to the components themselves. For instance, landslides and mudslides caused by heavy rainfall can destabilise tower

foundations, leading to infrastructure collapse. Therefore, monitoring the surface changes and geological hazards within the transmission corridor is a prerequisite for ensuring the safety of power facilities. Currently, most change detection methods focus on building change detection, with very few algorithms specifically designed for disaster detection in power infrastructure. Research on disaster changes caused by natural events such as floods, fires, typhoons, and landslides, and the detection of changes in the land surrounding transmission equipment, is crucial for estimating affected areas, assessing damage to facilities, and enabling intelligent disaster identification and evaluation.

The prevailing methodology for disaster change detection involves the comparative analysis of bi-temporal remote sensing imagery acquired before and after a hazardous event. However, due to the uncertainty of disaster occurrences and inconsistent imaging conditions of equipment used to capture disaster scenes, a primary constraint in this domain is the limited quantity and poor quality of available data. Datasets of pre-and post-disaster images are typically not only sparse but also suffer from significant contamination by irrelevant changes, which can obscure the actual disaster impact. This makes it difficult to fully extract features from pre-and post-change images, and training a high-performance algorithm purely based on disaster scene change detection data is challenging. Recent literature (Li et al., 2024) mentions that the performance improvement of deep learning-based evaluation metrics in the past three years is only 1%–2%, concurrently, the scale of these models has grown dramatically, as evidenced by parameter counts that have surged well into the hundreds of millions. This means that as model scales grow, limited annotated data restricts further improvement in change detection model performance. To mitigate this issue, the literature also points out that foundation models are designed to acquire a broad repository of generalisable knowledge via large-scale pre-training. This process is pivotal for mitigating the dependency on extensive, task-specific annotated data for downstream applications. In other words, pre-training on general knowledge can somewhat alleviate the limited availability of annotated data.

Currently, a few algorithms in change detection tasks have already employed basic models. Pre-trained on extensive, large-scale datasets, these vision foundation models (VFMs) now serve as powerful backbones for a diverse array of downstream tasks within the remote sensing domain. Compared to training models from scratch with randomly initialised parameters, using pre-trained basic models, which contain extensive general knowledge, significantly reduces data requirements. For example, algorithm (Ding et al., 2024) proposes the SAM-CD model, the model's architecture is centred on the FastSAM VFM, which serves as the primary backbone for visual feature extraction. The high-fidelity features derived from this process are instrumental in enhancing the performance of high-resolution remote sensing change detection; algorithm (Dong et al., 2024) introduces the ChangeCLIP framework based on the CLIP base model, which generates textual semantic prompts from images and combines them with image features to transform change detection into a multimodal task.

However, most current visual base models are trained on natural images, which differ significantly from remote sensing images: remote sensing imagery is characterised by a distinct set of properties, including a nadiral (overhead) perspective, unique spectral signatures that differ from the colour distributions in natural photos, and significant variations in scale and spatial resolution. This results in limited fine-tuning performance when directly applying pre-trained models for change detection (Wang et al., 2024). To address this issue, literature proposes a multi-task pre-training (MTP) approach, the

architecture is characterised by a shared encoder, which is responsible for learning a common and robust set of feature representations from the input. This is complemented by multiple task-specific decoders, each of which is independently trained to handle a distinct downstream task. The work described involves the large-scale pre-training of a VFM on an extensive corpus of remote sensing imagery and demonstrated excellent performance in various downstream remote sensing tasks, the model exhibits strong generalisation capabilities across a diverse suite of downstream tasks, including object detection, scene classification, semantic segmentation, and change detection.

The change detection model presented in this paper is designed as a feature-level change detection framework that sequentially extracts features from dual-time remote sensing images of power infrastructure. It effectively extracts time-level features from images while using a remote sensing visual base model to embed general knowledge from the remote sensing domain into the change detection features. The proposed methodology is specifically designed to mitigate the inherent domain discrepancy between natural images and remote sensing imagery. Moreover, the model employs a pyramid-based multi-scale feature extraction module, which allows the model to better preserve image details and global structural features, adapting well to pixel-level change detection tasks. Finally, the proposed method incorporates an attention-based alignment module to efficiently fuse features of different sizes, automatically extracting useful feature information from the base model and effectively merging it for feature extraction. The main contributions of this paper include the following two aspects:

1 A disaster-oriented twin-network framework utilising VFMs: we propose a specialised change detection framework for monitoring power infrastructure in disaster scenarios. Unlike traditional methods that train from scratch, we leverage a frozen, pre-trained VFM to extract robust multi-scale visual representations. This approach effectively addresses the challenge of limited annotated data in disaster datasets, (e.g., GVLM) by transferring general visual knowledge to the specific task of identifying geological hazards in complex terrains.

