MSSM-SCDNet: A Multiclass Semantic Change Detection Network Suitable for Coastal Areas Based on Multiband Spatial-spectral Attention Mechanism

Zhen Liu, Xue Sun, Jianchen Liu, Hao Liu, Yuhang Zhou, Fazhi Cheng, Yilong Zi, and Zhen Zhang

Abstract—**Coastal change detection holds significant importance in the management of marine resources, coastal city change analysis, coastal land planning, and utilization. Deep learning-based remote sensing semantic change detection has evolved into a crucial method for identifying alterations in coastal areas. However, commonly used land cover type annotations lack specific multiclass semantic change detection type annotations unique to coastal areas. Additionally, existing datasets for coastal change detection lack rich spectral details. Therefore, this study has created a finely annotated coastal high spatial resolution multiclass semantic change detection dataset, namely CHRM-SCD, which includes 5 land cover types and 20 semantic change types. This is the first high-resolution semantic change benchmark dataset for coastal areas based on Gaofen-2 imagery. Based on this dataset, a multiclass semantic change detection network based on multiband spatial-spectral attention mechanism has been proposed in this study to achieve multiclass semantic change detection in coastal areas. It achieves 89.20% overall accuracy, 81.48% mean intersection over union, and 50.26% separated kappa coefficient, showing improvements of 7.28%, 11.58%, and 21.39% over the BiSRNet method, respectively. The stability of this research method is also demonstrated on the semantic change detection dataset. The dataset developed in this study is applicable for tasks related to detecting changes in coastal areas. The proposed method demonstrates practical effectiveness in the field of multispectral high-resolution remote sensing for coastal change detection.**

Index Terms—**Coastal, deep learning, Gaofen-2 (GF), high spatial resolution, remote sensing, semantic change detection (SCD).**

I. Introduction

The coastal zone is a transitional area between terrestrial and aquatic environments, exhibiting characteristics of both terrestrial and marine ecosystems [1]. It holds significant importance in various aspects such as ecological environment, disaster assessment [2], [3], climate change [4], [5], coastline dynamics [6], [7], land use and land cover (LULC) [8], [9]. However, due to dual interventions by human activities and natural processes, such as storm surges, land reclamation, urban expansion, fisheries and aquaculture, and salt production, the coastal zone remains in a state of dynamic flux. Consequently, the detection of coastal zone changes has emerged as a crucial research focus. Remote sensing technology, characterized by its wide coverage, high temporal frequency, and cost-effectiveness, has become a pivotal approach for monitoring coastal zone changes.

In the past few decades, there has been significant progress in remote sensing change detection (CD) techniques, particularly in the domain of land use surveys. Broadly, methods for remote sensing CD can be classified into two primary methodologies: pixel-based CD (PBCD) [10], [11] and object-based CD (OBCD) [12], [13]. PBCD involves simple arithmetic operations, such as ratios or differences, applied to pre-processed remote sensing images acquired at different time periods. While this method is relatively straightforward, the computational complexity is high, and differences between image objects are amplified due to factors like radiometric errors or differing spatial resolutions. These disparities significantly impact the accuracy of CD. Therefore, PBCD is typically suitable for CD in medium to low spatial resolution remote sensing imagery [14]. With the advent of high and even ultra-high spatial resolution imagery, remote sensing CD methods have gradually transitioned from pixel-level analysis to an object-based approach. OBCD involves classifying remote sensing images into different object types based on the similarity between pixels and subsequently detecting changes between different time periods for these object types. This method takes into consideration spatial neighborhood, shape characteristics, and texture features of objects, to some extent addressing the limitations of pixel-level CD. However, the accuracy of OBCD depends on criteria for determining pixel similarity [15], and it still has shortcomings when dealing with highly heterogeneous classes, especially in high-resolution images.

In recent years, the application of deep learning in CD for remote sensing images has become a focal point in research. Broadly, there are two main approaches of CD in optical remote sensing imagery using deep learning. One approach involves

Manuscript received 7 January 2024; revised 4 June 2024 and 24 June 2024; accepted 30 June 2024. Date of publication 3 July 2024; date of current version 13 September 2024. This work was supported in part by the Natural Science Foundation of Shandong under Grant ZR2020MD022 and in part by the Qingdao Science and Technology Demonstration and Guidance Project under Grant 22-3-7-cspz-1-nsh. *(Corresponding author: Zhen Zhang.)*

Zhen Liu, Xue Sun, Hao Liu, Yuhang Zhou, Fazhi Cheng, and Yilong Zi are with the College of Ocean Science and Engineering, Shandong University of Science and Technology, Qingdao 266000, China (e-mail: zhliu01@126.com; sundaxue1010@163.com; liuhao@liuhaoo.com; 15681500190@163.com; 15053318722@163.com; ziyilong_py@163.com).

Jianchen Liu is with the College of Geodesy and Geomatics, Shandong University of Science and Technology, Qingdao 266000, China (e-mail: liujianchen@sdust.edu.cn).

Zhen Zhang is with the Faculty of Land and Resources Engineering, Kunming University of Science and Technology, Kunming 650031, China (e-mail: zhangzhen@kust.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3422901

first classifying the images and then discriminating changes based on the classification results. This method heavily relies on the accuracy of the classification model. Additionally, when comparing two sets of classification results, prediction errors can accumulate [16]. The other approach directly conducts CD using deep learning, where deep learning techniques are employed to directly generate change results between two temporal images, leading to a noticeable improvement in accuracy. However, the majority of established methods of deep learning-based CD are binary CD (BCD) methods [17], [18], [19], [20], [21]. These methods focus on determining whether land use types have changed but do not furnish details regarding the specific characteristics of the change. Understanding the specific changes in land cover types is crucial for large-scale land cover surveys. In recent years, an increasing number of researchers have made contributions to semantic CD (SCD) [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33]. SCD involves analyzing pixel-level "from-to" changes, providing detailed information on the specific changes in land cover types.

According to the structural attributes of deep learning networks, CD models can be classified into single-branch structure, dual-branch structure, and multitask structure. Within the single-branch structure, two CD approaches can be identified. The first involves combining two temporal images through band fusion and image differencing operations, resulting in the generation of disparity images, which are then fed into a deep network for the extraction of profound change features [19]. Nonetheless, this technique could potentially introduce noise to the procedure. Another strategy within the single-branch structure involves initially classifying the two temporal images using a single-branch framework and subsequently comparing the classification outcomes to derive change results [16]. This method runs the risk of overlooking the temporal relationship between the two images. The dual-branch structure employs two separate branch networks to capture change characteristics from the two temporal images. The extracted change features are then fused, and the fused features are subsequently fed into a deep network for further extraction of useful change features until the change results are obtained [21], [26], [34]. In order to perform SCD, the multitask structure has gradually gained traction. In contrast, the multitask structure utilizes three branches for CD. Two of these branches are specialized in extracting features or semantically classifying the two temporal images, whereas the third branch is dedicated to extracting binary change features. The semantic classification results from the first two branches are then used to mask the binary change results, yielding semantic change results [24], [25]. This adoption of the multitask structure has demonstrated improved CD accuracy [16].

As of the present, numerous CD datasets have been widely employed, encompassing BCD datasets such as WHU Building [35], Season-Varying [36], Google Data Set [37], LEVIR-CD [18], and others. Single-class SCD datasets include BDD [38], xDB [39], among others. Furthermore, there are multiclass SCD datasets like Hi-UCD [27], Hi-UCD mini [40], SECOND [31], and HRSCD [16]. While the spatial resolution of these datasets continues to improve, certain limitations persist. For instance, many datasets feature samples derived from three-band remote sensing images, lacking the richness of spectral details. Due to the complexity of dataset creation and other factors, there is a relative scarcity of multiclass SCD datasets. Furthermore, these datasets predominantly focus on labeling common land cover types, often lacking annotations specific to coastal zone regions.

Regarding coastal zone CD tasks, this article makes contributions in following two main aspects.

1) A finely annotated coastal zone high spatial resolution multiclass SCD (CHRM-SCD) dataset has been established in this study. This dataset represents the first high-resolution semantic change benchmark dataset for coastal areas based on Gaofen-2 (GF-2) imagery. It provides imagery with richer spectral information (four bands) and pixel-level semantic class annotations, encompassing unique coastal zone features, including coastal fence aquaculture areas.
2) This article introduces a multiband spatial-spectral attention mechanism multiclass SCD network (MSSM-SCDNet). This network leverages the GDAL library to incorporate multiband (with the number of bands $> = 4$) image inputs and integrates a channel and spatial attention module (SAM) into the multitask CD structure based on the Bi-SRNet [23]. This approach is aimed at achieving multiclass SCD in coastal zone regions.

This article commences by providing an overview of the research progress in remote sensing CD of LULC. The rest of the article is organized as follows. Section II delineates the study area and elucidates the dataset creation process. Section III elucidates the methodology proposed in this article. Section IV delineates the experimental environment configuration, evaluation metrics, and result analysis. Section V delves into experiments regarding the selection of parameter combinations and the applicability of this method to the SEmantic Change detectiON dataset (SECOND). Finally, Section VI concludes the article.

II. Dataset

A. Study Area

The study area (see Fig. 1) is situated in the West Coast New Area of Qingdao, China. It is located in the southwestern part of Qingdao City, at the southwestern tip of the Shandong Peninsula, adjacent to Jiaozhou Bay. Geographically, it lies between approximately 35°35' to 36°08' N and 119°30' to 120°11' E. The land area covers approximately 2096 km^2, while the sea area extends to around 5000 km^2. The region measures approximately 79.25 km diagonally from northeast to southwest and spans 62.36 km from east to west. It boasts a coastline of 282 km, comprising 83 km^2 of tidal flats, 42 islands, and 23 natural harbors along the coast. West Coast New Area belongs to the hilly region of Ludong, the territory of the mountains undulating, ravines crisscrossing. The western part is dominated by the Xiaozhu Mountain Range, while the eastern region is coastal, marked by a winding and intricate coastline, numerous islands, and an abundance of estuaries. The central portion is characterized by coastal plain deposits, with a topography that gradually descends from west to east. The West Coast New Area serves as a critical outlet for the Yellow River Basin and

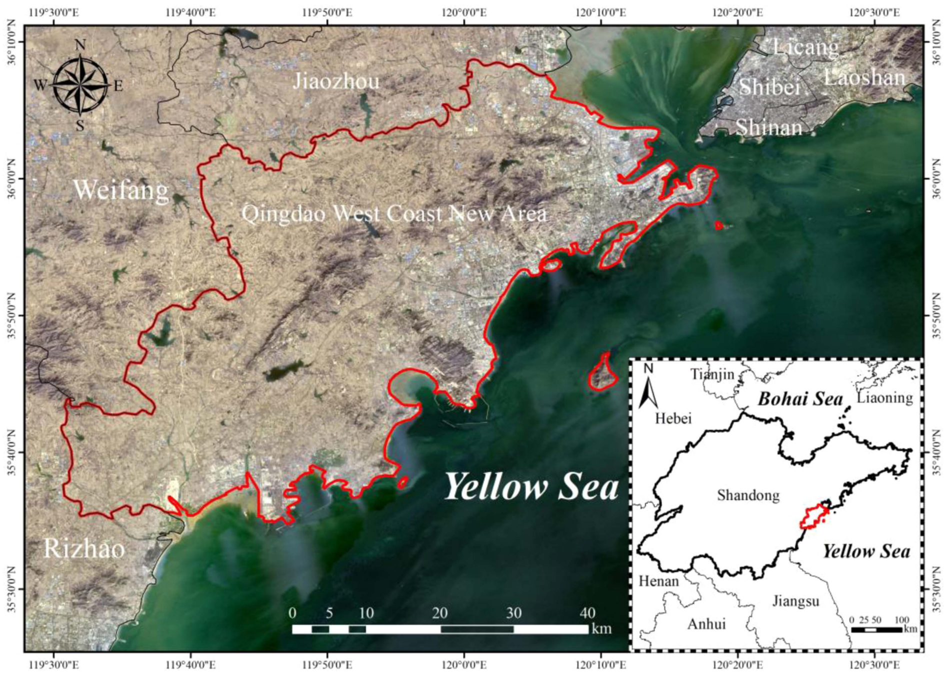


Fig. 1. Map illustrating the geographical location of the study area.

