Brain-Inspired Remote Sensing Interpretation: A Comprehensive Survey

Licheng Jiao, *Fellow, IEEE*, Zhongjian Huang, *Student Member, IEEE*, Xu Liu, *Member, IEEE*, Yuting Yang, *Graduate Student Member, IEEE*, Mengru Ma, Jiaxuan Zhao, *Graduate Student Member, IEEE*, Chao You, Biao Hou, *Member, IEEE*, Shuyuan Yang, *Senior Member, IEEE*, Fang Liu, *Senior Member, IEEE*, Wenping Ma, *Senior Member, IEEE*, Lingling Li, *Senior Member, IEEE*, Puhua Chen, *Senior Member, IEEE*, Zhixi Feng, *Member, IEEE*, Xu Tang, *Senior Member, IEEE*, Yuwei Guo, *Senior Member, IEEE*, Xiangrong Zhang, *Senior Member, IEEE*, Dou Quan, *Member, IEEE*, Shuang Wang, *Senior Member, IEEE*, Weibin Li, Jing Bai, *Senior Member, IEEE*, Yangyang Li, *Senior Member, IEEE*, Ronghua Shang, *Senior Member, IEEE*, and Jie Feng, *Senior Member, IEEE*

(Review Paper)

***Abstract*—Brain-inspired algorithms have become a new trend in next-generation artificial intelligence. Through research on brain science, the intelligence of remote sensing algorithms can be effectively improved. This article summarizes and analyzes the essential properties of brain cognize learning and the recent advance of remote sensing interpretation. First, this article introduces the structural composition and the properties of the brain. Then, five represent brain-inspired algorithms are studied, including multiscale geometry analysis, compressed sensing, attention mechanism, reinforcement learning, and transfer learning. Next, this article summarizes the data types of remote sensing, the development of typical applications of remote sensing interpretation, and the implementations of remote sensing, including datasets, software, and hardware. Finally, the top ten open problems and the future direction of brain-inspired remote sensing interpretation are discussed. This work aims to comprehensively review the brain mechanisms and the development of remote sensing and to motivate future research on brain-inspired remote sensing interpretation.**

***Index Terms*—Brain modeling, deep learning, image processing, remote sensing.**

Manuscript received 21 December 2022; revised 9 February 2023; accepted 17 February 2023. Date of publication 22 February 2023; date of current version 30 March 2023. This work was supported in part by the Key Scientific Technological Innovation Research Project by Ministry of Education, the State Key Program and the Foundation for Innovative Research Groups of the National Natural Science Foundation of China under Grant 61836009, in part by the Major Research Plan of the National Natural Science Foundation of China under Grant 91438201, Grant 91438103, and Grant 91838303, in part by the National Natural Science Foundation of China under Grant U22B2054, Grant U1701267, Grant 62076192, Grant 62006177, Grant 61902298, Grant 61573267, Grant 61906150, and Grant 62276199, in part by the 111 Project, the Program for Cheung Kong Scholars and Innovative Research Team in University, under Grant IRT 15R53, in part by the ST Innovation Project from the Chinese Ministry of Education, the Key Research and Development Program in Shaanxi Province of China under Grant 2019ZDLGY03-06, in part by the National Science Basic Research Plan in Shaanxi Province of China under Grant 2022JQ-607, in part by the China Postdoctoral fund under Grant 2022T150506, and in part by the Scientific Research Project of Education Department In Shaanxi Province of China under Grant 20JY023. *(Corresponding authors: Zhongjian Huang; Licheng Jiao.)*

The authors are with the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education of China, International Research Center of Intelligent Perception and Computation, School of Artificial Intelligence, Xidian University, Xi'an 710071, China (e-mail: lchjiao@mail.xidian.edu.cn; huangzj@stu.xidian.edu.cn; xuliu361@163.com; ytyang_1@stu.xidian.edu.cn; mengrumalearn@163.com; jiaxuanzhao@stu.xidian.edu.cn; cy_chaoyou@163.com; avcodec@163.com; syyang@xidian.edu.cn; f63liu@163.com; wpma@mail.xidian.edu.cn; llli@xidian.edu.cn; phchen@xidian.edu.cn; zxfeng@xidian.edu.cn; tangxu128@gmail.com; yuweiguo18@126.com; xrzhang@mail.xidian.edu.cn; dquan@stu.xidian.edu.cn; shwang.xd@gmail.com; weibinli@xidian.edu.cn; baijing@mail.xidian.edu.cn; yyli@xidian.edu.cn; rhshang@mail.xidian.edu.cn; jiefeng@xidian.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3247455

I. Introduction

REMOTE sensing is a technology that observes and detects the objects on the Earth by the sensors equipped on aircraft or satellites [1], [2]. It is a noncontact, long-distance detection technology that began in the 1960s [3]. It uses visible light, infrared, and electromagnetic waves radiated or the reflection by the target itself to perceive and identify the target at a long distance. The remote sensing data obtained by remote sensing technology enhances the ability of human beings to research the Earth [4]. At the same time, remote sensing applications involve many fields. It is widely used in various military and civilian areas, such as satellite surveillance, land and resources survey, land use and land cover, urban dynamic change monitoring, meteorological monitoring, environmental assessment and monitoring, and disaster investigation and evaluation. This dramatically expands the critical impact of remote sensing on human production and life [5].

Nowadays, we face many challenges in remote sensing interpretation. First, due to the quickening growth of unmanned aerial vehicles (UAV) and satellite technology in recent years, the amount of data has increased dramatically [6]. The spectral, spatial, and temporal dimensionalities of the data require more computing resources [7]. In addition, large, labeled datasets in remote sensing are not easily obtained. This restricts the use of larger models to improve the accuracy of the algorithms. Last but not least, the interpretability of algorithms is necessary for remote sensing interpretation [8].

In recent years, artificial intelligence technology has improved the accuracy and efficiency of remote sensing interpretation. Faced with massive, complex, and diverse remote sensing data,

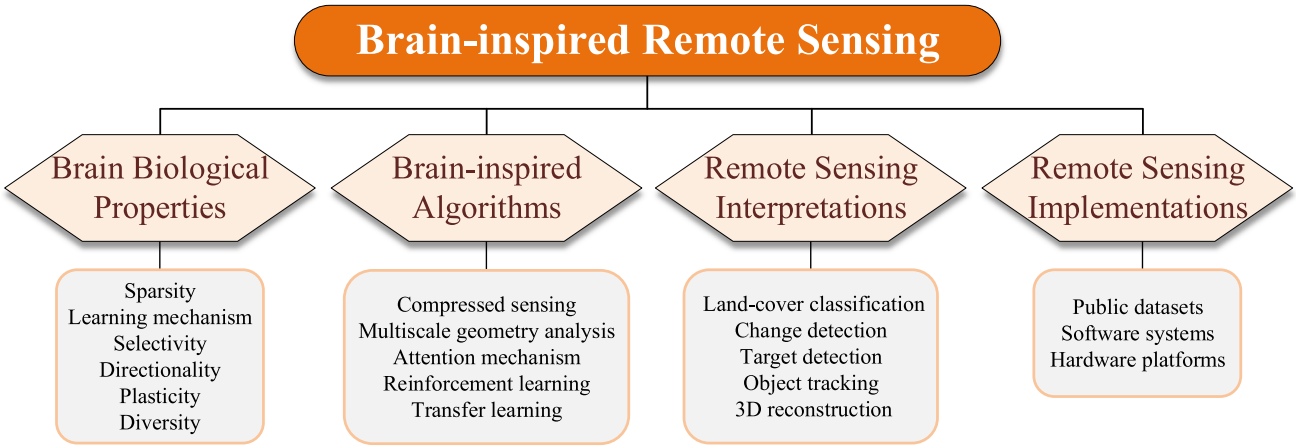


Fig. 1. Organizational framework of this review.

artificial intelligence has realized automatic feature extraction, parameter learning, and classification.

Artificial intelligence aims to study and develop computer algorithms that can handle tasks requiring human intelligence. Its development is closely related to brain science [9]. Brain science is to study the structure, function, and operation mechanism of the biological brain and further understand how the brain processes information, mines knowledge, and makes decisions. Artificial intelligence draws inspiration from brain science and designs intelligent algorithms.

In 1943, neuroscientist W.S. McCilloch and mathematician W. Pitts established the MP model, an abstract and simplified model constructed according to the structure and working principle of biological neurons. The so-called "simulated brain" was born [10]. In 1949, Hebbian learning was proposed. This algorithm is inspired by the dynamics of biological nervous systems. According to the study, a synapse between two neurons is strengthened when the neurons on either side of the synapse (input and output) have highly correlated outputs. Hebbian learning learns from this property and improves the weight between two highly correlated neurons during the learning process [11]. In 1958, perceptron was proposed to model the way information is stored and organized in the brain [12]. In 1983, physicist John Hopfield proposed a neural network for Associative Memory called the Hopfield network [13]. In 2006, Geoffrey Hinton proposed a multilayer neural network for data reduction, which opened the curtain of deep learning research [14].

The research on artificial intelligence is closely related to the brain. These algorithms, inspired by the structure and characteristics of the brain, continue to promote the development of artificial intelligence. Artificial intelligence is also constantly looking for new inspiration from biological brains.

A. Motivation

In recent years, as more and more diverse neural networks have been proposed, people have paid more attention to the design of brain-inspired algorithms, and many reviews of brain-inspired algorithms have been proposed. Hassabis et al. [15] analyzed the historical interaction between artificial intelligence and neuroscience fields, providing new perspectives to develop artificial intelligence. Yang et al. [16] provided a comprehensive review of the research of brain-inspired artificial intelligence and its related engineering technique. Strisciuglio and Petkov [17] focused on the relationship between research in neuroscience and advances in computer vision. Simeone et al. [18] organized a special section to introduce machine learning (ML) and signal processing algorithms for brain-inspired computing. Fan et al. [9] researched new brain imaging techniques to explore the secrets of brain science and built brain dynamic connectivity maps. Jiao et al. [19] discussed the main problems and applications of bio-inspired computation and recognition, introducing algorithm implementation, model simulation, and practical application of parameter setting. Tianyuan et al. [20] introduced the relationship between artificial intelligence and neuroscience, the research status of brain-inspired intelligence, and the profound influence of artificial intelligence in other fields.

The characteristics of the brain and brain-inspired algorithms are worth discussing. The brain-inspired algorithms are developed according to the research on the latest brain characteristics and improve performance, efficiency, and interpretability. This will provide a new perspective for remote sensing interpretation. In this review, we mainly investigate the features of the brain and introduce the related brain-inspired algorithms. In addition, the interpretation (data types and main applications of remote sensing) and implementation (public datasets, software, and hardware) are presented. We attempt to summarize the characteristics of the brain and discuss remote sensing tasks to provide readers with new perspectives on remote sensing data analysis and promote the design of brain-inspired algorithms. The main contributions of the present review can be summarized as follows.

1) We provide a comprehensive survey of brain structure and summarize the brain properties as sparsity, learning mechanism, selectivity, directionality, plasticity, and diversity.
2) This survey investigates five essential applications in remote sensing data interpretation, including object classification, target detection, change detection, video tracking, and 3-D reconstruction. These methods cover image tasks in remote sensing, as well as video and point cloud data developed in recent years.

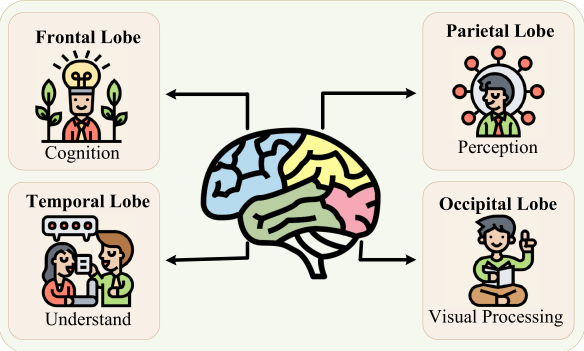


Fig. 2. Four major functional areas of the cerebral cortex.

3) The public datasets and an overview of related software and hardware are summarized.
4) Current challenges and future research directions are presented.

The rest of this article is organized is as follows (as shown in Fig. 1). The basic structure and characteristics of the brain and the brain-inspired algorithms are presented in Sections II and III. In Section IV, the data types, such as optical images, radar images, airborne light detection and ranging (LiDAR), and remote sensing videos, are summarized. In addition, the latest advances in the five applications of remote sensing are presented. In Section V, the public datasets, software platforms, and hardware resources required to implement the algorithms are discussed. We discuss the future challenges and directions of combining brain mechanisms with remote sensing interpretation in Section VI. Finally, Section VII concludes this article.

II. Theory of the Brain

A. Biological Structure of the Brain

The brain is the principal organ in the central nervous system. It is mainly composed of the cerebral cortex, cerebellum, diencephalon, and brainstem. Among them, the cerebral cortex is the most advanced part of conscious thinking and sensory processing, and it is also the main part of the brain. It has the ability to recognize, represent and learn. It contains four functional areas: temporal lobe, occipital lobe, frontal lobe, and parietal lobe [21]. The specific division is shown in Fig. 2.

1) *Occipital lobe:* It is the visual processing center of the brain, including low-level visuospatial processing (position, spatial frequency), color discrimination, and motion perception.
2) *Temporal lobe:* It is responsible for processing sensory input using visual memory, language, and emotional connections to derive higher level information.
3) *Parietal lobe:* It can process various sensory information, including touch, smell, taste, etc. It is also related to language and memory.
4) *Frontal lobe:* It is the most advanced part of brain development and has advanced cognitive functions. It is mainly responsible for the processes of movement, cognition, and thinking. It is capable of tasks, such as attention, judgment, thinking, analysis, calculation, and planning, and is related to human needs and emotions.

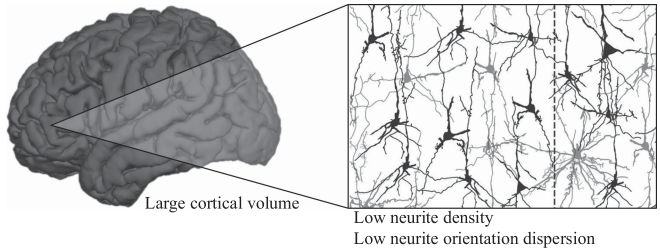


Fig. 3. Microstructure of neurons and neurites in the brain. It indicates a degree of sparsity in the cerebral cortex. (Image from [22]).

The cerebral cortex facilitates the development and computation of neural networks. Brain perception and cognition are the biological basis, providing new ideas for the efficient and accurate realization of artificial intelligence perception and understanding. Unfortunately, these natural biological properties are not fully considered in current neural network designs. Therefore, brain-inspired modeling and algorithm research is significant and can further promote the development of a new generation of artificial intelligence.

B. Biological Properties of the Brain

The research on the biological properties of the brain has opened a new window for brain-inspired remote sensing. Analyzing the biological properties of the human learning mechanism can help us establish a variety of algorithms to simulate the brain. Understanding the brain mechanism has recently been a significant new development trend in the international academic community. For the perception and cognition of knowledge, the brain mainly has biological properties, such as sparsity, learning mechanism, selectivity, directionality, plasticity, and diversity.

1) Sparsity: The biological brain, especially the human brain, is a hierarchical, sparse, and periodic structure [22], as shown in Fig. 3. Sparsity plays an important role in biological brains. Olshausen and Field [23] presented the neuron sparse coding theory. In 2007, Huber et al. [24] and Houweling and Brecht [25] tested the hypothesis of "sparse coding" of neurons with rat experiments. The processing of scene information by the biological retina is sparse, which makes learning more efficient. In the brain's primary visual cortex (V1), researchers in computational neuroscience believe that sparse coding is the main way of image representation in the visual system. The neurons in the V1 are also sparse in the dynamic processing and computation of information. Simultaneously, the neurons in the V4 area realize the representation of visual information through sparse coding. The higher the level, the larger the receptive field, that is, the information processing is from a local to a larger area. When the level is low, the area processed by the receptive field is smaller, and the sparsity is stronger, and vice versa.

2) Learning Mechanism: The human brain is good at rapid cross-task learning and generalized cognition. In 2011, Tenenbaum et al. [26] pointed out that the brain has a strong ability for abstract representation and can learn generalized knowledge from a small amount of data. In the brain, the region responsible

for cognition and learning is mainly the hippocampus. Cells in the hippocampus are interconnected into networks, each of which is defined by a more abstract grid of cells. Based on these abstract templates for expressing relationships and symbols, it is easy for the brain to directly apply the existing abstract templates and recombine them to understand new things when receiving external environmental stimuli or tasks.

The human brain stores a vast amount of knowledge about the world that underlies language, thought, and reasoning. There are two kinds of knowledge representation in the human brain, sensory and language derived. The ability to form memories is a key to learning and knowledge accumulation. In 2020, Josselyn and Tonegawa [27] explored evidence of engram cells as the basis of memory (especially in rodents), investigating how new information is integrated into existing knowledge memory.

3) Selectivity: Roelfsema [28] pointed out that the brain has the ability to pay attention to special things and autonomously control the attention area in a new environment.

Selective attention modulates neuronal activity in nearly all brain structures responsible for visual processing, including ventral pathways (from V1 through extrastriatal cortex (V2–V4) to inferotemporal cortex), dorsal pathways (from V1 to V2 to the middle and medial temporal lobes and parietal lobes responsible for motor information processing), prefrontal lobes, subcutaneous structural nuclei, such as lateral geniculate body, superior colliculus, occipital nucleus, dorsomedial thalamus, and reticular nucleus of thalamus, striatum, and substantia nigra reticularis.

At the same time, the brain receives a large amount of information. However, it cannot process all the information entering the system with the same degree of priority. Only some information can be filtered and processed through selective attention and enter consciousness. For example, the primary visual cortex can generate visual saliency maps in the very early stages of visual information processing to guide the distribution of spatial selective attention, regulate sensory input, and improve people's perception and behavior. In addition, selective attention has various regulatory effects on the neural representation of target stimuli, such as enhancing neuronal firing and firing synchronization, enhancing neuronal selectivity, enhancing neuronal signal-to-noise ratio, and moving and reducing neuronal receptive fields. Therefore, selective attention is a deeply sophisticated cognitive process that always coordinates the brain's cognitive processing.

4) Directionality: In 1971, O'Keefe discovered in the course of experiments that there are "place cells" in the hippocampus that can record location information, which can be selectively activated to give specific locations a special identity. In the mid-1980s and early 1990s, the "head orientation cells" were discovered that determine the orientation of the head, marking orientation with selective excitation. At the same time, the "grid cell" that can delineate a plane coordinate system was also discovered, which can record all the position information generated during the movement, etc. These cells cooperate with each other to create a 2-D map of the brain, the material basis for cognitive maps. In 2015, Finkelstein et al. [29] pointed out that there are azimuthal and oblique angle cells in the brain that can perceive direction and position information.

5) Plasticity: The brain will change the internal neural mechanism due to the needs of the external environment, that is to say, the brain is constantly assimilation and accommodating, so the brain has plasticity [30]. Brain plasticity refers to the ability of the brain to be modified by environment and experience. It can be divided into structural plasticity and functional plasticity. The structural plasticity of the brain means that the connections between synapses and neurons within the brain can establish new connections due to the influence of learning and experience, thereby affecting the behavior of individuals. It includes neuronal plasticity and synaptic plasticity. Functional plasticity can be understood in that through learning and training, the function of a representative area of the brain can be replaced by adjacent brain areas, and it is also manifested in the recovery of brain function in patients with brain injury to a certain extent after learning and training. Brain plasticity is closely related to learning and memory.

6) Diversity: The diversity of neurons is the basis for the complex and delicate functions of the brain. In 2021, Berg et al. [31] used techniques, such as patch clamp, to reveal the richness of neuronal types in the cortex. In 2021, Yao et al. [32] constructed the mouse primary motor cortex, characterized more than 56 neuron types, and analyzed the developmental mechanism of the diversity of interneurons in the human brain. He also discovered the interneuron precursor cell types that exist specifically in the human brain and revealed the richness and diversity of human brain interneurons compared with other species.

III. Theory of the Brain-Inspired Algorithms

In this section, we discuss related brain heuristic theories from the perspective of brain properties. First, multiscale geometric analysis and compressed sensing (CS) have been extensively studied due to sparsity in the brain. The attention characteristics inspired the combination of attention mechanism and deep neural network to create SENet [33], nonlocal [34], transformer [35], and other networks. The training of artificial intelligence algorithms is enriched by reinforcement learning and transfer learning, which draw on the brain's natural learning process. This section starts from the abovementioned brain-inspired algorithms. It combines algorithms in remote sensing to provide readers with new ideas for combining brain-inspired algorithms and remote sensing.

A. Compressed Sensing

CS is a breakthrough theory for information acquisition. When the sampling rate is substantially lower than the "Nyquist" sampling rate, CS can still accurately reconstruct sparse signals with high probability. It gets discrete samples of the signal with random sampling and reconstructs them using a nonlinear reconstruction technique. Its core idea is mainly based on the sparse structure of the signal and the uncorrelated characteristics of the signal [36], [37], [38].

The sampling method of CS is a simple operation correlating a signal with a particular set of waveforms. These waveforms are independent in the sparse space. The CS method can directly

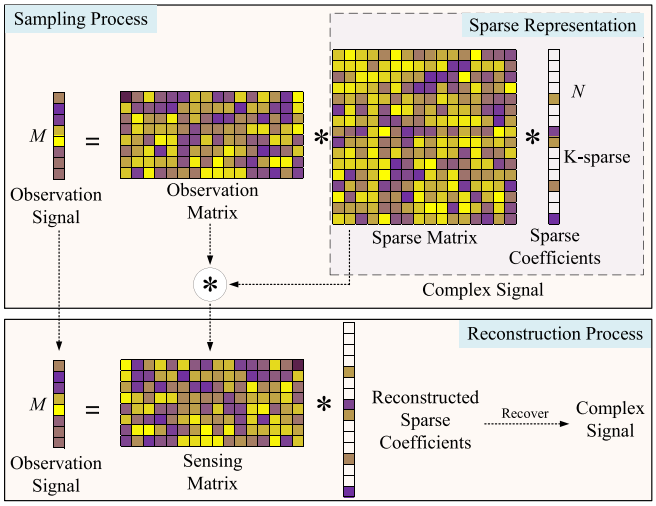


Fig. 4. Signal sampling and reconstruction processes of CS.

obtain compressed samples through the time domain transformation of the signal, which reduces the redundant information in the signal sampling process. The optimization algorithm is required to recover the original signal from the compressed samples. It is an underdetermined linear inverse problem where the signal is known to be sparse. Therefore, the prerequisite for realizing CS is that the signal is sparse in the frequency domain, and a random subsampling mechanism is adopted.

CS has two important operations to satisfy the above conditions: sparse representation and compressed observation. Sparse representation is the representation of complex signals as uncorrelated sparse signals. Compressed observation is to achieve random subsampling. Finally, sparse representation, compressed observation, and signal reconstruction constitute the three parts of the CS framework. To realize the CS, the sparseness of signals is the premise. The basis of CS is the compressed observation theory. The main components of CS are the reconstruction models and techniques [19].

The sampling and reconstruction processes of CS are shown in Fig. 4. In general, a complex signal can be represented as sparse coefficients, which satisfies the prerequisite of CS. Then, the observation signal is obtained by sampling with an observation matrix. During the reconstruction process, the observation signal and sensing matrix are known. A reconstruction algorithm is adopted to reconstruct the sparse coefficients. Finally, the complex signal can be recovered from the sparse coefficients.

1) Sparse Representation: The concept of sparse representation was first proposed in 1959 by Hubel and Wiesel [39] in their study of cellular receptive fields in the visual stripe cortex of cats. The experimental results established a precedent for sparse representation by showing that the receptive fields of cells in the "primary visual cortex" may provide a sparse response to visual perception information. In 1969, a sparse representation model based on Hebbian local learning principles was proposed [40]. The construction of the associative mechanism in the network structure benefits from the sparse representation's ability to maximize memory capacity. Houweling and Brecht et al. [25] conducted biological visual neurophysiological experiments that effectively supported the hypothesis of sparse neural coding.

According to the type of sparse matrix, the sparse representation methods of signals can be divided into the following three types: orthogonal transform basis method, multiscale geometric analysis method, and overcomplete dictionary method. To cover more signal types, the concept of the dictionary is proposed. Compared with the complete dictionary, the representation of the signal under the overcomplete dictionary is more sparse. The study of dictionary learning has grown in popularity in signal processing. There are two main ways to construct overcomplete dictionaries: using predefined analysis dictionaries (Heaviside, Gabor, Dirac, Fourier, and Wavelet dictionaries) or using dictionary learning algorithms (K-means, K-SVD algorithm, maximum likelihood estimation, and shift-invariant dictionary learning) [41].

2) Compression Measurement Matrix: The research focus of compressed observation theory is using a few nonadaptive observations to obtain enough signal information for reconstruction. Commonly used Gaussian random matrices and Bernoulli matrices belong to the category of random measurement matrices. Such matrices have high reconstruction accuracy but require large storage space and time complexity. Deterministic measurement matrices not only save storage space compared with random measurement matrices, but also are relatively easy to confirm whether they meet the Restricted Isometry Property criteria [42]. In addition, some deterministic measurement matrices can be obtained by applying a special structure. Corresponding fast algorithms can be designed to enhance the effectiveness of reconstruction. Partial Fourier matrices, structured measurement matrices, and partial Hadamard matrices are commonly used as deterministic matrices.

3) Sparse Reconstruction: Sparse reconstruction is an essential part of recovering the signal in CS. It needs to obtain the original signal through the compressed observation of the signal. Greedy, relaxation, and natural calculation methods are commonly used to solve the reconstruction problem.

The greedy method, also known as an iterative method, is an essential algorithm in solving sparse signal reconstruction problems. It uses an iterative method to approach the final solution gradually.

The convex relaxation reconstruction method is a kind of reconstruction method that has been widely studied and applied. It uses the l_1 norm to approximate the l_0 norm and simplifies the nonconvex optimization problem to the convex optimization problem. The convex optimization problem is easy to solve the reconstruction models.

Evolutionary algorithms have self-organization, self-adaptation, and self-learning capabilities. It can solve various complex problems that are difficult to solve in traditional computing methods without requiring complex reasoning calculations.

In the CS theory, signal sampling and compression can be performed simultaneously, discarding many redundant data during high-speed sampling. It dramatically reduces the sampling rate and computational cost of the sensor. As the key to CS theory, signal reconstruction is essential to solving NP-hard

problems. The evolutionary algorithm can be used to learn the optimal atomic combination in the dictionary direction, and the optimal atomic combination can be used to reconstruct the image. Meanwhile, the original optimization problem of CS is nonconvex and a combinatorial optimization problem. This solves the problem with the advantages of evolutionary algorithms and increases the flexibility and adaptability of the compressive sensing reconstruction algorithm.

CS is adopted to compress data usually in remote sensing. For example, hyperspectral images (HSIs) have high spectral resolution bringing a great challenge to the data storage and transmission [43]. Wang et al. [43] proposed a CS algorithm based on spectral unmixing. It samples the HSIs both spatially and spectrally and jointly optimizes the endmember extraction and abundance estimation. Xue et al. [44] designed a nonlocal tensor sparse and low-rank regularization approach for HSIs compressive sensing reconstruction. A subspace-based nonlocal tensor ring decomposition method is proposed for HSIs compressive sensing reconstruction [45]. Furthermore, Ghahremani et al. [46] leveraged the compressive sensing to pan-sharpen the low-resolution multispectral data with high-resolution panchromatic data.

B. Multiscale Geometry Analysis

Neuroscientists have shown that the receptive field of the mammalian visual cortex has local, directional, and band-pass characteristics [47]. The critical details in natural situations are only partially captured by neurons. Multiscale geometry uses the base functions to capture the partial detail of the signal. The base functions are rectangles, which can approximate the singular curve with the fewest coefficients and fully exploit the original function's geometric regularity. At the same time, the support interval's direction of base functions manifests the directionality of multiscale geometric analysis.

Multiscale geometry originated from wavelet analysis, beyond the wavelet analysis [48]. Wavelet analysis has achieved great success in various applications. The wavelet analysis can represent 1-D signals more sparsely than the Fourier analysis. However, in the case of 2-D or high-dimensional, wavelet analysis can only be formed into separable wavelets with limited directions, so it cannot achieve the optimal representation of high-dimensional signals. Multiscale geometric analysis is designed to solve this problem [49]. As shown in Fig. 5, a comparison of the contour representation with the wavelet analysis and multiscale geometric analysis is presented. The multiscale geometry analysis uses a more sparse representation to capture the 2-D contour.

Adaptive and nonadaptive are the categories under which the multiscale geometric analysis of pictures falls. The adaptive approach often starts with edge detection and uses the edge information to approximate the original function accurately. In fact, it is a combination of edge detection and image representation, such as Bandelet [51] and Wedgelet [52]. Nonadaptive methods do not use the geometric features of the image as a priori but directly decompose the image on a set of fixed base functions, eliminating the need for dependence on the image's structure.

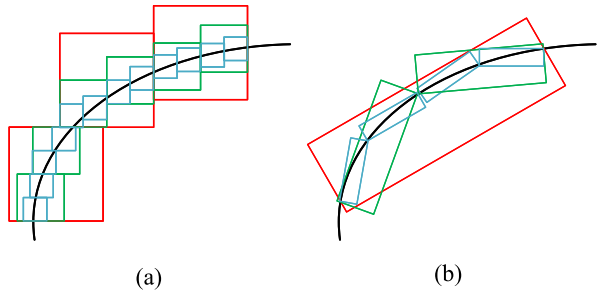


Fig. 5. Toy example of the contour representations to compare the difference between wavelet analysis and multiscale geometric analysis. The box in different colors presents a coefficient of the analysis on a specific scale. (a) Wavelet analysis. (b) Multiscale geometric analysis.

