Scientific Research Publishing

Journal of Computer and Communications, 2025, 13(4), 265-288
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

# Research on the Application of VMamba Scanning Algorithm in High-Resolution Remote Sensing Change Detection

**Rui Shi, Zhenchuan Wang**

School of Computer Science and Engineering, Sichuan University of Science & Engineering, Yibin, China
Email: shirui@suse.edu.cn, 1010568787@qq.com

**How to cite this paper:** Shi, R. and Wang, Z.C. (2025) Research on the Application of VMamba Scanning Algorithm in High-Resolution Remote Sensing Change Detection. *Journal of Computer and Communications*, **13**, 265-288.
https://doi.org/10.4236/jcc.2025.134017

**Received:** March 24, 2025
**Accepted:** April 26, 2025
**Published:** April 29, 2025



## Abstract

The VMamba (Visual State Space Model) is built upon the Mamba model by stacking Visual State Space (VSS) modules and utilizing the 2D Selective Scan (SS2D) module to extend the original Mamba model's capability from handling one-dimensional sequences to two-dimensional sequences. This enhancement broadens the application of the Mamba model to visual tasks. Compared to CNNs and Transformers, Mamba retains two significant advantages: long-sequence modeling and linear complexity, making it well-suited for high-resolution image tasks. While previous studies have explored its application in high-resolution remote sensing image processing, challenges such as high computational cost and slow training speed persist. The core issue arises from multiple sequence scans and the merging process after sequence processing, which slows down model training. This paper investigates the sequence scanning process and proposes multiple scanning algorithms. Specifically, we employ a unidirectional sequence scanning algorithm in high-resolution remote sensing change detection to reduce the number of scans in the scanning module, thereby accelerating model training. By evaluating its performance in classification and object detection tasks, we thoroughly test the feature extraction capabilities of these scanning algorithms in the VMamba model. Through comparative experiments in high-resolution remote sensing change detection, we demonstrate that our proposed unidirectional scanning algorithm achieves comparable or even superior performance with higher computational efficiency compared to omnidirectional scanning algorithms. Experimental results further suggest a potential correlation between the SS2D algorithm's feature extraction capability and its performance in remote sensing change detection. This study provides valuable insights for further research on Mamba-based remote sensing change detection algorithms.

**Keywords**

VMamba Algorithm, Remote Sensing Images, Change Detection, Mamba Network

## 1. Introduction

Bitemporal remote sensing change detection refers to the process of detecting and identifying changes in land cover by analyzing images of the same geographical area captured at different times by satellites, UAVs, or other remote sensing platforms. This includes binary classification tasks for detecting change areas and multi-class classification tasks for identifying change types [1]. As an essential method for monitoring surface changes, remote sensing change detection is widely applied in environmental monitoring, disaster response, urban planning, and other fields, providing critical decision-making support for environmental protection, resource management, urban development, strategic security, and land resource management.

Mamba [2] is a novel deep learning architecture based on the State Space Model (SSM), initially designed for natural language processing. The standard Mamba model operates on one-dimensional sequences. To extend its capability to visual tasks, Liu *et al.* [3] proposed VMamba, whose core component is the Visual State Space (VSS) module, composed of the 2D Selective Scan (SS2D) algorithm. VMamba integrates the advantages of CNNs [4] and Transformers [5], enabling both local feature extraction and global modeling with high computational efficiency. Compared to Transformers, VMamba has lower computational complexity, making it highly suitable for large-scale remote sensing data processing, especially for multitemporal imagery.

In current VMamba-based remote sensing change detection studies, the SS2D module typically adopts four-directional or omnidirectional sequence scanning methods. Although these methods theoretically enhance feature extraction, they also significantly increase the number of sequence scans and merging operations. This results in excessive memory consumption when processing image data and decreases model efficiency due to frequent sequence reads and merges.

To address these challenges, we propose multiple additional scanning algorithms and conduct a detailed performance study. By evaluating their feature extraction capabilities in classification and object detection tasks, we identify algorithms that balance strong feature extraction performance with computational efficiency. We implement these optimized algorithms within the RS-Mamba [6] benchmark model and test them using two public remote sensing change detection datasets: LEVIR-CD [7] and WHU-CD [8]. Our findings indicate that, compared to the omnidirectional scanning algorithm in RS-Mamba, the proposed unidirectional scanning algorithm achieves better efficiency. This discovery provides valuable insights for future research on VMamba applications in high-resolution

remote sensing change detection.

## 2. Relate Work

### 2.1. Mamba and VMamba

The core idea of the Mamba network is to process sequence data by introducing the State Space Model (SSM). The State Space Model is a mathematical model used to describe dynamic systems, typically applied in control theory and signal processing. Mamba combines SSM with deep learning, proposing a new approach for sequence modeling.

SSM is a mathematical model for describing dynamic systems and is widely used in fields such as signal processing, control systems, time series analysis, and deep learning. It represents the system's dynamic evolution through "state variables" and uses observation variables to describe measurable data. Its structure is shown in **Figure 1**.

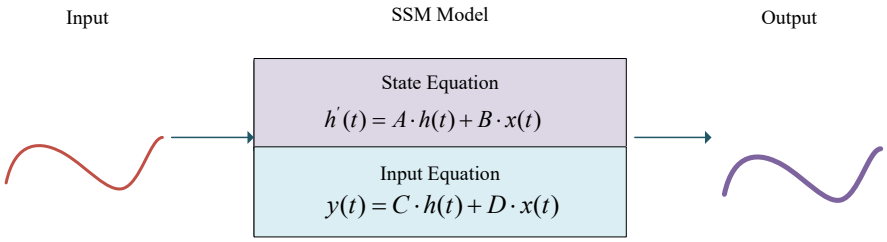


**Figure 1.** Algorithm flow of state-space model.

The State Space Model typically consists of two equations: the state equation, which describes how the system's state evolves over time, and the observation equation, which describes how observations are generated from the state. Let $h'(t)$ be the hidden system state and $y(t)$ be the observation data. The State Space Model can be described by **Equation (1)** and **Equation (2)**.

$$h'(t) = A \cdot h(t) + B \cdot x(t) \tag{1}$$

$$y(t) = C \cdot h(t) + D \cdot x(t) \tag{2}$$

Where $x(t)$ is the optional control input, $h(t)$ is the current system state, and *A*,B, C, and D are fixed parameter matrices that control the historical state, input information, output information, and state transition. Since this model describes the state changes of a system in continuous space over time, to apply it in real-world sequence modeling, Gu *et al.* [9] introduced the Structured State Space Sequence (S4) model.

The S4 model addresses the high computational and memory demands of traditional SSMs when handling long sequence data by introducing a new parameterization method. Instead of directly computing the SSM convolution (which is very time-consuming for long sequences), S4 calculates the values of its truncated generating function at unit roots, and then uses the inverse Fast Fourier Transform (FFT) to obtain the SSM's convolution kernel. This method reduces the

computational complexity from $O(N^2L)$ to $\tilde{O}(N+L)$, where $N$ is the sequence length and $L$ is the state dimension.

S5 (2023) [10] is an improved version of S4. S5 enhances the frequency domain computation method, reducing the overhead of generating function calculation and FFT transformation, further lowering the computational complexity. It also introduces sparsification techniques to reduce the number of non-zero elements in the state matrix A, thus lowering the computation cost of matrix multiplication. The specific HiPPO matrix used for initialization in S4 could not be diagonally processed in a numerically stable manner, so in S5, the state matrix is diagonalized, and an approximation of the diagonalized matrix is used to achieve HiPPO [11] initialization, maintaining computational efficiency while achieving good performance.

S6 (2024) [2] is a further improvement of S5, aiming to fully address the instability of SSM in deep learning training and provide stronger capabilities than the Transformer. Its computational diagram is shown in **Figure 2**. S6 introduces a selective mechanism that allows the model's parameters (such as Δ, *B*, and C) to be dynamically adjusted according to the input. This mechanism enables the model to selectively propagate or forget information based on the current input, making it better suited for handling discrete modalities (such as text) and tasks requiring content-dependent reasoning. S6 also incorporates hardware-aware computational optimization techniques such as Parallel Scan and Activation Recomputation. These techniques allow S6 to run efficiently on modern hardware such as GPUs, significantly improving computational efficiency. The computational complexity of S6 is linear with respect to the sequence length O(N), much lower than the quadratic complexity of the Transformer. S6 is a core component of the Mamba architecture, and the Mamba network consists of multiple stacked Mamba basic modules. The structure of the basic Mamba block is shown in **Figure 3**.

