

Land Cover Classification Based on Multimodal Remote Sensing Fusion

Wei Chen¹, Jiage Chen², Yuewu Wan^{2*}, Xining Liu¹, Mengya Cai¹, Jingguo Xu¹, Hongbo Cui¹, Mengdie Duan¹

¹ Sanya Institute of South China Sea Geology, Guangzhou Marine Geological Survey, China Geological Survey, Sanya 572024, China- (chengwei001, liuxining, caiyamen, xujinguo001, cuihongbo, duanmengdie)@mail.cgs.gov.cn

² National Geomatics Center of China, 100830 Beijing, China - (jiagechen, wyw)@ngcc.cn

Keywords: Land cover data, deep learning, Multi source image fusion

Abstract

Global high-precision and high timeliness land cover data is a fundamental and strategic resource for global strategic interest maintenance, global environmental change research, and sustainable development planning. However, due to difficulties in obtaining control and reference information from overseas, a single data source cannot effectively cover, and surface coverage classification faces significant challenges in information extraction. Based on this, this article proposes an intelligent interpretation method for typical elements based on multimodal fusion, starting from the characteristics of domestic remote sensing images. It also develops an optical SAR data conversion and complementarity strategy based on convolutional translation networks, as well as a typical element extraction algorithm. This solves the problems of sparse remote sensing images, limited effective observations, and difficult information recognition, thereby achieving automation of typical element information under dense observation time series. High precision extraction and analysis.

1. Introduction

Land cover data is a fundamental and strategic resource for global strategic interest maintenance, global environmental change research, and sustainable development planning. At present, domestic and foreign institutions have developed dozens of global surface coverage data products, whose spatial resolution is constantly improving, gradually increasing from the original coarse resolution (300-10km) to medium to high resolution (10-30m). However, its timing and accuracy still struggle to meet practical needs. Especially for cultivated land types with strong spatial heterogeneity and obvious seasonal dynamics, their accuracy is often greatly limited. The effectiveness of extracting cultivated land in areas with complex planting structures, fragmented landscapes, and strong heterogeneity is still unsatisfactory. The reason for this is that global surface coverage data products are mainly concentrated in a single data source, such as optical Landsat, Sentinel-2, and other data. However, a single remote sensing data source often leads to issues such as ineffective global spatiotemporal coverage, limited accuracy in automatic extraction, and difficulty in mining rich information on land cover types. Specifically, it mainly includes the following aspects:

(1) The development of global surface coverage data products involves the massive information processing of tens of thousands of remote sensing images. Image defects (or missing information) caused by clouds, terrain shadows, hardware settings, etc. are common, and a single data source is difficult to achieve high-quality global spatiotemporal continuous coverage of images. In recent years, with the rapid development of aerospace technology, remote sensing data with different imaging methods and resolutions has shown explosive growth, such as domestically produced high-resolution images and SAR images. The combined use of these massive and multimodal remote sensing data will enable effective coverage on a global

land scale, providing an important data foundation for all-weather, all element, and all process monitoring.

(2) Although the development of remote sensing earth observation technologies such as multispectral, high-resolution, and synthetic aperture radar has led to a diversified trend in remote sensing imaging methods. However, due to the mutual constraints between imaging indicators, the surface information observed by a single remote sensing system is often not comprehensive, especially in the global large-scale use of a single remote sensing data for accurate land cover extraction, which still poses challenges. When extracting cultivated land types from optical remote sensing data, the effect is poor in areas with complex planting structures, fragmented landscapes, and strong heterogeneity. The radar backscatter coefficient is sensitive to the dielectric properties of ground objects, and has different responses to the moisture content of vegetation and soil, as well as the geometric characteristics of the surface such as roughness, which can provide unique information different from optical images.

