



Deep learning for processing and analysis of remote sensing big data: a technical review

Xin Zhang, Ya'nan Zhou & Jiancheng Luo

To cite this article: Xin Zhang, Ya'nan Zhou & Jiancheng Luo (2021): Deep learning for processing and analysis of remote sensing big data: a technical review, Big Earth Data, DOI: [10.1080/20964471.2021.1964879](https://doi.org/10.1080/20964471.2021.1964879)

To link to this article: <https://doi.org/10.1080/20964471.2021.1964879>



© 2021 The Author(s). Published by Taylor & Francis Group and Science Press on behalf of the International Society for Digital Earth, supported by the CASEarth Strategic Priority Research Programme.



Published online: 30 Aug 2021.



Submit your article to this journal [↗](#)



Article views: 2652





View related articles [↗](#)



View Crossmark data [↗](#)

Deep learning for processing and analysis of remote sensing big data: a technical review

Xin Zhang ^a, Ya'nan Zhou ^b and Jiancheng Luo^a

^aState Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China; ^bCollege of Hydrology and Water Resources, Hohai University, Nanjing, China

ABSTRACT

In recent years, the rapid development of Earth observation technology has produced an increasing growth in remote sensing big data, posing serious challenges for effective and efficient processing and analysis. Meanwhile, there has been a massive rise in deep-learning-based algorithms for remote sensing tasks, providing a large opportunity for remote sensing big data. In this article, we initially summarize the features of remote sensing big data. Subsequently, following the pipeline of remote sensing tasks, a detailed and technical review is conducted to discuss how deep learning has been applied to the processing and analysis of remote sensing data, including geometric and radiometric processing, cloud masking, data fusion, object detection and extraction, land-use/cover classification, change detection and multitemporal analysis. Finally, we discussed technical challenges and concluded directions for future research in deep-learning-based applications for remote sensing big data.

ARTICLE HISTORY

Received 2 June 2021
Accepted 28 July 2021



KEYWORDS

Remote sensing; big data; deep learning; technical review

1. Introduction

Remote sensing (RS) is one of the most important methods for observing the Earth's surface, and plays an essential role in many fields, such as climate change (Yang et al., 2013), land and resource surveys (Zhou, Luo, Shen, Hu, & Yang, 2014), disaster monitoring and assessment (Rahman & Di, 2017), crop growth monitoring (Zhou et al., 2019), urbanization (Zhang & Huang, 2018b), and land-use/cover changes (Halefom, Teshome, Sisay, & Ahmad, 2018).

During recent decades, the rapid development of satellite and sensor technology has led to the explosive growth of data from various platforms and sensors, driving us into the era of remote sensing big data (RSBD). RSBD helps us to extend the domain of RS applications and presents new challenges for its processing and analysis, for example, how to extract global farmland accurately, quickly, and automatically from RSBD. On the other hand, the development of RS is accompanied by advancements in information technology and artificial intelligence. RS scientists introduce machine learning algorithms (such as artificial neural networks, support vector machines and random forests) to

CONTACT Ya'nan Zhou  zhouyn@hhu.edu.cn  College of Hydrology and Water Resources, Hohai University, Nanjing, China

© 2021 The Author(s). Published by Taylor & Francis Group and Science Press on behalf of the International Society for Digital Earth, supported by the CASEarth Strategic Priority Research Programme.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

improve the performance of RS applications. In particular, recently, there has been tremendous development in intelligent technology represented by deep learning which has wide applications in many fields. With the great ability of hierarchical learning of representative and discriminative features, deep learning has instigated a new wave of promising research on RSBD processing and analysis (Ma et al., 2019a; Zhang, Zhang, & Du, 2016a; Zhu et al., 2017).

Challenges and opportunities coexisting, deep learning techniques have been successful in almost all areas of RS applications. Until now, there have been few studies concerning reviews of deep learning applications for remote sensing (Zhang et al., 2016a; Zhu et al., 2017; Li, Zhang, & Xue et al., 2018bb; Ma et al., 2019a), but these reviews focused on the technical pertinence of RSBD and ignored technical analysis of the advantages and disadvantages of deep learning algorithms. In particular, new deep learning models (such as graph neural networks) further extend the remote sensing application fields of deep learning techniques. RSBD promotes multitemporal analysis in tasks related to the monitoring of Earth surface dynamics, a field which was ignored in previous reviews (Ma et al., 2019a). With an increasing number of novel studies on deep learning methods for RSBD, it appears that a more systematic analysis is necessary to obtain a technical understanding of the processing and analysis of RSBD.

Therefore, the motivation for this study was to conduct a technical review of almost all of the major and promising subareas in deep-learning-based processing and analysis of RSBD. The remainder of the paper is structured as follows: [Section 2](#) provides a technical overview of RSBD. [Section 3](#) recalls the general deep learning architectures or networks. [Section 4](#) outlines how to apply deep learning algorithms to RS applications, following a technical overview of the processing and analysis of RSBD. [Section 5](#) discusses the advantages and disadvantages of deep learning algorithms for RSBD. [Section 6](#) concludes the paper with a discussion on potential directions for further deep-learning-based studies for RSBD.

2. Remote sensing big data

The rapid development of sensor technology has accelerated the development of RS. Various sensors on satellites, space shuttles, unmanned aerial vehicles (UAVs), and ground observation stations have created space-air-ground Earth observation systems. First, due to the enormous demands for large-region and precise environment and resource applications, many satellites and constellations have been launched, such as the Landsat series of the United States, the Copernicus project of the European Union, and the major project of the National High-resolution Earth Observation System of China (Gaofen). Furthermore, private capital and commercial players are on the march toward RS applications. For example, Planet has launched approximately 200 small satellites,¹ producing a large amount of RS data every day. The optical Sentinel-2 A and B satellites alone produce approximately 9.54 TB of data on average per day according to the acquisition plan.² Second, due to its unique advantages, e.g. flexibility, high spatial resolution and data acquisition on demand, aerial RS has been recognized as an effective complement to traditional space-based RS platforms. The recent advance of UAV technology with small-sized and high-detection-precision sensors makes UAV-based RS a very popular and increasingly used technology, which also produces massive high-resolution images.

Third, owing to the development of new technologies in smartphones and wireless networks, an increasing number of ground observation stations have been established. They produce high-frequency on-the-spot observations, which further enriches the RS data sources. Generally, space-air-ground observation systems provide massive, multi-source, multimodal, multiscale, high-dimensional, dynamic-state, and heterogeneous RSBDs (Liu, 2015). Table 1 presents some common data sources of RSBD.

From big data, RSBD inherits the four “V” features (referred hereinafter as 4Vs): volume, variety, velocity, and value. In addition, RSBD can be described by its dimensions.

- Volume: An increasing number of sensors on satellites, space shuttles, and the ground are continuously observing Earth’s surface, accumulating massive RS data.
- Variety: RSBD consists of data from multiple sources (satellite, UAV, ground, etc.), and multimodal (LiDAR, radar, optical, etc.), multiresolution (from 1 cm to 100 km), and multitemporal (collected on irregular dates and periods) data, as well as data from different disciplines depending on application domains.

Table 1. Some common data sources of remote sensing big data.

Category	Data Source	Spatial (m)	Duration (year) Temporal (day)	Spectral
Satellite (optical)	Landsat 1-8	15 – 120	1972 – now 16 – 18	4 – 11
	MODIS	250 – 1000	1999 – now 1 – 2	36
	SPOT 1-7	1.5 – 20	1987 – now 26	4 – 5
	Sentinel-2	10 – 60	2015 – now 5 – 10	13
	Worldview 1-4	0.31 – 1.64	2007 – now 1 – 3	1 – 17
	HJ-1 (China)	30 – 300	2008 – now 4	4 – 8 ^a
	GaoFen-1/6 (China)	2 – 16	2013 – now 4	4 – 9
Satellite (SAR)	Sentinel-1	5 – 20	2014 – now 6 – 12	Single & Dual
	COSMO-SkyMed	1 – 15	2007 – now 1 – 16	
	RADASAT 1-2	3 – 100	1995 – now 1 – 24	Single & Dual & Quad
	ALOS-1-2	3 – 100	2006 – now 14 – 46	
	TerraSAR-X	1 – 16	2007 – now 11	
	GanFen-3 (China)	1 – 500	2016 – now <3	
UAV	Multispectral camera	0.02 – 0.2 m	on demand	3 – 4
Sensor Networks	LiDAR	/		1
	Hydrology	/	on demand	/
	(level, flow, sediment, precipitation, evaporation)			
	Meteorology	/	on demand	/
	(temperature, humidity, pressure, wind, cloud)			

^aNo consideration on the hyperspectral sensor on HJ-1 satellite

- **Velocity:** Earth surface, the target of Earth observation is changing, and RS data are generated at a rapid growth rate. Furthermore, an increasing number of tasks require (near) real-time processing and analysis to produce the latest information for decision support, e.g. hours can save hundreds of hectares of forest in a forest fire.
- **Value:** “Value” is an inherent quality of big data. Information retrieved from RSBD can predict changes in global climate, guide agricultural planting, assist urban planning, etc.

In addition to the common 4Vs features of general big data, RSBD presents the following characteristics:

- RS data are multiresolution and multiscale. Due to various sensor capabilities and sensing distances, the resolution of RS data ranges from hundreds of kilometres to a few centimetres, bringing an appreciable scale effect for processing and analysis.
- RS data are often multimodal, e.g. from optical (multi and hyperspectral), LiDAR, and synthetic aperture radar (SAR) sensors, where the imaging geometries and content are completely different.
- The time variable is becoming increasingly important. In addition to the pursuit of high spatial and spectral resolutions in the last century, the demand for high temporal resolution has recently increased. For example, constellations (e.g. satellites A and B of Sentinel-1) or virtual constellations (e.g. Landsat 8 and Sentinel-2) consisting of several similar satellites can achieve shorter revisit cycles. On the other hand, an increasing number of tasks put forward higher demands on imaging timeliness, such as crop monitoring and disaster assessment.
- From qualification (such as image classification and object extraction), RS applications advance toward quantification (such as parameter inversion and crop yield estimation).

These characteristics of RSBD make accurate, quick, and automatic processing and analysis of RDBD challenging for remote sensing applications.

3. General deep learning algorithms

First appearing in the 1960s, the perceptron was the basis of the earliest neural networks. It is a bioinspired model for binary classification that aims to mathematically formalize how a biological neuron works. Unfortunately, the perceptron cannot perform nonlinear classification. The backpropagation algorithm was beginning to train a multilayer artificial neural network (NN) in the 1980s, promoting tremendous development of NNs (Rumelhart, Hinton, & Williams, 1985). The NN can solve nonlinear problems, but it usually contains only one hidden layer (and is thus referred to as a shallow NN) because of difficulties in training a multilayer NN. By 2006, more sophisticated methodologies had been proposed to train the NN with more than one hidden layer (referred to as a deep NN) (Hinton & Salakhutdinov, 2006). By then, NNs had entered the era of deep learning.

Because regular grid pixels are suitable for convolution operations, convolutional neural networks (CNNs) have been successful in image processing tasks. Due to its ability to capture long-term dependencies, recurrent neural networks (RNNs) have been applied to sequential data. As an unsupervised framework, autoencoders (AEs) can be stacked to

Table 2. List of popular deep learning architectures in remote sensing tasks.

Model	Features and advantages	Variants or implementation	Tasks and references
CNN	Image processing and analysis	LeNet, R-CNN, VggNet, ResNet, U-Net	geometric processing (Wang et al., 2018d), cloud masking (Wu et al., 2020a), object extraction (Abdollahi, Pradhan, Shukla, Chakraborty, & Alamri, 2020), semantic segmentation (Waldner & Diakogiannis, 2020)
RNN	Sequential data and time-series analysis	LSTM, BiLSTM, GRU	sequential prediction (Zhou, Yang, & Feng et al., 2020), sequential classification (Zhou et al., 2019), change detection (Lyu, Lu, & Mou, 2016)
AE	Feature extraction and representation	Stacked AE, Sparse AE, Variational AE, Denoising AE	feature extraction (Shao et al., 2017; Zhou et al., 2020; Xing, Wang, Yang, & Jiao, 2018; Huang, Xiao, Wei, Liu, & Tang, 2015)
GAN	Data generation, unsupervised learning	Conditional GAN, f-divergence GAN, Cycle-consistent GAN	data fusion (Liu et al., 2018c), data reconstruction (Bermudez, Happ, Feitosa, & Oliveira, 2019)
GNN	network analysis of irregular data	RGNN, GCN, GAT, GAE, STGNN	image classification (Qin et al., 2018; Mou, Lu, Li, & Zhu, 2020; Wan et al., 2020)

learn deeper representations, while generative adversarial networks (GANs) learn the generative model of data distribution through adversarial methods (Gui, Sun, & Wen et al., 2020). Graph neural networks (GNNs) are fit for dealing with data with irregular structures (Zhou, Cui, & Zhang et al., 2018), e.g. graphs and networks. CNNs, RNNs, AEs, GANs, and GNNs have been common deep learning architectures. There are many variants and hybrids of these architectures. Table 2 includes the most popular models used in RS applications but is not a comprehensive list. The following sections discuss each of these architectures at a high level.

