Article

Self-Distillation-Based Polarimetric Image Classification with Noisy and Sparse Labels

Ningwei Wang [1] , Haixia Bi [1,*] , Fan Li [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.); lifan@mail.xjtu.edu.cn (F.L.); jhgao@mail.xjtu.edu.cn (J.G.)
2 School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China; xnolimited@hotmail.com
3 Department of Mathematics and Fundamental Research, Peng Cheng Laboratory, Shenzhen 518055, China
* Correspondence: haixia.bi@xjtu.edu.cn

Abstract: Polarimetric synthetic aperture radar (PolSAR) image classification, a field crucial in remote sensing, faces significant challenges due to the intricate expertise required for accurate annotation, leading to susceptibility to labeling inaccuracies. Compounding this challenge are the constraints posed by limited labeled samples and the perennial issue of class imbalance inherent in PolSAR image classification. Our research objectives are to address these challenges by developing a novel label correction mechanism, implementing self-distillation-based contrastive learning, and introducing a sample rebalancing loss function. To address the quandary of noisy labels, we proffer a novel label correction mechanism that capitalizes on inherent sample similarities to rectify erroneously labeled instances. In parallel, to mitigate the limitation of sparsely labeled data, this study delves into self-distillation-based contrastive learning, harnessing sample affinities for nuanced feature extraction. Moreover, we introduce a sample rebalancing loss function that adjusts class weights and augments data for small classes. Through extensive experiments on four benchmark PolSAR images, our approach demonstrates its effectiveness in addressing label inaccuracies, limited samples, and class imbalance. Through extensive experiments on four benchmark PolSAR images, our research substantiates the robustness of our proposed methodology, particularly in rectifying label discrepancies in contexts marked by sample paucity and imbalance. The empirical findings illuminate the superior efficacy of our approach, positioning it at the forefront of state-of-the-art PolSAR classification techniques.

Keywords: label correction; self-distillation contrastive learning; sample rebalancing; polarimetric synthetic aperture radar (PolSAR) image classification

1. Introduction

Polarimetric synthetic aperture radar (PolSAR) is an advanced and important remote sensing technique owing to its distinctive ability to transmit and receive electromagnetic waves across various polarization modes [1]. This unique capability enables PolSAR to provide richer information on the scattering properties of Earth's surface. Consequently, PolSAR image classification, which is oriented towards categorizing image pixels into corresponding terrain classes, becomes instrumental for a spectrum of applications ranging from sea monitoring and agriculture to geological mapping and strategic governmental decisions [2]. PolSAR image classification has evolved, leading to diverse methodologies categorized into three main types: (1) physical-scattering-mechanism-based methods [2–4], (2) statistics-based methods [5,6], and (3) machine-learning-based methods [7–10]. Deep learning, with its superior feature representation, has significantly advanced PolSAR image classification [11].

However, PolSAR classification faces challenges, particularly noisy and sparse labels. These distortions misguide the model to assimilate noise patterns instead of the authentic

features. Limited annotations further challenge model accuracy and generalization. This paper seeks to unravel the following conundrum: How to improve the accuracy and robustness of the DNN-based PolSAR image classification method in a weak label scenario, i.e., with noisy and sparse labels?

Addressing noisy labels has engendered the inception of two predominant methodologies [12–14]. The first strategy focuses on the identification and purgation of these erroneous labels prior to model training [15,16]. This rectification can be accomplished through manual scrutiny, clustering, or the deployment of outlier detection algorithms. The alternative approach pivots towards the direct training of noise-robust models on corrupted datasets [17,18]. This necessitates the modification of the conventional loss function, accounting for the noisy labels. Ensemble learning, epitomized by methodologies like bootstrapping [19], self-training [20], and co-teaching [21], emerges as a robust tool. Such strategies harness the predictive prowess of an array of models, thereby refining overarching performance.

