

# Class Imbalance in the Automatic Interpretation of Remote Sensing Images: A Review

Pengdi Chen<sup>1</sup>, Yuanrui Ren, Baoan Zhang, and Yuan Zhao

**Abstract**—Class imbalance is a very challenging problem in data science, affecting the development of several application fields. This problem also plagues the automatic interpretation of remote sensing images. Especially in tasks such as classification mapping, object detection, change detection, and scene classification, the classes of training samples required by machine learning exhibit uneven distribution, which seriously affects the accuracy of model training. Our meta-analysis is based on 171 journal papers retrieved and screened from the Web of Science database, covering publication years, highly productive countries, highly cited authors, remote sensing data types, data augmentation methods, and the distribution of the main application fields. The solution to the proposed problem involves three aspects: model innovation and optimization, loss function improvement, and data augmentation. Experiments on benchmark datasets have demonstrated the effectiveness of these methods. In terms of remote sensing task applications, we provide a comprehensive review and analysis of recent research cases on deep learning aimed at addressing the class imbalance problem. Finally, we discuss the synergistic relationship between models, loss functions, and data augmentation, summarize the current challenges in this field, as well as propose several ideas for addressing the class imbalance problem.

**Index Terms**—Automatic interpretation, class imbalance, deep learning, meta-analysis, remote sensing images.

## I. INTRODUCTION

**A**UTOMATIC interpretation of remote sensing imagery is achieved by integrating remote sensing data with image processing and analytical methods [1], typically employing traditional machine learning and deep learning methods. Among these, traditional machine learning methods include support vector machines, decision trees, random forests, and K-nearest neighbors, all of which rely on manually designed feature extraction. Deep learning utilizes deep neural networks to automatically learn and extract important features and has begun to dominate various visual remote sensing tasks including classification mapping, object detection, change detection, and scene classification. The realization of a given task usually

Received 18 October 2024; revised 22 January 2025 and 15 February 2025; accepted 24 March 2025. Date of publication 28 March 2025; date of current version 14 April 2025. This work was supported in part by the Gansu Natural Resources Department Science and Technology Program under Grant 202425 and in part by the Gansu Provincial Institute of Mapping. (Corresponding author: Pengdi Chen.)

Pengdi Chen and Yuanrui Ren are with the Lanzhou University, Lanzhou 730000, China (e-mail: chenpd21@lzu.edu.cn; renyr21@lzu.edu.cn).

Baoan Zhang and Yuan Zhao are with the Mapping Institution of Gansu Province, Lanzhou 730000, China (e-mail: zhangba1118@163.com; 653452779@qq.com).

Digital Object Identifier 10.1109/JSTARS.2025.3555567

requires the support of corresponding sample data. However, the sample data used in actual training often needs to address the problem of class imbalance. Class imbalance is when certain classes in a dataset have significantly fewer samples than others, resulting in an overall “long-tail” distribution characteristic [2]. In practical applications, models trained on imbalanced data tend to be biased toward the majority class [3].

In this article, we systematically review class imbalance solutions for the automatic interpretation of remote sensing images in the era of deep learning, covering meta-analysis, methodologies, and application fields. Through meta-analysis, the research dynamics in this field can be visualized. On this basis, the methodological review presents specific solutions to address class imbalance. Finally, the analysis of application fields demonstrates the scope of class imbalance’s influence.

### A. Background and Aims

Generally speaking, the issue of imbalance is widespread across various fields, such as computer vision and pattern recognition. This article focuses specifically on the imbalance problem in remote sensing. However, to fully appreciate the contributions of this work, it is essential to understand some of the previous solutions.

Over the past two decades, the issue of class imbalance has garnered increasing attention and has become prevalent across various application domains, including information security, biomedicine, manufacturing systems, and computer vision [4] (see Table I). To mitigate the bias toward majority classes, researchers may modify the training data to reduce imbalance or adjust the model’s learning or decision processes to enhance sensitivity to minority classes. Common approaches to address class imbalance can be classified into data-level, algorithm-level, and hybrid methods [4], [5].

Data-level methods utilize resampling techniques to alter the number of samples across categories, including undersampling, oversampling, and mixed-sampling [6]. Random undersampling (RUS) involves randomly selecting and removing samples from the majority class to achieve balance [7], [8]. However, this method may remove important samples that are useful for classification, decreasing majority class accuracy. Random oversampling (ROS) balances the class distribution in a dataset by randomly duplicating minority class samples [22], [23]. A representative ROS method is the synthetic minority over-sampling technique (SMOTE) and its variants [24], [25], which generate new synthetic samples through linear interpolation between

TABLE I  
APPLICATION FIELDS RELATED TO CLASS IMBALANCE

Number	Research Fields	Applications
1	Information Security	Network Intrusion Detection [9], Fraud Detection [10], Software Defect Detection [11]
2	Biomedicine	Medical Diagnosis [12], Protein Classification [13], Gene Expression Classification [14]
3	Manufacturing Systems	Anomaly Detection [15], Industrial System Prediction [16], Equipment Condition Monitoring [17]
4	Computer Vision	Semantic Segmentation [18], Object Detection [19], Change Detection [20], Scene Classification [21]

K-nearest neighbor samples to achieve balance in terms of sample quantity. ROS partially addresses the key feature loss issue present in RUS but the drawback is the increase in redundant data, which may lead to longer training times for the classifier. Hybrid sampling combines the advantages of RUS and ROS, typically using SMOTE to increase the number of minority class samples, reducing the overfitting problem during training, and then using the RUS method to mitigate class overlap or class noise [26], [27]. Algorithmic-level approaches include cost-sensitive learning, ensemble learning, and active learning [28]. Cost-sensitive learning integrates the classification costs of different classes into the model training process, improving the model's attention to classifying minority class samples [29], [30]. Ensemble learning constructs an optimal classification model by combining multiple classifiers and is often used with data resampling, cost-sensitive learning, and deep learning techniques [31], [32]. Active learning involves automatically selecting the most informative samples for labeling during the model training process, which can be classified into online active learning and deep active learning [33], [34]. Finally, hybrid methods combine sampling and algorithmic approaches to achieve optimal class rebalancing [35]. While these methods can significantly improve the problem of class imbalance, they still face challenges such as increased computational resources and time and the need for fine-tuning the model parameters.

The main aim of this article is to systematically review and analyze class imbalance methods for addressing the problem in the automated interpretation of remote sensing images based on deep learning techniques. In order to do this, we did the following.

- 1) We conducted a detailed bibliometric analysis of papers related to this work screened from 2017 to 2023. Data mining techniques were employed to analyze publication years, highly productive countries, highly cited authors, remote sensing data types, data augmentation methods, as well as the distribution of major application fields.
- 2) A comprehensive methodological and experimental analysis of the selected papers was conducted, and the solutions included model innovation and optimization, improvements in loss function design, and data augmentation preprocessing.
- 3) Based on the selected papers, we summarize the current remote sensing application fields where class imbalance exists, including classification mapping, object detection, change detection, and scene classification.
- 4) We discuss the article at the end, the synergistic relationship between models, loss functions, and data

augmentation, summarize the current challenges in this field, and propose several ideas for addressing the class imbalance problem.

### B. Comparison With Previous Reviews

In recent years, numerous review papers have emerged that address the problem of class imbalance. However, they primarily focus on other fields or specific aspects within a given field. Giorgio et al. [36] systematically reviewed and summarized the class imbalance problem in the manufacturing sector, noting that the solutions include data manipulation, machine learning, and deep learning. The rational use of these methods according to different tasks and scenarios is an effective solution to address the class imbalance problem. Johnson and Khoshgoftaar [4] focused on investigating the relationship between class imbalance and deep learning, examining factors such as data complexity, architecture of testing, performance interpretation, usability, and big data applications to better understand the effectiveness of deep learning when applied to imbalanced data. Zhang et al. [28] examined the solution of deep network recognition models to the long-tailed class imbalance from the direction of visual recognition, including three aspects: class rebalancing, information enhancement, and model improvement. Ghosh et al. [37] explored the applicability of traditional machine learning methods for addressing class imbalance in deep learning and, through experiments, assessed the effectiveness of network depth and regularization techniques in addressing this problem. Oksuz et al. [38] systematically explained the imbalance problem in object detection, covering class, scale, spatial, and object imbalance, and proposed potential solutions.

In the field of remote sensing, Xu et al. [39] mentioned the effect of class imbalance on animal detection in aerial and satellite images but they did not provide a comprehensive review of this problem. Huang et al. [40], in their study on semantic segmentation of remote sensing images, found that class imbalance significantly affects segmentation accuracy, particularly foreground–background imbalance, which hinders the extraction of individual classes. Sun et al. [41], in a few-shot learning research on remote sensing image interpretation, found that geographic object scale differences and foreground–background class imbalance are the main challenges deep learning faces in this field. They also reviewed current solutions, focusing on data augmentation and prior knowledge-based learning. Paoletti et al. [42] examined class imbalance in hyperspectral data classification for remote sensing, providing a comprehensive review of

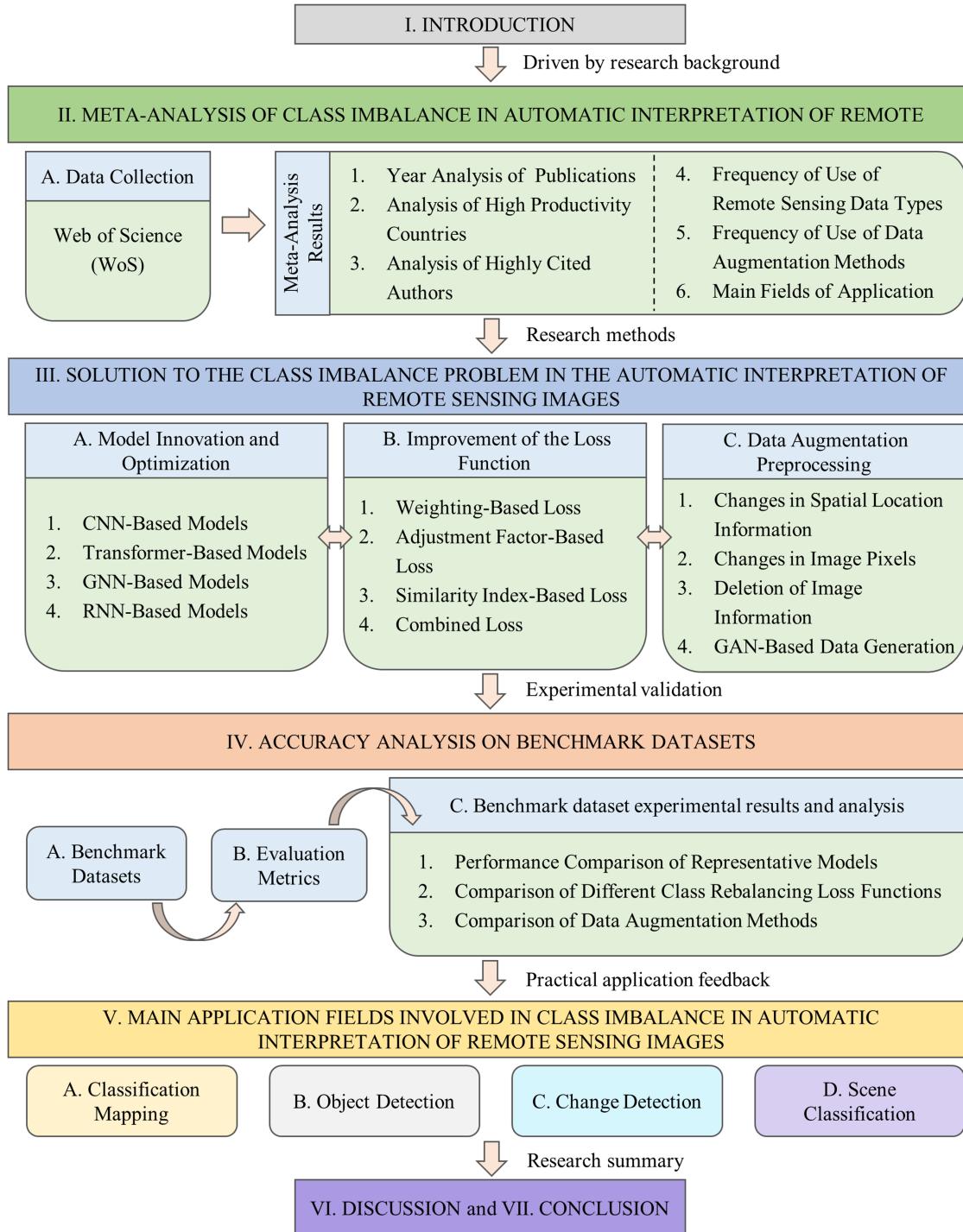


Fig. 1. Structure of this review.

oversampling techniques, and validated the substantial performance of these methods in addressing class imbalance through experimental comparisons. Despite numerous review studies providing valuable insights into addressing class imbalance from various perspectives, such as data, methods, and applications, it is evident that few investigations comprehensively and systematically focus on the problem of class imbalance in the automatic interpretation of remote sensing images using deep learning.

### C. Guide to Reading This Review

The structure of this review is shown in Fig. 1, and the organization of the remaining sections is as follows. Section II conducts a meta-analysis of relevant papers from the WoS database. Section III reviews methods for addressing class imbalance in the automatic interpretation of remote sensing imagery. Section IV analyzes and compares the performance of different methods on benchmark datasets. Section V reviews the

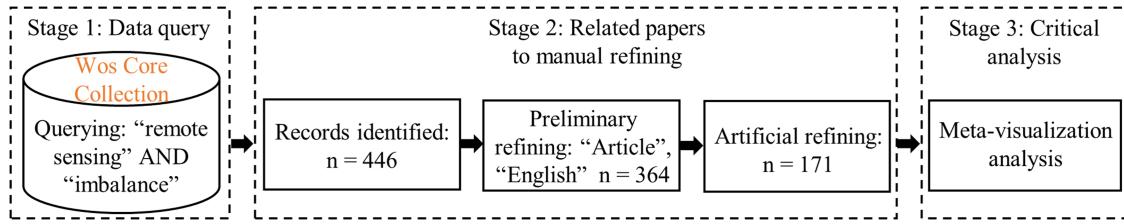


Fig. 2. Process of selecting and analyzing research papers.

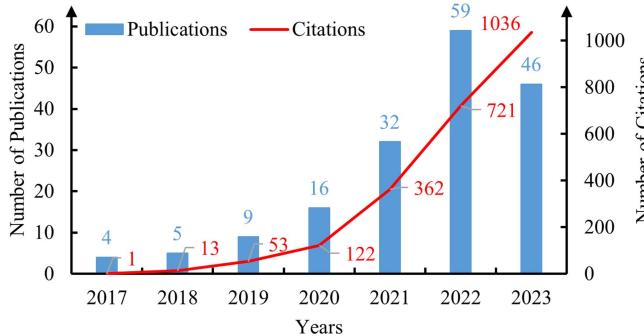


Fig. 3. Statistics on year of publications.

application of class imbalance research in four remote sensing tasks. Section VI discusses the synergy between models, loss functions, and data augmentation while summarizing potential solutions to the challenges. Section VII concludes the article.

## II. META-ANALYSIS OF CLASS IMBALANCE IN AUTOMATIC INTERPRETATION OF REMOTE SENSING IMAGES

### A. Data Collection

We searched for papers published in Web of Science (WoS) from 2017 to 2023 using the search terms “Remote sensing” AND “Imbalance.” This initial search resulted in 446 papers, which were then manually screened to exclude irrelevant literature, leading to a final selection of 171 eligible peer-reviewed papers for meta-analysis. The process of paper retrieval and analysis is illustrated in Fig. 2.

### B. Meta-Analysis Results

1) *Year Analysis of Publications:* The bar chart in Fig. 3 reflects the growth of research output over the years. Evidently, the number of publications related to the topic gradually increased from 4 papers in 2017 to 46 papers in 2023. Specifically, research on the class imbalance problem in the field of remote sensing consisted of only 4, 5, and 9 papers from 2017 to 2019, respectively. Starting in 2020, the number of publications began to rise significantly, peaking in 2022 with 59 papers published on the topic. In 2023, 46 papers were included, showing a decline compared to 2022 but still indicating an overall upward trend, suggesting that researchers are increasingly focusing on class imbalance problem in remote sensing image interpretation. The red line in Fig. 3 indicates the variation in citations across the

years, revealing an exponential growth trend in citations, with a sharp increase starting in 2020, when citations reached 122, an increase of 69 compared to 2019. This highlights that research on class imbalance problems in this field began to be emphasized from that year onward.

2) *Analysis of High Productivity Countries:* Among the reviewed publications, we counted the geospatial distribution and quantitative contributions to the research on the class imbalance problem in the automatic interpretation of remote sensing images in each country. As shown in Fig. 4, regarding spatial distribution, all publications are distributed in 26 countries, with the highest concentration in Asia and Europe, indicating a greater focus on class imbalance problem by researchers in these regions. Regarding collaboration, China and Australia have the most partnerships with other countries. China has established cooperation with 11 countries, which are mainly located in Europe. Australia has partnered with seven countries in Asia and Europe. Notably, the United Arab Emirates, Sweden, Greece, and Portugal have not collaborated with other countries.

3) *Analysis of Highly Cited Authors:* Table II presents the top 20 highly cited authors from the 171 publications, where P and C represent the number of publications and citations, respectively, with darker colors indicating more excellent corresponding publications and citations. The table shows that eight authors have 100 or more citations, and 12 have fewer than 100. From the timeline, the number of citations for each author has gradually increased since 2020, indicating that class imbalance in the automatic interpretation of remote sensing images is attracting increasing attention from scholars. Although X. Yang, Z. Shao, and Y. Liu only have 3, 3, and 1 paper(s), their citations reached 186, 176, and 172, respectively, so these three papers should be focused on. Notably, Y. Liu’s paper, “Building Change Detection for Remote Sensing Images Using a Dual-Task Constrained Deep Siamese Convolutional Network Model” has the highest single-paper citation count of 172 and should be considered a key reference.