3.1. Convolutional neural networks (CNN)

CNNs are designed to take advantage of the multidimensional grid structure of the input image (LeCun, Bengio, & Hinton, 2015). Thus, they have been extremely successful in processing RS data. Employing weight sharing and local connectivity, CNNs establish deep networks to learn features from the pixel level to the semantic level for applications. Standard CNNs (such as LeNet-5 in LeCun, Bottou, Bengio, & Haffner, 1998) are usually composed of three main types of neural layers: convolution, pooling, and fully connected layers. Every layer of a CNN transforms the input volume to an output volume of neuron activation, eventually leading to the final network with fully connected layers, resulting in a mapping of the input data to a one-dimensional (1D) feature vector. For a special task, the 1D feature is fed into a perceptron layer to assign class labels or compute probabilities of the given class in the input.

A convolutional layer is an essential part of a CNN, converting the input into a representation of a more abstract level. It utilizes convolutional kernels to slide across the input image, performs a convolution operation between each input region and the kernel, and generates feature maps. Usually, a pointwise nonlinearity activation function (for instance, ReLU, tanh, and sigmoid) is applied to the linear results of the convolutional layers to improve the capacity of nonlinear representation of CNNs. Typically, following

a convolutional layer, a pooling layer is used to reduce the spatial dimensions (width and height) of the feature maps. It summarizes the features from a convolutional layer to a more abstract level by sliding a two-dimensional (2D) filter over each band of the feature map and summarizing (e.g. finding the max and average) the features within the region covered by the filter. Such an operation leads to a loss of spatial information but simultaneously improves the generalization and robustness of the network. A fully connected layer is often added between the penultimate layer and the output layer to further model the nonlinear relationships of the input features. In a fully connected layer, neurons have full connections to all activation in the previous layer. Fully connected layers eventually convert the 2D feature maps into a 1D feature vector. The derived vector can either be fed forward into a certain number of categories for classification or can be considered a feature vector for further processing (Voulodimos, Doulamis, Doulamis, & Protopapadakis, 2018).

In addition to the three basic layers, various layers or blocks were designed to improve the performance of networks, such as a residual block for solving the underfitting problem of deep networks (He, Zhang, & Ren et al., 2016), a bottleneck block to reduce the computational workload, a normalization layer for accelerating training and weakening the internal covariate shift, and an attention block to dynamically adjust the connection weights (Wang & Shen, 2017).

3.2. *Recurrent neural networks*

By exploiting the local dependency of visual information, CNNs have demonstrated record-setting results in many image applications. However, CNNs rely on the assumption of independence among examples. This is unacceptable for data related in time or space, such as video, audio, and texts. RNNs are connectionist models with the ability to selectively pass information across sequence steps (Graves, Mohamed, & Hinton, 2013). Therefore, they can model sequential and time dependencies on multiple scales. The typical feature of the RNN architecture is a cyclic connection, which enables the RNN to possess the capacity to update the current state based on past states and current input data. Thus, RNNs have been widely adopted for sequential data.

However, RNNs consisting of sigma or tanh cells are unable to learn and connect the relevant information when the input gap is large. To handle “long-term dependencies”, long short-term memory (LSTM) was proposed (Hochreiter & Schmidhuber, 1997), in which “gates” were introduced to control dependencies. Each LSTM block is composed of a cell, the memory part of the unit, and three gates: an input gate, an output gate and a forget gate (also called keep gate) (Gers, Schraudolph, & Schmidhuber, 2002). The three gates control the flow of information through the cell. An LSTM unit can remember values over arbitrary time intervals. This LSTM design incorporates nonlinear, data-dependent controls into the RNN cell, which can be trained to ensure that the gradient of the objective function with respect to the state signal does not vanish. Thus, LSTM presents tremendous memory capacity and solves the vanishing gradient problem. It has been widely used in various kinds of sequential tasks, such as time-series classification and prediction.

The learning capacity of the LSTM is superior to that of the standard recurrent cell. However, additional parameters increase the computational burden. Therefore, the gated recurrent unit (GRU) was introduced in (Chung, Gulcehre, & Cho et al., 2014). The GRU cell integrates the forget gate and input gate of the LSTM cell to create an update gate. The GRU cell has only two gates: an update gate and a reset gate. Therefore, it could save one gating signal and the associated parameters. Since one gate is missing, the single GRU cell is less powerful than the original LSTM.

3.3. Autoencoder neural networks

An AE is designed for representation learning to discover latent structure within the input data and for feature selection and dimension reduction. It is composed of three components: an encoder, code, and decoder (Dong, Liao, Liu, & Kuang, 2018). The encoder compresses the input to generate the code, and the decoder reconstructs the input based on the code. A common property of AEs is that the size of the input and output layer is the same as a symmetric architecture (Hinton & Salakhutdinov, 2006). The underlying idea is to learn a mapping from an input pattern x to a new encoding $c = h(x)$, which ideally outputs the same pattern as it takes in for the input, i.e. $x \approx y = g(c)$. Hence, the encoding c , which usually has a lower dimension than x , allows us to reproduce (or code for) x .

By employing various forms of regularization in the general autoencoder architecture, many variants are proposed to ensure that the compressed representation represents a meaningful and generalizable latent space of the original data input. For instance, to exploit the inner structure of data, a sparse AE adds a penalty on the sparsity constraint of the hidden layer. This restriction forces the network to condense and store only important features of the data. In denoising AE, random noise is deliberately added to the input, and the network is forced to reconstruct the unadulterated input. This is because a good representation should be capable of capturing the stable structures in the form of dependencies and regularities characteristic of the unknown distribution. The learned representation is robust toward the slight disturbances observed.

AEs are simple networks that map an input x to a latent representation through one hidden layer (Zhu et al., 2017). A stacked or deep AE (SAE) is a neural network consisting of multiple layers of AEs, where the outputs of each layer are wired to the inputs of the following layer.

3.4. Generative adversarial networks

A GAN learns the generative model of data distribution through adversarial methods (Goodfellow, Pouget-Abadie, & Mirza et al., 2014). Structurally inspired by zero-sum game theory, GANs consist of two models: a generator and a discriminator. The generator tries to capture the distribution of true examples for new data example generation. The discriminator is usually a binary classifier, discriminating generated examples from the true examples as accurately as possible. The optimization of GANs is a min-max optimization problem. The optimization terminates at a saddle point that is a minimum concerning the generator and a maximum to the discriminator. Then, the generator can be thought to have captured the real distribution of true examples. Although these two models can be

implemented with any form of differentiable functions that map data from one space to the other, they are typically implemented by deep neural networks (e.g. CNNs, LSTMs).

GANs are excellent generative models. Many efforts have been devoted to obtaining better GANs through different optimization methods, such as conditional GANs, deep convolutional GANs, f-divergence GANs, and cycle-consistent GANs (Gui et al., 2020). These advantages allow GANs and their variants to be widely applied to many applications of RS, such as data fusion (Jiang et al., 2019; Ma et al., 2020) and cloud masking (Sun et al., 2017).

3.5. Graph neural networks

The success of deep learning (e.g. CNNs and RNNs) in many domains is partially attributed to latent feature representations from Euclidean data. However, such representation would lose efficacy in graphs. As graphs can be irregular, a graph may have a variable size of nodes, and nodes from a graph may have a different number of neighbours, resulting in some important operations (e.g. convolution) being easy to conduct in Euclidean space but difficult to apply to the graph domain (Wu, Pan, & Chen et al., 2020b; Zhou et al., 2018).

GNNs are proposed to better learn representations on graphs via feature propagation and aggregation (Wu et al., 2020b). Many efforts have been conducted on feature propagation and aggregation. GNNs can be generally categorized into five groups: recurrent graph neural networks (RGNNs), graph convolutional networks (GCNs), graph attention networks (GATs), graph autoencoders (GAEs), and spatial-temporal graph networks (STGNNs). RGNNs are mostly pioneer works of GNNs. They aim to learn node representations with recurrent neural architectures, assuming a node in a graph constantly exchanges information with its neighbours until a stable equilibrium is reached. Inspired by the progress of convolutional networks, GCNs redefine convolution for the non-Euclidean data domain (Kipf & Welling, 2016a). Another great source of inspiration was the development of attention mechanisms and full end-to-end attention models. Attention methods have been successfully applied to many graph convolutional networks, such as GAT (Veličković, Cucurull, & Casanova et al., 2017). In contrast to standard GCNs, which give the same weights to all neighbouring nodes when performing a convolution, GAT employs an attention mechanism to assign different weights to a node's neighbours. GAEs map nodes into a latent feature space and decode graph information from latent representations, which can be used to learn network embeddings or generate new graphs (Kipf & Welling, 2016b). Graphs in many real-world applications (such as traffic streams) are dynamic in terms of both graph structures and graph inputs. STGNNs capture the dynamic nature of graphs (Guo, Lin, & Feng et al., 2019; Yao, Wu, & Ke et al., 2018) by modelling the dynamic node inputs and assuming interdependency between connected nodes. Thus, STGNNs can be used for forecasting future node values or labels or predicting spatial-temporal graph labels.

Due to their convincing performance and high interpretability, GNNs have recently been widely applied in a wide range of problem domains across scene graph generation, scene classification (Gao, Shi, & Li et al., 2021; Liang, Deng, & Zeng, 2020), point clouds segmentation/classification (Wen, Li, Yao, Peng, & Chi, 2021; Widyaningrum, Bai, Fajari, &

Lindenbergh, 2021), text classification, traffic forecasting (Peng, Wang, Du, Bhuiyan, Ma, Liu & Yu, 2020), and event detection (Guo et al., 2019). However, applications of GNNs in RS are just beginning.

4. RSBD meets deep learning algorithms

Based on the technical overview of deep learning algorithms (Section 3) and practical demands in applications of RSBD (Section 2), we have found that deep learning can be applied to every aspect of RS data processing and analysis: from the traditional topics of image registration and object detection to the recent challenging tasks of high-level semantic segmentation and multitemporal analysis. In this section, we follow the pipeline of RS processing and analysis to review the advance of deep-learning-based RS.

4.1. Pipeline of processing and analysis for RSBD

RSBD has been successfully applied in many fields. Through carefully reviewing the flowchart of RS applications, from source data to target information and knowledge, we can split the technical pipeline of RS tasks into four major steps, including data acquisition, data processing, data analysis, and data application, as presented in Figure 1. In this procedure, important technical steps are data processing and analysis, in which deep learning algorithms have bright prospects.

In RS processing, tasks can be grouped into four main categories, including geometric (spatial) processing, radiometric (spectral) processing, cloud masking, and data fusion. After distributed RS data are received, radiometric processing is applied to eliminate radiometric effects from the atmosphere, terrain, and spectral differences. Second, geometric processing is conducted to locate RS data in the geographical space and to reconstruct three-dimensional (3D) models of scenes. Furthermore, according to application requirements, cloud masking is applied to RS images to delineate cloud-covered regions to extract valid data. For special applications, incorporating multisource and multimodal RS data and data fusion methods could produce fully covered, higher spatial-temporal resolution datasets, e.g. pan-sharpening.

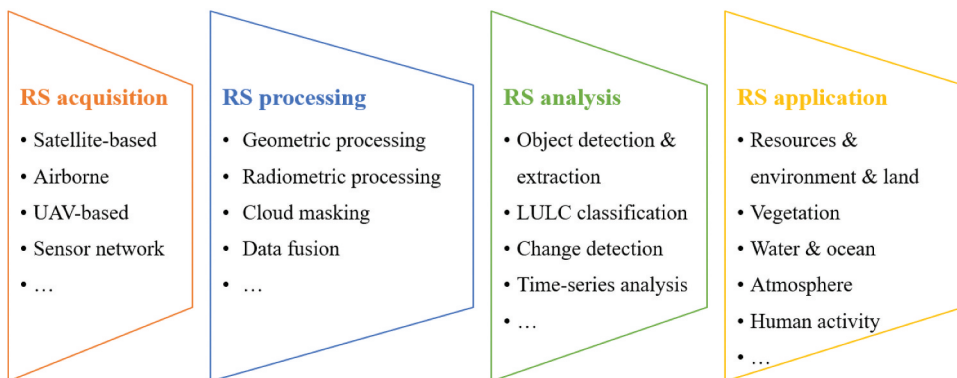


Figure 1. The general technical pipeline of processing and analysis for remote sensing big data.

Using the results from RS processing, RS analysis is carried out to extract thematic/object information and conclude scene knowledge, including object detection and extraction, land-use/cover (LULC) classification, change detection, and multitemporal analysis. Using object detection and extraction methods, researchers can identify, locate, and delineate interesting ground objects, such as ships, oil-spilling regions, dams, bridge constructions, airports, parking lots, power stations, and waterbodies. LULC classification labels pixels or regions in an image as one of several classes. Furthermore, change detection and multitemporal analysis are applied on two or more RS data points to detect change regions and present the change process of the ground surface.

It is worth noting that it is difficult to cover all available RS processing and analysis techniques in this pipeline, e.g. point cloud segmentation. However, this pipeline involved common and important algorithms. Thus, deep learning algorithms in RS would follow this pipeline, as discussed in the following subsection.

4.2. Remote sensing processing

4.2.1. Geometric (spatial) processing

RS geometric processing aims to locate every pixel in the geographic space, including geometric rectification, orthorectification, terrain correction, and 3D reconstruction, etc. The fundamental preliminary step of geometric processing is image registration, aligning multiple images captured by different sensors (multimodal registration) at different times (multitemporal registration) from different viewpoints (multiangle registration) into the same coordinate system. The typical procedure of image registration includes the following three steps: feature extraction (descriptor), feature matching, and image warping.

