Bitemporal Remote Sensing Change Detection With State-Space Models

Lukun Wang, *Senior Member, IEEE*, Qihang Sun, Jiaming Pei, *Graduate Student Member, IEEE*, Muhammad Attique Khan, *Member, IEEE*, Maryam M. Al Dabel, Yasser D. Al-Otaibi, and Ali Kashif Bashir, *Senior Member, IEEE*

***Abstract*—Change detection in very-high-resolution remote sensing images has gained significant attention, particularly with the rise of deep learning techniques such as convolutional neural networks and Transformers. The Mamba structure, successful in computer vision, has been applied to this domain, enhancing computational efficiency. However, much of the research focuses on improving global modeling, neglecting the role of local information crucial for change detection. Moreover, there remains a gap in understanding which structural modifications are more suited for the change detection task. This article investigates the impact of different scanning mechanisms within Mamba, evaluating five mainstream methods to optimize its performance in change detection. We propose local bitemporal change detection mamba (LBCDMamba), a novel architecture based on our proposed local–global selective scan module, which effectively integrates global and local information through a unified scanning strategy. To address the lack of fine-grained details in current models, we propose a multibranch patch attention module, which captures both local and global features by partitioning data into smaller patches. In addition, a bitemporal feature fusion module is proposed to fuse bitemporal features, improving temporal–spatial feature representation. Extensive experiments on three benchmark datasets demonstrate the superior performance of LBCDMamba, outperforming existing popular methods in change detection tasks. This work also provides new insights into optimizing Mamba for change detection, with potential applications across remote sensing and related fields.**

***Index Terms*—Change detection, feature fusion, mamba, remote sensing, scanning methods.**

Received 4 March 2025; revised 7 May 2025; accepted 29 May 2025. Date of publication 4 June 2025; date of current version 30 June 2025. *(Corresponding author: Jiaming Pei.)*

Lukun Wang and Qihang Sun are with the College of Intelligent Equipment, Shandong University of Science and Technology, Taian 271000, China (e-mail: wanglukun@sdust.edu.cn; qihang.sun@sdust.edu.cn).

Jiaming Pei is with the University of Sydney, Camperdown, NSW 2050, Australia (e-mail: jpei0906@uni.sydney.edu.au).

Muhammad Attique Khan is with the Center of AI, Prince Mohammad bin Fahd University, Al Khobar 31952, Saudi Arabia (e-mail: mkhan3@pmu.edu.sa).

Maryam M. Al Dabel is with the Department of Computer Science and Engineering, College of Computer Science and Engineering, University of Hafr Al Batin, Hafar Al Batin 39524, Saudi Arabia (e-mail: maldabel@uhb.edu.sa).

Yasser D. Al-Otaibi is with the Department of Information Systems, Faculty of Computing and Information Technology in Rabigh, King Abdulaziz University, Jeddah 21589, Saudi Arabia (e-mail: yalotaibi@kau.edu.sa).

Ali Kashif Bashir is with the Department of Computing and Mathematics, Manchester Metropolitan University, M1 5GD Manchester, U.K., and also with the Department of Computer Science and Information Technology, College of Engineering, Abu Dhabi University, Abu Dhabi 59911, United Arab Emirates (e-mail: dr.alikashif.b@ieee.org).

Digital Object Identifier 10.1109/JSTARS.2025.3576433

I. Introduction

With the development of the remote sensing imagery field, sufficient research conditions have been provided for related downstream tasks. One of these tasks is remote sensing change detection, which refers to identifying changes in content by comparing remote sensing images taken at different times. This change information is of great significance to urban planning, land use, and other areas [1]. However, remote sensing images usually have an ultrahigh resolution, but due to natural factors and technological complexity, they are often subject to external interference [2], [3], traditional methods face significant difficulties in achieving accurate detection.

With the development of deep learning, various structures have gradually been used to build detection models. For change detection, the extraction and processing of image features have a significant impact on the overall detection performance of the model. Initially, with the introduction of convolutional neural networks (CNNs) [4], [5], most models adopted a twin structure based on CNNs to extract features [6], [7], often using stacked convolutional modules to enrich the extracted features. However, such methods often lead to complex model structures. Some researchers have also conducted research based on CNN from semisupervised and other perspectives [8], [9], [10], achieving good results. Although CNNs excel at extracting local features, their ability to capture global dependencies is inherently limited by the size of the receptive field.

Subsequently, researchers turned their attention to the transformer structure, relying on it to build encoder–decoder models [11], [12], [13]. This approach has seen significant improvements compared to previous CNN-only architectures and has been widely used in this field, becoming the mainstream research direction for a considerable period. There have also been studies combining CNNs and transformers, which have achieved greater accuracy improvements compared to using a single structure, such as MCTNet [14] and EHCTNet [15]. However, despite these successes, the inherent quadratic complexity of transformers in computation has led to increased computational consumption, posing a challenge for model lightweight.

Recently, the Mamba structure has been proposed [16]. Compared to transformers, Mamba's significant advantage lies in its ability to focus on global features while reducing computational consumption, providing a new way to balance computational cost and accuracy improvement. The proposal of this structure has made new breakthroughs in the field of change

detection [17], [18], [19]. However, current research is primarily focused on its global modeling capabilities, often neglecting the importance of effectively integrating local detail information. In the field of change detection, global context information plays a vital role in capturing large-scale spatial structures and understanding the overall semantic layout of a scene, which is crucial for reducing false positives in cluttered or noisy environments. Conversely, local details are essential for identifying subtle or small-scale changes—such as the emergence or removal of fine-grained objects such as small buildings or roads—that may otherwise be overlooked by purely global models. Therefore, researchers have conducted extensive studies on these two aspects. For example, Ma et al. [20] designed the TFF and SFF modules to process detailed information. Noman et al. [21] proposed the ELGC-Net, which extracts local features through depthwise convolution and processes local and global features through an aggregation module. Xiao et al. [22] designed the DFC module to compensate for detailed information lost during the extraction process. Wu et al. [23] proposed CDXLSTM, which includes an XLSTM-based feature enhancement layer and has a strong global context perception. Huang et al. [24] proposed MFDS-Net, which employs a global semantic-based approach to achieve more refined descriptions of changes. Thus, effectively combining global and local information is critical to achieving precise and comprehensive change detection performance. There has been some exploration into the internal scanning structure of Mamba [25], [26], [27], which has led to noticeable performance improvements.

Yet, no work has so far discussed and analyzed the application effects of different scanning methods in the field of change detection. There is also no consensus on the optimal adaptability of the Mamba architecture for capturing local features in change detection. In change detection, there is still the issue of insufficient precision in detecting subtle changes. Accurate prediction relies on both local details and global context information of the image. Therefore, enabling the model to extract richer features is key to further development in the field of change detection. In this process, the computational cost of the model has also attracted the attention of researchers, and achieving a balance between computational cost and performance is also an important research perspective.

In this article, we explore various scanning structures in the Mamba architecture to assess their effects on change detection tasks. We propose local bitemporal change detection Mamba (LBCDMamba), which improves the integration of local and global information in Mamba models. Our key contributions are as follows.

1) After extensively exploring the effects of different scanning structures on the change detection task, we propose a novel change detection architecture, LBCDMamba, based on our discoveries. We also compared the application effects of various scanning methods in the Mamba architecture for the field of change detection. Based on the characteristics of high-resolution images and the demand for high-precision detection, we proposed a new scanning method. This provides a new perspective for the further exploration of Mamba applications and for other fields that require the processing of high-resolution images.
2) We present a multibranch patch attention (MBPA) module that integrates local and global features, and a bitemporal feature fusion (BTFF) module that enhances pixelwise fusion of multiscale bitemporal image details for improved temporal feature representation.
3) Through extensive experiments conducted on three widely used benchmark datasets, we demonstrate that LBCD-Mamba consistently outperforms existing state-of-the-art (SOTA) models, achieving superior performance across multiple evaluation metrics.

The rest of this article is organized as follows. Section II reviews the related work. Section III focuses on the overall architecture of the proposed model and the innovative approaches. Section IV presents the datasets used and the experimental setup, along with comparative experimental results and analysis. Section V explains the limitations of the method in the field of change detection and future research directions. Finally, Section VI concludes this article.

II. Related Works

A. CNN-Based Models

CNNs have become a cornerstone in change detection tasks due to their strong ability to extract spatial features from images. Daudt et al. [28] introduced FC-Siam-diff, which brought fully convolutional networks into the realm of change detection, providing a new approach to spatial feature extraction. Similarly, Fang et al. [29] proposed SNUNet-CD, utilizing information transmission to mitigate the challenge of local information loss inherent in deep neural networks. Ye et al. [30] proposed an end-to-end change detection method, utilizing 3-D convolution. Zhang et al. [7] detected changes in buildings and trees through a Siamese CNN. Zhang et al. [31] proposed CAMixer, applying both convolution and attention mechanisms in the model, being the first to explore a Transformer-like network for the interpretation of multitemporal SAR data.