The multi-modal image differences are particularly evident in heterogeneous time series, such as significant differences in response position, amplitude, and temporal trajectory between optical time series and SAR time series. The existing methods for combining optical and SAR data can be roughly divided into two categories, namely fusion based on single temporal phase and fusion based on time series images. The fusion based on single shot images focuses on building spatial mapping relationships, with the goal of constructing SAR and optical mapping factors for pixels at the same position based on cloud free data from the front and back temporal phases, ultimately determining information under cloud coverage (Gao et al., 2020; Wang et al., 2019). However, due to the different imaging mechanisms of sensors, it is difficult to construct a stable and reliable mapping relationship between optical and SAR images in a single scene, and the constructed relationship is also difficult to transfer to other time periods. The fusion based on

* Corresponding author

time series images focuses on the mapping relationship of the time dimension, greatly reducing the dependence on single temporal image features. Many studies have begun to explore the long-term reliable relationship between optical and SAR data. A recent study introduced a multi output Gaussian regression process model for Sentinel-1 and Sentinel-2 to generate complete leaf area index (LAI) time series (Claverie et al., 2018). However, this correspondence is greatly influenced by vegetation types or climate differences, making it difficult to construct a long-term and reliable mapping relationship. With the advent of the era of deep learning and artificial intelligence, deep neural networks with specific architectures have been able to automatically learn the deep features of ground objects. Zhao et al. (2020) achieved the prediction of complete NDVI from SAR long time series using a Long Short Term Memory Neural Network (LSTM) model. However, the model relies on the temporal and temporal connections before and after the time series, and SAR data is affected by noise, so the temporal and temporal connections are not obvious, which greatly limits the model's ability

Therefore, this article will combine advanced deep learning frameworks, starting from the deep fusion of multi-source temporal remote sensing data, effectively combining the data advantages of optics and SAR, achieving information complementarity, and thereby improving the accuracy of surface coverage mapping. Secondly, from the perspective of multimodal feature extraction and fusion of time-series remote sensing data, achieve rolling prediction of high-resolution remote sensing data. On this basis, achieve large-scale surface coverage mapping to alleviate the common lag problem of existing mapping methods.

The major contributions of this paper are summarized as follows: (1) our approach integrates information from multiple sources time-series data into a high spatiotemporal dataset, enabling effective fusion and mining of multi-source data. (2) In order to solve the difference between optical and SAR representation, this paper proposes a deep coupling model for integrating semantic features and visual features.

The remainder of this paper is organized as follows. Section 2 introduces the representation learning of land cover classification and the deep cross-modal coupling model in detail. Section 3 summarizes the experimental results. Finally, the conclusion is detailed in Section 4.

2. Methodology

In response to the core issues of how to improve the frequency of ground observations and integrate the advantages of multimodal remote sensing observations in the extraction of typical elements of land cover, this paper integrates multimodal remote sensing big data and field research data, and conducts research on the "cross platform multimodal" remote sensing data collaboration and automatic extraction of typical elements supported by deep learning. We have developed optical SAR data conversion and complementarity strategies based on convolutional translation networks, as well as typical feature extraction methods, aiming to solve the problems of sparse remote sensing images, limited effective observations, and difficult information recognition. This will enable the automation and high-precision extraction and analysis of typical feature information under dense observation time series, providing new ideas and means for dynamic monitoring of typical feature information.

Previous studies have shown that SAR backscattering and time-series feature decomposition can identify cultivated land types, but often cannot achieve satisfactory accuracy based on a single feature input. In order to achieve complementary advantages between different features, it is necessary to effectively couple SAR images with spectral features. Due to the significant difference between scattering features and spectral features, it is impossible to extract multidimensional depth features based on a single convolutional neural network. Therefore, two convolutional branch networks were designed to extract scattering features and spectral temporal features of SAR data, respectively. Under the action of the feature fusion module, different depth features extracted by multiple branches were automatically fused.

A dual branch convolutional neural network consists of two structurally similar neural network branches that extract depth features from scattering and spectral time-series data, each of which involves multiple convolutional layers, pooling layers, and nonlinear activation functions. Convolutional neural networks are used to extract deep features of images through a series of periodic convolution, pooling, and nonlinear transformation operations. During the convolution process, convolutional kernels are used to learn deep features from the input SAR image. The convolutional kernels of shallow convolutional layers can only learn low-level features of the input image due to their small receptive field. As the network layers deepen, the receptive field continues to increase. Furthermore, the model can learn deeper feature representations of SAR images. At the same time, during the convolution process, pooling and non-linear operations are inserted. Pooling enables the network to remove secondary features, retain primary features, and rapidly reduce the number of model parameters and computational complexity. Nonlinear operations increase the non-linear factors of neural networks, enabling the network model to achieve more complex feature extraction functions.