Feature extraction plays a crucial role in image registration because it decides what type of feature to use for matching. In traditional methods, numerous handcrafted feature descriptors achieve good results in RS images, such as SIFT, SURF, and ORB (Bay, Ess, Tuytelaars, & Van Gool, 2008; Lowe, 2004; Rublee, Rabaud, & Konolige et al., 2011). Due to incredible feature representation, deep learning algorithms (especially CNNs) can produce many multiscale candidate features (descriptors). Thus, various CNNs have been developed for feature extraction (Ye, Su, Xiao, Zhao, & Min, 2018; Zhu, Jiao, Ma, Liu, & Zhao, 2019) and have improved the performance of image registration. For example, Yang, Dan, and Yang (2018c) used a CNN to generate robust multiscale feature descriptors (that keep both convolutional information and localization capabilities) for multitemporal RS image registration.

To further utilize the end-to-end learning ability of deep learning, researchers have tried to design complete architectures unifying feature extraction and matching in a closed-loop framework to directly learn the geometric transformation between two images. By taking the image patch pairs as input and matching labels as output, Wang et al. (2018d) designed a direct mapping network in which hidden layers correspond to a feature extraction and the output layer corresponds to feature matching. This end-to-end architecture optimizes the learning mapping function through information feedback when training the network, permitting the results of feature matching to guide the process of feature extraction and then making the learned features more appropriate.

In addition to registration among homogenous images, deep learning algorithms can be applied for the registration of multimodal RS data, e.g. optical images and SAR data

(Merkle, Auer, Müller, & Reinartz, 2018; Hughes, Schmitt, & Zhu, 2018; Zhang, Ni, Yan, Xiang, Wu, Yang, & Bian, 2019d; Ma et al., 2019c). A common strategy of these methods is translating the input images into the same feature space, enabling the two images to share similar intensity or feature information. Subsequently, feature extraction and matching are carried out between the two translated images.

Recently, the latest learning-based techniques have been introduced into image registration, such as reinforcement learning and multitask learning. In Duan, Chiang, Leyk, Uhl, and Knoblock (2020), an automatic vector-to-raster alignment algorithm based on reinforcement learning was introduced to annotate precise locations of geographic features on scanned maps. This algorithm can also be applied to various features (roads, water lines, and railroads). Girard *et al.* utilized multitask learning to improve registration performance (Girard, Charpiat, & Tarabalka, 2018).

4.2.2. Radiometric (spectral) processing

Image radiometric processing maps or corrects pixel values from one measured space to another, with or without reference data, including physical parameter inversion, absolute and relative radiometric correction, and radiometric normalization. Due to the complications of physical processes, when the parameters of radioactive transfer models are unavailable, statistical-learning-based models (e.g. shallow neural networks, support vector machines) have been developed into practical methods (López-Serrano, López-Sánchez, Álvarez-González, & García-Gutiérrez, 2016; Seo, Kim, Eo, Park, & Park, 2017; Xu, Hou, & Tokola, 2012; Zhou & Grassotti, 2020b). In particular, deep learning techniques have further promoted learning-based radiometric processing. Ogut, Bosch-Lluis, and Reising (2019) applied a deep-learning-based calibration technique for the calibration of high-frequency airborne microwave and millimeter-wave radiometer instruments. Xu, Cervone, Franch, and Salvador (2020) presented an artificial intelligence/deep learning-based solution to characterize the atmosphere at different vantage points and to retrieve target spectral properties. A deep learning emulator approach for atmospheric correction was developed, suggesting that a deep learning model can be trained to emulate a complex physical process (Duffy et al., 2019).

Here, we have found two problems. First, current deep-learning-based radiometric processing is still in the traditional domain of machine learning and is not applied to high-level semantic representations from deep networks; there have been very few studies in this field. Let us analyze them from deep learning algorithms and the RS process. RS imaging is physically based, and there are already many radiative transport models. Studies on radiometric processing are desirable for quantitative and explicable RS process models. However, statistical-based deep learning is designed to find the correlativity between inputs and targets, which is inconsistent with the object of radiometric processing. Thus, it is difficult to employ deep learning to improve the understanding of the RS process, and current deep learning algorithms are not suitable for radiometric processing.

4.2.3. Cloud masking

Taking clouds as a special target, cloud masking is a part of object extraction and belongs to the domain of image analysis. In the pipeline of RS applications, cloud-covered regions are taken as invalid and cloud masking is an important step to delineate and extract cloud-free (valid) regions from RS images for the following analysis. Thus, cloud masking is

a hot research topic in RS applications. In Shi, Xie, and Zi et al. (2016), a simple CNN with four convolutional layers and two fully connected layers was designed for cloud masking on multispectral images, suggesting the promise of CNNs for cloud masking compared to traditional approaches. In Jeppesen's study (Jeppesen, Jacobsen, Inceoglu, & Toftegaard, 2019), a multitemporal approach based on the U-Net architecture for cloud masking was applied on the Landsat 8 Biome and SPARCS datasets, showing improvements compared to the FMask algorithm (Qiu, Zhu, & He, 2019). Shao et al. (2017) adopt a fuzzy stacked AE model to integrate the feature learning ability of AE networks and the detection ability of fuzzy functions for highly accurate cloud masking. By integrating the low-level spatial features and high-level semantic features simultaneously, a fully convolutional network was developed for distinguishing clouds and snow in RS images (Zhan et al., 2017). Based on superpixels, Xie, Shi, Shi, Yin, and Zhao (2017) designed a two-branch CNN to extract multiscale features from each superpixel and predict the superpixel as one of three classes: thick cloud, thin cloud, and noncloudy. For more precise boundaries of large-region clouds, Wu et al. (2020a) introduced group training and boundary optimization to improve the Mask R-CNN for cloud masking. Yuan, Meng, and Cheng et al. (2017, September) proposed a multitask (two tasks of cloud segmentation and cloud edge detection) deep neural network for accurate cloud detection in RS images.

Cloud masking shares the same challenge of limited training samples. One direction for solving this challenge is through weakly supervised learning, using block-level labels indicating only the presence or absence of clouds in one image block (Li et al., 2020e). Another way is transfer learning. Mateo-García, Laparra, López-Puigdollers, and Gómez-Chova (2020) presented an approach using labeled Landsat-8 datasets to train deep learning models for cloud detection that can be applied (or transferred) to Proba-V sensors.

In practice, it was inconvenient to download large-region images for cloud masking. It will be helpful to obtain cloud masking products directly from cloud platforms, such as Google Earth Engine (GEE) and Sentinel Hub.³ Through GEE services, a CNN-based cloud detection model was first trained locally then deployed on GEE, which makes it possible to directly detect clouds for Landsat-8 imagery in GEE (Yin, Ling, & Foody et al., 2020).

4.2.4. Data fusion

Since there are incompatible conflicts among spectral, spatial, and temporal resolutions for sensor imaging, it is difficult for a special sensor to simultaneously achieve high spatial, high spectral, and high temporal resolutions. For example, MODIS data have a shorter revisit cycle of 1 day but a lower spatial resolution of 250 m, while Landsat-8 images have a higher spatial resolution of 30 m but a longer revisit cycle of 16 days.

Data fusion is the process of combining multiple images (captured through various sensors under different parameter settings) to generate a single image which combines all the meaningful information from the individual images. Therefore, data fusion methods provide the composite (fused) image with complementary information. A typical example of RS data fusion is pan-sharpening, which indicates the fusion of a low-resolution multispectral image and a high-resolution panchromatic (PAN) image to achieve a high-resolution multispectral image (also known as spatial-spectral fusion). Another common fusion is spatial-temporal fusion, which incorporates high spatial-resolution images with multitemporal images to produce multitemporal high spatial-

resolution images. When RS data are captured by various sensors with different imaging modalities (including optical images, SAR, and LiDAR data), multimodal fusion is conducted.

(1) spatial-spectral fusion (pan-sharpening)

Inspired by the recent progress achieved in image superresolution and the better capability of deep learning networks in characterizing the complex relationship between the input and the target images many RS data fusion methods using deep learning techniques have been proposed.

The pioneering study of deep-learning-based pan-sharpening (Huang et al., 2015) employed a sparse denoising AE network to model the relationship from low-resolution to high-resolution image patches under the assumption that the multispectral and PAN images share the same relationship from low- to high-resolution versions. An improved study combined multiple stacked sparse AEs to develop a deep metric learning method to learn a refined geometric multi-manifold neighbour embedding (Xing et al., 2018). In addition to AE networks, CNNs are active in pan-sharpening (Yang, Fu, & Hu et al., 2017). Following the super-resolution procedure for natural images, Masi, Cozzolino, Verdoliva, and Scarpa (2016) proposed a CNN-based pan-sharpening method, in which the down-sampled multispectral and PAN images were stacked and fed into the network, and the output of the network with the original multispectral image was compared. Wei, Yuan, Shen, and Zhang (2017) used residual learning to develop a very deep CNN with 11 convolutional layers to improve the accuracy of pan-sharpening fusion. To reconstruct the spatial details in upsampled multispectral images, Benzenati, Kallel, and Kessentini (2020) and Liu et al. (2020a) proposed two-stage approaches. In the first stage, CNNs captured mid-level and high-level spatial features from PAN images. In the second stage, detailed injection models merged the spatial details extracted in the first stage into multispectral images. Liu et al. (2018c) proposed a GAN-based algorithm for pan-sharpening, in which a two-stream fusion architecture was designed to generate the desired high-resolution multispectral images, and a fully connected convolutional network serving as a discriminator was applied to distinguish “real” or “pan-sharpened” multispectral images.

It is worth noting that downscaling methods belong to this family. Direct methods apply superresolution CNNs on low-resolution images to produce high-resolution ($\times 2$, $\times 4$, and even $\times 8$) results (Haut et al., 2018; Lei, Shi, & Zou, 2017; Ma, Pan, Guo, & Lei, 2019b; Tuna, Unal, & Sertel, 2018). Practical approaches incorporate domain-specific knowledge from high-resolution thematic images relative to the target information as supplementary data to improve the performance of data fusion (Galar, Sesma, Ayala, Albizua, & Aranda, 2020; Liu & Weng, 2018a; Liu, Xu, & Zhang, 2019c; Shao, Cai, Fu, Hu, & Liu, 2019).

(2) spatial-temporal fusion

Multitemporal sequential data are an important tool to observe the change and evolution of the Earth’s surface. However, they are usually of low spatial resolution, e.g. MODIS data. Thus, spatial-temporal fusion techniques are used to produce dense temporal data with high spatial resolution (Li, Li, & He et al., 2020a). These techniques not only employ spatial scale information as in spatial-spectral fusion and spatial-temporal fusion but also take full advantage of temporal tendencies and dependencies in multitemporal images.

The pathbreaking study of deep learning spatial-temporal fusion is the deep-STEP method (Das & Ghosh, 2016). It was derived from a deep stacking network, in which

temporally evolved feature sets were introduced to incorporate the temporal change along with the spatial feature learning at hidden layers for spatial-temporal prediction. Many efforts have been devoted to developing various CNN frameworks for spatial-temporal fusion (Song, Liu, Wang, Hang, & Huang, 2018; Tan, Yue, Di, & Tang, 2018; Yin, Wu, & Foody et al., 2020) by using CNNs to learn complex nonlinear mapping at spatial scales. In these networks, the inputs are a pair of low-temporal high-spatial and high-temporal low-spatial images for training and another pair of high-temporal low-spatial images for prediction. Information is merged in the form of extracted CNN-based features, and then the merged features are reconstructed to the predicted image.

Additionally, to further capture temporal information, two-stream CNN frameworks were designed. Liu, Deng, Chanussot, Hong, and Zhao (2019d) proposed a two-stream CNN, considering the temporal dependence and temporal consistency among image sequences in the CNN-based superresolution process. Jia et al. (2020) combined both CNN-based forward and backward prediction branches into a two-stream framework for spatial-temporal data fusion, where temporal change-based and spatial information-based mapping were simultaneously presented, addressing the prediction of both phenological and land cover changes with better generalization ability and robustness.

(3) multi-modal fusion

Different from the above fusions using images from the same or similar imaging mechanism, multimodal fusion considers the nonlinear transformation and assimilation of different imaging information in multimodal fusion, i.e. fusion between optical images and SAR data.

In optical-SAR fusion, the common technique is to transform heterogeneous data into a similar feature space based on feature representations of deep learning. In Li, Lei, Sun, Li, and Kuang (2020d), a multimodal deep learning network incorporating a pseudo-Siamese CNN and attention-based channel selection module was proposed to fuse the deep features from optical images and SAR data for land cover classification. Additionally, a few studies introduced GANs to perform direct translations from SAR data to optical images to generate cloud-free optical images (Bermudez et al., 2019; Grohnfeldt, Schmitt, & Zhu, 2018; He & Yokoya, 2018). Furthermore, time-series data were employed to improve the effectiveness and stability in optical-SAR transformation. In Scarpa, Gargiulo, Mazza, and Gaetano (2018), a CNN was used to exploit optical and SAR joint time series to retrieve the normalized difference vegetation index of the missing optical image. Zhou et al. (2020) proposed a prediction-transformation-fusion time-series reconstruction method for cloud/shadow-covered optical images, incorporating multitemporal SAR data through LSTM networks and AEs.