TABLE I
GF-2 Satellite Images Selected in the Study

Serial Number	Imaging Time	Scene Number	Product Number	Serial Number	Imaging Time	Scene Number	Product Number
01	2017-01-16	3236285	2125905	14	2020-12-25	8516588	5341734
02	2017-01-16	3236286	2125917	15	2020-12-25	8516492	5341551
03	2017-01-16	3236287	2125913	16	2020-12-25	8516493	5341553
04	2018-04-19	4921161	3131208	17	2021-05-07	8956270	5634471
05	2018-04-19	4921163	3131204	18	2021-10-17	9467649	5965679
06	2018-04-19	4921164	3132307	19	2021-10-17	9467792	5965197
07	2018-11-12	5746397	3593720	20	2021-12-15	9665629	6138581
08	2019-01-20	6009535	3777747	21	2021-12-29	9712919	6178525
09	2019-03-25	6237235	3903767	22	2022-01-13	9763339	6217634
10	2019-10-28	7023707	4343549	23	2022-01-13	9763662	6217796
11	2019-10-28	7023708	4343552	24	2022-06-25	10308461	6548244
12	2019-12-31	7252703	4514565	25	2022-06-25	10308462	6548243
13	2020-12-25	8516490	5341547	26	2022-11-05	10744085	6883850

represents a vital endpoint of the eastern terminus of the Eurasian Continental Bridge.

B. Dataset Introduction

The GF-2 satellite, launched on August 19, 2014, as part of China's High-Resolution Earth Observation System, is an optical remote sensing satellite. It features a spatial resolution of better than 1 m for panchromatic images and 4 m for multispectral images (including near-infrared, red, green, and blue), with a 45-km swath width. GF-2 with 5 days revisiting period provides high-quality imagery for applications in land resources management, environmental monitoring, agriculture, forestry, and disaster mitigation.

In this study, a total of 26 high-resolution GF-2 remote sensing images were utilized for the study area. These images encompass both multispectral imagery with a 4-m spatial resolution and panchromatic imagery with a 1-m spatial resolution. A fusion of the images resulted in a 1-m spatial resolution. The temporal coverage spans from 2017 to 2022, encompassing all four seasons. The images selected for this article are clear and cloudless, with good quality. Comprehensive details are available in Table I.

C. Dataset Creation Process

The annotation process (see Fig. 2) for the dataset primarily consists of five steps: image preprocessing, image registration

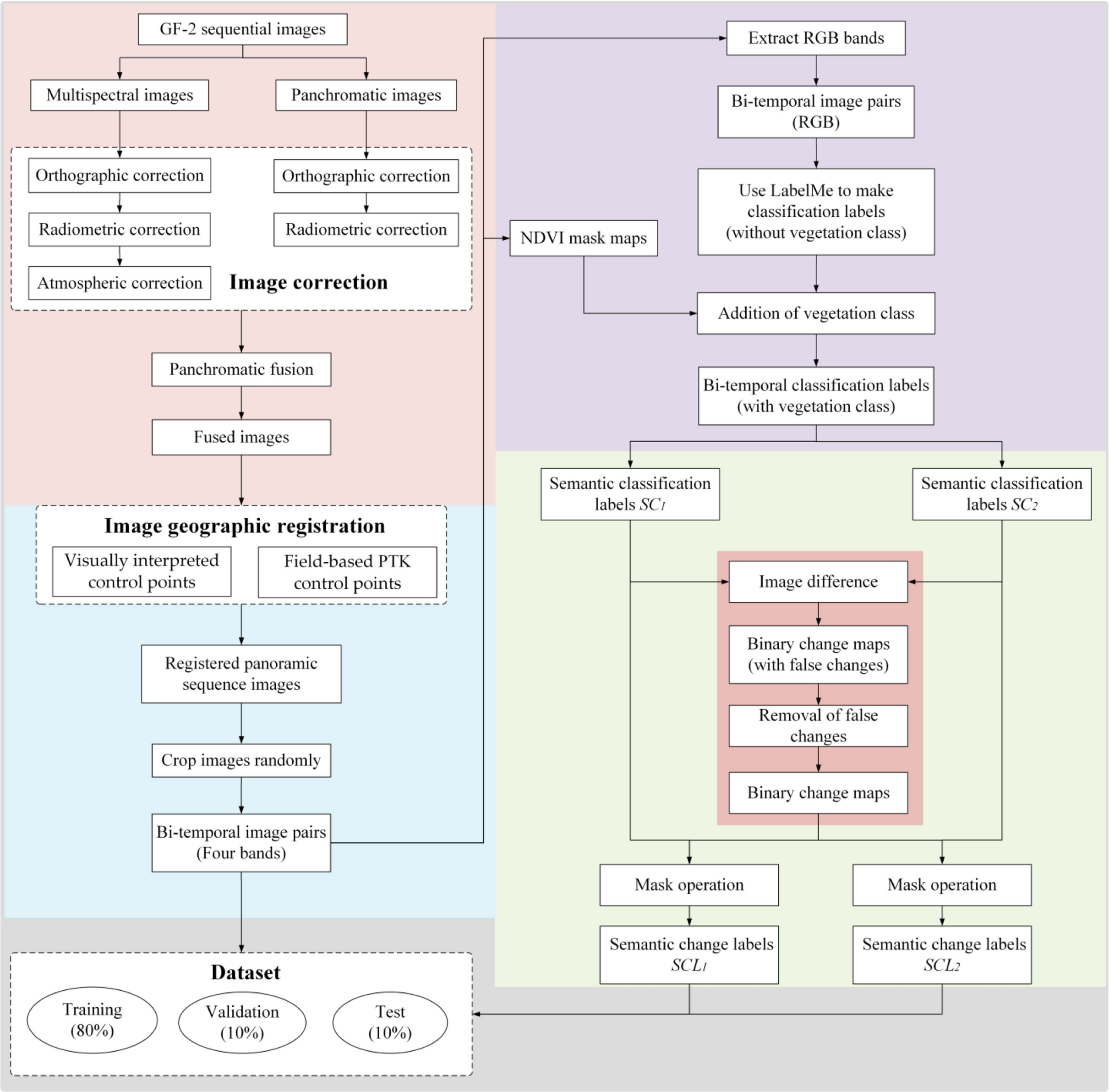

GF-2 sequential images
Multispectral images
Panchromatic images
Orthographic correction
Radiometric correction
Atmospheric correction
Orthographic correction
Radiometric correction
Image correction
Panchromatic fusion
Fused images
Image geographic registration
Visually interpreted control points
Field-based PTK control points
Registered panoramic sequence images
Crop images randomly
Bi-temporal image pairs (Four bands)
Dataset
Training (80%)
Validation (10%)
Test (10%)
Extract RGB bands
Bi-temporal image pairs (RGB)
Use LabelMe to make classification labels (without vegetation class)
NDVI mask maps
Addition of vegetation class
Bi-temporal classification labels (with vegetation class)
Semantic classification labels SC1
Semantic classification labels SC2
Image difference
Binary change maps (with false changes)
Removal of false changes
Binary change maps
Mask operation
Mask operation
Semantic change labels SCL1
Semantic change labels SCL2

Fig. 2. Annotation process for the CHRM-SCD dataset.

and cropping, sample annotation, removal of false changes, and generation of semantic change labels.

1) Image Preprocessing: The preprocessing of GF-2 remote sensing images in this study consists of three main steps: multispectral preprocessing, panchromatic preprocessing, and fusion of panchromatic and multispectral bands. Multispectral preprocessing includes orthographic correction, radiometric correction, and atmospheric correction, while panchromatic preprocessing involves orthographic and radiometric correction. The preprocessed panchromatic band and multispectral bands are fused to generate images with a spatial resolution of 1 m.

2) Image Registration and Cropping: Due to the spatial disparities that naturally occur in remote sensing images captured in different years, pixel-level remote sensing CD is adversely affected. The accuracy of image georegistration directly impacts the precision of CD results. In this study, a georegistration method for remote sensing image raster layers was employed. Six years of GF-2 images in the study area were selected for georegistration. Each pair of images involved a selection of approximately 200–400 ground control points, which included visually interpreted control points and field-based RTK control points. The goal was to ensure that the spatial registration error was within one pixel. Following the georegistration process, the images were randomly cropped without overlapping to generate 602 image pairs with a size of 512 × 512 for subsequent analysis.

3) Sample Annotation:

a) *Land cover class definition:* With reference to common land cover classes and supplemented by on-site field surveys, five land cover types are annotated, including ground, artificial objects (buildings, asphalt roads, playgrounds, and vessels), water (seawater, ponds, and lakes), coastal fence aquaculture areas, and vegetation. The detailed classification of each semantic class has been presented in Table II.

b) *Semantic classification annotation:* In order to improve the efficiency of dataset annotation, a team of seven people conducted the annotation, which took more than three months. A strict quality control strategy was developed to ensure the quality of dataset annotation. By referring to common land cover classes and conducting field surveys, the annotation criteria for each class were unified. Six people were responsible for dataset annotation, and the remaining one was responsible for checking and modifying the annotation results to ensure the uniformity of the classes in the labels. LabelMe, a professional and user-friendly pixel-level image annotation tool, was employed for annotating. Initially, the RGB bands of the cropped image pairs were extracted to generate bitemporal images in PNG format. LabelMe was then used to annotate four land cover types: ground (with vegetation), artificial objects, water, and coastal fence aquaculture areas. These four classes are represented by $C(i, j) = 1, 2, 3, 4$ respectively, where i represents the number of rows and j represents the number of columns of the image. Given the scattered distribution of vegetation, the delineation of boundaries between ground and vegetation in wooded areas posed a challenge. Moreover, there was a potential for vegetation within building shadow regions to be overlooked. Therefore, the key was to effectively separate the vegetation class from the ground class using the remote sensing method.

c) *Addition of vegetation class:* The Normalized Difference Vegetation Index (NDVI) can accurately detect vegetation class. First, calculate the NDVI values for the bitemporal images separately and generate two NDVI masks using a threshold, as depicted in (1) and (2). Then, pixel-wise comparison is made between the preannotated classification label map $C(i, j)$ and the NDVI mask map $\mathrm{Mask}(i, j)$. The "ground" ($C(i, j) = 1$) at positions where it is determined as "vegetation" ($\mathrm{Mask}(i, j) = 1$) in the mask map is changed to the vegetation class ($C(i, j) = 5$), while the class values at other pixel positions remain unchanged. From this, the vegetation class is separated from the ground class, resulting in bitemporal semantic classification labels $\mathrm{SC}(i, j)$ for five coastal land cover types, as depicted in (3)

$$\mathrm{NDVI} = \left(\mathrm{NIR} - R\right) / \left(\mathrm{NIR} + R\right) \tag{1}$$

$$\mathrm{Mask}_{1,2}\left(i, j\right) = \begin{cases} 1, & \text{vegetation} \\ 0, & \text{no vegetation} \end{cases} \tag{2}$$

where NIR represents the near-infrared band, an R corresponds to the red band; $\mathrm{Mask}_{1,2}(i, j)$ represents bitemporal NDVI mask maps

$$\mathrm{SC}_{1,2}\left(i, j\right) = \begin{cases} 5, & C_{1,2}\left(i, j\right) = 1 \text{ and } \mathrm{Mask}_{1,2}\left(i, j\right) = 1 \\ C_{1,2}\left(i, j\right), & \text{other} \end{cases} \tag{3}$$

where $C(i, j)$ represents the bitemporal semantic classes annotated using LabelMe, $\mathrm{SC}_{1,2}(i, j)$ denotes the generated bitemporal semantic classification labels. The possible values of $\mathrm{SC}_{1,2}(i, j)$ are 12,3,4,5, which represent ground, artificial object, water, coastal fence aquaculture area, and vegetation, respectively.