The standard Mamba design is intended for one-dimensional sequences. To handle visual tasks, Liu *et al.* [3] introduced VMamba. The core of VMamba is the visual state space (VSS) module, which consists of a 2D selective scanning algorithm (SS2D). SS2D is a four-direction scanning mechanism specifically designed for the spatial domain. The SS2D scanning process is shown in **Figure 4** and **Figure 5** shows

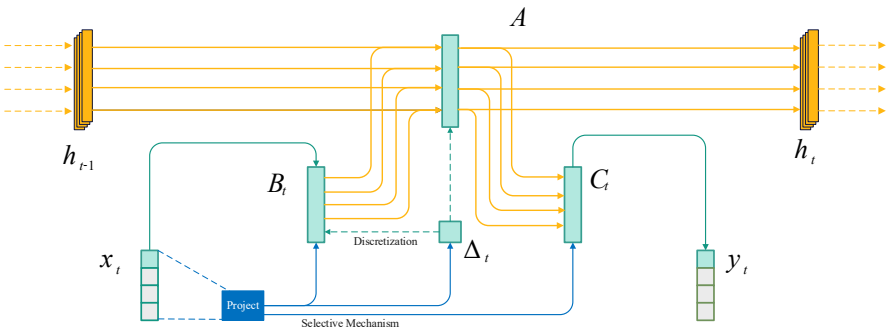


**Figure 2.** Algorithm flow of s6 selective state-space model.

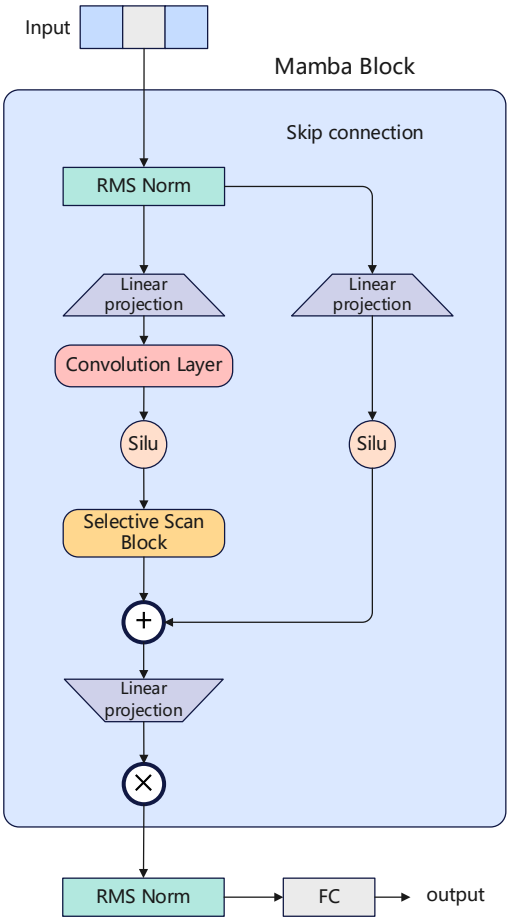


**Figure 3.** Mamba block architecture [2].

the VSS block structure. Its effectiveness in image feature extraction has been demonstrated. The Mamba network combines the advantages of CNNs and Transformers, allowing for global modeling of data while effectively addressing long-range dependencies, leading to better performance in image-related tasks.

RS-Mamba and CD-Mamba [12] are representative applications of Mamba in remote sensing change detection. In these models, omnidirectional and four-direction SS2D scanning blocks are used as feature encoders in feature extraction, respectively. They also use the same multi-scale Siamese network architecture. RS-Mamba attempts to replace the decoder with a CNN fully connected network to reduce computational costs.

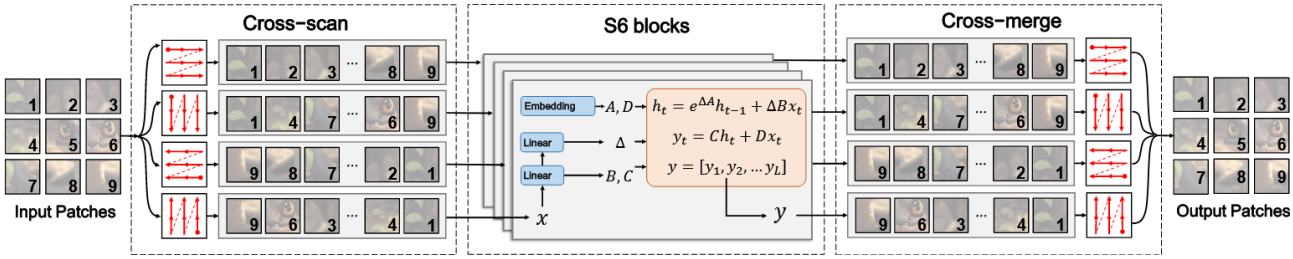


**Figure 4.** SS2D scanning process [3].

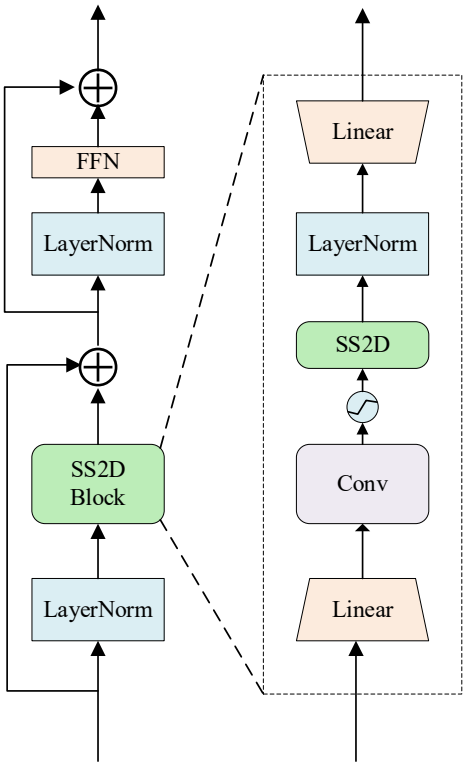


Figure 5. VSS block structure.

## 2.2. Evaluation Metrics

In our experiments, which include image classification, object detection, and change detection, the evaluation metrics include Precision (P), Accuracy (A), Overall Accuracy (OA), Recall (R), F1-Score (F1), Intersection over Union (IoU), and Kappa Coefficient [13].

To obtain the above metrics, the experiment introduces the confusion matrix. The confusion matrix is a table used to evaluate the performance of classification models, especially suitable for binary and multi-class problems. It shows the relationship between the model's predicted results and the actual labels, helping us intuitively understand the classifier's performance. The confusion matrix serves as the basis for calculating many classification evaluation metrics, including Accuracy, Precision, Recall, F1-Score, and Kappa Coefficient. The confusion matrix and related definitions are shown in Table 1.

Table 1 Confusion matrix structure.

| | | Prediction | |
|---|---|---|---|
| | | Positive | Negative |
| Ground Truth | Positive | True Positive(TP) | False Negative(FN) |
| | Negative | False Positive(FP) | True Negative(TN) |

Precision measures the proportion of true positive samples among the samples

predicted as positive by the model. The higher the precision, the fewer the false positives, indicating the model's ability to correctly predict positive samples. The calculation is given in **Equation** (3).

$$P = \frac{TP}{TP + FP} \tag{3}$$

Accuracy measures the proportion of correctly classified samples among all samples. The higher the accuracy, the more accurate the model's predictions. The calculation is given in **Equation** (4).

$$A = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

Recall measures the proportion of true positive samples among all actual positive samples. The higher the recall, the better the model's ability to detect positive samples. The calculation is given in **Equation** (5).

$$R = \frac{TP}{TP + FN} \tag{5}$$

F1-Score considers the harmonic mean of Precision and Recall, making it suitable for model evaluation when classes are imbalanced. The higher the F1-Score, the better the model's performance in change detection. The calculation is given in **Equation** (6).

$$F1 = 2 \times \frac{PR}{P + R} \tag{6}$$

IoU is used to measure the overlap between the predicted region and the true region. The higher the IoU, the better the prediction performance. The intersection refers to the area where the predicted box overlaps with the true box, while the union refers to the total area of the predicted and true boxes (*i.e.*, the area of their union). The calculation is given in **Equation** (7).

$$IoU = \frac{TP}{TP + FP + FN} \tag{7}$$

mIoU is a commonly used metric in semantic segmentation tasks, representing the average IoU of all categories. The higher the mIoU, the better the model. The calculation is given in **Equation** (8).

$$mIoU = \frac{1}{N} \sum_{i=1}^{N} \frac{TP_i}{TP_i + FP_i + FN_i} \tag{8}$$

*mAP* (mean Average Precision): *mAP* is a commonly used evaluation metric for object detection models, representing the average precision of the model at different *IoU* thresholds. In this study, COCO/bbox_mAP_50 is used to represent the average precision at an *IoU* threshold of 50% in the COCO dataset to evaluate the performance of object detection tasks. The process begins by integrating the P-R curve, then using interpolation to calculate *AP* in the COCO dataset. For each Recall value, the maximum Precision value is taken. The calculation is given in **Equation** (9), and the average *AP* of all categories is taken to obtain the *mAP*, as shown in **Equation** (10).

$$AP = \sum_{i=1}^{N} \left( R_{i+1} - R_i \right) \cdot P_{\text{interp}} \left( R_{i+1} \right) \tag{9}$$

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \tag{10}$$

The Kappa coefficient (also known as Cohen's Kappa) is a statistical measure used to evaluate the consistency of a classifier, as expressed in **Equation** (11). It is especially useful for assessing the match between predicted results and actual labels, considering the effect of random agreement. It represents the actual observed classification consistency, *i.e.*, the proportion of cases where the two classifications match, and represents the expected classification consistency based on random prediction. In a confusion matrix $N$, the calculation of is given in **Equation** (12), and the calculation of is given in **Equation** (13).

