ISPRS Journal of Photogrammetry and Remote Sensing 152 (2019) 166–177

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Review Article

Deep learning in remote sensing applications: A meta-analysis and review

Lei Ma[a,b,e,g,*], Yu Liu[c], Xueliang Zhang[a], Yuanxin Ye[d], Gaofei Yin[d], Brian Alan Johnson[f]

[a] School of Geography and Ocean Science, Nanjing University, Nanjing 210023, China
[b] Department of Geosciences, Texas Tech University, Lubbock, TX 79409, United States
[c] Department of Biomedical Engineering, Hefei University of Technology, Hefei 230009, China
[d] Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 610031, China
[e] Signal Processing in Earth Observation, Technical University of Munich, Munich 80333, Germany
[f] Natural Resources and Ecosystem Services, Institute for Global Environmental Strategies, 2018-11, Kamiyamaguchi, Hayama, Kanagawa 240-0115, Japan
[g] Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, 82234 Wessling, Germany

ARTICLE INFO

Keywords:
Deep learning (DL)
Remote sensing
LULC classification
Object detection
Scene classification

ABSTRACT

Deep learning (DL) algorithms have seen a massive rise in popularity for remote-sensing image analysis over the past few years. In this study, the major DL concepts pertinent to remote-sensing are introduced, and more than 200 publications in this field, most of which were published during the last two years, are reviewed and analyzed. Initially, a meta-analysis was conducted to analyze the status of remote sensing DL studies in terms of the study targets, DL model(s) used, image spatial resolution(s), type of study area, and level of classification accuracy achieved. Subsequently, a detailed review is conducted to describe/discuss how DL has been applied for remote sensing image analysis tasks including image fusion, image registration, scene classification, object detection, land use and land cover (LULC) classification, segmentation, and object-based image analysis (OBIA). This review covers nearly every application and technology in the field of remote sensing, ranging from preprocessing to mapping. Finally, a conclusion regarding the current state-of-the art methods, a critical conclusion on open challenges, and directions for future research are presented.

1. Introduction

Remote sensing of images has been successfully applied in many fields, such as classification and change detection. However, remote-sensing image processing involves a few preprocessing procedures in addition to classification and change detection; furthermore, it is highly dependent on the method that is applied. Hence, the remote-sensing community is always committed to developing remote-sensing methods for improving the performance of aspects, such as preprocessing, segmentation, and classification. Neural networks, the basis of deep learning (DL) algorithms, have been used in the remote sensing community for many years. However, prior to the development of DL, the remote-sensing community had shifted its focus from neural networks to support vector machine (SVM) and ensemble classifiers, e.g., random forest (RF), for image classification and other tasks (e.g. change detection). SVM received much attention due to its ability to handle high dimensionality data and perform well with limited training samples, among other things (Mountrakis et al., 2011), while RF gained popularity and ease of use (e.g. relatively insensitive to classification parameters) and generally high accuracy (Belgiu and Drăguţ, 2016). In more recent years, however, the advent of DL has led to renewed interest in neural networks. Since 2014, the remote-sensing community has shifted its attention to DL, and DL algorithms have achieved significant success at many image analysis tasks including land use and land cover (LULC) classification, scene classification, and object detection (Chen et al., 2014; Zou et al., 2015; Chen et al., 2015; Romero et al., 2016; Cheng et al., 2016; Marmanis et al., 2016; Yu et al., 2017; Kussul et al., 2017; Sharma et al., 2017; Vetrivel et al., 2018).

Until now, most studies reviewing DL have either been general reviews concerning the development of the DL algorithm (LeCun et al., 2015; Zhang et al, 2016), or detailed topical reviews for a few hot fields, e.g., speech recognition (Hinton et al., 2012) and medical image recognition (Litjens et al., 2017). Although a few studies involved a review of DL applications in remote sensing (Liu et al., 2018; Li et al., 2018a), many important areas of the field have not been researched. For example, the review by Liu et al. (2018) focused specifically on the application of DL for remote-sensing data fusion, and ignored its application in other important areas of remote sensing (e.g.,

* Corresponding author.
E-mail addresses: maleinju@gmail.com (L. Ma), yuliu@hfut.edu.cn (Y. Liu), zxl@nju.edu.cn (X. Zhang), yeyuanxin110@163.com (Y. Ye), yingf@swjtu.edu.cn (G. Yin), johnson@iges.or.jp (B.A. Johnson).

https://doi.org/10.1016/j.isprsjprs.2019.04.015
Received 19 November 2018; Received in revised form 21 March 2019; Accepted 22 April 2019
Available online 28 April 2019

classification). In other words, a systematic and comprehensive review concerning the application of DL in the field of remote sensing has not yet been conducted. Zhu et al. (2017) performed a broader review than others, but they focused on some relatively uncommon sub-areas of the remote sensing field (e.g., 3D modeling applications) while ignoring many other deserving (and more common) sub-areas; e.g., image classification applications. DL algorithms have been utilized in many diverse sub-areas within the remote sensing field, and the quantity of the related publications is currently increasing remarkably, so it appears that a more systematic (i.e. quantitative) analysis is necessary to get a comprehensive and objective understanding of the applications of DL for remote-sensing analysis. For example, the different applications for which DL was applied and the problems encountered in such studies are beneficial information for researchers interested in DL and remote sensing.

Therefore, the motivation for our study was to conduct a comprehensive review of almost all major sub-areas of the remote sensing field having connections with DL, including image fusion, image registration, scene classification, object detection, LULC classification, image segmentation, object-based image analysis (OBIA), and other tasks. Through a meta-analysis, we identified and categorized the publications related to DL and remote sensing, and summarized the main scientific advances noted in the literature. Finally, a conclusion, critical summary, and outlook toward future research are given.

2. DL works focusing on algorithmic advancements

DL is a learning algorithm based on neural networks (Schmidhuber, 2015). A neural network comprises neurons or units with certain activation a and parameters $\Theta = \{W, \gamma\}$ (Litjens et al., 2017). DL models (networks) are composed of many layers that transform input data (e.g., images) to outputs (e.g., categories) while learning progressively higher-level features (Litjens et al., 2017). Layers between the input and output are often referred to as "hidden" layers. A neural network containing multiple hidden layers is typically considered as a "deep" neural network—hence, the term "deep learning" (Litjens et al., 2017).

Backpropagation-based training of deep neural networks with many layers became an explicit research subject in the early 1990s (Hochreiter, 1991), but it was largely ignored by the machine-learning community at that time. Moreover, it did not receive much attention from the computer vision and remote sensing communities for quite a long period of time, mainly because deep neural networks (DNNs) were considered hard to train efficiently. In the new millennium, however, DNNs have finally attracted wide attention, mainly by outperforming alternative machine-learning algorithms in numerous application areas (Ning et al., 2005; Hinton et al., 2006; Ciresan et al., 2012). In fact, interest in deep feedforward networks was revived around 2006, and DL became practically feasible to some extent with the help of unsupervised learning (Bengio et al., 2007; Hinton and Salakhutdinov, 2006). As one significant early achievement, the convolutional neural network (CNN)-based AlexNet architecture won the popular ImageNet contest by a wide margin in 2012 (Krizhevsky et al., 2012). This was primarily because of its efficient use of graphics processing units (GPU), rectified linear units (ReLUs), and many training examples. After this contest, DL subsequently received much greater attention in different sub-fields of computer vision (Hinton et al., 2012; Litjens et al., 2017), and it was found to perform well for many applications in the field of remote sensing (Chen et al., 2014; Zhang et al., 2016; Zhu et al., 2017).

The remainder of this section is devoted to introducing several commonly used DL models in remote sensing, including the supervised CNN and recurrent neural network (RNN) models, and unsupervised autoencoders (AE) and deep belief networks (DBN) models, also including recently popular generative adversarial networks (GAN) model.

2.1. Convolutional neural networks

CNN, one of the most extensively used DL models, was originally designed to process data in the form of multiple arrays (LeCun et al., 2015). Because of this characteristic, it is well-suited for processing multiband remote-sensing image data in which pixels are arranged regularly. To be specific, the CNN consists mainly of three different types of hierarchical structures: convolution layers, pooling layers, and fully connected layers. At each layer, the input image is convolved with a set of K kernels $\mathrm{W}= \{W_1, W_2, \cdots, W_K\}$ and added biases $\gamma = \{b_1, \cdots, b_K\}$, each generating a new feature map X_k. These features are subjected to elementwise nonlinear transform $\sigma(\cdot)$, and the same process is repeated for every convolutional layer l:$\mathrm{X}_k^l = \sigma(W_k^{l-1} * X^{l-1} + b_k^{l-1})$. Compared with traditional MLPs (Multi-Layer Perceptrons), in CNNs the values of pixels within a neighborhood of a certain size are aggregated using a permutation invariant function, typically the max or mean operation. At the end of the convolutional stream of the network, fully connected layers (i.e., regular neural-network layers) are usually added, where weights are no longer shared (Litjens et al., 2017). Later, some popular CNN architectures, e.g., ALEXNET (Krizhevsky et al., 2012), VGG NETWORKS (Simonyan and Zisserman, 2014), RESNET (He et al., 2016) such as the fully convolutional network (Long et al., 2015) and also a recent development for GoogleNet called Inception-v4 (Szegedy et al., 2017) are discussed.

2.2. Recurrent neural networks

As an alternative widely used supervised learning model, the RNN model was traditionally used for a discrete sequence analysis. In an RNN, the input and output data can be of varying length. Therefore, certain tasks that involve sequential inputs, such as speech and language processing, often benefit more from the RNN. In fact, the most exciting application of backpropagation (if available) is to train RNNs. With the unfolding in time of the computation involved in the forward computation, RNNs will generate very deep feedforward networks to learn long-term dependencies like regular DNNs; thus, it is difficult to learn and store information for very long (Bengio et al, 1994). To address this problem, an explicit memory was used to augment the networks. Therefore, several specialized memory units have been developed—for example, the long short-term Memory cell (Hochreiter and Schmidhuber, 1997) and gated recurrent unit (Cho et al., 2014). With the development of their architecture and ways of training (Sutskever, 2012), RNNs have been successfully and extensively applied in predicting the next character in the text, or the next word in a sequence (Sutskever et al., 2011; Mikolov et al., 2013) and have been extended to other more complex tasks of remote-sensing images. (e.g., Ho Tong Minh et al., 2018; Mou et al., 2017; Ienco et al., 2017; Lyu et al., 2016).