4) *Frequency of Use of Remote Sensing Data Types:* Fig. 5 illustrates the types of remote sensing data used to study class imbalance problem across 171 publications, including multispectral, hyperspectral, synthetic aperture radar (SAR), and multimodal data. The multimodal data include combinations of multispectral with light detection and ranging data and multispectral with SAR data. It is worth noting that some articles in object detection tasks used both multispectral image datasets and SAR image datasets. Fig. 5 shows that multispectral data are

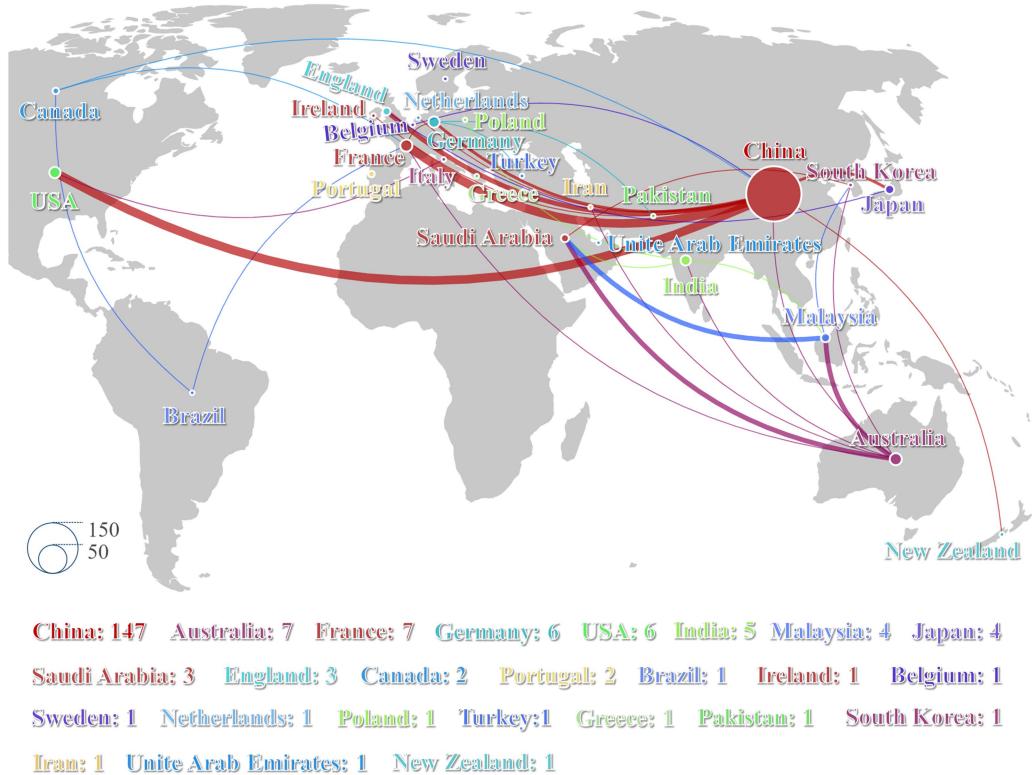


Fig. 4. Spatial distribution of high-productivity countries. The size of each bubble represents the total number of publications from that country, with larger bubbles indicating a higher publication volume. The thickness of the lines reflects the level of collaboration between countries, with thicker lines indicating more extensive cooperation.

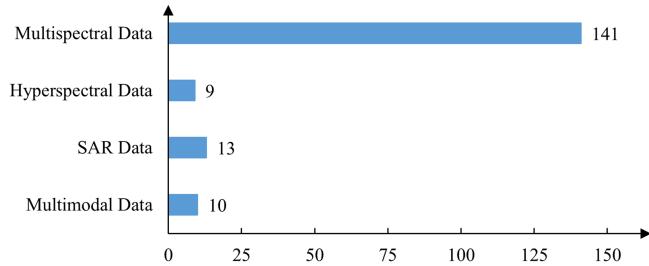


Fig. 5. Frequency of use of class imbalance data types in publications.

the most frequently used in studies of class imbalance, with 141 publications, followed by SAR and multimodal data, with 13 and 10 publications, respectively. Hyperspectral data are used less frequently, with only nine publications.

5) *Frequency of Use of Data Augmentation Methods:* Data augmentation methods can effectively increase the number of minority class samples and alleviate the class imbalance issue in datasets. Fig. 6 summarizes 16 data augmentation methods used across 171 publications. These include affine transformation methods—Translation, Scaling, Rotation, Flipping, and Cropping—as well as mixing methods such as Mosaic, Mixup, Copy-Paste, Random Shuffle, CutMix, RSI-Mix, and ChessMix. These two categories are classified as “Changes in Spatial Location Information” in Section III-C. Adding Noise and Color Jitter/Transformation fall under “Changes in Image Pixels,” while Cutout and GridMask are mask-based methods,

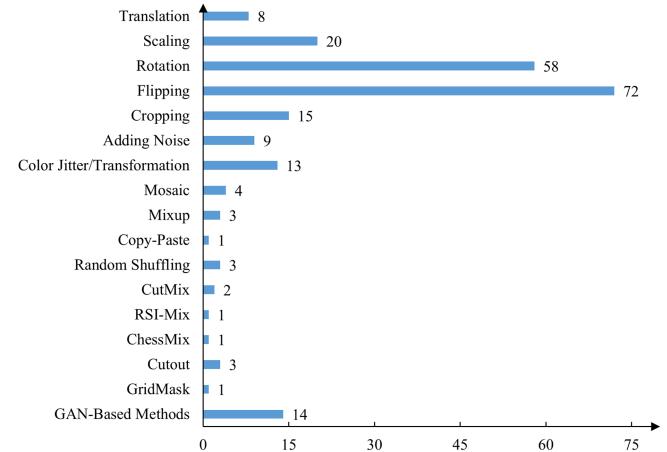


Fig. 6. Frequency of use of data augmentation methods in publications.

categorized as “Deletion of Image Information.” Generative adversarial network (GAN) based methods are classified under “GAN-Based Data Generation.”

As shown in Fig. 6, Rotation and Flipping are the most commonly used methods, appearing 58 and 72 times, respectively. This is likely due to their simple implementation. Following these, Scaling, Cropping, GAN-Based Methods, and Color Jitter/Transformation are used 20, 15, 14, and 13×, respectively. Methods based on hybrid approaches and masks appear less frequently.

TABLE II  
HIGHLY CITED AUTHORS

Authors	Publications (P)   Citations (C)														Total	
	2017		2018		2019		2020		2021		2022		2023			
	P	C	P	C	P	C	P	C	P	C	P	C	P	C		
Xue Yang	0	0	0	0	0	0	0	0	1	12	1	57	1	117	3	186
Zhenfeng Shao	0	0	0	0	0	0	1	4	1	59	0	61	1	52	3	176
Yi Liu	0	0	0	0	0	0	0	0	1	12	0	57	0	103	1	172
Xian Sun	0	0	2	1	0	11	0	17	2	34	3	56	2	45	9	164
Kun Fu	0	0	2	1	0	11	0	17	2	30	1	53	1	41	6	153
Jiangyun Li	0	0	1	0	0	9	0	16	1	19	0	38	0	47	2	129
Biswajeet Pradhan	0	0	0	0	0	0	1	1	2	22	0	39	0	42	3	104
Quanzhi An	0	0	0	0	1	0	0	12	0	32	0	31	0	25	1	100
Wenhui Diao	0	0	1	0	0	3	0	6	2	12	1	38	2	37	6	96
Yi Zhang	0	0	2	1	0	11	0	17	0	24	1	27	0	16	3	96
Abolfazl Abdollahi	0	0	0	0	0	0	1	1	1	19	0	30	0	35	2	85
Rongsheng Dong	0	0	0	0	1	3	0	8	0	19	0	19	0	30	1	79
Andrew Mellor	1	0	0	3	0	7	1	10	0	17	0	18	0	23	2	78
Zhiyong Xu	0	0	0	0	0	0	0	0	1	11	0	29	0	35	1	75
Xinghua Li	0	0	0	0	0	0	0	0	1	1	0	34	0	36	1	71
Hao Chen	0	0	0	0	0	0	0	0	1	5	0	23	0	33	1	61
Xun Gao	0	0	1	1	0	8	0	11	0	18	0	15	0	5	1	58
Tianwen Zhang	0	0	0	0	0	0	0	0	1	0	0	24	0	33	1	57
Luke Collins	0	0	0	0	0	0	1	3	0	15	0	18	0	20	1	56
Jan Hemmerling	0	0	0	0	0	0	0	0	1	0	0	24	0	31	1	55

The darker the color, the larger the corresponding value.

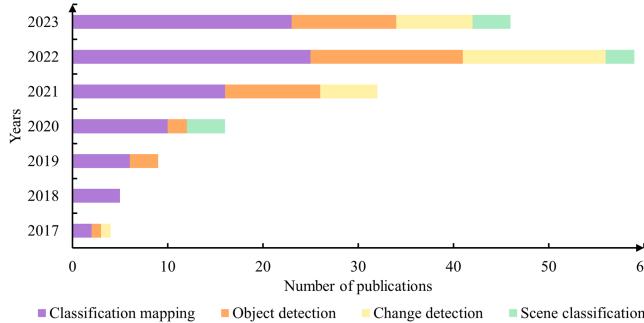


Fig. 7. Main application fields of class imbalance research.

6) *Main Fields of Application*: We classified the 171 selected publications by application area into four main types: classification mapping, object detection, change detection, and scene classification, each encompassing various subfields (see Fig. 7). In terms of publication volume, classification mapping has the highest number of papers, followed by object detection and change detection. In contrast, scene classification has the fewest, indicating that the primary attention in the field of remote sensing interpretation is in the first three categories, with relatively little research on the fourth category related to class imbalance. In terms of the year of publication, the number of papers in each field shows an increasing trend with the year, with the highest

number published in 2022 and the lowest in 2017. In 2018, no papers related to class imbalance appeared in object detection, and there is a lack of papers related to change detection in 2018–2020. Papers related to scene classification only appeared in 2020, 2022, and 2023, suggesting that such problems have received little attention in the field and that researchers should pay attention to this statistical detail.

### III. SOLUTION TO THE CLASS IMBALANCE PROBLEM IN THE AUTOMATIC INTERPRETATION OF REMOTE SENSING IMAGES

Class imbalance in automatic interpretation of remote sensing images is often influenced by disparities in the area, distribution characteristics, and object sizes of ground objects within the research area [43]. This results in differing pixel ratios for different objects, severely disrupting the feature extraction capabilities of deep neural networks [44]. The public sample dataset shown in Fig. 8 illustrates this class distribution imbalance problem.

In the training data, limitations due to differences in ground object distribution and the high cost of labeling lead to insufficient samples for some object classes with small area proportions. This class imbalance causes the majority classes to dominate model performance, thereby restricting the generalization ability of minority classes [45]. Given the unique nature of remote sensing imagery, traditional pattern recognition and machine learning methods for addressing class imbalance

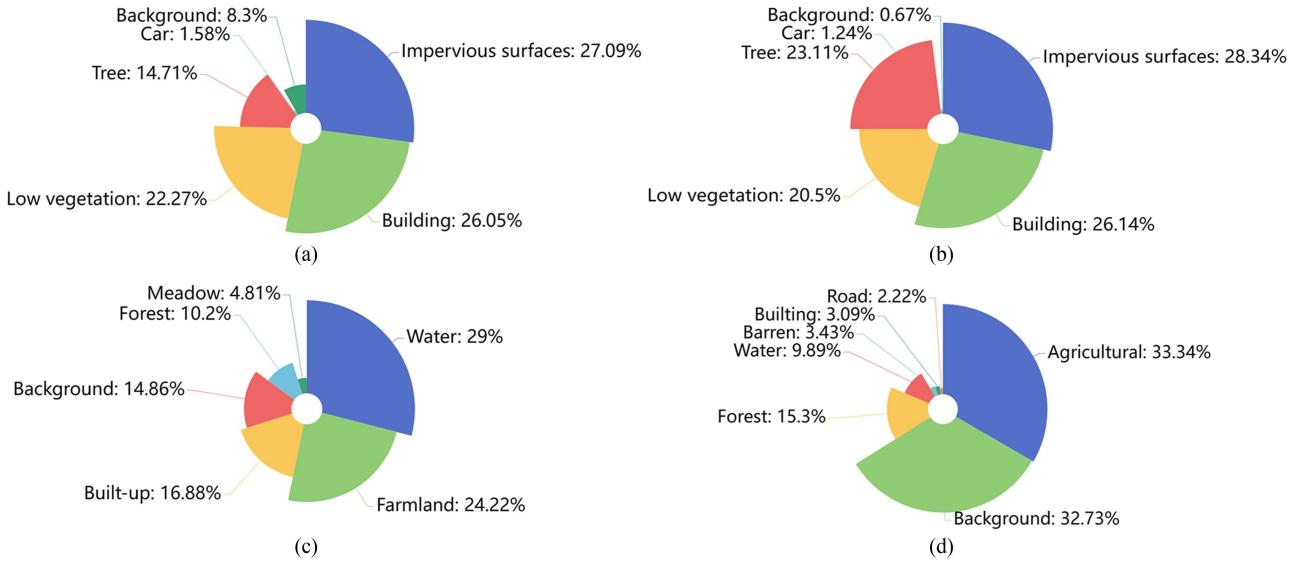


Fig. 8. Distribution of classes in the public dataset. (a)–(d) Potsdam, Vaihingen, GID, LoveDA-rural, respectively.

may not be suitable for this field. For example, the SMOTE method relies on Euclidean distance to select nearest neighbor samples, making it applicable only to numerical data [46]. Therefore, for image-level and pixel-level remote sensing tasks, it is difficult to use SMOTE to generate new data to balance minority classes. Based on existing literature, we classify class imbalance learning methods in remote sensing image interpretation into three types: model-based, loss function-based, and data augmentation-based, each containing different techniques (see Fig. 9).

#### A. Model Innovation and Optimization

In the automatic interpretation of remote sensing imagery, model-based solutions for addressing class imbalance typically use deep neural networks in conjunction with loss functions. The backbone of these deep neural networks includes architectures such as convolutional neural network (CNN), Transformer, graph neural network (GNN), and recurrent neural network (RNN) (see Table III).

1) *CNN-Based Models*: The most commonly used deep learning backbone for addressing class imbalance problems in remote sensing image interpretation tasks is the CNN model, which has been widely applied in various remote sensing applications such as classification mapping, object detection, change detection, and scene classification [47], [48], [49], [50] (see Fig. 10). The CNN-based model has the advantages of multiscale local connectivity, weight sharing, and multilayer structure [63]. It aims to learn the spatial features that best describe the object class or quantity, such as edges, corners, textures, or shapes. The multiscale feature extraction function it has allows multiple successive operations on training sample data of different scales and proportions as a way of fusing recognized low-level spatial features with high-level semantic features [64].

Some research works have been devoted to incorporating attention mechanisms, balanced weighted loss, and classifiers

integrating multiple underlying CNN into the network design to further mitigate the class imbalance problem. Incorporating the attention mechanism and the balance-weighted loss function allows the model to pay more attention to minority class samples or small object classes in the foreground during training, a more commonly used way to deal with class imbalance currently [65], [66]. Networks that integrate multiple base CNN classifiers utilize an ensemble learning method, which builds a robust classifier by combining several weak classifiers, thereby improving overall classification performance [67]. When dealing with class imbalance, ensemble learning can ensure that minority classes receive sufficient attention by adjusting class weights in each classifier or using different classifiers for training [68], [69], [70]. Another learning approach involving CNN models is transfer learning, which uses knowledge from a source domain to assist in learning within a target domain [71]. In transfer learning, by adjusting model parameters or feature representations, the model can adapt more quickly to the data distribution of the target domain, providing significant potential for addressing data scarcity and class imbalance problems [72], [73], [74].

2) *Transformer-Based Models*: Transformer is a deep learning model framework developed after CNN, which uses a multihead attention structure for global modeling by learning long-range dependencies between sequence elements [75] (see Fig. 11). Each head of the multihead attention has its learnable weights, and different class weighting values can be assigned to alleviate the class imbalance problem during the training process. Most research uses Transformer in conjunction with CNN [51], [52], [76], where the former extracts global features and the latter extracts local features, enhancing the model's ability to recognize small objects in imbalanced datasets by fusing both feature types.