The field of image processing has traditionally seen a surge of research focusing on mitigating the challenges posed by noisy labels. In the specific domain of PolSAR image classification, the investigation into noisy labels remains comparatively nascent. Ni et al. [22] pioneered an insightful difference distribution diagram, articulating the intrinsic probability of a training sample being untainted. This probabilistic assessment paved the way for distinguishing clean labels from their noisy counterparts. Further innovation was heralded by Hou et al. [23] through their generative classification framework, adeptly tackling both the predicaments of unfaithful limited labels and the perturbations introduced by outliers in PolSAR pixels. Nevertheless, contemporary algorithms harbor intrinsic limitations. In contexts enriched with labels, eliminating detected noisy labels might not inflict significant harm. Yet, in scenarios marked by label paucity, such removal intensifies the small-sample dilemma, leading to potential algorithmic performance deterioration. Furthermore, an evident lacuna remains, as these methodologies overlook the potential leverage that can be garnered from the inherent similarity between training samples, which is quintessential for labeling.

To address these challenges, our research introduces a relabeling mechanism. This endeavor is grounded in the pivotal assertion that the discriminative model features extracted from neighboring samples with the same label play a vital role in driving the relabeling mechanism's efficacy.

Parallelly, the PolSAR image classification domain grapples with the issue of label scarcity. With the progress of deep learning, many PolSAR image classification methods [24–28] have been proposed to alleviate this problem. Semisupervised learning [29–31] ambitiously seeks to optimize classifier generalization, leveraging both labeled and unlabeled data. Active learning [11,32], in its quest, adopts a selective approach to acquire salient samples for labeling, aiming for maximized learning efficiency. Transfer learning [33,34], drawing from affluent source domains, endeavors to uplift the performance in target domains characterized by data scarcity. Reinforcement learning [35,36], albeit less prevalent in PolSAR terrains, adopts a unique perspective, emphasizing sequential decision making and reward maximization.

Venturing into a distinct trajectory, self-supervised learning [20,24,25] exploits the data's inherent properties to formulate alternative guidance signals, often involving pretext tasks for model training. This paradigm notably circumvents the label reliance in semisupervised learning, human intervention in active learning, domain-specific insights in transfer learning, and environmental interactions in reinforcement learning. However, self-supervised learning's capability to harness the intrinsic label information positions it advantageously, enabling nuanced feature extraction. Such prowess is manifested through its "pseudolabel" generation, correlating closely with true labels, and thus fostering meaningful data interpretations without extensive manual annotations [24].

Contrastive learning, as an important branch of self-supervised learning, while achieving commendable success in natural image classifications, remains scarcely explored within

the domain of PolSAR images. TCSPANet, as delineated by [37], integrates a dual-stage methodology: Initially, TCNet, rooted in contrastive learning, facilitates unsupervised representation learning. Subsequently, a subpatch attention encoder (SPAE), structured upon the transformer paradigm, models the context within patch samples. In a distinct approach, Zhang et al. [26] introduced the PolSAR-specific contrastive learning network (PCLNet). This network employs an unsupervised pretraining phase, anchored on instance discrimination [38], to harness valuable representations from unlabeled PolSAR data. Further, the self-supervised PolSAR representation learning (SSPRL) method [25] draws inspiration from the accomplishments of BYOL [19]. It is pertinent to note the following differences: TCSPANet operates through a bifurcated framework encompassing TCNet and SPAE, PCLNet capitalizes on an instance-discrimination-based pretraining phase, and SSPRL deploys a twin network structure alongside positive pairs, aiming for optimal efficiency across varied domains.

DINO [39] distinguishes itself by leveraging an exponential moving average (EMA) and central updates to fortify knowledge distillation. Unlike SSPRL, DINO uses EMA to seamlessly integrate the parameters of the online network with its target counterpart, an innovation that curtails parameter oscillation, thereby augmenting model stability. Within the DINO architecture, the teacher model's output serves to refine a center vector, which subsequently modulates the teacher model's results. This innovative step considerably bolsters the training efficacy of the student model. Recognizing its potential, we meld it into our framework, aiming to address the persistent issue of limited PolSAR-labeled data availability.

