AEGL-Net: Adaptive Multiscale Global-Local Feature Fusion Network for Remote Sensing Change Detection

Zilu Ying*, Yijun Zhou*, Yikui Zhai*, *Senior Member, IEEE*, Hufei Zhu*, Hongsheng Zhang, *Senior Member, IEEE*, Pasquale Coscia, *Senior Member, IEEE*, Angelo Genovese, *Senior Member, IEEE*, Fabio Scotti, *Senior Member, IEEE*, Vincenzo Piuri, *Fellow, IEEE*, C. L. Philip Chen, *Fellow, IEEE*.

***Abstract*—With the rapid advancements in deep learning technology, the field of remote sensing change detection (RSCD) has witnessed significant improvements and innovations. In this context, bitemporal image processing, using features directly extracted by the backbone for subsequent fusion operations, may be obstructed by external environmental factors, potentially limiting the effective capture of complex feature variations. Moreover, overlooking local features during the fusion of bitemporal features can significantly affect the final detection results. As a result, achieving accurate change detection (CD) still encounters various challenges. To tackle these issues, this paper proposes a CD network (AEGL-Net) with Adaptive Multiscale Enhancement (AME) and Global-Local Feature Fusion (GLFF) modules. First, AME enhances features at each stage of backbone extraction through an adaptive strategy, balancing the enhancement of semantic information and texture details. Then, GLFF is used to fuse the bitemporal image features, which enhances the modeling of global dependencies while also fusing shared and context-aware weights to enhance the local features. Finally, the merged features are fed into the decoder to generate precise change maps. Experiments conducted with four open RSCD datasets (LEVIR-CD, S2Looking, SYSU-CD, and UAV-CD) demonstrate that our proposed AEGL-Net outperforms ten state-of-the-art models in the RSCD field. Our code is available at https://github.com/yikuizhai/AEGL-Net.**

***Index Terms*—Remote sensing change detection (RSCD), Adaptive Multiscale Enhancement (AME), Global-Local Feature Fusion (GLFF).**

This study was funded by Guangdong Basic and Applied Basic Research Foundation (No. 2021A1515011576), Guangdong Higher Education Innovation and Strengthening School Project (No. 2020ZDZX3031, No. 2022ZDZX1032, No. 2023ZDZX1029), Wuyi University Hong Kong and Macao Joint Research and Development Fund (No. 2022WGALH19), Guangdong Jiangmen Science and Technology Research Project (No. 2220002000246, No. 2023760300070008390).(Corresponding author: Yikui Zhai and Hongsheng Zhang)

Zilu Ying,Yijun Zhou,Yikui Zhai and Hufei Zhu are with the College of Electronics and Information Engineering, Wuyi University, Jiangmen, 529020, China(e-mail: ziluy@163.com, 17346700814@163.com, yikuizhai@163.com, zhuhufei@aliyun.com).

Hongsheng Zhang is with the Department of Geography, The University of Hong Kong, Hong Kong, China (e-mail: zhanghs@hku.hk).

Pasquale Coscia, Angelo Genovese, Vincenzo Piuri and Fabio Scotti are with the Department of Computer Science, Universita Degli Studi di Milano, 20133, Milan, Italy (e-mail: pasquale.coscia@unimi.it; angelo.genovese@unimi.it; vincenzo.piuri@unimi.it; fabio.scotti@unimi.it).

C. L. Philip Chen is with the Faculty of Computer Science and Engineering, South China University of Technology, Guangzhou, Guangdong 510006, China (e-mail: philip.chen@ieee.org).

*Contributed to this work equally.

Manuscript received April 19, 2021; revised August 16, 2021.

I. Introduction

REMOTE sensing change detection (RSCD) corresponds to identifying variations in changes between remotely sensed images of the coherent territory across different historical periods [1]. Currently, RSCD methodologies have found extensive application in land change science [2], disaster evaluation [3], urban planning [4], cropland observation [5], and ecological environment monitoring [6].

Current RSCD techniques can be categorized into conventional methods and deep learning-based approaches. Conventional RSCD methods can be further classified into three main types: transform-based approaches, algebra-based approaches, and classification-based approaches [7], [8], [9], [10]. Transformation-based methods commonly use techniques such as change vector analysis [11] and principal component analysis [12] to map image spectral combinations onto a specific feature space aimed at recognizing altered pixels. Algebra-based methods use techniques such as regression analysis [13] and difference normalization to create image variation maps, and then impose thresholds to categorize pixels into either modified or unmodified groups. Classification-based approaches commonly use strategies such as K-means clustering and support vector machines to classify changed and unchanged regions. These approaches depend on the quality of the remotely sensed imagery (such as weather changes, seasonal variations, and fluctuations in light intensity) and lack robustness.

In recent years, deep learning has demonstrated exceptional performance in the field of remote sensing, thanks to its powerful capabilities in information mining and feature extraction. As the diversity and complexity of remote sensing data continue to increase, traditional methods have increasingly shown limitations in handling high-dimensional data and addressing environmental changes. Deep learning, through automated feature learning, can extract high-level information from large volumes of remote sensing images without manual intervention. This makes it superior to traditional techniques in tasks such as object detection, data classification, and hyperspectral image super-resolution (HS-SR). For instance, UIU-Net [14], by embedding a smaller U-Net within a larger U-Net architecture, effectively enhances the detection of small objects in infrared images, yielding significant improvements.

Additionally, CCR-Net [15] achieves compact fusion representations of multimodal remote sensing data through a cross-modal reconstruction strategy, facilitating effective information exchange and thereby improving classification performance. DC-Net [16], by progressively fusing hyperspectral and multispectral information at the pixel, subpixel, image, and feature levels, significantly enhances HS-SR performance. In change detection (CD), early studies primarily used convolutional neural networks (CNNs) to construct foundational networks, which were further enhanced by advanced techniques to improve performance, such as the Siamese structure [17], multiscale mechanism [18], [19], and attentional mechanism [20], which contributed to the success of these RSCD networks. However, researchers have discovered that while CNN-based RSCD networks excel in end-to-end feature extraction, CNNs primarily sense and extract features in images through local convolutional operations, restricting their capacity to capture distant contextual information. The Transformer has gained recognition for its outstanding performance in Natural Language Processing (NLP) tasks. Vision Transformer (VIT) [21] was the first to apply the Transformer model to image classification tasks. Unlike previous work on CNNs, VIT divides an image into fixed-size blocks and treats the pixels of each block as positional encodings in a sequence. This approach enables the direct utilization of Transformer's self-attention mechanism to model the links between blocks, thereby efficiently capturing contextual information.

Although existing CNN- and Transformer-based CD models demonstrate strong performance in RSCD tasks, there remains room for improvement. First, bitemporal RS imagery is susceptible to ambient conditions such as weather changes, cyclical variations, and fluctuations in light intensity. As a result, most CD models, which rely solely on features extracted by the backbone for subsequent fusion, may fail to capture complex feature variations, leading to inaccuracies in representing subtle changes and potential loss of information. Second, while Transformer-based CD models effectively model global features, they tend to overlook certain local features of bitemporal RS imagery, which complicates the recognition of small objects or local changes, thereby affecting the model's effectiveness.

In response to the aforementioned issues, we explore our research motivation in detail from two perspectives. First, the acquisition of dual-temporal remote sensing imagery is often influenced by environmental factors, which can introduce noise unrelated to actual changes, thereby increasing the complexity of change detection tasks. Existing change detection models typically rely solely on the features output by the backbone network for fusion, which struggles to effectively capture complex feature variations across multiple scales. This limitation is particularly evident when detecting subtle changes, where there is a risk of distortion or information loss. In addition, traditional multi-scale methods typically adopt static strategies, making it difficult to adapt to multi-scale interference from noise and lighting in remote sensing imagery, which leads to issues such as loss of details or residual noise. To address this, we propose an adaptive multi-scale feature enhancement module, which dynamically evaluates the contribution weights of multi-scale features and adaptively enhances them. At the same time, it suppresses irrelevant features caused by external environmental disturbances (such as noise and lighting), thereby significantly improving detection robustness and accuracy in complex change scenarios. Second, existing RSCD methods often rely on simple feature fusion operations (such as single-stage concatenation or difference operations) or have limitations in the collaborative modeling of global and local features. Therefore, more attention should be given to enhancing the model's ability to fuse global and local features, in order to improve its adaptability and detection accuracy in complex change scenarios. Global information helps the model identify large-scale trends in changes, while local information is vital for detecting small objects or subtle changes. A lack of local information may result in missed detections of small objects or subtle changes, while the absence of global information can prevent the model from capturing spatial relationships between regions and overall change trends. To address this, we have designed a global and local feature fusion module. The global branch enhances the model's ability to perceive large-scale changes and improves its capacity to model global dependencies. The local branch, on the other hand, optimizes the detection of fine-grained changes by combining shared weights (By utilizing the same parameters to reduce model complexity, the model can effectively learn the features of dual-temporal imagery) with context-aware weights (Dynamic adjustment of weight distribution based on the contextual information of input features helps the model more accurately identify changes in different regions and objects) mechanisms. Through this fusion of global and local features, we effectively enhance the model's ability to perceive change regions, thereby improving the accuracy and robustness of change detection.

Building on the two motivations outlined above, we propose a CD network called AEGL-Net, which integrates a multi-scale and dual-branch attention mechanism. In AEGL-Net, we introduce the Adaptive Multi-Scale Enhancement (AME) module and the Global-Local Feature Fusion (GLFF) module. The key contributions of this work are as follows:

1) A innovative RSCD networks(AEGL-Net) is introduced, which incorporates an Adaptive Multiscale Enhancement Module (AME) and a Global-Local Feature Fusion (GLFF) Module to efficiently and accurately detect changed regions.
2) An Adaptive Multi-Scale Enhancement (AME) module is used to enhance features, which differs from previous methods that utilize multi-scale mechanisms by dynamically integrating multi-scale semantic and texture information extracted by the backbone into each scale.
3) A Global-Local Feature Fusion (GLFF) module is employed for the fusion of bitemporal features. The global branch boosts the model's capability of processing global information, thereby improving the network's modeling of global dependencies. Meanwhile, the local branch, by integrating shared weights and context-aware weights, enhances the representation of local features.
4) A detailed examination of the proposed method was carried out using four different RSCD datasets through

a sequence of rigorous experiments. The experimental outcomes indicated that the approache surpasses the current cutting-edge approaches in relation of both IOU and F1 metrics across all four datasets.

II. Relate Work

A. Traditional RSCD networks

Over the past few decades, numerous traditional RSCD networks have been developed, with transform, algebraic, and classification methods serving as prominent examples. Transform methods include change vector analysis (CVA) [11], principal component analysis (PCA) [12], and independent component analysis (ICA) [22]. Change Vector Analysis (CVA) technology identifies and quantifies change regions by comparing differences in corresponding pixel values between images taken at different times. Principal Component Analysis (PCA) obtains the principal change features from the bitemporal images, reducing dimensionality to minimize noise impact. Independent Component Analysis (ICA) is another linear transformation method that, unlike PCA, decorrelates and makes the output components statistically independent, thereby more effectively highlighting change regions. Algebraic methods include image disparity, image scaling, and image regression analysis [23], [24], it precisely locates subtle differences in bitemporal images through summation and subtraction operations. The summation and subtraction modules quantify channel changes and capture spatial differences, respectively, which are then analyzed in depth by the Transformer model. Finally, heterogeneous modulation blocks integrate and enhance the channel and spatial feature differences, such as k-means algorithm, and random forests, [25]. Traditional CD methods are straightforward but fail to deeply explore the features of bitemporal images. They rely heavily on the quality of RS images and lack robustness, making them unsuitable for handling CD tasks in complex scenarios.

B. Deep learning-based RSCD networks

Lately, deep learning-based RSCD has demonstrated strong feature extraction capabilities [26], [27], [28], [29]. RSCD networks, such as CNNs, have shown significant potential in the context of RSCD thanks to their powerful feature extraction capabilities [17]. For instance, Daudt et al. [17] proposed three networks based on the UNet [30] architecture: FC-EF, FC-conc, and FC-diff. The FC-EF network stitches bitemporal images in the channel dimension for incorporation as input. FC-conc processes images before and after changes separately by sharing parameters through a twin network and then splices them with the generated features at the decoding layer. FC-diff splices images with generated features at the FC-conc based on the disparity between the features of the before and after images, and then splices them with the generated features at the decoding layer. Li et al. [31] put forward a network based on a multi-scale framework, which inputs multi-level bitemporal features captured by CNN to the difference enhancement module, and then fuses the refined difference features before inputting them to the decoder to generate the change map. Lei et al. [19] proposed a multiscale decoupled convolution module that extracts multiscale features of change objects by means of cyclic multiscale convolution.

Due to the local convolutional operation of CNN, its receptive field is limited and cannot capture distant contextual details information. To tackle this problem, researchers have incorporated the Transformer model into the RSCD task to boost the model's ability to obtain remote contextual information. Chen et al. [32] were among the first to introduce an RSCD network based on the Transformer model (BIT), effectively capturing spatio-temporal context information between bitemporal images by combining the Transformer structure and semantic tokenization techniques. Zhao et al. [33] extracted features from bi-chronological images using single-stream and dual-stream encoders, achieving cross-phase fusion of the encoder-extracted features through the attention module, significantly improving the model's detection accuracy of the changed regions. Zhang et al. [34] utilized the Swin Transformer [35] to model the multiscale feature context information of the bitemporal phase image, with the decoder and Swin Transformer blocks recovering detailed information of the change area. Jiang et al. [36] proposed a Visual Change Transformer model that utilizes a graph neural network to model feature maps, optimizes the Top-K tokens, and integrates self-cross attention with the APA module to ultimately generate change maps. Wang et al. [37] introduced the CD Adder Network, which accurately identifies differences between dual-temporal remote sensing images through addition and subtraction operations, quantifies channel and spatial variations, and conducts in-depth analysis using a Transformer model. It is important to note that the effectiveness of deep learning-based RSCD methods heavily depends on feature representation, as the quality of this representation directly impacts detection performance. Therefore, enhancing feature representation is crucial for optimizing model performance. In this process, feature enhancement and fusion play critical roles. Feature enhancement aims to strengthen the representational capacity of features, thereby preventing overfitting and improving generalization. Meanwhile, feature fusion effectively integrates features from two temporal phases, fully extracting change-related information and enhancing the model's accuracy in detecting areas of change.

