Viser Technology Pte. Ltd.
Mechanical Engineering Science
DOI: 10.33142/mes.v7i2.17492

Research Article Open Access

A Survey of Remote Sensing Image Segmentation Based on Deep Learning

Shibo SUN, Yunzuo ZHANG*

School of Information science and Technology, Shijiazhuang Tiedao University, Shijiazhuang, Hebei, 050043, China

***Corresponding Author:** Yunzuo ZHANG, E-mail: zhangyunzuo888@sina.com*

Abstract

Remote sensing image segmentation has a wide range of applications in land cover classification, urban building recognition, crop monitoring, and other fields. In recent years, with the booming development of deep learning, remote sensing image segmentation models based on deep learning have gradually emerged and produced a large number of scientific research achievements. This article is based on deep learning and reviews the latest achievements in remote sensing image segmentation, exploring future development directions. Firstly, the basic concepts, characteristics, classification, tasks, and commonly used datasets of remote sensing images are presented. Secondly, the segmentation models based on deep learning were classified and summarized, and the principles, characteristics, and applications of various models were presented. Then, the key technologies involved in deep learning remote sensing image segmentation were introduced. Finally, the future development direction and application prospects of remote sensing image segmentation were discussed. This article reviews the latest research achievements in remote sensing image segmentation from the perspective of deep learning, which can provide reference and inspiration for the research of remote sensing image segmentation.

Keywords: *Remote sensing image segmentation; Deep learning; Split tasks; Model classification; Key technology*

1 Introduction

Remote sensing image segmentation is the process of dividing remote sensing images into different regions, and it is an important basis for remote sensing image analysis. The quantity, quality and diversity of remote sensing images have been improved with the updating of remote sensing platforms and sensors, bringing more data and higher requirements for remote sensing image segmentation. However, remote sensing image segmentation also faces challenges such as data scarcity, high annotation cost, large scale variation, class imbalance, complex background, etc., which affect the performance and adaptability of remote sensing image segmentation. In order to overcome these challenges, improve the accuracy and efficiency of remote sensing image segmentation, deep learning, a technique of automatic feature learning, has been widely applied and developed in the field of remote sensing image segmentation. Deep learning can use large amounts of data and complex network structures to automatically learn the features and patterns of remote sensing images, achieve end-to-end training and inference, and be applicable to various segmentation tasks. This paper reviews the concept, task, dataset, model classification, technical characteristics, development direction and application prospect of remote sensing image segmentation from the perspective of deep learning, and provides a reference perspective for the research and application of remote sensing image segmentation.

Remote sensing images are images of the earth's surface observed from a distance by sensors mounted on remote sensing platforms such as aircraft or artificial satellites. Remote sensing images have characteristics such as high spatial resolution, high spectral resolution, and high spatiotemporal coverage, and can reflect the physical, chemical, biological and other information of the earth's surface. Image segmentation is the basis of many visual understanding systems, and remote sensing image segmentation applies image segmentation techniques to the field of remote sensing, achieving pixel-level classification of remote sensing images, which has important applications in fields such as environmental monitoring, urban planning, land resource utilization, etc. Remote sensing image segmentation, as a special image segmentation task, encounters the following three main problems in the research process:

The objects of interest in remote sensing images have large scale variations, ranging from a few pixels to thousands of pixels, which leads to the multi-scale problem, making it difficult to locate and identify the objects of interest.

Received on July 12, 2025; Accepted on August 20, 2025

The background in remote sensing images is more complex and diverse, due to the influence of factors such as terrain, landform, season, illumination, etc., there are large intra-class differences and small inter-class differences in the background, [1] which leads to low class separability, making it easy for the objects of interest to be confused with the background.

The foreground objects in remote sensing images occupy a relatively small proportion, compared to the target objects in natural images, the foreground objects in remote sensing images often only occupy a small part of the image, [2] which leads to the foreground-background imbalance problem, making it easy for the objects of interest to be ignored or occluded.

2 Remote Sensing Images and Remote Sensing Image Segmentation Tasks

In this section, we mainly introduce the types of remote sensing images and the common tasks and datasets of remote sensing image segmentation.

2.1 Common types of remote sensing images and their characteristics

According to the different ways of sensor mounting, imaging, and sensing electromagnetic waves, remote sensing images can be divided into various types, each of which has its own characteristics and application fields. Here are some common types of remote sensing images and their characteristics:

Hyperspectral remote sensing images: Hyperspectral remote sensing images refer to remote sensing images that are imaged by remote sensing sensors in continuous narrow bands, usually containing hundreds to thousands of bands. Due to the complex characteristics of hyperspectral data, the accurate classification of hyperspectral data is challenging for traditional machine learning methods. In addition, the spectral information captured by hyperspectral imaging has a nonlinear relationship with the materials it corresponds to. In recent years, deep learning has been recognized as a powerful feature extraction tool that can effectively solve nonlinear problems. Driven by these successful applications, deep learning has also been introduced to hyperspectral remote sensing image classification and has shown good performance 3. Figure 1 shows an example of a hyperspectral remote sensing image.[3,4]

Synthetic aperture radar remote sensing images: Synthetic aperture radar remote sensing images refer to remote sensing images that are imaged by electromagnetic waves in the microwave band, usually obtained by synthetic aperture radar sensors mounted on platforms such as satellites or aircraft. Synthetic aperture radar remote sensing images can reflect the geometric features, roughness, dielectric constant, etc. of the objects, and are suitable for monitoring and analysis of terrain, landform, earthquake, landslide, etc. The characteristics of synthetic aperture radar remote sensing images are that they are not affected by weather, illumination, etc., and can achieve all-weather, all-day observation 4, but they also have problems such as complex scattering mechanism, low image quality, interference fringes, etc. Figure 2 shows an example of a synthetic aperture radar remote sensing image.