2.3. Autoencoders (AEs) and stacked autoencoders

Normally, autoencoders (AEs) are designed to learn a compressed and distributed dataset representation. The number of hidden units in one hidden layer is smaller compared with the input or the output, and this is the most important feature of AE. Therefore, an AE can accomplish the purpose of data compression and dimensionality reduction through one hidden layer. Hence, AEs are mostly used for the processing of feature hierarchy. AEs (Zhu et al., 2017) are simple networks that map an input x to a latent representation through one hidden layer h, and $h = f(\mathrm{W}x + \beta)$. Here W is a weight matrix to be estimated during training, β is a bias vector, and f indicates a nonlinear function. Subsequently, the reconstructed input γ could be expressed as$\gamma = f(\mathrm{W}'h + \beta')$, which was reconstructed by a reverse mapping and using the same weight to decode the latent representation, $\mathrm{W}' = \mathrm{W}^T$, $\beta' = \beta^T$.

As the name implies, a stacked AE (SAE) (or deep AEs) 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. It is formed by stacking AE layers. In the field of remote sensing, such multilayer AEs are usually used for feature representation and have produced good effects (Zabalza et al., 2016; Abdi et al., 2017; Gong et al., 2017; Chen et al., 2017; Hao et al., 2018), particularly in spectral–spatial feature learning (Abdi et al., 2017; Wang et al., 2018a).

2.4. Restricted Boltzmann machines and deep belief networks

A restricted Boltzmann machine (RBM) (Hinton, 2012) is a generative stochastic undirected neural network consisting of a visible layer x and a hidden layer h. The layers are connected, whereas the units in the layers are not connected. As two-layer networks, RBMs present a particular type of Markov random field. For a particular state (x, h) of visible and hidden units, an energy function is defined as the joint configuration of the units. Like SAEs, DBNs each consist of multiple layers of RBMs, except that the individual layers in DBNs are trained using the RBM model, instead of AEs in an unsupervised manner. Final fine-tuning is performed by adding a linear classifier to the top layer of the DBN and implementing a supervised optimization procedure. In fact, the DNNs are first trained in an unsupervised manner (pretraining) and then the stacked network is fine-tuned. This usually yields good results, for example, SAEs and DBNs.

2.5. Generative adversarial networks

Generative adversarial networks (GANs) (Goodfellow et al., 2014) have recently become a very popular category of unsupervised DL models. The GAN contains a system of two networks contesting with each other: a generative network and discriminative network. The generative network learns to map from a latent space to a particular data distribution of interest (e.g., images), while the discriminative network discriminates between the real data and the generated data produced by the generative network. The target of training the generative network is to "fool" the discriminative network by producing examples that appear realistic that have the true data distribution. The discriminative network is generally a standard convolutional network to produce probabilities. Both networks try to optimize a different and opposing loss function in a zero-zum game (Oliehoek et al., 2017). In the last three years, GANs have been successfully applied in many computer vision and image processing applications (Isola et al., 2017; Lin et al., 2017; Zhan et al., 2018).

3. Methods and data to review DL in remote sensing application

3.1. Study selection and data extraction by meta-analysis method

To identify articles concerning remote-sensing image analyses using DL, a title/abstract/keyword search was performed in the Scopus database using the search query: "deep learning" AND "remote sensing" (search date: June 2, 2018). After eliminating the review articles, 402 relevant documents were obtained, including 221 peer-reviewed articles and 181 conference papers. Subsequently, various types of information were extracted from each of the journal articles, including the "study area", "study target", "DL model used", and "accuracy" (Table 1). This database served as the basis for further statistical analysis. The intent was to reveal quantitatively the research status of DL for different remote sensing applications. When investigating the 221 articles in detail, 45 were determined to be unrelated to this meta-analysis and thus discarded. The final database contained the records of the remaining 176 relevant articles, and the parameters in Table 1 were extracted by reviewing each paper in detail. It is noted that here conference papers are not involved in this meta-analysis due to their low quality with very simple results.

Table 1
Twelve attributes used for DL and remote sensing.

#	Attribute	Categories
1	Remote sensing data	Type: Hyperspectral, SAR, Lidar, other
2	Study area	Type: Urban, agriculture, other
3	Study target	Type: Image fusion, scene classification or object detection, LULC classification, seg., other
4	DL model	Type: CNN, RNN, AE, DBN, other
5	Processing unit	Type: Pixel, object
6	Accuracy	Value
7	Training samples	Value
8	Area of study site	Value
9	Resolution	Value (high resolution, moderate, coarse)
10	Paper type	Type: Journal, conference
11	Year	Value
12	Journal	Value

3.2. Conference papers

Conference papers were excluded from our meta-analysis due to their (typically) lower level of academic rigor than peer-reviewed articles, and because many were likely expanded into journal papers after presenting at the conferences. However, from the 181 conference papers retrieved by the Scopus search, we found that the topic is now well represented at the major international remote sensing conferences (Table 2). Three major remote sensing academic societies, in particular, have been major outlets for contributions related to DL in the field of remote sensing; namely the IEEE Geoscience and Remote Sensing Society (IGARSS), the International Society for Photogrammetry and Remote Sensing (ISPRS), and the Society of Photo-Optical Instrumentation Engineers (SPIE). Another notable finding was that the first DL-related special issue appeared in the SPIE *Journal of Applied Remote Sensing* (Ball et al., 2017), followed by the "Deep Learning for Remotely Sensed Data" special issue of the *ISPRS Journal of Photogrammetry and Remote Sensing* in 2018.

3.3. Peer-reviewed articles

For the unfiltered 221 peer-reviewed journal papers, a majority of the articles were published in the 17 journals listed in Table 3 (the journal of only one publication is not showed here). In total, these 17 journals accounted for 171 articles, or 77% of all of the relevant peer-reviewed journal papers related to DL and remote sensing. In terms of the number of published articles, the top five journals were, in order, *IEEE Geoscience and Remote Sensing Letters*, *ISPRS Journal of Photogrammetry and Remote Sensing*, *Remote Sensing*, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, and *IEEE Transactions on Geoscience and Remote Sensing.*

Table 2
Conferences/workshops identified as relevant, and number of relevant papers.

Title of conference/workshop	#
International Geoscience and Remote Sensing Symposium (IGARSS)	33
International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences – ISPRS Archives	28
Proceedings of SPIE – The International Society for Optical Engineering	20
RSIP 2017 – International Workshop on Remote Sensing with Intelligent Processing, Proceedings	7
2017 Joint Urban Remote Sensing Event, JURSE 2017	6
Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing	6
Proceedings – International Conference on Image Processing, ICIP	4
38th Asian Conference on Remote Sensing – Space Applications: Touching Human Lives, ACRS 2017	3

Table 3
Journals identified as relevant, and number of relevant papers.

Name of journal	#
IEEE Geoscience and Remote Sensing Letters	33
ISPRS Journal of Photogrammetry and Remote Sensing	29
Remote Sensing	25
IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing	20
IEEE Transactions on Geoscience and Remote Sensing	18
Journal of Applied Remote Sensing	14
International Journal of Remote Sensing	8
ISPRS International Journal of Geo-Information	4
Journal of Hydrometeorology	3
Journal of Sensors	3
Geophysical Research Letters	2
Neurocomputing	2
Pattern Recognition	2
Photogrammetric Engineering and Remote Sensing	2
Proceedings of the IEEE	2
Remote Sensing of Environment	2
Soft Computing	2

4. Results and brief overview

A simple statistical analysis was conducted using the above data to analyze the current status and trends in the use of DL within the remote sensing field. This included calculating: (1) the number of conference papers and peer-reviewed journal articles published annually from 2008 to 2018, as shown in Fig. 1; based on the latest Scopus data from March 9, 2019); (2) the number of journal articles focusing on different sub-areas of remote sensing (i.e. different study targets), as shown in Fig. 2; (3) the usage frequency of different DL models, as shown in Fig. 3; (4) the spatial resolution of the remote-sensing images DL models were used to analyze, as shown in Fig. 4; and (5) the types of regions that were studied most frequently, as shown in Fig. 5. Finally, a statistical analysis was conducted to assess the level of accuracy that DL models achieved for three types of classification tasks: LULC classification, object detection, scene recognition, as shown in Fig. 6.

Due to the typically greater timeliness of conference papers (e.g. shorter paper length/faster peer-review process than journal papers), applications of DL in remote sensing were mainly presented in this format at the initial stage. In particular, the International Geoscience and Remote Sensing Symposium (IGARSS) and Proceedings of SPIE - The International Society for Optical Engineering, respectively, included 33 and 20 academic articles concerning the application of DL in remote sensing (as described in Table 2). As of 2018, however, the number of journal papers on this topic now outnumbers the number of conference papers, demonstrating the growing maturity of the research.

As shown in Fig. 2, most studies focused on LULC classification, object detection, and scene classification. For conducting these

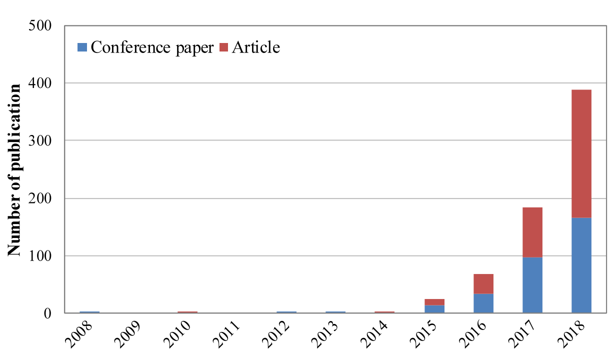

Fig. 1. Number of conference papers and articles in the Scopus database for a general search on ["deep learning" AND "remote sensing"].

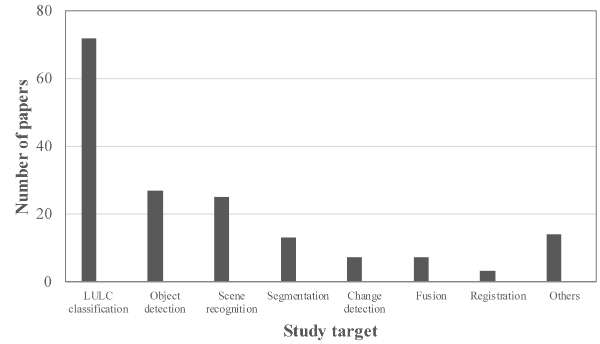

Fig. 2. Number of publications for different study targets.

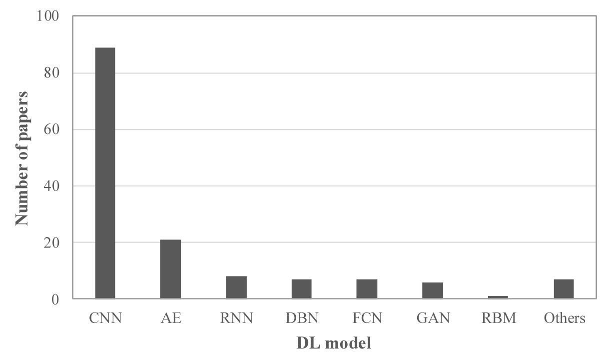

Fig. 3. Distribution of DL model used in the studies.