C. Feature enhancement

Feature enhancement in change detection improves the distinguishability and expressiveness of change features, effectively highlighting information in change regions while reducing noise interference. This enhancement ultimately increases the accuracy of model detection in these areas. Additionally, feature enhancement contributes to the robustness of the model and helps prevent overfitting. Li et al. [38] proposed a multi-scale differential feature enhancement network that generates multi-scale bi-temporal features through a shared weighted concatenated encoder. These features are then passed through a differential enhancement module to refine the difference features. Finally, the features are reconstructed into a change map via a decoder. The differential enhancement module

enhances information transfer between features at different scales using multi-layer encoders and transformer decoders. Lei et al. [39] designed a differential enhancement module that effectively learns the difference representations between the foreground and background, thereby reducing the interference from irrelevant changes in detection results. Song et al. [40] proposed a context and differential enhancement network, designing a content differential enhancement module to handle bi-temporal features at each layer of the encoder. By introducing a spatial attention mechanism, this module effectively enhances the focus on change regions, improves the expression of change difference features, and reduces irrelevant interference, thus increasing the robustness and accuracy of the model in complex environments. It is important to note that although these methods have yielded satisfactory results in change detection tasks, most focus on enhancing the difference features in bi-temporal images. While this approach helps capture change regions, it may overlook the feature information of each temporal phase. By solely enhancing the difference features, the detailed information of each temporal phase may not be fully utilized, which limits the model's ability to handle complex change scenarios.

D. Feature fusion

Feature fusion plays a crucial role in RSCD models. Its primary objective is to accurately capture differential information by merging the features of bitemporal remote sensing images, thereby enabling more effective identification of change regions. Le et al. [41] proposed a change detection network for remote sensing images that utilizes asymmetric convolutional residual blocks and multi-scale fusion. The encoder employs asymmetric convolutions to extract feature edges, demonstrating enhanced robustness to rotation, flipping, and aspect ratio distortions. The decoder further improves the detection accuracy of edge change regions by integrating multi-scale feature maps. Li et al. [42] designed a temporal interaction module, which distinguishes itself from previous methods that relied solely on single-stage concatenation or subtraction for feature fusion. This module fuses bi-temporal features by leveraging difference features and a series of arithmetic operations. Jiang et al. [43] proposed a novel joint learning framework that combines a fusion subnetwork and a differential subnetwork. The fusion subnetwork focuses on extracting and integrating bi-temporal features, while the differential subnetwork emphasizes processing the detailed differences in change regions. It is worth noting that, although these methods have achieved satisfactory results, most rely on simple feature fusion operations. When performing change detection in complex scenarios, this fusion approach remains insufficient, failing to capture subtle differences in changes. Therefore, greater attention should be given to the model's ability to fuse global and local features to enhance its adaptability and detection accuracy in complex change scenarios.

III. METHOD

In this part, we outline the overall structure of our put forward AEGL-Net in subsection A. To enhance the features of the bitemporal in subsection B, we introduce an AME module. For better integration of the features of the bitemporal, we present our proposed GLFF module in subsection C. Finally, we introduce the decoder and loss function of AEGL-Net in subsections D and E, respectively. In addition, we have conducted a theoretical analysis in subsection F to further enhance the theoretical depth of the manuscript.

A. AEGL-Net Overall Structure

This section describes the AEGL-Net structure, as shown in Fig. 1, which is based on the MobileNetV2 backbone, AME, GLFF, and decoder components.

We use a lightweight backbone, MobileNetV2, which is capable to efficiently extracting features from bitemporal RS images despite the fact that MobileNetV2 has fewer parameters compared to traditional deeper and more complex network structures [44]. This benefits from the deep separable convolution [45] and residual concatenation [44] in its structure, which are designed to effectively preserve important feature information.

Given the inputs of the bitemporal RS image $T1, T2 \in R^{3\times256\times256}$, the bitemporal features of the five-scale outputs are denoted as $\{Si\_T1, i = 1, 2, 3, 4, 5\}$ and $\{Si\_T2, i = 1, 2, 3, 4, 5\}$ (i represents the corresponding scale). To prevent the large amount of noise in the initial phase from interfering with the model, only the features of the last four phases are utilized. To further enhance the expressive power of these features, an AME is introduced to ameliorate and augment the features of the last four stages extracted from the backbone. Subsequently, the enhanced bi-temporal phase features are fed into the GLFF, where efficient feature fusion operations are performed. Finally, the fusion results of GLFF are input into the decoder, responsible for generating accurate change maps, thus enabling the precise identification and analysis of change regions.

B. Adaptive Multiscale Enhancement (AME)

In deep learning research, it is widely recognized that semantic information is essential for enhancing high-level features. This is crucial for object localization, as it aids in pinpointing an object's location and understanding its contextual relationships. On the other hand, the fine-grained details of object boundaries and textures predominantly make up the low-level features, providing a detailed description of the object's precise shape and surface properties [46], [47], [48]. From this perspective, much of the work on multiscale approaches has focused on employing a series of multiscale operations for the fusion of bitemporal features [49], [50], [51]. This approach overlooks the potential benefits of enhancing each stage of the feature acquisition flow from the backbone by incorporating both elementary and advanced features. This augmentation can enhance semantic information while preserving detailed information, making it more conducive to the subsequent fusion of bitemporal phase images. Multi-scale features are enhanced through a combination of convolution, normalization, and activation operations, progressively improving the expressive power of features at each scale. First, convolution is applied to

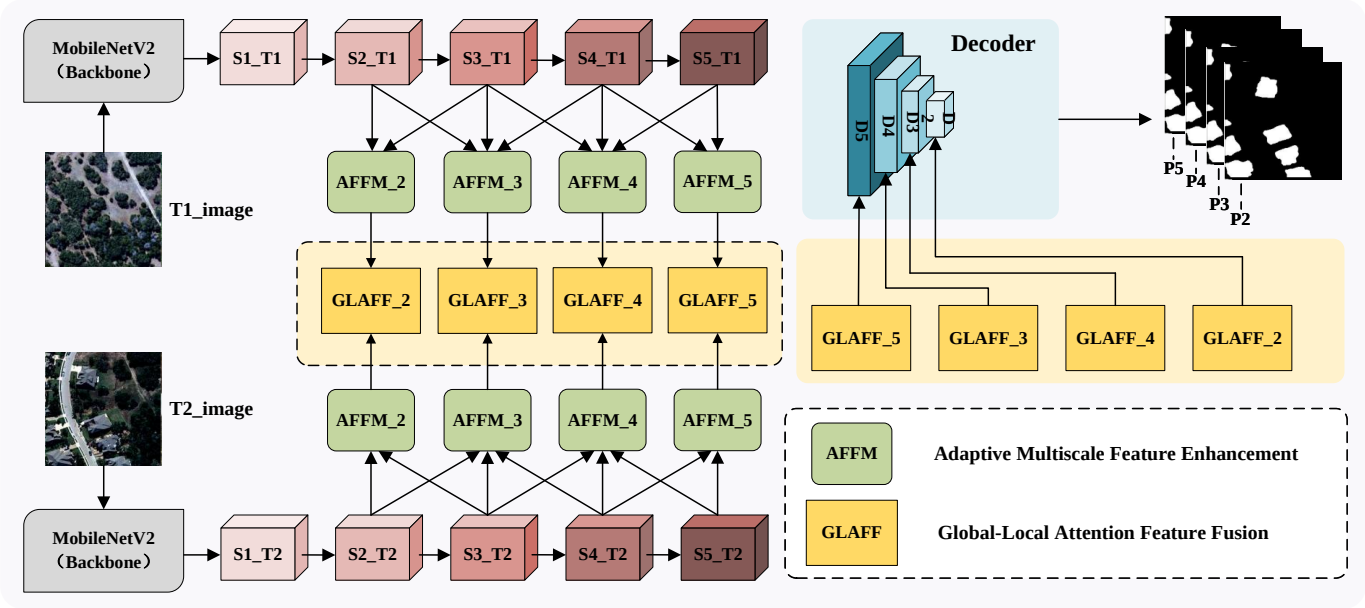


Fig. 1. AEGL-Net structure shows the generation of bitemporal images using the MobileNetV2 network to create multi-scale bitemporal features. The adjacent scales undergo feature enhancement through AME, followed by fusion of the enhanced bitemporal features using GLFF. The fused features from each scale are then fed into the decoder to produce the mask map. Subsequently, the change map from each scale undergoes upsampling and merging processes to generate the ultimate change map. Finally, upsampling and merging operations are conducted on each scale to produce the final change map.

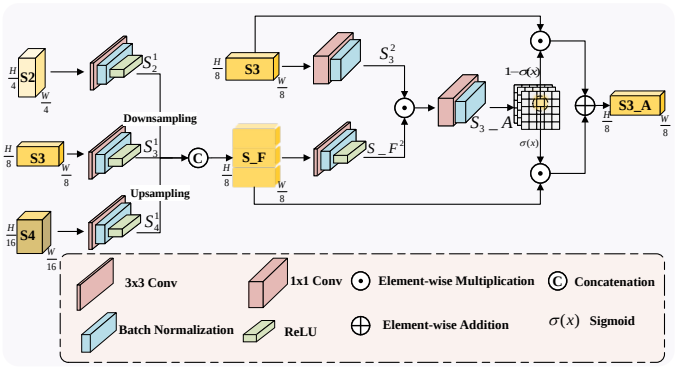


Fig. 2. AME Structure, the values of W and H are both 256.

extract features, followed by normalization to ensure consistency, and activation to enhance the model's representational capacity. Next, the features at the current scale are processed alongside the fused features through additional convolution, normalization, and activation operations to further strengthen the feature representation. Finally, the Sigmoid function is applied to evaluate the fused and current scale features. A value close to 1 indicates a significant contribution to the task (beneficial), while a value close to 0 indicates a smaller contribution (non-beneficial). During training, the model learns which features are most important for the change detection task through optimization techniques such as backpropagation, thereby achieving enhanced feature representation.

The structure of AME is displayed in Fig. 2. As an example, we use features extracted from the second, third, and fourth phases of the backbone. Feature maps $\{Si, i = 2, 3, 4\}$ are processed through the sequence of 3×3 convolution, batch normalization (BN), and ReLU layers to construct features S_2^1 and S_3^1. The 3×3 convolution effectively captures low-level features such as textures, angles, and edges at each stage of the image. The BN tier normalizes the output of the convolutional layer, accelerating convergence during network training and decreasing sensitivity to initialization. The ReLU activation function prevents gradient vanishing and enhances the model's ability to represent non-linearities. Then, upsampling is applied to S_4^1 and downsampling to S_2^1 to ensure their dimensions are consistent with S_3^1. Finally, feature $\{S_i^1, i = 2, 3, 4\}$ is concatenated with others to form the multi-scale feature S_F, which can be represented as:

$$S_2^1 = Down(ReLU(BN(Conv_{3\times3}(S_2)))) \tag{1}$$

$$S_3^1 = ReLU(BN(Conv_{3\times3}(S_3))) \tag{2}$$

$$S_4^1 = UP(ReLU(BN(Conv_{3\times3}(S_4)))) \tag{3}$$

$$S\_F = Cat(S_2^1, S_3^1, S_4^1) \tag{4}$$

where $Down(\cdot)$ stands the downsampling operation, $UP(\cdot)$ stands the upsampling operation, $Cat(\cdot)$ represents the concatenation operation, and S_F represents the aggregated multi-scale features.

After aggregating multi-scale features S_F, to achieve enhancement purposes distinct from directly adding S_F to $S3$. We first process S_F through a series of operations including a 3×3 convolution, BN tier, and ReLU to generate feature S_F^2. Meanwhile, $S3$ is processed through a 1×1 convolution and a BN tier to produce feature S_3^2. Subsequently, features S_F^2 and S_3^2 are element-wise multiplied, With the application of a 1×1 convolution and BN tier afterward to produce the enhanced feature S_3_A. Let the pixels from S_F and $S3$ in feature S_3_A be denoted as $\vec{V}_{S\_F}$ and $\vec{V}_{S3}$,

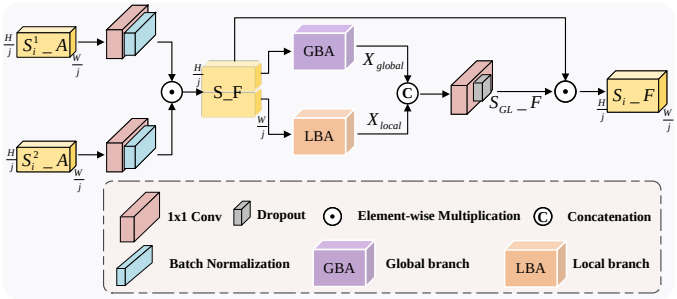


Fig. 3. GLFF Structure, $\{\frac{H}{j}, \frac{W}{j}, j = 4, 8, 16, 32\}$, the values of W and H are both 256, with the value of j corresponding to i. For example, when i = 2, j = 4, when i = 3, j = 8, and so on.

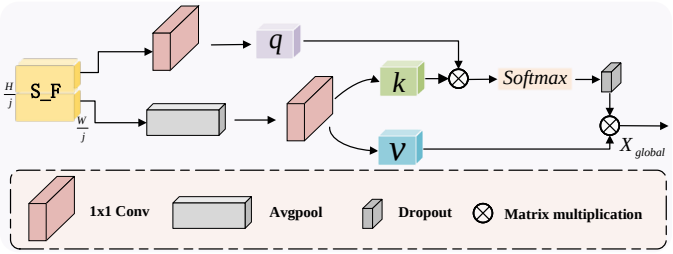


Fig. 4. GBA Structure, $\{\frac{H}{j}, \frac{W}{j}, j = 4, 8, 16, 32\}$, the values of W and H are both 256, with the value of j corresponding to i. For example, when i = 2, j = 4, when i = 3, j = 8, and so on.

respectively. The output of the Sigmoid function can then be expressed as:

$$S_3^2 = BN(Conv_{1\times1}(S3)) \tag{5}$$

$$S\_F^2 = UP(ReLU(BN(Conv_{3\times3}(S\_F)))) \tag{6}$$

$$S_3\_A = BN(Conv_{3\times3}(S\_F^2 \cdot S_3^2)) \tag{7}$$

$$\sigma = Sigmoid(S\_F(\vec{V}_{S\_F}) \cdot S3(\vec{V}_{S3})) \tag{8}$$

where σ represents the probability that two pixels in S_F and $S3$ belong to the same object.