4) Removal of False Changes: The bitemporal classification labels are differenced to obtain the binary change map $\mathrm{BC}(i, j)$, as depicted in (4). Due to the subjective nature of image annotation, the binary change map may contain some false changes. On one hand, during the classification annotation of bitemporal images, the boundaries between different classes are not perfectly aligned, which can lead to line-shaped false changes. On the other hand, there is a phenomenon of mislabeling the same class for the same location, resulting in block-shaped false changes. Due to the presence of false changes, it can interfere with the deep neural network's ability to identify change types. Therefore, it is necessary to remove false changes from the binary change maps. This study employs two methods for this purpose: image median filtering and manual removal. First, median filtering is applied to eliminate line-shaped false changes from the maps, followed by a manual removal to eliminate block-shaped false changes. Finally, the binary change maps without false changes are obtained. This process is illustrated in steps ④ to ⑥ in Fig. 3

$$\mathrm{BC}\left(i, j\right) = \begin{cases} 1, & \mathrm{SC}_1\left(i, j\right) \neq \mathrm{SC}_2\left(i, j\right) \\ 0, & \mathrm{SC}_1\left(i, j\right) = \mathrm{SC}_2\left(i, j\right) \end{cases} \tag{4}$$

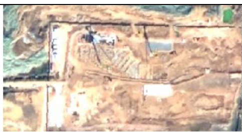
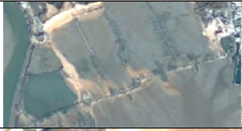
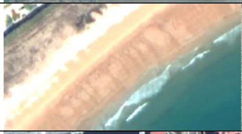
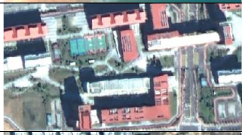
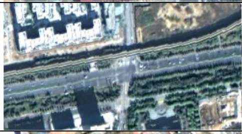
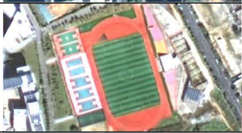
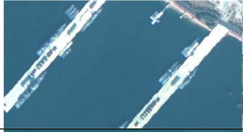
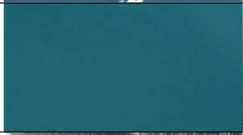
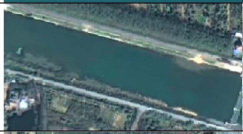
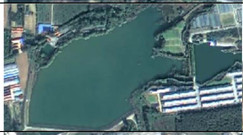
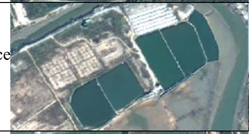
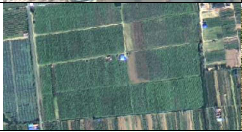
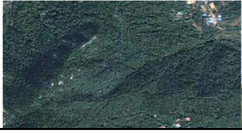
where $\mathrm{BC}(i, j)$ represents the binary change map.

5) Generation of Semantic Change Labels: The pretime classification labels are masked with the binary change maps to obtain the pretime semantic change labels $\mathrm{SCL}_1(i, j)$. Similarly, the post-time classification labels are masked with the binary change maps to obtain the post-time semantic change labels $\mathrm{SCL}_2(i, j)$, as depicted in (5). Different classes in the semantic change labels are assigned different colors, where white represents no change, gray represents the ground, red represents artificial objects, blue represents water, yellow represents coastal aquaculture areas, and green represents vegetation. The colors corresponding to each class are as depicted in (6). Finally, these 602 pairs of images and labels are randomly divided into three parts according to the ratio of 8:1:1, with the training set, validation set and test set accounting for 482 pairs, 60 pairs, and 60 pairs respectively

$$\mathrm{SCL}_{1,2}\left(i, j\right) = \begin{cases} \mathrm{SC}_{1,2}\left(i, j\right), & \mathrm{BC}\left(i, j\right) = 1 \\ 0, & \mathrm{BC}\left(i, j\right) = 0 \end{cases} \tag{5}$$

$$\mathrm{LabelColor}_{1,2}\left(i, j\right) = \begin{cases} \text{White}, & \mathrm{SCL}_{1,2}\left(i, j\right) = 0 \\ \text{Gray}, & \mathrm{SCL}_{1,2}\left(i, j\right) = 1 \\ \text{Red}, & \mathrm{SCL}_{1,2}\left(i, j\right) = 2 \\ \text{Blue}, & \mathrm{SCL}_{1,2}\left(i, j\right) = 3 \\ \text{Yellow}, & \mathrm{SCL}_{1,2}\left(i, j\right) = 4 \\ \text{Green}, & \mathrm{SCL}_{1,2}\left(i, j\right) = 5 \end{cases}. \tag{6}$$

TABLE II
LAND COVER CLASSES WITHIN CHRM-SCD DATASET

Semantic Classes	Description	GF-2 Images	Scenes
Ground	Bare ground		
	Tidal flat		
	Beach		
Artificial objects	Building		
	Asphalt road		
	Playground		
	Vessel		
Water	Seawater		
	Pond		
	Lake		
Coastal fence aquaculture areas	Aquaculture area with artificial fence by the sea.		
Vegetation	Meadow		
	Woodland		

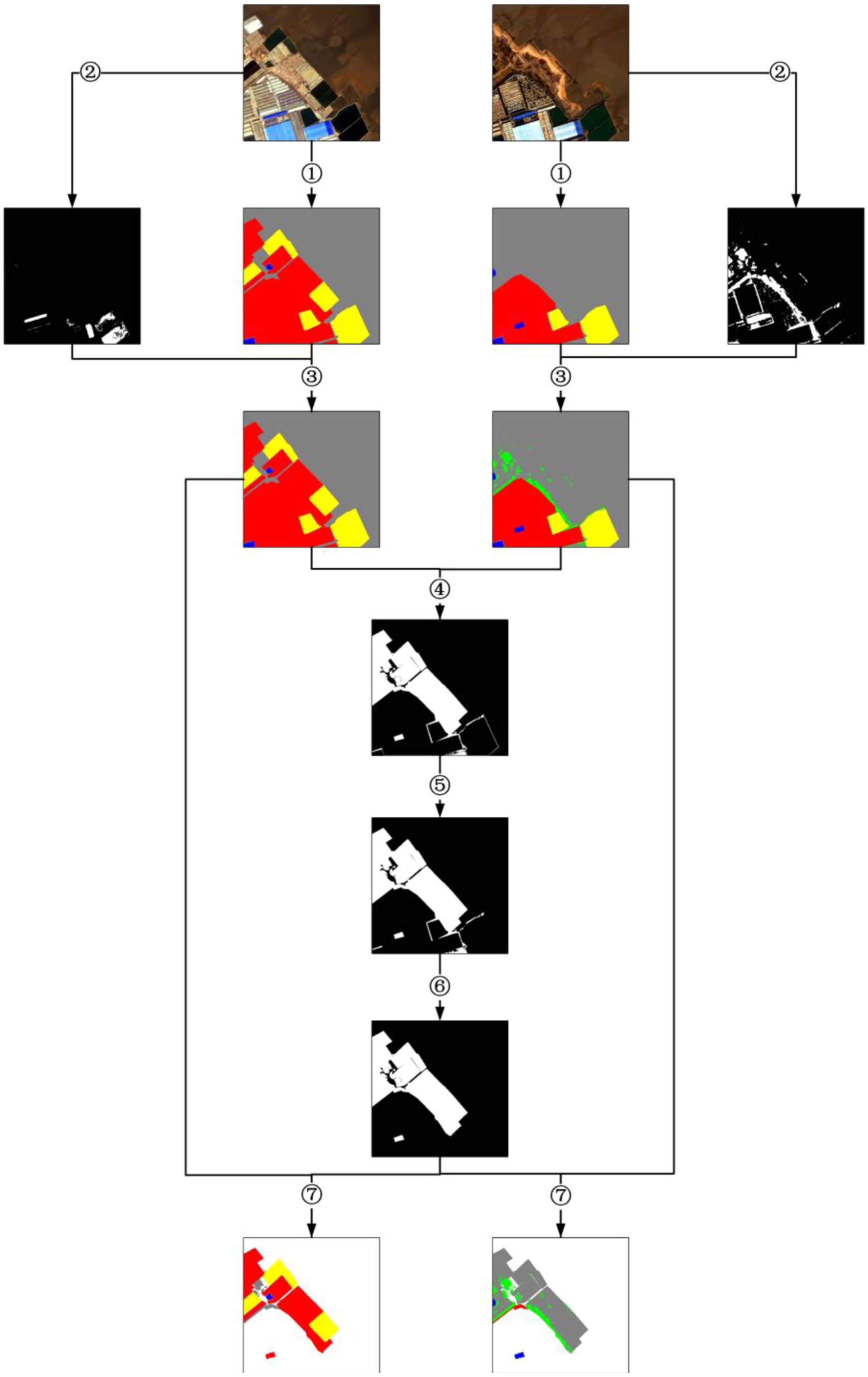


Fig. 3. Process of generating bitemporal semantic change labels: ① Annotate classes using LabelMe. ② Generate NDVI mask map. ③ Add vegetation class based on discriminant conditions. ④ Image difference. ⑤ Remove false changes through median filtering. ⑥ Manually remove false changes. ⑦ Generate semantic change labels through masking operations.

III. Method

Using the CHRM-SCD dataset as a foundation, this article extends the input data by leveraging the GDAL library to incorporate multiband (with a number of bands $> = 4$) images. Building upon the Bi-SRNet architecture, a multitask CD structure is introduced that incorporates channel and SAMs, referred to as the Convolutional Block Attention Module (CBAM) [41]. This approach aims to achieve multiclass SCD in coastal zone regions.

A. CBAM Block

CBAM combines the channel attention module (CAM) and the SAM so that each network branch can learn “which is important” on the channel axis and “where is important” on

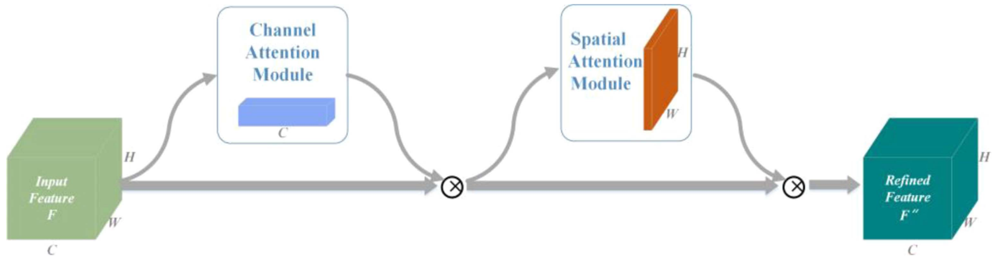


Fig. 4. CBAM block.

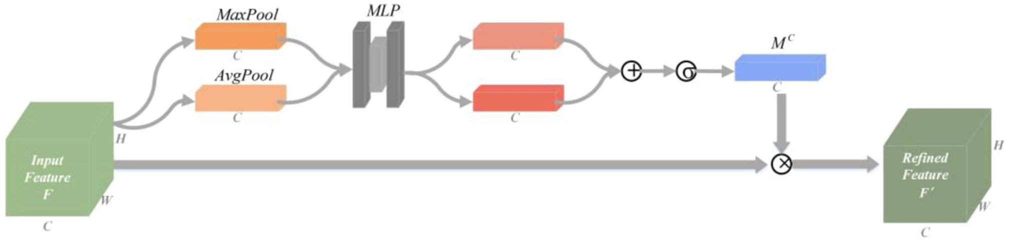


Fig. 5. CAM block.

the spatial axis. The CBAM block effectively helps information be transmitted in the network by learning which information should be emphasized or suppressed. It can achieve better results than attention modules that only focus on channels or spatial features [41]. Fig. 4 illustrates the architectural framework of CBAM. The input feature is processed sequentially through CAM and SAM. It can be seen from Fig. 4 that the configurations of the input feature F and the refined feature F'' remain unchanged. The CAM and SAM are described as follows.