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{11}$$

$$p_o = \sum_{i=1}^{k} P_{ii} \tag{12}$$

$$p_o = \frac{1}{N_{\text{total}}} \sum_{i=1}^{k} N_{ii} \tag{13}$$

The Kappa coefficient is used in classification tasks, especially when multiple categories are present, to evaluate the consistency of the model's predictions for each label. In this experiment, it may be used for evaluation in binary and multi-class classification tasks.

## 3. VMamba SS2D Scan Methods

### 3.1. Current Scan Methods and Scan Process

The data processing process in SS2D mainly consists of three steps: cross-scanning, selective scanning using S6 blocks, and cross-merging.

- Cross-Scanning: SS2D first expands the input patches along four different traversal paths. Each path scans the image and flattens the image patches into sequences. This step transforms the spatial information of the image into a sequential format, making it easier for subsequent processing.
- Selective Scanning: Each patch sequence is then processed in parallel through individual S6 blocks. The S6 blocks dynamically adjust based on the input image features, selectively propagating and forgetting information. This step enhances feature extraction capabilities while improving efficiency by parallel computation.
- Cross-Merging: The processed result sequences are reshaped and merged into the output image. During this stage, each pixel in the image integrates information from other pixels from four directions (both forward and backward). This allows SS2D to establish a global receptive field in the 2D space, improving the extraction of image features.

In VMamba, SS2D is a four-direction scanning mechanism designed specifically

for the spatial domain, which can effectively combine spatial information. RS-Mamba further introduces a diagonal scanning approach, creating a full-directional state space, which includes scanning sequences in eight directions. These eight directions consist of four forward directions and four backward directions, with each direction requiring an S6 block to extract features. The four forward directions shows in Figure 6. Although full-directional scanning can extract rich features, it is computationally expensive and slower. To address this issue, this paper proposes a new method that combines single-direction or multi-direction scanning, aiming to reduce computational cost by using different scanning algorithm combinations. The effectiveness of this method is tested on multiple practical tasks. Through experiments on Imagenet image classification, COCO object detection, and RS-Mamba remote sensing change detection, the feature extraction ability of the new scanning algorithm is evaluated.

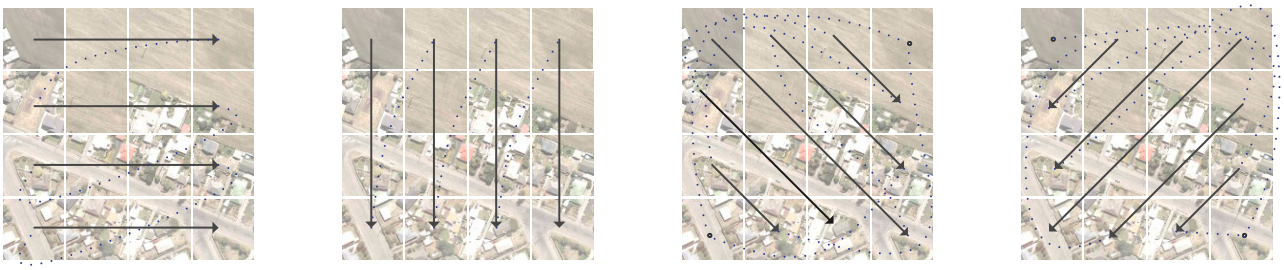


Figure 6. Omnidirectional scanning process.

By testing on these tasks, the relationship between feature extraction ability and remote sensing change detection tasks was determined. Finally, the model was tested on LEVIR-CD and WHU-CD remote sensing datasets, demonstrating that the improved Mamba algorithm performs better in remote sensing change detection tasks, making it more effective for remote sensing applications in change detection.

### 3.2. New Scan Methods

This paper proposes a total of seven completely new scanning algorithms and one combination, including four scanning algorithms and their combinations in RS-Mamba, resulting in a total of 13 different scanning methods. Compared to the original scanning algorithms, relevant experiments were conducted, and the scanning algorithm and the corresponding number of sequence scans (S6 scan sequences) for each method are listed in Table 2.

Table 2. Scanning method and number of scans.

| Scan Name | Scan Count | Scan Name | Scan Count |
|---|---|---|---|
| CrossScan_cir | 2 | CrossScan_para_diag | 2 |
| CrossScan_sn_hor | 2 | CrossScan_para_antidiag | 2 |

**Continued**

| | | | |
|---|---|---|---|
| CrossScan_sn_verti | 2 | CrossScan_para_hor | 2 |
| CrossScan_sn_diag | 2 | CrossScan_para_verti | 2 |
| CrossScan_sn_antidiag | 2 | CrossScan_fake_diag | 2 |
| CrossScan_fake_antidiag | 2 | CrossScan_sn_rec_diag | 8 |

Among these scanning algorithms, para_hor, para_verti, fake_diag, and fake_anti_diag are extracted from the original full-directional scanning algorithm. The sn_rec_diag is a serpentine scanning algorithm, which is a combination of four serpentine scanning methods. After the combination, it has the same number of S6 scanning blocks as the full-directional scanning algorithm. To make it easier to use, the three steps of SS2D scanning processing have been packaged into a separate file for easy use in any task using VMamba scanning algorithms. All scanning algorithms are robust and can be applied to images of any resolution. The scanning order of the seven new scanning methods proposed is shown in Figure 7.

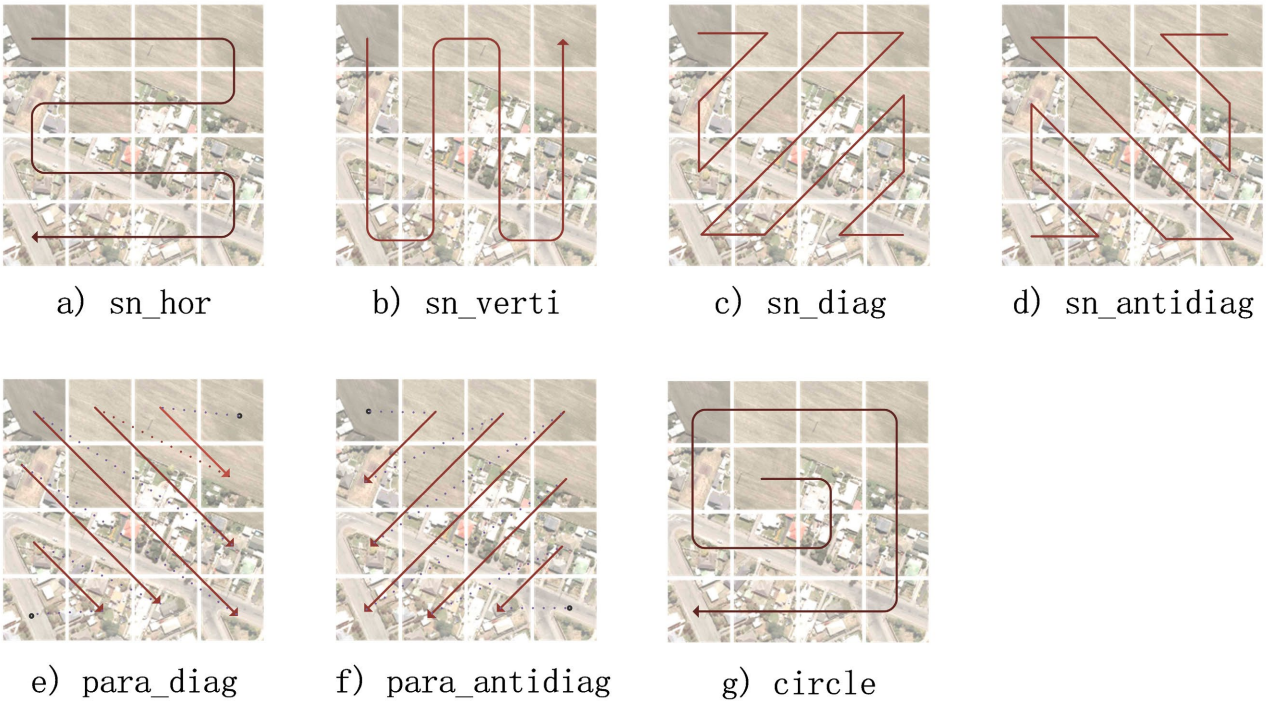


Figure 7. New scanning method diagram.

To assess the computational cost of the core scanning algorithms, we used the fvcore package tool to perform related calculations. The input consisted of two dual-temporal images, and based on the remote sensing change detection algorithm (Chapter 4) module we proposed, we only calculated the core scanning algorithm, specifically the PythonOp. SelectiveScanCore part. This part consists of 15 scanning modules. To objectively record the data, we used K FLOPs/Pixel as

the unit of measurement, representing the number of floating-point operations performed per pixel in this process. Table 3 shows the core scanning computational cost for all 13 scanning methods. From the table, it can be seen that the core computational cost is directly related to the number of scans. In the actual test process, we also found that the computational speed is somewhat related to the scanning method, but this will not be elaborated further here.