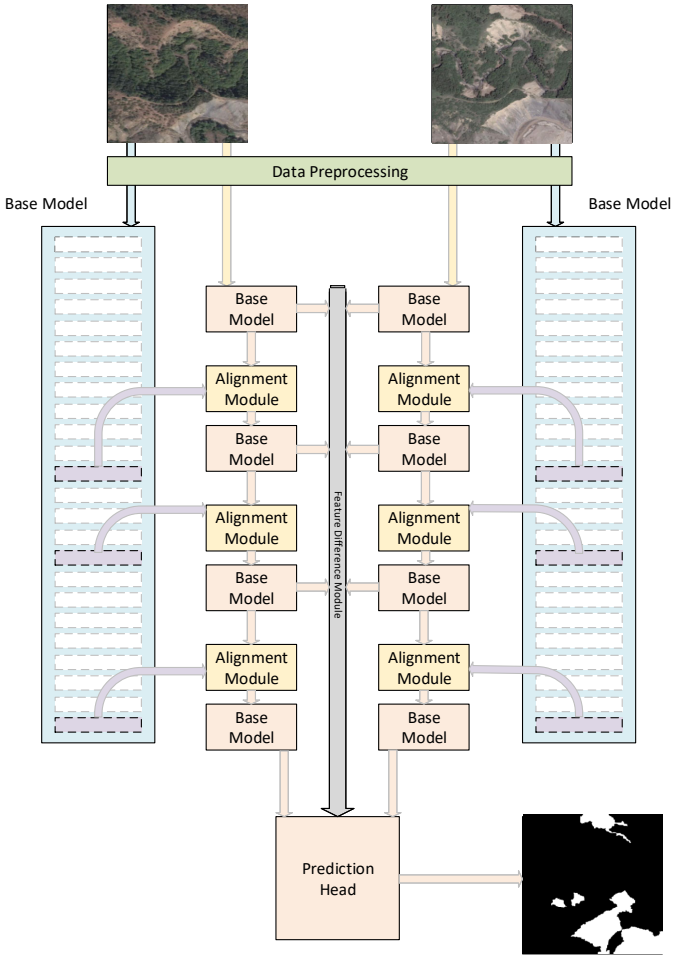
2 A novel correlation-aware feature alignment module for domain gap mitigation: we design a cosine similarity-based attention mechanism to effectively integrate general VFM features with domain-specific features. Different from existing methods such as the bi-temporal adapter network (BAN) or SAM-CD, which typically employ bridge modules for direct feature injection and often suffer from ‘negative transfer’ due to background noise, our approach introduces a strict semantic filtering mechanism. By utilising cosine similarity to measure the angular alignment (structural consistency) between VFM and domain features, our module dynamically filters out magnitude-based noise, (e.g., seasonal vegetation changes or illumination shifts) caused by the domain gap between natural and remote sensing images. This ensures that only the most structurally relevant general knowledge is transferred to guide the restoration of fine-grained boundary details in disaster scenarios.

## 2 Change detection model

The proposed change detection framework is architected around six core modules, as visually detailed in Figure 1: data pre-processing, base model, alignment module, base module, feature difference module, and prediction head. For data pre-processing, the raw

images were cropped and augmented during the experiment, which will be explained in detail in the experimental section. The frozen pre-trained model refers to the rotated varied-size window attention (RVSA) model (Wang et al., 2023), which is used to extract general knowledge from both pre-disaster and post-disaster scenes. The base module is responsible for extracting multi-scale, domain-specific knowledge from the raw images. The two types of features are aligned and fused through the alignment module, and then the feature difference module is used to integrate the fused features at different levels. Finally, the prediction head module processes the multi-scale change features and generates the change map. The following section will provide further details on the five modules.

**Figure 1** Overview of our proposed model (see online version for colours)




## *2.1 Pre-trained model*

The base model is used to extract multi-scale features from pre-and post-disaster power remote sensing images. After performing data pre-processing on the pre-and post-change

power remote sensing images, subsequently, the pre-processed images are input into the pre-trained foundational model, which is tasked with extracting a rich hierarchy of features at multiple scales. The RVSA model is employed here. Unlike standard vision transformers (ViT) that utilise fixed square windows, which often fragment continuous, irregularly shaped objects like landslides or winding mountain roads, RVSA introduces a flexible windowing mechanism. Given the varying orientations of remote sensing targets due to the bird's-eye view perspective, this model introduces an additional learnable rotation angle factor, extending the variable-size window attention in visual transformers. This design allows the network to adaptively align with the geometric structures of ground targets, ensuring more integral feature extraction. This extension generates adaptive scaling, translation, and rotation of the windows. Specifically, given the input features $I \in R^{C\times H\times W}$ (with $C$, $H$, and $W$ corresponding to the channel, height, and width dimensions of the input feature map, respectively), The input features are uniformly divided into different windows, and the features of each window can be represented as $I_p \in R^{C\times p\times p}$, (the value $p$ is the window size), and a total of $\frac{H}{p}\times\frac{W}{p}$ windows are obtained. Then, three linear layers are used to generate the query features, as well as the initial key and value features, represented as $Q_p$, $K_p$, and $V_p$ respectively. Next, the changes in the windows are predicted using $I_p$:

$$S_p, O_p, \Theta_p = Linear\left(LeakyReLU\left(GAP\left(I_p\right)\right)\right) \tag{1}$$

Here, GAP refers to global average pooling. $S_p = \{s_x, s_y \in \mathrm{R}^1\}$ are the scaling factors along the X and Y axes, respectively, while $O_p = \{o_x, o_y \in \mathrm{R}^1\}$ represent the translation offsets along the X and Y axes. In addition, $\Theta_p = \{\theta \in \mathrm{R}^1\}$ denotes the rotation angle. Taking the corner points of a window as an example:

$$\begin{bmatrix} x_{l/r} \\ y_{l/r} \end{bmatrix} = \begin{bmatrix} x^c \\ y^c \end{bmatrix} + \begin{bmatrix} x^r_{l/r} \\ y^r_{l/r} \end{bmatrix} \tag{2}$$

where $x_l$, $x_r$, $y_l$, $y_r$ are the coordinates defining the initial bounding box, specifically its top-left and bottom-right vertices, and $x^c$, $y^c$ are the coordinates of the window's centre point. $x^r_l$, $x^r_r$, $y^r_l$, $y^r_r$ are the horizontal and vertical offsets from the centre point to each of the corner points. The transformation of the window can be achieved by using the obtained scaling, translation, and rotation factors:

$$\begin{bmatrix} x_{l/r}{}' \\ y_{l/r}{}' \end{bmatrix} = \begin{bmatrix} x^c \\ y^c \end{bmatrix} + \begin{bmatrix} o_x \\ o_y \end{bmatrix} + \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x^r_{l/r}\cdot s_x \\ y^r_{l/r}\cdot s_y \end{bmatrix} \tag{3}$$

where $x_l'$, $x_r'$, $y_l'$, $y_r'$ are the coordinates of the corner points of the transformed window. New key and value features, $\boldsymbol{K}_{p'}$ and $\boldsymbol{V}_{p'}$, are sampled from the transformed window. These features are subsequently used to compute the self-attention score according to the following equation:

$$\boldsymbol{F}_p = SA\left(\boldsymbol{Q}_p, \boldsymbol{K}_{p'}, \boldsymbol{V}_{p'}\right) = softmax\left(\frac{\boldsymbol{Q}_p \boldsymbol{K}^T_{p'}}{\sqrt{C'}}\right)\boldsymbol{V}_{p'} \tag{4}$$

$\boldsymbol{F}_p \in R^{s^2 \times C'}$ is the feature output of a window's self-attention operation, and $c = hc'$, where $h$ is the number of self-attention (SA) heads. The shape of the final output features from the rotated variable-sized window attention is restored by concatenating the features of different self-attention heads along The final step involves reconstructing the output feature map through two concurrent operations: an aggregation of information along the channel dimension, and a spatial reassembly of the feature representations from the distinct attention windows along the spatial dimensions. A key architectural modification in our approach involves substituting the conventional multi-head self-attention (MHSA) mechanism of the standard ViT. In its place, we implement our proposed rotated variable-sized window attention.