We classify Transformers into Vision Transformer (ViT), Swin Transformer (SwinT), and Pyramid Vision Transformer (PVT) based on the backbone. ViT segments the image into multiple patches, which are then treated as sequential data input

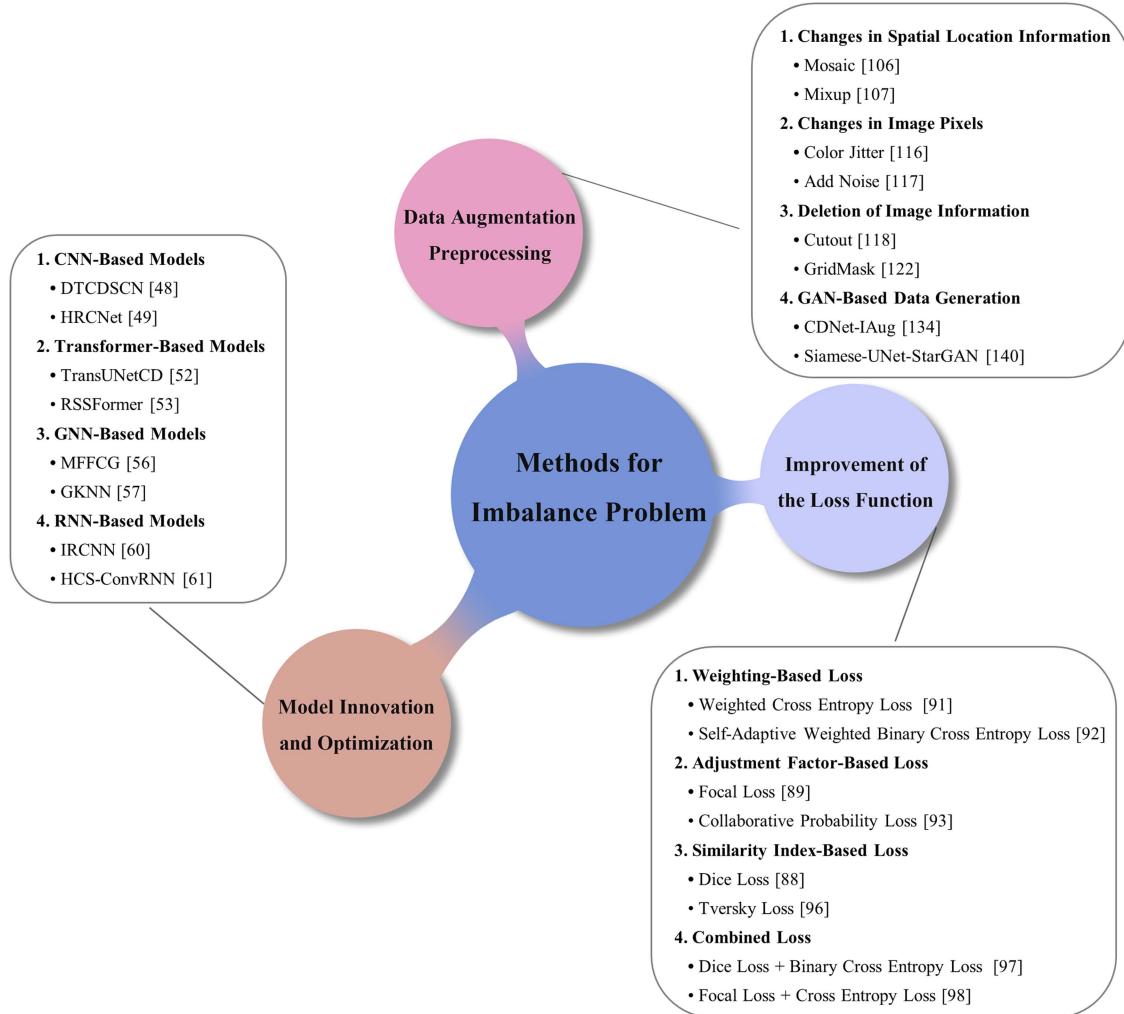


Fig. 9. Methodological classification for addressing class imbalance in remote sensing image interpretation.

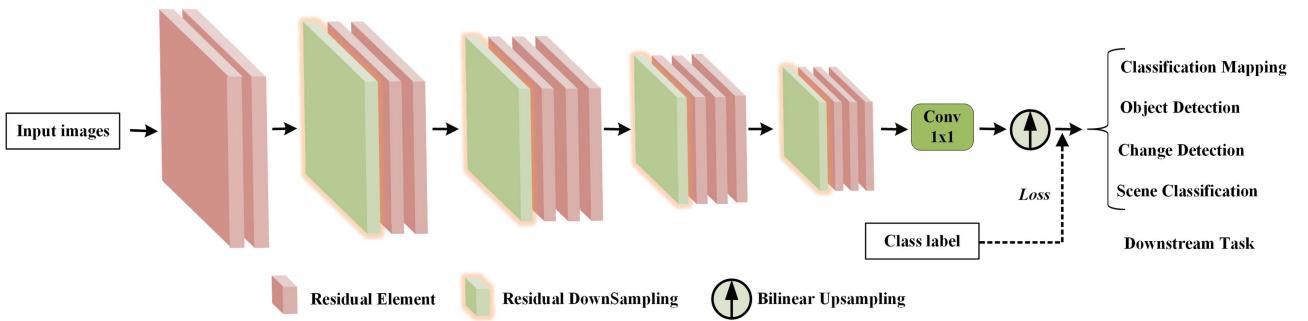


Fig. 10. Structure of the CNN Model.

into a standard Transformer model for training. It does not rely on traditional convolutional operations but captures image features exclusively through a self-attention mechanism [77]. SwinT is a hierarchical Transformer in which the range of self-attention is restricted to local windows, and these window sizes are gradually enlarged as the network depth is increased to capture local to global information effectively [78]. PVT combines a Transformer and pyramid structure, which capture image features at different resolutions by constructing multiscale

feature representations [79]. Each of these Transformer architectures has its advantages: ViT is simple and straightforward, well-suited for training on large-scale datasets; SwinT is more efficient for processing image data and adaptable to inputs of various sizes; PVT provides a multiscale perspective, making it ideal for fine-grained recognition tasks.

3) **GNN-Based Models:** GNN is a deep learning model specifically designed for processing graph data, exhibiting robust capabilities for modeling spatial relationships by

TABLE III  
SEVERAL REPRESENTATIVE MODELS OF AUTOMATIC INTERPRETATION ARCHITECTURE OF REMOTE SENSING IMAGES  
TO ADDRESS THE PROBLEM OF CLASS IMBALANCE

Category	Representative Method	Highlights
CNN	DTCDSN [47]	The model consists of a change detection network and two semantic segmentation networks, enabling the simultaneous completion of both tasks. The introduction of focal loss helps to mitigate the class imbalance problem.
	HRCNet [48]	The model addresses the problem of class scale imbalance and boundary uncertainty by fusing the advantages of HRNet, dual-attention module, boundary awareness, and loss function.
	MFCN [49]	The model utilizes a multiscale convolutional kernel to extract detailed features of ground objects. The proposed WBCE+Dice loss can further train the model from imbalanced samples.
	CFWS [50]	The model comprises object-based label extraction, label correction, and boundary-aware semantic segmentation. A classification-first, segmentation-later strategy is proposed to alleviate the class imbalance problem.
Transformer	TransUNetCD [51]	By leveraging the long-range feature dependencies of Transformers to overcome the limitations of UNet and incorporating a weighted loss function, the model effectively mitigates the impact of class imbalance.
	RSSFormer [52]	The model consists of an adaptive Transformer fusion module, a detail-aware attention layer, and a foreground saliency-guided loss that enables optimization of foreground–background balance.
	EGDE-Net [53]	The model is composed of an encoder structure built from multiple edge-guided Transformer modules and FDEM, along with a decoder structure formed by DAM. It also incorporates a hybrid loss function to perform change detection on imbalanced data.
	MMT [54]	The proposed mixed-mask attention mechanism effectively distinguishes between foreground and background information, while the integration of a multiscale learning strategy with Transformers efficiently leverages large-scale feature maps.
GNN	MFFCG [55]	By combining 3D-CNN with GATs, the model achieves efficient classification of imbalanced hyperspectral data through the dynamic adjustment of node weights.
	GKNN [56]	The model employs a graph kernel GNN to compensate for the imbalance between tail and head classes.
	DGRN [57]	By integrating graph reasoning with GCNs, the model eliminates the foreground–background imbalance, achieving accurate classification and localization.
	DCI-PGCN [58]	By utilizing the capability of GCN to more effectively distinguish class feature boundaries, the model achieves foreground landslide detection over extensive fields.
RNN	IRCNN [59]	The model consists of a multibranch twin CNN, irregular temporal distance LSTM, and fully connected layers, combined with a weighted BCE loss to address the class imbalance problem.
	HCS-ConvRNN [60]	The proposed hierarchical classification structure classifies pixels into relevant classes layer by layer. To reduce the effect of class imbalance, local loss values are weighted during training.
	LSI-RNN [61]	The model generates a gravitational field map by introducing a neural discriminant dimension reduction layer, which transforms the classification problem into a line regression problem and address the local blurring at the edges and the class imbalance at the corners.
	Densenet121-LSTM [62]	By harnessing the powerful feature extraction capabilities of CNN and the temporal processing abilities of LSTM, the model enhances the discriminability of small objects, while the new loss function alleviates the problem of label ratio imbalance.

considering the connections between nodes within a graph [80] (see Fig. 12). Unlike traditional CNNs, which are suited for regular grid data, GNN effectively handles irregular graph structures. Numerous GNN models have been developed to address the class imbalance problem in remote sensing, with graph convolutional network (GCN) and graph attention network (GAT) being the most widely used. GCN learns the feature representations of the nodes by performing convolutional operations on the graph structure, and it can effectively utilize the topological information of the graph data. In remote sensing image

interpretation, GCN deals with the class imbalance problem by adjusting the weights of node features or designing a suitable sampling strategy [58]. GAT is a GNN based on the attention mechanism, which can learn the node relationships in the graph data and achieve the effect of addressing the class imbalance by enhancing the attentional weights of the nodes of a few classes [55].

4) *RNN-Based Models*: RNN is a deep learning model for processing sequence data, which captures long-range dependencies between sequences by performing recursive operations on

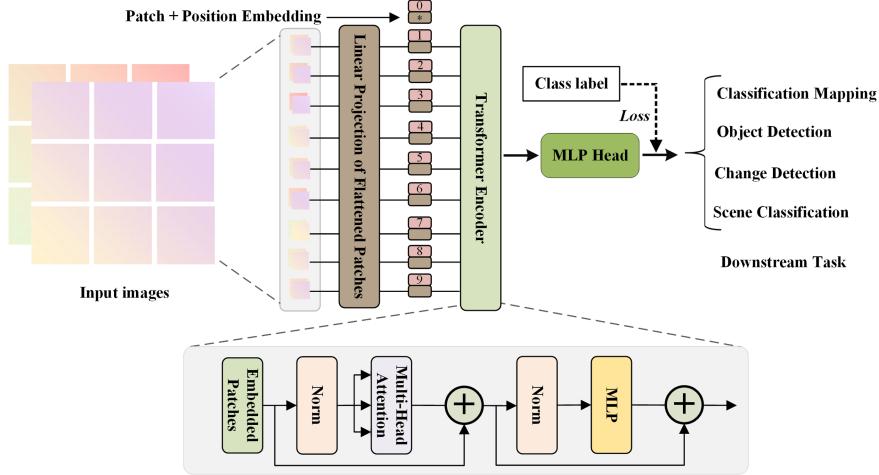


Fig. 11. Structure of ViT Model.

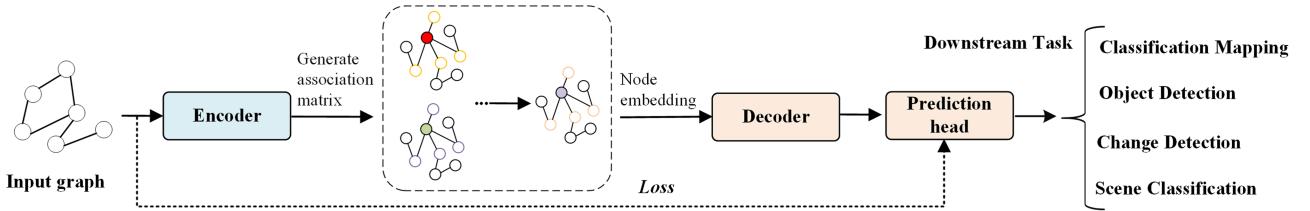


Fig. 12. Structure of GNN Model.

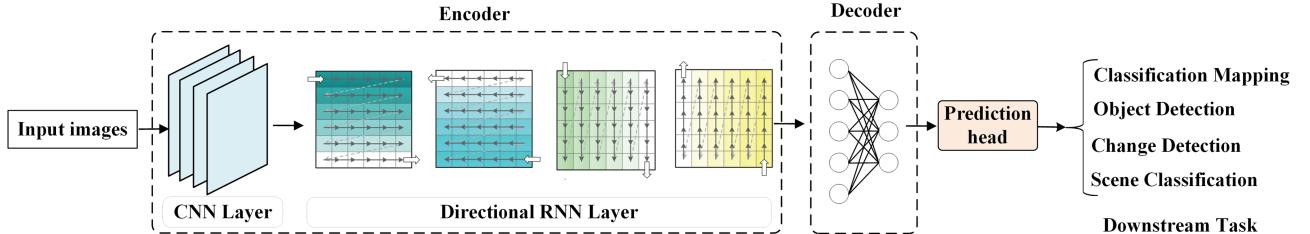


Fig. 13. Structure of RNN Model.

the data to achieve effective modeling and prediction of sequence data [81] (see Fig. 13). In remote sensing, RNNs are frequently applied to temporal data and hyperspectral image interpretation due to their advantages in handling time series images. Common RNN models include long short-term memory (LSTM) network and gated recurrent units [82].

Based on existing literature, RNN models addressing the class imbalance in remote sensing can be classified into CNN-RNN, Attention-RNN, and Transformer-RNN. CNN-RNN combines the strengths of CNN and RNN, where the model first employs CNNs to extract spatial features from images and then inputs these features into an RNN, incorporating designed loss functions or sampling strategies for sequential modeling, such as the Densenet121-LSTM [62] and GRMA-Net [83] models. Attention-RNN integrates the attention mechanism with RNN, enabling the RNN model to focus more on significant information. In the presence of class imbalance, Attention-RNN can enhance the recognition rate of minority classes by allocating them

greater attention, thereby improving overall classification performance, such as the ARNet [84] model. Transformer-RNN is an RNN model based on the self-attention mechanism, which is able to deal with the long-range dependence problem, and it can maintain the sequence processing capability while effectively dealing with class imbalance in remote sensing images, such as the RNN-Transformer [85] model.

### B. Improvement of the Loss Function

Commonly used loss functions in deep learning include binary cross-entropy loss (BCE), cross-entropy loss (CE) [86], Dice loss (DiL) [87], focus loss (FL) [88], L1 loss (L1L) [89], and so on. The basic idea of optimizing the loss function is to adjust the weight and importance of each class by designing an excellent balanced loss function to balance the loss contribution of each class [42]. Methods based on loss functions are among the most commonly used approaches to address class imbalance

TABLE IV  
REPRESENTATIVE LOSS FUNCTIONS FOR CLASS IMBALANCE

Category	Abbreviation	Loss function	Formulation	Articles
Weighting-Based Loss	WCEL	Weighted Cross-Entropy Loss	$WCEL = -\sum_{i=1}^N w_i \cdot y_i \log(p_i)$	[90]
	SWBCEL	Self-Adaptive Weighted Binary Cross-Entropy Loss	$SWBCEL = -\frac{1}{N} \sum_{i=1}^N [w_1 y_i \log(p_i) + w_2 (1 - y_i) \log(1 - p_i)]$	[91]
	MFBL	Median-frequency balance loss	$MFBL = -\sum_{i=1}^N \frac{1}{\log(1+\mu_k)} \cdot y_i \log(p_i)$	[44]
Adjustment Factor-Based Loss	FL	Focal Loss	$FL = -\sum_{i=1}^N (1 - p_i)^\gamma \cdot y_i \log(p_i)$	[88]
	CFL	Calibrated Focal Loss	$CFL = -\alpha \cdot w_c \log(p_i) + FL$	[92]
	CPL	Collaborative Probability Loss	$CPL = -y_0 \log p_b p_{m_0} - \sum_{i=1}^{N-1} y_i \log p_{m_i} - \sum_{i=1}^{N-1} y_i \log(1 - p_b) + \log Z$	[93]
Similarity Index-Based Loss	DiL	Dice Loss	$DiL = 1 - \frac{2 \sum_{i=1}^N y_i p_i + 1}{\sum_{i=1}^N y_i + \sum_{i=1}^N p_i + 1}$	[87]
	LCDiL	Log-Cosh Dice Loss	$LCDiL = \log\left(\frac{e^{DiL} + e^{-DiL}}{2}\right)$	[94]
	TL	Tversky Loss	$TL = 1 - \frac{\sum_{i=1}^N y_i p_i + 1}{\sum_{i=1}^N y_i p_i + \alpha \sum_{i=1}^N y_i (1 - p_i) + (1 - \alpha) \sum_{i=1}^N (1 - y_i) p_i + 1}$	[95]
Combined Loss	CL	Combined Loss	$CL = \lambda_1 \cdot DiL + \lambda_2 \cdot BCE$ $CL = FL + \lambda \cdot CEL$ $CL = BCE + TL$	[96] [97] [98]

in remote sensing image interpretation. The purpose of these loss functions is to improve the performance and generalization ability of the model by adjusting the weights of the categories or samples so that the model can pay more attention to those fewer but important classes or hard-to-categorize classes. We classify the existing loss functions for dealing with class imbalance problem into four types, including weighting-based loss, adjustment factor-based loss, and similarity index-based loss, and combined loss function. Table IV summarizes some representative loss functions for dealing with class imbalance problem.

1) *Weighting-Based Loss*: Weighting-based loss functions address the class imbalance by assigning different weights to different classes, and usually, classes with minority samples are given higher weights. WCEL [90] builds upon the CEL function by incorporating a weight coefficient for each class, resulting in higher weights for minority classes and lower weights for majority classes. SWBCEL [91] imposes weight coefficients on the BCE function to constrain the network, thus focusing training more on foreground regions. MFBL [44] increases the weights accounted for by a few classes by adding the log-inverse value of the ratio of the number of pixels of each class to the number of pixels of all markers to CEL.

2) *Adjustment Factor-Based Loss*: Adjustment factor-based loss functions adjust the loss values by introducing an adjustment factor into the base loss to better handle imbalanced data. FL [99] introduces a modulation factor to the CEL, enabling the model to focus more on misclassified samples during training, thus alleviating the imbalance between positive and negative samples. CFL [92] is a loss function that incorporates a calibrated term factor into the FL. The calibration term is essentially a pixel-by-pixel WCEL function whose weights are determined according to the proposed PCM. CFL forces the network to pay more attention to complex samples by

suppressing the loss of simple samples. CPL [93] is a loss function that integrates dual-branch outputs within the same probability framework, achieving critical feature extraction for the foreground by incorporating a normalization factor into the CEL. Dynamic gradient adjustment loss (DGAL) [100] introduces a gradient harmonization mechanism into FL, enhancing the recognition ability of foreground objects. Similarly, class imbalance loss (CIL) [101] improves FL by considering the distribution differences of class samples and introducing a more complex adjustment factor based on trigonometric functions to regulate the model training process. Balanced focal loss (BFL) [102] redefines the class balance factor in FL, incorporating the ratio of positive and negative samples, and can address both positive-negative imbalance and class imbalance problem simultaneously.