A pivotal concern in real-world datasets is the unequal distribution of object types, culminating in sample imbalance challenges. This imbalance frequently translates to suboptimal performance for minority classes. To address this, our research introduces a novel Self-Distillation-Based Correction Strategy (SDBCS), which integrates a label correction strategy, a sample rebalancing loss function, and data augmentation targeted for minority classes, enhancing overall classification accuracy. Our research proffers three pivotal contributions:

(1) We propose a new method using a feature distance matrix to correct label inaccuracies. This matrix, derived from contrastive learning principles, helps identify and rectify mislabeled samples by analyzing pixel similarities.
(2) We explore self-distillation learning to overcome the scarcity of labeled data in PolSAR. This approach utilizes inherent sample similarities for discriminative representation and achieves effective results, even with limited labels.
(3) Our strategy includes a rebalancing loss function and a data augmentation method for minority classes, significantly improving classification accuracy for minority classes.

2. Literature Review

2.1. Noisy Label Correction

The challenge of noisy labels in deep learning has become particularly critical in recent times. Models trained on noisy datasets can become susceptible to suboptimal representations, causing degraded performance in subsequent tasks. Addressing the noisy label issue, the research community has primarily focused on two solutions: (1) methods that train noise-resilient models directly on corrupted datasets and (2) methods that detect and rectify noisy labels before model training.

The former strategy involves modeling noise patterns directly, employing techniques such as robust loss functions [40,41], and noise corrections via noise transition matrices [15]. For instance, Ma et al. [18] developed a loss function that augments the resilience of DNNs against noisy labels. However, these methods often falter in the face of intricate noise patterns. Conversely, the latter strategy, gaining traction in recent years, particularly emphasizes sample selection. While some early approaches focused on curtailing the influence of noisy samples by training on selected clean subsets [42,43], more contemporary methods exploit semisupervised learning techniques [44]. Nonetheless, these techniques

frequently rely on assumptions about noise patterns, which can be detrimental if real-world noise deviates from these assumptions.

The intricacy of labeling PolSAR data, given the specialized expertise it demands, cannot be underestimated. This involves conferring precise class labels to specific pixels or regions within a PolSAR image, thereby setting the stage for frequent mislabeling. Such mislabeling, i.e., noisy labels, will inevitably undermine model performance. Notably, the differential distribution diagram delineated by [22] offered insights into clean sample probabilities, assisting in discerning between clean and noisy labels. Hou et al. [23] tackled the quandary of unreliable limited labels using a blended generative classification framework, wherein both labeled and unlabeled pixels were harnessed to derive high-level features.

2.2. Label Scarcity Problem with Contrastive Learning

PolSAR image classification, powered by supervised CNNs, has shown notable success. Yet, amassing large labeled datasets is both costly and time-intensive. Furthermore, limited training data can lead to model overfitting and reduced generalization. Given these issues, recent efforts, including label scarcity learning [45,46], aim to extract meaningful knowledge from minimal labeled samples. Specifically, methods under label scarcity learning, such as those cited, either harness learned optimization [47] or execute a feed-forward pass [48–50] without weight modifications. However, the methods employing a feed-forward pass often necessitate intricate inference protocols, reliance on RNN architectures, or task-specific fine-tuning [51,52].

Remarkable advancements in unsupervised representation learning have been realized via the advent of contrastive learning methodologies. By juxtaposing positive and negative samples in a self-supervised fashion, these strategies seek to derive salient data representations. For instance, the InstDisc [38] technique was the first to innovate a discrimination task, leveraging a memory bank to accumulate negative samples, thereby creating an expansive and consistent dictionary. Meanwhile, methods like CPC v1 [53], CMC [54], and MoCo v1 [55] have offered a multitude of contrasting and clustering tasks. Grill et al. [19] introduced BYOL, which employs one view's extracted feature to predict the feature of another view from the same instance, utilizing a momentum-based moving average for updating both encoder and representation. Yet, for all their success, contrastive learning techniques still grapple with achieving pinnacle accuracy on certain downstream assignments, particularly when benchmarked against supervised methods. Building upon prior successes, DINO emerged as a proposed solution to address these challenges, showcasing enhanced quality in learned representations. Notably, Caron et al. [39], drawing inspiration from BYOL, introduced several innovative techniques to elevate the performance metrics of self-supervised learning strategies.