Compared with optical and SAR sensors, LiDAR measures the distance or position of the ground surface. Fusion of optical images and LiDAR data would benefit the understanding of ground surfaces. To overcome the huge difference in the information form, two- or multibranch networks are commonly used tricks: one branch for extracting spectral features and the other branch for learning spatial correlation features. Finally, for further tasks, the features of the two branches will be flattened and stacked (Feng, Zhu, Yang, & Li, 2019b; Wang, Zhang, & Guo et al., 2019a).

With an increasing number of sensor networks producing massive discrete point-like data, recent studies try to incorporate sensor data with RS images to obtain finer and more accurate thematic maps. Taking advantage of the reliability of RS data and the

spatial-temporal dynamic characteristics of sensor data, Cao et al. (2020) proposed an end-to-end deep-learning-based approach to fuse remote and social sensing data to recognize urban region functions in high-density cities using CNNs and LSTM networks. Shen, Zhou, and Li et al. (2019) used a deep belief network to learn the complex relationships among RS data, social sensing data, meteorological data, and the spatial-temporal features of $PM_{2.5}$ for finer spatial-temporal mapping of $PM_{2.5}$ concentrations in urban areas.

4.3. Remote sensing analysis

4.3.1. Object detection and object extraction

Before discussing object detection and object extraction, it is essential to denote their differences, as both have similar RS applications and are frequently confused. In this study, object detection is the process of finding instances of objects in input images, e.g. ships, cars, airplanes, and wind generators. It produces one or more bounding boxes with the class label attached to each bounding box. However, object extraction, a further extension of object detection, delineates the presence of an object (such as a waterbody, road, and urban area) through pixelwise masks generated for each object in the image, which is close to instance segmentation in the computer vision field. This task is much harder than object detection.

(1) object detection

Inspired by the success of deep learning algorithms (e.g. the Regional-CNN and R-CNN frameworks) for object detection for natural images, many deep-learning-based object detection methods have been developed for RS data, indicating distinct advantages over traditional methods due to their better ability to learn high-level semantical representations (Li, Cheng, Bu, & You, 2018; Chen, Zhang, et al., 2018; Hu, Li, & Zhou et al., 2019b; Ren, Zhu, & Xiao, 2018; Long, Gong, Xiao, & Liu, 2017; Yan et al., 2019). However, there are several challenges for object detection, including rotation/scale invariance and limited labeled samples for training (Li, Wan, Cheng, Meng, & Han, 2020b).

Taking horizontal bounding boxes as regions of interest (RoIs), the R-CNN framework has good properties for the generalization of translation-invariant features but poor performance on rotation and scale variations (Cheng, Zhou, & Han, 2016; Long et al., 2017). Instead of using horizontal bounding boxes, oriented bounding boxes have been employed to eliminate the mismatching between rotated RoIs and corresponding objects (Liu, Wang, Weng, & Yang, 2016b; Xia, Bai, & Ding et al., 2018). Li, Zhang, Huang, and Yuille (2018c) presented rotation-insensitive region proposal networks by introducing multi-angle anchors, which can effectively handle the problem of geospatial object rotation variations. Cheng et al. (2016) introduced a rotation-invariant layer into existing CNN architectures to improve the performance of object detection. Cheng, Han, Zhou, and Xu (2019) explicitly imposed a rotation-invariant regularizer and a Fisher discrimination regularizer on the CNN-based feature to further boost object detection performance. Ding, Xue, and Long et al. (2018) combined a rotated RoI learner (transforming a horizontal RoI into a rotated RoI) and a rotated position-sensitive RoI align module (extracting rotation-invariant features) into a two-stage framework for detecting oriented

objects in aerial images. These methods are promising for detecting sparsely distributed rotated objects (Zhang, Guo, Zhu, & Yu, 2018c).

Multiscale RS objects are an important feature that often appear at very different scales in RS data. On the one hand, the scale variability is caused by image resolution. On the other hand, different object categories have large size differences. This brings a huge challenge when transferring deep learning algorithms for natural images into RS applications. Continuous efforts are being devoted to addressing this issue. In the hybrid CNN of Chen, Xiang, Liu, and Pan (2014), feature maps from the last convolutional layer and the max-pooling layer were divided into multiple blocks of variable receptive field sizes or max-pooling field sizes to extract multiscale features for vehicle detection in RS images. In a multiscale visual attention network, Wang, Bai, and Wang et al. (2018a) used a skip-connected encoder-decoder model to extract multiscale features from a full-size image. Deng et al. (2018) developed a multiscale object proposal network to ease the inconsistency between the size variability of objects and fixed filter receptive fields. The method was tested on several intermediate feature maps according to the certain scale ranges of different objects. In Zhang, Yuan, Feng, and Lu (2019e), multiscale convolutional features are extracted to represent the hierarchical spatial semantic information, and multiple fully connected layer features are stacked together to improve the rotation and scaling robustness.

To reduce the need for training samples, a few detection methods transferred pre-trained CNNs for object detection. Zhou, Cheng, Liu, Bu, and Hu (2016) presented a weakly supervised learning framework to train an object detector, using a pretrained CNN model to extract features and a negative bootstrapping scheme for faster convergence. Zhang, Shi, and Wu (2015) incorporated deep surrounding features from the pretrained CNN model with local features (histogram of oriented gradients) to hierarchically detect oil tanks. Li, Zhang, Lei, Wang, and Guo (2020c) employed transfer learning and fine-tuning approaches to train three well-established CNN-based models for detecting agricultural greenhouses from multisource RS images.

(2) object extraction

Object extraction can be formed as a biclass classification or instance segmentation, aiming to predict the pixelwise mask of each instance in RS images. These objects can be buildings, roads, waterbodies, and other regions of interest, such as urban canopy (Timilsina, Aryal, & Kirkpatrick, 2020), nutrient deficient areas (Dadsetan, Pichler, & Wilson et al., 2021).

Buildings are important targets in high spatial-resolution RS data. Many deep learning networks established for natural images have been improved for building extraction (Hui, Du, Ye, Qin, & Sui, 2018; Liu, Luo, & Huang et al., 2019a; Yang, Yu, Luo, & Chen, 2019; Yuan, 2016). Additionally, some useful tricks were used to obtain better results, including attention (Yang et al., 2018b), multitasking learning (Hui et al., 2018; Yang et al., 2018a), spatial pyramid pooling (Liu et al., 2019e) and dilated convolution (Ji, Wei, & Lu, 2019). For example, to deal with multiscale buildings, Liu et al. (2019b) designed a spatial residual inception module in a fully convolutional network to capture and aggregate multiscale contexts for semantic understanding by successively fusing multilevel features. Ji, Wei, and Lu (2018) developed a two-branch U-Net with shared weights, one for original images and the other for their downsampled counterparts. Furthermore, a few studies used

supplemental data (e.g. LiDAR data) to improve the performance of building extraction (Huang, Zhang, Xin, Sun, & Zhang, 2019; Maltezos, Doulamis, Doulamis, & Ioannidis, 2017). Inspired by GANs, Li, Yao, and Fang (2018a) developed a deep adversarial network for extracting building rooftops in RS images, in which the generator produced a pixelwise image classification map using a fully convolutional model, whereas the discriminator tends to enforce forms of high-order structural features learned from a ground-truth label map. Recently, for the precise delineation of boundaries, Shi, Li, and Zhu (2020) proposed a gated GCN, which enables the refinement of weak and coarse semantic predictions to generate sharp borders and fine-grained pixel-level classification.

Compared with building extraction, the task of road extraction faces more challenges, including its narrowness, sparsity, diversity, multiscale characteristics, and class imbalance. Thus, special blocks and models are designed. In Gao et al. (2018), a weighted balance loss function is presented to solve the class imbalance problem caused by the sparseness of roads. Zhang, Liu, and Wang (2018d) introduced rich skip connections in networks to facilitate information propagation, resulting in fewer parameters but better performance. Dilation convolution was employed in road extraction to enlarge the receptive field of feature points without reducing the resolution of feature maps (Zhou, Zhang, & Wu, 2018b). Further studies placed road extraction and other tasks into multitask learning frameworks, such as road centerline extraction (Lu et al., 2019), road edges (Liu et al., 2018d) and building extraction (Zhang & Wang, 2019f). Abdollahi et al. (2020) compared four types of deep-learning-based methods for road extraction, including the GAN model, deconvolutional networks, FCNs, and patch-based CNN models.

Waterbodies are common targets in RS images. Various deep learning frameworks have been developed for extracting waterbodies (Chen et al., 2018; Feng, Sui, Huang, Xu, & An, 2018; Isikdogan, Bovik, & Passalacqua, 2017; Li et al., 2019a; Song et al., 2020; Wang, Li, & Zeng et al., 2020b). It is worth mentioning that waterbodies consist of lakes, rivers, and sea. Thus, water body extraction from RS images also suffers from the same problems as building and road extraction, such as multiscale targets, class imbalance, and narrowness.

4.3.2. Land-use and land-cover image classification

Land-use/cover (LULC) classification is the foundational analysis for RS data, which has advanced from pixelwise to the object-based analysis paradigm. Pixelwise LULC classification, assigning land-type labels to each pixel in an image, has a similar technical procedure as semantic segmentation in computer vision. Recent pixelwise LULC classification using semantic segmentation networks has achieved great advances (Tong et al., 2020; Zhang et al., 2020a, 2019a). In particular, to further explore spatial and graph topological information, GNNs have been employed in LULC classification tasks on hyperspectral images (Mou et al., 2020; Qin et al., 2018; Wan, Gong, & Zhong et al., 2019; Wan et al., 2020; Wang, Ma, Chen, & Du, 2021), very high-resolution satellite images (Cui et al., 2021; Khan, Chaudhuri, Banerjee, & Chaudhuri, 2019; Liu, Kampffmeyer, & Jenssen et al., 2020b; Ouyang & Li, 2021), and time-series images (Censi et al., 2021).

Compared with pixelwise image analysis, object-based image analysis makes full use of various features of objects (regions or patches), such as spectral, textural, and geometric features (Tong, Xia, & Lu et al., 2018; Zhou, Li, Feng, Zhang, & Hu, 2017). From this perspective, by learning multiscale regional representations, deep learning algorithms

are innately suitable for object-based LULC classification. The traditional procedure of object-based classification usually consists of object generation and object classification. Current deep-learning-based studies focus on object classification under assumptions of available object maps. Taking patches with a size of 5×5 as objects, Sharma, Liu, Yang, and Shi (2017) proposed a deep patch-based CNN framework for medium-resolution RS data. Fu, Ma, Li, and Johnson (2018) generated patches using the barycenter of segmented objects (resulting from multiresolution segmentation) as centers and fixed the window at 32×32 pixels and 64×64 pixels and subsequently classified these patches through CNNs. Another strategy that combines deep learning with object-based classification, Tong et al. (2018) and Lv et al. (2018) rely on CNNs to predict the land type on a pixel level by patchwise classification and then vote to determine the segmented object's type on an object level by the pseudolabels of a pixel. Some studies try to utilize additional data to generate objects. Huang, Zhao, and Song (2018) employed a skeleton-based decomposition method that employed road networks to split the image into regular regions and fed them into a deep convolutional neural network for urban land-use classification.

When carefully reviewing these object-based LULC classifications using deep learning, we found that objects from traditional segmentation or other data sources suffer from problems of under/oversegmentation and rough boundaries, making it impossible to accurately represent LULC features. A few pioneering studies attempted to employ deep learning algorithms to generate ground objects with precise boundaries. For example, Waldner and Diakogiannis (2020) used a deep convolutional neural network with a fully connected U-Net backbone that features dilated convolutions and conditioned inference to extract field boundaries from RS images. Liu et al. (2020d) incorporated three CNNs in a hierarchical scheme to extract precise farmland parcels in mountainous areas. However, boundaries or edges in the image are low-level features. It is difficult for deep learning algorithms with convolution and pooling to produce pixelwise edges. Some tricks have been introduced to improve boundary accuracy, including dilated convolution, upsampling, and postprocessing on extracted edges using conditional random field and spatial analysis algorithms (Liu et al., 2019a; Marmanis et al., 2018; Pan, Zhao, & Xu, 2020).

4.3.3. *Change detection*

Change detection is the process of identifying the changes in the ground surface (such as flooding, vegetation growth, and urban construction) using multiple RS data obtained at different times. With the intensification of human activities and huge amounts of high-temporal resolution RS data available, change detection has been extensively researched in recent decades.

The mainstream method for change detection is to identify ground surface changes from source images. When stacking two images together, the change detection problem can be translated into biclass classification, delineating the changed region through deep learning algorithms. It usually consists of two main steps, including feature representation and semantic segmentation. For homogeneous images acquired from similar or the same sensor (e.g. optical images), the same CNNs can be applied to extract multiscale feature representations and delineate the change regions (Gao, Dong, Li, & Xu, 2016; Liu, Jiang, Zhang, & Zhang, 2020c; Mou, Bruzzone, & Zhu, 2018; Peng, Zhang, & Guan, 2019; Wang, Yuan, & Du et al., 2018b; Wang et al., 2018c; Zhang, Wei, Ji, & Lu, 2019b; Zhang, Xu, Chen,

Yan, & Sun, 2018a). For heterogeneous RS data (e.g. optical images and SAR data), transformer networks first transformed one or two data points into the same feature space. Then, these features were fed into the above networks designed for homogeneous images to identify the change regions (Liu, Gong, & Qin et al., 2016a; Zhang, Gong, Su, Liu, & Li, 2016b; Zhao, Wang, & Gong et al., 2017). Furthermore, to conclude and transfer change rules from multitemporal data for wider applications, Lyu et al. (2016) employed an improved LSTM model to learn efficient change rules with transferability to detect both binary and multiclass changes.