The represent algorithms are Ridgelet [53], Curvelet [54], and Contourlet [55].

The effort of fusing multiscale geometric analysis with neural networks is also growing with the emergence of deep learning. Contourlet CNN [56] is proposed to extract sparse and efficient representations of images. The contourlet transform (CT) is first used to extract the spectral features of the image and then fused with the spatial features extracted by the CNN network. Chen et al. [57] proposed ContourletNet to implement rain removal. It utilizes the multiscale, multidirectional, and hierarchical characteristics of CT to design a hierarchical multidirectional network, extracting multiple directional subbands and semantic subbands of different scales. The neural contourlet network [58] utilizes the CT to capture the geometric information of the spatial domain in the scene for depth estimation.

In remote sensing data analysis, the multiscale geometric analysis also plays an important role. For unsupervised change detection in SAR images, Zhang et al. [59] proposed adaptive contourlet fusion clustering. Aiming at the characteristics of polarimetric SAR, Li et al. [60] proposed a complex contourlet-CNN for PolSAR image classification. The method uses CT to help complex CNN capture abstract features of specific directions and frequency bands and can retrieve the region and direction information corresponding to the extracted features. Gao et al. [61] proposed a multiscale curvelet scattering network to improve the multiscale directional information of the scattering process.

C. Attention Mechanism

Selectivity in the brain is the core mechanism. Humans can quickly eliminate distractions and capture important information. Drawing on this mechanism, attention has become a significant component of neural network architecture. It has several uses in computer vision, statistical learning, speech recognition, and natural language processing.

The reason why the attention mechanism has received widespread attention is that, on the one hand, it stimulates the mechanism of the human brain. On the other hand, we can partially explain the neural network's performance and enhance the model's efficacy by visualizing the attention maps.

The recurrent neural network (RNN) structure was the first neural network to employ the attention mechanism as part of

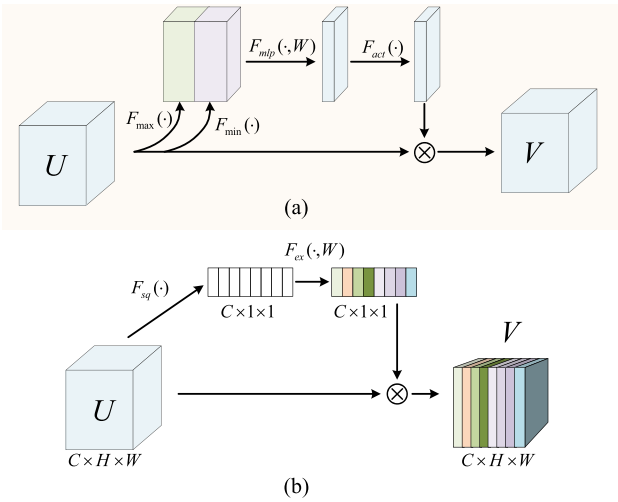


Fig. 6. Illustration of the spatial attention and channel attention. (a) Spatial attention proposed in [50]. (b) Channel attention proposed in [33].

the encoder–decoder framework of RNN to encode long input sentences [62]. It has steadily gotten into the field of computer vision in recent years with an increase in attention mechanism variations. Deep learning and visual attention techniques have been successfully combined in several studies. The main goal of the attention mechanism in computer vision is to train the model to concentrate on significant details while dismissing irrelevant ones. Current attention methods can be divided into spatial attention and channel attention (as shown in Fig. 6).

The fundamental idea behind the attention mechanism in the spatial domain is to apply the appropriate spatial transformation to the spatial domain information. It helps the neural network extract important information from the images. Each layer of a convolutional neural network will output a feature map. For convolutional neural networks to perform spatial attention, a weight matrix must be learned for each pixel in the feature map [63]. The weight matrix will be multiplied by the feature map to balance the influence of each pixel.

The fundamental concept of channel-based attention is to suppress the invalid or small effect features and highlight the effective features to improve performance. This is done by learning the feature weights on the channel domain through the network [33]. In particular, it automatically determines the relevance of each feature channel through learning and then increases beneficial features and suppresses features that are not useful for the present job. Usually, pure channel-based attention has the same weight in the spatial dimension. That is, the information in each channel is directly global average pooled, and the local information in the channel is ignored.

The attention mechanism can efficiently improve the target features in various remote sensing applications while simultaneously resolving the issue of redundant features in remote sensing data. In HSI classification, it is difficult for traditional convolutional neural networks to extract local features of HSIs. In order to strengthen the learning of local key features in the spatial domain and spectral domain of HSIs, the Resnet [64] introduces a HSI feature extraction method based on spatial-spectral attention on the basis of a convolutional network and uses a calculation to obtain the mask and identifies the features required for classification and improves the representation ability of hyperspectral. In remote sensing image instance segmentation, Zhang et al. [65] proposed a semantic attention module; using additional segmentation supervision for attention, the activation values of instances under complex remote sensing noise background are significantly improved.

D. Reinforcement Learning

The process of human learning knowledge is affected by the environment and historical experience. This learning process is the plasticity of the brain. In order to simulate this property, the learning process of reinforcement learning is designed as an interaction between the agent and the environment. The agent can learn by performing different actions and obtaining different rewards in the simulated environment [67]. Deep reinforcement learning integrates the powerful understanding ability of deep learning in perception problems, such as vision and the decision-making ability of reinforcement learning, and realizes end-to-end learning. The emergence of deep reinforcement learning has made reinforcement learning technology truly practical and can solve complex problems in real-world scenarios [68], [69].

Different from the goals of supervised and unsupervised learning, the problem to be solved by the algorithm is how the agent performs actions in the environment to obtain the maximum cumulative reward. $< A, S, R, P >$ is the classic quadruple in reinforcement learning. A represents all the agent's actions. S is the state of the world that the agent can perceive. R is a real value representing reward or punishment. P is the world the agent interacts with, known as the model. Specifically, the strategy refers to the choice of actions the agent will make when it is in state S. The reward signal defines the goal of the agent's learning. The value function is defined to judge whether the reward in interaction is good or bad. The model is a simulation of the natural world, and it models the environment's reaction after the agent samples it. In reinforcement learning, an agent observes where actions and rewards interact with the environment to complete a task.

In remote sensing, reinforcement learning determines sequential actions by maximizing cumulative feature rewards through interaction with the environment. Especially, when only a few labeled pixels are available, reinforcement learning can achieve relatively high accuracy without using any labeled training dataset. This is well suited for remote sensing tasks with fewer data, such as in SPRL [66]. As shown in Fig. 7, SPRL adopts reinforcement learning-based methods for polarimetric synthetic aperture radar (PolSAR) data classification. The pixels are set to "state" and "work" according to reinforcement learning, and their "action" is modified by interacting with the "environment." Design a spatially polarized "reward" function from the local neighborhood to explore spatial and polarized information for more accurate classification. This results in a self-evolving and model-free classifier with a simple principle robust to speckle noise in the data. By interacting with the environment, SPRL

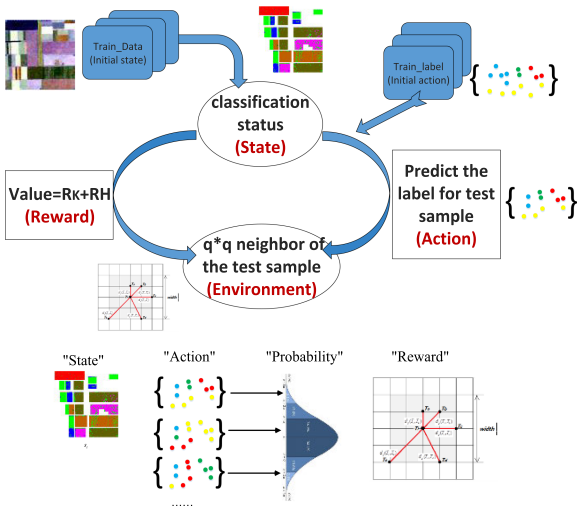


Fig. 7. Example flowchart of PolSAR classification with reinforcement learning. Image from [66].

networks can achieve high classification accuracy when only a few labeled pixels are available.

Similarly, for few-shot remote sensing data, an enhanced deep Q-network technique for classifying PolSAR images was put forth. It can provide valuable data by interacting with agents in a greedy manner [70]. Multilayer feature images and classification actions are correspondingly referred to in the network as environment states and agent actions. Certain conditions reward model predictions. Give the agent feedback by using an annotated sample set of data.

To detect the dense ships from the complex background, Fu et al. [71] proposed a ship rotation detection model based on feature fusion pyramid network-based deep reinforcement learning (FFPN-RL), which applies deep reinforcement learning to the tilted ship detection task. Angle prediction is made through three actions of the action set. Using different rotation angles in the action set makes it possible to achieve higher prediction accuracy and reduce the number of decision-making actions. The reward function encourages or penalizes angle-predicting agents with selected actions. The agent accumulates experience with the abovementioned rewards, learns from them, and ultimately chooses the appropriate action in each decision. As a result, the detecting network can produce inclined rectangular boxes for ships more efficiently.

E. Transfer Learning

As an essential ML method, transfer learning has been widely studied. It can simulate the human's learning ability of "inferring others" and transfer the knowledge learned in the past to new tasks, and speed up the cost of learning new tasks [72]. On the other hand, transfer learning can train ML methods of supervised learning using part of the labeled data, reducing dependence on a large amount of labeled data [73]. The primary trend in current transfer learning development is to use a large amount of labelled classification data to pretrain a benchmark network

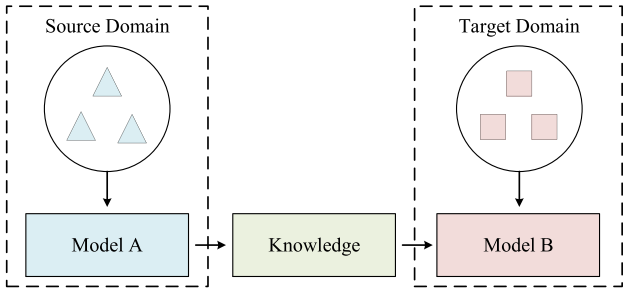


Fig. 8. Learning process of transfer learning.

and then use a small amount of labeled data to fine-tune the network for different tasks.

As shown in Fig. 8, the core idea of transfer learning is applying knowledge gained from one problem to another, a different but related problem. When performing transfer learning, the constraints of the pretrained model and setting an appropriate learning rate are important. Using pretrained networks may limit the architectures used with new datasets.

A lower learning rate is usually used for the weights of the convolutional network being fine-tuned compared with the randomly initialized one. It is possible to train a good classifier using the source domain data. However, the source domain model cannot classify the target domain data well due to subtle differences between the source and target domain data. A commonly used method is to align the feature distributions of the target domain and the source domain data. The target domain data can be classified using the model trained with the source domain data.

Domain adaptation [74] is a unique type of transfer learning that occurs when the data distributions in the source and target domains vary, but the two objectives are the same. Domain adaptation is currently a significant research hotspot in transfer learning. Its task is to learn a mapping that can simultaneously map the source and target domains to a common feature space so that the composite mapping can be simulated. Combine mappings learned only in the source domain and very close to mappings learned only in the target domain.

At present, there are many related studies combining transfer learning with remote sensing data. Xie et al. [75] proposed utilizing a transfer learning strategy to leverage nighttime light intensity to train a fully convolutional CNN model to forecast evening lights in daytime photos. The features learned are helpful for poverty prediction. Chen et al. [76] used a single deep convolutional neural network and limited training samples to perform transfer learning and improve the detection accuracy of aircraft in remote sensing data. A change detection-driven transfer learning method is proposed to leverage the time series images updating the land cover maps [77]. The method aims to leverage the existing knowledge of the source domain to define a reliable training set for the target domain. This is achieved by applying an unsupervised change detection method to the target and source domains and initializing the target domain training set by migrating the detected class labels of unchanged training samples from the source domain to the target domain.

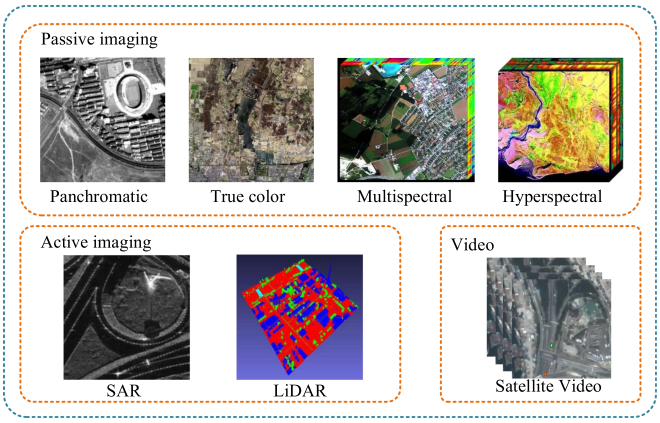


Fig. 9. Illustration of various data types. The data types can be divided into two categories: passive imaging and active imaging.

IV. Interpretations of Remote Sensing

A. Data Types of Remote Sensing

As artificial intelligence has advanced, it has increasingly become used in more and more applications with impressive results [78]. The field of remote sensing is no exception [79]. Intelligent interpretation of remote sensing is crucial to study in many areas, including environmental monitoring, land resources [80], crop monitoring [81] and yield estimation, forest carbon sink estimation [82], and national defense security [83]. Intelligent remote sensing interpretation is also an important requirement for national strategic development [8].

Remote sensing image refers to films or photos that record the size of electromagnetic waves of various ground objects, mainly divided into aerial photos [84] and satellite photos [85]. Remote sensing imaging methods mainly include aerial photography, aerial scanning, and microwave radar. Remote sensing images can be broadly separated into active and passive remote sensing based on various detecting techniques [86]. According to the capture spectral range of the sensor, it is divided into ultraviolet remote sensing, visible light remote sensing, infrared remote sensing, microwave remote sensing, and multiband remote sensing [87]. This section mainly summarizes the widely studied optical remote sensing images and radar images in the existing remote sensing data (as shown in Fig. 9), including optical remote sensing images [88], radar images [89], LiDAR point cloud data [90], and remote sensing videos [91].

1) Optical Images: Optical images are a kind of remote sensing data that obtains target information on different spectra by dividing the radiation of objects into several narrower spectral bands. The same objects have similar spectral characteristics [92]. The radiation energy of different objects in bands is different.

According to the number of captured spectral bands and the narrowness of the spectral bands, optical images can be roughly classified into three types: panchromatic, multispectral, and hyperspectral [93]. Generally, most satellites can take panchromatic and multispectral images.

Panchromatic images: Panchromatic images have only one grayscale image band, i.e., the brightness of a particular pixel is proportional to the pixel value. The pixel value is related to the intensity of solar radiation reflected by the target. Panchromatic images generally have a high spatial resolution, but their images have little spectral information [94].

Multispectral images: Multispectral imagery usually refers to three to ten spectral bands expressed in pixels. Each band can be acquired using a remote sensing radiometer [95]. An image with both the high GSD and abundant spectral information can be generated by properly fusing the panchromatic image with the multispectral image.

HSIs: While hyperspectral data contain very narrow bands (10–20 nm) [96], HSIs may have thousands of bands. For each band of hyperspectral data, imaging spectrometers are often required to acquire them. Compared with high-resolution, multispectral images, HSIs have high spectral resolution and abundant bands. It contains rich radiation, spatial and spectral information [97], and is a comprehensive carrier of various details. The areas of feature mapping and resource exploration have made extensive use of HSIs [98]. Unlike standard RGB images, HSIs are often multichannel images. Hyperspectral rich band information often contains richer features. We can select the band by the sensitivity of different ground objects to different bands to highlight certain objects [99].

2) Radar Images: Radar is an active microwave remote sensor that emits microwave radiation and receives electromagnetic waves reflected from a target [100]. The radar imaging system mainly includes five parts: a pulse generator, transmitter, radar antenna, receiver, and recorder. The pulse generator generates a high-power FM signal and repeatedly emits microwave pulses of a specific wavelength at a particular time interval through the transmitter. Commonly used radar images can be divided into synthetic aperture radar (SAR) and PolSAR.

SAR: SAR [101] is an active microwave imaging device. Its imaging principle forms the virtual antenna of the radar through the movement of the flight carrier, thereby obtaining high-azimuth resolution radar images. SAR can be divided into airborne and spaceborne according to aircraft type. Both have their advantages and uses. Airborne SAR has higher resolution, whereas spaceborne SAR can observe a wider area for a long time, has a global macroscopic effect, and is periodic. The cost is also lower than the airborne, so spaceborne SAR has been widely used. According to whether synthetic aperture processing is performed, imaging radar can be divided into real aperture radar (RAR) and SAR [102], [103] [as shown in Fig. 10(a)].

Real aperture imaging radar transmits a pulsed radio beam with a very narrow width to the side of the radar antenna (called the range direction) to the traveling direction of the aircraft (called the azimuth direction). The beam irradiates a long narrow ground strip perpendicular to the flight direction. Then, the radar antenna is converted into the receiving working state and receives the backscattered wave reflected from the target [104], [105]. As the vehicle travels, the emitted beam scans the surface in this continuous strip along the direction of flight. The radar image is created line by line [106].

The resolution of radar images includes distance resolution and azimuth resolution. Distance resolution refers to the resolution in the vertical flight direction. The azimuth resolution

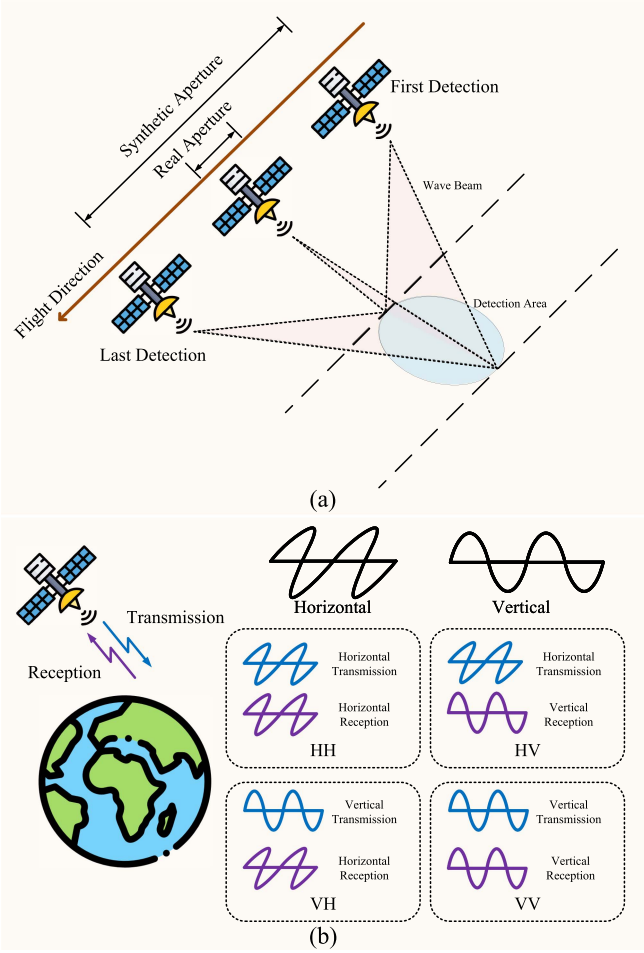


Fig. 10. (a) Imaging principles of SAR. (b) Polarization combinations of PolSAR.

refers to the resolution along the flight direction [107]. The distance resolution is mainly related to the pulse signal emitted by the radar system. The shorter the pulse duration, the higher the distance resolution. However, the transmission power will decrease if the pulse width is too small. In addition, the signal-to-noise ratio of the reflected pulse will also decrease, which is contradictory [108].

The basic principle of SAR is to use a small antenna as a single radiating unit to make it move continuously along a straight line. The reflected pulse of the same target at different positions performs related processing, which can obtain higher image resolution [108]. SAR is the same as RAR in the distance direction, using pulse compression to improve the resolution. In the azimuth direction, the resolution is improved by the principle of synthetic aperture [109]. While the position of the radiating element is constantly changing, the received signals can be recorded and processed to obtain the same effect as the observation with a longer virtual antenna length (synthetic aperture length) of the actual antenna.

By transmitting electromagnetic pulses and receiving target echoes for coherent imaging, SAR can shoot multipolarization, multiband, high-resolution images all day, all weather. It obtains backscattering information of ground objects to realize the task of Earth observation. Compared with optical and infrared remote sensing technologies, SAR belongs to microwave remote sensing [110]. It can not only obtain the Earth's surface information, such as topography and landforms, but also penetrate the surface to obtain underground, concealed, and high-resolution ground data in harsh environments.

PolSAR: PolSAR system [111] is developed based on the single-channel SAR system, which can provide multidimensional remote sensing information of targets. Compared with traditional single-channel SAR, polarimetric SAR not only utilizes the amplitude, phase, and frequency characteristics of target scattered echoes but also utilizes its polarization characteristics [112]. For example, the L-band with a longer wavelength can penetrate forests and surface vegetation coverage. It can be used in the military to discover hidden targets in jungles or shallowly buried surfaces [113].

By sending and receiving electromagnetic waves with various polarizations, PolSAR measures the polarization scattering properties of ground objects and builds up the polarization scattering matrix. The polarization of electromagnetic waves is sensitive to the target's physical characteristics, such as surface roughness, dielectric constant, geometry, and orientation. Thus, the polarization scattering matrix includes abundant target information.

PolSAR obtains polarization scattering matrixes by measuring the scattered echoes in each resolution unit on the ground [114]. The amplitude and phase properties of the target scattered echoes can be completely described using these polarization scattering matrices.

When the electric field of the electromagnetic wave is parallel to the scattering surface, the electromagnetic wave is called a horizontal (H) polarized wave. Similarly, the perpendicular one is called vertical (V) polarized waves. Therefore, PolSAR can be divided into four polarization modes based on the transmitting and receiving antenna's direction.

As shown in Fig. 10(b), there are four polarization combinations: VV, HH, VH, and HV. For example, VV polarization, namely vertical transmission/vertical reception, indicates that the polarized SAR transmitting antenna transmits vertical electromagnetic waves, and the receiving antenna also accepts vertical electromagnetic waves. By obtaining four basic polarization combinations (HH, HV, VH, and VV polarizations) [115], the received power value of the antenna in all possible polarization states can be accurately calculated.

In recent decades, PolSAR technology has developed rapidly, and its wide application has also received increasing attention [116]. At the same time, people's demands for SAR are growing, and they want to obtain images of the same target in several frequency bands, polarizations, and viewpoints. In addition, SAR miniaturization is also significant due to the need for military unmanned reconnaissance aircraft. Nowadays, PolSAR is one of the most sophisticated sensors used in remote sensing. It has many practical applications and importance in civil and military fields.

3) Airborne LiDAR: Airborne LiDAR [117] is a detection technology integrating attitude determination, laser, and high-precision GPS differential positioning technology.

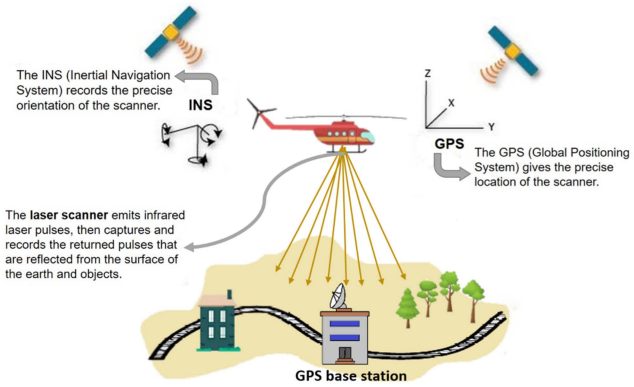


Fig. 11. Graphic depicting an airborne LiDAR system. Image from [118].

LiDAR determines the relative distance between the scanner and the object by measuring the signal travel time [119]. Compared with the data obtained by traditional photogrammetry, point cloud data can reflect terrain information more accurately. The data collected by airborne LiDAR are a series of discrete 3-D points with irregular spatial distribution, called "point cloud."

As shown in Fig. 11, airborne LiDAR systems mainly include laser scanners, inertial navigation systems (INS) [120], and dynamic differential GPS receivers. The laser scanner measures the distance from the launch point of the laser to the ground target. The inertial navigation system uses the inertial measurement unit (IMU) [121] to measure the attitude parameters of the aircraft's central optical axis scanning device. The dynamic differential GPS receiver is used to determine the spatial location of the launch point of the LiDAR.

After the airborne LiDAR system completes the laser scanning, the data obtained include the position, orientation, and laser scanning distance [122]. Among them, the position and orientation include differential GPS and IMU information. These data record the information of each laser pulse, including position, azimuth/angle, distance, time, intensity, echo, and other data obtained by the system during flight. The X, Y, and Z coordinates of the laser point in the WGS84 coordinate system can be calculated. These discrete points with precise 3-D coordinates are called the LiDAR point cloud [123].

The 3-D LiDAR point cloud data include information, such as the spatial 3-D coordinates of the point, echo intensity, echo times, and scanning angle [124]. In practical applications, the information frequently employed is the point cloud geometry, laser intensity, and laser echo data returned by emitted laser pulses. The laser echo signal is produced when a laser pulse is fired from a laser scanner and is then reflected or scattered by a ground point. The airborne LiDAR system may offer not only the 3-D coordinates of the target point but also intensity information of the laser echo signal [125]. Due to the different reflection characteristics of each material to the laser signal, the point cloud data can easily distinguish the boundaries of different objects for object classification.

4) Remote Sensing Videos: Remote sensing video [126] is usually divided into satellite video and UAV video according to the platform that carries the sensor. Satellite video is a kind of onboard video. It generally refers to the video obtained by satellites in the fields related to research and exploration of space. UAV video is the videos captured by UAV. The illustration of remote sensing videos is shown in Fig. 12.

(a)

(b)

Fig. 12. Illustration of remote sensing videos. (a) Satellite videos. (b) UAV videos.

Satellite Videos: Satellite imagery refers to a satellite platform that carries an image payload and can obtain images of ground target areas.

Satellite videos [127] can continuously image the target area for a long time, providing dynamic information and realizing long-term dynamic real-time monitoring. The camera is mounted on a microsatellite platform and consists of a telescopic objective lens, an area array focal plane detector, and an electronic processing circuit [128]. The telescopic objective lens images the ground scene within the 2-D field of view on the image plane, and after photoelectric conversion and electronic circuit processing of the area array detector located at the image plane, the remote sensing image of the ground scene is obtained. When the shutter that controls the exposure is opened, the light emitted by the ground scene is transmitted through the atmosphere and reaches the camera's entrance pupil. The telescopic objective focuses on the area array focal plane detector to obtain a frame of video of the target. As the satellite platform flies in orbit, there is relative movement between the camera and the ground scene. When the shutter is opened again, another frame of the target is obtained. This cycle continues to form a frame push process. In the process of frame push imaging, the exposure time is often more significant than the integration time corresponding to a single pixel. The captured image is prone to displacement in the direction along the track, that is, image movement and the image easily becomes blurred. The image movement compensation device, such as a reaction wheel or gyroscope, can be used to adjust the camera attitude

to eliminate or reduce the impact of image movement. After multiframe image compression, frame alignment algorithm, and other software processing, a continuous dynamic video is finally formed.

As a new method of acquiring image data for Earth observation, satellite remote sensing video can be applied to large-scale dynamic target change monitoring and its instantaneous characteristic analysis [129]. It reduces the time interval between adjacent image frames by adopting the "image recording" method for a specific area, which not only achieves large-scale coverage but also makes up for the limitation of the reentry period of traditional satellites. Compared with conventional remote sensing satellites, the target observation area of satellite remote sensing video is small, but the timeliness is good [130]. It can realize fixed-point and fixed-range remote sensing monitoring in small areas, which makes it have unique application advantages in some major engineering fields. For example, it can keep abreast of the progress and construction of major projects and provide real-time video information support for the impact on the surrounding ecological environment.

Compared with traditional video surveillance image data, satellite remote sensing image data have the following challenges [131].

1) In the process of satellite remote sensing image imaging, the slow movement of the sensor causes the displacement of buildings, trees, and other targets to change, resulting in many false moving targets, making the background more complicated.
2) Due to the limitation of the spatial resolution of satellite remote sensing imaging, the target is only a few to a dozen pixels in size in the image, and the contrast with the background is low, so it is impossible to obtain more detailed information of the target.
3) In the satellite videos, factors, such as illumination change, shadow movement, and others, lead to the dynamic changes in the background. Due to the low resolution, these dynamic changes are more likely regarded as the moving target causing false alarms. We directly apply traditional moving target detection methods in satellite videos resulting in false detection.

UAV Videos: UAV [132] is a kind of unmanned aerial vehicle. With the improvement of hardware performance and the development of image processing algorithms, the research on UAV vision has become a hotspot. Due to geographical restrictions, the advantages of large-scale, multiangle, high-resolution data can be obtained. It plays an increasingly important role in target tracking, image stitching, power line inspection, island monitoring, coastline inspection, postdisaster monitoring, and river flood season monitoring [132].

In addition to takeoff and landing, the flight state of the UAV can be roughly divided into the hovering state and the cruising state, and the videos obtained in these two states have different characteristics. The drone can shoot stable video in the hovering state. Still, the rotation of the wing and the influence of the external wind will cause the picture to shake, resulting in irregular motion of the video background. The UAV cruising state refers to the translational flight state of the UAV in forward and backward flight. In the video shot, the image has a large offset in a short period. In addition to the moving target, the background also has much movement.

Compared with satellite videos, UAV-borne image data has the following advantages.

1) Make up for the lack of timeliness of satellite remote sensing and ordinary aerial remote sensing, lack of maneuverability, and the lack of regional information due to limitations, such as weather conditions and time [133].
2) The drone images have high resolution and can obtain high-resolution panoramic images of the flight area. However, due to the long distance of satellite shooting, the resolution and accuracy of the image cannot be satisfied.
3) The UAV system has a low cost of use and simple maintenance and operation [134].
4) The UAV system can quickly acquire visible light and infrared imaging at medium and low altitudes, conduct fast and real-time ground inspection and monitoring, and record the current image status objectively and directly [135].