The pre-processed image pairs, both before and after transformation, are input into this pre-trained model. The entire pre-trained model consists of 24 layers. Here, layers 11, 15, and 23 are used as features at different scales. The deeper the layer, the more abstract the extracted features; the shallower the layer, the more fine-grained the extracted features.

## *2.2 Alignment module*

The proposed method employs an attention-based feature alignment module to effectively harmonise and fuse the multi-scale features derived from the pre-trained foundation model. Since the foundation model is pre-trained on general natural images, a significant domain gap exists when applied to remote sensing disaster scenes. Direct fusion of these features often leads to 'negative transfer' due to the complex background noise, (e.g., vegetation, shadows) unique to satellite imagery. Therefore, we utilise an attention mechanism to strictly measure the semantic consistency between features, selectively filtering out irrelevant general information while retaining valuable texture cues.

The features generated by the pre-trained model in module 1.1 and the features from the base module in module 1.3 may not have the same dimensions. Furthermore, not all pre-trained knowledge is necessarily effective. Specifically, two issues must be considered:

a Not all general knowledge is essential, so it is necessary to determine how to filter out useless information and retain valuable general features.

b The pre-trained model contains 24 layers, and the features in the base module are downsampled by at least a factor of 10, resulting in extremely low resolution.

Therefore, it is necessary to integrate these highly abstract features with the general change detection framework.

The proposed method employs an attention-based feature alignment module to effectively harmonise and fuse the multi-scale features derived from the pre-trained foundation model. Since the foundation model is pre-trained on general natural images, a significant domain gap exists when applied to remote sensing disaster scenes. Direct fusion (as seen in methods like BAN or SAM-CD) often leads to 'negative transfer' due to the complex background noise, (e.g., vegetation, shadows) unique to satellite imagery.

To address this, it is crucial to analyse the alignment from a representation learning perspective. We use $F_g \in \mathbb{R}^{H_g \times W_g \times C_g}$ to represent the general features extracted by the

pre-trained model and $F_s \in \mathbb{R}^{H_s \times W_s \times C_s}$ to represent the remote sensing domain features. First, a layer normalisation (LN) operation is applied to the domain-specific features:

$$F_g = Linear\left(LN\left(F_g\right)\right) \tag{5}$$

Theoretical justification for cosine similarity: the primary purpose of the LN step is to standardise the statistical distribution of $F_g$, bringing it into alignment with the feature space of $F_s$. Subsequently, to effectively distinguish valuable features from the general knowledge without introducing domain-specific noise, we specifically design a cosine similarity measure rather than standard additive attention or Euclidean distance. In disaster scenes, irrelevant changes (like seasonal colour shifts or shadow variations) often cause large magnitude spikes in feature vectors. Cosine similarity evaluates the angle (directional consistency) between $F_g$ and $F_s$, effectively ignoring these magnitude-based noises. It acts as a strict semantic gate: it only produces high attention weights when the general foundational features share structural and geometric consensus, (e.g., landslide boundaries) with the domain-specific features.

The correlation matrix $A \in \mathbb{R}^{H_s W_s \times H_g \times W_g}$ between $\overline{F}_g$ and $F_s$ is computed. Then, $A$ is divided by $\sqrt{C_s}$ and a softmax method is applied to obtain the attention weights $A^*$. Finally, a dot product operation is performed between $A^*$ and $\overline{F}_g$ to get the weighted features $F_s^*$:

$$F_s^* = softmax\left(\frac{F_s \times \overline{F}_g^T}{\sqrt{C_s}}\right)\overline{F}_g \tag{6}$$

The tensor $F_s$ is then reshaped to a tensor with dimensions $F_s^* \in R^{H_s \times W_s \times C_s}$. This both extracts the useful parts of the general features and solves the scale misalignment problem. The final output of the module is then generated via a residual connection. Here, $UpSample(\cdot)$ denotes upsampling using bilinear interpolation:

$$F_o = F_s^* + UpSample\left(\overline{F}_g\right) + F_s \tag{7}$$

The choice of cosine similarity over other measures (e.g., Euclidean distance) is rooted in its robustness to radiometric variations. In remote sensing, domain gaps often manifest as magnitude discrepancies in feature space due to differing sensors, illumination, or seasonal changes. While Euclidean distance is highly sensitive to the absolute magnitude of feature vectors, cosine similarity measures the angular alignment, focusing on the directional orientation of semantic structures. Formally, by mapping features onto a hypersphere via LN and evaluating their cosine correlation, the module acts as a cross-domain feature rectifier.

From a domain adaptation perspective, the alignment module performs selective knowledge transfer. Let the general features $F_g$ represent a broad manifold learned from natural images. Directly fusing $F_g$ into the disaster detection task may introduce 'out-of-distribution' noise. Our mechanism uses the domain-specific features $F_s$ as a query to project $F_g$ onto the target domain's manifold. High cosine similarity indicates that a general feature, (e.g., a generic edge detector) aligns with the structural consensus of the disaster scene, whereas low similarity effectively filters out irrelevant natural image priors. This formal filtering process ensures that only the structurally consistent

representations are propagated, thus mitigating the domain gap between natural and remote sensing imagery.

