International Journal of Applied Earth Observations and Geoinformation 112 (2022) 102926

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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

Deep learning in multimodal remote sensing data fusion: A comprehensive review

Jiaxin Li [a,c], Danfeng Hong [a], Lianru Gao [a,*], Jing Yao [a], Ke Zheng [d,a], Bing Zhang [b,c], Jocelyn Chanussot [e,b]

[a] Key Laboratory of Computational Optical Imaging Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China
[b] Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China
[c] College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China
[d] College of Geography and Environment, Liaocheng University, Liaocheng 252059, China
[e] University Grenoble Alpes, CNRS, Grenoble INP, GIPSA-Lab, Grenoble 38000, France

A R T I C L E I N F O

Keywords:
Artificial intelligence
Data fusion
Deep learning
Multimodal
Remote sensing

A B S T R A C T

With the extremely rapid advances in remote sensing (RS) technology, a great quantity of Earth observation (EO) data featuring considerable and complicated heterogeneity are readily available nowadays, which renders researchers an opportunity to tackle current geoscience applications in a fresh way. With the joint utilization of EO data, much research on multimodal RS data fusion has made tremendous progress in recent years, yet these developed traditional algorithms inevitably meet the performance bottleneck due to the lack of the ability to comprehensively analyze and interpret strongly heterogeneous data. Hence, this non-negligible limitation further arouses an intense demand for an alternative tool with powerful processing competence. Deep learning (DL), as a cutting-edge technology, has witnessed remarkable breakthroughs in numerous computer vision tasks owing to its impressive ability in data representation and reconstruction. Naturally, it has been successfully applied to the field of multimodal RS data fusion, yielding great improvement compared with traditional methods. This survey aims to present a systematic overview in DL-based multimodal RS data fusion. More specifically, some essential knowledge about this topic is first given. Subsequently, a literature survey is conducted to analyze the trends of this field. Some prevalent sub-fields in the multimodal RS data fusion are then reviewed in terms of the to-be-fused data modalities, i.e., spatiospectral, spatiotemporal, light detection and ranging-optical, synthetic aperture radar-optical, and RS-Geospatial Big Data fusion. Furthermore, We collect and summarize some valuable resources for the sake of the development in multimodal RS data fusion. Finally, the remaining challenges and potential future directions are highlighted.

1. Introduction

On account of the superiority in observing our Earth environment, RS has been playing an increasingly important role in various EO tasks (Hong et al., 2021b; Zhang et al., 2019a). With the ever-growing availability of multimodal RS data, researchers have easy access to the data which are suitable for the application at hand. Although a large amount of multimodal data become readily available, each modality can barely capture one or few specific properties and hence cannot fully describe the observed scenes, which poses a great constraint on subsequent applications. Naturally, multimodal RS data fusion is a feasible way to break out of the dilemma induced by unimodal data. By integrating the complementary information extracted from multimodal data, a more robust and reliable decision can be made in many tasks, such as change detection, LULC classification, etc.

Unlike multisource and multitemporal RS, the term of "modality" has been a lack of a clear and unified definition. In this paper, we attempted to give a detailed definition on basis of previous works (Gómez-Chova et al., 2015; Dalla Mura et al., 2015). Principally, RS data are characterized by two main factors, i.e., the technical specifications of the sensors and the actual acquisition condition. Specifically, the former determine the internal characteristics of the product, e.g., imaging mechanism and the resolutions. While, the latter controls the external properties, e.g., the acquisition time, observation angles, and mounted platforms. Thus, the aforementioned factors contribute to

* Corresponding author.
E-mail addresses: lijiaxin203@mails.ucas.ac.cn (J. Li), hongdf@aircas.ac.cn (D. Hong), gaolr@aircas.ac.cn (L. Gao), yaojing@aircas.ac.cn (J. Yao), zhengke@lcu.edu.cn (K. Zheng), zb@radi.ac.cn (B. Zhang), jocelyn@hi.is (J. Chanussot).

https://doi.org/10.1016/j.jag.2022.102926
Received 29 April 2022; Received in revised form 13 July 2022; Accepted 14 July 2022
Available online 26 July 2022

Table 1
List of the main abbreviations.

Abbreviation	Description	Abbreviation	Description
AE	Autoencoder	LULC	Land use and land cover
CS	Component substitution	LiDAR	Light detection and ranging
CNN	Convolutional neural network	MF	Matrix factorization
DHP	Deep hyperspectral prior	MRA	Multiresolution analysis
DL	Deep learning	MS	Multispectral
DI	Details injection	NDVI	Normalized difference vegetation index
EO	Earth observation	Pan	Panchromatic
EP	Extinction profile	POI	Points of interest
GAN	Generative adversarial network	RS	Remote sensing
GBD	Geospatial big data	SAR	Synthetic aperture radar
GNN	Graph neural network	TR	Tensor representation
HS	Hyperspectral	VO	Variational optimization
LST	Land surface temperature	ViT	Visual transformer

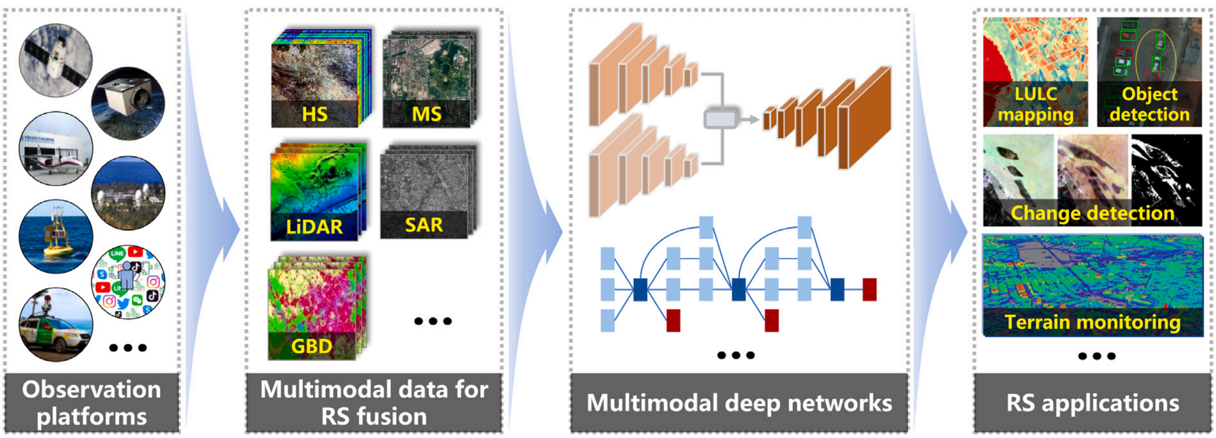

Fig. 1. An illustration of DL in multimodal RS data fusion.

the descriptions of the captured scene and can be described as "modality". Apparently, multimodal RS data fusion includes multisource RS data fusion and multitemporal RS data fusion.

Some typical RS modalities include Pan, MS, HS, LiDAR, SAR, infrared, night time light, and satellite video data. Very recently, GBD, as a new member in the RS family, have attracted growing attention in the EO tasks. To integrate the complementary information provided by these modalities, traditional methods have been intensively studied by designing handcrafted features based on domain-specific knowledge and exploiting rough fusion strategies, which inevitably impairs the fusion performance, especially for heterogeneous data (Hong et al., 2021a). Thanks to the growth of artificial intelligence, DL shows great potential in modeling a complicated relationship between input and output data by adaptively realizing the feature extraction and fusion in an automatic manner. Depending on the to-be-fused modalities and corresponding tasks, DL-based multimodal RS data fusion can be generalized into a unified framework (see Fig. 1). Accordingly, this review will focus on the methods proposed in each fusion subdomain along with a brief introduction in each modality and related tasks.

Currently, there exist some literature reviews regarding multimodal data fusion, which are summarized in Table 2 according to different modality fusion. Existing reviews either pay less attention to the direction of DL or only cover few sub-areas in multimodal RS data fusion, lacking a comprehensive and systematic description on this topic. The motivation of our survey is to give a comprehensive review of popular domains in DL-based multimodal RS data fusion, and further facilitate and promote the relevant research in this burgeoning domain. More specifically, literature related to this topic is collected and analyzed in Section 2, followed by Section 3, which elaborates on representative sub-fields in multimodal RS data fusion. In Section 4, some useful resources in respect of tutorials, datasets and codes are given. Finally, Section 5 provides remarks concerning the challenges and prospects. For the convenience of readers, main abbreviations used in this article are listed in Table 1.

2. Literature analysis

2.1. Data retrieval and collection

In this section, Web of Science and CiteSpace (Chen, 2006) are chosen as the main analysis tools. Taking the Query one in Table 3 for example, 691 results are initially returned from Web of Science Core Collection by using the advanced search: TS=("remote sensing") AND TS=("deep learning") AND TS=("fusion"). After only considering the "Article" document type, 598 papers published from 2015 to 2022 are included for the subsequent analysis.

2.2. Statistical analysis and results

2.2.1. Statistical analysis of articles published annually

The trend of related papers published in 2015–2022 is shown in Fig. 2. The bar chart suggests that growing attention has been paid to this burgeoning field with a steady increase in the number of publications. On the other hand, the upward trend in the line graph is consistent with that in the bar chart, which indicates that DL technologies have been playing an increasingly important role in the field of multimodal RS data fusion.

2.2.2. Statistical analysis of the distribution of publications in terms of countries and journals

Two pie charts showing the proportion of published papers by the top 10 countries and journals are displayed in Fig. 3(a) and Fig. 3(b), respectively. It can be seen that the top 10 countries take up about 90% of the total outputs, constituting the main pillar of this direction.

Table 2
Typical multimodal data fusion reviews.

	Domains	References	Descriptions
Homogeneous fusion	Pansharpening	Ranchin et al. (2003)	Introducing the methods belonging to ARSIS, along with giving a simple comparison
		Vivone et al. (2014)	Giving thorough descriptions and assessments of the methods belonging to CS and MRA families
		Meng et al. (2019)	Introducing the methods belonging to CS, MAR, and VO from the idea of meta-analysis
		Vivone et al. (2020)	Giving a systematic introduction and evaluation of the methods in the category of CS, MAR, VO, and ML
	HS pansharpening	Loncan et al. (2015)	Conducting a comprehensive analysis and evaluation in the methods from CS, MAR, hybrid, bayesian, and MF
	HS-MS fusion	Yokoya et al. (2017)	Extensive experiments are presented to assess the methods from CS, MRA, unmixing, and bayesian
		Dian et al. (2021b)	Studying the performance of methods from CS, MAR, MF, TR, and DL
	Spatiotemporal	Chen et al. (2015)	Discussing and evaluating four models from transformation/reconstruction/learning-based methods
		Zhu et al. (2018)	Reviewing the characteristics of five categories and their applications
		Belgiu and Stein (2019)	Introducing the methods in three categories, as well as the challenges and opportunities
		Li et al. (2020b)	Analyzing the performance of representative methods with their provided benchmark dataset
Heterogeneous fusion	HS-LiDAR	Man et al. (2014)	Summarizing the research on HS-LiDAR fusion for forest biomass estimation
		Kuras et al. (2021)	Giving an overview of HS-LiDAR fusion in the application of land cover classification
	SAR-optical	Kulkarni and Rege (2020)	Evaluating the performance of methods from CS and MRA in pixel-level
	RS-GBD	Li et al. (2021b)	Providing a review on RS-social media fusion and their distributed strategies
		Yin et al. (2021a)	Reviewing the fusion of RS-GBD in the application of urban land use mapping from feature-level and decision-level perspectives
Others		Wald (1999)	Setting up some definitions regrading data fusion
		Gómez-Chova et al. (2015)	Providing a review in seven data fusion applications for RS
		Lahat et al. (2015)	Summarizing the challenges in multimodal data fusion across various disciplines
		Dalla Mura et al. (2015)	Giving a comprehensive discussion on data fusion problems in RS by analyzing the Data Fusion Contests
		Ghassemian (2016)	Introducing the RS fusion methods in pixel/feature/decision-level and different evaluation criteria
		Schmitt and Zhu (2016)	Modeling the data fusion process, along with introducing some typical fusion scenarios in RS
		Li et al. (2017)	Introducing fusion methods in pixel-level and their major applications
		Liu et al. (2018)	Reviewing DL-based pixel-level fusion methods in digital photography, multi-modality imaging, and RS imagery
		Ghamisi et al. (2019)	Conducting a detailed review in spatiospectral, spatiotemporal, HS-LiDAR, etc
		Zhang et al. (2021d)	Reviewing DL-based fusion methods in digital photography, multi-modal image, sharpening fusion
		Kahraman and Bacher (2021)	Describing methods in HS-LiDAR and HS-SAR fusion

Table 3
Data retrieval results of WOS from 2015 to 2022.

Query	Contents	Original results	Refined results
Q1	(TS=("remote sensing") AND TS=("deep learning") AND TS=("fusion"))	691	598
Q2	(TS=("remote sensing") AND TS=("fusion"))	6483	4403

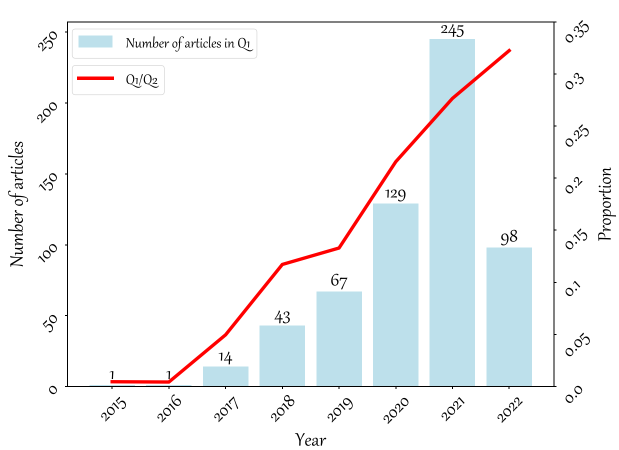

Fig. 2. Number of published articles annually in Q1 and its proportion on Q2.

More concretely, China makes a major contribution to the field, which accounts for more than half of all publications, followed by USA, which occupies about 10%.

Besides, *Remote Sensing*, *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, and *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* make up about half of the overall publications, with *Remote Sensing* ranking first.

2.2.3. Statistical analysis of the keywords in the literature

Fig. 4 exhibits the keywords appearing in the collected articles, where a bigger font size corresponds to a higher frequency. As the figure indicates, CNN is widely used in the field of DL-based multimodal RS data fusion. Besides, classification, cloud removal, and object detection become the main tasks in the fusion process, where MS, HS, LiDAR and SAR are the mainly-used data.