Compared with other relatively stable camera equipment, such as surveillance cameras on roads and shopping malls, the high mobility of drones can make data collection not limited by geographical areas. It has unique advantages in resource and environmental monitoring, forest fire monitoring, and rescue command in areas where vehicles and people cannot reach and has become more flexible. The image data obtained by aerial cameras, satellites, etc., carried by airships at high altitudes, using UAVs for moving target detection are more challenging. Table I lists the characteristics of UAV videos compared with satellite videos.

In general, video data contain richer information than individual images in terms of content or time [136]. In particular, satellites gradually begin to develop video functions, significantly expanding the source of video data.

B. Applications of Remote Sensing

Brain-inspired remote sensing interpretation is applied to all aspects of remote sensing data processing, effectively processing the replicated and diverse data of remote sensing. In this section, we summarize the development during recent years of five applications, including land-cover classification, change detection, target detection, object tracking, and 3-D reconstruction.

1) Land-Cover Classification: Land-cover classification, which is also known as semantic segmentation in nature image processing, is one of the most basic image analysis tasks in remote sensing. It classifies each pixel in the image and assigns a category to each pixel, achieving an understanding of the image content.

In 2015, Long et al. [137] first proposed fully convolutional networks (FCN) for semantic segmentation tasks. The FCN network replaces all the fully connected layers in the neural network with convolutional layers, realizing a network composed of all convolutional layers. Since the FCN network fails to make good use of multiscale features, in 2015, Ronneberger et al. [138] proposed the U-Net network. The

TABLE I
COMPARISON OF UAV VIDEOS AND SATELLITE VIDEOS

Video Type	Coverage area	Background	Illumination variation	Target size	Change of rotation	Change of target size
UAV Videos	$< 1\,km^2$	unstable	large	big	Y	Y
Satellite Video Videos	$50 - 90\,km^2$	stable	small	tiny	Y	N

U-Net network utilizes the skip connection operation to make full use of the multiscale features generated during the downsampling process and then obtains excellent segmentation results. Moreover, in 2017, Badrinarayanan et al. [139] proposed the SegNet network based on the U-Net network. The network performs nonlinear upsampling in the decoder using the pooling indices computed in the max-pooling step of the corresponding encoder. In the same year, Gao Huang et al. proposed DenseNet [140]. The convolutional layer of the DenseNet network connects each layer with each layer in a feedforward manner so that the layers close to the input and the output contain shorter connections to recover information lost during convolution, since both UNet and SegNet fail to fully utilize the local neighborhood information around pixels. Also, Chen et al. [141] proposed the DeepLabV3+ network. The DeepLab network utilizes atrous spatial pyramid pooling (SPP). Multiscale local receptive fields of pixels are fused while reducing resolution.

Compared with natural images, remote sensing images have the following characteristics.

1) The size of the same class objects varies widely, and the problem of size change needs to be solved.
2) Due to the fact that satellites shoot the ground at high altitudes, the obtained images are very wide in scope. The object occupies very few pixels, which generates the problem of sample imbalance.
3) When shooting in a large area, the same class of objects show a variety of different appearance because of weather, light, and other natural conditions.
4) Large-scale shooting is usually accompanied by low resolution, which makes each semantic region lacks morphological contour information.

These characteristics of remote sensing images impose higher requirements for land-cover classification.

Land-cover classification can be roughly divided into object-based and pixel-based methods. The object-based method divides the image into regions and classifies the regions according to the feature of the whole region. While the pixel-based method does not need region division and directly uses the characteristics of the pixels to classify directly. Due to the heterogeneity of medium- and low-resolution remote sensing images [as shown in Fig. 13(a) and 13(b)], each pixel is considered to be mixed and may contain more than one semantic category. Therefore, pixel-based classification methods are usually ineffective for medium-resolution and low-resolution remote sensing images, whereas object-based methods can achieve coarse image segmentation by classifying regions. There is less category mixing in high-resolution images, as shown in Fig. 13(c), where each pixel represents the characteristics of this area. Compared with the region-based method, the pixel-based method can give full

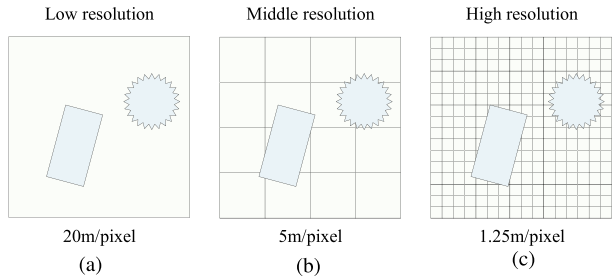


Fig. 13. (a) Low resolution: The pixels are significantly larger than the object. (b) Medium resolution: The pixel and the object are the same sizes. (c) High resolution: Pixels are significantly smaller than objects.

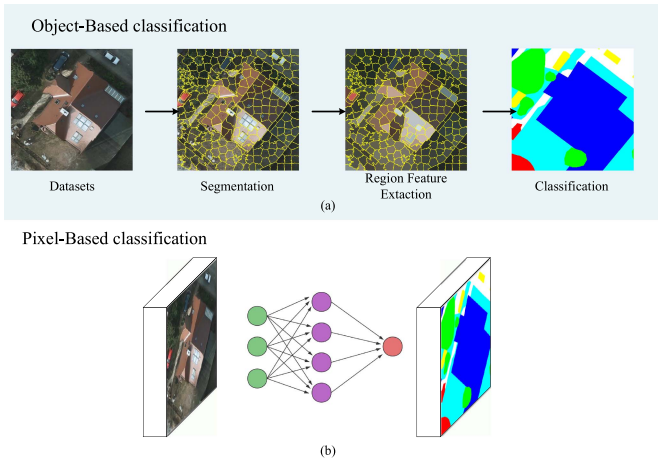


Fig. 14. (a) Object-based classification. (b) Pixel-based classification.

play to the characteristics of the pixel itself and perform the segmentation task more successfully.

Object-based classification: The core to be processed of the object-based classification method is the segment (segments), that is, the grouping of multiple pixels with the same attribute into an object. Unlike pixel-based classification methods, object-based methods divide remote sensing images into separate regions and evaluate their characteristics by spatial and spectral features. Object-based methods are also more similar to the human visual understanding process, understanding semantic information by considering the different properties and spatial arrangements of these objects and then intuitively identifying objects from images rather than individual pixels. Currently, object-based classification of features is also used in archaeology, exploration of glacial landforms, wetland mapping, and other applications. Object-based methods usually consist of three main parts: image segmentation, object feature extraction, and object classification, as shown in Fig. 14(a). The image segmentation part, the first step of the object-based method, divides the remote sensing image into multiple homogeneous segments

with segmentation algorithms, such as edge-based segmentation and region-based segmentation. The object feature extraction part makes up for the shortcomings of pixel-based methods, including features, such as shape, texture, and spectrum, are extracted. Finally, in the object classification part, different objects are classified by the classifier in their feature space.

In recent years, how to integrate deep learning and object-based land-cover classification has attracted the attention of many scholars. Zhang [142] proposed an object-based convolutional neural network (OCNN) method for land use classification. OCNN first segmented remote sensing images into linear-based objects and general objects and then sent them into the neural network for analysis. Timilsina et al. [143] presented a new method combining the object-based postclassification refinement method and CNNs, which takes optical and SAR data as input and uses the CNN network to obtain coarse results, which are extracted with the help of OBIA. Spatial, texture, and context features refine the coarse results. Zhang et al. [144] proposed a multilevel context-guided classification method (MLCG-OCNN) for high-resolution remote sensing images. Instead of using object and context blocks as input, MLCG-OCNN accurately identifies objects using high-level features learned from spectral patterns, geometric features, and object-level contextual information. The classification results for each object are then improved with pixel-level contextual guidance. Papadomanolaki et al. [145] introduced a novel object-based deep learning system that incorporates anisotropic diffusion data preprocessing and an extra loss to integrate object-based priors.

Pixel-based classification: Pixel-based approaches employ image pixels as the basic unit of analysis, and individual pixels are labeled as a single semantic category, such as vegetation, buildings, vehicles, or roads [as shown in Fig. 14(b)]. Early methods based on pixel-by-pixel classification mainly adopted k-means, support vector machines, neural networks, and other methods. With the improvement of remote sensing imaging technology, the resolution of remote sensing images has been greatly improved. The pixel-based method completes the segmentation task by clustering the pixels with similar features into the same category and assigning a category through the pixels' features.

Peng et al. [146] proposed cross fusion net (CFNet) based on UNet. The CFNet network fuses and predicts the multiscale features in a concatenated manner. In addition, the network designs a channel attention refinement module to select informative features and a cross fusion module to expand the low-level feature map of the receptive field to improve the segmentation accuracy of small-scale objects. Heidler et al. [147] proposed the HED-UNet network, which exploits the multiscale features generated in the decoding process to provide features for both semantic prediction and boundary prediction tasks.

Liu et al. [148] constructed an atrous convolution module based on atrous convolution in the DeepLabv3+ network, which can arbitrarily control the depth, width, group, and step of the module with different dilation rates to make full use of local features. Peng et al. [149] used a multiscale convolution kernel parallel method to make full use of the local information of the pixel. Dense skip connections are adopted to mitigate the consequences of the loss of high-level features in the image due to the nature of convolutional low-pass filtering. Shang et al. [150] proposed atrous convolution with different expansion rates, the global information, and self-information for extracting multiscale contextual information to solve the problem of object size discrepancy in remote sensing images. Wang et al. [151] proposed a dual-channel spectral-spatial fusion capsule generative adversarial network (DcCapsGAN) for HSI classification. DcCapsGAN utilizes a capsule and generative adversarial network structure to overcome the limitation of training size with high-dimensional features and the effectiveness of spectral-spatial exploitation.

A novel spectral spatial transformer-M that assembles spatial attention and extracts spectral features is proposed to improve performance for the class pixels located on the land-cover category boundary area [152]. Wang et al. [153] proposed an UNetFormer to model both global and local information for efficient semantic segmentation achieving up to 322.4 FPS with a 512×512 input. Inspired by multiscale vision transformer, He et al. [154] proposed a cross-spectral vision transformer to extract pixelwise multiscale features and enhance local details between neighboring spectral bands for HSI classification.

2) Change Detection: Remote sensing change detection (RSCD) refers to extracting and identifying different information between multitemporal images from the identical geographical area [155], [156]. As shown in Fig. 15, RSCD methods typically consist of the processes of remote sensing images preprocessing (alignment, correction, noise reduction, etc.), selection of suitable change detection method, and evaluating the results. Weismiller et al. [157] first performed change detection for coastal environments and since then a large number of studies have been conducted on RSCD. Nowadays, RSCD takes an active part in a variety of applications, including urbanization monitoring [158], damage assessment [159], and environmental monitoring [160]. According to the analysis units, the existing RS CD methods are classified into pixel-based, object-based, and scene-based, each of which has its own advantages and shortcomings [161]. In recent years, new approaches have also been developed to combine these analysis units in the process of change detection to better extract change information.

Pixel-based change detection: Since pixel represents the most basic unit of remote sensing image, early methods of RSCD mainly employed algebraic methods to evaluate every pixel of the given remote sensing images, such as the image difference method [162] and regression analysis method [163], [164]. Furthermore, RSCD can also be undertaken by means of pixel transformation, such as principal component analysis [165] and change vector analysis [164]. In pixel transformation methods, remote sensing images are transformed and combined with spatial projections and converted into different mathematical spaces for analysis in order to optimize various features further. Due to the unpredictability of high-frequency components in the high-resolution remote sensing image and errors of geometric alignment and radiometric correction in preprocessing, traditional pixel-based methods are hardly capable of modeling to apply to high-resolution remote sensing images [166]. Therefore, traditional pixel-based methods are typically adopted for low- and medium-resolution images [167], [168].

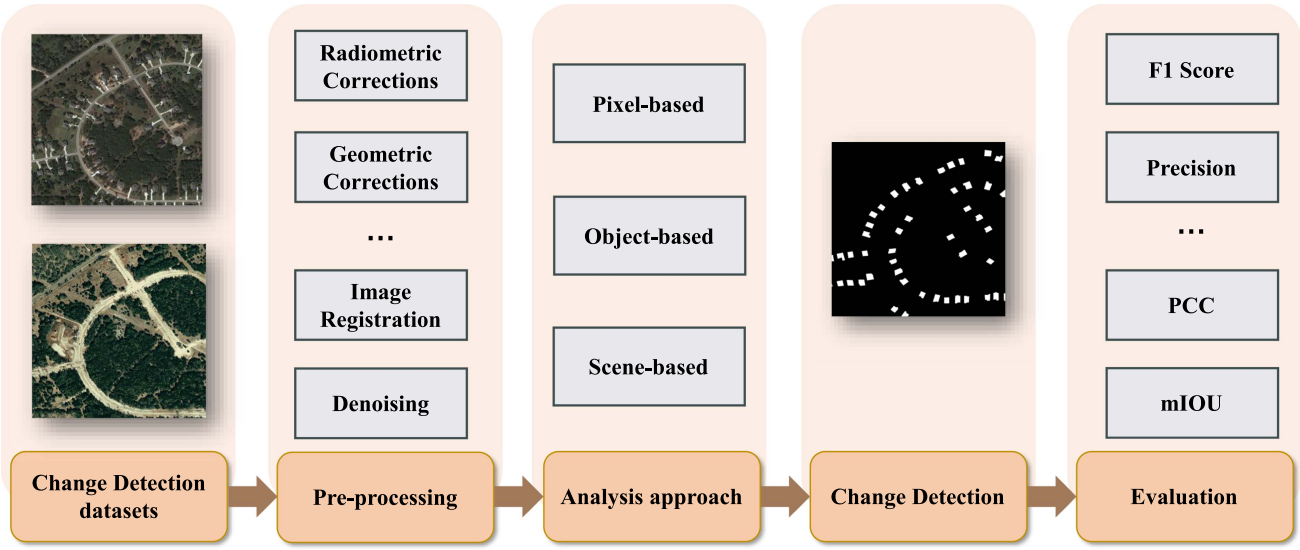


Fig. 15. General diagram of the RSCD process, including remote sensing image preprocessing, selection of suitable change detection methods, and results evaluation.

In addition, the pixel classification change detection method is another pixel-based change detection method that obtains the change matrix of an image by comparing two postclassification images, which reflect the change information in the study area [164]. Such methods include postclassification comparison, unsupervised change detection methods, and artificial neural network-based methods [167]. However, supervised approaches suffer from the difficulty of selecting high-quality datasets, whereas unsupervised approaches encounter difficulties in recognizing and labeling change objects and in selecting numbers of clusters [156], [164].

In recent years, the rise of deep learning has led to a large number of deep learning-based semantic segmentation methods being applied in pixel-based change detection and greatly eased the abovementioned difficulties. For example, Wang et al. [169] introduced a hybrid affinity matrix with fused subpixel representation and proposed a convolutional neural network framework for RSCD. Daudt et al. [170] used a FCN to perform change detection on multitemporal images on Earth observation images. As it has been proven that obtaining contextual information in multitemporal images and combining multiscale features of change regions provides an effective prediction of fine changes and improves the accuracy of change detection [171], research works combining multiscale features have been proposed. For example, Chen et al. [172] designed a multiscale feature convolution unit combined with deep siamese convolutional networks for supervised and unsupervised change detection. Moreover, aiming at further feature and information fusion. Zheng et al. [171] designed a cross-layer convolutional neural network (CLNet), which aggregates multilevel contextual information and multiscale features through two parallel branches.

Since CNN-based methods are not skilled in acquiring remote information in space, the transformer has also been introduced to remote sensing change detection. Chen et al. [173] proposed the dual-temporal image transformer (BIT), which expresses dual-temporal images as several labeled tokens, and the context is modeled in a compact token-based space–time with a transformer-based converter encoder. The learned global context-rich tokens are then fed back into the pixel space to enhance the original pixel-level features by the transformer-based decoder. A pure transformer network with a siamese U-shaped structure is also proposed to solve CD problems [174]. In addition, some scholars have also introduced graph convolutional networks [155], GANs, and DBNs, into pixel-based change detection [156].

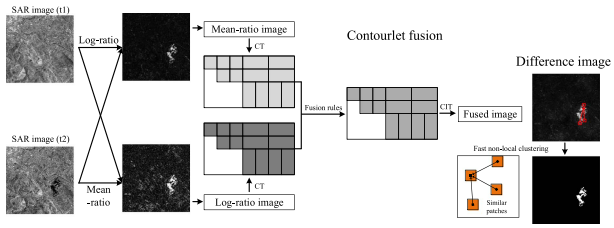


Fig. 16. ContourletFusion clustering (CFC) framework for change detection in SAR images. Image from [59].

Apart from RSCD based on active imaging, change detection in SAR images has also received attention from scholars. During SAR image change detection, since local pixels are coherent, it is critical to reduce the image of scattering noise while preserving the image details as local pixels are coherent. To address the challenges above, as shown in Fig. 16, Zhang et al. [59] presented the adaptive contourlet fusion clustering algorithm as well as a new FGFCM-based fast nonlocal clustering algorithm (FNLC) for SAR change detection, which leverages the change and invariant information from ratio difference images. Specifically, the contourlet fusion method of image fusion first decomposes two input ratio images by the CT, which in turn yields multiresolution and multidirectional decomposition coefficients. Then, different fusion rules are employed to fuse the low- and high-frequency coefficients of the input image, respectively. Finally, the fused coefficients were subjected to the contourlet inversion transform

to acquire the fusion image. In addition, the proposed FNLC method classifies the changed and unchanged areas in the fusion image, enhancing the performance of SAR images in terms of noise suppression.

Object-based change detection (OBCD): Similar to object-based classification, the analysis unit for OBCD is the object in images. Chen et al. [176] defined OBCD as a process of applying object-based analysis to identify variances in geographic objects at different times. Typically, it consists of the following steps: creating homogeneous regions (i.e., image objects) on the basis of image segmentation, extracting change information, and identifying change areas. The OBCD method is highly sensitive to the segmentation algorithm adopted and tends to disregard semantic information, as well as interobject information [177]. Also, the selection of the scale parameter (SP) used to control the object size is a fundamental step in OBCD. Traditional object generation methods based on mathematical approaches fail to solve these difficulties. Meanwhile, based on the accelerated growth of deep learning, OBCD methods have solved these difficulties to some extent.

In the process of object segmentation, both insufficient and excessive segmentation leads to the appearance of features that fail to reflect the real world and may produce useless objects, which may degrade performance [164]. The emergence of deep learning has made it possible to further fuse spatial features. Wang et al. [178] presented a method for change detection combining multiple feature integration methods, showing that multiple objects features yield higher accuracy in object-based methods with different segmentation scales and classifiers. In addition, superpixel segmentation methods are widely utilized to extract objects. Zhang et al. [177] proposed a superpixel enhanced CD network (ESCNet) for very-high-resolution (VHR) images to extract object information with a superpixel segmentation network. To further exploit the contextual information among objects, Zhan et al. [179] presented an unsupervised scale-driven network for VHR images with a multiscale decision fusion strategy. The network identifies change regions by fusing change detection results achieved by various scales from SVM-based classification. It also makes full use of the spatial contextual information of image objects. Zhang et al. [180] introduced the GCN model to remote sensing OBCD and constructed graph neural networks for objects to obtain contextual information between neighboring objects, enhancing performance and computational efficiency.

Bounding box selection is another object-based approach. Among such methods, object detection algorithms, such as Faster R-CNN backbone, are widely utilized, which consider the "changed regions" in the image as detection objects and the "unchanged regions" as background [161]. Zhang et al. [181] proposed a single-stage change detection model with a dual correlated attention-guided detector to enhance robustness. The input images are sent to a weight-sharing backbone to extract features at different scales. A constructed dual correlated attention module is following to refine the change-related features from the channel and spatial aspects and inhibit the uncorrelated features. Han et al. [182] proposed dual regions of interest networks, consisting of three functional blocks: a feature extraction

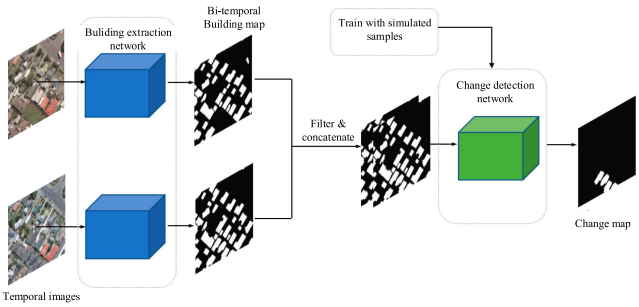


Fig. 17. Change detection framework built on a combination of pixel level and object level, where the building extraction network leverages the object-based Mask R-CNN and the pixel-based multiscale FCN, respectively (image from [175]).

network, a change proposal network, and a different judgment network, to improve feature representation and achieve better change discrimination. Priyanto et al. [183] applied faster R-CNN as a feature extractor to detect and monitor the number of changing floating net cages in fisheries and marine areas.

Furthermore, since both pixel-based and object-based methods hold their respective advantages, many scholars have combined them to achieve better performance. Lu et al. [184] proposed an unsupervised algorithm-level change detection fusion scheme that applies OBCD to improve the accuracy of the traditional pixel-based change detection algorithms. Ji et al. [175] employed mask R-CNN and MS-FCN to extract building features. As shown in Fig. 17, the building extraction network outputs object- and pixel-level building change maps, and feeds them to a self-trained building change detection network to compute building change maps. Han et al. [185] suggested a weighted Dempster–Shafer theory fusion method that generates OBCD by combining multiple pixel-based change detection results.

Scene-based change detection: Remote sensing scene level change detection (SLSCD) intends to analyze and identify land use changes in a given multitemporal remote sensing image of the same area from a semantic perspective [161]. Rather than pixel- and OBCD methods, SLSCD assigns land use/cover labels to image scenes, e.g., for industrial and residential areas. SLSCD is mainly deployed in the analysis of change at the semantic level, i.e., the shift in ground cover type, and is no longer focused solely on the question of whether the ground state has changed. A number of approaches have been proposed, which are broadly classified into traditional-based methods as well as deep learning methods.

Before the surge of deep learning, approaches utilizing handcrafted features were proposed successively, such as scale-invariant feature transformation and a bag of visual words (BOVW) models. Wu et al. [186] presented an SLSCD framework based on the BOVW model and a classification-based approach to extraction semantic change information, in which scene images are represented by the word frequencies of three kinds of multitemporal learned dictionaries. To further exploit the time-scale information and compensate for the weakness of manual features, some scholars introduced unsupervised methods to SLSCD. Wu et al. [187] proposed a method that combines

kernel slow feature analysis (KSFA), an unsupervised learning algorithm based on the fusion of KSFA and postclassification fusion, combining independent scene classification with change probability to identify scene changes and recognize transition types. Du et al. [188] proposed a latent Dirichlet allocation and multivariate alteration detection method for unsupervised scene change detection.

As a large number of remote sensing scene data samples with annotations are acquired, the traditional methods above-mentioned show low robustness for large-scale datasets, and the whole scheme of some traditional methods fails to perform joint optimization [189]. Following the growth of deep learning, a number of researchers have introduced deep learning to SLSCD to break through these difficulties. Wang et al. [190] proposed a scene change detection network named DCCANet. DCCANet extracts convolutional features through a CNN and uses deep typical correlation analysis (DCCA) to learn the nonlinear transformation of two view data, which enhances the temporal correlation of multitemporal correlation of the temporal images and obtains highly correlated features.

3) Target Detection: The research of remote sensing image target detection has a broad application perspective. It can monitor the traffic conditions of important areas [191], roads, ports, and airports, and then coordinate the detection of aircraft in airports [192], vehicles on roads [193], and ships in ports [194]. However, owing to the complex information of remote sensing images and the small size of targets, detection methods based on natural images cannot achieve good results on remote sensing images. Therefore, a large number of methods have been proposed for object detection tasks in remote sensing image interpretation. Object detection focuses on whether there are object instances from a defined class given the input information, and if so, returns the spatial location, extent, and class of each object through a bounding box [195]. With the development of deep learning, thanks to the powerful semantic representation ability of deep features extracted by neural networks, the performance of target detection has been rapidly improved. Generally, deep learning-based object detection methods are mainly divided into two categories: two-stage detection frameworks and one-stage detection frameworks [196]. The difference between them is shown in Fig. 18.

Two-stage detection frameworks: The two-stage detector first generates region proposals and then classifies the candidate boxes. For object detection in remote sensing images, besides the limitation of training samples, the biggest challenge is how to effectively deal with the change of object rotation [5]. Li et al. [197] constructed a region proposal network including additional multiangle anchors and a local contextual feature fusion network to better extract the rotation and appearance blur features of spatial objects in remote sensing images. In addition to extending directly on classic two-stage detectors, such as R-CNN and faster R-CNN, many scholars have also proposed other two-stage methods according to the characteristics of remote sensing images. Zou et al. [199] designed SVDNet based on a singular value decomposition algorithm, and adopted feature pooling operation and linear SVM classifier for ship verification. Bai et al. [198] proposed an object detection method

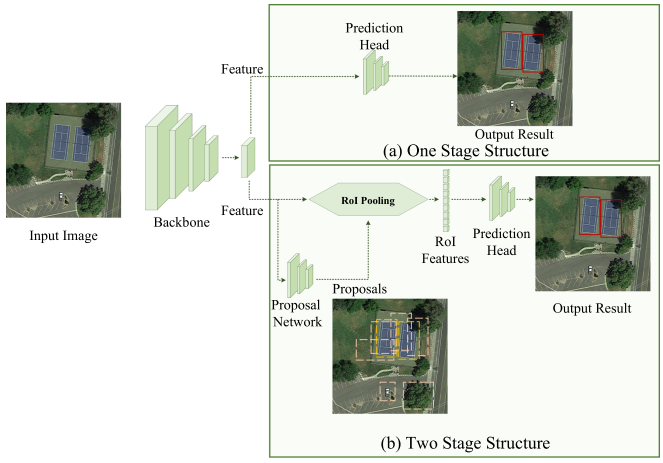


Fig. 18. One-stage and two-stage structures comparison. (a) One stage, (b) Two stage.

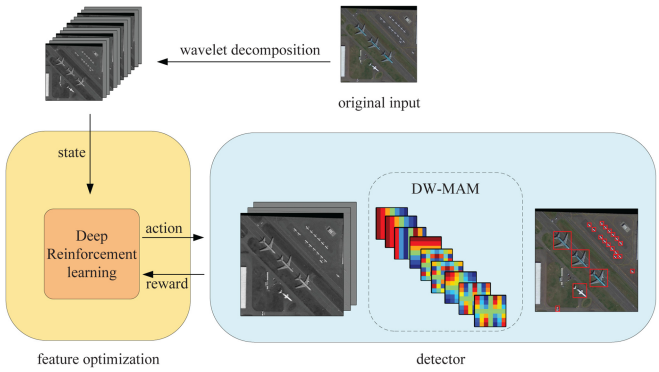


Fig. 19. Framework of the proposed method. (Image from [198]).

based on time–frequency analysis for large-scale remote sensing images with complex backgrounds. They utilized wavelet decomposition for time–frequency transformation, which was then combined with deep learning for feature optimization. A feature optimization method based on deep reinforcement learning is proposed to select the main time–frequency channels. In addition, a discrete wavelet multiscale attention mechanism is designed to enable the detector to focus on object regions instead of the background, effectively extracting multiscale and multidirectional features from remote sensing images (as shown in Fig. 19).

Object detection has come a long way recently. However, the widely adopted horizontal bounding box representation is not suitable for omnipresent directional objects, such as those in aerial images and scene text. Xu et al. [200] proposed a simple and effective framework to detect multidirectional objects (as shown in Fig. 20). Instead of directly regressing the four vertices, it slides the vertices of the horizontal bounding box on each corresponding edge to accurately describe a multidirectional object. Zhou et al. [201] proposed a correlation learning detector based on transformer. It fully leverages the position information and correlation among objects, predicting the rotated bounding boxes for dense objects in remote sensing images.

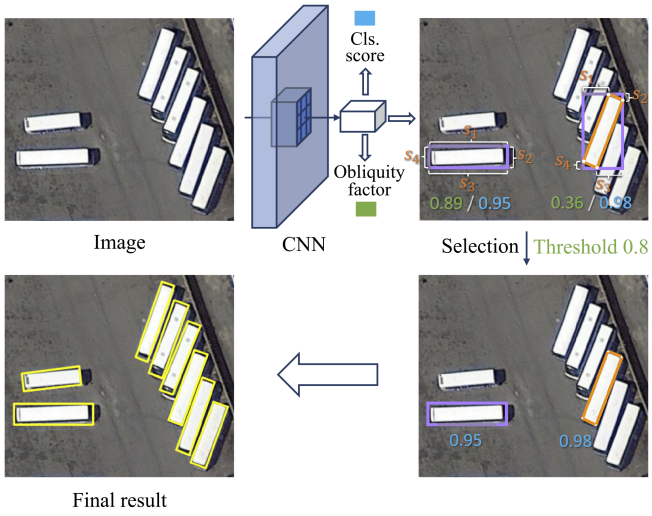


Fig. 20. Pipeline of the proposed method. (Image from [200]).

One-stage detection frameworks: The one-stage detection framework does not generate region proposals and obtains prediction results directly from the input information. Liu et al. [202] adopted the YOLOv2 architecture as the basic network for ship detection and proposed a remote sensing image ship detection framework for any direction. Based on RetinaNet, Yang et al. [203] proposed the R3det detector for the detection of rotating objects. The strategy combines the advantages of the high recall rate of horizontal anchors and the adaptability of rotating anchors to dense scenes and achieves feature alignment using a designed feature refinement module. Wu et al. [204] proposed the optical remote sensing imagery detector (ORSIm detector) with strong robustness using spatial frequency channel features, fast feature channel scaling, and other methods to make it capable of handling complex object deformation behavior in images.