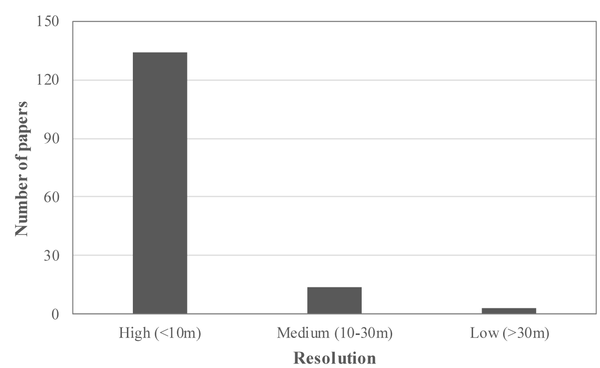

Fig. 4. Distribution of image spatial resolution used in the investigated case studies.

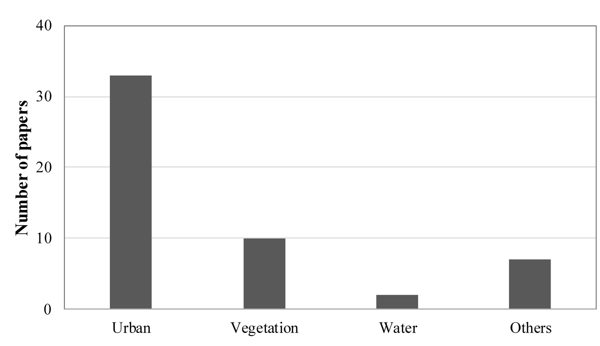

Fig. 5. Distribution of application area.

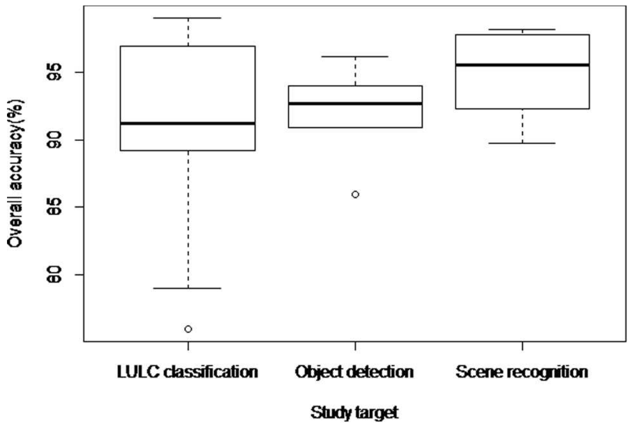

Fig. 6. Distribution of overall accuracies for three classification sub study targets (LULC classification, object detection, scene classification).

analyses, the studies mainly adopted publically-available benchmark image datasets. For example, LULC classification studies mostly focused on hyperspectral data or high-spatial-resolution images, i.e., the Reflective Optics System Imaging Spectrometer (ROSIS-03) "University of Pavia" dataset, the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) "Indian Pines" dataset (Chen et al., 2014, 2015; Mou et al., 2017), and the "Vaihingen" benchmark dataset (http://www2.isprs.org/commissions/comm3/wg4/semantic-labeling.html) (Marcos et al., 2018; Audebert et al., 2018). Therefore, there are a limited number of studies that have completely focused on the actual practical applications of DL for these remote sensing tasks. While the use of DL models for segmentation, fusion, and registration were less frequently investigated, many journal articles still highlighted that the models were relatively successful for these tasks (Kemker et al., 2018; Xing et al., 2018a; Wang et al., 2018b). This demonstrates that DL has significant and diverse prospects in the field of remote sensing, which distinguishes it from the other conventional classification algorithms in the analysis of remote-sensing images. Here, "others" in Fig. 2 consisted of tasks that only a few studies focused on, e.g. analysis of network media data (Li et al., 2017a), time series analysis (Das and Ghosh, 2016; Cai et al., 2018), and retrieval of precipitation data (Tao et al., 2016).

As shown in Fig. 3, the CNN model has been the most commonly used for remote-sensing image analysis, followed by the AE model (including the SAE model). The RNN, DBN, and GAN models, and particularly the RBM model, were much less commonly used. As earlier mentioned in Section 2.1, the higher popularity of CNN is likely because it has unique characteristics that make it highly suitable for processing multiband remote-sensing image data in which pixels are arranged regularly.

As shown in Fig. 4, DL models are most frequently used to analyze remote-sensing images with spatial resolutions of 10 m or finer. In fact, the image data cited in the more than 100 studies had a spatial resolution of finer than 2 m. This suggests that remote-sensing data with a high spatial resolution benefits more from DL, probably because such data contains rich spatial feature information. DL models have been shown to be quite successful in extracting high-level feature information from such data, which is very useful for various image analysis tasks.

Fig. 5 shows the number of published articles concerning DL in different types of study areas (e.g., "urban areas", "vegetated areas", and "water areas"). In this study, both cropland and forests are categorized under vegetated areas. Other types of study areas (including roads, snow, soil, and wetland) were very few in number, and are not listed here individually. As shown in Fig. 5, most of the studies pertained to urban areas. Here, the statistics do not include the studies that used the benchmark datasets, because the benchmark datasets can be classified into many types and have no specific application types.

Fig. 6 shows a box plot graph of classification accuracy regarding the types of study target, including the accuracy of scene classification, object detection, and LULC classification. Scene classification had the highest median classification accuracy (~95%), object detection has the second-highest median classification accuracy (~92%), and LULC classification has the lowest median classification accuracy (~91%). This can be ascribed to the studies' overwhelming use of benchmark datasets for object detection and scene classification, as these benchmark datasets were specifically developed for DL. The accuracy of DL algorithms (and all classification algorithms) for LULC classification is highly dependent on the number of LULC classes considered as well as the spectral/spatial distinction between classes. Most studies involving LULC classification attempted to map several LULC classes, so the median classification accuracy was slightly lower and the variability of classification accuracy was much higher compared with other two applications. Evidently, because of the uniqueness of its application in remote sensing, it is necessary to further develop and explore the application of DL in remote sensing, especially for LULC classification. Even so, this median accuracy (~91%) of the LULC classifications using DL is already significantly higher than the accuracy of other classifiers in supervised object-based image classification (including RF, SVM, decision trees, etc.). Because in our previous survey there is no classifier with a median accuracy of more than 90% for supervised object-based image classification (Ma et al., 2017).

5. Application of DL in remote sensing

This review briefly shows the statistical analysis of DL-based studies in the field of remote sensing. Fig. 7 provides a visual representation of the highest-frequency terms appearing in the title and abstract of the peer-reviewed literature, where higher-frequency results are in a larger font size. From this figure it appears that CNN is more popular than other DL models (matching the results of Fig. 3), and spectral-spatial features of images are utilized for analysis. The main focus of the studies is on classification tasks (i.e., LULC classification, object detection, and scene classification), but several other applications are worth mentioning including fusion, segmentation, change detection, and registration. Therefore, it appeared that DL could be applied to almost every step of the remote-sensing image processing. Subsequently, a taxonomy related to DL in remote sensing (Fig. 8) from preprocessing to accuracy assessment was drawn, and, as a result, the review given

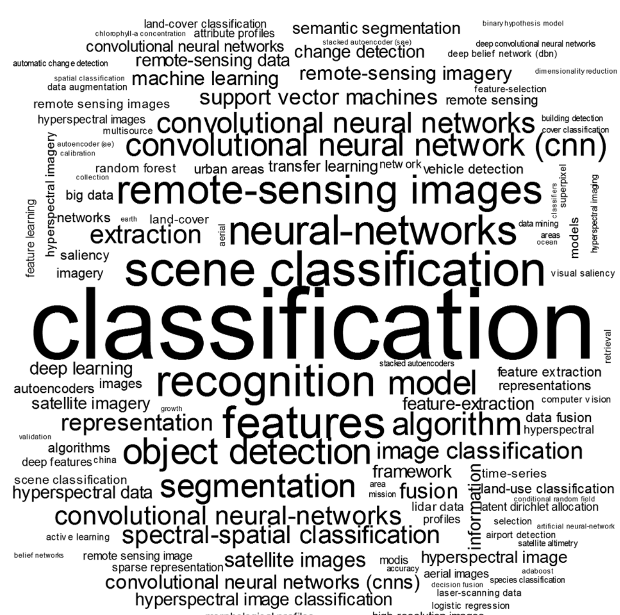

Fig. 7. Tag cloud of this review.

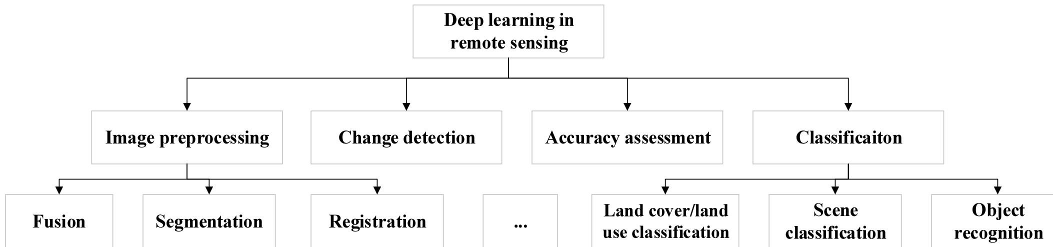

Fig. 8. The taxonomy containing four tasks: image preprocessing, classification, change detection, and accuracy assessment.

below mainly focused on these topics (shown in Fig. 8), to show the contribution and future works of DL for remote sensing application.

5.1. Image fusion

As a fundamental task in the field of remote sensing imagery, remote sensing image fusion techniques aim to obtain an image that simultaneously has high spectral and spatial resolutions. A typical example of remote sensing image fusion is known as pan-sharpening, which indicates the fusion of a low-resolution multi-spectral (MS) image and a high-resolution panchromatic (PAN) image to achieve a high-resolution MS image. Another instance is the fusion of a low-resolution hyper-spectral (HS) image and a high-resolution MS image to generate a high-resolution HS image. Remote sensing image fusion can be regarded as an image super-resolution problem for the low-resolution source image, aided by the high-resolution source image. Inspired by the progress achieved in DL-based image super-resolution, many DL based remote sensing image fusion methods have been proposed in the last few years, indicating the distinct advantages over traditional methods owing to the better capability of DL models in characterizing the complex relationship between the input and the target images (Liu et al., 2018).