B. Transformer-Based Models

Vaswani et al. [11] proposed transformer, transformer offers a more robust solution for global modeling by capturing long-range dependencies. In the field of change detection, research is gradually being conducted around it. For instance, Chen et al. [12] proposed BIT, effectively leveraging transformers to extract features and model spatial-temporal contextual information. In addition, Bandara et al. [13] introduced ChangeFormer, which integrates a transformer encoder with an MLP decoder in a unified architecture. Lu and Huang [32] proposed the relational change detection transformer based on the Transformer, achieving superior change detection performance on multiple datasets through a shared-weight backbone network, cross-attention module, and feature constraint module. Xu et al. [33] proposed UCDFormer and applied it to unsupervised change detection, taking into account the style and seasonal differences in bitemporal images, achieving good results in such applications. Jiang et al. [34] proposed VcT, constructing the backbone network based on the Transformer and combining it with graph neural networks for modeling to achieve the

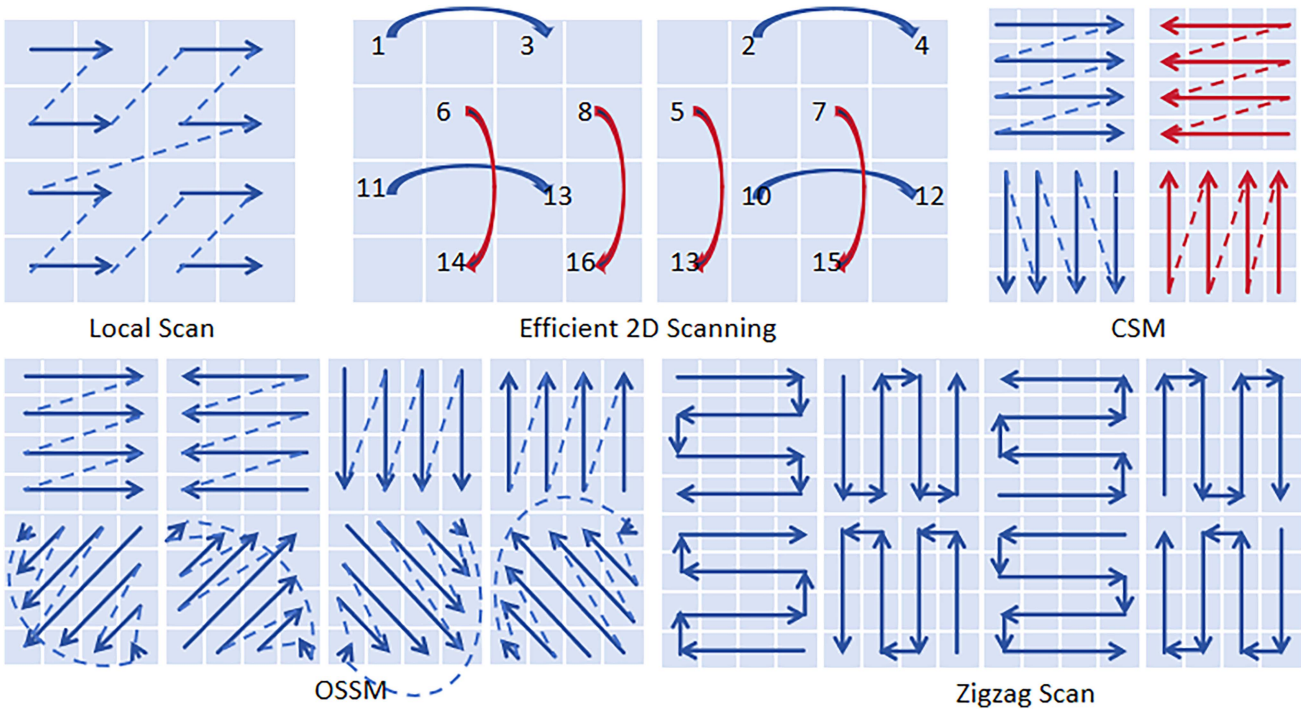

Fig. 1. Illustrations of five distinct scanning direction structures, namely Local Scan, Efficient Scan, CSM, OSSM, and Zigzag Scan.

prediction of change graphs. Zhu et al. [35] proposed ChangeViT, addressing the disadvantage of CNNs in dealing with large-scale changes.

Some researchers have attempted to combine CNNs with Transformers. Li et al. [14] proposed MCTNet, which fuses CNNs and Transformers in a multiscale manner and achieves good detection performance for changes of different sizes. Yang et al. [15] proposed EHCTNet, which enhances the hybrid of CNN and Transformer to achieve more complete detection of changed regions. Other researchers have focused on general functional modules based on backbone networks constructed by CNNs or Transformers. Wang et al. [36] proposed CAT, which is used to perceive feature differences and can be integrated with both Transformers and CNNs.

C. Mamba-Based Models

After the introduction of Mamba, it has been applied in multiple fields. Liu et al. [37] introduced VMamba, a visual backbone based on the Mamba structure that achieved exceptional results in classification, detection, and segmentation. Ma et al. [38] proposed BS-Mamba for black soil area assessment. Xie et al. [39] proposed ProMamba for the field of image segmentation, achieving accurate segmentation of polyps. Zhang et al. [40] proposed Point Cloud Mamba, which more effectively processed point cloud data and surpassed the SOTA methods at that time. Ghazaei and Aptoula [41] proposed a change detection model by combining a lightweight CNN encoder and Vision Mamba, effectively reducing the number of parameters and improving computational efficiency. Based on the Mamba structure, researchers have gradually applied it to the field of change detection. Wu et al. [42] proposed CD-Lamba, which employs local enhancement, a cross-temporal state-space model (SSM) scanning strategy, and enhanced interaction between segmentation windows to improve its performance in change detection. Paranjape et al. [19] proposed M-CD and introduced a different module to combine image features. Kuang and Ge [43] proposed 2DMCG based on a variant of Vision Mamba, enhancing Mamba's ability to capture 2-D spatial information. Zhang et al. [17] proposed CDMamba, effectively combining local and global features. Chen et al. [18] proposed ChangeMamba, which successfully applied the Mamba architecture to remote sensing change detection, achieving SOTA performance across multiple datasets. Zhao et al. [44] further enhanced this architecture with RS-Mamba, utilizing an omnidirectional scanning approach to improve spatial feature extraction and enhance global modeling capacity.

D. Selected Mamba Scanning Structures

As Mamba evolves in the field of imaging, most research has focused on employing different scanning methods [37], [44], [25], [26], [27] to learn spatial relational features of images. Despite some progress, the ideal scanning structure for change detection remains underexplored. This article aims to fill this gap by systematically evaluating five mainstream Mamba scanning structures, as illustrated in Fig. 1.

Local Scan [25] divides the input image into multiple independent small windows, where each window is scanned independently, thereby avoiding the loss of local dependencies when tiling spatial data. In addition, the scanning order within each window follows a left-to-right and top-to-bottom pattern. By combining window scanning with horizontal and vertical scanning directions, LocalScan enhances the ability to capture local

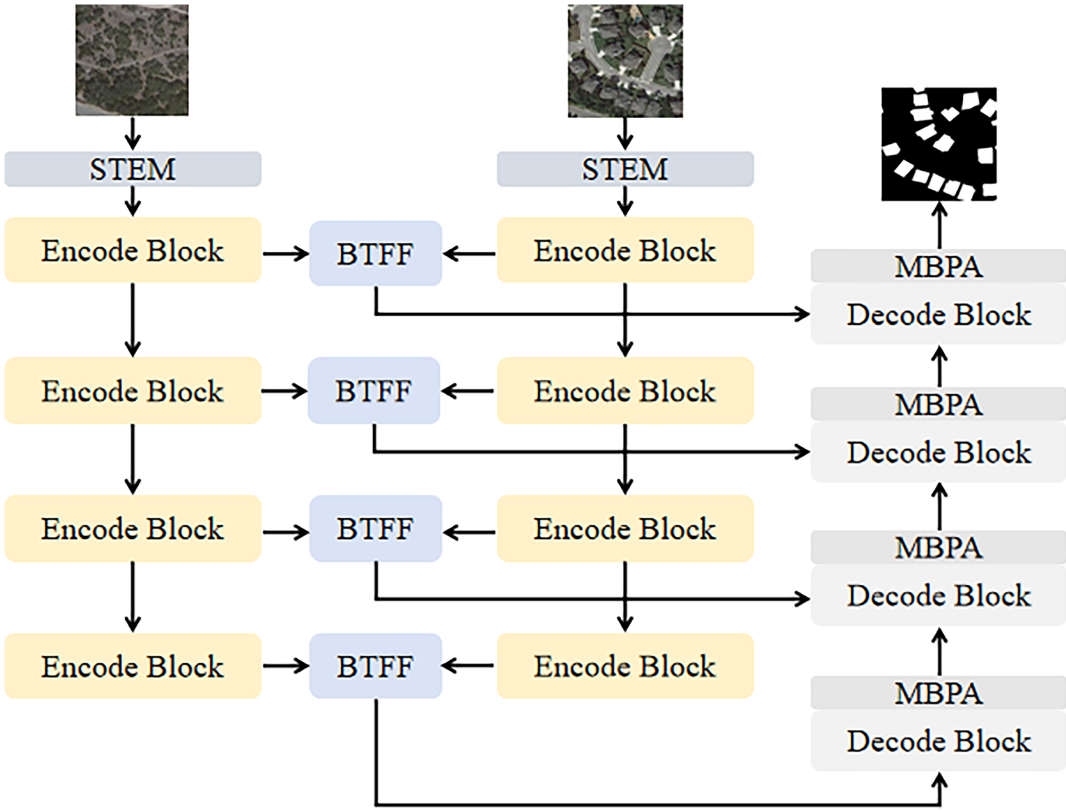

Fig. 2. Complete structure of LBCDMamba. The BTFF module and the MBPA module play crucial roles in the LBCDMamba. The input images are divided into multiple patches by the STEM module, and then, fed into the encoder. At each encoder stage, bitemporal features $\{F_i^1\}_{i=1}^4$ and $\{F_i^2\}_{i=1}^4$ are extracted. These features are further processed by the BTFF module to obtain more comprehensive global features $\{F_i^b\}_{i=1}^4$. In the decoder, each stage refines these features through the MBPA module to produce finer grained representations.

details while maintaining a global perspective. This approach allows for more precise modeling of local features without sacrificing the global context.