Table 3. Core computational complexity of different scanning algorithms (K FLOPs/Pixel).

| Scan Name | Complexity | Scan Name | Complexity |
|---|---|---|---|
| CrossScan_cir | 19.13 | CrossScan_para_diag | 19.13 |
| CrossScan_sn_hor | 19.13 | CrossScan_para_antidiag | 19.13 |
| CrossScan_sn_verti | 19.13 | CrossScan_para_hor | 19.13 |
| CrossScan_sn_diag | 19.13 | CrossScan_para_verti | 19.13 |
| CrossScan_sn_antidiag | 19.13 | CrossScan_fake_diag | 19.13 |
| CrossScan_fake_antidiag | 19.13 | CrossScan_sn_rec_diag | 76.56 |
| CrossScan_Org | 76.56 | | |

## 4. Experiments

### 4.1. ImageNet Image Classification Experiment

The goal of the ImageNet [14] image classification task is to recognize objects or scenes in an image and classify them into one of 1000 categories. For example, categories can include animals (such as dogs, cats, and birds) or objects (such as cars, airplanes, and chairs). In the ImageNet image classification experiment, single-label classification is typically used, meaning each image can only belong to one category (although multi-label classification may be used in some extended tasks). In the ImageNet experiment using VMamba for classification, the network structure used is shown in Figure 8, with default parameters, including data preprocessing methods. For details, please refer to the original text of the author.

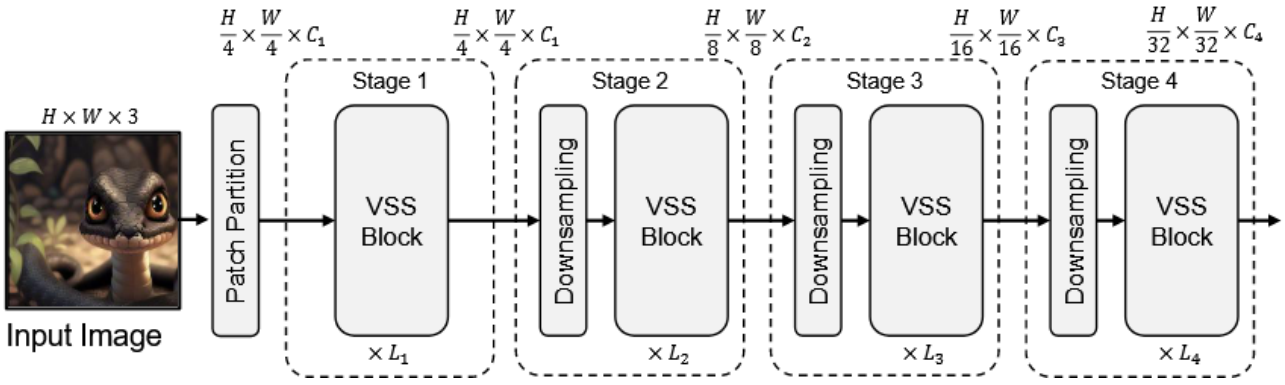


Figure 8. VMamba network architecture [15].

The experimental process involves using the training set and validation set.

Since the test dataset was not provided, our experiment process is as follows: after each epoch of training, validation is performed on the validation set, and the model parameters with the highest accuracy in the validation set are recorded for the second image segmentation experiment.

The data processing of this classification model begins with the original image, which is first passed through a convolution operation for Patch_Embedding. After normalizing the data, it is sent to the VSSM module composed of multiple layers of SSM 2D for processing. The processed feature data is then classified by a classification head. The classification head normalizes the input data, performs channel transformation, applies global average pooling, and flattens the data. Finally, a Linear layer produces the category prediction.

Below is the parameter comparison of various scanning algorithms' Top-1 accuracy (ACC) on ImageNet. To avoid overfitting, EMA results were not used. Models with EMA performed better on the test set than those without EMA. Compared to the standard training models, EMA increased Top-1 Accuracy. The purpose of this experiment is to verify the effectiveness of different scanning algorithms and their computational costs. Therefore, the results of the standard training model are used for comparison. The best model is recorded when the Top-1 ACC reaches the maximum value, which will be used for subsequent segmentation task validation. "Org" represents the original full-directional scanning algorithm results. The accuracy (ACC) results are shown in Figure 9.

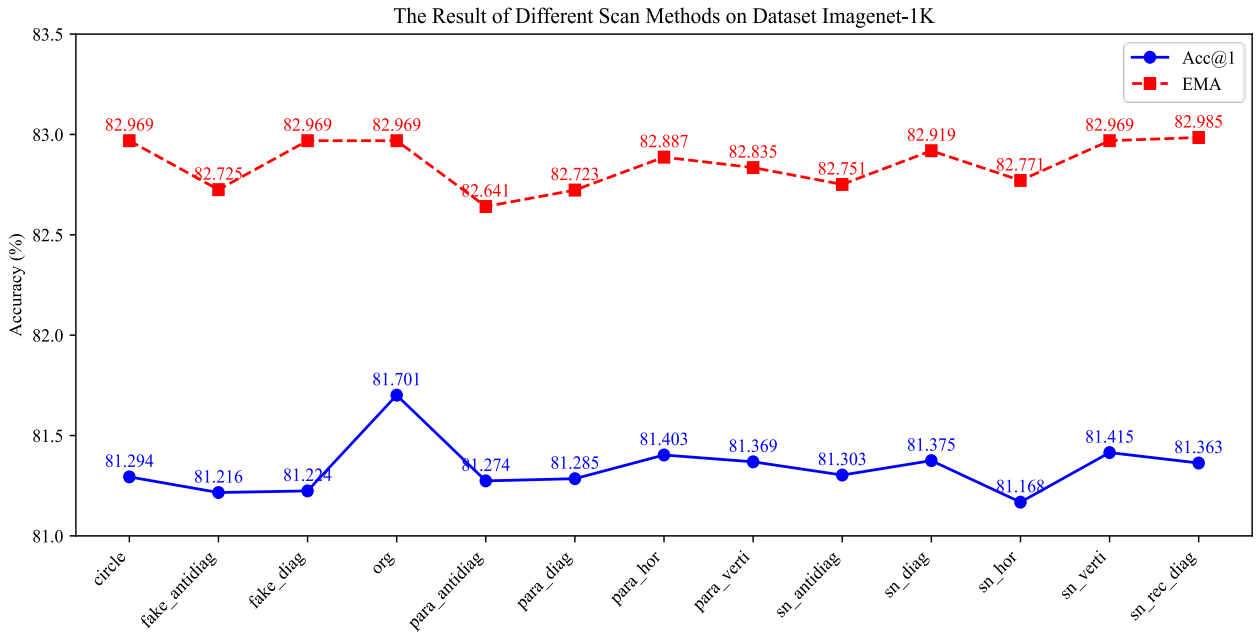


Figure 9. Accuracy of different algorithms in ımagenet-1k ımage classification.

From the figure, we can see that the original full-directional scanning algorithm (Org) indeed has stronger feature acquisition capabilities and performs exceptionally well in classification tasks. In this experiment, we can observe that even with

single-direction scanning algorithms, the following algorithms still perform better. The newly proposed algorithms sn_verti and para_hor both show good performance in Acc@1 and EMA Acc@1, outperforming other single-direction scanning algorithms.

### 4.2. COCO2017 Object Detection Experiment

COCO (Common Objects in Context) [16] is one of the most commonly used object detection datasets in computer vision, containing 80 object categories. It includes challenges such as multi-scale, complex backgrounds, and occlusions. In this experiment, the object detection testing is carried out using the OpenMMLab-developed libraries MMCV and MMDetection. These libraries are core components of the OpenMMLab ecosystem. MMCV serves as the foundational library for MMDetection [17], which is a dedicated library for object detection tasks, while MMCV provides many general tools and modules to support and enhance the development of MMDetection and other computer vision tasks.

In this experiment, I used the VSSM module, which is composed of my proposed scanning algorithms, as the feature extraction backbone network. The model parameters are pre-trained on the ImageNet dataset from the classification experiment, and together with other MMCV tools, the algorithm's transferability is validated. The operational logic for MMDetection in object detection tasks is shown in **Figure 10**.