Drawing on the idea of SSD [205], Ma et al. [206] presented an end-to-end scale-aware target detection framework for multicategory target detection tasks, such as large differences in the size of geospatial objects and dense distribution of geospatial objects in the same complex scene. The framework consists of a feature separation and remerging module, an offset error correction module, and a target saliency enhancement module. The feature separation and remerging module aim to eliminate the salient information of larger sized objects in the shallow feature map and highlight the features of small objects. Then, the effective detail features of larger sized targets are passed to the deep feature map to alleviate the problem of easy feature confusion between multiscale targets. The offset error correction module corrects the inconsistency of feature space layout between multilayer feature maps through the proposed offset loss function. The target saliency enhancement module enhances the target features of interest and suppresses background information through the proposed membership function. Finally, the multiscale feature maps containing fine target features are detected to obtain better detection performance (as shown in Fig. 21).

To address the challenge of complex background in remote sensing image target detection, Zhang et al. [207] proposed a foreground-aware remote sensing image target detection model, which enhanced the foreground awareness of the detector from the perspectives of feature relationship learning and network optimization. The method enhanced the discriminative ability of foreground regions in feature maps by building a foreground relation learning module and introducing a foreground anchor loss function to enable the network to focus on the optimization of foreground anchors. A dual network structure based on the transformer architecture is proposed to hierarchically embed the local features into global representations for object detection in remote sensing [208].

4) Object Tracking: Video object tracking is a fundamental prerequisite for scene content analysis and understanding of high-level vision tasks. As shown in Fig. 22, it detects and tracks objects in image sequences. Object tracking is the process of detecting and tracking objects in an image sequence, during which the object is specified in the first frame and further detected and tracked in the next frame of the video [209], [210]. The main purpose of object tracking in the field of remote sensing is to track objects of interest in optical satellite video, aerial video, and UAV video. Remote sensing object tracking is used in intelligent traffic flow monitoring [211], environmental monitoring [212], UAV detection [213], etc. In this section, we focus on discussing the recently emerging object tracking algorithms on satellite videos. Object tracking in satellite video is far different from the natural video. First, satellites have a wide range, usually covering several thousand square kilometers in a single video. Taking Jilin-1 as an example, its resolution is about 1m, so a video has a video size of several thousand by several thousand. In remote sensing videos, objects of interest are often a dozen pixels in size with few appearance features, and it is difficult to distinguish objects by appearance features in complex scenes accurately. Therefore, when designing remote sensing object tracking networks, it is necessary to compensate for the scarce appearance features through perspectives, such as motion models.

Single-object tracking: Generally, generative methods and discriminative methods are the two mainstream single-object tracking frameworks.

1) *Generative models:* Generative tracking methods typically learn a model representing an object in the current frame. In the next frame, a candidate object that is most similar to the object is selected as the tracking result. The model maximizes the similarity or minimizes the corresponding reconstruction error [214], [215]. The object models of early generative algorithms include the Gaussian mixture model, Bayesian network model, Markov model, etc. Wang et al. [216] proposed a high-resolution ship tracking method from coarse to fine. A constrained template matching method was introduced in this method. Frost and Tapamo [217] presented a ship tracking model with shape priors, using level set segmentation to improve the detection performance. Although generative techniques are effective in the majority of the aforementioned scenarios, most of the current approaches only pay attention to the characteristics of the object itself and ignore its correlation characteristics with the environment or other

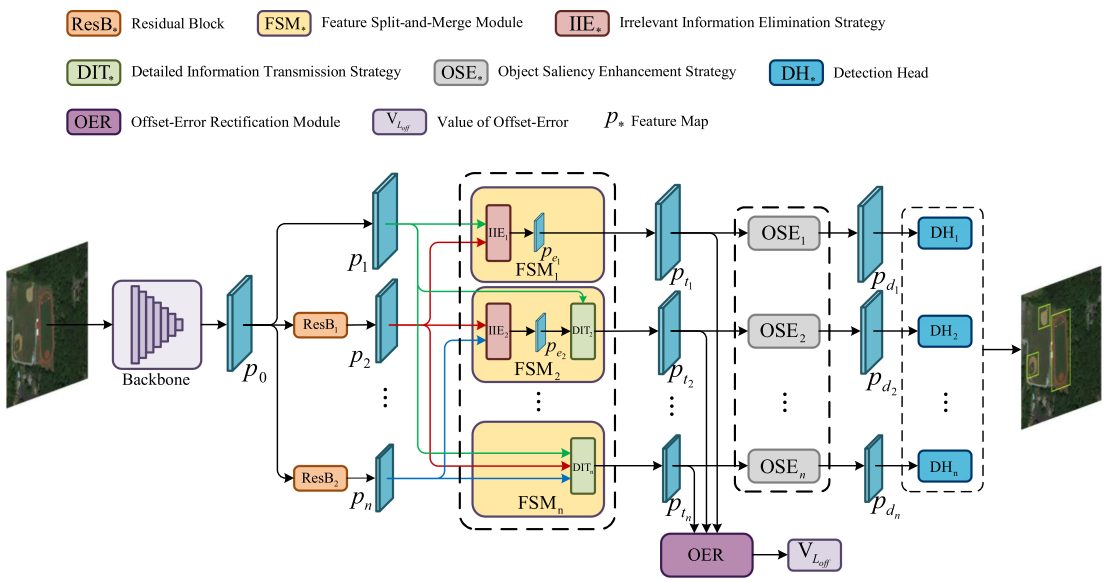


Fig. 21. Framework of split erg enhancement network (SME-Net). (Image from [206]).

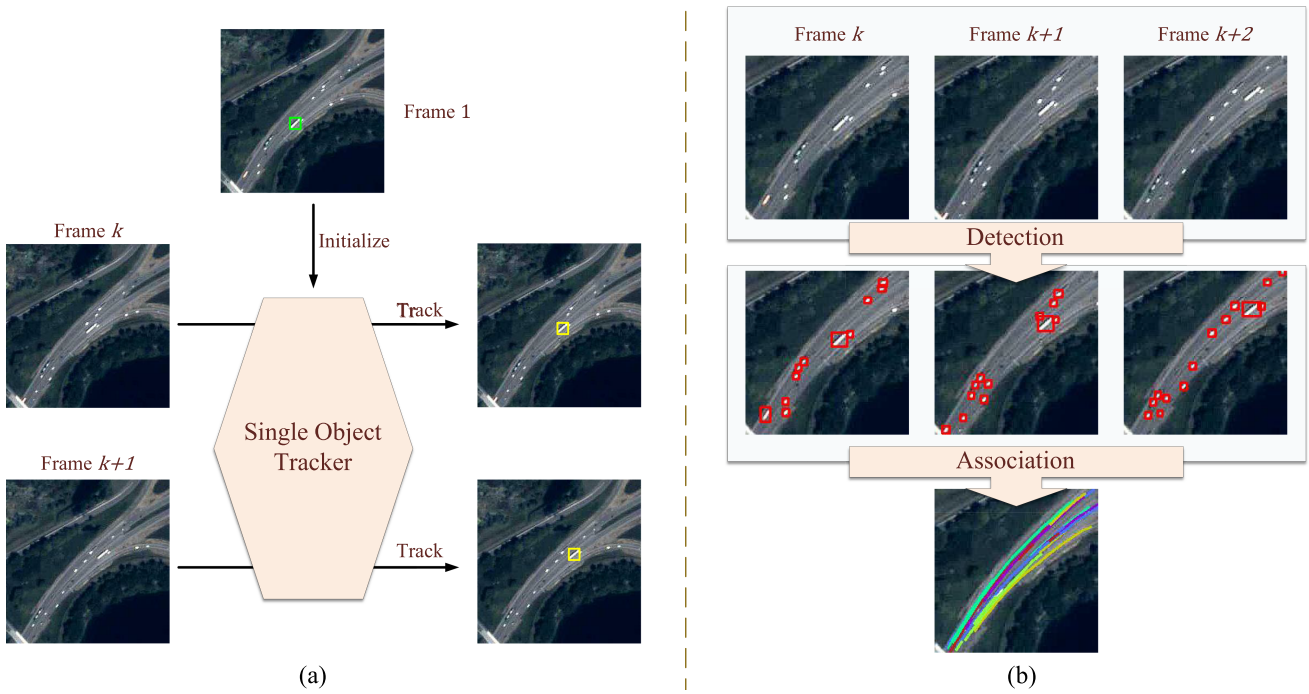


Fig. 22. (a) Single object tracking. The green box is the initial target bounding box to initial the tracker. Yellow boxes are the tracking results. (b) Multiple object tracking can divide into two steps: detection and association. The red boxes are the detection results. The trajectories in difference color represent individual objects.

nonobjects. As a result, since 2010, academics have given discriminative-based approaches more attention.

2) *Discriminative models:* Discriminant tracking methods usually treat tracking as a binary classification problem of distinguishing the object from the background, thereby selecting the object [215]. Currently, discriminative methods represented by the correlation filter and deep learning have achieved satisfactory results and are widely used. The tracker based on the correlation filter extracts the object features according to the object position of the first frame of the video and performs training and learning to obtain the correlation filter, and the extracted features are subjected to Fourier transform, multiplied with the correlation filter, and then inverse Fourier transform, which improves the computational efficiency [218]. Du et al. [219] used the kernelized correlation filter (KCF) tracker [220], a classical algorithm in correlation filtering, for remote sensing video object tracking. According to the characteristics of remote sensing images, KCF is combined with the three-frame difference method to obtain more accurate tracking results. Shao et al. [221] combined KCF and optical flow to propose a VCF tracker for satellite video object tracking.

Since the object lacks appearance features, VCF uses the optical flow map as the object's velocity feature map and uses KCF to track the object on the velocity feature. Also, the inertial mechanism is designed to prevent model tracking drift adaptively by adopting the characteristics of object motion. A correlation filter-based dual-flow tracker is proposed to explore the spatial-spectral feature fusion and motion model for small object tracking [222].

Fu et al. [223] proposed a DRCF tracker based on a double regularization strategy to solve the detrimental boundary effect in DCF-based visual object tracking and enhance the discriminative power of the filter. Xuan et al. [224] introduced a rotation adaptive correlation filter tracking algorithm to address the tracking stability problem caused by the rotation of the object by estimating the rotation angle of the object. From the perspective of features, Liu et al. added deep VGG features on the basis of manual features to extract object features, and expanded the correlation filtering and tracking method of satellite video. An occlusion judgment index is proposed, and the motion trajectory is used to compensate for the occlusion.

However, the algorithms of correlation filter-based trackers tend to use handcrafted features, which often face challenges when the object size is small and the background is complex. Deep learning techniques provide a new research trend. The object tracking algorithm framework based on deep learning obtains the region of interest extracted features from the predicted position of the previous frame rate, and then establishes a deep network-based discriminant model to obtain the tracking result of the current frame of the object.

Compared with the fixed object positioning method of correlation filtering, the deep learning network acquires the positioning ability of object tracking through learning, which makes the algorithm more flexible. The most straightforward implementation of deep learning is to apply the pretrained model directly to remote sensing video tracking. For example, Hu et al. [202] proposed a CRAM network that combines deep learning and optical flow method. Appearance features and motion features are extracted from optical images and optical flow images to alleviate the tracking drift problem. Feng et al. [225] combined the classical algorithm SiamRPN++ of the Siamese network with the frame difference method based on clustering and put forward CDF-SiamRPN++. In CDF-SiamRPN++, the difference map between adjacent frames is divided by the clustering method, which effectively suppresses the interference of environmental noise and retains effective motion information. Shao et al. [226] presented the HRSiam tracker, which combines the high-resolution feature extraction step by HRNet with the SiamRPN tracker. Since HRNet is capable of performing feature extraction and multiscale feature fusion while maintaining high resolution in parallel, applying the extracted high-resolution features to SiamRPN for object tracking leads to a powerful small-object tracking capability.

Song et al. [227] also proposed a tracker based on SiamRPN++. The tracker integrates spatial and channels attention to improve tracking accuracy. Li et al. [91] raised a CRFPF module to establish parallel branches to extract multiscale features, and a collaborative attention learning mechanism is designed to learn the relevant information enhancing the saliency of the objects. Also, an MBLT tracker is proposed to learn the motion and background of the object [228]. The tracking process of MBIT is shown in Fig. 23. First, the DCF tracker generates raw tracking results. Then, a prediction network based on FCN is proposed to estimate the location probability. Third, a feasible region is segmented by FLICM. Finally, the results of the abovementioned three modules are combined to predict the tracking results. To exploit the learning ability of the neural network, deep reinforcement learning is also introduced to track objects in satellite videos. Cui et al. [218] proposed an action decision–occlusion handling network to leverage the occlusion information and drive actions under occlusion.

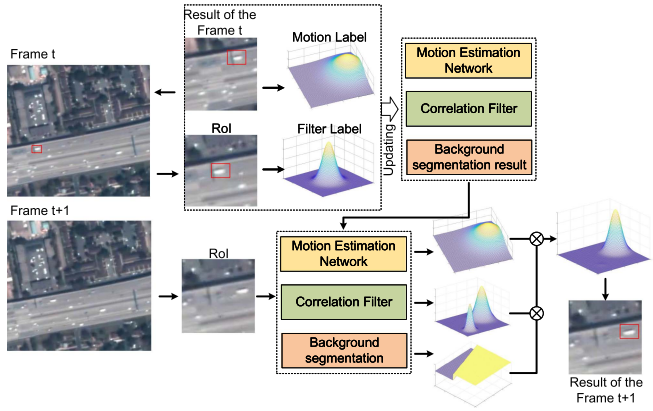


Fig. 23. Tracking process of MBLT.

Multiobject tracking: Compared with single-object tracking, multiobject tracking in remote sensing video has greater application prospects. Multiobject tracking in remote sensing video allows continuously monitor suspicious objects in the military and obtain enemy intelligence; for civilian use, it can monitor traffic flow for statistical analysis, and provide data support for urban management. In remote sensing video, multiobject tracking is divided into three categories: aircraft, ships, and vehicles. Because of the large object size of aircraft and the sparse and less obscured ships moving on the sea surface, few papers have performed multiobject tracking for these two types of objects. He et al. [229] designed algorithms for two types of objects, ships and aircraft in satellite video. This algorithm models multiobject tracking from a multitask learning perspective as a graph information inference process. Through the spatiotemporal relationship module of the graph, the algorithm can mine the potential higher order relationships in the graph.

Compared with the tracking of aircraft and ships, the multiobject tracking of vehicles has received extensive attention. Xiao et al. [230] considered the tracking problem a relational graph matching framework. A joint probability relational graph method is proposed to integrate the road map and the motion of the vehicle to obtain high detection and tracking accuracy in wide-area videos. Zhang et al. [231] proposed a two-step global data association algorithm: First, the local object trajectory of the vehicle is generated, and then, the local trajectory is merged into the global trajectory. The trajectory association model defines a trajectory transition matrix based on Kalman filtering to

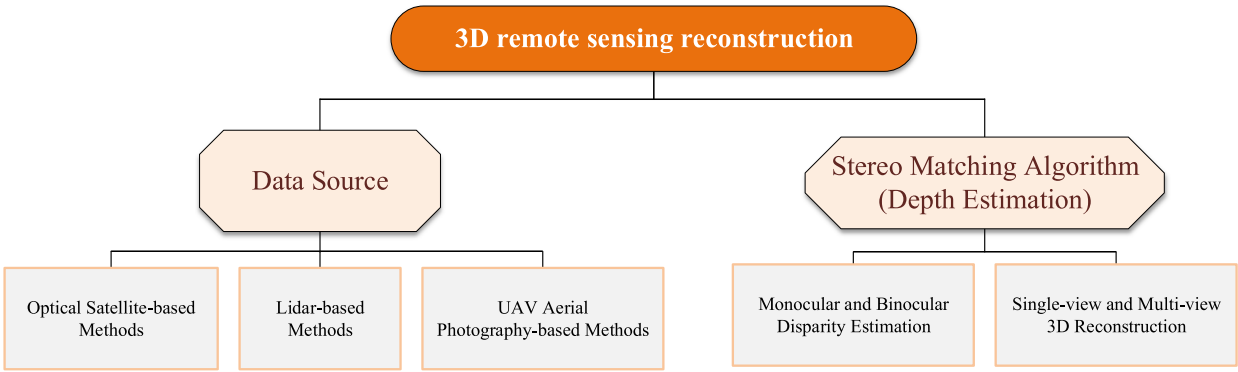


Fig. 24. 3-D reconstruction in remote sensing, including different source data-based remote sensing reconstruction and the stereo matching algorithms.

link trajectories with larger time intervals. At the same time, through the double-layer k shortest path optimization method, the approximate optimal solution to the association problem is obtained.

Ahmadi et al. [232] applied background subtraction to detect moving vehicles and estimate the trajectory, speed, and other information of the vehicle. Zhang et al. [233] also used background subtraction to detect moving vehicles and apply dynamic association methods to match the objects. Ao et al. [234] established a local noise model to distinguish vehicle objects through an exponential probability distribution. Jie et al. [235] proposed a cross-frame keypoint detection network (CKDNet) and a spatiotemporal motion information guided tracking network. CKDNet assists the detection of keypoints by collecting complementary information between frames and efficiently tracks densely arranged vehicles by building a two-branch long short-term memory. Wu et al. [236] presented slow feature and motion feature to guide the multiobject tracking, in which bounding box proposal-guided NMS modules based on SFs enhance the detection of regions of interest.

5) 3-D Reconstruction in Remote Sensing: 3-D reconstruction is a fundamental challenge in the wide remote sensing applications [237]. In this section, remote sensing 3-D reconstruction is mainly investigated according to data sources and stereo matching algorithms, as shown in Fig. 24.

Different source data-based remote sensing reconstruction: According to the data source, the existing remote sensing 3-D reconstruction can be divided into the optical satellite-based, LiDAR-based, and UAV aerial photography-based reconstruction methods [238], [239], [240].

The digital surface model and 3-D reconstruction based on optical satellite technology are also called visual stereo mapping. It mainly uses optical remote sensing satellites to perform high-precision ground stereoscopic observations to obtain ground models. Similar to optical satellite imagery, the ground imagery acquired by the UAV aerial photography method is also visual. It has been demonstrated as an efficient and reliable tool to generate high-precision reconstructions and models of topographic and historical landscape structures [241]. Langhammer et al. [241] used drones to obtain images of abandoned landscapes built for wood flow, and then performs 3-D reconstruction, which is of great significance for water resource management.

For the 3-D reconstruction of optical images, some self-supervised techniques can minimize the distance between the 2-D projection of the reconstruction result and the input image. Some unsupervised methods are based on generative adversarial networks to reconstruct 3-D shapes. By contrast, remote sensing images based on LiDAR scanning have high resolution and strong reliability. The seamless and accurate elevation data it obtains have many applications in the Earth sciences. The data obtained by the LiDAR point cloud device are point cloud data. Each point contains 3-D coordinate information and sometimes includes color information, reflection intensity information, echo frequency information, etc. In short, the digital elevation model, digital surface model,and digital orthophoto that can be generated by LiDAR are used in various aspects, such as urban 3-D modeling, natural disaster assessment, and resource survey.

In addition, the interferometric SAR tomography technology is also used to invert the scattering intensity of ground objects at different heights on the vertical ground, so as to perform 3-D radar imaging. Tomography technology makes it possible to reconstruct the vertical elevation and direction structure of ground objects and has great application potential in terrain mapping, forest parameter estimation, 3-D modeling of urban buildings, and imaging of historical relics.

Stereo matching: Stereo matching has specific research significance in 3-D reconstruction and has certain universality, so it has become a research hotspot of 3-D reconstruction. The general process of stereo matching is as follows. After the image is preprocessed, the idea of the global method (the path on the right in the abovementioned figure) is to use the global information to perform disparity optimization (disparity optimization). It seeks to find the optimal disparity result for each pixel so that the international and overall matching cost is minimized [242]. The disparity calculation in the local stereo matching algorithm is generally relatively simple, and the WTA winner-take-all theory is used to search for the disparity directly. Both methods need to perform parallax postprocessing after calculating the parallax drawing. After the disparity map is initially obtained, the results of the disparity map are judged, and possible matching errors are found and corrected. Common disparity postprocessing methods include left–right consistency detection, occlusion filling, and weighted median filtering.

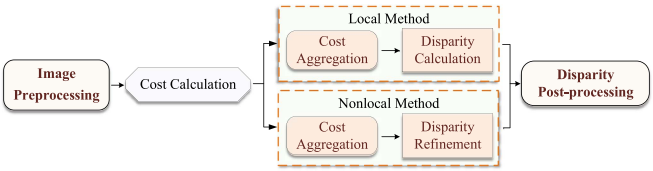


Fig. 25. General process of the stereo matching.

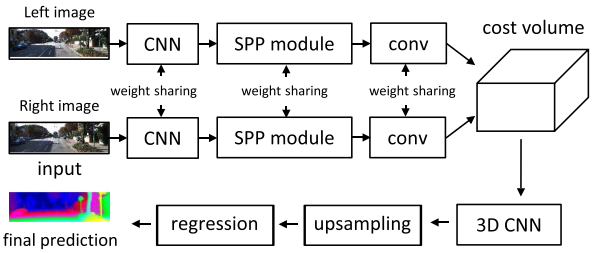


Fig. 26. Architecture overview of classical PSMNet.

Stereo matching has certain research significance in 3-D reconstruction, and it has certain universality, so it has become a research hotspot of 3-D reconstruction. The general process of the stereo matching is shown as Fig. 25. After the image is preprocessed, the classic idea of using global information for parallax optimization is to find the optimal parallax result for each pixel, so as to minimize the global and overall matching cost [242]. This step is called disparity optimization. Disparity calculation has become one of the research focuses in existing stereo matching. Depth and disparity can be directly converted to each other, so depth estimation has also become a research hotspot of stereo matching.

Depth estimation: Depth estimation, using one or only/multiple viewing angles of the RGB image, estimates the distance of each pixel in the image relative to the shooting source. It is a critical step in the task of scene reconstruction and understanding and is part of the 3-D reconstruction. In addition to the costly method of obtaining depth point clouds by using LiDAR or the reflection of structured light on the object's surface, the most common traditional depth estimation methods are monocular and binocular ranges. In contrast, the amount of calculation of the monocular ranging method is complex, and the accuracy is not as high as that of the binocular, and it is often used when the conditions are challenging. Deep learning has also continued to develop in depth estimation methods.

1) *Monocular and binocular disparity estimation:* There are mainly monocular estimation and binocular estimation methods. There are many common deep learning monocular ranging methods. For example, Facil et al. [243] proposed CAM-Convs convolution, which can take the camera parameters into account, so that the neural network can learn to calibrate the perception mode. Wang et al. [244] proposed a motion feature that considers one of the most important features of the human visual system. It employs an RNN to train with multi-view image reprojection techniques to improve monocular depth estimation. Tosi et al. [333] proposed monoResMatch, which combines features from different angles, keeps consistent with the input image, and performs stereo matching between two cues to infer from a single input image to the novel deep learning framework. The overview in [245] investigates deep learning binocular depth estimation methods and gives a comparison of 16 deep learning depth estimation methods, including the GANet [246], PSMNet [247], and SegStereo [248] in 2018 and 2019. In recent years, some relatively advanced methods have also appeared, such as PlaneMVS [249], Nerfingmvs [250], and in [251] and [252]. The architecture overview of PSMNet is shown in Fig. 26. It is a typical model for binocular disparity estimation. The left and right images are the model's input. CNN is taken as the feature extraction module along with the SPP module for feature harvesting. Then, the extracted features are concatenated together as the input of the cost volume module. Finally, a 3-D CNN with unsampling and regression module is designed for the cost volume regularization and disparity regression.
As for the disparity estimation in remote sensing [253], [254], [255]. Among them, Yu et al. [253] mainly uses 2-D discrete wavelet transform to enhance the local invariant features of the existing weighted α-shape (W α SH). It is used in remote sensing images with less affine distortion and less noise. Experiments perform that it can effectively alleviate the image matching problems of geometric distortion and radiation distortion in stereo remote sensing images. In addition, a novel edge-aware bidirectional pyramid stereo matching network is suggested in [254] to enhance performance in textureless regions while preserving the primary structure. It can effectively solve the problem of poor disparity estimation accuracy caused by occlusion areas of high-rise buildings and textureless areas. Jia et al. [255] tried to use CNN to match remote sensing stereo images of featureless areas, such as the lunar surface.
2) *Single-view and multiview 3-D reconstruction:* From the perspective of view, existing related algorithms can be divided into single-view and multiview 3-D reconstruction methods. Single-view 3-D reconstruction refers to the realization of 3-D reconstruction of an image or target given a single image. The majority of single-view 3-D understanding techniques currently in use employ an encoder–decoder structure, where the encoder converts the input image into a latent representation and the decoder must engage in complex analysis of the 3-D structure of the output space [256].
Although single-view 3-D reconstruction can generate different 3-D results (such as point clouds or meshes), it can also handle many disordered images. Remote sensing image reconstruction in three dimensions is essential for tracking changes to the Earth's surface. In [257] and [258], the authors orderly predict depth from a given image and estimate a single-view spherical map from depth under the same view. However, single-view 3-D reconstruction results usually lack completeness and accuracy, especially when there are obstacles or occluded areas.

TABLE II
SUMMARY OF PUBLIC DATASETS FOR VARIOUS REMOTE SENSING APPLICATIONS

Dataset Name	Image Size	Image Channel	Image Number	Category Number	Label Type	Spatial Resolution	Image Type	Year
Classification								
Washington DC [268]	600 * 600	191	1	7	–	–	Satellite Image	2012
WHU-RS [269]	1280 * 307	3\4	150	15	–	0.8–10 m	Aerial Image	2013
NWPU-RESISC45 [270]	256 * 256	3	300	8	–	–	Satellite Image	2017
GID [271]	7200 * 6800	3\4	150	15	–	0.8–10 m	Satellite Image	2018
Aeroscapes [272]	720 * 720	3	3269	11	–	–	Aerial Image	2018
UAVid [273]	4096 * 2160	3	300	8	–	–	Aerial Image	2020
DFC22 [274]	2000 * 2000	3	766	12	–	0.5 m	Aerial Image	2022
Change Detection								
WHU Building Change Detection Dataset [275]	32207 * 15354	3	1	1	–	0.075 m	Aerial Image	2018
LEVIR-CD [276]	1024 * 1024	3	637	1	–	0.5 m	Satellite Image	2020
OSCD dataset [277]	600 * 600	13	24	1	–	10–60 m	Satellite Image	2018
CDD Dataset [278]	256 * 256	3	16000	1	–	0.03–1 m	Satellite Image	2018
River HSIs dataset [279]	463 * 241	198	1	1	–	30 m	Satellite Image	2019
xBD [280]	1024 * 1024	3\4\8	11034	4	–	0.5 m	Satellite Image	2019
HRSCD [281]	10000 * 10000	3	291	6	–	0.5 m	Aerial Image	2019
DSIFN Dataset [282]	512 * 512	3	3988	1	–	2 m	Satellite Image	2020
Google dataset [283]	256 * 256	3	1067	1	–	0.55 m	Satellite Image	2020
SYSU-CD Dataset [284]	256 * 256	3	20000	1	–	0.5 m	Aerial Image	2021
Target Detection								
NWPU VHR-10 [285]	500 * 500-1100 * 1100	3	1510	10	HBB	0.08–2 m	Satellite Image	2014
RSOD [286]	1000 * 1000	3	976	4	HBB	0.3–3 m	Satellite Image	2017
TGRS HRRSD [287]	152 * 152-10569 * 10569	3	21761	13	HBB	0.15–1.2 m	Satellite Image	2017
VisDrone2019-DET [288]	1000 * 1000	3	10209	10	HBB	–	Aerial Image	2018
DIOR [289]	800 * 800	3	23463	20	HBB, OBB	0.5–30 m	Satellite Image	2019
iSAID [290]	800 * 800-13000 * 13000	3	2806	15	OBB	–	Satellite Image	2019
HRSID [291]	800 * 800	1	5604	1	HBB	0.5–3 m	Satellite Image	2020
VISO-Detection [264]	1000 * 1000	3	32825	4	HBB	0.5–1.1 m	Satellite Image	2021
DOTA [292]	800 * 800-20000 * 20000	3	11268	18	OBB	–	Aerial & Satellite Image	2021
Remote sensing Video								
Satellite Video MOD [293]	400 * 400 \600 * 400	1	1400	1	HBB	1.0 m	Satellite Video	2014
UAV123 [294]	720 * 1280	3	110000	12	HBB	–	UAV Video	2017
VisDrone2019-SOT [288]	3840 * 2160	3	1393000	10	HBB	1–1.2 m	UAV Video	2018
VisDrone2019-MOT [288]	3840 * 2160	3	39988	10	HBB	1–1.2 m	UAV Video	2018
UAVDT [295]	1080 * 540	3	37084	5	HBB	–	UAV Video	2018
VISO [264]	1000 * 1000	3	32825	4	HBB	0.5–1.1 m	Satellite Video	2021
SatSOT [296]	12000 * 5000	3	27664	4	HBB	–	Satellite Video	2022
SV248S [2]	4096 * 2160	3	156621	4	Polygon	0.92 m	Satellite Video	2022
Urban-level 3D Point Cloud								
Dataset Name	Area^2 (km^2)	Category Number	Instance Category Number	Point number	RGB	Sensors	Years	
DublinCity [297]	2	13	–	260 M	No	ALS	2019	
DALES [298]	10	8	–	505 M	No	ALS	2020	
LASDU [299]	1.02	5	–	3.12 M	No	ALS	2020	
Campus3D [300]	1.58	24	–	937.1 M	Yes	UAV Photogrammetry	2020	
Sensaturban [301]	4.4	13	–	2847 M	Yes	UAV Photogrammetry	2021	
STPLS3D [302]	17.25	18	14	–	Yes	UAV Photogrammetry	2022	

HBB: Horizontal bounding box. OBB: Oriented bounding box.