Lidar remote sensing images: Lidar remote sensing images refer to remote sensing images that are imaged by electromagnetic waves in the laser band, usually obtained by lidar sensors mounted on platforms such as satellites or aircraft. Lidar remote sensing images can reflect the elevation, morphology, structure, etc. of the objects, and are suitable for monitoring and analysis of surface elevation, buildings, vegetation, aerosols, etc.[5] The characteristics of lidar remote sensing images are that they have high accuracy, high resolution, high sensitivity, etc., and can achieve three-dimensional reconstruction, change detection, target recognition, etc., but they also have problems such as high cost, large data volume, complex processing, etc. Figure 3 shows an example of a lidar remote sensing image.

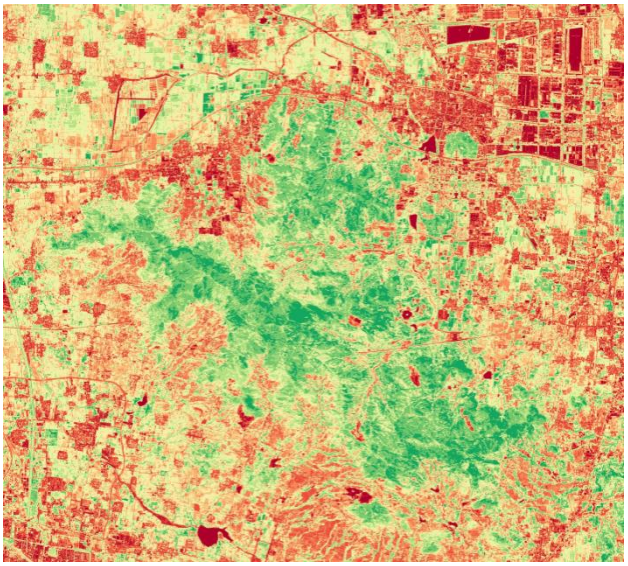

Figure 1 Xample of hyperspectral remote sensing images

Figure 2 Xample of Synthetic Aperture Radar Remote Sensing

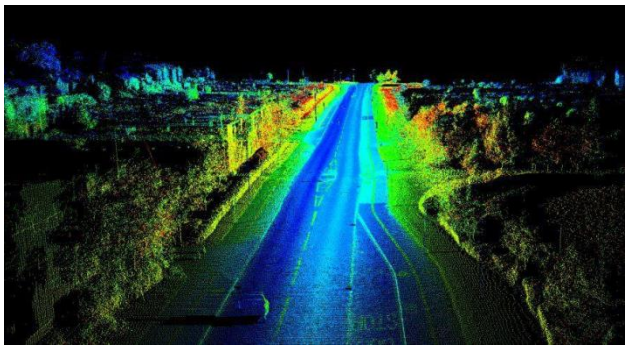

Figure 3 Example of LiDAR remote sensing images

2.2 Common tasks and data sets of remote sensing image segmentation

The common tasks and datasets of remote sensing image segmentation include:

Small object segmentation: Small objects occupy fewer pixels in remote sensing images, such as airplanes, vehicles, ships, etc. The difficulties of this task lie in the small size, irregular shape, complex background, and high similarity among the objects. [6,7]

Ship segmentation: Ship segmentation refers to the process of separating the ship objects from the water background in remote sensing images, which is the basis of maritime target monitoring. The difficulties of ship segmentation lie in the diverse shapes, different sizes, uneven colors, and low contrast with the water. [8]

Building segmentation: Building segmentation requires separating the building objects from the ground background in remote sensing images, which is an important means for urban planning and management. The difficulties of building segmentation lie in the complex shapes, different sizes, similar colors, and confusion with shadows and trees. [9]

Natural environment segmentation: Natural environment segmentation refers to the process of separating the natural environment objects from other backgrounds in remote sensing images, which have various types of objects, from clouds to vegetation to ice, with large differences. The difficulties of natural environment segmentation lie in the diverse classes, uneven distribution, fuzzy boundaries, and confusion with other objects. [10,11]

Road extraction: Road extraction refers to the process of separating the road objects from other backgrounds in remote sensing images, which is an important basis for traffic planning and management. The difficulties of road extraction lie in the different widths, complex shapes, uneven colors, and confusion with buildings and shadows. [12,13]

The commonly used datasets for remote sensing image segmentation include:

DOTA: DOTA is a large-scale aerial image dataset, containing 2806 images, with a total of 188282 small objects, divided into 15 categories, such as airplanes, ships, bridges, vehicles, etc. The images of DOTA come from Google Earth, with diverse perspectives, scales, backgrounds, and object densities.

SeaShips: SeaShips is a large-scale ship segmentation dataset, containing 31000 images, with a total of 59000 ship objects, divided into four categories, namely cargo ships, tankers, fishing boats, and speedboats. The images of SeaShips come from Planet Labs, with high resolution, multiple time phases, multiple angles, and multiple regions.

WHU Building Dataset: WHU Building Dataset is a high-resolution building segmentation dataset, containing 326 images, with more than 10000 building objects. The images of WHU Building Dataset come from QuickBird and WorldView-2, with different resolutions, angles, illuminations, and seasonal changes.

LandCoverNet: LandCoverNet is a global annual land cover classification dataset, containing 31000 images, each image is 256×256 pixels, divided into 7 categories, namely impervious surface, agriculture, forest, soil, water, wetland, and ice and snow.