Efficient Scan [26] differs from other methods that focus on scan order by utilizing varying dilation rates to skip a certain number of pixels on the feature map, thereby reducing the number of feature points that need to be processed. During the scanning process, pixels are selectively skipped based on the dilation rate, allowing for feature extraction and group reorganization. By reducing the number of feature points, the computational complexity is significantly lowered from $O(N)$ to $O(N/p^2)$, where p is the dilation rate. Efficient Scan enhances the model's global perception capability without sacrificing spatial resolution, making it suitable for lightweight models in resource-constrained computational environments while maintaining high resolution.

CSM [37] addresses the limitations of traditional SSMs in visual tasks due to the nonsequential nature of image data. CSM adopts a four-way scanning strategy to expand along both rows and columns: scanning from top-left to bottom-right, from bottom-right to top-left, from top-right to bottom-left, and from bottom-left to top-right. By unfolding image patches in different spatial directions, the features obtained from each direction are then fused to model the global contextual information of the image. This approach ensures comprehensive integration of information across different directions, allowing for more robust modeling of the image's global context.

OSSM [44] combines eight different scanning sequences to achieve global scanning. Specifically, it includes scanning from left to right, from right to left, from top to bottom, from bottom to top, two along the main diagonal, and two along the antidiagonal. The input image is scanned in all these directions, and the feature sequences obtained from the scans are stacked and input into the SSM block for processing. The processed features are then merged to extract comprehensive global spatial features.

Zigzag Scan [27] takes into account the spatial continuity that was not considered in traditional Mamba frameworks during image scanning. The Zigzag scan employs a "Z"-shaped structure to perform the image scan, preserving spatial continuity. This approach leverages inductive bias in visual data, enhancing the model's ability to model the data. The specific scanning schemes can be categorized into 1 to 8 types, depending on the starting point and direction of the scanning.

III. Proposed Method

Fig. 2 illustrates our change detection framework, LBCDMamba, which is constructed on the Mamba architecture. Given two very-high-resolution (VHR) images T_1 and T_2, their features are extracted through a vision transformer (ViT)-like stem module followed by encoders. Then, the BTFF module is proposed to fuse the multiscale features from both images. In the decoder stage, four MBPA modules are proposed to integrate local and global features, which are crucial for achieving the final result.

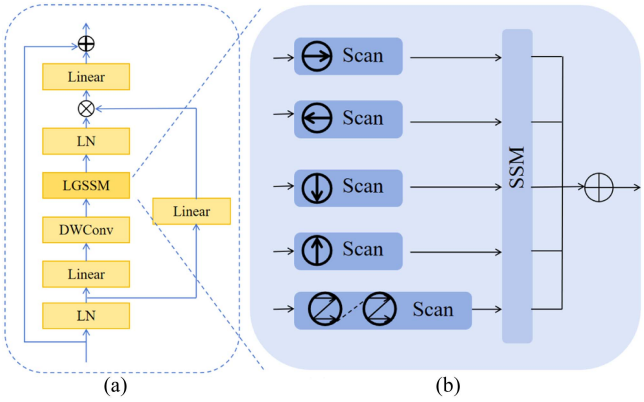

Fig. 3. Illustration of the proposed scanning structure. (a) LGSS block. (b) Local–global selective scan module (LGSSM).

A. Preliminaries

1) State-Space Models (SSMs): SSMs are mathematically defined as a 1-D function or sequence $x(t) \in \mathbb{R}$ is mapped to a response $y(t) \in \mathbb{R}$ through a hidden state $h(t) \in \mathbb{R}^N$, which is typically represented as a linear ordinary differential equation

$$h'(t) = Ah(t) + B(t) \tag{1}$$

$$y(t) = Ch(t) \tag{2}$$

where $A \in \mathbb{R}^{N\times N}$, $B \in \mathbb{R}^{N\times 1}$, and $C \in \mathbb{R}^{1\times N}$.

2) Discretization: The discretization process of SSMs involves converting continuous ordinary differential equations into discrete functions. Based on the discretization technique described in [45], the continuous parameters (A, B) can be discretized using the ZOH method

$$\bar{A} = e^{\Delta A} \tag{3}$$

$$\bar{B} \approx (\Delta A)(\Delta A)^{-1} AB = \Delta B \tag{4}$$

$$h(t) = \bar{A}h(t-1) + \bar{B}x(t) \tag{5}$$

$$y(t) = Ch(t) \tag{6}$$

where $\Delta \in \mathbb{R}^D$, $\bar{A} \in \mathbb{R}^{N\times N}$, $\bar{B} \in \mathbb{R}^{N\times 1}$, and $\bar{C} \in \mathbb{R}^{1\times N}$.

3) Output: The model computes the output in parallel via global convolution

$$\overline{K} = (C\overline{B}, C\overline{AB}, \ldots, C\overline{A}^{L-1}\overline{B}) \tag{7}$$

$$y = x * \overline{K} \tag{8}$$

where L represents the length of the input sequence x, and $\overline{K} \in \mathbb{R}^L$ is the SSM convolution kernel.

B. Architecture Overview

Considering the current limitations of the Mamba model in effectively integrating global and local information for change detection, we propose an innovative local–global selective scan module (LGSSM), as illustrated in Fig. 3(b). This module combines local window scanning with a four-directional strategy, thereby enhancing the model's ability to capture fine-grained local details while maintaining a global perspective.

To address the challenges in VHR remote sensing change detection, based on the scanning comparison experiments presented in Section IV, we propose a novel network framework named LBCDMamba, built upon the Mamba architecture, as illustrated in Fig. 2. Given a pair of bitemporal VHR remote sensing images T_1 and T_2, the inputs are first divided into multiple patches using a ViT-like stem module. These patches are then fed into the encoder to extract representative features. Considering the need for temporal feature interaction, the extracted features from T_1 and T_2 are processed through a BTFF module to generate a fused feature map, which is subsequently passed to the decoder to predict the final change detection result.

The encoder consists of four hierarchical stages, each composed of multiple local–global state-space (LGSS) blocks that incorporate the proposed LGSSM while maintaining consistent feature dimensionality. At each stage, the input is initially downsampled, followed by a sequence of LGSS blocks for progressive feature extraction. This process yields fine-grained bitemporal features $\{F_i^1\}_{i=1}^4$ and $\{F_i^2\}_{i=1}^4$ for stages $i = 1, 2, 3, 4$. Compared with the conventional VSS module, LGSSM enhances the capability of the model to integrate both local and global contextual cues via local scanning, thereby improving feature expressiveness. Our prior experiments have demonstrated the effectiveness of this strategy within the Mamba framework.

After feature extraction by the encoder, the BTFF module performs pixelwise fusion of the corresponding feature maps $\{F_i^1\}_{i=1}^4$ and $\{F_i^2\}_{i=1}^4$, rather than simple concatenation, to obtain a set of globally enhanced feature representations $\{F_i^b\}_{i=1}^4$ that capture richer semantic information across time.

The decoder also comprises four stages. In each stage, the fused bitemporal features are fed into the MBPA module, which partitions the features into multiple small patches and employs parallel branches to attend to both local and global regions. This structure enhances the expressive capacity of the feature maps by integrating multiscale contextual information. The initial decoding stage processes only the features from the two deepest encoder stages. Each stage employs a spatiotemporal interaction mechanism, where multiple parallel LGSS blocks are used to model the temporal relationships between $\{F_i^1\}_{i=1}^4$ and $\{F_i^2\}_{i=1}^4$, and the learned features are combined with those from the previous decoder stage.

During upsampling, the globally fused features $\{F_i^b\}_{i=1}^4$ are introduced to enhance semantic representation from a global perspective. The output features at each stage are then forwarded to the subsequent decoder layer to progressively refine the final change detection prediction.

C. LGSS Block

Given the limitations of the current Mamba model in effectively integrating global and local information for change detection, we propose an innovative selective scanning mechanism termed LGSSM. This mechanism combines the four-directional scanning strategy from CSM with a local scan, forming multiple distinct scanning branches capable of independently processing features from each input image. The outputs from these branches are subsequently fused into a unified representation within the SSM. By integrating both local and

directional global scanning, this method enhances the model's ability to capture fine-grained spatial details while preserving global contextual information.

As illustrated in Fig. 3(a), the LGSS block serves as a fundamental component of the proposed LBCDMamba model. In this module, the input feature map first undergoes layer normalization to ensure numerical stability during subsequent computations. The normalized features are then split into two information streams and processed along separate paths. One of these streams is passed through a linear projection, followed by a 3×3 depthwise convolution and a SiLU activation function. During this process, enhanced features are extracted and passed to the proposed LGSSM module, whose primary function is to jointly extract both local and global representations from the input, thereby enriching the expressive power of the learned features.