Multiview 3-D reconstruction alleviates and solves the above-mentioned problems to a certain extent. There are two main types of multiview reconstruction: one is to reconstruct stationary objects from images of two or more views, and the other is to reconstruct 3-D shapes of moving objects from video or multiple frames [259]. Multiview reconstruction is flexible and scalable, which can be adapted to large-scale scenarios. Of course, there are still numerous obstacles to overcome in order to accurately rebuild multiview depth maps in urban landscapes, such as the presence of repeating textures and texture-poor places. To address the aforementioned issues, Hu et al. [260] proposed a multiview 3-D reconstruction (IMGTR) method based on image triangles. Rupnik et al. [261] proposed to generate high-quality digital surface models by combining many depth maps that were calculated using a dense image matching method. It performs well at reconstructing surface discontinuities, repeating patterns, and nontextured surfaces.

V. IMPLEMENTATION OF REMOTE SENSING

A. Public Datasets

Deep learning algorithms have demonstrated excellent performance in various fields. This is inseparable from using large amounts of finely labelled data for neural network training. Researchers need to use labeled data to develop algorithms to meet different applications. Commonly used remote sensing datasets are summarized in Table II and categorized according to the tasks for which they are mainly applied.

B. Software Platforms

In recent years, Earth observation technology has developed tremendously, and large-scale remote sensing data are stored, recorded, and developed for free use by society and researchers [262], [263], [264]. However, traditional remote sensing interpretation methods require users to download and process data on local computers. For example, image processing platforms, such as the Environment for Visualizing Images (ENVI), can perform image enhancement, orthorectification, data fusion and transformation, knowledge-based decision tree classification, and other functions on the image after the user obtains the data. This platform is an offline software installed on a single machine that assists people in data preprocessing and simple image recognition tasks [265]. With the increased data, the computing power to store and interpret data locally is facing enormous challenges.

Platforms for remote sensing applications have started to move toward the cloud as the Internet has grown [266], [267]. The remote sensing platform deployed in the cloud has the following characteristics.

1) The cloud platform can provide abundant storage and computing resources. Users can efficiently process large-scale remote sensing data;

2) Computation-intensive tasks are performed through cloud servers, reducing the computing power requirements of the user's computer and lowering the threshold for software use.
3) Users can access the platform by any device which can access the Internet and perform tasks, such as remote sensing image processing and analysis anytime, anywhere.
4) Accessing the platform through web pages, users can obtain the latest data and update functions of platforms at any time and use the latest algorithms to process the latest data and improve work efficiency.

The abovementioned advantages are not available in traditional image processing tools. Therefore, various research institutes and companies have invested in constructing remote sensing cloud platforms. This platform can perform interpretation services in the cloud without downloading the data locally. These platforms integrate various tools and applications to provide users with a complete data acquisition and processing solution. From the perspective of usage, the mainstream platforms can be classified into two types: one is the remote sensing data cloud platform for professional users with programming tools. This platform requires users to use the provided application programming interfaces (API) for data manipulation and processing, such as Google Earth Engine (GEE). Through various flexible APIs, professionals can customize functions and algorithms for their own needs to achieve different functions. The other type is the remote sensing data cloud platform for ordinary users. This type of platform further encapsulates data and algorithms. Users only need to select or upload data in the corresponding format and select the task to be interpreted. The platform will be able to realize automatic algorithm processing and visualization of data and results, such as Remote Sensing Data Intelligent Interpretation Platform, SenseEarth, and so on. Through simple and convenient operation, ordinary practitioners can also interpret remote sensing data, which benefits the civilian promotion of remote sensing technology. In this section, we select the GEE platform and the Remote Sensing Data Intelligent Interpretation Platform for introduction and show the specific characteristics of these two types of platforms, respectively.

1) Google Earth Engine: GEE is a remote sensing interpretation cloud platform launched by Google in 2010. This platform is one of the most popular big data geographic information processing platforms. The platform provides users free services to discover, analyze, and visualize big geospatial data based on Google's computing infrastructure.

In GEE, different third-party network applications can be implemented through the interfaces provided by the platform. For researchers who use the GEE platform, it is essential to use the API provided by the platform. GEE provides APIs in two languages, JavaScript and Python, to meet the needs of most programmers. Through different APIs, users can easily access data, use various applications provided by GEE, and view the running results in real time. The platform is divided into three parts: Data catalog and Explorer, Code editor, and Timelapse.

Data catalog and explorer: The data catalog contains a significant amount of geospatial data, which collects numerous publicly accessible satellite images, including the Landsat, MODIS,

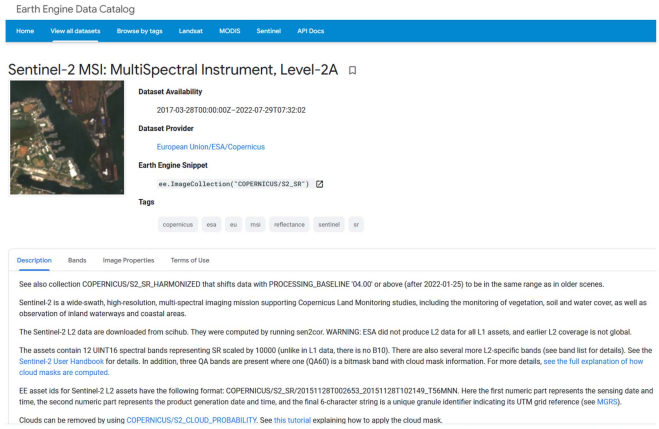


Fig. 27. Data catalog of GEE.

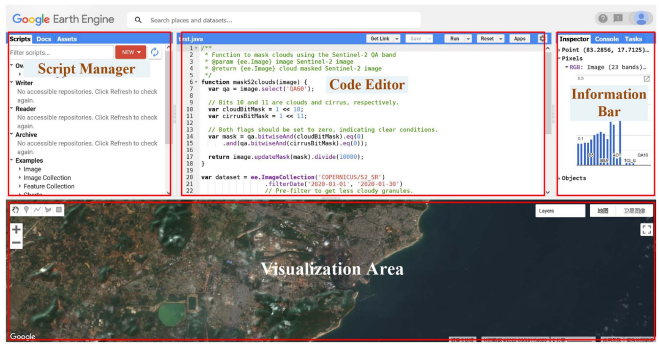


Fig. 28. Code editing platform of GEE.

and Sentinel images, as well as numerous atmospheric, meteorological, and vector datasets. The datasets cover various satellite and air systems for optical imagery, environmental variables, weather and climate forecasting, land cover, and socioeconomic.

Fig. 27 is the data content page captured by the multispectral instrument of the Sentinel-2 satellite in the data catalog. This content page shows visual thumbnails of the data, the time when the data are available, the dataset provider, the API used to access the data, and a detailed description of the data. Users can browse the data catalog, select the required dataset according to the dataset description, and use the provided API to obtain the data. The data can then be quickly visualized via explore, provided by GEE.

Code editing platform: The code editing platform is GEE's main platform for data acquisition, processing, analysis, and visualization. As shown in Fig. 28, the code editor is mainly divided into four functional blocks: visualization area, script manager, code editor, and information bar.

Below the page is the visual area of the code editing platform. This is the main area for user interaction, data, and result visualization. This area uses the world map as the base map to provide basic geographic location information. The data and code analysis results are displayed by stacking of multiple layers. Users can drag and zoom the results in the visualization area, mark the position by clicking, and so on. The location information of the marker will be displayed on the Inspector page of the information bar.

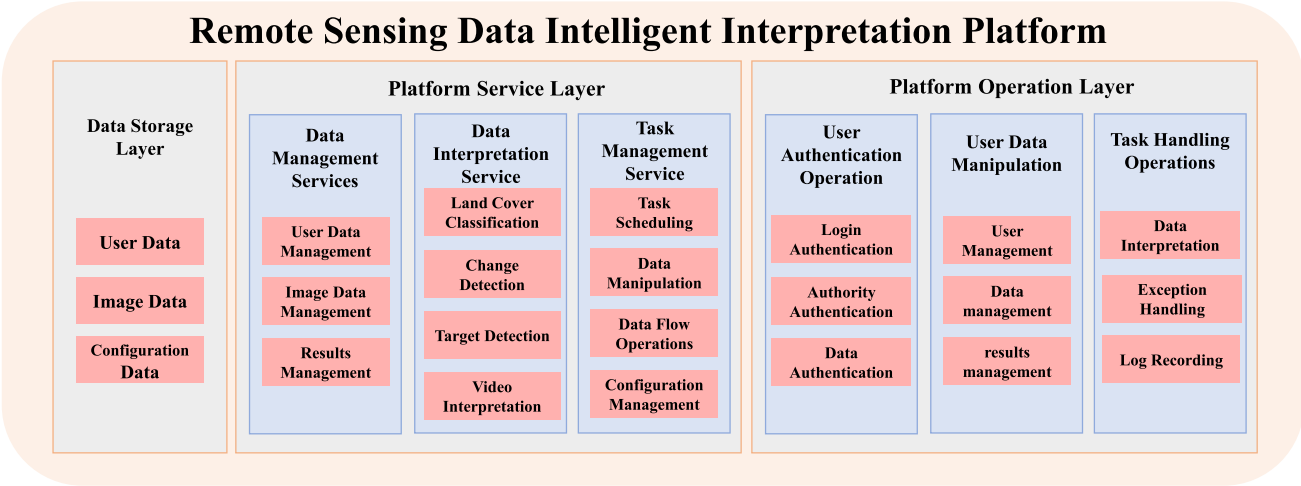


Fig. 29. Framework of remote sensing data intelligent interpretation platform.

On the left-hand side of the page is the script manager, which stores scripts edited by users and sample scripts provided by the GEE platform. Through the manager, users can select or delete their scripts. At the same time, the sample scripts provided by GEE cover image acquisition, preprocessing, visualization, drawing, etc., and provide demos, such as classification, climate modeling, terrain visualization, etc., to provide users with complete code usage demonstrations.

In the middle of the page is the code editor area through the cloud platform infrastructure provided by Google. By editing JavaScript and Python code, users do not need to consider the problem of the code running environment. Someone can run the code directly by clicking the "Run" button at the top of the page after writing the code.

On the right-hand side of the page is the information window. The info window includes Inspector, Console, etc. The Inspector displays information about the user's markers on the map. The Console will display the output print of the code running.

Simple mathematical operations to sophisticated image processing and ML functions are all available on the platform.

By writing code, users can fully utilize the functions of the GEE platform. The GEE platform provides rich data and API, the focus of its widespread use. However, since it is free and open to the public, computationally intensive tasks, such as deep learning, cannot be widely supported. Users are limited to varying degrees in training models, data acquisition, and designing new methods and functions.

Timelapse: Based on nearly 40 years of data stored on the GEE platform, the Timelapse project generates scalable video worldwide. The project stitches together one image annually into a video for each region, showing people the Earth's changes in time and space. In this project, we can record the most realistic records of natural and human activities, such as glacial fusion, bushfires, and urban development.

2) Remote Sensing Data Intelligent Interpretation Platform: Different from Google Earth Engine in Section V-B1, "Remote Sensing Data Intelligent Interpretation Platform" is designed to meet practitioners' need to interpret remote sensing data. By encapsulating the relevant functional blocks, users can straightforwardly operate the platform. With the help of artificial intelligence algorithms, the platform integrates available blocks, such as data interpretation, data management, and scene application, which realizes algorithm processing automation and data interpretation results visualization. The platform can perform real-time extraction and identification of target information from full-modal remote sensing data, such as panchromatic, visible, multispectral, hyperspectral, SAR images, and satellite videos. Currently, the platform has opened four primary functions, including land-cover classification, object detection and recognition, element change detection, and intelligent video interpretation, which offers practitioners technical assistance for processing data from remote sensing.

As shown in Fig. 29, the system architecture of the "Remote Sensing Data Intelligent Interpretation Platform" comprises three parts: data storage layer, platform service layer, and platform operation layer.

The data storage layer mainly contains user, configuration, and image data. User data record relevant information of users. Configuration records relevant information of remote sensing data. Image data include public datasets provided by the platform and private datasets uploaded by users that are only visible to owners. Image data are all stored in the cloud, which significantly reduce the pressure of user data storage and can quickly provide data support for interpretation tasks.

The platform service layer mainly includes three parts: data management service, data interpretation service, and task management service. The data management service manages the user data, image data, and interpretation results. It can operate the data in the cloud with the help of the instructions of the platform operation layer. The image interpretation service integrates various artificial intelligence algorithms. It determines the interpretation tasks through the platform operation layer and then efficiently completes tasks, such as land-cover classification, target detection, change detection, and video target tracking. The task management service is mainly responsible for data retrieval, parameter transfer, and task scheduling. When the user creates multiple tasks through the platform operation

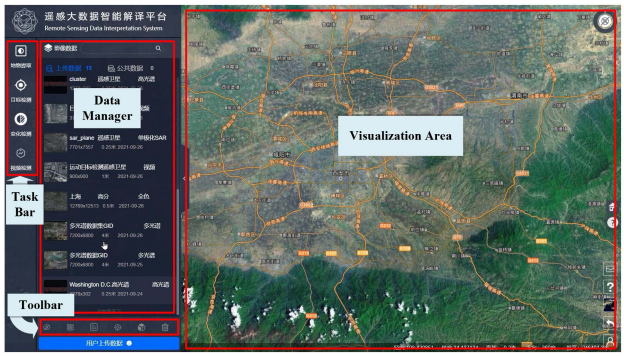


Fig. 30. Main page of big data intelligent interpretation platform.

layer, the layer needs to schedule the tasks and provide them with the corresponding initialization parameters and image data.

The platform operation layer mainly consists of user authentication, user data, and task processing operations. User authentication operations can use the user information stored in the cloud for authentication and give users operation privileges. User data operation can read and modify user data, image data, and interpret results in the data storage layer. The task processing layer is mainly responsible for assigning tasks to the platform operation layer and providing feedback on exception information and log information from the platform operation layer.

Based on the abovementioned architecture, the platform contains two critical systems: the User Interaction System (UIS) and Data Interpretation System (DIS).

The client is mainly the interface between the user and the platform. Users can access the client through a browser to perform data uploading, browsing, interpretation task execution, and analysis and display the result. The server is responsible for data storage management and performing different interpretation tasks.

UIS: The UIS is the core system for users to interact with the platform. Users can use the Internet to access web pages at any time and enter the UIS to perform interpretation tasks after logging in and authenticating. Fig. 30 shows the system operation page after login. The system operation page is divided into four areas: task module, data list, data display area, and function module. In the task module, users can choose the type of task they want to perform. In the data list, public datasets and privately uploaded remote sensing images are displayed in thumbnails; the data display area will display the remote sensing images selected by the user and the corresponding remote sensing images in real time. Interpret the result. Users can drag and zoom in this area for data browsing. The function options provide users with functions, such as "image transparency selection," "visualization channel selection," "image zooming," "interpretation result selection," and so on.

DIS: As the core of the Remote Sensing Data Intelligent Interpretation Platform, the DIS is mainly responsible for intelligently interpreting remote sensing images, efficiently and accurately mining the adequate information of remote sensing images, and providing users with real-time analysis services of remote sensing data. The platform contains four primary tasks: land-cover classification, object detection and recognition, element change detection, and intelligent video interpretation. Each task is divided into subtasks according to the target type and data source, such as SAR, visible, multispectral, and hyperspectral. Land-cover classifications are divided into road classification, water classification, building classification, and land-cover classification. Object detection and identification are divided into aircraft, bridge, and ship detection. Video intelligent interpretations are divided into single target tracking, multitarget tracking, and motion target detection. The available tasks are given in Table III. Fig. 31 shows the interpretation results of some tasks, such as change detection of SAR, ship detection of SAR, water classification of HSI, object tracking, and multiaircraft tracking.

TABLE III
AVAILABLE TASKS IN BIG DATA INTELLIGENT INTERPRETATION PLATFORM

Tasks	Sub-Tasks	Data
Land cover classification	Road extraction	Optical
	Water extraction	Optical
	Building extraction	Optical
Object detection	Airplane detection	Optical, SAR
	Bridge detection	Optical
	Ship detection	Optical, SAR
Change detection	Change detection	Optical, SAR
Video analysis	Single object tracking	Optical
	Multi-objects tracking	Optical
	Moving objects detection	Optical

C. Hardware Systems

In conventional research, researchers usually use multiple graphics processing units (GPUs) or computer clusters for algorithm research [302], [303], but they ignore the constraints of energy consumption and computing resources. Although many algorithms can achieve excellent results under GPU acceleration, there is still a long way to go from the requirements of the actual industry. Many complex models cannot be deployed on small devices or computed in real time, which are the main problems confusing many engineers.

In applying remote sensing algorithms, the research and development of hardware systems are more urgent. Currently, most remote sensing algorithms are calculated at ground computing stations, which significantly affects the application of remote sensing technology and the complete mining of remote sensing data. The main existing requirements are divided into three points.

1) *Real time:* Real time can also be called nondelay, which requires equipment to have a fixed processing time when processing data to ensure stable processing of data streams.
2) *Data volume:* The amount of data captured by remote sensing satellites are significant, and not all data can be sent to the ground for the processing. This requires hardware devices that can be mounted on aircraft and satellites for processing and only transmit essential data to improve data utilization efficiency.

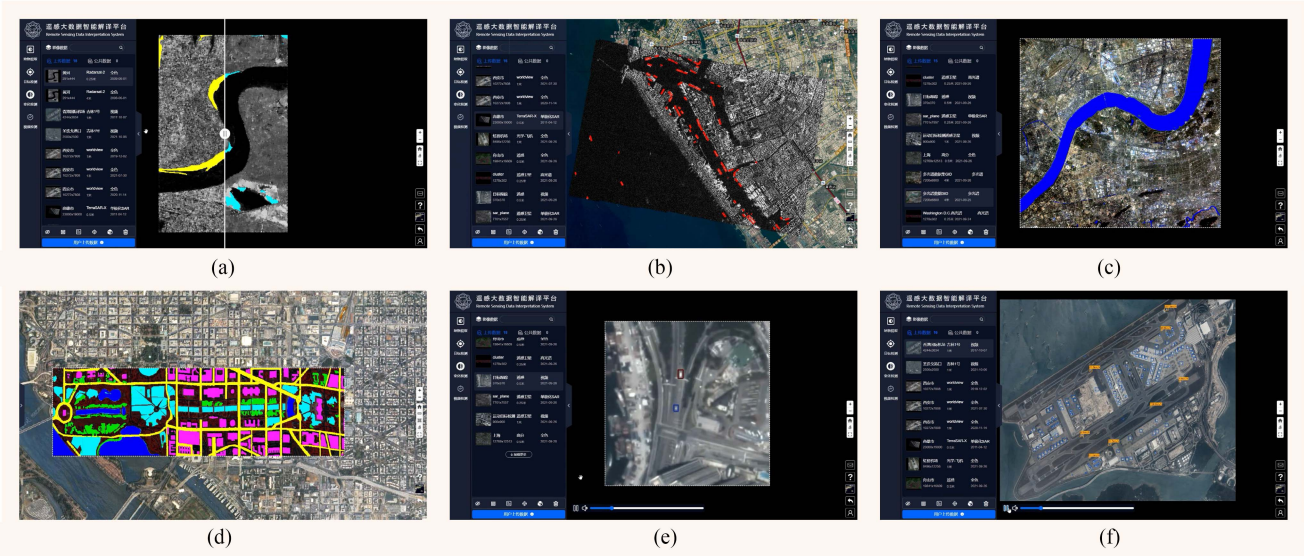


Fig. 31. Interpretation results of some tasks in big data intelligent interpretation platform. (a) Change detection of SAR. (b) Ship detection of SAR. (c) Water classification of MSI. (d) Landcover classification of HSI. (e) Object tracking. (f) Multiaircraft tracking.

3) *Power consumption:* Airborne and satellite-based devices require low power consumption due to batteries and other power supplies. Low power consumption can prolong the use of electricity. Therefore, this chapter summarizes the mainstream hardware platforms and selects field programmable gate array (FPGA) devices that are easy to develop, computationally stable and low power for further research.

1) Classification of Hardware Systems: All chips capable of running AI algorithms, including CPUs, can be called AI chips. In the traditional von Neumann structure, each instruction executed by the CPU needs to read data from memory and operate on the data according to that instruction [304]. From this feature, the primary responsibility of the CPU is not only data operations, but also executing commands, such as memory reading, instruction analysis, and branching. However, most AI algorithms, especially deep learning algorithms, usually require a lot of data processing. When the CPU executes the algorithm, the CPU is limited to serial execution, which will spend a lot of time reading and analyzing data/instructions. This is why algorithms cannot be suitable for parallel processing intensive data and cannot fully utilize the chip's potential. Therefore, the computing framework is usually performed heterogeneously, combining a CPU and a computing card. The CPU performs data reading and other operations on the data, and the computing card implements large-scale and intensive mathematical calculations. Generally speaking, AI chips refer to chips that are different from CPUs and are especially designed for acceleration according to the characteristics of artificial intelligence algorithms. According to the technical architecture, it can be divided into GPU, application-specific integrated circuit (ASIC), FPGA, and neuromorphic computing chip [305] (as shown in Fig. 32).

GPU: The GPU has a relatively straightforward architectural design. As a result of the majority of transistors forming several dedicated circuits and pipelines, the GPU outperforms the CPU in terms of computation performance. The GPU also has strong floating-point computing capabilities, which can help deep learning algorithms overcome the computing pressure and release the full potential of AI. GPU development has reached a relatively mature stage at this time. GPUs are being used by businesses, such as Google, Facebook, Microsoft, Twitter, and Baidu, to analyze image, video, and audio assets to improve search engines and image intelligence software. In addition, GPU is appropriate for various industries, such as VR/AR and unmanned driving. But GPUs also have some limitations. Training and inference are the two phases of the deep learning algorithmic process. The GPU platform is a productive platform for training algorithms. However, when processing a single input for inference, the benefits of parallel computing cannot be completely realized. The GPU also consumes a lot of power and cannot work independently. A CPU is required to schedule it to work.

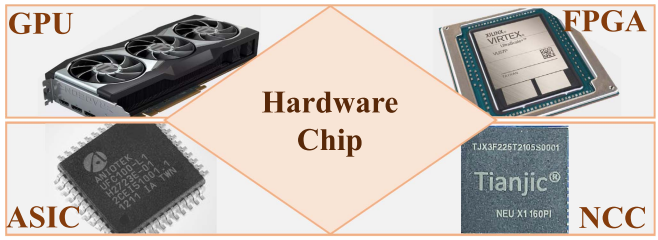


Fig. 32. Four mainstream hardware chips. GPU: Graphics processing unit. ASIC: Application-specific integrated circuit. FPGA: Field programmable gate array. NCC: Neuromorphic computing chip.

ASIC: The ASIC is a specialized customized chip designed to meet a particular requirement. For high-performance, low-power mobile applications, customized features benefit ASICs' performance-to-power ratio and have advantages in terms of reliability and integration. Google's TPU, Cambrian Chips, Horizon's BPU, and Amazon's Inferentia are all ASIC chips. Artificial intelligence applications are ideal for ASIC devices.

First, the fully customized circuit of ASIC can boost performance. Google's TPU is 30 to 80 times quicker than CPU and GPU solutions while using less power and space. Second, downstream demand encourages the specialization of artificial intelligence chips. Due to the real-time requirements and the privacy of training data, the computing of many application scenarios cannot wholly rely on the cloud. The local software and hardware must support it. However, the long design cycle of ASIC cannot accommodate the advancement of the algorithms that restrict its use.

FPGA: The full name of FPGA is "field programmable gate array." Two characteristics can be identified when comparing FPGA and CPU. First, the FPGA does not have the storage brought by memory and control. Thus the data reading is quicker. Second, it uses less energy because the FPGA does not need a reading command. At the same time, FPGA is different from GPU. FPGA provides more pronounced efficiency improvements in specific applications thanks to its parallel pipeline and data parallel processing capabilities. FPGA is frequently employed in the inference phase of deep learning algorithms because it is ideal for data processing on the hardware pipeline and has excellent operation performance. In addition, FPGA provides the advantages of design flexibility and speed over ASIC. The modification of the algorithms can be easily deployed in the FPGA without redesigning the circuit. Because of its flexibility and performance, it frequently replaces ASIC in various industries.

Neuromorphic computing chip: A neuromorphic computing chip is a circuit simulating the computing mechanism of the brain from a structural perspective. This technology is still in the development stage. Its research work can be further divided into two levels. One is the neural network level, which corresponds to the neuromorphic architecture and processor. Its memory, CPU, and communication components are fully integrated, and information processing is carried out locally, eliminating the usual speed bottleneck between computer memory and CPU. Neurons can readily and swiftly communicate with one another. These neurons will activate simultaneously as long as they receive other neurons' pulses (action potentials). The Truenorth chip from IBM and the Tianji chip from Tsinghua serve as examples. The second is the level of neurons and synapses, and the corresponding innovation is the level of components. For instance, the world's first artificial stochastic phase-change neurons, capable of achieving high-speed unsupervised learning, were produced by IBM Zurich Research Center [306]. Although neuromorphic computing chips are not yet completely developed and there is still some distance between large-scale applications, it has the potential to revolutionize computer architecture.

2) FPGA Structure and Advantages: As early as the 1960s, Gerald Estrin proposed the concept of reconfigurable computing. But it was not until 1985 that Xilinx introduced the first FPGA chips. Although the parallelism and power consumption of the FPGA platform is excellent, the platform has not received much attention due to its high reconfiguration cost and complicated programming. Unlike GPUs and CPUs under the Von Neumann-style architecture, although FPGAs are more difficult to develop, they still have many advantages. The following is discussed in five aspects [307].

1) *Development time and difficulty:* The development time and difficulty of FPGA are between ASIC and GPU. Usually, algorithms are developed directly on the GPU using mature algorithm frameworks. After an algorithm has been designed, the operators required by the algorithm must be prepared first, and then, the algorithm can be deployed on the FPGA. Modern deep networks are usually stacked with a series of fixed operations (such as convolution and pooling), so these commonly used operators can be used directly. Companies, such as Xilinx, provide corresponding deployment toolkits. Users can use the toolkit to directly convert the programmed algorithms of deep learning frameworks, such as TensorFlow, and deploy them on the FPGA, significantly reducing the time and difficulty of FPGA development.
2) *Flexibility:* FPGAs can provide a more flexible architecture. Its flexibility is mainly reflected in programmable computing resources and IO interfaces. The computing resources on the FPGA are programmable hardware resources of a mixture of DPS and block random access memory modules. Users can realize large-scale parallel computing by configuring data channels, single instruction multithreading, etc., to meet the needs of the required workloads. At the same time, any programmable IO connection allows the FPGA to connect to any device (network or storage device) without the help of the CPU to assist in data scheduling, dramatically improving the FPGA's flexibility in use.
3) *Real time:* The inside of the FPGA chip is realized by hardware through millions of logic units. The hardware connections between logical units represent the algorithm flow. In this way, the FPGA avoids the operation of reading the operation instruction. At the same time, through the connection to the storage on the hardware, each logic unit is directly configured with separate storage, which avoids the need to apply for memory, arbitration, and other operations in GPU computing, and further improves the stability of the FPGA. Combining the abovementioned two points, FPGA can ensure data reading and algorithms' execution at the hardware level. The performance of all algorithms can complete the calculation in a fixed clock cycle, which can effectively meet the needs of most real-time processing hardware systems.

3) Application of FPGA in Remote Sensing: According to the parallel processing method (as shown in Fig. 33), it can be divided into independent parallel processing of data blocks, internal serial calculations;, overall data processing, parallel internal calculations, and parallel processing of data blocks and parallel internal calculations.

(1) Data parallel, calculation serial: It is suitable for the weak correlation between each data block, which can be operated independently, and there is causality between the operations of each step. Remote sensing images can be used to observe a certain area, usually with large image width and high data volume. Remote sensing data can be used to examine a specific

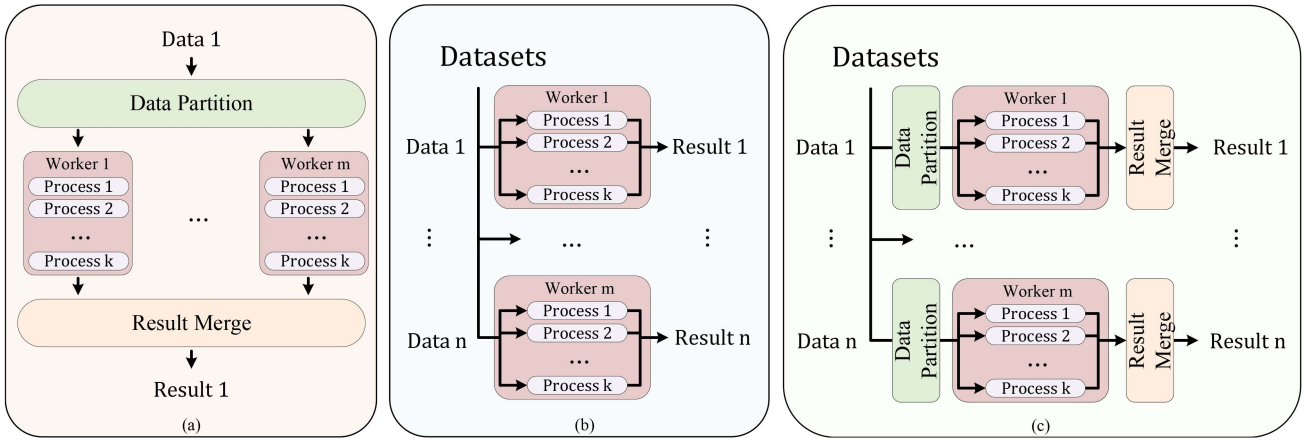


Fig. 33. Parallel processing types of FPGA. (a) Parallel processing of data blocks, serial internal calculations. (b) Overall data processing, parallel internal calculations. (c) Parallel processing of data blocks and parallel internal calculations.

location, often with big image width and high data volume. Data parallelism and computation serialization are popular parallel technologies that are basic and easily scalable. Li et al. [308] deployed the large-scale remote sensing real-time tree canopy detection algorithm on the FPGA and divides the original large-scale scene data into small blocks. It optimizes and adjusts the original method based on a maximum local filter to reduce the utilization of FPGA, reduce idle cycles, and achieve a balance of different resource utilization. Ortiz et al. [309] proposed a parallel endmember extraction method for on-orbit HSIs based on the Fast UNmixing algorithm. This method divides the original HSI into fixed-size subimages and iteratively extracts endmembers from the subimages. This technique can be applied broadly in various computer settings and is very scalable regarding varied processing performance and energy efficiency. In addition, the block-based partition scheme can provide higher fault tolerance, which is suitable for remote sensing satellite environments with high space radiation and vulnerable hardware. González et al. [310] implemented the target detection method based on the orthogonal projection operator ATGP-OSP on FPGA. This article analyzes the orthogonal projection operator, in which the operation of matrix inversion can be highly parallelized by the Gauss–Jordan elimination method. A memory access module is designed in the system, the delay of input and output communication is reduced by prefetching technology, and the operation efficiency is improved. Báscones et al. [311] applied low complexity predictive lossy compression to HSI compression. The image is processed in parallel in blocks, and the iterative optimization process of each spectral channel is highly streamlined. A large number of FIFOs are used, which significantly reduces the use of DSP at the expense of slightly increasing memory, compressing the HSI in real time that satisfies several quality requirements.