If σ is high, it means that adding $\vec{V}_{S\_F}$ to $S3$ enhances it, and vice versa. Thus, $S3$ can adaptively filter and absorb beneficial features from the multi-scale features in S_F, while suppressing irrelevant or useless ones for enhancement. The output of he third scale AME can be expressed as:

$$S3\_A = \sigma \cdot \vec{V}_{S\_F} + (1 - \sigma) \cdot \vec{V}_{S3} \tag{9}$$

C. Global-Local Feature Fusion (GLFF)

The input for the RSCD task consists of bitemporal images. However, capturing these images can be influenced by factors unrelated to the RSCD task, including seasonal variations, changes in lighting conditions, and building renovations. Additionally, in cases of foreground-background category imbalance, the model tends to predict more background categories, which can result in various issues. On one hand, the model may misclassify some foreground samples as background, leading to underdetection. On the other hand, the model may also misclassify some background samples as foreground, resulting in misdetection. All these challenges impact the accuracy and reliability of the model in RSCD tasks. To tackle these issues, we have developed a GLFF, as illustrated in Fig. 3. The global branch enhances the model's ability to process global information and improves its capacity to model global dependencies, while the local branch effectively integrates shared weights and context-aware weighting mechanisms. By fusing dual-temporal features and combining global and local feature extraction modules, the model's capability for remote sensing change detection is significantly improved. The global branch captures large-scale change information, thereby enhancing the model's ability to represent global dependencies. In contrast, the local branch dynamically adjusts feature weights through shared weights and context-aware weighting mechanisms. Through operations such as feature fusion, convolution, and residual connections, the expressive power of the features and the stability of the model are further enhanced, resulting in increased precision in change detection.

Firstly, $\{S_i^1\_A, i = 2, 3, 4, 5\}$ and $\{S_i^2\_A, i = 2, 3, 4, 5\}$ (i represents the corresponding scale)are multiplied by 1×1 convolution and BN layer to enhance the feature expression ability and reduce the internal variable offset. Then, the S_F is obtained by multiplying the bitemporal features. The S_F is inputted to the GBA and the LBA to obtain low-frequency global features and high-frequency local features. Subsequently, the information is fused into the dual-branch by using Concat to further enhance the extracted feature information. Further enhancement is achieved by using 1×1 convolution and Dropout to get the feature S_{GL}_F. The extracted feature information is further improved by 1×1 convolution and Dropout. Leveraging the excellent performance of the residual structure [44] in deep learning, the residual structure is used to multiply the S_F with the fused two-branch features to enhance feature utilization. The Global-Local attention feature fusion flow can be expressed as follows:

$$S\_F = BN(Conv_{1\times1}(S_i^1\_A)) \cdot BN(Conv_{1\times1}(S_i^2\_A)) \tag{10}$$

$$S_{GL}\_F = Dropout(Conv_{1\times1}(Cat(X_{global}, X_{local}))) \tag{11}$$

$$S_i\_F = S_{GL}\_F \cdot S\_F \tag{12}$$

where X_{global} stands the output of the Global Branch Attention (GBA), X_{local} represents the output of the Local Branch Attention (LBA), $Dropout(\cdot)$ represents the regularization operation, and S_i_F represents the fused features.

The Global Branching Attention (GBA) first enhances spatial invariance through operations such as convolution and global pooling, and then utilizes an attention mechanism to highlight key region features. By incorporating Dropout, it prevents overfitting, thereby effectively capturing global information. Specifically, the GBA structure is shown in Fig. 4. First, S_F maps the features through 1×1 convolutional layers and reshapes them into query (Q). S_F is then downsampled in spatial dimensions through AvgPool, which enhances spatial invariance of the features. Next, it maps the features using 1×1 convolutional layers and reshapes the key (K) and value (V). Subsequently, matrix multiplication is used between them to

generate an attention map through a softmax layer. To prevent overfitting, Dropout is applied before the attention map and V matrix multiplication. Finally, the result X_{global} is output. The process of GBA output X_{global} can be represented as:

$$Q = Conv_{1\times1}(S\_F) \tag{13}$$

$$K = Conv_{1\times1}(Avgpool(S\_F)) \tag{14}$$

$$V = Conv_{1\times1}(Avgpool(S\_F)) \tag{15}$$

$$X_{global} = V \otimes Avgpool(Softmax(Q \otimes K^T)) \tag{16}$$

where $Avgpool(\cdot)$ represents the average pooling operation and $\otimes$ represents matrix multiplication.

The Local Branching Attention (LBA) module effectively aggregates local information through depthwise separable convolutions, thereby improving the ability to extract local features. At the same time, the introduction of Swish and Tanh activation functions further enhances the representation of features in local change regions, significantly improving the detection capability of change areas. Specifically, the structure of LBA is illustrated in Fig. 5. Global branching enhances the global receptive field, effectively capturing low-frequency global information, but lacks the ability to process high-frequency local information. In LBA, we first map the features using 1×1 convolutional layers. Then, we aggregate the local information using Depthwise (DW) convolution, where the DW convolution weights are shared. After this, we reshape the features into Q, K, and V matrices. The resulting sum is then subjected to Hadamard multiplication. In contrast to some previous methods for capturing local information [52], [53], [54], [55], the only nonlinear operator in the process of generating context-aware weights for local self-attention is Softmax. However, we incorporate Swish in addition to Tanh, and the stronger nonlinear function can produce higher quality context-aware weights. The process of LBA output X_{local} can be represented as:

$$Q = DWconv(Conv_{1\times1}(S\_F)) \tag{17}$$

$$K = DWconv(Conv_{1\times1}(S\_F)) \tag{18}$$

$$V = DWconv(Conv_{1\times1}(S\_F)) \tag{19}$$

$$X_{local} = V \cdot Dropout(Tanh(Swish(Q \cdot K))) \tag{20}$$

where $DWconv(\cdot)$ represents a 5×5 DW convolution, $Tanh(\cdot)$ and $Swish(\cdot)$ represents two different activation functions.

D. Decoder

The decoder's structure is presented in Fig. 6, inspired by the decoder in the paper [56]. The decoder consists of four stages $\{Di, i = 2, 3, 4, 5\}$(i stands for the stage number of the decoder). The output of each stage contains both global and local information of the current stage $\{d_{pi}, i = 2, 3, 4, 5\}$(i represents the features outcome at the ith phase of the decoder). Additionally, we have the change mask map $\{P_i, i = 2, 3, 4, 5\}$(i represents the variation mask map outcome at the ith stage of the decoder) . To fully exploit the deep and shallow feature information from various stages, we feed the output of

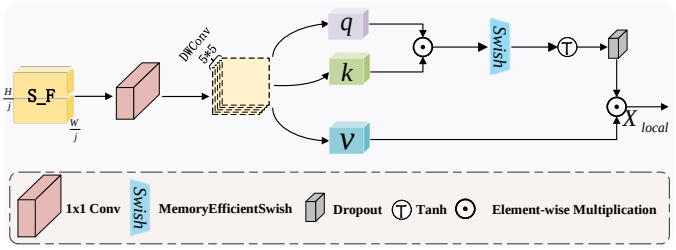


Fig. 5. LBA Structure, $\{\frac{H}{j}, \frac{W}{j}, j = 4, 8, 16, 32\}$, the values of W and H are both 256, with the value of j corresponding to i. For example, when i = 2, j = 4, when i = 3, j = 8, and so on.

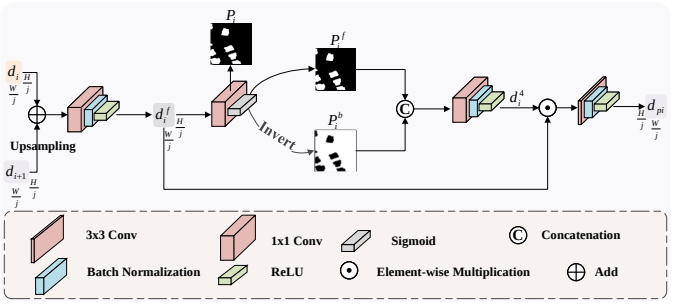


Fig. 6. Decoder Structure, $\{\frac{H}{j}, \frac{W}{j}, j = 4, 8, 16, 32\}$, the values of W and H are both 256, with the value of j corresponding to i. For example, when i = 2, j = 4, when i = 3, j = 8, and so on.

GLFF $\{S_i\_A, i = 2, 3, 4, 5\}$(i stands for the scale number) into the relevant stage of the decoder, such as S_5_A into D_5, and so forth. For instance, in the case of D_4, d_5 is up-sampled to match the resolution size of d_4. Then, d_5 and d_4 are combined and fused, resulting in the fused feature information d_4^f. This fused information is further enhanced through convolution, normalization, and activation operations. Subsequently, d_4^f is transformed into a variation mask map P_4.

In addition, we generate a change mask map P_4^f using the Sigmoid activation function for d_4^f after convolution. Simultaneously, we generate a reverse change mask map P_4^b through an inverse operation. The obtained P_4^f and P_4^b contain the context information of the current stage change object, after concatenating features P_4^f and P_4^b. We generate feature d_4^1 through convolution and other operations and then multiplying them with d_4^f through operations such as convolution to enhance feature utilization. Finally, we use 3×3 convolution and other operations to obtain more feature information in d_{p4}. The generation process of d_{p4} can be described as follows:

$$d_4^f = ReLU(BN(Conv_{1\times1}(d_4 + UP(d_5)))) \tag{21}$$

$$P_4 = Conv_{1\times1}(d_4^f) \tag{22}$$

$$P_4^f = Sigmoid(Conv_{1\times1}(d_4^f)) \tag{23}$$

$$P_4^b = Invert(Sigmoid(Conv_{1\times1}(d_4^f))) \tag{24}$$

$$d_i^4 = ReLU(BN(Conv_{1\times1}(Cat(P_4^f, P_4^b)))) \tag{25}$$

$$d_{p4} = ReLU(BN(Conv_{3\times3}(d_4^f \cdot d_i^4))) \tag{26}$$

where $Invert(\cdot)$ symbolizes the inversion operation, and $Cat(\cdot)$ symbolizes the concatenation operation.

E. Loss Function

Loss functions are vital in deep learning models, particularly for the RSCD task, which requires pixel-level prediction. To tackle the issue of class imbalance that can hinder the learning of changing object features, a hybrid loss function is utilized. This hybrid loss function combines the binary cross-entropy loss function L_{bce} with the dice loss function L_{dice} [57]. The total loss function L_{total} can be denotes as:

$$L_{bce}(f,r) = r \cdot \log f + (1-r) \cdot \log(1-f) \tag{27}$$

$$L_{dice}(f,r) = 1 - \frac{2 \cdot r \cdot f}{\|r\| + \|f\|} \tag{28}$$

$$L_{total}(f,r) = \sum_{i=2}^{5} L_{bce}(f_i,r) + L(f_i,r) \tag{29}$$

where f is the forecast change map and r is the real change map, $\|\cdot\|$ represents the ℓ_1 norm.

F. Theoretical Analysis

Threshold Effect of Sigmoid:

The Sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$ maps the input to the interval (0,1), forming a nonlinear "soft threshold" adjustment mechanism. The derivative of the Sigmoid function, $\sigma'(x) = \sigma(x)(1-\sigma(x))$, always lies within (0,0.25], providing a stable gradient for backpropagation. This ensures the stability of the optimization process for σ, further supporting the effectiveness of the threshold effect. Additionally, we conduct a theoretical analysis from the perspectives of convex optimization and the bias-variance trade-off in statistical learning.

1) Stability from the convex optimization perspective Define the objective function $L(\sigma)$ as the mean squared error between the enhanced feature S3_A and the true change feature Y:

$$L(\sigma) = E[\|Y - (\sigma \cdot \vec{V}_{S\_F} + (1-\sigma) \cdot \vec{V}_{S3})\|_2^2] \tag{30}$$

where E[] denotes the expectation operation, and $\|\|_2$ represents the L2 norm.

The second derivative of the objective function is:

$$\frac{\partial^2 L}{\partial \sigma^2} = 2 \cdot E[\|\vec{V}_{S\_F} - \vec{V}_{S3}\|_2^2] \tag{31}$$

Since the second derivative of $L(\sigma)$, $\frac{\partial^2 L}{\partial \sigma^2} > 0$ it indicates that $L(\sigma)$ is a strictly convex function with a unique global minimum. By ensuring $\sigma \in (0,1)$ through the sigmoid function, the stability of the optimization process is guaranteed. For example, in the LEVIR-CD dataset, when the difference between multi-scale features (deep semantics) and features at the current scale (shallow texture) is significant, the strict convexity ensures that the model can quickly converge to the optimal weight distribution, avoiding being trapped in local extremes.

2) Bias-Variance Trade-off in Statistical Learning. The mean squared error (MSE) of the augmented features, denoted as $MSE(S3\_A)$, is expressed as follows:

$$Bias = \sigma \cdot (E[\vec{V_{S\_F}}] - Y) + (1-\sigma) \cdot (E[\vec{V}_{S3}] - Y) \tag{32}$$

$$Variance = \sigma^2 \cdot Var[\vec{V_{S\_F}}] + (1-\sigma)^2 \cdot Var[\vec{V_{S3}}] \tag{33}$$

$$MSE(S3\_A) = Bias^2 + Variance \tag{34}$$

where Bias represents the bias term, Variance denotes the variance term, and Var() is the variance operator.

From the perspective of the bias-variance trade-off in statistics, optimization of model performance is achieved by balancing bias and variance. For instance, when multi-scale features exhibit high variance due to factors like illumination or noise, σ approaches 0, prioritizing the use of smooth features at the current scale (e.g., unobstructed local textures). Conversely, when current-scale features are affected by shadows or noise, σ tends toward 1, relying instead on the global consistency of multi-scale features. Experiments involving the addition of random illumination and noise (Table VIII) validated the robustness of AEGL-Net against illumination and noise perturbations.

Dynamic Weight Allocation:

In AEGL-Net, the fusion of global and local features effectively enhances the model's capability to perceive change regions. Specifically, local features dynamically adjust weight distribution based on the contextual information of input features, enabling the model to more accurately identify changes in diverse regions and objects. The following presents a theoretical analysis from the perspectives of dynamic adaptability and locality constraints.