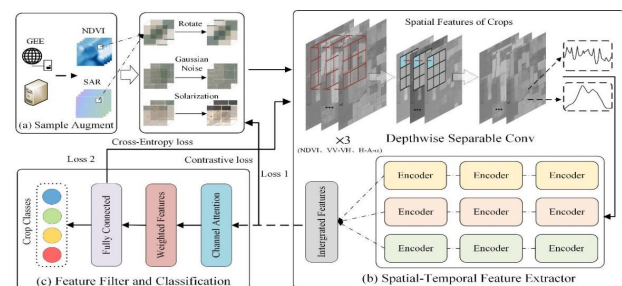


Figure 1. Flowchart of the proposed method.

2.1 Deep Learning Model Construction

In order to construct classification samples from optical SAR time series images, we randomly extract fixed size square image blocks in the spatial domain along the time axis, and use the data corresponding to their central pixels as classification labels. Subsequently, data augmentation techniques (Solarization, Gaussian noise, and rotation) are used to enhance the generated sample set (each pair of samples includes the original sample and corresponding enhanced samples, with other samples as their negative samples). Then, perform feature extraction on all samples for the input model. For each branch of the multi branch spatiotemporal feature extraction module, the deep spatial features of the samples are first extracted through convolutional layers, and then the corresponding temporal features are further extracted through the improved Transformer

module. In order to reduce the computational complexity of the model, a more lightweight depthwise separable convolutional network was adopted. It consists of two parts: channel convolution and point by point convolution.

Assuming all samples (including data enhanced samples) can be represented as $[S_1, S_2, S_3]$, where represents different data types (S_1 represents NDVI time series, S_2 represents backward scattering data of Sentinel 1, and S_3 represents polarization decomposition data of Sentinel 1). For a sample set x , its time dimension is M . Firstly, use M convolution kernels of size to extract the corresponding spatial features, where k is the size of the convolution kernel, M is the length of the time dimension of each input sample, H and W represent the length and width of the sample. The specific spatial feature extraction calculation process can be expressed as formula 1.

$$DConv = [d_1, d_2, \dots, d_M], d_s = \sum_{i=0}^H \sum_{j=0}^W Dw_{i,j}^l \cdot x_{i+k,j+k}^l \quad (1)$$

Where d_s represents the s -th spatial feature map, $Dw_{i,j}^l$ represents the weight matrix of the l -th layer in the first module, (i, j) represents the coordinate position of the output feature map in space, and $x_{i,j}^l$ represents the input feature corresponding to the l -th layer position. Next, perform point convolution to generate integrated feature maps. Using E convolution kernels of $1 \times 1 \times M$ size, the features of each pixel are weighted and combined in the time dimension.

2.2 Land Cover Classification

Due to the absolute disadvantage of the number of training samples in practical engineering applications relative to the total number of pixels in the entire remote sensing image, the model has added a self supervised contrastive learning scheme to stabilize the training process. Specifically, the model stabilizes the training process by making the sample features in different transformations as close as possible, while keeping the corresponding features of different samples as far apart as possible. Among them, cosine similarity is used to calculate the similarity between different features, and the calculation process is as follows:

$$sim(z_i, z_j) = \frac{z_i \cdot z_j}{\|z_i\| \|z_j\|} \quad (2)$$

On this basis, a loss function L_1 in the model was constructed for all sample features extracted from the training dataset, as shown below:

$$L_1 = -\log \frac{\exp(sim(z_i, z_j) / \tau)}{\sum_{k=1}^{2R} 1_{[k \neq i]} \exp(sim(z_i, z_k) / \tau)} \quad (3)$$

Among them, z_k represents the negative sample depth features from different samples, and τ represents the temperature coefficient. R is the number of all samples in each batch of training.

In order to construct an efficient multimodal feature classifier, we calculate the equivalent weights of fused features based on Channel Attention (CAAtt) to measure the contribution of

different features to model classification. Unlike the MSA module, CAAtt calculates the weights of different channels through a ReLU form gating mechanism, rather than initializing three learnable matrix parameters. Meanwhile, it combines two fully connected layer functions to reduce the complexity of model computation.