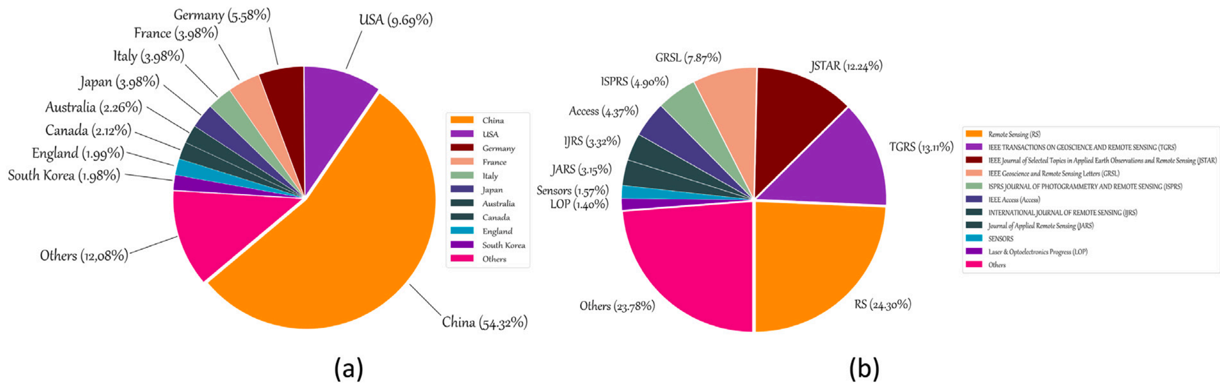

Fig. 3. Proportion of published articles by top 10 (a) countries and (b) journals.

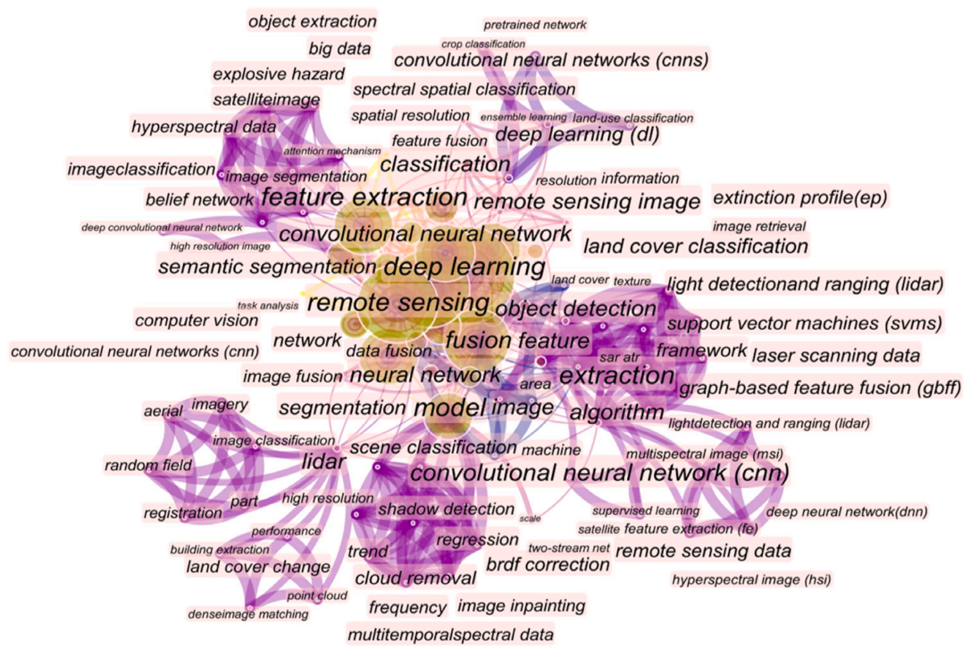

Fig. 4. A visualization of the keyword co-occurrence network.

3. A review of DL-based multimodal remote sensing data fusion methods

This paper divide existing methods into two main groups, i.e., homogeneous fusion and heterogeneous fusion. Specifically, homogeneous fusion refers to pansharpening, HS pansharpening, HS-MS fusion, and spatiotemporal fusion, while heterogeneous fusion includes LiDAR-optical, SAR-optical, and RS-GBD fusion. Since the aforementioned sub-fields develop quite diversely, different criteria are adopted to introduce each subdomain, as shown in Fig. 5. For the convenience of readers, we also list some classic literature in each direction.

3.1. Homogeneous fusion

The homogeneous fusion, including spatiospectral fusion (i.e., pansharpening, HS pansharpening, and HS-MS fusion) and spatiotemporal fusion, is primarily committed to solving the trade-off in spatial–spectral and spatial–temporal resolutions happening in the optical images due to the imaging mechanism. This section will introduce typical methods proposed in these domains.

3.1.1. Pansharpening

Pansharpening refers to the fusion of MS and Pan to generate a high spatial resolution MS image. In general, AE, CNN, and GAN are commonly-used network architectures for DL-based pansharpening.

- **Supervised methods**

It is well-known that supervised methods perform the pansharpening by linking the observations with the references. Usually, the input data need to be simulated by spatially downsampling the original data. Huang et al. (2015) propose the first DL-based method in dealing with pansharpening problem, where a sparse denoising AE is adopted to learn the transformation in Pan domain, and then the observed MS is input into the pretrained AE to generate the final output. Following this milestone work, many methods are successively proposed by treating pansharpening as an image super-resolution problem (Azarang and Ghassemian, 2017; Xing et al., 2018). Apart from AE structure, CNN is also extensively used and can be categorized into three major groups, i.e., single-branch, multi-branch, and hybrid network. Methods belonging to the first group simply concatenate input Pan and upsampled MS or their pre-processed versions into a new component as the input of networks. For example, Masi et al. (2016) propose the first CNN-based pansharpening methods with three convolutional layers by

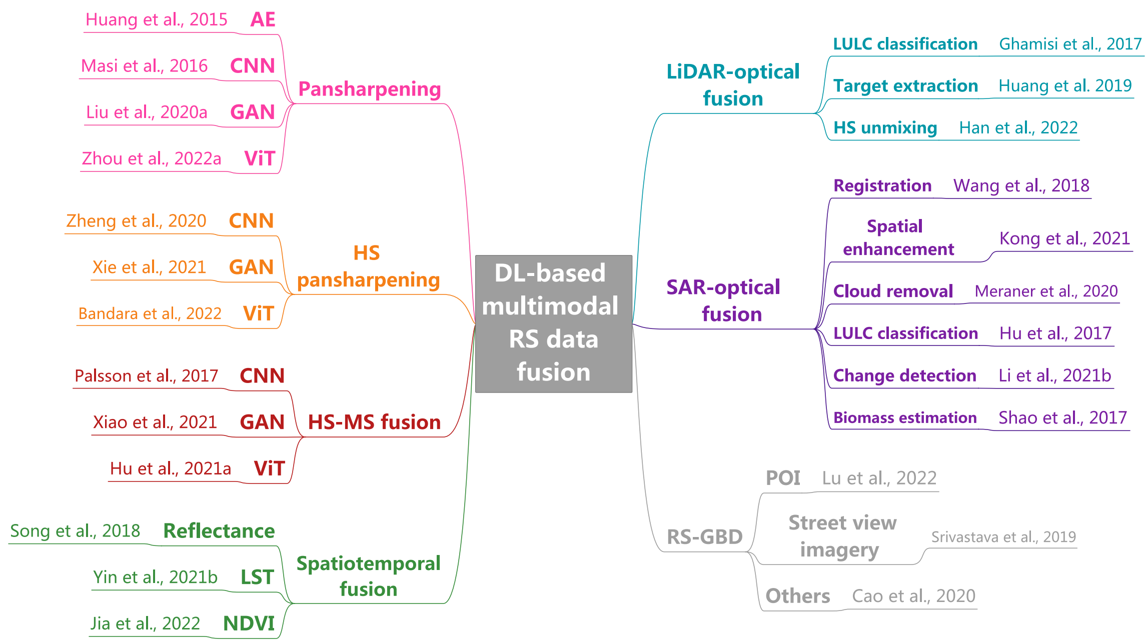

Fig. 5. The taxonomy of DL-based multimodal RS data fusion in this paper.

adapting the SRCNN architecture. Later, numerous methods inspired by this pioneer work are presented, in which residual learning and dense connection are commonly used (Wei et al., 2017; Yang et al., 2017; Scarpa et al., 2018; Yuan et al., 2018; Peng et al., 2020; Fu et al., 2020; Lei et al., 2021). However, simply stacking pre-interpolated MS with Pan as the input of networks not only ignores individual features but also raises an extra computational burden. Hence, instead of treating the two modalities equally, multi-branch networks apply different sub-networks to separately extract the modality-specific features (Shao and Cai, 2018; Zhang et al., 2019b; Liu et al., 2020a; Chen et al., 2021; Zhang and Ma, 2021; Xing et al., 2020; Yang et al., 2022a). Hybrid network-based methods provide a cutting-edge solution to pansharpening by embracing the conception of traditional methods, i.e., DI-based methods (He et al., 2019a; Deng et al., 2020) and VO-based methods (Shen et al., 2019; Cao et al., 2021; Tian et al., 2021), and therefore effectively merging the strengths in both domains. Different from CNN, GAN-based methods treat pansharpening as an image generation problem by establishing a adversarial game between a generator and a discriminator network. The first GAN-based pansharpening method designs a two-branch generator network (Liu et al., 2020b), and then different loss functions and new network structures are explored to extract more discriminative features (Shao et al., 2019; Ozcelik et al., 2020; Gastineau et al., 2022). ViT is recently introduced into pansharpening due to its ability in capturing long-range information (Zhou et al., 2021a, 2022a).

- **Unsupervised methods**

Scale-related problems may occur in supervised methods since they are often trained at a lower resolution. However, unsupervised methods implement the training and testing processes at the original scale without the need to simulate the references. Hence, the key lies in precisely establishing the relationships between the input data and the fused product by designing proper loss functions, i.e., the degraded fusion result should be identical to input Pan and MS in the spatial and spectral domains, respectively. For example, Ma et al. (2020) utilize a spatial adversarial loss to represent the spatial information hidden in the output of the generator. Besides, other widely-used loss functions include gradient loss (Seo et al., 2020), perceptual loss (Zhou et al., 2020), and non-reference loss (Zhou et al., 2021b; Luo et al., 2020).

3.1.2. HS pansharpening

Similar to pansharpening, HS pansharpening intends to combine spectral information in HS with spatial information in Pan to produce a HS image with high spatial resolution.

- **Supervised methods**

Supervised methods aim to learn the transformation from inputs to target data which do not exist in real world, and thus simulation experiments are usually implemented. Specifically, a pair of low spatial resolution HS and low spectral resolution MS is generated by spatially and spectrally degrading the observed HS, respectively. By doing so, the two simulated images are regarded as the inputs of networks and the original HS serves as references.

Like those pioneer works in pansharpening, CNN and GAN are naturally applicable to HS pansharpening task. Zheng et al. (2020) propose a single-branch CNN-based method, where multiple channel-spatial-attention blocks are cascaded to adaptively extract informative features. Inspired by the representative work, DHP is further enhanced by adding spatial-related constraints to optimize the procedure of HS upsampling (Bandara et al., 2022). To recover the missing information hidden in the inputs, residual-style networks are broadly utilized in two-branch HS pansharpening networks. Especially, He et al. (2019b) clearly exhibit the superiority of skip connection in terms of training efficiency. There are also enormous efforts aiming to tackle specific problems, such as spectral-fidelity (He et al., 2020; Guan and Lam, 2021), pansharpening with arbitrary resolution enhancement (He et al., 2021b), and arbitrary spectral bands (Qu et al., 2022a). The hybrid networks, such as DI-embedded methods (Dong et al., 2021c) and VO-embedded methods (Xie et al., 2020), can adaptively learn the spatial details and deep priors that need a explicit modeling by traditional methods. Additionally, Dong et al. (2021d) directly unfold the iterative optimization algorithm into a end-to-end network, where degradation models are considered to exploit the prior information. Following the idea presented in pansharpening, GAN is successfully applied to HS pansharpening with various designs of the discriminators. A typical example given by Xie et al. (2021) utilizes a spatial discriminator to restrain the difference between input Pan and the spectrally down-sampled version of the generated output, where the generator network is trained in the high frequency. Other commonly-used discriminators include the spectral discriminator (Dong et al., 2021b) and the spatial–spectral discriminator (Dong et al., 2021a). Transformer also finds its application in HS pansharpening by Bandara and Patel (2022), in which modality-specific feature extractor are designed to capture textural details for subsequent spectral details fusion.

- **Unsupervised methods**

The unsupervised HS pansharpening is rarely studied compared with pansharpening. One possible reason is that the input Pan and MS share similar spectral coverage, while there exists a big discrepancy between Pan and HS in the spectral range, which leads to the difficulty in preserving spatial information. A tentative work by Nie et al. (2022) utilizes a gradient and a high-frequency loss to model the spatial relationship, where an initialized image is first generated by the ratio estimation strategy.

3.1.3. HS-MS fusion

Pansharpening related works can be regarded as special cases of HS-MS fusion which aims to attain HS product with high spatial resolution by fusing paired HS-MS images. Therefore, many DL-based pansharpening methods can be transferred to tackle HS-MS fusion with necessary modifications. Following this, typical methods will be introduced in accordance with the same taxonomy in pansharpening.

- **Supervised methods**

The supervised HS-MS fusion follows the same scheme of HS pansharpening by replacing the input Pan with MS. Single-branch HS-MS fusion methods are put forward with classic structures, such as 3-D CNN (Palsson et al., 2017), residual network (Han and Chen, 2019), dense connection network (Han et al., 2018), and three-component network (Zhang et al., 2021a), etc. Compared with these single-branch work that directly upsamples HS to the same resolution as MS, multi-branch methods adopt an alternative strategy to relax this problem, i.e., by gradually upsampling HS through the operation of deconvolution or pixel shuffle, where spatial information extracted from MS is injected into the corresponding scale (Xu et al., 2020a; Han et al., 2019; Zhou et al., 2019). Recently, interpretable networks combined with conventional models show great potential on this task, with examples either incorporate DI models into networks to adaptively learn detailed images (Sun et al., 2021; Lu et al., 2021), or design networks to automatically learn the observation models (Wang et al., 2021b, 2019) and deep priors (Dian et al., 2018; Wang et al., 2021a) in preparation for the subsequent fusion. The deep unrolling methodology is also employed in HS-MS fusion, which effectively links the DL- and VO-based methods by unrolling the iterative optimization procedure into network training steps (Shen et al., 2022; Xie et al., 2022, 2019; Wei et al., 2020; Yang et al., 2022b). Besides the prevalent CNN model, Xiao et al. (2021) introduce a physical-based GAN method by embedding degradation models into the generator, where the output generated by degradation models are input into the discriminator for a further spatial–spectral enhancement. Transformer is also introduced for HS-MS fusion (Hu et al., 2021a), where the structured embedding matrix is sent into a transformer encoder to learn the residual map.

- **Unsupervised methods**

Unsupervised HS-MS fusion methods only requires a pair of HS-MS images as the input of networks and the fused HS can be obtained when the optimization of network is completed. These methods roughly comprise of two categories, i.e., encoding-decoding-based and generation-constraint-based methods. The former class assumes that the target image can be represented by the multiplication of two matrices with each matrix standing for a explicit physical meaning, where AE is usually employed to model the aforementioned procedure. The first work is proposed by Qu et al. (2018), where weights of the decoder are shared by two AEs. Along this line, several successful methods sharing the similar idea are proposed lately (Zheng et al., 2021; Yao et al., 2020; Liu et al., 2022b). The latter aims to directly generate the target image through an elaborately designed generator with an initialized image as the input. In order to obtain a better reconstruction, extra information and constraints are needed to guide the network training. To be more specific, the input image can be the MS image at hand (Fu et al., 2019; Han et al., 2019a; Li et al., 2022), a random tensor (Uezato et al., 2020; Liu et al., 2021b), and a specially learned code (Zhang et al., 2021b, 2020b).