1) CAM Block: Considering the impact of the difference in feature weights of different channels in the network on the coastal zone SCD results, the CAM block focuses on learning "which channels are important." The weight assigned to each channel can be interpreted as its relative importance. The CAM block captures the significance of each channel in the channel attention map. Multiplying input feature maps with corresponding weights can highlight channels relevant to changes while suppressing irrelevant channels. The implementation of CAM is shown in Fig. 5. First, the input feature maps F undergo compression through the application of both the maximum pooling (MaxPool) and average pooling (AvgPool) operations along the spatial axis. And the input feature maps with dimensions C×H×W are squeezed into two feature vectors with dimensions C×1×1 after the pooling operation. Then, these two feature vectors are pushed to a shared multilayer perceptron (MLP). After passing the shared MLP, the two vectors are summed element by element. Ultimately, a sigmoid (σ) function is adopted to assign the attention weights across all channels, and then the channel attention map M^C is obtained. At this time, the weights (between 0 and1) of all channels of input feature maps F are received. After that, these weights are multiplied with the original input feature maps F to produce the refined feature maps F'. The channel attention map M^C is calculated as follows

$$M^C = \sigma \left(\text{MLP}\left(\text{MaxPool}\left(F\right)\right) + \text{MLP}\left(\text{AvgPool}\left(F\right)\right)\right). \quad (7)$$

2) SAM Block: Similarly, considering the impact of the difference in feature weights of different positions in the network on the coastal zone SCD results, the SAM block concentrates on learning "which positions are important." The weight assigned to each position can be understood as its relative importance. The spatial attention map encodes the significance of each pixel position. The network continuously approaches the network prediction value to the ground truth (GT) by minimizing the loss function operation. The SAM can automatically adjust the weight of each pixel position after training. Higher weights are assigned to the pixel positions that have changed, whereas lower weights are assigned to the unchanged pixel positions. Multiplying the input feature maps with the corresponding spatial attention map weights, features at changed pixel positions are emphasized while features at unchanged pixel positions are suppressed. Fig. 6 illustrates the implementation of SAM. For the input feature maps F', the MaxPool operation and the AvgPool operation are performed on the channel axis of each feature point. Then, these two results are stacked ([;] indicates a stacking operation). Next, perform a convolution operation $f^{7\times7}$ with a filter with dimensions 7×7, and take a sigmoid (σ) function for creating a spatial attention map M^S. At this time, the weight (between 0 and1) of each feature point position of the input feature maps is obtained. After that, the input feature maps F' are simply multiplied by these weights to acquire the refined feature maps F''. The spatial attention map M^S can be expressed as follows:

$$M^S = \sigma\left(f^{7\times7}\left(\left[\text{MaxPool}\left(F'\right); \text{AvgPool}\left(F'\right)\right]\right)\right). \quad (8)$$

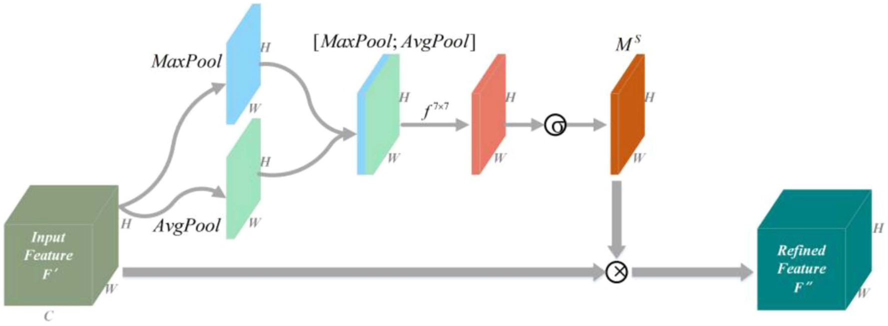


Fig. 6. SAM block.

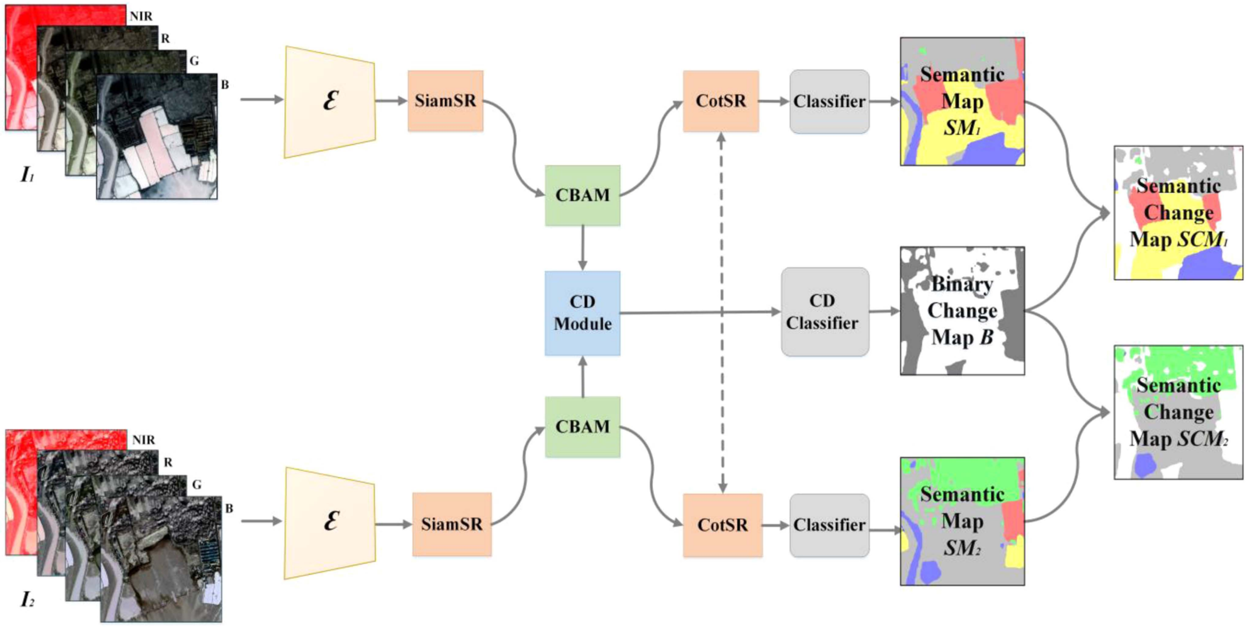


Fig. 7. Structure of the proposed MSSM-SCDNet for coastal zone SCD.

B. Coastal SCD Network MSSM-SCDNet Structure

To highlight the changed information and suppress invariant information in coastal zone, the MSSM-SCDNet is proposed in this article by integrating the attention modules. The architectural framework is illustrated in Fig. 7. Input two temporal images I_1 and I_2 to the network, MSSM-SCDNet first uses two fully convolutional network encoders ε to extract semantic features Y_1 and Y_2. Y_1 and Y_2 are subsequently processed by two siamese semantic reasoning (SiamSR) blocks, which integrate semantic focus into two temporal branches to emphasize features. The weights of the two encoders ε and the two SiamSR blocks are shared to alleviate the overfitting problem. The two emphasized features Y_1' and Y_2' are further pushed to the CBAM blocks. The generated features Y_1'' and Y_2'' are pushed to the cross temporal semantic reasoning blocks, which can learn cross-temporal semantic consistency to highlight features in unchanged regions. Next, the two output features Y_1''' and Y_2''' are mapped into the semantic maps SM_1 and SM_2 through the classifiers. Meanwhile, the CD module projects the unaligned information in Y_1'' and Y_2'' to the binary change map B through the CD classifier. The above three feature maps are all output through convolutional layers of size 1×1, and their weights are not shared. Finally, the semantic map SM_1 and the binary change map B are masked to generate the semantic change map SCM_1. The semantic map SM_2 and binary change map B are masked to produce the semantic change map SCM_2. See formulas (9)–(13) for the above realization process. Three loss functions are used in this article to optimize the MSSM-SCDNet network: semantic class loss ℓ_{sem}, binary change loss ℓ_{bc}, and semantic consistency loss ℓ_{sc} [24]. ℓ_{sem} is the multiclass cross-entropy loss calculated between semantic segmentation maps SM_1, SM_2 and semantic change labels SCM_1, SCM_2. ℓ_{bc} is the binary cross-entropy loss calculated between the binary change map B predicted by the network and the binary change label BC. ℓ_{sc} is used to correlate the loss between SM_1, SM_2 and BC

$$Y_1 = \varepsilon\left(I_1\right), Y_2 = \varepsilon\left(I_2\right) \tag{9}$$

$$Y_1' = \mathrm{SiamSR}\left(Y_1\right), Y_2' = \mathrm{SiamSR}\left(Y_2\right) \tag{10}$$

TABLE III
LIBRARY FUNCTIONS AND THEIR CORRESPONDING VERSIONS

Library Functions	**Version**
pip	22.2.2
numpy	1.21.5
scikit-image	0.18.1
opencv-contrib-python	4.6.0.66
torchvision	0.13.1
tensorboardx	2.2

$$Y_1^{''} = \mathrm{CBAM}\left(Y_1'\right), Y_2^{''} = \mathrm{CBAM}\left(Y_2'\right) \tag{11}$$

$$\mathrm{SM}_1, \mathrm{SM}_2 = \mathrm{CotSR}\left(Y_1^{''}, Y_2^{''}\right) \tag{12}$$

$$\mathrm{SCM}_1 = \mathrm{Mask}\left(B, \mathrm{SM}_1\right), \mathrm{SCM}_2 = \mathrm{Mask}\left(B, \mathrm{SM}_2\right). \tag{13}$$

IV. Experimental Results

This section outlines the environmental configuration, evaluation metrics, and presents the experimental results analysis conducted using the CHRM-SCD dataset.

A. Environmental Configuration

The deep learning model framework utilized in this study was PyTorch 1.12.1 operating on the Windows 10 platform, with Python version 3.8. Other prominent library functions are detailed in Table III. The experiments were performed using a workstation that was outfitted with a GeForce RTX 3090 Ti GPU. The optimization technique employed in this research was the stochastic gradient descent with Nesterov momentum.

B. Evaluation Metrics

Three evaluation metrics are employed to gauge the accuracy of coastal zone SCD, comprising overall accuracy (OA), mean intersection over union (mIoU), and the separated kappa (SeK) coefficient.