1) Dynamic Adaptability. The local branch captures local neighborhood feature information via depthwise separable convolution, generating query Q, key K, and value V. The context-aware weight w is produced through the nonlinear activation of Q and K, with its generation process defined as:

$$w = Tanh(Swish(Q \cdot K)) \tag{35}$$

For any input feature x, the context-aware weight $w_{i,j}$ of the local branch output X_{local} is expressed as:

$$w_{i,j} = Tanh(Swish(\sum_{k \in N(i)} w_k \cdot x(k) \cdot \sum_{l \in N(j)} w_l \cdot x(l) \cdot)) \tag{36}$$

where N(i) denotes the neighborhood centered on pixel i, and w_k represents the shared weight of DWConv.

The Swish function $Swish(z) = z \cdot \sigma(z)$ (where σ is the sigmoid function) exhibits non-monotonic properties. Let matrix $M = Q \cdot K$, with its element $m_{i,j} = \sum_{k \in N(i)} w_k \cdot x(k) \cdot \sum_{l \in N(j)} w_l \cdot x(l)$. Swish operates element-wise on M, retaining only weights of strongly correlated features with $m_{i,j} > 0$ Based on feature correlation strength, it achieves "invalid correlation filtering" aligning with the optimization goal of "retaining features with high mutual information" in information theory.

The Tanh function maps inputs to $[-1,1]$, normalizing the output of $Swish(Q \cdot K)$. Let $z = Swish(Q \cdot K)$, after Tanh transformation, $\| w \|_\infty \leq 1$ ensuring stable element values in the weight matrix $W = [w_{i,j}]$. From an optimization theory perspective, this normalization satisfies the Lipschitz condition for gradient updates. That is, there exists a constant L such that the weight update step Δ_W satisfies $\| \Delta w \| \leq \eta \cdot L \cdot \| \nabla \ell \|$ ($\| \nabla \ell \|$ is the gradient norm, and η is the learning rate), guaranteeing training stability.

Due to the Swish function's characteristics, w_k is non-zero only when $\sum_{k\in N(i)} w_k \cdot x(k) \cdot \sum_{l\in N(j)} w_l \cdot x(l) > 0$. Let the rank of matrix $Q \cdot K$ be r. its non-zero elements are distributed across at most r rows/columns. Thus, the non-zero element proportion satisfies $\frac{nnz(W)}{N} \leq \frac{rank(Q\cdot K)}{N} \cdot \max_{i,j}(w_{i,j})$ where nnz(W) is the number of non-zero elements in weight matrix W, and N is the total number of matrix elements. This inequality demonstrates that dynamic weights retain key local features via sparsity, reduce redundant correlations, achieve efficient feature space compression, and preserve task-relevant mutual information.

2)Locality Constraints. The value of the output X_{local} from the local branch relies on local neighborhood information. For any input feature x, X_{local} is expressed as:

$$V(j) = DWconv_{5\times5}(Conv_{1\times1}(x)) \tag{37}$$

$$X_{local}(i) = \sum_{j\in N(i)} w_{i,j} \cdot V(j) \tag{38}$$

where V(j) involves the neighborhood of j. Therefore, V(j) contains the local contextual information of j. According to the definition of locality constraints, when $j \notin N(i)$, $w_{i,j} = 0$. Thus, $X_{local}(i)$ only aggregates features within the neighborhood N(i), that is: $X_{local}(i) = \sum_{j\notin N(i)} 0 \cdot V(j) + \sum_{j\notin N(i)} w_{i,j} \cdot V(j)$.This characteristic theoretically excludes the interference of distant irrelevant features, ensuring that dynamic weights focus on the features of the change region.

IV. Experiments

To fully evaluate the effectiveness of AEGL-Net, we carried out comprehensive experiments on four widely recognized RSCD datasets that cover various scenarios and challenges, fully reflecting the adaptability and stability of our model. We contrasted AEGL-Net with the current best approach on these datasets, and the results of the comparison experiments clearly illustrate the advantages of our model in terms of performance. Subsequently, we supply detailed depictions of the experiments and evaluation metrics of the RSCD task. To present the experimental results more clearly, we also conducted visual analyses, enabling a clear comparison of our model's performance on different datasets with other methods. Lastly, we undertook ablation experiments to more thoroughly authenticate the performance of AEGL-Net.

A. Data Description

We provide an overview of four influential datasets: LEVIR-CD, S2Looking, SYSU-CD, and UAV-CD, for evaluating the performance of AEGL-Net. These datasets embrace raw bitemporal images and their respective ground truth.

1) LEVIR-CD [58]: It is collected from Google Earth, consists of building change detection data spanning from 2002 to 2018 and includes various types of structures. This dataset includes 637 image pairs, where each image has a resolution of 1024 × 1024 pixels and an optical resolution of 0.5 meters. The images are subsequently cropped to a size of 256 × 256 pixels.
2) S2Looking [59]: It comprises 5,000 rural bitemporal images and 65,920 instances of global change, each with a size of 256 × 256 pixels. The images exhibit characteristics such as wide viewing angles and significant lighting variations.
3) SYSU-CD [60]: It comprising 20,000 pairs of aerial images from Hong Kong, collected between 2007 and 2014, each image is 256 × 256 pixels in size with a resolution of 0.5 meters.
4) UAV-CD[1] : Based on the UAV-BCD [61], it further enhanced sample diversity and increased the sample size by including land changes alongside building changes, expanding the total from 2024 to 2660 samples. Specifically, UAV-CD consists of a total of 2660 pairs of 768 × 768 pixel , spatial resolution of 0.06m.

B. Evaluation Metrics

To meticulously assess the AEGL-Net model's performance in the RSCD task, we apply four prevalent evaluation metrics: precision (P), recall (R), F1 score (F1), and Intersection over Union (IOU). Among these metrics, IOU and F1 are considered the primary evaluation criteria. P measures the fraction of true positives among all the samples that the model has identified as positive. R measures the fraction of true positive predictions among all the actual positive samples identified by the model. F1 merges the P and R metrics of the model, with a higher F1 reflecting better performance in the RSCD task. IOU indicates the ratio of the intersection of the predicted and actual change areas to their combined area. Each metric is calculated as follows:

$$Precision = \frac{T_P}{T_P + F_P} \tag{39}$$

$$Recall = \frac{T_P}{T_P + F_N} \tag{40}$$

$$IOU = \frac{T_P}{T_P + F_N + F_P} \tag{41}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{42}$$

where T_P stands the total number of pixels correctly recognized as a change, F_P stands the total count of unchanged pixels that were falsely classified as changed, and F_N refers to the number of pixels that experienced a change but were mistakenly categorized as unchanged.

C. Implementation Details

A single NVIDIA Geforce RTX 3060 (12G) GPU was employed for all the experiments conducted in this study. In the course of training, the PyTorch framework and Adam optimizer are utilized, with momentum set to 0.9 and weight decay configured at 0.0001, parameters β_1, β_2 of 0.9 and 0.99, where the initial learning rate is 0.0005, the batch size is 32, respectively. a poly learning strategy is used, i.e., $(1 - (cur\_iteration/\max\_iteration))^{power} \times lr$, where

[1]https://github.com/yikuizhai/UAV-CD

power and max_iteration. The model was trained for 40000 iterations on each dataset. Dropout was used to prevent overfitting. By randomly dropping a portion of neurons during the training process, it reduces the model's excessive reliance on specific neurons, thereby preventing overfitting.

D. Comparison with SOTA Models

The proposed model was compared with ten SOTA models on four different datasets, which include FC-EF [17], FC-Conc [17], FC-Diff [17], BIT [32], ChangeFormer [62], FCCDN [63], ICIF-Net [54], DMINet [64], VcT [36], and S2-CD [37]. It is important to note that in the RSCD task, the IOU and F1 values serve as key metrics for evaluating the model's performance. These metrics provide an intuitive reflection of the accuracy and robustness of the model. Therefore, we have chosen IOU and F1 values as the primary criteria for assessing model quality. To ensure a fair comparison, we have re-implemented these models based on their open-source codes. The following sections offer a detailed overview of these models:

1) FC-EF: It is a structure, based on the U-Net model, which has been adjusted to accept image pairs spliced together in the channel dimension.
2) FC-Conc: It is a fully convolutional Siamese network that identifies change regions by combining the feature maps of bitemporal images. The network consists of two weight-sharing branches that analyze images from different temporal phases.
3) FC-Diff: It is a fully convolutional Siamese network that identifies change regions by computing the variance between bitemporal image features. The network comprises two branches with shared weights that analyze images from different temporal phases.
4) BIT: It is an early model that introduces the concept of Transformer models into change detection. This model represents bitemporal images as tokens, leveraging a Transformer encoder to model the contextual relationships. Contextual tokens learned during the process are then reintroduced to the pixel space, where a Transformer decoder is applied to enhance the original features.
5) ChangeFormer: It is a Transformer-based Siamese network architecture that utilizes an advanced Vision Transformer model as the encoder. By incorporating a hierarchical Transformer encoder and an MLP decoder within the Siamese network framework.
6) FCCDN: It employs a dual encoder-decoder architecture along with a non-local feature pyramid for multi-scale feature fusion, while a densely connected module enhances the integration of temporal features. Furthermore, self-supervised learning is utilized to enhance the model's performance and generalization capabilities.
7) ICIF-Net: It represents a feature fusion network that combines mechanisms for intra-scale cross-interaction with those for inter-scale interaction, utilizing CNN and Transformer models to extract both local and global features. The network concludes by fusing global-local features through attention-based inter-scale fusion.
8) DMINet: It adopts a cross-period joint attention mechanism, integrating self-attention and cross-attention mechanisms through the JointAtt module to effectively guide image features and suppress interference. It utilizes difference feature acquisition and multi-level difference aggregation technology to achieve accurate detection of image changes.
9) VcT: It is a Visual Change Transformer network designed for CD. This network treats feature map pixels as graph nodes and models structured information using a graph neural network. It utilizes clustering algorithms to mine and refine Top-K tokens. Additionally, it employs self-cross-attention mechanisms along with the Anchor-Prompt Attention (APA) learning module to enhance features. In the final stage, a prediction head generates detailed change maps.
10) S2-CD: It precisely locates subtle differences in bitemporal images through summation and subtraction operations. The summation and subtraction modules quantify channel changes and capture spatial differences, respectively, which are then analyzed in depth by the Transformer model. Finally, heterogeneous modulation blocks integrate and enhance the channel and spatial feature differences.

1) Experiments on LEVIR-CD: Table I details the results for LEVIR-CD. The FCCDN model, among the ten baseline models, achieves the highest scores with IoU and F1 at 83.52%, and 91.02%, respectively. In comparison, our proposed AEGL-Net model achieves 84.55%, and 91.63% for these metrics, surpassing FCCDN. Specifically, our model outperforms FCCDN by 1.03% in IoU, and 0.61% in F1. Fig. 7 shows our method effectively reduces false positives and false negatives in LEVIR-CD, with AEGL-Net outperforming advanced methods. In (a), (c), (e), most methods miss change areas (blue regions) under cluttered environments, illumination variations, or dense changes, while ours accurately detects them, demonstrating superior detail-handling and detection performance.

2) Experiments on S2Looking: Table I details the results for S2Looking, the offered AEGL-Net model outperforms the DMINet model, which is the best among the ten baseline models,with reference to both IoU and F1. Specifically, AEGL-Net achieves improvements of +1.91% in IoU and +1.76% in F1. Fig. 8 on S2Looking shows AEGL-Net's superior detection performance: our method accurately identifies changed areas across (a)–(f), uniquely detecting small changes surrounded by complex environments in (a), (c), and (e), where state-of-the-art methods fail.

3) Experiments on SYSU-CD: Table I details the results for SYSU-CD, the AEGL-Net model surpasses the other ten baseline models, including the best-performing DMINet model, in both IoU and F1, which are key performance indicators. Specifically, our model achieves a 2.17% betterment in IoU and a 1.53% betterment in F1. These results clearly prove that AEGL-Net performs better than other baseline models in accurately detecting change areas in the SYSU-CD. Fig. 9 on SYSU-CD shows our method's consistent superiority: it reduces false positives (red in (a)), minimizes missed detections

TABLE I
QUANTITATIVE EVALUATION RESULTS OF THE LEVIR-CD, S2LOOKING, SYSU-CD, AND UAV-CD DATASET ARE PRESENTED. THE BEST VALUES ARE HIGHLIGHTED IN BOLD, AND ALL VALUES ARE EXPRESSED AS PERCENTAGES

Method	LEVIR-CD				S2Looking				SYSU-CD				UAV-CD			
	IOU	F1	R	P	IOU	F1	R	P	IOU	F1	R	P	IOU	F1	R	P
FC-EF	58.66	73.94	70.99	77.16	11.79	21.09	15.36	33.64	54.87	70.86	69.74	72.01	46.15	63.15	57.49	70.06
FC-Conc	73.93	85.01	85.17	84.85	16.22	27.92	22.08	37.97	60.05	75.04	78.38	71.98	44.75	61.83	55.47	69.84
FC-Diff	76.69	86.81	84.80	88.91	16.68	28.59	23.64	36.15	51.65	68.12	59.98	78.81	48.53	65.35	57.31	76.00
BIT	81.72	89.94	89.55	90.32	45.88	62.90	55.38	72.80	63.31	77.53	75.03	80.21	54.09	70.21	65.94	75.07
ChangeFormer	82.01	90.11	88.77	91.50	43.91	61.03	53.00	71.92	60.91	75.70	71.15	80.88	47.73	64.61	62.31	67.09
FCCDN	83.52	91.02	88.31	**93.91**	41.29	58.45	48.66	73.16	64.23	78.22	79.22	77.24	57.54	73.05	70.48	75.81
ICIF-Net	81.73	89.96	89.22	90.69	38.85	55.96	45.86	71.77	62.44	76.88	72.95	81.26	56.05	71.83	67.55	76.70
DMINet	83.44	90.97	89.82	92.16	46.45	63.43	55.09	**74.76**	67.58	80.65	76.16	**85.71**	57.61	73.11	69.42	77.20
VcT	81.28	89.67	86.75	92.80	45.38	62.43	55.63	71.12	65.50	79.15	77.29	81.11	54.03	70.16	63.87	77.82
S2-CD	81.31	89.69	87.78	91.68	43.47	60.59	51.14	74.34	64.25	78.23	76.11	80.47	56.58	72.27	67.25	78.09
AEGL-Net(Ours)	**84.55**	**91.63**	**90.78**	92.48	**48.36**	**65.19**	**60.96**	70.05	**69.75**	**82.18**	**81.40**	82.96	**60.23**	**75.18**	**72.15**	**78.47**