Despite the evident potential of contrastive learning in generic image classification, its application remains conspicuously underrepresented in PolSAR imagery. Noteworthy explorations by Cui et al. [37] and Zhang et al. [25,26] have begun harnessing the merits of methods such as SimCLR, InstDisc, and BYOL for self-supervised PolSAR representation learning. These trailblazers proposed an avant-garde, self-supervised PolSAR representation learning paradigm, underscoring the potential synergy between contrastive learning and PolSAR imagery, especially in scenarios punctuated by label paucity.

To summarize, despite the widespread use of deep learning in PolSAR image classification, its effectiveness heavily relies on extensive annotations. This study aims to bridge the noticeable gap in applying contrastive learning within the PolSAR context. Our work differentiates itself by introducing a label correction strategy that utilizes inherent similarities among training samples to correct erroneous labels, which effectively solves the dilemma of noisy labels. Furthermore, we integrate self-distillation-based contrastive learning and a sample rebalancing loss function into an integrated framework, remarkably improving the classification performance on the PolSAR dataset, which presents label scarcity and class imbalance challenges.

3. Methodology

3.1. Overview of Our Method

In the subsequent sections, we delineate our methodology, beginning with the establishment of pertinent notations, followed by an exposition of the proposed framework. Given a PolSAR image, the PolSAR feature data are represented as $X \in R^{H \times W \times D}$, where H and W are the height and width of the PolSAR image, respectively, and D signifies the dimension of the chosen raw feature vector. The objective of our approach is to allocate a class label to each pixel in the image.

Figure 1 encapsulates the architecture of our proposed model, integrating modules for self-distillation-based feature extraction, label correction, and classification. Our approach commences with a finite set of randomly chosen pixels possessing noisy labels. In the initial phase, a convolutional neural network (CNN) is trained employing self-distillation-based deep representation learning. Following this, a global distance matrix is constructed, facilitating the identification of pixels bearing the highest resemblance for each sample. The labeling process then ensues, wherein labels are attributed based on the prevalence of a particular label within each cohort of similar pixels. Conclusively, to address class imbalances, a sample rebalancing loss function is introduced, which duly modulates the weights designated to varying classes, thereby refining classification accuracy.
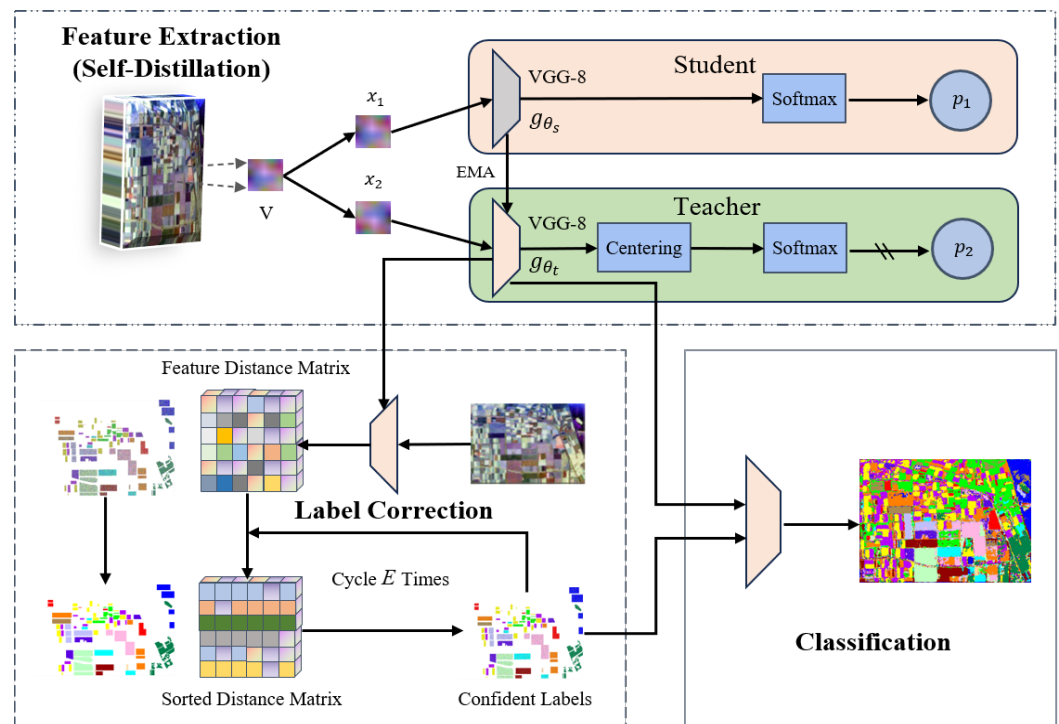

Figure 1. The proposed methodological pipeline encompasses three distinct modules: self-distillation-based feature extraction, label correction, and classification.