The SpaceNet Datasets: The SpaceNet Datasets are a large-scale remote sensing image dataset, containing high-resolution satellite images and corresponding road network labels of multiple cities. The images of SpaceNet come from DigitalGlobe, with different resolutions, angles, illuminations, and seasonal changes. Table 1 lists the image sizes, categories, and numbers of each dataset.

Table 1 Introduction to Common Datasets and Their Content

Dataset name	Image size	Number of categories	Number of images
DOTA-v1.0	800～20 000	15	2806
DOTA-v1.5	800～20 000	16	2806
DOTA-2.0	800～20 000	18	11268
SeaShips	1920×1080	6	31455
WHU Building Dataset	800～20 000	1	2806
LandCoverNet	256×256	7	31000
The SpaceNet Datasets	900～1300	18	21346

3 Classification of Remote Sensing Segmentation Models Based on Deep Learning

In this section, we classify the segmentation models based on deep learning according to different network structures. The application fields and representative networks of each type of model are summarized in Table 2.

3.1 Segmentation models based on CNN

Convolutional neural networks (CNN) are neural networks that use convolution operations to extract image features, which have advantages such as translation invariance, parameter sharing, and sparse connection, and are suitable for processing image data. CNN networks were very popular in the early deep

learning models, and segmentation models based on CNN were one of the important breakthroughs of deep learning in the field of image segmentation, which can achieve end-to-end training and inference, and are applicable to various segmentation tasks. Common backbone networks such as VGG[14], ResNet[15], GoogLeNet[16] and others adopt the convolutional structure of CNN. In recent years, the rise of Tramsformer has brought new ideas and methods for image segmentation. Zhang[17] and others combined Tramsformer and CNN in their research and applied it to high-resolution remote sensing image semantic segmentation, and the method proposed was very close to the state-of-the-art methods in terms of overall accuracy.

3.2 Segmentation models based on GAN

Generative adversarial networks (GAN) are a technique that uses two competing neural networks to generate new data, where one network is called a generator, responsible for generating fake data, and the other network is called a discriminator, responsible for judging the authenticity of the data. Figure 4 shows the structure of the GAN network. Segmentation models based on GAN usually combine the traditional multi-class cross-entropy loss with the adversarial network, first pre-train the adversarial network, and then use the adversarial loss to fine-tune the segmentation network. Segmentation models based on GAN can use the generator's ability to enhance or reconstruct remote sensing images, thereby improving the quality and effect of segmentation. Tasar[18] and others proposed a color mapping-based method, using ColorMapGANs to transform the source domain remote sensing images into the target domain style, and then used a pre-trained segmentation network to segment the transformed images. The method achieved significant performance improvement on two high-resolution remote sensing datasets. In addition, GAN is also often used to solve the problem of domain transfer feature mismatch. Zhang[19] and others proposed a method of using GAN structure combined with cheap available data to train the model for the segmentation task under unsupervised conditions, which can effectively reduce the domain gap.

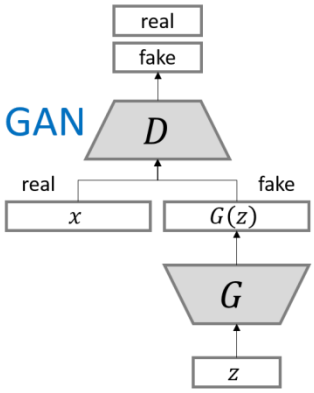

Figure 4 GAN network architecture diagram

3.3 Segmentation models based on Tramsformer

Tramsformer is a neural network that uses self-attention to achieve sequence-to-sequence mapping, which has advantages such as parallel computing, long-distance dependence, and position encoding. Segmentation models based on Tramsformer use Tramsformer to achieve image segmentation, which usually divide the input image into multiple sub-regions, and then use the encoder and decoder of Tramsformer to extract and reconstruct the features of each sub-region. This type of model can use the self-attention mechanism to perform global context understanding of remote sensing images, and improve the consistency and accuracy of segmentation. Figure 5 shows the structure of the Tramsformer network. Robin[20] and others proposed the Segmenter model based on Vision Tramsformer, which achieved excellent results in semantic segmentation. Liu[21] and others proposed a new visual Tramsformer, which can serve as a general backbone for computer vision, and its performance greatly exceeded the previous state-of-the-art techniques. He[22] and others embedded Swin-Tramsformer into the UNet network to form a new dual-encoder structure.

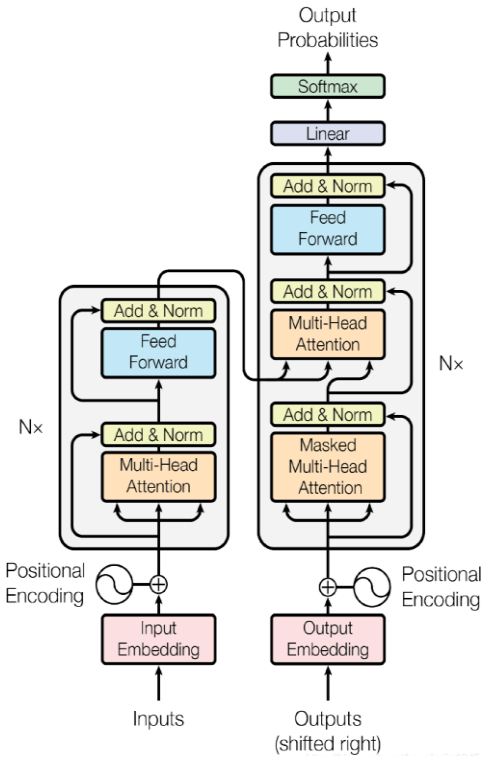

Figure 5 Tramsformer network architecture diagram

3.4 Segmentation models based on pre-trained models

Segmentation models based on pre-trained models are an important way to alleviate the problem of large-scale labeled data scarcity, which usually use pre-trained models as encoders or feature extractors, and then add a decoder or segmentation head after them. This