## *2.3 Base module*

The general change detection backbone operates in a two-stage process. First, it sequentially extracts a hierarchy of multi-scale features. Following this, it progressively fuses the aligned, multi-scale feature representations that have been enriched by the pre-trained foundation model;

For the pre-processed bi-temporal images obtained from module 1.1, a multi-level transformer encoder is also used to extract multi-level features. The network learns a rich feature hierarchy. Features extracted from shallow layers are characterised by high spatial resolution, preserving fine-grained details such as edges and textures. In contrast, features from deeper layers have a lower spatial resolution but encode more abstract, semantic information representing the global context of the image. Specifically, given an input feature $I \in R^{3\times H\times W}$ (where 3 represents the image is in RGB three-channel format), each base module will output a tensor of features with dimension $F_l \in \mathrm{R}^{C_l\times\frac{H}{2^{l+1}}\times\frac{W}{2^{l+1}}}$. Here, $l$ represents the layer number, with a value of $\{1, 2, 3, 4\}$. This part uses four base modules to extract four layers of features. After each base module, downsampling is used and the number of channels is increased, which means $C_{l+1} > C_l$. The resulting $F_l$ from each step will be used as the domain feature $F_s$ in module 1.2, which is then fused with the general feature $F_g$ to obtain the output $F_o$. The output $F_o$ is used as the input $F_{l+1}$ for the next base module. For convenience, we use $F_o = Mix(F_g, F_l)$ to represent the process of module 1.2. The pre-processed original image passes through four base modules to obtain four types of features at different scales, and then module 1.2 is used to automatically align the features.

Details of the base module: when processing the features of layer $l$, $F_l \in \mathrm{R}^{C_l\times\frac{H}{2^{l+1}}\times\frac{W}{2^{l+1}}}$, we first downsample them to obtain $F_l^* \in \mathrm{R}^{C_{l+1}\times\frac{H}{2^{l+2}}\times\frac{W}{2^{l+2}}}$. A convolutional module is used here to perform downsampling. The initial downsampling uses a convolution with a kernel size of 7×7, a stride of 4, and a padding of 3. Spatial downsampling is achieved through three subsequent convolutional layers. Each of these layers employs a 3×3 kernel with a stride of 2 and a padding of 1 to reduce the feature map dimensions. $Conv_{DS_l}$ represents this downsampling process.

$$F_l^* = Conv_{DS_l}(F_l), l \in \{1, 2, 3, 4\} \tag{8}$$

The main building block of each base module is the self-attention module. The attention mechanism used in the original transformer is as follows:

$$Attn(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_{\text{head}}}}\right)V \tag{9}$$

where $Q$, $V$, and $K$ represent the query, value and key, respectively. The three features have the same dimension, RHW×C. Formula (9) has a high computational complexity of $O((HW)2)$. Before applying the attention mechanism, the $HW$ dimension is first reduced, as shown in formulas (10)–(11).

$$Q = Reshape(Q) \tag{10}$$

$$Q = Linear(Q) \tag{11}$$

For the original feature $Q \in \mathrm{R}^{HW \times C}$, the $Reshape(Q)$ operation converts $Q$ into a tensor with features $Q \in \mathrm{R}^{\frac{HW}{r} \times C \cdot r}$, where $r$ is a hyperparameter representing the scaling ratio. Then, a linear layer $Linear(Q)$ remaps $Q \in \mathrm{R}^{\frac{HW}{r} \times C \cdot r}$ to features $Q \in \mathrm{R}^{\frac{HW}{r} \times C}$. This results in a feature with a reduced dimension for $Q$. The same process is applied to $K$ and $V$. This reduces the computational complexity to $O\left(\frac{(HW)^2}{r}\right)$.

Since the $HW$ dimension is flattened into a one-dimensional vector in the attention mechanism, to incorporate positional information, two multi-layer perceptrons (MLPs) and a convolution are used to add positional information:

$$F_{out} = MLP\left(GELU\left(CONV2D_{3\times3}\left(MLP\left(F_{in}\right)\right)\right)\right) + F_{in} \tag{12}$$

where $F_{in}$ represents the self-attention feature, and GELU stands for the Gaussian error linear unit activation. In contrast to the static positional encoding employed by the standard ViT, which is tied to a fixed input resolution, our proposed encoding scheme is dynamic and can flexibly adapt to images of varying resolutions.

For the base module, the method for passing the input from one base module to the next is as follows: first, downsampling is performed to generate the corresponding $Q_j$, $K_j$, $V_j$, where $j$ denotes the number of multi-attention heads. The three features are then dimensionally reduced. The MHSA yields the output $F_l^*$. Finally, positional information is added and combined with general knowledge to get the input for the next base module:

$$F_l^* = Conv_{DS_l}\left(F_l\right) \tag{13}$$

$$\begin{aligned} Q_j &= F_l^* W_j^q \\ K_j &= F_l^* W_j^k \\ V_j &= F_l^* W_j^v \end{aligned} \tag{14}$$

$$\begin{aligned} Q_j &= \&Linear\left(Reshape\left(Q_j\right)\right) \\ K_j &= Linear\left(Reshape\left(K_j\right)\right) \\ V_j &= Linear\left(Reshape\left(V_j\right)\right) \end{aligned} \tag{15}$$

$$\begin{aligned} head_j \& &= Attn\left(Q_j, K_j, V_j\right) \\ \& &= Softmax\left(\frac{Q_j K_j^T}{\sqrt{d_{\text{head}}}}\right) V_j \end{aligned} \tag{16}$$

$$F_l^* = Concat\left(head_1, ..., head_j\right) W^o \tag{17}$$

$$F_l^* = MLP\left(GELU\left(Conv\left(MLP\left(F_l^*\right)\right)\right)\right) + F_l^* \tag{18}$$

$$F_{l+1} = Mix\left(F_g, F_l^*\right) \tag{19}$$

## *2.4 Feature difference module*

The feature difference module is used to process multi-scale feature maps from different temporal phases. In the previous module 1.3, two temporal phases, before and after the change, yielded four features of different sizes, $F_l^{pre}, F_l^{post}, l \in \{1, 2, 3, 4\}$. The difference features are obtained using the following method, where 2D convolution, rectified linear units (ReLU), and batch normalisation are used to fuse the results. Here, $\varphi$ denotes the concatenation of tensors along the channel dimension.