3. Experiment

3.1 Study Areas and Data Preparation

Based on the spatial distribution of surface cover types, two overseas experimental areas were selected. The first experimental zone is located in the central part of Thailand, in the western part of the Nakhon Rath Plateau, about 259 kilometers away from the capital Bangkok. It covers an area of approximately 20494 square kilometers and is the largest province in Thailand, containing rich red ceramic culture and history. Dong Phraya Yen is the birthplace of multiple rivers in the northeast, with a dry and hot climate. The experimental area selected for this area has an area of approximately 13403.87km². The second pilot zone is located in Punjab Pradesh, Ludhiana, Pakistan. It is adjacent to Pakistan to the west. Punjab means the land of the five rivers, which refers to the confluence of the five tributaries of the Indus River, namely the Jhelum River, the Janab River, the Ravi River, the Bias River, and the Satryj River. Covering an area of 50400 square kilometers. The population is approximately 16.79 million, mostly Sikh. The capital is Chandigarh. 95% of the entire area is plain, gently dipping from northeast to southwest. In the north, there are tributaries of the Siwalik Mountains, with an altitude of 300-600 meters. The Ravi River, Bias River, and Sutlej River pass through the area. Belonging to the subtropical continental climate, there are drastic changes in cold and heat.

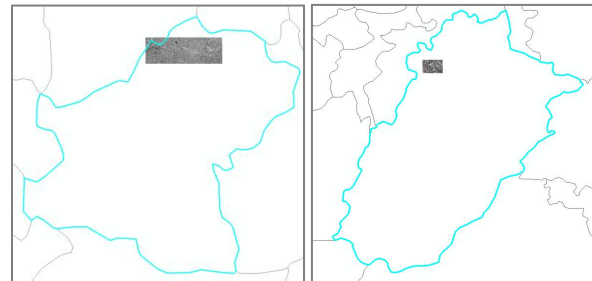


Figure 2. Study areas located in Thailand and Pakistan

We obtained domestic optical remote sensing images ZY-3 or GF, foreign optical Sentinel-2 images, and foreign SAR Sentinel-1 images covering the experimental area. The training sample data is mainly based on the classification results of global 10 meter surface cover data, with a time span between 2017 and 2020. The typical types of surface cover mainly include seven major types: cultivated land, forest land, shrub land, grassland, artificial surface, desert and bare surface, and water body.

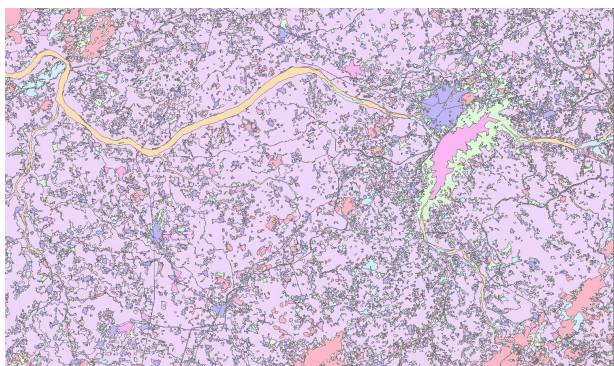


Figure 3. land cover data

Remote sensing images require a certain degree of preprocessing before processing, in order to achieve better radiometric and geometric accuracy, improve image interpretability and visual effects, and provide necessary benchmarks for data input into deep learning models. Specifically, the preprocessing in surface coverage information extraction includes some basic remote sensing image preprocessing steps, such as radiometric correction, geometric correction, etc. In addition, in order to ensure that the size of remote sensing image data generally exceeds the input size of commonly used deep learning models, and to achieve a unified and appropriate input image size for each model, it is necessary to embed and crop remote sensing images to the same size.

Currently, Sentinel-1 radar data Level-0 and Level-1 level archived data can be downloaded for free. When we use the Level-1 ground distance data of Sentinel 1 in different modes for research, the preprocessing generally uses the Sentinel series data processing software SNAP developed by the European Space Agency and the remote sensing image processing platform ENVI 5.3. The main preprocessing steps include: applying orbital files, removing edge noise, removing thermal noise, speckle filtering, radiometric calibration, terrain correction, geocoding, and image cropping. Although thermal noise removal has been performed when generating GRD grade product data, there is still a large amount of thermal noise in GRD products that requires multiple repeated processing. Except for image cropping using ENVI 5.3 software, all other operations are performed using SNAP software.