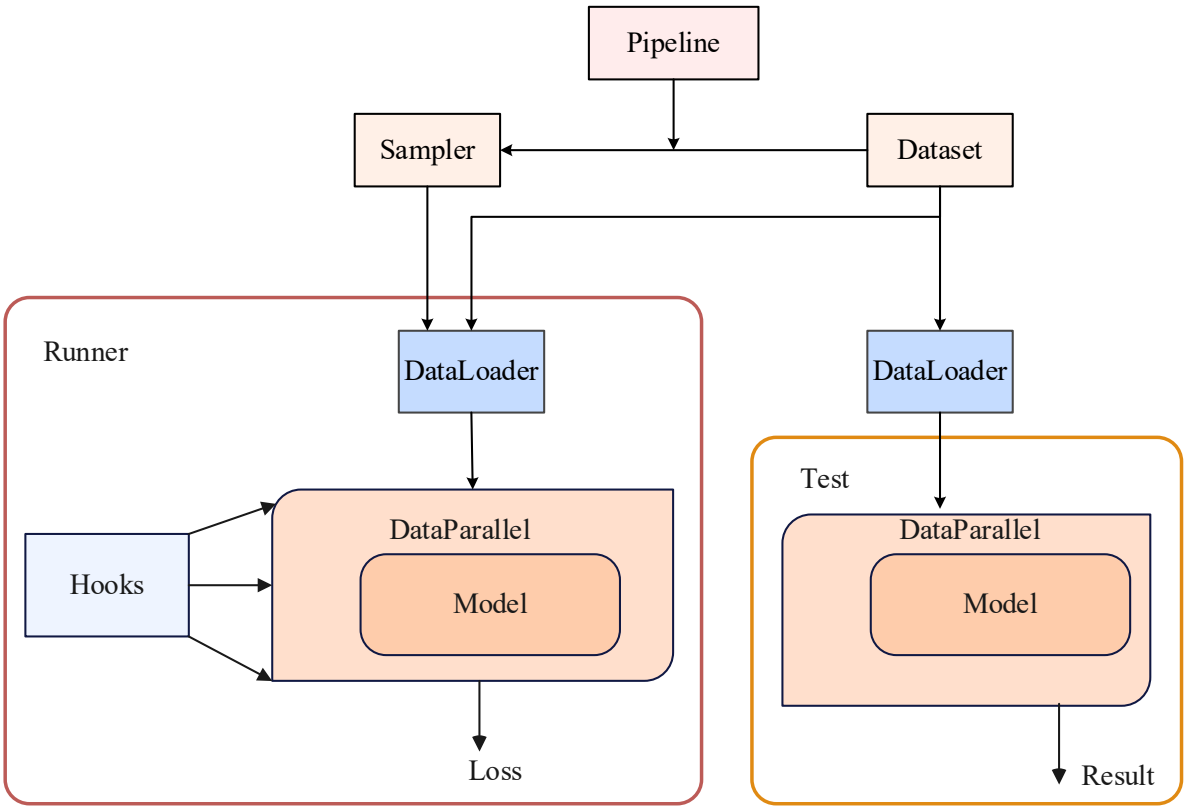


**Figure 10.** MMDET Tool Running Logic.

This experiment leverages the detection network from the VMamba paper, where part of the configuration, including the Backbone pre-trained file, differs from the original. The best parameters from the various scanning algorithms in the ImageNet classification task are used as the Backbone pre-trained files. The model is fine-tuned on the COCO2017 dataset for 12 epochs. The best model parameters from these 12 epochs are tested on the COCO2017 test dataset. The performance of each scanning algorithm is evaluated based on the final model's performance. For specific methods and processes related to object detection, readers can refer to other sources. Figure 11 shows the general process of training and testing the model. The goal of this experiment is to verify the transferability of the model by comparing the convergence speed and performance of different algorithms, thereby demonstrating the effectiveness differences of the algorithms.

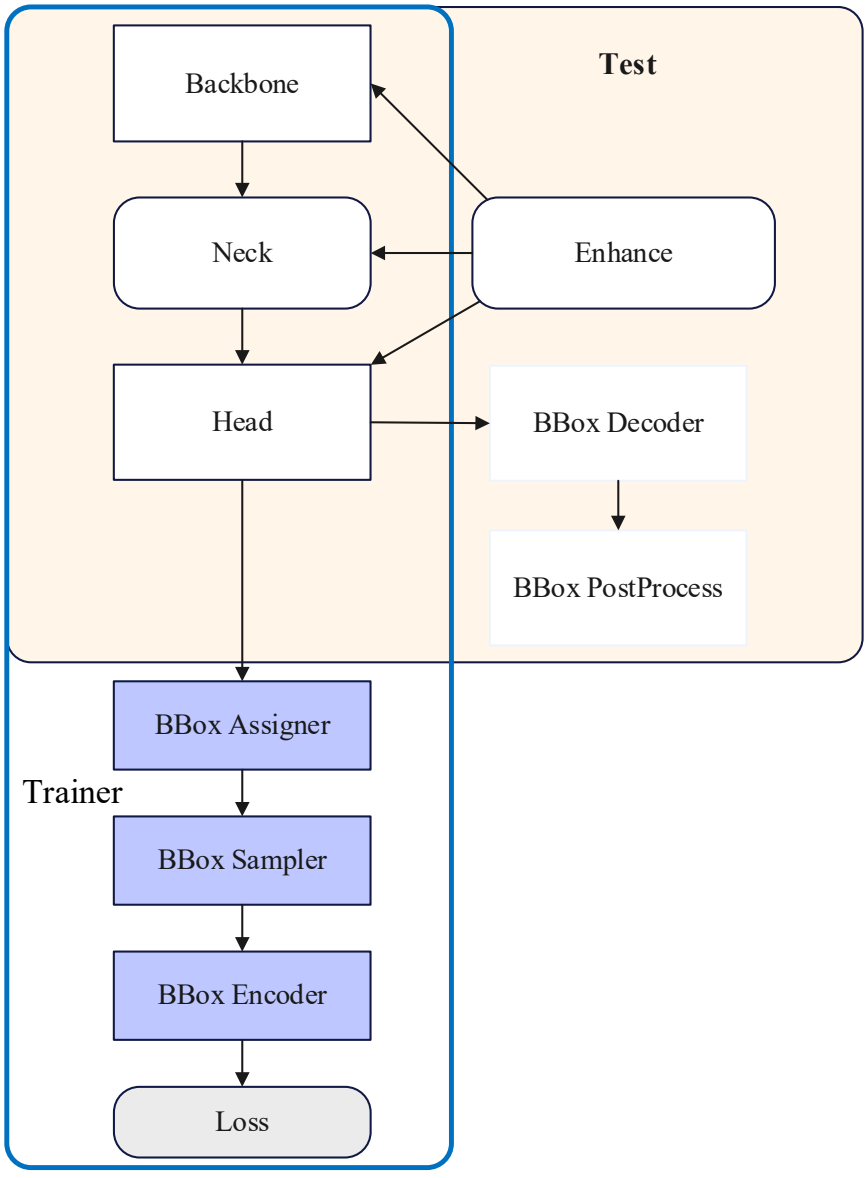


Figure 11. MMDet object detection model workflow.

The object detection in this experiment adopts a two-stage approach. The process involves using the backbone network MM_VSSM for feature extraction after preprocessing the data. The pre-trained model parameters are used in this experiment. The FPN (Feature Pyramid Network) [18] is used for fusing features of different scales. The Region Proposal Network (RPN) generates candidate bounding boxes, and RoI-Align extracts features from these candidate boxes. The bbox_roi_extractor is responsible for extracting ROI features, which are later used by the bbox_head to regress the bounding box coordinates and determine the object location. The bbox_head is used for classifying the candidate boxes and regressing the bounding boxes. The mask_head component generates object segmentation masks. The model computes classification loss, bounding box regression loss, and mask loss based on the predicted results and ground truth labels. During the evaluation phase, the BBox postprocess module removes duplicate detection boxes using operations such as non-maximum suppression (NMS), box adjustment (e.g., sorting by category and confidence), and threshold filtering (e.g., removing boxes with low confidence), and outputs the final segmentation result after binarization.

In the configuration, we modified the image data preprocessing and resampling size to 800 × 600 and set the batch size to 18. Since the goal of this experiment is to validate the feature extraction capabilities of SS2D scanning algorithms in the Backbone, we made adjustments to accommodate the hardware constraints. The configurations, including the original full-directional scanning algorithm, were compared. Other default configuration parameters can be found in the target detection section of the VMamba source code's configuration file.

Below are the performance results of the 13 scanning algorithms on the

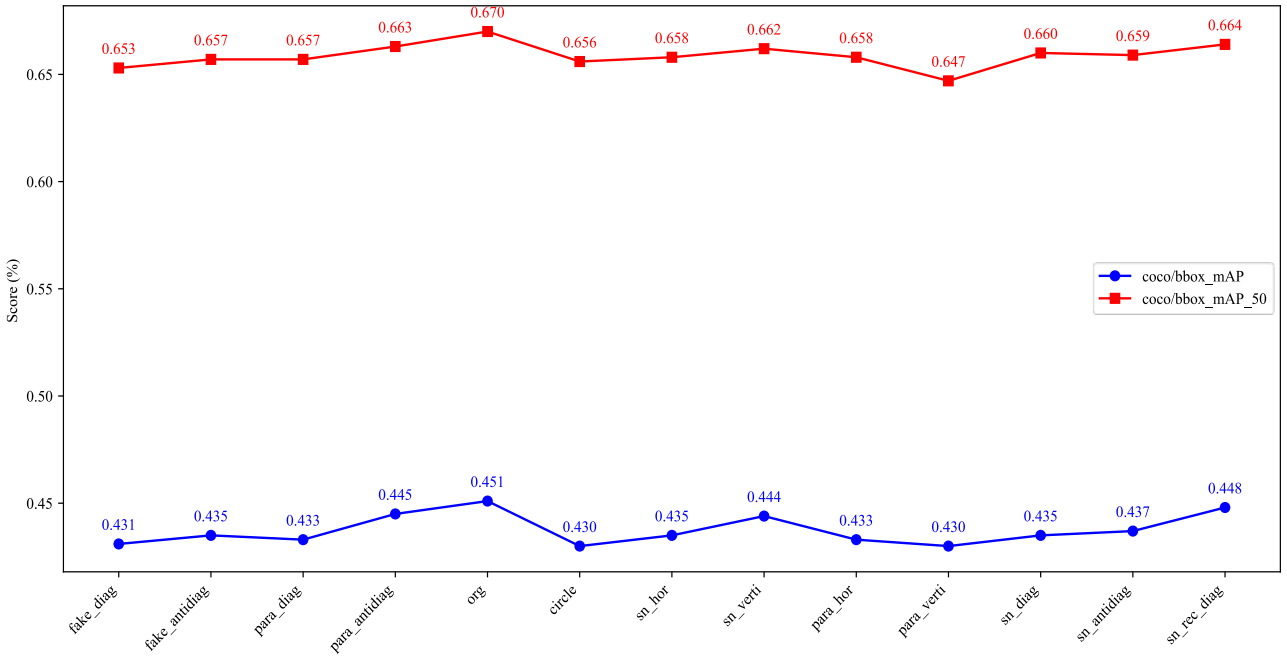


**Figure 12.** Performance of different algorithms in COCO 2017 object detection.

COCO2017 object detection task validation set. We use COCO/bbox_mAP and COCO/bbox_mAP_50 as evaluation metrics, with COCO/bbox_mAP_50 being the primary metric for assessing the detection ability of each algorithm. Figure 12 presents the test results of various scanning algorithms.