Since obtaining enough labeled samples for supervised training is usually time-consuming and labor-intensive, many efforts have been made to achieve deep-learning-based change detection in an unsupervised or semisupervised manner (Gong, Yang, & Zhang, 2017; Liu et al., 2016a; Zhang et al., 2016b; Zhao et al., 2017). Encoding multitemporal images such as a graph, Saha, Mou, and Zhu et al. (2020) applied a GCN to propagate labeled information from a few training samples to unlabeled samples for semisupervised change detection. Du, Ru, Wu, and Zhang (2019) used slow feature analysis theory to suppress the unchanged components and highlight the changed components of the transformed features from symmetric deep networks to find unchanged pixels with high confidence as training samples. Li et al. (2019b) employed false labels generated by spatial fuzzy clustering to convert a supervised training process into an unsupervised learning process for unsupervised change detection.

4.3.4. Multi-temporal analysis

With an increasing amount of RS data, multitemporal analysis has become an important technique in periodic and dynamic RS applications. Due to the similarity between RS sequential data with video and audio, many deep learning frameworks in computer vision can be used for RS multitemporal analysis.

With the ability to capture long-term dependencies, RNNs have been widely employed for multitemporal analysis in RS applications. Liu et al. (2020d) designed a two-layer LSTM network for time-series sequence classification. Sun, Di, and Fang (2019) designed an LSTM-based framework to take advantage of the temporal pattern of crop growth across image time series to improve the accuracy of crop mapping. Based on parcel maps, Zhou et al. (2019) proposed an LSTM-based time-series analysis method using multitemporal SAR data for crop classification, achieving a 5.0% improvement in overall accuracy compared to those of traditional methods. Furthermore, Zhao et al. (2019) compared several RNN-relative architectures for crop type classification using multitemporal Sentinel-2 data, including a 1D CNN, an LSTM, a GRU, and a novel self-attention network. In these studies, RNN-based applications consist of two separated steps: extracting spatial features and constructing time-series data and time-series analysis. There is a lack of end-to-end frameworks.

The convolutional LSTM (ConvLSTM) network connecting a convolutional network with a recurrent network was proposed to simultaneously learn the spatial and temporal feature representation for spatial-temporal classification and prediction (Shi, Chen, & Wang et al., 2015; Teimouri, Dyrmann, & Jørgensen, 2019; Zhang et al., 2020b). In applications of hyperspectral images, researchers modeled long-term dependencies in the spectral domain and employed ConvLSTM networks to extract more discriminative spatial-spectral features to achieve end-to-end frameworks (Feng, Wu, & Chen et al.,

2019a; Hu, Li, & Pan et al., 2019a; Hu et al., 2020; Wang, Li, & Deng et al., 2020a). The recent success of the attention mechanism in sequential data inspired its applications in multi-temporal analysis. For example, Feng et al. (2020) proposed an attention-based recurrent convolutional neural network for accurate vegetable mapping from multitemporal UAV red-green-blue imagery.

From the perspective of data format, time-series data can be taken as a special case of images, where height = 1. Thus, CNNs and variants used in image domains can be easily transferred for multitemporal analysis, e.g. temporal convolutional neural networks (TCNs) (Hewage, Behera, & Trovati et al., 2020; Lea, Vidal, & Reiter et al., 2016; Yan, Mu, Wang, Ranjan, & Zomaya, 2020b). TCN is a framework that employs casual convolutions and dilations and is thus adaptive for sequential data with its temporality and large receptive fields. Yan, Chen, Chen, and Liang (2020a) proposed a time series prediction approach that employs TCNs to carry out multistep prediction of land cover from dense time series RS images.

4.3.5. Other analysis tasks

In addition to the above conventional application areas in RS image analysis, deep learning has been successful in many interesting fields, such as:

- (1) Number counting. In a large-scale region, e.g. globally, it is difficult to count the number of special objects, such as trees, wild animals (Peng, Wang, Liao, Shao, Sun, Yue & Ye, 2020) and whales (Guirado, Tabik, Rivas, Alcaraz-Segura, & Herrera, 2019), in RS images for traditional methods. Brandt et al. (2020) used deep learning to map the crown size of each tree more than 3 m² in size over a land area that spans 1.3 million km² in the West African Sahara, Sahel, and subhumid zones, exploring their role in mitigating degradation, climate change and poverty.
- (2) Macroscopic indicators. Some studies applied deep learning techniques on RSBD to retrieve information on the volume occupancy of oil storage tanks (Wang, Li, Yu, & Liu, 2019b), fossil-fuel power plants (Zhang & Deng, 2019c), and solar power plants (Hou, Wang, & Hu et al., 2019) to calculate macroeconomic indicators, measuring international trade, production intensity, and material reserves.

5. Discussions

Through the above “review on deep learning algorithms in processing and analysis of RSBD”, we argued that deep learning technology has provided powerful tools for processing and analysis of RSBD based on its great feature learning and expression in the spatial, spectral and temporal domains. Details are as follows.

- (1) First, deep learning algorithms mostly achieve state-of-the-art performance in the above processing and analysis tasks. This allows remote sensing methods to be applied for practical applications (Zhu et al., 2017; Li et al., 2018b; Ma et al., 2019a). Such methods also respond to the requirements of high precision in the processing and analysis of RSBD.
- (2) Using multilayer structures, deep learning networks transfer multimodal RSBD into similar feature spaces, presenting great capabilities for data fusion (Shen et al.,

2019; Zhou et al., 2020). This provides new ideas for cooperative analysis of multi-source, multimodal, multiscale, and heterogeneous RSBDs.

- (3) Deep learning could provide end-to-end frameworks from raw data to target results (Cao et al., 2020; Wang et al., 2018d). This advantage could reduce human-computer interaction and greatly improve the efficiency and automation level of RSBD tasks.

However, due to complicated scenes and imaging processes, there are some disadvantages and challenges for deep learning in the processing and analysis of RSBD, such as:

- (1) Lack of specialized networks for RSBD. Current studies mostly introduce and improve deep learning networks from the field of computer vision (Cheng et al., 2016; Tong et al., 2020). Compared with natural scene images, RS data are multi-resolution, multitemporal, multispectral, multiview, and multitarget. Therefore, according to the characteristics of RSBD, the question of whether we can design and construct specialized deep networks to improve the processing and analysis of RSBD needs to be answered.
- (2) Limited computing resources for RSBD. Thousands of feature representations in deep learning take up more computing resources than traditional methods. For example, a simple deep learning network may contain millions of parameters that need to be solved. In consideration of massive data from satellites, UAVs, and sensor networks, current deep-learning-based RS applications are still in task-driven stages and have a long way to data-driven stages.
- (3) Inadequacy labeled samples or datasets. Deep learning networks can learn high-level abstract feature representations from RS images, and their performance relies on large numbers of training samples (much more than that for traditional methods) (Li et al., 2020b). However, there is a lack of enough training samples because of the difficulties of collecting labeled data. Although sample augmentation techniques and pretrained models (Zhou et al., 2016) could release the demand of samples, their effects are limited because of the lack of domain-specific knowledge and scaling effects. Thus, how to retain the representation learning performance of deep learning methods with fewer adequate training samples remains a major challenge.
- (4) Lack of physical mechanisms and poor transferability. Deep learning is a statistical fitting of the input data and the output result, and there is a lack of capacity to reveal the mechanisms of remote sensing imaging and understanding. Thus, it may not be appropriate for remote sensing physical processes, e.g. radiometric processing and parameter inversion. Taking further account of great regional variations in the land surface, it is difficult to transfer a trained deep-learning model for a local region to other regions, especially when learning high-level feature representations from limited training samples. For example, crop classification models trained in South China would fail in North China due to phenological differences. Current deep learning algorithms mostly apply to local regions (Zhou et al., 2019), and there is a lack of general transferable models.

6. Conclusions and future trends

In this article, we review the state-of-the-art deep-learning techniques in RSBD. First, we detailed the characters of RSBD. Then, after a brief overview of deep learning development and the pipeline of processing and analysis for RSBD, we comprehensively summarized deep learning techniques in RS processing and analysis, including geometric processing, radiometric processing, cloud masking, data fusion, object detection and extraction, LULC classification, change detection, and multitemporal analysis. Finally, the advantages and disadvantages of deep learning for processing and analyzing RSBD were discussed. Although deep learning techniques have achieved enormous success in almost all RS applications, this field is still relatively young, and there are many pending challenges and potential future work, such as:

- (1) Semisupervised or unsupervised learning methods. Deep learning algorithms need too many labeled samples for training, and collecting RS samples is time-consuming and difficult. Thus, with limited samples or even no sample available, how can deep learning algorithms be applied to process and analyze RS data?
- (2) Multitemporal analysis techniques. Compared with spatial and spectral analysis, temporal analysis is a young direction in RS. With an increasing number of multitemporal RS data available, how can deep learning be employed to indicate the evolution rules and predict the future of the ground surface?
- (3) End-to-end frameworks. Developing end-to-end frameworks to reduce the human-computer interaction and to largely improve the efficiency and automation level of RS applications.
- (4) Specialized deep learning models. From the mechanisms of RS and land surface processes, studying the characteristics of RS and its processing would guide the design of specialized deep learning models for RS and further improve RSBD applications in breadth and depth.

Finally, we hope this review reveals new possibilities for readers to further explore the remaining issues in developing novel deep-learning-based approaches for RSBD.

Notes

1. <https://www.planet.com/products/planet-imagery/> [2021-03-26]
2. <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2/acquisition-plans> [2021-03-26]
3. <https://medium.com/sentinel-hub/cloud-masks-at-your-service-6e5b2cb2ce8a> [2021-03-36]

Acknowledgments

The authors thank the anonymous reviewers for providing helpful and constructive comments that improved the manuscript substantially.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported in part by the National Key Research and Development Program under Grant [2017YFB0504201], the National Natural Science Foundation of China under Grant Nos. [42071316, 61473286 and 401201460], Open Fund of State Key Laboratory of Remote Sensing Science under Grant No. [OFSLRSS201919], the Fundamental Research Funds for the Central Universities under Grant No. [B200202008].

Notes on contributors



Xin Zhang received the Ph.D. degree in cartography and geography information system from the Institute of Geographic Sciences and Natural Resources, Chinese Academy of Sciences (CAS) in 2004. He is a professor in Aerospace Information Research Institute, CAS. His research interests include remote sensing big data with intelligent algorithms, and service of remote sensing information.




Ya'Nan Zhou received the Ph.D. degree in cartography and geography information system from the University of Chinese Academy of Sciences in 2014, and the B.S. degree in Geodesy Engineering from Wuhan University in 2009. He is an associate professor in Hohai University, China. His research interests include thematic information retrieval from satellite images and artificial intelligence applications in remote sensing.



Jiancheng Luo received the Ph.D. degree in cartography and geography information system from the Institute of Geographic Sciences and Natural Resources, Chinese Academy of Sciences (CAS) in 1999, and the B.S. degree in remote sensing from Zhejiang University in 1991. He is currently a professor in Aerospace Information Research Institute, CAS. His research interest is artificial intelligence techniques in remote sensing.

ORCID

Xin Zhang  <http://orcid.org/0000-0003-0394-7972>

Ya'nan Zhou  <http://orcid.org/0000-0002-4880-6439>

Data availability statement

Data sharing is not applicable to this article as no new data were created or analysed in this study.