(2) Data are processed as a whole, and the internal calculation is parallel: It applies to the relationship between each data block, and each step operation can be performed independently. González et al. [312] proposed a method to implement pixel purity index PPI on FPGA. The calculation of endmember string projection in the PPI method is independent and can be executed simultaneously, so it is very suitable for parallel processing. In addition, the calculation of the dot product in the endmember string projection can also be performed on a pixel-by-pixel basis. That is, data parallelism can be realized. However, since this method requires additional computing resources to process intermediate results, which makes the clock cycle longer, only the process of each endmember string and pixel dot product in the operation process is performed in parallel.

(3) Data are divided into blocks, and operations are parallelized: This method can theoretically utilize computing resources most efficiently. However, data distribution and integration costs must be considered in practical applications. Lei et al. [313] further analyzed the ATGP method based on data parallelism and proposed a vectorization method of operator matrix. The operation of vector projection is calculated in parallel so that the update of the operator in a vector only needs to be in one step. The computation time is significantly reduced. The execution of convolutional neural networks exhibits a high degree of parallelism. Pixels at different locations can be processed in parallel, whereas standard convolutional layers contain multiple filters. But due to hardware limitations, it is impossible to utilize all parallel modes fully. Therefore, the authors in [314], [315], and [316] divided the filters into multiple groups for operation. When computed, the grouped filters are moved along the channel dimension, and intermediate results are stored in the accumulation buffer until the end of the channel gets the convolution result at the current position.

Moreover, the channel convolution operation abovementioned is carried out simultaneously at the different pixels, and the result is the feature map that the current convolution layer has processed. Zhang et al. [317] proposed an independent dual-channel DDR hierarchical storage scheme for storing and reading weight parameters and feature data. The scheme uses ping-pong buffering technology to avoid the conflict between the storage of output feature maps of each layer and the access of input feature maps.

The algorithm processing efficiency on hardware is improved to solve the problem that FPGA storage and bandwidth are

challenging to match in the parallel implementation of the CNN network. It solves the problem of poorly matching FGPA storage and processing bandwidth, which improves the efficiency of arithmetic processing on the hardware. Zhang et al. [318] proposed a three-level memory access architecture, including off-chip memory, on-chip buffer, and local storage. The CNN's parameters are stored in off-chip memory. The convolution processing engine receives picture data from the input buffer. There is no way to set up enough hardware modules to calculate the entire layer at once due to the limitations of hardware logic and memory resources. Each convolutional layer often has several convolution process engines, each with a local memory for storing intermediate results.

VI. Top Ten Open Problems

The application of deep neural networks in remote sensing has become a major trend. However, modern deep learning still has many unsolvable problems. Since humans can deal with all kinds of complex tasks dynamically, brain-inspired algorithms are new research paradigms. With the study of the idea of brain properties, it can effectively make up for the current problems of deep learning. By reviewing the brain properties and current development of the remote sensing image interpretation, we summarize ten future research directions and challenges.

A. How to Design Brain-Inspired Algorithms That Mimic Brain Structure?

The structure of the human brain is hierarchical, sparse, and periodic. At present, the algorithms designed in the field of remote sensing follow a fixed structure. For example, convolution is widely used in image processing tasks to extract features, which realizes a simple simulation of the bottom layer of human brain vision. In addition, the connection of neural networks are dense. In the human brain, however, the underlying visual layers are sparse. The design of a neural network can partly meet the task requirements, but it is still far from the brain structure.

The spiking neural network [319] is a neural network that further simulates the structure of the human brain. It accumulates on neurons through information flow to achieve signal activation and inhibition. At the same time, this structure is closer to the structure of the human brain, thereby realizing sparse connections in information processing. Capsule network [320] also models neurons, representing pose information of features through vectors.

These algorithms that mimic brain structure have been extensively studied in natural data. However, due to the complex characteristics of remote sensing data, brain-inspired algorithms still need further exploration in of remote sensing.

B. Interpretability of Brain-Inspired Remote Sensing Algorithms

Currently, using neural networks to improve the accuracy and efficiency of algorithms is the mainstream method. However, the inner mechanism of neural networks and the choice of parameters have not been well studied. This leads to the fact that the results of the algorithms are not completely credible and reliable in the actual environment. Therefore, the core research of brain-inspired remote sensing is to mimic the cognition, perception, and other abilities related to the human brain to propose the algorithms with high interpretability.

There are very little research works on the interpretability of existing remote sensing algorithms. Hong et al. [321] discussed the development of interpretable hyperspectral artificial intelligence algorithms from the perspective of nonconvex modeling optimization. Many shallow algorithms can be explained by combining them with knowledge of physics. However, the interpretability research of deep algorithms is still a very difficult problem. Guo et al. [322] used the interpretable CNN framework [323] to prune network. This class of methods adds additional losses to the filters in the network to achieve interpretable learning for different classes. In addition, the transformer leverages the attention to build the neural network. It also shows the ability, such as our brains, to successfully handle a disordered flow of information [324]. Furthermore, the attention map is also shown interpretability.

These studies can improve the interpretability of the algorithm to a certain extent. Future remote sensing algorithms still need to combine remote sensing algorithms with brain properties and physical knowledge to improve interpretability.

C. Constructing the Causal Reasoning Ability of Brain-Inspired Remote Sensing Algorithms

The brain is a complex, intelligent structure using knowledge and facts to reason and make conclusions. It makes inferences about things based on perceptions acquired by different organs. These abilities all boil down to causal reasoning. As an emerging theory, causal inference has gradually formed its theoretical system to guide the algorithm design of artificial intelligence.

Currently, in the interpretation of remote sensing data, there are also many researchers trying to add the ability of reasoning to the design of the algorithm. Mou et al. [325] designed a spatial correlation module to construct long-range correlations of objects in the scene. This module can provide relation-enhanced feature representation to improve the accuracy of semantic segmentation. Cao et al. [326] also tried to model and reason about global relational information. This method improves the performance of HSI denoising from the perspective of spatial pixels and channels. The relational reasoning network [327] was proposed in Salient Object Detection in optical remote sensing image. These methods all focus on designing a network structure, constructing the relationship between feature channels, and realizing reasoning about the data. Therefore, brain-inspired algorithms based on reasoning are still in the early stage.

Now deep learning needs to move forward from data-based to knowledge-based. As an essential way to utilize knowledge, causal inference is the focus of brain-inspired algorithm research. The causal inference has three important hierarchies: association, intervention, and counterfactual. These theories formulate the reasoning and decision of human brains. Combining these theories with remote sensing data interpretation

tasks will effectively promote the performance of remote sensing interpretation tasks and improve the interpretability of remote sensing algorithms.

D. Generalization Ability of Remote Sensing Algorithms

Remote sensing data have diverse and complex characteristics, but current algorithms can only handle the task of a single dataset. Even processing the same task, a model cannot be applied to data captured at different ground sample distances (GSD), spectral resolutions, and times. Therefore, it is a waste of resources to train a model to adapt to different data repeatedly. The human brain has strong learning and generalization capabilities. By imitating the learning and memory capabilities of the human brain, we can design dynamic networks for learning and utilizing a variety of data and improve the migration ability of the algorithm in a variety of data.

At the same time, remote sensing image interpretation involves various tasks, such as classification, detection, tracking, and so on. Most algorithms are designed to deal with a single task. However, there is a certain correlation between each task. The brain can use the knowledge of relevant tasks to assist the interpretation of the current task, thereby improving accuracy and speed. For example, the knowledge of the relationship between planes and airports can help us ignore the irrelevant area, achieving rapid localization of the planes. The fusion of these tasks requires a unified brain-inspired remote sensing to perform joint learning of multiple tasks and simulate the mechanism of human information utilization to realize the complementarity of each task.

From another perspective, the remote sensing data collected are always a small set compared with the entire Earth. In the open world, the performance of algorithms is still difficult to estimate and suffers. The human brain has the ability to discriminate unknown types of objects. For unknown objects or categories, it can give the uncertainty of the result so that different strategies can be applied to the uncertain data. This estimation of uncertainty is of great significance in the practical use of remote sensing algorithms. In the natural field, there have been many studies related to open-set data. Such algorithms can identify unknown samples and separate them into unknown classes [328], [329]. Therefore, the algorithm design of the open set is also an important part of the design of brain-inspired remote sensing. It requires the algorithms to face the data from the open world outside the training set, with the ability of self-adaptation, self-induction, self-learning, and the ability to deal with uncertain results. Judgment can predict reasonable results according to the geographical conditions of different regions and regions.

E. How to Implement a Remote Sensing Algorithm With Temporal Memory and Self-Learning?

The observation of remote sensing information is a continuous process. The satellites capture the images in a certain periodicity. By regularly capturing local areas, a series of temporal observations are formed. Existing remote sensing algorithms usually only consider the performance of interpretation in a single image, or obtain the changed area through two images. However, geographic information is in a time-series relationship and continuous change. Only interpreting a single image does not have the ability to predict future changes. Therefore, designing memory capabilities and autonomous learning in the algorithm is the exploration direction of future brain-inspired remote sensing. Based on brain-inspired algorithms that memorize and learn from continuous data, it is possible to predict future situations. According to the prediction results, we can dynamically adjust the capturing frequency of satellites in different areas, realize more intensive observation of high-risk areas, and improve the ability of remote sensing algorithms for disaster early warning.

F. How to Utilize Large-Scale Unlabeled Remote Sensing Data?

We have acquired a large amount of remote sensing data with the increasing number of satellites. However, modern deep learning algorithms rely on massive amounts of labeled data for supervised training. This requires a lot of manpower and resources. In order to utilize a large amount of unlabeled data, semisupervised and self-supervised learning has become a new research trend.

Semisupervised learning combines supervised learning and unsupervised learning. It uses a small amount of labeled data to train a basic model to explore a large amount of unlabeled data. Self-supervised learning is to use the consistency of multiple views of data to train the network. It constructs multiple views of a single target by random augmentation or other strategies and brings considerable performance.

In the field of remote sensing, multisource data naturally constitutes a multiview representation of a target, meeting the need for unsupervised and self-supervised. While using unlabeled data, the interference caused by natural factors, such as cloud occlusion and multisource data matching errors, also needs to be considered.

G. How to Integrate Multimodal Dynamic Data for Interpretation?

In order to monitor the Earth comprehensively, satellites carry sensors with various GSD and imaging methods. The diverse data collected by these sensors bring great challenges to the design of algorithms.

At present, it is mainly to use a data fusion algorithm to improve the performance of the model by using multimodal data, which has been widely studied. Grayscale and HSIs are typical examples of data fusion. Grayscale images have high GSD but only contain a single spectrum. HSIs has high spectral resolution with low GSD. Therefore, these two kinds of data can achieve better complementarity.

Data fusion can effectively improve the performance of the algorithms. With the improvement of shooting technology, dynamic data, such as optical satellite videos and SAR remote sensing videos, have also been developed. In the future, how to

realize the fusion of dynamic multimodal data will be a problem deserving of study.

H. Big Model of Remote Sensing

With the development of deep learning, Big Models have demonstrated an unprecedented ability to understand and create, breaking the limitation that traditional AI can only handle a single task, bringing humans one step closer to the goal of general artificial intelligence. In 2020, OpenAI released a pre-training model GPT-3 [330] with 175 billion parameters. It can not only write articles, answer questions, and translate, but also have the ability to have multiple rounds of dialogue, coding, and mathematical calculations. However, there are still many technical difficulties in realizing the versatility of all modalities and all tasks for Big Model. At the same time, due to the limitation of computing resources, its training and application are quite challenging.

There are less studies on Big Model of remote sensing. Using the reasoning ability of Big Model, it is possible to fully mine various remote sensing data and realize the connection of various tasks. The goal of establishing a Big Model of remote sensing is to solve the problem of fusion and utilization of remote sensing data captured in different modalities, different GSD, and at different times and has the ability to cover a series of remote sensing applications, such as classification, detection, and tracking.

The emergence of Big Model has broken our understanding of algorithms. However, its expensive calculation is not practical at this time. The way to use the Big Model is a crucial problem for remote sensing. In the future, knowledge distillation, model pruning and other technologies can be used to extract the learned understanding ability of Big Model into a small model for specific tasks, thereby improving the learning generalization ability of special models.

I. Security of Remote Sensing Algorithm During Training and Inference

Nowadays, we leverage more and more data to train a large model. The security of remote sensing algorithms is also a worthy issue. The security of remote sensing algorithms is mainly divided into two aspects. On the one hand, it is necessary to use a large number of remote sensing data in different regions for training to improve the generalization ability of models. Due to the particularity of remote sensing data, many remote sensing data contain sensitive information related to countries or companies. Many studies have proved that the network may leak data during the training process [331]. Therefore, it is urgent to study how to design and ensure the security of data during training and realize the federated learning of multiparty training of remote sensing data.

On the other hand, when forward inferring the model, the ability to resist external attacks also needs to be paid attention to. In natural scenarios, many neural network attack studies have shown that fixed neural networks are prone to misjudgment due to minor disturbances. The same situation also exists in remote sensing algorithms. If this attack appears in remote sensing algorithms that automate decision-making, it would have serious implications. Small perturbations do not affect the human brain's judgment of objects. Therefore, remote sensing algorithms need to simulate the memory and associative abilities of the human brain to achieve robustness to attacks.

J. Brain-Inspired Remote Sensing Software and Hardware Systems

As the commercial satellite industry has matured, remote sensing data interpretation have become more than just a need for professionals. Most of the existing remote sensing data platform software requires professionals to design and operate corresponding algorithms for different tasks and data. These limitations restrict the widespread civilian use of remote sensing algorithms. Therefore, remote sensing data interpretation software requires algorithms to cover a variety of tasks, apply to different data and put forward requirements for the ease of use of the software. The remote sensing software system designed based on the abovementioned requirements can provide a comprehensive interpretation of data through simple operations. Users can choose to view tasks, such as object classification, target detection, and interpretation results, of any category according to actual needs.

In terms of hardware systems, on-orbit processing of data can more effectively improve data utilization and save data transmission bandwidth. In this review, we introduce the FPGA and its application in remote sensing. In order to run the algorithm directly on the aircrafts or satellites, we can choose to deploy the algorithms on the space-grade FPGA so as to ensure the stability of the system in extreme environments. However, the computing power and extremely low power consumption of the neuromorphic computing chips are more worth looking forward. For example, TianjicX [332] has realized the experiment of a cat-and-mouse game under the condition of ultra-low power consumption and low delay. The total dynamic power consumption of the chip in the experiment is only 0.6 W. When the remote sensing algorithm is used in the neuromorphic computing chips, the on-orbit satellite can process data in real time with ultra-low power consumption. Only the data with research value will be transmitted back to the ground after preprocessing. This improves the efficiency of data collection and analysis. However, the research on neuromorphic computing chips is still in its infancy, and remote sensing algorithms still need further research to be deployed into neuromorphic computing chips. The neuromorphic computing chips still need further research to improve the stability of the chips in space so as to meet the needs of on-orbit data analysis.

VII. Conclusion

In this survey, we systematically discussed the brain-inspired algorithms in remote sensing. We first summarize the structure and properties of the brain. These properties include six aspects: sparsity, learning, selectivity, directionality, plasticity, and diversity, which can effectively guide readers to think about brain-inspired remote sensing interpretation algorithms from the characteristics. Further, we summarize the data types and

development of five tasks in remote sensing, i.e., object classification, object detection, change detection, object tracking, and 3-D reconstruction. At the same time, the public datasets, the software platforms, and hardware systems are also discussed. The development of brain-inspired algorithms in remote sensing is still not fully explored, and it will help us overcome future challenges.

REFERENCES

[1] M. Mahmud, M. Reba, J. Wei, and N. A. Razak, "Remote sensing entropy to assess the sustainability of rainfall in tropical catchment," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 117, no. 1, 2018, Art. no. 012023.

[2] Y. Li et al., "Deep learning-based object tracking in satellite videos: A comprehensive survey with a new dataset," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 4, pp. 181–212, Dec. 2022.

[3] A. F. Goetz, B. N. Rock, and L. C. Rowan, "Remote sensing for exploration; An overview," *Econ. Geol.*, vol. 78, no. 4, pp. 573–590, 1983.

[4] D. Pasetto et al., "Integration of satellite remote sensing data in ecosystem modelling at local scales: Practices and trends," *Methods Ecol. Evol.*, vol. 9, no. 8, pp. 1810–1821, 2018.

[5] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS J. Photogrammetry Remote Sens.*, vol. 152, pp. 166–177, 2019.

[6] X. Li, X. Yao, and Y. Fang, "Building-A-Nets: Robust building extraction from high-resolution remote sensing images with adversarial networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 10, pp. 3680–3687, Oct. 2018.

[7] M. Reichstein et al., "Deep learning and process understanding for data-driven Earth system science," *Nature*, vol. 566, no. 7743, pp. 195–204, 2019.

[8] L. Zhang and L. Zhang, "Artificial intelligence for remote sensing data analysis: A review of challenges and opportunities," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 2, pp. 270–294, Jun. 2022.

[9] J. Fan, L. Fang, J. Wu, Y. Guo, and Q. Dai, "From brain science to artificial intelligence," *Engineering*, vol. 6, no. 3, pp. 248–252, 2020.

[10] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bull. Math. Biophys.*, vol. 5, no. 4, pp. 115–133, 1943.

[11] D. O. Hebb, *The Organization of Behavior: A Neuropsychological Theory*. New York, NY, USA: Psychol. Press, 2005.

[12] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychol. Rev.*, vol. 65, no. 6, pp. 386–408, 1958.

[13] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Nat. Acad. Sci. USA*, vol. 79, no. 8, pp. 2554–2558, 1982.

[14] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[15] D. Hassabis, D. Kumaran, C. Summerfield, and M. Botvinick, "Neuroscience-inspired artificial intelligence," *Neuron*, vol. 95, no. 2, pp. 245–258, 2017.

[16] S. Yang, X. Hao, B. Deng, X. Wei, H. Li, and J. Wang, "A survey of brain-inspired artificial intelligence and its engineering," *Life Res.*, vol. 1, no. 1, pp. 23–29, 2018.

[17] N. Strisciuglio and N. Petkov, "Brain-inspired algorithms for processing of visual data," in *Proc. Int. Workshop Brain-Inspired Comput.*, 2019, pp. 105–115.

[18] O. Simeone et al., "Learning algorithms and signal processing for brain-inspired computing [From the guest editors]," *IEEE Signal Process. Mag.*, vol. 36, no. 6, pp. 12–15, Nov. 2019.

[19] L. Jiao, R. Shang, F. Liu, and W. Zhang, *Brain and Nature-Inspired Learning, Computation and Recognition*. Cambridge, MA, USA: Elsevier, 2020.

[20] J. Tianyuan, F. Chaoqiong, W. Lina, W. Liya, and W. Xia, "Brain-inspired artificial intelligence: Advances and applications," *Chin. Spaceflight*, vol. 22, no. 1, pp. 12–19, 2021.

[21] G. Fleming, "Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. (Brain, vol. lx, p. 389, Dec. 1937.) Penfield, W., and Boldrey, E.," *J. Ment. Sci.*, vol. 84, no. 352, pp. 868–868, 1938.

[22] E. Genç et al., "Diffusion markers of dendritic density and arborization in gray matter predict differences in intelligence," *Nature Commun.*, vol. 9, no. 1, pp. 1–11, 2018.

[23] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[24] D. Huber et al., "Sparse optical microstimulation in barrel cortex drives learned behaviour in freely moving mice," *Nature*, vol. 451, no. 7174, pp. 61–64, 2008.

[25] A. R. Houweling and M. Brecht, "Behavioural report of single neuron stimulation in somatosensory cortex," *Nature*, vol. 451, no. 7174, pp. 65–68, 2008.

[26] J. B. Tenenbaum, C. Kemp, T. L. Griffiths, and N. D. Goodman, "How to grow a mind: Statistics, structure, and abstraction," *Science*, vol. 331, no. 6022, pp. 1279–1285, 2011.

[27] S. A. Josselyn and S. Tonegawa, "Memory engrams: Recalling the past and imagining the future," *Science*, vol. 367, no. 6473, 2020, Art. no. eaaw4325.

[28] P. R. Roelfsema, "Attention–voluntary control of brain cells," *Science*, vol. 332, no. 6037, pp. 1512–1513, 2011.

[29] A. Finkelstein, D. Derdikman, A. Rubin, J. N. Foerster, L. Las, and N. Ulanovsky, "Three-dimensional head-direction coding in the bat brain," *Nature*, vol. 517, no. 7533, pp. 159–164, 2015.

[30] S. Ikeda, "Functional recovery on stroke, stroke model and rehabilitation (progress in regenerative medicine)," *Japanese J. Psychosomatic Med.*, vol. 53, no. 8, pp. 742–747, 2013.

[31] J. Berg et al., "Human neocortical expansion involves glutamatergic neuron diversification," *Nature*, vol. 598, no. 7879, pp. 151–158, 2021.

[32] Z. Yao et al., "A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex," *Nature*, vol. 598, no. 7879, pp. 103–110, 2021.

[33] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.

[34] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7794–7803.

[35] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6000–6010.

[36] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[37] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[38] L. Jiao, S. Yang, F. Liu, and B. Hou, "Review and prospect of compressed perception," *Acta Electronica Sinica*, vol. 39, no. 7, 2011, Art. no. 1651.

[39] D. H. Hubel and T. N. Wiesel, "Receptive fields of single neurones in the cat's striate cortex," *J. Physiol.*, vol. 148, no. 3, pp. 574–591, 1959.

[40] D. J. Willshaw, O. P. Buneman, and H. C. Longuet-Higgins, "Non-holographic associative memory," *Nature*, vol. 222, no. 5197, pp. 960–962, 1969.

[41] L. Jiao, Y. Yang, F. Liu, S. Yang, and B. Hou, "The new generation brain-inspired sparse learning: A comprehensive survey," *IEEE Trans. Artif. Intell.*, vol. 3, no. 6, pp. 887–907, Dec. 2022.

[42] J. Bourgain, S. Dilworth, K. Ford, S. Konyagin, and D. Kutzarova, "Explicit constructions of rip matrices and related problems," *Duke Math. J.*, vol. 159, no. 1, pp. 145–185, 2011.

[43] L. Wang, Y. Feng, Y. Gao, Z. Wang, and M. He, "Compressed sensing reconstruction of hyperspectral images based on spectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1266–1284, Apr. 2018.

[44] J. Xue, Y. Zhao, W. Liao, and J. C.-W. Chan, "Nonlocal tensor sparse representation and low-rank regularization for hyperspectral image compressive sensing reconstruction," *Remote Sens.*, vol. 11, no. 2, 2019, Art. no. 193.

[45] Y. Chen, T.-Z. Huang, W. He, N. Yokoya, and X.-L. Zhao, "Hyperspectral image compressive sensing reconstruction using subspace-based nonlocal tensor ring decomposition," *IEEE Trans. Image Process.*, vol. 29, pp. 6813–6828, 2020.

[46] M. Ghahremani, Y. Liu, P. Yuen, and A. Behera, "Remote sensing image fusion via compressive sensing," *ISPRS J. Photogrammetry Remote Sens.*, vol. 152, pp. 34–48, 2019.

[47] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat," *J. Neuriophysiol.*, vol. 28, no. 2, pp. 229–289, 1965.

[48] L. Jiao, H. Biao, S. Wang, and F. Liu, "Image multiscale geometric analysis: Theory and applications beyond wavelets," *Xi'an Univ. Electron. Sci. Technol.*, vol. 1, pp. 124–132, 2008.

[49] L. Jiao and S. Tan, "Development and prospect of image multiscale geometric analysis," *Acta Electonica Sinica*, vol. 31, no. S1, 2003, Art. no. 1975.

[50] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.

[51] E. L. Pennec and S. Mallat, "Non linear image approximation with bandelets," CMAP/École Polytechnique, Palaiseau, France, Tech. Rep., 2003.

[52] D. L. Donoho, "Wedgelets: Nearly minimax estimation of edges," *Ann. Statist.*, vol. 27, no. 3, pp. 859–897, 1999.

[53] E. J. Candes, "Ridgelets: Theory and applications," Ph.D. dissertation, Dept. Stat., Stanford Univ., Stanford, CA, USA, 1998.

[54] E. J. Candes and D. L. Donoho, "Curvelets: A surprisingly effective nonadaptive representation for objects with edges," Dept. Statist., Stanford Univ., CA, USA, Tech. Rep. GEN_1999-28, 2000.

[55] M. N. Do and M. Vetterli, "Contourlets: A directional multiresolution image representation," in *Proc. Int. Conf. Image Process.*, vol. 1, 2002, pp. 357–360.

[56] M. Liu, L. Jiao, X. Liu, L. Li, F. Liu, and S. Yang, "C-CNN: Contourlet convolutional neural networks," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 32, no. 6, pp. 2636–2649, Jun. 2021.

[57] W.-T. Chen, C.-C. Tsai, H.-Y. Fang, I. Chen, J.-J. Ding, and S.-Y. Kuo, "ContourletNet: A generalized rain removal architecture using multi-direction hierarchical representation," in *Proc. Brit. Mach. Vis. Conf.*, 2021.

[58] Z. Shen, C. Lin, L. Nie, K. Liao, and Y. Zhao, "Neural contourlet network for monocular 360 depth estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 12, pp. 8574–8585, Dec. 2022.

[59] W. Zhang, L. Jiao, F. Liu, S. Yang, and J. Liu, "Adaptive contourlet fusion clustering for SAR image change detection," *IEEE Trans. Image Process.*, vol. 31, pp. 2295–2308, 2022.

[60] L. Li, L. Ma, L. Jiao, F. Liu, Q. Sun, and J. Zhao, "Complex contourlet-CNN for polarimetric SAR image classification," *Pattern Recognit.*, vol. 100, 2020, Art. no. 107110.

[61] J. Gao, L. Jiao, F. Liu, S. Yang, B. Hou, and X. Liu, "Multiscale curvelet scattering network," *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1–15, 2021, doi: 10.1109/TNNLS.2021.3118221.

[62] F. Wang and D. M. J. Tax, "Survey on the attention based RNN model and its applications in computer vision," 2016, *arXiv:1601.06823*.

[63] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 2017–2025.

[64] J. M. Haut, M. E. Paoletti, J. Plaza, A. Plaza, and J. Li, "Visual attention-driven hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 8065–8080, Oct. 2019.

[65] T. Zhang et al., "Semantic attention and scale complementary network for instance segmentation in remote sensing images," *IEEE Trans. Cybern.*, vol. 52, no. 10, pp. 10999–11013, Oct. 2022.

[66] M. Wang, Z. Wang, C. Yang, S. Yang, and Y. Gao, "Polarimetric SAR data classification via reinforcement learning," *IEEE Access*, vol. 7, pp. 137629–137637, 2019.

[67] Y. Li, "Deep reinforcement learning: An overview," 2017, *arXiv: 1701.07274*.

[68] V. Mnih et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[69] T. P. Lillicrap et al., "Continuous control with deep reinforcement learning," 2015, *arXiv:1509.02971*.

[70] K. Huang, W. Nie, and N. Luo, "Fully polarized SAR imagery classification based on deep reinforcement learning method using multiple polarimetric features," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 10, pp. 3719–3730, Oct. 2019.

[71] K. Fu et al., "A ship rotation detection model in remote sensing images based on feature fusion pyramid network and deep reinforcement learning," *Remote Sens.*, vol. 10, no. 12, 2018, Art. no. 1922.

[72] F. Zhuang et al., "A comprehensive survey on transfer learning," *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021.

[73] M. Rostami, S. Kolouri, E. Eaton, and K. Kim, "Deep transfer learning for few-shot SAR image classification," *Remote Sens.*, vol. 11, no. 11, 2019, Art. no. 1374.

[74] J. Peng, Y. Huang, W. Sun, N. Chen, Y. Ning, and Q. Du, "Domain adaptation in remote sensing image classification: A survey," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 9842–9859, Nov. 2022, doi: 10.1109/JSTARS.2022.3220875.

