remote sensing

MDPI

Review

A Survey on Deep Learning for Few-Shot PolSAR Image Classification

Ningwei Wang [1,†], Weiqiang Jin [1,†], Haixia Bi [1,*], Chen Xu [2,3] and Jinghuai Gao [1]

[1] School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an 710049, China; ningweiwang@stu.xjtu.edu.cn (N.W.); weiqiangjin@stu.xjtu.edu.cn (W.J.); jhgao@mail.xjtu.edu.cn (J.G.)
[2] School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China; xuch01@pcl.ac.cn
[3] Department of Mathematics and Fundamental Research, Peng Cheng Laboratory, Shenzhen 518000, China
* Correspondence: haixia.bi@xjtu.edu.cn
† These authors contributed equally to this work.

Abstract: Few-shot classification of polarimetric synthetic aperture radar (PolSAR) images is a challenging task due to the scarcity of labeled data and the complex scattering properties of PolSAR data. Traditional deep learning models often suffer from overfitting and catastrophic forgetting in such settings. Recent advancements have explored innovative approaches, including data augmentation, transfer learning, meta-learning, and multimodal fusion, to address these limitations. Data augmentation methods enhance the diversity of training samples, with advanced techniques like generative adversarial networks (GANs) generating realistic synthetic data that reflect PolSAR's polarimetric characteristics. Transfer learning leverages pre-trained models and domain adaptation techniques to improve classification across diverse conditions with minimal labeled samples. Meta-learning enhances model adaptability by learning generalizable representations from limited data. Multimodal methods integrate complementary data sources, such as optical imagery, to enrich feature representation. This survey provides a comprehensive review of these strategies, focusing on their advantages, limitations, and potential applications in PolSAR classification. We also identify key trends, such as the increasing role of hybrid models combining multiple paradigms and the growing emphasis on explainability and domain-specific customization. By synthesizing SOTA approaches, this survey offers insights into future directions for advancing few-shot PolSAR classification.

Keywords: polarimetric SAR; image classification; deep learning; few-shot learning; survey

Citation: Wang, N.; Jin, W.; Bi, H.; Xu, C.; Gao, J. A Survey on Deep Learning for Few-Shot PolSAR Image Classification. *Remote Sens.* **2024**, *16*, 4632. https://doi.org/10.3390/rs16244632

Academic Editor: Timo Balz

Received: 6 November 2024
Revised: 5 December 2024
Accepted: 9 December 2024
Published: 11 December 2024

1. Introduction

Deep learning and computer vision are used in various applications such as image classification, object detection in industrial production, medical image analysis, action recognition, and remote sensing [1–8]. PolSAR is an essential tool in remote sensing, offering the ability to collect high-resolution images under all weather and lighting conditions. PolSAR captures the backscattering characteristics of various objects through multiple polarization channels, enabling applications across a range of sectors, including land cover classification, vegetation monitoring, urban planning, and disaster management [9–13]. PolSAR image classification is a multilevel process that begins with extracting features from images and progresses to classifying them into categories [14]. As shown in Figure 1, in preprocessing, steps like radiometric calibration, geometric correction, and speckle noise reduction are applied to ensure data consistency and clarity. Preprocessing steps such as radiometric calibration, geometric correction, and speckle noise reduction ensure the consistency and clarity of PolSAR data. Speckle noise, modeled as multiplicative noise, can be expressed as follows:

$$I = I_0 \cdot N, \tag{1}$$

where I denotes the observed intensity, I_0 denotes the true intensity, and N is speckle noise. Adaptive filters such as the refined Lee filter [15] minimize noise while preserving edge

details. Geometric corrections are critical for aligning PolSAR data with geographic coordinates, accounting for distortions caused by topography and sensor geometry. Radiometric corrections ensure that the pixel intensity values accurately represent backscattering coefficients by compensating for variations due to sensor characteristics or environmental factors. The methods discussed in the book by Cumming and Wong [16] provide a comprehensive framework for implementing these corrections. For example, geometric corrections utilize sensor models and ground control points to correct distortions:

$$x' = T(x), \tag{2}$$

where x represents the original pixel coordinates, $T(\cdot)$ denotes the transformation function, and x' denotes the corrected coordinates. Radiometric calibration involves normalizing intensities based on system gains and atmospheric losses:

$$\sigma^0 = \frac{P}{R^2}, \tag{3}$$

where σ^0 denotes the normalized backscatter coefficient, P denotes the measured power, and R denotes the slant range. This ensures the data are comparable across acquisitions and sensors.

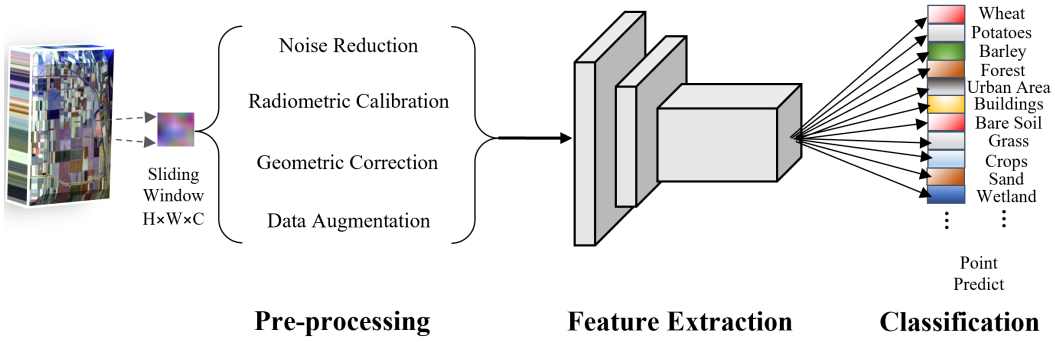

Figure 1. PolSAR image classification process.

The next step involves feature extraction, where polarimetric decompositions, such as Freeman–Durden or Cloude–Pottier, transform the complex-valued data into interpretable features that reflect surface properties like roughness and orientation. Polarimetric decomposition transforms complex-valued PolSAR data into interpretable features:

- Freeman–Durden decomposition [17]: Separates scattering contributions into volume scattering (P_V), double-bounce scattering (P_D), and surface scattering (P_S), as follows:

$$S = P_V + P_D + P_S, \tag{4}$$

  where each component is derived from the covariance matrix C.
- Cloude–Pottier decomposition [18]: Computes polarimetric entropy (H) and anisotropy (A) using eigenvalues $(\lambda_1, \lambda_2, \lambda_3)$ of the coherency matrix T, as follows:

$$H = -\sum_{i=1}^{3} \frac{\lambda_i}{\Lambda} \log\left(\frac{\lambda_i}{\Lambda}\right), \quad A = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2}, \tag{5}$$

  where $\Lambda = \lambda_1 + \lambda_2 + \lambda_3$.

These methods reveal physical properties such as surface roughness, orientation, and randomness of scattering mechanisms.

Since PolSAR data are often high-dimensional, techniques like principal component analysis or CNNs are applied to reduce dimensionality, thus lowering computational demands while preserving essential information. PolSAR data's high dimensionality necessitates efficient dimensionality reduction techniques:

- Principal component analysis (PCA) [19]: Projects data X onto a lower-dimensional subspace:
$$Z = XW, \tag{6}$$
where W contains eigenvectors of the covariance matrix of X.
- Deep learning-based feature extraction [20]: CNNs reduce dimensionality by extracting hierarchical features. For example, ResNet [21] uses residual connections, as follows:
$$y = f(x) + x, \tag{7}$$
where $f(x)$ represents the learned transformation within a residual block [21].

After feature extraction, a classification model, such as an SVM, random forest, or a DL network like a CNN, is trained on labeled samples. PolSAR classification has evolved from traditional machine learning models to deep learning approaches:

- Support vector machine (SVM): Optimizes a hyperplane that separates classes with maximum margin [22]:
$$\min_{w,b} \frac{1}{2}\|w\|^2 \quad \text{s.t.} \quad y_i(w \cdot x_i + b) \geq 1, \forall i. \tag{8}$$
- Deep neural networks: CNNs and architectures such as ResNet and Vision Transformers extract spatial and spectral features to classify PolSAR data [23].

DL models may employ transfer learning from pre-trained networks to improve classification accuracy on limited training data. Once trained, the classifier assigns labels to each pixel or region of the PolSAR image based on learned features, resulting in a classified map where each area corresponds to specific terrain types such as urban areas, forests, or water bodies. The versatility of PolSAR data makes it especially valuable for areas with frequent cloud cover, such as tropical regions, or for applications requiring continuous monitoring, such as in agriculture or defense operations. PolSAR image classification benefits significantly from combining physical modeling and data-driven approaches. Physical modeling provides domain-specific priors, such as scattering mechanisms and polarimetric decompositions, which enhance the interpretability and robustness of classification models. Meanwhile, data-driven approaches such as deep learning enable scalable feature extraction and generalization. By integrating these two paradigms, classification frameworks can leverage the interpretability of physical models while capitalizing on the flexibility of data-driven techniques, addressing challenges such as domain shifts, noise robustness, and limited labeled data. However, PolSAR data introduce unique challenges that must be addressed to achieve effective classification.

One major challenge is speckle noise, an inherent issue with coherent radar imaging systems, which manifests as granular interference across the imagery [24]. This noise complicates the task of pixel-level classification by reducing spatial consistency and overall classification accuracy. Techniques such as Pauli and Freeman–Durden decompositions attempt to mitigate speckle but often sacrifice spatial detail. Additionally, the complexity of PolSAR's multi-polarization data creates a high-dimensional feature space. Each pixel contains information from several channels, necessitating models that can accurately capture both local and global relationships within the data. Conventional CNNs, designed primarily for RGB images, struggle with such intricacies. As a result, architectures such as 3D-CNNs and networks integrating coherent and non-coherent decomposition features have been proposed to handle these challenges better [25]. PolSAR classification is further complicated by significant intra-class variability and inter-class similarity. Variability within classes arises due to environmental factors such as differences in observation angles, polarization modes, and weather conditions, which alter the backscattering characteristics of the same terrain type [26]. At the same time, certain land cover types, such as urban areas and bare soil, may exhibit similar scattering responses, leading to inter-class similarity.

These issues complicate classification tasks by increasing the overlap in feature space and reducing model separability.

DL has shown great promise in overcoming some of these limitations through its ability to automatically extract features. However, its reliance on large labeled datasets presents a critical bottleneck. Annotating PolSAR data is labor-intensive and requires expertise, which makes large-scale dataset creation expensive and time-consuming. The issue is further compounded by domain shifts, as PolSAR images collected from different satellites or environments exhibit significant variations that reduce model performance on new datasets [27]. Effective domain adaptation and TL techniques are essential but remain underexplored in PolSAR classification. Moreover, deep networks, such as CNNs and Transformer-based architectures, often contain millions of parameters. Without sufficient labeled data, these models tend to memorize the training set rather than generalize to new samples, resulting in poor performance on unseen datasets. This issue is particularly severe in PolSAR, where intra-class variability and inter-class similarity further hinder generalization. Additionally, PolSAR data's high dimensionality, with multiple polarization channels per pixel, leads to sparse feature spaces. This makes it challenging to extract relevant information without overfitting.

To address these challenges, FSL has emerged as a promising strategy. FSL enables models to generalize from limited labeled samples by simulating real-world conditions during training through episodic learning. Techniques such as prototypical networks [28] and relation networks [29], which learn from small support sets, align closely with the needs of PolSAR classification, where annotated datasets are typically small. Prototypical networks define the prototype of class c_k as follows:

$$c_k = \frac{1}{|S_k|} \sum_{z_i \in S_k} z_i, \tag{9}$$

where S_k denotes the support set for class k. Relation networks learn similarity scores between query samples q and support samples s using the following formula: $r(q, s) = \text{sim}(f(q), f(s))$, where $f(\cdot)$ represents a feature extractor. FSL also leverages meta-learning [30] and contrastive learning [31] approaches to acquire transferable features that can be generalized across different domains, reducing the impact of domain shifts. Unsupervised pre-training on unlabeled PolSAR datasets has further improved the effectiveness of FSL by enabling models to learn robust representations, which can be fine-tuned on minimal labeled data. FSL is particularly valuable for detecting rare or underrepresented classes in PolSAR datasets, such as specific terrain types or vegetation covers. Techniques like graph-based FSL models and attention-weighted networks have shown promise by leveraging neighborhood information and spatial correlations between pixels for enhanced classification performance.

The integration of FSL into PolSAR classification offers robust performance across diverse environmental conditions and sensor systems, making it an essential technique for both research and operational applications. As FSL models learn to handle rare or underrepresented terrain types, they open up new possibilities for applications such as vegetation monitoring and land cover analysis, especially in challenging environments where traditional DL models struggle. With advancements in domain adaptation, feature fusion, and meta-learning, FSL is expected to play an increasingly important role in making PolSAR-based solutions more practical and scalable across various remote sensing tasks. A list of key acronyms used in this article is summarized in Table 1. The key contributions of this work are summarized as follows:

- Comprehensive analysis of few-shot learning methods for PolSAR classification: We provide a detailed survey of existing FSL techniques, including data augmentation, transfer learning, meta-learning, and multimodal learning, highlighting their strengths and limitations for PolSAR classification.

- Insightful discussion of challenges and future trends: We analyze the unique challenges of few-shot PolSAR classification, including data scarcity, speckle noise, and cross-domain variability, and propose future research directions, such as physics-informed deep learning.
- Quantitative and qualitative benchmarking: Extensive experiments on standard PolSAR datasets such as Flevoland, San Francisco, and Oberpfaffenhofen evaluate the effectiveness of various methods discussed in this survey. Metrics such as OA, Kappa coefficient, and F1 score are employed to compare and analyze the performance of SOTA approaches.

In this paper, we systematically guide readers through the problem domain, methodologies, and experimental results, beginning with an introduction to PolSAR's role in remote sensing, its advantages, challenges, and the relevance of few-shot learning. We provide formal definitions of few-shot learning, descriptions of widely used PolSAR datasets, and evaluation metrics. We then review existing methodologies, including data augmentation, transfer learning, meta-learning, and multimodal learning, discussing their applicability and limitations. Quantitative comparisons on benchmark datasets and qualitative analyses of the strengths and weaknesses of these methods are presented. Finally, we highlight current challenges, such as data imbalance and speckle noise, discuss emerging trends like self-supervised and physics-informed learning, and conclude with a summary of our findings and recommendations for future research.

Table 1. A list of key acronyms.