3) *Similarity Index-Based Loss*: Similarity index-based loss functions utilize the similarity index to measure the similarity between samples, thus accounting for intra- and interclass differences in the loss function. DiL [103] utilizes the Dice coefficients to measure the degree of overlap between predicted and true regions. The Dice coefficient provides a balanced evaluation in the context of data imbalance and is commonly employed in binary semantic segmentation tasks. LCDiL [94] combines DiL with the Log-Cosh loss function, aiming to deal with sparse segmentation in unbalanced datasets. TL [95] is a generalization of the DiL function that allows for adjusting hyperparameters to control the tradeoff between false positives and false negatives, making TL particularly effective for handling imbalanced datasets.

4) *Combined Loss*: Constituting a joint loss function by integrating the above-related loss functions is also a current common method to address the class imbalance problem. Joint loss by combining the advantages of multiple loss functions in order to improve the processing capability of imbalanced

TABLE V  
DATA AUGMENTATION PREPROCESSING METHODS

Category	Data Augmentation Methods
Changes in Spatial Location Information	Translation, Cropping, Rotation, Mosaic [105], Mixup [106], CutMix [107], AugMix [108], GridMix [109], PuzzleMix [110], ReMixMatch [111], HRMix [112], Copy-Paste[113],etc.
Changes in Image Pixels	Adjust the Color Space of Image [114], Color Jitter [115], Add Noise [116], etc.
Deletion of Image Information	Cutout [117, 118], Random Erasing [119], Hide-and-Seek [120], GridMask [121], etc.
GAN-Based Data Generation	Pix2Pix[122], StarGAN[123], CycleGAN[124], GanGAN[125], etc.

data, such as DiL+BCEL [96], FL+CEL [97], BCE+TL [98], WBCEL+DiL[49], Smooth Loss + CEL[74], and so on.

### C. Data Augmentation Preprocessing

In remote sensing image interpretation, one key to addressing the class imbalance problem is to use data augmentation methods to improve the model's generalization ability and classification performance by appropriately transforming and expanding the training data to balance the distribution of samples between different classes [104]. These augmentation techniques include changes in spatial location information, changes in image pixels, deletion of image information, and data generation based on GANs. As given in Table V, by introducing various data augmentation techniques, the original dataset can be effectively expanded to generate new training samples, thereby increasing the number of minority class samples and enhancing the model's ability to recognize the minority class.

1) *Changes in Spatial Location Information*: By applying affine transformations such as Translation, Cropping, and Rotation, as well as mixed methods such as Mosaic, Mixup, and CutMix, the spatial location information of samples can be altered. This approach simulates remote sensing images from different angles and scales, which enhances the model's recognition capabilities across various perspectives and increases the diversity of training samples [126].

Mosaic [105] method first selects four different original images and then splices parts of the images and labels information onto the newly generated blank image according to certain rules. In this process, an attempt is made to increase the frequency of minority class occurrences and reduce the model's underfitting of the tail instances. Mixup [106], [127], [128] generates new virtual samples by linearly interpolating the features and labels of the two training samples. This design approach can introduce a priori knowledge to the model that the linear interpolation of features corresponds to the linear interpolation of labels, which reduces the overfitting of the model to the training samples and improves the model's generalization abilities and classification performances. CutMix [107], [129] generates new samples by partially regionally cropping and splicing features and labels of two training samples. It offers advantages such as increased

sample diversity, retention of semantic information in images, and addressing class imbalance problem.

2) *Changes in Image Pixels*: Differentiated training samples can be generated by adjusting the color space of the image, applying color perturbation, and adding noise [130]. This approach simulates images under varying lighting conditions, enhancing the model's adaptability to changes in illumination. Color space augmentation is used to control the brightness or darkness of the image and can be applied to each channel value to increase the number of training samples [114]. Color perturbations involve randomly altering the image's brightness, contrast, saturation, and hue, enabling the model to adapt to variations in lighting conditions [115]. Injecting noise is another data augmentation technique that aids neural networks in learning robust features, making them more resilient against adversarial attacks [116], [131].

3) *Deletion of Image Information*: Removing part of the image information by random deletion, occlusion, blur, and other methods, or removing some local information of the object by occlusion and deletion, can help the model better learn the essential characteristics of the data [130]. Common processing methods include Cutout [117], [118], Random Erasing [119], Hide-and-Seek [120], GridMask [121], and so on.

The core idea of Cutout [117], [118] is to randomly delete a rectangular region in the image and fill it with a pixel value of 0 to simulate occlusion. It can be considered a particular form of Dropout, with the critical difference being that the input is changed from features to images. This method offers advantages such as reducing overfitting, increasing model robustness, and being simple to implement. The Random Erasing [119] method operates on a similar principle to Cutout but the erased region's length and width are random as are the pixel values used for filling. This approach enriches the diversity of the training data, which helps improve the model's generalization abilities. The basic idea of the Hide-and-Seek [120] method is to randomly divide the image into grids of uniform size and erase part of the grids at a certain rate during the model training process. However, it inevitably leads to the complete erasure of some small objects, forcing the deep network to learn relevant features from a global perspective. The augmentation methods mentioned above select masked regions randomly, which can result in either completely erasing or fully retaining important locations. To balance this problem, GridMask [121] generates a structured grid and then removes the image information in the grid. This process is equivalent to adding a regularization term to the network, which avoids the overfitting of the network and only needs to be augmented at data input without changing the network architecture. FenceMask, proposed by Li et al. [132], is a regular sparse grid mask method, which can efficiently retain small object classes in the image by uniformly obscuring the whole image with rectangles similar to the shape of a fence.

4) *GAN-Based Data Generation*: GAN and its variants have been extensively researched and applied to the automatic interpretation of remote sensing imagery [133], [134], [135], [136]. GAN consists of two components: generator and discriminator. The generator is responsible for generating data, while the discriminator determines whether the input data are real or

TABLE VI  
SOME BENCHMARK DATASETS IN REMOTE SENSING IMAGE INTERPRETATION

Remote Sensing Task	Benchmark Dataset	Description
Classification Mapping	ISPRS Vaihingen [141] (Multi-class)	The dataset consists of 33 images with a spatial resolution of 0.9 m, and each image has an average size of $2494 \times 2064$ pixels. The corresponding labels consist of five foreground classes and one background class. Semantic segmentation is commonly performed on TOP images composed of near-infrared, red, and green bands.
	iSAID[142] (Multi-class)	The dataset consists of 2806 high-resolution images, containing 15 classes and a total of 655 451 object instances. The image sizes range from $12028 \times 5014$ to $455 \times 387$ pixels.
	Massachusetts Roads and Buildings [143] (Binary-class)	The road dataset consists of 1171 aerial images, each with a size of $1500 \times 1500$ pixels. The road scenes cover urban, suburban, and rural areas. The building dataset consists of 151 aerial images, each with a size of $1500 \times 1500$ pixels.
Object Detection	HRSID [144] (Binary-class)	The dataset consists of 5604 high-resolution SAR images, containing 16,951 ship instances.
	SSDD[145] (Binary-class)	The dataset consists of 1160 SAR images with spatial resolutions ranging from 1 to 15 m, containing 2456 ship objects.
	UCAS-AOD [146] (Multi-class)	The dataset consists of 2420 images, containing 14,596 instances of airplanes and cars.
	NUDT-AOSR15 [101] (Multi-class)	The dataset consists of 98 images with a spatial resolution of 0.6 m, each with a size of $12544 \times 12544$ pixels, and contains 15 ship classes.
Change Detection	LEVIR-CD [147] (Binary class)	The dataset, used for building change detection, consists of 637 images with a spatial resolution of 0.5 m, and each image has a size of $1024 \times 1024$ pixels.
	DigitalGlobe [148] (Binary-class)	The dataset consists of 16,000 very high-resolution images with spatial resolutions ranging from 0.3 to 1 m, containing a large variety of change features across different classes and scales.
	xBD [149] (Binary-class)	The dataset consists of 22,068 high-resolution images, each with a size of $1024 \times 1024$ pixels, containing before-and-after images of changes across 19 different events.
	WHU-CD [150] (Binary-class)	The dataset consists of two images with a spatial resolution of 0.2 m, each with a size of $15\,354 \times 32\,507$ pixels, containing 220 000 buildings.
Scene Classification	UCM [151] (Multi-label)	The dataset consists of 2100 high-resolution images, each of which is $256 \times 256$ in size, and contains 21 scene classes with 100 remote sensing images in each class.
	HR-SAR [152] (Multi-label)	The dataset contains 46 collections from urban and nonurban areas, with each SAR image having a pixel spacing of 1.25 m, a spatial resolution of approximately 2.9 m, and a size of $160 \times 160$ pixels.
	EuroSAT [153] (Single-label)	The dataset consists of 27 000 images with a spatial resolution of 10 m, each with a size of $64 \times 64$ pixels, and contains 10 different land cover categories.

“Binary-class” refers to images focusing on only two classes, and “multi-class” refers to images focusing on three or more classes. “Single-label” refers to describing an image scene with a single class, while “multi-label” refers to describing an image scene with two or more classes.

generated by the generator. In the presence of class imbalance, the generator can be trained to generate more minority class samples, thereby balancing the class distribution in the dataset.

Conditional generative adversarial networks, as an extension of GAN, have been applied in various models, including Pix2Pix, CycleGAN, StarGAN, and supervised/semisupervised GAN [137]. Chen et al. [133] addressed the issues of data scarcity and class imbalance in building change detection by first using an improved GanGAN to generate a large number of building objects. The newly generated samples were then added to the original dataset, and the proposed CDNet model was used for change detection. Experimental results confirmed the effectiveness of this approach. Shumilo et al. [138] first used Pix2Pix to perform remote sensing data augmentation and then used UNet to perform semantic segmentation. Xue et al. [139] used an improved StarGAN to generate disaster images and then used Siamese-UNet for segmentation evaluation. Li et al. [140] proposed a GAN-based image-level sample pair generation model (ISPG) to address the performance deficiencies in building change detection caused by data

scarcity and class imbalance. ISPG combines a newly designed label-translation GAN, multiscale adversarial loss, and feature-matching loss to improve the quality of remote sensing image generation.

#### IV. ACCURACY ANALYSIS ON BENCHMARK DATASETS

##### A. Benchmark Datasets

In recent years, various remote sensing image interpretation datasets have been released in the Earth observation community, and researchers have conducted numerous experiments based on these datasets. However, most of these datasets exhibit class imbalance problem. Table VI summarizes several widely used benchmark datasets, with data types including multispectral and SAR imagery, and features rich class diversity and scene variety. These datasets cover four types of tasks: classification mapping, object detection, change detection, and scene classification. In Section IV-C, methods for addressing class imbalance problem will be compared and analyzed based on these datasets.

TABLE VII

EVALUATION METRICS, TP, TN, FP, AND FN DENOTE TRUE POSITIVES, TRUE NEGATIVES, FALSE POSITIVES, AND FALSE NEGATIVES, RESPECTIVELY

Evaluation Metrics	Range of Application	Description	Formulas
Precision	Classification Mapping; Object Detection; Change Detection; Scene Classification	Precision refers to the proportion of detected positive samples that are actually positive.	$Precision = \frac{TP}{TP+FP}$
Recall	Classification Mapping; Object Detection; Change Detection; Scene Classification	Recall refers to the proportion of detections in samples that are actually positive.	$Recall = \frac{TP}{TP+FN}$
F1 or mF1	Classification Mapping; Object Detection; Change Detection; Scene Classification	F1 is the reconciled average of Precision and Recall. mF1 is the mean of all class F1.	$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$ $mF1 = \frac{1}{N} \sum_{i=1}^N F1_i$
IoU or mIoU	Classification Mapping; Object Detection; Change Detection	IoU is the ratio of the intersection and concatenation of the predicted and true labels. The mIoU is the mean of all classes of IoU.	$IoU = \frac{\text{Area of Intersection}}{\text{Area of Union}}$ or $IoU = \frac{TP}{TP + FP + FN}$ $mIoU = \frac{1}{N} \sum_{i=1}^N IoU_i$
AP or mAP	Object Detection;	AP is an evaluation metric calculated based on the area under the precision-recall curve for each class. The mAP is the average of AP values across different classes. It is commonly used to assess the overall performance of multiclass object detection tasks.	$AP = \int_0^1 P(r) dr$ $mAP = \frac{1}{N} \sum_{i=1}^N AP_i$
OA or ACC	Classification Mapping; Object Detection; Change Detection; Scene Classification	The calculation formulas for OA and ACC are the same, both being computed as the ratio of the number of correctly classified classes (or images) to the total number of classes (or total number of images).	$OA, ACC = \frac{TP}{TP + FP + TN + FN}$

TABLE VIII  
PERFORMANCE COMPARISON OF DIFFERENT MODELS ON THE ISPRS VAIHINGEN DATASET

Type	Method	F1					mF1	OA	mIoU
		Imp. Surf.	Building	Low. veg.	Tree	Car			
CNN	SPANet [154]	93.5	96.2	86.8	90.9	90.6	91.6	91.8	83.8
	IKD-Net [155]	96.1	90.5	<b>87.2</b>	<b>92.0</b>	92.5	91.7	92.1	-
	MFAFNet [156]	97.0	95.8	85.5	90.5	89.2	91.6	93.7	84.8
Transformer	CIAPNet [157]	<b>97.1</b>	96.2	85.9	90.3	91.5	<b>92.2</b>	<b>93.8</b>	<b>85.7</b>
	RSSFormer [52]	93.7	<b>96.9</b>	81.3	91.8	89.2	90.6	90.8	-
	DD-Transformer [158]	93.2	95.5	84.6	89.8	88.9	90.4	91.4	82.7
GNN	PGR-DL [159]	93.7	95.8	86.5	90.8	90.1	91.4	91.8	84.2
	SGA-Net [160]	93.2	95.5	82.6	88.4	<b>92.8</b>	90.5	-	82.6
	SCG-GCN [161]	92.4	94.8	83.9	89.7	88.0	89.8	90.4	-
RNN	FSFM-PHN [162]	96.9	95.9	84.3	89.9	90.6	91.5	93.3	84.7

"Imp. Surf." refers to impervious surfaces, and "Low. veg." refers to low vegetation. The bold numbers represent the best values for the corresponding evaluation metrics. (Unit: %).

### B. Evaluation Metrics

In remote sensing image interpretation, various metrics need to be utilized to evaluate the extent to which different methods address the class imbalance problem. Table VII summarizes the quantitative metrics employed in this study to evaluate the performance of different methods. These metrics include Precision, Recall, F1 score, mF1 score, Intersection over Union (IoU), Mean IoU (mIoU), Average Precision (AP), Mean AP (mAP), Overall Accuracy (OA), and Accuracy (ACC). F1, IoU, and AP

are used for binary-class or single-label datasets, whereas for multi-class or multi-label datasets, mF1, mIoU, and mAP are considered.

### C. Benchmark Dataset Experimental Results and Analysis

1) *Performance Comparison of Representative Models*: Table VIII summarizes the performance of several representative deep-learning semantic segmentation models on the ISPRS Vaihingen dataset, where "car" is the minority class. Due to the

TABLE IX  
PERFORMANCE COMPARISON OF DIFFERENT LOSS FUNCTIONS

Remote Sensing Task	Method	Benchmark Dataset	Loss Function	IoU or mIoU	F1 or mF1	OA or ACC	AP or mAP
Classification Mapping	FCN [92]	ISPRS Vaihingen	CEL	78.3	86.6	88.9	-
			CFL*	<b>79.0</b>	<b>87.5</b>	<b>89.2</b>	-
	FactSeg [93]	iSAID	CEL	64.0	-	-	-
			CPL*	<b>64.8</b>	-	-	-
	D-LinkNet [94]	Massachusetts Roads	BCE	61.3	76.0	97.9	-
			DiL	63.4	77.6	<b>98.2</b>	-
			LCDiL	<b>63.8</b>	<b>77.9</b>	<b>98.2</b>	-
			TL	61.3	76.0	98.0	-
			CL (BCE + DiL)	62.3	76.7	98.1	-
Object Detection	CenterNet [100]	UCAS-AOD	L1L	-	-	-	95.6
			FL	-	-	-	95.9
			DGAL*	-	-	-	<b>96.4</b>
	R2CNN_Res101 [101]	NUDT-AOSR15	CEL	-	77.9	-	79.1
			CIL*	-	<b>80.9</b>	-	<b>82.9</b>
Change Detection	CDNet [133]	LEVIR-CD	CEL	-	87.5	-	-
			WCEL	-	87.9	-	-
			FL	-	85.6	-	-
			DiL	-	<b>88.1</b>	-	-
	AERNet [91]	LEVIR-CD	BCE	90.0	89.6	-	-
			SWBCE*	<b>91.1</b>	<b>90.8</b>	-	-
	MFCN [49]	DigitalGlobe	BCE	85.4	85.4	96.6	-
Scene Classification	CAGRN [102]	UCM	FL	87.3	87.6	97.1	-
			WBCEL	88.6	88.9	97.4	-
			CL* (WBCEL + DiL)	<b>90.7</b>	<b>91.1</b>	<b>98.0</b>	-
	ResNet-18 [74]	HR-SAR	BCE	-	95.3	-	-
			FL	-	95.7	-	-
			BFL*	-	<b>96.0</b>	-	-
			CEL	-	70.9	-	-
			CL* (Smooth Loss + CEL)	-	<b>72.0</b>	-	-

“\*” indicates loss functions newly proposed in the corresponding papers. CEL, BCE, and L1L are traditional loss functions, while the other losses are class rebalancing loss functions. For dataset information, refer to Table VIII. The bold numbers represent the best values for the corresponding evaluation metrics (Unit: %).

limited number of experiments involving RNN-based models on this dataset, only one such model is listed in Table VIII. Overall, all four types of models demonstrate high classification performance on this dataset. Among them, the Transformer-based CIAPNet model achieves the highest mF1, OA, and mIoU values of 92.2%, 93.8%, and 85.7%, respectively. In descending order of performance, other models include the CNN-based MFAFNet model, the RNN-based FSFM-PHN model, and the GNN-based PGR-DL model. In terms of mF1, CIAPNet outperforms the other models by 0.6%, 0.7%, and 0.8%, respectively. Regarding the F1 score for each class, all models achieve results above 80%, indicating strong performance. Specifically, the highest F1 score for the “car” class is achieved by SGA-Net at 92.8%, followed by IKD-Net and CIAPNet, with scores of 92.5% and 91.5%, respectively.