type of segmentation model can use the ability of pre-trained models to effectively extract features from remote sensing images, thereby improving the efficiency and effect of segmentation. Early pre-trained models include VGG, ResNet, DenseNet[23], etc., which have relatively few parameters. Li[24] and others used ResNeXt-101 instead of ResNet as the backbone in their model, enhancing the feature extraction ability. With the continuous improvement of data collection technology, pre-trained models based on large-scale datasets have been applied to the remote sensing field. ViT-G12[25] is one of the pre-trained models with a considerable amount of parameters. ViT-G12 was trained on the Million-AID dataset, with a parameter amount of 2.4B, and has a strong feature extraction ability.

4 Key Technologies for Remote Sensing Image Segmentation

The In this section, we summarize the key technologies in remote sensing image segmentation. Table 3 lists the representative models and the datasets they use based on the following six key technologies. As mentioned in Section 2, remote sensing image types and segmentation tasks are diverse, so different models use different key technologies. For example, small object segmentation tasks tend to pay more attention to semantic information, and long-distance dependence is particularly important when dealing with such tasks; while feature fusion operations can preserve spatial details and better complete boundary recognition. Remote sensing image segmentation key technologies can be classified according to different focuses, and this section will introduce the following six types:

Attention mechanism: Attention mechanism can make the network automatically focus on the most important parts when processing images, thereby improving the accuracy and robustness of segmentation. Attention mechanism is widely used in segmentation tasks, which can be divided into spatial attention and channel attention, the former focuses on different positions in the image, and the latter focuses on different features in the image. Figures 6 and 7 show the module structures of spatial attention and channel attention. Segmentation techniques based on attention mechanism can solve the problems of target size difference, complex background, occlusion, etc. in remote sensing images, and improve the details and boundaries of segmentation. Lei Ding[35] and others used adaptive attention mechanism in their research, which bridged the gap between high-level and low-level features, and kept the spatial details and semantic information during the feature fusion process. Although attention mechanism can improve the accuracy and robustness of segmentation, it also increases the computation and parameter amount of the network, resulting in slower training and inference speed of the network. In addition, attention mechanism may also over-focus on some specific areas and ignore other information, leading to overfitting problems.

Table 2 Application fields and representative models of various segmentation models

Category	Common areas of application	Representative model
CNN-based segmentation model	Feature classification, land cover monitoring, urban planning, etc	ConvNeXtV2[26] RepLKNet[27]
GAN-based segmentation model	Image enhancement, image reconstruction, image style conversion, etc	Pix2Pix[28] SPGAN-DA[29] PU-GAN[30]
Segmentation model based on Tramsformer	Remote sensing scene understanding, remote sensing target detection, remote sensing video analysis, etc	Conv2Former[31] Swin-Tramsformer[32]
Based on pre-trained segmentation models	Zero-shot segmentation of remote sensing images, multi-source fusion of remote sensing images, and cross-domain migration of remote sensing images	CMID[33] DGCC-EB[34]

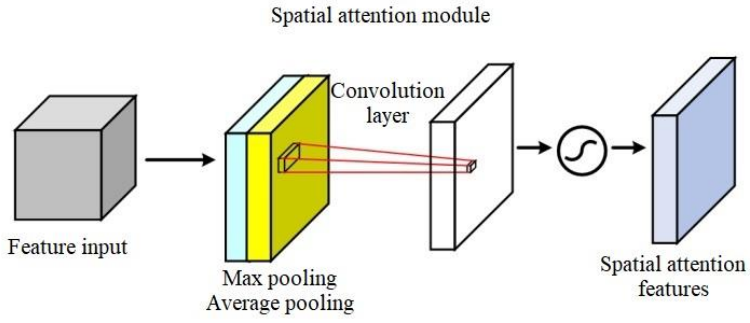

Figure 6 Spatial attention module

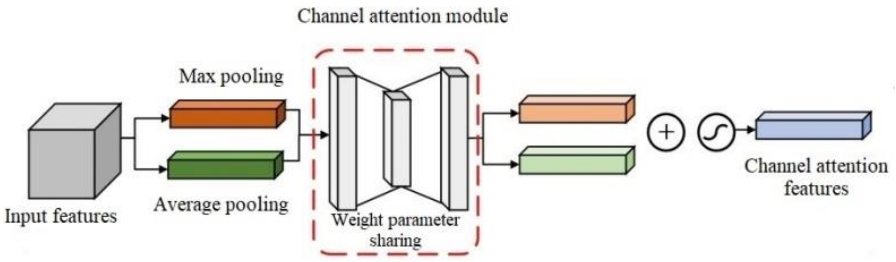

Figure 7 Channel attention module

Dilated convolution: Dilated convolution is a convolution operation that introduces dilation in the convolution kernel, which can increase the receptive field of the convolution kernel without increasing the parameters and computation, and capture more context information. Segmentation techniques based on dilated convolution can solve the problems of target detail loss, resolution reduction, etc. in remote sensing images, and improve the accuracy and clarity of segmentation. Figure 8 shows the dilated convolution model with dilation rates of 1, 2, and 4. Common segmentation techniques based on dilated convolution models include DeepLab[36], DenseASPP[37], etc. Zhao[38] and others used dilated convolution to enlarge the receptive field, and at the same time reduced the number of downsampling layers as much as possible to prevent the loss of detail information. Although dilated convolution plays a huge role in enhancing the receptive field, the size and stride of dilated convolution, the distribution and arrangement of dilated convolution, etc. are also problems that need to be solved. In addition, dilated convolution itself may also cause uneven receptive field of the network, and jagged phenomenon of the boundary.