Full Form	Abbreviation	Full Form	Abbreviation
Average accuracy	AA	Contrastive learning	CL
Convolutional neural network	CNN	Data augmentation	DA
Deep learning	DL	Few-shot cross-domain	FSCD
Few-shot learning	FSL	Graph neural network	GNN
K-nearest neighbors	KNN	Meta-learning	ML
One-shot learning	OSL	Overall accuracy	OA
Polarimetric synthetic aperture radar	PolSAR	Relation network	RN
Self-supervised learning	SSL	Support vector machine	SVM
Transfer learning	TL	Zero-shot learning	ZSL

2. Definition, Datasets, and Evaluation Metrics

2.1. Few-Shot Image Classification Definition

Few-shot image classification refers to the task of enabling a model to accurately classify new categories with only a few labeled examples per class. FSL is typically framed in an N-way K-shot setting, where the model must classify among N different classes with only K labeled samples per class. For example, in a 5-way 1-shot task, the model needs to distinguish between five classes given just one labeled example per class. Depending on the value of K, few-shot classification can include specific cases like OSL, where K equals 1, and the model learns to classify with a single labeled sample per class. Another extension is zero-shot learning [32], where the model must classify samples from classes it has never seen before, relying on auxiliary information like semantic descriptions or relationships between classes to perform the task.

FSL aims to develop models that generalize well to unseen classes by focusing on learning transferable features. This ability is crucial for applications like PolSAR, where domain shifts—such as changes in environmental conditions or data collected from different satellites—can significantly affect model performance. By enabling accurate classification with minimal labeled data, FSL helps overcome the practical challenges of PolSAR classification, offering a scalable solution for identifying rare terrain types, monitoring vegetation, and supporting other remote sensing tasks with limited annotations.

Figure 2 illustrates a knowledge graph that represents key concepts related to FSL, PolSAR image classification, and deep learning. Each node in the graph corresponds to a specific concept, technique, or approach relevant to PolSAR classification, such as "GANs" and "feature diversity" under data augmentation, or "prototypical networks" and "cosine similarity" linked to meta-learning approaches. The edges between nodes represent relationships or interactions, with connections such as the one between "Transfer Learning" and "Pre-trained Models" highlighting the reliance on pre-trained models as a foundational component of transfer learning techniques. To enhance clarity, the nodes are color-coded according to their methodological categories. The central node, representing PolSAR classification, connects to critical techniques including domain adaptation, meta-learning, and transfer learning, emphasizing the interdisciplinary nature of Few-Shot PolSAR classification. The graph further highlights advanced methods, such as GNNs, cross-modal learning, and attention mechanisms, which are pivotal for addressing challenges like high-dimensional data and inter-class similarity in PolSAR classification. This comprehensive representation effectively captures the relationships and dependencies among various techniques and approaches in the field.

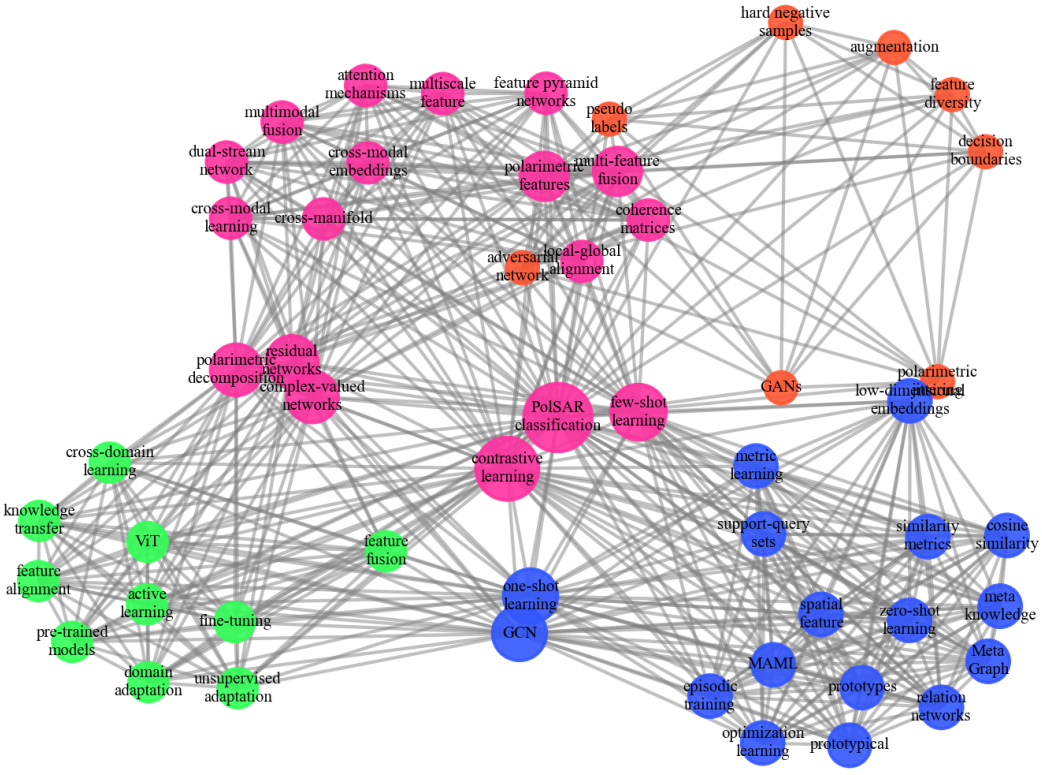

Figure 2. Knowledge graph of key concepts in Few-Shot PolSAR image classification. Nodes represent key concepts, methods, or techniques, while edges indicate the relationships or dependencies between them. Node colors correspond to methodological categories: orange for data augmentation-based methods, green for transfer learning-based methods, blue for meta-learning-based methods, and pink for multimodal-based methods. The graph was constructed based on literature analysis and keyword extraction, with relationships derived from established dependencies in the field.

2.2. PolSAR Datasets

PolSAR datasets are essential resources for few-shot classification tasks. PolSAR datasets can be broadly classified into two categories based on their defining characteristics: location-based and sensor-based. Location-based datasets (e.g., Flevoland, San Francisco) are named after the geographic regions they represent, often serving region-specific studies. In contrast, sensor-based datasets (e.g., RADARSAT-2, Sentinel-1) derive their names from the sensors that acquired them, offering diverse spatial and temporal coverage. These datasets capture diverse terrain types, urban structures, and environmental conditions,

making them useful for a wide range of applications, including land cover classification, vegetation monitoring, and disaster management [33]. This section provides an overview of several commonly used PolSAR datasets, along with their key characteristics, summarized in Table 2.

Table 2. Common PolSAR datasets and their characteristics.

Dataset	Category	Size (Pixels)	Resolution (m)	Bands (Polarizations)	Classes
Flevoland	Location-Based	750 × 1024	6 to 12	HH, HV, VV	15
San Francisco	Location-Based	1300 × 1300	5	HH, HV, VH, VV	5
Oberpfaffenhofen	Location-Based	1300 × 1024	1.5	HH, HV, VV	3
RADARSAT-2	Sensor-Based	Variable	3 to 50	HH, HV, VH, VV	10+
AIRSAR	Sensor-Based	750 × 1024	12.5	HH, HV, VH, VV	9
UAVSAR	Sensor-Based	Variable	1	HH, HV, VH, VV	10+
Sentinel-1	Sensor-Based	Variable	10	HH, HV (Dual)	5+

The Flevoland dataset is widely used for agricultural studies, providing detailed information on 15 crop and land cover types, including wheat, potatoes, beans, and water bodies. The San Francisco dataset, collected over the urban Bay Area with RADARSAT-2, features five key classes such as water, vegetation, and both low- and high-density urban regions. These urban settings offer a challenging testbed for models due to their structural diversity and variability.

The Oberpfaffenhofen dataset, acquired using the ESAR L-band sensor, covers three well-defined terrain categories: built-up areas, woodland, and open areas. Its high-resolution imagery makes it a valuable benchmark for multi-class classification in remote sensing. The AIRSAR and UAVSAR datasets from NASA provide high-resolution, fully polarimetric data for diverse environments, including wetlands, forests, and urban areas, making them ideal for multi-domain classification tasks.

RADARSAT-2 data are used extensively for real-world monitoring applications such as environmental assessment and disaster response, with its flexible acquisition modes supporting multiple resolutions and polarizations. Finally, Sentinel-1 data, though limited to dual-polarization, provide frequent revisits and global coverage, making these data essential for dynamic monitoring tasks.

These datasets highlight the importance of models capable of generalizing across varied environments and sensors. FSL presents a promising solution by enabling effective performance with limited labeled data, offering scalability and robustness across a range of PolSAR classification scenarios.

2.3. Evaluation Metrics for Few-Shot PolSAR Image Classification

Evaluating the performance of DL models for few-shot PolSAR image classification requires metrics that reflect both classification accuracy and the model's ability to generalize to new classes and environments. The following section outlines key metrics commonly used in the literature to assess the effectiveness of few-shot PolSAR classification models.

Overall accuracy (OA) [34] measures the proportion of correctly classified samples over the total number of samples across all classes. This metric provides a high-level view of model performance but can be biased toward majority classes, making it less informative when dealing with imbalanced datasets or rare classes [35].

$$OA = \frac{1}{n}\sum_{i=1}^{n} 1(\hat{y}_i = y_i), \tag{10}$$

where n is the total number of samples, y_i is the true label, and $\hat{y}_i$ is the predicted label.

Average accuracy (AA) [36] addresses the class imbalance issue by computing the mean accuracy for each class independently and then averaging the results. It ensures that smaller or rare classes are given equal weight in the final evaluation, which is crucial for applications like PolSAR terrain classification, where some classes (e.g., rare vegetation types) have few samples.

$$AA = \frac{1}{N}\sum_{j=1}^{N}\frac{TP_j}{TP_j + FN_j}, \tag{11}$$

where N is the total number of classes, TP_j is the number of true positives for class j, and FN_j is the number of false negatives. The formula for AA is equivalent to calculating the mean of the per-class OA, which is given by the following:

$$Mean(OA) = \frac{1}{N}\sum_{j=1}^{N}OA_j = \frac{1}{N}\sum_{j=1}^{N}\frac{TP_j}{TP_j + FN_j}. \tag{12}$$

This confirms that AA and the mean of per-class OA values are identical. The key difference lies in the interpretation: AA ensures that each class contributes equally to the final metric, addressing class imbalance issues.

The Kappa coefficient [37] evaluates the agreement between the predicted and true classifications, adjusting for the agreement occurring by chance. It is particularly useful for PolSAR applications where the number of classes and data imbalance can skew simpler accuracy metrics [38].

$$\kappa = \frac{OA - p_e}{1 - p_e}, \tag{13}$$

where OA is the overall accuracy, and p_e is the expected agreement by chance, given by the following:

$$p_e = \sum_{i=1}^{N}\left(\frac{n_i^{\text{pred}} \times n_i^{\text{true}}}{n^2}\right). \tag{14}$$

Here, N is the total number of classes, n_i^{pred} is the number of samples predicted as class i, n_i^{true} is the number of samples truly belonging to class i, and n is the total number of samples.

The F1 score [39] balances precision and recall, providing a harmonic mean between the two. In the few-shot setting, both Macro-F1 [40] (which computes the metric for each class and takes their unweighted average) and Micro-F1 [41] (which aggregates contributions of all classes) are relevant, especially in scenarios with rare or underrepresented classes [42].

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}. \tag{15}$$

$$Precision = \frac{TP}{TP + FP}. \tag{16}$$

$$Recall = \frac{TP}{TP + FN}. \tag{17}$$

The Micro-F1 score is calculated by aggregating the TP, FP, and FN across all classes, and then computing the overall precision and recall, as follows:

$$Micro\text{-}F1 = 2 \times \frac{Precision_{micro} \times Recall_{micro}}{Precision_{micro} + Recall_{micro}}$$

$$Precision_{micro} = \frac{\sum_i TP_i}{\sum_i TP_i + \sum_i FP_i}, \quad Recall_{micro} = \frac{\sum_i TP_i}{\sum_i TP_i + \sum_i FN_i}$$

Here, the sums are computed across all classes. Micro-F1 is more sensitive to the overall distribution of correct and incorrect predictions, making it ideal for evaluating models on the entire dataset, especially when there is a class imbalance.

In contrast, the Macro-F1 score computes the F1 score for each class independently and then averages the results. This approach treats each class equally, regardless of the number of samples, and is especially useful when assessing how well a model performs across all classes. The Macro-F1 score is defined as follows:

$$Macro\text{-}F1 = \frac{1}{N}\sum_{j=1}^{N}\left(2 \times \frac{Precision_j \times Recall_j}{Precision_j + Recall_j}\right)$$

where $Precision_j$ and $Recall_j$ are computed for each class j, and N is the total number of classes. Unlike Micro-F1, the Macro-F1 score provides an average performance across all classes and is particularly useful when the importance of smaller or rare classes is significant.

A confusion matrix [43] provides detailed insights into classification errors by showing the distribution of true and predicted labels for each class. For PolSAR data, confusion matrices help identify specific pairs of classes that are often confused due to similar scattering characteristics (e.g., forest and shrubland). This information is critical for refining feature extraction techniques.

In FSL, models are evaluated on episodic tasks, where each episode mimics the target scenario by randomly sampling support and query sets. Episodic accuracy [44] measures the proportion of correct predictions across multiple episodes and reflects the model's ability to generalize from a few samples [45]:

$$Episodic\ Accuracy = \frac{1}{M}\sum_{m=1}^{M} 1(\hat{y}_{m,i} = y_{m,i}), \tag{18}$$

where M is the total number of episodes, $y_{m,i}$ is the true label in episode m, and $\hat{y}_{m,i}$ is the predicted label.

The area under the receiver operating characteristic curve (AUC-ROC) [46] measures the ability of the model to distinguish between positive and negative classes across varying thresholds. In PolSAR classification, this metric can be particularly informative for rare terrain types or binary classification tasks, such as distinguishing between urban and non-urban areas [47].

These metrics collectively provide a comprehensive framework for evaluating the performance of DL models in few-shot PolSAR image classification. They not only capture classification accuracy but also address key aspects like class imbalance, generalization across domains, real-time applicability, and task-specific performance. The selection of appropriate metrics depends on the specific requirements of the task, such as the importance of rare class detection, cross-domain generalization, or inference speed.

3. Few-Shot Learning for PolSAR Image Classification

3.1. Data Augmentation-Based Methods

Data augmentation plays a crucial role in overcoming the limitations posed by small datasets in FSL. In PolSAR image classification, the scarcity of labeled data is particularly problematic due to the complex nature of polarimetric data and the high cost of obtaining annotations. DA addresses this by generating synthetic samples through various transformations, which enhance the diversity of the training set and improve model robustness. This section explores the recent advancements in DA techniques applied to PolSAR FSL.