1) OA is a common evaluation metric for semantic segmentation and CD tasks. Express $P = \{p_{ij}\}$ (*i*, $j \in$ {0, 1 ..., N}, 0 means no change, N is the total number of change classes) as a confusion matrix, where p_{ij} indicates the overall count of pixels misclassified by the network, that is, the network predicted class is *i*, but the actual class is *j*; p_{ii} indicates the overall count of pixels that the network predicted correctly. The formula for calculating OA is as follows:

$$\mathrm{OA} = \sum_{i=0}^{N} p_{ii} / \sum_{i=0}^{N} \sum_{j=0}^{N} p_{ij}. \tag{14}$$

2) The mIoU is the standard measure of semantic segmentation. It is utilized to quantify the correlation between the actual value and the predicted value. A greater correlation corresponds to a larger mIoU. Here, mIoU is the average of the intersection over union for no change regions (IoUn) and the intersection over union for all change regions (IoUy):

$$\mathrm{mIoU} = \left(\mathrm{IoU}_n + \mathrm{IoU}_y\right)/2 \tag{15}$$

$$\mathrm{IoU}_n = p_{00} / \left(\sum_{i=0}^{N} p_{i0} + \sum_{j=0}^{N} p_{0j} - p_{00}\right) \tag{16}$$

$$\mathrm{IoU}_y = \sum_{i=1}^{N} \sum_{j=1}^{N} p_{ij} / \left(\sum_{i=0}^{N} \sum_{j=0}^{N} p_{ij} - p_{00}\right). \tag{17}$$

3) The SeK coefficient reflects the consistency between the predicted value and the real value. The larger the SeK is, the closer the predicted value is to the real value. The SeK coefficient is obtained according to $\hat{p}_{ij} = p_{ij}$ in the confusion matrix $\hat{P} = \{\hat{p}_{ij}\}$, and $\hat{p}_{00} = 0$ is excluded to eliminate the true unchanged pixels whose number occupies the majority. SeK is expressed as

$$\mathrm{SeK} = e^{IoU_y - 1} \cdot K \tag{18}$$

$$K = \left(s_0 - s_e\right) / \left(1 - s_e\right) \tag{19}$$

$$S_0 = \sum_{i=0}^{N} \hat{p}_{ii} / \sum_{i=0}^{N} \sum_{j=0}^{N} \hat{p}_{ij} \tag{20}$$

$$S_e = \sum_{i=0}^{N} \left(\sum_{j=0}^{N} \hat{p}_{ij} \times \sum_{j=0}^{N} \hat{p}_{ji}\right) / \left(\sum_{i=0}^{N} \sum_{j=0}^{N} \hat{p}_{ij}\right)^2. \tag{21}$$

C. Experimental Results Analysis

To validate the role of CBAM, as well as the CAM and SAM components, this study conducted ablation experiments. Bi-SRNet served as the baseline network, while BiSR-CAMNet involved the incorporation of CAM into Bi-SRNet, and BiSR-SAMNet incorporated SAM into the baseline. The proposed coastal zone SCD network, MSSM-SCDNet, introduced CBAM into the Bi-SRNet framework. The impact of each component on the results of coastal zone SCD is delineated in Table IV.

As evident from Table IV, networks augmented with attention modules exhibit improvements in various metrics compared to the baseline network Bi-SRNet. BiSR-CAMNet, featuring channel attention mechanisms, demonstrated notable enhancements of 3.98%, 5.87%, and 10.28% in OA, mIoU, and Sek metrics, respectively. Similarly, BiSR-SAMNet, incorporating spatial attention mechanisms, yielded remarkable gains of 5.85%, 9.21%, and 16.62% in OA, mIoU, and Sek metrics, respectively. Introducing both channel and spatial combined attention mechanisms in MSSM-SCDNet resulted in significant improvements of 7.28%, 11.58%, and 21.39% in OA, mIoU, and Sek metrics, respectively.

The results of these networks in the experiments are illustrated in Fig. 8. In group ①, neither Bi-SRNet nor BiSR-CAMNet identified changes related to the missing coastal fence aquaculture area. Although BiSR-SAMNet detected this change, the

TABLE IV
QUANTITATIVE EVALUATION RESULTS OF THE ABLATION STUDY

Methods	**Accuracy**		
	OA(%)	**mIoU(%)**	**Sek(%)**
Bi-SRNet	81.92	69.90	28.87
BiSR-CAMNet	85.90	75.77	39.15
BiSR-SAMNet	87.77	79.11	45.49
MSSM-SCDNet (proposed)	89.20	81.48	50.26

TABLE V
COMPARISON OF THE RESULTS PROVIDED BY DIFFERENT PARAMETER GROUPS

Method	**Groups**	**Parameters**			**Accuracy**		
		batch	**lr**	**epoch**	**OA(%)**	**mIoU(%)**	**Sek(%)**
MSSM-SCDNet	(1)	4	0.001	90	80.30	68.61	26.60
	(2)	4	0.01	90	87.46	78.76	44.93
	(3)	4	0.1	97	86.66	77.80	42.72
	(4)	8	0.001	91	77.62	64.22	19.98
	(5)	8	0.01	95	86.26	77.02	41.52
	(6)	8	0.1	98	89.19	81.21	49.76
	(7)	16	0.001	91	75.01	61.00	14.27
	(8)	16	0.01	92	84.12	73.96	35.79
	(9)	16	0.1	94	89.20	81.48	50.26

recognized aquaculture area was relatively small. In group ②, all four networks exhibited good performance in recognizing the large prechange coastal fence aquaculture, with MSSM-SCDNet providing the closest approximation to the ideal outcome for incomplete upper-boundary aquaculture area. The recognition of unchanged areas appeared to be more accurate with MSSM-SCDNet, as indicated by the buildings on the left. Bi-SRNet showed suboptimal performance in identifying post-change tidal flats. In group ③, for the continuous identification of asphalt roads, both Bi-SRNet and BiSR-CAMNet exhibited subpar performance, while MSSM-SCDNet demonstrated noticeable improvements. In group ④, regarding the identification of scattered vegetation and building gaps, Bi-SRNet yielded unsatisfactory results, while BiSR-CAMNet, BiSR-SAMNet, and MSSM-SCDNet successively demonstrated improved recognition. It is evident that MSSM-SCDNet excels in the identification of unchanged areas. In summary, when considering CD across various coastal land cover types, the proposed network in this article demonstrates excellent performance in SCD for diverse coastal classes.

V. DISCUSSION

A. *Comparative Experimental Analysis of Different Parameter Groups*

In order to determine the optimal parameter combination of MSSM-SCDNet for Coastal Zone SCD, this article selected three values for batch size and three values for learning rate (lr) for pairwise groups, resulting in a total of nine distinct parameter groups, denoted as [batch; lr]: [4; 0.001], [4; 0.01], [4; 0.1], [8; 0.001], [8; 0.01], [8; 0.1], [16; 0.001], [16; 0.01], and [16; 0.1]. To ensure that each set of experiments received sufficient training, no specific number of training epochs was defined. Instead, experiments were terminated when there was no improvement in the performance metrics for ten consecutive epochs. Following multiple iterations of experimentation, the best results from each group were selected for comparison. The comparative results of various performance metrics are presented in Table V. It is evident that when the batch size is 16, the lr is 0.1, and the epochs are 94, MSSM-SCDNet achieves optimal performance, as shown in group (9) in Table V. At this time, the OA, mIoU, and Sek metrics reach values of 89.20%, 81.48%, and 50.26%, respectively.

The comparative experimental results for each parameter group are illustrated in Figs. 9–12. In Fig. 9, concerning the identification of newly constructed buildings, groups (1), (4), and (7) exhibit poor performance, whereas groups (6) and (9) demonstrate significant effectiveness in distinguishing individual buildings from the surrounding ground. A potential observation has emerged, suggesting that when batch sizes remain constant, an increase in the lr may contribute to a more stable recognition of architectural details. In Fig. 10, with respect to the continuity and boundary recognition of newly paved asphalt roads, group (9) closely approximates the desired outcome, while groups (4) and (7) exhibit noticeable disconnections, and the other groups do not achieve a uniform identification of road surfaces. In Fig. 11, group (4) confuses the aquaculture area with seawater. Group (7) fails to identify changes within the aquaculture area, group (1) exhibits insufficient recognition of it, and group (3) displays redundancy in its identification. In Fig. 12, groups (1), (4), and (7) lack sensitivity in identifying scattered vegetation. Groups (2), (3), (5), (6), and (8) fall short in capturing architectural details. However, group (9) achieves the best recognition of both vegetation and buildings. It is evident

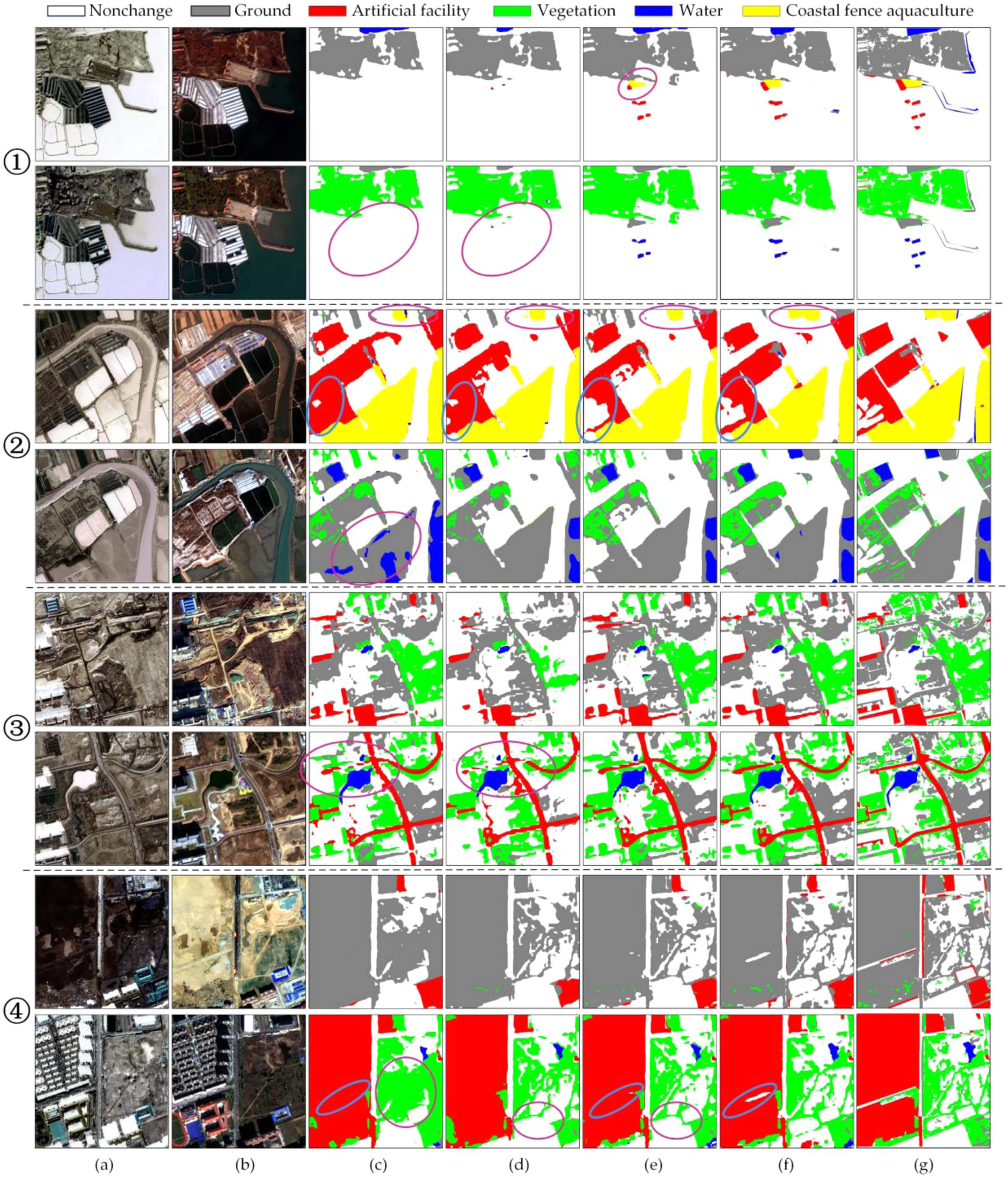


Fig. 8. (a) Bitemporal images (GF-2). (b) Bitemporal images (RGB). (c) Bi-SRNet. (d) BiSR-CAMNet. (e) BiSR-SAMNet. (f) MSSM-SCDNet (proposed). (g) GT.

that an optimal parameter group can significantly enhance the effectiveness of deep learning networks for coastal zone SCD.