From the results, we can see that in terms of mAP_50 (mean Average Precision at an Intersection over Union threshold of 0.5), the accuracy of algorithms like org, para_antidiag, sn_verti, sn_diag, and sn_rec_diag is greater than 0.66. However, when evaluating mAP_50, the algorithms org, para_antidiag, sn_verti, and sn_rec_diag also perform well. Therefore, the algorithms with overall good performance are org, para_antidiag, sn_verti, and sn_rec_diag.

Combining the ImageNet classification performance, it can be concluded that sn_verti in the single-direction scanning algorithms has stable feature extraction capabilities. Although it does not perform as well as the full-directional scanning algorithm (org), it exhibits better algorithm efficiency.

### 4.3. Comparative Experiments on SS2D Change Detection

The goal of this experiment is to validate the practical effectiveness of various scanning algorithms in remote sensing image change detection tasks. In the RS-Mamba network, the omnidirectional selective scanning method was proposed for the first time and introduced into VHR (Very High Resolution) remote sensing dense prediction tasks, making it the foundational network for this experiment. By replacing the SS2D scanning algorithm in the model with the algorithm proposed in this paper, experiments are conducted on the WHU-CD and LEVIR-CD datasets. The experimental results, combined with previous classification and object detection experiments, help determine whether the feature extraction capability is directly related to the performance in remote sensing change detection, and assess the specific performance of the algorithm in this domain.

The change detection network in RS-Mamba adopts a Siamese network architecture. The dual-temporal VHR remote sensing images are first converted into dual-temporal image patch sequences via patch embedding. These sequences are then fed into a dual-temporal encoder with shared weights to extract features. The shared-weight encoder in RS-Mamba consists of five stages, each containing multiple OSS (Object-Specific Scan) blocks, and the encoder itself is made up of four encoder blocks. After feature extraction by the shared-weight encoder, the features of the dual-temporal images with the same size are concatenated along the channel dimension and fused through convolutions, integrating the information of the dual-temporal VHR remote sensing images to effectively segment the changed objects. The fused features are upsampled in the decoder and concatenated with the corresponding fused features from skip connections and convolutions. After two convolution operations for dimensionality reduction and upsampling to restore the original resolution, a final convolution operation is applied for classification output. The overall structure is illustrated in Figure 13.

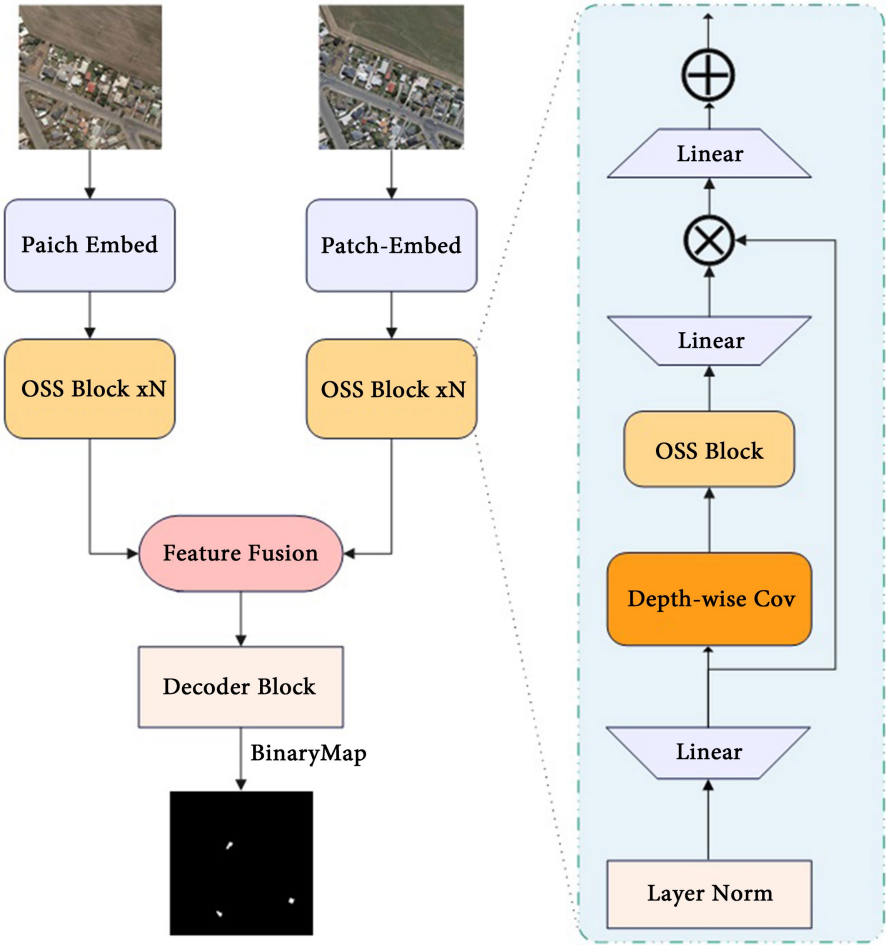


Figure 13. RS-Mamba network architecture [6].

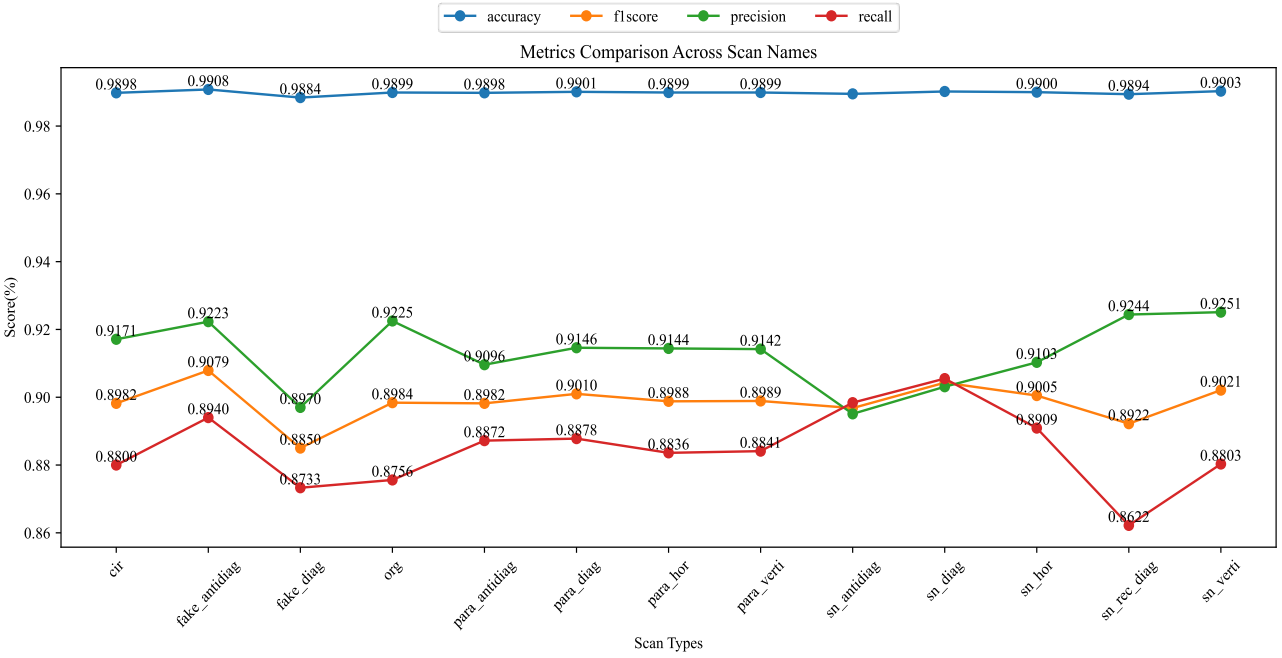


Figure 14. Test results of different scanning methods in RS-Mamba.

In this experiment, we replace the SS2D scanning algorithm with the relevant content from this study and verify the potential performance of different scanning algorithms in VHR remote sensing dense prediction tasks. The batch size is set to 4, and after 300 epochs, the model with the highest F1 score is selected. The validation data from the LEVIR dataset is used to assess the detection performance of different scanning algorithms. Evaluation metrics A, F1, Precision (P), and Recall (R) are used to describe the test results, as shown in Figure 14, where the key algorithmic metrics are annotated on the graph. The results for other algorithms are presented in Table 4, with the best-performing parameters highlighted in red and the second-best parameters in bold.