3.1.1. Traditional Data Augmentation Techniques

Traditional DA methods, such as flipping, rotating, and scaling, are frequently used to create variations of the existing data. For PolSAR images, these transformations are applied to manipulate spatial features and polarimetric channels. For example, Ashwin and Ansal [39] applied simple transformations like rotations and translations to PolSAR

datasets to improve model generalization across different terrain types. Zhang et al. [48] proposed a multi-task DL framework designed to efficiently utilize complex-valued PolSAR image data for classification. The core innovation of this approach lies in the extraction and separate processing of amplitude and phase information from PolSAR data, which creates a richer set of data representations. These diverse representations act as an implicit data augmentation strategy by improving the model's ability to handle the variability and complexity of PolSAR data. Additionally, a depthwise separable convolution is introduced to enhance phase-related feature extraction, which further increases the model's robustness and generalization to unseen classes. Wang et al.'s [49] proposed DA method enhances model performance by redistributing labeled data across different image regions. This approach mitigates the model's over-reliance on specific areas, thereby reducing issues such as label skew and class imbalance. By applying transformations to shuffle and merge labeled data, the model gains a more comprehensive understanding of various spatial regions, improving its generalization capability and resilience to noise and occlusions. While these basic techniques increase the number of training samples, they may not fully capture the complexity of PolSAR data, such as speckle noise. Therefore, specialized augmentation methods that consider the unique characteristics of PolSAR data are necessary [50].

3.1.2. Advanced Augmentation Techniques with GANs

Generative adversarial networks (GANs) have become a powerful tool for generating realistic synthetic samples in FSL. Recent work has adapted GANs for PolSAR DA by training the generator to create polarimetric images that mimic the properties of real PolSAR data. Physical modeling plays a role in ensuring that GAN-generated samples remain realistic and domain-relevant. This ensures that synthetic data captures the statistical and physical consistency of real-world PolSAR images. Similarly, feature augmentation methods, such as polarimetric jittering, introduce controlled variations in features like phase and amplitude while preserving underlying scattering mechanisms. These physically grounded augmentations expand the diversity of training data while maintaining interpretability and robustness. As shown in Figure 3, the process begins with inputting PolSAR data represented by a tensor containing a pixel and its surrounding neighborhoods. A noise vector and a label vector are then generated and combined to form a connected vector, which the generator (G) transforms into synthetic data. Both real and synthetic (fake) data are then fed into the discriminator (D) for feature extraction and normalization. The discriminator's outputs undergo a normalization step before being used to calculate the loss functions D_{loss} and G_{loss}. Finally, the combined loss is backpropagated to optimize both the generator and discriminator networks, enhancing the model's ability to generate realistic PolSAR samples tailored to specific classification or clustering tasks. Liu et al. [51] introduced task-oriented GAN (TOGAN), a GAN framework specifically designed for PolSAR image classification and clustering. TOGAN goes beyond traditional GANs by incorporating a task-specific network (TaskNet), which ensures that the generated PolSAR samples are tailored to improve the performance of specific tasks, such as classification or clustering. This task-oriented approach enhances the variability and relevance of the generated data, helping classifiers and clustering algorithms generalize better, even with limited labeled data.

Yang et al. [52] employed a GAN to generate hard negative samples near decision boundaries, enhancing the model's ability to distinguish between challenging examples. The proposed approach demonstrates significant performance improvements on PolSAR datasets, particularly in complex boundary regions with inconsistent terrain types. Xie et al. [53] propose a method that leverages auxiliary classifier generative adversarial network (ACGAN) to generate realistic PolSAR samples that enhance FSL. By incorporating an auxiliary classifier, the generator produces class-specific synthetic samples, significantly improving classification performance. Dong et al. [54] introduced the polarimetric scattering characteristics-guided adversarial network (PSCAN) for unsupervised PolSAR image classification, incorporating several key innovations in adversarial learning. A major in-

novation is the integration of polarimetric scattering characteristics into the adversarial learning process, improving feature discriminability by constructing pseudo-labels based on scattering mechanisms rather than automatic labeling. This approach enhances feature transferability while ensuring the semantic alignment between source and target domains. Additionally, PSCAN incorporates an auxiliary task that utilizes these pseudo-labels to further refine the learned features. Song and Xu [55] introduced a ZSL framework using deep generative networks to construct a continuous SAR target feature space. This enables the generation of synthetic data for unseen classes, improving classification in cases where labeled samples are unavailable. By learning orientation-invariant features, their model enhances generalization, making it a valuable addition to GAN-based augmentation.

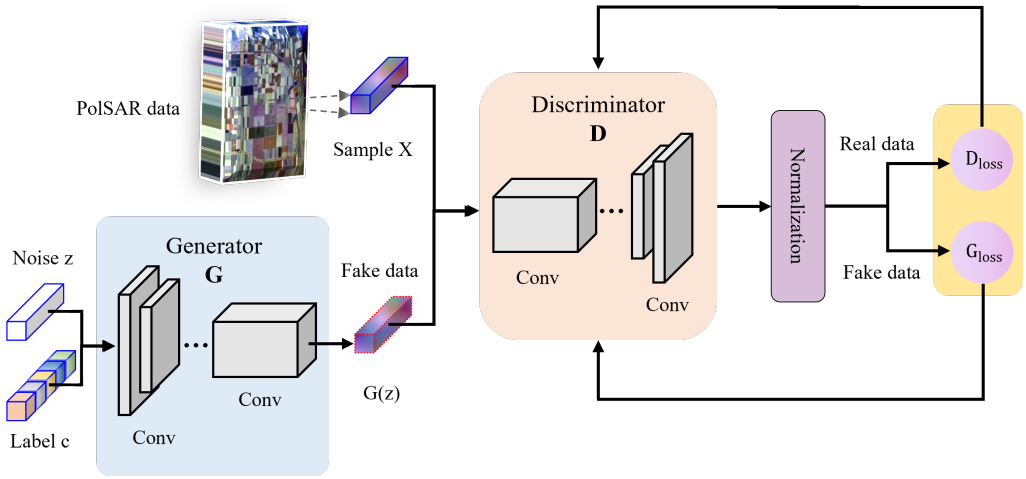

Figure 3. Overview of the GAN-based PolSAR data augmentation pipeline, illustrating the flow from raw PolSAR input to the final loss optimization stage for both the generator and discriminator.

3.2. Transfer Learning-Based Methods

TL has emerged as a critical tool for addressing the challenges posed by limited labeled data in PolSAR FSL. By transferring knowledge from models trained on large datasets, this approach enables effective PolSAR classification with only a few labeled samples. This section explores different strategies for applying TL in PolSAR FSL, highlighting recent research and methodologies.

3.2.1. Pre-Trained Models for Feature Extraction

A widely used approach in PolSAR FSL involves leveraging pre-trained models for feature extraction, which reduces the reliance on large labeled datasets. These models, often initialized with features from external datasets, are fine-tuned to adapt to PolSAR data. Incorporating physically derived features, such as those from polarimetric decompositions (e.g., Cloude–Pottier, Freeman–Durden), ensures the inclusion of domain-specific knowledge like scattering mechanisms and surface characteristics. During fine-tuning, physical modeling further aids in aligning features across domains, enhancing adaptability to target tasks.

Wang et al. [56] proposed a ViT-based model for PolSAR land cover classification. Pre-trained using a masked autoencoder on unlabeled PolSAR data, the model effectively captures long-range dependencies, improving classification accuracy on the Flevoland and Hainan datasets, even with limited labeled data. Cattoia et al. [57] proposed a transcoding-based pre-training approach, where a network is pre-trained to translate PolSAR data into optical images using a regression network, conditional GAN, or cycle-GAN. The learned features are then fine-tuned for semantic segmentation. The pre-trained models, particularly the conditional GAN, demonstrated significant improvements in accuracy, especially in scenarios with limited labeled data. Han et al. [58] presented a DL-based approach for PolSAR image classification by utilizing deep features extracted

through CNNs. The authors proposed a framework that integrates pre-trained CNNs to extract discriminative features from PolSAR images, followed by a classifier for land cover classification. Xie et al. [59] proposed a channel adaptive complex-valued fully convolutional network (TF-CVFCN) with a novel TL method to address the challenges of limited labeled data in PolSAR terrain classification. The key innovation lies in a new TL strategy that allows the flexible design of the network structure and independent layer freezing, enabling faster convergence and better performance. Additionally, the TF-CVFCN framework fully utilizes complex-valued PolSAR data to enhance feature extraction and incorporates lightweight deep separable convolutions optimized through a multi-objective particle swarm optimizer.

3.2.2. Domain Adaptation for PolSAR Classification

Domain adaptation addresses the domain shift problem, where discrepancies between source and target datasets degrade the model's performance. For PolSAR data, domain shifts can arise from different geographic regions, sensor types, or imaging conditions. Domain adaptation techniques help align feature distributions across domains to ensure the model generalizes well in the target domain. This subsection explores various domain adaptation strategies applied to PolSAR few-shot learning.

Gui et al. [60] introduced the unsupervised generalized zero-shot domain adaptation (uGZSDA) framework for PolSAR classification. The method addresses class distribution shifts between domains by using scattering component semantics (SCS) and a stacked autoencoder (SAE). This approach enhances classification accuracy in target domains without requiring labeled data. Gui et al. [61] proposed a general feature paradigm (GFP) for unsupervised cross-domain PolSAR image classification, addressing challenges posed by domain shifts due to differences in sensors, imaging angles, and land cover distributions. The innovation lies in a four-step feature transformation process that optimizes interclass aggregation and reduces domain shifts. GFP is compatible with typical domain adaptation methods, facilitating its application across various cross-domain PolSAR tasks.

Li et al. [62] proposed an online active extreme learning machine (OA-ELM) with a novel discrepancy sampling strategy for PolSAR image classification. This method combines online sequential learning and active learning, significantly improving classification accuracy and efficiency. The discrepancy sampling strategy directly uses the non-probabilistic outputs of ELM to select informative samples. Dong et al. [63] introduced a causal inference-guided feature enhancement framework for PolSAR image classification. The key innovation lies in using a structural causal model (SCM) to identify and intervene in variables affecting feature discriminability and generalizability. By applying backdoor adjustment, the framework eliminates confounding factors in the feature extraction process, preventing overfitting and enhancing feature discriminability. Qin et al. [64] proposed a relational-based transductive transfer learning method for PolSAR image classification. The innovation lies in a three-phase time-series clustering algorithm that leverages spatiotemporal relational knowledge, enabling accurate label transfer without relying on labeled data from the target domain. Gui et al. [65] proposed an eigenvalue statistical component-based PU-Learning (ESC-PUL) framework for the extraction of built-up areas in PolSAR images, particularly addressing challenges related to domain shifts and orientation angles. The key innovation is the integration of ESC with PU-Learning, enabling robust extraction of built-up areas using only positive samples while avoiding the need for negative samples. Additionally, a subspace alignment technique is incorporated to address domain shifts between different PolSAR datasets, facilitating cross-domain analysis.

Hua et al. [66] proposed an unsupervised domain adaptation framework utilizing a coordinate attention mechanism to enhance spatial feature extraction and introducing a weighted clustering algorithm to generate reliable pseudo-labels for the target domain. An adversarial network with a bi-classifier is used to align data distribution between domains. Cao et al. [67] proposed a complex-valued cross-domain few-shot learning network (CVCDFSL) to address the challenge of few labeled samples and domain shifts in

PolSAR image classification. The method leverages complex-valued data representations and cross-domain learning techniques, employing a novel feature alignment strategy in the complex domain to achieve superior classification performance in cross-domain few-shot settings. Sun et al. [68] introduced a method for domain adaptation in PolSAR land classification utilizing linear discriminative Laplacian Eigenmaps (LDLEs). This approach projects source and target domain data into a shared subspace where their distributions are aligned, preserving both discriminative and geometric structures. Sun et al. [69] proposed a scatter matrix-based domain adaptation framework for classifying bi-temporal PolSAR images. This method optimizes a scatter matrix to align feature distributions between source and target domains, improving the adaptability of classifiers trained on historical data to classify new, unlabeled temporal data.

3.2.3. Self-Supervised Learning for Transfer Learning

Recent advancements in SSL have significantly enhanced the potential of TL in PolSAR FSL. Self-supervised pre-training enables the model to learn robust feature representations from large amounts of unlabeled PolSAR data by solving pretext tasks, such as CL or clustering. The model can then be fine-tuned on a few labeled samples, providing a substantial performance boost. In the SSL framework, as depicted in Figure 4, a shared backbone network is employed for both the teacher and student models. The student network is updated directly, while the teacher network is updated via the exponential moving average of the student parameters. Two augmented views of the same input PolSAR data are generated, denoted as x_1 and x_2. The student model processes x_1 to predict a probability distribution p_1, while the teacher model processes x_2 to generate p_2 after centering. This setup enables the model to maximize similarity between corresponding representations, which aligns with the principles of mutual learning and enhances robustness to variations in input data. The lower section of the figure illustrates the classification process, where the pre-trained SSL model is fine-tuned on a small labeled dataset, resulting in refined land cover classification outputs. This dual-phase structure—pre-training with self-supervised tasks followed by supervised fine-tuning—captures complex scattering mechanisms and target decompositions in PolSAR data, leading to improved classification accuracy with few-shot labeled data.

Ren et al. [70] introduced a mutual information-based SSL framework aimed at improving PolSAR land cover classification with minimal labeled data. The key innovation lies in exploiting the inherent properties of PolSAR data without relying on manual labels. The model maximizes mutual information between multi-modal features of the same pixel while enhancing differences across different pixels, capturing complex scattering mechanisms in a self-supervised manner. Zhang et al. [71] proposed MLR-SimSiam, a contrastive pre-training model designed for few-shot PolSAR image classification. The innovation lies in two key components: polarimetric jittering (PJ), which creates challenging positive samples by leveraging polarimetric target decomposition, and the mutual learning regularizer (MLR), which bridges the semantic gap between different augmented data representations, improving robustness and generalization. Kuang et al. [72] proposed a complex-valued SSL framework integrating an attention mechanism to focus on informative channels. The method incorporates both amplitude and phase information, combined with a noise injection data augmentation strategy tailored to PolSAR noise characteristics, enhancing model robustness and reducing reliance on labeled data.