A careful review of these articles reveals that these models, to varying extents, incorporate one or more of the following structures: multiscale feature extraction and fusion, short- and

long-range context modeling, and attention mechanisms. Models that integrate these structures significantly improve the interpretation ability of minority classes, thereby alleviating the impact of class imbalance. Therefore, when designing models to address the class imbalance problem, researchers should appropriately integrate these effective structures.

2) *Comparison of Different Class Rebalancing Loss Functions:* Table IX summarizes the impact of traditional loss functions and various class rebalancing loss functions on interpretation results across different benchmark datasets. In classification mapping tasks, the experimental results on the ISPRS Vaihingen dataset reported by Bai et al. [92] show that the proposed CFL function effectively enhances model performance. Compared to CEL, CFL improves mIoU, mF1, and OA by 0.7%, 0.9%, and 0.3%, respectively. Similarly, on the iSAID dataset, the CPL function proposed in [93] improves mIoU by 0.8% compared to CEL. In [94], performance tests of different class rebalancing loss functions on the Massachusetts Roads

dataset reveal that LCDiL yields the best results. Overall, these types of loss functions demonstrate a clear advantage over BCEL in optimizing model performance. Compared to BCEL, DiL, LCDiL, and CL improve the F1 score by 1.6%, 1.9%, and 0.7%, respectively.

In object detection tasks, the DGAL function proposed in [100] outperforms both FL and L1L. Experiments on the UCAS-AOD dataset show that compared to L1L, both FL and DGAL improve mAP by 0.3% and 0.8%, respectively. Subsequently, on the NUDT-AOSR15 dataset, the CIL function proposed in [101] improves mAP by 3.8% compared to CEL.

In change detection tasks, Chen et al. [133] compared the performance of CEL with other class rebalancing loss functions on the LEVIR-CD dataset. The results show that, except for FL, all other loss functions achieved improved F1 scores, with DiL showing the best performance, increasing by 0.6%. Zhang et al. [91] demonstrated through experiments on the LEVIR-CD dataset that the proposed SWBCEL outperforms BCEL, with improvements of 1.1% and 1.2% in IoU and F1, respectively. Li et al. [49] compared the performance of different loss functions on the DigitalGlobe dataset, and the results indicate that the newly proposed CL function performs the best. Compared to BCEL, CL improved IoU, F1, and OA by 5.3%, 5.7%, and 1.4%, respectively, while other class rebalancing loss functions also showed some improvements.

In scene classification tasks, Bi et al. [102] validated the newly proposed BFL function on the UCM dataset. The results show that, compared to BCEL, both FL and BFL improve F1 by 0.4% and 0.7%, respectively. Similarly, Huang et al. [74] conducted experiments on the HR-SAR dataset, and the results indicate that the proposed CL function outperforms CEL, with an improvement of 1.9% in F1.

From the above experimental analysis, it is evident that class rebalancing loss functions significantly outperform traditional loss functions in addressing the class imbalance problem. Although the performance of each loss function varies across different methods and datasets, there is an overall trend of gradual improvement.

*3) Comparison of Data Augmentation Methods:* Table X summarizes the application effects of different data augmentation methods in class-imbalanced image interpretation tasks. In classification mapping tasks, the experimental results by Pereira and Dos Santos [163] and Gong et al. [164] on the ISPRS Vaihingen dataset show that the newly proposed ChessMix and MSAug data augmentation methods significantly improve classification results, with mIoU increasing by 1.2% and 6.1%, respectively. Similarly, Lv et al. [165] validated the effectiveness of the G-sGAN method on the Massachusetts Buildings dataset, where F1, IoU, and OA improved by 0.9%, 1.2%, and 0.3%, respectively.

In object detection tasks, the experiments by Suo et al. [113] and Guo and Zhou [166] on the SSDD dataset show that the MDBA and BoxPaste data augmentation methods outperform traditional affine transformations (such as flipping and rotation) in terms of performance improvement. Specifically, MDBA and BoxPaste methods increase AP by 0.8% and 1.9%, respectively. In addition, the experiments by Guo et al. [167] on the HRSID

dataset demonstrate that the GMDA method outperforms affine transformations, with GMDA increasing AP by 1.7%, while affine transformations only led to a 0.2% improvement.

In change detection tasks, the experimental results by Rui et al. [139] on the xBD dataset show that the DisasterGAN method performs the best, followed by CutMix and affine transformation methods. These three methods lead to F1 score improvements of 7.8%, 5.2%, and 4.0%, respectively. In addition, the experiments by Oubara et al. [168] on the WHU-CD dataset demonstrate that the proposed c-GAN method significantly enhances the model's detection performance, with F1, IoU, and OA improving by 16.5%, 21.5%, and 0.4%, respectively. Similarly, Chen et al. [133] on the LEVIR-CD dataset shows that the GAN-based IAug method improves the detection F1 score by 1.5%.

In scene classification tasks, the experimental results by Lu et al. [169] on the EuroSAT dataset demonstrate that the newly proposed RSI-Mix method outperforms the Cutout method, with ACC increasing by 1.2% and 0.3%, respectively. Subsequently, the experiments by Pan et al. [170] and Han et al. [171] on the UCM dataset show that the GAN-based Diversity-GAN and GAN-RSIGM-VAT methods significantly outperform affine transformation methods. These three methods achieve OA improvements of 5.2%, 11.8%, and 0.3%, respectively.

The above analysis shows that image interpretation results with data augmentation methods are significantly better than those without data augmentation. Specifically, GAN-based augmentation methods yield the best performance, followed by methods such as CutMix and Cutout, while affine transformation methods show relatively lower effectiveness. Therefore, to improve the recognition ability of minority classes or small objects, GAN-based data augmentation methods should be prioritized.

## V. MAIN APPLICATION FIELDS INVOLVED IN CLASS IMBALANCE IN AUTOMATIC INTERPRETATION OF REMOTE SENSING IMAGES

### A. Classification Mapping

In remote sensing image classification mapping, the uneven distribution of different geographic objects on the Earth's surface directly leads to significant disparities in the number of samples for each class in the training data, making class imbalance a common problem [172]. Therefore, addressing the class imbalance problem is crucial for improving the accuracy and reliability of mapping. The existing classification mapping is mainly based on deep semantic segmentation technology. It adopts feature down-sampling encoder and feature up-sampling mapping decoder architectures to assign a specific category to each image pixel to achieve accurate recognition and localization of objects in the entire image [173]. Commonly used frameworks include CNN, RNN, GAN, and Transformer.

Satellite imagery is affected by complex geographic spatial distributions, sensor variations, and seasonal influences, which increase intraclass and interclass differences and exacerbate foreground–background class imbalance, limiting the segmentation of small objects [3]. Zheng et al. [174] proposed the FarSeg

TABLE X  
PERFORMANCE COMPARISON OF DIFFERENT DATA AUGMENTATION METHODS

Remote Sensing Task	Method	Benchmark Dataset	Data Augmentation	Without (w/o) or With (w)	F1 or mF1	IoU or mIoU	OA or ACC	AP or mAP
Classification Mapping	FCN-Res50 [163]	ISPRS Vaihingen	- ChessMix <sup>*, 1</sup>	w/o w	- -	60.1 <b>61.3</b>	84.8 <b>85.6</b>	- -
	PFNet [164]	ISPRS Vaihingen	- MSAug <sup>*, 1, 3</sup>	w/o w	76.8 <b>82.3</b>	65.1 <b>71.2</b>	- -	- -
	UNet++ [165]	Massachusetts Buildings	G-sGAN <sup>*, 4</sup>	w/o w	71.7 <b>72.6</b>	57.0 <b>58.2</b>	89.8 <b>90.1</b>	- -
Object Detection	YOLOX-Nano [166]	SSDD	- Rotation, Flipping <sup>1</sup> MDBA <sup>*, 1, 3</sup>	w/o w w	89.9 90.4 <b>91.5</b>	- - -	- - -	90.7 91.0 <b>91.8</b>
	ATSS [113]	SSDD	- BoxPaste <sup>*, 1</sup>	w/o w	- -	- -	- -	94.4 <b>96.3</b>
	GMDR-Net [167]	HRSID	- Rotation, Flipping <sup>1</sup> GMDA <sup>*, 1</sup>	w/o w w	- - -	- - -	- - -	74.6 74.8 <b>76.3</b>
Change Detection	Siamese-UNet [139]	xBD	- Rotation, Flipping, Cropping <sup>1</sup> CutMix <sup>1</sup> DisasterGAN <sup>*, 4</sup>	w/o w w w	70.3 74.3 75.5 <b>78.1</b>	- - - -	- - - -	- - - -
	Adv-CDNet [168]	WHU-CD	- Copy-Paste, Rotation, Color Jitter <sup>1, 2</sup> c-GAN <sup>*, 4</sup>	w/o w w	65.3 69.6 <b>81.8</b>	50.4 53.4 <b>71.9</b>	98.2 98.2 <b>98.6</b>	- - -
	CDNet [133]	LEVIR-CD	- IAug <sup>*, 4</sup>	w/o w	87.5 <b>89.0</b>	- -	- -	- -
Scene Classification	Resnet-50 [169]	EuroSAT	- Cutout <sup>3</sup> RSI-Mix <sup>1</sup>	w/o w w	- - -	- - -	96.8 97.1 <b>98.0</b>	- - -
	VGG-16 [170]	UCM	- Rotation, Flipping <sup>1</sup> Diversity-GAN <sup>*, 4</sup>	w/o w w	- - -	- - -	89.7 90.0 <b>94.9</b>	- - -
	VGG-10 [171]	UCM	- GAN-RSIGM-VAT <sup>*, 4</sup>	w/o w	65.8 <b>77.7</b>	- -	66.2 <b>78.0</b>	- -

“\*\*” indicates newly proposed data augmentation methods. “1, 2, 3, 4” correspond to: changes in spatial location information, changes in image pixels, deletion of image information, and GAN-based data generation. For related dataset information, refer to Table VIII. Bold values represent the best performance for each method (Unit: %).

model, which contains two core modules: first, the foreground–background relationship module that exploits symbiotic relationships between geographic objects to enhance foreground feature recognition, and second, the foreground-aware optimization module that uses dynamic weight loss strategies to suppress the background and focus on the foreground. Niu et al. [159] proposed a semantic segmentation framework that combines graph inference and decoupling learning to address the foreground–background imbalance. Combining the two can make full use of the multiscale information of the image to explicitly model the foreground and the object boundaries. Ma et al. [93] proposed the FactSeg model to address the small object and foreground–background imbalance problem. It first enhances foreground features through foreground activation and, second, designs a small object mining strategy that automatically selects effective samples for training to improve the recognition of small objects.

Li et al. [175] proposed the CSRL-Net model, which introduces a jointly weighted loss function that combines a reweighting and logarithmic adjustment strategy that allows the model to balance the learning bias between classes by focusing more on the tail classes during the training process. Similarly, the RSSFormer [52] model incorporates a foreground saliency-guided loss to direct the network’s attention towards hard samples with lower foreground saliency responses, achieving balanced optimization across classes. Xu et al. [176] introduced the unsupervised domain adaptation CaGAN model, which can explicitly align images from the source and target domains while effectively alleviating class imbalance. The ResUNet-a model proposed by Diakogiannis et al. [177] uses geometric data augmentation, such as rotation and scaling and an improved DL function to address the class height imbalance problem in semantic segmentation.

However, the aforementioned methods are largely based on optical remote sensing images, which face limitations in obtaining high-quality data under complex weather conditions (e.g., clouds, fog) or during nighttime, severely affecting their applicability. In contrast, SAR has demonstrated significant advantages in extreme environments due to its active imaging capabilities [178]. Polarimetric SAR (PolSAR), in particular, captures more high-resolution Earth observation object information through various polarization modes and has gradually been applied to remote sensing image classification tasks [179], [180], [181]. Nevertheless, SAR-based classification also faces challenges such as label scarcity, noise interference, and class imbalance. To address these problems, Kuang et al. [182] proposed a PolSAR image contrastive learning (PiCL) method based on domain knowledge of PolSAR. In PiCL, the positive sample learning strategy is employed to achieve domain-specific representation learning, while pseudolabel generation based on contextual consistency and weighted feature-level synthetic data oversampling techniques are used to address the class imbalance problem. Similarly, Wang et al. [183] proposed a comprehensive solution to the aforementioned issues in PolSAR image classification. This solution includes using a label correction mechanism to resolve the noise label dilemma, a self-distillation-based contrastive learning method to alleviate the limitation of sparse labels, and the introduction of a sample rebalancing loss function to enhance the learning ability of minority classes, thereby mitigating the impact of class imbalance. Furthermore, Liu et al. [184] proposed a task-oriented GAN model to address small sample sizes and class imbalance. They first utilize GAN to generate realistic PolSAR samples to increase the training sample size and then apply a specific task network for image classification.

Although methods based on data augmentation, loss functions, and model modules have achieved excellent results in addressing the class imbalance problem, when there are multiple small sample classes in an image, data augmentation methods alone may lead to further class imbalances. Existing loss function-based weighting methods mainly use sample frequencies to assign class weights and do not consider the differences in the classification difficulty of the samples in each class, thus not adapting well to the model's dynamic training process.

### B. Object Detection

Due to the special characteristics of remote sensing data, object classes often present themselves at varying scales [185]. Object detection, as an essential branch of computer vision, has the main task of identifying and locating various geographic objects of interest in an image, which usually includes classification, localization, and detection processes. Traditional object detection algorithms frequently exhibit significant bias when confronted with class imbalance, often leaning toward overfitting to majority classes while neglecting minority classes [38]. This bias not only reduces the generalization ability of the model but also limits its application scope in complex and changing remote sensing scenarios.

Despite the significant progress achieved by deep learning techniques in remote sensing image object detection, the

problem of class imbalance remains one of the bottlenecks hindering its development [186]. In small object detection, this often manifests as an imbalance between the foreground and background. Due to the influence of geometric scale, small objects in the foreground usually occupy very few pixels and are easily disturbed by irrelevant information in the background [174], [187]. To overcome this challenge, researchers have proposed various solution strategies, ranging from data-level augmentation processing to model-level constructs, which attempt to address class imbalance from different perspectives, with model-level approaches showing particular potential due to their direct effect on the learning process [188]. For example, by incorporating attention mechanisms, models can focus more on feature learning for minority classes. In addition, transfer learning can be utilized to transfer knowledge from other domains to the object detection task, thereby mitigating the effects of class imbalance. Applying class rebalancing loss functions can also enhance the model's focus on the learning process of minority classes. Literature [105] incorporates the instance balancing Mosaic method and the class rebalancing loss function in the proposed detection model. The former provides rich tail-class features to the model through resampling to alleviate the imbalance of data distribution. The latter can provide the model with more balanced loss information according to the learning difficulty of the instances, which improves the detection accuracy of the tail instances. Literature [189] integrates the class balancing criterion and adaptive weight loss function in the active learning approach, which is crucial in addressing the long-tail problem. Liu et al. [190] proposed the ABNet model to improve the detection of remote sensing foreground objects by embedding an attentional mechanism, which plays an active role in addressing the scale difference and class distribution imbalance. Rasna et al. [191] utilized transfer learning and data augmentation techniques in their proposed object detection network to capture semantic and contextual information related to imbalanced classes during the training phase. Zhang et al. [192] proposed an entropy-based resampling method to address the class imbalance problem in SAR image object detection. The variance-weighted information entropy introduced in this method considers both the class imbalance and the difficulty imbalance of specific instances, thereby accounting for their impact on object detection. Liu et al. [193] utilized evidence learning to obtain the epistemic uncertainty as a descriptor of sample bias and combined it with contrastive learning to correct the learning bias under intra-class imbalance by using the uncertainty labels of samples. This approach effectively improved the performance of SAR ship detection.

Although various methods have been proposed to address the problem of class imbalance in object detection, such as data augmentation, which alleviates class imbalance to some extent, it may introduce additional noise and overfitting risks. Model-based and loss function-based approaches typically require careful tuning of hyperparameters and exhibit varying sensitivity to different datasets and tasks. Consequently, these methods' generalization ability and robustness still need to be improved.

### C. Change Detection

Remote sensing image change detection is identifying surface changes by analyzing image data of the same area acquired at different points in time [194]. In practical applications, class imbalance may lead to the missed detection of critical changes, such as small areas of environmental damage or early-stage disaster changes. To overcome these problems, Chen et al. [133] and Li et al. [140] first used GAN to generate a large number of bitemporal images containing various buildings, and then trained a simple change detection model on an enhanced dataset, effectively reducing the risk of class imbalance. Literature [195] combines the multiscale attention mechanism and contrast loss function in the proposed MASNet model, which can learn the features of changed and unchanged samples on multiple scales. Zhang et al. [65] proposed a global-awareness-based GAS-Net model to address the problem of the relationship between foreground and background. Given the large proportion of redundant pixels in the unchanged region of the data, the discriminative ability of the less pixel-changed region in the foreground should be focused on in the model construction. In addition, Hou et al. [196] proposed a deep semisupervised learning framework (DCG-CRe), which effectively mitigates the class imbalance problem in SAR image change detection by combining a CNN network, approximate ranking clustering algorithm, and a class rebalancing pseudolabel filtering strategy. Cui et al. [197], in the sample learning and prediction phase of their proposed automated SAR change detection framework, integrated a class-balanced antinoise change detection network. By combining focal loss and mean absolute error loss, they effectively alleviated the impact of class imbalance.