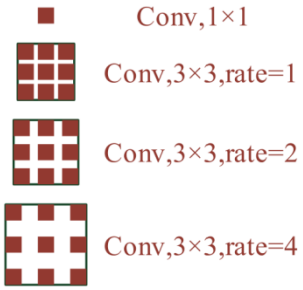

Figure 8 Dilated convolution with different magnifications

Feature fusion: Feature fusion uses different levels of feature information for effective fusion and utilization, thereby improving the details and boundaries of segmentation. Feature fusion techniques can solve the problems of target detail loss, resolution reduction, boundary blur, etc. in remote sensing images, and improve the accuracy of segmentation. There are three common ways of feature fusion, namely addition, concatenation, and fusion based on attention mechanism. Figure 9 shows the fusion process of high-level features and low-level features under the channel attention weight matrix. Common segmentation models that use feature fusion techniques include U-Net[39], RefineNet[40], etc. In this process, the way of feature fusion is very important, and inappropriate fusion methods may lead to noise introduction[41] Peng[42] and others proposed a cross-fusion module in their model, which used high-level feature maps to compensate for the receptive field of low-level feature maps, making the low-level feature maps have a similar receptive field to the high-level feature maps, thus significantly improving the semantic information capture ability of the low-level feature maps.

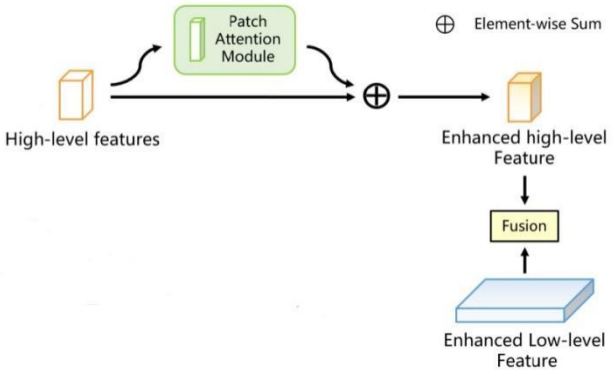

Figure 9 Fusion based on attention mechanism

Multi-task learning: Multi-task learning performs multiple segmentation tasks simultaneously, such as semantic segmentation, instance segmentation, boundary segmentation, etc., by sharing features and optimizing objectives, thereby improving the consistency and accuracy of segmentation. Multi-task learning segmentation techniques can solve the problems of target diversity, class imbalance, semantic ambiguity, etc. in remote sensing images, and improve the robustness and generalization of segmentation. For example, boundary segmentation and object segmentation tasks can promote each other[43] Li[44] and others used adaptive weight mechanism for multi-task learning, and separated the boundary information from the semantic features, and then used the corresponding loss to supervise.

Encoder-decoder structure: Encoder-decoder structure is a common structure in deep learning models, which uses two sub-networks, encoder and decoder, to perform feature extraction and feature reconstruction, respectively, to achieve pixel-level prediction of the

input image. Encoder-decoder structure segmentation technique is one of the important breakthroughs of deep learning in the field of image segmentation, which can achieve end-to-end training and inference, and is applicable to various segmentation tasks. Common segmentation techniques based on encoder-decoder structure models include U-Net, SegNet[45], etc. Liu[46] and others used encoder-decoder structure as the backbone network of their model.

Contrastive learning: Contrastive learning is a technique that uses positive and negative sample pairs in images to learn the feature representation of images by maximizing the similarity of positive sample pairs and minimizing the similarity of negative sample pairs. Figure 10 shows the feature extraction based on the existing support image to help the unknown image segmentation. Contrastive learning can solve the problems of target annotation insufficiency, unsupervised learning, self-supervised learning, etc. in remote sensing images to some extent, and improve the scalability and adaptability of segmentation. Tang[47] and others used the contrastive learning method of the self-supervised learning framework to improve the model's representation ability at the local pixel level.

5 Prospect

In this section, we mainly discuss the future development direction and application prospect of remote sensing image segmentation from a technical perspective.

5.1 Future development direction of remote sensing image segmentation

Remote sensing image segmentation has important applications in land use classification, urban planning, environmental monitoring, resource management, and other fields. The progress of remote sensing image segmentation technology has improved the accuracy and efficiency of remote sensing image segmentation. However, remote sensing image segmentation also faces problems such as data scarcity, high annotation cost, large scale variation, class imbalance, complex background, etc., which limit the performance and generalization ability of remote sensing image segmentation. To solve these problems, future research directions can be explored from the following three aspects:

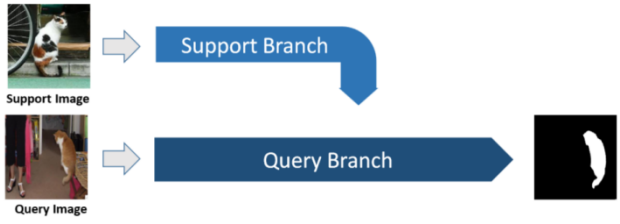

Figure 20 Contrastive learning

Table 3 Common key technologies and their representative models

Key technology	Applicable scenarios	Network structure	Training dataset
Attention mechanisms	Differences in target size, complex background, occlusion, etc.	DANet[48]	Cityscapes
		SENet[49]	ImageNet
		AGNet[50]	HDR+ Burst Photography Dataset
Dilated convolution	Loss of target detail, reduced resolution, etc	DeepLab	PASCAL VOC
		DenseASPP	Cityscapes
		DPC	COCO
Feature fusion	Loss of target detail, reduced resolution, blurred boundaries, etc	U-Net	Cell Tracking Challenge Dataset
		PSPNet[51]	ADE20K
		FPN[52]	COCO
Multi-task learning	Goal diversity, category imbalance, semantic ambiguity, etc	Mask R-CNN[53]	COCO
		FarSeg[54]	BraTS 2018
		DSSNet[55]	PASCAL VOC
Encoder-decoder structure	Various segmentation tasks	SegNet	CamVid
		DeconvNet[56]	PASCAL VOC
		FCN	PASCAL VOC
Contrastive learning	Insufficient target labeling, unsupervised learning, self-supervised learning, etc	MoCo[57]	ImageNet
		SimCLR[58]	ImageNet
		BYOL[59]	ImageNet

Multi-source data fusion: Remote sensing images contain multiple types of data, such as optical data, radar data, hyperspectral data, infrared data, etc., and different types of data have different features and advantages. By fusing different types of data, the robustness and accuracy of remote sensing image segmentation can be improved by using their complementarity. Multi-source data fusion methods include feature-level fusion and decision-level fusion, the former is to fuse different types of data at the feature extraction stage, and the latter is to fuse different types of data at the segmentation result stage. The key problem of multi-source data fusion is how to design effective fusion strategies, balance the differences and consistency between different data, and how to deal with the incompleteness of data.

Low-sample and zero-sample learning: One of the difficulties of remote sensing image segmentation is class imbalance, because the land cover classes in remote sensing images have diversity and complexity, and some classes have few samples, or even do not appear in the training data, which makes it difficult for the model to learn the features of these classes, affecting the accuracy of segmentation. Facing this problem, low-sample and zero-sample learning are promising research directions.

Low-sample learning uses a small number of samples to learn, and improves the generalization ability of the model by using transfer learning. Zero-sample learning uses auxiliary information, such as semantic information or attribute information, to learn, and establishes the association between classes, to achieve the recognition of unseen classes. The key problem of low-sample and zero-sample learning is how to design effective feature extraction and feature matching methods, use information from different sources, and how to deal with the differences and similarities between classes.

Interpretability and reliability: One of the objectives of remote sensing image segmentation is to provide reliable information support for remote sensing applications, therefore, the results of remote sensing image segmentation require not only accuracy, but also interpretability and reliability. However, the wide application of deep learning makes the remote sensing image segmentation model more and more complex, making the internal mechanism and output results of the model difficult to understand and verify, which may cause users' distrust. To improve the interpretability and reliability of remote sensing image segmentation, research needs to be conducted from three aspects: model, data, and result. Model aspect, it is necessary to design interpretable model structure, or provide model visualization and explanation methods, to reveal the working principle and key factors of the model. Data aspect, it is necessary to provide data quality assessment and uncertainty analysis, to evaluate the reliability and applicability of data. Result aspect, it is necessary to provide result confidence assessment and error detection, to evaluate the credibility and risk of the result.

5.2 Application prospect of remote sensing image segmentation

With the development of remote sensing technology, the application level of remote sensing image segmentation will also be continuously improved, providing more intelligent remote sensing services for various industries. The following are some application scenarios of remote sensing image segmentation:

Land cover change detection: Land cover refers to the natural and artificial coverings on the earth's surface, such as water, vegetation, bare soil, buildings, etc., which is an important indicator of the physical, chemical and biological processes on the earth's surface. The change of land cover will affect climate change, ecosystem, resource utilization, urban development, and other aspects, therefore, monitoring the change of land cover is of great significance for understanding and managing the earth environment. Remote sensing image segmentation can segment remote sensing images of different time phases, extract the categories and ranges of land cover, and then compare and analyze the changes of different categories, thus realizing the detection of land cover change.

Urban planning and management: Remote sensing image segmentation can finely extract and classify the elements of the city, such as buildings, roads, green spaces, water bodies, etc., to obtain the information of the city's spatial structure, functional distribution, ecological environment, etc., and provide scientific basis and reference for urban planning and management. Remote sensing image segmentation can also monitor and evaluate the changes of the city dynamically, and analyze the characteristics of the city's expansion, renewal, density, morphology, etc., and provide guidance and suggestions for the optimization and adjustment of the city.

Disaster monitoring and assessment: Remote sensing image segmentation can segment the remote sensing images before and after the disaster, and extract the information of the disaster type, range, degree, impact, etc., thus realizing the rapid identification, quantitative assessment and loss analysis of the disaster. Remote sensing image segmentation can also track and predict the evolution process of the disaster, and analyze the dynamic changes, development trends and potential risks of the disaster, and provide support and basis for disaster early warning and prevention.

Ecological environment protection: Remote sensing image segmentation can segment the various elements of the ecological environment, such as vegetation, soil, water, etc., and extract the information of the ecological environment structure, function, quality, service, etc., thus obtaining the information of the ecological environment status, change, problem, etc., and provide scientific data and methods for ecological environment protection. Remote sensing image segmentation can also monitor and evaluate the restoration and improvement of the ecological environment, and analyze the restoration effect, improvement measures, optimization schemes, etc., and provide decision basis and suggestions for ecological environment protection.