After LGSSM processing, the output feature map is again normalized via a second layer normalization step, and then fused, with the output of the other information stream. The fused features are passed through another linear transformation to enable a deeper mixing of semantic information. Finally, a residual connection adds the original input features back to the transformed output, which helps stabilize gradient flow and enhances the model's training capacity.

As shown in Fig. 3(b), the LGSSM effectively integrates the four-directional scan with the local scanning mechanism to simultaneously extract global and local features from the input imagery. Specifically, it constructs multiple scan branches—each designed to independently process different spatial aspects of the input data. These independently processed features are then aggregated within the LGSS block into a unified output. This hybrid scanning design significantly enhances the model's capability to detect subtle spatial changes while maintaining awareness of the broader scene context.

Within the LGSS block, the features processed through each individual scan branch are merged into a unified output through a fusion operation. This architectural design enables LGSSM to efficiently combine local window scanning with directional global strategies, thereby improving the model's capacity to capture both localized details and long-range dependencies. Global information is essential for recognizing large-scale scene-level changes, while local scanning ensures that small, fine-grained variations are not overlooked—an especially critical requirement in remote sensing applications where such changes may signify meaningful land surface transformations. Thus, by jointly modeling local and global dependencies, LGSSM becomes particularly well-suited for high-precision remote sensing change detection, allowing the model to focus on both coarse-scale and fine-scale alterations within complex imagery.

D. BTFF Module

Considering spatial–temporal relationships and scale variations in dual-temporal images, we introduce a novel BTFF module. This module applies convolution and batch normalization to the features extracted by the encoder. Elementwise multiplication is used to integrate the features and generate a similarity map for weighted fusion. The Sigmoid function is employed to compute attention coefficients, which dynamically assign weights and prioritize key regions. The structural design of this module is illustrated in Fig. 4.

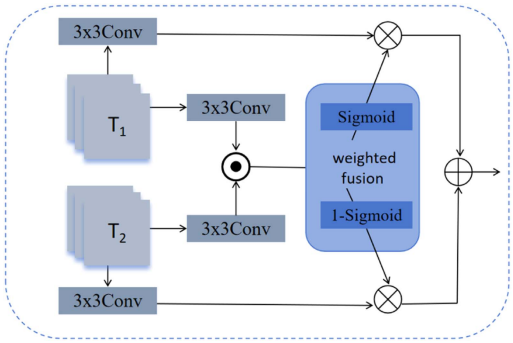

Fig. 4. Operation process of the BTFF module.

Initially, convolution operations are performed to extract contextual information from the bitemporal images. Subsequently, the features at each pixel location are fused through a weighted mechanism, where the attention coefficients are calculated using the Sigmoid function. This allows the module to determine how to allocate weights between the features of T_1 and T_2 adaptively

$$\alpha_i = \text{Sigmoid}\left(f(T_1[i]) \cdot f(T_2[i])\right) \tag{9}$$

Here, $f(T_1[i])$ and $f(T_2[i])$ denote the feature representations of the ith pixel extracted from T_1 and T_2, respectively. The symbol "$\cdot$" denotes elementwise multiplication. α_i represents the attention coefficient corresponding to the ith pixel, with its value constrained within the range $[0, 1]$. It is used to control the degree of reliance on the features from each temporal instance.

Using the computed attention coefficient α_i, the features from T_1 and T_2 are fused in a weighted manner to generate the final feature representation. The fusion process is formally defined as

$$\text{Out}[i] = \alpha_i \cdot T_1[i] + (1 - \alpha_i) \cdot T_2[i]. \tag{10}$$

Based on the fused feature map Out$[i]$, the BTFF module further computes the differences between the two temporal instances. These differences are then processed through a Sigmoid function to produce the change feature map

$$\text{Change Map} = \text{Sigmoid}(\|T_1 - T_2\| + \|F_{\text{out}} - T_1\|). \tag{11}$$

Here, $\|T_1 - T_2\|$ represents the difference between the two temporal inputs T_1 and T_2, while $\|F_{\text{out}} - T_1\|$ denotes the difference between the fused features and the features from T_1. The final output feature map is obtained through the weighted fusion operation.

E. MBPA Module

In the current Mamba models applied to change detection, there is often a focus on global modeling, which neglects the critical information provided by detail cues. To address this, we propose the MBPA module, which extracts features through three parallel branches: the local branch, the global branch,

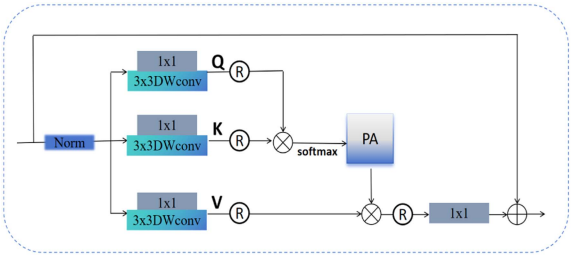

Fig. 5. Operation process of the MBPA module.

and the dilated convolution branch. By leveraging this parallel architecture, the mechanism effectively captures both global and local information present in the image. Meanwhile, it adaptively enhances discriminative features and suppresses irrelevant content through attention modulation. The structure of this module is illustrated in Fig. 5.

Let the input feature map be denoted as $Y \in \mathbb{R}^{H\times W\times C}$, which is first processed using layer normalization to obtain a normalized feature representation. Subsequently, 1×1 convolutions with learnable weights W_c^Q, W_c^K, and W_c^V, together with 3×3 depthwise convolutions parameterized by W_d^Q, W_d^K, and W_d^V, are employed to compute the Q, K, and V feature maps. Each branch extracts features at a distinct spatial scale using convolutional operations, allowing the module to adapt to changes of varying sizes and capture both fine-grained local details and broader contextual cues. The attention mechanism is applied within each branch to dynamically weight features. The computation process is formulated as follows:

$$Q = W_d^Q \cdot W_c^Q \cdot Y \tag{12}$$

$$K = W_d^K \cdot W_c^K \cdot Y \tag{13}$$

$$V = W_d^V \cdot W_c^V \cdot Y. \tag{14}$$

The 1×1 convolution is responsible for aggregating information along the channel dimension. To enhance local context representation, each channel is further processed by a 3×3 depthwise convolution prior to the computation of feature covariance. This depthwise convolution enables the incorporation of spatial context while preserving local feature integrity.

Subsequently, the attention map A is obtained by computing the dot product between the Q and K, which encodes global contextual dependencies

$$A = \text{Softmax}\left(\frac{K^\top Q}{\alpha}\right). \tag{15}$$

In this context, α is a learnable scaling parameter that controls the magnitude of the dot product. This attention feature map highlights the regions of change through weighted enhancement while simultaneously suppressing background noise interference.

Subsequently, as shown in (16), the attention map $\mathbf{A}$ is multiplied with the value matrix $\mathbf{V}$ to obtain the weighted feature map $\mathbf{X}$. This process adaptively adjusts the weights of the features to emphasize important regions, thereby assisting the MBPA module in focusing on areas of change.

Finally, a residual connection is employed to integrate the input feature map $\mathbf{X}$ with the weighted output $\hat{\mathbf{X}}$ through elementwise addition, ensuring effective information propagation and preventing information loss in deeper network layers.

$$\hat{X} = A \cdot V \tag{16}$$

$$\hat{X}_{\text{final}} = \hat{X} + X. \tag{17}$$

The final input feature map $\mathbf{X}$ corresponds to the original feature map $\mathbf{Y}$ after undergoing a series of transformations through the attention mechanism.

F. Optimization With Combined Loss

Since this study aims to explore which type of scanning structure is more suitable for change detection in remote sensing imagery, and considering that change detection is essentially a specific form of semantic segmentation [46], the commonly used cross-entropy (CE) loss is adopted as the loss function. The objective is to minimize the discrepancy between the predicted class probabilities and the ground truth labels. The CE loss function is defined as follows:

$$\mathcal{L}_{\text{ce}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{x=0}^{1}\tilde{Y}_i(x)\log\left(P_i(x)\right). \tag{18}$$

Here, $\tilde{Y}_i$ is the one-hot encoded representation of Y_i, where each sample is assigned a class label (1 for change, 0 for no change). P_i denotes the predicted probability for binary change detection, obtained through the *softmax* activation function, indicating the likelihood that each pixel belongs to either the change or no-change class.

In change detection tasks, pixels belonging to changed areas are typically considered positive samples, while those in unchanged areas are treated as negative samples. However, due to the extreme class imbalance between changed and unchanged pixels in remote sensing images, the CE loss function is prone to bias the model toward the majority class, potentially ignoring minority classes. This class imbalance problem becomes particularly severe when the dataset is unevenly distributed, which may lead to a decline in model performance. To address this issue, we incorporate the Lovász-softmax loss [47] to effectively mitigate the performance degradation caused by sample imbalance during training. In remote sensing change detection, changed pixels are often much fewer than unchanged pixels, and conventional loss functions may fail to capture the distinctive features of the changed regions. The Lovász-softmax loss optimizes the objective function in a way that better handles class imbalance, enhances the model's ability to learn from minority classes (i.e., change regions), and consequently, improves performance on imbalanced datasets. The overall loss can be represented as follows:

$$L_{\text{final}} = \gamma_1 L_{\text{ce}} + \gamma_2 L_{\text{lov}} \tag{19}$$

where L_{ce} is used for category recognition. L_{lov} is used to mitigate the impact of the sample imbalance between changed and unchanged pixels. Loss items factors, $\{\gamma_1, \gamma_2\}$ are set as $\{1.0, 0.75\}$.