### D. Scene Classification

Remote sensing image scene classification is based on assigning a corresponding semantic class to each scene image block based on the image content [198]. These scene image patches typically refer to local image segments manually cropped from large-scale, semantically clear remote sensing images (e.g.,  $512 \times 512$ ). Scene classification differs from pixel-level and object-level classification in that it does not require fine-grained pixel labels as support. Instead, it only requires assigning a corresponding class name as a semantic label to each scene image patch [199].

Early remote sensing image scene classification primarily focused on single-label scenes, with the goal of assigning a single class to each image [198]. This classification approach overlooked the influence of other classes within the neighborhood and concentrated solely on the classification performance of the target class. With advanced remote sensing technology, high-resolution remote sensing imagery offers rich spectral, shape, and texture features. However, the complexity of the data, the scarcity of labeled samples, and the presence of class imbalance pose challenges for scene classification. Existing methods primarily utilize deep network models combined with loss functions and data augmentation techniques to address the impact of the aforementioned problem on single-label scene classification [200], [201], [202]. Wang et al. [203] proposed

the MGDNet model to address the class imbalance problem in scene classification. The model was designed with a class imbalance pseudolabel selection method to assess the reliability of unlabeled samples, which combined with the designed diversity component feature loss function improves the judgment of local features. Yessou et al. [204] compared the effects of seven loss functions on scene classification, concluding that FL and WCEL are more suitable as loss functions when training datasets exhibit class imbalance. Zhao et al. [205] introduced a hierarchical distillation framework with integrated self-calibrated sampling learning, which significantly improves the feature representation of middle and tail data in remote sensing image scene classification, thereby mitigating the impact of class imbalance. Wang et al. [206] proposed an MGSNet model to address the imbalance between the target and background in scene images. The model employs a target-background separation strategy, using background information as auxiliary data. By incorporating contrastive regularization, it enhances the diversity of features that the network focuses on, thereby improving the separation between target and background information.

However, real-world remote sensing images often contain multiple scene types, and single-label descriptions are no longer sufficient to meet the semantic information needs of land cover. In this context, multilabel remote sensing image scene classification tasks have emerged, aiming to assign multiple scene labels to an input image [207], [208]. Multilabel classification tasks have gained attention in remote sensing image interpretation due to their more comprehensive semantic descriptions and wide applications [209], [210], [211]. Nevertheless, issues such as class imbalance, noise, and label data scarcity continue to hinder progress in multilabel classification tasks [102], [209], [212]. To address these imbalance issues, Bi et al. [102] proposed a CAGRN model, which introduces a novel cross-attention-driven representation learning method. This method dynamically models interclass dependencies using specific relational graphs and explicitly establishes feature-label connections. In addition, a newly designed BFL rebalances the treatment based on the imbalance levels of different classes. Wang et al. [209] adopted a strategy combining data, models, and loss functions to mitigate the impact of class imbalance. First, they employed a soft balancing sampling strategy during the data preprocessing stage. Then, they effectively integrated the extracted multilabel semantic attributes and image features using the proposed MLSFF framework, and adjusted the focus on different classes using focal loss. Ji et al. [62] applied the newly proposed max-pooling loss function and CNN-RNN model to multilabel scene classification, significantly alleviating the class imbalance problem and effectively improving the overall classification performance of the model. Yessou et al. [204] compared the effects of seven loss functions on multilabel scene classification and concluded that, in the presence of class imbalance in the training dataset, FL and WCEL are more convenient loss functions to use.

In summary, remote sensing image scene classification is often influenced by intraclass and interclass differences, imbalanced samples, and a lack of annotated datasets. Existing remote sensing datasets may not encompass all types of surface scenes, especially rare or difficult-to-acquire classes. When deep

networks are trained under such unbalanced data, the result is a decrease in classification accuracy for a minority class due to the poor adaptability of the network. Therefore, how to improve the classification accuracy of minority classes while keeping the performance of majority classes unchanged is the key to subsequent research.

## VI. DISCUSSION

### A. Importance of the Collaboration With Model, Loss Function, and Sample Dataset to Address the Class Imbalance Problem

In semantic segmentation models, incorporating functional modules for short- and long-range contextual modeling, multi-scale feature extraction and fusion, and attention mechanisms is a key technological approach to addressing the class imbalance problem. This can significantly improve the classification of minority classes. Class imbalance is a common problem in remote sensing datasets, where minority classes are often distributed at a local scale with fuzzy boundaries, making it difficult to accurately extract objects using only local features. By introducing long-range dependence modeling, the model can effectively capture global features and achieve precise class boundary delineation with the combination between it and short-range detail modeling. Furthermore, the multiscale feature extraction and fusion within the model is crucial for improving recognition in minority classes. In the dataset, objects often exhibit diverse features along the scale dimension, and minority classes are particularly prone to being masked by majority classes, leading to low recognition rates. When embedding multiscale feature extraction and fusion module, the model can extract detailed semantic information in a variety of scales, thus enhancing its adaptability to objects of different sizes. In addition, the attention mechanism further enhances the model's feature representation for minority classes. The attention mechanism enhances the model's focus on key regions by dynamically assigning weights, especially for capturing semantic features of minority class. For example, in the models listed in Table VIII, they more or less considered all three structures mentioned above. Among them, CIAPNet, which is characterized by all three structures simultaneously, achieved the highest mF1 (92.2%), OA (93.8%) and mIoU (85.7%) among the listed models, and the “car” class F1 also achieved 91.5%.

The loss function plays a crucial role in addressing the CI problem. Through a rational design of loss functions, the model can be effectively guided to focus on the minority class, thus achieving balanced learning across various classes. While the traditional loss function is susceptible to the effects of class imbalance in favor of the majority classes, the newly designed class rebalancing loss and its combined loss function can mitigate the effect of this problem to a certain extent. The experiments listed in Table IX have supported this statement.

The preparation of a sample dataset directly affects the performance of semantic segmentation model in handling the class imbalance problem. Adjusting the class distribution among land cover types to increase the samples of minority classes can significantly improve the recurrence rate for minority classes

in the training course. Data augmentation strategies provide the model with more samples of minority class. Furthermore, the model can achieve more balanced learning of both minority and majority classes in the course of training by optimizing the distribution of them, thus reducing the bias introduced by insufficient samples or imbalanced class distribution. This suggests that the preparation of the sample dataset should not only ensure a reasonable quantity of samples but also focus on balancing class distribution to leverage the model's potential, thereby overcoming the limitations imposed by class imbalance. The experiments listed in Table X further support this statement.

Therefore, addressing the CI problem in semantic segmentation requires a collaborative effort among the model, loss function, and dataset preparation to achieve optimal interpretation results.

### B. Challenges and Prospects

The problem of class imbalance in the automatic interpretation of remote sensing images is an important and complex challenge that directly affects the application of tasks such as classification mapping, object detection, change detection, and scene classification. In reviewing this problem, we find that researchers have proposed various solution strategies, which have their advantages but also have certain limitations. Therefore, research on class imbalance based on deep learning still has considerable room for development, with many challenges remaining in areas such as limited training data, domain shift, model architecture, evaluation metrics for class imbalance, model complexity, and interpretability. In order to obtain better image interpretation results, future research can explore the following aspects.

- 1) Emphasize model innovation and optimization, including the development of class rebalancing loss functions.
- 2) The self-supervised learning method based on domain adaptation plays an active role in addressing the problems of class imbalance and domain shift caused by missing sample datasets. Therefore, the method can be combined with GAN to construct a new model for generating new samples of remote sensing images with real spatial structure to address the class imbalance problem at the data level.
- 3) The model architecture can also incorporate the latest techniques, such as Mamba [213] and SAM [214], to assist in improving the accuracy of minority class samples.
- 4) Developing new evaluation metrics for class imbalance should also be a focus. Effective metrics can accelerate convergence and optimize model parameters during training, whereas current metrics such as F1 and IoU do not clearly reflect the magnitude of fluctuations between classes.

## VII. CONCLUSION

This article aims to investigate the importance of addressing class imbalance in automatic remote sensing image interpretation, and to provide a systematic review covering meta-analysis, methodologies, and application fields. To this end, we first

searched and screened 171 key papers related to the review based on the Wos database for the meta-analysis, covering publication years, high productivity countries, highly cited authors, remote sensing data types, data augmentation methods, and the distribution of the main application fields. On this basis, the solutions to the class imbalance problem in four remote sensing image interpretation fields, namely classification mapping, object detection, change detection, and scene classification, were summarized and discussed in detail. The scope includes model innovation and optimization, loss function improvement, and data augmentation, with deep neural networks being the most widely used at the model level. Experiments on benchmark datasets further demonstrate the effectiveness of these approaches. Finally, we discuss the synergistic relationship between models, loss functions, and data augmentation, summarize the current challenges in this field, as well as propose several ideas for addressing the class imbalance problem.

We hope that this review will be of some assistance to researchers exploring solutions to class imbalance in the automatic interpretation of remote sensing images.

#### ACKNOWLEDGMENT

The authors would like to thank the supercomputing center of Lanzhou University.

#### REFERENCES

- [1] Y. S. Li, X. W. Li, Y. J. Zhang, D. F. Peng, and L. Bruzzone, "Cost-efficient information extraction from massive remote sensing data: When weakly supervised deep learning meets remote sensing big data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 120, 2023, Art. no. 103345.
- [2] A. A. Khan, O. Chaudhari, and R. Chandra, "A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation," *Expert Syst. With Appl.*, vol. 244, 2024, Art. no. 122778.
- [3] C. X. Jing et al., "Interclass similarity transfer for imbalanced aerial scene classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 3502105.
- [4] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, no. 1, pp. 1–54, 2019.
- [5] C. Y. Huang and H. L. Dai, "Learning from class-imbalanced data: Review of data driven methods and algorithm driven methods," *Data Sci. Finance Econ.*, vol. 1, no. 1, pp. 21–36, 2021.
- [6] E. Rendon, R. Alejo, C. Castorena, F. J. Isidro-Ortega, and E. E. Grandia-Gutierrez, "Data sampling methods to deal with the Big Data multi-class imbalance problem," *Appl. Sci.-Basel*, vol. 10, no. 4, 2020, Art. no. 1276.
- [7] Q. Dai, J. W. Liu, and Y. H. Shi, "Class-overlap undersampling based on Schur decomposition for Class-imbalance problems," *Expert Syst. With Appl.*, vol. 221, 2023, Art. no. 119735.
- [8] Y. Xie, X. Huang, F. Qin, F. Li, and X. Ding, "A majority affiliation based under-sampling method for class imbalance problem," *Inf. Sci.*, vol. 662, 2024, Art. no. 120263.
- [9] H. J. Ren, Y. H. Tang, W. Y. Dong, S. Ren, and L. H. Jiang, "DUEN: Dynamic ensemble handling class imbalance in network intrusion detection," *Expert Syst. With Appl.*, vol. 229, 2023, Art. no. 120420.
- [10] Z. C. Li, M. Huang, G. J. Liu, and C. J. Jiang, "A hybrid method with dynamic weighted entropy for handling the problem of class imbalance with overlap in credit card fraud detection," *Expert Syst. With Appl.*, vol. 175, 2021, Art. no. 114750.
- [11] K. Bhandari, K. Kumar, and A. L. Sangal, "Data quality issues in software fault prediction: A systematic literature review," *Artif. Intell. Rev.*, vol. 56, no. 8, pp. 7839–7908, 2023.
- [12] S. Shilaskar, A. Ghatol, and P. Chatur, "Medical decision support system for extremely imbalanced datasets," *Inf. Sci.*, vol. 384, pp. 205–219, 2017.
- [13] P. Rana, A. Sowmya, E. Meijering, and Y. Song, "Imbalanced classification for protein subcellular localization with multilabel oversampling," *Bioinformatics*, vol. 39, no. 1, 2023, Art. no. btac841.
- [14] S. Al-Azani, O. S. Alkhnbashi, E. Ramadan, and M. Alfarraj, "Gene expression-based cancer classification for handling the class imbalance problem and curse of dimensionality," *Int. J. Mol. Sci.*, vol. 25, no. 4, 2024, Art. no. 2102.
- [15] M. L. He, M. Petering, P. LaCasse, W. Otieno, and F. Maturana, "Learning with supervised data for anomaly detection in smart manufacturing," *Int. J. Comput. Integr. Manuf.*, vol. 36, no. 9, pp. 1331–1344, 2023.
- [16] P. Lade, R. Ghosh, and S. Srinivasan, "Manufacturing analytics and industrial Internet of Things," *IEEE Intell. Syst.*, vol. 32, no. 3, pp. 74–79, May/Jun. 2017.
- [17] Y. G. Yuan, J. A. Wei, H. S. Huang, W. D. Jiao, J. X. Wang, and H. L. Chen, "Review of resampling techniques for the treatment of imbalanced industrial data classification in equipment condition monitoring," *Eng. Appl. Artif. Intell.*, vol. 126, 2023, Art. no. 106911.
- [18] J. Yang, B. Matsushita, and H. Zhang, "Improving building rooftop segmentation accuracy through the optimization of UNet basic elements and image foreground-background balance," *ISPRS J. Photogramm. Remote Sens.*, vol. 201, pp. 123–137, 2023.
- [19] C. F. Pan, R. S. Li, W. Liu, W. J. Lu, C. Y. Niu, and Q. F. Bao, "Remote sensing image ship detection based on dynamic adjusting labels strategy," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4702621.
- [20] R. Zhang, H. Zhang, X. Ning, X. Huang, J. Wang, and W. Cui, "Global-aware siamese network for change detection on remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 199, pp. 61–72, 2023.
- [21] Y. Zhang, X. Zheng, and X. Lu, "Remote sensing image retrieval by deep attention hashing with distance-adaptive ranking," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 4301–4311, 2023.
- [22] T. J. Li, Y. X. Wang, L. C. Liu, L. Chen, and C. L. P. Chen, "Subspace-based minority oversampling for imbalance classification," *Inf. Sci.*, vol. 621, pp. 371–388, 2023.
- [23] R. Z. Zhang, S. W. Lu, B. K. Yan, P. L. Yu, and X. Q. Tang, "A density-based oversampling approach for class imbalance and data overlap," *Comput. Ind. Eng.*, vol. 186, 2023, Art. no. 109747.
- [24] N. V. Chawla, K. W. Bowyer, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.
- [25] Y. H. Quan, X. Zhong, W. Feng, J. C. W. Chan, Q. Li, and M. D. Xing, "SMOTE-based weighted deep rotation forest for the imbalanced hyperspectral data classification," *Remote Sens.*, vol. 13, no. 3, 2021, Art. no. 464.
- [26] F. Li, B. Wang, P. Wang, M. F. Jiang, and Y. M. Li, "An imbalanced ensemble learning method based on dual clustering and stage-wise hybrid sampling," *Appl. Intell.*, vol. 53, no. 18, pp. 21167–21191, 2023.
- [27] M. X. Lu, L. T. Tay, and J. Mohamad-Saleh, "Landslide susceptibility analysis using random forest model with SMOTE-ENN resampling algorithm," *Geomatics Natural Hazards Risk*, vol. 15, no. 1, 2024, Art. no. 2314565.
- [28] Y. Zhang, B. Kang, B. Hooi, S. Yan, and J. Feng, "Deep long-tailed learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 9, pp. 10795–10816, Sep. 2023.
- [29] K. H. Kim and S. Y. Sohn, "Hybrid neural network with cost-sensitive support vector machine for class-imbalanced multimodal data," *Neural Netw.*, vol. 130, pp. 176–184, 2020.
- [30] Z. J. Ren et al., "Adaptive cost-sensitive learning: Improving the convergence of intelligent diagnosis models under imbalanced data," *Knowl.-Based Syst.*, vol. 241, 2022, Art. no. 108296.
- [31] D. Devi, S. K. Biswas, and B. Purkayastha, "Correlation-based oversampling aided cost sensitive ensemble learning technique for treatment of class imbalance," *J. Exp. Theor. Artif. Intell.*, vol. 34, no. 1, pp. 143–174, 2022.
- [32] B. S. Raghuvanshi and S. Shukla, "SMOTE based class-specific extreme learning machine for imbalanced learning," *Knowl.-Based Syst.*, vol. 187, 2020, Art. no. 104814.
- [33] J. Zhang, S. Shao, S. Verma, and R. Nowak, "Algorithm selection for deep active learning with imbalanced datasets," *Adv. Neural Inf. Process. Syst.*, vol. 36, pp. 9614–9647, 2024.
- [34] Y. F. Zhang et al., "Online adaptive asymmetric active learning with limited budgets," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 6, pp. 2680–2692, Jun. 2021.
- [35] B. Krawczyk, "Learning from imbalanced data: Open challenges and future directions," *Prog. Artif. Intell.*, vol. 5, no. 4, pp. 221–232, 2016.