[75] M. Xie, N. Jean, M. Burke, D. Lobell, and S. Ermon, "Transfer learning from deep features for remote sensing and poverty mapping," in *Proc. 30th AAAI Conf. Artif. Intell.*, 2016, pp. 3929–3935.

[76] Z. Chen, T. Zhang, and C. Ouyang, "End-to-end airplane detection using transfer learning in remote sensing images," *Remote Sens.*, vol. 10, no. 1, 2018, Art. no. 139.

[77] B. Demir, F. Bovolo, and L. Bruzzone, "Updating land-cover maps by classification of image time series: A novel change-detection-driven transfer learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 300–312, Jan. 2012.

[78] Z. Sun et al., "A review of Earth artificial intelligence," *Comput. Geosci.*, vol. 159, 2022, Art. no. 105034.

[79] B. Zhang et al., "Progress and challenges in intelligent remote sensing satellite systems," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1814–1822, Feb. 2022, doi: 10.1109/JSTARS.2022.3148139.

[80] G. Grasso, D. Zane, and R. Dragone, "Field and remote sensors for environmental health and food safety diagnostics: An open challenge," *Biosensors*, vol, 12, no. 5, 2022, Art. no. 285.

[81] M. Awais et al., "UAV-based remote sensing in plant stress imagine using high-resolution thermal sensor for digital agriculture practices: A meta-review," *Int. J. Environ. Sci. Technol.*, vol. 20, pp. 1135–1152, 2022.

[82] R. W. Albuquerque et al., "Mapping key indicators of forest restoration in the amazon using a low-cost drone and artificial intelligence," *Remote Sens.*, vol. 14, no. 4, 2022, Art. no. 830.

[83] Y. Liu, D. Zeng, Y. Hu, and S. Zhou, "Visualization analysis of satellite intelligence based on scientific knowledge graph," in *Proc. Int. Conf. Electron. Inf. Eng., Big Data, Comput. Technol.*, 2022, vol. 12256, pp. 324–330.

[84] C. Mücher, S. Los, G. Franke, and C. Kamphuis, "Detection, identification and posture recognition of cattle with satellites, aerial photography and UAVs using deep learning techniques," *Int. J. Remote Sens.*, vol. 43, no. 7, pp. 2377–2392, 2022.

[85] A. Dakir, F. Barramou, and O. B. Alami, "Opportunities for artificial intelligence in precision agriculture using satellite remote sensing," in *Geospatial Intelligence*. Cham, Switzerland: Springer, 2022, pp. 107–117.

[86] R. Wang, K. Wu, Q. He, Y. He, Y. Gu, and S. Wu, "A novel method of monitoring surface subsidence law based on probability integral model combined with active and passive remote sensing data," *Remote Sens.*, vol. 14, no. 2, 2022, Art. no. 299.

[87] 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, 2022, Art. no. 871.

[88] X. Zhou et al., "Edge-guided recurrent positioning network for salient object detection in optical remote sensing images," *IEEE Trans. Cybern.*, vol. 53, no. 1, pp. 539–552, Jan. 2023.

[89] A. H. Oveis, E. Giusti, S. Ghio, and M. Martorella, "Extended open-max approach for the classification of radar images with a rejection option," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 1, pp. 196–208, Feb. 2023.

[90] A. Xiao, J. Huang, D. Guan, F. Zhan, and S. Lu, "Transfer learning from synthetic to real LiDAR point cloud for semantic segmentation," in *Proc. AAAI Conf. Artif. Intell.*, 2022, pp. 2795–2803.

[91] X. Li et al., "A collaborative learning tracking network for remote sensing videos," *IEEE Trans. Cybern.*, vol. 53, no. 3, pp. 1954–1967, Mar. 2023.

[92] C. Hu, L. Qi, Y. Xie, S. Zhang, and B. B. Barnes, "Spectral characteristics of sea snot reflectance observed from satellites: Implications for remote sensing of marine debris," *Remote Sens. Environ.*, vol. 269, 2022, Art. no. 112842.

[93] M. Wang, G. Xie, Z. Zhang, Y. Wang, S. Xiang, and Y. Pi, "Smoothing filter-based panchromatic spectral decomposition for multispectral and hyperspectral image pansharpening," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 3612–3625, Apr. 2022.

[94] L. Liu et al., "CASR-Net: A color-aware super-resolution network for panchromatic image," *Eng. Appl. Artif. Intell.*, vol. 114, 2022, Art. no. 105084.

[95] A. Gerace et al., "In-flight performance of the multi-band uncooled radiometer instrument (MURI) thermal sensor," *Remote Sens. Environ.*, vol. 279, 2022, Art. no. 113086.

[96] L. Lu, Z. Gong, Y. Liang, and S. Liang, "Retrieval of chlorophyll-a concentrations of class II water bodies of inland lakes and reservoirs based on ZY1-0D satellite hyperspectral data," *Remote Sens.*, vol. 14, no. 8, 2022, Art. no. 1842.

[97] J. Yin, C. Qi, W. Huang, Q. Chen, and J. Qu, "Multibranch 3D-dense attention network for hyperspectral image classification," *IEEE Access*, vol. 10, pp. 71886–71898, 2022.

[98] M. Wang, L. Ying, C. Guan, and H. Li, "Feature recognition and classification based on hyperspectral data mining," *Geosci. Remote Sens.*, vol. 5, no. 1, pp. 11–15, 2022.

[99] H. Fu et al., "A novel band selection and spatial noise reduction method for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022.

[100] M. Kurum et al., "A UAS-based RF testbed for water utilization in agroecosystems," in *Proc. Auton. Air Ground Sens. Syst. Agricultural Optim. Phenotyping VI*, 2021, vol. 11747, pp. 74–88.

[101] S. Paul and M. J. Akhtar, "Novel metasurface lens-based RF sensor structure for SAR microwave imaging of layered media," *IEEE Sensors J.*, vol. 21, no. 16, pp. 17827–17837, Aug. 2021.

[102] X. Tuo, Y. Zhang, Y. Huang, and J. Yang, "A fast sparse Azimuth super-resolution imaging method of real aperture radar based on iterative reweighted least squares with linear sketching," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2928–2941, Feb. 2021.

[103] D. Mao et al., "Angular superresolution of real aperture radar using online detect-before-reconstruct framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, Dec. 2021.

[104] J. Yang, *Bistatic Synthetic Aperture Radar*. Amsterdam, The Netherlands: Elsevier, 2022.

[105] A. Belous, "Antennas and antenna devices for radar location and radio communication," in *Handbook of Microwave and Radar Engineering*. Cham, Switzerland: Springer, 2021, pp. 167–333.

[106] S. V. Kalinin et al., "Machine learning in scanning transmission electron microscopy," *Nature Rev. Methods Primers*, vol. 2, no. 1, pp. 1–28, 2022.

[107] C.-C. Chen and H. C. Andrews, "Target-motion-induced radar imaging," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-16, no. 1, pp. 2–14, Jan. 1980.

[108] T. T. Mar and S. S. Y. Mon, "Pulse compression method for radar signal processing," *Int. J. Sci. Eng. Appl.*, vol. 3, pp. 31–35, 2014.

[109] Y. K. Chan and V. Koo, "An introduction to synthetic aperture radar (SAR)," *Prog. Electromagn. Res. B*, vol. 2, pp. 27–60, 2008.

[110] A. Singh, G. K. Meena, S. Kumar, and K. Gaurav, "Evaluation of the penetration depth of L-and S-band (NISAR mission) microwave SAR signals into ground," in *Proc. URSI Asia-Pacific Radio Sci. Conf.*, pp. 1 2019.

[111] J. Duan, L. Zhang, M. Xing, Y. Wu, and M. Wu, "Polarimetric target decomposition based on attributed scattering center model for synthetic aperture radar targets," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 12, pp. 2095–2099, Dec. 2014.

[112] X. Xu and R. M. Narayanan, "FOPEN SAR imaging using UWB step-frequency and random noise waveforms," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 37, no. 4, pp. 1287–1300, Oct. 2001.

[113] R. A. Williamson and P. R. Nickens, *Science and Technology in Historic Preservation*. New york, NY, USA: Springer, 2000.

[114] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 1, pp. 6–43, Mar. 2013.

[115] H.S. Srivastava, "Interaction of multifrequency multi-polarized DLR ESAR data with various targets: A case study with C, L and P bands acquired at all the four linear (VV, VH, HH & HV) polarizations," in *Proc. JEP-MW Conf.*, 2007, pp. 15–16.

[116] J. Hu, D. Hong, and X. X. Zhu, "MIMA: MAPPER-induced manifold alignment for semi-supervised fusion of optical image and polarimetric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9025–9040, Nov. 2019.

[117] S. E. Reutebuch, H.-E. Andersen, and R. J. McGaughey, "Light detection and ranging (LIDAR): An emerging tool for multiple resource inventory," *J. Forestry*, vol. 103, no. 6, pp. 286–292, 2005.

[118] K. Wang, T. Wang, and X. Liu, "A review: Individual tree species classification using integrated airborne LiDAR and optical imagery with a focus on the urban environment," *Forests*, vol. 10, no. 1, 2019, Art. no. 1. [Online]. Available: https://www.mdpi.com/1999-4907/10/1/1

[119] A. Wehr and U. Lohr, "Airborne laser scanning: An introduction and overview," *ISPRS J. Photogrammetry Remote Sens.*, vol. 54, no. 2/3, pp. 68–82, 1999.

[120] U. Onyekpe, V. Palade, and S. Kanarachos, "Learning to localise automated vehicles in challenging environments using inertial navigation systems (INS)," *Appl. Sci.*, vol. 11, no. 3, 2021, Art. no. 1270.

[121] V. B. Semwal, N. Gaud, P. Lalwani, V. Bijalwan, and A. K. Alok, "Pattern identification of different human joints for different human walking styles using inertial measurement unit (IMU) sensor," *Artif. Intell. Rev.*, vol. 55, no. 2, pp. 1149–1169, 2022.

[122] M. Błaszczyk, M. Laska, A. Sivertsen, and S. D. Jawak, "Combined use of aerial photogrammetry and terrestrial laser scanning for detecting geomorphological changes in Hornsund, Svalbard," *Remote Sens.*, vol. 14, no. 3, 2022, Art. no. 601.

[123] W. Sun, J. Wang, F. Jin, and Y. Yang, "A quality improvement method for 3D laser slam point clouds based on geometric primitives of the scan scene," *Int. J. Remote Sens.*, vol. 42, no. 1, pp. 378–388, 2021.

[124] A. Li, X. Liu, J. Sun, and Z. Lu, "Risley-prism-based multi-beam scanning LiDAR for high-resolution three-dimensional imaging," *Opt. Lasers Eng.*, vol. 150, 2022, Art. no. 106836.

[125] W. Wagner, M. Hollaus, C. Briese, and V. Ducic, "3D vegetation mapping using small-footprint full-waveform airborne laser scanners," *Int. J. Remote Sens.*, vol. 29, no. 5, pp. 1433–1452, 2008.

[126] F. Bi, M. Lei, Y. Wang, and D. Huang, "Remote sensing target tracking in UAV aerial video based on saliency enhanced MDnet," *IEEE Access*, vol. 7, pp. 76731–76740, 2019.

[127] T. Yang et al., "Small moving vehicle detection in a satellite video of an urban area," *Sensors*, vol. 16, no. 9, 2016, Art. no. 1528.

[128] N. Kerle, L. L. Janssen, and G. C. Huurneman, *Principles of Remote Sensing* (ITC Educational Textbook Series). Enschede, The Netherlands: ITC, 2004.

[129] H. Ye, H. Guo, G. Liu, and Y. Ren, "Observation scope and spatial coverage analysis for earth observation from a moon-based platform," *Int. J. Remote Sens.*, vol. 39, no. 18, pp. 5809–5833, 2018.

[130] Y. Wang, T. Wang, G. Zhang, Q. Cheng, and J.-Q. Wu, "Small target tracking in satellite videos using background compensation," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7010–7021, Oct. 2020.

[131] D. Zhou et al., "Satellite remote sensing of surface urban heat islands: Progress, challenges, and perspectives," *Remote Sens.*, vol. 11, no. 1, 2018, Art. no. 48.

[132] Z. Yang et al., "UAV remote sensing applications in marine monitoring: Knowledge visualization and review," *Sci. Total Environ.*, vol. 838, 2022, Art. no. 155939.

[133] H. Zhang, L. Wang, T. Tian, and J. Yin, "A review of unmanned aerial vehicle low-altitude remote sensing (UAV-LARS) use in agricultural monitoring in China," *Remote Sens.*, vol. 13, no. 6, 2021, Art. no. 1221.

[134] R. Clothier and R. Walker, "Determination and evaluation of UAV safety objectives," in *Proc. 21st Int. Conf. Unmanned Air Veh. Syst.*, 2006, pp. 18.1–18.16.

[135] K. Kanistras, G. Martins, M. J. Rutherford, and K. P. Valavanis, "A survey of unmanned aerial vehicles (UAVs) for traffic monitoring," in *Proc. IEEE Int. Conf. Unmanned Aircr. Syst.*, 2013, pp. 221–234.

[136] X. Yang, L. Liu, N. Wang, and X. Gao, "A two-stream dynamic pyramid representation model for video-based person re-identification," *IEEE Trans. Image Process.*, vol. 30, pp. 6266–6276, Jul. 2021.

[137] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

[138] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2015, pp. 234–241.

[139] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[140] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.

[141] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.

[142] C. Zhang, "An object-based convolutional neural network (OCNN) for urban land use classification," *Remote Sens. Environ.*, vol. 216, pp. 57–70, 2018.

[143] S. Timilsina, J. Aryal, and J. B. Kirkpatrick, "Mapping urban tree cover changes using object-based convolution neural network (OB-CNN)," *Remote Sens.*, vol. 12, no. 18, 2020, Art. no. 3017.

[144] C. Zhang, P. Yue, D. Tapete, B. Shangguan, M. Wang, and Z. Wu, "A multi-level context-guided classification method with object-based convolutional neural network for land cover classification using very high resolution remote sensing images," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 88, 2020, Art. no. 102086.

[145] M. Papadomanolaki, M. Vakalopoulou, and K. Karantzalos, "A novel object-based deep learning framework for semantic segmentation of very high-resolution remote sensing data: Comparison with convolutional and fully convolutional networks," *Remote Sens.*, vol. 11, no. 6, 2019, Art. no. 684.

[146] C. Peng, K. Zhang, Y. Ma, and J. Ma, "Cross fusion net: A fast semantic segmentation network for small-scale semantic information capturing in aerial scenes," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2021.

[147] K. Heidler, L. Mou, C. Baumhoer, A. Dietz, and X. X. Zhu, "HED-UNet: Combined segmentation and edge detection for monitoring the antarctic coastline," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2021.

[148] Q. Liu, M. Kampffmeyer, R. Jenssen, and A.-B. Salberg, "Dense dilated convolutions' merging network for land cover classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 6309–6320, Sep. 2020.

[149] C. Peng, Y. Li, L. Jiao, Y. Chen, and R. Shang, "Densely based multi-scale and multi-modal fully convolutional networks for high-resolution remote-sensing image semantic segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 8, pp. 2612–2626, Aug. 2019.

[150] R. Shang, J. Zhang, L. Jiao, Y. Li, N. Marturi, and R. Stolkin, "Multi-scale adaptive feature fusion network for semantic segmentation in remote sensing images," *Remote Sens.*, vol. 12, no. 5, 2020, Art. no. 872.

[151] J. Wang, S. Guo, R. Huang, L. Li, X. Zhang, and L. Jiao, "Dual-channel capsule generation adversarial network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, Jan. 2022.

[152] J. Bai et al., "Hyperspectral image classification based on multibranch attention transformer networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, Aug. 2022.

[153] L. Wang et al., "UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 190, pp. 196–214, 2022.

[154] W. He, W. Huang, S. Liao, Z. Xu, and J. Yan, "CSiT: A multiscale vision transformer for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 9266–9277, Oct. 2022.

[155] X. Tang et al., "An unsupervised remote sensing change detection method based on multiscale graph convolutional network and metric learning," *IEEE Trans. Geosci. Remote Sens.*, Sep. 2021.

[156] Y. Afaq and A. Manocha, "Analysis on change detection techniques for remote sensing applications: A review," *Ecological Informat.*, vol. 63, 2021, Art. no. 101310.

[157] R. Weismiller, S. Kristof, D. Scholz, P. Anuta, and S. Momin, "Change detection in coastal zone environments," *Photogrammetric Eng. Remote Sens.*, vol. 43, no. 12, pp. 1533–1539, 1977.

[158] I. R. Hegazy and M. R. Kaloop, "Monitoring urban growth and land use change detection with GIS and remote sensing techniques in Daqahlia governorate Egypt," *Int. J. Sustain. Built Environ.*, vol. 4, no. 1, pp. 117–124, 2015.

[159] Z. Zheng, Y. Zhong, J. Wang, A. Ma, and L. Zhang, "Building damage assessment for rapid disaster response with a deep object-based semantic change detection framework: From natural disasters to man-made disasters," *Remote Sens. Environ.*, vol. 265, 2021, Art. no. 112636.

[160] S. Mishra, P. Shrivastava, and P. Dhurvey, "Change detection techniques in remote sensing: A review," *Int. J. Wirel. Mobile Commun. Ind. Syst.*, vol. 4, no. 1, pp. 1–8, 2017.

[161] H. Jiang et al., "A survey on deep learning-based change detection from high-resolution remote sensing images," *Remote Sens.*, vol. 14, no. 7, 2022, Art. no. 1552.

[162] N. Quarmby and J. Cushnie, "Monitoring urban land cover changes at the urban fringe from SPOT HRV imagery in south-east England," *Int. J. Remote Sens.*, vol. 10, no. 6, pp. 953–963, 1989.

[163] R. Kumar, S. Nandy, R. Agarwal, and S. Kushwaha, "Forest cover dynamics analysis and prediction modeling using logistic regression model," *Ecological Indicators*, vol. 45, pp. 444–455, 2014.

[164] M. Hussain, D. Chen, A. Cheng, H. Wei, and D. Stanley, "Change detection from remotely sensed images: From pixel-based to object-based approaches," *ISPRS J. Photogrammetry Remote Sens.*, vol. 80, pp. 91–106, 2013.

[165] J. Richards, "Thematic mapping from multitemporal image data using the principal components transformation," *Remote Sens. Environ.*, vol. 16, no. 1, pp. 35–46, 1984.

[166] M. Wang, H. Zhang, W. Sun, S. Li, F. Wang, and G. Yang, "A coarse-to-fine deep learning based land use change detection method for high-resolution remote sensing images," *Remote Sens.*, vol. 12, no. 12, 2020, Art. no. 1933.

[167] A. Asokan and J. Anitha, "Change detection techniques for remote sensing applications: A survey," *Earth Sci. Informat.*, vol. 12, no. 2, pp. 143–160, 2019.

[168] T. Blaschke and J. Strobl, "What's wrong with pixels? Some recent developments interfacing remote sensing and GIS," *Zeitschrift für Geoinformationssysteme*, vol. 14, pp. 12–17, 2001.

[169] Q. Wang, Z. Yuan, Q. Du, and X. Li, "GETNET: A general end-to-end 2-D CNN framework for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 3–13, Jan. 2018.

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

[171] Z. Zheng, Y. Wan, Y. Zhang, S. Xiang, D. Peng, and B. Zhang, "CLNet: Cross-layer convolutional neural network for change detection in optical remote sensing imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 175, pp. 247–267, 2021.

[172] H. Chen, C. Wu, B. Du, and L. Zhang, "Deep Siamese multi-scale convolutional network for change detection in multi-temporal VHR images," in *Proc. IEEE 10th Int. Workshop Anal. Multitemporal Remote Sens. Images*, 2019, pp. 1–4.

[173] H. Chen, Z. Qi, and Z. Shi, "Remote sensing image change detection with transformers," *IEEE Trans. Geosci. Remote Sens.*, Jul. 2021.

[174] C. Zhang, L. Wang, S. Cheng, and Y. Li, "SwinSUNet: Pure transformer network for remote sensing image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, Mar. 2022.

[175] S. Ji, Y. Shen, M. Lu, and Y. Zhang, "Building instance change detection from large-scale aerial images using convolutional neural networks and simulated samples," *Remote Sens.*, vol. 11, no. 11, 2019, Art. no. 1343.

[176] G. Chen, G. J. Hay, L. M. Carvalho, and M. A. Wulder, "Object-based change detection," *Int. J. Remote Sens.*, vol. 33, no. 14, pp. 4434–4457, 2012.

[177] H. Zhang, M. Lin, G. Yang, and L. Zhang, "ESCNet: An end-to-end superpixel-enhanced change detection network for very-high-resolution remote sensing images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 1, pp. 28–42, Jan. 2023.

[178] X. Wang, S. Liu, P. Du, H. Liang, J. Xia, and Y. Li, "Object-based change detection in urban areas from high spatial resolution images based on multiple features and ensemble learning," *Remote Sens.*, vol. 10, no. 2, 2018, Art. no. 276.

[179] T. Zhan, M. Gong, X. Jiang, and M. Zhang, "Unsupervised scale-driven change detection with deep spatial–spectral features for VHR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5653–5665, Aug. 2020.

[180] X. Zhang, X. Tan, G. Chen, K. Zhu, P. Liao, and T. Wang, "Object-based classification framework of remote sensing images with graph convolutional networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, Apr. 2021.

[181] L. Zhang, X. Hu, M. Zhang, Z. Shu, and H. Zhou, "Object-level change detection with a dual correlation attention-guided detector," *ISPRS J. Photogrammetry Remote Sens.*, vol. 177, pp. 147–160, 2021.

[182] P. Han, C. Ma, Q. Li, P. Leng, S. Bu, and K. Li, "Aerial image change detection using dual regions of interest networks," *Neurocomputing*, vol. 349, pp. 190–201, 2019.

[183] I. Priyanto, C. A. Hartanto, and A. M. Arymurthy, "Change detection of floating net cages quantities utilizing faster R-CNN," in *Proc. IEEE 3rd Int. Conf. Comput. Informat. Eng.*, 2020, pp. 140–145.

[184] J. Lu, J. Li, G. Chen, L. Zhao, B. Xiong, and G. Kuang, "Improving pixel-based change detection accuracy using an object-based approach in multitemporal SAR flood images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 7, pp. 3486–3496, Jul. 2015.

[185] Y. Han, A. Javed, S. Jung, and S. Liu, "Object-based change detection of very high resolution images by fusing pixel-based change detection results using weighted Dempster–Shafer theory," *Remote Sens.*, vol. 12, no. 6, 2020, Art. no. 983.

[186] C. Wu, L. Zhang, and L. Zhang, "A scene change detection framework for multi-temporal very high resolution remote sensing images," *Signal Process.*, vol. 124, pp. 184–197, 2016.

[187] C. Wu, L. Zhang, and B. Du, "Kernel slow feature analysis for scene change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 2367–2384, Apr. 2017.

[188] B. Du, Y. Wang, C. Wu, and L. Zhang, "Unsupervised scene change detection via latent Dirichlet allocation and multivariate alteration detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 4676–4689, Dec. 2018.

[189] 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, 2021, Art. no. 102465.

[190] Y. Wang, B. Du, L. Ru, C. Wu, and H. Luo, "Scene change detection via deep convolution canonical correlation analysis neural network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 198–201.

[191] C. Ke, "Military object detection using multiple information extracted from hyperspectral imagery," in *Proc. Int. Conf. Prog. Informat. Comput.*, 2017, pp. 124–128.

[192] Z. Deng, H. Sun, S. Zhou, J. Zhao, L. Lei, and H. Zou, "Multi-scale object detection in remote sensing imagery with convolutional neural networks," *ISPRS J. Photogrammetry Remote Sens.*, vol. 145, pp. 3–22, 2018.

[193] Y. Dong, F. Chen, S. Han, and H. Liu, "Ship object detection of remote sensing image based on visual attention," *Remote Sens.*, vol. 13, no. 16, 2021, Art. no. 3192.

[194] H. Li, "An overview on remote sensing image classification methods with a focus on support vector machine," in *Proc. IEEE Int. Conf. Signal Process. Mach. Learn.*, 2021, pp. 50–56.

[195] L. Liu et al., "Deep learning for generic object detection: A survey," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 261–318, 2020.

[196] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proc. IEEE*, vol. 111, no. 3, 2023, pp. 257–276, doi: 10.1109/JPROC.2023.3238524.

[197] K. Li, G. Cheng, S. Bu, and X. You, "Rotation-insensitive and context-augmented object detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2337–2348, Apr. 2018.

[198] J. Bai et al., "Object detection in large-scale remote-sensing images based on time-frequency analysis and feature optimization," *IEEE Trans. Geosci. Remote Sens.*, Oct. 2021.

[199] Z. Zou and Z. Shi, "Ship detection in spaceborne optical image with SVD networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 5832–5845, Oct. 2016.

[200] Y. Xu et al., "Gliding vertex on the horizontal bounding box for multi-oriented object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 4, pp. 1452–1459, Apr. 2021.

[201] Y. Zhou, S. Chen, J. Zhao, R. Yao, Y. Xue, and A. E. Saddik, "CLT-Det: Correlation learning based on transformer for detecting dense objects in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, Sep. 2022.

[202] W. Liu, L. Ma, and H. Chen, "Arbitrary-oriented ship detection framework in optical remote-sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 6, pp. 937–941, Jun. 2018.

[203] X. Yang, Q. Liu, J. Yan, A. Li, Z. Zhang, and G. Yu, "R3Det: Refined single-stage detector with feature refinement for rotating object," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 3163–3171.

[204] X. Wu, D. Hong, J. Tian, J. Chanussot, W. Li, and R. Tao, "ORSim detector: A novel object detection framework in optical remote sensing imagery using spatial-frequency channel features," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 5146–5158, Jul. 2019.

[205] W. Liu et al., "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.

[206] W. Ma et al., "Feature split-merge-enhancement network for remote sensing object detection," *IEEE Trans. Geosci. Remote Sens.*, Jan. 2022.

[207] T. Zhang et al., "Foreground refinement network for rotated object detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, Sep. 2021, doi: 10.1109/TGRS.2021.3109145.

[208] J. Xue, D. He, M. Liu, and Q. Shi, "Dual network structure with interweaved global-local feature hierarchy for transformer-based object detection in remote sensing image," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6856–6866, Aug. 2022.

[209] Z. Soleimanitaleb, M. A. Keyvanrad, and A. Jafari, "Object tracking methods: A review," in *Proc. IEEE 9th Int. Conf. Comput. Knowl. Eng.*, 2019, pp. 282–288.

[210] L. Jiao, D. Wang, Y. Bai, P. Chen, and F. Liu, "Deep learning in visual tracking: A review," *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1–20, 2021, doi: 10.1109/TNNLS.2021.3136907.

[211] E. Macioszek and A. Kurek, "Extracting road traffic volume in the city before and during COVID-19 through video remote sensing," *Remote Sens.*, vol. 13, no. 12, 2021, Art. no. 2329.

[212] V. V. Klemas, "Coastal and environmental remote sensing from unmanned aerial vehicles: An overview," *J. Coastal Res.*, vol. 31, no. 5, pp. 1260–1267, 2015.

[213] J. Li, D. H. Ye, M. Kolsch, J. P. Wachs, and C. A. Bouman, "Fast and robust UAV to UAV detection and tracking from video," *IEEE Trans. Emerg. Topics Comput.*, vol. 10, no. 3, pp. 1519–1531, Jul.–Sep. 2022.

[214] T. Xu, Z.-H. Feng, X.-J. Wu, and J. Kittler, "Learning adaptive discriminative correlation filters via temporal consistency preserving spatial feature selection for robust visual object tracking," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5596–5609, Nov. 2019.

[215] Z. Li, W. Wei, T. Zhang, M. Wang, S. Hou, and X. Peng, "Online multi-expert learning for visual tracking," *IEEE Trans. Image Process.*, vol. 29, pp. 934–946, Aug. 2019.

[216] H. Wang et al., "Vision based long range object detection and tracking for unmanned surface vehicle," in *Proc. IEEE 7th Int. Conf. Cybern. Intell. Syst. Conf. Robot., Automat. Mechatronics*, 2015, pp. 101–105.

[217] D. Frost and J.-R. Tapamo, "Detection and tracking of moving objects in a maritime environment using level set with shape priors," *EURASIP J. Image Video Process.*, vol. 2013, no. 1, 2013, Art. no. 42.

[218] Y. Cui, B. Hou, Q. Wu, B. Ren, S. Wang, and L. Jiao, "Remote sensing object tracking with deep reinforcement learning under occlusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, Jul. 2021.

[219] B. Du, Y. Sun, S. Cai, C. Wu, and Q. Du, "Object tracking in satellite videos by fusing the kernel correlation filter and the three-frame-difference algorithm," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 168–172, Feb. 2018.

[220] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 583–596, Mar. 2015.

[221] J. Shao, B. Du, C. Wu, and L. Zhang, "Tracking objects from satellite videos: A velocity feature based correlation filter," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7860–7871, Oct. 2019.

[222] Y. Chen, Y. Tang, Z. Yin, T. Han, B. Zou, and H. Feng, "Single object tracking in satellite videos: A correlation filter-based dual-flow tracker," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6687–6698, Jun. 2022.

[223] C. Fu, J. Xu, F. Lin, F. Guo, T. Liu, and Z. Zhang, "Object saliency-aware dual regularized correlation filter for real-time aerial tracking," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8940–8951, Dec. 2020.

[224] S. Xuan et al., "Rotation adaptive correlation filter for moving object tracking in satellite videos," *Neurocomputing*, vol. 438, pp. 94–106, 2021.

[225] J. Feng, B. Hui, Y. Liang, Q. Yao, and X. Zhang, "Improved siamRPN++ with clustering-based frame differencing for object tracking of remote sensing videos," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 4163–4166.