IV. Experiments

A. Datasets Introduction

We conducted experimental evaluations on three publicly available benchmark remote sensing datasets: LEVIR-CD, LEVIR-CD+ [48], and WHU-CD [49]. The first two datasets primarily focus on diverse building changes in urban areas, emphasizing fine-grained variations within dense scenes. In contrast, WHU-CD includes a variety of complex land cover scenarios and mainly targets large-scale change objects that are generally sparse and irregular in distribution. Detailed information about these datasets is provided as follows.

1) LEVIR-CD: This dataset features 637 high-resolution image pairs from Google Earth, showing building changes across the U.S. over 5–14 years. Each image has a size of 1024×1024 pixels with 0.5-m resolution, encompassing various changes, such as building additions and demolitions, road expansions, vegetation changes, etc. The urban-centric nature of LEVIR-CD and its well-annotated binary change masks make it an ideal benchmark for evaluating model performance in structured and high-density environments. Following the official protocol, the dataset is divided into nonoverlapping patches of 256×256 pixels, which are randomly allocated to the training, validation, and testing sets in a ratio of 7:1:2.

2) LEVIR-CD+: This dataset is an extended version of the LEVIR-CD, comprising 985 pairs of VHR images obtained from Google Earth. This dataset retains the images from LEVIR-CD and highlights a diverse array of buildings such as urban homes, compact garages, and expansive warehouses, featuring 31 333 building instances. This dataset introduces additional urban scenes with more diverse building types, providing a more comprehensive evaluation of the model's robustness and generalization capability in complex urban environments. Following the official protocol, the dataset is divided into nonoverlapping patches of 256×256 pixels, with 10 192 samples used for training and 5 568 for testing.

3) WHU-CD: This dataset comprises a pair of high-resolution images from New Zealand, measuring 32507×15354 pixels with a resolution of 0.2 m per pixel. Captured in April 2012 and April 2016, it covers an area of 20.5 km^2. This dataset contains 12 796 buildings in the image acquired in 2012 and 16 077 buildings in the corresponding image acquired in 2016, reflecting a wide range of complex land-cover and structural changes. Following commonly adopted practices in recent studies, the image pair is divided into nonoverlapping patches of 256×256 pixels and split into training, validation, and testing sets with a ratio of 6:2:2.

B. Experimental Setup

1) Implementation Details: All experiments were conducted on a workstation equipped with an Intel(R) Xeon(R) Platinum 8352 V CPU at 2.10 GHz and a GeForce RTX 4090 GPU(with 24-GB memory). The proposed LBCDMamba was implemented using Python 3.8 and PyTorch 2.0.0. We optimized the network using the AdamW optimizer [50], and hyperparameter settings were summarized as follows. The initial learning rate, weight decay, and momentum were set to 0.0001, 0.0005, and 0.9, respectively, without employing any learning rate scheduling strategies. The batch size was set to 6, and training was conducted for 120 000 iterations. Data augmentation techniques, including random horizontal flipping, vertical flipping, and 90°rotation, were applied to enhance model generalization.

2) Evaluation Metrics: In our experiments, we employed five evaluation metrics to assess the model performance: recall (Rec), precision (Pre), overall accuracy (OA), F1-score (F1), intersection over union (IoU), and Cohen's Kappa. Pre measures the proportion of correctly predicted change pixels among all pixels predicted as changed. Rec indicates the proportion of actual change pixels that were successfully detected. OA reflects the ratio of correctly classified pixels to the total number of pixels. F1 is the harmonic mean of Pre and Rec, balancing both metrics. IoU evaluates the degree of overlap between the predicted change regions and the ground truth. Cohen's Kappa is used to evaluate the agreement between predicted and ground truth change labels. Notably, Cohen's Kappa is only employed in the ablation study of different scanning structures. The detailed formulas for these metrics are presented as follows:

$$\text{Pre} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{20}$$

$$\text{Rec} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{21}$$

$$\text{F1} = \frac{2 \cdot \text{Pre} \cdot \text{Rec}}{\text{Pre} + \text{Rec}} \tag{22}$$

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{23}$$

$$\text{OA} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{24}$$

$$\text{kappa} = = \frac{P_o - P_e}{1 - P_e} \tag{25}$$

where TP, TN, FP, and FN denote the numbers of true positives, true negatives, false positives, and false negatives, respectively. P_o denotes the observed agreement between the prediction and ground truth, while P_e represents the expected agreement occurring by random chance.

C. Comparative Experiments

To rigorously evaluate the effectiveness of the proposed LBCDMamba, we conducted a comparative study against several representative and SOTA change detection methods. Our benchmark tests utilized the same datasets with identical data splits and uniform data settings. The comparative models are grouped into three categories: first, CNN-based approaches such as FC-Siam-Conc [28], HANet [51], SNUNet [29], and HCGMNet [52]; second, Transformer-based strategies including ChangeFormer [13] and BIT [12]; and third, Mamba-inspired methods such as ChangeMamba [18], RSM-CD [44], and MambaCD [17].

For the comparison methods, we utilized the performance metrics reported in their original publications whenever available. If such metrics were not provided, we trained and evaluated

TABLE I
PERFORMANCE COMPARISON ON LEVIR-CD DATASET

Method	Rec.	Pre.	OA	F1	IoU
FC-Siam-Conc [28]	76.77	91.99	98.49	83.69	71.96
HANet [51]	89.36	91.21	99.02	90.28	82.27
SNUNet [29]	87.17	89.18	98.82	88.16	78.83
HCGMNet [52]	90.61	**92.96**	99.18	91.77	84.79
ChangeFormer [13]	88.80	92.05	99.04	90.40	82.48
BIT [12]	89.37	89.24	98.92	89.31	80.68
ChangeMamba [18]	90.41	91.15	99.06	90.74	83.12
RSM-CD [44]	89.73	92.52	–	91.10	83.66
MambaCD [17]	90.08	91.43	99.06	90.75	83.07
Ours	**91.64**	92.92	**99.26**	**92.25**	**85.71**

The highest score is marked in bold. All the scores are described in percentage (%).

TABLE II
PERFORMANCE COMPARISON ON LEVIR-CD+ DATASET

Method	Rec.	Pre.	OA	F1	IoU
FC-Siam-Conc [28]	78.49	78.39	98.24	78.44	64.53
HANet [51]	75.53	79.70	98.22	77.56	63.34
SNUNet [29]	78.73	71.07	97.83	74.70	59.62
HCGMNet [52]	81.94	82.81	98.57	82.37	70.03
ChangeFormer [13]	79.97	81.34	98.44	80.65	67.58
BIT [12]	81.84	85.02	98.67	83.40	71.53
ChangeMamba [18]	81.08	83.36	98.57	82.20	69.79
RSM-CD [44]	80.27	84.49	–	82.32	69.96
MambaCD [17]	81.00	85.11	98.65	83.01	70.95
Ours	**83.34**	**86.42**	**98.79**	**84.85**	**73.43**

The highest score is marked in bold. All the scores are described in percentage (%).

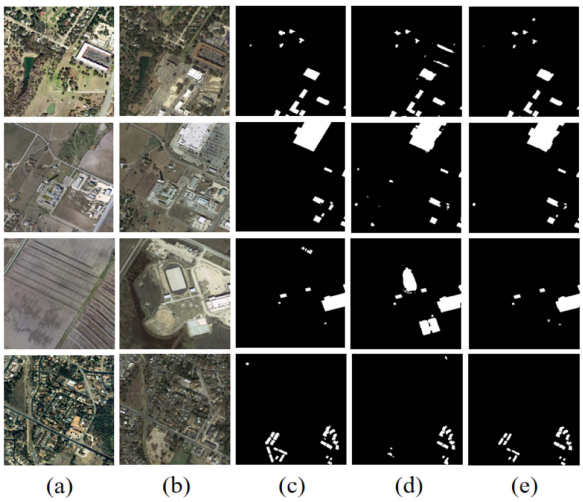

Fig. 6. Some inference results of LBCDMamba on the LEVIR-CD dataset. (a) T1 images. (b) T2 images. (c) Ground-truth images. (d) Baseline. (e) Ours.

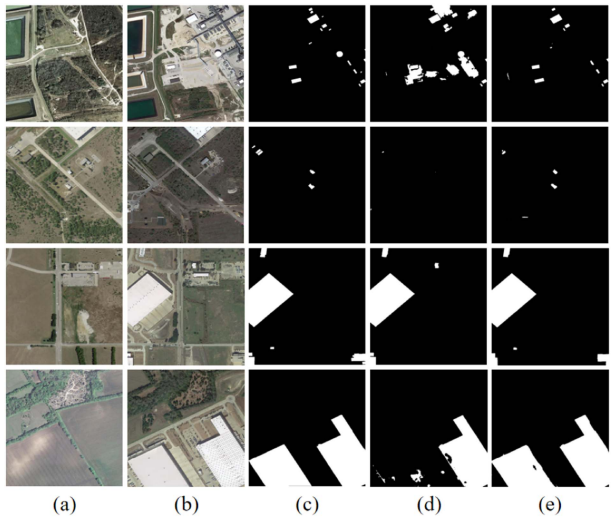

Fig. 7. Some inference results of LBCDMamba on the LEVIR-CD+ dataset. (a) T1 images. (b) T2 images. (c) Ground-truth images. (d) Baseline. (e) Ours.

the models using their official code repositories under consistent conditions, including the same loss function and data augmentation strategies, to ensure a fair comparison.