(1) *MS and PAN image fusion (pan-sharpening)*: The pioneering work on DL-based pan-sharpening introduces a SAE model to generate a direct mapping between the low- and high-resolution MS image patches (Huang et al., 2015). The PAN images are employed to train the model based on the assumption that the MS and PAN images share the same relationship between low- and high-resolution versions. An improved work on SAE-based pan-sharpening adopts a deep metric learning approach to learn the refined geometric multi-manifold embedding via multiple stacked sparse autoencoders (Xing et al., 2018a). In addition to SAE, the CNN has recently emerged as a very active DL model in remote sensing image fusion. Masi et al. (2016) proposed a CNN-based pan-sharpening method by applying a three-layer convolutional network, which is originally used for natural image super-resolution. The source MS image and PAN image are stacked together as the input, and the target high-resolution MS image is adopted as the output, leading to an end-to-end mapping. Since then, several improved works on CNN-based pan-sharpening have been proposed. Wei et al. (2017) proposed a much deeper network that contains 11 convolutional layers for pan-sharpening through residual learning. Yuan et al. (2018) introduced a multiscale and multidepth CNN architecture for fusion of MS and PAN images. Shao and Cai (2018) proposed a two-branch convolutional network (one for MS image and the other for PAN image) for pan-sharpening to adequately extract spectral and spatial features. Scarpa et al. (2018) integrated a target-adaptive tuning phase into the convolutional network for pan-sharpening, aiming to overcome the difficulty of insufficient training data for a specific fusion task.

(2) *HS and MS image fusion*: Palsson et al. (2017) firstly proposed a DL-based HS and PAN image fusion method through a 3D-CNN. To ensure the computational efficiency, PCA is applied to reduce the dimension of the HS image before it is fed to the network, and the final fused HS image is reconstructed by performing the inverse PCA transformation on the network output. Yang et al. (2018) presented the HS and MS image fusion method with a two-branch convolutional network, which is applied to characterize the relationship between the spectrums of low-resolution and high-resolution HS images. The input of one branch is a pixel in the up-scaled low-resolution HS image while the other is its corresponding spatial neighborhood in the high-resolution MS image. Dian et al. (2018) proposed the HS and MS image fusion method by combining a deep convolutional network and a model-based approach. The model-based approach plays both roles of pre-processing and post-processing, for the CNN-based module.

5.2. Image registration

Image registration is a method of aligning two or more images captured by different sensors, at different times or from different viewpoints (Zitova & Flusser, 2003; Ye et al., 2017). It is a fundamental preliminary step for many remote sensing analysis tasks, for example, image fusion, change detection, image mosaic, etc. In general, image registration includes the following four steps (Zitova and Flusser, 2003): (1) feature extraction, (2) feature matching, (3) transformation model estimation and, (4) image resampling. Feature extraction plays a crucial role in image registration because it decides what type of feature is to be used for image matching. Since deep learning, as a fully data-driven scheme, can automatically learn the features from images, it has been applied to remote sensing image registration recently. Most of the image registration methods using deep learning are based on a Siamese network (Bromley et al., 1994; Merkle et al., 2017; He et al., 2018; Hughes et al., 2018a; Wang et al., 2018b). The basic idea of these methods is to train a deep neural network consisting of two parts. The first part is used to extract features from image patches by training a siamese or pseudo-siamese network. Subsequently, the second part is used to measure the similarity between these features for image matching. In addition, the GANs are applied to image matching and registration (Merkle et al., 2018; Hughes et al., 2018b). These methods first translate an image into the other one by training the GANs, enabling the two images to have similar intensity or feature information. Subsequently, feature extraction and the matching are carried out between the two artificially generated images, which effectively improves the performance of image matching. Besides image-to-image registration, DL is also applied for image-to-map registration. Zampieri et al. (2018), designed a chain of scale-specific neural networks, to register cadastral maps of buildings as well as road polylines onto aerial images. Subsequently, Girard et al., improved the scale-specific neural networks by introducing the multi-task learning, which improved the registration performance (Girard et al., 2018). In general, remote sensing image registration by DL has become an active research subject in recent years. However, at present there are no public training datasets for remote sensing image registration, and training samples have to be made manually before applying the DL technique. Owing to the diversity of remote sensing data, i.e., images captured at different resolution, at different times (or by different modalities), it will be an important challenge and laborious task to establish large public training datasets for image registration.

5.3. Scene classification and object detection

Proceding to scene classification and object detection, firstly it is essential to know their differences, as both have similar remote-sensing applications and are frequently confused. In this study, scene classification is defined as a procedure to determine the image categories from numerous pictures—for example, agricultural scenes, forest scenes, and beach scenes (Zou et al., 2015)—and the training samples are a series of labeled pictures. However, object detection aims to detect different objects in a single image scene — for example, airplanes (Zhong et al., 2018), cars (Ding et al., 2018), and urban villages (Li et al., 2017b) — and the training samples are the pixels in a fixed-sized window or patch.

For scene classification, Cheng et al. (2018) imposed a metric learning regularization term on CNN-derived features to optimize a new discriminative objective function for training the CNN model, and they successfully applied it to scene classification. To the best of our knowledge, most of the scholars are focusing on developing the DL algorithm for remote sensing, and prefer to use stable benchmark datasets (Xia et al., 2017), including the RSSCN7 dataset (Zou et al., 2015; Yang et al., 2017), UC-Merced dataset (Zhao and Du, 2016; Scott et al., 2017), and WHU-RS dataset (Xia et al., 2010; Zhong et al., 2016; Han et al., 2017), and there are few studies related to the use of real data. Therefore, it appears necessary to focus greater attention on how to apply the scene classification techniques to implement remote-sensing applications practically. For example, typically the application of a scene classification technique to the accuracy assessment of remote-sensing image classification is advantageous, as pointed out by Xing et al. (2018b) (the details are shown in Section 5.7). In addition, compared with the benchmark datasets used in scene classification, the volumes of available remote-sensing datasets as training set are particularly limited in practical application. Therefore, the performance of DL methods in remote-sensing scene classification is restricted. Subsequently, many studies have begun to exploit how to extend training samples (e.g., transfer learning and data augmentation) to enhance the efficiency of DL in remote sensing. For example, Marmani et al. (2016) exploited a pretrained CNN to extract an initial set of representations. Afterwards, Han et al. (2017) proposed a semi-supervised generative framework to generate a new training set, by combining pretrained CNN and SVM classifiers. Scott et al. (2017) employed the transfer learning to preserve the deep visual feature extraction learned over an image corpus, from a different image domain. Yu et al. (2017) applied three operations (flip, translation, and rotation) to generate augmented data and obtain a more descriptive deep model by using these augmented data as training data.

For object detection from remote-sensing images, in addition to the limitation of training samples, the biggest challenge is to effectively deal with the problem of object rotation variations (Cheng et al., 2016; Yu et al., 2016). Therefore, Cheng et al. (2016) proposed learning a new rotation-invariant layer on the basis of the existing CNN architectures to advance the performance of multiclass object detection, which included the detection of ships and airplanes. However, in this study, a benchmark dataset was used for object detection, i.e., the NWPU VHR-10 dataset (Cheng and Han, 2016). Nevertheless, studies have been carried out to detect target objects for many practical applications in remote sensing — for example, Vetrivel et al. (2018) integrated CNN features and 3D point cloud features from oblique aerial images to detect buildings damaged by an earthquake. Li et al. (2017b) proposed a novel unsupervised DL method to detect urban villages from high-resolution images by learning a data-driven feature. Ding et al. (2018) enhanced the structure of the base VGG16-Net algorithm for improving the precision of airplane and car detection, whereas Chen et al. (2018a) used transfer learning to implement end-to-end airplane detection by improving the efficiency of training samples at the fine-tuning phase. In addition, Wu et al. (2016) used a DBN model to extract aircraft by integrating object shapes. Because of the ultra-high-resolution characteristics of unmanned aerial vehicles (UAVs), Kellenberger et al. (2018) even tried using CNNs to detect mammals in UAV images, and Li et al. (2017c) used CNNs to detect oil palm trees in QuickBird images.

As mentioned above, we found that current studies preferred to extract certain specific type(s) of objects (airplanes, cars, etc.) from high-resolution images through a fixed window size, either through scene classification or object detection. However, more data and object types of objects are encountered in practical remote-sensing applications—for example, medium-resolution Landsat data and Sentinel data. Therefore, how to design the effective algorithms to overcome the difficulties emerging from different-scale objects (the different type of objects often appears at different scales in remote-sensing images, and also the same object can have variable size in different-scale remote-sensing images) is an urgent problem in both subfields (Deng et al., 2018).

5.4. Land-use and land-cover (LULC) classification

As stated earlier, many DL studies have focused on scene classification, probably because numerous benchmark datasets can be used for this application in DL. Our review of the LULC classification studies in this section does not include these scene classification studies because they do not pertain to the actual LULC mapping in the remote-sensing images. For example, several studies (Luus et al., 2015; Wang et al., 2017a; Weng et al., 2017) used the UC Merced dataset, but all of them were named land-use classification. In general, it was found that DL algorithms were typically utilized for classifying high-resolution images, owing to the fine structural information (i.e. spatial details) of LULC objects in these types of images. Although there are also many freely available medium-resolution (10 m-30 m) satellite images for LULC mapping, including Landsat and Sentinel-2 data, it is difficult to apply conventional DL algorithms directly to these images because of the lack of such fine structures (Sharma et al., 2017). Therefore, Sharma et al. (2017) proposed a patch-based CNN which was more suitable than the pixel-based CNN for medium images. In addition, while most studies focused on classification of optical images, Lv et al. (2015) used a DBN model to classify synthetic-aperture radar images.

Early studies concerning LULC classification based on DL mostly focused on feature representation or learning, while the final classification used other simpler classifiers (Tuia et al., 2015). This was mainly because deep features show powerful ability in image representation and understanding (Li et al., 2018b), by providing high-level spatial information which is created by hierarchical structures (Liang and Li, 2016) rather than generating low-level features (Zhao and Du, 2016). for example, Romero et al. (2016) proposed unsupervised deep feature extraction for classification by using greedy layerwise unsupervised pretraining. Additionally, Zhao and Du (2016) presented a multiscale CNN algorithm to learn spatial-related deep features to classify hyperspectral remote imagery.

Regarding the DL model used in LULC classification, CNN has been received the greatest attention at all times (Maggiori et al., 2017). Initially, however, DBNs (Chen et al., 2015; Tuia et al., 2015; Liu P et al., 2017) and AEs (Chen, et al., 2014) were also used quite frequently. There were also number of studies in which specific improvements to the DL model(s) were proposed for the purpose of more accurate LULC mapping. For example, Marcos et al. (2018) proposed a CNN architecture with rotation equivariance and applied it to two sub-decimeter land cover semantic-labeling benchmarks (Vaihingen dataset). With the increased presence of CNN applications in remote sensing, Zhu et al. (2018) for the first time introduced the GAN as a regularization technique to the CNN model in the classification of hyperspectral remote-sensing images. The result of this prevented the overfitting phenomenon in the training process. Moreover, they proposed two strategies to process spectral features and spectral-plus-spatial features—i.e., 1D-GAN for spectral vectors and 3D-GAN for spectral-and-spatial features. Experimental results revealed that this method could significantly improve LULC classification accuracy compared with the traditional CNN. Following Zhu et al. (2018), Hamida et al. (2018) for the second time proposed using the 3D-DL strategy to process spectral-spatial features. It combines the traditional CNN network with the twist of applying 3D convolution operations to enable joint spectral-and-spatial information processing, while traditional 1D convolution operators only inspect the spectral content of the data. Experimental results revealed that, in the CNN model, the 3D-DL strategy appears to be a generally accepted method to process spectral-spatial features in better manner.