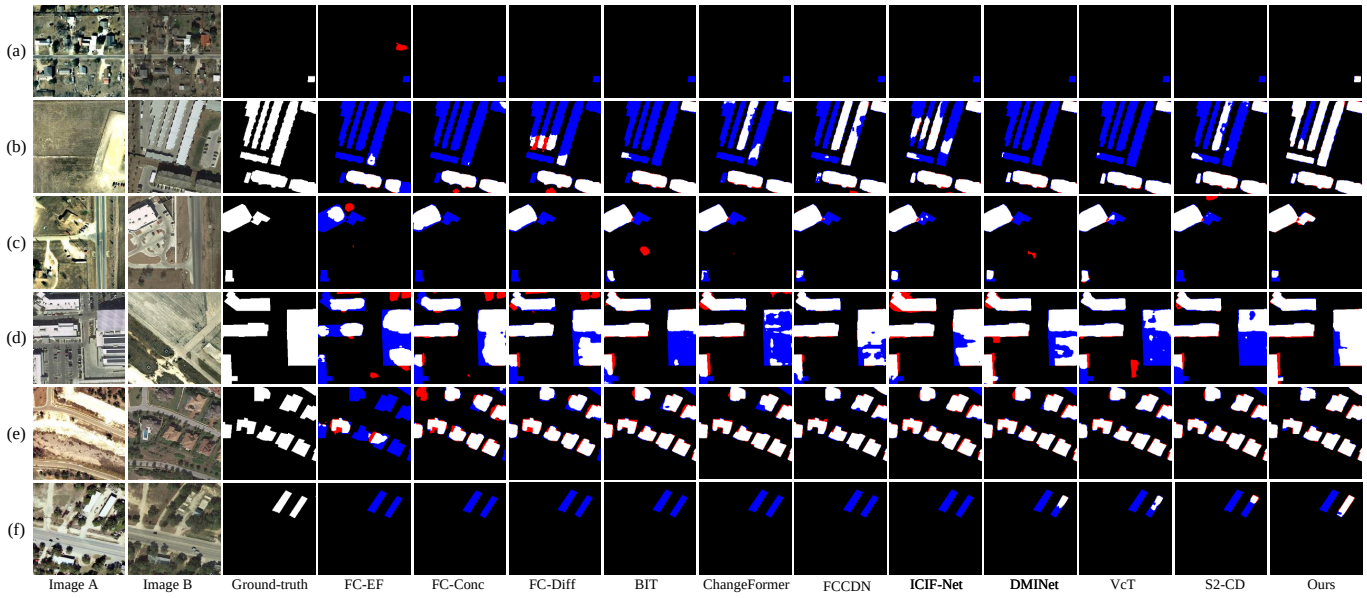


Fig. 7. Visual comparison of AEGL-Net with other advanced methods on the LEVIR-CD. In the Fig, white represents true positives, black represents true negatives, blue represents false negatives, and red represents false positives.

in dense environments (d), and excels in handling changed objects' edges and structures (e), outperforming state-of-the-art methods.

4) Experiments on UAV-CD: Table I details the results for UAV-CD, the AEGL-Net model outperforms ten other benchmark models, including the top-performing DMINet model, in both key performance metrics: IOU and F1 score. Specifically, our model leads by 2.62% in IOU and by 2.07% in F1 score. These results clearly indicate that AEGL-Net exhibits stronger performance compared to other benchmarks in identifying change regions within the UAV-CD. Additionally, Fig. 10 demonstrates the effectiveness of our method. Compared with state-of-the-art methods, which exhibit numerous blue-marked missed detection areas, our approach significantly minimizes such omissions and achieves a notably lower miss rate, reflecting its superior detection capability.

To further assess the efficiency of AEGL-Net, we compared it with several other models across four datasets: LEVIR-CD, S2Looking, SYSU-CD, and UAV-CD. The comparison focused on model parameters, floating-point operations (FLOPs), and inference time. As shown in Table II, AEGL-Net outperforms most of the compared models in terms of both parameter size and inference time. Although our method has slightly more parameters than FC-EF, FC-Conc, and FC-Diff, it demonstrates a significant performance improvement. Moreover, AEGL-Net achieves the lowest number of floating-point operations among all the models compared. Therefore, considering parameters, FLOPs, inference time, and overall performance, AEGL-Net clearly outperforms the other models.

In a nutshell, the visualization results presented in Fig. 7, 8, 9, 10 confirm that our model, through the AME module, enhances feature representation, enabling more accurate detection of change region boundaries. Additionally, the use of the GLFF module allows for the modeling of global features while

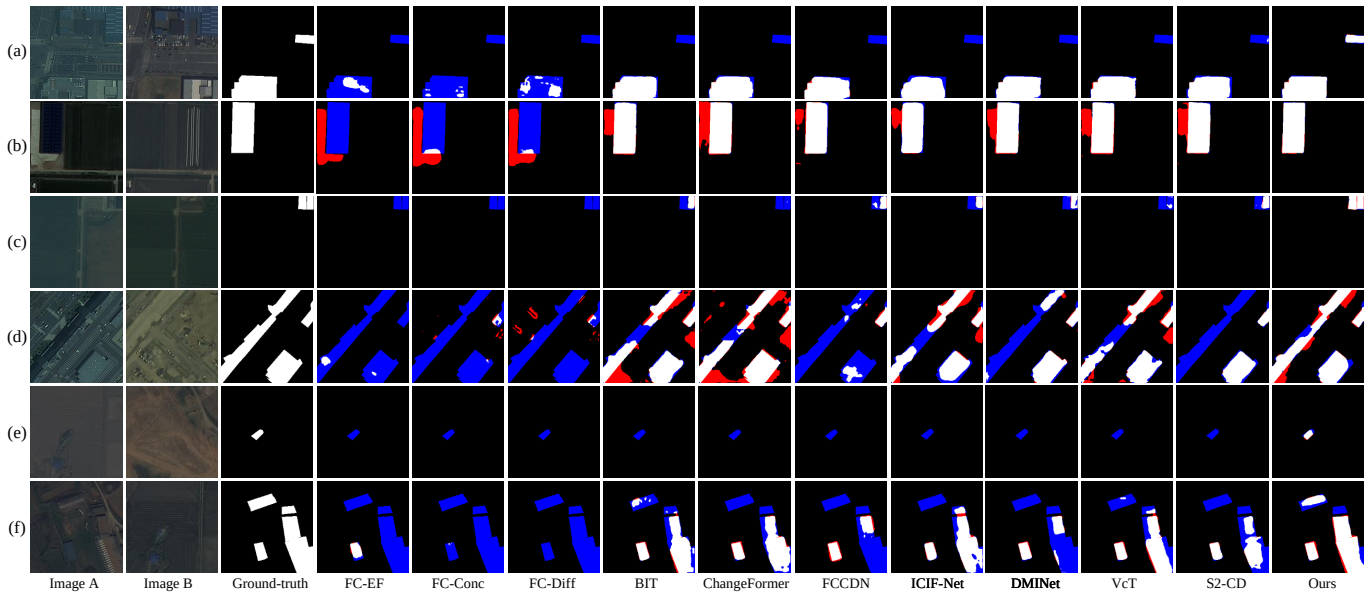


Fig. 8. Visual comparison of AEGL-Net with other advanced methods on the S2Looking. In the Fig, white represents true positives, black represents true negatives, blue represents false negatives, and red represents false positives.

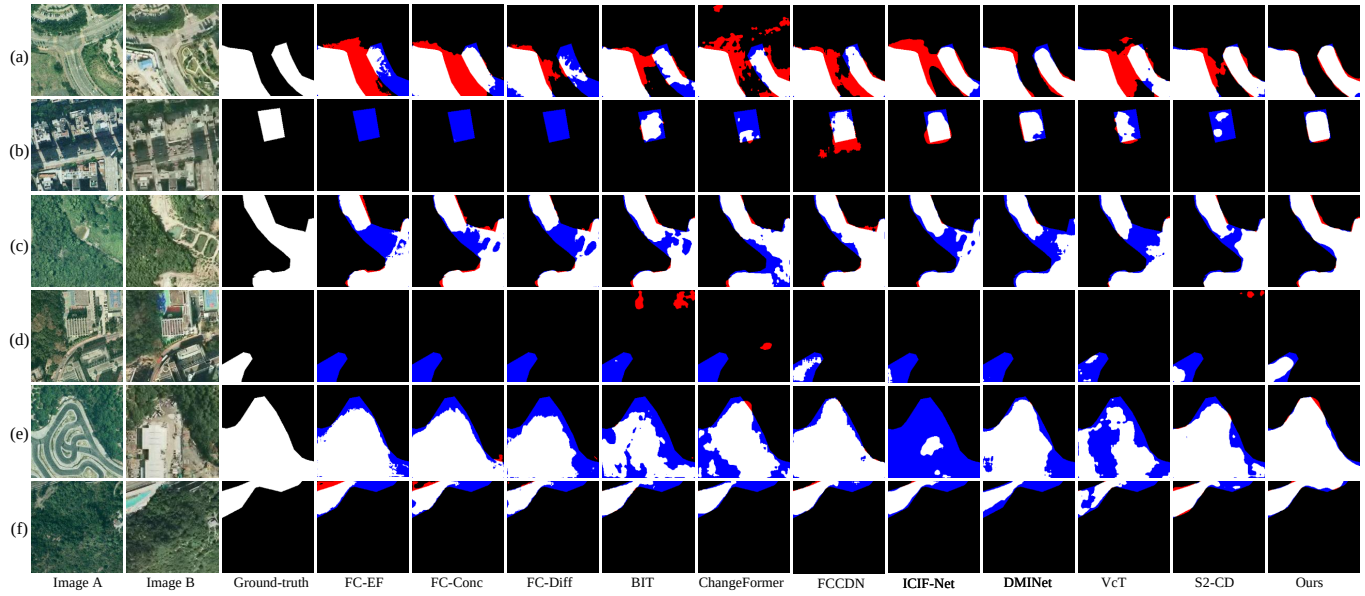


Fig. 9. Visual comparison of AEGL-Net with other advanced methods on the SYSU-CD. In the Fig, white represents true positives, black represents true negatives, blue represents false negatives, and red represents false positives.

retaining local feature information, thereby enabling precise identification of small objects and localized changes.

E. Ablation Study

In this section, to validate the efficacy of AME and GLFF, we performed ablation experiments on the LEVIR-CD, S2Looking, SYSU-CD, and UAV-CD. Table III provides the results, and the visualization results are displayed in Fig. 11. The baseline model, which does not include the AME and GLFF modules, employs 1×1 convolutions and element-wise addition to perform alignment and feature fusion operations.

1) Ablation of AME: The purpose of AME is to enable adaptive integration of multi-scale semantic and texture information extracted from the backbone network into features at each stage. The AME module enhances the ability to capture and utilize semantic details and texture variations at different scales, thereby improving the features at each scale. The results of experiments with the baseline model incorporating AME, referred to as baseline+AME, on the four datasets are shown in Table III. Specifically, after incorporating AME, the F1 for baseline+AME on the LEVIR-CD, S2Looking, SYSU-CD, and UAV-CD are 91.13%, 63.59%, 81.07%, and 73.28%, respectively, representing increases of 1.07%, 0.95%, 2.4%,

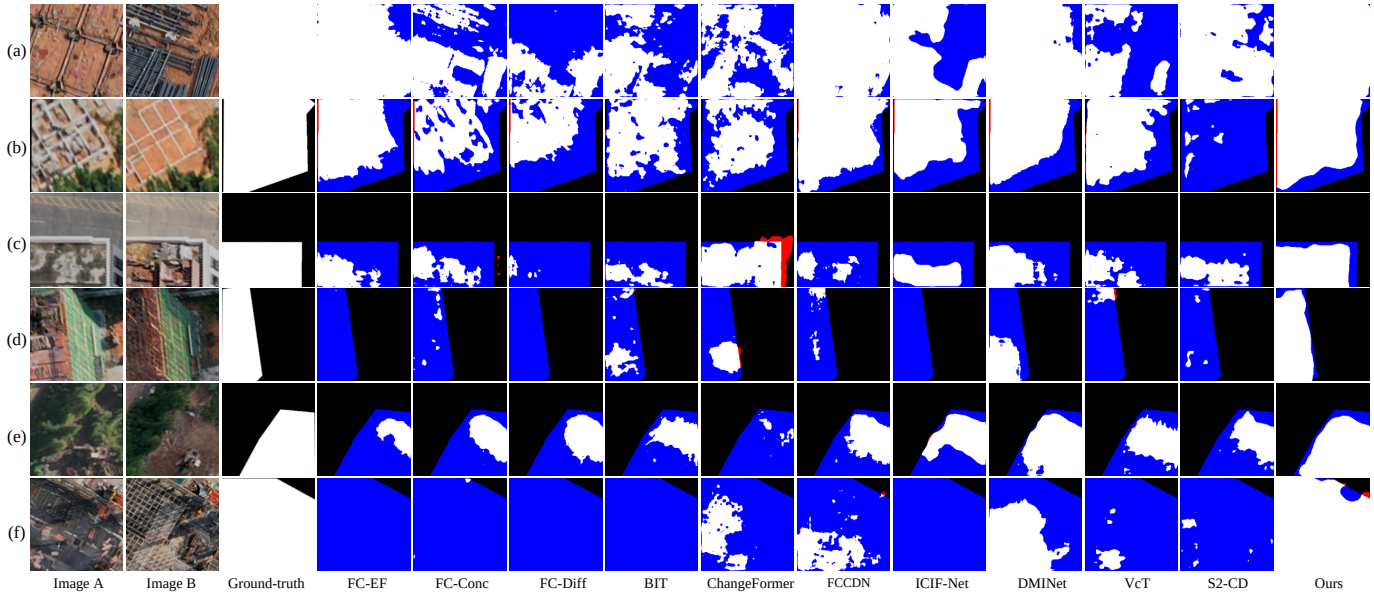


Fig. 10. Visual comparison of AEGL-Net with other advanced methods on the UAV-CD. In the Fig, white represents true positives, black represents true negatives, blue represents false negatives, and red represents false positives.