Table 4. Test results of All scanning methods in RS-mamba.

| | A | F1 | P | R |
|---|---|---|---|---|
| cir | 0.9898 | 0.8982 | 0.9171 | 0.8800 |
| fake_antidiag | **0.9908** | **0.9079** | 0.9223 | 0.8940 |
| fake_diag | 0.9884 | 0.8850 | 0.8970 | 0.8733 |
| org | 0.9899 | 0.8984 | 0.9225 | 0.8756 |
| Org* | - | 0.9110 | 0.9252 | 0.8973 |
| para_antidiag | 0.9898 | 0.8982 | 0.9096 | 0.8872 |
| para_diag | 0.9901 | 0.9010 | 0.9146 | 0.8878 |
| para_hor | 0.9899 | 0.8988 | 0.9144 | 0.8836 |
| para_verti | 0.9899 | 0.8989 | 0.9142 | 0.8841 |
| sn_antidiag | 0.9895 | 0.8968 | 0.8951 | **0.8984** |
| sn_diag | 0.9902 | **0.9043** | 0.9031 | **0.9055** |
| sn_hor | 0.9900 | 0.9005 | 0.9103 | 0.8909 |
| sn_rec_diag | 0.9894 | 0.8922 | **0.9244** | 0.8622 |
| sn_verti | **0.9903** | 0.9021 | **0.9251** | 0.8803 |

From the figures and tables, we can see that under the same environment, the original omnidirectional scanning algorithm performs worse than many unidirectional scanning algorithms in the remote sensing change detection task on the LEVIR-CD dataset. The sn_rec_diag algorithm, proposed in this study, performs similarly to the original omnidirectional algorithm, suggesting that the choice of scanning method significantly impacts change detection performance. Among the important parameters, the proposed scanning algorithms generally perform well in RS-Mamba. Algorithms such as fake_antidiag, sn_verti, and sn_diag show better performance, especially in the F1 score, on core metrics. To balance algorithm stability and efficiency, this paper selects the RS-Mamba model with the sn_verti unidirectional scanning algorithm for comparative experiments across multiple datasets against other algorithms.

### 4.4. Comparative Experiments on Remote Sensing Change Detection

This study uses RS-Mamba as the experimental benchmark to compare the performance of common remote sensing change detection models on the LEVIR-CD and WHU-CD datasets. The parameters of the SSM module are kept consistent with those in RS-Mamba. The baseline model's performance on these datasets is reproduced in our local experimental environment, while the test results of other models are cited from their original papers. During training on different datasets, the number of epochs is set to 200, with varying batch sizes: 8 for LEVIR-CD and 64 for WHU-CD256. The optimizer used is Adam, with the learning rate linearly increasing to 1e-3 within the first 1000 steps.

The comparison of key parameters between commonly used methods and ENRS-Mamba on the LEVIR-CD and WHU-CD datasets is shown in Table 5 and Table 6, where RS-Mamba represents the results obtained in our experimental environment.

To provide a more objective evaluation of the overall performance of different algorithms, Table 7 presents the parameter size and computational cost of some statistically competitive methods.

Table 5. Performance of common algorithms in change detection on LEVIR-CD dataset.

| Methods | P | F1 | R | IoU |
|---|---|---|---|---|
| RS-Mamba(org) | 0.9225 | 0.8984 | 0.8750 | 0.8156 |
| FC-EF [8] | 0.8691 | 0.8340 | 0.8017 | 0.7153 |
| FC-Siam-Diff [8] | 0.8953 | 0.8631 | 0.8331 | 0.7591 |
| MTCNet [19] | 0.9087 | **0.9024** | **0.8962** | 0.8222 |
| BIT [20] | 0.8924 | 0.8930 | **0.8937** | 0.8068 |
| ChangeFormer [21] | 0.9205 | **0.9040** | **0.8880** | **0.8247** |
| RS-Mamba (sn_verti) | **0.9251** | **0.9021** | **0.8803** | **0.8235** |

Table 6. Performance of common algorithms in change detection on WHU-CD.

| Methods | P | F1 | R | IoU |
|---|---|---|---|---|
| RS-Mamba(org) | 0.9233 | 0.9122 | 0.9014 | 0.8386 |
| FC-EF [8] | 0.7886 | 0.7875 | 0.7864 | 0.6495 |
| FC-Siam-Diff [8] | 0.8473 | 0.8600 | 0.8731 | 0.7544 |
| MTCNet [19] | 0.7510 | 0.8265 | **0.9190** | 0.7043 |
| BIT [20] | 0.8664 | 0.8398 | 0.8148 | 0.7239 |
| MSCANet [22] | 0.9110 | 0.9047 | 0.8986 | 0.8260 |
| RS-Mamba (sn_verti) | **0.9363** | **0.9170** | 0.8985 | **0.8467** |

Table 7. Algorithm efficiency of representative remote sensing change detection methods.

| Methods | Params(M) | FLOPs(G) |
|---|---|---|
| RS-Mamba [6] | 51.95 | 26.46 |

**Continued**

| | | |
|---|---|---|
| MambaBCD-Tiny [12] | 17.13 | 45.74 |
| SNUNet [23] | 12.03 | 27.44 |
| ChangeFormerV3 [21] | 24.30 | 33.68 |
| RS-Mamba (sn_verti) | 44.71 | 16.82 |

Figure 15 and Figure 16 show the actual results of the model structure tested on the LEVIR-CD and WHU-CD datasets with default configuration parameters.


Figure 15. Example of Test Results on LEVIR-CD (Resolution 1024*1024).

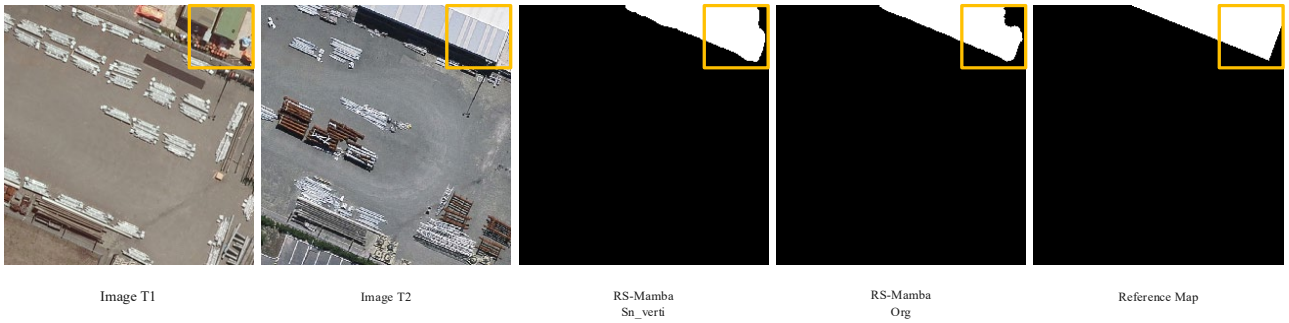


Figure 16. Example of Test Results on WHU-CD (Resolution 256*256).

From the figures, it is evident that compared to RS-Mamba, which employs an omnidirectional scanning-based change detection model, the proposed algorithm demonstrates superior performance. This improvement is mainly reflected in fewer noise artifacts and more complete classification results, leading to an overall enhancement in change detection performance.

Among the comparison models, Table 5 and Table 6 present the performance of representative models on remote sensing change detection datasets. Representative CNN-based models include FC-EF, FC-Siam-Diff, and DSIFN, while Transformer-based models include ChangeFormer and MTCNet. ChangeFormer is a hybrid model combining CNN and Transformer networks, while RS-Mamba represents Mamba-based models.

From the tabulated data, it can be observed that overall, Transformer-based or hybrid models outperform pure CNN-based models, with Mamba-based models achieving the best performance. Although the proposed method did not achieve

the best performance in LEVIR-CD change detection by only replacing the scanning algorithm, it significantly reduces computational complexity compared to models with similar performance levels. Moreover, in WHU-CD (with a resolution of 256×256), the proposed method demonstrates a leading advantage in remote sensing change detection. These experimental results validate the effectiveness of the proposed unidirectional scanning algorithm, which exhibits competitive advantages in both computational efficiency and detection accuracy.

### 4.5. High-Resolution Image Comparison Experiment

To evaluate the algorithm's capability in handling high-resolution images, we utilized the original WHU-CD dataset and conducted two sets of tests by partitioning it into different resolutions. In the comparative experiments presented in this study, we used the dataset with a resolution of 256 × 256. To assess the change detection performance of the algorithm on high-resolution remote sensing images, the original WHU-CD dataset was non-overlappingly segmented at different resolutions, specifically at 1024 × 1024 and 1536 × 1536.

The original WHU-CD images were captured at a resolution of 32,507 × 15,354 pixels. The partitioned dataset was divided into training, testing, and validation sets in a 6:2:2 ratio. Additionally, due to significant differences in GPU memory consumption between the RS-Mamba network using the original omnidirectional scanning algorithm and the scanning algorithm proposed in this study at the same resolution, the memory usage of the original omnidirectional scanning algorithm was found to be more than twice that of the proposed scanning algorithm. Consequently, different batch sizes were used for model training under different algorithms, as detailed in Table 8. At a resolution of 1536 × 1536, the original omnidirectional scanning algorithm encountered out-of-memory (OOM) issues when the batch size was set to 2. Therefore, at this resolution, only the proposed algorithm was tested, and the results from this part should be considered as a reference. The experimental results under different resolution settings are presented in Table 9 and Table 10. The testing results on the WHU-CD dataset at a resolution of 1024 × 1024 are illustrated in Figure 17.