3.2 Experimental Designs

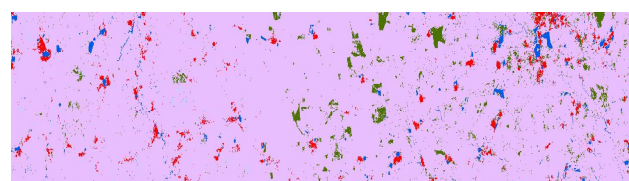
Firstly, we used real label data to calculate the proportion of different land cover types in the experimental area. Then, we divided the image into two parts, with the upper part used for training and the lower part used for prediction. We adopt a random sampling strategy and capture fixed sized patches at the center point. The extracted samples are divided into completely independent training, testing, and validation sets in a ratio of 5:3:2.

For the proposed model, the sample size is set to 16. In each branch, we use 256 convolution kernels with sizes of 33 and 11 * 256 to extract spatial features from the input samples. When extracting temporal features, based on experimental results and experience, the depth of the encoder is set to 3, the number of MSA heads is set to 8, and the input node of FFN is set to 1296. Finally, the combination of spatiotemporal features from different branches is input into the feature converter to output surface category information. At the same time, the above models all used the same Python framework during the training process. The optimizer used in all models is Adam, which is set to beta by default_1=0.9, beta_2=0.999, epsilon=1e-08. The

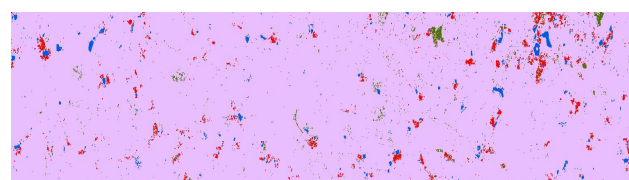
initial learning rate is set to 0.0001. Batch_Set the size to 400 and the training epoch to 70. The model training was completed on a server equipped with CentOS 7.6 operating system, two NVIDIA Tesla 100 16G graphics cards, and an Intel (R) Xeon (R) Gold 5118 CPU. The Python version used is 3.7.

3.3 Experimental Results

We collected and processed multimodal data from central Thailand in the experimental area, including optical ZY3 and foreign SAR Sentinel-1. Using the multimodal data fusion model method proposed in this project, ZY3+Sentinel-1 (hereinafter referred to as Combination 1) and Sentinel-1 (hereinafter referred to as Combination 2) were inputted into the model to obtain the classification results of typical land cover features under two combination strategies, as shown in the figure. In order to further verify the reliability of the surface cover classification results, a quantitative evaluation of the surface cover classification results for each period is conducted. We used real label data to calculate the proportion of different surface cover types in the experimental area. Then, we divided the image into two parts, with the upper part used for training and the lower part used for prediction. We adopt a random sampling strategy and extract patches of fixed size from the center point. The extracted samples are divided into completely independent training, testing, and validation sets in a 5:3:2 ratio. The distribution quantity of test and validation sets for each surface coverage type is shown in the table.



(a) ZY3+Sentinel-1



(b) Sentinel-1

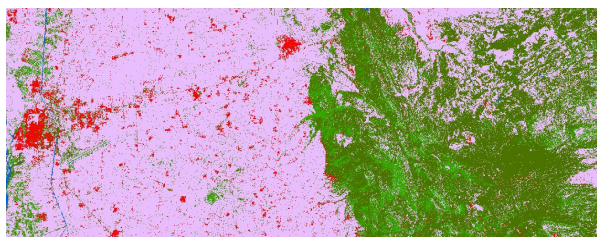
Figure 4. Classifications results of land cover.

