


## Review

# Land Use and Land Cover Classification Meets Deep Learning: A Review

Shengyu Zhao <sup>1</sup>, Kaiwen Tu <sup>1</sup>, Shutong Ye <sup>1</sup>, Hao Tang <sup>1</sup>, Yaocong Hu <sup>2</sup> and Chao Xie <sup>1,3,\*</sup> <sup>1</sup> College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing 210037, China<sup>2</sup> School of Electrical Engineering, Anhui Polytechnic University, Wuhu 241000, China<sup>3</sup> College of Landscape Architecture, Nanjing Forestry University, Nanjing 210037, China

\* Correspondence: chaoxie@njfu.edu.cn

**Abstract:** As one of the important components of Earth observation technology, land use and land cover (LULC) image classification plays an essential role. It uses remote sensing techniques to classify specific categories of ground cover as a means of analyzing and understanding the natural attributes of the Earth's surface and the state of land use. It provides important information for applications in environmental protection, urban planning, and land resource management. However, remote sensing images are usually high-dimensional data and have limited available labeled samples, so performing the LULC classification task faces great challenges. In recent years, due to the emergence of deep learning technology, remote sensing data processing methods based on deep learning have achieved remarkable results, bringing new possibilities for the research and development of LULC classification. In this paper, we present a systematic review of deep-learning-based LULC classification, mainly covering the following five aspects: (1) introduction of the main components of five typical deep learning networks, how they work, and their unique benefits; (2) summary of two baseline datasets for LULC classification (pixel-level, patch-level) and performance metrics for evaluating different models (OA, AA, F1, and MIOU); (3) review of deep learning strategies in LULC classification studies, including convolutional neural networks (CNNs), autoencoders (AEs), generative adversarial networks (GANs), and recurrent neural networks (RNNs); (4) challenges faced by LULC classification and processing schemes under limited training samples; (5) outlooks on the future development of deep-learning-based LULC classification.



**Citation:** Zhao, S.; Tu, K.; Ye, S.; Tang, H.; Hu, Y.; Xie, C. Land Use and Land Cover Classification Meets Deep Learning: A Review. *Sensors* **2023**, *23*, 8966. <https://doi.org/10.3390/s23218966>

Academic Editor: Sijia Li

Received: 27 September 2023

Revised: 24 October 2023

Accepted: 2 November 2023

Published: 3 November 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** deep learning; LULC; image classification; remote sensing

## 1. Introduction

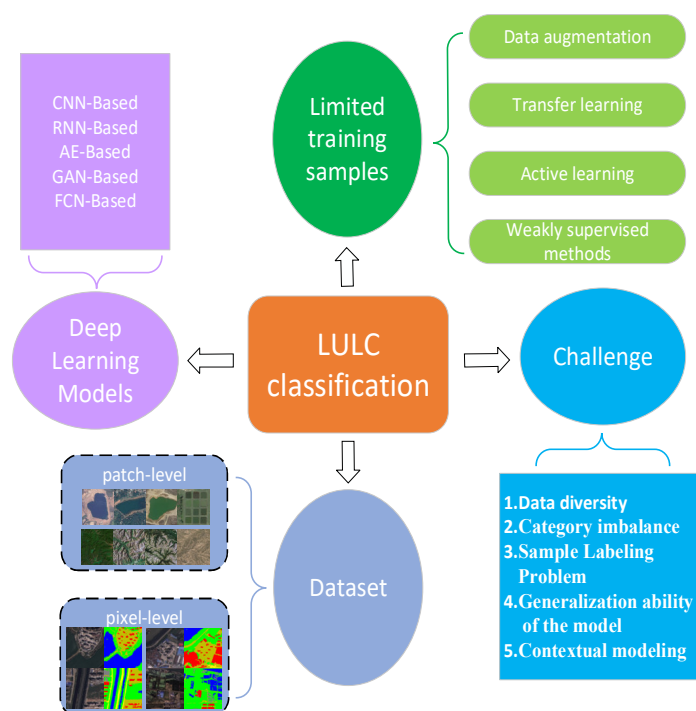
Land use and land cover (LULC) is the expression of the transformation of the natural ecosystem of land into an artificial ecosystem and the naturally occurring or human-induced cover on the surface. It serves an important role in the fields of disaster management, urban planning, environmental protection, and agricultural production, and has been a popular theme in Earth observation research. Remote sensing images have become the main data source for LULC classification due to their advantages, such as wide coverage and continuous monitoring to obtain time series data. Since the introduction of deep learning, excellent results have been achieved in the field of image processing, and thus, LULC classification has become a popular research topic [1–3].

The initial classification technique was visual interpretation classification, which was judged by the expertise of the interpreter. The advantage is that the accuracy is high, generally higher than the computer classification accuracy, so the visual interpretation classification is still applied to high-precision, high-resolution remote sensing image classification [4]. However, the biggest disadvantage is poor repeatability and poor timeliness. With the development of computer technology, machine learning is widely used in LULC classification. Traditional classification techniques are divided into two types: unsupervised and supervised classification. Unsupervised classification is represented by K-Means

and Expectation Maximization, which can distinguish the categories for data processing; however, the attributes of the classification results are uncertain. Common supervised classification algorithms include support vector machine (SVM), decision tree, maximum likelihood classification, and so on [5–8]. The parameters of the discriminant function are derived from the known image element data, and then the discriminant function is used to classify the unknown image elements. Traditional classification techniques are characterized by good repeatability and timeliness compared to visual interpretation; however, the accuracy of the classification will be greatly reduced after changing the data or study area [9].

In contrast to classic machine learning algorithms, deep learning (DL) demonstrates unique advantages in image classification. While traditional machine learning algorithms require the manual design of features for classification tasks, deep learning eliminates the need for manual intervention; it automatically learns and extracts features relevant to the target task, and this automatic feature extraction capability endows deep learning models with strong robustness and makes it easier to migrate the models across different datasets [10–14]. Deep learning algorithms can learn from large-scale data and discover potential patterns and regularities, thus improving the accuracy and effectiveness of LULC classification [9].

The rest of this paper is organized as follows: Section 2 describes the composition and construction of five commonly used deep learning models (CNN, RNN, AE, GAN, and FCN). Section 3 systematically summarizes the two benchmark datasets (patch-level and pixel-level) used for LULC classification, along with performance metrics. Section 4 describes the LULC classification based on a deep learning approach. Section 5 lists the challenges of LULC classification, such as category imbalance, sample labeling problems, etc. Section 6 discusses four solutions to the LULC classification problem with limited training samples (transfer learning, data augmentation, active learning, weak supervision methods). Section 7 provides an outlook for future research on deep learning in LULC classification. It concludes in Section 8. The general framework of this review is shown in Figure 1.



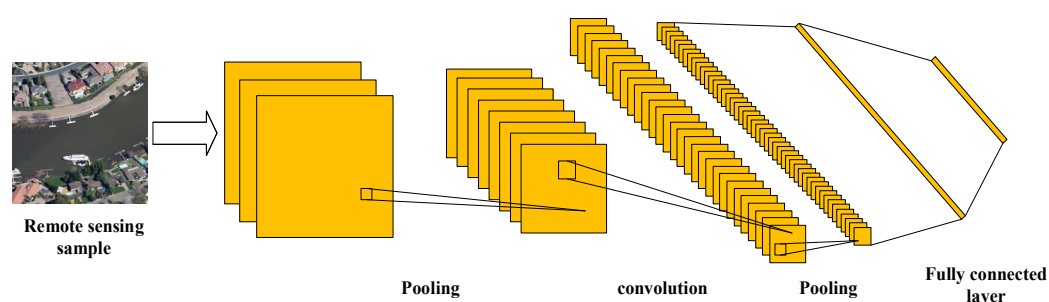
**Figure 1.** LULC classification general framework diagram.

## 2. Typical DL Models

Deep learning [15], a unique branch in the machine learning field, solves complex problems by performing computations through multiple layers of successive neural networks, i.e., mimicking human brain mechanisms to interpret data. The key to DL is to build deep neural networks by increasing the number of network layers so as to obtain a powerful performance that allows the algorithm to cope with more complex tasks. Moreover, DL plays a crucial role in processing and analyzing large-scale data by utilizing efficient algorithms for unsupervised or semi-supervised feature learning, as well as hierarchical feature extraction, thus getting rid of the tedious process of manually extracting features [16]. Commonly used DL models are convolutional neural networks, generative adversarial networks, recurrent neural networks, autoencoders, and full convolutional neural networks.

### 2.1. Convolutional Neural Networks

Convolutional neural network (CNN) is a neural network with convolutional operational feedback and deep structure in the network. Simply put, it is a system that simulates the vision of the human brain. It is a multilayer neural network structure consisting of three parts: input layer, intermediate layer, and output layer, where the intermediate layer contains some hidden layers [17,18]. After convolutional operations, the multilayer neural network combines the main features of the image to gradually generate high-level features. This process extracts features with high-level semantic information from local information, realizing multi-level information transfer and gradual integration. By combining feature extraction with classification recognition, the image recognition task can be carried out. The intermediate layer of CNN is also composed of three parts: convolutional layer, pooling layer, and fully connected layer. It is developed and researched based on neural networks. Compared to traditional neural networks, the unique feature of CNN is the convolution operation in the convolution layer. The input image goes through the convolutional layer to extract local features, and then goes to the pooling layer to downsize the extracted local features where the two layers, the convolutional layer and pooling layer, can be stacked multiple times repeatedly. The input 2D image information is transferred layer by layer to the fully connected layer to associate the high-level features obtained from the learning of convolutional and pooling layers, with the output categories making the final classification decision and finally outputting the result [19–22]. Figure 2 shows the graphical representation of CNN.

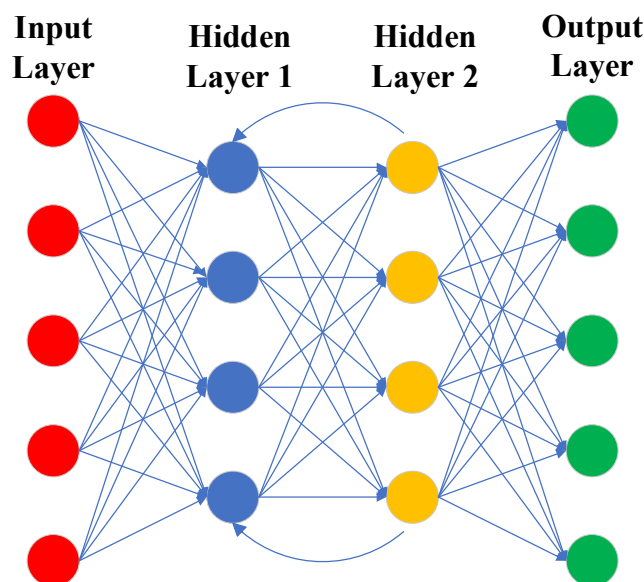


**Figure 2.** Graphical representation of CNN.

### 2.2. Recurrent Neural Networks

Recurrent neural network (RNN) [23–25], as one of the branches of deep learning networks, usually refers to temporal recurrent neural networks, which are mainly used to process sequential data, with the aim of associating the current output of a sequence with the previous information to form a recurrent connection. The most important feature of RNN is the hidden state throughout the input and output of the network. It is used as an input to the RNN along with the input vectors, which are updated and then used as an output to the RNN along with the output vectors [26,27]. At this point, the updated hidden state that is output will be used as part of the next input, thus preserving the

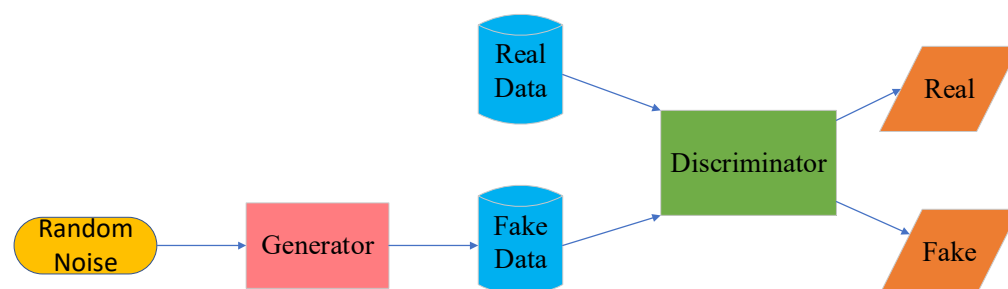
previous information. Remote sensing images usually contain both temporal and spatial information, and RNNs are often applied to LULC classification due to their powerful temporal processing capabilities. Figure 3 shows the graphical depiction of RNN.