3.1.4. Spatiotemporal fusion

Apart from the trade-off in spatial–spectral resolutions, there also exists a contradiction in spatial–temporal domain, i.e., images with high spatial resolution at the same area captured by current satellite platforms are usually obtained with a long time interval, and vice versa, which greatly hampers the practical applications such as change detection. Therefore, spatiotemporal fusion aims to produce temporally dense products with fine spatial resolution by fusing one or multiple pairs of coarse/fine images (e.g., MODIS-Landsat pairs) and a coarse spatial resolution image at the predicted time. This section introduces some typical methods in terms of their predicted land surface variables, e.g., reflectance, LST, NDVI, etc.

A large majority of DL-based methods are designed for the reflectance images, where CNN prevail among all models. Inspired by the super-resolution problem, Song et al. (2018) propose the pioneering work, where a nonlinear mapping and a super-resolution network are learned to generate the predicted image. However, simply treating spatiotemporal fusion as a super-resolution problem inevitably impairs the performance due to the lack of the exploration in temporal information and hence many methods simultaneously exploiting the information underlying the spatial and temporal domains are proposed (Tan et al., 2018, 2019; Li et al., 2020a). Especially, Liu et al. (2019) exploit temporal dependence and temporal consistent in the training process by incorporating the temporal information into the loss function, and hence obtain remarkable improvement. Compared with CNN, there are a few GAN-based methods that aim to generate outputs by optimizing a min–max problem. Zhang et al. (2021c) proposed a DL-based end-to-end trainable network in solving spatiotemporal fusion problem, where a two-stage framework are designed to gradually recover the predicted image. However, all the discussed methods require at least three images as inputs in the predicted stage, which may not be easily satisfied in practice. Thus, Tan et al. (2022) proposed a conditional GAN-based methods embedded with normalization techniques to eliminate the restriction on the number of input images.

Compared with above models, DL-based methods originally designed for LST or NDVI are relatively scarce. Though some literature adopt reflectance-oriented methods to generate products of other land surface variables and obtain good performance, there still exist differences between these variables. Facing this problem, Yin et al. (2021b) propose a LST-oriented methods by considering the temporal consistency, where two final outputs generated by multiscale CNN are fused together according to a novel weight function. As for NDVI products, Jia et al. (2022) propose a multitask framework with a super-resolution net and a fusion net, where a time-constraint loss function is introduced to alleviate the time consistency assumption.

3.2. Heterogeneous fusion

Different from homogeneous fusion which aims to generate an outcome with high spectral, spatial, or temporal resolution based on pixel-level fusion, heterogeneous fusion mainly refers to the integration in LiDAR-optical, SAR-optical, RS-GBD, etc. Since the imaging mechanisms of these data are totally different, feature-level and decision-level are widely adopted.

3.2.1. LiDAR-optical fusion

LiDAR-optical fusion can be applied to many tasks, e.g., registration, pansharpening, target extraction, estimation of forest biomass (Zhang and Lin, 2017). Since it is hard to give a thorough and detailed introduction concerning all aspects, we focus on one particular domain, i.e., HS-LiDAR data fusion in the application of LULC classification, and give some examples employed in other tasks.

HS data have been widely used in the classification task by virtue of its rich spectral information, but the performance inevitably meets the bottleneck in the situation where spectral information is not sufficient to discriminate the targets (Hong et al., 2020a). Luckily, the LiDAR system is capable of acquiring 3-D spatial geometry, which compensates for the shortage in HS, and hence the joint utilization of HS and LiDAR data in identifying materials becomes a hot spot in recent years. Ghamisi et al. (2017) pioneer the first DL-based HS-LiDAR fusion network, where features of input data are extracted by EPs and then integrated by two fusion strategies for the consequent DL-based classifier. Though great improvement is achieved compared with traditional methods, the way in feature extraction and feature fusion is simple and rough, which limits further improvement to some extent. Inspired by this milestone, many advanced methods have been proposed, aiming at improving the two critical steps. For the feature extraction, a typical example is given by Chen et al. (2017) who utilize a two-branch network to separately extract spectral-spatial-elevation features and then a fully connected layer is used to integrate these heterogeneous features for the final classification. Other particularly designed features extraction networks include a three-branch network (Li et al., 2018), a dual-tunnel network (Xu et al., 2018; Zhao et al., 2020), and a encoder–decoder translation network (Zhang et al., 2020a). For the feature fusion, Feng et al. (2019) incorporate Squeeze-and-Excitation networks into the fusion step to adaptively realize the feature calibration. Other novel fusion strategies are also proposed, such as cross-attention module (Mohla et al., 2020), a reconstruction-based network (Hong et al., 2022), a feature-decision combined fusion network (Hang et al., 2020), and a graph fusion network (Du et al., 2021). Instead of directly utilizing HS-LiDAR data for the classification, Hang et al. (2022) propose a novel strategy to deal with the issue of limited training samples in HS classification. Specifically, paired HS-LiDAR data are first utilized to extract useful features, and then a fine-tuning strategy is designed to transfer these features for HS classification with limited samples.

Researchers in LiDAR-optical fusion also pay attention to target extraction, such as buildings, roads, impervious surfaces, etc. Huang et al. (2019) propose a encoder–decoder network embedded with a gated feature labeling unit to identify the buildings and non-buildings areas. Algorithms in extracting roads and impervious surfaces are also proposed by Parajuli et al. (2018) and Sun et al. (2019), respectively. Very recently, Han et al. (2022) propose the first DL-based multimodal unmixing network, where the height information from LiDAR extracted by the squeeze-and-excitation attention module is used to guide the unmixing process in HS.

3.2.2. SAR-optical fusion

Different from optical images, SAR system is designed to collect backscatter signals of ground objects that can not only reflect the information of RADAR system parameters but also embody the physical and geometric characteristics of the observed scenes (Liu et al., 2021a). Although SAR data can provide complementary knowledge for optical images, it is highly prone to speckle noise that may heavily restrict its practical potential. The joint use of SAR and optical data becomes a feasible solution to realize better understanding and analysis of targets of interest.

According to which level the fusion is carried out, we can divide SAR-optical data fusion into three categories, namely, pixel-level, feature-level, and decision-level. Though there exists a large gap between SAR and optical data in the imaging mechanism, it is feasible to synthetically generate an optical product with abundant textural and structural information with the aid of SAR image through a pixel-level fusion. In that case, registration becomes extremely crucial and many DL-based registration methods between SAR and optical data are proposed, such as the siamese CNN (Zhang et al., 2019c), and the self-learning and transferable network (Wang et al., 2018). After obtaining a pair of co-registered SAR-optical data, many traditional methods originally designed for pansharpening are extended for the SAR-optical pixel-level fusion. Kong et al. (2021) propose a GAN-based network containing a U-shaped generator and a convolutional discriminator, where extensive losses are taken into consideration to fully eliminate the speckle noise and preserve abundant structure information. In addition, optical images are easily subject to atmospheric conditions, where the cloud cover critically impairs the spectral and spatial information. Luckily, SAR is almost insensitive to these factors thanks to its independence from weather conditions. Thus, many pixel-level-based methods are designed to generate a cloud-free optical image from the corresponding cloud-corrupted optical image with the help of an auxiliary SAR data at the same area (Gao et al., 2020; Grohnfeldt et al., 2018). Among them, Meraner et al. (2020) adopt a simple residual structure to directly learn the mapping from the input data pairs to the cloud-free target and demonstrate its superiority even in the situation where scenes are covered by thick clouds. Very recently, Li et al. (2022b) propose the first SAR-optical spatiotemporal fusion method to recover vegetation NDVI in cloudy regions in the aid of transformer.

In addition to pixel-level fusion, high level fusion for applications like LULC classification also catches considerable interest using SAR-optical data. Hu et al. (2017) propose the first DL-based HS-SAR data fusion network, in which a simple yet effective two-branch architecture is used to separately extract heterogeneous features for the final convolutional fusion. Nevertheless, the efficiency of such a straightforward feature extraction remains limited without considering information redundancy. Hence, a novel BN technique constrained by the sparse constraint is devised to reduce the unnecessary features and make the network generalize better (Li et al., 2022a). At the same time, Wang et al. (2022b) propose a cross-attention aided module to realize feature fusion while capturing the long-range dependencies of input data. In addition to the tasks mentioned above, SAR-optical fusion has also been applied to change detection (Li et al., 2021), biomass estimation (Shao et al., 2017), etc.

3.2.3. RS-GBD fusion

GBD contain a wide range of sources from social media, geographic information systems, mobile phones, etc, which greatly contribute to the understanding in our living environment. More specifically, RS exhibit a strong ability in capturing physical attributes of a large-scale earth surface from a global view. On the other hand, the information provided by GBD is highly associated with human behaviors, which gives abundant socioeconomic descriptions as a supplement to RS. Notably there exists a big gap between GBD and RS in the data structures, therefore current popular dual-branch network that is widely used to extract modality-specific features cannot be directly employed to the fusion of GBD and RS data. This section sorts out some successful examples in RS-GBD fusion according to the category of GBD used in the fusion process, such as street view imagery, POI, vehicle trajectory data, etc.

POI refer to the objects that can be abstracted into a point, such as theaters, bus stops, and houses. Different from RS data, each POI generally contains name, coordinate and some other geographic information, which can be easily gleaned by electronic maps, such as OpenStreetMap. Since the attributes of each POI have close correlation with functional facilities, the integration between POI and RS poses a new opportunity to the task in urban functional zone classification. Very recently, Lu et al. (2022) propose a unified DL-based method to jointly exploit characteristic features underlying POI and RS. Concretely, POI are firstly converted into a distance heatmap to meet the input requirement of CNN, and then two modules are used for feature extraction and spatial relation exploration respectively. Other related algorithms with different structures are also proposed, such as a deep multi-scale network (Xu et al., 2020b; Bao et al., 2020) and a bi-branch network (Fan et al., 2021). Besides the aforementioned task in urban functional zone classification, population mapping also gains tremendous help from POI. For example, Cheng et al. (2021) first transform the multimodal data, including POI, road network, and RS images, into a high-dimensional tensor representation as the input of networks, and then a dual-stream model is employed to extract spatial and attribute feature for the population estimation.

In addition to POI, street view imagery is another important data source that can be gathered from social media (e.g., Twitter, Instagram, and Weibo) and street view cars (e.g., Google, Baidu, and Gaode). Different from RS data, it gives fine-grained pictures along the street networks from human's view, and hence provides a diverse and complementary descriptions about our surroundings (Lefèvre et al., 2017). A typical example is given by Srivastava et al. (2019) who utilize the RS and Google street view data to realize urban land use classification. More concretely, a two-branch structured network is used to separately extract features from both modalities which are then stacked into a new feature for the later classification. It is worth mentioning that authors propose a novel solution to deal with a tricky situation where one modality data are missing during the testing phase. Since labeling samples for supervised classifiers is always a costly and time consuming task, Chi et al. (2017) propose a novel system aided by social media photos and deep learning to reduce labeling costs, successfully realizing RS image classification.

Besides, other kinds of GBD also gather great attention in the fusion task. Various citizen-related data, e.g., Taxi trajectory, time-series electricity, and user visit data, are utilized to identify the urban functional areas (Qian et al., 2020; Cao et al., 2020; Yao et al., 2022). Additionally, Liu et al. (2022c) design two AEs to separately extract modality-specific and cross-modal representation from trajectory and RS data, achieving outstanding performance improvement in road extraction. Mantsis et al. (2022) use the snow-related twitters along with Sentinel-1 images to realize snow depth estimation. He et al. (2021a) employ a two-branch network to extract information from RS and Tencent user density to estimate the proportion of mixed land use.

4. List of resources

With a massive number of multimodal RS data available, DL-based technologies have witnessed considerable breakthroughs in data fusion. Numerous DL models and related algorithms using various multimodal data are springing up, which provides endless inspirations for people who take up the research on DL-based multimodal RS data fusion. For the sake of developments and communications on this domain, we collect and summarize some relevant resources, including tutorials for the beginners, available multimodal RS data used in the literature, and open-source codes provided by the authors.

4.1. Tutorials

We further give some materials and references for beginners who are willing to work on DL-related RS tasks, as listed in Table 4. The references in the category of RS can give readers a quick and comprehensive view in the features, principles and applications of different modalities from RS. Materials in DL introduce some widely-used models that constitute the pillars of almost all DL-based algorithms. Following that, we recommend five classic references in RS & AI, which aims to present some successful applications in RS achieved by DL. From the aforementioned tutorials along with their citations, readers can have a basic knowledge of relevant backgrounds in preparation for the further research.

4.2. Available multimodal RS data

To comprehensively evaluate the existing algorithms and select suitable models for practical applications, available multimodal RS datasets are indispensable taches of the whole fusion procedure. Thanks to the Data Fusion Technical Committee (DFTC) of the IEEE Geoscience and Remote Sensing Society, a Data Fusion Contest is held annually since 2006, which provides researchers valuable multimodal RS datasets and promotes the development in data fusion domain. Nowadays, these available datasets have been widely used in the literature for the methodology evaluation. More information can be found in Dalla Mura et al. (2015) and Kahraman and Bacher (2021) which provide a detailed summary of these datasets and their applications. Hence, in this section we collect available datasets in Table 5 except the aforementioned datasets provided by DFTC, contributing to the RS community.

4.3. Open-source codes in DL-based multimodal RS data fusion

For the researchers who already have some background knowledge in this domain and are ready to design their own algorithms, the open-source codes can provide them tremendous help. In that case, we search and summarize available codes from GitHub and authors' homepages in Table 6 for the sake of comparison between different approaches.

5. Problems and prospects

A great deal of progress has been made recently in the DL-based multimodal RS data fusion. However, there still exists some problems remaining to be solved. This section aims to point out current challenges faced by the fast-growing domain and presents prospects for the future directions.

5.1. From well-registered to non-registered

Image registration is a fundamental prerequisite for many RS tasks, such as data fusion and change detection. Since the accuracy of registration between two modalities has a non-negligible influence on the image fusion, aligning to-be-fused data with high precision become an extremely important step before the fusion process, especially for the pixel-level fusion. Since there are many platforms equipped with Pan and MS sensors simultaneously, paired Pan-MS images are easily obtained under the same atmospheric environment and at the same acquisition time, which greatly reduce the registration difficulty. on the contrary, it is rather harder to obtain paired HS-Pan or HS-MS images under the same situation, so data registration becomes a crucial task compared with Pan-MS. However, much attention has been paid to designing advanced fusion algorithms by assuming that the input data are perfectly co-registered, and thus ignoring the importance of such a preprocessing. Only a few DL-based fusion work focus on the multitask by jointly realizing image registration and fusion. Very recently, Zheng et al. (2022) make an attempt to realize the registration and fusion tasks in an end-to-end unsupervised fusion network, where the inputs are a pair of unregistered HS-MS data. In the future, it is advisable to pay more attention to the registration step and incorporate this preprocessing into the fusion process.

5.2. From image-oriented to application-oriented quality assessment

Quality assessment for the output product is an indispensable part of the whole fusion process. The evaluation for high level fusion, i.e., feature-level and decision-level, generally depends on the performance of subsequent applications, such as classification, target detection, and change detection. However, the assessment for pixel-level fusion is usually implemented by calculating related indexes from

Table 4
Some tutorials for beginners.