$$F_l^{diff} = BN\left(ReLU\left(Conv\left(\varphi\left(F_l^{pre}, F_l^{post}\right)\right)\right)\right) \tag{20}$$

## *2.5 Prediction head*

The prediction head operates on the computed difference features to produce the final pixel-level change map, which serves as the ultimate output of the change detection model. The prediction head takes as input the four multi-scale difference features generated by the preceding module (Section 1.4). It then fuses these hierarchical representations to produce the final disaster change prediction map. As the initial step in the fusion process [see equations (21)–(23)], a 1×1 convolutional layer is applied to each feature map. This operation projects the features into a new space, harmonising their channel dimensions to a uniform number. Then, to ensure all four feature maps share a common spatial dimension, bilinear interpolation is employed to upsample each one to a uniform resolution. Finally, following a channel-wise concatenation of the four feature maps, a 1×1 convolutional layer is employed. The purpose of this convolution is to project the resulting high-dimensional tensor back to a uniform channel depth, as described by the following equation:

$$F_l^{diff} = Conv_{1\times1}\left(F_l^{diff}\right), l \in \{1, 2, 3, 4\} \tag{21}$$

$$F_l^{diff} = Bilinear\left(F_l^{diff}\right), l \in \{1, 2, 3, 4\} \tag{22}$$

$$F = Conv_{1\times1}\left(Concat\left(F_1^{diff}, ..., F_4^{diff}\right)\right) \tag{23}$$

Finally, the fused feature map $F$ is upsampled to the size of the original input image. A MLP is then used to process it according to the prediction classes, outputting a change map (CM). The upsampling employs a transposed convolution with a stride of 2 and a kernel size of 3. There are two prediction classes: changed and unchanged.

Finally, a pixel-wise cross-entropy loss function is used, which is defined as follows:

$$L_{ce} = -\frac{1}{H\times W}\sum_{h=1}^{H}\sum_{w=1}^{W}\sum_{c=1}^{N_c} y_{h,w,c}\log\left(p_{h,w,c}\right) \tag{24}$$

where $H$ and $W$ represent the height and width of the prediction map output and $N_c$ denotes the number of classes. $y_{h,w,c}$ is the binary indicator (0 or 1) indicating if class $c$ is

the correct classification for the pixel at position ($h$, $w$), and $p_{h,w,c}$ is the predicted probability that the pixel belongs to class $c$.

## 3 Experimental analysis

### *3.1 Dataset and pre-processing*

In the field of change detection, several high-quality datasets have been made publicly available, such as LEVIR-CD, SYSU-CD, and GZ-CD. While these datasets have significantly advanced the field, they primarily focus on urban building changes, characterised by man-made structures with regular geometric boundaries and distinct edges. However, the scope of this study is specifically oriented towards disaster monitoring for power infrastructure, where the changes are caused by natural phenomena, (e.g., landslides, mudslides) rather than urban construction. Unlike building changes, disaster-induced changes in complex terrains exhibit highly irregular shapes, blurred boundaries, and significant background interference from vegetation and topography. Consequently, standard building change detection datasets do not adequately represent the challenges faced in disaster scenarios.

To rigorously validate the efficacy of our proposed method in this specific domain, we selected the global very-high-resolution landslide mapping (GVLM) dataset (Zhang et al., 2023). The GVLM dataset represents a significant contribution to the field as the first publicly accessible, global-scale repository for very-high-resolution landslide mapping. A key characteristic of this dataset is its fine spatial resolution of 0.59 metres. It covers landslide sites in 17 countries or regions worldwide, including landslides of various types, causes, and terrains. The total data coverage area reaches 166.78 square kilometres.

The original GVLM dataset covers landslide sites in 17 countries or regions globally. Each landslide site includes a pre-landslide image, a post-landslide image, and an annotated change label. To manage the high spatial resolution and inconsistent dimensions of the original remote sensing scenes, we implemented a tiling-based pre-processing strategy. Each bi-temporal image pair was systematically cropped into a series of overlapping patches, each with a uniform size of 512×512 pixels, ensuring that the same crop location was used for each pair. After cropping, 1,986 image pairs were obtained and then split into training, testing, and validation sets with a ratio of 6:2:2. To fully leverage this dataset, data augmentation was performed before the images were input to the model, which specifically includes the following steps are taken:

a Data augmentation is performed, with images undergoing random rotation and translation. Specifically, there is a 50% probability of a random rotation with an angle ranging from –20 to 20 degrees. Similarly, there is a 50% probability of horizontal or vertical flipping to increase the diversity of the training data and reduce the risk of model overfitting.

b The images and their labels are randomly cropped to a size of 512×512. The maximum proportion of the target class within the cropped area is set not to exceed 0.75, ensuring that the cropped region contains both the target object and the background, which helps the model learn more comprehensive features.

c The temporal order of the pre-and post-disaster images is randomly swapped with a 50% probability. This effectively augments the temporal data and helps the model learn time-agnostic features.

d A series of image transformations are applied to further enhance data diversity. These include adjusting the image's brightness (randomly increasing or decreasing it by 10 from the current brightness), contrast (adjusting it between 0.8 and 1.2), saturation, and hue (randomly changing it by 10 units). These transformations simulate images under different lighting conditions, thereby improving the model's robustness.

e The processed image data and their annotations are organised and packaged into a format suitable for model input. This includes converting the image data to a specific data type, (e.g., a tensor) and organising the annotation data to match the model's input requirements, ensuring the training process can proceed smoothly.

### *3.2 Parameter settings and experimental platform*

Experimental setup: all experiments were conducted using the PyTorch deep learning framework. Model training and inference were performed on a single NVIDIA Quadro RTX 5000 graphics processing unit (GPU).