**Figure 3.** Graphical depiction of RNN.

### 2.3. Generating Adversarial Networks

Generative adversarial network (GAN) is one of the most promising unsupervised learnings in recent years. It contains two kinds of neural networks, generator network and discriminator network, and the two neural networks learn against each other. During the training process, the generator constantly adjusts its parameters to allow the generated samples to fool the discriminator [28]. The discriminator, to more accurately determine whether the samples are real samples or generator samples, also constantly adjusts the parameters. In this way, in iterative optimization, the final generator-generated samples are realistic enough so the discriminator cannot accurately distinguish. GAN, due to its unique adversarial learning mechanism, can solve the problems caused by category imbalance, cross-region adaptive learning, and data augmentation, thus improving the accuracy and robustness of the network. Figure 4 shows the graphical representation of GAN [29].



**Figure 4.** Graphical representation of GAN.

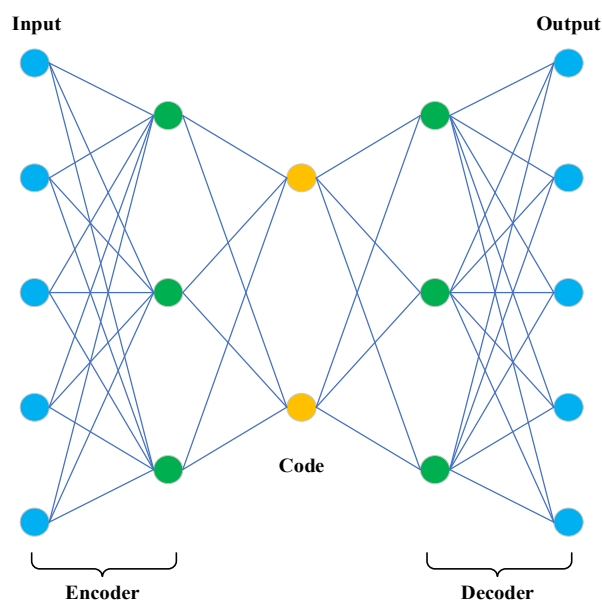
### 2.4. Autoencoder

Autoencoder (AE) is an unsupervised deep learning model consisting of an encoder and a decoder. The functionality is similar to PCA; however, in contrast to PCA, AE not only handles linear mapping but also performs nonlinear mappings for better performance. The input data will be mapped to a lower dimension than the original data for the coded representation while the encoder reconstructs the data as much as possible so that the coded representation maps back to the original dimension. Thus, the ultimate goal of AE is

to reconstruct the output similar to the original data, i.e., to minimize the reconstruction error of the decoder as much as possible [30]. When the number of layers of AE is increased to multiple layers, it is a stacked autoencoder (SAE), also called a depth autoencoder. It consists of an input layer, a hidden layer stack, and an output layer. The training process is divided into two phases, as in AE: the encoding phase and decoding phase. It can be expressed by the following equation:

$$\begin{cases} \text{Hidden} = \text{AF}(W_{\text{Hidden}} \times x + B_{\text{Hidden}}) \\ \text{Output} = \text{AF}(W_{\text{Output}} \times x + B_{\text{Output}}) \end{cases} \quad (1)$$

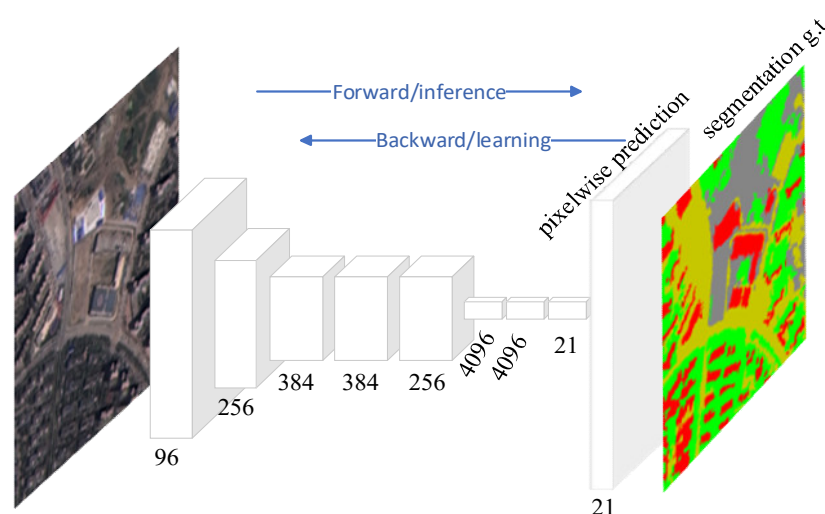
$W_{\text{Hidden}}$  and  $W_{\text{Output}}$  denote the weight of input to hidden and hidden to output, respectively.  $B_{\text{Hidden}}$  denotes hidden layer bias and  $B_{\text{Output}}$  denotes output layer bias. AF is the activation function. SAE is widely used in LULC classification due to the fact that it can help to extract the low-dimensional features in remote sensing images and capture the key information. Figure 5 shows the graphical depiction of AE.



**Figure 5.** Graphical depiction of AE.

### 2.5. Fully Convolutional Neural Networks

Fully convolutional neural networks (FCNs) belong to a special variant of CNNs and are mainly used for image semantic segmentation. Unlike CNN, it does not contain fully connected layers, but replaces fully connected layers with fully convolutional layers. Some FCNs also contain some local perceptual layers, e.g., transposed convolutional layers, upsampling layers, etc. Unlike traditional CNN-based semantic segmentation methods that take image chunks as input, FCN samples a more direct approach, instead taking the entire image as input and manually labeling the graph as output. This end-to-end training approach greatly improves the computational efficiency of the network as well as more accurate segmentation results, making FCN stand out in semantic segmentation tasks. FCN can be a favorable tool for LULC classification because it not only provides pixel-level prediction, it also captures the spatial information in remote sensing images well in order to achieve multi-scale feature learning and other advantages, so as to achieve efficient and accurate classification [31,32]. Figure 6 shows an illustration of FCN, where an input image of arbitrary size is first subjected to feature extraction through multiple convolutional and pooling layers in which the feature map resolution is gradually reduced while higher semantic features are obtained. Secondly, pixel level segmentation prediction is generated by transposed convolutional layer operations to restore the image resolution to the input state. Finally, the image is outputted.



**Figure 6.** Graphical representation of FCN.

### 3. Datasets and Performance Metrics

In deep learning, the accuracy of classification relies heavily on high-quality datasets as well as proper class categorization. Especially with the introduction of CNNs, it becomes crucial to have a large amount of training data to support the training of the network. With the continuous development of deep learning, various deep neural network models are applied to LULC classification, which makes comparing the gaps between models a research hotspot, so it is important to construct an open-source sample dataset. In recent years, researchers and organizations have released a variety of LULC classification datasets. These datasets cover a rich variety of data samples, including multi-scale images, data collected by different sensors, and multimodal features, providing a more reliable basis for data comparison in related studies. These datasets are broadly categorized into two types: patch-level datasets and pixel-level datasets. Patch-level datasets refer to the assignment of a fixed-size image to a specific feature class, and are mostly used for remote sensing image scene classification. Pixel-level samples, on the other hand, where each pixel point is considered a sample and assigned to the respective category, are mostly used for semantic segmentation of remote sensing images.

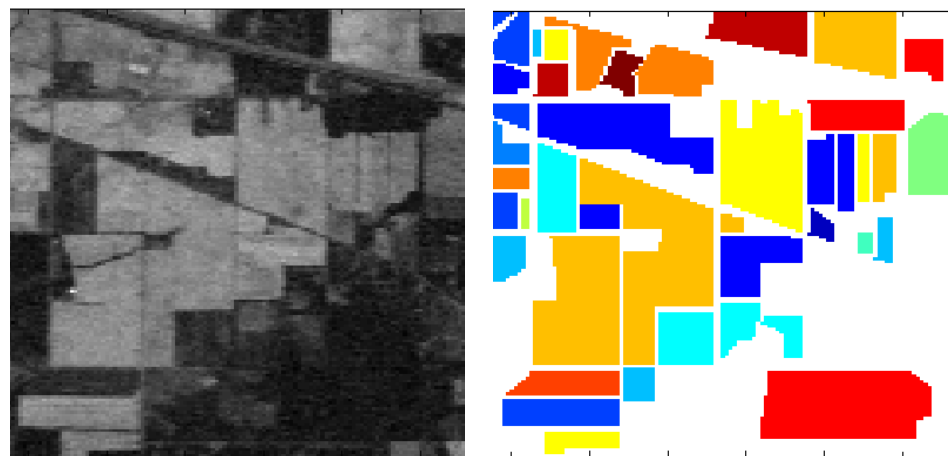
#### 3.1. Pixel-Level Datasets

Pixel-level datasets are used for LULC classification whose ultimate goal is the segmented map, i.e., land cover mapping. Such pixel-level datasets are mostly used as training semantic segmentation models, which is a huge amount of work involved in segmenting an image into several constituent regions with specific labeling of all pixels covered by the segmented regions. Therefore, producing such datasets can take a lot of time and labor on the part of the researcher. Four pixel-level benchmark datasets for LULC classification are described below.

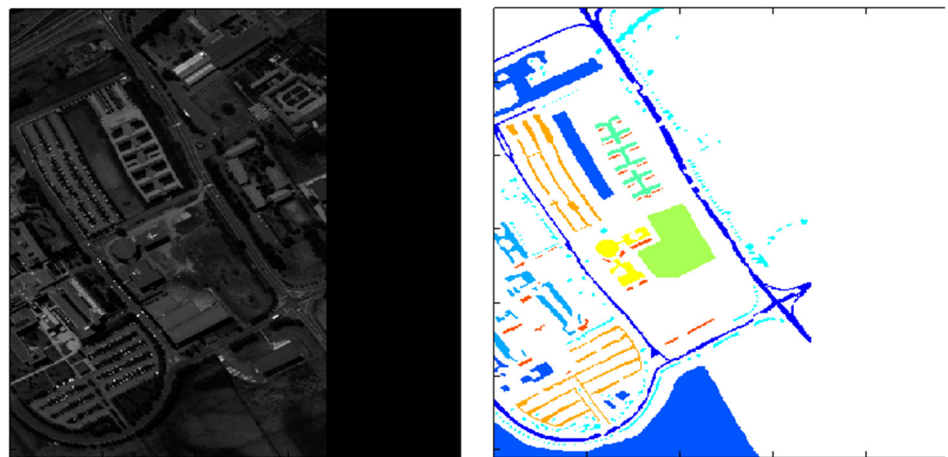
- The Indian Pines [33,34] was created by NASA in 2015 and was the first public dataset to be proposed for land cover classification. The AVIRIS sensor (224 spectral bands) imaged Indian Pines, Indiana, and the sample dataset covered 145 by 145 pixels with 16 land cover classes and a spatial resolution of 20 m. The Indian Pines dataset is small in size but still representative of developing new algorithms and methods. For evaluating algorithms, a ratio of 80:20 or 70:30 is usually chosen to divide the training and test sets (Figure 7).
- Pavia University [34] was published in 2015, and the image was captured by a ROSIS sensor with a size of  $610 \times 340$  pixels. There are nine different land cover classes, including pasture, bare soil, grassland, etc., containing a total of 103 spectral bands. A ratio of 80:20 or 70:30 is usually chosen to divide the training and test sets to evaluate the algorithms (Figure 8).



- LoveDA [35] is an adaptive ground cover dataset created by the RSIDEA team at Wuhan University in 2021 for land cover mapping. The 5987 images of this dataset are taken from the data collected by GoogleEarth in three areas, Wuhan, Nanjing, and Changzhou, and each image is  $1024 \times 1024$  pixels in size with a spatial resolution of 0.3 m after correction processing. Due to the late publication of the LoveDA dataset, it has had less impact compared to some of the classic semantic segmentation datasets, such as Indian Pines and Pavia University (Figure 9).
- LandCoverNet [36] was created in 2020 as a public dataset for land cover classification for the world. These data are taken from Sentinel-1/2 and Landsat-8 multispectral satellite images acquired in 2018. This dataset has a total of 1980 samples with a size of  $256 \times 256$  pixels and a spatial resolution of 10 m. It includes seven feature classes, namely, artificial bare ground, natural bare ground, cultivated vegetation, woody vegetation, semi-natural vegetation, water, and permanent snow/ice (Figure 10).