[226] J. Shao, B. Du, C. Wu, M. Gong, and T. Liu, "HRSiam: High-resolution siamese network, towards space-borne satellite video tracking," *IEEE Trans. Image Process.*, vol. 30, pp. 3056–3068, Feb. 2021.

[227] W. Song et al., "A joint siamese attention-aware network for vehicle object tracking in satellite videos," *IEEE Trans. Geosci. Remote Sens.*, Jun. 2022.

[228] W. Zhang, L. Jiao, F. Liu, L. Li, X. Liu, and J. Liu, "MBLT: Learning motion and background for vehicle tracking in satellite videos," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Sep. 2021.

[229] Q. He, X. Sun, Z. Yan, B. Li, and K. Fu, "Multi-object tracking in satellite videos with graph-based multitask modeling," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, Feb. 2022.

[230] J. Xiao, H. Cheng, H. Sawhney, and F. Han, "Vehicle detection and tracking in wide field-of-view aerial video," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 679–684.

[231] J. Zhang, X. Jia, J. Hu, and K. Tan, "Satellite multi-vehicle tracking under inconsistent detection conditions by bilevel k-shortest paths optimization," in *Proc. IEEE Digit. Image Comput., Techn. Appl.*, 2018, pp. 1–8.

[232] S. A. Ahmadi, A. Ghorbanian, and A. Mohammadzadeh, "Moving vehicle detection, tracking and traffic parameter estimation from a satellite video: A perspective on a smarter city," *Int. J. remote Sens.*, vol. 40, no. 22, pp. 8379–8394, 2019.

[233] J. Zhang, X. Zhang, X. Tang, Z. Huang, and L. Jiao, "Vehicle detection and tracking in remote sensing satellite vidio based on dynamic association," in *Proc. IEEE 10th Int. Workshop Anal. Multitemporal Remote Sens. Images*, 2019, pp. 1–4.
[234] W. Ao, Y. Fu, X. Hou, and F. Xu, "Needles in a haystack: Tracking city-scale moving vehicles from continuously moving satellite," *IEEE Trans. Image Process.*, vol. 29, pp. 1944–1957, Oct. 2020.
[235] J. Feng et al., "Cross-frame keypoint-based and spatial motion information-guided networks for moving vehicle detection and tracking in satellite videos," *ISPRS J. Photogrammetry Remote Sens.*, vol. 177, pp. 116–130, 2021.
[236] J. Wu, X. Su, Q. Yuan, H. Shen, and L. Zhang, "Multivehicle object tracking in satellite video enhanced by slow features and motion features," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–26, Dec. 2022.
[237] X. X. Zhu et al., "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, 2017.
[238] Y. Xiang, R. Mottaghi, and S. Savarese, "Beyond PASCAL: A benchmark for 3D object detection in the wild," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2014, pp. 75–82.
[239] Y. Xiang et al., "ObjectNet3D: A large scale database for 3D object recognition," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 160–176.
[240] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 3354–3361.
[241] J. Langhammer, B. Janský, J. Kocum, and R. Minařík, "3-D reconstruction of an abandoned Montane reservoir using UAV photogrammetry, aerial LiDAR and field survey," *Appl. Geogr.*, vol. 98, pp. 9–21, 2018.
[242] C. Lu, H. Uchiyama, D. Thomas, A. Shimada, and R.-I. Taniguchi, "Sparse cost volume for efficient stereo matching," *Remote Sens.*, vol. 10, no. 11, 2018, Art. no. 1844.
[243] J. M. Facil, B. Ummenhofer, H. Zhou, L. Montesano, T. Brox, and J. Civera, "CAM-Convs: Camera-aware multi-scale convolutions for single-view depth," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 11826–11835.
[244] R. Wang, S. M. Pizer, and J.-M. Frahm, "Recurrent neural network for (un-) supervised learning of monocular video visual odometry and depth," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 5555–5564.
[245] H. Laga, L. V. Jospin, F. Boussaid, and M. Bennamoun, "A survey on deep learning techniques for stereo-based depth estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 4, pp. 1738–1764, Apr. 2022.
[246] M. Morley, R. Atkinson, D. Savić, and G. Walters, "GANet: Genetic algorithm platform for pipe network optimisation," *Adv. Eng. Softw.*, vol. 32, no. 6, pp. 467–475, 2001.
[247] J.-R. Chang and Y.-S. Chen, "Pyramid stereo matching network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 5410–5418.
[248] G. Yang, H. Zhao, J. Shi, Z. Deng, and J. Jia, "SegStereo: Exploiting semantic information for disparity estimation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 636–651.
[249] J. Liu et al., "PlaneMVS: 3D plane reconstruction from multi-view stereo," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 8665–8675.
[250] Y. Wei, S. Liu, Y. Rao, W. Zhao, J. Lu, and J. Zhou, "NerfingMVS: Guided optimization of neural radiance fields for indoor multi-view stereo," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 5610–5619.
[251] X. Wang et al., "Multi-view stereo in the deep learning era: A comprehensive revfiew," *Displays*, vol. 70, 2021, Art. no. 102102.
[252] R. Peng, R. Wang, Z. Wang, Y. Lai, and R. Wang, "Rethinking depth estimation for multi-view stereo: A unified representation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 8645–8654.
[253] M. Yu, K. Deng, H. Yang, and C. Qin, "Improved wαsh feature matching based on 2D-DWT for stereo remote sensing images," *Sensors*, vol. 18, no. 10, 2018, Art. no. 3494.
[254] R. Tao, Y. Xiang, and H. You, "An edge-sense bidirectional pyramid network for stereo matching of VHR remote sensing images," *Remote Sens.*, vol. 12, no. 24, 2020, Art. no. 4025.
[255] Q. Jia, X. Wan, B. Hei, and S. Li, "DispNet based stereo matching for planetary scene depth estimation using remote sensing images," in *Proc. IEEE 10th IAPR Workshop Pattern Recognit. Remote Sens.*, 2018, pp. 1–5.
[256] M. Tatarchenko, S. R. Richter, R. Ranftl, Z. Li, V. Koltun, and T. Brox, "What do single-view 3D reconstruction networks learn?," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 3405–3414.
[257] X. Zhang, Z. Zhang, C. Zhang, J. Tenenbaum, B. Freeman, and J. Wu, "Learning to reconstruct shapes from unseen classes," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 2263–2274.
[258] J. Pan, J. Li, X. Han, and K. Jia, "Residual MeshNet: Learning to deform meshes for single-view 3D reconstruction," in *Proc. IEEE Int. Conf. 3D Vis.*, 2018, pp. 719–727.
[259] E. Stathopoulou, M. Welponer, and F. Remondino, "Open-source image-based 3D reconstruction pipelines: Review, comparison and evaluation," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. XLII-2/W17, pp. 331–338, 2019.
[260] Z. Hu, Y. Hou, P. Tao, and J. Shan, "IMGTR: Image-triangle based multi-view 3D reconstruction for urban scenes," *ISPRS J. Photogrammetry Remote Sens.*, vol. 181, pp. 191–204, 2021.
[261] E. Rupnik, M. Pierrot-Deseilligny, and A. Delorme, "3D reconstruction from multi-view VHR-satellite images in MicMac," *ISPRS J. Photogrammetry Remote Sens.*, vol. 139, pp. 201–211, 2018.
[262] X. Roynard, J.-E. Deschaud, and F. Goulette, "Paris-Lille-3D: A large and high-quality ground-truth urban point cloud dataset for automatic segmentation and classification," *Int. J. Robot. Res.*, vol. 37, no. 6, pp. 545–557, 2018.
[263] G.-S. Xia et al., "DOTA: A large-scale dataset for object detection in aerial images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3974–3983.
[264] Q. Yin et al., "Detecting and tracking small and dense moving objects in satellite videos: A benchmark," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–18, Nov. 2022.
[265] M. J. Canty, *Image Analysis, Classification and Change Detection in Remote Sensing: With Algorithms for ENVI/IDL and Python*. Boca Raton, FL, USA: CRC Press, 2014.
[266] M. Amani et al., "Google Earth engine cloud computing platform for remote sensing Big Data applications: A comprehensive review," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5326–5350, 2020.
[267] P. Wang, J. Wang, Y. Chen, and G. Ni, "Rapid processing of remote sensing images based on cloud computing," *Future Gener. Comput. Syst.*, vol. 29, no. 8, pp. 1963–1968, 2013.
[268] M. D. Lewis et al., "Disease outbreak detection system using syndromic data in the greater Washington DC area," *Amer. J. Prev. Med.*, vol. 23, no. 3, pp. 180–186, 2002.
[269] G. Sheng, W. Yang, T. Xu, and H. Sun, "High-resolution satellite scene classification using a sparse coding based multiple feature combination," *Int. J. remote Sens.*, vol. 33, no. 8, pp. 2395–2412, 2012.
[270] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and State of the Art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.
[271] X.-Y. Tong et al., "Land-cover classification with high-resolution remote sensing images using transferable deep models," *Remote Sens. Environ.*, vol. 237, 2020, Art. no. 111322.
[272] I. Nigam, C. Huang, and D. Ramanan, "Ensemble knowledge transfer for semantic segmentation," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2018, pp. 1499–1508.
[273] Y. Lyu, G. Vosselman, G.-S. Xia, A. Yilmaz, and M. Y. Yang, "UAVid: A semantic segmentation dataset for UAV imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 165, pp. 108–119, 2020.
[274] R. Hänsch et al., "The 2022 IEEE GRSS data fusion contest: Semisupervised learning [Technical committees]," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 1, pp. 334–337, Mar. 2022.
[275] 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.
[276] H. Chen and Z. Shi, "A spatial-temporal attention-based method and a new dataset for remote sensing image change detection," *Remote Sens.*, vol. 12, no. 10, 2020, Art. no. 1662. [Online]. Available: https://www.mdpi.com/2072-4292/12/10/1662
[277] R. C. Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, "Urban change detection for multispectral Earth observation using convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 2115–2118.
[278] M. Lebedev, Y. V. Vizilter, O. Vygolov, V. Knyaz, and A. Y. Rubis, "Change detection in remote sensing images using conditional adversarial networks," *Int. Arch. Photogrammetry Remote Sens. Spatial Inf. Sci.*, vol. 42, no. 2, pp. 565–571, 2018.
[279] R. Gupta et al., "Creating xBD: A dataset for assessing building damage from satellite imagery," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 10–17.

[280] R. C. Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, "Multitask learning for large-scale semantic change detection," *Comput. Vis. Image Understanding*, vol. 187, 2019, Art. no. 102783.

[281] C. Zhang et al., "A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 166, pp. 183–200, 2020.

[282] 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.

[283] 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. 2021.

[284] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.

[285] Y. Long, Y. Gong, Z. Xiao, and Q. Liu, "Accurate object localization in remote sensing images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2486–2498, May 2017.

[286] Y. Zhang, Y. Yuan, Y. Feng, and X. Lu, "Hierarchical and robust convolutional neural network for very high-resolution remote sensing object detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5535–5548, Aug. 2019.

[287] P. Zhu et al., "Detection and tracking meet drones challenge," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 7380–7399, Nov. 2022.

[288] K. Li, G. Wan, G. Cheng, L. Meng, and J. Han, "Object detection in optical remote sensing images: A survey and a new benchmark," *ISPRS J. Photogrammetry Remote Sens.*, vol. 159, pp. 296–307, 2020.

[289] S. W. Zamir et al., "iSAID: A large-scale dataset for instance segmentation in aerial images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 28–37.

[290] S. Wei, X. Zeng, Q. Qu, M. Wang, H. Su, and J. Shi, "HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation," *IEEE Access*, vol. 8, pp. 120234–120254, 2020.

[291] J. Ding et al., "Object detection in aerial images: A large-scale benchmark and challenges," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 7778–7796, Nov. 2022.

[292] J. Zhang, X. Jia, and J. Hu, "Error bounded foreground and background modeling for moving object detection in satellite videos," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2659–2669, Apr. 2020.

[293] M. Mueller, N. Smith, and B. Ghanem, "A benchmark and simulator for UAV tracking," in *Proc. 14th Eur. Conf. Comput. Vis.*, 2016, pp. 445–461.

[294] D. Du et al., "The unmanned aerial vehicle benchmark: Object detection and tracking," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 370–386.

[295] M. Zhao, S. Li, S. Xuan, L. Kou, S. Gong, and Z. Zhou, "SatSOT: A benchmark dataset for satellite video single object tracking," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5617611.

[296] S. Zolanvari et al., "DublinCity: Annotated LiDAR point cloud and its applications," pp. 1–13, Sep. 2019.

[297] N. Varney, V. K. Asari, and Q. Graehling, "DALES: A large-scale aerial LiDAR data set for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 186–187.

[298] Z. Ye et al., "LASDU: A large-scale aerial LiDAR dataset for semantic labeling in dense urban areas," *ISPRS Int. J. Geo-Inf.*, vol. 9, no. 7, 2020, Art. no. 450.

[299] X. Li et al., "Campus3D: A photogrammetry point cloud benchmark for hierarchical understanding of outdoor scene," in *Proc. 28th ACM Int. Conf. Multimedia*, 2020, pp. 238–246.

[300] Q. Hu, B. Yang, S. Khalid, W. Xiao, N. Trigoni, and A. Markham, "Towards semantic segmentation of urban-scale 3D point clouds: A dataset, benchmarks and challenges," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 4977–4987.

[301] M. Chen et al., "STPLS3D: A large-scale synthetic and real aerial photogrammetry 3D point cloud dataset," in *Proc. 33rd Brit. Mach. Vis. Conf.*, Nov. 21–24, 2022, [Online]. Available: https://bmvc2022.mpi-inf.mpg.de/0429.pdf

[302] Y. E. Wang, G.-Y. Wei, and D. Brooks, "Benchmarking TPU, GPU, and CPU platforms for deep learning," 2019, *arXiv:1907.10701*.

[303] J. Gu et al., "Tiresias: A GPU cluster manager for distributed deep learning," in *Proc. 16th USENIX Symp. Networked Syst. Des. Implementation*, 2019, pp. 485–500.

[304] L. Jiao et al., *FPGA Based Deep Neural Network Design and Implementation*. Xi'an, China: Xidian Univ. Press, 2020.

[305] B. Li, J. Gu, and W. Jiang, "Artificial intelligence (AI) chip technology review," in *Proc. IEEE Int. Conf. Mach. Learn., Big Data Bus. Intell.*, 2019, pp. 114–117.

[306] T. Tuma, A. Pantazi, M. Le Gallo, A. Sebastian, and E. Eleftheriou, "Stochastic phase-change neurons," *Nature Nanotechnol.*, vol. 11, no. 8, pp. 693–699, 2016.

[307] R. Wu, X. Guo, J. Du, and J. Li, "Accelerating neural network inference on FPGA-based platforms—A survey," *Electronics*, vol. 10, no. 9, 2021, Art. no. 1025.

[308] W. Li et al., "A real-time tree crown detection approach for large-scale remote sensing images on FPGAs," *Remote Sens.*, vol. 11, no. 9, 2019, Art. no. 1025.

[309] A. G. Ortiz et al., "A runtime-scalable and hardware-accelerated approach to on-board linear unmixing of hyperspectral images," *Remote Sens.*, vol. 10, no. 11, 2018, Art. no. 1790.

[310] C. V. González, S. Bernabe, D. Mozos, and A. Plaza, "FPGA implementation of an algorithm for automatically detecting targets in remotely sensed hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4334–4343, Sep. 2016.

[311] D. Báscones, C. González, and D. Mozos, "An extremely pipelined FPGA implementation of a lossy hyperspectral image compression algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7435–7447, Oct. 2020.

[312] C. V. González, J. Resano, D. Mozos, A. Plaza, and D. Valencia, "FPGA implementation of the pixel purity index algorithm for remotely sensed hyperspectral image analysis," *EURASIP J. Adv. Signal Process.*, vol. 2010, 2010, Art. no. 969806.

[313] J. Lei et al., "A novel FPGA-based architecture for fast automatic target detection in hyperspectral images," *Remote Sens.*, vol. 11, no. 2, 2019, Art. no. 146.

[314] N. Zhang, X. Wei, H. Chen, and W. Liu, "FPGA implementation for CNN-based optical remote sensing object detection," *Electronics*, vol. 10, no. 3, 2021, Art. no. 282.

[315] S. Liu and W. Luk, "Towards an efficient accelerator for DNN-based remote sensing image segmentation on FPGAs," in *Proc. IEEE 29th Int. Conf. Field Programmable Log. Appl.*, 2019, pp. 187–193.

[316] X. Wei, W. Liu, L. Chen, L. Ma, H. Chen, and Y. Zhuang, "FPGA-based hybrid-type implementation of quantized neural networks for remote sensing applications," *Sensors*, vol. 19, no. 4, 2019, Art. no. 924.

[317] X. Zhang, X. Wei, Q. Sang, H. Chen, and Y. Xie, "An efficient FPGA-based implementation for quantized remote sensing image scene classification network," *Electronics*, vol. 9, no. 9, 2020, Art. no. 1344.

[318] N. Zhang, X. Wei, L. Chen, and H. Chen, "Three-level memory access architecture for FPGA-based real-time remote sensing image processing system," in *Proc. IEEE Int. Conf. Signal, Inf. Data Process.*, 2019, pp. 1–6.

[319] W. Maass, "Networks of spiking neurons: The third generation of neural network models," *Neural Netw.*, vol. 10, no. 9, pp. 1659–1671, 1997.

[320] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 3859–3869.

[321] D. Hong et al., "Interpretable hyperspectral artificial intelligence: When nonconvex modeling meets hyperspectral remote sensing," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 2, pp. 52–87, Jun. 2021.

[322] X. Guo, B. Hou, B. Ren, Z. Ren, and L. Jiao, "Network pruning for remote sensing images classification based on interpretable CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, May 2021.

[323] Q. Zhang, Y. N. Wu, and S.-C. Zhu, "Interpretable convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8827–8836.

[324] Y. Tang and D. Ha, "The sensory neuron as a transformer: Permutation-invariant neural networks for reinforcement learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, vol. 34, pp. 22574–22587.

[325] L. Mou, Y. Hua, and X. X. Zhu, "A relation-augmented fully convolutional network for semantic segmentation in aerial scenes," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 12416–12425.

[326] X. Cao, X. Fu, C. Xu, and D. Meng, "Deep spatial-spectral global reasoning network for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5504714.

[327] R. Cong, Y. Zhang, L. Fang, J. Li, Y. Zhao, and S. Kwong, "RRNet: Relational reasoning network with parallel multiscale attention for salient object detection in optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, Oct. 2021.

[328] L. P. Jain, W. J. Scheirer, and T. E. Boult, "Multi-class open set recognition using probability of inclusion," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 393–409.
[329] W. J. Scheirer, L. P. Jain, and T. E. Boult, "Probability models for open set recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 11, pp. 2317–2324, Nov. 2014.
[330] T. Brown et al., "Language models are few-shot learners," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, vol. 33, pp. 1877–1901.
[331] L. Zhu, Z. Liu, and S. Han, "Deep leakage from gradients," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 14774–14784.
[332] S. Ma et al., "Neuromorphic computing chip with spatiotemporal elasticity for multi-intelligent-tasking robots," *Sci. Robot.*, vol. 7, no. 67, 2022, Art. no. eabk2948.
[333] F. Tosi, F. Aleotti, M. Poggi, and S. Mattoccia, "Learning monocular depth estimation infusing traditional stereo knowledge," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 9799–9809.

Licheng Jiao (Fellow, IEEE) received the B.S.degree from Shanghai Jiaotong University, Shanghai, China, in 1982 and the M.S. and PhD degree from Xi'an Jiaotong University, Xi.an, China, in 1984 and 1990, respectively.

Since 1992, he has been a distinguished professor with the school of Electronic Engineering, Xidian University, Xi.an, where he is currently the Director of Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education of China. He has been a foreign member of the academia European and the Russian academy of natural sciences. His research interests include machine learning, deep learnig, natural computation, remote sensing, image processing, and intelligent information processing.

Prof. Jiao is the Chairperson of the Awards and Recognition Committee, the Vice Board Chairperson of the Chinese Association of Artificial Intelligence, the Foreign member of the Academia European, the Foreign member of the Russian Academy of Natural Sciences, the fellow of IEEE, The Institution of Engineering and Technology (IET), Chinese Association for Artificial Intelligence (CAAI), China Computer Federation (CCF) and Chinese Association of Automation (CAA), a Councilor of the Chinese Institute of Electronics (CIE), a committee member of the Chinese Committee of Neural Networks, and an expert of the Academic Degrees Committee of the State Council.

Zhongjian Huang (Student Member, IEEE) received the B.S. degree in intelligent science and technology in 2018 from Xidian University, Xian, China, where he is currently working toward the Ph.D. degree in computer science and technology.

He is currently a Member of Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, International Research Center for Intelligent Perception and Computation, and Joint International Research Laboratory of Intelligent Perception and Computation, Xidian University. His current research interests include video tracking and satellite videos analysis.

Xu Liu (Member, IEEE) received the B.S. degree in mathematics and applied mathematics from the North University of China, Taiyuan, China, in 2013, and the Ph.D. degree in electronic circuit and system from Xidian University, Xi'an, China, in 2019.

He is currently an Associate Professor of Huashan elite and Postdoctoral Researcher with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, School of Artificial Intelligence, Xidian University, Xi'an, China. From 2015 to 2019, he is the Chair of IEEE Xidian University Student Branch. His current research interests include machine learning and image processing.

Yuting Yang (Graduate Student Member, IEEE) received the B.S. degree in electronic information science and technology from Northwest University, Xi'an, China, in 2018. She is currently working toward the Ph.D. degree in computer science and technology with Xidian University, Xi'an, China.

She is currently a Member of Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, International Research Center for Intelligent Perception and Computation, and Joint International Research Laboratory of Intelligent Perception and Computation, Xidian University. Her research interests include computer vision, the interpretability of deep learning, and multiscale geometric analysis.

Mengru Ma received the B.S. degree in communication engineering from the Hebei University of Engineering, Han Dan, China, in 2019. She is currently working toward the M.S. degree in computer science and technology with the School of Artificial Intelligence, Xidian University, Xi'an, China.

Her research interests include deep learning, image interpretation, and multiresolution remote sensing images fusion classification.

Jiaxuan Zhao (Graduate Student Member, IEEE) received the B.S. degree in materials science and engineering in 2019 from Xidian University, Xi'an, China, where she is currently working toward the Ph.D. degree in computer science and technology with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, School of Artificial Intelligence.

Her research interests include multimodal fusion, evolutionary computing, and image understanding.

Chao You received the B.E. degree in aerospace science and technology from Xidian University, Xi'an, China, in 2019.

Since then, he has been directly taking doctoral programs with the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education of China, School of Artificial Intelligence, Xidian University. His research interests include deep learning, remote sensing image processing, and graph neural network.

Biao Hou (Member, IEEE) received the B.S. and M.S. degrees in mathematics from Northwest University, Xi'an, China, in 1996 and 1999, respectively, and the Ph.D. degree in circuits and systems from Xidian University, Xi'an, China, in 2003.

Since 2003, he has been with the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education, Xidian University, where he is currently a Professor. His research interests include compressive sensing and synthetic aperture radar image interpretation.

Shuyuan Yang (Senior Member, IEEE) received the B.A. degree in electrical engineering and the M.S. and Ph.D. degrees in circuit and system from Xidian University, Xi'an, China, in 2000, 2003, and 2005, respectively.

She has been a Professor with the School of Artificial Intelligence, Xidian University. Her research interests include machine learning and multiscale geometric analysis.

Fang Liu (Senior Member, IEEE) received the B.S. degree in computer science and technology from Xi'an Jiaotong University, Xi'an, China, in 1984, and the M.S. degree in computer science and technology from Xidian University, Xi'an, China, in 1995.

She is currently a Professor with the School of Computer Science, Xidian University. Her research interests include signal and image processing, synthetic aperture radar image processing, multiscale geometry analysis, learning theory and algorithms, optimization problems, and data mining.

Wenping Ma (Senior Member, IEEE) received the B.S. degree in computer science and technology and the Ph.D. degree in pattern recognition and intelligent systems from Xidian University, Xi'an, China, in 2003 and 2008, respectively.

She is currently an Associate Professor with the School of Artificial Intelligence, Xidian University. Her research interests include natural computing and intelligent image processing.

Dr. Ma is a Member of CIE.

Lingling Li (Senior Member, IEEE) received the B.S. degree in electronic and information engineering and the Ph.D. degree in intelligent information processing from Xidian University, Xi'an, China, in 2011 and 2017, respectively.

From 2013 to 2014, she was an Exchange Ph.D. Student with the Intelligent Systems Group, Department of Computer Science and Artificial Intelligence, University of the Basque Country UPV/EHU, Leioa, Spain. She is currently an Associate Professor with the School of Artificial Intelligence, Xidian University. Her research interests include quantum evolutionary optimization, and deep learning.

Puhua Chen (Senior Member, IEEE) received the B.S. degree in environmental engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2009, and the Ph.D. degree in circuit and system from Xidian University, Xi'an, China, in 2016.

She is currently a Lecturer with the School of Artificial Intelligence, Xidian University. Her research interests include machine learning, pattern recognition, and remote sensing image interpretation.

Zhixi Feng (Member, IEEE) received the B.A. degree in automation from the Lanzhou University of Technology, Lanzhou, China, in 2012, and the Ph.D. degree in intelligent information processing from Xidian University, Xi'an, China, in 2018.

He is currently an Associate Professor of artificial intelligence with Xidian University. His research interests include machine learning and remote sensing information processing.

Xu Tang (Senior Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in electronic circuit and system from Xidian University, Xi'an, China, in 2007, 2010, and 2017, respectively.

From 2015 to 2016, he was a Joint Ph.D. Student along with Prof. W. J. Emery with the University of Colorado at Boulder, Boulder, CO, USA. He is currently an Associate Professor with the Key Laboratory of Intelligent Perception and Image Understanding, Ministry of Education, Xidian University. His research interests include remote sensing image content-based retrieval and reranking, hyperspectral image processing, remote sensing scene classification, and object detection.

Yuwei Guo (Senior Member, IEEE) was born in Shaanxi, China, in March 1988. She is currently working toward the M.S. and Ph.D. degrees in circuit and system with Xidian University, Xi'an, China.

She is currently an Associate Professor with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Xidian University. Her research interests include rough set theory, data mining, and image processing.

Xiangrong Zhang (Senior Member, IEEE) received the B.S. and M.S. degrees in computer application technology from the School of Computer Science, Xidian University, Xi'an, China, in 1999 and 2003, respectively, and the Ph.D. degree in pattern recognition and intelligent system from the School of Electronic Engineering, Xidian University, in 2006.

She is currently a Professor with the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education, Xidian University. From January 2015 to March 2016, she was a Visiting Scientist with the Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology. Her research interests include pattern recognition, machine learning, and remote sensing image analysis and understanding.

Dou Quan (Member, IEEE) received the B.S. degree in intelligent science and technology and the Ph.D. degree in electronic circuit and system from Xidian University, Xi'an, China, in 2015 and 2021, respectively.

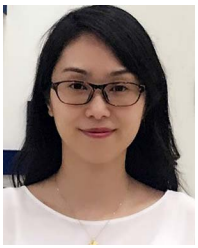
From 2019 to 2020, she was a Joint Ph.D. along with Prof. Jocelyn Chanussot with the Research Center of Inria Grenoble-Rhone-Alpes, Montbonnot-Saint-Martin, France. She is currently a Lecturer with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Xidian University. Her research interests include machine learning, deep learning and metric learning, image matching, image registration, and image classification.

Shuang Wang (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in circuits and systems from Xidian University, Xi'an, China, in 2000, 2003, and 2007, respectively.

She is currently a Professor with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Xidian University. Her research interests include sparse representation, image processing, synthetic aperture radar (SAR) automatic target recognition, remote sensing image captioning, and polarimetric SAR data analysis.

Weibin Li received the B.S. and M.S. degrees in mathematics from Northwest University, China, in 1998 and 2000, respectively, and the Ph.D. degree in circuits and systems from Xidian University, Xi.an, China, in 2004.

He is currently a Professor with Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China at Xidian University, Xi.an, China. His research interests are in the area of Spatio-temporal intelligence. His research interest includes GNSS navigation system, remote sensing image processing, Industrial Internet and industrial intelligence.



Jing Bai (Senior Member, IEEE) received the B.S. degree in electronic and information engineering from Zhengzhou University, Zhengzhou, China, in 2004, and the Ph.D. degree in pattern recognition and intelligent systems from Xidian University, Xi'an, China, in 2009.

She is currently a Professor with Xidian University. Her research interests include image processing, machine learning, and intelligent information processing.

Yangyang Li (Senior Member, IEEE) received the B.S. and M.S. degrees in computer science and technology and the Ph.D. degree in pattern recognition and intelligent system from Xidian University, Xi'an, China, in 2001, 2004, and 2007, respectively.

She is currently a Professor with the School of Artificial Intelligence, Xidian University. Her research interests include quantum-inspired evolutionary computation, artificial immune systems, and deep learning.

Ronghua Shang (Senior Member, IEEE) received the B.S. degree in information and computation science and the Ph.D. degree in pattern recognition and intelligent systems from Xidian University, Xi'an, China, in 2003 and 2008, respectively.

She is currently a Professor with Xidian University. Her research interests include evolutionary computation, image processing, and data mining.

Jie Feng (Senior Member, IEEE) received the B.S. degree in electronic information engineering from Chang'an University, Xi'an, China, in 2008, and the Ph.D. degree in electronic circuit and system from Xidian University, Xi'an, China, in 2014.

She is currently an Associate Professor with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University. Her research interests include remote sensing image processing, deep learning, and machine learning.