Firstly, from the comparison results of the images, it can be seen that the combination of optics and SAR can effectively extract land cover, while the use of pure SAR has the least spatial detail information in the image, which reflects farmland and artificial land better. Furthermore, as shown in Table 4, the accuracy evaluation results show that the classification accuracy of cultivated land is the highest under both combinations, both above 97%, followed by forest land and artificial surface. From the reflection of land types under different combinations, the combination of optics and SAR has a higher effect on artificial land than other combinations. The extraction of pure SAR images has the highest accuracy on cultivated land, while other land types are significantly lower than other combinations. In addition, due to the absence of shrublands, deserts, and exposed surfaces in the study area, accuracy evaluation and analysis will not be conducted temporarily.

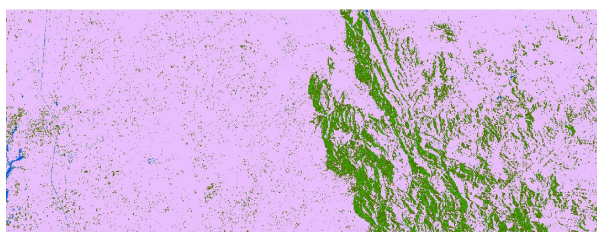
class	Methods/ F-1
	Our method
Cropland	88%
forest	84%
grassland	80%
Artificial surface	92%
waterbody	93%
Overall Accuracy	88%
Kappa	0.85

Table 1. Classification results of land cover data

We collected and processed multimodal data from Punjab, Pakistan, including foreign optical Sentinel-2 and foreign SAR Sentinel-1 in experimental zone 2. Using the proposed multimodal data fusion model method, ZY3+Sentinel-1 (hereinafter referred to as Combination 1) and Sentinel-1 (hereinafter referred to as Combination 2) were inputted into the model to obtain the classification results of typical land cover features under two combination strategies. In order to further verify the reliability of the surface cover classification results, a quantitative evaluation of the surface cover classification results for each period is conducted. We adopt a random sampling strategy and extract patches of fixed size from the center point. The extracted samples are divided into completely independent training, testing, and validation sets in a 5:3:2 ratio. The distribution quantity of test and validation sets for each surface coverage type is shown in the table.



(a) ZY3+Sentinel-1



(b) Sentinel-1

Figure 5. Classifications results of land cover.

In this experimental area, it can be seen that the ground cover classification results extracted by Combination One are roughly equivalent in spatial details and information richness, without large areas, and can reflect the true distribution of land features in the image. However, in the combination of two SAR images, the artificial surface was almost unable to be extracted, which may be due to the lack of clear boundaries in the echo signals of the artificial surface in SAR images, which is similar to the performance of surrounding land types. Furthermore, from the table, it can be seen that the accuracy evaluation results show that the classification accuracy of cultivated land under the two combinations is the highest, both above 94%, followed by water bodies and artificial surfaces. From the reflection of land types under different combinations, the combination of optics and

SAR has higher accuracy in water bodies than other combinations. SAR image extraction is only sensitive to farmland and water bodies, and other land types are significantly lower than other combinations.

class	Methods/ F-1
	Our method
Cropland	84%
forest	88%
grassland	82%
Artificial surface	91%
waterbody	83%
Overall Accuracy	86%
Kappa	0.83

Table 2. Classification results of land cover data

4. Conclusion

We evaluated our proposed method on a High resolution remote sense dataset of land cover data. The experimental results show that our proposed method outperforms traditional methods and achieves high accuracy in identifying land cover classification. In addition, we conducted a sensitivity analysis to evaluate the impact of different factors on the performance of our method. The results show that our method is robust to changes in the input data and the network architecture.

The proposed method has several advantages over traditional methods. First, it is automated and does not require manual surveys, which can save time and cost. Second, it can provide a more comprehensive and accurate understanding of the spatial distribution of land cover data. In future work, we plan to explore the use of additional data sources, such as INSAR data and Hyperspectral data, to further improve the accuracy of our method. Overall, our proposed method provides a promising approach for exploring land cover based on multi-source remote sensing images and cross-modal networks.

Acknowledgements

This research was supported by the Special funding for science and technology in Science and Technology City, Sanya Yazhou Bay (Grant No. SKJC-2022-010-03), and International Science & Technology Cooperation Program of Hainan Province (GHYF2024002).

