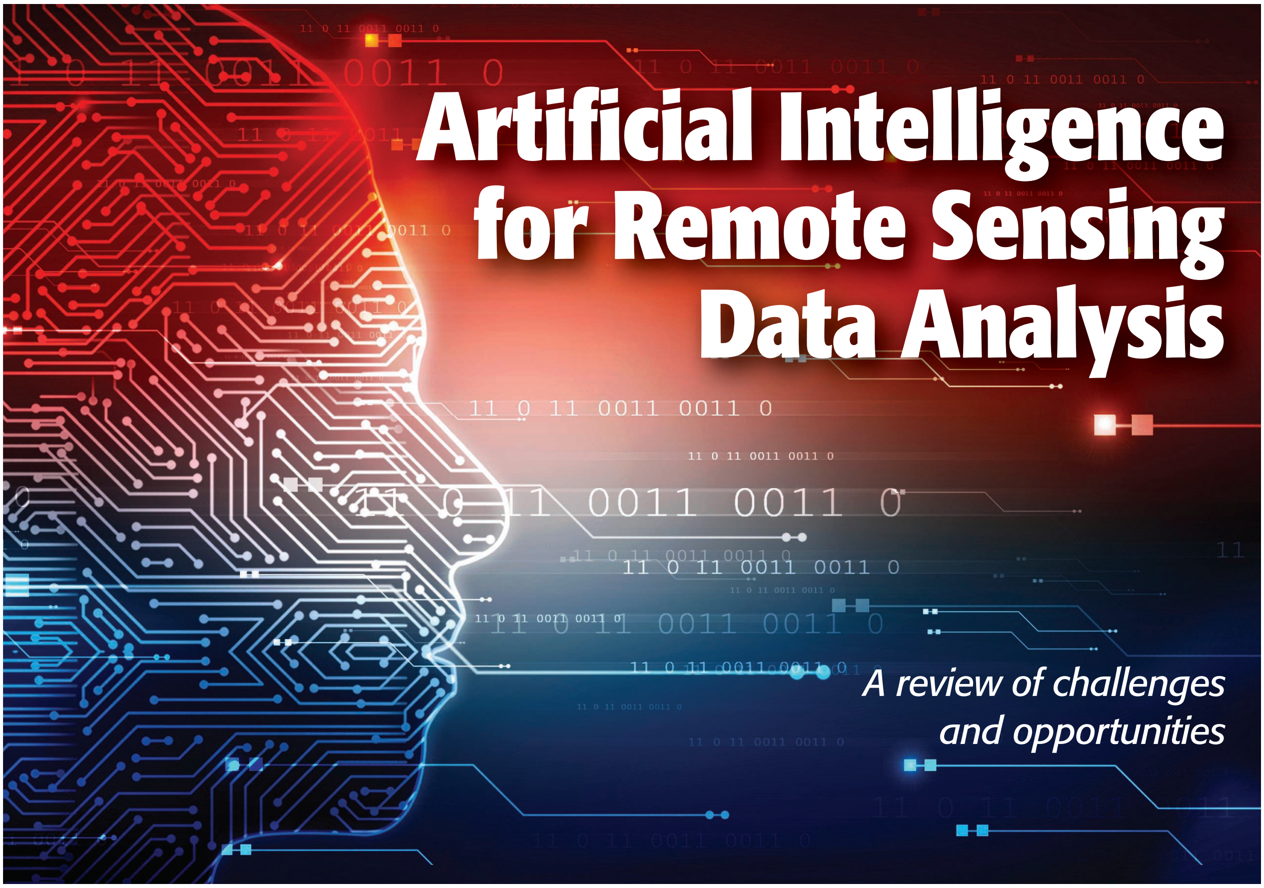

Artificial Intelligence for Remote Sensing Data Analysis

A review of challenges and opportunities

LEFEI ZHANG AND LIANGPEI ZHANG

Artificial intelligence (AI) plays a growing role in remote sensing (RS). Applications of AI, particularly machine learning algorithms, range from initial image processing to high-level data understanding and knowledge discovery. AI techniques have emerged as a powerful strategy for analyzing RS data and led to remarkable breakthroughs in all RS fields. Given this period of breathtaking evolution, this work aims to provide a comprehensive review of the recent achievements of AI algorithms and applications in RS data analysis. The review includes more than 270 research papers, covering the following major aspects of AI innovation for RS: machine learning, computational intelligence, AI explicability, data mining, natural language processing (NLP), and AI security. We conclude this review by identifying promising directions for future research.

INTRODUCTION

With the development of big data, deep learning (DL), large-scale GPUs, and AI chips, the world has moved into the era of AI in the past two decades. In the IEEE Geoscience and Remote Sensing Society (GRSS), AI approaches are being increasingly used to extract information and knowledge from the continuously growing stream of RS data. AI-related algorithms have been actively applied to process and analyze RS images (e.g., classification and detection) in the past five years (see Figure 1). Table 1 shows representative applications of AI approaches to RS scientific tasks, mainly in three aspects: classification and detection, regression, and state prediction [1]. Each aspect includes various RS scientific tasks with specific AI approaches applied to RS data. From this table, DL architectures [e.g., convolutional neural networks (CNNs)] have achieved promising results in various areas and have thus become the most important topic in recent publications and in this review.

The popularity and importance of AI approaches have led to several previous literature reviews to which this work is connected, as summarized in "Connection With Previous Reviews" section. To the best of our knowledge, this work is the first to provide a comprehensive review on AI techniques for RS in terms of the algorithm, methodology, and application. In particular, studies in the past five years are investigated, and the latest progress is discussed. This review focuses not only machine learning and DL (which provide key algorithms and methodologies) but also on different AI facets, i.e., the computational intelligence and explicability

Digital Object Identifier 10.1109/MGRS.2022.3145854
Date of current version: 13 April 2022

of AI (the algorithm level too), and applications (including data mining, NLP, and AI security). The proposed taxonomy may provide a framework to understand current research and identify open challenges for the future.

CONNECTION WITH PREVIOUS REVIEWS

Remarkable literature reviews have been previously published (see Table 2). As such, the present review mainly focuses on recent studies, given the rapid development of AI techniques. In these studies, the most frequently reviewed topic is DL, which has attracted considerable interest in nearly all AI aspects. Milestone reviews and high-impact papers on DL have emerged for RS, of which image analysis is basically a part of machine learning applications. Given that deep NNs (DNNs) have previously performed effectively in machine learning application-related tasks [13], including image classification and object detection, they can be easily introduced and transferred into the RS society to address similar tasks, such as classification and detection in RS images. In these areas, results are likewise promising.

Apart from DL, other aspects are presented in the AI community and are also investigated and used for RS. For example, Zhong et al. presented an overview of the applications of computational intelligence technologies in RS, which can provide possible solutions to challenges in optical RS image processing inspired by biological systems [29]. Plaza et al. organized a special issue that provides a unique perspective on the combination of spatial technologies and social media, sensing and positioning, and monitoring under a big data processing framework [41], which can be viewed as data mining for RS. Recently, the authors in [1] and [21] reviewed the development of DL in a geoscientific context to obtain further process understanding of Earth-system science problems and environmental RS applications.

Table 2 illustrates that all of the aforementioned excellent surveys cover and focus on only one (or at most two) aspect of AI techniques in RS. In the present review, we attempt to provide a comprehensive review that includes as many known aspects of where AI techniques meet RS applications. These aspects are not limited to the locations of

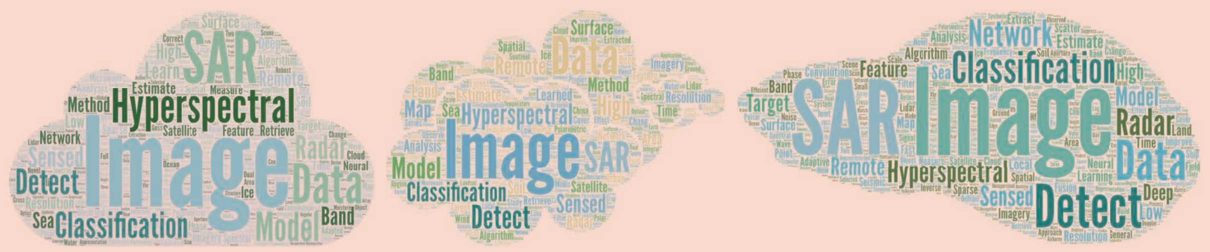

FIGURE 1. The most frequent keywords in IEEE GRSS publications from 2016 to 2020. (a) *IEEE Transactions on Geoscience and Remote Sensing*, (b) *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, and (c) *IEEE Geoscience and Remote Sensing Letters*. The size of each word is proportional to the frequency of that keyword.

TABLE 1. REPRESENTATIVE APPLICATIONS OF AI APPROACHES TO RS TASKS.

SCIENTIFIC TASK	AI APPROACH	RS DATA	REFERENCE
		Classification and Detection	
Land use classification	Object-based convolutional neural network (CNNs)	Vexcel UltraCam Xp digital aerial camera images	[2]
Region of interest detection	Multiview learning, CNNs	*GeoEye-1* images and *SPOT-5* images	[3]
Hail storm detection	Merged deep neural networks	*GOES* images and *MERRA-2* data	[4]
Drug crop site identification	CNNs	*IKONOS* images	[5]
Animal detection	Active learning, CNNs	UAV data sets	[6]
		Regression	
Leaf-area index estimation	Nonlinear, autoregressive networks	HJ-1A/1B CCD images	[7]
Crop-yield estimation	Spiking NNs	e-MODIS NDVI data	[8]
Aerosol optical depth	Backpropagation artificial NN	*Meteosat-10* and *Meteosat-11*	[9]
		State Prediction	
Earthquake detection	RNNs	Seismic waveforms	[10]
Temperature forecasting	Deep belief networks	MOD11A1, MOD13A2, and MCD12Q1	[11]
Rainfall prediction	SVM	Wide-angle high-resolution sky imaging system images and surface weather parameters	[12]

UAV: unmanned aerial vehicle.

where the related reviews are available but also include a few emerging directions, including NLP and AI security for RS, which are clarified in the following sections.

SCOPE

Considerable research has focused on AI techniques for RS. Accordingly, selection criteria are necessary to limit our focus to well-known journals in this field (e.g., IEEE GRSS publications *Remote Sensing* and *Journal of the International Society for Photogrammetry and Remote Sensing*) and major progress in the last five years (2016–current). Despite numerous proposals of AI-based methods for RS, we are unaware of any comprehensive recent survey. A thorough review and summary of existing works are essential for further progress in GRSS, particularly for researchers who hope to enter the field.

TABLE 2. A SUMMARY OF RELATED REVIEWS SINCE 2016.

YEAR	SURVEY TITLE	REFERENCE	VENUE	CONNECTION WITH THIS REVIEW
2022	"Land Cover Change Detection Techniques: Very-High-Resolution Optical Images: A Review"	[14]	*GRSM*	Machine learning
2022	"Low-Rank and Sparse Representation for Hyperspectral Image Processing: A Review"	[15]	*GRSM*	Machine learning
2022	"Change Detection From Very-High-Spatial-Resolution Optical Remote Sensing Images: Methods, Applications, and Future Directions"	[16]	*GRSM*	Machine learning
2021	"Interpretable Hyperspectral Artificial Intelligence: When Nonconvex Modeling Meets Hyperspectral Remote Sensing"	[17]	*GRSM*	AI
2021	"Classification of Remote Sensing Data With Morphological Attribute Profiles: A Decade of Advances"	[18]	*GRSM*	Machine learning
2020	"Remote Sensing Image Scene Classification Meets Deep Learning: Challenges, Methods, Benchmarks, and Opportunities"	[19]	*JSTARS*	Machine learning
2020	"Object Detection in Optical Remote Sensing Images: a Survey and a New Benchmark"	[20]	*ISPRS*	Machine learning
2020	"Deep Learning in Environmental Remote Sensing: Achievements and Challenges"	[21]	*RSE*	Machine learning
2020	"Feature Extraction for Hyperspectral Imagery: The Evolution From Shallow to Deep (Overview and Toolbox)"	[22]	*GRSM*	Machine learning
2019	"Deep Learning and Process Understanding for Data-Driven Earth System Science"	[1]	*Nature*	Machine learning
2019	"Deep Learning for Hyperspectral Image Classification: An Overview"	[23]	*TGRS*	Machine learning
2019	"Deep Learning for Classification of Hyperspectral Data: A Comparative Review"	[24]	*GRSM*	Machine learning
2019	"A Review of Change Detection in Multitemporal Hyperspectral Images: Current Techniques, Applications, and Challenges"	[25]	*GRSM*	Machine learning
2019	"Hyperspectral Band Selection: A Review"	[26]	*GRSM*	Machine learning
2019	"Deep Learning in Remote Sensing Applications: A Meta-Analysis and Review"	[27]	*ISPRS*	Machine learning
2019	"Deep Learning Classifiers for Hyperspectral Imaging: A Review"	[28]	*ISPRS*	Machine learning
2018	"Computational Intelligence in Optical Remote Sensing Image Processing"	[29]	*ASC*	Computational intelligence
2018	"New Frontiers in Spectral-Spatial Hyperspectral Image Classification: The Latest Advances Based on Mathematical Morphology, Markov Random Fields, Segmentation, Sparse Representation, and Deep Learning"	[30]	*GRSM*	Machine learning
2018	"Recent Advances on Spectral-Spatial Hyperspectral Image Classification: An Overview and New Guidelines"	[31]	*TGRS*	Machine learning
2017	"Remote Sensing Image Scene Classification: Benchmark and State of the Art"	[32]	*PIEEE*	Machine learning
2017	"Spatial Technology and Social Media in Remote Sensing: A Survey"	[33]	*PIEEE*	Data mining
2017	"Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources"	[34]	*GRSM*	Machine learning
2017	"Multiple Kernel Learning for Hyperspectral Image Classification: A Review"	[35]	*TGRS*	Machine learning
2017	"A Review of Supervised Object-Based Land-Cover Image Classification"	[36]	*ISPRS*	Machine learning
2016	"Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art"	[37]	*GRSM*	Machine learning
2016	"Domain Adaptation for the Classification of Remote Sensing Data: An Overview of Recent Advances"	[38]	*GRSM*	Machine learning
2016	"A Survey on Object Detection in Optical Remote Sensing Images"	[39]	*ISPRS*	Machine learning
2016	"Random Forest in Remote Sensing: A Review of Applications and Future Directions"	[40]	*ISPRS*	Machine learning

GRSM: IEEE Geoscience and Remote Sensing Magazine; JSTARS: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing ISPRS: Journal of the International Society for Photogrammetry and Remote Sensing; RSE: Remote Sensing of Environment; TGRS: IEEE Transactions on Geoscience and Remote Sensing: ASC: Applied Soft Computing; PIEEE: Proceedings of the IEEE.

MACHINE LEARNING

OVERVIEW AND ALGORITHMS BEFORE THE DL ERA

Machine learning approaches are increasingly being employed to extract patterns and insights into the continuously growing stream of RS data. Various machine learning algorithms have been successfully applied to RS problems, which can be generalized into two basic tasks, namely, classification and regression.

The general formulation of machine learning in RS can be defined as $y = f(x)$, where x is the input data and y is the expected output. Given a prior set of $\{x_i, y_i\} \mid i = 1, \dots, N$, also known as *training samples*, machine learning algorithms find a mapping function $f(\cdot)$ that can satisfy each $\hat{y}_i = f(x_i)$ as close to y_i as possible. This procedure is called *training*. Subsequently, any y_j other than y_i can be determined by $y_j = f(x_j)$, which is known as *testing* (or *predicting*). This idea has been used for various RS tasks [37].

Table 3 lists several milestones of machine learning algorithms in RS. In earlier years, support vector machine (SVM) was a popular technique for RS data analysis [42], including image classification, target detection, and spectral unmixing. The SVM technique remains active in research, even in this DL era, due to its solid theoretical basis (structural risk minimization) and easy expandability to tackle open questions in machine learning (e.g., small sample size and nonindependent identically distributed conditions [38]). Subsequently, manifold learning [43], which explores the geometric and topological structures of RS data, has provided new opportunities in RS. In addition, random forest (RF), given its simple operation and solid performance, is one of the standard algorithms in RS [44]. In 2011, Chen et al. introduced sparse representation (SR) for RS image classification and target detection [45], [46] and has raised a stream to study the sparsity of RS data in subsequent years [47]. Thereafter, RS data analysis has moved into the DL era since 2014; a similar transition in the machine learning community occurred in 2012 [48].