**Figure 7.** Schematic diagram of the Indian Pines.

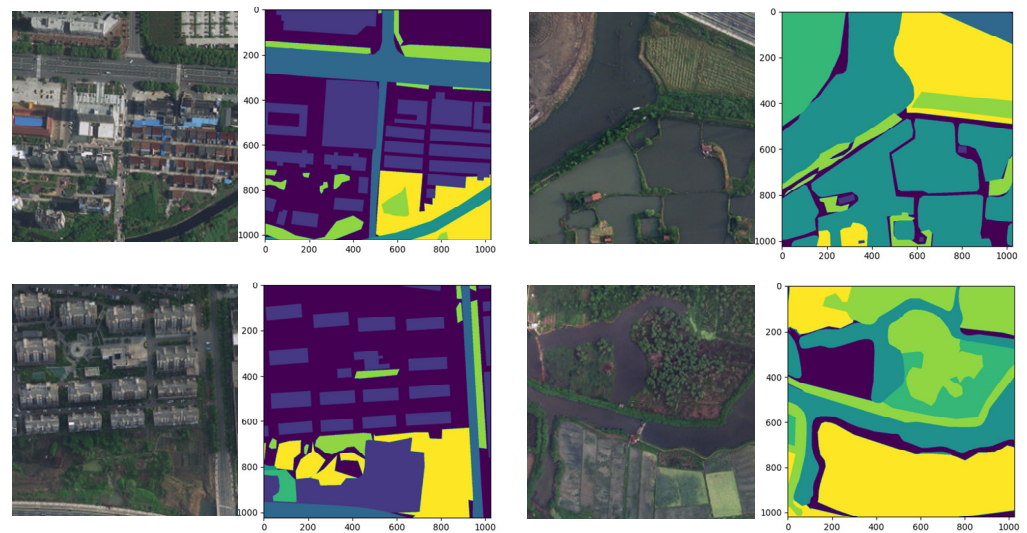


**Figure 8.** Schematic diagram of Pavia University.

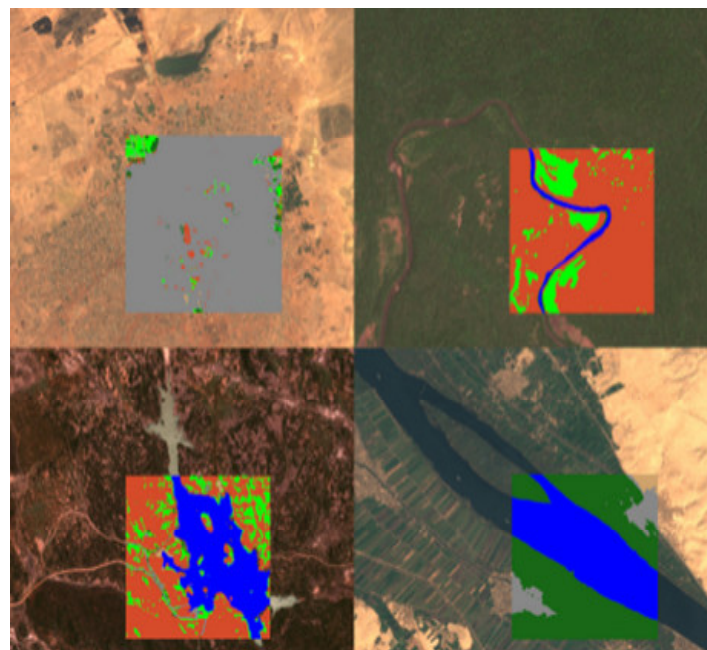
In recent years, LULC mapping has achieved remarkable results in various fields, such as ecological change detection and land use trend prediction, which has increased the interest of scholars in various countries in making sample datasets, resulting in an increasing number of pixel-level benchmark datasets.

Several commonly used pixel-level benchmark datasets are shown in Table 1. It can be seen that most of the datasets are hyperspectral and multispectral data, which brings diverse spectral characteristics of features to the model and thus improves the classification accuracy. However, most of the sample sets have a relatively small number of feature classes, which may be due to the high degree of similarity between some of the feature

classes and, hence, their grouping together. For example, there are color similarities between artificial and natural bare ground in the LandCoverNet dataset, and without segmentation, the model's LULC classification in new areas may be confusing, leading to a decrease in the model's generalization ability. Also, from the table, we see that the pixel-level LULC classification datasets are small in size, which may lead to an imbalance in the number of samples contained between different categories, making the model more biased towards learning categories with a large number of samples. Thus, categories with fewer samples may be far less accurate in classification than those with more samples. Too few samples also will lead to low generalization ability of the model, as the model is trained with only a small number of samples and captures a small number of features, so it is difficult to accurately differentiate between categories on new remote sensing images.



**Figure 9.** Schematic diagram of LoveDA.



**Figure 10.** Schematic diagram of LandCoverNet.



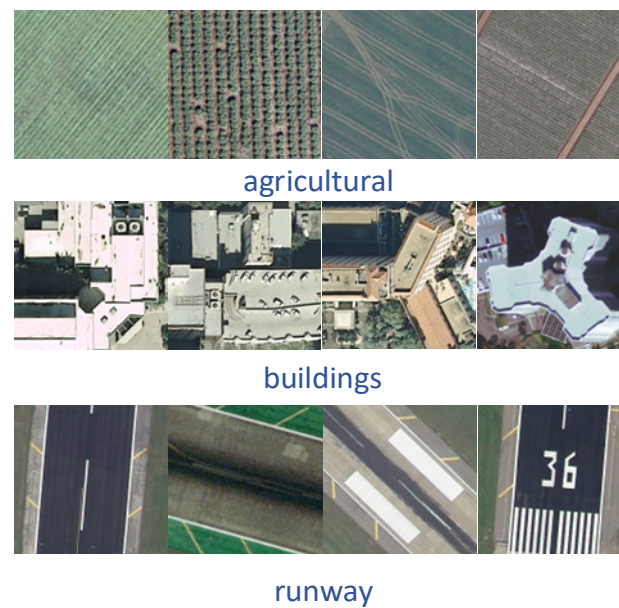
**Table 1.** Pixel-level benchmark datasets.

Dataset	Data Sources	Number	Image Dimensions	Spatial Resolution	Number of Bands	Number of Categories	Year
SPARCS [37]	Landsat8	80	1000 × 1000	30	10	7	2014
LoveDA [35]	GoogleEarth	5987	1024 × 1024	0.3	3	7	2021
Kennedy Space Center [34]	AVIRIS	1	614 × 512	18	224	13	2014
Pavia University [34]	ROSIS	1	610 × 340	1.3	103	9	2015
GID [38]	GF-2	150	6800 × 7200	1/4	4	15	2020
Indian Pines [33,34]	AVIRIS	1	145 × 145	20	224	16	2015
LandCoverNet [36]	Sentinel-1/2 Landsat-8	1980	256 × 256	10	10	7	2020

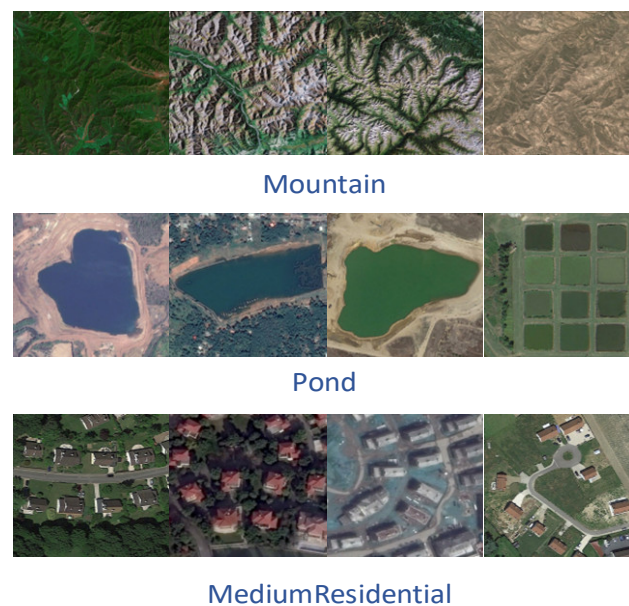
### 3.2. Patch-Level Datasets

Instead of labeling and categorizing each pixel, patch-level datasets are produced in image segments, i.e., entire blocks of images are labeled into a scene category. These image segments have a fixed size and location and thus are similar to remote sensing scene recognition datasets. Compared to the process of making pixel-level datasets, the production of patch-level datasets is much simpler, and it does not require a lot of labor to label the images. Therefore, the disadvantage becomes very obvious: such labeling is coarse and does not capture the details inside the image blocks. Four patch-level benchmark datasets for LULC classification are described below.

- The UC Merced [39] dataset was proposed by researchers at the University of California, Merced, in 2010 and contains 21 scene categories. Each scene category has 100 images with a region of  $256 \times 256$  pixels, where each pixel has a spatial resolution of 0.3 m. The 2100 images were collected from a variety of regions, including Los Angeles, Houston, and Miami, and cover a wide range of land-use types in the U.S. region. The vast majority of land use remote sensing classification experiments use the UC Merced dataset for data comparison, and again, it was one of the first datasets to be proposed. Most of the experiments divide the UC Merced dataset into an 80:20 or 50:50 ratio between training and test sets for algorithm evaluation (Figure 11).
- The AID [40] dataset was proposed in 2017, and was created by a research team from Wuhan University from images collected from Google Earth Images, covering many countries and regions around the world. The dataset contains 30 scene categories, each category contains 220 to 420 images, and the entire dataset has a total of 10,000 images, all of which are  $600 \times 600$  pixels in size and have a spatial resolution of 8 m to 0.5 m. Because the images of the AID dataset are collected from multiple sensors, the algorithm is usually evaluated by dividing the training set and test set into 20:80 and 50:50 ratios (Figure 12).
- The NWPU-RESISC45 [41] is a dataset proposed by a research team from Northwestern Polytechnical University in 2017, with 45 scene categories, each with 700 images, all of which have a size of  $256 \times 256$  pixels. Most of the images in this dataset have a spatial resolution of 30 m to 0.2 m, with lower spatial resolution for snow-covered mountain, lake, and island images. Because the NWPU-RESISC45 dataset contains a large number of scene categories, there exist some categories with high similarity between them, which makes it more challenging to develop new algorithms and methods (Figure 13).
- The EuroSAT [42] dataset is a large-scale dataset containing 27,000 images and was presented in 2019. It consists of images captured by the Sentinel-2 satellite and contains 13 spectral bands. A total of 10 scene categories are included, namely, cities, forests, farmland, grasslands, lakes, rivers, coasts, deserts, mountains, and industrial areas. Each category has 2000–3000 images, each with a region of  $64 \times 64$  pixels. When evaluating the algorithm, a ratio of 80:20 is usually chosen to divide the training and test sets (Figure 14).



**Figure 11.** Schematic diagram of the UC Merced.

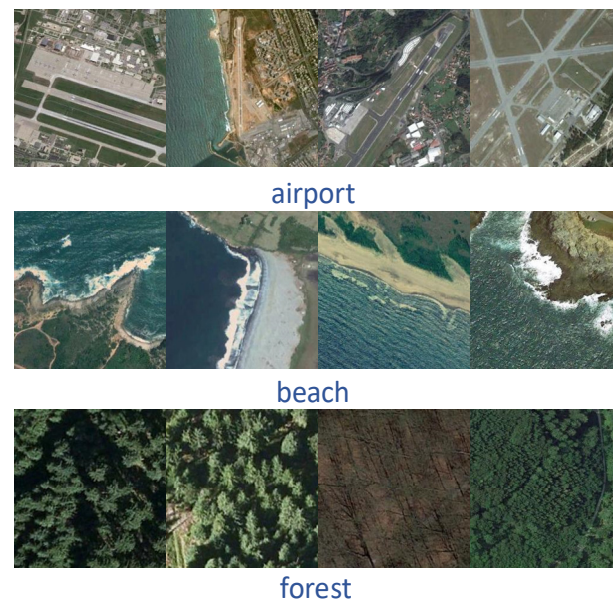


**Figure 12.** Schematic diagram of the AID.

Table 2 lists several common patch-level datasets. As can be seen from the table, most of these datasets are dominated by Sentinel-2 (e.g., BigEarthNet, EuroSAT) and Google Earth satellite imagery (AID, RSD46-WHU, etc.), with high spatial resolution, mostly 0.3–2 m. Of these, all contain only three spectral bands, except for two datasets released in 2019, BigEarthNet and EuroSAT, which have 13 spectral bands. These extra spectral bands can provide diverse data, enabling remote sensing images to provide detailed and comprehensive information on surface features. Therefore, a number of spectral bands greater than 3 may yield more accurate classification.