References

- Chen J, Lu M, Chen X, et al., 2013: A spectral gradient difference based approach for land cover change detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 85:1-12.
- Claverie M, Ju J C, Jeffrey G M, et al., 2018: The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote sensing of environment*, 219: 145-161.
- Claverie M, Vermote E, Franch B, et al., 2015: Evaluation of medium spatial resolution BRDF-Adjustment techniques using multi-angular SPOT4 (Take5) acquisitions. *Remote Sensing*, 7 (9): 12057-12075.

- Dong C, Loy C C, He K M, et al., 2014: Learning a deep convolutional network for image super-resolution. In European conference on computer vision, 184-199.
- Gao J H, Yuan Q Q, Li J, et al., 2020: Cloud removal with fusion of high resolution optical and sar images using generative adversarial networks. *Remote Sensing*, 12 (1): 191.
- Gómez C , White J C , Wulder M A., 2016: Optical remotely sensed time series data for land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 116:55-72.
- Goldewijk K K., 2001: Estimating global land use change over the past 300 years: The HYDE Database. *Global Biogeochemical Cycles*, 15(2): 417-433.
- Gong M, Zhang P, Su L, et al., 2016: Coupled dictionary learning for change detection from multi-source data. *IEEE Transactions on Geoscience and Remote Sensing*, 54(12) : 7077 -7091.
- Griffiths P, Nendel C, Hostert P., 2019: Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping. *Remote sensing of environment*, 220: 135-151.
- Hussain M C, Loveland T R., 2012: A review of large area monitoring of land cover change using Landsat data. *Remote Sensing of Environment*, 122:66-74.
- Hussain Ma, Chen DM, Cheng M, et al., 2013: Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS Journal of Photogrammetry and Remote Sensing*, 80, 91-106.
- Kennedy Robert E, Yang Z Q, Cohen W B., 2010: Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr—Temporal segmentation algorithms. *Remote Sensing of Environment*, 114 (12): 2897-2910.
- Liu J, Gong MG, Qin K, et al., 2018: A deep convolutional coupling network for change detection based on heterogeneous optical and radar images. *IEEE Transactions on Neural Networks and Learning Systems*, 29(3): 545-559.
- Liu, JY, Kuang WH, Zhang ZX, et al., 2014: Spatiotemporal characteristics, patterns, and causes of land-use changes in China since the late 1980s. *Journal of Geographical Sciences*, 24(2): 195-210.
- Roy David P, Ju JC, Lewis Philip, et al., 2008: Multi-temporal MODIS-Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data. *Remote Sensing of Environment*, 112 (6): 3112-3130.
- Roy David P, Zhang HK, Ju JC, et al., 2016: A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance. *Remote Sensing of Environment*, 176: 255-271.
- Shao ZF, Cai JJ, Fu P, et al., 2019: Deep learning-based fusion of Landsat-8 and Sentinel-2 images for a harmonized surface reflectance product. *Remote Sensing of Environment*, 235: 111425.
- Tewkesbury A P, Comber A J, Tate N J, et al., 2015: A critical synthesis of remotely sensed optical image change detection techniques. *Remote Sensing of Environment*, 160, 1-14.
- Vogelmann J E, Gallant A L, Shi H, et al., 2016: Perspectives on monitoring gradual change across the continuity of Landsat sensors using time-series data. *Remote Sensing of Environment*, 185:258-270.
- Wang L, Xu X, Yu Y, et al., 2019: SAR-to-optical image translation using supervised cycle-consistent adversarial networks. *IEEE Access*, 7: 129136-129149.
- Wang QM, Blackburn G A, Onojeghuo A O, et al., 2017: Fusion of Landsat 8 OLI and Sentinel-2 MSI data. *IEEE Transactions on Geoscience and Remote Sensing*, 55 (7): 3885-3899.
- Willis K S., 2015: Remote sensing change detection for ecological monitoring in United States protected areas[J]. *Biological Conservation*, 182 : 233-242.
- Zhao WZ, Qu Y, Chen JG, et al., 2020: Deeply synergistic optical and SAR time series for crop dynamic monitoring. *Remote Sensing of Environment*, 247: 111952.
- Zhao W, Wang Z, Gong M, et al., 2017: Discriminative feature learning for unsupervised change detection in heterogeneous images based on a coupled neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 55(12) : 7066 -7080.