B. Analysis of the Applicability of MSSM-SCDNet to the SECOND Dataset

SECOND [31] is a benchmark dataset for SCD. It is composed of bitemporal high-resolution optical images collected by some aerial platforms and sensors, including RGB channels. The spatial resolution of the images is between 0.5 and 3 m. There are a total of 4662 pairs of bitemporal images, each with the same size of 512 × 512 pixels. The dataset provides semantic change labels of bitemporal images. Each label is marked with a change category and six land cover classes, including nonchange, ground (impervious surface or bare land), playgrounds, buildings, water, trees, and low vegetation. The comparison between SECOND

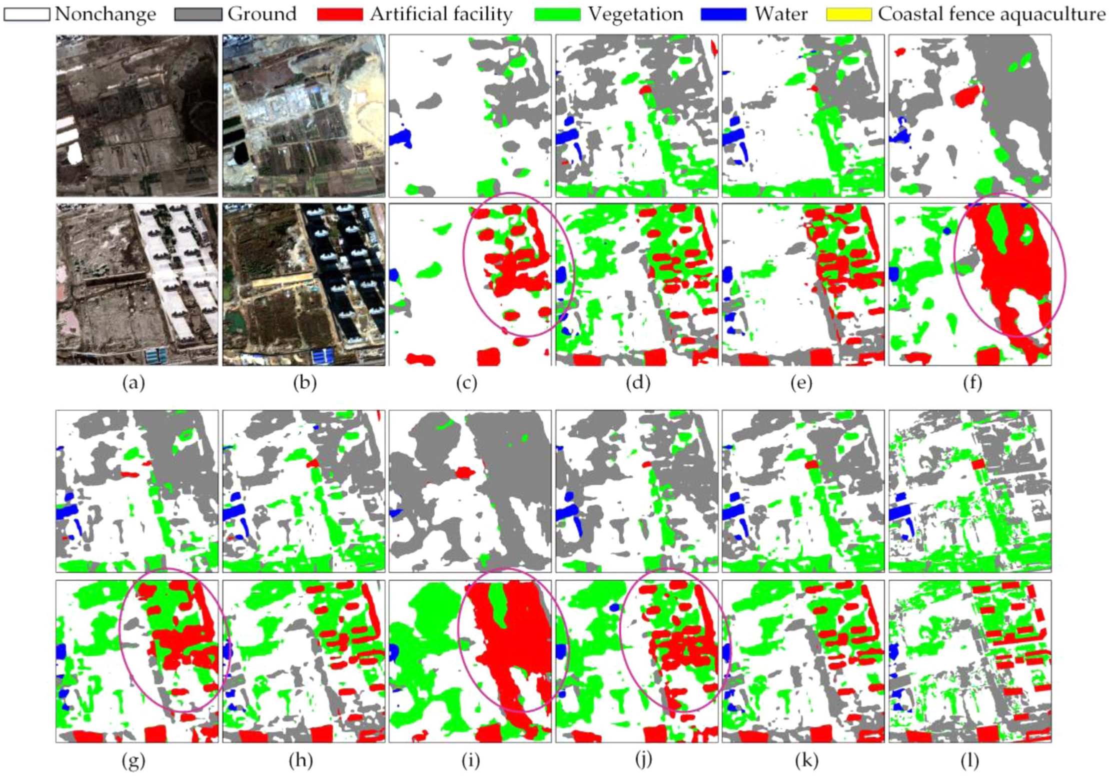


Fig. 9. (a) Bitemporal images (GF-2). (b) Bitemporal images (RGB). (c) Group (1). (d) Group (2). (e) Group (3). (f) Group (4). (g) Group (5). (h) Group (6). (i) Group (7). (j) Group (8). (k) Group (9). (l) GT.

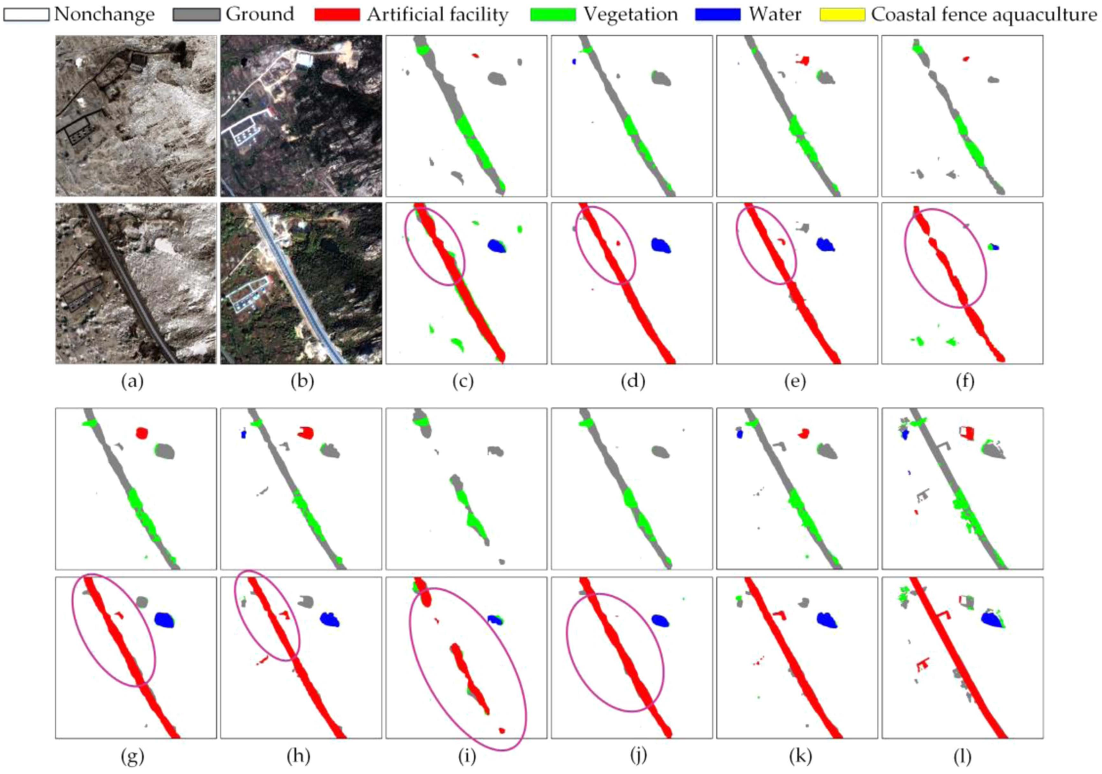


Fig. 10. (a) Bitemporal images (GF-2). (b) Bitemporal images (RGB). (c) Group (1). (d) Group (2). (e) Group (3). (f) Group (4). (g) Group (5). (h) Group (6). (i) Group (7). (j) Group (8). (k) Group (9). (l) GT.

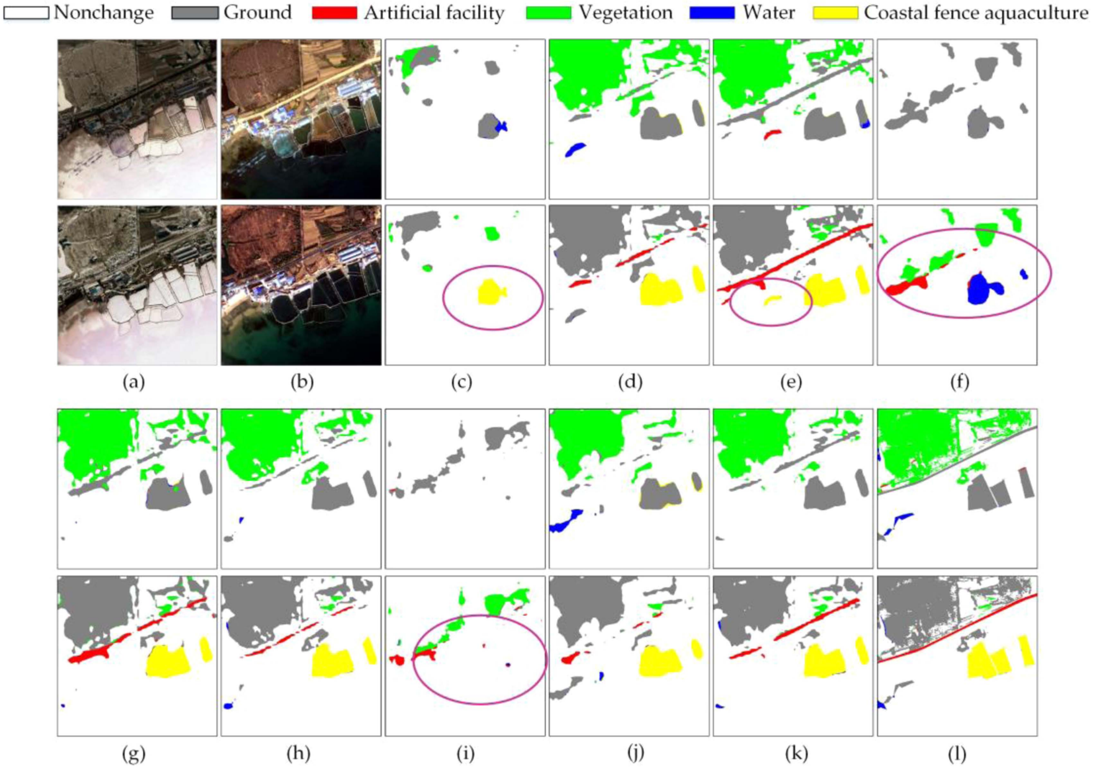


Fig. 11. (a) Bitemporal images (GF-2). (b) Bitemporal images (RGB). (c) Group (1). (d) Group (2). (e) Group (3). (f) Group (4). (g) Group (5). (h) Group (6). (i) Group (7). (j) Group (8). (k) Group (9). (l) GT.

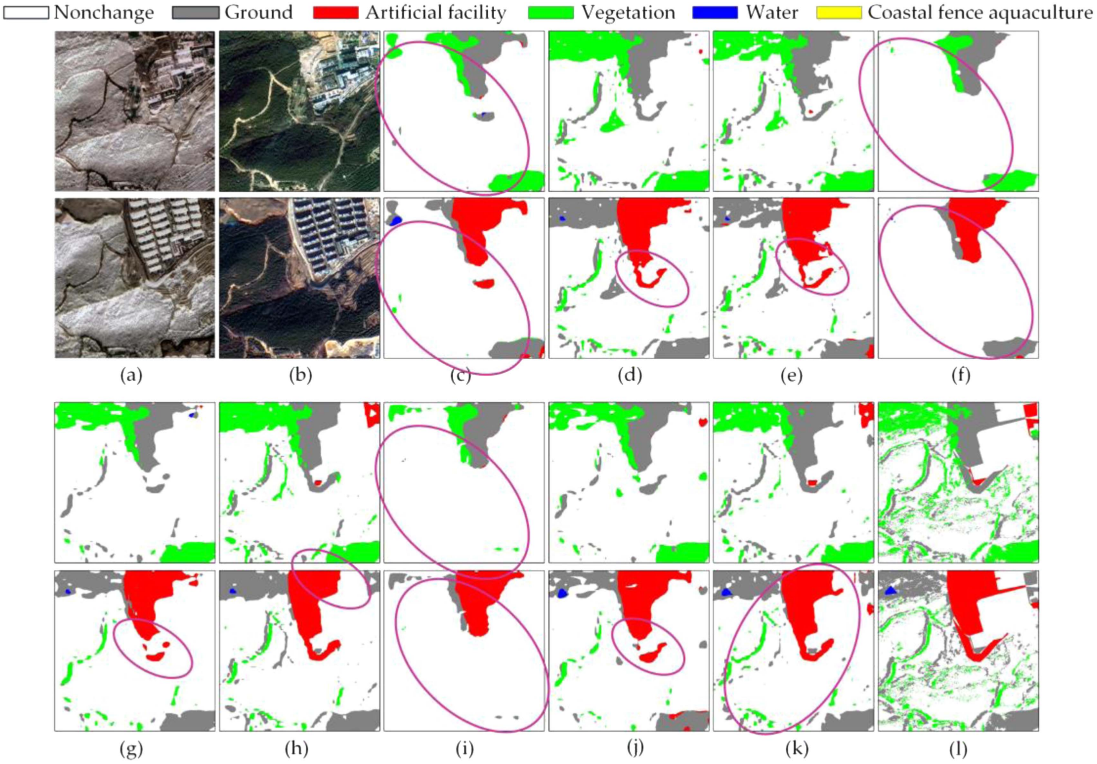


Fig. 12. (a) Bitemporal images (GF-2). (b) Bitemporal images (RGB). (c) Group (1). (d) Group (2). (e) Group (3). (f) Group (4). (g) Group (5). (h) Group (6). (i) Group (7). (j) Group (8). (k) Group (9). (l) GT.