1) Experimental Results on LEVIR-CD: Table I illustrates the quantitative comparison on the LEVIR-CD dataset. As observed, although LBCDMamba yields a slightly lower precision compared to HCGMNet, it surpasses all competing methods in key performance metrics, including Rec, F1-score (92.25%), IoU (85.71%), and overall accuracy. This clearly demonstrates the effectiveness of the proposed method in change detection tasks. Notably, even when compared with representative CNN-, Transformer-, and Mamba-based methods, LBCDMamba achieves consistent improvements in F1-score by 0.46%, 1.85%, and 1.15%, respectively.

Fig. 6 presents some qualitative comparisons results on the LEVIR-CD dataset. For both dense and subtle building changes, the proposed LBCDMamba demonstrates superior capability in accurately identifying changed regions compared to the baseline method. By enhancing the representation of both global and local features, LBCDMamba effectively mitigates false negatives and false positives, thereby improving the precision of building change detection.

2) Experimental Results on LEVIR-CD+: Table II shows the quantitative comparison results on the LEVIR-CD+ dataset. Experimental results demonstrate that our method achieves optimal performance in terms of Rec, Pre, OA, F1, and IoU, indicating its strong capability in accurately detecting changes in both densely and sparsely distributed target regions. This comprehensive improvement highlights the effectiveness of our approach in enhancing overall model performance for change detection tasks. Notably, LBCDMamba exceeds other methods with the Mamba structure by nearly 3% in IoU, further confirming the effectiveness of our designed Mamba structure.

Fig. 7 presents some qualitative comparisons on the LEVIR-CD+ dataset. The LEVIR-CD+ dataset comprises more complex and diverse building change types, posing higher demands on the model's generalization capability and boundary delineation accuracy. The visualization results demonstrate that LBCDMamba significantly outperforms the baseline method in terms of change region localization and boundary detail restoration. By leveraging superior global–local modeling capabilities in conjunction with a BTFF mechanism, our method achieves precise localization of object boundary changes while effectively reducing false positives and missed detections.

TABLE III
PERFORMANCE COMPARISON ON WHU-CD DATASET

Method	Rec.	Pre.	OA	F1	IoU
FC-Siam-Conc [28]	87.72	84.02	98.94	85.83	75.18
HANet [51]	88.30	88.01	99.16	88.16	78.82
SNUNet [29]	87.36	88.04	99.10	87.70	78.09
HCGMNet [52]	90.31	93.93	99.45	92.08	85.33
ChangeFormer [13]	85.55	88.25	99.05	86.88	76.80
BIT [12]	90.33	89.70	99.26	90.01	81.84
ChangeMamba [18]	91.49	94.91	99.47	93.18	88.07
RSM-CD [44]	90.42	93.37	–	91.87	84.96
MambaCD [17]	92.01	95.58	99.51	93.76	88.26
Ours	**92.54**	**95.75**	**99.57**	**94.13**	**88.94**

The highest score is marked in bold. All the scores are described in percentage (%).

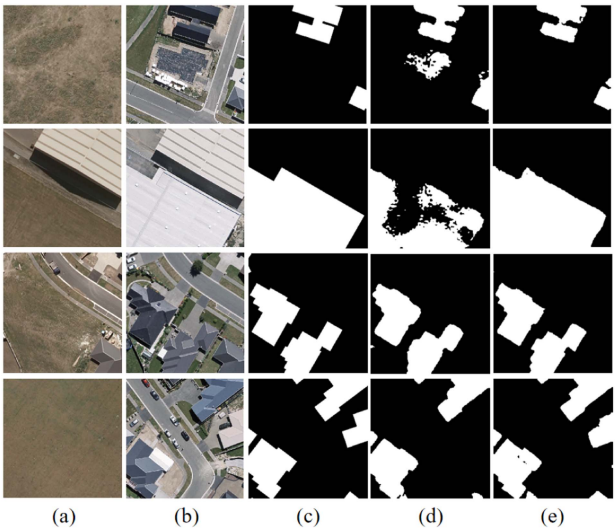

Fig. 8. Some inference results of LBCDMamba on the WHU-CD dataset. (a) T1 images. (b) T2 images. (c) Ground-truth images. (d) Baseline. (e) Ours.

3) Experimental Results on WHU-CD: Table III presents the quantitative comparison results on the WHU-CD dataset in comparison with other methods. LBCDMamba demonstrates superior performance across all the metrics. Notably, LBCDMamba achieves 94.13% in F1-score and 88.94% in IoU. These enhancements result from the model's ability to selectively scan and integrate spatial and temporal data from VHR imagery, ensuring precise segmentation.

Fig. 8 presents some qualitative comparisons on the WHU-CD dataset.our method demonstrates superior change detection performance on the WHU-CD dataset compared to the Baseline method. It achieves more accurate localization of changed regions, with smoother and more precise boundary delineation, significantly reducing false alarms and missed detections. In particular, our method shows strong capability in recovering fine-grained changes, especially in complex building structures. This improvement is primarily attributed to our proposed method's effective integration of local and global information, which enhances spatial modeling and temporal feature fusion, leading to a more robust and detailed change representation.

D. Complexity Analysis

As shown in Table IV, we compare the parameter size and computational cost (FLOPs) of our proposed LBCDMamba with several representative change detection models. The proposed LBCDMamba contains only 28.57 M parameters and requires 16.39 GFLOPs, which is significantly more efficient than recent Transformer-based and Mamba-based architectures. For example, compared to ChangeFormer, which has 33.61 M parameters and 213.13 GFLOPs, LBCDMamba reduces the FLOPs by over 90% while also decreasing the parameter count by approximately 15% . Likewise, when compared to ChangeMamba, which contains 49.94 M parameters and 28.70 GFLOPs, LBCDMamba reduces the computational cost by about 43%, demonstrating superior efficiency. Despite having a significantly lower computational burden, our model achieves competitive or even superior performance in accuracy metrics, indicating a better tradeoff between model complexity and detection accuracy.

TABLE IV
COMPARISON RESULTS OF COMPUTATIONAL EFFICIENCY ACROSS DIFFERENT MODELS

Method	Params(M)	FLOPs(G)
FC-Siam-Conc [28]	1.55	2.99
HANet [51]	2.61	17.67
SNUNet [29]	12.04	54.83
HCGMNet [52]	47.32	318.42
ChangeFormer [13]	33.61	213.13
BIT [12]	9.02	23.06
ChangeMamba [18]	49.94	28.70
RSM-CD [44]	49.97	17.43
LBCDMamba(ours)	28.57	16.39

E. Ablation Studies

1) Effectiveness of Different Scan Structures: Table V shows the change detection performance with different mamba scanning strategies on the LEVIR-CD+ dataset, such as CSM [37], OSSM [44], Local Scan [25], Zigzag Scan [27], Efficient Scan [26], as well as our proposed LGSSM. The experimental results demonstrate that the proposed scanning strategy outperforms all compared methods, achieving an F1-score of 83.86% and an IoU of 72.21% . Notably, compared with other scanning mechanisms, our LGSSM improves the F1-score by 3.40%, 2.32%, 1.66%, 0.83%, and 0.56%, respectively. As illustrated in Fig. 1, Zigzag Scan [27], constrained by computational resources, adopts a single-directional strategy but performs poorly in the complex and diverse scenes encountered in change detection tasks. Efficient Scan [26] incorporates dilated convolutions and jump sampling to reduce the computational cost, making it suitable for large-scale image processing in resource-constrained environments. However, this approach is less robust in detecting fine-grained changes. CSM [37] enhances global feature integration through four-directional scanning but lacks the capacity for detailed local feature representation, leading to missed and false detections. OSSM [44] captures large-scale spatial features via bidirectional selective scanning, improving global modeling capacity. Nevertheless, it still struggles with local detail modeling and poses challenges for practical deployment due to its high computational overhead. Local Scan [25] preserves local dependencies via small-window scanning, yet its lack of global context modeling results in limited performance. In contrast, our proposed LGSSM integrates global and local

TABLE V
COMPARATIVE EXPERIMENTAL RESULTS OF THE DIFFERENTIAL SCANNING STRUCTURES

Scan Method	Rec.	Pre.	OA	F1	IoU	KC
Zigzag Scan	79.88	80.64	98.43	80.26	67.23	79.25
Efficient Scan	81.20	81.88	98.50	81.54	68.83	80.76
CSM	81.08	83.36	98.57	82.20	69.79	81.46
OSSM	81.64	84.47	98.64	83.03	70.98	82.32
Local Scan	81.88	84.77	98.66	83.30	71.38	82.60
LGSSM	81.95	85.87	98.72	83.86	72.21	83.20

TABLE VI
EFFECTIVENESS OF EACH COMPONENT OF LBCDMAMBA ON LEVIR-CD+ DATASET

Baseline	LGSSM	BMPA	BTFF	F1	IoU
✓				82.20	69.79
✓	✓			83.86	72.21
✓	✓	✓		84.27	72.89
✓	✓		✓	84.51	73.17
✓	✓	✓	✓	**84.85**	**73.43**

"✓" means appending. (%)

scanning strategies to achieve more efficient information aggregation and feature modeling. This design not only enhances the model's ability to preserve global contextual information while capturing local details, but also demonstrates strong adaptability in handling complex structures and varying change scales in remote sensing scenarios, resulting in an F1-score improvement of 1.66% and an IoU gain of 2.42% over traditional scanning methods.