A supervised DL model usually requires a large number of training samples. In the field of remote-sensing, labeling the observed data to prepare training samples for each LULC class is highly time- and/or cost-intensive. Therefore, some augmentation techniques were developed for increasing the size and/or efficiency of the training dataset, e.g., transfer learning (Lyu et al., 2018) and active learning (Liu et al., 2017a) techniques. Along these lines, Shin et al. (2016) used transfer learning to overcome the restrictions on the number of training samples, while Fu et al. (2018) used active learning for multispectral (object-based) image classification, and Liu et al. (2017a) implemented it for hyperspectral image classification. In addition, a large number of studies are beginning to focus on implementing a high-precision DL model by using a small number of samples (Pan et al., 2017; Yu et al., 2018). For example, Yu et al. (2018) proposed an unsupervised convolutional feature fusion network to formulate an easy-to-train but effective CNN representation of remote-sensing images. Lin et al. (2017) proposed an unsupervised model based on GANs, using only unlabeled data for learning a representation.

Regarding the data source used for DL in LULC classification, most studies used only a single image scene, although some used multisource remote-sensing data (e.g., hyperspectral and high-resolution image data, lidar data, and topographic data) (Xu et al., 2018a). Time-series remote sensing imagery, however, was rarely analyzed using DL algorithms. Lenco et al. (2017) used multi-temporal remote sensing data (Pléiades imagers) and RNNs to perform LULC classification, but only three dates of imagery were used for this analysis (July and September in 2012, and March in 2013). This approach differs from the general time-series analysis approaches in the field of remote sensing. Hence, it is necessary to further conduct time series analysis concerning the application of DL in open-source remote-sensing images with large quantities of inventory data (e.g., Landsat or Sentinel data). For example, for the purpose of classification through remote-sensing time series, is it possible to import long time series information of pixels or objects to the DL model (e.g., RNN), similarly to importing the hyperspectral feature series of pixels or objects? Mou et al. (2017) had already used RNNs for analyzing hyperspectral pixels as sequential data and then determined pixel categories, because RNNs are mainly designed to handle sequential data.

DL has been widely applied in diverse study areas, including urban areas, vegetated areas, forest areas, and wetlands (Shao et al., 2017; Isikdogan et al., 2017; Audebert et al., 2018; Ho Tong Minh et al., 2018; Liu and Abd-Elrahman, 2018). As one of the most important fields of research in remote sensing, urban remote-sensing mapping has received much attention. Although numerous methods have been proposed for urban LULC classification, the accuracy and efficiency of these methods cannot always meet the needs of urban management and analysis in practice (Huang et al., 2018), and the area scope processed by these techniques is often small. In particular, studies concerning high-resolution remote-sensing images are mostly characterized by methodological innovation. From the perspective of practical application analysis, Lyu et al. (2018) recently used RNN DL model to minimize seasonal urban spectral variance, and consequently ensured a high classification accuracy of approximately 96% for urban LULC maps of one year. Moreover, their approach integrated a transfer learning strategy, thus attaining the possibility to reuse the training samples in different periods for conducting urban LULC mapping and change analysis. Huang et al. (2018) used units of a regular size to represent the irregular mapping units derived from street blocks, and then generated a practical land use map by deep CNNs. Yao et al. (2017) used a similar method to detect urban land use distributions by integrating Baidu points of interest and a Word2Vec model. However, their units of processing were generated based on road networks and cannot represent the actual LULC objects accurately. In addition, there are studies concerning the recognition of trees along urban streets through DL (Branson et al., 2018). Making thorough use of such Internet data as Google Maps, DL allows full play to its advantage in target detection.

Regarding the application of DL for LULC classification, a few new methods specific to different data features have been developed in addition to conventional approaches. In the field of classification of agricultural remote-sensing images, Cai et al. (2018) built a crop classification model by applying DNNs to time-series images. In this model, the ReLU is used as the activation function. In the study, time series images were skillfully divided by time steps, and then spectral values at different time points were used as input information of the DNN model, thereby, allowing it to take full advantage of both the time series information and the features of the DNN model. This strategy opens a new path to the currently prevailing classification of land cover in time-series through the DNN model. A larger number of improved DNN models are expected to be extensively applied in processing of the time-series remote-sensing images. Furthermore, Liu and Abd-Elrahman (2018) proposed a Multiview OBIA-framework-based DCNN. It trains the DCNN model for classification based on the spectral differences between the images acquired by UAVs from different angles of view. This method is similar to the idea concerning training the DNN model through time series information (Cai et al., 2018). The two instances above provide a novel idea concerning LULC classification through DL.

5.5. Semantic segmentation

Semantic segmentation aims to assign land cover labels to each pixel in an image. Facilitated by deep CNNs, especially by the end-to-end fully convolutional network (FCN) (Long et al., 2015), interest in semantic segmentation of remote sensing images has increased in recent years. It is noted that semantic segmentation discussed in this part refers to the dense prediction of pixel labels rather than pixel-by-pixel classification.

State-of-the-art semantic segmentation frameworks for remote sensing images are sequentially composed of encoder and decoder subnetworks. The main advantage of DCNN for semantic segmentation is its capability to explore the multi-level context information over very large receptive fields. However, it comes at a cost of low spatial resolution for the segmentation result, in which blurry class boundaries and the loss of object details become a challenge. To deal with this problem, four main strategies have been adopted in the remote sensing domain: (1) Developing no-downsampling encoder network by atrous convolution (Sherrah, 2016) or combing features from multiple resolutions (Zhang et al., 2017a; Chen et al., 2018b; Marmani et al., 2016); (2) Improving the decoder network by designing symmetric unconvoluted layers and skip connections (Liu et al., 2017b; Chen et al., 2018c; Kemker et al., 2018); (3) Using an ensemble of multiple networks with different initializations or different structures (Zhang et al., 2017b; Marmani et al., 2016); and (4) Post-processing the semantic segmentation results by using a probabilistic graph model (Sherrah, 2016; Liu et al., 2017b; Zhao et al., 2017a), by fusing segments produced by unsupervised segmentation (Zhao et al., 2017a), by use of an overlay strategy (Liu et al., 2017b; Chen et al., 2018c), or by using a filtering method (Xu et al., 2018b). Although active efforts have been imposed on these aspects, it remains a challenging issue to balance the trade-off between strong downsampling (which allows for richer context information extraction) and accurate boundary localization (which requires local details) for semantic segmentation of remote sensing images.

In addition to the above challenge, other problems specific to semantic segmentation of remote sensing images have been addressed in remote sensing domain, including: (1) Overcoming the difficulty of the high variety of geographic objects in very high spatial resolution images (Wang et al., 2017b); (2) Addressing the problem of class balancing caused by small objects by designing medium frequency balancing strategy (Liu et al., 2017b); and (3) Exploring multi-spectral bands or multi-modal data using hybrid networks (Sherrah, 2016; Audebert et al., 2018; Marmani et al., 2016) or training new network instead of fine tuning trained network in computer vision domain (Kemker et al., 2018; Xu et al., 2018b).

According to the semantic segmentation results reported on benchmark datasets, e.g. the ISPRS datasets (Rottensteiner et al., 2012), DCNN-based semantic segmentation can achieve excellent segmentation accuracies, indicating its substantial potential. However, the results on datasets with higher levels of difficulty, e.g., RIT-18 (Kemker et al., 2018) show that the improvement of networks for semantic segmentation need to be further explored in the future. DCNN-based semantic segmentation is not limited to LULC mapping or information extraction from 2D remote sensing images. Indeed, some of the pioneering works explored other interesting applications of semantic segmentation, such as the semantic segmentation of large-scale 3D scenes (Zhang et al., 2018a), and the extraction of building footprints and building heights (Zhuo et al., 2018), which hints at expanding the applications of DCNN-based semantic segmentation in the future.

5.6. Object-based image analysis

When different DL models are used for object classification, the processing methodology varies significantly because of the unit of the polygon. For example, Zhang et al. (2017c) directly imported the spectral, spatial, and texture features of objects to an SAE model, thereby training the parameters of the network. In contrast, Zhao et al. (2017b) used a patch-based CNN method to combine the deep features with object-based classification. Normally, present object-based classification through the CNNs is essentially required, first to the generate patches, and then represent the types of objects through pixel classification (Zhao et al., 2017b; Fu et al., 2018) except the methods of patch generation and segmentation that vary to a certain extent. For example, Guo et al. (2018) created pre-segment objects through graph-based segmentation, fused similar objects through a selective search method, and then extracted bounding boxes of potential ground objects. Subsequently, it was possible to add the bounding boxes to the training dataset to use the DL classification model, that perform the pixelwise classification. To apply CNN to urban mapping, Huang et al. (2018) processed the irregular mapping units derived from street blocks to generate a practical land-use map. However, the processing units were generated based on road networks, making it impossible to accurately represent all of the LULC features present. Recently, Fu et al. (2018) segmented objects through multiresolution segmentation, and generated patches using the barycenters of segmentation objects as centers and fixing the window at 32×32 and 64×64, and subsequently classified the patches through CNNs. The classification results obtained by them were closer to actual land cover. Based on the patch strategy, Zhang et al. (2018b) proposed the object-based CNN method for urban land use classification by using a standard CNN model. The method proposed by Zhang et al. (2018b) was similar to that proposed by Fu et al. (2018). The study by Fu et al. (2018) further analyzed the impact of different patch sizes (from 16 to 144 at a regular step of 16). They argued that a larger patch size is more beneficial to the classification by an object-based CNN. Especially, as another novel strategy that combines the DL with OBIA, Tong et al. (2018) now rely on DNNs to predict the type of land use on a pixel level by patch-wise classification, and then vote to determine the segmented object's type on an object level by the pseudo-labels of pixel.

5.7. Others

In addition to the above conventional application areas in remote-sensing image analysis, DL can also be applied in other areas (e.g., acquisition of validation data for classification accuracy, acquisition of classification training sample data, and data prediction) in a novel manner because of its high accuracy.

Xing et al. (2018b) were the first to apply CCNs to precision evaluation for LULC mapping, using geotagged photos for land cover validation through DL. Specifically, they used a VGG-16 network to first automatically tag the LULC types represented in the photos. Later, they used the photos as validation samples to compare with the results of a remote-sensing image-based LULC classification at the associated positions, and finally to calculate the classification accuracy of the derived LULC map. This method can significantly reduce the complexity and labor required for the conventional validation process, particularly for the validation of classification results of land cover within the given scope of a large area.