	Aspects	References	Descriptions
Tutorials	RS	Bioucas-Dias et al. (2013)	Introducing basic concepts and features of HS and its relevant topics
		Rasti et al. (2020)	Providing a review of feature extraction approaches in HS
		Moreira et al. (2013)	Giving principles and theories of SAR and its techniques and applications
	DL	Schmidhuber (2015)	Reviewing deep supervised learning, unsupervised learning, and reinforcement learning
		Liu et al. (2017)	Introducing typical DL architectures and their applications
		Zhang et al. (2018)	Reviewing DL models and their applications in analyzing big data
	RS & AI	Zhang et al. (2019a)	Introducing three main development stages for RS and focusing on DL for RS big data
		Zhang et al. (2016)	Introducing typical DL models and their applications in RS tasks
		Zhu et al. (2017)	Reviewing DL models and related algorithms in RS domains followed by a list of resources
		Hong et al. (2021b)	Giving a survey of nonconvex modeling toward interpretable AI models in HS
		Ma et al. (2019)	Conducting a literature survey by meta-analysis method and introducing relevant applications

Table 5
Non-exhaustive list of multimodal RS datasets.

	Source	Reference	Descriptions	Link
Pansharpening	Ikonos, QuickBird, Gaofen-1, and WorldView-2/3/4	Meng et al. (2021)	2,270 pairs of HR Pan/LR MS images from different kinds of remote sensing satellites	http://www.escience.cn/people/fshao/database.html
	GeoEye-1,WorldView-2/3/3, and SPOT-7,Pléiades-1B	Vivone et al. (2021)	14 pairs of Pan-MS images collected over heterogeneous landscapes by different satellites	https://resources.maxar.com/product-samples/pansharpening-benchmark-dataset
HS pansharpening	PRISMA	None	4 pairs of HR Pan/LR HS images provided by WHISPER	https://openremotesensing.net/hyperspectral-pansharpening-challenge/
Spatiotemporal	Landsat8, MODIS	Li et al. (2020b)	27,27, and 29 pairs of Landsat-MODIS images from 3 different datasets	https://drive.google.com/open?id=1yzw-4TaY6GcLPIRNFBpchETrFKno30he
	Landsat5/7, MODIS	Emelyanova et al. (2013)	14 and 17 pairs of Landsat-MODIS images from 2 dataset	https://data.csiro.au/collection/csiro:5846 and https://data.csiro.au/collection/csiro:5847
LULC classification	Sentinel-2, ITRES CASI-1500	Hong et al. (2021c)	HS-MS scene with 349 × 1905 pixels covering the University of Houston	https://github.com/danfenghong/ISPRS_S2FL
	EnMAP, Sentinel-1	Hong et al. (2021c)	HS-SAR scene with 1723 × 476 pixels covering the Berlin urban and its neighboring area	https://github.com/danfenghong/ISPRS_S2FL
	HySpex, Sentinel-1, and DLR-3 K system	Hong et al. (2021c)	HS-SAR-DSM with 332 × 485 pixels over Augsburg	https://github.com/danfenghong/ISPRS_S2FL
	ITRES CASI-1500, ALTM	Gader et al. (2013)	HS-LiDAR with 325 × 220 pixels over the University of Southern Mississippi Gulf Park Campus	https://github.com/GatorSense/MUUFLGulfport/tree/master/MUUFLGulfportSceneLabels
Objection extraction	USGS, OSM, state, and federal agencies	Huang et al. (2019)	Orthophotos, LiDAR point clouds, and ground-truth building masks	http://dx.doi.org/10.6084/m9.figshare.3504413
	Commission II/4 of the ISPRS	Hosseinpour et al. (2022)	Orthophotos with corresponding DSM and labels	https://www.isprs.org/education/benchmarks/UrbanSemLab/semantic-labeling.aspx
	TLCGIS	Parajuli et al. (2018)	RGB images, LiDAR-derived depth images, and road masks	https://bitbucket.org/biswas/fusion_lidar_images/src/master/

spatial and spectral domains, and it can be divided into two categories, i.e., quality with reference and quality with no reference. For the first class, some widely-used indexes, such as SSIM, SAM, and ERGAS, are calculated between the fused product and the reference image. However, the existing indexes are not sufficient enough to exhibit and compare various methods in a comprehensive and fair way, which inevitably hinder users from selecting appropriate methods for real-world applications. Very recently, Zhu et al. (2022) propose a novel framework for the quality assessment of spatiotemporal products, which not only takes the spatial and spectral errors into consideration but also the characteristics of input data and land surfaces. On the other hand, it is very likely that the reference images are not readily available in practice, so designing an index without the requirement of the reference is urgently needed. Liu et al. (2015) proposed a non-reference index for the pansharpening by using Gaussian scale space which is more consistent with human visual system. Besides, some researchers adopt application-oriented evaluating indicators to judge the performance of pansharpening methods, for example, Qu et al. (2017) evaluate the pansharpening approaches by comparing anomaly detection performance in their pansharpened outputs. In general, it is more desirable to employ application-related indexes to evaluate different algorithms since the purpose of fusion is to combine complementary information for a better decision in a specific application. Therefore, it is a good way for DL-based fusion methods to incorporate the application-related indexes into theirs loss functions to guide the network to learn representative outputs which are more suitable for the subsequent applications.

Table 6
Open-source codes in DL-based multimodal RS data fusion.

	Categories	Name	References	Languages/Frameworks	Links
Pansharpening	Supervised	A-PNN	Scarpa et al. (2018)	Theano	https://github.com/sergiovitale/pansharpening-cnn-python-version
		PanNet	Yang et al. (2017)	Chainer	https://github.com/oyam/PanNet-Landsat
		PNN	Masi et al. (2016)	MATLAB	http://www.grip.unina.it/research/85-image-enhancement/93-pnn.html
		GTP-PNet	Zhang and Ma (2021)	TensorFlow	https://github.com/HaoZhang1018/GTP-PNet
		SDPNet	Xu et al. (2021)	TensorFlow	https://github.com/hanna-xu/SDPNet-for-pansharpening
		TFNet	Liu et al. (2020a)	PyTorch	https://github.com/liouxy/tfnet_pytorch
		DiCNN	He et al. (2019a)	TensorFlow	https://github.com/whyLemon/Pansharpening-via-Detail-Injection-Based-Convolutional-Neural-Networks-
		Fusion-Net	Deng et al. (2020)	TensorFlow	https://github.com/liangjiandeng/FusionNet
		TDNet	Zhang et al. (2022a)	PyTorch	https://github.com/liangjiandeng/TDNet
		VO+Net	Wu et al. (2021a)	MATLAB	https://github.com/liangjiandeng/VOFF
		VP-Net	Tian et al. (2021)	TensorFlow	https://github.com/likun97/VP-Net
		DL-VM	Shen et al. (2019)	MATLAB	https://github.com/WHU-SGG-RS-Pro-Group/DL_VM
		MDSSC-GAN	Gastineau et al. (2022)	TensorFlow	https://github.com/agastineau/MDSSC-GAN_SAM
		PanColorGAN	Ozcelik et al. (2020)	PyTorch	https://github.com/ozcelikfu/PanColorGAN
		PSGAN	Liu et al. (2020b)	PyTorch	https://github.com/zhysora/PSGan-Family
		RED-cGAN	Shao et al. (2019)	TensorFlow	https://github.com/Deep-Imaging-Group/RED-cGAN
		ArbRPN	Chen et al. (2022)	PyTorch	https://github.com/Lihui-Chen/ArbRPN
		PanFormer	Zhou et al. (2022a)	PyTorch	https://github.com/zhysora/PanFormer
	Unsupervised	UCGAN	Zhou et al. (2022b)	PyTorch	https://github.com/zhysora/UCGAN
		Pan-GAN	Ma et al. (2020)	TensorFlow	https://github.com/yuwei998/PanGAN
		PercepPan	Zhou et al. (2020)	PyTorch	https://github.com/wasaCheney/PercepPan
		PGMAN	Zhou et al. (2021b)	PyTorch	https://github.com/zhysora/PGMAN
		ZeRGAN	Diao et al. (2022)	PyTorch	https://github.com/RSMagneto/ZeRGAN
HS pansharpening	Supervised	DIP-HyperKite	Bandara et al. (2022)	PyTorch	https://github.com/wgcban/DIP-HyperKite
		DBDENet	Qu et al. (2022a)	PyTorch	https://github.com/Jiahuiqu/DBDENet/tree/111528a82e579faaf02d4ffd3ea0df0e51de2efb
		MDA-Net	Guan and Lam (2021)	PyTorch	https://github.com/pyguan88/MDA-Net
		MSSL	Qu et al. (2022)	PyTorch	https://github.com/Jiahuiqu/MSSL
		MoG-DCN	Dong et al. (2021d)	PyTorch	https://github.com/chengerr/Model-Guided-Deep-Hyperspectral-Image-Super-resolution
		Pgnet	Li et al. (2022c)	PyTorch	https://github.com/rs-lsl/Pgnet
		HyperTransformer	Bandara and Patel (2022)	PyTorch	https://github.com/wgcban/HyperTransformer
HS-MS	Supervised	SSR-NET	Zhang et al. (2021a)	PyTorch	https://github.com/hw2hwei/SSRNET
		HSRnet	Hu et al. (2021b)	TensorFlow	https://github.com/liangjiandeng/HSRnet
		PZRes-Net	Zhu et al. (2021)	PyTorch	https://github.com/zbzhzhy/PZRes-Net
		Two-CNN-Fu	Yang et al. (2018)	Caffe	https://github.com/polwork/Hyperspectral-and-Multispectral-fusion-via-Two-branch-CNN
		ADMM-HFNet	Shen et al. (2022)	TensorFlow	https://github.com/liuofficial/ADMM-HFNet
		MHF-Net	Xie et al. (2022)	TensorFlow	https://github.com/XieQi2015/MHF-net
		CNN-Fus	Dian et al. (2021a)	MATLAB	https://github.com/renweidian/CNN-FUS
		DHIF-Net	Huang et al. (2022)	PyTorch	https://github.com/TaoHuang95/DHIF-Net
		DHSIS	Dian et al. (2018)	MATLAB+Keras	https://github.com/renweidian/DHSIS
		EDBIN	Wang et al. (2021b)	TensorFlow	https://github.com/wwhappylife/Deep-Blind-Hyperspectral-Image-Fusion
		SpfNet	Liu et al. (2022a)	TensorFlow	https://github.com/liuofficial/SpfNet
		TONWMD	Shen et al. (2020)	TensorFlow	https://github.com/liuofficial/TONWMD
		TSFN	Wang et al. (2021a)	MATLAB+PyTorch	https://github.com/xiuheng-wang/Sylvester_TSFN_MDC_HSI_superresolution

(continued on next page)

Table 6 (*continued*).

	Categories	Name	References	Languages/Frameworks	Links
		Fusformer	Hu et al. (2021a)	PyTorch	https://github.com/J-FHu/Fusformer
	Unsupervised	CUCaNet	Yao et al. (2020)	PyTorch	https://github.com/danfenghong/ECCV2020_CUCaNet
		HyCoNet	Zheng et al. (2021)	PyTorch	https://github.com/saber-zero/HyperFusion
		MIAE	Liu et al. (2022b)	PyTorch	https://github.com/liuofficial/MIAE/tree/c880d1df15f022f78fc8305e436aa0bfae378135
		NonRegSRNet	Zheng et al. (2022)	PyTorch	https://github.com/saber-zero/NonRegSRNet
		u^2-MDN	Qu et al. (2022b)	TensorFlow	https://github.com/yingutk/u2MDN
		uSDN	Qu et al. (2018)	TensorFlow	https://github.com/aicip/uSDN
		HSI-CSR	Fu et al. (2019)	Caffe	https://github.com/ColinTaoZhang/HSI-SR
		DBSR	Zhang et al. (2021b)	PyTorch	https://github.com/JiangtaoNie/DBSR
		GDD	Uezato et al. (2020)	PyTorch	https://github.com/tuezato/guided-deep-decoder
		UAL	Zhang et al. (2020b)	PyTorch	https://github.com/JiangtaoNie/UAL-CVPR2020
		UDALN	Li et al. (2022)	PyTorch	https://github.com/JiaxinLiCAS/UDALN_GRSL
		RAFnet	Lu et al. (2020)	TensorFlow	https://github.com/RuiyingLu/RAFnet
		Rec_HSISR_PixAwaRefin	Wei et al. (2022)	PyTorch	https://github.com/JiangtaoNie/Rec_HSISR_PixAwaRefin
Spatiotemporal	CNN	DCSTFN	Tan et al. (2018)	TensorFlow	https://github.com/theonegis/rs-data-fusion
		EDCSTFN	Tan et al. (2019)	PyTorch	https://github.com/theonegis/edcstfn
	GAN	GANSTFM	Tan et al. (2022)	PyTorch	https://github.com/theonegis/ganstfm
LiDAR-optical	LULC classification	AM^3Net	Wang et al. (2022c)	PyTorch	https://github.com/Cimy-wang/AM3Net_Multimodal_Data_Fusion
		CCR-Net	Wu et al. (2022)	TensorFlow	https://github.com/danfenghong/IEEE_TGRS_CCR-Net
		EndNet	Hong et al. (2022)	TensorFlow	https://github.com/danfenghong/IEEE_GRSL_EndNet
		FusAtNet	Mohla et al. (2020)	Keras	https://github.com/ShivamP1993/FusAtNet
		HRWN	Zhao et al. (2020)	Keras	https://github.com/xudongzhao461/HRWN
		IP-CNN	Zhang et al. (2022b)	Keras	https://github.com/HelloPiPi/IP-CNN-code
		MAHiDFNet	Wang et al. (2022a)	Keras	https://github.com/SYFYN0317/-MAHiDFNet
		MDL-RS	Hong et al. (2021a)	TensorFlow	https://github.com/danfenghong/IEEE_TGRS_MDL-RS
		RNPRF-RNDFF-RNPMF	Ge et al. (2021)	Keras	https://github.com/gechiru/RNPRF-RNDFF-RNPMF
		S^2ENet	Fang et al. (2022)	PyTorch	https://github.com/likyoo/Multimodal-Remote-Sensing-Toolkit
		two-branch CNN	Xu et al. (2018)	Keras	https://github.com/Hsuxu/Two-branch-CNN-Multisource-RS-classification
	Target extraction	CMGFNet	Hosseinpour et al. (2022)	PyTorch	https://github.com/hamidreza2015/CMGFNet-Building_Extraction
		GRRNet	Huang et al. (2019)	Caffe	https://github.com/CHUANQIFENG/GRRNet
	Unmixing	MUNet	Han et al. (2022)	PyTorch	https://github.com/hanzhu97702/IEEE_TGRS_MUNet
RS-GBD	POI data	UnifiedDL-UFZ	Lu et al. (2022)	PyTorch	https://github.com/GeoX-Lab/UnifiedDL-UFZ-extraction
	User density data	CF-CNN	He et al. (2021a)	Keras	https://github.com/SysuHe/MultiSourceData_CFCNN

5.3. From two-modality to multi-modality

With the quick development of multiple sensors on airborne and spaceborne platforms, the availability of modalities becomes more diverse. Currently, most of DL-based fusion algorithms are designed for only two-modality, limiting the application ability of multi-modality. As a result, *how to effectively utilize more modality data and fully exhibit their potentials as well as further make the performance bottleneck is a remaining challenge in the multimodal data fusion task.* More importantly, with more and more modality data easily accessible, future researches could consider developing a unified DL-based framework which could deal with arbitrary number of modalities as the inputs.