During training, we trained the model using the AdamW optimiser. The weight decay was configured to 0.05. The initial learning rate was set to 0.001. We employed a learning rate schedule that combines linear warm-up with cosine annealing decay. The batch size was set to 4 for each iteration.

### *3.3 Comparative algorithms*

In the experimental section, the proposed model is compared with several excellent change detection algorithms from recent years to validate its effectiveness. The algorithms used for comparison are introduced as follows:

a ChangeFormer (Bandara and Patel, 2022): this algorithm proposes a new transformer-based Siamese network architecture, ChangeFormer, The proposed architecture is composed of two primary components tailored for change detection in remote sensing images. It features a powerful Transformer encoder, responsible for learning robust, context-aware feature representations, and a computationally efficient, lightweight MLP decoder that generates the final change map.

b BiT (Noman et al., 2024): the bi-temporal image transformer (BiT) introduces a novel approach by conceptualising change detection as a tokenisation process. It is designed to abstract the high-level semantic information of temporal changes, encoding them into a learned vocabulary of discrete 'visual words' or semantic tokens. The core of the BiT framework is a transformer encoder, which operates on these semantic tokens to perform high-level spatiotemporal modelling. Subsequently, these contextually-enriched tokens are projected back into the pixel space. In the final stage, a transformer decoder leverages these feature-rich projected tokens to refine the initial pixel-level features and produce the final change map.

c SAM-CD (Ding et al., 2024): this algorithm proposes a change detection model called SAM-CD, the core of this approach is the utilisation of a VFM, which serves as a powerful feature extraction backbone. By harnessing the VFM's robust, pre-trained representations, the model significantly enhances the accuracy of change detection in high-resolution remote sensing images. The algorithm uses a FastSAM visual encoder and introduces a convolutional adapter and a semantic learning branch to enhance the model's ability to recognise changes in remote sensing images.

d BAN: this algorithm proposes a new method called the BAN to improve remote sensing change detection. BAN leverages the extensive knowledge of a base model by extracting general features from a frozen base model and injecting them into a bi-temporal adapter branch via a bridge module to extract task-or domain-specific features.

e Changer (Fang et al., 2023): this algorithm proposes a new generic change detection architecture, MetaChanger, and, based on it, introduces two derivative models: ChangerAD and ChangerEx. These two derivative models adopt aggregation-distribution and feature-swapping strategies, respectively. To better fuse and align bi-temporal features, a flow-based dual-alignment fusion (FDAF) module is introduced.

### *3.4 Evaluation metrics*

In the experiments, five evaluation metrics are adopted to measure the prediction results of the change detection algorithm: precision (P), recall (R), intersection over union (IoU), overall accuracy (OA), and the comprehensive evaluation metric F1. The calculation formulas are shown in (25)–(30), where a true positive (*TP*) is defined as an instance that is correctly classified as belonging to the positive class; false positive (*FP*) indicates an incorrect positive prediction (the model predicts positive but the label is negative); true negative (*TN*) indicates a correct negative prediction (the model predicts negative and the label is negative); and false negative (*FN*) indicates an incorrect negative prediction (the model predicts negative but the label is positive).

$$Precision = \frac{TP}{TP + FP} \tag{26}$$

$$Recall = \frac{TP}{TP + FN} \tag{27}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{28}$$

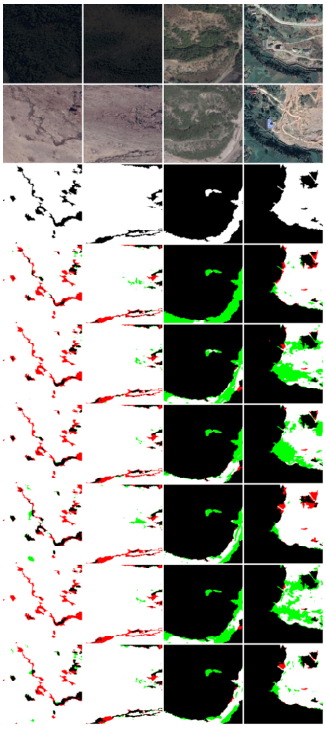
$$OA = \frac{TP + TN}{TP + TN + FP + FN} \tag{29}$$

$$F_1 = \frac{2}{Recall^{-1} + Precision^{-1}} \tag{30}$$

### *3.5 Experimental results and analysis*

The quantitative performance of our proposed method was benchmarked against several leading algorithms on the GVLM dataset, with the comparative results summarised in Table 1. Our model demonstrates superior performance, achieving the highest scores across all five evaluation metrics: it holds a leading position on the three comprehensive metrics – IoU, OA, and F1 – and ranks second on the remaining two. This indicates that the proposed method effectively utilises the general remote sensing features provided by the foundation vision model and uses them to guide the base module's application of bi-temporal image features.

**Figure 2** The proposed method's comparative visualisation results on the GVLM dataset against state-of-the-art algorithms are presented (see online version for colours)



Notes: Each row from top to bottom represents, in sequence: T1 images, T2 images, ground truth, ChangeFormer, BIT, SAMCD, BAN, Changer, and the proposed method. The rendered colours represent true positives (white), false positives (red), true negatives (black), and false negatives (green).

Figure 2 depicts a visual comparison of the proposed method against state-of-the-art algorithms on the GVLM dataset. By analysing the error maps (where red represents false positives and green represents false negatives), we observe common modes of failure in the baseline methods. Algorithms like ChangeFormer and BiT struggle with background interference, often misclassifying seasonal vegetation growth or road construction as landslides (visible as extensive red noise in the third and fourth rows). Firstly, algorithms like ChangeFormer and BiT struggle with background interference, often misclassifying

seasonal vegetation growth or road construction as landslides (visible as extensive red noise in the third and fourth rows). From the sample shown in the second column, it can be observed that the proposed method can accurately extract the landslide boundary at the bottom. The third-column sample shows that our method has an advantage in terms of the completeness of the detection area, producing relatively complete detection results. The fourth column shows a challenging sample, where some algorithms, influenced by roads and human activity areas, failed to detect the central change region. In contrast, the proposed method accurately locates the change area and excludes the influence of these factors.