- [36] A. de Giorgio, G. Cola, and L. H. Wang, "Systematic review of class imbalance problems in manufacturing," *J. Manuf. Syst.*, vol. 71, pp. 620–644, 2023.
- [37] K. Ghosh, C. Bellinger, R. Corizzo, P. Branco, B. Krawczyk, and N. Japkowicz, "The class imbalance problem in deep learning," *Mach. Learn.*, vol. 113, no. 7, pp. 4845–4901, 2022.
- [38] K. Oksuz, B. C. Cam, S. Kalkan, and E. Akbas, "Imbalance problems in object detection: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3388–3415, Oct. 2021.
- [39] Z. Y. Xu, T. J. Wang, A. K. Skidmore, and R. Lamprey, "A review of deep learning techniques for detecting animals in aerial and satellite images," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 128, 2024, Art. no. 103732.
- [40] L. W. Huang, B. T. Jiang, S. Y. Lv, Y. B. Liu, and Y. Fu, "Deep-learning-based semantic segmentation of remote sensing images: A survey," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 8370–8396, 2024.
- [41] X. Sun, B. Wang, Z. R. Wang, H. Li, H. C. Li, and K. Fu, "Research progress on few-shot learning for remote sensing image interpretation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2387–2402, 2021.
- [42] M. E. Paoletti, O. Mogollon-Gutierrez, S. Moreno-Alvarez, J. C. Sancho, and J. M. Haut, "A comprehensive survey of imbalance correction techniques for hyperspectral data classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 5297–5314, 2023.
- [43] Z. Q. Chen, Y. H. Shang, A. Python, Y. X. Cai, and J. W. Yin, "DB-BlendMask: Decomposed attention and balanced BlendMask for instance segmentation of high-resolution remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2022.
- [44] X.-Y. Tong, G.-S. Xia, and X. X. Zhu, "Enabling country-scale land cover mapping with meter-resolution satellite imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 196, pp. 178–196, 2023.
- [45] Y. Ren et al., "Full convolutional neural network based on multi-scale feature fusion for the class imbalance remote sensing image classification," *Remote Sens.*, vol. 12, no. 21, 2020, Art. no. 3547.
- [46] J. F. Chen, G. Chen, B. Fang, J. J. Wang, and L. Z. Wang, "Class-aware domain adaptation for coastal land cover mapping using optical remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 11800–11813, 2021.
- [47] Y. Liu, C. Pang, Z. Q. Zhan, X. M. Zhang, and X. Yang, "Building change detection for remote sensing images using a dual-task constrained deep siamese convolutional network model," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 5, pp. 811–815, May 2021.
- [48] Z. Y. Xu, W. C. Zhang, T. X. Zhang, and J. Y. Li, "HRCNet: High-resolution context extraction network for semantic segmentation of remote sensing images," *Remote Sens.*, vol. 13, no. 1, 2021, Art. no. 71.
- [49] X. H. Li, M. Z. He, H. F. Li, and H. F. Shen, "A combined loss-based multiscale fully convolutional network for high-resolution remote sensing image change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 8017505.
- [50] Y. X. Cao and X. Huang, "A coarse-to-fine weakly supervised learning method for green plastic cover segmentation using high-resolution remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 188, pp. 157–176, 2022.
- [51] Q. Y. Li, R. F. Zhong, X. Du, and Y. Du, "TransUNetCD: A hybrid transformer network for change detection in optical remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5622519.
- [52] R. T. Xu, C. W. Wang, J. G. Zhang, S. B. Xu, W. L. Meng, and X. P. Zhang, "RSSFormer: Foreground saliency enhancement for remote sensing land-cover segmentation," *IEEE Trans. Image Process.*, vol. 32, pp. 1052–1064, 2023.
- [53] Z. L. Chen et al., "EGDE-Net: A building change detection method for high-resolution remote sensing imagery based on edge guidance and differential enhancement," *ISPRS J. Photogramm. Remote Sens.*, vol. 191, pp. 203–222, 2022.
- [54] Z. Xu, J. Geng, and W. Jiang, "MMT: Mixed-mask transformer for remote sensing image semantic segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5613415.
- [55] U. A. Bhatti et al., "MFFCG-Multi feature fusion for hyperspectral image classification using graph attention network," *Expert Syst. With Appl.*, vol. 229, 2023, Art. no. 120496.
- [56] W. Cui et al., "Long-tailed effect study in remote sensing semantic segmentation based on graph Kernel principles," *Remote Sens.*, vol. 16, no. 8, 2024, Art. no. 1398.
- [57] J. Y. Shi et al., "A double-head global reasoning network for object detection of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5402216.
- [58] W. M. Li, Y. B. Fu, S. S. Fan, M. R. Xin, and H. Y. Bai, "DCI-PGCN: Dual-channel interaction portable graph convolutional network for landslide detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4403616.
- [59] B. Yang, L. Qin, J. Q. Liu, and X. X. Liu, "IRCNN: An irregular-time-distaned recurrent convolutional neural network for change detection in satellite time series," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 2503905.
- [60] J. Y. Li, B. Zhang, and X. Huang, "A hierarchical category structure based convolutional recurrent neural network (HCS-ConvRNN) for land-cover classification using dense MODIS Time-Series data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 108, 2022, Art. no. 102744.
- [61] Z. P. Liu, H. Tang, and W. Huang, "Building outline delineation from VHR remote sensing images using the convolutional recurrent neural network embedded with line segment information," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4705713.
- [62] J. C. Ji, W. P. Jing, G. S. Chen, J. B. Lin, and H. B. Song, "Multi-label remote sensing image classification with latent semantic dependencies," *Remote Sens.*, vol. 12, no. 7, 2020, Art. no. 1110.
- [63] J. Song, S. H. Gao, Y. Q. Zhu, and C. Y. Ma, "A survey of remote sensing image classification based on CNNs," *Big Earth Data*, vol. 3, no. 3, pp. 232–254, 2019.
- [64] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, "Review on convolutional neural networks (CNN) in vegetation remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 173, pp. 24–49, 2021.
- [65] R. Q. Zhang, H. C. Zhang, X. G. Ning, X. Huang, J. M. Wang, and W. Cui, "Global-aware siamese network for change detection on remote sensing images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 199, pp. 61–72, 2023.
- [66] G. K. Xue, Y. K. Liu, Y. W. Huang, M. S. Li, and G. P. Yang, "AANet: An attention-based alignment semantic segmentation network for high spatial resolution remote sensing images," *Int. J. Remote Sens.*, vol. 43, no. 13, pp. 4836–4852, 2022.
- [67] M. A. Ganaie, M. H. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, "Ensemble deep learning: A review," *Eng. Appl. Artif. Intell.*, vol. 115, 2022, Art. no. 105151.
- [68] J. J. Li, Y. Z. Meng, D. Y. Dorjee, X. B. Wei, Z. Y. Zhang, and W. Zhang, "Automatic road extraction from remote sensing imagery using ensemble learning and postprocessing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10535–10547, 2021.
- [69] B. Q. Chen, L. J. Wang, X. J. Fan, W. H. Bo, X. B. Yang, and T. Tjahjadi, "Semi-FCMNet: Semi-supervised learning for forest cover mapping from satellite imagery via ensemble self-training and perturbation," *Remote Sens.*, vol. 15, no. 16, 2023, Art. no. 4012.
- [70] A. Körez, N. Barışçı, A. Cetin, and U. Ergün, "Weighted ensemble object detection with optimized coefficients for remote sensing images," *ISPRS Int. J. Geo-Inf.*, vol. 9, no. 6, 2020, Art. no. 370.
- [71] Y. C. Ma, S. Chen, S. Ermon, and D. B. Lovell, "Transfer learning in environmental remote sensing," *Remote Sens. Environ.*, vol. 301, 2024, Art. no. 113924.
- [72] C. S. Zhou, J. S. Zhang, J. M. Liu, C. X. Zhang, G. Shi, and J. Y. Hu, "Bayesian transfer learning for object detection in optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 7705–7719, Nov. 2020.
- [73] H. K. Tang, H. L. Wang, and X. P. Zhang, "Multi-class change detection of remote sensing images based on class rebalancing," *Int. J. Digit. Earth*, vol. 15, no. 1, pp. 1377–1394, 2022.
- [74] Z. L. Huang, C. O. Dumitru, Z. X. Pan, B. Lei, and M. Datcu, "Classification of large-scale high-resolution sar images with deep transfer learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 1, pp. 107–111, Jan. 2021.
- [75] A. A. Aleissae et al., "Transformers in remote sensing: A Survey," *Remote Sens.*, vol. 15, no. 7, 2023, Art. no. 1860.
- [76] L. B. Wang et al., "UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 190, pp. 196–214, 2022.
- [77] K. J. Xu, P. F. Deng, and H. Huang, "Vision transformer: An excellent teacher for guiding small networks in remote sensing image scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5618715.

- [78] L. Gao et al., "STransFuse: Fusing swin transformer and convolutional neural network for remote sensing image semantic segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10990–11003, 2021.
- [79] P. Y. Lv, L. S. Ma, Q. M. Li, and F. Du, "ShapeFormer: A shape-enhanced vision transformer model for optical remote sensing image landslide detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 2681–2689, 2023.
- [80] S. Ouyang and Y. S. Li, "Combining deep semantic segmentation network and graph convolutional neural network for semantic segmentation of remote sensing imagery," *Remote Sens.*, vol. 13, no. 1, 2021, Art. no. 119.
- [81] B. Liu, X. C. Yu, A. Z. Yu, P. Q. Zhang, and G. Wan, "Spectral-spatial classification of hyperspectral imagery based on recurrent neural networks," *Remote Sens. Lett.*, vol. 9, no. 12, pp. 1118–1127, 2018.
- [82] R. L. Hang, Q. S. Liu, D. F. Hong, and P. Ghamisi, "Cascaded recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5384–5394, Aug. 2019.
- [83] B. Y. Li, Y. L. Guo, J. G. Yang, L. G. Wang, Y. Q. Wang, and W. An, "Gated recurrent multiattention network for VHR remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5606113.
- [84] Q. Wang, S. T. Liu, J. Chanussot, and X. L. Li, "Scene classification with recurrent attention of VHR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1155–1167, Feb. 2019.
- [85] W. L. Zhou, S. I. Kamata, H. P. Wang, and X. Xue, "Multiscanning-based RNN-transformer for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5512319.
- [86] A. Mao, M. Mohri, and Y. Zhong, "Cross-entropy loss functions: Theoretical analysis and applications," in *Proc. Int. Conf. Mach. Learn.*, 2023, pp. 23803–23828.
- [87] F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3d Vis.*, 2016, pp. 565–571.
- [88] T. Y. Lin, P. Goyal, R. Girshick, K. M. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.
- [89] X. Yang et al., "SCRDet: Towards more robust detection for small, cluttered and rotated objects," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 8232–8241.
- [90] J. Li, W. Ding, H. Li, and C. Liu, "Semantic segmentation for high-resolution aerial imagery using multi-skip network and Markov random fields," in *Proc. IEEE Int. Conf. Unmanned Syst.*, 2017, pp. 12–17.
- [91] J. D. Zhang et al., "AERNet: An attention-guided edge refinement network and a dataset for remote sensing building change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5617116.
- [92] H. W. Bai, J. Cheng, Y. Z. Su, S. Y. Liu, and X. Liu, "Calibrated focal loss for semantic labeling of high-resolution remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6531–6547, 2022.
- [93] A. Ma, J. Wang, Y. Zhong, and Z. Zheng, "FactSeg: Foreground activation-driven small object semantic segmentation in large-scale remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5606216.
- [94] H. Z. Xu, H. J. He, Y. Zhang, L. F. Ma, and J. A. T. Li, "A comparative study of loss functions for road segmentation in remotely sensed road datasets," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 116, 2023, Art. no. 103159.
- [95] S. Farhadpour, T. A. Warner, and A. E. Maxwell, "Selecting and interpreting multiclass loss and accuracy assessment metrics for classifications with class imbalance: Guidance and best practices," *Remote Sens.*, vol. 16, no. 3, 2024, Art. no. 533.
- [96] P. P. Zhu, H. Xu, and X. B. Luo, "MDAFormer: Multi-level difference aggregation transformer for change detection of VHR optical imagery," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 118, 2023, Art. no. 103256.
- [97] J. F. Feng, E. Tang, M. M. Zeng, Z. J. Gu, P. L. Kou, and W. Zheng, "Improving visual question answering for remote sensing via alternate-guided attention and combined loss," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 122, 2023, Art. no. 103427.
- [98] F. Z. Cui and J. Jiang, "MTSCD-Net: A network based on multi-task learning for semantic change detection of bitemporal remote sensing images," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 118, 2023, Art. no. 103294.
- [99] M. Yuan et al., "MCAFNet: A multiscale channel attention fusion network for semantic segmentation of remote sensing images," *Remote Sens.*, vol. 15, no. 2, 2023, Art. no. 361.
- [100] P. Wang, Y. X. Niu, J. Wang, F. Ma, and C. X. Zhang, "Arbitrarily oriented dense object detection based on center point network in remote sensing images," *Remote Sens.*, vol. 14, no. 7, 2022, Art. no. 1536.
- [101] L. B. Zhang, C. G. Zhang, S. N. Quan, H. X. Xiao, G. Y. Kuang, and L. Liu, "A class imbalance loss for imbalanced object recognition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2778–2792, 2020.
- [102] H. X. Bi, H. H. Chang, X. T. Wang, and D. F. Hong, "Cross-attention-driven adaptive graph relational network for multilabel remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5224414.
- [103] D. F. Hong et al., "Cross-city matters: A multimodal remote sensing benchmark dataset for cross-city semantic segmentation using high-resolution domain adaptation networks," *Remote Sens. Environ.*, vol. 299, 2023, Art. no. 113856.
- [104] G. Li, Q. Gao, M. Yang, and X. Gao, "Active learning based on similarity level histogram and adaptive-scale sampling for very high resolution image classification," *Neural Netw.*, vol. 167, pp. 22–35, 2023.
- [105] J. X. Yang, M. M. Yu, S. H. Li, J. Zhang, and S. Z. Hu, "Long-tailed object detection for multimodal remote sensing images," *Remote Sens.*, vol. 15, no. 18, 2023, Art. no. 4539.
- [106] J. M. Liu, S. J. Li, C. S. Zhou, X. Y. Cao, Y. Gao, and B. Wang, "SRAF-Net: A scene-relevant anchor-free object detection network in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5405914.
- [107] X. Q. Lu et al., "Weak-to-strong consistency learning for semisupervised image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5510715.
- [108] D. Hendrycks, N. Mu, E. D. Cubuk, B. Zoph, J. Gilmer, and B. Lakshminarayanan, "AugMix: A simple data processing method to improve robustness and uncertainty," 2019, *arXiv:1912.02781*.
- [109] K. Baek, D. Bang, and H. Shim, "GridMix: Strong regularization through local context mapping," *Pattern Recognit.*, vol. 109, 2021, Art. no. 107594.
- [110] J. H. Kim, W. Choo, and H. O. Song, "Puzzle mix: Exploiting saliency and local statistics for optimal mixup," *Int. Conf. Mach. Learn.*, vol. 119, pp. 5275–5285, 2020.
- [111] D. Berthelot et al., "Remixmatch: Semi-supervised learning with distribution alignment and augmentation anchoring," 2019, *arXiv:1911.09785*.
- [112] X. Huang, W. R. Wang, J. Y. Li, L. G. Wang, and X. Xie, "A stepwise refining image-level weakly supervised semantic segmentation method for detecting exposed surface for buildings (ESB) from very high-resolution remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5400517.
- [113] Z. L. Suo, Y. B. Zhao, S. Chen, and Y. L. Hu, "BoxPaste: An effective data augmentation method for SAR ship detection," *Remote Sens.*, vol. 14, no. 22, 2022, Art. no. 5761.
- [114] L. Q. Huang, W. B. Zhao, A. W. C. Liew, and Y. You, "An evidential combination method with multi-color spaces for remote sensing image scene classification," *Inf. Fusion*, vol. 93, pp. 209–226, 2023.
- [115] R. Ramos and B. Martins, "Using neural encoder-decoder models with continuous outputs for remote sensing image captioning," *IEEE Access*, vol. 10, pp. 24852–24863, 2022.
- [116] H. X. Dou, X. S. Lu, C. Wang, H. Z. Shen, Y. W. Zhuo, and L. J. Deng, "PatchMask: A data augmentation strategy with Gaussian noise in hyperspectral images," *Remote Sens.*, vol. 14, no. 24, 2022, Art. no. 6308.
- [117] N. Q. Liu, X. Xu, T. Celik, Z. X. Gan, and H. C. Li, "Transformation-invariant network for few-shot object detection in remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5625314.
- [118] T. DeVries and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," 2017, *arXiv:1708.04552*.
- [119] Z. Zhong, L. Zheng, G. L. Kang, S. Z. Li, and Y. Yang, "Random erasing data augmentation," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, pp. 13001–13008.
- [120] K. K. Singh and Y. Jae Lee, "Hide-and-seek: Forcing a network to be meticulous for weakly-supervised object and action localization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 3524–3533.
- [121] P. Chen, S. Liu, H. Zhao, and J. Jia, "Gridmask data augmentation," 2020, *arXiv:2001.04086*.
- [122] P. Isola, J. Y. Zhu, T. H. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5967–5976.
- [123] Y. Choi, M. Choi, M. Kim, J. W. Ha, S. Kim, and J. Choo, "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8789–8797.