These datasets mentioned above are large in scale, covering many regions and scenes around the world, with rich data diversity. At the same time, the images in these datasets have high spatial resolution and can be capable of high-precision feature classification tasks. Despite the large size of these datasets, there is still the problem of the relatively limited amount of data compared to the complexity and diversity of reality. On the other hand, some datasets, e.g., EuroSAT, have too few feature categories, making it difficult for the

network to capture subtle differences between categories and reducing the generalization ability of the model.



**Figure 13.** Schematic diagram of the NWPU-RESISC45.



**Figure 14.** Schematic diagram of the EuroSAT.

**Table 2.** Patch-level benchmark datasets.

Dataset	Data Sources	Number	Image Dimensions	Spatial Resolution	Number of Bands	Number of Categories	Year
BigEarthNet [42,43]	Sentinel-2	590,326	120 × 120	10	13	43	2019
RESISC45 [41]	WorldView-2	31,500	256 × 256	2	3	45	2016
EuroSAT [42]	Sentinel-2	27,000	64 × 64	10/20/60	13	10	2019
AID [40]	Google Earth	10,000	600 × 600	2	3	30	2017
UC MERCED [39]	Aerial imagery	2100	256 × 256	0.3	3	21	2010
OPTIMAL-31 [44]	Google Earth	1860	256 × 256	-	3	31	2019
RSD46-WHU [45]	GoogleEarth, Tianditu	117,000	256 × 256	0.5–2	3	46	2017

In general, the pixel-level samples lack a large and high-quality dataset like the patch-level. Moreover, the areas covered are relatively homogeneous and only suitable for specific regions, and the generalizability needs to be improved. The patch-level dataset has fewer spectral bands, which limits the model's ability to extract spectral information from features, thus limiting the scope of the application. Therefore, some hyperspectral and multispectral data need to be included. Hyperspectral/multispectral images can provide a more accurate estimation of mixing ratios, which leads to more accurate mixing analysis results. On the other hand, hyperspectral/multispectral imagery also allows for better differentiation of spectral characteristics between different features. The reflectance of different features at different wavelengths has unique characteristics, and hyperspectral imagery can capture these subtle differences, thus providing a better ability to identify and classify substances. This is important for applications such as LULC classification, vegetation type identification, and water monitoring.

### 3.3. Performance Indicators

The main performance metrics commonly used to evaluate LULC classification models are overall accuracy (OA), average accuracy (AA), F1-score (F1), and mean intersection and parallel ratio (MIOU) [46].

- Overall accuracy (OA) is the most used performance metric in LULC classification to measure the rate at which the model correctly predicts the samples in the dataset, which is represented by Equation (2).

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

where TP denotes true positive, TN denotes true negative, FP denotes false positive, and FN denotes false negative.

- Average accuracy (AA) indicates the average of the rate of correct predictions by the model in each category sample, as represented by Equation (3).

$$AA = \frac{\sum_i (\text{recall})}{i} \quad (3)$$

where recall denotes the recall rate.

- F1-score (F1) is the average value after calculating the precision rate and recall rate reconciliation; the larger the F1, the better the model performance, represented by Equation (4).

$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

where precision denotes the rate of accuracy.

- Mean intersection and unity ratio (MIOU) represents the average of the prediction accuracies of all the categories in the dataset and is commonly used to evaluate semantic segmentation, As expressed by Equation (5).

$$MIOU = \frac{1}{k+1} \sum_{i=0}^k \frac{TP}{FN + FP + TP} \quad (5)$$

where k denotes the number of categories and k + 1 denotes the number of categories plus the background class.

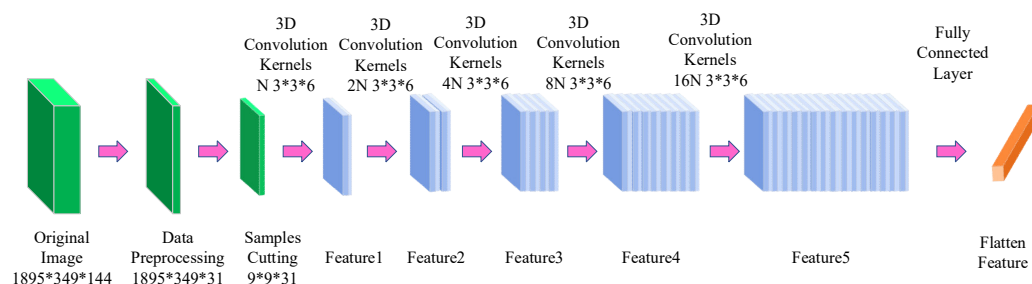
## 4. Deep-Learning-Based LULC Classification

### 4.1. Convolutional Neural Networks

CNNs can effectively capture textural, spatial, edge, and contextual features of feature classes in remote sensing images and can be adapted to remote sensing images at different scales, which makes them very suitable for processing LULC classification tasks. However,



the training of CNN relies on a large number of labeled samples, which is prone to overfitting and difficulty in dealing with some small-sized features in the case of scarce labeled samples. Temenos et al. [47] proposed a novel neural network framework that utilizes the SHAP deep interpreter to feed the results classified by the CNN to improve the accuracy of the final classification results. Pei [48] and others worked on the accurate classification of forest types and proposed a novel convolutional neural network, MSG-GCN. The network architecture is characterized by several features: First, to reduce the computational complexity, image features are extracted using multi-scale convolutional kernels incorporating different sensory fields. Second, to fully express the multi-scale, as well as the edge features, the authors use the Multi-scale Graph Convolutional Network (MSGCN) as a transition module. Again, to avoid the use of redundant information, the skip connection in U-Net++ is replaced with local attention. Finally, fusion is performed on the decoding layer, which synthesizes the high-level and low-level feature information and improves the accuracy of the model classification. The authors conducted comparative experiments with MSG-GCN with some state-of-the-art semantic segmentation methods, such as U-Net [49], FCN [50], and U-Net++ [51]. The results show that MSG-GCN can achieve 85.23% of the highest OA and 78.08% of the highest Kappa. Also, it can achieve 43.74% of the IoU in natural mixed forests, which is 9.34% and 10.07% higher than that of U-Net and U-Net++, and it can draw accurate maps to facilitate the management of forests. Ma [52] and others focused on developing spatial information for airborne hyperspectral data, for which a model called 3D-1D-CNN was proposed. The model can well cope with the challenges of feature extraction in complex urban areas, especially those affected by cloud shadows. Specifically, all the parameters extracted from the hyperspectral data are fused and partitioned into many stereo blocks to be passed to the 3D-CNN classifier, which contains texture features, spectral composition parameters, and vegetation indices, and their fusion provides rich information for feature extraction, enabling the model to extract features of complex urban areas more accurately. In the comparison experiments of methods such as 3D-1D-CNN, 1D-CNN, and 3D-CNN (see Figure 15), it was found that 3D-1D-CNN achieved the highest OA of 96.32%.

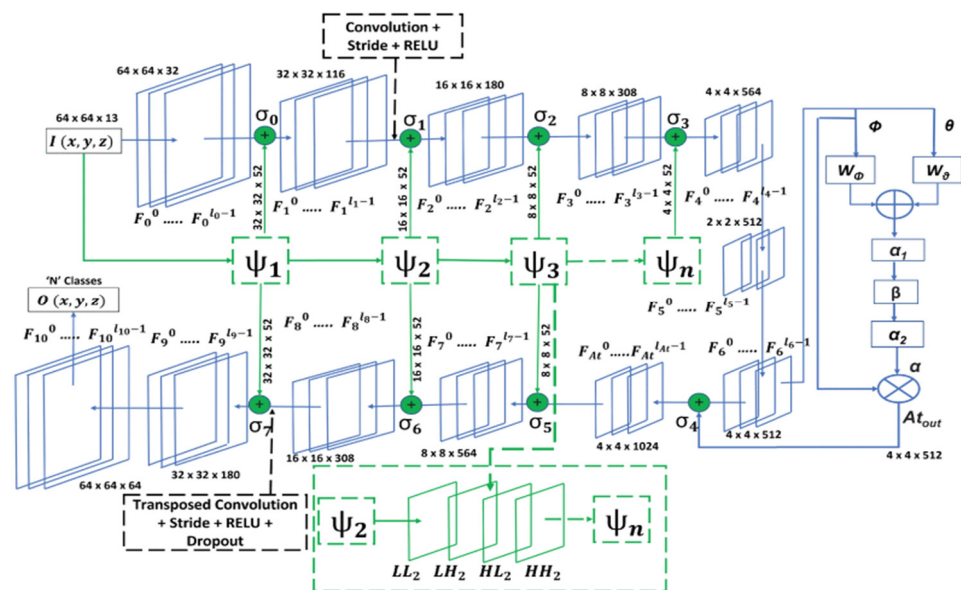


**Figure 15.** 3D-CNN architecture diagram.

For complex scenarios, Khan et al. [53] proposed a multi-branch framework consisting of two deep learning networks, DenseNet [22] and FCN. The DenseNet network is used to extract the contextual features of the input image, and FCN is used to extract the local features of the input image. Finally, the features extracted by the two deep learning networks are fused and optimized with loss function. The network framework achieves the highest OA of 99.52%, 96.37%, and 94.75% on the UC Merced [39], Siri-WHU [54], and EuroSAT [42] datasets. Compared to methods such as GoogleNet, ResNet-50, MOEA, and ResNet-101, the authors' method achieves higher accuracy and superior performance. Xia et al. [55] proposed a deep learning model based on unsupervised domain adaptation (UDA), which is capable of adapting to large-scale land cover mapping. Specifically, it solves the deep convolutional neural network (DCNN) biased mislabeling problem by using Siamese networks [56], and combines dynamic pseudo-label assignment and class balancing strategies, thus realizing adaptive domain joint learning. The authors compared their method with several UDA methods applicable to large-scale cities, such as



AdaptSeg [57], AdvEnt [58], CLAN [59], and FADA [60], and after incorporating dynamic pseudo-label assignments, respectively, their method achieves the highest OA under sparse samples with 81.14%, 51.20%, 40.81%, and 40.81% of the OA, mF1, and MIOU. The highest OA, mF1, and MIOU of 80.85%, 55.33, and 43.99 were achieved with dense samples. Wang [61] and his team focused on research on crop classification, and they found that the existing crop classification methods could not be applied to multi-scenario challenges, so they proposed a deep learning method called Cropformer, which is based on a new type of transformer and combines it with convolutional neural networks. The method differs from existing crop classification methods in that it accomplishes both global and local feature extraction. The research team utilized Cropformer for grouping crop experiments in multiple scenarios. Experimental results show that Cropformer is more accurate and efficient than existing crop classification methods. It can extract additional information from unlabeled data to build up a priori knowledge for later crop classification. This a priori knowledge makes Cropformer better at handling the task of feature classification and more widely applicable. Singh et al. [62] developed an attention-based convolutional neural network (WIANet) architecture (see Figure 16) that overcomes the lack of spectral context modeling in CNNs when processing multispectral images. The authors introduced the Wavelet Transform (WT) [63], as well as an attention mechanism to the U-Net-based architecture to improve the network's ability to discriminate spectral features, with the aim of more accurately distinguishing between classes whose spectral features have a high degree of similarity. Their methods achieved 92.43%, 91.80%, and 84.69% OA on the EuroSAT [42], NaSC-TG2 [64], and PatternNet [65] datasets, respectively, and achieved F1 on the EuroSAT dataset, which is higher than methods such as EfficientNet-B1 [66], RegNetX120 [67], RegNetY120 [67], and others.