References

- Abdollahi, A., Pradhan, B., Shukla, N., Chakraborty, S., & Alamri, A. (2020). Deep learning approaches applied to remote sensing datasets for road extraction: A state-of-the-art review. *Remote Sensing*, 12(9), 1444.
- Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3), 346–359.
- Benzenati, T., Kallel, A., & Kessentini, Y. (2020). Two stages pan-sharpening details injection approach based on very deep residual networks. *IEEE Transactions on Geoscience and Remote Sensing*, 59(6), 4984–4992.
- Bermudez, J. D., Happ, P. N., Feitosa, R. Q., & Oliveira, D. A. B. (2019). Synthesis of multispectral optical images from SAR/optical multitemporal data using conditional generative adversarial networks. *IEEE Geoscience and Remote Sensing Letters*, 16(8), 1220–1224.
- Brandt, M., Tucker, C. J., Kariryaa, A., Rasmussen, K., Abel, C., Small, J., . . . Fensholt, R. (2020). An unexpectedly large count of trees in the West African Sahara and Sahel. *Nature*, 587(7832), 78–82.
- Cao, R., Tu, W., Yang, C., Li, Q., Liu, J., Zhu, J., . . . Qiu, G. (2020). Deep learning-based remote and social sensing data fusion for urban region function recognition. *ISPRS Journal of Photogrammetry and Remote Sensing*, 163, 82–97.
- Censi, A. M., Ienco, D., Gbodjo, Y. J. E., Pensa, R. G., Interdonato, R., & Gaetano, R. (2021). Attentive spatial temporal graph CNN for land cover mapping from multi temporal remote sensing data. *IEEE Access*, 9, 23070–23082.
- Chen, X., Xiang, S., Liu, C. L., & Pan, C.-H. (2014). Vehicle detection in satellite images by hybrid deep convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*, 11(10), 1797–1801.
- Chen, Y., Fan, R., Yang, X., Wang, J., & Latif, A. (2018). Extraction of urban water bodies from high-resolution remote-sensing imagery using deep learning. *Water*, 10(5), 585.
- Chen, Z., Zhang, T., & Ouyang, C. (2018). End-to-end airplane detection using transfer learning in remote sensing images. *Remote Sensing*, 10(1), 139.
- Cheng, G., Han, J., Zhou, P., & Xu, D. (2019). Learning rotation-invariant and fisher discriminative convolutional neural networks for object detection. *IEEE Transactions on Image Processing*, 28(1), 265–278.
- Cheng, G., Zhou, P., & Han, J. (2016). Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 54(12), 7405–7415.
- Chung, J., Gulcehre, C., Cho, K. H. (2014). *Empirical evaluation of gated recurrent neural networks on sequence modeling*. arXiv preprint arXiv:1412.3555.
- Cui, W., He, X., Yao, M., Wang, Z., Hao, Y., Li, J., . . . Cui, W. (2021). Knowledge and spatial pyramid distance-based gated graph attention network for remote sensing semantic segmentation. *Remote Sensing*, 13(7), 1312.
- Dadsetan, S., Pichler, D., Wilson, D., Hovakimyan, N., & Hobbs, J. (2021). Superpixels and graph convolutional neural networks for efficient detection of nutrient deficiency stress from aerial imagery. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021: 2950–2959.
- Das, M., & Ghosh, S. K. (2016). Deep-STEP: A deep learning approach for spatiotemporal prediction of remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 13(12), 1984–1988.
- Deng, Z., Sun, H., Zhou, S., Zhao, J., Lei, L., & Zou, H. (2018). Multi-scale object detection in remote sensing imagery with convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145, 3–22.
- Ding, J., Xue, N., Long, Y., Xia, G., & Lu, Q. (2018). *Learning roi transformer for detecting oriented objects in aerial images*. arXiv preprint arXiv:1812.00155.
- Dong, G., Liao, G., Liu, H., & Kuang, G. (2018). A review of the autoencoder and its variants: A comparative perspective from target recognition in synthetic-aperture radar images. *IEEE Geoscience and Remote Sensing Magazine*, 6(3), 44–68.

- Du, B., Ru, L., Wu, C., & Zhang, L. (2019). Unsupervised deep slow feature analysis for change detection in multi-temporal remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 57(12), 9976–9992.
- Duan, W., Chiang, Y. Y., Leyk, S., Uhl, J. H., & Knoblock, C. A. (2020). Automatic alignment of contemporary vector data and georeferenced historical maps using reinforcement learning. *International Journal of Geographical Information Science*, 34(4), 824–849.
- Duffy, K. M., Vandal, T., Li, S., Ganguly, S., Nemani, R., & Ganguly, A. R. (2019). DeepEmSat: Deep emulation for satellite data mining. *Frontiers in Big Data*, 2, 42.
- Feng, J., Wu, X., Chen, J., Zhang, X., Tang, X., & Li, D. (2019). Joint multilayer spatial-spectral classification of hyperspectral images based on CNN and ConvLSTM. IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2019a: 588–591.
- Feng, Q., Yang, J., Liu, Y., Ou, C., Zhu, D., Niu, B., . . . Li, B. (2020). Multi-temporal unmanned aerial vehicle remote sensing for vegetable mapping using an attention-based recurrent convolutional neural network. *Remote Sensing*, 12(10), 1668.
- Feng, Q., Zhu, D., Yang, J., & Li, B. (2019b). Multisource hyperspectral and lidar data fusion for urban land-use mapping based on a modified two-branch convolutional neural network. *ISPRS International Journal of Geo-Information*, 8(1), 28.
- Feng, W., Sui, H., Huang, W., Xu, C., & An, K. (2018). Water body extraction from very high-resolution remote sensing imagery using deep U-Net and a superpixel-based conditional random field model. *IEEE Geoscience and Remote Sensing Letters*, 16(4), 618–622.
- Fu, T., Ma, L., Li, M., & Johnson, B. A. (2018). Using convolutional neural network to identify irregular segmentation objects from very high-resolution remote sensing imagery. *Journal of Applied Remote Sensing*, 12(2), 025010.
- Galar, M., Sesma, R., Ayala, C., Albizua, L., & Aranda, C. (2020). Super-resolution of Sentinel-2 images using convolutional neural networks and real ground truth data. *Remote Sensing*, 12(18), 2941.
- Gao, F., Dong, J., Li, B., & Xu, Q. (2016). Automatic change detection in synthetic aperture radar images based on PCANet. *IEEE Geoscience and Remote Sensing Letters*, 13(12), 1792–1796.
- Gao, X., Sun, X., Zhang, Y., Yan, M., Xu, G., Sun, H., . . . Fu, K. (2018). An end-to-end neural network for road extraction from remote sensing imagery by multiple feature pyramid network. *IEEE Access*, 6, 39401–39414.
- Gao, Y., Shi, J., & Wang, R. (2021). Remote sensing scene classification based on high-order graph convolutional network. *European Journal of Remote Sensing*, 141–155.
- Gers, F. A., Schraudolph, N. N., & Schmidhuber, J. (2002). Learning precise timing with LSTM recurrent networks. *Journal of Machine Learning Research*, 3(Aug), 115–143.
- Girard, N., Charpiat, G., & Tarabalka, Y. Aligning and updating cadaster maps with aerial images by multi-task, multi-resolution deep learning. Asian Conference on Computer Vision. Springer, Cham, 2018: 675–690.
- Gong, M., Yang, H., & Zhang, P. (2017). Feature learning and change feature classification based on deep learning for ternary change detection in SAR images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 129, 212–225.
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 2672–2680.
- Graves, A., Mohamed, A., & Hinton, G. Speech recognition with deep recurrent neural networks. (2013). IEEE international conference on acoustics, speech and signal processing. IEEE, 2013: 6645–6649. Vancouver, Canada.
- Grohnfeldt, C., Schmitt, M., & Zhu, X. A conditional generative adversarial network to fuse sar and multispectral optical data for cloud removal from Sentinel-2 images. IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2018: 1726–1729. Valencia, Spain.
- Gui, J., Sun, Z., Wen, Y., Tao, D., & Ye, J. (2020). A review on generative adversarial networks: Algorithms, theory, and applications. arXiv preprint arXiv:2001.06937.

- Guirado, E., Tabik, S., Rivas, M. L., Alcaraz-Segura, D., & Herrera, F. (2019). Whale counting in satellite and aerial images with deep learning. *Scientific Reports*, 9(1), 1–12.
- Guo, S., Lin, Y., Feng, N., Song, C., & Wan, H. (2019). Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2019, 33: 922–929.
- Halefom, A., Teshome, A., Sisay, E., & Ahmad, I. (2018). Dynamics of land use and land cover change using remote sensing and GIS: A case study of Debre Tabor Town, South Gondar, Ethiopia. *Journal of Geographic Information System*, 10(2), 165.
- Haut, J. M., Fernandez-Beltran, R., Paoletti, M. E., Plaza, J., Plaza, A., & Pla, F. (2018). A new deep generative network for unsupervised remote sensing single-image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 56(11), 6792–6810.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016: 770–778.
- He, W., & Yokoya, N. (2018). Multi-temporal sentinel-1 and-2 data fusion for optical image simulation. *ISPRS International Journal of Geo-Information*, 7(10), 389.
- Hewage, P., Behera, A., Trovati, M., Pereira, E., Ghahremani, M., Palmieri, F., & Liu, Y. (2020). Temporal convolutional neural (TCN) network for an effective weather forecasting using time-series data from the local weather station. *Soft Computing*, 1–30.
- Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313(5786), 504–507.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- Hou, X., Wang, B., Hu, W., Yin, L., & Wu, H. (2019). *SolarNet: A deep learning framework to map solar power plants in China from satellite imagery*. arXiv preprint arXiv:1912.03685.
- Hu, W., Li, H., Pan, L., Li, W., Tao, R., & Du, Q. (2019a). *Feature extraction and classification based on spatial-spectral convlstm neural network for hyperspectral images*. arXiv preprint arXiv:1905.03577.
- Hu, W. S., Li, H. C., Pan, L., Li, W., Tao, R., & Du, Q. (2020). Spatial-spectral feature extraction via deep ConvLSTM neural networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 58(6), 4237–4250.
- Hu, Y., Li, X., Zhou, N., Yang, L., & Peng, L. (2019b). A sample update-based convolutional neural network framework for object detection in large-area remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 16(6), 947–951.
- Huang, B., Zhao, B., & Song, Y. (2018). Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. *Remote Sensing of Environment*, 214, 73–86.
- Huang, J., Zhang, X., Xin, Q., Sun, Y., & Zhang, P. (2019). Automatic building extraction from high-resolution aerial images and LiDAR data using gated residual refinement network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 151, 91–105.
- Huang, W., Xiao, L., Wei, Z., Liu, H., & Tang, S. (2015). A new pan-sharpening method with deep neural networks. *IEEE Geoscience and Remote Sensing Letters*, 12(5), 1037–1041.
- Hughes, L. H., Schmitt, M., & Zhu, X. X. (2018). Mining hard negative samples for SAR-optical image matching using generative adversarial networks. *Remote Sensing*, 10(10), 1552.
- Hui, J., Du, M., Ye, X., Qin, Q., & Sui, J. (2018). Effective building extraction from high-resolution remote sensing images with multitask driven deep neural network. *IEEE Geoscience and Remote Sensing Letters*, 16(5), 786–790.
- Isikdogan, F., Bovik, A. C., & Passalacqua, P. (2017). Surface water mapping by deep learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(11), 4909–4918.
- Jeppesen, J. H., Jacobsen, R. H., Inceoglu, F., & Toftegaard, T. S. (2019). A cloud detection algorithm for satellite imagery based on deep learning. *Remote Sensing of Environment*, 229, 247–259.
- Ji, S., Wei, S., & Lu, M. (2018). Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1), 574–586.

- Ji, S., Wei, S., & Lu, M. (2019). A scale robust convolutional neural network for automatic building extraction from aerial and satellite imagery. *International Journal of Remote Sensing*, 40(9), 3308–3322.
- Jia, D., Song, C., Cheng, C., Shen, S., Ning, L., & Hui, C. (2020). A novel deep learning-based spatiotemporal fusion method for combining satellite images with different resolutions using a two-stream convolutional neural network. *Remote Sensing*, 12(4), 698.
- Jiang, K., Wang, Z., Yi, P., Wang, G., Lu, T., & Jiang, J. (2019). Edge-enhanced GAN for remote sensing image superresolution. *IEEE Transactions on Geoscience and Remote Sensing*, 57(8), 5799–5812.
- Khan, N., Chaudhuri, U., Banerjee, B., & Chaudhuri, S. (2019). Graph convolutional network for multi-label VHR remote sensing scene recognition. *Neurocomputing*, 357, 36–46.
- Kipf, T. N., & Welling, M. (2016a). *Semi-supervised classification with graph convolutional networks*. arXiv preprint arXiv:1609.02907.
- Kipf, T. N., & Welling, M. (2016b). *Variational graph auto-encoders*. arXiv preprint arXiv:1611.07308.
- Lea, C., Vidal, R., Reiter, A., & Hager, G. D. (2016). Temporal convolutional networks: A unified approach to action segmentation. European Conference on Computer Vision. Springer, Cham, 2016: 47–54.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.
- Lei, S., Shi, Z., & Zou, Z. (2017). Super-resolution for remote sensing images via local-global combined network. *IEEE Geoscience and Remote Sensing Letters*, 14(8), 1243–1247.
- Li, J., Li, Y., He, L., Chen, J., & Plaza, A. (2020a). Spatio-temporal fusion for remote sensing data: An overview and new benchmark. *Information Sciences*, 63(140301), 1–140301.
- Li, K., Cheng, G., Bu, S., & You, X. (2018). Rotation-insensitive and context-augmented object detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 56(4), 2337–2348.
- Li, K., Wan, G., Cheng, G., Meng, L., & Han, J. (2020b). Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing*, 159, 296–307.
- Li, L., Yan, Z., Shen, Q., Cheng, G., Gao, L., & Zhang, B. (2019a). Water body extraction from very high spatial resolution remote sensing data based on fully convolutional networks. *Remote Sensing*, 11(10), 1162.
- Li, M., Zhang, Z., Lei, L., Wang, X., & Guo, X. (2020c). Agricultural greenhouses detection in high-resolution satellite images based on convolutional neural networks: Comparison of faster R-CNN, YOLO v3 and SSD. *Sensors*, 20(17), 4938.
- Li, X., Lei, L., Sun, Y., Li, M., & Kuang, G. (2020d). Multimodal bilinear fusion network with second-order attention-based channel selection for land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 1011–1026.
- Li, X., Yao, X., & Fang, Y. (2018a). Building-a-nets: Robust building extraction from high-resolution remote sensing images with adversarial networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(10), 3680–3687.
- Li, Y., Chen, W., Zhang, Y., Tao, C., Xiao, R., & Tan, Y. (2020e). Accurate cloud detection in high-resolution remote sensing imagery by weakly supervised deep learning. *Remote Sensing of Environment*, 250, 112045.
- Li, Y., Peng, C., Chen, Y., Jiao, L., Zhou, L., & Shang, R. (2019b). A deep learning method for change detection in synthetic aperture radar images. *IEEE Transactions on Geoscience and Remote Sensing*, 57(8), 5751–5763.
- Li, Y., Zhang, H., Xue, Y., Jiang, Y., & Shen, Q. (2018b). Deep learning for remote sensing image classification: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(6), e1264.
- Li, Y., Zhang, Y., Huang, X., & Yuille, A. L. (2018c). Deep networks under scene-level supervision for multi-class geospatial object detection from remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146, 182–196.