DL

Outstanding reviews on the topic of DL in RS have been published in recent years [1], [22]–[24], [27], [28], [34], [37]. In the following, we subdivide the DL approaches applied to RS into three aspects, namely, convolutional-, generative-, and sequence-based approaches.

CONVOLUTIONAL-BASED METHODS

In 2012, CNNs achieved a record-breaking image classification accuracy in computer vision [48], and since then have been applied with great success to the detection, segmentation, and recognition of objects and regions in images. Technically, the convolutional operation is designed to process data in the form of multiple arrays, such as a multispectral RS image composed of 2D arrays containing pixel intensities in all spectral channels. Such a hierarchical convolutional structure can learn data representations with multiple levels of abstraction [13]. Thus, CNNs have been used in various RS data analytical tasks, including fusion, denoising, pixel classification, object detection, road detection, change detection, scene classification, and retrieval (see Table 4).

FUSION

RS image fusion is an effective means to obtain superior images with high spatial, temporal, and spectral resolutions by merging complementary information [73]. Fusion can be generalized as a machine learning problem by considering the input low-resolution image pairs as x and the corresponding output high-resolution image as y. Although most of the SR-based fusion algorithms perform well [74], CNN-based methods have achieved state-of-the-art accuracy. Based on CNNs, Song et al. suggested a spatiotemporal fusion method under the application background of massive RS data [56].

DENOISING

RS image denoising aims to remove noises (including random noise, stripe noise, and dead pixels) in images to

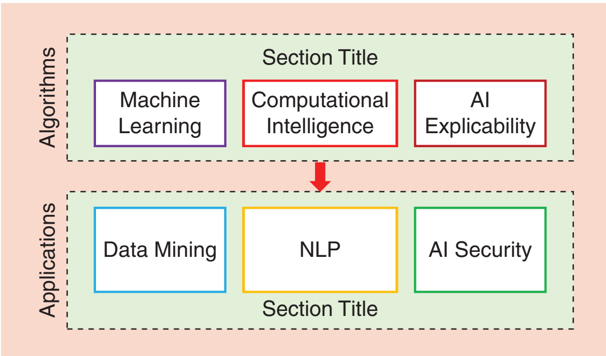

FIGURE 2. The organizational chart of this review.

TABLE 3. THE MILESTONES OF MACHINE LEARNING ALGORITHMS IN RS. THE TIME PERIOD UP TO 2014 WAS DOMINATED BY FEATURE EXTRACTION AND CLASSIFICATION WITH SHALLOW MODELS. A TRANSITION OCCURRED IN 2014, WITH THE DEVELOPMENT OF DL. AFTERWARD, STATE-OF-THE-ART METHODS WERE DOMINATED BY DL. THE MOST LISTED METHODS WERE PUBLISHED IN *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING* AND HIGHLY CITED PAPERS.

ALGORITHMS BEFORE DL				FOCUS OF THIS REVIEW				
SVM [49]	ML [50]	RF [44]	SR [45]	SAE [51]	DBN [52]	CNN [53]	GAN [54]	Transformer [55]
2004	2005	2006	2011	2014	2015	2016	2018	2020

SVM: support vector machine; ML: manifold learning; RF: random forest; SR: sparse representation; DBN: deep belief network; GAN: generative adversarial network; SAE: stacked auto-encoder.

improve their quality. This issue is addressed in most of the image processing-based algorithms by introducing total-variation and low-rank regularizations for single-image denoising [75]. CNN-based methods have outperformed many mainstream denoising techniques in quantitative evaluation. With a combined spatial-spectral deep CNN, Yuan et al. recommended learning a nonlinear, end-to-end mapping between noisy and clean images [58].

PIXEL CLASSIFICATION

Figure 1 demonstrates that a pixel-based classification is one of the most popular topics in the geosciences and RS community. A few technical reviews have focused on DL for RS image classification [23], [24], [28], while other reviews have discussed recent RS image classification methods [22], [30], [31] and DL for RS [34], [37]. Similarly, DL for RS image classification has been investigated in detail.

TABLE 4. A SUMMARY OF THE PROPERTIES OF REPRESENTATIVE CONVOLUTIONAL-BASED METHODS IN RS.

RS TASKS	REPRESENTATIVE ALGORITHM	REFERENCE	DATA SET	BACKBONE	HIGHLIGHTS
Fusion	STFDCNN	[56]	CIA LGC	Two five-layer CNNs	A novel spatiotemporal fusion method based on a CNN
—	Spatial-spectral fusion CNN	[57]	QuickBird *WorldView-2 IKONOS*	RCNN with 17 blocks	Combines a CNN with a variational model
Denoising	HSID-CNN	[58]	Washington, D.C. Indian Pines Pavia University	Sixteen-layer RCNN	Learning a mapping between noisy and clean RS images with a combined spatial-spectral CNN
Pixel classification	Deep FE Architectures	[53]	Indian Pines KSC Pavia University	CNNs with different layers	1D, 2D, and 3D CNNs
—	Spectral-spatial FE	[59]	Pavia Center Pavia University	Five-layer CNN	A CNN is utilized to find spatial-related features at high levels
—	Fully convolutional architecture	[60]	Boston Forez	Five-layer, fully connected convolutions	Fully convolutional architecture
—	Deep pixel-pair features	[61]	Indian Pines Salinas Pavia University	Ten-layer convolutions	Learns pixel-pair method features to increase the discriminative power for classification performance
—	Spectral-spatial residual network	[62]	Indian Pines KSC Pavia University	Thirteen-layer 3D convolutions with residual blocks	Adopts residual blocks to improve classification accuracy and learn features from spectral signatures and spatial contexts
—	Fast CNN	[63]	Indian Pines Pavia University	3D CNN with different layers	GPU implementation for simultaneously learning spatial and spectral information to reduce computation time and increase accuracy
Object detection	Rotation-invariant CNN	[64]	NWPU VHR-10	AlexNet	Learning a new rotation-invariant layer on the basis of existing CNN architectures to advance the performance of object detection
—	Multiscale, multiclass object detection	[65]	NWPU VHR-10 aircraft aerial-vehicle SAR-ship	Faster R-CNN	Simultaneously detecting multiclass objects
Road detection	Cascaded, end-to-end CNN	[66]	Google	Ten-layer convs	Simultaneously coping with road detection and centerline extraction tasks
—	Single patch-based CNN	[67]	Massachusetts Abu Dhabi	Six-layer CNN	Uses a deeper patch-based segmentation
Change detection	GETNET	[68]	Earth Observation-1 farmland countryside Poyang lake river	Four-layer convs	Presents a general, end-to-end 2D CNN to learn discriminative features from multisource data
Scene classification	Gradient-boosting random CNN	[69]	UC Merced Sydney	LeNet	Effectively combines many DNNs for scene classification
—	Transferring deep CNNs	[70]	UC Merced WHU-RS	AlexNet CaffeNet VGGNet PlacesNet	Transfer features from pretrained CNNs for RS scene classification
—	Discriminative CNNs	[71]	UC Merced AID NWPU-RESISC45	AlexNet VGGNet-16 GoogLeNet	Train the CNN with an additional discriminative objective function
Retrieval	Deep hashing NNs	[72]	SAT-4	Three-layer convolution	DHNNs are composed of deep feature and hashing learning neural networks and can be optimized in an end-to-end manner

RCNN: recurrent CNN; KSC: Kennedy Space Center; UC: University of California; STFDCNN: spatio-temporal fusion deep CNN;CIA LGC: Coleambally irrigation area Lower Gwydir Catchment; HSID CNN: hyperspectral image deep CNN; FE: feature extraction; GETNET: general end-to-end 2D CNN; DHNNs: deep hashing NNs.

OBJECT DETECTION

Object detection is also a fundamental problem in RS image analysis. DL-based RS image object detection methods and benchmark data sets have been described in comprehensive overviews [39] and [20].

ROAD DETECTION

CNNs have also been used for road detection in RS images. Cheng et al. proposed a cascaded end-to-end CNN to simultaneously cope with road detection and centerline extraction tasks [66]. Lan et al. used a UNet-like structure with dilated convolutional operations to enlarge the effective receptive field and find discriminative features for road segmentation from RS data [76].

CHANGE DETECTION

Current techniques, applications, and challenges in change detection in multitemporal hyperspectral images are reviewed in [25], where a CNN is considered a promising approach to address open issues (i.e., to consider spatial correlation in change detection). In addition, technical reviews have recently focused on very high-resolution optical images change detection [14], [16].

SCENE CLASSIFICATION

An RS image scene classification task is nearly similar to image classification in computer vision. For such tasks, CNNs are commonly used. Hu et al. experimentally confirmed that pretrained CNNs on computer vision data sets can be efficiently transferred for RS scene classification [69]. A systematic survey of DL for RS image scene classification, including methods, benchmarks, and opportunities, is provided [19].

RETRIEVAL

In recent years, the performance of RS image retrieval has also been improved due to the strong capacity of CNNs. Liu et al. recommended deep feature and adversarial hash learning models for content-based RS image retrieval [77], while Li et al. proposed deep hashing NNs for large-scale RS image retrieval [72].

Furthermore, various convolutional operations exist in RS tasks in addition to the standard CNNs discussed previously. These operations, such as 3D [53], adaptive [78], dilated [79], and graph [80] convolutions, have recently been developed but provide impressive performance. These novel convolutional styles are rich in CNN modalities and help efficiently abstract desired information from RS images.

GENERATIVE-BASED METHODS

Generative models are one of the most important branches of machine learning. However, they received less attention in the early period of DL because CNNs have achieved substantial, surprising performance for computer vision-related tasks. Deep belief networks (DBNs) [81] have been successfully used for RS tasks, including pixel classification [52], object detection [82], and scene classification [83]. Nonetheless, DBN-based approaches may not reach state-of-the-art operation due to the following possible reasons: 1) these methods follow a fully connected network structure, which is not as effective as locally connected CNNs, and 2) the hidden units must be Bool variables in pretraining, thereby limiting their capacity in fine-tune optimization.

An autoencoder (AE) is an encoding-decoding symmetrical network structure trained to copy its input to its output. In the GRSS, AEs have been well applied on the following aspects: 1) Pixel classification. Chen et al. pioneered the use of stacked AEs for hyperspectral RS data classification by considering spectral- and spatial-dominated information [51]. Recently, improved AEs, such as recursive [84], discriminative stacked [85], 3D convolutional [86], and variational [87], have been suggested and proven superior in several experiments. 2) Scene classification. Given its strong capability for feature learning, AEs have been introduced into RS image scene classification. Zhang et al. proposed an unsupervised feature learning with sparse AE to transform an input RS image into a feature vector for SVM classification [88]. 3) Spectral unmixing. The advantages of stack or denoising AEs help to overcome the effects of noises and outliers, leading to their remarkable robustness for such tasks [89]. An untied denoising AE with sparsity [90] and stacked nonnegative sparse AEs [91] have been offered for robust, hyperspectral unmixing. A convolutional operation is also suggested for combination with AE to efficiently incorporate spatial with spectral information [92], [93].

Generative adversarial networks (GANs) have recently become a major research topic. Theoretically, GANs consist of a competing generator and discriminator [94]. In [54], the usefulness and effectiveness of GANs for hyperspectral RS images are demonstrated for the first time. Specifically, a 1D GAN (spectral classifier) and a 3D GAN (spectral-spatial classifier) are well designed, and the generated adversarial samples are used with real training samples to improve their performance. To date, the applications of GANs in RS include the following:

1) Tackling insufficient training samples in classification, including explicitly generating additional samples [95]–[97] and indirectly improving the discrimination and generalization capabilities of classification networks [98]. This idea has also been reported to address RS image change detection [99], scene classification [100] and retrieval [72].
2) Dealing with low-level, image-generation-related tasks, including hyperspectral RS image pan sharpening [102], RS image superresolution [103], and cross-domain generation [104]. Although GANs have opened new opportunities in the RS community for the aforementioned challenges and demonstrated considerable potential for the analysis of complex RS images, the data generated by GANs are not real and, at times, even unreliable. Further details of this fact are provided in the "AI Security" section.

SEQUENCE-BASED METHODS

Recurrent NNs (RNNs) were originally used for modeling the long-range dependencies of sequential data [105] and may also be applied to images and data with large dimensions (e.g., RS data involving spectrum and time). Recently, a few studies have attempted to consider hyperspectral RS images as sequential data, and thus, RNNs have been used to learn features for pixel classification. Wu and Saurabh suggested the use of convolutional RNNs to extract spectral features from hyperspectral RS images [106]. Hang et al. proposed cascaded RNNs for hyperspectral RS image classification [107]. In [108] and [109], variants of RNNs using long short-term memory (LSTM) units are designed to learn spectral-spatial features from RS images. A spatial-sequential RNN [110] and a spectral-spatial RNN [111] have also been developed to include spatial information.

Given their capacity for processing sequential data, RNNs have also been introduced for multitemporal RS image analysis. Song et al. offered a recurrent, 3D, fully convolutional network [112], and Chen et al. suggested a deep, Siamese, convolutional multilayer RNN [113] for RS image change detection. RNNs have achieved considerable achievements in NLP, rather than in computer vision. Accordingly, a few preliminary explorations have been conducted on NLP-related tasks in RS (e.g., image captioning and description based on RNNs), which is reviewed in the NLP section. More recently, transformers, a novel and effective backbone, has been introduced to the RS image classification task. The authors in [55] proposed bidirectional encoder representations from transformers and showed that they outperformed any other CNN-based models. Similarly, Hong et al. recommended a SpectralFormer to rethink RS image classification from a sequential perspective with transformers and achieved a significant improvement in comparison with state-of-the-art backbone networks [114]. Although sequence-based methods have presented promising performance in RS image analysis, current studies focus on limited aspects. The application of RNNs in RS video sequence processing (e.g., object tracking) may be another potential research direction [115].

DATA SETS AND TRAINING SAMPLES

Machine learning is the key technique in AI. However, using such algorithms in RS data requires high-quality data sets, particularly well-labeled ones. Publicly available data sets with validation data are important for researchers to verify their developed algorithms and compare them with state-of-the-art methods. In various standard RS applications, the frequently employed data sets are summarized in Table 5.

Given the aforementioned rich number of data sets, high-quality training samples are the other critical factor in machine learning. Before the DL era, the following two common facts related to training samples could lead to poor results in test data: 1) the training sample size is small, such that the model cannot be optimally learned and 2) the training and test samples are not independently identically distributed, such that the learned model may not work as accurately as on the training data. The aforementioned phenomena frequently occur in the machine learning of RS data. The first issue is addressed by designing machine learning techniques for small sample sizes (e.g., semisupervised learning [151], active learning [152], and feature-dimension reduction [153]). Meanwhile, domain adaptation [38] and transfer learning [154] are widely used to relieve the second problem.

In the DL era, the number of training samples is always far from sufficient to support the training procedure because deep NNs have millions of parameters to be optimized. In accordance with existing literature, the three major strategies for DL with small sample sizes are training sample generation, few-shot learning, and transfer of knowledge from other data sets.

TRAINING SAMPLE GENERATION

In DL-based methods, sample generation is a commonly used technique that can extend the training set. In [53], the authors proposed two simple, yet effective ways (i.e., changing radiation- and mixture-based virtual samples) to create training samples from the perspective of RS imaging. However, these approaches ignore the real distribution of training samples in the feature space. The inherent superiority of GANs in data distribution learning provides feasible approaches to handle high-feature heterogeneity in sample generation, such as generating adversarial samples

TABLE 5. A SUMMARY OF DATA SETS FOR VARIOUS RS APPLICATIONS.

RS APPLICATIONS	REPRESENTATIVE DATA SETS
Data fusion	AHB [116], Tianjin [116], Daxing [116], hyperspectral imaging-lidar [117], and Houston [118]
Pixel classification	Washington, D.C. [119], Pavia University and Pavia Center [120], Indian Pine [119], Salinas Valley [120], KSC [120], Botswana [120], Urban [121], Xiongan [122], GID [123], DeepGlobe [124], Aeroscapes [125], and SEN12MS [126]
Change detection	River [68], MtS-WH [128], SVCD [129], OSCD [130], Urban Atlas [131]
Target recognition	LEVIR [132], DIOR [20], RSOD [133], NWPU VHR-10 [64], VEDAI [134], COWC [135], DOTA [136], ITCVD [137], DIUx xView [138], HRSC [139], and OpenSARShip [140]
Scene classification	OPTIMAL-31 [141], UC Merced [142], SAT-4 and SAT-6 [143], SIRI-WHU [144], AID [145], NWPU-RESISC45 [32], PatternNet [146], RSI-CB [147], AID++ [148], RSD46-WHU [149], and BigEarthNet [150]

KSC: Kennedy Space Center; UC: University of California.