**Figure 16.** WIANet network architecture.

#### 4.2. Generating Adversarial Networks

GAN is widely used for land cover classification in the case of sparse samples due to its unique adversarial mechanism that generates realistic images to act as samples and to increase the training samples of the network, thus improving the accuracy of classification. GAN can generate images for categories with fewer sample data, solving the problems of low classification accuracy and overfitting in some categories caused by unbalanced sample data. In addition, the images generated by GAN are cleaner, reducing the negative impact of noise in the original data. Ansith et al. [68] changed the input to the generator in the GAN by using an encoder to convert all remote sensing images in the dataset into potential vectors, replacing the original potential vectors generated from the network noise

signal. This GAN structure, combined with an encoder, recognizes more details and thus achieves higher accuracy even when the samples are sparse. The authors' method achieved 97.68% of the highest OA and 97.43% of the F1 on the AID dataset, and 95.68% of the OA and 95.38% of the F1 on the UC Merced dataset. Ma [69] and others proposed a generative adversarial network called SPG-GAN, which aims to allow GAN to generate samples with category labels and thus achieve higher classification accuracy. They trained the network in the direction of labeling information by including label information on the inputs of the generator and the discriminator. Moreover, the generator sample module is added to the generator and discriminator iterations, respectively, so that the final labeled samples generated are of higher quality and rich in more features and spatial details. SPG-GAN achieves the highest OA of 79.88% and 65.13% on the UC Merced and AID datasets, respectively. Miao et al. [70] introduced an additional convolution (involution) layer on the generator and discriminator layers of the GAN, respectively, to strengthen the feature extraction capability of the network. To narrow the gap between labeled and unlabeled samples, the authors introduced a consistency loss function on the Siamese network that performs semi-supervised classification. The method achieved the highest OA of 94.25% and 92.03% on the AID and RESICS-45 datasets, respectively. Xu et al. [71] proposed a new type of generative adversarial network called Lie Group (as shown in Figure 17) to cope with problems such as the decline in the learning ability of deep learning models due to the scarcity of samples. Like the method proposed by Miao et al. [70], the inputs of this network model are also composed of category information and data samples such that the newly generated data samples contain category information. The difference is that they introduced an object-scale sample generation strategy to ensure that the new data samples contain more feature information. Experiments show that the network still achieves highly accurate classification even with sparse samples. Because the scale of the receptive field of the FCN network is fixed, it is difficult for the model to judge objects that are much smaller than the current receptive field scale, thus causing misclassification. For this reason, Wang et al. [72] proposed a GAN that combines a multi-scale receptive field, which improves the performance of the model in the case of fewer labeled samples by introducing a space pyramid pooling (ASPP) module to obtain the multi-scale features and by utilizing semi-supervised training of the GAN, which alleviates a large amount of labeling and improves the performance of the model in the case of fewer labeled samples. This improved GAN achieves 75.25%, 78.66%, and 80.71% MIOU on the CCF2015 dataset with 1/8, 1/4, and 1/2 the amount of data, respectively, higher than several state-of-the-art semantic segmentation methods compared with it.

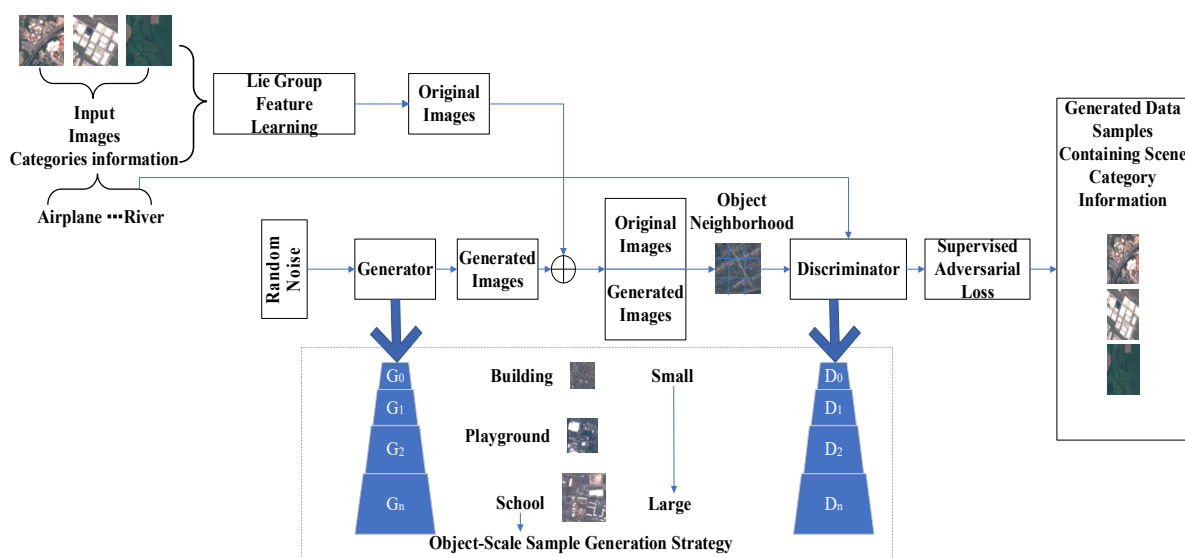


Figure 17. Novel Lie Group generative adversarial learning network.

#### 4.3. Recurrent Neural Networks

Although convolutional neural networks are most common in LULC, RNNs are better in special cases, especially when dealing with time-series data, where they can accurately capture the sequence information. The RNN is hidden in each time step due to its special hidden state, and this hidden state is propagative and can be passed from the previous time step to the next, whereby the RNN can capture rich contextual information. In addition, RNN can adaptively extract features when training data, eliminating the complicated process of designing a feature extractor. Ma et al. [25] incorporated Long Short-Term Memory (LSTM) into the RNN network to improve the efficiency of RNN computing similarity. Specifically, the method of object segmentation makes the network search for the target pixel from the whole map to the segmented map, then filters out the line segments with similar features, and finally extracts similar pixels from these line segments, which greatly shortens the time of similarity calculation. Also, for target pixels, the study considers both local and non-local segments, bringing more spatial information to the network. Tang et al. [73] proposed a recurrent neural network called DRRNN, which learns iteratively by utilizing class correlation features (CCFs) [74], and at each iteration, CCFs use spatial information to correct misclassification, in this way learning repeatedly until they reach an optimal state to discontinue the iteration. Tao [75] and others combined CNN and RNN to propose a network called SIC-Net, which allows the network to utilize the spatial information of remote sensing images to mine more discriminative features. SIC-Net is divided into two parts, the first part is used to convert inter-patch context information into inter-pixel context information, and the second part is used to model inter-pixel context information and focuses on the extraction of local and remote spatial features. This method has a great advantage in scene classification in the case of spectral confusion. Sohail et al. [76] also considered combining RNN with CNN and proposed a new neural network called MulNet. It is composed of three parts: the first part learns multi-scale local spectral spatial features of hyperspectral images by introducing a three-dimensional residual network (3DResNet), the second part employs a Feature Fusion Module (FFM) to aggregate these spectral spatial features that have been sampled at different ratios, and the third part of the RNN generates discriminative features based on these fused features as a way to achieve more accurate classification. MulNet achieved the highest OA of 91.45% and 87.15% on the Pavia University and Salinas datasets, respectively. Zhang et al. [77] integrated the non-local spatial sequence (NLSS) method into RNN and proposed a network called NLSS-RNN. The network is divided into three parts, the first part extracts low-level features from hyperspectral images. The second part extracts Local Spatial Sequence (LSS) features from low-level features using the NLSS method and preserves NLSS information. In the third part, the LSS features containing NLSS information are used as inputs to the RNN network as a way to generate high-level features. Finally, the advanced features are fed into the classifier for classification. This network that generates advanced features from low to high achieves the highest OA of 98.75%, 99.77%, and 97.23% on the Indian Pines, University of Pavia, and Salinas datasets, respectively, which is better than some of the state-of-the-art classification networks.

#### 4.4. Autoencoder

Autoencoders have excellent downscaling and are, therefore, often used to process hyperspectral data in LULC. AE has a unique encoding–decoding mechanism to mine the potential features in the original data, as well as to reconstruct the original data, which helps the model to learn the data at a deeper level to generate discriminative features with higher discriminative properties, thus realizing more accurate classification. As an unsupervised classification method, AE is also often used in the case of limited labeled samples, which can learn feature characteristics of features on unlabeled samples and then use the unsupervised pre-training results for supervised classification, effectively alleviating the problem of sample scarcity. Ibañez et al. [78] proposed a new spectral converter (MAEST) to alleviate the state of transformer networks where it is difficult to

properly characterize continuous spectral features due to the factor of noise. MAEST is composed of a reconstruction path and a classification path. The role of the reconstruction path is to reconstruct the original image using the AE. The encoder in the AE is used to extract the potential features of the unmasked spectra, and the decoder uses the potential features to reconstruct the masked spectral information. At the same time, the encoder can extract the most robust spectral features after this process to eliminate the negative effects of noise. The role of the classification path is to integrate the features obtained from the reconstruction path onto the transformer network and utilize the more robust features for more accurate classification. MAEST achieved the highest OA of 84.15%, 91.06%, and 88.55% on the Indian, Pavia University, and Houston 2013 datasets, respectively. Chen et al. [79] proposed a novel stacked autoencoder for processing Polarized Synthetic Aperture Radar (PolSAR) images, called MDPL-PolSAR. Specifically, a multilayer projective dictionary pair learning (MDPL) was developed to extract high-dimensional features from PolSAR, and then these high-dimensional features were used as inputs to the SAEs, which were adapted and learned by the SAEs to acquire the features with high discriminative properties, thus realizing more accurate classification. Also, for PolSAR images, Liu et al. [80] developed a novel autoencoder SAE\_MOEA/D, which is a combination of the Multi-Objective Evolutionary (MOEA) algorithm developed by the authors and a variant of SAE. MOEA can adaptively mine the optimal parameters, eliminating the time-consuming step of human exploration of the optimal parameters. Variants of SAE can adaptively construct the number of network layers according to the dataset to accommodate different PolSAR images. The method is robust and largely eliminates time-consuming and laborious work. Mughees et al. [81] proposed a novel classification method based on the spectral space of hyperspectral images (HSI). The method utilizes SAE as a classifier to extract the spectral features of HSI to obtain deeper spectral features. The space is then segmented using the Hidden Markov Random Field (HMRF) technique to obtain spatial features. Finally, the spectral and spatial features are fused to achieve more accurate classification. The method achieved 98.72% and 90.08% OA on the Pavia University and Indian Pine datasets, respectively. Mughees et al. [82] also improved the SAE and proposed a novel network called HVSAE. This network can adaptively segment the spatial structure of HSI, and similar spectral information is preserved in the segmented structure. These segmented spatial structures are then processed using SAE to obtain more discriminative features based on the similar spectral features in them. HVSAE achieved 90.08% and 98.98% OA on Indian Pine and Pavia University datasets, respectively. Chen et al. [83] developed a novel dimensionality reduction algorithm called SPCA, which allows the network to effectively extract spatial features in HSI. The authors used two spatial neighborhoods, one large and one small, to extract the spatial information; the large spatial neighborhood ensured sufficient spatial information, while in the small spatial neighborhood, spatial information was effectively utilized to reduce misclassification. The spectral-spatial features extracted by SPCA were then classified by SAE. The method achieved 99.27% OA on the Pavia University dataset. Mughees et al. [84] proposed a novel HSI classification method. The method consists of three parts. The first part utilizes SAE to extract deep spectral features of HSI. The second part utilizes the boundary leveling technique to segment the HSI and extracts the spatial features of the HSI in the process. The third part uses the majority voting (MV) strategy to fuse the spectral and spatial features before classifying them. The method achieves 89.35% and 98.64% OA on Pavia University and Indian Pine datasets, respectively.

## 5. LULC Challenges

1. Data diversity: Remote sensing data may change according to different times, different regions, different sensors, different climates, etc., so our model needs to be constantly updated to adapt to the variability caused by data diversity.
2. Category imbalance: When making LULC datasets, some categories will have too many or too few data samples due to factors such as differences in geographical

distribution, human intervention, and labeling difficulties. Therefore, the model tends to favor the categories with more data when training these data, while the categories with fewer samples lead to poorer classification accuracy due to too few training samples.