Zhang et al. [73] proposed a self-supervised PolSAR representation learning (SSPRL) model, designed to enhance PolSAR land cover classification, particularly in FSL scenarios. The core innovation of this model is its ability to learn robust feature representations from unlabeled PolSAR data using a CL framework without relying on negative samples. The SSPRL framework includes a dynamic convolutional (DyConv) encoder and employs a novel positive sample generation (PSG) method tailored for PolSAR data. Additionally, the model integrates a mix-up regularization strategy, further improving the generalization of learned features. Qiu et al. [74] proposed an FSL method for PolSAR ship detection

by integrating polarimetric feature selection with improved contrastive self-supervised learning (CSSL). The method leverages eight polarimetric feature extraction techniques and pre-trains the backbone network with CSSL, enhancing representation without negative samples. A multi-scale feature fusion module (MFFM) boosts feature learning, and the mix-up auxiliary pathway (MUAP) serves as regularization. The model is fine-tuned using a few labeled samples for ship detection. Wang et al. [75] proposed a dual-branch PolSAR image classification model that combines generative SSL with local feature extraction to address the challenge of limited labeled data in PolSAR image classification. The key innovation lies in the integration of a graph-masked autoencoder (GraphMAE) for learning superpixel-level polarimetric representations, alongside a CNN-based pixel branch to capture fine-grained pixel-level features. By fusing the features extracted from both branches, the model achieves superior classification accuracy and clearer boundary delineation compared to single-branch models.

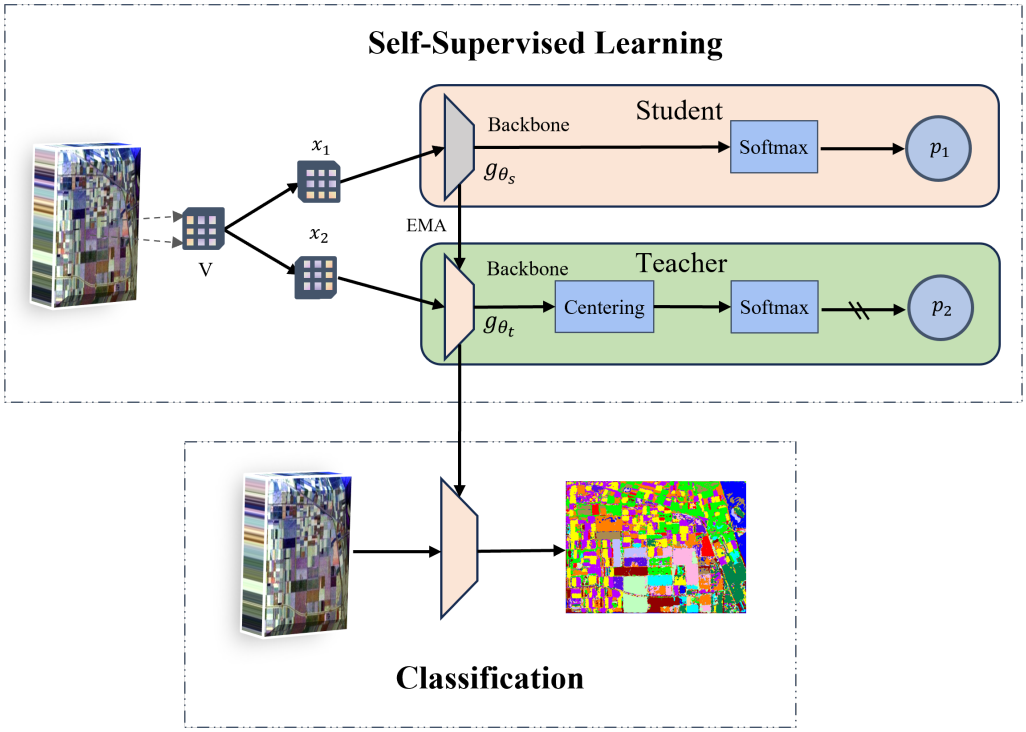

Figure 4. Overview of the self-supervised learning framework for PolSAR image classification.

Dong et al. [76] introduced a ViT-based framework for PolSAR image classification, marking the first application of ViT in this domain. The key innovation is leveraging self-attention to capture long-range interactions between image patches. The framework supports both supervised and unsupervised learning, with a novel contrastive learning strategy simplifying unsupervised pre-training. Cai et al. [77] introduced heterogeneous network-based contrastive learning (HCLNet) to leverage multi-features of PolSAR data by combining physical and statistical features for unsupervised high-level representation. The heterogeneous architecture integrates 1D and 2D CNNs to learn complementary features, addressing scattering confusion and enhancing land cover distinction. Darvishnezhad and Sebt [78] introduced a self-supervised ensemble learning framework (SSELF) for land use and land cover classification. The framework integrates EfficientNet-B0 for feature extraction and employs Deep Curriculum Learning to rank image patches by complexity. Ensemble learning strategies further enhance classification accuracy by leveraging features across various spatial scales and polarimetric bands.

3.3. Meta-Learning-Based Methods

ML has gained significant attention for its ability to learn how to learn, particularly in few-shot classification tasks. Unlike traditional methods that rely heavily on large labeled datasets, ML focuses on acquiring meta-knowledge from previously seen tasks and using this knowledge to generalize to new, unseen tasks with limited labeled data. In ML, as shown in Figure 5, tasks are divided into support and query sets across both training and testing stages, enabling models to learn generalizable features. During training, the model learns from diverse tasks generated from datasets with more categories, such as the Flevoland 15 dataset. Each task provides a support set (e.g., "Bean", "Building", "Bare soil") and a query set from the same categories. This episodic training strategy builds meta-knowledge, allowing the model to handle new tasks with minimal labeled samples. In the testing stage, unseen tasks from the Flevoland 7 and Oberpfaffenhofen datasets test the model's generalization. Each task consists of distinct support and query sets, challenging the model to classify categories such as "Linen", "Canola", and "Woodland", which differ from the training categories. This setup reflects the FSL approach, where models leverage meta-knowledge to classify unseen classes effectively with limited samples, even across different domains.

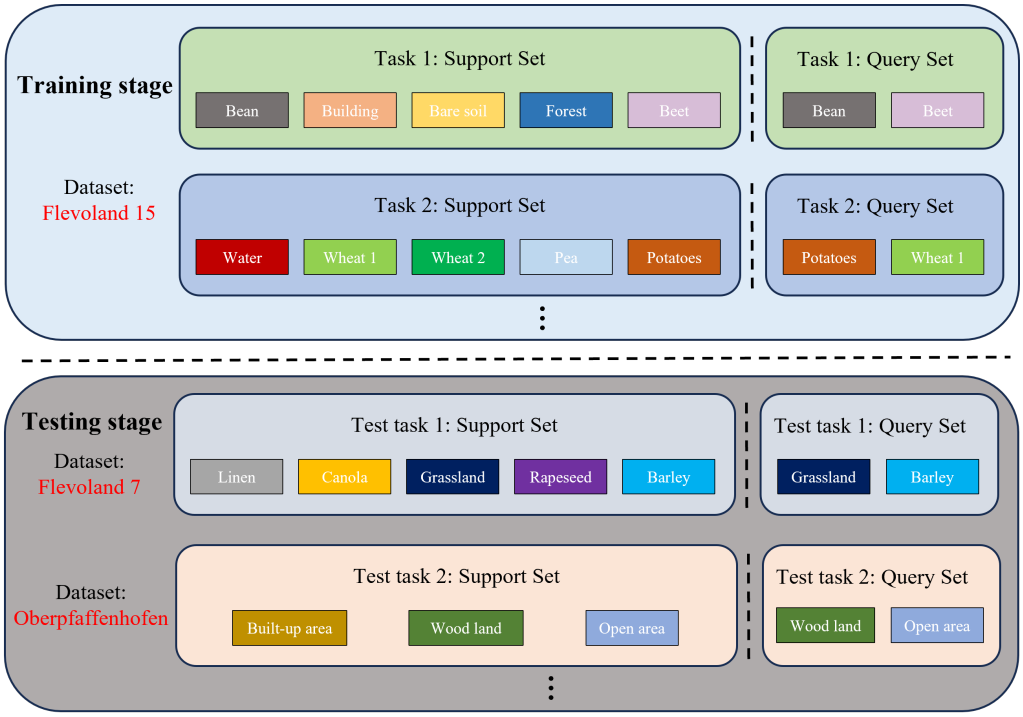

Figure 5. Data division in meta-learning-based few-shot PolSAR classification, showing the training and testing stages with support and query sets for each task.

In metric-based meta-learning (e.g., prototypical networks), physical features derived from PolSAR decomposition methods are used to define class prototypes. For example, class embeddings may incorporate features such as polarimetric entropy (H) or anisotropy (A), reflecting the physical separability of scattering mechanisms (e.g., urban vs. vegetation vs. water). By embedding these physically grounded features into the embedding space, meta-learning improves both the interpretability and robustness of class similarity metrics, enabling models to generalize effectively across tasks with minimal labeled data. Similarly, optimization-based methods, such as model-agnostic meta-learning (MAML), leverage physical priors by pretraining initialization parameters using scattering features, ensuring rapid and reliable adaptation to new tasks.

3.3.1. Metric-Based Meta-Learning

One of the most popular categories in ML is metric-based learning. This approach learns a distance metric that allows a model to classify new data points based on their proximity to labeled examples in a learned embedding space. Ni et al. [79] proposed a metric learning-based approach for fine-grained PolSAR image classification, addressing class similarity challenges in complex scenes. The method utilizes magnet loss, which enhances feature discrimination by clustering samples and penalizing class distribution overlaps. This improves the ability to capture subtle differences between similar classes, crucial for fine-grained classification. Alternative classifiers like kNN and SVM are used to replace softmax, further boosting classification accuracy. Yang et al. [80] proposed a meta-graph representation learning (MGRL) framework for PolSAR image classification, particularly addressing the challenges of FSL across different platforms and categories. The key innovation lies in combining a dual-path graph convolutional network (DGCN) with ML to capture both global and local spatial dependencies in PolSAR images. The framework integrates a trumpet convolutional network (TCN) to extract local scattering features and a GCN to model global topological structures. Zhang et al. [81] proposed a metric-based meta-learning framework (MML-FSTC) for few-shot PolSAR terrain classification. The model consists of two key components: transferable knowledge learning (TKL), which uses a 3D convolutional neural network (3DCNN) to extract transferable features from base labeled samples, and meta-knowledge learning (MKL), which fine-tunes the pre-trained 3DCNN through few-shot training episodes. These episodes use support and query sets to learn the metric-based cosine distance function, enabling the model to classify new terrain types with minimal labeled data. Shang et al. [82] introduced a novel spatial feature-based convolutional neural network (SF-CNN) to address the challenges of limited labeled data and speckle noise in PolSAR image classification. The key innovation of SF-CNN lies in its dual-branch architecture, which enables input expansion by processing groups of samples rather than individual ones, thereby effectively mitigating the issue of insufficient training data. Additionally, the network enhances the discriminative power of the extracted features by maximizing inter-class distances and minimizing intra-class distances within a low-dimensional feature space.

Dong et al. [83] proposed a metric learning-based method for extracting collapsed buildings from post-earthquake PolSAR imagery using eight building-related features such as entropy (H), scattering angle (α), anisotropy (A), and Yamaguchi decomposition scattering powers. The method improves the information-theoretic metric learning (ITML) algorithm by projecting these features into a lower-dimensional space to reduce the effects of topography and aspect angles. A kNN classifier is then used to classify buildings as collapsed or intact. Zhang et al. [84] introduced a novel feature evaluation method, called integrating multiple metrics with collinearity considered (IMC), for mapping forest above-ground biomass (AGB) using PolSAR data. The IMC method fuses various feature evaluation metrics, including linear (Pearson correlation), nonlinear (importance from random forests), and physical (sensitivity index) metrics, while addressing information redundancy among features. Prototypical networks [28] and relation networks [29] are two prominent examples of metric-based ML. Cao et al. [44] proposed a metric-based ML approach as part of their complex-valued cross-domain FSL classification (CCFSLC) framework for PolSAR image classification. This method is designed to handle domain shifts and the scarcity of labeled data across domains. Specifically, they employed cosine similarity as a metric to measure the similarity between support and query features. In each episode of ML, a support set of labeled samples and a query set of unlabeled samples are used to refine the model. By minimizing the distance between the support and query features, the model learns to generalize effectively to new, unseen classes. Hua et al. [85] developed a relation network framework enhanced by a spatial weighted attention network (SWANet). The method leverages metric-based ML principles to calculate similarity scores between query and support samples. SWANet employs spatial attention mechanisms to improve the representation of polarimetric and spatial features, which is crucial in

distinguishing complex land covers in PolSAR images. This combination of RN and attention modules significantly improves classification accuracy.

3.3.2. Optimization-Based Meta-Learning

Optimization-based ML methods have shown great promise for adapting models to new tasks with minimal labeled data and few gradient updates. These methods are particularly effective in FSL, where rapid adaptation is required due to the scarcity of labeled samples. One of the most well-known optimization-based methods is model-agnostic meta-learning (MAML), which has been widely applied to various tasks, including natural image classification.

In the context of PolSAR image classification, where unique challenges such as speckle noise and low intraclass variability complicate model training, MAML has been explored as a baseline approach. Zhang et al. [27] employed MAML in their few-shot PolSAR experiments, finding that while MAML is effective at rapid task adaptation, it struggles to achieve optimal performance due to the specific characteristics of PolSAR data. The limited intraclass diversity and complex polarimetric scattering mechanisms hinder MAML's ability to fully generalize to new classes. To address this issue, they proposed a PolSAR-specific contrastive learning network (PCLNet) that significantly outperformed MAML by better capturing these unique data characteristics. Huang et al. [86] introduced a deep reinforcement learning-based method for fully polarized SAR image classification, which leverages multiple polarimetric features for improved accuracy. The key innovation of this approach is the integration of a deep Q-network with polarimetric feature extraction methods, allowing the model to interactively learn by exploring the classification environment. The agent selects actions based on polarimetric features and rewards, enabling efficient classification even with small sample sizes. Liu et al. [87] proposed an Adaptive Graph Convolutional Network (AdapGCN) for PolSAR image classification, introducing innovations that enhance pixel-level classification. A key contribution is the design of pixel-centered subgraphs, enabling each pixel to aggregate information from both local and nonlocal neighbors for more adaptive and context-aware feature extraction. The framework also employs data-adaptive and spatial-adaptive kernels, integrated into the graph convolution process to model data structure and spatial relationships effectively. These kernels improve pixel communication while reducing computational complexity. Additionally, a multiscale learning strategy is used to capture complex relationships across different spatial scales.