5.4. From multimodal to crossmodal learning

Though multimodal data with diverse features contribute to our understanding in the world, it is more likely that some modality data are absent in practical scenarios. For example, SAR and MS data are available on a global scale. In contrast, HS data are more hard to collect on account of the limitation of sensors, which may lead to the shortage in some areas. Hence, *how to transfer the information hidden in the area with multimodal data into the scenario where one modality is missing* is a typical issue that crossmodal learning aims to deal with. A representative DL-based method tackling this practical problem is proposed by Hong et al. (2020b), where a limited number of HS-MS or HS-SAR pairs are used in the training phase to realize a large-scale classification task in an area only covered by one modality data, i.e., MS or SAR. In the future, it is believed that this critical domain will catch more attention in the RS fusion community under the influence of RS big data and DL.

5.5. From single-platform to cross-platform

Current observation platforms have branched out into ground-based, airborne, and spaceborne domains, which provides users with endless cross-platform data. Especially, unmanned aerial vehicles have attracted growing attention in the RS community on account of their high mobility, showing great potential in many tasks (Wu et al., 2021b). Though images from the cross-platform enables us to observe the Earth environment from a new perspective, these data not only exhibit totally diverse features in the spatial scale but also show a difference in the acquisition time, which becomes a non-negligible obstacle for the fusion procedure. Hence, *How to break through the barrier exists among different platforms and achieve an effective information interaction is a direction that needs to be studied in the future.*

5.6. From black-box to interpretable DL

Though DL has witnessed numerous breakthroughs in recent years, it is often accused of an inexplicable black-box learning procedure. Unlike the traditional methods which have clear physical and mathematical meanings, DL-based methods extract high-level features which are hard to explain. As discussed in Section 3.1.1, many model-driven DL-based methods are successively proposed to design a totally interpretable networks with each module presenting a specific operation. The combination between model-driven and data-driven methods poses a new view to understand the workflow in the black-box network and also points out a solution to make the black-box transparent. However, this solution is limited to spatiospectral fusion domain and hard to apply to feature-level and decision-level fusion. Therefore, high-level fusion still stays at the stage where researchers pay much attention to the feature extraction and feature fusion instead of fully understanding what the network really learns. However, understanding the features learned by each hidden layer contributes to designing more effective network structures in mining discriminative features and hence promoting the performance in the high-level tasks.

6. Conclusion

The ever-growing number of multimodal RS data poses not only a challenge but also an opportunity to the EO tasks. By jointly utilizing their complementary features, great breakthroughs have been witnessed over the recent years. Particularly, artificial intelligence-related technologies has demonstrated their advantages over traditional methods on account of their superiority in the feature extraction. Driven by aforementioned RS big data and cutting-edge tools, DL-based multimodal RS data fusion becomes a significant topic in the RS community. Therefore, this review gives a comprehensive introduction on this fast-growing domain, including a literature analysis, a systematic summary in several prevalent sub-fields in RS fusion, a list of available resources, and the prospects for the future development. Specifically, we focus on the second part, i.e., DL-based methods in different fusion subdomains, and give a detailed study in terms of used models, tasks, and data types. Finally, we are encouraged to find that DL has been applied to every corner of multimodal RS data fusion and obtained tremendous and promising achievements in recent years, which provides researchers more confidences to conduct in-depth study in the future.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Key Research and Development Program of China [2021YFB3900502]; the National Natural Science Foundation of China [42030111]; MIAI@Grenoble Alpes [ANR-19-P3IA-0003]; and the AXA Research Fund.

References

Azarang, A., Ghassemian, H., 2017. A new pansharpening method using multi resolution analysis framework and deep neural networks. In: 2017 3rd International Conference on Pattern Recognition and Image Analysis. IPRIA, IEEE, pp. 1–6. http://dx.doi.org/10.1109/PRIA.2017.7983017.

Bandara, W.G.C., Patel, V.M., 2022. HyperTransformer: A textural and spectral feature fusion transformer for pansharpening. arXiv preprint arXiv:2203.02503.

Bandara, W.G.C., Valanarasu, J.M.J., Patel, V.M., 2022. Hyperspectral pansharpening based on improved deep image prior and residual reconstruction. IEEE Trans. Geosci. Remote Sens. 60, 1–16. http://dx.doi.org/10.1109/TGRS.2021.3139292.

Bao, H., Ming, D., Guo, Y., Zhang, K., Zhou, K., Du, S., 2020. DFCNN-based semantic recognition of urban functional zones by integrating remote sensing data and POI data. Remote Sens. 12 (7), 1088. http://dx.doi.org/10.3390/rs12071088.

Belgiu, M., Stein, A., 2019. Spatiotemporal image fusion in remote sensing. Remote Sens. 11 (7), 818. http://dx.doi.org/10.3390/rs11070818.

Bioucas-Dias, J.M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N., Chanussot, J., 2013. Hyperspectral remote sensing data analysis and future challenges. IEEE Geosci. Remote Sens. Mag. 1 (2), 6–36. http://dx.doi.org/10.1109/MGRS.2013.2244672.

Cao, X., Fu, X., Hong, D., Xu, Z., Meng, D., 2021. PanCSC-Net: A model-driven deep unfolding method for pansharpening. IEEE Trans. Geosci. Remote Sens. 60, 1–13. http://dx.doi.org/10.1109/TGRS.2021.3115501.

Cao, R., Tu, W., Yang, C., Li, Q., Liu, J., Zhu, J., Zhang, Q., Li, Q., Qiu, G., 2020. Deep learning-based remote and social sensing data fusion for urban region function recognition. ISPRS J. Photogramm. Remote Sens. 163, 82–97. http://dx.doi.org/10.1016/j.isprsjprs.2020.02.014.

Chen, C., 2006. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. J. Am. Soc. Inf. Sci. Technol. 57 (3), 359–377. http://dx.doi.org/10.1002/asi.20317.

Chen, B., Huang, B., Xu, B., 2015. Comparison of spatiotemporal fusion models: A review. Remote Sens. 7 (2), 1798–1835. http://dx.doi.org/10.3390/rs70201798.

Chen, L., Lai, Z., Vivone, G., Jeon, G., Chanussot, J., Yang, X., 2022. ArbRPN: A bidirectional recurrent pansharpening network for multispectral images with arbitrary numbers of bands. IEEE Trans. Geosci. Remote Sens. 60, 1–18. http://dx.doi.org/10.1109/TGRS.2021.3131228.

Chen, Y., Li, C., Ghamisi, P., Jia, X., Gu, Y., 2017. Deep fusion of remote sensing data for accurate classification. IEEE Geosci. Remote Sens. Lett. 14 (8), 1253–1257. http://dx.doi.org/10.1109/LGRS.2017.2704625.

Chen, S., Qi, H., Nan, K., 2021. Pansharpening via super-resolution iterative residual network with a cross-scale learning strategy. IEEE Trans. Geosci. Remote Sens. 60, 1–16. http://dx.doi.org/10.1109/TGRS.2021.3138096.

Cheng, L., Wang, L., Feng, R., Yan, J., 2021. Remote sensing and social sensing data fusion for fine-resolution population mapping with a multimodel neural network. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 5973–5987. http://dx.doi.org/10.1109/JSTARS.2021.3086139.

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. http://dx.doi.org/10.1109/JPROC.2017.2730585.

Dalla Mura, M., Prasad, S., Pacifici, F., Gamba, P., Chanussot, J., Benediktsson, J.A., 2015. Challenges and opportunities of multimodality and data fusion in remote sensing. Proc. IEEE 103 (9), 1585–1601. http://dx.doi.org/10.1109/JPROC.2015.2462751.

Deng, L.-J., Vivone, G., Jin, C., Chanussot, J., 2020. Detail injection-based deep convolutional neural networks for pansharpening. IEEE Trans. Geosci. Remote Sens. 59 (8), 6995–7010. http://dx.doi.org/10.1109/TGRS.2020.3031366.

Dian, R., Li, S., Guo, A., Fang, L., 2018. Deep hyperspectral image sharpening. IEEE Trans. Neural Netw. Learn. Syst. 29 (11), 5345–5355. http://dx.doi.org/10.1109/TNNLS.2018.2798162.

Dian, R., Li, S., Kang, X., 2021a. Regularizing hyperspectral and multispectral image fusion by CNN denoiser. IEEE Trans. Neural Netw. Learn. Syst. 32 (3), 1124–1135. http://dx.doi.org/10.1109/TNNLS.2020.2980398.

Dian, R., Li, S., Sun, B., Guo, A., 2021b. Recent advances and new guidelines on hyperspectral and multispectral image fusion. Inf. Fusion 69, 40–51. http://dx.doi.org/10.1016/j.inffus.2020.11.001.

Diao, W., Zhang, F., Sun, J., Xing, Y., Zhang, K., Bruzzone, L., 2022. ZeRGAN: Zero-reference GAN for fusion of multispectral and panchromatic images. IEEE Trans. Neural Netw. Learn. Syst. http://dx.doi.org/10.1109/TNNLS.2021.3137373.

Dong, W., Hou, S., Xiao, S., Qu, J., Du, Q., Li, Y., 2021a. Generative dual-adversarial network with spectral fidelity and spatial enhancement for hyperspectral pansharpening. IEEE Trans. Neural Netw. Learn. Syst. http://dx.doi.org/10.1109/TNNLS.2021.3084745.

Dong, W., Yang, Y., Qu, J., Xie, W., Li, Y., 2021b. Fusion of hyperspectral and panchromatic images using generative adversarial network and image segmentation. IEEE Trans. Geosci. Remote Sens. 60, 1–13. http://dx.doi.org/10.1109/TGRS.2021.3078711.

Dong, W., Zhang, T., Qu, J., Xiao, S., Liang, J., Li, Y., 2021c. Laplacian pyramid dense network for hyperspectral pansharpening. IEEE Trans. Geosci. Remote Sens. 60, 1–13. http://dx.doi.org/10.1109/TGRS.2021.3076768.

Dong, W., Zhou, C., Wu, F., Wu, J., Shi, G., Li, X., 2021d. Model-guided deep hyperspectral image super-resolution. IEEE Trans. Image Process. 30, 5754–5768. http://dx.doi.org/10.1109/TIP.2021.3078058.

Du, X., Zheng, X., Lu, X., Doudkin, A.A., 2021. Multisource remote sensing data classification with graph fusion network. IEEE Trans. Geosci. Remote Sens. 59 (12), 10062–10072. http://dx.doi.org/10.1109/TGRS.2020.3047130.

Emelyanova, I.V., McVicar, T.R., Van Niel, T.G., Li, L.T., Van Dijk, A.I., 2013. Assessing the accuracy of blending Landsat–MODIS surface reflectances in two landscapes with contrasting spatial and temporal dynamics: A framework for algorithm selection. Remote Sens. Environ. 133, 193–209. http://dx.doi.org/10.1016/j.rse.2013.02.007.

Fan, R., Feng, R., Han, W., Wang, L., 2021. Urban functional zone mapping with a bibranch neural network via fusing remote sensing and social sensing data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 11737–11749. http://dx.doi.org/10.1109/JSTARS.2021.3127246.

Fang, S., Li, K., Li, Z., 2022. S2ENet: Spatial-spectral cross-modal enhancement network for classification of hyperspectral and LiDAR data. IEEE Geosci. Remote Sens. Lett. 19, 1–5. http://dx.doi.org/10.1109/LGRS.2021.3121028.

Feng, Q., Zhu, D., Yang, J., Li, B., 2019. Multisource hyperspectral and lidar data fusion for urban land-use mapping based on a modified two-branch convolutional neural network. ISPRS Int. J. Geo-Inf. 8 (1), 28. http://dx.doi.org/10.3390/ijgi8010028.

Fu, X., Wang, W., Huang, Y., Ding, X., Paisley, J., 2020. Deep multiscale detail networks for multiband spectral image sharpening. IEEE Trans. Neural Netw. Learn. Syst. 32 (5), 2090–2104. http://dx.doi.org/10.1109/TNNLS.2020.2996498.

Fu, Y., Zhang, T., Zheng, Y., Zhang, D., Huang, H., 2019. Hyperspectral image super-resolution with optimized RGB guidance. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11653–11662. http://dx.doi.org/10.1109/CVPR.2019.01193.

Gader, P., Zare, A., Close, R., Aitken, J., Tuell, G., 2013. MUUFL Gulfport Hyperspectral and LiDAR Airborne Data Set. Technical Report Rep. REP-2013-570, University of Florida, Gainesville, FL.

Gao, J., Yuan, Q., Li, J., Zhang, H., Su, X., 2020. Cloud removal with fusion of high resolution optical and SAR images using generative adversarial networks. Remote Sens. 12 (1), 191. http://dx.doi.org/10.3390/rs12010191.

Gastineau, A., Aujol, J.-F., Berthoumieu, Y., Germain, C., 2022. Generative adversarial network for pansharpening with spectral and spatial discriminators. IEEE Trans. Geosci. Remote Sens. 60, 1–11. http://dx.doi.org/10.1109/TGRS.2021.3060958.

Ge, C., Du, Q., Sun, W., Wang, K., Li, J., Li, Y., 2021. Deep residual network-based fusion framework for hyperspectral and LiDAR data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 2458–2472. http://dx.doi.org/10.1109/JSTARS.2021.3054392.

Ghamisi, P., Höfle, B., Zhu, X.X., 2017. Hyperspectral and LiDAR data fusion using extinction profiles and deep convolutional neural network. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10 (6), 3011–3024. http://dx.doi.org/10.1109/JSTARS.2016.2634863.

Ghamisi, P., Rasti, B., Yokoya, N., Wang, Q., Hofle, B., Bruzzone, L., Bovolo, F., Chi, M., Anders, K., Gloaguen, R., et al., 2019. Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art. IEEE Geosci. Remote Sens. Mag. 7 (1), 6–39. http://dx.doi.org/10.1109/MGRS.2018.2890023.

Ghassemian, H., 2016. A review of remote sensing image fusion methods. Inf. Fusion 32, 75–89. http://dx.doi.org/10.1016/j.inffus.2016.03.003.

Gómez-Chova, L., Tuia, D., Moser, G., Camps-Valls, G., 2015. Multimodal classification of remote sensing images: A review and future directions. Proc. IEEE 103 (9), 1560–1584. http://dx.doi.org/10.1109/JPROC.2015.2449668.

Grohnfeldt, C., Schmitt, M., Zhu, X., 2018. A conditional generative adversarial network to fuse SAR and multispectral optical data for cloud removal from sentinel-2 images. In: IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 1726–1729. http://dx.doi.org/10.1109/IGARSS.2018.8519215.

Guan, P., Lam, E.Y., 2021. Multistage dual-attention guided fusion network for hyperspectral pansharpening. IEEE Trans. Geosci. Remote Sens. 60, 1–14. http://dx.doi.org/10.1109/TGRS.2021.3114552.

Han, X.-H., Chen, Y.-W., 2019. Deep residual network of spectral and spatial fusion for hyperspectral image super-resolution. In: 2019 IEEE Fifth International Conference on Multimedia Big Data. BigMM, IEEE, pp. 266–270. http://dx.doi.org/10.1109/BigMM.2019.00-13.