Regarding the acquisition of training samples for land cover mapping, DL has the advantage of high scene recognition precision. Because of this advantage, and the availability of big social network media data (e.g., photos with location information), Chi et al. (2017) expressed a view that large quantities of tagged data could be automatically generated from location-based social media photos, thereby increasing the number of training samples available for remote-sensing image classification.

Considering the high-precision performance of DL, some studies built a relationship between multisource spatial data through a DL model and then predicted locational space data based on the known spatial data. For example, Ghamisi and Yokoya (2018) established an image-to-digital-surface-model (DSM) translation rule by a conditional generative adversarial network to simulate the DSM from a single optical image. Li et al. (2017d) estimated ground-level PM2.5 by fusing satellite and station observations, and they considered geographical distance and spatiotemporally correlated PM2.5 in a DBN.

Furthermore, DL is also applied to other specific areas related to remote sensing, including compression artifact reduction (Zhang et al., 2018c), network media data analysis (Li et al., 2017a), time series analysis (Das and Ghosh, 2016; Cai et al., 2018), and retrieval of precipitation data (Tao et al., 2016). In particular, the time series analysis is a promising application area of DL in remote sensing for classification and change detection, while a study by Das and Ghosh (2016) has proved the capability of DL in spatiotemporal prediction of remote-sensing data. Unfortunately, the studies in this regard are few. Since Lyu et al. (2016) first exploited recurrent neural networks in DL for bitemporal change detection using Landsat data, there have been few studies in this regard (Cai et al., 2018).

6. Conclusions

In this study, publications related to DL in almost all sub-areas of the remote sensing field were systematically analyzed through a meta-analysis. Subsequently, several main subfields using DL in remote sensing were summarized, including image fusion, image registration, scene classification, object detection, LULC classification, segmentation, OBIA, and even accuracy assessment. Eventually, a deeper review was conducted to describe and discuss the use of DL algorithms in all of these subfields, which differentiates our study from previous reviews on DL and remote sensing. Therefore, from this review, one can find the various remote sensing applications where DL is being used, and the specific opportunities and challenges for different subfields.

For scene classification, object detection, semantic segmentation, and LULC classification, a supervised DL model (e.g., CNN) must be based on large quantities of training samples. In practice, the acquisition cost of training samples is relatively high, and therefore some augmentation techniques are desirable for increasing the size or efficiency of training datasets, e.g., through transfer learning or active learning. Unsupervised DL models are also attractive for overcoming the training data limitations because they allow for training of the deep network using rich unlabeled data available in the world, e.g., the GAN model introduced recently.

For comparing different DL algorithms (or DL algorithms with other algorithms), it appears necessary to create standard (benchmark) datasets for different remote-sensing image processing tasks and applications. For medium-resolution and low-resolution satellite image data, in particular, there is a lack of benchmark datasets, while several high-resolution benchmark datasets exist (PatternNet, UCMD, aerial image dataset) (Zhou et al., 2018). This lack of benchmark datasets for some types of images and applications may be restricting the development of DL in these areas. Therefore, for medium-resolution and low-resolution remote-sensing images and other remote-sensing data with common features, it is important to create general remote-sensing ground reference datasets suited to different practical applications (e.g., classification), or develop new methods to extend this reference data (Lanaras et al., 2018), because DL has been massively applied to other sorts of remote-sensing data (Sharma et al., 2017; Chen et al., 2017).

For LULC classification, DL showed superprecision performance compared with traditional classifiers (e.g., RF and SVM). Nevertheless, the performance of DL in LULC classification is still inferior compared with scene classification and object detection. This may be attributed to the frequent use of benchmark datasets in scene classification and object detection for previous studies. Therefore, developments in DL for LULC classification are expected because of the diversity of remote-sensing images, and this will have further application in practical remote-sensing images; LULC classification also used standard hyperspectral data or high-resolution images repeatedly in previous publications, instead of real remote-sensing data.

DL models has also been adapted within the remote sensing field to allow for their implementation for non-standard image processing tasks, e.g., object-based image analysis and time series analysis. For object-based image classification, a patch-based strategy is a generally accepted method (Huang et al., 2018; Fu et al., 2018), which integrates CNNs with OBIA. The critical issue with this approach is how to determine the values of the relevant parameters (e.g., patch size), because classification accuracy is largely affected by these parameter values. Time series analysis, a common processing task in remote sensing, has been the focus of very few studies involving DL. Therefore, it is necessary to further explore the application potential of DL in time series analysis (e.g., Landsat or Sentinel), in particular, as DL actually possesses some advantages for the processing of time series data, e.g., the RNN was traditionally applied to a discrete sequence analysis.

For remote sensing image preprocessing, DL models including SAEs and CNNs have been considerably successful recently in remote sensing image fusion, and therefore, more DL models such as RNNs and GANs are expected to be introduced into this field for further developments. The important limitation of DL in image registration is the lack of available public training datasets, which should be a future endeavor of the remote sensing community. For semantic segmentation, the improvement of the network structure, especially the decoder network, remains a challenge to strike the balance between the global context and the local details. Furthermore, in addition to experiments on benchmark datasets, the applications of real remote sensing images deserve further research in the future.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 41701374, 61701160, 41601366), Natural Science Foundation of Jiangsu Province of China (No. BK20170640), National Key R&D Program of China (No. 2017YFB0504200), and the funding provided by the Alexander von Humboldt Foundation of Germany. Sincere thanks to anonymous reviewers and members of the editorial team, for the comments and contributions.

References

Abdi, G., Samadzadegan, F., Reinartz, P., 2017. Spectral-spatial feature learning for hyperspectral imagery classification using deep stacked sparse autoencoder. J. Appl. Remote Sens. 11 (4), 042604.
Audebert, N., Le Saux, B., Lefèvre, S., 2018. Beyond RGB: very high resolution urban remote sensing with multimodal deep networks. ISPRS J. Photogramm. Remote Sens. 140, 20–32.
Ball, J.E., Anderson, D.T., Chan, C.S., 2017. Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community. J. Appl. Remote Sens. 11 (4), 42609.
Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: a review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31.
Bengio, Y., Simard, P., Frasconi, P., 1994. Learning long-term dependencies with gradient descent is difficult. IEEE Trans. Neural Networks 5, 157–166.
Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H., 2007. Greedy layer-wise training of deep networks. Proc. Adv. Neural Inf. Process. Syst. 19, 153–160.
Hamida, B.A., Benoit, A., Lambert, P., 2018. 3-D deep learning approach for remote sensing image classification. IEEE Trans. Geosci. Remote Sens. 56 (8), 4420–4434.
Branson, S., Wegner, J.D., Hall, D., Lang, N., Schindler, K., Perona, P., 2018. From Google Maps to a fine-grained catalog of street trees. ISPRS J. Photogramm. Remote Sens. 135, 13–30.
Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., Shah, R., 1994. Signature verification using a “siamese” time delay neural network. Adv. Neural Inf. Process. Syst. 6 (NIPS 1993) 737–744.
Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B., Li, Z., 2018. A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. Remote Sens. Environ. 210, 35–47.
Chen, Y., Lin, Z., Zhao, X., Wang, G., Gu, Y., 2014. Deep learning-based classification of hyperspectral data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7 (6), 2094–2107.
Chen, Y., Zhao, X., Jia, X., 2015. Spectral-spatial classification of hyperspectral data based on deep belief network. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8 (6SI), 2381–2392.
Chen, Y., Jiao, L., Li, Y., Zhao, J., 2017. Multilayer projective dictionary pair learning and sparse autoencoder for PolSAR image classification. IEEE Trans. Geosci. Remote Sens. 55 (12), 6683–6694.
Chen, Z., Zhang, T., Ouyang, C., 2018a. End-to-end airplane detection using transfer learning in remote sensing images. Remote Sens. 10 (1), 139.
Chen, K., Fu, K., Yan, M., Gao, X., Sun, X., Wei, X., 2018b. Semantic segmentation of aerial images with shuffling convolutional neural networks. IEEE Geosci. Remote Sens. Lett. 15 (2), 173–177.
Chen, G., Zhang, X., Wang, Q., Dai, F., Gong, Y., Zhu, K., 2018c. Symmetrical dense-shortcut deep fully convolutional networks for semantic segmentation of very-high-resolution remote sensing images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11 (5), 1633–1644.
Cheng, G., Zhou, P., Han, J., 2016. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. IEEE Trans. Geosci. Remote Sens. 54 (12), 7405–7415.
Cheng, G., Han, J., 2016. A survey on object detection in optical remote sensing images. ISPRS J. Photogramm. Remote Sens. 117, 11–28.
Cheng, G., Yang, C., Yao, X., Guo, L., Han, J., 2018. When deep learning meets metric learning: remote sensing image scene classification via learning discriminative CNNs. IEEE Trans. Geosci. Remote Sens. 56 (5), 2811–2821.
Chi, M., Sun, Z., Qin, Y., Shen, J., Benediktsson, J.A., 2017. A novel methodology to label urban remote sensing images based on location-based social media photos. Proc. IEEE 105 (10), 1926–1936.
Cho, K., Van Merri¨enboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv:1406.1078.
Ciresan, D., Meier, U., Masci, J., Schmidhuber, J., 2012. Multi-column deep neural network for traffic sign classification. Neural Networks 32, 333–338.
Das, M., Ghosh, S.K., 2016. Deep-step: a deep learning approach for spatiotemporal prediction of remote sensing data. IEEE Geosci. Remote Sens. Lett. 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 J. Photogramm. Remote Sens. 145, 3–22.
Dian, R., Li, S., Guo, A., Fang, L., 2018. Deep hyperspectral image sharpening. IEEE Trans. Neural Networks Learn. Syst. 29, 5345–5355.
Ding, P., Zhang, Y., Deng, W.J., Jia, P., Kuijper, A., 2018. A light and faster regional convolutional neural network for object detection in optical remote sensing images. ISPRS J. Photogramm. Remote Sens. 141, 208–218.
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. J. Appl. Remote Sens. 12, 0250102.
Girard, N., Charpiat, G., Tarabalka, Y., 2018. Aligning and updating cadaster maps with aerial images by multi-task, multi-resolution deep learning. <https://www.lri.fr/~gcharpia/ACCV2018.pdf>.
Ghamisi, P., Yokoya, N., 2018. IMG2DSM: height simulation from single imagery using