Table 8. Batch size during model training for different algorithms at various resolutions.

| | 1024*1024 | 1536*1536 |
|---|---|---|
| RS-Mamba(org) | 4 | — |
| RS-Mamba(sn_verti) | 8 | 2 |

Table 9. Comparison test results on WHU-CD with 1024*1024 resolution.

| | P | F1 | R | IoU | Kappa |
|---|---|---|---|---|---|
| RS-Mamba(org) | 0.7412 | 0.7011 | 0.6651 | 0.5394 | 0.6931 |
| RS-Mamba(sn_verti) | **0.7870** | **0.7164** | **0.6574** | **0.6474** | **0.7111** |

Table 10. Experimental results on WHU-CD with 1536*1536 resolution.

| | P | F1 | R | IoU | Kappa |
|---|---|---|---|---|---|
| RS-Mamba(org) | — | — | — | — | — |
| RS-Mamba(sn_verti) | **0.7189** | **0.7247** | **0.7306** | **0.5768** | **0.7140** |


Figure 17. Example of test results on WHU-CD dataset with 1024*1024 resolution.

From Table 9 and Table 10, it can be observed that the proposed method outperforms the omnidirectional scanning approach in high-resolution images. In more challenging high-resolution remote sensing images, the proposed method exhibits lower hardware dependency while maintaining stable performance without degradation, enabling the model to make more accurate judgments.

As shown in Figure 17, the proposed method significantly reduces false detections in high-resolution remote sensing change detection. However, it also results in greater loss of edge details and fine structures. Nevertheless, the overall performance remains superior.

## 5. Conclusion and Analysis

From the results of the ImageNet classification experiment and the COCO dataset object detection experiment, the following conclusions can be drawn:

a) Omnidirectional scanning algorithms have stronger feature extraction capabilities, whereas unidirectional scanning algorithms are weaker in terms of feature extraction performance.

b) The results from remote sensing change detection tasks show that feature extraction capability is a necessary but not sufficient condition for better change detection performance. In other words, stronger feature extraction capability does not always guarantee better detection results in remote sensing change detection tasks. However, for good performance in change detection, strong feature extraction capability is essential. In this study, the scanning algorithm needs to demonstrate good classification and object detection capabilities, which would then translate into good detection performance in remote sensing change detection.

c) The unidirectional scanning algorithms proposed in this paper outperform the omnidirectional scanning algorithm in remote sensing change detection. Despite the computational cost being only 1/4 that of the omnidirectional scanning algorithm, the proposed unidirectional algorithms achieve slightly better perfor-

mance on the LEVIR-CD dataset and better performance on the WHU-CD dataset, thus proving the effectiveness of the algorithms.

To address the issue of excessive computational cost and memory usage in VMamba when processing image data, this paper proposes several unidirectional scanning algorithms for VMamba SS2D and analyzes the performance of these algorithms through classification and object detection tasks. Testing in RS-Mamba demonstrates the effectiveness and computational efficiency of the proposed unidirectional scanning algorithms in remote sensing change detection. These algorithms offer valuable reference for future research. However, the limitations of unidirectional scanning algorithms have been acknowledged. Future research could focus on optimizing these algorithms further to achieve feature extraction capabilities similar to omnidirectional scanning algorithms or even four-directional scanning algorithms, but with reduced computational costs. This presents a promising research direction.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

[1] Afaq, Y. and Manocha, A. (2021) Analysis on Change Detection Techniques for Remote Sensing Applications: A Review. *Ecological Informatics*, **63**, Article 101310. https://doi.org/10.1016/j.ecoinf.2021.101310

[2] Gu, A. and Dao, T. (2024) Mamba: Linear-Time Sequence Modeling with Selective State Spaces. https://doi.org/10.48550/arXiv.2312.00752

[3] Liu, Y., Tian, Y.J., *et al.* (2024) VMamba: Visual State Space Model. https://doi.org/10.48550/ARXIV.2401.10166

[4] Shimura, K., Li, J. and Fukumoto, F. (2018) HFT-CNN: Learning Hierarchical Category Structure for Multi-Label Short Text Categorization. *Proceedings of the* 2018 *Conference on Empirical Methods in Natural Language Processing*, Brussels, October-November, 2018, 811-816. https://doi.org/10.18653/v1/d18-1093

[5] Vaswani, A., Shazeer, N., Parmar, N., *et al.* (2023) Attention Is All You Need. https://doi.org/10.48550/arXiv.1706.03762

[6] Zhao, S., Chen, H., Zhang, X., Xiao, P., Bai, L. and Ouyang, W. (2024) RS-Mamba for Large Remote Sensing Image Dense Prediction. *IEEE Transactions on Geoscience and Remote Sensing*, **62**, 1-14. https://doi.org/10.1109/tgrs.2024.3425540

[7] Chen, H. and Shi, Z. (2020) A Spatial-Temporal Attention-Based Method and a New Dataset for Remote Sensing Image Change Detection. *Remote Sensing*, **12**, Article 1662. https://doi.org/10.3390/rs12101662

[8] Ji, S., Wei, S. and Lu, M. (2019) Fully Convolutional Networks for Multisource Building Extraction from an Open Aerial and Satellite Imagery Data Set. *IEEE Transactions on Geoscience and Remote Sensing*, **57**, 574-586. https://doi.org/10.1109/tgrs.2018.2858817

[9] Gu, A., Goel, K. and Ré, C. (2022) Efficiently Modeling Long Sequences with Structured State Spaces. https://doi.org/10.48550/arXiv.2111.00396

[10] Smith, J.T.H., Warrington, A. and Linderman, S. (2022) Simplified State Space Layers

for Sequence Modeling. *The* 11*th International Conference on Learning Representations*, Kigali, 21 September 2022, 1-35. https://openreview.net/forum?id=Ai8Hw3AXqks

[11] Gu, A., Dao, T., Ermon, S., Rudra, A. and Re, C. (2020) HiPPO: Recurrent Memory with Optimal Polynomial Projections. https://doi.org/10.48550/arXiv.2008.07669

[12] Chen, H., Song, J., Han, C., Xia, J. and Yokoya, N. (2024) Changemamba: Remote Sensing Change Detection with Spatiotemporal State Space Model. *IEEE Transactions on Geoscience and Remote Sensing*, **62**, 1-20. https://doi.org/10.1109/tgrs.2024.3417253

[13] McHugh, M.L. (2012) Interrater Reliability: The Kappa Statistic. *Biochemia Medica*, **22**, 276-282. https://doi.org/10.11613/bm.2012.031

[14] Deng, J., Dong, W., Socher, R., Li, L., Li, K. and Li, F.-F. (2009) ImageNet: A Large-Scale Hierarchical Image Database. 2009 *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, 20-25 June 2009, 248-255. https://doi.org/10.1109/cvpr.2009.5206848

[15] Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W. and Wang, X. (2024) Vision Mamba: Efficient Visual Representation Learning with Bidirectional State Space Model. https://doi.org/10.48550/ARXIV.2401.09417

[16] Lin, T., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., *et al*. (2014) Microsoft COCO: Common Objects in Context. In: *Lecture Notes in Computer Science*, Springer, 740-755. https://doi.org/10.1007/978-3-319-10602-1_48

[17] Chen, K., Wang, J.Q., Pang, J.M., *et al*. (2019) MMDetection: Open MMLab Detection Toolbox and Benchmark. https://doi.org/10.48550/arXiv.1906.07155

[18] Lin, T., Dollar, P., Girshick, R., He, K., Hariharan, B. and Belongie, S. (2017) Feature Pyramid Networks for Object Detection. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 936-944. https://doi.org/10.1109/cvpr.2017.106

[19] Wang, W., Tan, X., Zhang, P. and Wang, X. (2022) A CBAM Based Multiscale Transformer Fusion Approach for Remote Sensing Image Change Detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **15**, 6817-6825. https://doi.org/10.1109/jstars.2022.3198517

[20] Chen, H., Qi, Z. and Shi, Z. (2022) Remote Sensing Image Change Detection with Transformers. *IEEE Transactions on Geoscience and Remote Sensing*, **60**, 1-14. https://doi.org/10.1109/tgrs.2021.3095166

[21] Bandara, W.G.C. and Patel, V.M. (2022) A Transformer-Based Siamese Network for Change Detection. 2022 *IEEE International Geoscience and Remote Sensing Symposium*, Kuala Lumpur, 17-22 July 2022, 207-210. https://doi.org/10.1109/igarss46834.2022.9883686

[22] Liu, M., Chai, Z., Deng, H. and Liu, R. (2022) A CNN-Transformer Network with Multiscale Context Aggregation for Fine-Grained Cropland Change Detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **15**, 4297-4306. https://doi.org/10.1109/jstars.2022.3177235

[23] Fang, S., Li, K., Shao, J. and Li, Z. (2022) SNUNet-CD: A Densely Connected Siamese Network for Change Detection of VHR Images. *IEEE Geoscience and Remote Sensing Letters*, **19**, 1-5. https://doi.org/10.1109/lgrs.2021.3056416