- [124] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2223–2232.
- [125] T. Park, M. Y. Liu, T. C. Wang, and J. Y. Zhu, "Semantic image synthesis with spatially-adaptive normalization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 2332–2341.
- [126] Y. Y. Ren et al., "Full convolutional neural network based on multi-scale feature fusion for the class imbalance remote sensing image classification," *Remote Sens.*, vol. 12, no. 21, 2020, Art. no. 3547.
- [127] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," 2017, *arXiv:1710.09412*.
- [128] Z. H. You, J. X. Wang, S. B. Chen, J. Tang, and B. Luo, "FMWDCT: Foreground mixup into weighted dual-network cross training for semi-supervised remote sensing road extraction," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 5570–5579, 2022.
- [129] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "Cut-Mix: Regularization strategy to train strong classifiers with localizable features," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 6022–6031.
- [130] X. J. Hao, L. Liu, R. J. Yang, L. Z. Y. Yin, L. Zhang, and X. H. Li, "A review of data augmentation methods of remote sensing image target recognition," *Remote Sens.*, vol. 15, no. 3, 2023, Art. no. 827.
- [131] S. Pande, A. Banerjee, S. Kumar, B. Banerjee, and S. Chaudhuri, "An adversarial approach to discriminative modality distillation for remote sensing image classification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops*, 2019, pp. 4571–4580.
- [132] P. Li, X. Li, and X. Long, "Fencemask: A data augmentation approach for pre-extracted image features," 2020, *arXiv:2006.07877*.
- [133] H. Chen, W. Y. Li, and Z. W. Shi, "Adversarial instance augmentation for building change detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5603216.
- [134] Y. L. Gao, Y. Feng, and X. M. Yu, "Hyperspectral target detection with an auxiliary generative adversarial network," *Remote Sens.*, vol. 13, no. 21, 2021, Art. no. 4454.
- [135] H. Chen, Z. H. Li, J. J. Wu, W. Xiong, and C. Du, "SemiRoadExNet: A semi-supervised network for road extraction from remote sensing imagery via adversarial learning," *ISPRS J. Photogramm. Remote Sens.*, vol. 198, pp. 169–183, 2023.
- [136] Z. Wang, N. Xu, B. H. Wang, Y. H. Liu, and S. W. Zhang, "Urban building extraction from high-resolution remote sensing imagery based on multi-scale recurrent conditional generative adversarial network," *GIScience Remote Sens.*, vol. 59, no. 1, pp. 861–884, 2022.
- [137] Y. Z. Lu, D. Chen, E. Olaniyi, and Y. B. Huang, "Generative adversarial networks (GANs) for image augmentation in agriculture: A systematic review," *Comput. Electron. Agriculture*, vol. 200, 2022, Art. no. 107208.
- [138] L. Shumilo, A. Okhrimenko, N. Kussul, S. Drozd, and O. Shkalikov, "Generative adversarial network augmentation for solving the training data imbalance problem in crop classification," *Remote Sens. Lett.*, vol. 14, no. 11, pp. 1131–1140, 2023.
- [139] X. Rui, Y. Cao, X. Yuan, Y. Kang, and W. G. Song, "DisasterGAN: Generative adversarial networks for remote sensing disaster image generation," *Remote Sens.*, vol. 13, no. 21, 2021, Art. no. 4284.
- [140] Y. T. Li, H. Chen, S. Dong, Y. Zhuang, and L. L. Li, "Multi-temporal sample pair generation for building change detection promotion in optical remote sensing domain based on generative adversarial network," *Remote Sens.*, vol. 15, no. 9, 2023, Art. no. 2470.
- [141] F. Rottensteiner et al., "The ISPRS benchmark on urban object classification and 3D building reconstruction," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.; I-3*, vol. 1, no. 1, pp. 293–298, 2012.
- [142] S. Waqas Zamir et al., "iSAID: A large-scale dataset for instance segmentation in aerial images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recogn. Workshops*, 2019, pp. 28–37.
- [143] V. Mnih, *Machine Learning for Aerial Image Labeling*. Toronto, ON, Canada: Univ. of Toronto, 2013, pp. 28–37.
- [144] S. J. Wei, X. F. Zeng, Q. Z. Qu, M. Wang, H. Su, and J. Shi, "HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation," *IEEE Access*, vol. 8, pp. 120234–120254, 2020.
- [145] J. W. Li, C. W. Qu, and J. Q. Shao, "Ship detection in SAR images based on an improved faster R-CNN," in *Proc. SAR Big Data Era, Models, Methods Appl.*, 2017, pp. 1–6.
- [146] H. G. Zhu, X. G. Chen, W. Q. Dai, K. Fu, Q. X. Ye, and J. B. Jiao, "Orientation robust object detection in aerial images using deep convolutional neural network," in *Proc. IEEE Int. Conf. Image Process.*, 2015, pp. 3735–3739.
- [147] H. Chen and Z. W. Shi, "A spatial-temporal attention-based method and a new dataset for remote sensing image change detection," *Remote Sens.*, vol. 12, no. 10, 2020, Art. no. 1662.
- [148] M. Lebedev, Y. V. Vizilter, O. Vygolov, V. A. Knyaz, and A. Y. Rubis, "Change detection in remote sensing images using conditional adversarial networks," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 42, pp. 565–571, 2018.
- [149] R. Gupta et al., "Creating xRD: A dataset for assessing building damage from satellite imagery," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recogn. Workshops*, 2019, pp. 10–17.
- [150] S. P. Ji, S. Q. Wei, and M. Lu, "Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 574–586, Jan. 2019.
- [151] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geogr. Inf. Syst.*, 2010, pp. 270–279.
- [152] C. O. Dumitru, G. Schwarz, and M. Datcu, "Land cover semantic annotation derived from high-resolution SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2215–2232, Jun. 2016.
- [153] P. Helber, B. Bischke, A. Dengel, and D. Borth, "EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 7, pp. 2217–2226, Jul. 2019.
- [154] J. L. Hou, Z. Guo, Y. C. Feng, Y. M. Wu, and W. H. Diao, "SPANet: Spatial adaptive convolution based content-aware network for aerial image semantic segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 2192–2204, 2023.
- [155] Y. M. Wang, Y. Wan, Y. J. Zhang, B. Zhang, and Z. Gao, "Imbalance knowledge-driven multi-modal network for land-cover semantic segmentation using aerial images and LiDAR point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 202, pp. 385–404, 2023.
- [156] Y. Y. Dang, Y. Gao, and B. Liu, "MFAFNet: A multiscale fully attention fusion network for remote sensing image semantic segmentation," *IEEE Access*, vol. 12, pp. 123388–123400, 2024.
- [157] T. Liu, S. L. Cheng, and J. Yuan, "Category-based interactive attention and perception fusion network for semantic segmentation of remote sensing images," *Remote Sens.*, vol. 16, no. 20, 2024, Art. no. 3864.
- [158] J. Zhang, M. W. Shao, Y. C. Wan, L. Z. Meng, X. Y. Cao, and S. G. Wang, "Boundary-aware spatial and frequency dual-domain transformer for remote sensing urban images segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–18, 2024.
- [159] R. G. Niu, X. Sun, Y. Tian, W. H. Diao, Y. C. Feng, and K. Fu, "Improving semantic segmentation in aerial imagery via graph reasoning and disentangled learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5611918.
- [160] W. J. Zi, W. Xiong, H. Chen, J. Li, and N. Jing, "SGA-Net: Self-constructing graph attention neural network for semantic segmentation of remote sensing images," *Remote Sens.*, vol. 13, no. 21, 2021, Art. no. 4201.
- [161] Q. H. Liu, M. Kampffmeyer, R. Jenssen, and A. B. Salberg, "Self-constructing graph neural networks to model long-range pixel dependencies for semantic segmentation of remote sensing images," *Int. J. Remote Sens.*, vol. 42, no. 16, pp. 6187–6211, 2021.
- [162] T. Zhou, H. He, Y. Wang, and Y. Liao, "Improved gated recurrent units together with fusion for semantic segmentation of remote sensing images based on parallel hybrid network," *Multimedia Syst.*, vol. 3, no. 1, 2025, Art. no. 76.
- [163] M. B. Pereira and J. A. Dos Santos, "ChessMix: Spatial context data augmentation for remote sensing semantic segmentation," in *Proc. SIBGRAPI Conf. Graph., Patterns Images*, 2021, pp. 278–285.
- [164] Z. Gong, L. J. Duan, F. J. Xiao, and Y. X. Wang, "MSAug: Multi-strategy augmentation for rare classes in semantic segmentation of remote sensing images," *Displays*, vol. 84, 2024, Art. no. 102779.
- [165] N. Lv et al., "Remote sensing data augmentation through adversarial training," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 9318–9333, 2021.
- [166] Y. Y. Guo and L. Y. Zhou, "MEA-Net: A lightweight SAR ship detection model for imbalanced datasets," *Remote Sens.*, vol. 14, no. 18, 2022, Art. no. 4438.

- [167] Y. Y. Guo, Y. M. He, M. M. Gao, and L. Y. Zhou, "GMDR-Net: A lightweight OBB-based SAR ship detection model based on Gaussian mixture data augmentation and distance rotation IOU loss," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 11931–11942, 2024.
- [168] A. Oubara, F. L. Wu, R. Maleki, B. Y. Ma, A. Amamra, and G. L. Yang, "Enhancing adversarial learning-based change detection in imbalanced datasets using artificial image generation and attention mechanism," *ISPRS Int. J. Geo-Inf.*, vol. 13, no. 4, 2024, Art. no. 125.
- [169] X. Lu, C. Zhang, Q. Ye, C. Wang, C. Yang, and Q. Wang, "RSI-Mix: Data augmentation method for remote sensing image classification," in *Proc. 7th Int. Conf. Intell. Comput. Signal Process.*, 2022, pp. 1982–1985.
- [170] X. Pan, J. Zhao, and J. Xu, "A scene images diversity improvement generative adversarial network for remote sensing image scene classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 10, pp. 1692–1696, 2020.
- [171] W. Han et al., "Sample generation based on a supervised Wasserstein generative adversarial network for high-resolution remote-sensing scene classification," *Inf. Sci.*, vol. 539, pp. 177–194, 2020.
- [172] X. D. Zhang et al., "Geospatial object detection on high resolution remote sensing imagery based on double multi-scale feature pyramid network," *Remote Sens.*, vol. 11, no. 7, 2019, Art. no. 755.
- [173] B. Ren et al., "A dual-stream high resolution network: Deep fusion of GF-2 and GF-3 data for land cover classification," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 112, 2022, Art. no. 102896.
- [174] Z. Zheng, Y. F. Zhong, J. J. Wang, and A. L. Ma, "Foreground-aware relation network for geospatial object segmentation in high spatial resolution remote sensing imagery," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2020, pp. 4095–4104.
- [175] J. S. Li, S. J. Zhang, Q. Han, and Y. Y. Sun, "CSRL-Net: Contextual self-rasterization learning network with joint weight loss for remote sensing image semantic segmentation," *Int. J. Remote Sens.*, vol. 44, no. 23, pp. 7515–7542, 2023.
- [176] Q. Xu, X. Yuan, and C. Ouyang, "Class-aware domain adaptation for semantic segmentation of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4500317.
- [177] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data," *ISPRS J. Photogramm. Remote Sens.*, vol. 162, pp. 94–114, 2020.
- [178] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 1, pp. 6–43, Mar. 2013.
- [179] H. X. Bi, F. Xu, Z. Q. Wei, Y. Xue, and Z. B. Xu, "An active deep learning approach for minimally supervised PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9378–9395, Nov. 2019.
- [180] H. X. Bi, J. Sun, and Z. B. Xu, "Unsupervised PolSAR image classification using discriminative clustering," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3531–3544, 2017.
- [181] N. W. Wang, W. Q. Jin, H. X. Bi, C. Xu, and J. H. Gao, "A survey on deep learning for few-shot PolSAR image classification," *Remote Sens.*, vol. 16, no. 24, 2024, Art. no. 4632.
- [182] Z. Z. Kuang, H. X. Bi, F. Li, C. Xu, and J. Sun, "Polarimetry-inspired contrastive learning for class-imbalanced PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–19, 2024.
- [183] N. W. Wang, H. X. Bi, F. Li, C. Xu, and J. H. Gao, "Self-distillation-based polarimetric image classification with noisy and sparse labels," *Remote Sens.*, vol. 15, no. 24, 2023, Art. no. 5751.
- [184] F. Liu, L. C. Jiao, and X. Tang, "Task-oriented GAN for PolSAR image classification and clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2707–2719, Sep. 2019.
- [185] U. Kanjir, H. Greidanus, and K. Ostir, "Vessel detection and classification from spaceborne optical images: A literature survey," *Remote Sens. Environ.*, vol. 207, pp. 1–26, 2018.
- [186] Z. Li et al., "Deep learning-based object detection techniques for remote sensing images: A survey," *Remote Sens.*, vol. 14, no. 10, 2022, Art. no. 2385.
- [187] X. Hu, P. Zhang, and Y. Ban, "Large-scale burn severity mapping in multispectral imagery using deep semantic segmentation models," *ISPRS J. Photogramm. Remote Sens.*, vol. 196, pp. 228–240, 2023.
- [188] X. R. Zhang et al., "Remote sensing object detection meets deep learning: A metareview of challenges and advances," *IEEE Geosci. Remote Sens. Mag.*, vol. 11, no. 4, pp. 8–44, Dec. 2023.
- [189] D. Liang, J. W. Zhang, Y. P. Tang, and S. J. Huang, "MUS-CDB: Mixed uncertainty sampling with class distribution balancing for active annotation in aerial object detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5613013.
- [190] Y. F. Liu, Q. Li, Y. Yuan, Q. Du, and Q. Wang, "ABNet: Adaptive balanced network for multiscale object detection in remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5614914.
- [191] R. A. Amit and C. K. Mohan, "A robust airport runway detection network based on R-CNN using remote sensing images," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 36, no. 11, pp. 4–20, Nov. 2021.
- [192] C. Q. Zhang, Y. Deng, M. Z. Chong, Z. W. Zhang, and Y. H. Tan, "Entropy-based re-sampling method on SAR class imbalance target detection," *ISPRS J. Photogramm. Remote Sens.*, vol. 209, pp. 432–447, 2024.
- [193] Y. B. Liu, G. Yan, F. Ma, Y. S. Zhou, and F. Zhang, "SAR ship detection based on explainable evidence learning under intraclass imbalance," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–15, 2024.
- [194] H. W. Jiang et al., "A survey on deep learning-based change detection from high-resolution remote sensing images," *Remote Sens.*, vol. 14, no. 7, 2022, Art. no. 1552.
- [195] J. K. Li, S. Y. Zhu, Y. Y. Gao, G. X. Zhang, and Y. M. Xu, "Change detection for high-resolution remote sensing images based on a multi-scale attention siamese network," *Remote Sens.*, vol. 14, no. 14, 2022, Art. no. 3464.
- [196] X. Hou et al., "Deep collaborative learning with class-rebalancing for semi-supervised change detection in SAR images," *Knowl.-Based Syst.*, vol. 264, 2023, Art. no. 110281.
- [197] B. Cui et al., "Enhanced edge information and prototype constrained clustering for SAR change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5206116.
- [198] G. Cheng, J. W. Han, and X. Q. Lu, "Remote sensing image scene classification: Benchmark and State of the Art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, 2017.
- [199] G. Cheng, X. X. Xie, J. W. Han, L. Guo, and G. S. Xia, "Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3735–3756, 2020.
- [200] W. M. Li et al., "Classification of high-spatial-resolution remote sensing scenes method using transfer learning and deep convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1986–1995, 2020.
- [201] L. L. Fan, H. W. Zhao, and H. Y. Zhao, "Distribution consistency loss for large-scale remote sensing image retrieval," *Remote Sens.*, vol. 12, no. 1, 2020, Art. no. 175.
- [202] X. L. Qian et al., "Generating and sifting pseudolabeled samples for improving the performance of remote sensing image scene classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4925–4933, 2020.
- [203] W. Miao, J. Geng, and W. Jiang, "Multigranularity decoupling network with pseudolabel selection for remote sensing image scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5603813.
- [204] H. Yessou, G. Sumbul, and B. Demir, "A comparative study of deep learning loss functions for multi-label remote sensing image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 1349–1352.
- [205] W. Zhao, J. Liu, Y. Liu, F. Zhao, Y. He, and H. Lu, "Teaching teachers first and then student: Hierarchical distillation to improve long-tailed object recognition in aerial images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5624412.
- [206] J. J. Wang, W. Li, M. M. Zhang, R. Tao, and J. Chanussot, "Remote-sensing scene classification via multistage self-guided separation network," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5615312.
- [207] G. Sumbul and B. Demir, "A deep multi-attention driven approach for multi-label remote sensing image classification," *IEEE Access*, vol. 8, pp. 95934–95946, 2020.
- [208] P. Li, P. Chen, and D. Z. Zhang, "Cross-modal feature representation learning and label graph mining in a residual multi-attentional CNN-LSTM network for multi-label aerial scene classification," *Remote Sens.*, vol. 14, no. 10, 2022, Art. no. 2424.
- [209] S. Wang et al., "Multi-label semantic feature fusion for remote sensing image captioning," *ISPRS J. Photogramm. Remote Sens.*, vol. 184, pp. 1–18, 2022.
- [210] J. Kang, R. Fernandez-Beltran, D. F. Hong, J. Chanussot, and A. Plaza, "Graph relation network: Modeling relations between scenes for multilabel remote-sensing image classification and retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4355–4369, May 2021.
- [211] T. C. Song, S. F. Bai, F. Yang, C. Q. Gao, H. A. Chen, and J. Li, "Exploring hybrid contrastive learning and scene-to-label information for multilabel remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5631214.

- [212] A. K. Aksoy, M. Ravanbakhsh, and B. Demir, "Multi-label noise robust collaborative learning for remote sensing image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 5, pp. 6438–6451, May 2024.
- [213] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," 2023, *arXiv:2312.00752*.
- [214] A. Kirillov et al., "Segment anything," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2023, pp. 4015–4026.



**Pengdi Chen** received the M.Sc. degree in surveying and mapping engineering from the Kunming University of Science and Technology, Kunming, China, in 2021. He is currently working toward the Ph.D. degree with Lanzhou University, Lanzhou, China.

His research interests include remote sensing image interpretation, computer vision, and arid oasis land cover classification.



**Yuanrui Ren** received the B.Sc. degree in geographic information science from Central China Normal University, Wuhan, China, in 2021. She is currently working toward the M.Sc. degree with Lanzhou University, Lanzhou, China.

Her research interests include remote sensing, deep learning, and semantic segmentation.



**Baoan Zhang** received the M.Sc. degree in land resource management from Gansu Agricultural University, Lanzhou, China, in 2004.

He is currently a Senior Engineer with the Mapping Institution of Gansu Province, Lanzhou, China. His research interests include aerial photogrammetry and remote sensing image processing.



**Yuan Zhao** received the M.Sc. degree in cartography and geography information system from Nanjing University, Nanjing, China, in 2008.

He is currently a Senior Engineer with the Mapping Institution of Gansu Province, Lanzhou, China. His research interests include aerial photogrammetry and remote sensing image processing.