- Liang, J., Deng, Y., & Zeng, D. (2020). A deep neural network combined CNN and GCN for remote sensing scene classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 4325–4338.
- Liu, H., Luo, J., Huang, B., Hu, X., Sun, Y., Yang, Y., Xu, N., & Zhou, N. (2019a). DE-Net: Deep encoding network for building extraction from high-resolution remote sensing imagery. *Remote Sensing*, 11(20), 2380.
- Liu, H., & Weng, Q. (2018a). Scaling effect of fused ASTER-MODIS land surface temperature in an urban environment. *Sensors*, 18(11), 4058.
- Liu, J., Gong, M., Qin, K., & Zhang, P. (2016a). A deep convolutional coupling network for change detection based on heterogeneous optical and radar images. *IEEE Transactions on Neural Networks and Learning Systems*, 29(3), 545–559.
- Liu, L., Wang, J., Zhang, E., Li, B., Zhu, X., Zhang, Y., & Peng, J. (2020a). Shallow–deep convolutional network and spectral-discrimination-based detail injection for multispectral imagery pan-sharpening. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 1772–1783.
- Liu, P. (2015). A survey of remote-sensing big data. *Frontiers in Environmental Science*, 3, 45.
- Liu, P., Liu, X., Liu, M., Shi, Q., Yang, J., Xu, X., & Zhang, Y. (2019b). Building footprint extraction from high-resolution images via spatial residual inception convolutional neural network. *Remote Sensing*, 11(7), 830.
- Liu, Q., Kampffmeyer, M., Jenssen, R., & Salberg, A. (2020b). SCG-Net: Self-constructing graph neural networks for semantic segmentation. arXiv preprint arXiv:2009.01599.
- Liu, Q., Xu, L., & Zhang, Z. The downscaling of the SMOS global sea surface salinity product based on MODIS data using a deep convolution network approach. Proceedings of the 2019 3rd International Conference on Advances in Image Processing. 2019c: 97–100. Chengdu, China.
- Liu, R., Jiang, D., Zhang, L., & Zhang, Z. (2020c). Deep depthwise separable convolutional network for change detection in optical aerial images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 1109–1118.
- Liu, W., Wang, J., Luo, J., Wu, Z., Chen, J., Zhou, Y., . . . Yang, Y. (2020d). Farmland parcel mapping in mountain areas using time-series SAR data and VHR optical images. *Remote Sensing*, 12(22), 3733.
- Liu, X., Deng, C., Chanussot, J., Hong, D., & Zhao, B. (2019d). StfNet: A two-stream convolutional neural network for spatiotemporal image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9), 6552–6564.
- Liu, Y., Chen, X., Wang, Z., Wang, Z. J., Ward, R. K., & Wang, X. (2018b). Deep learning for pixel-level image fusion: Recent advances and future prospects. *Information Fusion*, 42, 158–173.
- Liu, Y., Gross, L., Li, Z., Li, X., Fan, X., & Qi, W. (2019e). Automatic building extraction on high-resolution remote sensing imagery using deep convolutional encoder-decoder with spatial pyramid pooling. *IEEE Access*, 7, 128774–128786.
- Liu, Y., Yao, J., Lu, X., Xia, M., Wang, X., & Liu, Y. (2018c). Roadnet: Learning to comprehensively analyze road networks in complex urban scenes from high-resolution remotely sensed images. *IEEE Transactions on Geoscience and Remote Sensing*, 57(4), 2043–2056.
- Liu, Z., Wang, H., Weng, L., & Yang, Y. (2016b). Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds. *IEEE Geoscience and Remote Sensing Letters*, 13(8), 1074–1078.
- Long, Y., Gong, Y., Xiao, Z., & Liu, Q. (2017). Accurate object localization in remote sensing images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 55(5), 2486–2498.
- López-Serrano, P. M., López-Sánchez, C. A., Álvarez-González, J. G., & García-Gutiérrez, J. (2016). A comparison of machine learning techniques applied to landsat-5 TM spectral data for biomass estimation. *Canadian Journal of Remote Sensing*, 42(6), 690–705.
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91–110.
- Lu, X., Zhong, Y., Zheng, Z., Liu, Y., Zhao, J., Ma, A., & Yang, J. (2019). Multi-scale and multi-task deep learning framework for automatic road extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 57(11), 9362–9377.

- Lv, X., Ming, D., Lu, T., Zhou, K., Wang, M., & Bao, H. (2018). A new method for region-based majority voting CNNs for very high resolution image classification. *Remote Sensing*, 10(12), 1946.
- Lyu, H., Lu, H., & Mou, L. (2016). Learning a transferable change rule from a recurrent neural network for land cover change detection. *Remote Sensing*, 8(6), 506.
- Ma, J., Yu, W., Chen, C., Liang, P., Guo, X., & Jiang, J. (2020). Pan-GAN: An unsupervised pan-sharpening method for remote sensing image fusion. *Information Fusion*, 62, 110–120.
- Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019a). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152, 166–177.
- Ma, W., Pan, Z., Guo, J., & Lei, B. (2019b). Achieving super-resolution remote sensing images via the wavelet transform combined with the recursive res-net. *IEEE Transactions on Geoscience and Remote Sensing*, 57(6), 3512–3527.
- Ma, W., Zhang, J., Wu, Y., Jiao, L., Zhu, H., & Zhao, W. (2019c). A novel two-step registration method for remote sensing images based on deep and local features. *IEEE Transactions on Geoscience and Remote Sensing*, 57(7), 4834–4843.
- Maltezos, E., Doulamis, N., Doulamis, A., & Ioannidis, C. (2017). Deep convolutional neural networks for building extraction from orthoimages and dense image matching point clouds. *Journal of Applied Remote Sensing*, 11(4), 042620.
- Marmanis, D., Schindler, K., Wegner, J. D., Galliani, S., Datcu, M., & Stilla, U. (2018). Classification with an edge: Improving semantic image segmentation with boundary detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 135, 158–172.
- Masi, G., Cozzolino, D., Verdoliva, L., & Scarpa, G. (2016). Pansharpening by convolutional neural networks. *Remote Sensing*, 8(7), 594.
- Mateo-García, G., Laparra, V., López-Puigdollers, D., & Gómez-Chova, L. (2020). Transferring deep learning models for cloud detection between Landsat-8 and Proba-V. *ISPRS Journal of Photogrammetry and Remote Sensing*, 160, 1–17.
- Merkle, N., Auer, S., Müller, R., & Reinartz, P. (2018). Exploring the potential of conditional adversarial networks for optical and SAR image matching. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(6), 1811–1820.
- Mou, L., Bruzzone, L., & Zhu, X. X. (2018). Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2), 924–935.
- Mou, L., Lu, X., Li, X., & Zhu, X. X. (2020). Nonlocal graph convolutional networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 58(12), 8246–8257.
- Ogut, M., Bosch-Lluis, X., & Reising, S. C. (2019). Deep learning calibration of the high-frequency airborne microwave and millimeter-wave radiometer (HAMMR) instrument. *IEEE Transactions on Geoscience and Remote Sensing*, 58(5), 3391–3399.
- Ouyang, S., & Li, Y. (2021). Combining deep semantic segmentation network and graph convolutional neural network for semantic segmentation of remote sensing imagery. *Remote Sensing*, 13(1), 119.
- Pan, X., Zhao, J., & Xu, J. (2020). An end-to-end and localized post-processing method for correcting high-resolution remote sensing classification result images. *Remote Sensing*, 12(5), 852.
- Peng, D., Zhang, Y., & Guan, H. (2019). End-to-end change detection for high resolution satellite images using improved UNet++. *Remote Sensing*, 11(11), 1382.
- Peng, H., Wang, H., Du, B., Bhuiyan, M. Z. A., Ma, H., Liu, J., & Yu, P. S. (2020). Spatial temporal incidence dynamic graph neural networks for traffic flow forecasting. *Information Sciences*, 521, 277–290.
- Peng, J., Wang, D., Liao, X., Shao, Q., Sun, Z., Yue, H., & Ye, H. (2020). Wild animal survey using UAS imagery and deep learning: Modified Faster R-CNN for kiang detection in Tibetan Plateau. *ISPRS Journal of Photogrammetry and Remote Sensing*, 169, 364–376.
- Qin, A., Shang, Z., Tian, J., Wang, Y., Zhang, T., & Tang, Y. Y. (2018). Spectral-spatial graph convolutional networks for semisupervised hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 16(2), 241–245.

- Qiu, S., Zhu, Z., & He, B. (2019). Fmask 4.0: Improved cloud and cloud shadow detection in Landsats 4–8 and Sentinel-2 imagery. *Remote Sensing of Environment*, 231, 111205.
- Rahman, M. S., & Di, L. (2017). The state of the art of spaceborne remote sensing in flood management. *Natural Hazards*, 85(2), 1223–1248.
- Ren, Y., Zhu, C., & Xiao, S. (2018). Small object detection in optical remote sensing images via modified faster R-CNN. *Applied Sciences*, 8(5), 813.
- Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. 2011 International conference on computer vision. IEEE, 2011: 2564–2571.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1985). *Learning internal representations by error propagation[R]*. California Univ San Diego La Jolla Inst for Cognitive Science. La Jolla, CA, USA.
- Saha, S., Mou, L., Zhu, X., & Bovolo, F. (2020). Semisupervised change detection using graph convolutional network. *IEEE Geoscience and Remote Sensing Letters*. doi:10.1109/LGRS.2020.2985340
- Scarpa, G., Gargiulo, M., Mazza, A., & Gaetano, R. (2018). A cnn-based fusion method for feature extraction from sentinel data. *Remote Sensing*, 10(2), 236.
- Seo, D. K., Kim, Y. H., Eo, Y. D., Park, W., & Park, H. (2017). Generation of radiometric, phenological normalized image based on random forest regression for change detection. *Remote Sensing*, 9(11), 1163.
- Shao, Z., Cai, J., Fu, P., Hu, L., & Liu, T. (2019). Deep learning-based fusion of Landsat-8 and Sentinel-2 images for a harmonized surface reflectance product. *Remote Sensing of Environment*, 235, 111425.
- Shao, Z., Deng, J., Wang, L., Fan, Y., Sumari, N., & Cheng, Q. (2017). Fuzzy autoencode based cloud detection for remote sensing imagery. *Remote Sensing*, 9(4), 311.
- Sharma, A., Liu, X., Yang, X., & Shi, D. (2017). A patch-based convolutional neural network for remote sensing image classification. *Neural Networks*, 95, 19–28.
- Shen, H., Zhou, M., Li, T., & Zeng, C. (2019). Integration of remote sensing and social sensing data in a deep learning framework for Hourly Urban PM2. 5 mapping. *International Journal of Environmental Research and Public Health*, 16(21), 4102.
- Shi, M., Xie, F., Zi, Y., & Yin, J. (2016). Cloud detection of remote sensing images by deep learning. 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, 2016: 701–704.
- Shi, X., Chen, Z., Wang, H., & Yeung D. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems*, 28, 802–810.
- Shi, Y., Li, Q., & Zhu, X. X. (2020). Building segmentation through a gated graph convolutional neural network with deep structured feature embedding. *ISPRS Journal of Photogrammetry and Remote Sensing*, 159, 184–197.
- Song, H., Liu, Q., Wang, G., Hang, R., & Huang, B. (2018). Spatiotemporal satellite image fusion using deep convolutional neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(3), 821–829.
- Song, S., Liu, J., Liu, Y., Feng, G., Han, H., Yao, Y., & Du, M. (2020). Intelligent object recognition of urban water bodies based on deep learning for multi-source and multi-temporal high spatial resolution remote sensing imagery. *Sensors*, 20(2), 397.
- Sun, L., Mi, X., Wei, J., Wang, J., Tian, X., Yu, H., & Gan, P. (2017). A cloud detection algorithm-generating method for remote sensing data at visible to short-wave infrared wavelengths. *ISPRS Journal of Photogrammetry and Remote Sensing*, 124, 70–88.
- Sun, Z., Di, L., & Fang, H. (2019). Using long short-term memory recurrent neural network in land cover classification on Landsat and Cropland data layer time series. *International Journal of Remote Sensing*, 40(2), 593–614.
- Tan, Z., Yue, P., Di, L., & Tang, J. (2018). Deriving high spatiotemporal remote sensing images using deep convolutional network. *Remote Sensing*, 10(7), 1066.
- Teimouri, N., Dyrmann, M., & Jørgensen, R. N. (2019). A novel spatio-temporal FCN-LSTM network for recognizing various crop types using multi-temporal radar images. *Remote Sensing*, 11(8), 990.