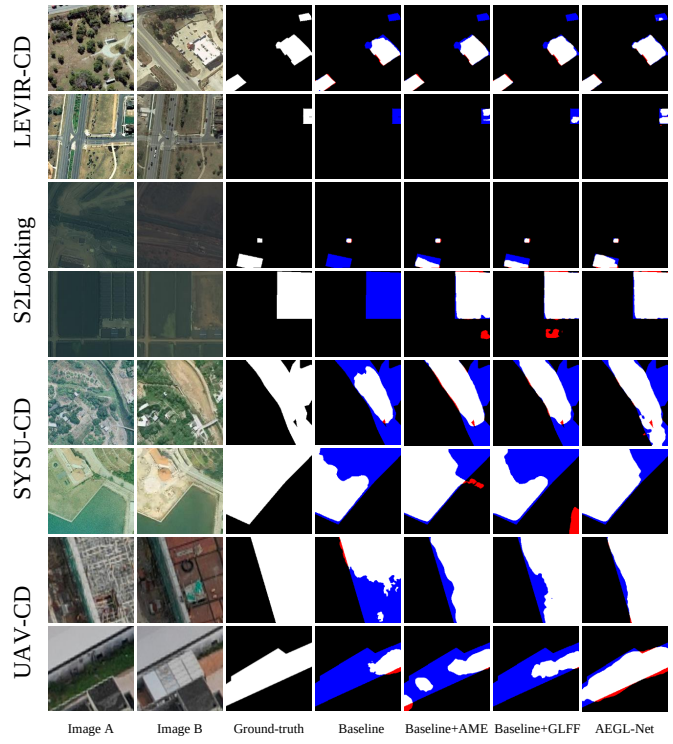


Fig. 11. Visualization results of ablation experiments for different modules in AEGL-Net. In the figure, white represents true positives, black represents true negatives, blue represents false negatives, and red represents false positives.

TABLE II
PARAMETERS(PARAMS), FLOATING-POINT OPERATIONS PER SECOND(FLOPS) AND INFERENCE TIME ON THE LEVIR-CD, S2LOOKING, SYSU-CD, AND UAV-CD DATASETS

Method	Params(M)	FLOPs(G)	**LEVIR-CD**	**S2Looking**	**SYSU-CD**	**UAV-CD**
			Inference Time(S)			
FC-EF	1.35	3.57	3.60	26.25	7.09	4.24
FC-Conc	1.54	5.32	4.45	31.06	8.47	5.01
FC-Diff	1.35	4.72	4.35	32.69	8.78	5.07
BIT	11.95	3.55	24.45	188.94	47.11	27.71
ChangeFormer	43.97	202.79	106.68	838.04	209.51	125.72
FCCDN	6.31	6.66	14.83	161.95	39.31	27.84
ICIF-Net	25.85	25.36	47.27	375.22	90.27	52.87
DMINet	6.75	14.42	23.50	179.44	45.43	27.18
VcT	12.40	10.64	61.00	495.27	129.01	74.98
S2-CD	12.16	9.26	21.69	166.42	42.79	25.59
AEGL-Net	3.85	1.26	17.11	131.47	33.29	20.05

and 1.7% compared to the baseline model. The IoU also show improvements of 1.78%, 1.02%, 3.32%, and 2.09%, respectively. Visualization results in Fig. 11 demonstrate that adding the AME module to the baseline model allows for the detection of more change areas (blue). For example, in Fig. 11, on the SYSU-CD, despite some false detections, the model can accurately identify more change areas compared to the baseline. Overall, both the improvements in IoU and F1 shown in Table III and the visualization results in Fig. 11 indicate that the AME module effectively enhances feature representation.

2) Ablation of GLFF: The purpose of GLFF is to enhance the modeling of global dependencies when fusing bitemporal features in AEGL-Net. GLFF effectively integrates shared weights and context-aware weights to enhance local features, which helps reduce false negatives for minor changes. The results of experiments with the baseline model incorporating GLFF, referred to as baseline+GLFF, on the four datasets are shown in Table III. Specifically, after incorporating GLFF, the IoU and F1 for baseline+GLFF are 1.78% and 1.07% higher, respectively, on the LEVIR-CD, 0.59% and 0.55% higher on the S2Looking, and 1.78% and 1.29% higher on the SYSU-CD, and 2.55% and 2.06% higher on the UAV-CD as opposed to the baseline. visual display results are presented in Fig. 11. The introduction of GLFF into the baseline model

TABLE III
QUANTITATIVE EVALUATION RESULTS OF THE ABLATION EXPERIMENTS FOR AME AND GLFF ON THE LEVIR-CD, S2LOOKING, SYSU-CD, AND UAV-CD. THE BEST VALUES ARE HIGHLIGHTED IN BOLD, AND ALL VALUES ARE EXPRESSED AS PERCENTAGES

Method	LEVIR-CD				S2Looking				SYSU-CD				UAV-CD			
	IOU	F1	R	P	IOU	F1	R	P	IOU	F1	R	P	IOU	F1	R	P
baseline	81.92	90.06	88.38	91.81	45.60	62.64	59.15	66.57	64.84	78.67	75.23	82.44	55.73	71.58	74.02	69.29
baseline+AME	83.70	91.13	89.92	92.36	46.62	63.59	60.69	66.78	68.16	81.07	77.30	**85.23**	57.82	73.28	70.58	76.19
baseline+GLFF	83.70	91.13	89.72	**92.58**	46.19	63.19	59.55	67.29	66.62	79.96	78.93	81.03	58.28	73.64	**75.44**	71.93
AEGL-Net	**84.55**	**91.63**	**90.78**	92.48	**48.36**	**65.19**	**60.96**	**70.05**	**69.75**	**82.18**	**81.40**	82.96	**60.23**	**75.18**	72.15	**78.47**

TABLE IV
QUANTITATIVE EVALUATION RESULTS OF ABLATION EXPERIMENTS WITH LBA AND GBA ON THE LEVIR-CD, S2LOOKING, SYSU-CD, AND UAV-CD.THE BEST VALUES ARE HIGHLIGHTED IN BOLD, ALL VALUES ARE EXPRESSED AS PERCENTAGES

Method	LEVIR-CD				S2Looking				SYSU-CD				UAV-CD			
	IOU	F1	R	P	IOU	F1	R	P	IOU	F1	R	P	IOU	F1	R	P
Without LBA	84.10	91.36	89.77	**93.02**	47.72	64.61	**62.31**	67.09	66.78	80.08	74.88	**86.07**	57.86	73.30	72.60	74.01
Without GBA	83.88	91.24	89.98	92.53	47.79	64.67	59.65	**70.61**	67.20	80.38	76.51	84.66	58.33	73.68	**73.33**	74.04
AEGL-Net	**84.55**	**91.63**	**90.78**	92.48	**48.36**	**65.19**	60.96	70.05	**69.75**	**82.18**	**81.40**	82.96	**60.23**	**75.18**	72.15	**78.47**

improves the accuracy in identifying and localizing change areas. Additionally, Fig. 11 indicate improved edge detection capabilities. The improvements in experimental metrics shown in Table III and the visualization results in Fig. 11 demonstrate that integrating bitemporal features with GLFF and utilizing global-local attention enhances both local and global feature learning.

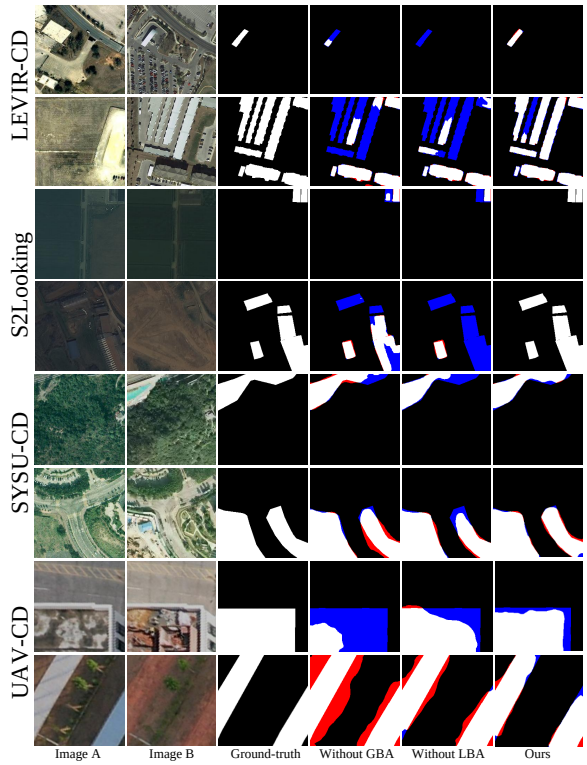


Fig. 12. Visualization results of ablation experiments With LBA and GBA On in AEGL-Net. In the figure, white represents true positives, black represents true negatives, blue represents false negatives, and red represents false positives.

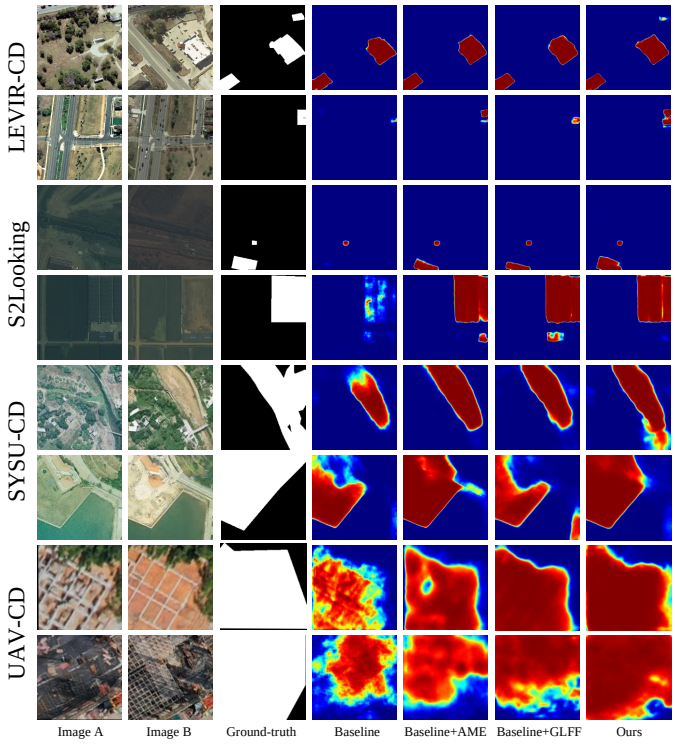


Fig. 13. Heatmap results of ablation experiments for different modules in AEGL-Net. Red areas indicate regions of higher attention, while blue areas indicate regions of lower attention.

From the visualizations of the four datasets in Fig. 11, we observe that incorporating either the AME or GLFF module into the baseline model improves the identification of change areas. However, the results displayed in Fig. 11, while demonstrating progress in reducing missed detections, also introduce some false detections. When both AME and GLFF modules are used together, the model not only accurately identifies more change areas but also effectively reduces false detections. To further illustrate the impact of AME and GLFF modules,

TABLE V
QUANTITATIVE EVALUATION RESULTS OF ABLATION EXPERIMENTS WITH DIFFERENT ACTIVATION FUNCTIONS ON THE LEVIR-CD, S2LOOKING, SYSU-CD, AND UAV-CD.THE BEST VALUES ARE HIGHLIGHTED IN BOLD, ALL VALUES ARE EXPRESSED AS PERCENTAGES

Method	LEVIR-CD				S2Looking				SYSU-CD				UAV-CD			
	IOU	F1	R	P	IOU	F1	R	P	IOU	F1	R	P	IOU	F1	R	P
Softmax	84.05	91.33	90.06	92.65	46.06	63.07	60.91	65.38	66.71	80.03	76.67	83.7	56.48	72.19	70.52	73.94
Tanh	84.12	91.37	90.68	92.07	47.25	64.18	**61.14**	67.54	66.87	80.15	75.04	**85.99**	58.93	74.16	**75.68**	72.70
Swish	84.13	91.38	**91.09**	**91.67**	47.60	64.52	60.82	68.70	66.95	80.20	75.45	85.60	59.17	74.35	74.36	74.34
Swish+Tanh	**84.55**	**91.63**	90.78	92.48	**48.36**	**65.19**	60.96	**70.05**	**69.75**	**82.18**	**81.40**	82.96	**60.23**	**75.18**	72.15	**78.47**

TABLE VI
QUANTITATIVE EVALUATION RESULTS OF ABLATION EXPERIMENTS WITH DIFFERENT BACKBONES ON THE LEVIR-CD, S2LOOKING, SYSU-CD, AND UAV-CD.THE BEST VALUES ARE HIGHLIGHTED IN BOLD, ALL VALUES ARE EXPRESSED AS PERCENTAGES

Backbone	Params	LEVIR-CD				S2Looking				SYSU-CD				UAV-CD			
		IOU	F1	R	P	IOU	F1	R	P	IOU	F1	R	P	IOU	F1	R	P
ResNet18	17.64M	83.82	91.20	**91.00**	91.40	47.07	64.01	60.56	67.89	68.11	81.03	79.76	82.35	55.53	71.41	68.53	74.54
ResNet34	27.75M	84.24	91.45	90.34	**92.57**	46.98	63.93	60.62	67.61	67.74	80.77	76.23	**85.88**	58.64	73.93	69.08	**79.51**
ResNet50	44.47M	84.00	91.31	90.19	92.45	43.22	60.35	56.92	64.23	68.79	81.51	78.70	84.52	59.99	75.00	**75.29**	84.52
MobileNetV2	3.85M	**84.55**	**91.63**	90.78	92.48	**48.36**	**65.19**	**60.96**	**70.05**	**69.75**	**82.18**	**81.40**	82.96	**60.23**	**75.18**	72.15	78.47

TABLE VII
QUANTITATIVE EVALUATION RESULTS OF ABLATION EXPERIMENTS WITH DIFFERENT BATCHSIZE ON THE LEVIR-CD, S2LOOKING, SYSU-CD, AND UAV-CD.THE BEST VALUES ARE HIGHLIGHTED IN BOLD, ALL VALUES ARE EXPRESSED AS PERCENTAGES

Method	LEVIR-CD				S2Looking				SYSU-CD				UAV-CD			
	IOU	F1	R	P	IOU	F1	R	P	IOU	F1	R	P	IOU	F1	R	P
batchsize=64	84.16	91.40	89.58	**93.30**	48.13	64.99	**62.42**	67.77	68.46	81.28	78.69	84.04	57.89	73.33	**75.79**	71.03
batchsize=24	84.15	91.39	**90.79**	91.99	47.07	64.01	59.10	69.81	68.44	81.26	76.73	**86.36**	59.12	74.31	72.46	76.25
batchsize=16	84.18	91.41	90.45	92.39	45.95	62.97	59.83	66.45	67.88	80.87	79.61	85.64	59.49	74.60	72.71	76.59
batchsize=8	83.27	90.87	90.49	91.26	43.22	60.35	59.44	61.29	68.36	81.20	80.36	82.06	57.58	73.08	76.29	70.13
batchsize=32	**84.55**	**91.63**	90.78	92.48	**48.36**	**65.19**	60.96	**70.05**	**69.75**	**82.18**	**81.40**	82.96	**60.23**	**75.18**	72.15	**78.47**

Fig. 13 presents heatmaps showing the changes in the model's attention areas before and after incorporating these modules. Fig. 13 indicates that models with AME or GLFF modules are more effective in focusing on change areas compared to the baseline model. When both AME and GLFF modules are introduced, the attention areas of the model closely align with the actual change areas. Fig. 13 shows that while models with AME or GLFF modules significantly focus on change areas, they also attend to some non-change areas. However, the simultaneous use of AME and GLFF modules not only increases the focus on change areas but also eliminates the attention to non-change areas observed with individual module usage. These results demonstrate that both AME and GLFF modules play crucial roles in AEGL-Net.