2) Effectiveness of Different Components in LBCDMamba: To gain a deeper understanding of our method, we conducted ablation experiments on the LEVIR-CD+ dataset to evaluate the impact of each component on model performance. The experimental results, as shown in Table VI, indicate that whether adding components individually or in pairs, all configurations outperform the baseline model.

In the ablation study, the introduction of LGSSM into the baseline significantly improved model performance, indicating that it enhances the ability to capture local information from a global perspective, thereby optimizing the integration of global and local information. Building upon this, the addition of the MBPA module further improved the F1 score and IoU by 0.41% and 0.68%, respectively. This demonstrates that MBPA's multibranch structure, designed to extract both global context and fine-grained spatial features, is beneficial for detecting complex change patterns. Furthermore, the standalone incorporation of the BTFF module led to additional performance gains, with F1 and IoU increasing by 0.65% and 0.96%, respectively, highlighting the advantages of pixelwise fusion of bitemporal features. When MBPA and BTFF were combined, the model achieved a notable improvement over the baseline, with increases of 2.65% in F1 score and 3.64% in IoU. This suggests that the two modules provide complementary strengths—MBPA enhances feature representation while BTFF strengthens temporal interaction and suppresses irrelevant content—an effect more clearly reflected in the quantitative and qualitative analyses in Section IV. Overall, the integration of each key component provides a powerful mechanism for robust change detection, further validating the effectiveness of the proposed LBCDMamba framework.

V. DISCUSSION

The proposed LBCDMamba method demonstrates superior detection performance in the bitemporal remote sensing change detection task and exhibits good generalization across diverse land cover scenes. However, there are still some limitations. As shown in Figs. 6–8, our method effectively avoids boundary blurring and fragmentation of the change regions, significantly reducing false negatives and false positives. Nevertheless, the model's performance declines in areas where the transition between change and nonchange regions occurs, and in regions where it is difficult to distinguish color differences between the bitemporal images. To address these issues, future research could focus on improving edge detection techniques, as well as investigating the impact of factors such as color differences and lighting changes on model performance to enhance its effectiveness. Furthermore, we suggest exploring how to combine supervised and unsupervised learning methods to reduce dependency on large labeled datasets, improving model performance and generalization. Furthermore, research should focus on utilizing deeper Mamba models to fully exploit the potential of SSM and explore more suitable Mamba architectures for bitemporal remote sensing image change detection tasks. Given its robust performance and strong generalization across diverse scenarios, the proposed method also shows potential for practical deployment in real-world applications such as urban development monitoring, postdisaster damage assessment, agricultural change analysis, and infrastructure management, where timely and accurate change information is essential for decision making.

VI. CONCLUSION

In this article, we propose the LBCDMamba model for dual-temporal remote sensing change detection tasks to address the limitations of CNN-based methods in global modeling capability and the secondary computational complexity issues of Transformer-based approaches. Specifically, we compare the performance of five mainstream scanning mechanisms: CSM, Local Scan, OSSM, Zigzag Scan, and efficient 2-D scanning. The results show that the combination of global scanning and the Local Scan mechanism significantly improves the performance of CD tasks on ultrahigh-resolution remote sensing images. The proposed LBCDMamba network effectively integrates global

and local information by combining global scanning and local scanning. This approach captures local dependencies in tiled data from a global perspective, mitigating the issue of losing fine-grained details. In addition, the MBPA module is employed, utilizing a multibranch feature extraction mechanism to parallelly capture both global and local information, and an attention mechanism is applied to adaptively enhance the extracted features. This enables the model to focus on high-resolution features while balancing the interaction of features across various sizes. Moreover, considering the simple interaction capabilities of current dual-temporal images, we introduce the BTFF module to account for the spatial and scale variations of different objects, ensuring accurate global modeling of bitemporal features. Ablation studies were conducted to validate the importance of each key component. Extensive experiments demonstrate that the proposed method achieves superior performance on both the LEVIR-CD and LEVIR-CD+ datasets. These findings not only advance the application of the Mamba structure in remote sensing but also pave the way for further research into enhancing change detection capabilities.

REFERENCES

[1] G. Cheng et al., "Change detection methods for remote sensing in the last decade: A comprehensive review," *Remote Sens.*, vol. 16, no. 13, 2024, Art. no. 2355.

[2] F. A. Al-Wassai and N. Kalyankar, "Major limitations of satellite images," 2013, *arXiv:1307.2434*.

[3] B. Rasti, Y. Chang, E. Dalsasso, L. Denis, and P. Ghamisi, "Image restoration for remote sensing: Overview and toolbox," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 2, pp. 201–230, Jun. 2022.

[4] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol. Cybern.*, vol. 36, no. 4, pp. 193–202, 1980.

[5] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[6] Y. Liu, C. Pang, Z. Zhan, X. Zhang, and X. Yang, "Building change detection for remote sensing images using a dual-task constrained deep siamese convolutional network model," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 5, pp. 811–815, May 2021.

[7] Z. Zhang, G. Vosselman, M. Gerke, D. Tuia, and M. Y. Yang, "Change detection between multimodal remote sensing data using siamese CNN," 2018, *arXiv:1807.09562*.

[8] W. Yan, P. Yan, and L. Cao, "Unsupervised convolutional neural network fusion approach for change detection in remote sensing images," 2023, *arXiv:2311.03679*.

[9] C. Han, C. Wu, M. Hu, J. Li, and H. Chen, "C2F-semiCD: A coarse-to-fine semi-supervised change detection method based on consistency regularization in high-resolution remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4702621.

[10] R. Yadav, A. Nascetti, and Y. Ban, "Context-aware change detection with semi-supervised learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2023, pp. 5754–5757.

[11] A. Vaswani et al., "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 5998–6008.

[12] H. Chen, Z. Qi, and Z. Shi, "Remote sensing image change detection with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5607514.

[13] W. G. C. Bandara and V. M. Patel, "A transformer-based siamese network for change detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2022, pp. 207–210.

[14] W. Li, L. Xue, X. Wang, and G. Li, "MCTNet: A multi-scale CNN-transformer network for change detection in optical remote sensing images," in *Proc. 26th Int. Conf. Inf. Fusion*, 2023, pp. 1–5.

[15] J. Yang, H. Wan, and Z. Shang, "Enhanced hybrid CNN and transformer network for remote sensing image change detection," *Sci. Rep.*, vol. 15, no. 1, 2025, Art. no. 10161.

[16] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," 2023, *arXiv:2312.00752*.

[17] H. Zhang et al., "CDMamba: Remote sensing image change detection with mamba," 2024, *arXiv:2406.04207*.

[18] H. Chen, J. Song, C. Han, J. Xia, and N. Yokoya, "ChangeMamba: Remote sensing change detection with spatiotemporal state space model," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4409720.

[19] J. N. Paranjape, C. De Melo, and V. M. Patel, "A mamba-based siamese network for remote sensing change detection," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2025, pp. 1186–1196.

[20] X. Ma et al., "STNet: Spatial and temporal feature fusion network for change detection in remote sensing images," in *Proc. IEEE Int. Conf. Multimedia Expo.*, 2023, pp. 2195–2200.

[21] M. Noman, M. Fiaz, H. Cholakkal, S. Khan, and F. S. Khan, "ELGC-Net: Efficient local–global context aggregation for remote sensing change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4701611.

[22] Y. Xiao, B. Luo, J. Liu, X. Su, and W. Wang, "Bi-temporal gaussian feature dependency guided change detection in remote sensing images," 2024, *arXiv:2410.09539*.

[23] Z. Wu, X. Ma, R. Lian, K. Zheng, and W. Zhang, "CDxLSTM: Boosting remote sensing change detection with extended long short-term memory," *IEEE Geosci. Remote Sens. Lett.*, vol. 22, 2025, Art. no. 3002005.

[24] Z. Huang, Z. Fu, J. Song, G. Yuan, and J. Li, "MFDS-Net: Multi-scale feature depth-supervised network for remote sensing change detection with global semantic and detail information," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, 2024, Art no. 7507905.

[25] T. Huang et al., "Localmamba: Visual state space model with windowed selective scan," in *Proc. Eur. Conf. Comput. Vis.*, Cham, 2025, pp. 12–22.

[26] X. Pei et al., "Efficientvmamba: Atrous selective scan for light weight visual mamba," in *Proc. AAAI Conf. Artif. Intell.*, 2025, vol. 39, no. 6, pp. 6443–6451.

[27] V. T. Hu et al., "Zigma: A dit-style zigzag mamba diffusion model," in *Proc. Eur. Conf. Comput. Vis.*, Cham, 2024, pp. 148–166.

[28] R. C. Daudt, B. Le Saux, and A. Boulch, "Fully convolutional Siamese networks for change detection," in *Proc. IEEE 25th Int. Conf. Image Process.*, 2018, pp. 4063–4067.