3.4. Multimodal-Based Methods

Multimodal learning involves combining information from different data sources or modalities to build more comprehensive and robust models [88]. By integrating data from multiple sources or modalities, these methods aim to exploit complementary information that can enhance classification performance when labeled PolSAR data are scarce. The multimodal process depicted in Figure 6 illustrates the integration of multiple features for few-shot PolSAR image classification. In this framework, multi-modal features, including coherence matrices and various target decomposition methods such as Pauli, Freeman, and $H/A/\alpha$, are extracted to enhance classification. These features (labeled as x_1 and x_2) are processed through separate feature extraction pipelines, allowing each modality to capture distinct characteristics. By employing a combination of these features, the model can better represent complex terrain types in PolSAR images. Each feature extraction module independently processes the respective modality data, with a focus on capturing complementary information that might be absent in single-modality data. The final step is the loss computation ($L_{i,j}$), highlighted in the pink block. This loss function integrates the multimodal features to guide the training process, ensuring that the model optimally leverages complementary information from both input types.

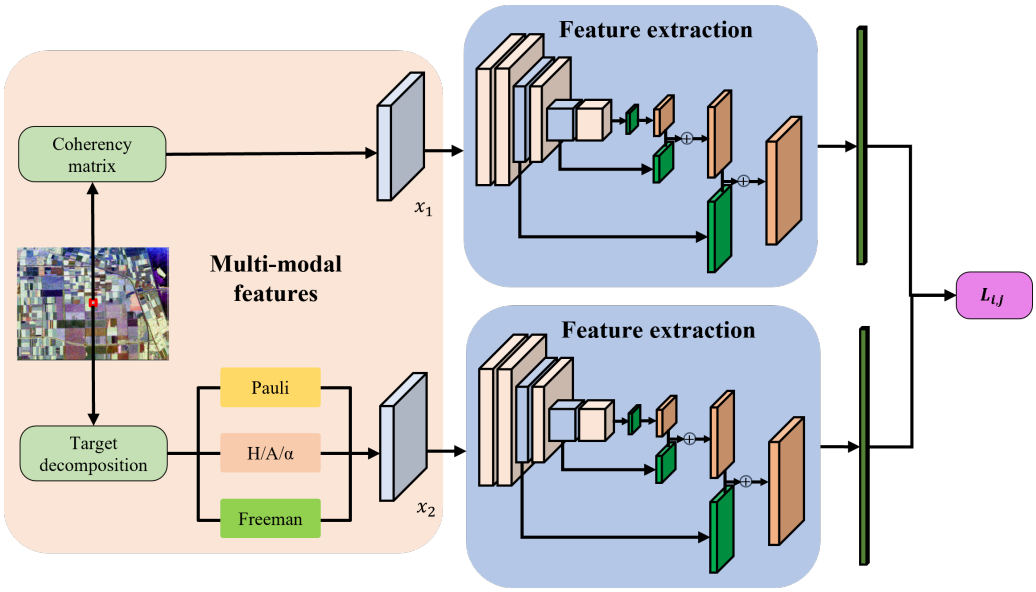

Figure 6. The diagram illustrates the multimodal feature extraction and fusion process. Coherency matrices and target decomposition features (Pauli, Freeman, $H/A/\alpha$) are separately processed through feature extraction pipelines (blue background) to capture complementary spatial, polarimetric, and semantic information. The final loss computation ($L_{i,j}$) integrates these features to optimize classification performance.

Multimodal frameworks integrate PolSAR features, such as polarimetric decompositions, with spectral or topographic features from other modalities. Physical modeling plays a crucial role in ensuring that fused features respect domain-specific constraints. For instance, polarimetric entropy or scattering coefficients derived from PolSAR decompositions are aligned with spectral features from optical data based on shared physical properties, such as surface reflectance or vegetation indices. Cross-modal alignment modules incorporate these constraints to enhance the coherence of fused representations, ensuring that the model captures complementary information without compromising physical consistency. This section explores various multimodal approaches applied to few-shot PolSAR classification.

3.4.1. Cross-Modal Learning

Cross-modal learning enables models to utilize data from different modalities (e.g., combining PolSAR data with optical or textual data) to improve few-shot classification tasks. In traditional unimodal approaches, PolSAR classification relies solely on polarimetric data, which may not fully capture all relevant information, particularly in complex environments. However, by integrating additional modalities, the model can learn more comprehensive representations of the target objects or scenes.

Quan et al. [89] proposed a novel cross-modal feature learning approach that effectively fuses SAR and optical data to improve land cover classification. The innovation lies in the use of a dual-stream network, where SAR and optical branches process respective modality features. A cross-modal interaction module enhances the integration of these complementary features, improving the model's ability to classify land cover types with distinct data characteristics. This method significantly outperforms traditional single-modal approaches. Dong and Hänsch [90] proposed a multimodal SSL framework for semantic analysis of PolSAR imagery, aiming to enhance segmentation performance by leveraging both SAR and optical data. The key innovation of this method lies in the extension of the DINO framework (self-distillation with no labels) using Vision Transformers and incorporating multimodal data (SAR and optical) during the pretraining phase. This allows the model to learn more diverse and generalized representations, which significantly improves semantic segmentation results, especially in scenarios with limited labeled data. Li et al. [91]

proposed the multi-feature dual-stage cross manifold attention network (MF-DCMANet) for PolSAR target recognition, introducing several key innovations in feature extraction and fusion. One of the primary innovations is the combination of monogenic features and polarization features through a dual-stage cross-fusion process. The cross-feature network (CFN) is utilized to extract mid-level semantic information, while the cross-manifold attention (CMA) module captures nonlinear relationships between features on the Grassmann manifold, enhancing the model's ability to represent rich and fine-grained information. Furthermore, the use of local attention windows improves local feature representation while reducing computational costs. Jin et al. [92] introduced a cross-modal contrastive learning (CMCL) method designed for remote sensing image classification. This approach incorporates both intra-modal contrastive learning (IMCL) and CMCL, optimizing them jointly to capture semantic consistency within and between modalities. The innovation lies in the hybrid cross-modal fusion module (HCFM), which compactly integrates local and global cross-modal features. This allows for more effective classification of multimodal remote sensing images (MRSI) by capturing global dependencies and complementary information across modalities. Hong et al. [93] introduced X-ModalNet, a semi-supervised deep cross-modal framework that integrates multispectral imagery (MSI), synthetic aperture radar (SAR), and hyperspectral imagery (HSI) for improved remote sensing classification. The key innovation lies in its three-module architecture: a self-adversarial (SA) module to enhance feature robustness, an interactive learning (IL) module to facilitate multimodal feature fusion, and a label propagation (LP) module to improve classification in semi-supervised scenarios. Wang et al. [94] proposed a novel cross-modal graph knowledge representation and distillation learning (CGKR-DL) framework for land cover classification. The key innovation lies in leveraging both GCNs and CNNs to capture the remote topology and local feature structures of multimodal remote sensing data. This method addresses the limitations of traditional CNN-based cross-modal distillation methods by introducing a multi-granularity graph distillation module, which facilitates stable and effective knowledge transfer. The framework integrates a feature distillation module based on graph discrimination to guide knowledge transfer from teacher to student networks, enabling accurate classification even when some modal data are missing.

3.4.2. Multimodal Fusion

In PolSAR FSL, multimodal fusion techniques can be applied to combine the strengths of multiple data sources [95]. Ren et al. [96] proposed a multimodal sparse representation (MSR) framework for PolSAR image classification, focusing on feature fusion across different modalities. The key innovation lies in combining polarimetric data features, target decomposition features, and texture features to form a comprehensive multimodal representation. The framework utilizes multimodal manifold regularizations to preserve the intrinsic structure of the data from various modalities. The fusion process projects these high-dimensional multimodal features into a low-dimensional space, improving classification performance. Sebt and Darvishnezhad [97] proposed a feature fusion method based on a local binary graph for PolSAR image classification, integrating features extracted from both CNNs and GCNs. The key innovation of this approach is the introduction of a mini-batch GCN that reduces computational costs while handling large-scale PolSAR datasets. By employing a local graph-based fusion method, the framework enhances classification accuracy by fusing the distinct features extracted by CNNs and GCNs, while reducing redundant information through local spatial neighborhood modeling. Shi et al. [98] introduced a novel approach that integrates a double-channel convolutional network (DCCNN) and an edge-preserving Markov Random Field (MRF) to enhance classification accuracy in PolSAR images. The key innovation lies in combining a Wishart-based complex matrix subnetwork with a multi-feature subnetwork to jointly capture statistical characteristics and high-level semantic features. Additionally, the MRF component specifically preserves edge details and reduces speckle noise, leading to superior classification performance in heterogeneous terrain. Yang et al. [99] proposed a novel method for PolSAR image classification

that integrates ResBlock and the convolutional block attention module (CBAM) to enhance the extraction of mid-level features and polarimetric information from PolSAR data. The key innovation of this approach lies in utilizing CBAM to adaptively focus on the most important spatial and channel features, improving classification accuracy while reducing computational complexity. Additionally, a feature selection method based on channel attention is introduced to minimize redundant information and enhance efficiency. Geng et al. [100] proposed a feature-enhanced superpixel hypergraph neural network (FESHNN) for PolSAR image classification, which innovatively combines both spatial and polarimetric correlations to improve classification performance. The key innovation lies in the use of a superpixel hypergraph neural network (SHNN) that fuses spatial features and polarimetric features through the construction of a hypergraph. This hypergraph aggregates information from neighboring vertices, allowing for a better representation of both local and global features. Additionally, a feature enhancement module is incorporated to refine pixel-level and superpixel-level features, further enhancing the model's discrimination ability.

Wang et al. [101] introduced a multiscale superpixel-guided weighted graph convolutional network (MSGWGCN) for PolSAR image classification, addressing the challenge of misclassification due to various object sizes and complex boundaries. The key innovation lies in the use of multiscale superpixel segmentation to capture both local and global spatial features, which are further processed by a weighted graph convolutional network that adapts to land cover objects of different sizes. Additionally, a multiscale feature cascade fusion module fuses pixel-level features at different scales, significantly enhancing classification accuracy and boundary preservation. Jamali et al. [102] proposed a local window attention Transformer (PolSARFormer) for PolSAR image classification, combining both CNNs and ViT to effectively classify PolSAR imagery. The key innovation of this method lies in utilizing a Local Window Attention mechanism, which restricts the receptive field of each query token to its local neighborhood, significantly reducing the computational cost associated with traditional self-attention models like Swin Transformer. Ren et al. [103] introduced a low-rank constrained multimodal tensor representation (LR-MTR) method for PolSAR scene classification. The key innovation lies in utilizing multiple target decomposition methods (e.g., Freeman, $H/A/\alpha$, Pauli) to form pseudo-color images, which are then integrated into a tensor representation. This multimodal fusion, constrained by low-rank tensor norms, captures cross-modal information while reducing redundancy. The LR-MTR method improves classification accuracy by preserving the intrinsic relationships between modalities. Liu et al. [104] introduced a self-trained deep forest (STDF) framework for urban impervious surface area extraction in arid areas using limited samples, leveraging both multispectral and PolSAR imagery. The primary innovation lies in the fusion of multispectral features from Sentinel-2 images and polarimetric features from GaoFen-3 PolSAR data, combined with a self-training mechanism that enhances classification accuracy despite the limited availability of labeled samples. The framework employs a deep forest model, which is particularly effective for small datasets due to its adaptive model complexity and ability to reduce overfitting. Hua et al. [105] proposed a feature fusion network for PolSAR image classification that integrates both physical and deep features to enhance classification accuracy, especially with limited labeled samples. The key innovation lies in the introduction of a learnable feature fusion module (LFFM) and an improved feature pyramid network (IFPN). The LFFM autonomously selects the most relevant features during the fusion process, while the IFPN captures multiscale feature representations by combining high- and low-level features. These innovations improve the interpretability and performance of the model by leveraging the complementary characteristics of physical and spectral features. Wang et al. [106] introduce a novel approach for fusing SAR and optical data to improve land cover classification. The key innovation lies in using a dual-stream network, where ResNet extracts deep features from optical images, and PidiNet captures edge features from SAR data. The framework also incorporates the iAFF fusion module to facilitate interaction between low- and high-level features from both modalities. Additionally, the ASPP module is used to enhance global feature dependency

by managing interactions between high-level features. This method significantly improves classification performance by combining complementary information from both SAR and optical imagery.

Fernandez-Beltran et al. [107] proposed a hierarchical multimodal probabilistic latent semantic analysis (HMpLSA) approach for remote sensing image fusion, specifically integrating SAR and multispectral imaging (MSI) data. The novelty of this framework lies in its two-level latent topic architecture. The first level generates individual semantic representations from SAR and MSI, while the second level fuses these representations into a common semantic space, uncovering inter-sensor feature patterns at a higher abstraction level. This dual-path fusion approach enables more effective unsupervised land cover classification by preserving the multimodal structure and enhancing semantic pattern discovery. Yang et al. [108] introduced a novel CNN-based polarimetric decomposition feature selection framework for PolSAR image classification, offering several key innovations in feature selection. A significant contribution is the use of a 1D CNN to efficiently select high-quality features from polarimetric target decomposition, which are essential for terrain classification. By employing the Kullback–Leibler distance (KLD), the framework optimizes feature selection by avoiding random selection and improving efficiency. Unlike traditional methods, this approach considers the performance of feature combinations rather than individual contributions. Cao et al. [109] introduced the DFAF-Net, a dual-frequency PolSAR image classification network that innovatively uses frequency-aware attention blocks (FABs) and an adaptive feature fusion block (AFFB). The key advancements of this method lie in its ability to leverage complementary information from dual-frequency PolSAR data. The FAB modules generate attention masks specific to each frequency, enhancing the representation of polarimetric and spatial features by focusing on critical frequency-related details. The AFFB module adaptively fuses these frequency-specific features, significantly improving the distinction between similar land cover classes. Chu et al. [110] proposed a novel DL approach for fusing PolSAR and optical images in land-cover semantic segmentation tasks. The key innovation lies in the spatial dense channel attention module (SDCAM), which captures local channel relationships at different spatial positions. This enables effective feature fusion between PolSAR and optical images by enhancing the local channel correlations, a critical factor for accurately distinguishing complex land-cover types. The dual-stream Siamese encoder allows for separate feature extraction from each modality, which is then fused using the SDCAM. This approach, coupled with a symmetric skip-connection decoder, leads to improved segmentation performance by leveraging the complementary strengths of both PolSAR and optical imagery.

4. Comparison of Different Few-Shot PolSAR Image Classification

4.1. Comparison of Performance on Benchmark Datasets

Table 3 evaluates and compares the performance of various classification techniques on a benchmark Polarimetric SAR (PolSAR) dataset, Flevoland II. It includes both patch-based (PB) methods, direct segmentation (DS) methods, and data augmentation (DA)-based methods. The approaches range from traditional classifiers like SVM and CNN [111] to advanced deep learning architectures such as U-Net [112], SETR [113], and PolSARMixer [49]. The results demonstrate the significant advantage of data augmentation in improving classification accuracy and robustness.