Han, Z., Hong, D., Gao, L., Yao, J., Zhang, B., Chanussot, J., 2022. Multimodal hyperspectral unmixing: Insights from attention networks. IEEE Trans. Geosci. Remote Sens. 60, 1–13. http://dx.doi.org/10.1109/TGRS.2022.3155794.

Han, X.-H., Shi, B., Zheng, Y., 2018. SSF-CNN: Spatial and spectral fusion with CNN for hyperspectral image super-resolution. In: 2018 25th IEEE International Conference on Image Processing. ICIP, IEEE, pp. 2506–2510. http://dx.doi.org/10.1109/ICIP.2018.8451142.

Han, X., Yu, J., Luo, J., Sun, W., 2019a. Hyperspectral and multispectral image fusion using cluster-based multi-branch BP neural networks. Remote Sens. 11 (10), 1173. http://dx.doi.org/10.3390/rs11101173.

Han, X.-H., Zheng, Y., Chen, Y.-W., 2019. Multi-level and multi-scale spatial and spectral fusion CNN for hyperspectral image super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. pp. 4330–4339. http://dx.doi.org/10.1109/ICCVW.2019.00533.

Hang, R., Li, Z., Ghamisi, P., Hong, D., Xia, G., Liu, Q., 2020. Classification of hyperspectral and LiDAR data using coupled CNNs. IEEE Trans. Geosci. Remote Sens. 58 (7), 4939–4950. http://dx.doi.org/10.1109/TGRS.2020.2969024.

Hang, R., Qian, X., Liu, Q., 2022. Cross-modality contrastive learning for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. http://dx.doi.org/10.1109/TGRS.2022.3188529.

He, J., Li, X., Liu, P., Wu, X., Zhang, J., Zhang, D., Liu, X., Yao, Y., 2021a. Accurate estimation of the proportion of mixed land use at the street-block level by integrating high spatial resolution images and geospatial big data. IEEE Trans. Geosci. Remote Sens. 59 (8), 6357–6370. http://dx.doi.org/10.1109/TGRS.2020.3028622.

He, L., Rao, Y., Li, J., Chanussot, J., Plaza, A., Zhu, J., Li, B., 2019a. Pansharpening via detail injection based convolutional neural networks. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12 (4), 1188–1204. http://dx.doi.org/10.1109/JSTARS.2019.2898574.

He, L., Zhu, J., Li, J., Meng, D., Chanussot, J., Plaza, A., 2020. Spectral-fidelity convolutional neural networks for hyperspectral pansharpening. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 5898–5914. http://dx.doi.org/10.1109/JSTARS.2020.3025040.

He, L., Zhu, J., Li, J., Plaza, A., Chanussot, J., Li, B., 2019b. HyperPNN: Hyperspectral pansharpening via spectrally predictive convolutional neural networks. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12 (8), 3092–3100. http://dx.doi.org/10.1109/JSTARS.2019.2917584.

He, L., Zhu, J., Li, J., Plaza, A., Chanussot, J., Yu, Z., 2021b. CNN-based hyperspectral pansharpening with arbitrary resolution. IEEE Trans. Geosci. Remote Sens. 60, 1–21. http://dx.doi.org/10.1109/TGRS.2021.3132997.

Hong, D., Gao, L., Hang, R., Zhang, B., Chanussot, J., 2022. Deep encoder-decoder networks for classification of hyperspectral and LiDAR data. IEEE Geosci. Remote Sens. Lett. 19, 1–5. http://dx.doi.org/10.1109/LGRS.2020.3017414.

Hong, D., Gao, L., Yao, J., Zhang, B., Plaza, A., Chanussot, J., 2020a. Graph convolutional networks for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 59 (7), 5966–5978. http://dx.doi.org/10.1109/TGRS.2020.3015157.

Hong, D., Gao, L., Yokoya, N., Yao, J., Chanussot, J., Du, Q., Zhang, B., 2021a. More diverse means better: Multimodal deep learning meets remote-sensing imagery classification. IEEE Trans. Geosci. Remote Sens. 59 (5), 4340–4354. http://dx.doi.org/10.1109/TGRS.2020.3016820.

Hong, D., He, W., Yokoya, N., Yao, J., Gao, L., Zhang, L., Chanussot, J., Zhu, X., 2021b. Interpretable hyperspectral artificial intelligence: When nonconvex modeling meets hyperspectral remote sensing. IEEE Geosci. Remote Sens. Mag. 9 (2), 52–87. http://dx.doi.org/10.1109/MGRS.2021.3064051.

Hong, D., Hu, J., Yao, J., Chanussot, J., Zhu, X.X., 2021c. Multimodal remote sensing benchmark datasets for land cover classification with a shared and specific feature learning model. ISPRS J. Photogramm. Remote Sens. 178, 68–80. http://dx.doi.org/10.1016/j.isprsjprs.2021.05.011.

Hong, D., Yokoya, N., Xia, G.-S., Chanussot, J., Zhu, X.X., 2020b. X-ModalNet: A semi-supervised deep cross-modal network for classification of remote sensing data. ISPRS J. Photogramm. Remote Sens. 167, 12–23. http://dx.doi.org/10.1016/j.isprsjprs.2020.06.014.

Hosseinpour, H., Samadzadegan, F., Javan, F.D., 2022. CMGFNet: A deep cross-modal gated fusion network for building extraction from very high-resolution remote sensing images. ISPRS J. Photogramm. Remote Sens. 184, 96–115. http://dx.doi.org/10.1016/j.isprsjprs.2021.12.007.

Hu, J.-F., Huang, T.-Z., Deng, L.-J., 2021a. Fusformer: A transformer-based fusion approach for hyperspectral image super-resolution. arXiv preprint arXiv:2109.02079.

Hu, J.-F., Huang, T.-Z., Deng, L.-J., Jiang, T.-X., Vivone, G., Chanussot, J., 2021b. Hyperspectral image super-resolution via deep spatiospectral attention convolutional neural networks. IEEE Trans. Neural Netw. Learn. Syst. http://dx.doi.org/10.1109/TNNLS.2021.3084682.

Hu, J., Mou, L., Schmitt, A., Zhu, X.X., 2017. FusioNet: A two-stream convolutional neural network for urban scene classification using polsar and hyperspectral data. In: 2017 Joint Urban Remote Sensing Event. JURSE, IEEE, pp. 1–4. http://dx.doi.org/10.1109/JURSE.2017.7924565.

Huang, T., Dong, W., Wu, J., Li, L., Li, X., Shi, G., 2022. Deep hyperspectral image fusion network with iterative spatio-spectral regularization. IEEE Trans. Comput. Imaging 8, 201–214. http://dx.doi.org/10.1109/TCI.2022.3152700.

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. http://dx.doi.org/10.1109/LGRS.2014.2376034.

Huang, J., Zhang, X., Xin, Q., Sun, Y., Zhang, P., 2019. Automatic building extraction from high-resolution aerial images and LiDAR data using gated residual refinement network. ISPRS J. Photogramm. Remote Sens. 151, 91–105. http://dx.doi.org/10.1016/j.isprsjprs.2019.02.019.

Jia, D., Cheng, C., Shen, S., Ning, L., 2022. Multi-task deep learning framework for spatiotemporal fusion of NDVI. IEEE Trans. Geosci. Remote Sens. 60, 1–13. http://dx.doi.org/10.1109/TGRS.2021.3140144.

Kahraman, S., Bacher, R., 2021. A comprehensive review of hyperspectral data fusion with lidar and sar data. Annu. Rev. Control 51, 236–253. http://dx.doi.org/10.1016/j.arcontrol.2021.03.003.

Kong, Y., Hong, F., Leung, H., Peng, X., 2021. A fusion method of optical image and SAR image based on dense-UGAN and Gram–Schmidt transformation. Remote Sens. 13 (21), 4274. http://dx.doi.org/10.3390/rs13214274.

Kulkarni, S.C., Rege, P.P., 2020. Pixel level fusion techniques for SAR and optical images: A review. Inf. Fusion 59, 13–29. http://dx.doi.org/10.1016/j.inffus.2020.01.003.

Kuras, A., Brell, M., Rizzi, J., Burud, I., 2021. Hyperspectral and lidar data applied to the urban land cover machine learning and neural-network-based classification: A review. Remote Sens. 13 (17), 3393. http://dx.doi.org/10.3390/rs13173393.

Lahat, D., Adali, T., Jutten, C., 2015. Multimodal data fusion: an overview of methods, challenges, and prospects. Proc. IEEE 103 (9), 1449–1477. http://dx.doi.org/10.1109/JPROC.2015.2460697.

Lefèvre, S., Tuia, D., Wegner, J.D., Produit, T., Nassar, A.S., 2017. Toward seamless multiview scene analysis from satellite to street level. Proc. IEEE 105 (10), 1884–1899. http://dx.doi.org/10.1109/JPROC.2017.2684300.

Lei, D., Chen, H., Zhang, L., Li, W., 2021. NLRNet: An efficient nonlocal attention ResNet for pansharpening. IEEE Trans. Geosci. Remote Sens. 60, 1–13. http://dx.doi.org/10.1109/TGRS.2021.3067097.

Li, X., Du, Z., Huang, Y., Tan, Z., 2021. A deep translation (GAN) based change detection network for optical and SAR remote sensing images. ISPRS J. Photogramm. Remote Sens. 179, 14–34. http://dx.doi.org/10.1016/j.isprsjprs.2021.07.007.

Li, W., Gao, Y., Zhang, M., Tao, R., Du, Q., 2022a. Asymmetric feature fusion network for hyperspectral and SAR image classification. IEEE Trans. Neural Netw. Learn. Syst. http://dx.doi.org/10.1109/TNNLS.2022.3149394.

Li, H., Ghamisi, P., Soergel, U., Zhu, X.X., 2018. Hyperspectral and LiDAR fusion using deep three-stream convolutional neural networks. Remote Sens. 10 (10), 1649. http://dx.doi.org/10.3390/rs10101649.

Li, S., Kang, X., Fang, L., Hu, J., Yin, H., 2017. Pixel-level image fusion: A survey of the state of the art. Inf. Fusion 33, 100–112. http://dx.doi.org/10.1016/j.inffus.2016.05.004.

Li, Y., Li, J., He, L., Chen, J., Plaza, A., 2020a. A new sensor bias-driven spatio-temporal fusion model based on convolutional neural networks. Sci. China Inf. Sci. 63 (4), 1–16. http://dx.doi.org/10.1007/s11432-019-2805-y.

Li, J., Li, Y., He, L., Chen, J., Plaza, A., 2020b. Spatio-temporal fusion for remote sensing data: An overview and new benchmark. Sci. China Inf. Sci. 63 (4), 1–17. http://dx.doi.org/10.1007/s11432-019-2785-y.

Li, J., Li, C., Xu, W., Feng, H., Zhao, F., Long, H., Meng, Y., Chen, W., Yang, H., Yang, G., 2022b. Fusion of optical and SAR images based on deep learning to reconstruct vegetation NDVI time series in cloud-prone regions. Int. J. Appl. Earth Obs. Geoinf. 112, 102818. http://dx.doi.org/10.1016/j.jag.2022.102818.

Li, J., Liu, Z., Lei, X., Wang, L., 2021b. Distributed fusion of heterogeneous remote sensing and social media data: A review and new developments. Proc. IEEE 109 (8), 1350–1363. http://dx.doi.org/10.1109/JPROC.2021.3079176.

Li, S., Tian, Y., Xia, H., Liu, Q., 2022c. Unmixing based PAN guided fusion network for hyperspectral imagery. IEEE Trans. Geosci. Remote Sens. 60, 1–17. http://dx.doi.org/10.1109/TGRS.2022.3141765.

Li, J., Zheng, K., Yao, J., Gao, L., Hong, D., 2022. Deep unsupervised blind hyperspectral and multispectral data fusion. IEEE Geosci. Remote Sens. Lett. 19, 1–5. http://dx.doi.org/10.1109/LGRS.2022.3151779.

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. http://dx.doi.org/10.1016/j.inffus.2017.10.007.

Liu, X., Deng, C., Chanussot, J., Hong, D., Zhao, B., 2019. StfNet: A two-stream convolutional neural network for spatiotemporal image fusion. IEEE Trans. Geosci. Remote Sens. 57 (9), 6552–6564. http://dx.doi.org/10.1109/TGRS.2019.2907310.

Liu, X., Hong, D., Chanussot, J., Zhao, B., Ghamisi, P., 2021a. Modality translation in remote sensing time series. IEEE Trans. Geosci. Remote Sens. 60, 1–14. http://dx.doi.org/10.1109/TGRS.2021.3079294.

Liu, J., Huang, J., Liu, S., Li, H., Zhou, Q., Liu, J., 2015. Human visual system consistent quality assessment for remote sensing image fusion. ISPRS J. Photogramm. Remote Sens. 105, 79–90. http://dx.doi.org/10.1016/j.isprsjprs.2014.12.018.

Liu, X., Liu, Q., Wang, Y., 2020a. Remote sensing image fusion based on two-stream fusion network. Inf. Fusion 55, 1–15. http://dx.doi.org/10.1016/j.inffus.2019.07.010.

Liu, S., Miao, S., Su, J., Li, B., Hu, W., Zhang, Y.-D., 2021b. UMAG-Net: A new unsupervised multiattention-guided network for hyperspectral and multispectral image fusion. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 7373–7385. http://dx.doi.org/10.1109/JSTARS.2021.3097178.

Liu, J., Shen, D., Wu, Z., Xiao, L., Sun, J., Yan, H., 2022a. Patch-aware deep hyperspectral and multispectral image fusion by unfolding subspace-based optimization model. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15, 1024–1038. http://dx.doi.org/10.1109/JSTARS.2022.3140211.

Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., Alsaadi, F.E., 2017. A survey of deep neural network architectures and their applications. Neurocomputing 234, 11–26. http://dx.doi.org/10.1016/j.neucom.2016.12.038.

Liu, J., Wu, Z., Xiao, L., Wu, X.-J., 2022b. Model inspired autoencoder for unsupervised hyperspectral image super-resolution. IEEE Trans. Geosci. Remote Sens. 60, 1–12. http://dx.doi.org/10.1109/TGRS.2022.3143156.

Liu, L., Yang, Z., Li, G., Wang, K., Chen, T., Lin, L., 2022c. Aerial images meet crowdsourced trajectories: a new approach to robust road extraction. IEEE Trans. Neural Netw. Learn. Syst. http://dx.doi.org/10.1109/TNNLS.2022.3141821.

Liu, Q., Zhou, H., Xu, Q., Liu, X., Wang, Y., 2020b. PSGAN: A generative adversarial network for remote sensing image pan-sharpening. IEEE Trans. Geosci. Remote Sens. 59 (12), 10227–10242. http://dx.doi.org/10.1109/TGRS.2020.3042974.

Loncan, L., De Almeida, L.B., Bioucas-Dias, J.M., Briottet, X., Chanussot, J., Dobigeon, N., Fabre, S., Liao, W., Licciardi, G.A., Simoes, M., et al., 2015. Hyperspectral pansharpening: A review. IEEE Geosci. Remote Sens. Mag. 3 (3), 27–46. http://dx.doi.org/10.1109/MGRS.2015.2440094.

Lu, R., Chen, B., Cheng, Z., Wang, P., 2020. RAFnet: Recurrent attention fusion network of hyperspectral and multispectral images. Signal Process. 177, 107737. http://dx.doi.org/10.1016/j.sigpro.2020.107737.