[54], generating and sifting labeled samples [155], and supervised Wasserstein GANs [156]. With the aim to produce large-scale training data in developing sophisticated DL approaches, the authors in [157] provided the SEN12MS data set that contains 180,662 patch triplets of corresponding *Sentinel-1* dual-polarized, synthetic aperture radar (SAR) data; *Sentinel-2* multispectral images; and Moderate-Resolution Imaging Spectrometer (MODIS)-derived land cover maps. These approaches effectively improve the performance in common tasks, such as pixel classification, semantic segmentation, and scene classification.

FEW-SHOT LEARNING

Another group of methods used for mitigating the lack of data in training deep NNs is few-shot learning [158]. In contrast to training sample generation, few-shot learning approaches seek to learn the general latent representation from unlabeled images and transfer the model to new tasks/targets using a few labeled samples. Few-shot learning techniques are introduced in consideration of the sparsity of RS data. For example, deep few-shot learning [159], a deep-relation network [160], and feature-merged, single-shot detection [161] are proposed for classification and detection with few labeled samples. Zero-shot learning, which can deal with unseen land cover classification problems, is also employed in [162] and [163].

TRANSFERRING KNOWLEDGE FROM OTHER DATA SETS

Transfer learning differs from few-shot learning; that is, the model before transfer is for a specific task, whereas the one for few-shot learning has no specific target. A widely investigated area in transfer learning is domain adaptation, where the models before and after transfer handle the same task with different labels and scenes. In [164] and [165], two domain-adaptation networks are developed for pixel and scene classification. In particular, the unsupervised domain adaptation, which need not require any training data in the target domain, has been explored for RS image semantic segmentation [166], [167]. Aside from domain adaptation, studies also investigate conventional transfer learning, in which the target task is different from that for which the model was originally trained. Hu et al. first explored the efficiency of pretrained convolutional features to handle different tasks [70]. On this basis, Yang et al. [169] and He et al. [170] used features from multiple CNNs to enhance transferring capability. Transfer learning has also been adopted for change detection [171], object detection [172], [173], and image retrieval [174].

COMPUTATIONAL INTELLIGENCE

Computational intelligence helps AI algorithms and models find their theoretical optimal solution (or suboptimal in practice), which is crucial for RS data analysis. In the next sections, we discuss the recent progress of computational intelligence in RS image analysis in the following three aspects: evolutionary algorithms (EAs), NAS strategies, and quantum computing.

EAs

RS image analytical problems often cause considerable challenges, such as high-dimensional data, complex data structures, and nonlinear optimization. EAs are regarded as powerful approaches to knowledge discovery [175] and can provide possible solutions to these problems, which are inspired by biological systems [29]. Table 6 lists the representative EAs for RS image analysis.

In the field of endmember extraction (EE), a substantial number of EAs have recently been introduced to consider this problem as a combinational optimization. Zhang et al. previously introduced discrete particle swarm optimization (DPSO) [176] and ant colony optimization algorithms [194] to obtain the minimum root-mean-square error (RMSE) between the original and its remixed image. On the basis of

TABLE 6. A SUMMARY OF EAs FOR RS IMAGE ANALYSIS.

APPLICATIONS	REPRESENTATIVE ALGORITHM	REFERENCE	EAS
Endmember extraction	DPSO	[176]	PSO algorithm
–	QPSO, IQPSO	[177], [178]	QPSO algorithm
–	ADEE	[179]	DE algorithm
–	ACO algorithm	[176]	ACO algorithm
–	MODPSO, improved-version MODPSO	[180], [181]	Multiobjective PSO algorithm
–	$(\mu + \lambda)$-mode	[182]	Multiobjective DE algorithm
–	Bi-MOEE	[183]	MOEA based on decomposition (MOEA/D)
Unmixing	BiPSO	[184]	PSO algorithm
–	MOSU	[185]	Multiobjective cooperative co-EA
–	Tp-MOSU, CM-MOSU	[186], [187]	MOEA/D
–	MOMSM	[188]	Multiobjective memetic algorithm
Band selection	GWO	[189]	GWO algorithm
–	3FA-ELM	[190]	Firefly algorithm
–	PSO-FCM	[191]	PSO algorithm
–	MOBS	[192]	MOEA
–	APBI	[193]	MOEA/D

DPSO: discrete particle swarm optimization; QPSO: quantum-behaved PSO; IQPSO: improved-version QPSO; ADEE: adaptive DE; ACO: ant colony optimization; MOEA: multiobjective EA; MODPSO: multiobject DPSO; Bi-MOEE: bilinear mixture model-based multiobjective EE; BiPSO: biswarm particle swarm optimization; MOSU: multiobjective sparse unmixing; Tp-MOSU: two-phase MOSU; CM-MOSU: classification-based model for MOSU; GWO: gray wolf optimizer; MOBS: multiobjective optimization band selection; APBI: adaptive penalty-based boundary; MOMSM: multiobjective subpixel mapping; 3FA-ELM: firefly-algorithm-inspired framework with band selection and extreme learning machine; PSO-FCM: fuzzy clustering with particle swarm optimization.

quantum behavior, quantum-behaved PSO (QPSO) [177] and its improved-version (I) IQPSO [178] are also suggested. Given that the differential evolution (DE) algorithm is efficient for continuous optimization, an adaptive DE is proposed to achieve RMSE minimization in the feasible domain [179]. Recently, various researchers have considered EE as a multiobjective optimization problem and recommended multiobjective EAs to handle EE, such as a multiobjective DPSO (MODPSO) [180], its improved-version MODPSO [181], and the $(\mu + \lambda)$ multiobjective DE algorithm [182]. In addition, Jiang et al. designed a bilinear mixture model-based multiobjective EE that can simultaneously select real and virtual endmembers from an extended spectral library [183]. More recently, multiobjective PSO algorithms have also been used for endmember bundle extraction [195].

In the hyperspectral RS image unmixing field, Luo et al. utilized a biswarm PSO bilinear unmixing technique to simultaneously estimate endmember signatures and their corresponding abundance [184]. Similar to the aforementioned EE, the unmixing task can also be regarded as a multiobjective optimization; therefore, various multiobjective-based EAs are proposed for unmixing, such as multiobjective sparse unmixing (MOSU) [185], two-phase MOSU [186], and classification-based models for MOSU [187]. After spectral unmixing, the multiobjective optimization technique can be further used for hyperspectral RS image subpixel mapping [188].

In the RS image band selection field, a gray wolf optimizer is offered by designing the fitness function on the basis of class separability measures and accuracy rate [189]. The firefly algorithm is also successfully adopted, and its performance is evaluated through extreme learning machine-based image classification [190]. Moreover, the PSO algorithm has been reported for unsupervised hyperspectral band selection [191]. Similar to the EE and unmixing mentioned previously, multiobjective optimization is also a powerful tool for band selection, as validated in the works on multiobjective optimization band selection [192] and adaptive penalty-based boundary intersection [193].

NAS STRATEGIES

DL has achieved remarkable progress in various RS tasks, but most of the effective neural architectures are manually designed. Meanwhile, a manual search is unfeasible due to the infinite possible choices of network architecture. Accordingly, NAS strategies [196], which automate architectural engineering, have become important research topics in machine learning. Current NAS strategies are divided into three categories, namely, reinforcement learning [196], EA [197], and gradient based [198]. Thus, several NAS methods have been applied to computer vision tasks, such as image denoising [199], image recognition [200], object detection [201], and semantic segmentation [202].

A pioneering method that uses the autodesigned idea for hyperspectral RS image classification is an auto-CNN [203]. The search space of an auto-CNN contains several operations, such as convolutional filters and nonlinear operations. An auto-CNN is searched on the basis of gradient descent to find a suitable deep architecture. Subsequently, Zhang et al. proposed a 3D asymmetric NN search algorithm to automatically search for efficient architectures for hyperspectral imaging classifications [204]. PolSAR-tailored Differentiable Architecture Search promoted polarimetric SAR RS image classification by adopting NAS techniques in which the hyperparameters (i.e., the spatial size and output depth of convolutional kernels) are included in the tailored search space [205].

Recently, studies have also used RS image scene classification based on NAS strategies [206], [207]. RSNet is a RS image recognition framework that adopts a two-stage cascade optimization strategy [208]. A hierarchical, basic search space is designed, including a module-level space for the basic structure block, and a transition-level space for spatial-resolution transformation. Then, a well-performed architecture based on gradient descent is searched. In the training stage, the searched RSNet is retrained to optimize the model's parameters, a switchable recognition module is devised before the retraining stage to enhance the generalization capability of RSNet for different RS recognition tasks (i.e., land and scene cover recognition). These different recognition tasks have diverse output requirements. In the land cover classification task, atrous spatial pyramid pooling is applied to process the output for a segmentation probability map that uses multiscale contextual information. In the scene classification task, the architecture output is processed with global average pooling to obtain a classification probability vector. Such task-driven architecture training made the RSNet generalizable for different RS image recognition tasks. This work can also provide lightweight, deep NN architectures suitable for RS image recognition, which can be further extended to in-orbit satellite data processing and hyperspectral image analysis.

QUANTUM COMPUTING

Quantum algorithms seek ways to accelerate the solution of computational problems using a quantum computer [209]. Given that a qubit is in highly entangled quantum states, quantum computing has potential reasons to surpass classical computing [210]. Therefore, preliminary research has introduced quantum computing into RS data analysis. Du et al. proposed the IQPSO algorithm for EE by combining quantum mechanics with PSO to guarantee algorithm convergence [178]. The authors in [211] leveraged the idea of a statistical ensemble to improve the quality of quantum annealing (QA)-based binary compressive sensing. Otgonbaatar and Datcu attempted to find a quantum algorithmic approach for the feature extraction of texture using a QA computer [212]. In [213], the authors presented two quantum-assisted approaches for the registration of MODIS images, that is, a mapping of image matching to a quadratic-unconstrained binary optimization problem, and a machine learning approach that uses generative models trained using QA statistics that can obtain subpixel accuracy. Pepe et al. sought to develop novel imaging instruments for space objects on the basis of the intrinsic quantum correlation properties of astronomical light sources [214]. Gawron and Lewinski investigated the application of quantum circuit-based NN classifiers for multispectral data classification to obtain land cover information [215]. More recently, a quantum-enhanced deep-learning model is proposed for lithology interpretation [216].

AI EXPLICABILITY

As discussed in the previous sections, advanced AI algorithms may well learn the complicated (e.g., multisource, high-dimensional, and nonlinear) characteristics of RS data. Highly accurate predictions can often be achieved on RS data by modeling large, labeled data sets. However, many data-driven AI algorithms cannot provide a clear physical meaning of the learned features of RS data and lack the theoretical explanation between the input observation and output prediction. In particular, in hyperspectral RS, due to the spectral variabilities caused by various degradation mechanisms, redundancy of high-dimensional signals, and the complex practical cases underlying hyperspectral products, extracting useful diagnostic information to understand the environment is difficult [17]. Therefore, explainable and physics-combined models are crucial to the practical development of AI models in RS data to better characterize more complex real scenes and improve model efficiency through the efficient exploitation of interpretable information.

EXPLAINABLE MODELS

Many machine learning techniques do not demonstrate how data features take effect and why predictions are obtained. Recently, preliminary research has emerged to focus on this issue. Camps-Valls et al. believed that DL necessarily has to become more interpretable and amenable to scrutiny, either by imposing sparse-promoting and knowledge-based priors directly, or by developing hybrid models that explore the subspace of the most physically plausible solutions [217]. In [218], the authors developed an explicitly explainable CNN model that allows assessment of the connections between landscape scenes and the presence of specific landcover types. Meher and Kothari used an interpretable, rule-based, fuzzy, extreme learning machine for RS image classification [219]. Guo et al. suggested a heuristic channel-pruning algorithm for RS image classification and advocated a rectified, interpretable network to provide filters with clarified knowledge representations [220]. The authors in [221] introduced a biologically interpretable two-stage DNN to improve performance on several tasks, including classification and recognition. Li et al. introduced a trainable, soft-attention, mechanism-based, deep-dilated CNN automatic seismic facies analysis. The generated spatial-spectral attention maps revealed a relationship between geological depositions and seismic spectral responses [222].

For crop-yield prediction, Wolanin et al. studied the activation maps of hidden units in CNNs from RS data and found that the learned features mainly related to the length of the growing season and the temperature and light conditions during this period [223]. In [224], the authors focused on understanding LSTM models by looking at the distribution of hidden units, the effect of increasing network complexity, and relative importance of input covariates for crop-yield estimation. Adsuara et al. presented a method that can explicitly describe variable relations by identifying the most expressive and simplest ordinary differential equations explaining data to model-relevant components of the biosphere [225]. Xiong et al. introduced sparsity constrained nonnegative matrix factorization-Net for a hyperspectral unmixing task built by unrolling an Lp sparsity-constrained, nonnegative matrix factorization model belonging to linear-spectral mixture model families, thereby providing high physical interpretability [226].

PHYSICS-COMBINED MODELS

Data-driven AI algorithms may violate key physical constraints and cannot generalize outside of their training set. Therefore, practical researchers still prefer models with a clear physical meaning rather than simply data fitting. Reichstein et al. envisioned various synergies between physical and data-driven models, with the ultimate goal of generating hybrid modeling approaches [1]. In principle, physical approaches are directly interpretable and offer the potential of extrapolation beyond observed conditions, whereas data-driven approaches are highly flexible and amenable to finding unexpected patterns. For geoscientific research, data-driven AI approaches cannot replace, but rather, strongly complement and enrich, physical modeling [1]. In [21], the authors held the same belief that DL cannot completely replace physical models, and that their combination might open a promising door for environmental RS. In practice, Svendsen et al. developed a physics-based interpretable model for Earth-observation time-series modeling, analysis, and understanding by incorporating physical knowledge encoded in differential equations into a multioutput Gaussian process (GP) convolution model [280]. Camps-Valls et al. introduced three different schemes based on GP modeling in the interplay between physics and machine learning, with the focus on the Earth system [227]. The authors in [228] adopted an NN architecture of loss function to enforce physical constraints and improve its generalization ability by leveraging the Clausius–Clapeyron equation to rescale inputs and outputs of parametrization physically. Tian et al. devised an interpretable deep network for variational pansharpening in which all the modules had clear physical meanings (e.g., the prior is based on a nonlinear operator, a data-fidelity term, and the unrolling of the variable-splitting method) [229].

DATA MINING

Recently, the use of data-driven methods in building intelligent systems has been broadly investigated. AI models can achieve encouraging results in many tasks because of the large amount of data collected in various approaches. As is known to all, various machine learning and DL algorithms have different mapping functions between two sets of data. RS AI algorithms are the same. According to different types of data mappings, several successful applications of data mining from RS data are summarized in the following five sections.

SCENE CLASSIFICATION AND OBJECT DETECTION

Scene classification and object detection make up a large set of machine learning tasks in RS intelligent systems, mainly including the mapping from input data—mostly RS images—to predefined classes or categories. Successful applications in different tasks have been found. Khlopenkov and Bedka used matching learning to detect intense convective storms [230] and infrared input images from low-Earth orbiting satellites to develop an improved infrared algorithm with a pattern recognition method. Saad et al. proposed a stacked denoising AE to automatically detect earthquake arrivals [231]. In [232], the authors used a classification network with social media photographs collected from public venues to help in urban land use classification. Liu et al. [233] and Zhang et al. [234] carry out region-of-interest detection based on salient feature extractions and distinctiveness. Ferreira et al. invented a data-driven fusion approach to identify drug crops from RS images [5]. The authors in [6] implemented efficient animal detection in unmanned aerial vehicle (UAV) imagery using deep CNNs and active learning.