conditional generative adversarial net. IEEE Geosci. Remote Sens. Lett. 15 (5), 794–798.
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 J. Photogramm. Remote Sens. 129, 212–225.
Goodfellow, I., Abadie, J., Mirza, M., Xu, B., Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets, arXiv: 1406.2661v1.
Guo, R., Liu, J., Li, N., Liu, S., Chen, F., Cheng, B., Duan, J., Li, X., Ma, C., 2018. Pixel-wise classification method for high resolution remote sensing imagery using deep neural networks. ISPRS Int. J. Geo-Inf. 7 (3), 110.
Hao, S., Wang, W., Ye, Y., Nie, T., Bruzzone, L., 2018. Two-stream deep architecture for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 56 (4), 2349–2361.
Han, W., Feng, R., Wang, L., Cheng, Y., 2017. A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification. ISPRS J. Photogramm. Remote Sens. 145, 23–43.
Hinton, G.E., Salakhutdinov, R., 2006. Reducing the dimensionality of data with neural networks. Science 313, 504–507.
Hinton, G.E., Osindero, S., Teh, Y.-W., 2006. A fast learning algorithm for deep belief nets. Neural Comp. 18, 1527–1554.
Hinton, G.E., 2012. A practical guide to training restricted boltzmann machines. In: Montavon, G., Orr, G.B., Müller, KR. (Eds.), Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science, vol 7700. Springer, Berlin, Heidelberg.
Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A., Jaitly, N., Kingsbury, B., 2012. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Signal Process Mag. 29 (6), 82–97.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778.
He, H., Chen, M., Chen, T., Li, D., 2018. Matching of remote sensing images with complex background variations via siamese convolutional neural network. Remote Sens. 10 (2), 355.
Hochreiter, S. 1991. Untersuchungen zu dynamischen neuronalen Netzen (Diploma thesis), Institut für Informatik, Lehrstuhl Prof. Brauer, Technische Universit München, Advisor: J. Schmidhuber.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9, 1735–1780.
Ho Tong Minh, D., Ienco, D., Gaetano, R., Lalande, N., Ndikumana, E., Osman, F., Maurel, P., 2018. Deep recurrent neural networks for winter vegetation quality mapping via multitemporal SAR sentinel-1. IEEE Geosci. Remote Sens. Lett. 15 (3), 464–468.
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 Sens. Environ. 214, 73–86.
Huang, W., Xiao, L., Wei, Z., Liu, H., Tang, S., 2015. A new pan-sharpening method with deep neural networks. IEEE Geosci. Remote Sens. Lett. 12 (5), 1037–1041.
Hughes, L.H., Schmitt, M., Mou, L., Wang, Y., Zhu, X., 2018a. identifying corresponding patches in SAR and optical images with a Pseudo-Siamese CNN. IEEE Geosci. Remote Sens. Lett. 15 (5), 784–788.
Hughes, L., Schmitt, M., Zhu, X., 2018b. Mining hard negative samples for sar-optical image matching using generative adversarial networks. Remote Sens. 10 (10), 1552.
Ienco, D., Gaetano, R., Dupaquier, C., Maurel, P., 2017. Land cover classification via multitemporal spatial data by deep recurrent neural networks. IEEE Geosci. Remote Sens. Lett. 14 (10), 1685–1689.
Isikdogan, F., Bovik, A.C., Passalacqua, P., 2017. Surface water mapping by deep learning. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10 (11), 4909–4918.
Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. arXiv preprint arXiv:1611.07004.
Kemker, R., Salvaggio, C., Kanan, C., 2018. Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. ISPRS J. Photogramm. Remote Sens. 145, 60–77.
Kellenberger, B., Marcos, D., Tuia, D., 2018. Detecting mammals in UAV images: best practices to address a substantially imbalanced dataset with deep learning. Remote Sens. Environ. 216, 139–153.
Krizhevsky, A., Sutskever, I., Hinton, G., 2012. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems. pp. 1097–1105.
Kussul, N., Lavreniuk, M., Skakun, S., Shelestov, A., 2017. Deep learning classification of land cover and crop types using remote sensing data. IEEE Geosci. Remote Sens. Lett. 14 (5), 778–782.
Lanaras, C., Bioucas-Dias, J., Galliani, S., Baltsavias, E., Schindler, K., 2018. Super-resolution of Sentinel-2 images: learning a globally applicable deep neural network. ISPRS J. Photogramm. Remote Sens. 146, 305–319.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.
Li, Y., Zhang, H., Xue, X., Jiang, Y., Shen, Q., 2018a. Deep learning for remote sensing image classification: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, e1264.
Li, P., Ren, P., Zhang, X., Wang, Q., Zhu, X., Wang, L., 2018b. Region-wise deep feature representation for remote sensing images. Remote Sens. 10 (6), 871.
Li, J., He, Z., Plaza, J., Li, S., Chen, J., Wu, H., Wang, Y., Liu, Y., 2017a. Social media: new perspectives to improve remote sensing for emergency response. Proc. IEEE 105 (10), 1900–1912.
Li, Y., Huang, X., Liu, H., 2017b. Unsupervised deep feature learning for urban village detection from high-resolution remote sensing images. Photogramm. Eng. Remote Sens. 83 (8), 567–579.
Li, W., Fu, H., Yu, L., Cracknell, A., 2017c. Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. Remote Sens. 9 (1), 22.
Li, T., Shen, H., Yuan, Q., Zhang, X., Zhang, L., 2017d. Estimating ground-level pm2.5 by fusing satellite and station observations: a geo-intelligent deep learning approach. Geophys. Res. Lett. 44 (23), 11985–11993.
Liang, H., Li, Q., 2016. Hyperspectral imagery classification using sparse representations of convolutional neural network features. Remote Sens. 8 (2), 99.
Lin, D., Fu, K., Wang, Y., Xu, G., Sun, X., 2017. MARTA GANs: unsupervised representation learning for remote sensing image classification. IEEE Geosci. Remote Sens. Lett. 14 (11), 2092–2096.
Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Sanchez, C.I., 2017. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88.
Liu, T., Abd-Elrahman, A., 2018. Deep convolutional neural network training enrichment using multi-view object-based analysis of Unmanned Aerial systems imagery for wetlands classification. ISPRS J. Photogramm. Remote Sens. 139, 154–170.
Liu, P., Zhang, H., Eom, K.B., 2017a. Active deep learning for classification of hyperspectral images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10 (2), 712–724.
Liu, Y., Minh Nguyen, D., Deligiannis, N., Ding, W., Munteanu, A., 2017b. Hourglass-shapenetwork based semantic segmentation for high resolution aerial imagery. Remote Sens. 9 (6), 522.
Liu, Y., Chen, X., Wang, Z., Wang, Z.J., Ward, R.K., Wang, X., 2018. Deep learning for pixel-level image fusion: Recent advances and future prospects. Inf. Fusion 42, 158–173.
Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proc IEEE Int. Conf. Computer Vision and Pattern Recognition CVPR, pp. 3431–3440.
Luus, F.P.S., Salmon, B.P., Van Den Bergh, F., Maharaj, B.T.J., 2015. Multiview deep learning for land-use classification. IEEE Geosci. Remote Sens. Lett. 12 (12), 2448–2452.
Lv, Q., Dou, Y., Niu, X., Xu, J., Xu, J., Xia, F., 2015. Urban land use and land cover classification using remotely sensed SAR data through deep belief networks. J. Sens. 2015 538063(10).
Lyu, H., Lu, H., Mou, L., 2016. Learning a transferable change rule from a recurrent neural network for land cover change detection. Remote Sens. 8 (6), 506.
Lyu, H., 2018. Long-term annual mapping of four cities on different continents by applying a deep information learning method to landsat data. Remote Sens. 10 (3), 471.
Ma, L., Li, M., Ma, X., Cheng, L., Du, P., Liu, Y., 2017. A review of supervised object-based land-cover image classification. ISPRS J. Photogramm. Remote Sens. 130, 277–293.
Masi, G., Cozzolino, D., Verdoliva, L., Scarpa, G., 2016. Pansharpening by convolutional neural networks. Remote Sens. 8 (7), 594.
Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. Convolutional neural networks for large-scale remote-sensing image classification. IEEE Trans. Geosci. Remote Sens. 55 (2), 645–657.
Marcos, D., Volpi, M., Kellenberger, B., Tuia, D., 2018. Land cover mapping at very high resolution with rotation equivariant CNNs: towards small yet accurate models. ISPRS J. Photogramm. Remote Sens. 145, 96–107.
Marmanis, D., Datcu, M., Esch, T., Stilla, U., 2016. Deep learning earth observation classification using ImageNet pretrained networks. IEEE Geosci. Remote Sens. Lett. 13 (1), 105–109.
Merkle, N., Luo, W., Auer, S., Muller, R., Urtasun, R., 2017. Exploiting deep matching and sar data for the geo-localization accuracy improvement of optical satellite images. Remote Sens. 9 (6), 18.
Merkle, N., Auer, S., Müller, R., Reinartz, P., 2018. Exploring the potential of conditional adversarial networks for optical and SAR image matching. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11 (6), 1811–1820.
Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J., 2013. Distributed representations of words and phrases and their compositionality. Proc. Adv. Neural Inf. Process. Syst. 26, 3111–3119.
Mou, L., Ghamisi, P., Zhu, X.X., 2017. Deep recurrent neural networks for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 55 (7), 3639–3655.
Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: a review. ISPRS J. Photogramm. Remote Sens. 66 (3), 247–259.
Ning, F., Delhomme, D., LeCun, Y., Piano, F., Bottou, L., Barbano, P.E., 2005. Toward automatic phenotyping of developing embryos from videos. IEEE Trans. Image Process. 14, 1360–1371.
Oliehoek, F.A., Savani, R., Gallego-Posada, J., Van der Pol, E., De Jong, E.D., Groß, R., 2017. GANGs: Generative adversarial network games. ArXiv e-prints arXiv:1712.00679v2.
Pan, B., Shi, Z., Xu, X., 2017. MugNet: Deep learning for hyperspectral image classification using limited samples. ISPRS J. Photogramm. Remote Sens. 145, 108–119.
Palsson, F., Sveinsson, J.R., Ulfarsson, M.O., 2017. Multispectral and hyperspectral image fusion using a 3-d-convolutional neural network. IEEE Geosci. Remote Sens. Lett. 14 (5), 639–643.
Romero, A., Gatta, C., Camps-Valls, G., 2016. Unsupervised deep feature extraction for remote sensing image classification. IEEE Trans. Geosci. Remote Sens. 54 (3), 1349–1362.
Rottensteiner, F., Sohn, G., Jung, J., Gerke, M., Baillard, C., Benitez, S., Breitkopf, U., 2012. The ISPRS benchmark on urban object classification and 3D building reconstruction. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume I-3, 2012 XXII ISPRS Congress, Melbourne, Australia, pp. 293–298.
Scarpa, G., Vitale, S., Cozzolino, D., 2018. Target-adaptive CNN-based pansharpening. IEEE Trans. Geosci. Remote Sens. 56 (9), 5443–5457.
Schmidhuber, J., 2015. Deep learning in neural networks: an overview. Neural Networks 61, 85–117.
Scott, G.J., England, M.R., Starms, W.A., Marcum, R.A., Davis, C.H., 2017. Training deep convolutional neural networks for land-cover classification of high-resolution imagery. IEEE Geosci. Remote Sens. Lett. 14 (4), 549–553.