Lu, W., Tao, C., Li, H., Qi, J., Li, Y., 2022. A unified deep learning framework for urban functional zone extraction based on multi-source heterogeneous data. Remote Sens. Environ. 270, 112830. http://dx.doi.org/10.1016/j.rse.2021.112830.

Lu, X., Yang, D., Jia, F., Zhao, Y., 2021. Coupled convolutional neural network-based detail injection method for hyperspectral and multispectral image fusion. Applied Sciences 11 (1), 288. http://dx.doi.org/10.3390/app11010288.

Luo, S., Zhou, S., Feng, Y., Xie, J., 2020. Pansharpening via unsupervised convolutional neural networks. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 4295–4310. http://dx.doi.org/10.1109/JSTARS.2020.3008047.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. ISPRS J. Photogramm. Remote Sens. 152, 166–177. http://dx.doi.org/10.1016/j.isprsjprs.2019.04.015.

Ma, J., Yu, W., Chen, C., Liang, P., Guo, X., Jiang, J., 2020. Pan-GAN: An unsupervised pan-sharpening method for remote sensing image fusion. Inf. Fusion 62, 110–120. http://dx.doi.org/10.1016/j.inffus.2020.04.006.

Man, Q., Dong, P., Guo, H., Liu, G., Shi, R., 2014. Light detection and ranging and hyperspectral data for estimation of forest biomass: a review. J. Appl. Remote Sens. 8 (1), 081598. http://dx.doi.org/10.1117/1.JRS.8.081598.

Mantsis, D.F., Bakratsas, M., Andreadis, S., Karsisto, P., Moumtzidou, A., Gialampoukidis, I., Karppinen, A., Vrochidis, S., Kompatsiaris, I., 2022. Multimodal fusion of sentinel 1 images and social media data for snow depth estimation. IEEE Geosci. Remote Sens. Lett. 19, 1–5. http://dx.doi.org/10.1109/LGRS.2020.3031866.

Masi, G., Cozzolino, D., Verdoliva, L., Scarpa, G., 2016. Pansharpening by convolutional neural networks. Remote Sens. 8 (7), 594. http://dx.doi.org/10.3390/rs8070594.

Meng, X., Shen, H., Li, H., Zhang, L., Fu, R., 2019. Review of the pansharpening methods for remote sensing images based on the idea of meta-analysis: Practical discussion and challenges. Inf. Fusion 46, 102–113. http://dx.doi.org/10.1016/j.inffus.2018.05.006.

Meng, X., Xiong, Y., Shao, F., Shen, H., Sun, W., Yang, G., Yuan, Q., Fu, R., Zhang, H., 2021. A large-scale benchmark data set for evaluating pansharpening performance: Overview and implementation. IEEE Geosci. Remote Sens. Mag. 9 (1), 18–52. http://dx.doi.org/10.1109/MGRS.2020.2976696.

Meraner, A., Ebel, P., Zhu, X.X., Schmitt, M., 2020. Cloud removal in sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion. ISPRS J. Photogramm. Remote Sens. 166, 333–346. http://dx.doi.org/10.1016/j.isprsjprs.2020.05.013.

Mohla, S., Pande, S., Banerjee, B., Chaudhuri, S., 2020. Fusatnet: Dual attention based spectrospatial multimodal fusion network for hyperspectral and lidar classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 416–425. http://dx.doi.org/10.1109/CVPRW50498.2020.00054.

Moreira, A., Prats-Iraola, P., Younis, M., Krieger, G., Hajnsek, I., Papathanassiou, K.P., 2013. A tutorial on synthetic aperture radar. IEEE Geosci. Remote Sens. Mag. 1 (1), 6–43. http://dx.doi.org/10.1109/MGRS.2013.2248301.

Nie, J., Xu, Q., Pan, J., 2022. Unsupervised hyperspectral pansharpening by ratio estimation and residual attention network. IEEE Geosci. Remote Sens. Lett. 19, 1–5. http://dx.doi.org/10.1109/LGRS.2022.3149166.

Ozcelik, F., Alganci, U., Sertel, E., Unal, G., 2020. Rethinking CNN-based pansharpening: Guided colorization of panchromatic images via GANS. IEEE Trans. Geosci. Remote Sens. 59 (4), 3486–3501. http://dx.doi.org/10.1109/TGRS.2020.3010441.

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. http://dx.doi.org/10.1109/LGRS.2017.2668299.

Parajuli, B., Kumar, P., Mukherjee, T., Pasiliao, E., Jambawalikar, S., 2018. Fusion of aerial lidar and images for road segmentation with deep cnn. In: Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. pp. 548–551. http://dx.doi.org/10.1145/3274895.3274993.

Peng, J., Liu, L., Wang, J., Zhang, E., Zhu, X., Zhang, Y., Feng, J., Jiao, L., 2020. Psmd-net: A novel pan-sharpening method based on a multiscale dense network. IEEE Trans. Geosci. Remote Sens. 59 (6), 4957–4971. http://dx.doi.org/10.1109/TGRS.2020.3020162.

Qian, Z., Liu, X., Tao, F., Zhou, T., 2020. Identification of urban functional areas by coupling satellite images and taxi GPS trajectories. Remote Sens. 12 (15), 2449. http://dx.doi.org/10.3390/rs12152449.

Qu, J., Hou, S., Dong, W., Xiao, S., Du, Q., Li, Y., 2022a. A dual-branch detail extraction network for hyperspectral pansharpening. IEEE Trans. Geosci. Remote Sens. 60, 1–21. http://dx.doi.org/10.1109/TGRS.2021.3132997.

Qu, Y., Qi, H., Ayhan, B., Kwan, C., Kidd, R., 2017. Does multispectral/hyperspectral pansharpening improve the performance of anomaly detection? In: 2017 IEEE International Geoscience and Remote Sensing Symposium. IGARSS, IEEE, pp. 6130–6133. http://dx.doi.org/10.1109/IGARSS.2017.8128408.

Qu, Y., Qi, H., Kwan, C., 2018. Unsupervised sparse dirichlet-net for hyperspectral image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2511–2520. http://dx.doi.org/10.1109/CVPR.2018.00266.

Qu, Y., Qi, H., Kwan, C., Yokoya, N., Chanussot, J., 2022b. Unsupervised and unregistered hyperspectral image super-resolution with mutual Dirichlet-Net. IEEE Trans. Geosci. Remote Sens. 60, 1–18. http://dx.doi.org/10.1109/TGRS.2021.3079518.

Qu, J., Shi, Y., Xie, W., Li, Y., Wu, X., Du, Q., 2022. MSSL: Hyperspectral and panchromatic images fusion via multiresolution spatial–spectral feature learning networks. IEEE Trans. Geosci. Remote Sens. 60, 1–13. http://dx.doi.org/10.1109/TGRS.2021.3066374.

Ranchin, T., Aiazzi, B., Alparone, L., Baronti, S., Wald, L., 2003. Image fusion—The ARSIS concept and some successful implementation schemes. ISPRS J. Photogramm. Remote Sens. 58 (1–2), 4–18. http://dx.doi.org/10.1016/S0924-2716(03)00013-3.

Rasti, B., Hong, D., Hang, R., Ghamisi, P., Kang, X., Chanussot, J., Benediktsson, J.A., 2020. Feature extraction for hyperspectral imagery: The evolution from shallow to deep: Overview and toolbox. IEEE Geosci. Remote Sens. Mag. 8 (4), 60–88. http://dx.doi.org/10.1109/MGRS.2020.2979764.

Scarpa, G., Vitale, S., Cozzolino, D., 2018. Target-adaptive CNN-based pansharpening. IEEE Trans. Geosci. Remote Sens. 56 (9), 5443–5457. http://dx.doi.org/10.1109/TGRS.2018.2817393.

Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural Netw. 61, 85–117. http://dx.doi.org/10.1016/j.neunet.2014.09.003.

Schmitt, M., Zhu, X.X., 2016. Data fusion and remote sensing: An ever-growing relationship. IEEE Geosci. Remote Sens. Mag. 4 (4), 6–23. http://dx.doi.org/10.1109/MGRS.2016.2561021.

Seo, S., Choi, J.-S., Lee, J., Kim, H.-H., Seo, D., Jeong, J., Kim, M., 2020. UP-SNet: Unsupervised pan-sharpening network with registration learning between panchromatic and multi-spectral images. IEEE Access 8, 201199–201217. http://dx.doi.org/10.1109/ACCESS.2020.3035802.

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. http://dx.doi.org/10.1109/JSTARS.2018.2805923.

Shao, Z., Lu, Z., Ran, M., Fang, L., Zhou, J., Zhang, Y., 2019. Residual encoder–decoder conditional generative adversarial network for pansharpening. IEEE Geosci. Remote Sens. Lett. 17 (9), 1573–1577. http://dx.doi.org/10.1109/LGRS.2019.2949745.

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. http://dx.doi.org/10.1109/JSTARS.2017.2748341.

Shen, H., Jiang, M., Li, J., Yuan, Q., Wei, Y., Zhang, L., 2019. Spatial–spectral fusion by combining deep learning and variational model. IEEE Trans. Geosci. Remote Sens. 57 (8), 6169–6181. http://dx.doi.org/10.1109/TGRS.2019.2904659.

Shen, D., Liu, J., Wu, Z., Yang, J., Xiao, L., 2022. ADMM-HFNet: A matrix decomposition-based deep approach for hyperspectral image fusion. IEEE Trans. Geosci. Remote Sens. 60, 1–17. http://dx.doi.org/10.1109/TGRS.2021.3112181.

Shen, D., Liu, J., Xiao, Z., Yang, J., Xiao, L., 2020. A twice optimizing net with matrix decomposition for hyperspectral and multispectral image fusion. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 4095–4110. http://dx.doi.org/10.1109/JSTARS.2020.3009250.

Song, H., Liu, Q., Wang, G., Hang, R., Huang, B., 2018. Spatiotemporal satellite image fusion using deep convolutional neural networks. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11 (3), 821–829. http://dx.doi.org/10.1109/JSTARS.2018.2797894.

Srivastava, S., Vargas-Muñoz, J.E., Tuia, D., 2019. Understanding urban landuse from the above and ground perspectives: A deep learning, multimodal solution. Remote Sens. Environ. 228, 129–143. http://dx.doi.org/10.1016/j.rse.2019.04.014.

Sun, W., Ren, K., Meng, X., Xiao, C., Yang, G., Peng, J., 2021. A band divide-and-conquer multispectral and hyperspectral image fusion method. IEEE Trans. Geosci. Remote Sens. 60, 1–13. http://dx.doi.org/10.1109/TGRS.2020.3046321.

Sun, Z., Zhao, X., Wu, M., Wang, C., 2019. Extracting urban impervious surface from worldview-2 and airborne LiDAR data using 3D convolutional neural networks. J. Indian Soc. Remote Sensing 47 (3), 401–412. http://dx.doi.org/10.1007/s12524-018-0917-5.

Tan, Z., Di, L., Zhang, M., Guo, L., Gao, M., 2019. An enhanced deep convolutional model for spatiotemporal image fusion. Remote Sens. 11 (24), 2898. http://dx.doi.org/10.3390/rs11242898.

Tan, Z., Gao, M., Li, X., Jiang, L., 2022. A flexible reference-insensitive spatiotemporal fusion model for remote sensing images using conditional generative adversarial network. IEEE Trans. Geosci. Remote Sens. 60, 1–13. http://dx.doi.org/10.1109/TGRS.2021.3050551.

Tan, Z., Yue, P., Di, L., Tang, J., 2018. Deriving high spatiotemporal remote sensing images using deep convolutional network. Remote Sens. 10 (7), 1066. http://dx.doi.org/10.3390/rs10071066.

Tian, X., Li, K., Wang, Z., Ma, J., 2021. VP-net: An interpretable deep network for variational pansharpening. IEEE Trans. Geosci. Remote Sens. 60, 1–16. http://dx.doi.org/10.1109/TGRS.2021.3089868.

Uezato, T., Hong, D., Yokoya, N., He, W., 2020. Guided deep decoder: Unsupervised image pair fusion. In: European Conference on Computer Vision. Springer, pp. 87–102. http://dx.doi.org/10.1007/978-3-030-58539-6_6.

Vivone, G., Alparone, L., Chanussot, J., Dalla Mura, M., Garzelli, A., Licciardi, G.A., Restaino, R., Wald, L., 2014. A critical comparison among pansharpening algorithms. IEEE Trans. Geosci. Remote Sens. 53 (5), 2565–2586. http://dx.doi.org/10.1109/TGRS.2014.2361734.

Vivone, G., Dalla Mura, M., Garzelli, A., Pacifici, F., 2021. A benchmarking protocol for pansharpening: Dataset, preprocessing, and quality assessment. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 6102–6118. http://dx.doi.org/10.1109/JSTARS.2021.3086877.

Vivone, G., Dalla Mura, M., Garzelli, A., Restaino, R., Scarpa, G., Ulfarsson, M.O., Alparone, L., Chanussot, J., 2020. A new benchmark based on recent advances in multispectral pansharpening: Revisiting pansharpening with classical and emerging pansharpening methods. IEEE Geosci. Remote Sens. Mag. 9 (1), 53–81. http://dx.doi.org/10.1109/MGRS.2020.3019315.

Wald, L., 1999. Some terms of reference in data fusion. IEEE Trans. Geosci. Remote Sens. 37 (3), 1190–1193. http://dx.doi.org/10.1109/36.763269.

Wang, X., Chen, J., Wei, Q., Richard, C., 2021a. Hyperspectral image super-resolution via deep prior regularization with parameter estimation. IEEE Trans. Circuits Syst. Video Technol. 32 (4), 1708–1723. http://dx.doi.org/10.1109/TCSVT.2021.3078559.

Wang, X., Feng, Y., Song, R., Mu, Z., Song, C., 2022a. Multi-attentive hierarchical dense fusion net for fusion classification of hyperspectral and LiDAR data. Inf. Fusion 82, 1–18. http://dx.doi.org/10.1016/j.inffus.2021.12.008.

Wang, W., Fu, X., Zeng, W., Sun, L., Zhan, R., Huang, Y., Ding, X., 2021b. Enhanced deep blind hyperspectral image fusion. IEEE Trans. Neural Netw. Learn. Syst. http://dx.doi.org/10.1109/TNNLS.2021.3105543.

Wang, J., Li, W., Gao, Y., Zhang, M., Tao, R., Du, Q., 2022b. Hyperspectral and SAR image classification via multiscale interactive fusion network. IEEE Trans. Neural Netw. Learn. Syst. http://dx.doi.org/10.1109/TNNLS.2022.3171572.

Wang, J., Li, J., Shi, Y., Lai, J., Tan, X., 2022c. AM3Net: Adaptive mutual-learning-based multimodal data fusion network. IEEE Trans. Circuits Syst. Video Technol. http://dx.doi.org/10.1109/TCSVT.2022.3148257.

Wang, S., Quan, D., Liang, X., Ning, M., Guo, Y., Jiao, L., 2018. A deep learning framework for remote sensing image registration. ISPRS J. Photogramm. Remote Sens. 145, 148–164. http://dx.doi.org/10.1016/j.isprsjprs.2017.12.012.

Wang, W., Zeng, W., Huang, Y., Ding, X., Paisley, J., 2019. Deep blind hyperspectral image fusion. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4149–4158. http://dx.doi.org/10.1109/ICCV.2019.00425.