6 Conclusion

In the past decade, deep learning has achieved explosive development, greatly stimulating the research and application of deep learning in remote sensing image segmentation tasks. This paper introduces some basic concepts of remote sensing image segmentation, as well as its common tasks and datasets, and reviews the research status of remote sensing image segmentation models and techniques based on deep learning. Finally, several possible directions for future research are discussed. We hope that this research can provide valuable insights for researchers and inspire researchers to make more progress.

References

[1] L. Huang, B. Jiang. Deep Learning-based Semantic

Segmentation of Remote Sensing Images: A Survey [J]. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 2023:1-28.
[2] Q. An, Z. Pan. DRBox-v2: An Improved Detector With Rotatable Boxes for Target Detection in SAR Images [J]. IEEE Trans. Geosci. Remote Sens., 2019, 57(11): 8333-8349.
[3] S. Li, W. Song. Deep Learning for Hyperspectral Image Classification: An Overview [J]. IEEE Trans. Geosci. Remote Sens., 2019, 57(9): 6690-6709.
[4] R. Shang, M. Liu. Region-Level SAR Image Segmentation Based on Edge Feature and Label Assistance [J]. IEEE Trans. Geosci. Remote Sens., 202260: 1–16.
[5] F. Gaudfrin, O. Pujol. A New Lidar Technique Based on Supercontinuum Laser Sources for Aerosol Soundings: Simulations and Measurements—The PERFALIS Code and the COLIBRIS Instrument [J]. IEEE Trans. Geosci. Remote Sens., 2023, 61: 1-21.
[6] A. Ma, J. Wang. FactSeg: Foreground Activation-Driven Small Object Semantic Seg-mentation in Large-Scale Remote Sensing Imagery [J]. IEEE Trans. Geosci. Remote Sens., 2022, 60: 1-16.
[7] J.-H. Kim, Y. Hwang. GAN-Based Synthetic Data Augmentation for Infrared Small Target Detection [J]. IEEE Trans. Geosci. Remote Sens., 2022, 60: 1-12.
[8] N. Wang, B. Li. Ship Detection in Spaceborne Infrared Image Based on Lightweight CNN and Multisource Feature Cascade Decision [J]. IEEE Trans. Geosci. Remote Sens., 2021, 59(5): 4324-4339.
[9] J. Kang, et al. DisOptNet: Distilling Semantic Knowledge From Optical Images for Weather-Independent Building Segmentation [J]. IEEE Trans. Geosci. Remote Sens., 2022, 60: 1-15.
[10] K. Heidler, L. Mou. HED-UNet: Combined Segmentation and Edge Detection for Monitoring the Antarctic Coastline [J]. IEEE Trans. Geosci. Remote Sens., 2022, 60: 1-14.
[11] Z. Lu, et al. An Iterative Classification and Semantic Segmentation Network for Old Landslide Detection Using High-Resolution Remote Sensing Images [J]. IEEE Trans. Geosci. Remote Sens., 2023, 61: 1-13.
[12] DiResNet: Direction-Aware Residual Network for Road Extraction in VHR Remote Sensing Images.
[13] Y. Wei, S. Ji. Scribble-Based Weakly Supervised Deep Learning for Road Surface Extraction From Remote Sensing Images [J]. IEEE Trans. Geosci. Remote Sens., 2022, 60: 1-12.
[14] K. Simonyan, A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition [J]. arXiv, 2015.
[15] K. He, X. Zhang. Deep Residual Learning for Image Recognition [J]. arXiv, 2015.
[16] C. Szegedy, et al. Going deeper with convolutions [C]. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE: Boston, MA, USA.
[17] C. Zhang, W. Jiang. Transformer and CNN Hybrid Deep Neural Network for Semantic Segmentation of Very-High-Resolution Remote Sensing Imagery [J]. IEEE Trans. Geosci. Remote Sens., 2022, 60: 1-20.
[18] O. Tasar, S.L. Happy. ColorMapGAN: Unsupervised Domain Adaptation for Semantic Segmentation Using Color Mapping Generative Adversarial Networks [J]. IEEE Trans. Geosci. Remote Sens., 2020, 58(10): 7178-7193.
[19] L. Zhang, M. Lan. Stagewise Unsupervised Domain Adaptation With Adversarial Self-Training for Road Segmentation of Remote-Sensing Images [J]. IEEE Trans. Geosci. Remote Sens., 2022, 60: 1-13.
[20] R. Strudel, R. Garcia. Segmenter: Transformer for Semantic Segmentation [C]. In 2021 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE: Montreal, QC, Canada.
[21] Z. Liu, et al. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows [C]. In 2021 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE: Montreal, QC, Canada.
[22] C. Zhang, L. Wang. SwinSUNet: Pure Transformer Network for Remote Sensing Image Change Detection [J]. IEEE Trans. Geosci. Remote Sens., 2022, 60: 1-13.
[23] Y. Zhu, S. Newsam. DenseNet for dense flow [C]. In 2017 IEEE International Conference on Image Processing (ICIP), IEEE: Beijing.
[24] R. Li, et al. Multiattention Network for Semantic Segmentation of Fine-Resolution Remote Sensing Images [J]. IEEE Trans. Geosci. Remote Sens., 2022, 60: 1-13.
[25] K. Cha, J. Seo. A Billion-scale Foundation Model for Remote Sensing Images [J]. arXiv, 2023.
[26] S. Woo, et al. ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoders [J]. arXiv, 2023.
[27] X. Ding, X. Zhang. Scaling Up Your Kernels to 31x31: Revisiting Large Kernel Design in CNNs.
[28] P. Isola, J.-Y. Zhu. Image-to-Image Translation with Conditional Adversarial Networks [J]. arXiv, 2018.
[29] Y. Li, T. Shi. SPGAN-DA: Semantic-Preserved Generative Adversarial Network for Domain Adaptive Remote Sensing Image Semantic Segmentation [J]. IEEE Trans. Geosci. Remote Sens., 2023, 61: 1-17.
[30] L. Zhou, H. Yu. PU-GAN: A One-Step 2-D InSAR Phase Unwrapping Based on Conditional Generative Adversarial Network [J]. IEEE Trans. Geosci. Remote Sens., 2022, 60: 1-10.
[31] Q. Hou, C.-Z. Lu. Conv2Former: A Simple Transformer-Style ConvNet for Visual Recognition [J]. arXiv, 2022.
[32] Z. Liu, et al. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows [C]. In 2021 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE: Montreal, QC, Canada.
[33] D. Muhtar, X. Zhang. CMID: A Unified Self-Supervised Learning Framework for Remote Sensing Image Understanding [J]. IEEE Trans. Geosci. Remote Sens., 2023, 61: 1-17.
[34] H. Zhang, et al. DGCC-EB: Deep Global Context Construction With an Enabled Boundary for Land Use Mapping of CSMA [J]. IEEE Trans. Geosci. Remote Sens., 2022, 60: 1-15.
[35] L. Ding, H. Tang. LANet: Local Attention Embedding to Improve the Semantic Segmentation of Remote Sensing Images [J]. IEEE Trans. Geosci. Remote Sens., 2021, 59(1):