3. **Sample labeling problem:** The labeling of LULC training samples not only requires researchers to have a high level of expertise and be familiar with the characteristics of the vegetation, urban, and water categories, it also requires a lot of manpower and time to classify the category areas and classifications. Some features are highly similar to each other, e.g., swamps and wetlands, and researchers may make subjective errors in judgment that lead to relatively fuzzy boundary delineation between categories, thus reducing the classification accuracy of the model.
4. **Generalization ability of the model:** A model may achieve high classification accuracy in a specific region; however, if it is changed to a region or area in other countries, the generalization ability of the model will be reduced due to different distribution of features, differences in the labeling of the samples, and differences in the characteristics of the categories, among other factors.
5. **Contextual modeling:** Remote sensing data are spatially correlated, i.e., there are dependencies between data in nearby areas. However, traditional deep learning methods do not mine the contextual information between data well, making the model unable to learn and understand the data comprehensively, which reduces the classification accuracy of the model.

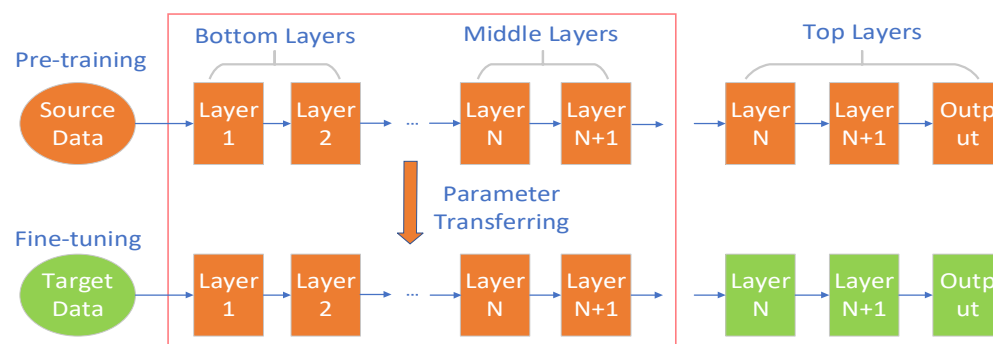
## 6. LULC Classification with Limited Samples

The problem of limited training samples is one of the main challenges faced in LULC classification, and in this section, we will discuss the solutions to this challenge in terms of four methods: transfer learning, data augmentation, weakly supervised learning, and active learning.

### 6.1. Transfer Learning

Transfer learning (TL) is the application of knowledge learned in the old domain (source domain) to the new domain (target domain), which compensates for the negative impact of sample scarcity and improves the generalization ability of the model, leading to more accurate classification. Therefore, TL is often used to deal with LULC under limited samples. Domain adaptation (DA), as a subfield in TL, is even more favored by researchers because it can solve the discrepancy problem between the source and target domains. Liu et al. [85] proposed a DA-based SFnet-DA network to solve the problem of insufficient labeled samples in practical applications. Specifically, an adversarial training consisting of domain classifiers and feature extractors was designed so that the parameters obtained from training the labeled dataset can be directly used for the prediction of unlabeled images. SFnet-DA achieves 97.13% F1 on the WHDLD dataset, which is higher and faster to process than excellent semantic segmentation networks such as U-Net and Deeplab V3+. Traditional DA methods make it difficult to balance the critical classes in the target domain, which may result in some classes in the target domain lacking labeled samples for training, so an unsupervised DA method was proposed by Soto et al. [86]. Specifically, a pseudo-label generation scheme called CVA was designed, with which the feature alignment of the target and source domains is balanced to avoid over-biasing to one category. Bai et al. [87] proposed a novel model that enables DA to directly align features in the source and target domains, which contains two DA branches and a semantic segmentation network. The two DA branches, one for adapting the representation space and one for adapting the prediction results, create feedback on each other to obtain more accurate classification results. Figure 18 illustrates the transfer learning.

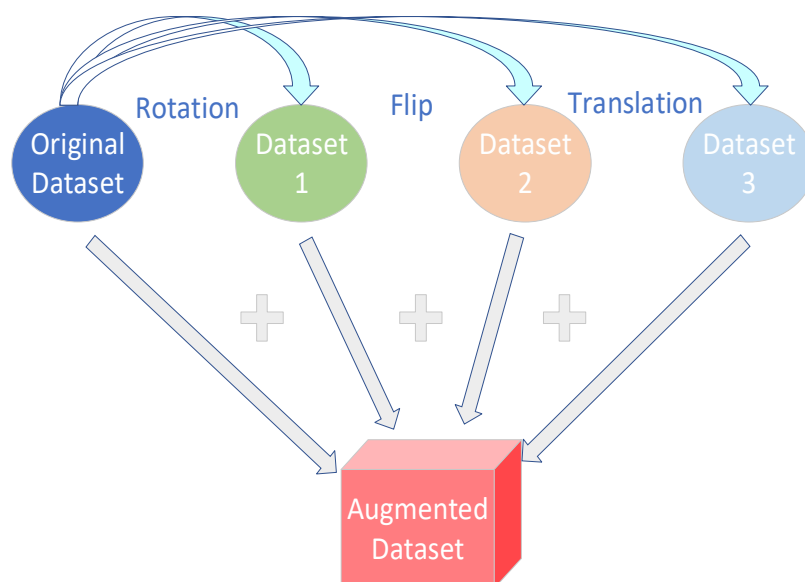




**Figure 18.** Transfer learning illustration.

### 6.2. Data Augmentation

Data augmentation is one of the important means of solving the finite sample problem, which involves various transformations such as rotating, scaling, and cropping of the original data as a way of generating new training samples and increasing the training volume of the model. In addition, data augmentation makes the training samples more diverse and improves the generalization ability of the model. Scott et al. [88] trained a DCNN model by augmenting and expanding the dataset with various transpositions and rotations of the training samples to obtain an upright view of each category. Experiments showed a significant improvement in classification accuracy using the data-augmented dataset. Yi et al. [89] proposed a novel data augmentation method that optimizes the network model by adding an attention mechanism to the GAN, replacing the loss function and adding a convolutional kernel, which enhances the GAN's ability to learn global features, improves the stability of the model, and results in a higher quality of samples generated by the GAN. Figure 19 illustrates the data augmentation.

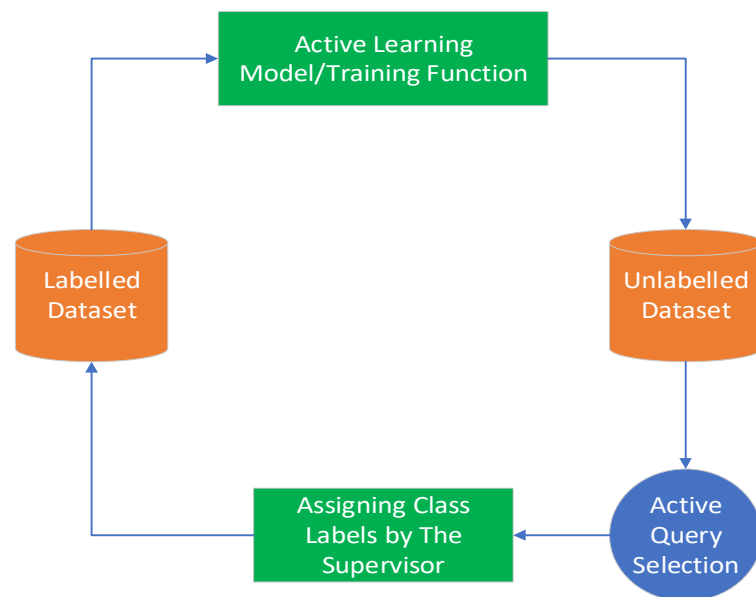


**Figure 19.** Data augmentation illustration.

### 6.3. Active Learning

As a kind of iterative learning, active learning (AL) adaptively selects higher quality samples from unlabeled samples to train the model and improve the performance of the model and iterates by selecting informative and low redundancy samples from them for labeling, thus reducing the need for training samples. Haut et al. [90] proposed a new model that is composed of Bayesian convolutional neural network (B-CNN) and AL. B-CNN can be adapted to train with fewer samples and extracts useful data from CNNs

with different architectures, preventing the generation of overfitting and improving the robustness of the network. One of the roles of AL is to use the probability function to analyze those unlabeled samples and extract a lot of useful information for the model. Lei et al. [91] proposed a deep network method incorporating AL to overcome the problem of the limited number of training samples due to the complex and tedious labeling work of HSI. The network extracts uncertainty information from unlabeled samples, which is then incorporated into the original training samples for augmentation purposes and constitutes a new dataset. Experiments show that the method performs well on several public datasets. Figure 20 illustrates active learning.



**Figure 20.** Active learning illustration.

#### 6.4. Weak Supervision Methods

Traditional deep learning methods need to rely on a large number of training samples to achieve accurate classification. However, the labeling of remote sensing images is time-consuming and labor-intensive, and the limited training samples make it difficult for the model to capture deep and complex relationships. Therefore, weakly supervised learning, which is less dependent on labeled samples, becomes one of the most effective methods to solve the sample scarcity problem. In the following sections, we will explore the use of zero-shot learning and few-shot learning methods to deal with the limited sample problem in LULC.

##### 6.4.1. Zero-Shot Learning

Zero-shot learning (ZSL) allows knowledge migration, i.e., learning semantic information from visible categories and transferring it to invisible categories, thus reducing the need for labeled samples. Li et al. [92] proposed a novel zero-shot method called ZSSC, which constructs a guided graph through semantic vectors to sort out the connections between visible and invisible categories. A label propagation algorithm incorporating unsupervised domain adaptation was developed for migrating the knowledge learned from visible categories. The results show that this method is significantly better than other ZSL-based methods. Li et al. [93] proposed an end-to-end ZSL approach that bridges the class structure gap between the semantic space and the visual image space, allowing the model to efficiently utilize semantic representations to classify visual samples. An interpretable constraint was designed to improve the robustness of the network and make the network learning more stable. After a large number of experiments, it was shown that the method can effectively solve the problem caused by the scarcity of labeled samples.

#### 6.4.2. Few-Shot Learning

Few-shot learning aims to utilize a very small number of training samples to achieve effective training and fast learning. Chen [94] and others worked on how to improve the accuracy of scene classification in the presence of sample scarcity. To this end, they proposed a few-shot classification method called DN4AM. Specifically, the influence of semantically irrelevant objects is weakened by using the attention graph associated with each category, which makes the model more focused on classification-related information and improves the accuracy of model classification. The authors used the NWPU-RESISC45, UC Merced, and WHU-RS19 datasets to evaluate the performance of DN4AM (Figure 21). The results show that DN4AM demonstrates excellent performance and achieves high-precision classification even when the labeled samples are scarce. Jiang et al. [95] proposed a graph-based few-shot learning method that is composed of two models: feature learning and feature fusion. The feature learning model is used to transform the extracted multi-scale features into graph-based features, allowing the model to better utilize the spatial relationships between remote sensing images. Feature fusion models are used to integrate graph-based multi-scale features to obtain more semantic information as a way to achieve more accurate classification in few-shot conditions.

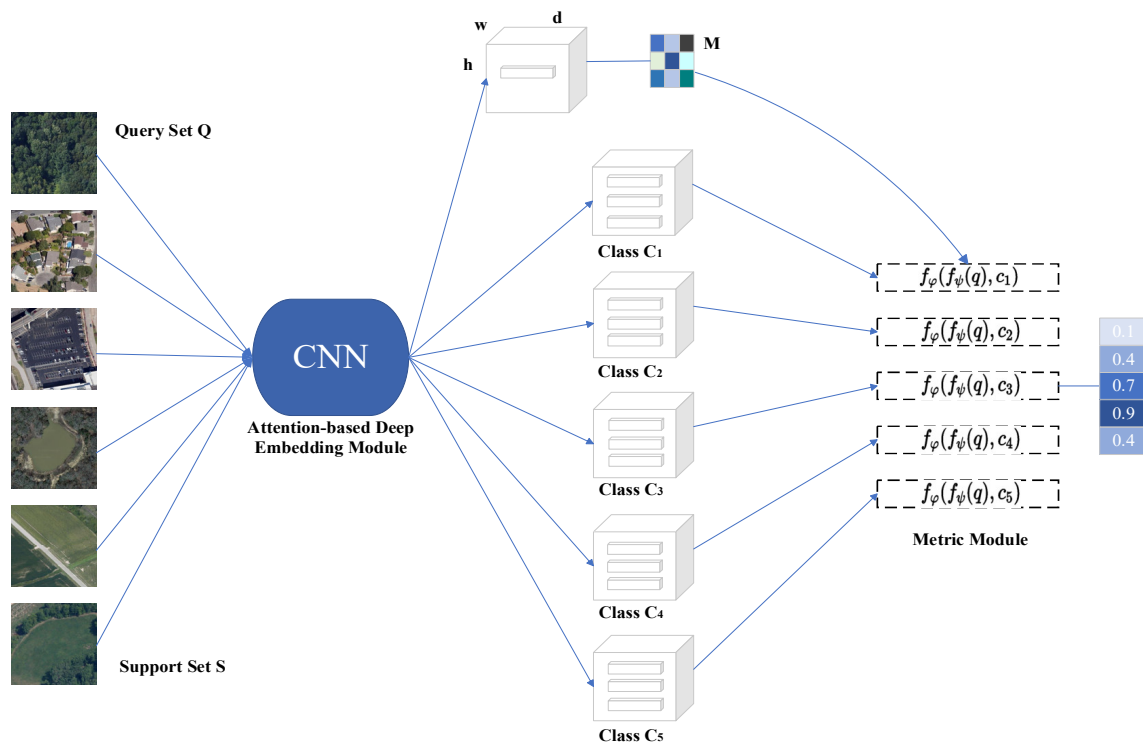


Figure 21. DN4AM architecture.