DATA RECONSTRUCTION AND INFORMATION RESTORATION

Missing data and absent information are often encountered in RS tasks because of difficulty in accessing high-quality satellite data or other sources. However, with the development of DL, methods have been explored for data reconstruction and information restoration. In such tasks, the mapping is from input data, commonly RS images, to missing values or recovered information for a certain purpose. The typical tasks are population distribution estimation and scene reconstruction. Wang et al. proposed fusing location-based social media data with RS image data for human crowd flood estimation [235]. Heterogeneous social media data are used to build a statistical framework for probabilistic density estimation in RS. Complementary data, such as from social media, can also be used to adjust nighttime light (NTL) images, which record light distribution in an area from an aerial camera. Nonetheless, these raw images can be affected by environmental factors and are difficult to use directly. To address this issue, the authors in [236] recommended the use of social media data, specifically the density of geotagged tweets, to improve the quality of NTL images. Chen et al. provided another example of using additional data in RS. They used volunteered geographic information as an extra data source to label training sets for a deep convolutional network. The extra data are considered weak supervision signals for deep NNs and are used to train the network for refined labeling [237].

ENVIRONMENTAL PARAMETER EXTRACTION

The machine learning task of environmental parameter extraction is used to train a model to summarize specific environmental parameters for an area in RS scenes. The typical environmental parameters include aerosol optical depth, leaf-area index, and so on. In environmental parameter extraction models, the data mapping is from input RS images to specific environmental parameters. Liang and Sun used back-propagation NNs to implement the retrieval of aerosol optical depth from the *Landsat-8* Operational Land Imager's observations over Beijing. The experimental results confirm that the data-driven AI method is superior to the traditional deep blue algorithm in aerosol optical depth extraction [238]. Zhang et al. used hemispherical photos collected by UAVs as input, and combined classifications with stepwise regression and linear regression [239]. The model can automatically summarize the forest leaf-area index values with the combined use of canopy cover and height information.

REGRESSION ANALYSIS

In various RS tasks, regression is often used to predict the future according to the provided data. For example, in temperature forecasting and rainfall prediction, with previous data of rainfall or temperature change, the model finds trends and patterns in time series and then uses these to predict future temperature or rainfall data. Therefore, the mapping of regression tasks occurs between several time-series data. Nimish and Bharath utilized an artificial NN (ANN) architecture for a yearly regression analysis of the land surface temperature time-series data from 1991 to 2017 [240]. From the time series, the ANN model learns the developing pattern of land surface temperature according to years and predicts future temperatures. Similarly, Kusiak et al. also used an ANN for a regression analysis on rainfall time-series data [241]. In authors in [12] analyzed various parameters that affect precipitation in the atmosphere, along with a detailed feature-correlation study. Based on the findings of the feature-correlation study, an optimum set of features are used in a data-driven AI algorithm to obtain precise, hourly rainfall predictions daily.

STATE ESTIMATION AND EVENT LOCALIZATION

Disaster estimation and geographical factor change detection are among the key targets of geosciences. Finding that floods are characterized by social media data, McDougall [242] and Poser and Dransch [243] integrated social media data in flood-estimation models, including digital-elevation models, flood masks, and hydraulic modeling. The benefits from data mining technologies are then demonstrated. Moreover, Cervone et al. utilized additional information from photos, videos, and news in pixel-level classification models (e.g., decision trees) to assist with flood estimation [244]. To achieve a crop-yield estimation, the authors in [8] used spiking NNs, the final output of which was a Shandong province winter wheat crop mask. More recently, Yang et al. [245] combined a crop growth model and an RF algorithm to estimate winter wheat yield.

NLP

The technologies discussed in the previous sections focus on processing and understanding RS images collected using sensors but cannot describe the semantic meaning or contents of RS images from a higher level. Accordingly, the images serve as only indirect sources of knowledge, and the filtering image contents, organizing information, and drawings indicate conclusions that still require human effort from domain experts. In the GRSS, additional direct approaches have been proposed to describe RS image information to alleviate human labor and simplify the understanding of RS images. Language models, which can directly generate sentences in a natural language, have attracted considerable attention. In the following sections, we introduce several tasks that use language models to understand RS images, as structured in Table 7.

DESCRIPTION GENERATION FOR RS IMAGES

Recently, several studies have investigated methods to describe RS images. A fundamental category of these tasks is RS image captioning [246]–[248], specifically for predicting one or a few comprehensive sentences that precisely describe certain elements, especially those of people interested in a given RS image. For this goal, several methods have combined vision models with language ones or prefix templates to produce descriptions.

The methods for generating descriptions for RS images can be divided into three groups. The first is template based, which fills in fixed templates with keywords given by a computer vision model. These methods simplify complex natural language generation and focus on retrieving information from RS images but cannot produce complex and flexible contents. Shi and Zou explored the description of RS images based on templates [247]. Two major differences between natural and RS images (i.e., multilevel semantic and semantic ambiguity) are emphasized and exploited through a multitask learning framework to achieve template-based captioning.

Another group of description-generation models is retrieval based, which aims to determine appropriate descriptions from a candidate pool. Similar to template-based methods, retrieval-based ones are short in generating complex descriptions. Wang et al. determined the inability of previous approaches to model the relationships among objects in images and proposed using retrieval-based techniques to describe the contents in RS images [257]. The image and language inputs are mapped into a shared embedding space to allow retrieval based on distance.

In contrast to the aforementioned groups of methods, template-free approaches comprise another way to generate descriptions for RS images but do not assume any constraints on the sentences to generate; therefore, such methods have better potential for producing complex, diverse, and detailed descriptions. Instead of limiting the solution search space by reducing the variety of outputs, template-free procedures use language-generation models (e.g., RNNs) to directly infer descriptions from image features, thereby producing the output word by word. Nevertheless, the difficulty in these methods is the generation of precise and concrete sentences due to the large solution search space. Qu et al. [246] and Zhang et al. [248] proposed models that caption RS images by using architectures based on the combination of vision and language models. In these models, a convolutional network is used to extract features from RS images, and a natural language description is generated using a recurrent network.

Observing that various RS images can lead to errors or omissions during the FE phase of captioning, thereby affecting the performance, the authors in [258] suggested a denoising-based, multiscale, feature-fusion mechanism to obtain a clear and explicit feature representation and enhance captioning performance. Wang et al. determined that earlier means generate incorrect descriptions due to scene ambiguity in RS image captioning [249] and therefore

TABLE 7. A SUMMARY OF NLP FOR RS DATA ANALYSIS.

RS TASKS	REPRESENTATIVE ALGORITHM	REFERENCE	HIGHLIGHTS
Description generation for RS images	CNN-LSTM model	[246]	Building the first framework for RS image captioning
–	Multitask CNN	[247]	Multitask learning framework for template-based captioning
–	RNNLM	[248]	CNN-LSTM network for RS image captioning
–	Topic-retrieval network	[249]	Topic-retrieval network for RS image captioning
RS image question answering	RSI-VQA	[250]	Directly retrieving answers from RS images
–	Mutual-attention inception network	[251]	Building an RSI-VQA data set
RS image generation with sequential data	RRS-GAN	[252]	Using GANs to recover the RS appearance with a description
–	Audio GAN	[253]	Using audio data to reproduce RS images
RS image with audio sequence	DCM retrieval	[254]	Cross-modal retrieval via sound and image data
–	JLVA retrieval	[255]	Joint retrieval of audio and images
–	CTT recognition	[256]	Sound as an extra information source to assist with recognition

RNNLM: recurrent neural network language model; RSI-VQA: visual question answering for remote sensing data; RRS-GAN: retro-remote sensing generative adversarial networks; DCM: deep cross-modal; JLVA: jointly learning of visual and auditory; CTT: cross-task transfer.

incorporated topic information with image features during the forward process by creating a novel retrieval-topic recurrent memory unit based on an LSTM network. In [259], the authors found that environmental sound can be useful to indicate the attention or topic of an image, then added an extra branch to encode audio information in the previous architecture. More recently, Wang et al. proposed a general word-sentence network with a word extractor and sentence generator for RS image captioning with clear explainability [260].

RS IMAGE QUESTION ANSWERING

Although captioning methods have allowed for automatic descriptions of RS images, these approaches output general results, which summarize image information. However, at times, people focus only on the information of interest, and the general descriptions become uninformative. As a result, RS image question answering has been developed [250], [261], [262]. Question-answering tasks query sentences as input and produce corresponding answers on the basis of the information for RS images. Accordingly, a model can accurately output results that the user cares most for and help filter information. Given that the methods and data sets must be well defined due to the high complexity of visual question-answering (VQA) tasks, its application in RS data has been neglected for a long time. Recently, Lobry et al. recommended describing RS images in the form of question answering because they considered it a direct way to retrieve desired information from RS images. A method is created to generate two RS image VQA data sets and techniques to solve VQA questions and support model development [250]. Zheng et al. also shared a mutual attention inception network and an RSIVQA data set with more than 37,000 images with questions and corresponding answers for RS image VQA [251].

RS IMAGE GENERATION WITH SEQUENTIAL DATA

The inversion of captioning tasks is text-to-image generation, which considers description sentences as inputs to reproduce the visual appearance of target contents and has been shown to be useful in the computer vision community [263]. In the field of RS, several studies have explored RS image-generation methods. Bejiga et al. [252] first used a conditional image generator model to reconstruct RS data from ancient text documents: a recurrent model is used to encode description sentences into a latent space and utilized a GAN to synthesize an image on the basis of the latent code. Zhao and Shi developed a multistage, structured GAN to structurally synthesize RS images, given text descriptions [264]. Another task with a similar pipeline is audio-to-image generation. Specifically, models generate the corresponding images with audio data as inputs. To generate RS images, the authors in [253] suggested first translating audio data into a sequential representation (e.g., sentences) and then producing images using text-to-image-generation models.

RS IMAGE GENERATION WITH AUDIO SEQUENCE

In addition to text documents, audio data are also an important source of extra information that can be used to assist the applications of RS data. In comparison with text data that describe the high-level information already present in RS images, audio data provide current or additional information. Such data add further details to the scene presented in RS images, and consequently, its use is also explored to help understand RS images. At times, RS images are paired with audio data (i.e., sound documents recording the environment presented in RS images). However, image and sound data are usually separately stored due to format limitations. Accordingly, an effective retrieval mechanism is desired. To create such a system, Guo et al. offered a multimodal system that maps audio and image data into the same feature space and retrieves them on the basis of similarities in embedding features [255]. Chen et al. suggested using NNs to map paired image-audio data to the same feature space, and allocated hash-like codes to achieve improved efficiency and minimal quantization errors during the calculation [254]. More recently, Ning et al. proposed semantics-consistent representation learning with the consideration of pairwise, intramodality, and nonpaired intermodality relationships for RS image-voice data sets cross-modal retrieval [265]. Aside from information retrieval, audio information is also used to help improve the models for RS scene recognition. Hu et al. developed ADVANCE, a multimodal data set that provides paired aerial image and audio data, which can benefit the recognition of aerial images in several approaches [256].

AI SECURITY

Security and reliability are important factors when addressing RS tasks. Although AI techniques have remarkably improved the interpretation performance of RS data in the past few decades, its potential risks have been observed. For example, the deployed AI system might be attacked by specific deception algorithms [266], known as *adversarial attacks*, which can generate subtle perturbations imperceptible to a human observer but may greatly mislead state-of-the-art AI systems to make wrong predictions. Correspondingly, algorithms that can help resist such attacks are named *adversarial defenses*. In this section, we first provide a brief review of the development of adversarial attack algorithms. Then, representative adversarial defense methods are introduced. Finally, related research in the RS field is summarized.

ADVERSARIAL ATTACKS

Although DL models have achieved remarkable results in many challenging computer vision tasks, Szegedy et al. [267] first revealed that these state-of-the-art machine learning algorithms are vulnerable to adversarial attacks in image classification tasks. Adversarial examples can be generated by simply adding subtle perturbations to original input images. Although these adversarial examples may look the same as the original clean images to the human vision system, the experimental result in [267] shows that deep NNs can misclassify adversarial examples into an incorrect category with high confidence.

In [268], Xu et al. presented an example of adversarial attacks on DL models for RS scene classification, which demonstrated that the minimal difference between adversarial images and the original ones cannot be detected by the human vision system but can fool deep NNs with high confidence. This phenomenon may undoubtedly limit the practical deployment of DL models in the safety-critical RS field.

The detailed idea of adversarial attacks is first illustrated in [267]. Later, a more efficient algorithm and a few of its extensions are proposed [269], [270]. Although existing literature mainly assumes that adversarial examples can be fed directly into a machine learning classifier, this is not always the case for systems operating in the physical world. In many cases, AI systems may use signals from cameras and other sensors as input. A natural question is thus whether adversarial attacks can also be implemented in real-world scenarios. In [271], Kurakin et al. discovered that machine learning systems are also vulnerable to such adversarial examples, even in physical-world scenarios.

ADVERSARIAL DEFENSES

The defense for adversarial attacks can be grouped into two categories, the first of which is improved training schemes. Adversarial examples can fool state-of-the-art deep NNs because they have never been seen in the training set by classifiers, despite looking almost the same as the original clean training samples to a human observer. A natural adversarial defense is to extend the training set with these generated adversarial examples and train the deep NNs with a mixed clean and adversarial samples. This scheme is known as *adversarial training* [267], [269], which can not only resist adversarial attacks but also help regularize a deep NN [269]. The second category is improved network architectures. Here, recent studies have also attempted to directly improve network architectures to acquire immunity from adversarial attacks. Gu and Rigazio [272] first explored the influence of network topology on adversarial attacks. The representative algorithms in this category include defense-GAN [273] and Jacobian regularization [274].

THREAT OF ADVERSARIAL EXAMPLES IN RS

In a preliminary analysis of adversarial attacks in the context of RS image classification tasks, In [266], the authors discovered that adversarial examples also occur in the field of RS. The experimental results indicate that adversarial attacks on a small patch inside a satellite image can also cause a deep NN to make an incorrect interpretation. Chen et al. showed that the vulnerability of a model to adversarial examples is related to RS image data sets; those with high variety are minimally susceptible to adversarial attacks [275], [276]. Furthermore, the severity of the adversarial problem has a negative relationship with the richness of feature information for optical data [277]. Xu et al. investigated the adversarial training strategy to alleviate the vulnerability of deep models to adversarial examples in RS tasks. Similar to [269], the original clean RS images are first used to train a deep NN. Adversarial examples are then generated using fast gradient sign method. Adversarial training is achieved by simultaneously minimizing the cross-entropy loss from the original clean images and the adversarial examples. The results demonstrate that such a simple training strategy is also effective in increasing the resistibility of deep models to adversarial examples for RS scene classification tasks [268].

Adversarial examples are also found in SAR images. The experimental results in [278] showed that adversarial SAR images are relatively effective in misleading CNN models, as indicated by the obtained high attack success rate. Li et al. proposed a sample boundary-based, adversarial example selectivity distance to explain the attack selectivity of adversarial SAR images [279]. To conduct more efficient adversarial attacks, the authors in [101] further offered an accelerated adversarial attack algorithm for SAR images.

Pioneering research has been carried out on adversarial defenses of RS image classification tasks. Cheng et al. proposed an effective defense framework specified for RS image scene classification in which a new training framework is designed to train the classifier by introducing the generated examples with random and unknown attacks [127]. The authors in [168] presented a soft-threshold-defense method to determine whether or not an image is an adversarial example, and experimental results confirmed that the incorrect prediction rates of a few attack algorithms declined in several scenarios.

CONCLUSIONS AND REMARKS

In this technical review, we systematically discussed state-of-the-art AI techniques in RS data analysis. In detail, this review covered six fundamental facets of AI in RS from the most attractive topic (i.e., machine learning) to recently emerging areas (e.g., NAS strategies, AI explicability, and security). The latest research showed that AI approaches are increasingly used to extract information and analytical data from the continuously growing stream of RS data. However, applications of AI approaches to problems in RS remain in their infancy, and many open questions are worthy of further investigation. The challenges specific to RS tasks are expected to further stimulate the development of methodologies. A few potentially interesting topics include the following:

- *AI approaches for real-world RS tasks*—The following four issues merit notice under this topic: 1) Numerous current studies are built and evaluated on simulated data sets or a limited size of real-world data sets. The accuracy of RS image pixel classification on existing standard data sets has reached 100%, even before the DL era. Nevertheless, the performance remains far from satisfactory when these classification algorithms are applied in practice due to reconsideration of many factors. 2) The current machine learning algorithms are not as robust as expected. For example, various algorithms may work well on the training data set but fail to present a good result on other similar real-world data sets. This situation may be due to parameters that need careful tuning or a certain difference between two data sets that must be addressed. 3) DL algorithms are skilled in fitting data from the observed input to the given ground truth by using hierarchical and complicated mapping functions. The latent RS physical models must also be included for optimal understanding of RS data. 4) In this review, most of the studies focused only on the RS images that can be directly applied to analysis by using CNNs developed by computer vision researchers. However, other heterogeneous data (e.g., social media data in the "Data Mining" section and the caption and audio data in the "NLP" section) may also be explored to emphasize that AI approaches can handle RS big data analysis.