The main reason for these results is that our method is based on a pyramidal multi-scale feature extraction module, which captures rich feature information at different levels. Under the guidance of the base model and attention mechanism, it suppresses the influence of irrelevant factors, thus providing a fine-grained and accurate change map for the change detection task.

**Table 1** The quantitative results of different CD methods on GVLM (%)

| *Algorithm* | *P* | *R* | *IoU* | *OA* | $F_1$ | *Params(M)* |
|---|---|---|---|---|---|---|
| ChangeFormer | *90.07* | 88.59 | 80.71 | 98.61 | 89.33 | 41.03 |
| BIT | 87.04 | 85.95 | 76.20 | 98.24 | 86.49 | 11.95 |
| SAM-CD | 88.67 | 87.69 | 78.86 | 98.46 | 88.18 | 330.97 |
| BAN | 88.78 | *90.87* | *81.51* | *98.65* | *89.81* | 28.31 |
| Changer | 89.51 | 83.49 | 76.05 | 98.28 | 86.40 | 11.46 |
| The proposed method | *90.05* | *90.16* | *81.99* | *98.70* | *90.11* | 231.74 |

### *3.6 Analysis of the influence of base model parameters*

When using the base model, two questions need to be considered:

1 Since the proposed method uses a base model, should the base model be frozen during training.

2 The adopted base model is composed of 24 stacked basic modules.

The proposed method relies on the abstract general features it extracts, while the base module extracts features at four scales. Is it necessary to fuse the feature information from the base model in the relatively early first layer. Table 2 provides five quantitative results for different base model parameters and features on the GVLM dataset. Here, 'four-layer features' refers to injecting layers 7, 11, 15, and 23 of the pre-trained model into layers 1–4 of the base module, respectively, while 'three-layer features' refers to injecting layers 11, 15, and 23 of the pre-trained model into layers 2–4 of the base module, respectively. The results show that not freezing the pre-trained model has a significant negative impact on the results, with a substantial decrease across all metrics. This indicates that an unfrozen pre-trained model affects the base model's effectiveness in extracting general knowledge and should not be involved in gradient descent. Similarly, the general knowledge provided by the base model should not be fused in the first base module. This is because the early base modules are responsible for extracting fine-grained local feature information, and fusing high-level abstract features would interfere with this process, thereby impacting the OA.

**Table 2** Research on the parameters and features of the foundation model (%)

| *Algorithm* | *P* | *R* | *IoU* | *OA* | $F_1$ |
|---|---|---|---|---|---|
| Four-layer features and an unfrozen base model | 88.88 | 87.90 | 79.19 | 98.49 | 88.39 |
| Three-layer features and an unfrozen base model | 89.80 | 88.99 | 80.82 | 98.62 | 89.39 |
| Four-layer features and a frozen base model | *89.95* | *89.16* | *81.08* | *98.64* | *89.55* |
| Three-layer features and a frozen base model | *90.05* | *90.16* | *81.99* | *98.70* | *90.11* |

Sensitivity analysis further justifies our key design choices. Regarding attention mechanism design, cosine similarity proves more robust than standard dot-product attention by focusing on angular structural consensus, which effectively mitigates domain-induced radiometric noise. Additionally, we found a three-layer fusion (layers 11, 15, 23) to be optimal, as incorporating earlier layers introduces redundant low-level noise that degrades performance. Freezing the backbone remains essential to prevent catastrophic forgetting and manifold collapse on limited disaster data, ensuring the model functions as a stable, lightweight adapter for domain-specific tasks.

### *3.7 Ablation study of core components*

To rigorously validate the individual contribution of each core component in our framework, we conducted a systematic ablation study on the GVLM dataset, with the state-of-the-art ChangeFormer model serving as the baseline. The results are summarised in Table 3.

**Table 3** Ablation study of the proposed components on GVLM dataset (%)

| *Model* | *VFM backbone* | *Multi-scale* | *Alignment* | *P* | *R* | *IoU* | *OA* | *F1* |
|---|---|---|---|---|---|---|---|---|
| Baseline | | | | 87.10 | 86.60 | 76.75 | 98.25 | 86.85 |
| +VFM | √ | | | 89.05 | 88.43 | 79.76 | 98.50 | 88.74 |
| +Multi-scale | √ | √ | | 89.70 | 89.60 | 81.25 | 98.65 | 89.65 |
| Proposed | √ | √ | √ | *90.05* | *90.16* | *81.99* | *98.70* | *90.11* |

Starting from the ChangeFormer baseline (86.85% F1), integrating a frozen VFM backbone improved the F1 to 88.74%, highlighting the value of pre-trained knowledge. Adding the multi-scale module further increased the score to 89.65% by enhancing boundary capture. Finally, our alignment module yielded the peak performance (90.11% F1, 81.99% IoU), confirming its efficacy in filtering domain noise and bridging the gap beyond simple feature stacking.

## 4 Conclusions

In this paper, we proposed a change detection framework tailored for monitoring power transmission line corridors in disaster scenarios, effectively integrating general visual knowledge with domain-specific features via an attention-based alignment module. Experimental results on the GVLM dataset demonstrate the proposed method's superiority, achieving the highest IoU (81.99%) and F1 (90.11%) scores while accurately

delineating irregular landslide boundaries compared to state-of-the-art algorithms. In a plausible real-world application setting, this framework is designed to be deployed on high-performance servers analysing data collected from unmanned aerial vehicles (UAVs) or optical satellites. By automatically mapping geological hazards (landslides) along transmission corridors, the system serves as an early warning mechanism, allowing power grid operators to identify threatened towers and dispatch maintenance crews for structural reinforcement before a collapse occurs. However, the reliance on a large-scale VFM results in high computational complexity, which limits deployment on resource-constrained edge devices, and the method's generalisation across diverse disaster types remains to be fully validated. Future research will focus on developing lightweight models via knowledge distillation for real-time applications and exploring semi-supervised learning frameworks to address the scarcity of annotated disaster data.