426-435.
[36] L.-C. Chen, G. Papandreou. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs [J]. arXiv, 2017.
[37] M. Yang, K. Yu. DenseASPP for Semantic Segmentation in Street Scenes [C]. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE: Salt Lake City, UT, USA.
[38] Q. Zhao, J. Liu. Semantic Segmentation With Attention Mechanism for Remote Sensing Images [J]. IEEE Trans. Geosci. Remote Sens., 2022, 60: 1-13.
[39] O. Ronneberger, P. Fischer. U-Net: Convolutional Networks for Biomedical Image Segmentation [J]. arXiv, 2015.
[40] G. Lin, A. Milan. RefineNet: Multi-path Refinement Networks for High-Resolution Semantic Segmentation [C]. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE: Honolulu, HI.
[41] J. Fu, X. Sun. An Anchor-Free Method Based on Feature Balancing and Refinement Network for Multiscale Ship Detection in SAR Images [J]. IEEE Trans. Geosci. Remote Sens., 2021, 59(2): 1331-1344.
[42] C. Peng, K. Zhang. Cross Fusion Net: A Fast Semantic Segmentation Network for Small-Scale Semantic Information Capturing in Aerial Scenes [J]. IEEE Trans. Geosci. Remote Sens., 2022, 60: 1-13.
[43] J. Zheng, A. Shao. Remote Sensing Semantic Segmentation via Boundary Supervi-sion-Aided Multiscale Channelwise Cross Attention Network [J]. IEEE Trans. Geosci. Remote Sens., 2023, 61: 1-14.
[44] A. Li, L. Jiao. Multitask Semantic Boundary Awareness Network for Remote Sensing Image Segmentation [J]. IEEE Trans. Geosci. Remote Sens., 2022, 60: 1-14.
[45] V. Badrinarayanan, A. Kendall. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation [J]. IEEE Trans. Pattern Anal. Mach. Intell., 2017, 39(12): 2481-2495.
[46] W. Liu, F. Su. Bispace Domain Adaptation Network for Remotely Sensed Semantic Segmentation [J]. IEEE Trans. Geosci. Remote Sens., 2020: 1-11.
[47] M. Tang, K. Georgiou. Semantic Segmentation in Aerial Imagery Using Multi-level Contrastive Learning with Local Consistency [C]. In 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), IEEE: Waikoloa, HI, USA.
[48] H. Xue, C. Liu. DANet: Divergent Activation for Weakly Supervised Object Localization [C]. In 2019 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE: Seoul, Korea (South).
[49] J. Hu, L. Shen. Squeeze-and-Excitation Networks [J]. arXiv, 2019.
[50] S. Zhang, et al. Attention Guided Network for Retinal Image Segmentation [J]. arXiv, 2019.
[51] H. Zhao, J. Shi. Pyramid Scene Parsing Network [J]. arXiv, 2017.
[52] T.-Y. Lin, P. Dollár. Feature Pyramid Networks for Object Detection [J]. arXiv, 2017.
[53] K. He, G. Gkioxari. Mask R-CNN [J]. arXiv, 2018.
[54] Z. Zheng, Y. Zhong. Foreground-Aware Relation Network for Geospatial Object Segmentation in High Spatial Resolution Remote Sensing Imagery.
[55] B. Pan, X. Xu. DSSNet: A Simple Dilated Semantic Segmentation Network for Hyperspectral Imagery Classification [J]. IEEE Geosci. Remote Sens. Lett., 2020, 17(11): 1968-1972.
[56] H. Noh, S. Hong. Learning Deconvolution Network for Semantic Segmentation [J]. 2015.
[57] K. He, H. Fan. Momentum Contrast for Unsupervised Visual Representation Learning [J]. arXiv, 2020.
[58] T. Chen, S. Kornblith. A Simple Framework for Contrastive Learning of Visual Representations [J]. arXiv, 2020.
[59] J.-B. Grill. Bootstrap your own latent: A new approach to self-supervised Learning [J]. 2020.