- *Explainable AI algorithms for RS data*—RS big data exacerbate the problems of nontransparency and unexplainability of current AI. These problems are becoming a barrier between the latest AI techniques and several RS applications. Explainable AI is a crucial step toward the practical deployment of AI models in the GRSS. The possible ways forward for RS applications can be explored with explainable AI, such as the following: 1) Establishing hybrid models that incorporate physics and process understanding not only to achieve physical consistency but to also gain credibility with domain users. For example, improving parameterization, such as learning differential equations from data, surrogate modeling and emulation, replacing (and optimizing) layers with mechanistic models, and physical model output calibration with DL, are feasible interactions. In this way, the scientific interplay between theory and observation and between hypothesis generation and theory-driven hypothesis testing can continue. 2) Learning the basic functions in the hidden layers. Several techniques have been developed to examine the interpretability of deep models, such as feature visualization and characterization of the network architecture, feature attribution to analyze how each input contributes to a particular prediction, and model distillation, which explains an NN with a simple model. The analysis of RS systems allows for further understanding and utilization of RS data.
- *Security-related AI approaches for RS data*—To date, awareness has been raised on adversarial examples occurring in the field of RS. Apart from existing literature, possible research directions deserve special attention in the future, including the following: 1) An analysis of different interpretation tasks with various types of data. Although image recognition tasks have been analyzed for red, green, and blue images, adversarial attacks may also threaten semantic segmentation, object detection, and other related tasks. The influence of adversarial examples on other types of RS data thus also deserves further study. 2) Specific adversarial attack/defense algorithms designed for RS data. Given that existing literature directly adopts the traditional adversarial attack/defense methods for RS tasks, effective attacks/defenses may be achieved on the basis of the characteristics of RS data. 3) Adversarial example detection. Another possible coping strategy for adversarial attacks is adversarial example detection, which directly detects whether a given RS sample is attacked by adversarial perturbations. 4) Physical adversarial attacks and defenses. Although adversarial examples are proven to be generated in digital images, how to implement adversarial attacks and defenses in the physical world for RS remains an open issue.

ACKNOWLEDGMENTS

The authors would like to thank the handling editor and anonymous reviewers for their careful reading and helpful remarks. This work was supported by the National Natural Science Foundation of China under grant 62122060 and the special fund of Hubei Luojia Laboratory under grant 220100014. The corresponding author is Liangpei Zhang.

AUTHOR INFORMATION

Lefei Zhang (zhanglefei@whu.edu.cn) received his B.S. and Ph.D. degrees from Wuhan University, China, in 2008 and 2013, respectively. He is a professor with the School of Computer Science, Wuhan University, and also with the Hubei Luojia Laboratory, Wuhan, 430079, China. He was a Big Data Institute visitor with the Department of Statistical Science, University College London, U.K., in 2016, and a Hong Kong scholar with the Department of Computing, Hong Kong Polytechnic University, China, in 2017. He is chair of the IEEE Geoscience and Remote Sensing Society Wuhan Chapter. He serves as an associate editor for *IEEE Geoscience and Remote Sensing Letters*. His research interests include pattern recognition, image processing, and remote sensing. He is a Senior Member of IEEE.

Liangpei Zhang (zlp62@whu.edu.cn) received his B.S. degree from Hunan Normal University, Changsha, China, his M.S. degree from Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, and his Ph.D. degree from Wuhan University, Wuhan, China, in 1982, 1988, and 1998, respectively. He is a Changjiang Scholar Chair Professor, appointed by the Ministry of Education of China, in the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, 430079, China. Previously, he was a principal scientist for the China State Key Basic Research Project (2011–2016) and was appointed by the Ministry of National Science and Technology of China to lead the remote sensing program in China. He has published more than 470 SCI papers in top-tier journals and has been regularly recognized as a highly cited researcher (Clarivate Analytics/Thomson Reuters) in recent years. He was a recipient of the 2020 GRSS David Landgrebe Award and the 2020 IEEE Geoscience and Remote Sensing Society (GRSS) Transactions Prize Paper Award. He was the founding chair of IEEE GRSS Wuhan Chapter. He serves as an associate editor for *IEEE Transactions on Geoscience and Remote Sensing*. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence. He is a Fellow of IEEE.

REFERENCES

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

[2] C. Zhang *et al.*, "An object-based convolutional neural network (OCNN) for urban land use classification," *Remote Sens. Environ.*, vol. 216, pp. 57–70, Oct. 2018, doi: 10.1016/j.rse.2018.06.034.

[3] J. Ma, L. Zhang, and Y. Sun, "ROI extraction based on multiview learning and attention mechanism for unbalanced remote sensing data set," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 6210–6223, 2020, doi: 10.1109/TGRS.2020.2975380.

[4] M. Pullman, I. Gurung, M. Maskey, R. Ramachandran, and S. A. Christopher, "Applying deep learning to hail detection: A case study," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 10,218–10,225, 2019, doi: 10.1109/TGRS.2019.2931944.

[5] A. Ferreira *et al.*, "Eyes in the skies: A data-driven fusion approach to identifying drug crops from remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4773–4786, 2019, doi: 10.1109/JSTARS.2019.2917024.

[6] B. Kellenberger, D. Marcos, S. Lobry, and D. Tuia, "Half a percent of labels is enough: Efficient animal detection in UAV imagery using deep CNNs and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9524–9533, 2019, doi: 10.1109/TGRS.2019.2927393.

[7] B. Chen *et al.*, "Spatio-temporal prediction of leaf area index of rubber plantation using HJ-1A/1B CCD images and recurrent neural network," *ISPRS J. Photogrammetry Remote Sens.*, vol. 102, pp. 148–160, Apr. 2015, doi: 10.1016/j.isprsjprs.2014.12.011.

[8] P. Bose, N. K. Kasabov, L. Bruzzone, and R. N. Hartono, "Spiking neural networks for crop yield estimation based on spatiotemporal analysis of image time series," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6563–6573, 2016, doi: 10.1109/TGRS.2016.2586602.

[9] S. Kolios and N. Hatzianastassiou, "Quantitative aerosol optical depth detection during dust outbreaks from meteosat imagery using an artificial neural network model," *Remote Sens.*, vol. 11, no. 9, p. 1022, 2019, doi: 10.3390/rs11091022.

[10] T. Chin, K. Chen, D. Chen, and D. Lin, "Intelligent real-time earthquake detection by recurrent neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5440–5449, 2020, doi: 10.1109/TGRS.2020.2966012.

[11] H. Gao, Y. Yang, C. Li, L. Gao, and B. Zhang, "Multiscale residual network with mixed depthwise convolution for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3396–3408, 2021, doi: 10.1109/TGRS.2020.3008286.

[12] S. Manandhar, S. Dev, Y. H. Lee, Y. S. Meng, and S. Winkler, "A data-driven approach for accurate rainfall prediction," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9323–9331, 2019, doi: 10.1109/TGRS.2019.2926110.

[13] Y. Lecun, Y. Bengio, and G. E. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[14] L. ZhiYong, T. Liu, J. A. Benediktsson, and N. Falco, "Land cover change detection techniques: Very-high-resolution optical images: A review," *IEEE Geosci. Remote Sens. Mag. (replaces Newslett.)*, early access, 2022, doi: 10.1109/MGRS.2021.3088865.

[15] J. Peng *et al.*, "Low-rank and sparse representation for hyperspectral image processing: A review," *IEEE Geosci. Remote Sens. Mag. (replaces Newslett.)*, early access, 2022, doi: 10.1109/MGRS.2021.3075491.

[16] D. Wen *et al.*, "Change detection from very-high-spatial-resolution optical remote sensing images: Methods, applications, and future directions," *IEEE Geosci. Remote Sens. Mag. (replaces Newslett.)*, early access, 2022, doi: 10.1109/MGRS.2021.3063465.

[17] D. Hong *et al.*, "Interpretable hyperspectral artificial intelligence: When nonconvex modeling meets hyperspectral remote sensing," *IEEE Geosci. Remote Sens. Mag. (replaces Newslett.)*, vol. 9, no. 2, pp. 52–87, 2021, doi: 10.1109/MGRS.2021.3064051.

[18] D. S. Maia, M.-T. Pham, E. Aptoula, F. Guiotte, and S. Lefvre, "Classification of remote sensing data with morphological attribute profiles: A decade of advances," *IEEE Geosci. Remote Sens. Mag. (replaces Newslett.)*, vol. 9, no. 3, pp. 43–71, 2021, doi: 10.1109/MGRS.2021.3051859.

[19] G. Cheng, X. Xie, J. Han, L. Guo, and G.-S. Xia, "Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3735–3756, Jun. 2020, doi: 10.1109/JSTARS.2020.3005403.

[20] 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, doi: 10.1016/j.isprsjprs.2019.11.023.

[21] Q. Yuan *et al.*, "Deep learning in environmental remote sensing: Achievements and challenges," *Remote Sens. Environ.*, vol. 241, no. 111716, pp. 1–24, 2020, doi: 10.1016/j.rse.2020.111716.

[22] B. Rasti, D. Hong, R. Hang, P. Ghamisi, and J. A. Benediktsson, "Feature extraction for hyperspectral imagery: The evolution from shallow to deep (overview and toolbox," *IEEE Geosci. Remote Sens. Mag. (replaces Newslett.)*, vol. 8, no. 4, pp. 60–88, 2020, doi: 10.1109/MGRS.2020.2979764.

[23] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, 2019, doi: 10.1109/TGRS.2019.2907932.

[24] N. Audebert, B. L. Saux, and S. Lefevre, "Deep learning for classification of hyperspectral data: A comparative review," *IEEE Geosci. Remote Sens. Mag. (replaces Newslett.)*, vol. 7, no. 2, pp. 159–173, 2019, doi: 10.1109/MGRS.2019.2912563.

[25] S. Liu, D. Marinelli, L. Bruzzone, and F. Bovolo, "A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges," *IEEE Geosci. Remote Sens. Mag. (replaces Newslett.)*, vol. 7, no. 2, pp. 140–158, 2019, doi: 10.1109/MGRS.2019.2898520.

[26] W. Sun and Q. Du, “Hyperspectral band selection: A review,” *IEEE Geosci. Remote Sens. Mag. (replaces Newslett.)*, vol. 7, no. 2, pp. 118–139, 2019, doi: 10.1109/MGRS.2019.2911100.

[27] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Lin, and B. Johnson, “Deep learning in remote sensing applications: A meta-analysis and review,” *ISPRS J. Photogrammetry Remote Sens.*, vol. 152, pp. 166–177, Jun. 2019, doi: 10.1016/j.isprsjprs.2019.04.015.

[28] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, “Deep learning classifiers for hyperspectral imaging: A review,” *ISPRS J. Photogrammetry Remote Sens.*, vol. 158, pp. 279–317, Dec. 2019, doi: 10.1016/j.isprsjprs.2019.09.006.

[29] Y. Zhong, A. Ma, Y. S. Ong, Z. Zhu, and L. Zhang, “Computational intelligence in optical remote sensing image processing,” *Appl. Soft Comput.*, vol. 64, pp. 75–93, Mar. 2018, doi: 10.1016/j.asoc.2017.11.045.

[30] P. Ghamisi *et al.*, “New frontiers in spectral spatial hyperspectral image classification: The latest advances based on mathematical morphology, Markov random fields, segmentation, sparse representation, and deep learning,” *IEEE Geosci. Remote Sens. Mag. (replaces Newslett.)*, vol. 6, no. 3, pp. 10–43, 2018, doi: 10.1109/MGRS.2018.2854840.

[31] L. He, J. Li, C. Liu, and S. Li, “Recent advances on spectral spatial hyperspectral image classification: An overview and new guidelines,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1579–1597, 2018, doi: 10.1109/TGRS.2017.2765364.

[32] 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, 2017, doi: 10.1109/JPROC.2017.2675998.

[33] J. Li, J. A. Benediktsson, B. Zhang, T. Yang, and A. Plaza, “Spatial technology and social media in remote sensing: A survey,” *Proc. IEEE*, vol. 105, no. 10, pp. 1855–1864, 2017, doi: 10.1109/JPROC.2017.2729890.

[34] X. X. Zhu *et al.*, “Deep learning in remote sensing: A comprehensive review and list of resources,” *IEEE Geosci. Remote Sens. Mag. (replaces Newslett.)*, vol. 5, no. 4, pp. 8–36, 2017, doi: 10.1109/MGRS.2017.2762307.

[35] Y. Gu, J. Chanussot, X. Jia, and J. A. Benediktsson, “Multiple kernel learning for hyperspectral image classification: A review,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6547–6565, 2017, doi: 10.1109/TGRS.2017.2729882.

[36] L. Ma, M. Li, X. Ma, L. Cheng, P. Du, and Y. Liu, “A review of supervised object-based land-cover image classification,” *ISPRS J. Photogrammetry Remote Sens.*, vol. 130, pp. 277–293, Aug. 2017, doi: 10.1016/j.isprsjprs.2017.06.001.

[37] L. Zhang, L. Zhang, and B. Du, “Deep learning for remote sensing data: A technical tutorial on the state of the art,” *IEEE Geosci. Remote Sens. Mag. (replaces Newslett.)*, vol. 4, no. 2, pp. 22–40, 2016, doi: 10.1109/MGRS.2016.2540798.

[38] D. Tuia, C. Persello, and L. Bruzzone, “Domain adaptation for the classification of remote sensing data: An overview of recent advances,” *IEEE Geosci. Remote Sens. Mag. (replaces Newslett.)*, vol. 4, no. 2, pp. 41–57, 2016, doi: 10.1109/MGRS.2016.2548504.

[39] G. Cheng and J. Han, “A survey on object detection in optical remote sensing images,” *ISPRS J. Photogrammetry Remote Sens.*, vol. 117, pp. 11–28, Jul. 2016, doi: 10.1016/j.isprsjprs.2016.03.014.

[40] M. Belgiu and L. Drgu, “Random forest in remote sensing: A review of applications and future directions,” *ISPRS J. Photogrammetry Remote Sens,*, vol. 114, no. 114, pp. 24–31, 2016, doi: 10.1016/j.isprsjprs.2016.01.011.

[41] A. Plaza, J. A. Benediktsson, J. Li, T. Yang, and B. Zhang, “Spatial technology and social media [Scanning the Issue],” *Proc. IEEE*, vol. 105, no. 10, pp. 1851–1854, 2017, doi: 10.1109/JPROC.2017.2733939.

[42] G. Mountrakis, J. Im, and C. Ogole, “Support vector machines in remote sensing: A review,” *ISPRS J. Photogrammetry Remote Sens.*, vol. 66, no. 3, pp. 247–259, 2011, doi: 10.1016/j.isprsjprs.2010.11.001.

[43] D. Lunga, S. Prasad, M. M. Crawford, and O. Ersoy, “Manifold-learning-based feature extraction for classification of hyperspectral data: A review of advances in manifold learning,” *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 55–66, 2014, doi: 10.1109/MSP.2013.2279894.

[44] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, “Random forests for land cover classification,” *Pattern Recog. Lett.*, vol. 27, no. 4, pp. 294–300, 2006, doi: 10.1016/j.patrec.2005.08.011.

[45] Y. Chen, N. M. Nasrabadi, and T. D. Tran, “Hyperspectral image classification using dictionary-based sparse representation,” *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, 2011, doi: 10.1109/TGRS.2011.2129595.

[46] Y. Chen, N. M. Nasrabadi, and T. D. Tran, “Sparse representation for target detection in hyperspectral imagery,” *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 629–640, 2011, doi: 10.1109/JSTSP.2011.2113170.

[47] R. M. Willett, M. F. Duarte, M. A. Davenport, and R. G. Baraniuk, “Sparsity and structure in hyperspectral imaging: Sensing, reconstruction, and target detection,” *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 116–126, 2014, doi: 10.1109/MSP.2013.2279507.

[48] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[49] F. Melgani and L. Bruzzone, “Classification of hyperspectral remote sensing images with support vector machines,” *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, 2004, doi: 10.1109/TGRS.2004.831865.