3) Ablation of LBA and GBA: To gauge the efficacy of the global and local branches in the GLFF module, we conducted experiments as detailed in Table IV. The results imply that AEGL-Net consistently outperforms models without the LBA or GBA modules across the LEVIR-CD, S2Looking, SYSU-CD, and UAV-CD. Relying solely on global features tends to overlook detailed changes, leading to decreased detection accuracy. Conversely, the absence of local feature analysis restricts the model's performance in detecting small or subtle changes, adversely affecting overall detection accuracy. On the other hand, relying exclusively on local features ignores the broader contextual information of the image, making it challenging to identify changes over extensive areas, which may result in inconsistent and erroneous detections. The lack of global features increases the risk of false positives and missed detections, reducing the model's ability to handle complex scenarios. In addition, the visualization results in Fig. 12 show that, compared to AEGL-Net, the absence of GBA or LBA leads to more false negatives and false positives, further validating the effectiveness of GBA and LBA.

4) Ablation of Activation Functions: To evaluate the effectiveness of the Swish and Tanh activation functions used in LBA, we conducted experiments as shown in Table V. The results indicate that replacing either Swish or Tanh with Softmax leads to a significant decrease in performance across the LEVIR-CD, S2Looking, SYSU-CD, and UAV-CD. While using Swish or Tanh alone results in slightly better performance compared to Softmax on these datasets, it is still significantly inferior to AEGL-Net. The performance metrics presented in Table V also confirm that the non-monotonic nature of Swish enhances non-linearity, which is beneficial for capturing complex changes. Additionally, the smoothness of Tanh improves gradient flow, mitigates the vanishing gradient problem, and enhances model stability.

TABLE VIII
QUANTITATIVE EVALUATION RESULTS OF VARIABILITY EXPERIMENTS ON LEVIR-CD, S2LOOKING, SYSU-CD, AND UAV-CD UNDER NOISY OR LIGHTING CONDITIONS

Method	LEVIR-CD				S2Looking				SYSU-CD				UAV-CD			
	IOU	F1	R	P	IOU	F1	R	P	IOU	F1	R	P	IOU	F1	R	P
baseline	81.92	90.06	88.38	91.81	45.60	62.64	59.15	66.57	64.84	78.67	75.23	82.44	55.73	71.58	74.02	69.29
baseline+noise	81.20	89.62	88.61	90.66	41.05	58.20	54.76	62.11	54.68	70.70	71.86	69.57	52.47	68.83	61.36	78.35
baseline+light	81.09	89.56	88.59	90.55	43.50	60.63	55.60	66.65	52.41	68.77	78.76	61.03	51.72	68.18	65.41	71.19
baseline+noise+light	78.82	88.15	86.31	90.08	35.88	52.81	46.38	61.30	46.93	63.88	60.19	68.05	44.54	61.63	61.05	62.22
AEGL-Net+noise	84.45	91.57	90.29	92.88	47.54	64.44	59.92	69.71	67.65	80.70	75.67	86.46	58.37	73.71	70.56	77.16
AEGL-Net+light	84.44	91.56	90.26	92.90	46.90	63.85	59.09	69.44	68.55	81.34	77.54	85.53	56.80	72.45	68.13	77.36
AEGL-Net+noise+light	84.08	91.35	90.15	92.58	45.71	62.74	60.60	65.04	67.47	80.58	77.49	83.92	55.61	71.47	79.05	65.22
AEGL-Net	84.55	91.63	90.78	92.48	48.36	65.19	60.96	70.05	69.75	82.18	81.40	82.96	60.23	75.18	72.15	78.47

F. Discussion

In this section, we will thoroughly examine the performance of AEGL-Net under different conditions. Specifically, we analyze the variations in AEGL-Net's performance when different Backbone networks are used, as well as the model's training effectiveness and generalization ability with varying Batch sizes. Additionally, we discuss AEGL-Net's performance under the influence of environmental factors such as noise and lighting variations. To validate the robustness and stability of AEGL-Net.

1) Scalable Experiment: We conducted further validation of the proposed method using commonly employed backbones in the RSCD field, including ResNet18, ResNet34, and ResNet50. Due to memory limitations on the experimental machine when using ResNet50, a batch size of 16 was used, while other experiments used a batch size of 32. As shown in Table VI, our proposed method, despite having only 3.85M parameters, outperforms models using ResNet18, ResNet34, and ResNet50 as backbones in terms of IoU and F1 on the LEVIR-CD, S2Looking, SYSU-CD, and UAV-CD. This further underscores the advantages of AEGL-Net. Additionally, we examined the effect of varying batch sizes on model performance by setting the batch sizes to 64, 24, 16, and 8, and comparing the results with AEGL-Net (batch size = 32). As shown in Table VII, the model's performance was slightly inferior to that of AEGL-Net when the batch size was either greater than or less than 32, further indicating that AEGL-Net achieves optimal performance with a batch size of 32.

2) Variabilities Experiment: Inspired by [65], we manually introduced noise and illumination variations into four datasets to further evaluate the robustness and performance of AEGL-Net under different variability conditions. The experimental results are presented in Table VIII. As indicated in the table, AEGL-Net outperforms the baseline model in terms of IoU by 5.26%, 9.83%, 20.54%, and 11.07% across the four datasets, respectively. Additionally, it achieves improvements in the F1 of 3.20%, 9.93%, 16.70%, and 9.84%. These results demonstrate that AEGL-Net significantly enhances change detection performance in the presence of noise and illumination variations, showcasing strong resistance to interference. Furthermore, a detailed analysis of the data in Table VIII reveals that AEGL-Net consistently surpasses the baseline even when noise or illumination changes are applied independently. This suggests that AEGL-Net not only excels under various variability conditions but also maintains robust detection accuracy despite local image quality degradation.

V. CONCLUSION

In this research, we propose a model incorporating Adaptive Multi-Scale Enhancement (AME) and Global-Local Feature Fusion (GLFF) modules for remote sensing change detection (RSCD). This network leverages AME to refine multi-scale features extracted by the backbone, thereby improving the capture and utilization of semantic details and texture changes across various scales.. Additionally, we introduce GLFF to effectively fuse bitemporal features. Unlike other attention mechanisms, this module enhances global dependency modeling through global-local attention while also integrating shared and context-aware weights to enhance local features. A wide range of tests was conducted using four public datasets, and the findings indicate that the approach we introduced surpasses other leading-edge change detection techniques in terms of performance.

However, as a fully supervised model, AEGL-Net's performance is contingent on having labeled training data. In the absence of such labels, the model cannot be trained, which limits its applicability, especially in scenarios requiring rapid responses or where data labeling is challenging. This dependency necessitates ample labeled data to facilitate comprehensive training, making it unsuitable for unsupervised or semi-supervised learning contexts. Furthermore, given the widespread application of large-scale models, our future work will focus on integrating these large-scale models with self-supervised, semi-supervised, and unsupervised RSCD networks to address complex challenges encountered in practical applications effectively.

REFERENCES

[1] M. Mandal and S. K. Vipparthi, "An Empirical Review of Deep Learning Frameworks for Change Detection: Model Design, Experimental Frameworks, Challenges and Research Needs," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 6101–6122, Jul. 2022.

[2] L. Li, W. Zhan, W. Ju, J. Peñuelas, Z. Zhu, S. Peng, X. Zhu, Z. Liu, Y. Zhou, J. Li, J. Lai, F. Huang, G. Yin, Y. Fu, M. Li, and C. Yu, "Competition between biogeochemical drivers and land-cover changes determines urban greening or browning," *Remote Sensing of Environment*, vol. 287, p. 113481, Mar. 2023.

[3] Z. Zheng, Y. Zhong, J. Wang, A. Ma, and L. Zhang, "Building damage assessment for rapid disaster response with a deep object-based semantic change detection framework: From natural disasters to man-made disasters," *Remote Sensing of Environment*, vol. 265, p. 112636, Nov. 2021.

[4] O. Yousif and Y. Ban, "Improving Urban Change Detection From Multitemporal SAR Images Using PCA-NLM," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 4, pp. 2032–2041, Apr. 2013.

[5] E. L. Bullock, S. P. Healey, Z. Yang, R. Houborg, N. Gorelick, X. Tang, and C. Andrianirina, "Timeliness in forest change monitoring: A new assessment framework demonstrated using sentinel-1 and a continuous change detection algorithm," *Remote Sensing of Environment*, vol. 276, p. 113043, 2022.

[6] C. A. Mucher, K. T. Steinnocher, F. P. Kressler, and C. Heunks, "Land cover characterization and change detection for environmental monitoring of pan-Europe," *International Journal of Remote Sensing*, vol. 21, no. 6-7, pp. 1159–1181, Jan. 2000.

[7] L. Wan, Y. Xiang, and H. You, "A Post-Classification Comparison Method for SAR and Optical Images Change Detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 7, pp. 1026–1030, Jul. 2019.

[8] X. Peng, R. Zhong, Z. Li, and Q. Li, "Optical Remote Sensing Image Change Detection Based on Attention Mechanism and Image Difference," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 9, pp. 7296–7307, Sep. 2021.

[9] M. Gong, J. Zhao, J. Liu, Q. Miao, and L. Jiao, "Change Detection in Synthetic Aperture Radar Images Based on Deep Neural Networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 1, pp. 125–138, Jan. 2016.

[10] S. Saha, F. Bovolo, and L. Bruzzone, "Unsupervised Deep Change Vector Analysis for Multiple-Change Detection in VHR Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 6, pp. 3677–3693, Jun. 2019.

[11] L. Bruzzone and D. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 3, pp. 1171–1182, May 2000.

[12] G. Byrne, P. Crapper, and K. Mayo, "Monitoring land-cover change by principal component analysis of multitemporal landsat data," *Remote Sensing of Environment*, vol. 10, no. 3, pp. 175–184, Nov. 1980.

[13] T. Zhan, Y. Sun, Y. Tang, Y. Xu, and Z. Wu, "Tensor Regression and Image Fusion-Based Change Detection Using Hyperspectral and Multispectral Images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 9794–9802, 2021.

[14] X. Wu, D. Hong, and J. Chanussot, "UIU-Net: U-Net in U-Net for Infrared Small Object Detection," *IEEE Transactions on Image Processing*, vol. 32, pp. 364–376, 2023.

[15] X. Wu, D. Hong, and J. Chanussot, "Convolutional neural networks for multimodal remote sensing data classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–10, 2021.

[16] D. Hong, J. Yao, C. Li, D. Meng, N. Yokoya, and J. Chanussot, "Decoupled-and-Coupled Networks: Self-Supervised Hyperspectral Image Super-Resolution With Subpixel Fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–12, 2023.

[17] R. Caye Daudt, B. Le Saux, and A. Boulch, "Fully Convolutional Siamese Networks for Change Detection," in *2018 25th IEEE International Conference on Image Processing (ICIP)*. Athens: IEEE, Oct. 2018, pp. 4063–4067.

[18] 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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 3867–3878, 2023.

[19] T. Lei, X. Geng, H. Ning, Z. Lv, M. Gong, Y. Jin, and A. K. Nandi, "Ultralightweight Spatial–Spectral Feature Cooperation Network for Change Detection in Remote Sensing Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2023.

[20] J. Chen, Z. Yuan, J. Peng, L. Chen, H. Huang, J. Zhu, Y. Liu, and H. Li, "DASNet: Dual Attentive Fully Convolutional Siamese Networks for Change Detection in High-Resolution Satellite Images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 1194–1206, 2021.

[21] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE," 2021.

[22] S. Marchesi and L. Bruzzone, "ICA and kernel ICA for change detection in multispectral remote sensing images," in *2009 IEEE International Geoscience and Remote Sensing Symposium*, vol. 2, Jul. 2009, pp. II–980–II–983.

[23] Y. Sun, L. Lei, X. Tan, D. Guan, J. Wu, and G. Kuang, "Structured graph based image regression for unsupervised multimodal change detection," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 185, pp. 16–31, Mar. 2022.

[24] Y. Sun, L. Lei, D. Guan, M. Li, and G. Kuang, "Sparse-Constrained Adaptive Structure Consistency-Based Unsupervised Image Regression for Heterogeneous Remote-Sensing Change Detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.

[25] T. Xiao, Y. Wan, J. Chen, W. Shi, J. Qin, and D. Li, "Multiresolution-Based Rough Fuzzy Possibilistic C-Means Clustering Method for Land Cover Change Detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 570–580, 2023.

[26] Z. Yuan, L. Mou, Z. Xiong, and X. X. Zhu, "Change Detection Meets Visual Question Answering," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[27] B. Fang, G. Chen, G. Ouyang, J. Chen, R. Kou, and L. Wang, "Content-Invariant Dual Learning for Change Detection in Remote Sensing Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–17, 2022.