For U-Net and SETR, incorporating data augmentation (DA) improves their performance compared to the DS approach. U-Net-DA achieves a significant improvement in mIoU and mDice compared to U-Net-DS. SETR-DA benefits notably from data augmentation, with clear gains across all metrics. Traditional methods (e.g., SVM-PB and CNN-PB) have considerably lower performance metrics, especially for classes like maize and onions, showcasing their limitations in handling PolSAR data's complex feature space. Direct segmentation methods (U-Net-DS, SETR-DS) often face challenges in global inference and boundary resolution, as evidenced by their relatively lower metrics compared to DA-based methods.

Table 3. Classification results of different methods in the Flevoland II dataset.

Class	SVM-PB	CNN-PB [111]	U-Net-DS [112]	SETR-DS [113]	U-Net-DA [112]	SETR-DA [113]	PolSARMixer [49]
Potato	99.70	99.52	99.39	99.46	99.86	99.10	99.72
Fruit	99.98	100.0	99.04	92.63	96.99	99.58	99.49
Oats	88.89	100.0	97.67	88.94	95.06	99.54	97.52
Beet	97.62	94.25	98.28	99.48	99.50	97.94	99.51
Barley	99.58	97.61	99.63	99.41	99.76	98.73	99.68
Onions	31.93	90.47	86.13	67.43	85.92	97.69	97.26
Wheat	99.76	96.67	99.85	99.95	99.76	98.67	99.67
Beans	94.05	92.38	100.0	100.0	90.33	93.51	97.84
Peas	99.93	100.0	99.75	86.57	97.47	97.94	98.97
Maize	79.21	98.10	1.57	0.00	93.10	95.95	97.56
Flax	98.50	94.76	99.95	79.50	99.17	98.93	99.30
Rapeseed	99.31	99.53	99.80	99.90	99.84	98.38	99.73
Grass	93.88	97.71	98.15	94.86	98.03	97.75	98.71
Lucerne	93.91	96.67	97.76	94.70	97.72	99.12	98.71
OA	97.69	96.80	98.16	96.67	99.12	99.21	99.51
AA	91.17	96.98	91.21	85.92	96.61	98.75	98.83
mIoU	86.69	87.78	82.79	76.27	94.89	97.96	98.14
mDice	91.23	90.39	84.35	80.28	97.29	98.96	99.06
Kappa	0.9438	0.9674	0.9780	0.9603	0.9896	0.9906	0.9940

Among all methods, PolSARMixer achieves the best overall performance, with the highest overall accuracy (OA: 99.51%), average accuracy (AA: 98.83%), mean intersection over union (mIoU: 98.14%), mean Dice coefficient (mDice: 99.06%), and Kappa coefficient (0.9940). These results confirm the critical role of data augmentation in few-shot PolSAR classification tasks, effectively mitigating class imbalance and improving the generalization of classification models. The accuracy metrics in the table align with the referenced study's findings, making them suitable for inclusion in a comprehensive survey.

The comparative results of different methods for few-shot classification on the San Francisco dataset are presented in Table 4. Traditional methods, such as KNN and SVM, exhibit relatively low performance, with OA values of 60.98% and 62.09%, respectively, under the five-shot condition. In contrast, methods leveraging advanced SSL and transfer learning strategies achieve significantly higher accuracy. Among SSL-based methods, SimCLR [114] and RSS (ResNet-based SimSiam [115]) demonstrate improved performance, achieving OA values of 79.17% and 79.59%, respectively, for five-shot classification. However, methods integrating transfer learning, such as MI-SSL [70], PCLNet [73], and SSPRL [27], further enhance classification accuracy. Notably, the MLRSS-RC (ResNet-based MLR-SimSiam with CBAM) method achieves the best results, with an OA of 82.95% and a Kappa coefficient of 79.07% under the five-shot condition, and an OA of 88.73% and a Kappa coefficient of 84.64% under the ten-shot condition. These results underscore the effectiveness of combining transfer learning and self-supervised frameworks in addressing the challenges of few-shot PolSAR image classification.

To evaluate the effectiveness of multimodal approaches for PolSAR classification, we compared the performance of various methods on the Flevoland I dataset. Specifically, the method LI-PS-SF in [105] incorporates both physical and deep features, showing significant improvements in classification accuracy. The comparison is detailed in Table 5, which presents the OA and Kappa for each method. The dataset consists of 15 different terrain classes, with 10 labeled samples per class, totaling 150 labeled samples.

Table 4. Few-shot experimental results of different methods on the San Francisco dataset. The best performance for each backbone is highlighted in bold.

Method	5 Samples per Class		10 Samples per Class	
	OA	Kappa	OA	Kappa
KNN	60.98	51.23	65.70	57.12
SVM	62.09	52.62	69.30	61.62
CNN [111]	75.63	69.54	80.01	75.01
MLRSS-C [71] (CNN-based)	**76.47**	**72.72**	**82.27**	**78.23**
SimCLR [114]	79.17	73.96	82.22	77.77
RSS [115]	79.59	74.49	81.21	76.51
MLRSS-R [71] (ResNet-based)	**79.85**	**75.13**	**85.96**	**80.62**
MI-SSL [70]	80.54	73.64	87.61	83.21
PCLNet [73]	82.65	78.31	84.27	80.34
SSPRL [27]	82.10	77.59	88.51	83.09
MLRSS-RC [71]	**82.95**	**79.07**	**88.73**	**84.64**

Table 5. OA (%) and Kappa of the Flevoland I dataset with different methods.

Method	LI-PS-SF [105]	ProtoNet [116]	PFDD [117]	WT [118]	TMST [119]	PASGS [120]	SDBCS [31]	Super-RF [121]	MPCNN [122]
Training ratio	10 samples per class					1%		10%	
Total number	150					1677		16771	
Stem beans	97.90	99.76	80.97	99.64	96.40	93.77	93.48	96.77	99.79
Rapeseed	93.25	59.55	79.32	84.43	81.95	70.55	84.00	53.13	93.55
Bare soil	100	100	86.61	99.20	99.31	81.03	99.94	94.57	99.84
Potatoes	96.69	93.14	88.01	78.00	65.31	92.52	94.29	96.04	96.04
Beet	97.10	97.63	92.78	97.65	93.45	86.85	79.91	95.70	98.00
Wheat 2	86.05	85.57	65.10	98.76	72.48	81.72	92.03	79.93	95.97
Peas	100	98.62	88.82	97.57	92.29	94.66	85.62	98.64	99.00
Wheat 3	99.51	98.44	95.39	97.46	90.05	95.65	97.52	99.39	93.88
Lucerne	97.63	95.72	94.88	99.00	95.07	93.01	91.53	96.63	98.45
Barley	100	99.92	97.79	99.67	95.64	95.08	99.68	100	40.25
Wheat	88.14	94.40	80.09	89.58	87.09	93.82	96.88	99.05	95.52
Grasses	88.95	89.18	69.35	88.98	72.13	70.87	88.53	84.03	84.03
Forest	99.73	98.23	83.05	96.76	90.32	99.26	97.23	95.88	99.97
Water	100	97.88	97.63	93.13	96.30	90.56	83.14	100	91.27
Building	84.76	90.61	86.53	82.59	76.87	40.76	84.24	0	96.01
OA	96.01	92.93	86.30	92.50	87.01	89.73	91.87	92.18	91.87
Kappa	94.34	92.63	84.51	91.77	85.42	88.83	91.13	91.44	91.13

The results demonstrate that the LI-PS-SF significantly outperforms other SOTA methods. In particular, it achieves a 9.71% improvement in OA over PFDD and 3.51% over WT, both of which are notable DL methods. Moreover, compared with semi-supervised methods such as TMST, the proposed approach achieves a higher OA by 9.00%. Even when compared with approaches that utilize significantly more labeled data, such as PASGS (which uses 1% of training samples, totaling 1677 samples) and Super-RF (which uses 10% of training samples, totaling 16,771 samples), the proposed method still achieves higher OA by 6.28% and 3.83%, respectively.

Table 6 presents the classification results of different methods when the target-domain dataset is Germany. The evaluated models include CV-CNN [123], RCV-CNN [111], PCLNet [27], SSPRL [73], CLIN [124], MKL [81], and CCFSLC [44], along with its corresponding real-valued model, RCFSLC [44]. The key metrics used for comparison are OA, AA, recall, F1 score, and Kappa coefficient.

Table 6. Comparison of classification results of different methods when the target-domain dataset is Oberpfaffenhofen.

Category	CV-CNN [123]	RCV-CNN [111]	PCLNet [27]	SSPRL [73]	CLIN [124]	MKL [81]	RCFSLC [44]	CCFSLC [44]
Built-up area	52.83	47.99	75.15	76.03	63.98	61.51	56.76	70.62
Woodland	49.43	98.44	74.79	75.94	94.05	93.74	92.10	94.06
Open area	96.64	92.89	96.23	93.66	92.72	96.01	97.48	95.12
OA (%)	76.80	82.71	86.93	85.92	85.80	86.96	86.28	88.80
AA (%)	66.30	79.78	82.06	81.88	83.61	83.76	82.11	86.60
Recall (%)	67.80	79.56	83.40	82.94	81.65	84.06	82.68	86.11
F1 (%)	66.90	76.67	82.67	82.36	81.86	82.69	81.05	86.03
Kappa	0.5952	0.7098	0.7749	0.7585	0.7611	0.7771	0.7645	0.8087

As seen in Table 6, CCFSLC achieves the highest Overall Accuracy (OA) of 88.80%, outperforming other methods by a significant margin. It also attains the highest AA of 86.60% and the highest Kappa coefficient of 0.8087, indicating superior classification performance and reliability. In category-specific accuracy, CCFSLC shows strong performance across all categories, particularly in the "Woodland" and "Built-up area" categories.

4.2. Qualitative Comparison of Different Algorithms

Table 7 presents a comparative analysis of different methods used in PolSAR few-shot image classification, categorized into transfer learning, meta-learning, data augmentation, and multimodal learning paradigms. Each method has distinct advantages and disadvantages, which should be considered when selecting an appropriate strategy for few-shot PolSAR image classification.

Data augmentation helps address the scarcity of labeled PolSAR data by generating synthetic samples, enhancing model robustness. Traditional techniques, such as flipping and rotating, are simple and effective but fail to capture the complexity of PolSAR data. Advanced methods, like GANs, produce more realistic data at the cost of higher computational demands. However, despite increasing sample diversity, both traditional and advanced methods often struggle to generalize to new, unseen classes, particularly when the augmented data does not adequately represent the true distribution of PolSAR images.

Meta-learning is specifically designed for FSL scenarios, excelling at generalizing to new classes with limited labeled data. Metric-based approaches classify new data by measuring proximity to labeled examples, making them highly effective for such tasks. However, the complexity of meta-learning models poses significant challenges during implementation and training, particularly for PolSAR-specific features that require careful tuning.

Transfer learning mitigates overfitting by leveraging knowledge from a source domain, typically a large dataset, to improve classification performance in the target domain with minimal labeled data. This approach reduces the need for large datasets by adapting models trained on different tasks. However, it is susceptible to negative transfer, where the source knowledge does not align with PolSAR-specific characteristics, potentially harming performance.

Multimodal learning integrates information from multiple data sources, enhancing feature representation and improving classification accuracy, particularly when polarimetric data alone is insufficient. However, this approach requires sophisticated models and significant computational resources, making implementation and training more challenging.

In conclusion, the learning paradigm choice depends heavily on the specific characteristics of the PolSAR classification task, the availability of labeled data, and the computational

resources at hand. Future research should focus on hybrid methods that combine the strengths of these paradigms while addressing their respective limitations.

Table 7. Comparative analysis of different methods in each paradigm of few-shot image classification algorithms.

Category	Method	Advantages	Disadvantages
Transfer learning	Pre-trained	Easy to implement, good feature selection and transformation	Data distribution is often different
	Self-supervised	Alleviates overfitting	Prone to overfitting
	Domain adaption	Alleviates overfitting	The number of iterations should be fewer
Meta-learning	Optimization-based	Makes models learn new tasks quickly	Extensive calculations and high memory consumption
	Metric-based	Easy to calculate	Weak interpretability and high memory consumption
Data augmentation	Data generation-based	Increases sample numbers	Cannot completely solve overfitting
	Feature enhancement-based	Increases feature numbers	Easy to be disturbed by noise
Multimodal	Knowledge transfer-based	Learns better feature representation	Easy to be disturbed by noise during the fusion process
	Metric-based	Simple calculation and high accuracy	Weak interpretability and high memory consumption

5. Challenges and Future Trends in Few-Shot PolSAR Image Classification

Few-shot PolSAR image classification faces several challenges that hinder its scalability and performance. In this section, we identify key challenges and provide corresponding future trends to address them systematically.

1. Data scarcity and imbalance: One of the primary challenges in PolSAR image classification is the limited availability of labeled data, especially for rare terrain types, leading to biased models and poor generalization.

Future trend: To mitigate this, advanced semi-supervised and unsupervised learning methods should be developed. For instance, integrating self-supervised learning (SSL) with domain-specific physical priors can help extract meaningful features from unlabeled PolSAR datasets and improve classification accuracy across all terrain types.

2. Speckle noise and data variability: Speckle noise significantly reduces the spatial consistency of PolSAR images, complicating classification tasks. Additionally, domain shifts caused by sensor differences, environmental factors, or acquisition conditions further degrade model performance.

Future trend: Physics-informed deep learning models that incorporate noise models specific to PolSAR data can enhance noise robustness. Furthermore, domain-invariant feature learning and adversarial domain adaptation techniques should be employed to handle data variability and improve cross-domain generalization.

3. High dimensionality and computational complexity: PolSAR data contains high-dimensional information across multiple polarization channels, leading to computational inefficiencies and increased risks of overfitting.

Future trend: Lightweight architectures, such as efficient convolutional neural networks and Transformer-based models tailored for PolSAR data, should be developed. Dimensionality reduction techniques leveraging polarimetric decomposition can also reduce computational complexity while preserving critical information.

4. Lack of generalization across modalities: While multimodal approaches offer enhanced feature representations, integrating PolSAR data with other modalities such as

optical or hyperspectral data remains challenging due to differences in resolution, noise, and data formats.

Future trend: Cross-modal fusion techniques should align features across modalities while respecting their unique physical characteristics. Techniques like graph-based learning and contrastive learning are promising for ensuring coherent multimodal representations.

5. Interpretability and explainability: Deep learning models often operate as black boxes, limiting interpretability, especially in critical applications like disaster management or defense.