[50] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, “Exploiting manifold geometry in hyperspectral imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 441–454, 2005, doi: 10.1109/TGRS.2004.842292.

[51] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, “Deep learning-based classification of hyperspectral data,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, 2014, doi: 10.1109/JSTARS.2014.2329330.

[52] Y. Chen, X. Zhao, and X. Jia, “Spectral spatial classification of hyperspectral data based on deep belief network,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2381–2392, 2015, doi: 10.1109/JSTARS.2015.2388577.

[53] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, “Deep feature extraction and classification of hyperspectral images based on convolutional neural networks,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, 2016, doi: 10.1109/TGRS.2016.2584107.

[54] L. Zhu, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Generative adversarial networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5046–5063, 2018, doi: 10.1109/TGRS.2018.2805286.

[55] J. He, L. Zhao, H. Yang, M. Zhang, and W. Li, "HSI-BERT: Hyperspectral image classification using the bidirectional encoder representation from transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 165–178, 2020, doi: 10.1109/TGRS.2019.2934760.

[56] H. Song, Q. Liu, G. Wang, R. Hang, and B. Huang, "Spatiotemporal satellite image fusion using deep convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 821–829, 2018, doi: 10.1109/JSTARS.2018.2797894.

[57] H. Shen, M. Jiang, J. Li, Q. Yuan, Y. Wei, and L. Zhang, "Spatial spectral fusion by combining deep learning and variational model," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 6169–6181, 2019, doi: 10.1109/TGRS.2019.2904659.

[58] Q. Yuan, Q. Zhang, J. Li, H. Shen, and L. Zhang, "Hyperspectral image denoising employing a spatial spectral deep residual convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1205–1218, 2019, doi: 10.1109/TGRS.2018.2865197.

[59] W. Zhao and S. Du, "Spectral spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, 2016, doi: 10.1109/TGRS.2016.2543748.

[60] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional neural networks for large-scale remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 645–657, 2017, doi: 10.1109/TGRS.2016.2612821.

[61] W. Li, G. Wu, F. Zhang, and Q. Du, "Hyperspectral image classification using deep pixel-pair features," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 844–853, 2017, doi: 10.1109/TGRS.2016.2616355.

[62] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral spatial residual network for hyperspectral image classification: A 3-d deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 847–858, 2018, doi: 10.1109/TGRS.2017.2755542.

[63] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "A new deep convolutional neural network for fast hyperspectral image classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 145, pp. 120–147, 2017, doi: 10.1016/j.isprsjprs.2017.11.021.

[64] 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, 2016, doi: 10.1109/TGRS.2016.2601622.

[65] Z. Deng, H. Sun, S. Zhou, J. Zhao, L. Lei, and H. Zou, "Multiscale object detection in remote sensing imagery with convolutional neural networks," *ISPRS J. Photogrammetry Remote Sens.*, vol. 145, pp. 3–22, 2018, doi: 10.1016/j.isprsjprs.2018.04.003.

[66] G. Cheng, Y. Wang, S. Xu, H. Wang, S. Xiang, and C. Pan, "Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3322–3337, 2017, doi: 10.1109/TGRS.2017.2669341.

[67] R. Alshehhi, P. R. Marpu, W. L. Woon, and M. D. Mura, "Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks," *ISPRS J. Photogrammetry Remote Sens.*, vol. 130, pp. 139–149, 2017, doi: 10.1016/j.isprsjprs.2017.05.002.

[68] 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, 2019, doi: 10.1109/TGRS.2018.2849692.

[69] F. Zhang, B. Du, and L. Zhang, "Scene classification via a gradient boosting random convolutional network framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1793–1802, 2016, doi: 10.1109/TGRS.2015.2488681.

[70] F. Hu, G. S. Xia, J. Hu, and L. Zhang, "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," *Remote Sens.*, vol. 7, no. 11, pp. 14,680–14,707, 2015, doi: 10.3390/rs71114680.

[71] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, "When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2811–2821, 2018, doi: 10.1109/TGRS.2017.2783902.

[72] Y. Li, Y. Zhang, X. Huang, H. Zhu, and J. Ma, "Large-scale remote sensing image retrieval by deep hashing neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 950–965, 2018, doi: 10.1109/TGRS.2017.2756911.

[73] H. Shen, X. Meng, and L. Zhang, "An integrated framework for the spatio temporal spectral fusion of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7135–7148, 2016, doi: 10.1109/TGRS.2016.2596290.

[74] B. Huang and H. Song, "Spatiotemporal reflectance fusion via sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3707–3716, 2012, doi: 10.1109/TGRS.2012.2186638.

[75] Q. Yuan, L. Zhang, and H. Shen, "Hyperspectral image denoising employing a spectral spatial adaptive total variation model," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3660–3677, 2012, doi: 10.1109/TGRS.2012.2185054.

[76] M. Lan, Y. Zhang, L. Zhang, and B. Du, "Global context based automatic road segmentation via dilated convolutional neural network," *Inf. Sci.*, vol. 535, pp. 156–171, Oct. 2020, doi: 10.1016/j.ins.2020.05.062.

[77] C. Liu, J. Ma, X. Tang, F. Liu, X. Zhang, and L. Jiao, "Deep hash learning for remote sensing image retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3420–3443, 2021, doi: 10.1109/TGRS.2020.3007533.

[78] X. Chen *et al.*, "Adaptive effective receptive field convolution for semantic segmentation of VHR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3532–3546, 2021, doi: 10.1109/TGRS.2020.3009143.

[79] K. Jin, Y. Chen, B. Xu, J. Yin, and J. Yang, "A patch-to-pixel convolutional neural network for small ship detection with PolSAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 6623–6638, 2020, doi: 10.1109/TGRS.2020.2978268.

[80] S. Wan, C. Gong, P. Zhong, B. Du, L. Zhang, and J. Yang, "Multiscale dynamic graph convolutional network for hyperspectral

image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3162–3177, 2020, doi: 10.1109/TGRS.2019.2949180.

[81] G. E. Hinton, S. Osindero, and Y. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006, doi: 10.1162/neco.2006.18.7.1527.

[82] W. Diao, X. Sun, X. Zheng, F. Dou, H. Wang, and K. Fu, "Efficient saliency-based object detection in remote sensing images using deep belief networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 2, pp. 137–141, 2016, doi: 10.1109/LGRS.2015.2498644.

[83] Q. Zou, L. Ni, T. Zhang, and Q. Wang, "Deep learning-based feature selection for remote sensing scene classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 11, pp. 2321–2325, 2015, doi: 10.1109/LGRS.2015.2475299.

[84] X. Zhang, Y. Liang, C. Li, N. Huyan, L. Jiao, and H. Zhou, "Recursive autoencoders-based unsupervised feature learning for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 1928–1932, 2017, doi: 10.1109/LGRS.2017.2737823.

[85] P. Zhou, J. Han, G. Cheng, and B. Zhang, "Learning compact and discriminative stacked autoencoder for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4823–4833, 2019, doi: 10.1109/TGRS.2019.2893180.

[86] S. Mei, J. Ji, Y. Geng, Z. Zhang, X. Li, and Q. Du, "Unsupervised spatial spectral feature learning by 3d convolutional autoencoder for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6808–6820, 2019, doi: 10.1109/TGRS.2019.2908756.

[87] X. Wang, K. Tan, Y. Chen, and Q. Du, "Hyperspectral image classification based on 3-d octave convolution with spatial spectral attention network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2430–2447, 2021, doi: 10.1109/TGRS.2020.3005431.

[88] F. Zhang, B. Du, and L. Zhang, "Saliency-guided unsupervised feature learning for scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2175–2184, 2015, doi: 10.1109/TGRS.2014.2357078.

[89] Y. Su, J. Li, A. Plaza, A. Marinoni, P. Gamba, and S. Chakravortty, "Daen: Deep autoencoder networks for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4309–4321, 2019, doi: 10.1109/TGRS.2018.2890633.

[90] Y. Qu and H. Qi, "UDAS: An untied denoising autoencoder with sparsity for spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1698–1712, 2019, doi: 10.1109/TGRS.2018.2868690.

[91] Y. Su, A. Marinoni, J. Li, J. Plaza, and P. Gamba, "Stacked nonnegative sparse autoencoders for robust hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 9, pp. 1427–1431, 2018, doi: 10.1109/LGRS.2018.2841400.

[92] F. Khajehrayeni and H. Ghassemian, "Hyperspectral unmixing using deep convolutional autoencoders in a supervised scenario," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 567–576, Feb. 2020, doi: 10.1109/JSTARS.2020.2966512.

[93] B. Palsson, M. O. Ulfarsson, and J. R. Sveinsson, "Convolutional autoencoder for spectral-spatial hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 535–549, 2021, doi: 10.1109/TGRS.2020.2992743.

[94] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[95] Y. Zhan, D. Hu, Y. Wang, and X. Yu, "Semisupervised hyperspectral image classification based on generative adversarial networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 212–216, 2018, doi: 10.1109/LGRS.2017.2780890.

[96] J. Feng, H. Yu, L. Wang, X. Cao, X. Zhang, and L. Jiao, "Classification of hyperspectral images based on multiclass spatial spectral generative adversarial networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5329–5343, 2019, doi: 10.1109/TGRS.2019.2899057.

[97] Q. Xiao, B. Liu, Z. Li, W. Ni, Z. Yang, and L. Li, "Progressive data augmentation method for remote sensing ship image classification based on imaging simulation system and neural style transfer," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 9176–9186, Sep. 2021, doi: 10.1109/JSTARS.2021.3109600.

[98] R. Hang, F. Zhou, Q. Liu, and P. Ghamisi, "Classification of hyperspectral images via multitask generative adversarial networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1424–1436, 2021, doi: 10.1109/TGRS.2020.3003341.

[99] H. Chen, W. Li, and Z. Shi, "Adversarial instance augmentation for building change detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, Mar. 2021, doi: 10.1109/TGRS.2021.3066802.

[100] Y. Yu, X. Li, and F. Liu, "Attention GANs: Unsupervised deep feature learning for aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 519–531, 2020, doi: 10.1109/TGRS.2019.2937830.

[101] C. Du, C. Huo, L. Zhang, B. Chen, and Y. Yuan, "Fast CW: A fast adversarial attack algorithm to fool SAR target recognition with deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Feb. 2021, doi: 10.1109/LGRS.2021.3058011.

[102] W. Xie, Y. Cui, Y. Li, J. Lei, Q. Du, and J. Li, "HPGAN: Hyperspectral pansharpening using 3-d generative adversarial networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 463–477, 2021, doi: 10.1109/TGRS.2020.2994238.

[103] K. Jiang, Z. Wang, P. Yi, G. Wang, T. Lu, and J. Jiang, "Edge-enhanced GAN for remote sensing image superresolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5799–5812, 2019, doi: 10.1109/TGRS.2019.2902431.

[104] M. F. Reyes, S. Auer, N. Merkle, C. Henry, and M. Schmitt, "Sar-to-optical image translation based on conditional generative adversarial networks optimization, opportunities and limits," *Remote Sens.*, vol. 11, no. 17, p. 2067, 2019, doi: 10.3390/rs11172067.

[105] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986, doi: 10.1038/323533a0.

[106] H. Wu and P. Saurabh, "Convolutional recurrent neural networks for hyperspectral data classification," *Remote Sens.*, vol. 9, no. 3, p. 298, 2017, doi: 10.3390/rs9030298.

[107] R. Hang, Q. Liu, D. Hong, and P. Ghamisi, "Cascaded recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5384–5394, 2019, doi: 10.1109/TGRS.2019.2899129.

[108] F. Zhou, R. Hang, Q. Liu, and X. Yuan, "Hyperspectral image classification using spectral spatial LSTMs," *Neurocomputing*, vol. 328, pp. 39–47, Feb. 2019, doi: 10.1016/j.neucom.2018.02.105.

[109] Y. Xu, L. Zhang, B. Du, and F. Zhang, "Spectral spatial unified networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5893–5909, 2018, doi: 10.1109/TGRS.2018.2827407.

[110] X. Zhang, Y. Sun, K. Jiang, C. Li, L. Jiao, and H. Zhou, "Spatial sequential recurrent neural network for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 4141–4155, 2018, doi: 10.1109/JSTARS.2018.2844873.

[111] B. Liu, X. Yu, A. Yu, P. Zhang, and G. Wan, "Spectral spatial classification of hyperspectral imagery based on recurrent neural networks," *Remote Sens. Lett.*, vol. 9, no. 12, pp. 1118–1127, 2018, doi: 10.1080/2150704X.2018.1511933.

[112] A. Song, A. Choi, Y. Han, and Y. Kim, "Change detection in hyperspectral images using recurrent 3d fully convolutional networks," *Remote Sens.*, vol. 10, no. 11, p. 1827, 2018, doi: 10.3390/rs10111827.

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

[114] D. Hong *et al.*, "Spectralformer: Rethinking hyperspectral image classification with transformers," *IEEE Trans. Geosci. Remote Sens.*, early access, 2022, doi: 10.1109/TGRS.2021.3130716.

[115] J. Ding, L. Wen, C. Zhong, and O. Loffeld, "Video SAR moving target indication using deep neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7194–7204, 2020, doi: 10.1109/TGRS.2020.2980419.

[116] J. Li, Y. Li, L. He, J. Chen, and A. Plaza, "Spatio-temporal fusion for remote sensing data: An overview and new benchmark," *Sci. China Inform. Sci.*, vol. 63, no. 4, pp. 1–17, 2020, doi: 10.1007/s11432-019-2785-y.

[117] S. Mohla, S. Pande, B. Banerjee, and S. Chaudhuri, "FusAtNet: Dual attention based spectrospatial multimodal fusion network for hyperspectral and LiDAR classification," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit. Workshops*, 2020, pp. 416–425, doi: 10.1109/CVPRW50498.2020.00054.

[118] "2013 IEEE GRSS data fusion contest," Hyperspectral Image Analysis Lab. https://hyperspectral.ee.uh.edu/?page_id=459

[119] "Multispec—A freeware multispectral image data analysis system," Purdue Univ. https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html

[120] "Hyperspectral remote sensing scenes," Grupo de Inteligencia Computacional (GCE). http://www.ehu.eus/ccwintco/index.php/Hyperspectral_ Remote_ Sensing_ Scenes

[121] "Hydice urban hypercube," U.S. Army Corps of Engineers. http://www.tec.army.mil/Hypercube/

[122] "Aerial hyperspectral remote sensing image dataset of Horseshoe Bay Village in Xiongan." http://www.hrs-cas.com/a/share/shujuchanpin/2019/0501/1049.html

[123] X.-Y. Tong *et al.*, "Land-cover classification with high-resolution remote sensing images using transferable deep models," *Remote Sens. Environ.*, vol. 237, no. 111322, pp. 1–20, 2020, doi: 10.1016/j.rse.2019.111322.

[124] I. Demir *et al.*, "Deepglobe 2018: A challenge to parse the earth through satellite images," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit. Workshops*, 2018, pp. 172–181, doi: 10.1109/CVPRW.2018.00031.

[125] I. Nigam, C. Huang, and D. Ramanan, "Ensemble knowledge transfer for semantic segmentation," in *Proc. IEEE Winter Conf. Appl. Comput. Vision*, 2018, pp. 1499–1508, doi: 10.1109/WACV.2018.00168.

[126] "The sen12ms dataset" https://mediatum.ub.tum.de/1474000

[127] G. Cheng, X. Sun, K. Li, L. Guo, and J. Han, "Perturbation-seeking generative adversarial networks: A defense framework for remote sensing image scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, May 2021, doi: 10.1109/TGRS.2021.3081421.

[128] L. Ru, C. Wu, B. Du, and L. Zhang, "Deep canonical correlation analysis network for scene change detection of multi-temporal VHR imagery," in *Proc. IEEE Int. Workshop Anal. Multitemporal Remote Sens. Images*, 2019, pp. 1–4, doi: 10.1109/MultiTemp.2019.8866943.

[129] M. A. Lebedev, Y. V. Vizilter, O. V. Vygolov, V. A. Knyaz, and A. Y. Rubis, "Change detection in remote sensing images using conditional adversarial networks," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. XLII-2, no. 2, pp. 565–571, 2018, doi: 10.5194/isprs-archives-XLII-2-565-2018.

[130] R. C. Daudt, B. L. 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, doi: 10.1109/IGARSS.2018.8518015.