[28] J. Lei, Y. Gu, W. Xie, Y. Li, and Q. Du, "Boundary Extraction Constrained Siamese Network for Remote Sensing Image Change Detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[29] Y. Zhai, W. Li, T. Xian, X. Jia, H. Zhang, Z. Tan, J. Zhou, J. Zeng, and C. L. Philip Chen, "CAS-Net: Comparison-Based Attention Siamese Network for Change Detection With an Open High-Resolution UAV Image Dataset," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–17, 2024.

[30] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, vol. 9351, pp. 234–241.

[31] H. Li, X. Liu, H. Li, Z. Dong, and X. Xiao, "MDFENet: A Multiscale Difference Feature Enhancement Network for Remote Sensing Change Detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 3104–3115, 2023.

[32] H. Chen, Z. Qi, and Z. Shi, "Remote Sensing Image Change Detection With Transformers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.

[33] Y. Zhao, P. Chen, Z. Chen, Y. Bai, Z. Zhao, and X. Yang, "A Triple-Stream Network With Cross-Stage Feature Fusion for High-Resolution Image Change Detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–17, 2023.

[34] C. Zhang, L. Wang, S. Cheng, and Y. Li, "SwinSUNet: Pure Transformer Network for Remote Sensing Image Change Detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[35] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal, QC, Canada: IEEE, Oct. 2021, pp. 9992–10 002.

[36] B. Jiang, Z. Wang, X. Wang, Z. Zhang, L. Chen, X. Wang, and B. Luo, "VcT: Visual Change Transformer for Remote Sensing Image Change Detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2023.

[37] L. Wang, Y. Fang, Z. Li, C. Wu, M. Xu, and M. Shao, "Summator–Subtractor Network: Modeling Spatial and Channel Differences for Change Detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–12, 2024.

[38] H. Li, X. Liu, H. Li, Z. Dong, and X. Xiao, "Mdfenet: A multiscale difference feature enhancement network for remote sensing change detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 3104–3115, 2023.

[39] T. Lei, J. Wang, H. Ning, X. Wang, D. Xue, Q. Wang, and A. K. Nandi, "Difference enhancement and spatial–spectral nonlocal network for change detection in vhr remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2021.

[40] D. Song, Y. Dong, and X. Li, "Context and difference enhancement network for change detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 9457–9467, 2022.

[41] W. Le, L. Huang, B.-H. Tang, Q. Tian, and M. Wang, "Acmfnet: Asymmetric convolutional feature enhancement and multiscale fusion network for change detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024.

[42] Z. Li, C. Tang, L. Wang, and A. Y. Zomaya, "Remote sensing change detection via temporal feature interaction and guided refinement," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2022.

[43] K. Jiang, W. Zhang, J. Liu, F. Liu, and L. Xiao, "Joint variation learning of fusion and difference features for change detection in remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–18, 2022.

[44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 770–778.

[45] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI: IEEE, Jul. 2017, pp. 1800–1807.

[46] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning Deep Features for Discriminative Localization," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 2921–2929.

[47] G. Li, Z. Liu, Z. Bai, W. Lin, and H. Ling, "Lightweight Salient Object Detection in Optical Remote Sensing Images via Feature Correlation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022.

[48] P. Zhang, D. Wang, H. Lu, H. Wang, and X. Ruan, "Amulet: Aggregating Multi-level Convolutional Features for Salient Object Detection," in *2017 IEEE International Conference on Computer Vision (ICCV)*. Venice: IEEE, Oct. 2017, pp. 202–211.

[49] Q. Guo, J. Zhang, S. Zhu, C. Zhong, and Y. Zhang, "Deep Multiscale Siamese Network With Parallel Convolutional Structure and Self-Attention for Change Detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022.

[50] F. Luo, T. Zhou, J. Liu, T. Guo, X. Gong, and J. Ren, "Multiscale Diff-Changed Feature Fusion Network for Hyperspectral Image Change Detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–13, 2023.

[51] Z. Ying, Z. Tan, Y. Zhai, X. Jia, W. Li, J. Zeng, A. Genovese, V. Piuri, and F. Scotti, "DGMA2-Net: A Difference-Guided Multiscale Aggregation Attention Network for Remote Sensing Change Detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–16, 2024.

[52] Z. Fu, J. Li, L. Ren, and Z. Chen, "SLDDNet: Stagewise Short and Long Distance Dependency Network for Remote Sensing Change Detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–19, 2023.

[53] L. Song, M. Xia, L. Weng, H. Lin, M. Qian, and B. Chen, "Axial Cross Attention Meets CNN: Bibranch Fusion Network for Change Detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 21–32, 2023.

[54] Y. Feng, H. Xu, J. Jiang, H. Liu, and J. Zheng, "ICIF-Net: Intra-Scale Cross-Interaction and Inter-Scale Feature Fusion Network for Bitemporal Remote Sensing Images Change Detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[55] X. Tang, T. Zhang, J. Ma, X. Zhang, F. Liu, and L. Jiao, "WNet: W-Shaped Hierarchical Network for Remote-Sensing Image Change Detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2023.

[56] Z. Li, C. Tang, X. Liu, W. Zhang, J. Dou, L. Wang, and A. Y. Zomaya, "Lightweight Remote Sensing Change Detection With Progressive Feature Aggregation and Supervised Attention," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–12, 2023.

[57] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*. Stanford, CA, USA: IEEE, Oct. 2016, pp. 565–571.

[58] H. Chen and Z. Shi, "A Spatial-Temporal Attention-Based Method and a New Dataset for Remote Sensing Image Change Detection," *Remote Sensing*, vol. 12, no. 10, p. 1662, May 2020.

[59] L. Shen, Y. Lu, H. Chen, H. Wei, D. Xie, J. Yue, R. Chen, S. Lv, and B. Jiang, "S2Looking: A Satellite Side-Looking Dataset for Building Change Detection," *Remote Sensing*, vol. 13, no. 24, p. 5094, Dec. 2021.

[60] Q. Shi, M. Liu, S. Li, X. Liu, F. Wang, and L. Zhang, "A Deeply Supervised Attention Metric-Based Network and an Open Aerial Image Dataset for Remote Sensing Change Detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2022.

[61] Z. Ying, Z. Tan, W. Li, J. Zhou, Z. Liang, and Y. Zhai, "UAV-BCD: A UAV Building Change Detection Dataset," in *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, Jul. 2023, pp. 1012–1015.

[62] W. G. C. Bandara and V. M. Patel, "A Transformer-Based Siamese Network for Change Detection," in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, Jul. 2022, pp. 207–210.

[63] P. Chen, B. Zhang, D. Hong, Z. Chen, X. Yang, and B. Li, "Fccdn: Feature constraint network for vhr image change detection," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 187, pp. 101–119, 2022.

[64] Y. Feng, J. Jiang, H. Xu, and J. Zheng, "Change Detection on Remote Sensing Images Using Dual-Branch Multilevel Intertemporal Network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–15, 2023.

[65] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An Augmented Linear Mixing Model to Address Spectral Variability for Hyperspectral Unmixing," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.

Zilu Ying received the B.S., M.S., and Ph.D. degrees in electrical information engineering from Beihang University, Beijing, China, in 1985, 1988, and 2009, respectively.

He is a Full Professor with Wuyi University, Jiangmen, China. He is an Executive Director of the Guangdong Society of Image and Graphics and a member of the Signal Processing Branch of the Chinese Institute of Electronics. His research interests include biometric extraction and pattern recognition.

Yijun Zhou received the B.S. degree from Nanchang Institute of Science and Technology, in 2023. He is currentlyworking toward the master's degree with the Schoolof Electronics and Information Engineering, WuyiUniversity. His research interests include computervision, pattern recognition, and image processing.

Yikui Zhai (Senior Member, IEEE) received his Ph.D. degree in signal and information processing from Beihang University, Beijing, China, in June 2013.

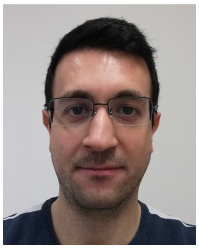
Since October 2007, he has been working with Wuyi University, Jiangmen, China, where he is a Full Professor now. He is also Associate Dean of the School of Electronics and Information Engineering in Wuyi University, since 2021. He has been a Visiting Scholar with Department of Computer Science, the Universita degli Studi di Milano, Italy, during June 2016 to June 2017, August 2023 and January 2024. His research interests include: Image Processing, Deep Learning, Optical Character Recognition, Object Detection, UAV Change Detection, Self Supervise Learning.

Hufei Zhu received the B.E. degree in electical engineering from the University of Science and Technology of China, Hefei, China, in 1998, the M.E. degree in computing from the National University of Singapore, Singapore, in 2004, and the Ph.D. degree in electrical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2014. He is currently a Research Fellow with the National Engineering Laboratory for Big Data System Computing Technology, Shenzhen University, Shenzhen, China, and a Distinguished Professor with the School of Electronics and Information Engineering, Wuyi University, Jiangmen, Guangdong, China.

His research activity has led to more than 80 patents granted in China, USA, Europe, Russia, Japan, South Korea, India, and so on, and has also led to numerous publications in leading international journals and conference proceedings. His current research interests include neural networks, signal processing, and wireless communications.

Hongsheng Zhang (Senior Member, IEEE) received the B.Eng. degree in computer science and technology and the M.Eng. degree in computer applications technology from South China Normal University, Guangzhou, China, in 2007 and 2010, respectively, and the Ph.D. degree in earth system and geoinformation science from The Chinese University of Hong Kong, Hong Kong, China, in 2013.

He is currently an Assistant Professor with the Department of Geography, The University of Hong Kong, Hong Kong, China. His research interests include remote-sensing applications in tropical and subtropical areas, with a focus on the urban environment and coastal sustainability monitoring, using multisource remote sensing data fusion and image pattern recognition techniques.



Pasquale Coscia (Senior Member, IEEE) received the Ph.D. degree in Industrial and Information Engineering from the Universita degli Studi della Campania Luigi Vanvitelli, Italy, in 2019.

From 2019 to 2022, he was a post-doctoral researcher at the Universita degli Studi di Padova, Italy. He is currently an Assistant Professor at the Department of Computer Science of the Universita degli Studi di Milano, Italy. He is Co-chair of the Intelligent Measurement Systems Technical Committee (TC-22) of the IEEE Instrumentation and Measurement Society since 2023. His research activities are focused on theoretical, methodological, and applied aspects of computational intelligence for signal and image processing.

Angelo Genovese (Senior Member, IEEE) received the Ph.D. degree in computer science from the Universita degli Studi di Milan, Crema, Italy, in 2014. He has been a postdoctoral Research Fellow in computer science with the Universita degli Studi di Milan since 2014.

He has been a Visiting Researcher with the University of Toronto, Toronto, ON, Canada. Original results have been published in over 30 papers in international journals, proceedings of international conferences, books, and book chapters. His current research interests include signal and image processing, three-dimensional reconstruction, computational intelligence technologies for biometric systems, industrial and environmental monitoring systems, and design methodologies and algorithms for self-adaptive systems.

Fabio Scotti (Senior Member, IEEE) received the Ph.D. degree in computer engineering from the Politecnico di Milano, Milan, Italy, in 2003.

He was an Assistant Professor with the Department of Information Technologies, Universita degli Studi di Milan, Milan, from 2002 to 2015, where he was an Associate Professor with the Department of Computer Science from 2015 to 2020. He has been a Full Professor with the Universita degli Studi di Milan since 2020. Original results have been published in over 150 papers in international journals, proceedings of international conferences, books, book chapters, and patents. His current research interests include biometric systems, machine learning and computational intelligence, signal and image processing, theory and applications of neural networks, 3-D reconstruction, industrial applications, intelligent measurement systems, and high-level system design.

Prof. Scotti is an Associate Editor of the IEEE Transactions on Human–Machine Systems and the IEEE Open Journal of Signal Processing. He is serving as a Book Editor (Area Editor, section Less-Constrained Biometrics) of the Encyclopedia of Cryptography, Security, and Privacy (3rd Edition, Springer). He has been an Associate Editor of the IEEE Transactions on Information Forensics and Security, Soft Computing (Springer) and a Guest Coeditor of the IEEE Transactions on Instrumentation and Measurement.

Vincenzo Piuri (Fellow, IEEE) received the M.S. and Ph.D. degrees in computer engineering from Politecnico di Milan, Milan, Italy, in 1984 and 1988, respectively.

He was the Department Chair with the University of Milan, Milan, from 2007 to 2012, where he has been a Full Professor since 2000. He was an Associate Professor with Politecnico di Milan from 1992 to 2000, a Visiting Professor with The University of Texas at Austin, Austin, TX, USA from 1996 to 1999, and a Visiting Researcher with George Mason University, Fairfax, VA, USA, from 2012 to 2016. He founded a startup company, Sensuresrl, Bergamo, Italy, in the area of intelligent systems for industrial applications (leading it from 2007 to 2010) and was active in industrial research projects with several companies. His main research and industrial application interests are intelligent systems, computational intelligence, pattern analysis and recognition, machine learning, signal and image processing, biometrics, intelligent measurement systems, industrial applications, distributed processing systems, Internet-of-Things, cloud computing, fault tolerance, application-specifi digital processing architectures, and arithmetic architectures.

Dr. Piuri is an ACM Fellow.

C. L. Philip Chen (Fellow, IEEE) received the M.S. degree in electrical engineering from the University of Michigan at Ann Arbor, Ann Arbor, MI, USA, in 1985, and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1988.

He was a tenured Professor, the Department Head, and an Associate Dean of two different universities in the U.S. for 23 years. He is currently the Head of the School of Computer Science and Engineering, South China University of Technology, Guangzhou, Guangdong, China. His current research interests include systems, cybernetics, and computational intelligence.

Dr. Chen received the 2016 Outstanding Electrical and Computer Engineers Award from his alma mater, Purdue University. He has been the Editor-in-Chief of the IEEE TRANSACTION ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS since 2014 and an associate editor of several IEEE TRANSACTIONS. He was the Chair of TC 9.1 Economic and Business Systems of International Federation of Automatic Control from 2015 to 2017 and also a Program Evaluator of the Accreditation Board of Engineering and Technology Education of the U.S. for computer engineering, electrical engineering, and software engineering programs. He was the IEEE SMC Society President from 2012 to 2013 and a Vice President of Chinese Association of Automation (CAA). He is a Fellow of AAAS, IAPR, CAA, and HKIE.