Wei, W., Nie, J., Li, Y., Zhang, L., Zhang, Y., 2020. Deep recursive network for hyperspectral image super-resolution. IEEE Trans. Comput. Imaging 6, 1233–1244. http://dx.doi.org/10.1109/TCI.2020.3014451.

Wei, W., Nie, J., Zhang, L., Zhang, Y., 2022. Unsupervised recurrent hyperspectral imagery super-resolution using pixel-aware refinement. IEEE Trans. Geosci. Remote Sens. 60, 1–15. http://dx.doi.org/10.1109/TGRS.2020.3039534.

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. http://dx.doi.org/10.1109/LGRS.2017.2736020.

Wu, X., Hong, D., Chanussot, J., 2022. Convolutional neural networks for multimodal remote sensing data classification. IEEE Trans. Geosci. Remote Sens. 60, 1–10. http://dx.doi.org/10.1109/TGRS.2021.3124913.

Wu, Z.-C., Huang, T.-Z., Deng, L.-J., Hu, J.-F., Vivone, G., 2021a. VO+ net: An adaptive approach using variational optimization and deep learning for panchromatic sharpening. IEEE Trans. Geosci. Remote Sens. 60, 1–16. http://dx.doi.org/10.1109/TGRS.2021.3066425.

Wu, X., Li, W., Hong, D., Tao, R., Du, Q., 2021b. Deep learning for unmanned aerial vehicle-based object detection and tracking: a survey. IEEE Geosci. Remote Sens. Mag. 10 (1), 91–124. http://dx.doi.org/10.1109/MGRS.2021.3115137.

Xiao, J., Li, J., Yuan, Q., Jiang, M., Zhang, L., 2021. Physics-based GAN with iterative refinement unit for hyperspectral and multispectral image fusion. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 6827–6841. http://dx.doi.org/10.1109/JSTARS.2021.3075727.

Xie, W., Cui, Y., Li, Y., Lei, J., Du, Q., Li, J., 2021. HPGAN: Hyperspectral pansharpening using 3-D generative adversarial networks. IEEE Trans. Geosci. Remote Sens. 59 (1), 463–477. http://dx.doi.org/10.1109/TGRS.2020.2994238.

Xie, W., Lei, J., Cui, Y., Li, Y., Du, Q., 2020. Hyperspectral pansharpening with deep priors. IEEE Trans. Neural Netw. Learn. Syst. 31 (5), 1529–1543. http://dx.doi.org/10.1109/TNNLS.2019.2920857.

Xie, Q., Zhou, M., Zhao, Q., Meng, D., Zuo, W., Xu, Z., 2019. Multispectral and hyperspectral image fusion by MS/HS fusion net. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1585–1594. http://dx.doi.org/10.1109/CVPR.2019.00168.

Xie, Q., Zhou, M., Zhao, Q., Xu, Z., Meng, D., 2022. MHF-net: An interpretable deep network for multispectral and hyperspectral image fusion. IEEE Trans. Pattern Anal. Mach. Intell. 44 (3), 1457–1473. http://dx.doi.org/10.1109/TPAMI.2020.3015691.

Xing, Y., Wang, M., Yang, S., Jiao, L., 2018. Pan-sharpening via deep metric learning. ISPRS J. Photogramm. Remote Sens. 145, 165–183. http://dx.doi.org/10.1016/j.isprsjprs.2018.01.016.

Xing, Y., Yang, S., Feng, Z., Jiao, L., 2020. Dual-collaborative fusion model for multispectral and panchromatic image fusion. IEEE Trans. Geosci. Remote Sens. 60, 1–15. http://dx.doi.org/10.1109/TGRS.2020.3036625.

Xu, S., Amira, O., Liu, J., Zhang, C.-X., Zhang, J., Li, G., 2020a. HAM-MFN: Hyperspectral and multispectral image multiscale fusion network with RAP loss. IEEE Trans. Geosci. Remote Sens. 58 (7), 4618–4628. http://dx.doi.org/10.1109/TGRS.2020.2964777.

Xu, X., Li, W., Ran, Q., Du, Q., Gao, L., Zhang, B., 2018. Multisource remote sensing data classification based on convolutional neural network. IEEE Trans. Geosci. Remote Sens. 56 (2), 937–949. http://dx.doi.org/10.1109/TGRS.2017.2756851.

Xu, H., Ma, J., Shao, Z., Zhang, H., Jiang, J., Guo, X., 2021. SDPNet: A deep network for pan-sharpening with enhanced information representation. IEEE Trans. Geosci. Remote Sens. 59 (5), 4120–4134. http://dx.doi.org/10.1109/TGRS.2020.3022482.

Xu, S., Qing, L., Han, L., Liu, M., Peng, Y., Shen, L., 2020b. A new remote sensing images and point-of-interest fused (RPF) model for sensing urban functional regions. Remote Sens. 12 (6), 1032. http://dx.doi.org/10.3390/rs12061032.

Yang, J., Fu, X., Hu, Y., Huang, Y., Ding, X., Paisley, J., 2017. PanNet: A deep network architecture for pan-sharpening. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1753–1761. http://dx.doi.org/10.1109/ICCV.2017.193.

Yang, Y., Tu, W., Huang, S., Lu, H., Wan, W., Gan, L., 2022a. Dual-stream convolutional neural network with residual information enhancement for pansharpening. IEEE Trans. Geosci. Remote Sens. 60, 1–16. http://dx.doi.org/10.1109/TGRS.2021.3098752.

Yang, J., Xiao, L., Zhao, Y.-Q., Chan, J.C.-W., 2022b. Variational regularization network with attentive deep prior for hyperspectral–multispectral image fusion. IEEE Trans. Geosci. Remote Sens. 60, 1–17. http://dx.doi.org/10.1109/TGRS.2021.3080697.

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 (5), 800. http://dx.doi.org/10.3390/rs10050800.

Yao, J., Hong, D., Chanussot, J., Meng, D., Zhu, X., Xu, Z., 2020. Cross-attention in coupled unmixing nets for unsupervised hyperspectral super-resolution. In: European Conference on Computer Vision. Springer, pp. 208–224. http://dx.doi.org/10.1007/978-3-030-58526-6_13.

Yao, Y., Yan, X., Luo, P., Liang, Y., Ren, S., Hu, Y., Han, J., Guan, Q., 2022. Classifying land-use patterns by integrating time-series electricity data and high-spatial resolution remote sensing imagery. Int. J. Appl. Earth Obs. Geoinf. 106, 102664. http://dx.doi.org/10.1016/j.jag.2021.102664.

Yin, J., Dong, J., Hamm, N.A., Li, Z., Wang, J., Xing, H., Fu, P., 2021a. Integrating remote sensing and geospatial big data for urban land use mapping: A review. Int. J. Appl. Earth Obs. Geoinf. 103 (1), 102514. http://dx.doi.org/10.1016/j.jag.2021.102514.

Yin, Z., Wu, P., Foody, G.M., Wu, Y., Liu, Z., Du, Y., Ling, F., 2021b. Spatiotemporal fusion of land surface temperature based on a convolutional neural network. IEEE Trans. Geosci. Remote Sens. 59 (2), 1808–1822. http://dx.doi.org/10.1109/TGRS.2020.2999943.

Yokoya, N., Grohnfeldt, C., Chanussot, J., 2017. Hyperspectral and multispectral data fusion: A comparative review of the recent literature. IEEE Geosci. Remote Sens. Mag. 5 (2), 29–56. http://dx.doi.org/10.1109/MGRS.2016.2637824.

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. http://dx.doi.org/10.1109/JSTARS.2018.2794888.

Zhang, B., Chen, Z., Peng, D., Benediktsson, J.A., Liu, B., Zou, L., Li, J., Plaza, A., 2019a. Remotely sensed big data: Evolution in model development for information extraction [point of view]. Proc. IEEE 107 (12), 2294–2301. http://dx.doi.org/10.1109/JPROC.2019.2948454.

Zhang, T.-J., Deng, L.-J., Huang, T.-Z., Chanussot, J., Vivone, G., 2022a. A triple-double convolutional neural network for panchromatic sharpening. IEEE Trans. Neural Netw. Learn. Syst. http://dx.doi.org/10.1109/TNNLS.2022.3155655.

Zhang, X., Huang, W., Wang, Q., Li, X., 2021a. SSR-NET: Spatial–spectral reconstruction network for hyperspectral and multispectral image fusion. IEEE Trans. Geosci. Remote Sens. 59 (7), 5953–5965. http://dx.doi.org/10.1109/TGRS.2020.3018732.

Zhang, M., Li, W., Du, Q., Gao, L., Zhang, B., 2020a. Feature extraction for classification of hyperspectral and LiDAR data using patch-to-patch CNN. IEEE Trans. Cybern. 50 (1), 100–111. http://dx.doi.org/10.1109/TCYB.2018.2864670.

Zhang, M., Li, W., Tao, R., Li, H., Du, Q., 2022b. Information fusion for classification of hyperspectral and LiDAR data using IP-CNN. IEEE Trans. Geosci. Remote Sens. 60, 1–12. http://dx.doi.org/10.1109/TGRS.2021.3093334.

Zhang, J., Lin, X., 2017. Advances in fusion of optical imagery and LiDAR point cloud applied to photogrammetry and remote sensing. Int. J. Image Data Fusion 8 (1), 1–31. http://dx.doi.org/10.1080/19479832.2016.1160960.

Zhang, Y., Liu, C., Sun, M., Ou, Y., 2019b. Pan-sharpening using an efficient bidirectional pyramid network. IEEE Trans. Geosci. Remote Sens. 57 (8), 5549–5563. http://dx.doi.org/10.1109/TGRS.2019.2900419.

Zhang, H., Ma, J., 2021. GTP-PNet: A residual learning network based on gradient transformation prior for pansharpening. ISPRS J. Photogramm. Remote Sens. 172, 223–239. http://dx.doi.org/10.1016/j.isprsjprs.2020.12.014.

Zhang, H., Ni, W., Yan, W., Xiang, D., Wu, J., Yang, X., Bian, H., 2019c. Registration of multimodal remote sensing image based on deep fully convolutional neural network. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12 (8), 3028–3042. http://dx.doi.org/10.1109/JSTARS.2019.2916560.

Zhang, L., Nie, J., Wei, W., Li, Y., Zhang, Y., 2021b. Deep blind hyperspectral image super-resolution. IEEE Trans. Neural Netw. Learn. Syst. 32 (6), 2388–2400. http://dx.doi.org/10.1109/TNNLS.2020.3005234.

Zhang, L., Nie, J., Wei, W., Zhang, Y., Liao, S., Shao, L., 2020b. Unsupervised adaptation learning for hyperspectral imagery super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3070–3079. http://dx.doi.org/10.1109/CVPR42600.2020.00314.

Zhang, H., Song, Y., Han, C., Zhang, L., 2021c. Remote sensing image spatiotemporal fusion using a generative adversarial network. IEEE Trans. Geosci. Remote Sens. 59 (5), 4273–4286. http://dx.doi.org/10.1109/TGRS.2020.3010530.

Zhang, H., Xu, H., Tian, X., Jiang, J., Ma, J., 2021d. Image fusion meets deep learning: A survey and perspective. Inf. Fusion 76, 323–336. http://dx.doi.org/10.1016/j.inffus.2021.06.008.

Zhang, Q., Yang, L.T., Chen, Z., Li, P., 2018. A survey on deep learning for big data. Inf. Fusion 42, 146–157. http://dx.doi.org/10.1016/j.inffus.2017.10.006.

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. http://dx.doi.org/10.1109/MGRS.2016.2540798.

Zhao, X., Tao, R., Li, W., Li, H.-C., Du, Q., Liao, W., Philips, W., 2020. Joint classification of hyperspectral and LiDAR data using hierarchical random walk and deep CNN architecture. IEEE Trans. Geosci. Remote Sens. 58 (10), 7355–7370. http://dx.doi.org/10.1109/TGRS.2020.2982064.

Zheng, K., Gao, L., Hong, D., Zhang, B., Chanussot, J., 2022. NonRegSRNet: A nonrigid registration hyperspectral super-resolution network. IEEE Trans. Geosci. Remote Sens. 60, 1–16. http://dx.doi.org/10.1109/TGRS.2021.3135501.

Zheng, K., Gao, L., Liao, W., Hong, D., Zhang, B., Cui, X., Chanussot, J., 2021. Coupled convolutional neural network with adaptive response function learning for unsupervised hyperspectral super resolution. IEEE Trans. Geosci. Remote Sens. 59 (3), 2487–2502. http://dx.doi.org/10.1109/TGRS.2020.3006534.

Zheng, Y., Li, J., Li, Y., Guo, J., Wu, X., Chanussot, J., 2020. Hyperspectral pansharpening using deep prior and dual attention residual network. IEEE Trans. Geosci. Remote Sens. 58 (11), 8059–8076. http://dx.doi.org/10.1109/TGRS.2020.2986313.

Zhou, M., Fu, X., Huang, J., Zhao, F., Liu, A., Wang, R., 2021a. Effective pan-sharpening with transformer and invertible neural network. IEEE Trans. Geosci. Remote Sens. 60, 1–14. http://dx.doi.org/10.1109/TNNLS.2022.3155655.

Zhou, F., Hang, R., Liu, Q., Yuan, X., 2019. Pyramid fully convolutional network for hyperspectral and multispectral image fusion. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12 (5), 1549–1558. http://dx.doi.org/10.1109/JSTARS.2019.2910990.

Zhou, H., Liu, Q., Wang, Y., 2021b. Pgman: An unsupervised generative multiadversarial network for pansharpening. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 6316–6327. http://dx.doi.org/10.1109/JSTARS.2021.3090252.

Zhou, H., Liu, Q., Wang, Y., 2022a. PanFormer: a transformer based model for pan-sharpening. arXiv preprint arXiv:2203.02916.

Zhou, H., Liu, Q., Weng, D., Wang, Y., 2022b. Unsupervised cycle-consistent generative adversarial networks for pan-sharpening. IEEE Trans. Geosci. Remote Sens. 60, 1–14. http://dx.doi.org/10.1109/TGRS.2022.3166528.

Zhou, C., Zhang, J., Liu, J., Zhang, C., Fei, R., Xu, S., 2020. PercepPan: Towards unsupervised pan-sharpening based on perceptual loss. Remote Sens. 12 (14), 2318. http://dx.doi.org/10.3390/rs12142318.

Zhu, X., Cai, F., Tian, J., Williams, T.K.-A., 2018. Spatiotemporal fusion of multisource remote sensing data: Literature survey, taxonomy, principles, applications, and future directions. Remote Sens. 10 (4), 527. http://dx.doi.org/10.3390/rs10040527.

Zhu, Z., Hou, J., Chen, J., Zeng, H., Zhou, J., 2021. Hyperspectral image super-resolution via deep progressive zero-centric residual learning. IEEE Trans. Image Process. 30, 1423–1438. http://dx.doi.org/10.1109/TIP.2020.3044214.

Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geosci. Remote Sens. Mag. 5 (4), 8–36. http://dx.doi.org/10.1109/MGRS.2017.2762307.

Zhu, X., Zhan, W., Zhou, J., Chen, X., Liang, Z., Xu, S., Chen, J., 2022. A novel framework to assess all-round performances of spatiotemporal fusion models. Remote Sens. Environ. 274, 113002. http://dx.doi.org/10.1016/j.rse.2022.113002.