[29] S. Fang, K. Li, J. Shao, and Z. Li, "SNUNet-CD: A densely connected siamese network for change detection of VHR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 8007805.

[30] Y. Ye, M. Wang, L. Zhou, G. Lei, J. Fan, and Y. Qin, "Adjacent-level feature cross-fusion with 3-D CNN for remote sensing image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5618214.

[31] H. Zhang, Z. Lin, F. Gao, J. Dong, Q. Du, and H.-C. Li, "Convolution and attention mixer for synthetic aperture radar image change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 4012105.

[32] K. Lu and X. Huang, "RCDT: Relational remote sensing change detection with transformer," 2022, *arXiv:2212.04869*.

[33] Q. Xu, Y. Shi, J. Guo, C. Ouyang, and X. X. Zhu, "UCDFormer: Unsupervised change detection using a transformer-driven image translation," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5619917.

[34] B. Jiang et al., "VcT: Visual change transformer for remote sensing image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 2005214.

[35] D. Zhu, X. Huang, H. Huang, Z. Shao, and Q. Cheng, "ChangeViT: Unleashing plain vision transformers for change detection," 2024, *arXiv:2406.12847*.

[36] D. Wang, L. Jiao, J. Chen, S. Yang, and F. Liu, "Changes-aware transformer: Learning generalized changes representation," 2023, *arXiv:2309.13619*.

[37] Y. Liu et al., "Vmamba: Visual state space model," *Adv. Neural Inf. Process. Syst.*, vol. 37, pp. 103031–103063, 2024.

[38] X. Ma, Z. Lv, C. Ma, T. Zhang, Y. Xin, and K. Zhan, "BS-Mamba for black-soil area detection on the Qinghai-Tibetan plateau," *J. Appl. Remote Sens.*, vol. 19, no. 2, pp. 28502–28502, 2025.

[39] J. Xie, R. Liao, Z. Zhang, S. Yi, Y. Zhu, and G. Luo, "ProMamba: Prompt-Mamba for polyp segmentation," 2024, *arXiv:2403.13660*.

[40] T. Zhang et al., "Point cloud mamba: Point cloud learning via state space model," in *Proc. AAAI Conf. Artif. Intell.*, 2025, vol. 39, no. 10, pp. 10121–10130.

[41] E. Ghazaei and E. Aptoula, "Change state space models for remote sensing change detection," 2025, *arXiv:2504.11080*.

[42] Z. Wu et al., "CD-Lamba: Boosting remote sensing change detection via a cross-temporal locally adaptive state space model," 2025, *arXiv:2501.15455*.
[43] J. Kaung and H. Ge, "2DMCG: 2DMambawith change flow guidance for change detection in remote sensing," 2025, *arXiv:2503.00521*.
[44] S. Zhao, H. Chen, X. Zhang, P. Xiao, L. Bai, and W. Ouyang, "RS-Mamba for large remote sensing image dense prediction," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5633314.
[45] A. Gupta, A. Gu, and J. Berant, "Diagonal state spaces are as effective as structured state spaces," *Adv. Neural Inf. Process. Syst.*, vol. 35, pp. 22982–22994, 2022.
[46] H. Chen, J. Song, C. Wu, B. Du, and N. Yokoya, "Exchange means change: An unsupervised single-temporal change detection framework based on intra- and inter-image patch exchange," *ISPRS J. Photogrammetry Remote Sens.*, vol. 206, pp. 87–105, 2023.
[47] M. Berman, A. R. Triki, and M. B. Blaschko, "The Lovasz-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4413–4421.
[48] H. Chen and Z. Shi, "A spatial-temporal attention-based method and a new dataset for remote sensing image change detection," *Remote Sens.*, vol. 12, no. 10, 2020, Art. no. 1662.
[49] S. Ji, S. Wei, and M. Lu, "Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 574–586, Jan. 2019.
[50] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2017, *arXiv:1711.05101*.
[51] C. Han, C. Wu, H. Guo, M. Hu, and H. Chen, "HANet: A hierarchical attention network for change detection with bitemporal very-high-resolution remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 3867–3878, 2023.
[52] C. Han, C. Wu, and B. Du, "HCGMNet: A hierarchical change guiding map network for change detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2023, pp. 5511–5514.

Lukun Wang (Senior Member, IEEE) received the Ph.D. degree in computer application technology from the Ocean University of China, Qingdao, China, in 2016.

He is currently a Master Supervisor and an Associate Professor with the School of Intelligent Equipment, Shandong University of Science and Technology, Taian, China. He is also the Assistant to the Dean of the School of Intelligent Equipment, Shandong University of Science and Technology, Taian, China, where he is also the Director of the Three Innovation Center. He is also the Head of the Computer Vision and Pattern Recognition Team, Shandong University of Science and Technology. He has many years of experience in scientific research and engineering technology, innovation, and entrepreneurship. His research interests include artificial intelligence, Big Data, the Internet of Things, machine learning, and information security.

Dr. Wang is an especially invited reviewer of the IEEE.

Qihang Sun received the B.S. degree in information management and information systems in 2021 from the Shandong University of Science and Technology, Jinan, China. He is currently pursuing the M.S. degree in network and information security with Shandong University of Science and Technology, Tai'an, China.

His research interests include computer vision and remote sensing.

Jiaming Pei (Graduate Student Member, IEEE) is currently working toward the Ph.D. degree in computer science with the University of Sydney, Sydney, NSW, Australia.

From 2021 to 2022, he visited the Southwestern University of Finance and Economics, Chengdu, China. He has authored or coauthored and worked on some papers in the refereed journals and conferences, such as International Conference on Learning Representations, IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, IEEE TRANSACTIONS ON CONSUMER ELECTRONICS, and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. His research interests include the application of data mining and federated learning.

Muhammad Attique Khan (Member, IEEE) received the master's and Ph.D. degrees in human activity recognition for application of video surveillance and skin lesion classification using deep learning from COMSATS University Islamabad, Islamabad, Pakistan, in 2018 and 2022, respectively.

He is currently an Assistant Professor with the Artificial Intelligence Department, Prince Mohammad bin Fahd University, Al Khobar, Saudi Arabia. He has more than 350 publications that have more than 18000 citations and an impact factor of more than 1250 with h-index 76 and i-index 250. His main research interests include medical imaging, Internet of Things, security with artificial intelligence, magnetic resonance imaging analysis, video surveillance, human gait recognition, and agriculture plants using deep learning.

Dr. Khan is the reviewer of several reputed journals, such as IEEE TRANSACTION ON INDUSTRIAL INFORMATICS, IEEE TRANSACTION OF NEURAL NETWORKS, *Pattern Recognition Letters*, *Multimedia Tools and Application*, *Computers and Electronics in Agriculture*, *IET Image Processing, Biomedical Signal Processing Control*, *IET Computer Vision*, *EURASIP Journal of Image and Video Processing*, IEEE ACCESS, *MDPI Sensors*, *MDPI Electronics*, *MDPI Applied Sciences*, *MDPI Diagnostics*, and *MDPI Cancers*.

Maryam M. Al Dabel received the Ph.D. degree in computer science from the University of Sheffield, Sheffield, U.K., in 2016.

She is currently an Assistant Professor of artificial intelligence and machine learning with the College of Computer Science and Engineering, University of Hafr Al Batin, Hafar Al-Batin, Saudi Arabia, where she is also a Vice Dean with Computer Science and Engineering College.

Yasser D. Al-Otaibi received the Ph.D. degree in information systems from Griffith University, Brisbane, Australia, in 2018.

He is currently an Assistant Professor with the Department of Information Systems, Faculty of Computing and Information Technology in Rabigh, King Abdulaziz University, Jeddah, Saudi Arabia. His current research interests include information technology adoption and acceptance, wireless sensor networks, and Internet of Things.

Ali Kashif Bashir (Senior Member, IEEE) received the Ph.D. degree in computer science and information technology from Korea University, Seoul, South Korea, in 2012.

He is a Chair Professor of computer networks and cybersecurity with Manchester Metropolitan University, Manchester, U.K. At Manchester Met, he leads the SISTEMS: Secure and Intelligence Systems Research Group, the Future Networks Lab, the Turing Network's AI Safety and Security Taskforce, and the cybersecurity pathway's line management. He collaborated with NTT, Japan; KEPCO, South Korea; ITER, South Korea; SK Telecom, etc., on world-leading initiatives, and has authored and coauthored more than 450 technology-leading articles resolving several interdisciplinary research problems, receiving more than 15 K citations.

Dr. Bashir has managed £10 million in strategic funds, obtained more than 4 million in additional funding from several government bodies, delivered more than 100 keynote speeches, and chaired more than 50 conferences and workshops. He is a Member of several technical societies and a Distinguished Speaker of ACM. He is listed as an IEEE Featured Author in 2021. He was the recipient of the Clarivate Highly Cited Researcher Award in 2023 and 2024 and the Excellent Editor Award from IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING in 2024. He is the Editor-in-Chief (EIC) of IEEE Technology, Policy and Ethics, and a Senior Editor for IEEE TRANSACTIONS ON CONSUMER ELECTRONICS (TCE), Deputy EIC of IEEE TCE Letters, and an Associate Editor for several reputed journals. He is also an Ethics and Plagiarism Investigation Committee Member of a few IEEE Journals.