TABLE VI
Comparison Information Between SECOND and CHRM-SCD Datasets

Dataset	Resolution	Number of Bands	Image Size	Number of Classes	Classified Information
SECOND	0.5-3m	3	512×512	6	Ground Playground Building Water Tree Low vegetation
CHRM-SCD	1m	4	512×512	5	Ground Artificial objects Water Coastal fence aquaculture Vegetation

TABLE VII
Quantitative Evaluation Results of Different Networks

Methods	Accuracy		
	OA(%)	mIoU(%)	Sek(%)
FC-EF	85.18	64.25	9.98
FC-Siam-conc	86.92	68.86	16.36
FC-Siam-diff	86.86	68.96	16.25
UNet++	85.18	63.83	9.90
ResNet-LSTM	86.77	67.16	15.96
IFN	86.47	68.45	14.25
HRSCD-str.4	86.62	71.15	18.80
MSSM-SCDNet (proposed)	87.66	72.88	21.84

and CHRM-SCD dataset constructed in this article is shown in Table VI.

To more comprehensively and objectively identify the performance of the proposed coastal SCD network MSSM-SCDNet architecture, this article further compares it with CD methods proposed by other researchers.

1) FC-EF, FC-Siam-conc, and FC-Siam-diff [42]: FC-EF is a BCD method using one encoder–decoder structure. Both FC-Siam-conc and FC-Siam-diff use the siamese encoder structure.
2) UNet++ [19]: It is an effective semantic segmentation encoder–decoder architecture. It inherits the structure of UNet and draws on the dense connection method of DenseNet.
3) ResNet-LSTM [43]: The network architecture integrates a convolutional neural network and a recurrent neural network to complete the CD task. Its encoder is changed to ResNet34 [44].
4) IFN [21]: The network uses an encoder and an attention-based decoder.
5) HRSCD-str.4 [16]: This is an SCD method that consists of residual block and triple encoder–decoder branches.

In the above methods, method 5) belongs to SCD, which can detect the change of “from-to.” Approaches 2), 3), and 4) belong to BCD. The last convolutional layer of these three networks has been modified to detect multiclass changes [23]. Table VII displays the evaluation metrics for all methods. Among them, the Sek values of FC-EF and UNet++ are low, which is probably because the networks do not process semantic information and change information separately. Both FC-Siam-conc and FC-Siam-diff use a decoder to stitch semantic features, and ResNet-LSTM takes time modeling into account, so the accuracy of all metrics is relatively high. Among the comparison methods, the two metric values (mIoU and Sek) of HRSCD-str.4 are the highest, which benefits from the strategy of BCD and land cover mapping tasks in this network as well as the skip connection between branches. The network MSSM-SCDNet proposed in this article achieved 87.66%, 72.88%, and 21.84% in OA, mIoU and Sek, respectively. The accuracy of all metrics exceeds all the above comparison methods.

To evaluate the methods more intuitively, two groups of SECOND coastal zone test data are selected for comparison, as shown in Fig. 13. It can be seen from the figure that in group ①, UNet++ and IFN fail to recognize water in the changed image. Water and low vegetation are confused. Although ResNet-LSTM identifies some water changes, it falls short in detecting complete water boundaries. Moreover, these three methods do not perform well for CD in some key areas

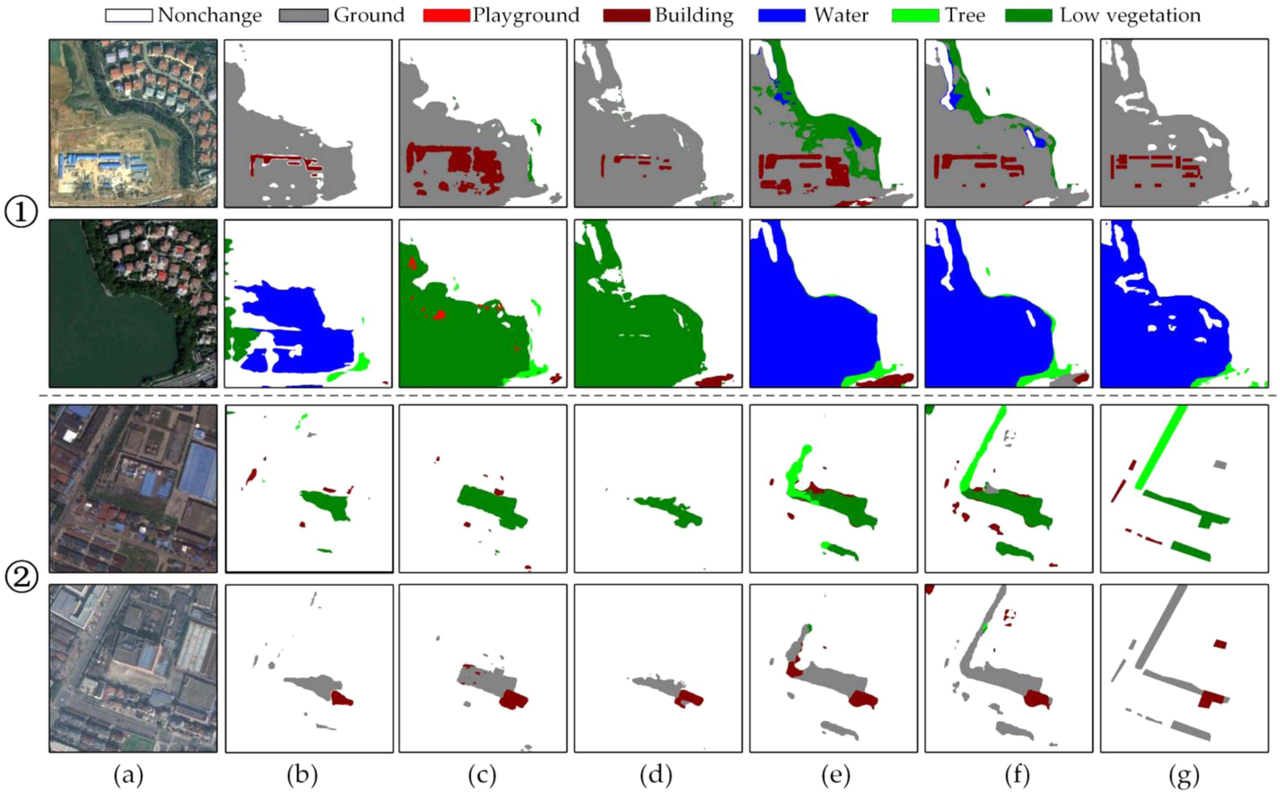


Fig. 13. (a) Bitemporal images. (b) ResNet-LSTM. (c) UNet++. (d) IFN. (e) HRSCD-str.4. (f) MSSM-SCDNet (proposed). (g) GT.

(such as trees) in group ②. In contrast, HRSCD-str.4 significantly improves on the previous three methods, effectively identifying water changes in group ①. However, it is not ideal in the boundary detection of multiple change classes (such as trees and asphalt roads) in group ②. The above changes can be well captured by MSSM-SCDNet. Moreover, the proposed method exhibits a closer alignment with the desired results in the boundary detection of coastal areas, encompassing structures like buildings, trees, and asphalt roads. This implementation is closely related to the weight assignment of CBAM to the channel and spatial position of the input feature.

VI. Conclusion

Coastal zone CD holds significant importance for monitoring coastal ecological environments, managing coastal resources, and facilitating comprehensive coastal governance. Remote sensing technology, characterized by its wide coverage, high temporal frequency, and cost-effectiveness, has emerged as a vital tool for coastal zone CD. However, existing remote sensing datasets for coastal zone CD often lack detailed spectral features. Due to the complexity involved in dataset creation, there is a relative scarcity of multiclass SCD datasets specifically tailored for coastal zones. Therefore, this study has developed a finely annotated CHRM-SCD dataset, referred to as CHRM-SCD. To our knowledge, this represents the premier high-resolution semantic change benchmark dataset for coastal zones based on GF-2 imagery. Leveraging this dataset, the present research introduces an MSSM-SCDNet, denoted as MSSM-SCDNet, designed for detecting multiclass semantic changes in coastal areas. MSSM-SCDNet achieves impressive performance with an OA of 89.20%, a mIoU of 81.48%, and a Sek index of 50.26%. These metrics represent substantial improvements of 7.28%, 11.58%, and 21.39%, respectively, over the BiSRNet baseline. The proposed approach also demonstrates robust performance on the SECOND dataset, underscoring its stability. The experimental results indicate that MSSM-SCDNet can accurately identify changes in the coastal fence aquaculture area while maintaining sensitivity to unchanged regions. Furthermore, the network exhibits closer to ideal performance in boundary detection for coastal features such as buildings, vegetation, and asphalt roads. As such, this research methodology proves highly applicable to the task of SCD in coastal zones with high spatial resolution.

Acknowledgment

The authors would like to thank China High-resolution Earth Observation System Shandong Center of Data and Application and the open-source SECOND dataset.

References

[1] N. Saintilan et al., “Widespread retreat of coastal habitat is likely at warming levels above 1.5°C,” *Nature*, vol. 621, no. 7977, pp. 112–119, Sep. 2023, doi: 10.1038/s41586-023-06448-z.

[2] X. Liu, Y. Liu, Z. Wang, X. Yang, X. Zeng, and D. Meng, “Comprehensive assessment of vulnerability to storm surges in coastal China: Towards a prefecture-level cities perspective,” *Remote Sens.*, vol. 15, no. 19, Oct. 2023, Art. no. 4828, doi: 10.3390/rs15194828.

[3] H. Xu et al., "Spatial assessment of coastal flood risk due to sea level rise in China's coastal zone through the 21st century," *Front. Mar. Sci.*, vol. 9, Aug. 2022, Art. no. 945901, doi: 10.3389/fmars.2022.945901.

[4] A. F. Arbogast, W. A. Lovis, K. G. McKeehan, and G. W. Monaghan, "A 5500-year record of coastal dune evolution along the shores of Lake Michigan in the North American Great Lakes: The relationship of lake-level fluctuations and climate," *Quaternary Sci. Rev.*, vol. 307, May 2023, Art. no. 108042, doi: 10.1016/j.quascirev.2023.108042.

[5] K. Valentine, E. R. Herbert, D. C. Walters, Y. Chen, A. J. Smith, and M. L. Kirwan, "Climate-driven tradeoffs between landscape connectivity and the maintenance of the coastal carbon sink," *Nat. Commun.*, vol. 14, no. 1, Mar. 2023, Art. no. 1137, doi: 10.1038/s41467-023-36803-7.

[6] Y. Cui, F. Yan, B. He, C. Ju, and F. Su, "Characteristics of shoreline changes around the South China Sea from 1980 to 2020," *Front. Mar. Sci.*, vol. 9, Oct. 2022, Art. no. 1005284, doi: 10.3389/fmars.2022.1005284.

[7] J. Murray, E. Adam, S. Woodborne, D. Miller, S. Xulu, and M. Evans, "Monitoring shoreline changes along the Southwestern Coast of South Africa from 1937 to 2020 using varied remote sensing data and approaches," *Remote Sens.*, vol. 15, no. 2, Jan. 2023, Art. no. 317, doi: 10.3390/rs15020317.