[131] Z. Dobesova, "Experiment in finding look-alike european cities using urban atlas data," *ISPRS Int. J. Geo-Inform.*, vol. 9, no. 6, pp. 406, 2020, doi: 10.3390/ijgi9060406.

[132] Z. Zou and Z. Shi, "Random access memories: A new paradigm for target detection in high resolution aerial remote sensing images," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1100–1111, 2018, doi: 10.1109/TIP.2017.2773199.

[133] 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, 2017, doi: 10.1109/TGRS.2016.2645610.

[134] S. Razakarivony and F. Jurie, "Vehicle detection in aerial imagery: A small target detection benchmark," *J. Vis. Commun. Image Represent.*, vol. 34, pp. 187–203, Jan. 2016, doi: 10.1016/j.jvcir.2015.11.002.

[135] T. N. Mundhenk, G. Konjevod, W. A. Sakla, and K. Boakye, "A large contextual dataset for classification, detection and counting of cars with deep learning," in *Proc. Conf. European Conf. Comput. Vision*, vol. 9907, pp. 785–800, Sep. 2016, doi: 10.1007/978-3-319-46487-9_48.

[136] G.-S. Xia *et al.*, "DOTA: A large-scale dataset for object detection in aerial images," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 3974–3983, doi: 10.1109/CVPR.2018.00418.

[137] M. Y. Yang, W. Liao, X. Li, and B. Rosenhahn, "Deep learning for vehicle detection in aerial images," in *Proc. IEEE Int. Conf. Image Process.*, 2018, pp. 3079–3083, doi: 10.1109/ICIP.2018.8451454.

[138] Diux xview 2018 detection challenge. [Online]. Available: http://xviewdataset.org/

[139] Z. Liu, L. Yuan, L. Weng, and Y. Yang, "A high resolution optical satellite image dataset for ship recognition and some new baselines," in *Proc. Int. Conf. Pattern Recognit. Appl. Methods*, 2017, pp. 324–331, doi: 10.5220/0006120603240331.

[140] L. Huang *et al.*, "Opensarship: A dataset dedicated to sentinel-1 ship interpretation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 1, pp. 195–208, 2018, doi: 10.1109/JSTARS.2017.2755672.

[141] Q. Wang, S. Liu, J. Chanussot, and X. Li, "Scene classification with recurrent attention of VHR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1155–1167, 2019, doi: 10.1109/TGRS.2018.2864987.

[142] Y. Yang and S. D. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2010, pp. 270–279.

[143] S. Basu, S. Ganguly, S. Mukhopadhyay, R. DiBiano, M. Karki, and R. R. Nemani, "DeepSAT: A learning framework for satellite imagery," *Proc. ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, no. 37, pp. 1–10, 2015.

[144] Q. Zhu, Y. Zhong, B. Zhao, G.-S. Xia, and L. Zhang, "Bag-of-visual-words scene classifier with local and global features for high spatial resolution remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 6, pp. 747–751, 2016, doi: 10.1109/LGRS.2015.2513443.

[145] G.-S. Xia *et al.*, "Aid: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, 2017, doi: 10.1109/TGRS.2017.2685945.

[146] W. Zhou, S. Newsam, C. Li, and Z. Shao, "PatternNet: A benchmark dataset for performance evaluation of remote sensing image retrieval," *ISPRS J. Photogrammetry Remote Sens.*, vol. 145, pp. 197–209, Nov. 2018, doi: 10.1016/j.isprsjprs.2018.01.004.

[147] H. Li *et al.*, "RSI-CB: A large-scale remote sensing image classification benchmark using crowdsourced data," *Sensors*, vol. 20, no. 6, p. 1594, 2020, doi: 10.3390/s20061594.

[148] P. Jin, G.-S. Xia, F. Hu, Q. Lu, and L. Zhang, "AID++: An updated version of aid on scene classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 4721–4724.

[149] RSD46-WHU. [Online]. Available: http://www.lmars.whu.edu.cn/prof_web/xiaozhifeng/dataset.html

[150] G. Sumbul, M. Charfuelan, B. Demir, and V. Markl, "BigEarthNet: A large-scale benchmark archive for remote sensing image understanding," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 5901–5904, doi: 10.1109/IGARSS.2019.8900532.

[151] G. Camps-Valls, T. V. B. Marsheva, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3044–3054, 2007, doi: 10.1109/TGRS.2007.895416.

[152] D. Tuia, M. Volpi, L. Copa, M. Kanevski, and J. Munoz-Mari, "A survey of active learning algorithms for supervised remote sensing image classification," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 606–617, 2011, doi: 10.1109/JSTSP.2011.2139193.

[153] X. Jia, B. Kuo, and M. M. Crawford, "Feature mining for hyperspectral image classification," *Proc. IEEE*, vol. 101, no. 3, pp. 676–697, 2013, doi: 10.1109/JPROC.2012.2229082.

[154] L. Zhang, L. Zhang, D. Tao, and X. Huang, "Sparse transfer manifold embedding for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1030–1043, 2014, doi: 10.1109/TGRS.2013.2246837.

[155] D. Ma, P. Tang, and L. Zhao, "SiftingGAN: Generating and sifting labeled samples to improve the remote sensing image scene classification baseline in vitro," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 7, pp. 1046–1050, 2019, doi: 10.1109/LGRS.2018.2890413.

[156] W. Han, L. Wang, R. Feng, L. Gao, and P. Liu, "Sample generation based on a supervised Wasserstein generative adversarial network for high-resolution remote-sensing scene classification," *Inf. Sci.*, vol. 539, pp. 177–194, Oct. 2020, doi: 10.1016/j.ins.2020.06.018.

[157] M. Schmitt, L. H. Hughes, C. Qiu, and X. X. Zhu, "SEN12MS – A curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion," *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. IV-2/W7, pp. 153–160, 2019, doi: 10.5194/isprs-annals-IV-2-W7-153-2019.

[158] Q. Song, F. Xu, X. X. Zhu, and Y.-Q. Jin, "Learning to generate SAR images with adversarial autoencoder," *IEEE Trans. Geosci. Remote Sens.*, 2022, doi: 10.1109/TGRS.2021.3086817.

[159] B. Liu, X. Yu, A. Yu, P. Zhang, G. Wan, and R. Wang, "Deep few-shot learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2290–2304, 2019, doi: 10.1109/TGRS.2018.2872830.

[160] K. Gao, B. Liu, X. Yu, J. Qin, and X. Tan, "Deep relation network for hyperspectral image few-shot classification," *Remote Sens.*, vol. 12, no. 6, p. 923, 2020, doi: 10.3390/rs12060923.

[161] P. Wang, X. Sun, W. Diao, and K. Fu, "FMSSD: Feature-merged single-shot detection for multiscale objects in large-scale remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3377–3390, 2020, doi: 10.1109/TGRS.2019.2954328.

[162] B. Pradhan, H. A. H. Al-Najjar, I. Maher, I. Tsang, and A. M. Alamri, "Unseen land cover classification from high-resolution orthophotos using integration of zero-shot learning and convolutional neural networks," *Remote Sens.*, vol. 12, no. 10, p. 1676, 2020, doi: 10.3390/rs12101676.

[163] Q. Song and F. Xu, "Zero-shot learning of SAR target feature space with deep generative neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2245–2249, 2017, doi: 10.1109/LGRS.2017.2758900.

[164] E. Othman, Y. Bazi, F. Melgani, H. Alhichri, N. Alajlan, and M. Zuair, "Domain adaptation network for cross-scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4441–4456, 2017, doi: 10.1109/TGRS.2017.2692281.

[165] S. Song, H. Yu, Z. Miao, Q. Zhang, Y. Lin, and S. Wang, "Domain adaptation for convolutional neural networks-based remote sensing scene classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 8, pp. 1324–1328, 2019, doi: 10.1109/LGRS.2019.2896411.

[166] O. Tasar, A. Giros, Y. Tarabalka, P. Alliez, and S. Clerc, "Daugnet: Unsupervised, multisource, multitarget, and life-long domain adaptation for semantic segmentation of satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1067–1081, 2020, doi: 10.1109/TGRS.2020.3006161.

[167] L. Zhang, M. Lan, J. Zhang, and D. Tao, "Stagewise unsupervised domain adaptation with adversarial self-training for road segmentation of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, 2022, doi: 10.1109/TGRS.2021.3104032.

[168] L. Chen, J. Xiao, P. Zou, and H. Li, "Lie to me: A soft threshold defense method for adversarial examples of remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Feb. 2021, doi: 10.1109/LGRS.2021.3096244.

[169] J. Yang, Y. Zhao, and J. C. Chan, "Learning and transferring deep joint spectral spatial features for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4729–4742, 2017, doi: 10.1109/TGRS.2017.2698503.

[170] X. He, Y. Chen, and P. Ghamisi, "Heterogeneous transfer learning for hyperspectral image classification based on convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3246–3263, 2020, doi: 10.1109/TGRS.2019.2951445.

[171] M. Yang, L. Jiao, F. Liu, B. Hou, and S. Yang, "Transferred deep learning-based change detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6960–6973, 2019, doi: 10.1109/TGRS.2019.2909781.

[172] C. Zhou, J. Zhang, J. Liu, C. Zhang, G. Shi, and J. Hu, "Bayesian transfer learning for object detection in optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 7705–7719, 2020, doi: 10.1109/TGRS.2020.2983201.

[173] R. Dong, D. Xu, J. Zhao, L. Jiao, and J. An, "SIG-NMS-based faster R-CNN combining transfer learning for small target detection in VHR optical remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8534–8545, 2019, doi: 10.1109/TGRS.2019.2921396.

[174] Y. Liu, L. Ding, C. Chen, and Y. Liu, "Similarity-based unsupervised deep transfer learning for remote sensing image retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 7872–7889, 2020, doi: 10.1109/TGRS.2020.2984703.

[175] A. E. Eiben and J. Smith, "From evolutionary computation to the evolution of things," *Nature*, vol. 521, no. 7553, pp. 476–482, 2015, doi: 10.1038/nature14544.

[176] B. Zhang, X. Sun, L. Gao, and L. Yang, "Endmember extraction of hyperspectral remote sensing images based on the discrete particle swarm optimization algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4173–4176, 2011, doi: 10.1109/TGRS.2011.2131145.

[177] R. Liu, L. Zhang, and B. Du, "A novel endmember extraction method for hyperspectral imagery based on quantum-behaved particle swarm optimization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 4, pp. 1610–1631, 2017, doi: 10.1109/JSTARS.2016.2640274.

[178] B. Du, Q. Wei, and R. Liu, "An improved quantum-behaved particle swarm optimization for endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 6003–6017, 2019, doi: 10.1109/TGRS.2019.2903875.

[179] Y. Zhong, L. Zhao, and L. Zhang, "An adaptive differential evolution endmember extraction algorithm for hyperspectral remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 6, pp. 1061–1065, 2014, doi: 10.1109/LGRS.2013.2285476.

[180] R. Liu, B. Du, and L. Zhang, "Multiobjective endmember extraction for hyperspectral image," *Remote Sens.*, vol. 9, no. 6, p. 558, 2017, doi: 10.3390/rs9060558.

[181] L. Tong, B. Du, R. Liu, and L. Zhang, "An improved multiobjective discrete particle swarm optimization for hyperspectral endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7872–7882, 2019, doi: 10.1109/TGRS.2019.2917001.

[182] L. Tong, B. Du, R. Liu, L. Zhang, and K. C. Tan, "Hyperspectral endmember extraction by ($\mu + \lambda$) multiobjective differential evolution algorithm based on ranking multiple mutations," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2352–2364, 2021, doi: 10.1109/TGRS.2020.3004307.

[183] X. Jiang, M. Gong, T. Zhan, and M. Zhang, "Multiobjective endmember extraction based on bilinear mixture model," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 8192–8210, 2020, doi: 10.1109/TGRS.2020.2988519.

[184] W. Luo *et al.*, "A new algorithm for bilinear spectral unmixing of hyperspectral images using particle swarm optimization," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 12, pp. 5776–5790, 2016, doi: 10.1109/JSTARS.2016.2602882.

[185] M. Gong, H. Li, E. Luo, J. Liu, and J. Liu, "A multiobjective cooperative coevolutionary algorithm for hyperspectral sparse unmixing," *IEEE Trans. Evol. Comput.*, vol. 21, no. 2, pp. 234–248, 2017, doi: 10.1109/TEVC.2016.2598858.

[186] X. Jiang, M. Gong, H. Li, M. Zhang, and J. Li, "A two-phase multiobjective sparse unmixing approach for hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 508–523, 2018, doi: 10.1109/TGRS.2017.2751060.

[187] X. Xu, Z. Shi, B. Pan, and X. Li, "A classification-based model for multi-objective hyperspectral sparse unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9612–9625, 2019, doi: 10.1109/TGRS.2019.2928021.

[188] M. Song, Y. Zhong, A. Ma, X. Xu, and L. Zhang, "Multiobjective subpixel mapping with multiple shifted hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 8176–8191, 2020, doi: 10.1109/TGRS.2020.2987910.

[189] S. A. Medjahed, T. A. Saadi, A. Benyettou, and M. Ouali, "Gray wolf optimizer for hyperspectral band selection," *Appl. Soft Comput.*, vol. 40, pp. 178–186, Mar. 2016, doi: 10.1016/j.asoc.2015.09.045.

[190] H. Su, Y. Cai, and Q. Du, "Firefly-algorithm-inspired framework with band selection and extreme learning machine for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 1, pp. 309–320, 2017, doi: 10.1109/JSTARS.2016.2591004.

[191] M. Zhang, J. Ma, and M. Gong, "Unsupervised hyperspectral band selection by fuzzy clustering with particle swarm optimization," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 773–777, 2017, doi: 10.1109/LGRS.2017.2681118.

[192] M. Gong, M. Zhang, and Y. Yuan, "Unsupervised band selection based on evolutionary multiobjective optimization for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 544–557, 2016, doi: 10.1109/TGRS.2015.2461653.

[193] B. Pan, Z. Shi, and X. Xu, "Analysis for the weakly pareto optimum in multiobjective-based hyperspectral band selection,"

IEEE Trans. Geosci. Remote Sens., vol. 57, no. 6, pp. 3729–3740, 2019, doi: 10.1109/TGRS.2018.2886853.

[194] B. Zhang, X. Sun, L. Gao, and L. Yang, "Endmember extraction of hyperspectral remote sensing images based on the ant colony optimization (ACO) algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 7, pp. 2635–2646, 2011, doi: 10.1109/TGRS.2011.2108305.

[195] R. Liu and X. Zhu, "Endmember bundle extraction based on multiobjective optimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 10, pp. 8630–8645, 2021, doi: 10.1109/TGRS.2020.3037249.

[196] B. Zoph and Q. V. 1, "Neural architecture search with reinforcement learning," in *Proc. Int. Conf. Learning Representations*, 2017, pp. 1–16, doi: 10.1109/ICAIIC48513.2020.9065031.

[197] E. Real *et al.*, "Large-scale evolution of image classifiers," in *Proc. Int. Conf. Machine Learning*, 2017, pp. 2902–2911.

[198] X. Dong and Y. Yang, "Searching for a robust neural architecture in four GPU hours," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2019, pp. 1761–1770.

[199] H. Zhang, Y. Li, H. Chen, and C. Shen, "Memory-efficient hierarchical neural architecture search for image denoising," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2020, pp. 3657–3666, doi: 10.1109/CVPR42600.2020.00371.

[200] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 8697–8710, doi: 10.1109/CVPR.2018.00907.

[201] C. Jiang, H. Xu, W. Zhang, X. Liang, and Z. Li, "SP-NAS: Serial-to-parallel backbone search for object detection," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2020, pp. 11,863–11,872, doi: 10.1109/CVPR42600.2020.01188.

[202] C. Liu *et al.*, "Auto-deeplab: Hierarchical neural architecture search for semantic image segmentation," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2019, pp. 82–92, doi: 10.1109/CVPR.2019.00017.

[203] Y. Chen, K. Zhu, L. Zhu, X. He, P. Ghamisi, and J. A. Benediktsson, "Automatic design of convolutional neural network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 7048–7066, 2019, doi: 10.1109/TGRS.2019.2910603.

[204] H. Zhang, C. Gong, Y. Bai, Z. Bai, and Y. Li, "3-D-anas: 3-D asymmetric neural architecture search for fast hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, doi: 10.1109/TGRS.2021.3079123.