#### 7. Prospects

Combined with the most advanced deep learning technologies and innovations in recent years, LULC classification will face the following important trends:

- Cross-modal data fusion: With the rapid development of deep learning methods in the field of LULC classification, more research will be conducted in the future to fuse effective information from multiple modalities (e.g., texture, spectral, and temporal information, etc.) of remote sensing imagery and combine them with deep learning models to improve the accuracy and robustness of the network model classification.
- High-resolution data: The resolution of remote sensing image data will increase in the future, which puts more requirements on LULC classification. Therefore, one of the focuses of future research is to develop better algorithms and models to adapt

to high-resolution data to capture the details and changes in feature characteristics more accurately.

- **Expert knowledge:** Although deep learning and other technologies have achieved excellent results in LULC classification, it is still very important to integrate the experience and knowledge of human experts. It can not only be converted into an interpretable form of the model to improve the explanation ability of the classification results but also correct the misclassification of the model to achieve higher robustness and accuracy, which is in line with the needs of practical applications.
- **Transfer learning and adaptive learning:** The combination of the two can help solve the problem of low generalization ability of models due to domain differences. Future research can explore how to narrow the gap between the source and target domains due to transfer learning and optimize the model using adaptive learning to enable higher adaptation on the target domain for higher classification performance and generalization ability.

In the future, LULC classification will continuously improve the generalization ability and robustness of the model, as well as the accuracy and explainability of the classification with the help of cross-modal data fusion, high-resolution data, transfer learning and adaptive learning, among other technical means. This will provide more accurate and reliable information support for resource management, environmental protection, and sustainable development.

## 8. Conclusions

LULC classification plays an important role in urban and rural land planning, disaster prediction, monitoring of environmental change, and sustainable development. Early LULC classification methods based on low-level feature extraction unfolding can achieve good results on small-scale datasets, yet have limited results for large-scale as well as complex datasets. Therefore, as the variety and number of LULC sample datasets continue to expand, deep-learning-based classification methods have stepped onto the stage, bringing automation, adaptability, and accuracy to the LULC task, allowing us to better learn and manage surface resources. In this paper, we briefly introduce some commonly used deep learning models, including; CNN, RNN, AE, GAN, and FCN. After that, we systematically summarize the common public datasets commonly used for LULC classification in terms of samples and categorize them as pixel-level datasets and patch-level datasets, respectively. We also summarize the evaluation criteria used in LULC classification. In addition, we overview the excellent performance demonstrated by deep learning methods in LULC classification and briefly clarify the advantages and disadvantages of each method. Despite the fact that DL is very effective in handling LULC classification tasks, there are some limitations and challenges faced by LULC classification that prevent DL from realizing its full potential. Thus, in this paper, we summarize the challenges faced by LULC classification, elaborate on the treatment scheme in the case of scarcity of labeled samples, and finally point out the corresponding future directions to face these challenges.

**Author Contributions:** Conceptualization, S.Z. and C.X.; methodology, S.Z.; software, S.Z. and K.T.; validation, K.T. and S.Y.; formal analysis, S.Y. and H.T.; investigation, H.T. and K.T.; resources, C.X.; data curation, Y.H.; writing—original draft preparation, S.Z.; writing—review and editing, S.Z. and C.X.; visualization, Y.H.; supervision, C.X.; project administration, C.X.; funding acquisition, C.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded in part by the National Natural Science Foundation of China under Grants 61901221 and 62203012, in part by the Postgraduate Research and Practice Innovation Program of Jiangsu Province under Grant SJCX23\_0322, in part by the Nanjing Forestry University College Student Practice and Innovation Training Program under Grant 2023NFUSPITP0077, in part by the State Visiting Scholar Program of China Scholarship Council under Grant 202208320239, and in part by the National Key Research and Development Program of China under Grant 2019YFD1100404.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The training data presented in the study are openly available at <https://aistudio.baidu.com/datasetoverview>.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Chen, B.; Xu, B.; Zhu, Z.; Yuan, C.; Suen, H.P.; Guo, J.; Xu, N.; Li, W.; Zhao, Y.; Yang, J. Stable classification with limited sample: Transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. *Sci. Bull.* **2019**, *64*, 3.
- Zhao, C. Advances of research and application in remote sensing for agriculture. *Nongye Jixie Xuebao Trans. Chin. Soc. Agric. Mach.* **2014**, *45*, 277–293.
- Schmitt, M.; Hughes, L.H.; Qiu, C.; Zhu, X.X. SEN12MS—A Curated Dataset of Georeferenced Multi-Spectral Sentinel-1/2 Imagery for Deep Learning and Data Fusion. *arXiv* **2019**, arXiv:1906.07789. [[CrossRef](#)]
- Li, Y.; Xia, H.; Liu, Y.; Ji, K.; Huo, L.; Ni, C. Research on Morphological Indicator Extraction Method of Pinus massoniana Lamb. Based on 3D Reconstruction. *Forests* **2023**, *14*, 1726. [[CrossRef](#)]
- Feng, Q.; Liu, J.; Gong, J. UAV remote sensing for urban vegetation mapping using random forest and texture analysis. *Remote Sens.* **2015**, *7*, 1074–1094.
- Feng, Q.; Liu, J.; Gong, J. Urban flood mapping based on unmanned aerial vehicle remote sensing and random forest classifier—A case of Yuyao, China. *Water* **2015**, *7*, 1437–1455. [[CrossRef](#)]
- Gong, W.; Lian, J.; Zhu, Y. Capacitive flexible haptic sensor based on micro-cylindrical structure dielectric layer and its decoupling study. *Measurement* **2023**, *223*, 113785. [[CrossRef](#)]
- Chapelle, O.; Vapnik, V.; Bousquet, O.; Mukherjee, S. Choosing multiple parameters for support vector machines. *Mach. Learn.* **2002**, *46*, 131–159. [[CrossRef](#)]
- Xie, C.; Zhu, H.; Fei, Y. Deep coordinate attention network for single image super-resolution. *IET Image Process.* **2022**, *16*, 273–284. [[CrossRef](#)]
- Adam, E.; Mutanga, O.; Odindi, J.; Abdel-Rahman, E.M. Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: Evaluating the performance of random forest and support vector machines classifiers. *Int. J. Remote Sens.* **2014**, *35*, 3440–3458. [[CrossRef](#)]
- Zhang, L.; Zhang, L.; Du, B. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 22–40. [[CrossRef](#)]
- Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.-S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [[CrossRef](#)]
- Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [[CrossRef](#)]
- Deren, L.; Liangpei, Z.; Guisong, X. Automatic analysis and mining of remote sensing big data. *Acta Geod. Et Cartogr. Sin.* **2014**, *43*, 1211.
- LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
- Wu, K.; Jia, Z.; Duan, Q. The Detection of Kiwifruit Sunscald Using Spectral Reflectance Data Combined with Machine Learning and CNNs. *Agronomy* **2023**, *13*, 2137. [[CrossRef](#)]
- Chen, K.; Li, W.; Chen, J.; Zou, Z.; Shi, Z. Resolution-agnostic remote sensing scene classification with implicit neural representations. *IEEE Geosci. Remote Sens. Lett.* **2022**, *20*, 6000305. [[CrossRef](#)]
- Li, B.; Wang, Q.-W.; Liang, J.-H.; Zhu, E.-Z.; Zhou, R.-Q. SquconvNet: Deep Sequencer Convolutional Network for Hyperspectral Image Classification. *Remote Sens.* **2023**, *15*, 983. [[CrossRef](#)]
- Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
- Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
- He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
- Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
- Mou, L.; Ghamisi, P.; Zhu, X.X. Deep recurrent neural networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3639–3655. [[CrossRef](#)]
- Hang, R.; Liu, Q.; Hong, D.; Ghamisi, P. Cascaded recurrent neural networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5384–5394. [[CrossRef](#)]
- Ma, A.; Filippi, A.M.; Wang, Z.; Yin, Z. Hyperspectral image classification using similarity measurements-based deep recurrent neural networks. *Remote Sens.* **2019**, *11*, 194. [[CrossRef](#)]
- Fan, X.; Chen, L.; Xu, X.; Yan, C.; Fan, J.; Li, X. Land Cover Classification of Remote Sensing Images Based on Hierarchical Convolutional Recurrent Neural Network. *Forests* **2023**, *14*, 1881. [[CrossRef](#)]