[8] J. Louisor, O. Brivois, P. Mouillon, A. Maspataud, P. Belz, and J.-M. Laloue, "Coastal flood modeling to explore adaptive Coastal management scenarios and land-use changes under sea level rise," *Front. Mar. Sci.*, vol. 9, May 2022, Art. no. 710086, doi: 10.3389/fmars.2022.710086.

[9] D. Yang, W. Luan, Y. Li, Z. Zhang, and C. Tian, "Multi-scenario simulation of land use and land cover based on shared socioeconomic pathways: The case of coastal special economic zones in China," *J. Environ. Manage.*, vol. 335, Jun. 2023, Art. no. 117536, doi: 10.1016/j.jenvman.2023.117536.

[10] S. B. Al Rawashdeh, "Evaluation of the differencing pixel-by-pixel change detection method in mapping irrigated areas in dry zones," *Int. J. Remote Sens.*, vol. 32, no. 8, pp. 2173–2184, Mar. 2011, doi: 10.1080/01431161003674634.

[11] R. Koller and C. Samimi, "Deforestation in the Miombo woodlands: A pixel-based semi-automated change detection method," *Int. J. Remote Sens.*, vol. 32, no. 22, pp. 7631–7649, Nov. 2011, doi: 10.1080/01431161.2010.527390.

[12] K. V. Mitkari, M. K. Arora, and R. K. Tiwari, "Detecting glacier surface changes using object-based change detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2018, pp. 5180–5183, doi: 10.1109/IGARSS.2018.8519230.

[13] X. Shi, L. Lu, S. Yang, G. Huang, and Z. Zhao, "Object-oriented change detection based on weighted polarimetric scattering differences on POLSAR images," in *The Int. Arch. Photogramm., Remote Sensing Spatial Inf. Sci.*, Jun. 2015, pp. 149–154, doi: 10.5194/isprsarchives-XL-7-W4-149-2015.

[14] A. Shafique, G. Cao, Z. Khan, M. Asad, and M. Aslam, "Deep learning-based change detection in remote sensing images: A review," *Remote Sens.*, vol. 14, no. 4, Feb. 2022, Art. no. 871, doi: 10.3390/rs14040871.

[15] K. Johansen, L. A. Arroyo, S. Phinn, and C. Witte, "Comparison of geo-object based and pixel-based change detection of riparian environments using high spatial resolution multi-spectral imagery," *Photogramm. Eng. Remote Sens.*, vol. 76, no. 2, pp. 123–136, Feb. 2010, doi: 10.14358/PERS.76.2.123.

[16] R. Caye Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, "Multitask learning for large-scale semantic change detection," *Comput. Vis. Image Understanding*, vol. 187, Oct. 2019, Art. no. 102783, doi: 10.1016/j.cviu.2019.07.003.

[17] H. Chen, Z. Qi, and Z. Shi, "Remote sensing image change detection with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, Sep. 2022, doi: 10.1109/TGRS.2021.3095166.

[18] H. Chen and Z. Shi, "A spatial-temporal attention-based method and a new dataset for remote sensing image change detection," *Remote Sens.*, vol. 12, no. 10, May 2020, Art. no. 1662, doi: 10.3390/rs12101662.

[19] D. Peng, Y. Zhang, and H. Guan, "End-to-End change detection for high resolution satellite images using improved UNet++," *Remote Sens.*, vol. 11, no. 11, Jun. 2019, Art. no. 1382, doi: 10.3390/rs11111382.

[20] 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 Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, Jun. 2022, doi: 10.1109/TGRS.2021.3085870.

[21] C. Zhang et al., "A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 166, pp. 183–200, Aug. 2020, doi: 10.1016/j.isprsjprs.2020.06.003.

[22] F. Cui and J. Jiang, "MTSCD-Net: A network based on multi-task learning for semantic change detection of bitemporal remote sensing images," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 118, Apr. 2023, Art. no. 103294, doi: 10.1016/j.jag.2023.103294.

[23] L. Ding, H. Guo, S. Liu, L. Mou, J. Zhang, and L. Bruzzone, "Bi-temporal semantic reasoning for the semantic change detection in HR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, Feb. 2022, doi: 10.1109/TGRS.2022.3154390.

[24] L. Ding, J. Zhang, K. Zhang, H. Guo, B. Liu, and L. Bruzzone, "Joint Spatio-temporal modeling for the semantic change detection in remote sensing images," Dec. 2022, Accessed: Dec. 30, 2022. [Online]. Available: http://arxiv.org/abs/2212.05245

[25] Y. Niu, H. Guo, J. Lu, L. Ding, and D. Yu, "SMNet: Symmetric multi-task network for semantic change detection in Remote sensing images based on CNN and Transformer," *Remote Sens.*, vol. 15, no. 4, Feb. 2023, Art. no. 949, doi: 10.3390/rs15040949.

[26] D. Peng, L. Bruzzone, Y. Zhang, H. Guan, and P. He, "SCDNET: A novel convolutional network for semantic change detection in high resolution optical remote sensing imagery," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 103, Dec. 2021, Art. no. 102465, doi: 10.1016/j.jag.2021.102465.

[27] S. Tian, Y. Zhong, Z. Zheng, A. Ma, X. Tan, and L. Zhang, "Large-scale deep learning based binary and semantic change detection in ultra high resolution remote sensing imagery: From benchmark datasets to urban application," *ISPRS J. Photogramm. Remote Sens.*, vol. 193, pp. 164–186, Nov. 2022, doi: 10.1016/j.isprsjprs.2022.08.012.

[28] X.-S. Wei, Y.-Y. Xu, C.-L. Zhang, G.-S. Xia, and Y.-X. Peng, "CAT: A coarse-to-fine attention tree for semantic change detection," *Vis. Intell.*, vol. 1, no. 1, May 2023, Art. no. 3, doi: 10.1007/s44267-023-00004-z.

[29] H. Xia, Y. Tian, L. Zhang, and S. Li, "A deep Siamese postclassification fusion network for semantic change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, Apr. 2022, doi: 10.1109/TGRS.2022.3171067.

[30] S. Xiang, M. Wang, X. Jiang, G. Xie, Z. Zhang, and P. Tang, "Dual-task semantic change detection for remote sensing images using the generative change field module," *Remote Sens.*, vol. 13, no. 16, Aug. 2021, Art. no. 3336, doi: 10.3390/rs13163336.

[31] K. Yang et al., "Asymmetric Siamese networks for semantic change detection in aerial images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–18, Oct. 2021, doi: 10.1109/TGRS.2021.3113912.

[32] M. Zhao et al., "Spatially and semantically enhanced Siamese network for semantic change detection in high-resolution remote sensing images," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 2563–2573, Mar. 2022, doi: 10.1109/JSTARS.2022.3159528.

[33] Z. Zheng, Y. Zhong, S. Tian, A. Ma, and L. Zhang, "ChangeMask: Deep multi-task encoder-transformer-decoder architecture for semantic change detection," *ISPRS J. Photogramm. Remote Sens.*, vol. 183, pp. 228–239, Jan. 2022, doi: 10.1016/j.isprsjprs.2021.10.015.

[34] H. Chen, C. Wu, B. Du, L. Zhang, and L. Wang, "Change detection in multisource VHR images via deep Siamese convolutional multiple-layers recurrent neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2848–2864, Apr. 2020, doi: 10.1109/TGRS.2019.2956756.

[35] S. Ji, S. Wei, and M. Lu, "Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 574–586, Jan. 2019, doi: 10.1109/TGRS.2018.2858817.

[36] M. A. Lebedev, Y. V. Vizilter, O. V. Vygolov, V. A. Knyaz, and A. Y. Rubis, "Change detection in remote sensing images using conditional adversarial networks," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. XLII–2, pp. 565–571, May 2018, doi: 10.5194/isprs-archives-XLII-2-565-2018.

[37] D. Peng, L. Bruzzone, Y. Zhang, H. Guan, H. Ding, and X. Huang, "SemiCDNet: A semisupervised convolutional neural network for change detection in high resolution remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5891–5906, Jul. 2021, doi: 10.1109/TGRS.2020.3011913.

[38] B. Adriano et al., "Learning from multimodal and multitemporal earth observation data for building damage mapping," *ISPRS J. Photogramm. Remote Sens.*, vol. 175, pp. 132–143, May 2021, doi: 10.1016/j.isprsjprs.2021.02.016.

[39] R. Gupta et al., "xBD: A dataset for assessing building damage from satellite imagery," Nov. 2019. Accessed: Sep. 26, 2023. [Online]. Available: http://arxiv.org/abs/1911.09296

[40] S. Tian, A. Ma, Z. Zheng, and Y. Zhong, "Hi-UCD: A large-scale dataset for urban semantic change detection in remote sensing imagery," Dec. 2020. Accessed: Sep. 26, 2023. [Online]. Available: http://arxiv.org/abs/2011.03247

[41] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “CBAM: Convolutional block attention module,” Jul. 2018. Accessed: Dec. 07, 2022. [Online]. Available: http://arxiv.org/abs/1807.06521

[42] R. C. Daudt, B. L. Saux, and A. Boulch, “Fully convolutional Siamese networks for change detection,” Oct. 2018. doi: 10.48550/arXiv.1810.08462.

[43] L. Mou, L. Bruzzone, and X. X. Zhu, “Learning spectral-spatialtemporal features via a recurrent convolutional neural network for change detection in multispectral imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 924–935, 2018.

[44] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

Zhen Liu received the Ph.D. degree in marine information detection and processing from the Ocean University of China, Qingdao, China, in 2013.

He is currently an Associate Professor with the College of Ocean Science and Engineering, Shandong University of Science and Technology, Qingdao. His research interests include remote sensing of bathymetry and coastal zone.

Xue Sun received the B.E. degree in digital media technology from the College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, China, in 2019. She is currently working toward the M.S. degree in electronic information with the College of Ocean Science and Engineering, Shandong University of Science and Technology, Qingdao.

Her research interests include remote sensing and deep learning.

Jianchen Liu was born in Jiamusi, China, in 1987. He received the M.S. degree in geomatics engineering from the Shandong University of Science and Technology, Qingdao, China, in 2013, and the Ph.D. degree in photogrammetry and remote sensing from the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China, in 2017.

He is currently an Associate Professor with the College of Geodesy and Geomatics, Shandong University of Science and Technology. His research interests include unmanned aerial vehicle photogrammetry, computer stereovision, and 3-D modeling by oblique images.

Hao Liu received the B.S. degree in remote sensing science and technology in 2021 from the College of Ocean Science and Engineering, Shandong University of Science and Technology, Qingdao, China, where he is currently working toward the M.S. degree in electronic information with the College of Ocean Science and Engineering.

His current research interests include remote sensing and deep learning.

Yuhang Zhou received the B.E. degree in remote sensing science and technology from the College of Earth and Science, Chengdu University of Technology, Chengdu, China, in 2023. He is currently working toward the M.S. degree in naval architecture and ocean engineering with the College of Ocean Science and Engineering, Shandong University of Science and Technolog, Qingdao, China.

His current research interests include hyperspectral remote sensing and LiDAR.

Fazhi Cheng received the B.E. degree in marine technology in 2023 from the College of Ocean Science and Engineering, Shandong University of Science and Technology, Qingdao, China, where he is currently working toward the M.S. degree in electronic information.

His current research interests include remote sensing and LiDAR.

Yilong Zi received the B.E. degree in intelligent science and technology in 2023 from the College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, China, where he is currently working toward the M.S. degree in naval architecture and ocean engineering.

His current research interests include lidar point cloud processing and deep learning.

Zhen Zhang received the B.S. degree in remote sensing science and technology and the M.S. and Ph.D. degrees in photogrammetry and remote sensing from the Shandong University of Science and Technology, Qingdao, China, in 2012, 2016, and 2021, respectively.

He is currently a instructor with the Kunming University of Science and Technology. His research interests include hyperspectral remote sensing and wetland remote sensing.