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.
Shao, Z., Zhang, L., Wang, L., 2017. Stacked sparse autoencoder modeling using the synergy of airborne lidar and satellite optical and sar data to map forest above-ground biomass. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10 (12), 5569–5582.
Shao, Z., Cai, J., 2018. Remote sensing image fusion with deep convolutional neural network. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11 (5), 1656–1669.
Sherrah, J., 2016. Fully convolutional networks for dense semantic labelling of high-resolution aerial imagery. arXiv preprint arXiv:1606.02585.
Shin, H., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M., 2016. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans. Med. Imaging 35, 1285–1298.
Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv. Available: https://arxiv.org/pdf/1409.1556.pdf.
Sutskever, I., Martens, J., Hinton, G.E., 2011. Generating text with recurrent neural networks. In: Proc. 28th International Conference on Machine Learning, pp. 1017–1024.
Sutskever, I., 2012. Training recurrent neural networks. PhD thesis. Univ. Toronto.
Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A., 2017. Inception-v4, inception-resnet and the impact of residual connections on learning. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17), pp. 4278–4284.
Tao, Y., Gao, X., Hsu, K., Sorooshian, S., Ihler, A., 2016. A deep neural network modeling framework to reduce bias in satellite precipitation products. J. Hydrometeor. 17, 931–945.
Tong, X.Y., Xia, G.S., Lu, Q.K., Shen, H.F., Li, S.Y., You, S.C., Zhang, L.P., 2018. Learning transferable deep models for land-use classification with high-resolution remote sensing images. https://arxiv.org/abs/1807.05713.
Tuia, D., Flamary, R., Courty, N., 2015. Multiclass feature learning for hyperspectral image classification: Sparse and hierarchical solutions. ISPRS J. Photogramm. Remote Sens. 105, 272–285.
Vetrivel, A., Gerke, M., Kerle, N., Nex, F., Vosselman, G., 2018. Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning. ISPRS J. Photogramm. Remote Sens. 140, 45–59.
Wang, C., Zhang, L., Wei, W., Zhang, Y., 2018a. When low rank representation based hyperspectral imagery classification meets segmented stacked denoising auto-encoder based spatial-spectral feature. Remote Sens. 10 (2), 284.
Wang, S., Quan, D., Liang, X., Ning, M., Guo, Y., Jiao, L., 2018b. A deep learning framework for remote sensing image registration. ISPRS J. Photogramm. Remote Sens. 145, 148–164.
Wang, J., Luo, C., Huang, H., Zhao, H., Wang, S., 2017a. Transferring pre-trained deep CNNs for remote scene classification with general features learned from linear PCA network. Remote Sens. 9 (3), 225.
Wang, H., Wang, Y., Zhang, Q., Xiang, S., Pan, C., 2017b. Gated convolutional neural network for semantic segmentation in high-resolution images. Remote Sens. 9 (5), 446.
Wei, Y., Yuan, Q., Shen, H., Zhang, L., 2017. Boosting the accuracy of multispectral image pansharpening by learning a deep residual network. IEEE Geosci. Remote Sens. Lett. 14 (10), 1795–1799.
Weng, Q., Mao, Z., Lin, J., Guo, W., 2017. Land-use classification via extreme learning classifier based on deep convolutional features. IEEE Geosci. Remote Sens. Lett. 14 (5), 704–708.
Wu, Q., Diao, W., Dou, F., Sun, X., Zheng, X., Fu, K., Zhao, F., 2016. Shape-based object extraction in high-resolution remote-sensing images using deep Boltzmann machine. Int. J. Remote Sens. 37 (24), 6012–6022.
Xia, G., Hu, J., Hu, F., Shi, B., Bai, X., Zhong, Y., Lu, X., 2017. AID: A benchmark data set for performance evaluation of aerial scene classification. IEEE Trans. Geosci. Remote Sens. 55 (7), 3965–3981.
Xia, G.S., Yang, W., Delon, J., Gousseau, Y., Sun, H., Ma, H., 2010. Structural high-resolution satellite image indexing. In ISPRS TC VII Symposium-100 Years ISPRS, vol. 38, pp. 298-303.
Xing, Y., Wang, M., Yang, S., Jiao, L., 2018a. Pan-sharpening via deep metric learning. ISPRS J. Photogramm. Remote Sens. 145, 165–183.
Xing, H., Meng, Y., Wang, Z., Fan, K., Hou, D., 2018b. Exploring geo-tagged photos for land cover validation with deep learning. ISPRS J. Photogramm. Remote Sens. 141, 237–251.
Xu, X., Li, W., Ran, Q., Du, Q., Gao, L., Zhang, B., 2018a. Multisource remote sensing data classification based on convolutional neural network. IEEE Trans. Geosci. Remote Sens. 56 (2), 937–949.
Xu, Y., Wu, L., Xie, Z., Chen, Z., 2018b. Building extraction in very high resolution remote sensing imagery using deep learning and guided filters. Remote Sens. 10 (1), 144.
Yao, Y., Li, X., Liu, X., Liu, P., Liang, Z., Zhang, J., Mai, K., 2017. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. Int. J. Geog. Inf. Sci. 31 (4), 825–848.
Yang, X., Liu, W., Tao, D., Cheng, J., Li, S., 2017. Multiview canonical correlation analysis networks for remote sensing image recognition. IEEE Geosci. Remote Sens. Lett. 14 (10), 1855–1859.
Yang, J., Zhao, Y.Q., Chan, J.C.W., 2018. Hyperspectral and multispectral image fusion via deep two-branches convolutional neural network. Remote Sens. 10, 800.
Ye, Y., Shan, J., Bruzzone, L., Shen, L., 2017. Robust registration of multimodal remote sensing images based on structural similarity. IEEE Trans. Geosci. Remote Sens. 55 (5), 2941–2958.
Yu, Y., Guan, H., Zai, D., Ji, Z., 2016. Rotation-and-scale-invariant airplane detection in high-resolution satellite images based on deep-Hough-forests. ISPRS J. Photogramm. Remote Sens. 112, 50–64.
Yu, X., Wu, X., Luo, C., 2017. Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework. GIScience Remote Sens. 54 (5), 741–758.
Yu, Y., Gong, Z., Wang, C., Zhong, P., 2018. An unsupervised convolutional feature fusion network for deep representation of remote sensing images. IEEE Geosci. Remote Sens. Lett. 15 (1), 23–27.
Yuan, Q., Wei, Y., Meng, X., Shen, H., Zhang, L., 2018. A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11 (3), 978–989.
Zabalza, J., Ren, J., Zheng, J., Zhao, H., Qing, C., Yang, Z., Du, P., Marshall, S., 2016. Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging. Neurocomputing 185, 1–10.
Zampieri, A., Charpiat, G., Girard, N., Tarabalka, Y., 2018. Multimodal image alignment through a multiscale chain of neural networks with application to remote sensing. http://openaccess.thecvf.com/content_ECCV_2018/papers/Armand_Zampieri_Multimodal_image_alignment_ECCV_2018_paper.pdf.
Zitova, B., Flusser, J., 2003. Image registration methods: a survey. Image Vision Comput. 21 (11), 977–1000.
Zhao, W., Du, S., 2016. Learning multiscale and deep representations for classifying remotely sensed imagery. ISPRS J. Photogramm. Remote Sens. 113, 155–165.
Zhao, W., Du, S., Wang, Q., Emery, W.J., 2017a. Contextually guided very-high-resolution imagery classification with semantic segments. ISPRS J. Photogramm. Remote Sens. 132, 48–60.
Zhao, W., Du, S., Emery, W.J., 2017b. Object-based convolutional neural network for high-resolution imagery classification. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10 (7), 3386–3396.
Zhan, Y., Hu, D., Wang, Y., Yu, X., 2018. Semisupervised hyperspectral image classification based on Generative Adversarial Networks. IEEE Geosci. Remote Sens. Lett. 15 (2), 212–216.
Zhang, L., Zhang, L., Du, B., 2016. Deep learning for remote sensing data: a technical tutorial on the state of the art. IEEE Geosci. Remote Sens. Mag. 4 (2), 22–40.
Zhang, M., Hu, X., Zhao, L., Lv, Y., Luo, M., Pang, S., 2017a. Learning dual multi-scale manifold ranking for semantic segmentation of high-resolution images. Remote Sens. 9 (5), 500.
Zhang, M., Hu, X., Zhao, L., Pang, S., Gong, J., Luo, M., 2017b. Translation-aware semantic segmentation via conditional least-square generative adversarial networks. J. Appl. Remote Sens. 11 (4), 042622.
Zhang, X., Chen, G., Wang, W., Wang, Q., Dai, F., 2017c. Object-based land-cover supervised classification for very-high-resolution UAV images using stacked denoising autoencoders. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10 (7), 3373–3385.
Zhang, R., Li, G., Li, M., Wang, L., 2018a. Fusion of images and point clouds for the semantic segmentation of large-scale 3D scenes based on deep learning. ISPRS J. Photogramm. Remote Sens. 143, 85–96.
Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., Atkinsona, P.M., 2018b. An object-based convolutional neural network OCNN for urban land use classification. Remote Sens. Environ. 216, 57–70.
Zhang, B., Gu, J., Chen, C., Han, J., Su, X., Cao, X., Liu, J., 2018c. One-two-one networks for compression artifacts reduction in remote sensing. ISPRS J. Photogramm. Remote Sens. 145, 184–196.
Zhong, Y., Han, X., Zhang, L., 2018. Multi-class geospatial object detection based on a position-sensitive balancing framework for high spatial resolution remote sensing imagery. ISPRS J. Photogramm. Remote Sens. 138, 281–294.
Zhong, Y., Fei, F., Zhang, L., 2016. Large patch convolutional neural networks for the scene classification of high spatial resolution imagery. J. Appl. Remote Sens. 10 (2), 025006.
Zhu, L., Chen, Y., Ghamisi, P., Benediktsson, J.A., 2018. Generative adversarial networks for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 56 (9), 5046–5063.
Zhu, X.X., Tuia, D., Mou, L., Xia, G., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: a comprehensive review and list of resources. IEEE Geosci. Remote Sens. Mag. 5 (4), 8–36.
Zhuo, X., Fraundorfer, F., Kurz, F., Reinartz, P., 2018. Optimization of openstreetmap building footprints based on semantic information of oblique UAV images. Remote Sens. 10 (4), 624.
Zou, Q., Ni, L., Zhang, T., Wang, Q., 2015. Deep learning based feature selection for remote sensing scene classification. IEEE Geosci. Remote Sens. Lett. 12 (11), 2321–2325.
Zhou, W., Newsam, S., Li, C., Shao, Z., 2018. PatternNet: A benchmark dataset for performance evaluation of remote sensing image retrieval. ISPRS J. Photogramm. Remote Sens. 145, 197–209.