27. Zhao, L.; Ji, S. CNN, RNN, or ViT? An Evaluation of Different Deep Learning Architectures for Spatio-Temporal Representation of Sentinel Time Series. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *16*, 44–56. [\[CrossRef\]](#)
28. Sun, C.; Zhang, X.; Meng, H.; Cao, X.; Zhang, J.; Jiao, L. Dual-Branch Spectral-Spatial Adversarial Representation Learning for Hyperspectral Image Classification with Few Labeled Samples. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 1–15. [\[CrossRef\]](#)
29. Dieste, Á.G.; Argüello, F.; Heras, D.B. ResBaGAN: A Residual Balancing GAN with Data Augmentation for Forest Mapping. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 6428–6447. [\[CrossRef\]](#)
30. Jaiswal, G.; Rani, R.; Mangotra, H.; Sharma, A. Integration of hyperspectral imaging and autoencoders: Benefits, applications, hyperparameter tuning and challenges. *Comput. Sci. Rev.* **2023**, *50*, 100584. [\[CrossRef\]](#)
31. Xu, C.; Gao, M.; Yan, J.; Jin, Y.; Yang, G.; Wu, W. MP-Net: An efficient and precise multi-layer pyramid crop classification network for remote sensing images. *Comput. Electron. Agric.* **2023**, *212*, 108065. [\[CrossRef\]](#)
32. Yang, J.; Du, B.; Zhang, L. From center to surrounding: An interactive learning framework for hyperspectral image classification. *ISPRS J. Photogramm. Remote Sens.* **2023**, *197*, 145–166. [\[CrossRef\]](#)
33. Baumgardner, M.F.; Biehl, L.L.; Landgrebe, D.A. 220 band aviris hyperspectral image data set: June 12, 1992 indian pine test site 3. *Purdue Univ. Res. Repos.* **2015**, *10*, 991.
34. del Pais Vasco, U. Available online: <http://www.ehu.es/ccwintco/index.php/Hyperspectral-Remote-Sensing-Scenes> (accessed on 25 August 2012).
35. Wang, J.; Zheng, Z.; Ma, A.; Lu, X.; Zhong, Y. LoveDA: A remote sensing land-cover dataset for domain adaptive semantic segmentation. *arXiv* **2021**, arXiv:2110.08733.
36. Alemohammad, H.; Booth, K. LandCoverNet: A global benchmark land cover classification training dataset. *arXiv* **2020**, arXiv:2012.03111.
37. Hughes, M.J.; Hayes, D.J. Automated detection of cloud and cloud shadow in single-date Landsat imagery using neural networks and spatial post-processing. *Remote Sens.* **2014**, *6*, 4907–4926. [\[CrossRef\]](#)
38. Tong, X.-Y.; Xia, G.-S.; Lu, Q.; Shen, H.; Li, S.; You, S.; Zhang, L. Land-cover classification with high-resolution remote sensing images using transferable deep models. *Remote Sens. Environ.* **2020**, *237*, 111322. [\[CrossRef\]](#)
39. Yang, Y.; Newsam, S. Bag-of-visual-words and spatial extensions for land-use classification. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, San Jose, CA, USA, 2–5 November 2010; pp. 270–279.
40. Xia, G.-S.; Hu, J.; Hu, F.; Shi, B.; Bai, X.; Zhong, Y.; Zhang, L.; Lu, X. AID: A benchmark data set for performance evaluation of aerial scene classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3965–3981. [\[CrossRef\]](#)
41. Cheng, G.; Han, J.; Lu, X. Remote sensing image scene classification: Benchmark and state of the art. *Proc. IEEE* **2017**, *105*, 1865–1883. [\[CrossRef\]](#)
42. Helber, P.; Bischke, B.; Dengel, A.; Borth, D. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 2217–2226. [\[CrossRef\]](#)
43. Sumbul, G.; Charfuelan, M.; Demir, B.; Markl, V. Bigearthnet: A large-scale benchmark archive for remote sensing image understanding. In Proceedings of the IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium, San Jose, CA, USA, 2–5 November 2019; pp. 5901–5904.
44. Alhichri, H.; Alswayed, A.S.; Bazi, Y.; Ammour, N.; Alajlan, N.A. Classification of remote sensing images using EfficientNet-B3 CNN model with attention. *IEEE Access* **2021**, *9*, 14078–14094. [\[CrossRef\]](#)
45. Zhang, P.; Bai, Y.; Wang, D.; Bai, B.; Li, Y. A meta-learning framework for few-shot classification of remote sensing scene. In Proceedings of the ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021; pp. 4590–4594.
46. Moharram, M.A.; Sundaram, D.M. Land Use and Land Cover Classification with Hyperspectral Data: A comprehensive review of methods, challenges and future directions. *Neurocomputing* **2023**, *536*, 90–113. [\[CrossRef\]](#)
47. Temenos, A.; Temenos, N.; Kaselimi, M.; Doulamis, A.; Doulamis, N. Interpretable deep learning framework for land use and land cover classification in remote sensing using SHAP. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 8500105. [\[CrossRef\]](#)
48. Pei, H.; Owari, T.; Tsuyuki, S.; Zhong, Y. Application of a Novel Multiscale Global Graph Convolutional Neural Network to Improve the Accuracy of Forest Type Classification Using Aerial Photographs. *Remote Sens.* **2023**, *15*, 1001. [\[CrossRef\]](#)
49. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Proceedings, Part III 18. pp. 234–241.
50. Shao, Z.; Zhou, W.; Deng, X.; Zhang, M.; Cheng, Q. Multilabel remote sensing image retrieval based on fully convolutional network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 318–328. [\[CrossRef\]](#)
51. Shao, Z.; Yang, K.; Zhou, W. Performance evaluation of single-label and multi-label remote sensing image retrieval using a dense labeling dataset. *Remote Sens.* **2018**, *10*, 964.
52. Ma, X.; Man, Q.; Yang, X.; Dong, P.; Yang, Z.; Wu, J.; Liu, C. Urban Feature Extraction within a Complex Urban Area with an Improved 3D-CNN Using Airborne Hyperspectral Data. *Remote Sens.* **2023**, *15*, 992. [\[CrossRef\]](#)
53. Khan, S.D.; Basalamah, S. Multi-Branch Deep Learning Framework for Land Scene Classification in Satellite Imagery. *Remote Sens.* **2023**, *15*, 3408.

54. Zhao, B.; Zhong, Y.; Xia, G.-S.; Zhang, L. Dirichlet-derived multiple topic scene classification model for high spatial resolution remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2015**, *54*, 2108–2123. [\[CrossRef\]](#)
55. Tong, X.-Y.; Xia, G.-S.; Zhu, X.X. Enabling country-scale land cover mapping with meter-resolution satellite imagery. *ISPRS J. Photogramm. Remote Sens.* **2023**, *196*, 178–196. [\[CrossRef\]](#)
56. Zagoruyko, S.; Komodakis, N. Learning to compare image patches via convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 4353–4361.
57. Tsai, Y.-H.; Hung, W.-C.; Schuster, S.; Sohn, K.; Yang, M.-H.; Chandraker, M. Learning to adapt structured output space for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7472–7481.
58. Vu, T.-H.; Jain, H.; Bucher, M.; Cord, M.; Pérez, P. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2517–2526.
59. Luo, Y.; Zheng, L.; Guan, T.; Yu, J.; Yang, Y. Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2507–2516.
60. Wang, H.; Shen, T.; Zhang, W.; Duan, L.-Y.; Mei, T. Classes matter: A fine-grained adversarial approach to cross-domain semantic segmentation. In Proceedings of the European Conference on Computer Vision, Online, 23–28 August 2020; pp. 642–659.
61. Wang, H.; Chang, W.; Yao, Y.; Yao, Z.; Zhao, Y.; Li, S.; Liu, Z.; Zhang, X. Cropformer: A new generalized deep learning classification approach for multi-scenario crop classification. *Front. Plant Sci.* **2023**, *14*, 1130659. [\[CrossRef\]](#)
62. Singh, A.; Bruzzone, L. WIANet: A Wavelet-Inspired Attention-Based Convolution Neural Network for Land Cover Classification. *IEEE Geosci. Remote Sens. Lett.* **2022**, *20*, 5000305. [\[CrossRef\]](#)
63. Mallat, S.G. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11*, 674–693. [\[CrossRef\]](#)
64. Zhou, Z.; Li, S.; Wu, W.; Guo, W.; Li, X.; Xia, G.; Zhao, Z. NaSC-TG2: Natural scene classification with Tiangong-2 remotely sensed imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 3228–3242. [\[CrossRef\]](#)
65. Zhou, W.; Newsam, S.; Li, C.; Shao, Z. PatternNet: A benchmark dataset for performance evaluation of remote sensing image retrieval. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 197–209. [\[CrossRef\]](#)
66. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 10–15 June 2019; pp. 6105–6114.
67. Radosavovic, I.; Kosaraju, R.P.; Girshick, R.; He, K.; Dollár, P. Designing network design spaces. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10428–10436.
68. Ansith, S.; Bini, A. Land use classification of high resolution remote sensing images using an encoder based modified GAN architecture. *Displays* **2022**, *74*, 102229.
69. Ma, A.; Yu, N.; Zheng, Z.; Zhong, Y.; Zhang, L. A supervised progressive growing generative adversarial network for remote sensing image scene classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5618818. [\[CrossRef\]](#)
70. Miao, W.; Geng, J.; Jiang, W. Semi-supervised remote-sensing image scene classification using representation consistency siamese network. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5616614. [\[CrossRef\]](#)
71. Xu, C.; Shu, J.; Zhu, G. Adversarial Remote Sensing Scene Classification Based on Lie Group Feature Learning. *Remote Sens.* **2023**, *15*, 914. [\[CrossRef\]](#)
72. Wang, J.; Liu, B.; Zhou, Y.; Zhao, J.; Xia, S.; Yang, Y.; Zhang, M.; Ming, L.M. Semisupervised multiscale generative adversarial network for semantic segmentation of remote sensing image. *IEEE Geosci. Remote Sens. Lett.* **2020**, *19*, 8003805. [\[CrossRef\]](#)
73. Tang, Y.; Qiu, F.; Wang, B.; Wu, D.; Jing, L.; Sun, Z. A deep relearning method based on the recurrent neural network for land cover classification. *GIScience Remote Sens.* **2022**, *59*, 1344–1366. [\[CrossRef\]](#)
74. Huang, X.; Lu, Q.; Zhang, L.; Plaza, A. New postprocessing methods for remote sensing image classification: A systematic study. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 7140–7159. [\[CrossRef\]](#)
75. Tao, C.; Lu, W.; Qi, J.; Wang, H. Spatial information considered network for scene classification. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 984–988. [\[CrossRef\]](#)
76. Sohail, M.; Chen, Z.; Yang, B.; Liu, G. Multiscale spectral-spatial feature learning for hyperspectral image classification. *Displays* **2022**, *74*, 102278. [\[CrossRef\]](#)
77. Zhang, X.; Sun, Y.; Jiang, K.; Li, C.; Jiao, L.; Zhou, H. Spatial sequential recurrent neural network for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 4141–4155. [\[CrossRef\]](#)
78. Ibanez, D.; Fernandez-Beltran, R.; Pla, F.; Yokoya, N. Masked auto-encoding spectral-spatial transformer for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5542614. [\[CrossRef\]](#)
79. Chen, Y.; Jiao, L.; Li, Y.; Zhao, J. Multilayer projective dictionary pair learning and sparse autoencoder for PolSAR image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 6683–6694. [\[CrossRef\]](#)
80. Liu, G.; Li, Y.; Jiao, L.; Chen, Y.; Shang, R. Multiobjective evolutionary algorithm assisted stacked autoencoder for PolSAR image classification. *Swarm Evol. Comput.* **2021**, *60*, 100794. [\[CrossRef\]](#)

81. Mughees, A.; Tao, L. Hyperspectral image classification based on deep auto-encoder and hidden Markov random field. In Proceedings of the 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Guilin, China, 29–31 July 2017; pp. 59–65.
82. Mughees, A.; Tao, L. Hyper-voxel based deep learning for hyperspectral image classification. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 840–844.
83. Chen, C.; Zhang, J.; Li, T.; Yan, Q.; Xun, L. Spectral and Multi-Spatial-Feature Based Deep Learning for Hyperspectral Remote Sensing Image Classification. In Proceedings of the 2018 IEEE International Conference on Real-time Computing and Robotics (RCAR), Kandima, Maldives, 1–5 August 2018; pp. 421–426.
84. Mughees, A.; Tao, L. Efficient deep auto-encoder learning for the classification of hyperspectral images. In Proceedings of the 2016 International Conference on Virtual Reality and Visualization (ICVRV), Hangzhou, China, 24–26 September 2016; pp. 44–51.
85. Liu, J.; Wang, Y. Water body extraction in remote sensing imagery using domain adaptation-based network embedding selective self-attention and multi-scale feature fusion. *Remote Sens.* **2022**, *14*, 3538. [\[CrossRef\]](#)
86. Soto, P.J.; Costa, G.A.; Feitosa, R.Q.; Ortega, M.X.; Bermudez, J.D.; Turnes, J.N. Domain-adversarial neural networks for deforestation detection in tropical forests. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 2504505. [\[CrossRef\]](#)
87. Bai, L.; Du, S.; Zhang, X.; Wang, H.; Liu, B.; Ouyang, S. Domain adaptation for remote sensing image semantic segmentation: An integrated approach of contrastive learning and adversarial learning. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5628313. [\[CrossRef\]](#)
88. Scott, G.J.; England, M.R.; Starns, W.A.; Marcum, R.A.; Davis, C.H. Training deep convolutional neural networks for land-cover classification of high-resolution imagery. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 549–553. [\[CrossRef\]](#)
89. Yi, Y.; You, Y.; Li, C.; Zhou, W. EFM-Net: An Essential Feature Mining Network for Target Fine-Grained Classification in Optical Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5606416. [\[CrossRef\]](#)
90. Haut, J.M.; Paoletti, M.E.; Plaza, J.; Li, J.; Plaza, A. Active learning with convolutional neural networks for hyperspectral image classification using a new Bayesian approach. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6440–6461. [\[CrossRef\]](#)
91. Lei, Z.; Zeng, Y.; Liu, P.; Su, X. Active deep learning for hyperspectral image classification with uncertainty learning. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 5502405. [\[CrossRef\]](#)
92. Li, A.; Lu, Z.; Wang, L.; Xiang, T.; Wen, J.-R. Zero-shot scene classification for high spatial resolution remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4157–4167. [\[CrossRef\]](#)
93. Li, Y.; Zhu, Z.; Yu, J.-G.; Zhang, Y. Learning deep cross-modal embedding networks for zero-shot remote sensing image scene classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 10590–10603. [\[CrossRef\]](#)
94. Chen, Y.; Li, Y.; Mao, H.; Chai, X.; Jiao, L. A Novel Deep Nearest Neighbor Neural Network for Few-Shot Remote Sensing Image Scene Classification. *Remote Sens.* **2023**, *15*, 666. [\[CrossRef\]](#)
95. Jiang, N.; Shi, H.; Geng, J. Multi-Scale Graph-Based Feature Fusion for Few-Shot Remote Sensing Image Scene Classification. *Remote Sens.* **2022**, *14*, 5550. [\[CrossRef\]](#)

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.