[205] H. Dong, B. Zou, L. Zhang, and S. Zhang, "Automatic design of CNNs via differentiable neural architecture search for PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 6362–6375, 2020, doi: 10.1109/TGRS.2020.2976694.

[206] C. Broni-Bediako, Y. Murata, L. H. B. Mormille, and M. Atsumi, "Searching for CNN architectures for remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, doi: 10.1109/TGRS.2021.3097938.

[207] A. Ma, Y. Wan, Y. Zhong, J. Wang, and L. Zhang, "SceneNet: Remote sensing scene classification deep learning network using multi-objective neural evolution architecture search," *ISPRS J. Photogrammetry Remote Sens.*, vol. 172, pp. 171–188, Feb. 2021, doi: 10.1016/j.isprsjprs.2020.11.025.

[208] J. Wang, Y. Zhong, Z. Zheng, A. Ma, and L. Zhang, "Rsnet: The search for remote sensing deep neural networks in recognition tasks," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2520–2534, 2021, doi: 10.1109/TGRS.2020.3001401.

[209] A. W. Harrow and A. Montanaro, "Quantum computational supremacy," *Nature*, vol. 549, no. 7671, pp. 203–209, 2017, doi: 10.1038/nature23458.

[210] P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," *SIAM Rev.*, vol. 41, no. 2, pp. 303–332, 1999, doi: 10.1137/S0036144598347011.

[211] R. Ayanzadeh, M. Halem, and T. Finin, "An ensemble approach for compressive sensing with quantum annealers," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 3517–3520, doi: 10.1109/IGARSS39084.2020.9323509.

[212] S. Otgonbaatar and M. Datcu, "Quantum annealing approach: Feature extraction and segmentation of synthetic aperture radar image," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 3692–3695, doi: 10.1109/IGARSS39084.2020.9323504.

[213] C. Pelissier, T. Ames, and J. L. Moigne, "Quantum assisted image registration," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 3696–3699, doi: 10.1109/IGARSS39084.2020.9324720.

[214] F. V. Pepe, A. Scagliola, A. Garuccio, and M. D'Angelo, "Quantum imaging for space objects," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 3700–3702, doi: 10.1109/IGARSS39084.2020.9324462.

[215] P. Gawron and S. Lewinski, "Multi-spectral image classification with quantum neural network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 3513–3516, doi: 10.1109/IGARSS39084.2020.9323065.

[216] N. Liu, T. Huang, J. Gao, Z. Xu, D. Wang, and F. Li, "Quantum-enhanced deep learning-based lithology interpretation from well logs," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, Jun. 2021, doi: 10.1109/TGRS.2021.3085340.

[217] G. Camps-Valls, M. Reichstein, X. Zhu, and D. Tuia, "Advancing deep learning for earth sciences: From hybrid modeling to interpretability," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 3979–3982, doi: 10.1109/IGARSS39084.2020.9323558.

[218] A. Levering, D. Marcos, S. Lobry, and D. Tuia, "Interpretable scenicness from sentinel-2 imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 3983–3986, doi: 10.1109/IGARSS39084.2020.9323706.

[219] S. K. Meher and N. S. Kothari, "Interpretable rule-based fuzzy elm and domain adaptation for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5907–5919, 2021, doi: 10.1109/TGRS.2020.3024796.

[220] 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, 2022, doi: 10.1109/TGRS.2021.3077062.

[221] Y. Shi *et al.*, "A biologically interpretable two-stage deep neural network (BIT-DNN) for vegetation recognition from hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, 2022.

[222] F. Li, H. Zhou, Z. Wang, and X. Wu, "ADDCNN: An attention-based deep dilated convolutional neural network for seismic facies analysis with interpretable spatialspectral maps," *IEEE*

Trans. Geosci. Remote Sens., vol. 59, no. 2, pp. 1733–1744, 2021, doi: 10.1109/TGRS.2020.2999365.

[223] A. Wolanin *et al.*, "Estimating and understanding crop yields with explainable deep learning in the Indian wheat belt," *Environ. Res. Lett.*, vol. 15, no. 2, 2020, doi: 10.1088/1748-9326/ab68ac.

[224] A. Perez-Suay *et al.*, "Interpretability of recurrent neural networks in remote sensing," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 3991–3994, doi: 10.1109/IGARSS39084.2020.9323898.

[225] J. E. Adsuara *et al.*, "Discovering differential equations from earth observation data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 3999–4002, doi: 10.1109/IGARSS39084.2020.9324639.

[226] F. Xiong, J. Zhou, S. Tao, J. Lu, and Y. Qian, "SNMF-Net: Learning a deep alternating neural network for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, May 2021, doi: 10.1109/TGRS.2021.3081177.

[227] G. Camps-Valls *et al.*, "Physics-aware Gaussian processes in remote sensing," *Appl. Soft Comput.*, vol. 68, pp. 69–82, Jul. 2018, doi: 10.1016/j.asoc.2018.03.021.

[228] T. Beucler, M. Pritchard, P. Gentine, and S. Rasp, "Towards physically-consistent, data-driven models of convection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 3987–3990, doi: 10.1109/IGARSS39084.2020.9324569.

[229] X. Tian, K. Li, Z. Wang, and J. Ma, "VP-Net: An interpretable deep network for variational pansharpening," *IEEE Trans. Geosci. Remote Sens.*, 2022, doi: 10.1109/TGRS.2021.3089868.

[230] K. V. Khlopenkov and K. M. Bedka, "Development of pattern recognition algorithms to detect intense convective storms from multispectral satellite imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 4850–4852, doi: 10.1109/IGARSS.2018.8518596.

[231] O. M. Saad, K. Inoue, A. Shalaby, L. Samy, and M. S. Sayed, "Automatic arrival time detection for earthquakes based on stacked denoising autoencoder," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 11, pp. 1687–1691, 2018, doi: 10.1109/LGRS.2018.2861218.

[232] F. Fang, X. Yuan, L. Wang, Y. Liu, and Z. Luo, "Urban land-use classification from photographs," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 12, pp. 1927–1931, 2018, doi: 10.1109/LGRS.2018.2864282.

[233] G. Liu, L. Qi, Y. Tie, and L. Ma, "Region-of-interest detection based on statistical distinctiveness for panchromatic remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 2, pp. 271–275, 2019, doi: 10.1109/LGRS.2018.2870935.

[234] L. Zhang, X. Lv, J. Chen, and L. Zhang, "Region of interest detection based on salient feature clustering for remote sensing images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 88–91, doi: 10.1109/IGARSS.2016.7729013.

[235] H. Wang, E. Skau, H. Krim, and G. Cervone, "Fusing heterogeneous data: A case for remote sensing and social media," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 12, pp. 6956–6968, 2018, doi: 10.1109/TGRS.2018.2846199.

[236] N. Zhao, W. Zhang, Y. Liu, E. L. Samson, Y. Chen, and G. Cao, "Improving nighttime light imagery with location-based social media data," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2161–2172, 2019, doi: 10.1109/TGRS.2018.2871788.

[237] J. Chen, Y. Zhou, A. Zipf, and H. Fan, "Deep learning from multiple crowds: A case study of humanitarian mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1713–1722, 2019, doi: 10.1109/TGRS.2018.2868748.

[238] T. Liang and L. Sun, "Retrieval of aerosol optical depth (AOD) from the Landsat8 Oli observations over Beijing," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 5493–5496, doi: 10.1109/IGARSS39084.2020.9323832.

[239] D. Zhang, J. Liu, W. Ni, G. Sun, Z. Zhang, Q. Liu, and Q. Wang, "Estimation of forest leaf area index using height and canopy cover information extracted from unmanned aerial vehicle stereo imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 2, pp. 471–481, 2019, doi: 10.1109/JSTARS.2019.2891519.

[240] G. Nimish and H. A. Bharath, "Forecasting land surface temperature using artificial neural network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 4387–4390, doi: 10.1109/IGARSS39084.2020.9323745.

[241] A. Kusiak, X. Wei, A. P. Verma, and E. Roz, "Modeling and prediction of rainfall using radar reflectivity data: A data-mining approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2337–2342, 2013, doi: 10.1109/TGRS.2012.2210429.

[242] K. McDougall, "Using volunteered information to map the Queensland floods," in *Proc. Surveying Spatial Sci. Conf.: Innov. Action: Working Smarter*, 2011, pp. 13–23.

[243] K. Poser and D. Dransch, "Volunteered geographic information for disaster management with application to rapid flood damage estimation," *Geomatica*, vol. 64, no. 1, pp. 89–98, 2010.

[244] G. Cervone, E. Sava, Q. Huang, E. Schnebele, J. Harrison, and N. Waters, "Using twitter for tasking remote-sensing data collection and damage assessment: 2013 boulder flood case study," *Int. J. Remote Sens.*, vol. 37, no. 1, pp. 100–124, 2016, doi: 10.1080/01431161.2015.1117684.

[245] S. Yang *et al.*, "Integration of crop growth model and random forest for winter wheat yield estimation from UAV hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6253–6269, 2021, doi: 10.1109/JSTARS.2021.3089203.

[246] B. Qu, X. Li, D. Tao, and X. Lu, "Deep semantic understanding of high-resolution remote sensing image," in *Proc. Int. Conf. Comput., Inf. Telecommun. Syst.*, 2016, pp. 1–5, doi: 10.1109/CITS.2016.7546397.

[247] Z. Shi and Z. Zou, "Can a machine generate humanlike language descriptions for a remote sensing image?" *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3623–3634, 2017, doi: 10.1109/TGRS.2017.2677464.

[248] X. Zhang, X. Li, J. An, L. Gao, B. Hou, and C. Li, "Natural language description of remote sensing images based on deep learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 4798–4801, doi: 10.1109/IGARSS.2017.8128075.

[249] B. Wang, X. Zheng, B. Qu, and X. Lu, "Retrieval topic recurrent memory network for remote sensing image captioning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 256–270, Feb. 2020, doi: 10.1109/JSTARS.2019.2959208.

[250] S. Lobry, D. Marcos, J. Murray, and D. Tuia, "RSVQA: Visual question answering for remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8555–8566, 2020, doi: 10.1109/TGRS.2020.2988782.

[251] X. Zheng, B. Wang, X. Du, and X. Lu, "Mutual attention inception network for remote sensing visual question answering,"

IEEE Trans. Geosci. Remote Sens., 2022, doi: 10.1109/TGRS.2021.3079918.

[252] M. B. Bejiga, F. Melgani, and A. Vascotto, "Retro-remote sensing: Generating images from ancient texts," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 3, pp. 950–960, 2019, doi: 10.1109/JSTARS.2019.2895693.

[253] Z. Zheng, J. Chen, X. Zheng, and X. Lu, "Remote sensing image generation from audio," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 6, pp. 994–998, 2021, doi: 10.1109/LGRS.2020.2992324.

[254] Y. Chen, X. Lu, and S. Wang, "Deep cross-modal image-voice retrieval in remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7049–7061, 2020, doi: 10.1109/TGRS.2020.2979273.

[255] M. Guo, C. Zhou, and J. Liu, "Jointly learning of visual and auditory: A new approach for RS image and audio cross-modal retrieval," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 11, pp. 4644–4654, 2019, doi: 10.1109/JSTARS.2019.2949220.

[256] D. Hu *et al.*, "Cross-task transfer for geotagged audiovisual aerial scene recognition," in *Proc. European Conf. Comput. Vision*, 2020, pp. 68–84.

[257] B. Wang, X. Lu, X. Zheng, and X. Li, "Semantic descriptions of high-resolution remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 8, pp. 1274–1278, 2019, doi: 10.1109/LGRS.2019.2893772.

[258] W. Huang, Q. Wang, and X. Li, "Denoising-based multiscale feature fusion for remote sensing image captioning," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 3, pp. 436–440, 2021, doi: 10.1109/LGRS.2020.2980933.

[259] X. Lu, B. Wang, and X. Zheng, "Sound active attention framework for remote sensing image captioning," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1985–2000, 2020, doi: 10.1109/TGRS.2019.2951636.

[260] Q. Wang, W. Huang, X. Zhang, and X. Li, "Word-sentence framework for remote sensing image captioning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10,532–10,543, 2021, doi: 10.1109/TGRS.2020.3044054.

[261] S. Lobry, J. Murray, D. Marcos, and D. Tuia, "Visual question answering from remote sensing images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 4951–4954, doi: 10.1109/TGRS.2020.2988782.

[262] A. V. Potnis, R. C. Shinde, and S. S. Durbha, "Towards natural language question answering over earth observation linked data using attention-based neural machine translation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 577–580, doi: 10.1109/IGARSS39084.2020.9323183.

[263] E. Mansimov, E. Parisotto, J. L. Ba, and R. Salakhutdinov, "Generating images from captions with attention," in *Proc. Int. Conf. Learning Representations*, 2016, pp. 1–12.

[264] R. Zhao and Z. Shi, "Text-to-remote-sensing-image generation with structured generative adversarial networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Mar. 2021, doi: 10.1109/LGRS.2021.3068391.

[265] H. Ning, B. Zhao, and Y. Yuan, "Semantics-consistent representation learning for remote sensing image-voice retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Mar. 2021, doi: 10.1109/TGRS.2021.3060705.

[266] W. Czaja, N. Fendley, M. Pekala, C. Ratto, and I.-J. Wang, "Adversarial examples in remote sensing," in *Proc. ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2018, pp. 408–411.

[267] C. Szegedy *et al.*, "Intriguing properties of neural networks," in *Proc. Int. Conf. Learning Representations*, 2014, pp. 1–10.

[268] Y. Xu, B. Du, and L. Zhang, "Assessing the threat of adversarial examples on deep neural networks for remote sensing scene classification: Attacks and defenses," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1604–1617, 2021, doi: 10.1109/TGRS.2020.2999962.

[269] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in *Proc. Int. Conf. Learning Representations*, 2015, pp. 1–11.

[270] T. Miyato, S. Maeda, M. Koyama, and S. Ishii, "Virtual adversarial training: A regularization method for supervised and semi-supervised learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 1979–1993, 2019, doi: 10.1109/TPAMI.2018.2858821.

[271] A. Kurakin, I. Goodfellow, and S. Bengio, "Adversarial examples in the physical world," in *Proc. Int. Conf. Learning Representations*, 2017, pp. 1–14.

[272] S. Gu and L. Rigazio, "Towards deep neural network architectures robust to adversarial examples," in *Proc. Int. Conf. Learning Representations*, 2015, pp. 1–9.

[273] P. Samangouei, M. Kabkab, and R. Chellappa, "Defense-GAN: Protecting classifiers against adversarial attacks using generative models," in *Proc. Int. Conf. Learning Representations*, 2018, pp. 1–17.

[274] J. Hoffman, D. A. Roberts, and S. Yaida, "Robust learning with Jacobian regularization," in *Proc. Int. Conf. Learning Representations*, 2020, pp. 1–21.

[275] L. Chen, G. Zhu, Q. Li, and H. Li, "Adversarial example in remote sensing image recognition," 2019, *arXiv:1910.13222 [cs.CV]*.

[276] L. Chen *et al.*, "Attack selectivity of adversarial examples in remote sensing image scene classification," *IEEE Access*, vol. 8, pp. 137,477–137,489, Jul. 2020, doi: 10.1109/ACCESS.2020.3011639.

[277] L. Chen, Z. Xu, Q. Li, J. Peng, S. Wang, and H. Li, "An empirical study of adversarial examples on remote sensing image scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7419–7433, 2021, doi: 10.1109/TGRS.2021.3051641.

[278] T. Huang, Q. Zhang, J. Liu, R. Hou, X. Wang, and Y. Li, "Adversarial attacks on deep-learning-based sar image target recognition," *J. Netw. Comput. Appl.*, vol. 162, no. 102632, pp. 1–12, 2020, doi: 10.1016/j.jnca.2020.102632.

[279] H. Li *et al.*, "Adversarial examples for CNN-based SAR image classification: An experience study," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1333–1347, 2020, doi: 10.1109/JSTARS.2020.3038683.

[280] D. H. Svendsen, M. Piles, J. Muñoz-Marí, D. Luengo, L. Martino and G. Camps-Valls, "Integrating domain knowledge in data-driven earth observation with process cConvolutions," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2022, Art no. 4401715, doi: 10.1109/TGRS.2021.3059550.

GRS