- Timilsina, S., Aryal, J., & Kirkpatrick, J. B. (2020). Mapping urban tree cover changes using object-based convolution neural network (OB-CNN). *Remote Sensing*, 12(18), 3017.
- Tong, X., Xia, G., Lu, Q., Shen, H., Li, S., You, S., & Zhang, L. (2018). *Learning transferable deep models for land-use classification with high-resolution remote sensing images*. arXiv preprint arXiv:1807.05713.
- Tong, X. Y., Xia, G. S., Lu, Q., Shen, H., Li, S., You, S., & Zhang, L. (2020). Land-cover classification with high-resolution remote sensing images using transferable deep models. *Remote Sensing of Environment*, 237, 111322.
- Tuna, C., Unal, G., & Sertel, E. (2018). Single-frame super resolution of remote-sensing images by convolutional neural networks. *International Journal of Remote Sensing*, 39(8), 2463–2479.
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2017). *Graph attention networks*. arXiv preprint arXiv:1710.10903.
- Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational Intelligence and Neuroscience*, 2018, 1–13.
- Waldner, F., & Diakogiannis, F. I. (2020). Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network. *Remote Sensing of Environment*, 245, 111741.
- Wan, S., Gong, C., Zhong, P., Du, B., Zhang, L., & Yang, J. (2019). Multiscale dynamic graph convolutional network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 58(5), 3162–3177.
- Wan, S., Gong, C., Zhong, P., Du, B., Zhang, L., & Yang, J. (2020). Multiscale dynamic graph convolutional network for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 58(5), 3162–3177.
- Wang, C., Bai, X., Wang, S., Zhou, J., & Ren, P. (2018a). Multiscale visual attention networks for object detection in VHR remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 16(2), 310–314.
- Wang, J., Zhang, J., Guo, Q., & Li, T. (2019). Fusion of hyperspectral and lidar data based on dual-branch convolutional neural network. IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2019a: 3388–3391.
- Wang, Q., Yuan, Z., Du, Q., & Li, X. (2018b). GETNET: A general end-to-end 2-D CNN framework for hyperspectral image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1), 3–13.
- Wang, Q., Zhang, X., Chen, G., Dai, F., Gong, Y., & Zhu, K. (2018c). Change detection based on faster R-CNN for high-resolution remote sensing images. *Remote Sensing Letters*, 9(10), 923–932.
- Wang, S., Quan, D., Liang, X., Ning, M., Guo, Y., & Jiao, L. (2018d). A deep learning framework for remote sensing image registration. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145, 148–164.
- Wang, T., Li, Y., Yu, S., & Liu, Y. (2019b). Estimating the volume of oil tanks based on high-resolution remote sensing images. *Remote Sensing*, 11(7), 793.
- Wang, W., Li, H., Deng, Y., Shao, L., Lu, X., & Du, Q. (2020). Generative adversarial capsule network with ConvLSTM for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*.
- Wang, W., Ma, L., Chen, M., & Du, Q. (2021). Joint correlation alignment-based graph neural network for domain adaptation of multitemporal hyperspectral remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 3170–3184.
- Wang, W., & Shen, J. (2017). Deep visual attention prediction. *IEEE Transactions on Image Processing*, 27(5), 2368–2378.
- Wei, Y., Yuan, Q., Shen, H., & Zhang, L. (2017). Boosting the accuracy of multispectral image pansharpening by learning a deep residual network. *IEEE Geoscience and Remote Sensing Letters*, 14(10), 1795–1799.
- Wen, C., Li, X., Yao, X., Peng, L., & Chi, T. (2021). Airborne LiDAR point cloud classification with global-local graph attention convolution neural network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 173, 181–194.
- Widyaningrum, E., Bai, Q., Fajari, M. K., & Lindenbergh, R. C. (2021). Airborne laser scanning point cloud classification using the DGCNN deep learning method. *Remote Sensing*, 13(5), 859.

- Wu, W., Gao, X., Fan, J., Xia, L., Luo, J., & Zhou, Y. (2020a). Improved mask R-CNN-based cloud masking method for remote sensing images. *International Journal of Remote Sensing*, 41(23), 8910–8933.
- Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P.S. (2020b). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems* 32(1): 4–24.
- Xia, G., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., & Zhang, L. (2018). DOTA: A large-scale dataset for object detection in aerial images. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018: 3974–3983.
- Xie, F., Shi, M., Shi, Z., Yin, J., & Zhao, D. (2017). Multilevel cloud detection in remote sensing images based on deep learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(8), 3631–3640.
- Xing, Y., Wang, M., Yang, S., & Jiao, L. (2018). Pan-sharpening via deep metric learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145, 165–183.
- Xu, F., Cervone, G., Franch, G., & Salvador, M. (2020). Multiple geometry atmospheric correction for image spectroscopy using deep learning. *Journal of Applied Remote Sensing*, 14(2), 024518.
- Xu, Q., Hou, Z., & Tokola, T. (2012). Relative radiometric correction of multi-temporal ALOS AVNIR-2 data for the estimation of forest attributes. *ISPRS Journal of Photogrammetry and Remote Sensing*, 68, 69–78.
- Yan, J., Chen, X., Chen, Y., & Liang, D. (2020a). Multistep prediction of land cover from dense time series remote sensing images with temporal convolutional networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 5149–5161.
- Yan, J., Mu, L., Wang, L., Ranjan, R., & Zomaya, A. Y. (2020b). temporal convolutional networks for the advance prediction of ENSO. *Scientific Reports*, 10(1), 1–15.
- Yan, J., Wang, H., Yan, M., Diao, W., Sun, X., & Li, H. (2019). IoU-adaptive deformable R-CNN: Make full use of IoU for multi-class object detection in remote sensing imagery. *Remote Sensing*, 11(3), 286.
- Yang, H., Wu, P., Yao, X., Wu, Y., Wang, B., & Xu, Y. (2018b). Building extraction in very high resolution imagery by dense-attention networks. *Remote Sensing*, 10(11), 1768.
- Yang, H., Yu, B., Luo, J., & Chen, F. (2019). Semantic segmentation of high spatial resolution images with deep neural networks. *GIScience & Remote Sensing*, 56(5), 749–768.
- Yang, H. L., Yuan, J., Lunga, D., Laverdiere, M., Rose, A., & Bhaduri, B. (2018a). Building extraction at scale using convolutional neural network: Mapping of the united states. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(8), 2600–2614.
- Yang, J., Fu, X., Hu, Y., Huang, Y., Ding, X., & Paisley, J. (2017). PanNet: A deep network architecture for pan-sharpening. *Proceedings of the IEEE International Conference on Computer Vision*. 2017: 5449–5457.
- Yang, J., Gong, P., Fu, R., Zhang, M., Chen, J., Liang, S., . . . Dickinson, R. (2013). The role of satellite remote sensing in climate change studies. *Nature Climate Change*, 3(10), 875–883.
- Yang, Z., Dan, T., & Yang, Y. (2018c). Multi-temporal remote sensing image registration using deep convolutional features. *IEEE Access*, 6, 38544–38555.
- Yao, H., Wu, F., Ke, J., Tang, X., Jia, Y., Lu, S., Gong, P., Ye, J., & Li, Z. (2018). *Deep multi-view spatial-temporal network for taxi demand prediction*. arXiv preprint arXiv:1802.08714.
- Ye, F., Su, Y., Xiao, H., Zhao, X., & Min, W. (2018). Remote sensing image registration using convolutional neural network features. *IEEE Geoscience and Remote Sensing Letters*, 15(2), 232–236.
- Yin, Z., Ling, F., Foody, G. M., Li, X., & Du, Y. (2020). Cloud detection in Landsat-8 imagery in Google Earth Engine based on a deep convolutional neural network. *Remote Sensing Letters*, 11(12), 1181–1190.
- Yin, Z., Wu, P., Foody, G. M., Wu, Y., Liu, Z., Du, Y., & Ling, F. (2020). Spatiotemporal fusion of land surface temperature based on a convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 59(2), 1808–1822.
- Yuan, J. (2016). *Automatic building extraction in aerial scenes using convolutional networks*. arXiv preprint arXiv:1602.06564.
- Yuan, K., Meng, G., Cheng, D., Bai, J., Xiang, S., & Pan, C. (2017, September). Efficient cloud detection in remote sensing images using edge-aware segmentation network and easy-to-hard training

- strategy. In [2017 IEEE International Conference on Image Processing \(ICIP\)](#) (pp. 61–65). IEEE. Beijing, China.
- Zhan, Y., Wang, J., Shi, J., Cheng, G., Yao, L., & Sun, W. (2017). Distinguishing cloud and snow in satellite images via deep convolutional network. *IEEE Geoscience and Remote Sensing Letters*, 14(10), 1785–1789.
- Zhang, C., Harrison, P. A., Pan, X., Li, H., Sargent, I., & Atkinson, P. M. (2020a). Scale Sequence Joint Deep Learning (SS-JDL) for land use and land cover classification. *Remote Sensing of Environment*, 237, 111593.
- Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., & Atkinson, P. M. (2019a). Joint deep learning for land cover and land use classification. *Remote Sensing of Environment*, 221, 173–187.
- Zhang, C., Wei, S., Ji, S., & Lu, M. (2019b). Detecting large-scale urban land cover changes from very high resolution remote sensing images using cnn-based classification. *ISPRS International Journal of Geo-Information*, 8(4), 189.
- Zhang, G., Lu, H., Dong, J., Poslad, S., Li, R., Zhang, X., & Rui, X. (2020b). A framework to predict high-resolution spatiotemporal PM2.5 distributions using a deep-learning model: A case study of Shijiazhuang, China. *Remote Sensing*, 12(17), 2825.
- Zhang, H., & Deng, Q. (2019c). Deep learning based fossil-fuel power plant monitoring in high resolution remote sensing images: A comparative study. *Remote Sensing*, 11(9), 1117.
- Zhang, H., Ni, W., Yan, W., Xiang, D., Wu, J., Yang, X., & Bian, H. (2019d). Registration of multimodal remote sensing image based on deep fully convolutional neural network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(8), 3028–3042.
- Zhang, L., Shi, Z., & Wu, J. (2015). A hierarchical oil tank detector with deep surrounding features for high-resolution optical satellite imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(10), 4895–4909.
- Zhang, L., Zhang, L., & Du, B. (2016a). Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, 4(2), 22–40.
- Zhang, M., Xu, G., Chen, K., Yan, M., & Sun, X. (2018a). Triplet-based semantic relation learning for aerial remote sensing image change detection. *IEEE Geoscience and Remote Sensing Letters*, 16(2), 266–270.
- Zhang, P., Gong, M., Su, L., Liu, J., & Li, Z. (2016b). Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 116, 24–41.
- Zhang, T., & Huang, X. (2018b). Monitoring of urban impervious surfaces using time series of high-resolution remote sensing images in rapidly urbanized areas: A case study of Shenzhen. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(8), 2692–2708.
- Zhang, Y., Yuan, Y., Feng, Y., & Lu, X. (2019e). Hierarchical and robust convolutional neural network for very high-resolution remote sensing object detection. *IEEE Transactions on Geoscience and Remote Sensing*, 57(8), 5535–5548.
- Zhang, Z., Guo, W., Zhu, S., & Yu, W. (2018c). Toward arbitrary-oriented ship detection with rotated region proposal and discrimination networks. *IEEE Geoscience and Remote Sensing Letters*, 15(11), 1745–1749.
- Zhang, Z., Liu, Q., & Wang, Y. (2018d). Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5), 749–753.
- Zhang, Z., & Wang, Y. (2019f). JointNet: A common neural network for road and building extraction. *Remote Sensing*, 11(6), 696.
- Zhao, H., Chen, Z., Jiang, H., Jing, W., Sun, L., & Feng, M. (2019). Evaluation of three deep learning models for early crop classification using Sentinel-1A imagery time series—A case study in Zhanjiang, China. *Remote Sensing*, 11(22), 2673.
- Zhao, W., Wang, Z., Gong, M., & Liu, J. (2017). Discriminative feature learning for unsupervised change detection in heterogeneous images based on a coupled neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 55(12), 7066–7080.
- Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., & Sun, M. (2018). Graph Neural Networks: A Review of Methods and Applications. [arXiv preprint arXiv:1812.08434](#).

- Zhou, L., Zhang, C., & Wu, M. D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. CVPR Workshops. 2018b: 182–186. Salt Lake City, UT, USA.
- Zhou, P., Cheng, G., Liu, Z., Bu, S., & Hu, X. (2016). Weakly supervised target detection in remote sensing images based on transferred deep features and negative bootstrapping. *Multidimensional Systems and Signal Processing*, 27(4), 925–944.
- Zhou, Y., & Grassotti, C. (2020a). Development of a machine learning-based radiometric bias correction for NOAA's microwave integrated retrieval system (MIRS). *Remote Sensing*, 12(19), 3160.
- Zhou, Y., Li, J., Feng, L., Zhang, X., & Hu, X. (2017). Adaptive scale selection for multiscale segmentation of satellite images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(8), 3641–3651.
- Zhou, Y., Luo, J., Feng, L., Yang, Y., Chen, Y., & Wu, W. (2019). Long-short-term-memory-based crop classification using high-resolution optical images and multi-temporal SAR data. *GIScience & Remote Sensing*, 56(8), 1170–1191.
- Zhou, Y., Luo, J., Shen, Z., Hu, X., & Yang, H. (2014). Multiscale water body extraction in urban environments from satellite images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(10), 4301–4312.
- Zhou, Y. N., Yang, X., Feng, L., Wu, W., Wu, T., Luo, J., & Zhang, X. (2020). Superpixel-based time-series reconstruction for optical images incorporating SAR data using autoencoder networks. *GIScience & Remote Sensing*, 57(8), 1005–1025.
- Zhu, H., Jiao, L., Ma, W., Liu, F., & Zhao, W. (2019). A novel neural network for remote sensing image matching. *IEEE Transactions on Neural Networks and Learning Systems*, 30(9), 2853–2865.
- Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8–36.