Future trend: The integration of explainable AI (XAI) techniques into PolSAR classification frameworks is essential. For instance, attention mechanisms and saliency maps can highlight important polarimetric features, improving both model interpretability and user trust.

6. Conclusions

In this survey, we explored the recent advancements in deep learning methods for few-shot PolSAR image classification. The inherent challenges of PolSAR data, such as limited labeled samples, domain shifts, and complex polarimetric features, make few-shot learning a promising solution for this task.

Data augmentation techniques, both traditional and advanced methods using GANs, provide a means to expand limited PolSAR datasets by generating synthetic samples. While traditional methods like flipping and rotating help improve generalization, GAN-based approaches allow for more sophisticated data generation that can enhance the model's ability to distinguish between fine-grained terrain types and handle the complexity of polarimetric data. Transfer learning, particularly through the use of pre-trained models like ResNet or ViTs, has been shown to improve classification performance by leveraging features learned from large external datasets. This technique helps mitigate the overfitting problem inherent in training deep networks with limited data and can be further refined through domain adaptation techniques that align the feature distributions between different PolSAR datasets. Meta-learning-based methods, especially metric-based approaches such as prototypical networks and RNs, have demonstrated strong performance in few-shot PolSAR tasks by learning transferable metrics that generalize across new classes with minimal labeled data. Optimization-based meta-learning methods like MAML also provide rapid adaptation to new tasks, though their performance may be hindered by the unique characteristics of PolSAR data. By integrating multiple modalities, such as optical and PolSAR data, multimodal-based methods offer improved feature representation, leading to enhanced classification accuracy. These approaches capture complementary information from different sources, though they come with increased computational complexity and training challenges.

In conclusion, the choice of learning paradigm for few-shot PolSAR image classification depends on task-specific requirements, data availability, and computational constraints. While transfer learning and data augmentation offer scalable solutions to the limited data problem, meta-learning provides specialized methods for few-shot tasks. Multimodal learning, though resource-intensive, can significantly enhance classification performance. Future research should focus on hybrid models that combine the strengths of these approaches while addressing their limitations, aiming for robust and scalable solutions to real-world PolSAR classification problems.

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

Funding: This work was supported by the NSFC under grant nos. 42201394 and 12326615, the National Key R&D Program of China under grant no. 2022YFA1003800, the Major Key Project of PCL under grant no. PCL2024A06, and the Qinchuangyuan High-level Innovation and Entrepreneurial Talent Program under grant no. QCYRCXM-2022-30.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Shi, C.; Zhang, X.; Sun, J.; Wang, L. Remote sensing scene image classification based on self-compensating convolution neural network. *Remote Sens.* **2022**, *14*, 545. [CrossRef]
2. Zou, Z.; Chen, K.; Shi, Z.; Guo, Y.; Ye, J. Object detection in 20 years: A survey. *Proc. IEEE* **2023**, *111*, 257–276. [CrossRef]
3. Dong, H.; Zou, B.; Zhang, L.; Zhang, S. Automatic design of CNNs via differentiable neural architecture search for PolSAR image classification. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 6362–6375. [CrossRef]
4. Bi, H.; Xu, F.; Wei, Z.; Han, Y.; Cui, Y.; Xue, Y.; Xu, Z. Unsupervised PolSAR image factorization with deep convolutional networks. In Proceedings of the IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; IEEE: New York, NY, USA, 2019; pp. 1061–1064.
5. Liu, J.; Zhang, T.; Xiong, H. PolSAR Ship Targets Generation via the Polarimetric Feature Guided Denoising Diffusion Probabilistic Model. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 1–5. [CrossRef]
6. Wan, H.; Tang, P.; Tian, B.; Yu, H.; Jin, C.; Zhao, B.; Wang, H. Water Extraction in PolSAR Image Based on Superpixel and Graph Convolutional Network. *Appl. Sci.* **2023**, *13*, 2610. [CrossRef]
7. Zeng, X.; Wang, Z.; Wang, Y.; Rong, X.; Guo, P.; Gao, X.; Sun, X. SemiPSCN: Polarization Semantic Constraint Network for Semi-Supervised Segmentation in Large-Scale and Complex-Valued PolSAR Images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *62*, 1–18. [CrossRef]
8. Guan, X.; Li, F.; Bi, H.; Gong, L. Rank Learning Based Full-Resolution Quality Evaluation Method for Pansharpened Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 833–846. [CrossRef]
9. Hua, W.; Wang, Y.; Yang, S.; Jin, X. PolSAR Image Classification Based on Multi-Modal Contrastive Fully Convolutional Network. *Remote Sens.* **2024**, *16*, 296. [CrossRef]
10. Parikh, H.; Patel, S.; Patel, V. Modeling PolSAR classification using convolutional neural network with homogeneity based kernel selection. *Model. Earth Syst. Environ.* **2023**, *9*, 3801–3813. [CrossRef]
11. Jin, K.; Chen, Y.; Xu, B.; Yin, J.; Wang, X.; Yang, J. A patch-to-pixel convolutional neural network for small ship detection with PolSAR images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 6623–6638. [CrossRef]
12. Gao, G.; Bai, Q.; Zhang, C.; Zhang, L.; Yao, L. Dualistic cascade convolutional neural network dedicated to fully PolSAR image ship detection. *Isprs J. Photogramm. Remote Sens.* **2023**, *202*, 663–681. [CrossRef]
13. Liu, K.; Kuang, Z.; Jing, S.; Bi, H.; Li, Y.; Xu, C. Diffusion-Based Generative Self-Supervised Model for Few-Shot PolSAR Image Classification. In Proceedings of the IGARSS 2024–2024 IEEE International Geoscience and Remote Sensing Symposium, Athens, Greece, 7–12 July 2024; IEEE: New York, NY, USA, 2024; pp. 11127–11131.
14. Kuang, Z.; Bi, H.; Li, F.; Xu, C.; Sun, J. Polarimetry-inspired Contrastive Learning for Class-imbalanced PolSAR Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–19. [CrossRef]
15. Lee, J.S.; Jurkevich, L.; Dewaele, P.; Wambacq, P.; Oosterlinck, A. Speckle filtering of synthetic aperture radar images: A review. *Remote Sens. Rev.* **1994**, *8*, 313–340. [CrossRef]
16. Cumming, I.G.; Wong, F.H. Digital processing of synthetic aperture radar data. *Artech House* **2005**, *1*, 108–110.
17. Freeman, A.; Durden, S.L. A three-component scattering model for polarimetric SAR data. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 963–973. [CrossRef]
18. Cloude, S.R.; Pottier, E. An entropy based classification scheme for land applications of polarimetric SAR. *IEEE Trans. Geosci. Remote Sens.* **1997**, *35*, 68–78. [CrossRef]
19. Jolliffe, I.T. *Principal Component Analysis for Special Types of Data*; Springer: Berlin/Heidelberg, Germany, 2002.
20. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
21. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
22. Cortes, C. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
23. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]
24. López-Martínez, C.; Fabregas, X. Polarimetric SAR speckle noise model. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 2232–2242. [CrossRef]
25. Ding, L.; Zheng, K.; Lin, D.; Chen, Y.; Liu, B.; Li, J.; Bruzzone, L. MP-ResNet: Multipath residual network for the semantic segmentation of high-resolution PolSAR images. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [CrossRef]

26. Karachristos, K.; Koukiou, G.; Anastassopoulos, V. Fusion of Coherent and Non-Coherent Pol-SAR Features for Land Cover Classification. *Electronics* **2024**, *13*, 634. [CrossRef]
27. Zhang, L.; Zhang, S.; Zou, B.; Dong, H. Unsupervised deep representation learning and few-shot classification of PolSAR images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *60*, 1–16. [CrossRef]
28. Snell, J.; Swersky, K.; Zemel, R. Prototypical networks for few-shot learning. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4077–4087.
29. Sung, F.; Yang, Y.; Zhang, L.; Xiang, T.; Torr, P.H.; Hospedales, T.M. Learning to compare: Relation network for few-shot learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Salt Lake City, UT, USA, 18–23 June, 2018; pp. 1199–1208.
30. Hospedales, T.; Antoniou, A.; Micaelli, P.; Storkey, A. Meta-learning in neural networks: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 5149–5169. [CrossRef]
31. Wang, N.; Bi, H.; Li, F.; Xu, C.; Gao, J. Self-distillation-based polarimetric image classification with noisy and sparse labels. *Remote Sens.* **2023**, *15*, 5751. [CrossRef]
32. Gui, R.; Xu, X.; Wang, L.; Yang, R.; Pu, F. A generalized zero-shot learning framework for PolSAR land cover classification. *Remote Sens.* **2018**, *10*, 1307. [CrossRef]
33. Bi, H.; Chang, H.; Wang, X.; Hong, D. Cross-Attention-Driven Adaptive Graph Relational Network for Multi-Label Remote Sensing Scene Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–14. [CrossRef]
34. Malliot, H.A. Digital terrain elevation mapping system. In Proceedings of the 1996 IEEE Aerospace Applications Conference. Proceedings, Aspen, Colorado, 3–9 February 1996; IEEE: New York, NY, USA, 1996; Volume 4, pp. 91–105.
35. Hossin, M.; Sulaiman, M.N. A review on evaluation metrics for data classification evaluations. *Int. J. Data Min. Knowl. Manag. Process* **2015**, *5*, 1.
36. Cui, Y.; Jia, M.; Lin, T.Y.; Song, Y.; Belongie, S. Class-balanced loss based on effective number of samples. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June, 2019; pp. 9268–9277.
37. Larrañaga, A.; Álvarez-Mozos, J. On the added value of quad-pol data in a multi-temporal crop classification framework based on RADARSAT-2 imagery. *Remote Sens.* **2016**, *8*, 335. [CrossRef]
38. Cohen, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. [CrossRef]
39. Ashwin, P.; Ansal, K. Pol-SAR image classification using multifarious stratification stratagem in machine learning. *J. Intell. Fuzzy Syst.* **2024**, *37*, 1–20. [CrossRef]
40. Ioannidou, M.; Koukos, A.; Sitokonstantinou, V.; Papoutsis, I.; Kontoes, C. Assessing the added value of Sentinel-1 PolSAR data for crop classification. *Remote Sens.* **2022**, *14*, 5739. [CrossRef]
41. Yuan, J.; Lv, X.; Dou, F.; Yao, J. Change analysis in urban areas based on statistical features and temporal clustering using TerraSAR-X time-series images. *Remote Sens.* **2019**, *11*, 926. [CrossRef]
42. Powers, D.M. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv* **2020**, arXiv:2010.16061.
43. Parikh, H.; Patel, S.; Patel, V. Classification of SAR and PolSAR images using deep learning: A review. *Int. J. Image Data Fusion* **2020**, *11*, 1–32. [CrossRef]
44. Cao, Y.; Wu, Z.; Chen, J.; Huang, Z.; Yang, L. Cross-Domain PolSAR Image Classification Using Complex-Valued Few-Shot Learning Network. *IEEE Trans. Aerosp. Electron. Syst.* **2024**, 1–16. [CrossRef]
45. Vinyals, O.; Blundell, C.; Lillicrap, T.; Wierstra, D. Matching networks for one shot learning. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 3637–3645.
46. Lu, Y.; Yang, C.; Meng, Z. Lithology discrimination using Sentinel-1 dual-pol data and SRTM data. *Remote Sens.* **2021**, *13*, 1280. [CrossRef]
47. Hanley, J.A.; McNeil, B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **1982**, *143*, 29–36. [CrossRef] [PubMed]
48. Zhang, L.; Dong, H.; Zou, B. Efficiently utilizing complex-valued PolSAR image data via a multi-task deep learning framework. *Isprs J. Photogramm. Remote Sens.* **2019**, *157*, 59–72. [CrossRef]
49. Wang, Z.; Wang, Z.; Qiu, X.; Zhang, Z. Global Polarimetric Synthetic Aperture Radar Image Segmentation with Data Augmentation and Hybrid Architecture Model. *Remote Sens.* **2024**, *16*, 380. [CrossRef]
50. Ghamisi, P.; Plaza, J.; Chen, Y.; Li, J.; Plaza, A.J. Advanced spectral classifiers for hyperspectral images: A review. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–32. [CrossRef]
51. Liu, F.; Jiao, L.; Tang, X. Task-oriented GAN for PolSAR image classification and clustering. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 2707–2719. [CrossRef] [PubMed]
52. Yang, C.; Hou, B.; Chanussot, J.; Hu, Y.; Ren, B.; Wang, S.; Jiao, L. N-Cluster loss and hard sample generative deep metric learning for PolSAR image classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–16. [CrossRef]
53. Xie, W.; Yang, X.; Wang, R.; Zhao, F. Polsar Image Classification Via Auxiliary Classifier Generative Adversarial Network. In Proceedings of the IGARSS 2022–2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; IEEE: New York, NY, USA, 2022; pp. 1205–1208.
54. Dong, H.; Si, L.; Qiang, W.; Miao, W.; Zheng, C.; Wu, Y.; Zhang, L. A polarimetric scattering characteristics-guided adversarial learning approach for unsupervised PolSAR image classification. *Remote Sens.* **2023**, *15*, 1782. [CrossRef]