## Acknowledgements

This work is supported by 'National Grand' Scientific and Technological Project of Guangdong Power Grid [Research and Application of Typhoon Disaster Assessment Technology for Power Transmission and Distribution Lines Using Multi-Drone Collaborative and Multi-Source Integrated Mapping, No. 030600KC23090016 (GDKJXM20231035)].

## Declarations

All authors declare that they have no conflicts of interest.

## References

Bandara, W.G.C. and Patel, V.M. (2022) 'A transformer-based Siamese network for change detection', *IEEE International Symposium on Geoscience and Remote Sensing (IGARSS)*, DOI: 10.48550/arXiv.2201.01293.

Bandara, W.G.C. and Patel, V.M. (2022) 'A transformer-based siamese network for change detection', *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp.207–210.

Bikis, A., Engdaw, M., Pandey, D. et al. (2025) 'The impact of urbanization on land use land cover change using geographic information system and remote sensing: a case of Mizan Aman City Southwest Ethiopia', *Scientific Reports*, Vol. 15, No. 1, p.12014.

Ding, L., Zhu, K., Peng, D. et al. (2024) 'Adapting segment anything model for change detection in VHR remote sensing images', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 62, Art. No. 5611411, pp.1–11.

Dong, S., Wang, L., Du, B. et al. (2024) 'ChangeCLIP: remote sensing change detection with multimodal vision-language representation learning', *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 208, pp.53–69.

Dulal, R. and Dulal, R. (2026) 'Brain tumour identification using improved YOLOv8', *International Journal of Complexity in Applied Science and Technology*, Vol. 2, No. 1, pp.15–38.

Fang, S., Li, K. and Li, Z. (2024) 'Changer: feature interaction is what you need for change detection', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 61, Art. No. 4400211, pp.1–11.

Fang, S., Li, K., Shao, J. et al. (2023) 'Changer: feature interaction is what you need for change detection', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 61, Art. No. 4400211, pp.1–11.

Gu, L., Xu, S-Q. and Zhu, L-Q. (2020) 'Detection of building changes in remote sensing images via FlowS-Unet', *Acta Automatica Sinica*, Vol. 46, No. 6, pp.1291–1300.

Haque, M.Z., Tazvia, M. and Kuna, A.S. (2026) 'PreStroke_ML: a machine learning approach to heat stroke prediction', *International Journal of Complexity in Applied Science and Technology*, Vol. 2, No. 1, pp.97–107.

Jasim, A.T. (2025) 'Assessing LULC dynamics in Kirkuk City, Iraq using Landsat imagery and maximum likelihood classification', *DYSONA-Applied Science*, Vol. 6, No. 1, pp.113–119.

Ji, S., Tian, S. and Zhang, C. (2020) 'Urban land cover classification and change detection using fully atrous convolutional neural network', *Geomatics and Information Science of Wuhan University*, Vol. 45, No. 2, pp.233–241.

Kumari, S., Kumar, K. and Gaurav, A. (2026) 'A comprehensive review of machine Learning techniques for detecting fraud in banking and payment services', *International Journal of Complexity in Applied Science and Technology*, Vol. 2, No. 1, pp.76–96.

Lei, T., Zhai, Y-J., Xu, Y-T., Wang, Y-B. and Gong, M-G. (2024) 'Edge guided and dynamically deformable transformer network for remote sensing images change detection', *Acta Electronica Sinica*, Vol. 52, No. 1, pp.107–117.

Li, H., Xiao, P., Feng, X. et al. (2017) 'Using land long-term data records to map land cover changes in china over 1981-2010', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 10, No. 4, pp.1372–1389.

Li, K., Cao, X. and Meng, D. (2024) 'A new learning paradigm for foundation model-based remote-sensing change detection', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 62, Art. No. 5608611, pp.1–12.

Li, Z., Tang, C., Liu, X. et al. (2023) 'Lightweight remote sensing change detection with progressive feature aggregation and supervised attention', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 61, Art. No. 5602712, pp.1–12.

Liang, Z., Li, X., Deng, P. et al. (2022) 'Remote sensing image change detection fusion method integrating multi-scale feature attention', *Acta Geodaetica et Cartographica Sinica*, Vol. 51, No. 5, pp.668–676.

Lin, M., Yang, G. and Zhang, H. (2023) 'Transition is a process: pair-to-video change detection networks for very high resolution remote sensing images', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 32, pp.57–71.

Noman, M., Fiaz, M., Cholakkal, H. et al. (2024) 'Remote sensing change detection with transformers trained from scratch', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 62, Art. No. 4405314, pp.1–14.

Paranjape, J.N., De Melo, C. and Patel, V.M. (2025) 'A mamba-based Siamese network for remote sensing change detection', *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, IEEE, pp.1186–1196.

Tian, Q., Qin, K., Chen, J. et al. (2020) 'Building change detection for aerial images based on attention pyramid network', *Acta Optica Sinica*, Vol. 40, No. 21, p.2110002.

Wang, D., Zhang, J., Xu, M. et al. (2024) 'MTP: advancing remote sensing foundation model via multitask pretraining', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 17, Art. No. 10547432, pp.11632–11654.

Wang, D., Zhang, Q., Xu, Y. et al. (2023) 'Advancing plain vision transformer toward remote sensing foundation model', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 61, Art. No. 4408115, pp.1–15.

Zhang, H., Chen, H., Zhou, C. et al. (2024) 'BiFA: remote sensing image change detection with bitemporal feature alignment', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 62, Art. No. 5626217, pp.1–17.

Zhang, X., Yu, W., Pun, M-O. et al. (2023) 'Cross-domain landslide mapping from large-scale remote sensing images using prototype-guided domain-aware progressive representation learning', *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 197, pp.1–17, DOI: 10.1016/j.isprsjprs.2023.01.015.

Zhao, S., Zhang, X., Xiao, P. et al. (2023) 'Exchanging dual-encoder-decoder: a new strategy for change detection with semantic guidance and spatial localization', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 61, Art. No. 4407816, pp.1–16.