55. Song, Q.; Xu, F. Zero-shot learning of SAR target feature space with deep generative neural networks. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2245–2249. [CrossRef]
56. Wang, H.; Xing, C.; Yin, J.; Yang, J. Land cover classification for polarimetric SAR images based on vision transformer. *Remote Sens.* **2022**, *14*, 4656. [CrossRef]
57. Cattoi, A.; Bruzzone, L.; Haensch, R. Transcoding-based pre-training of semantic segmentation networks for PolSAR images. In Proceedings of the EUSAR 2022; 14th European Conference on Synthetic Aperture Radar, VDE, Leipzig, Germany, 25–27 July 2022; pp. 1–5.
58. Han, P.; Men, L.; Chen, Z. PolSAR Image Classification Based on Deep Features. In Proceedings of the China High Resolution Earth Observation Conference, Changsha, China, 17–19 November 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 392–403.
59. Xie, W.; Li, T.; Sun, H. Channel adaptive CVFCN using a new transfer method for PolSAR terrain classification. *Int. J. Remote Sens.* **2024**, *45*, 6521–6543. [CrossRef]
60. Gui, R.; Xu, X.; Yang, R.; Deng, K.; Hu, J. Generalized zero-shot domain adaptation for unsupervised cross-domain PolSAR image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *15*, 270–283. [CrossRef]
61. Gui, R.; Xu, X.; Yang, R.; Xu, Z.; Wang, L.; Pu, F. A general feature paradigm for unsupervised cross-domain PolSAR image classification. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [CrossRef]
62. Li, L.; Zeng, J.; Jiao, L.; Liang, P.; Liu, F.; Yang, S. Online active extreme learning machine with discrepancy sampling for PolSAR classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 2027–2041. [CrossRef]
63. Dong, H.; Si, L.; Qiang, W.; Zhang, L.; Yu, J.; Wu, Y.; Zheng, C.; Sun, F. A Novel Causal Inference Guided Feature Enhancement Framework for PolSAR Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *62*, 1–16. [CrossRef]
64. Qin, X.; Yang, J.; Li, P.; Sun, W.; Liu, W. A novel relational-based transductive transfer learning method for PolSAR images via time-series clustering. *Remote Sens.* **2019**, *11*, 1358. [CrossRef]
65. Gui, R.; Xu, X.; Wang, L.; Yang, R.; Pu, F. Eigenvalue statistical components-based PU-learning for PolSAR built-up areas extraction and cross-domain analysis. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 3192–3203. [CrossRef]
66. Hua, W.; Liu, L.; Sun, N.; Jin, X. A CA_Based Weighted Clustering Adversarial Network for Unsupervised Domain Adaptation PolSAR Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 1–5.
67. Cao, Y.; Zhu, D.; Wu, Z.; Chen, J.; Huang, Z.; Yang, L. Complex-valued cross-domain few-shot learning network for PolSAR image classification. In *IET Conference Proceedings CP874*; The Institution of Engineering and Technology: Stevenage, UK, 2023; Volume 2023, pp. 1726–1730.
68. Sun, W.; Li, P.; Yang, J.; Shi, L.; Zhao, L. Domain adaptation for polsar land classification using linear discriminative Laplacian Eigenmaps. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Worth, TX, USA, 23–28 July 2017; IEEE: New York, NY, USA, 2017; pp. 3278–3281.
69. Sun, W.; Li, P.; Du, B.; Yang, J.; Tian, L.; Li, M.; Zhao, L. Scatter matrix based domain adaptation for bi-temporal polarimetric SAR images. *Remote Sens.* **2020**, *12*, 658. [CrossRef]
70. Ren, B.; Zhao, Y.; Hou, B.; Chanussot, J.; Jiao, L. A mutual information-based self-supervised learning model for PolSAR land cover classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 9224–9237. [CrossRef]
71. Zhang, L.; Han, F.; Li, T.; Dai, H.; Zou, B.; Dong, H. MLR-SimSiam: A Contrastive Pre-training Model based on Polarimetric Jittering and Mutual Learning Regularizer for PolSAR Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 1–5. [CrossRef]
72. Kuang, Z.; Bi, H.; Li, F. Complex-valued self-supervised PolSAR image classification integrating attention mechanism. In Proceedings of the IGARSS 2023–2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 16–21 July 2023; IEEE: New York, NY, USA, 2023; pp. 5958–5961.
73. Zhang, W.; Pan, Z.; Hu, Y. Exploring PolSAR images representation via self-supervised learning and its application on few-shot classification. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
74. Qiu, W.; Pan, Z.; Yang, J. Few-shot polsar ship detection based on polarimetric features selection and improved contrastive self-supervised learning. *Remote Sens.* **2023**, *15*, 1874. [CrossRef]
75. Wang, Y.; Guo, Z.; Bi, H.; Hong, D.; Xu, C. Dual-Branch PolSAR Image Classification Based On Graphmae And Local Feature Extraction. In Proceedings of the IGARSS 2024–2024 IEEE International Geoscience and Remote Sensing Symposium, Athens, Greece, 7–12 July 2024; IEEE: New York, NY, USA, 2024; pp. 8866–8870.
76. Dong, H.; Zhang, L.; Zou, B. Exploring vision transformers for polarimetric SAR image classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–15. [CrossRef]
77. Cai, J.; Ma, Y.; Feng, Z.; Yang, S. Heterogeneous Network Based Contrastive Learning Method for PolSAR Land Cover Classification. *arXiv* **2024**, arXiv:2403.19902. [CrossRef]
78. Darvishnezhad, M.; Sebt, M.A. A novel self-supervised ensemble learning framework for land use and land cover classification of polarimetric synthetic aperture radar images. *IET Radar Sonar Navig.* **2024**, *18*, 379–409. [CrossRef]
79. Ni, J.; Jia, Y.; Yin, Q.; Zhou, Y.; Zhang, F. Metric learning based fine-grained classification for PolSAR imagery. In Proceedings of the IGARSS 2020–2020 IEEE International Geoscience and Remote Sensing Symposium, Virtual, 26 September 26–2 October 2020; IEEE: New York, NY, USA, 2020; pp. 716–719.
80. Yang, S.; Li, R.; Li, Z.; Meng, H.; Feng, Z.; He, G. Meta-Graph Representative Learning for PolSAR Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–12.

81. Zhang, P.; Liu, C.; Chang, X.; Li, Y.; Li, M. Metric-based meta-learning model for few-shot PolSAR image terrain classification. In Proceedings of the 2021 CIE International Conference on Radar (Radar), Haikou, China, 15–19 December 2021; IEEE: New York, NY, USA, 2021; pp. 2529–2533.
82. Shang, R.; Wang, J.; Jiao, L.; Yang, X.; Li, Y. Spatial feature-based convolutional neural network for PolSAR image classification. *Appl. Soft Comput.* **2022**, *123*, 108922. [CrossRef]
83. Dong, H.; Xu, X.; Gui, R.; Song, C.; Sui, H. Metric learning based collapsed building extraction from post-earthquake PolSAR imagery. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; IEEE: New York, NY, USA, 2016; pp. 4742–4745.
84. Zhang, T.; Long, J.; Lin, H.; Liu, Z.; Ye, Z.; Zheng, H. A Novel Feature Evaluation Method in Mapping Forest AGB by Fusing Multiple Evaluation Metrics Using PolSAR Data. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 1–5. [CrossRef]
85. Hua, W.; Zhang, Y.; Zhang, C.; Jin, X. PolSAR image classification based on relation network with SWANet. *Remote Sens.* **2023**, *15*, 2025. [CrossRef]
86. Huang, K.; Nie, W.; Luo, N. Fully polarized SAR imagery classification based on deep reinforcement learning method using multiple polarimetric features. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 3719–3730. [CrossRef]
87. Liu, F.; Wang, J.; Tang, X.; Liu, J.; Zhang, X.; Xiao, L. Adaptive graph convolutional network for PolSAR image classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–14. [CrossRef]
88. Baltrušaitis, T.; Ahuja, C.; Morency, L.P. Multimodal machine learning: A survey and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *41*, 423–443. [CrossRef] [PubMed]
89. Quan, Y.; Zhang, R.; Li, J.; Ji, S.; Guo, H.; Yu, A. Learning SAR-Optical Cross Modal Features for Land Cover Classification. *Remote Sens.* **2024**, *16*, 431. [CrossRef]
90. Dong, Y.; Hänsch, R. Multimodal self-supervised learning for semantic analysis of PolSAR imagery. In Proceedings of the IGARSS 2023–2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 16–21 July 2023; IEEE: New York, NY, USA, 2023; pp. 1704–1707.
91. Li, F.; Zhang, C.; Zhang, X.; Li, Y. MF-DCMANet: A multi-feature dual-stage cross manifold attention network for PolSAR target recognition. *Remote Sens.* **2023**, *15*, 2292. [CrossRef]
92. Jin, Y.; Deligiannis, A.; Fuentes-Michel, J.C.; Vossiek, M. Cross-modal supervision-based multitask learning with automotive radar raw data. *IEEE Trans. Intell. Veh.* **2023**, *8*, 3012–3025. [CrossRef]
93. Hong, D.; Yokoya, N.; Xia, G.S.; Chanussot, J.; Zhu, X.X. X-ModalNet: A semi-supervised deep cross-modal network for classification of remote sensing data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *167*, 12–23. [CrossRef]
94. Wang, W.; Liu, F.; Liao, W.; Xiao, L. Cross-modal graph knowledge representation and distillation learning for land cover classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–18. [CrossRef]
95. Chang, H.; Bi, H.; Li, F.; Xu, C.; Chanussot, J.; Hong, D. Deep Symmetric Fusion Transformer for Multimodal Remote Sensing Data Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–15. [CrossRef]
96. Ren, B.; Hou, B.; Wen, Z.; Xie, W.; Jiao, L. PolSAR image classification via multimodal sparse representation-based feature fusion. *Int. J. Remote Sens.* **2018**, *39*, 7861–7880. [CrossRef]
97. Sebt, M.A.; Darvishnezhad, M. Feature fusion method based on local binary graph for PolSAR image classification. *IET Radar Sonar Navig.* **2023**, *17*, 939–954. [CrossRef]
98. Shi, J.; Nie, M.; Ji, S.; Shi, C.; Liu, H.; Jin, H. Polarimetric SAR Image Classification Based on Double-Channel Convolution Network and Edge-preserving Markov Random Field. *Remote Sens.* 2023, *preprints*.
99. Yang, Z.; Zhang, Q.; Chen, W.; Chen, C. PolSAR Image Classification Based on Resblock Combined with Attention Model. In Proceedings of the 2021 IEEE 6th International Conference on Signal and Image Processing (ICSIP), Nanjing, China, 22–24 October 2021; IEEE: New York, NY, USA, 2021; pp. 340–344.
100. Geng, J.; Wang, R.; Jiang, W. Polarimetric SAR image classification based on feature enhanced superpixel hypergraph neural network. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–12. [CrossRef]
101. Wang, R.; Nie, Y.; Geng, J. Multiscale superpixel-guided weighted graph convolutional network for polarimetric SAR image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 3727–3741. [CrossRef]
102. Jamali, A.; Roy, S.K.; Bhattacharya, A.; Ghamisi, P. Local window attention transformer for polarimetric SAR image classification. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 1–5. [CrossRef]
103. Ren, B.; Chen, M.; Hou, B.; Hong, D.; Ma, S.; Chanussot, J.; Jiao, L. PolSAR Scene Classification via Low-Rank Constrained Multimodal Tensor Representation. *Remote Sens.* **2022**, *14*, 3117. [CrossRef]
104. Liu, X.; Samat, A.; Li, E.; Wang, W.; Abuduwaili, J. Self-Trained Deep Forest with Limited Samples for Urban Impervious Surface Area Extraction in Arid Area Using Multispectral and PolSAR Imageries. *Sensors* **2022**, *22*, 6844. [CrossRef] [PubMed]
105. Hua, W.; Hou, Q.; Jin, X.; Liu, L.; Sun, N.; Meng, Z. A Feature Fusion Network for PolSAR Image Classification Based on Physical Features and Deep Features. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 1–5. [CrossRef]
106. Wang, Y.; Zhang, W.; Chen, W.; Chen, C.; Liang, Z. MFFnet: Multimodal Feature Fusion Network for Synthetic Aperture Radar and Optical Image Land Cover Classification. *Remote Sens.* **2024**, *16*, 2459. [CrossRef]
107. Fernandez-Beltran, R.; Haut, J.M.; Paoletti, M.E.; Plaza, J.; Plaza, A.; Pla, F. Remote sensing image fusion using hierarchical multimodal probabilistic latent semantic analysis. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 4982–4993. [CrossRef]

108. Yang, C.; Hou, B.; Ren, B.; Hu, Y.; Jiao, L. CNN-based polarimetric decomposition feature selection for PolSAR image classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 8796–8812. [CrossRef]
109. Cao, Y.; Wu, Y.; Li, M.; Liang, W.; Hu, X. DFAF-Net: A dual-frequency PolSAR image classification network based on frequency-aware attention and adaptive feature fusion. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–18. [CrossRef]
110. Chu, B.; Chen, J.; Chen, J.; Pei, X.; Yang, W.; Gao, F.; Wang, S. SDCAFNet: A deep convolutional neural network for land-cover semantic segmentation with the fusion of PolSAR and optical images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 8928–8942. [CrossRef]
111. Xie, W.; Ma, G.; Zhao, F.; Liu, H.; Zhang, L. PolSAR image classification via a novel semi-supervised recurrent complex-valued convolution neural network. *Neurocomputing* **2020**, *388*, 255–268. [CrossRef]
112. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: Proceedings of the 18th International Conference, Munich, Germany, 5–9 October 2015*; Proceedings, Part III 18; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
113. Zheng, S.; Lu, J.; Zhao, H.; Zhu, X.; Luo, Z.; Wang, Y.; Fu, Y.; Feng, J.; Xiang, T.; Torr, P.H.; et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 6881–6890.
114. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A simple framework for contrastive learning of visual representations. In Proceedings of the International conference on machine learning. PMLR, Virtual , 13–18 July 2020; pp. 1597–1607.
115. Chen, X.; He, K. Exploring simple siamese representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 15750–15758.
116. Chen, D.; Chen, Y.; Li, Y.; Mao, F.; He, Y.; Xue, H. Self-supervised learning for few-shot image classification. In Proceedings of the ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Virtual, 6–12 June 2021; IEEE: New York, NY, USA, 2021; pp. 1745–1749.
117. Chen, S.W.; Tao, C.S. PolSAR image classification using polarimetric-feature-driven deep convolutional neural network. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 627–631. [CrossRef]
118. Jamali, A.; Mahdianpari, M.; Mohammadimanesh, F.; Bhattacharya, A.; Homayouni, S. PolSAR image classification based on deep convolutional neural networks using wavelet transformation. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
119. Wang, S.; Guo, Y.; Hua, W.; Liu, X.; Song, G.; Hou, B.; Jiao, L. Semi-supervised PolSAR image classification based on improved tri-training with a minimum spanning tree. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 8583–8597. [CrossRef]
120. Ni, J.; Xiang, D.; Lin, Z.; López-Martínez, C.; Hu, W.; Zhang, F. DNN-based PolSAR image classification on noisy labels. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 3697–3713. [CrossRef]
121. Chen, Q.; Cao, W.; Shang, J.; Liu, J.; Liu, X. Superpixel-based cropland classification of SAR image with statistical texture and polarization features. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [CrossRef]
122. Cui, Y.; Liu, F.; Jiao, L.; Guo, Y.; Liang, X.; Li, L.; Yang, S.; Qian, X. Polarimetric multipath convolutional neural network for PolSAR image classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–18. [CrossRef]
123. Zhang, Z.; Wang, H.; Xu, F.; Jin, Y.Q. Complex-valued convolutional neural network and its application in polarimetric SAR image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 7177–7188. [CrossRef]
124. Xin, X.; Li, M.; Wu, Y.; Zheng, M.; Zhang, P.; Xu, D.; Wang, J. Semi-supervised classification of dual-frequency polsar image using joint feature learning and cross label-information network. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–16. [CrossRef]