3.2. Raw Feature Extraction

We initiate by procuring the unprocessed polarimetric attributes, serving as the foundational input for our methodology. The resultant 6D feature set, symbolized as RF-*i* for *i* in the range 1 to 6, is derived from the complex coherency polarimetric matrix T, constructed using the Pauli basis of the PolSAR scattering matrix [56]. These attributes encapsulate critical information about the scattering mechanisms and are crucial for effective PolSAR image analysis.

As illustrated in Table 1, within this 6D feature set, RF-1 represents the total polarimetric power, known as SPAN (SPAN = $T_{11} + T_{22} + T_{33}$), expressed in decibel units. This feature provides a baseline measure of the total reflected energy, fundamental in understanding

the overall scattering characteristics of the observed scene. RF-2 and RF-3 symbolize the normalized power ratios of T_{22} and T_{33}, respectively. In the coherency matrix T, T_{22}, and T_{33} represent the power received in different polarization channels, such as horizontal–horizontal (HH) or vertical–vertical (VV), depending on the orientation of the PolSAR system. These elements are essential for analyzing the scattering behavior of different surface types in PolSAR imagery. By normalizing these power values against the SPAN, we obtain a relative measurement that is more robust to variations in absolute signal strength. RF-4 to RF-6 denote the relative correlation coefficients linked to the cross-polarization components T_{12}, T_{13}, and T_{23}. These coefficients measure the degree of correlation between different polarimetric channels, providing insights into the geometrical and dielectric properties of the scattering targets. They are particularly useful in distinguishing various surface types and man-made structures, which often exhibit unique polarimetric signatures.

Table 1. Raw polarimetric features employed in the proposed method.

Designation	Description
$RF\text{-}1 = 10\log_{10}(\text{SPAN})$	Polarimetric total power in decibel
$RF\text{-}2 = \frac{T_{22}}{\text{SPAN}}$	Normalized ratio of power T_{22}
$RF\text{-}3 = \frac{T_{33}}{\text{SPAN}}$	Normalized ratio of power T_{33}
$RF\text{-}4 = \frac{|T_{12}|}{\sqrt{T_{11} \cdot T_{22}}}$	Relative correlation coefficient of T_{12}
$RF\text{-}5 = \frac{|T_{13}|}{\sqrt{T_{11} \cdot T_{33}}}$	Relative correlation coefficient of T_{13}
$RF\text{-}6 = \frac{|T_{23}|}{\sqrt{T_{22} \cdot T_{33}}}$	Relative correlation coefficient of T_{23}

The necessity of this raw feature extraction process stems from its capability to convert complex and multidimensional PolSAR data into a format that is interpretable and applicable to machine learning algorithms. Features like T_{22} and T_{33} help in understanding the scattering behavior of different surfaces, crucial for accurate image classification. The feature extraction process thus translates PolSAR data into a form that machine learning algorithms can more effectively process and analyze. The selection of these particular features is informed by their established effectiveness in extracting meaningful information from PolSAR data, as highlighted in the existing literature [57]. These features assist in distinguishing different surface types and physical properties in the observed area, enhancing the classification accuracy. By employing these specific features, our approach not only capitalizes on the intrinsic properties of PolSAR data but also significantly enhances the potential for precise and robust classification outcomes. The scaling of RF-2 through RF-6 to the interval [0, 1] ensures uniformity in feature magnitude, which aids the learning algorithm in effectively processing and interpreting the data. This methodical approach to feature extraction lays a solid foundation for the subsequent machine learning processes, enabling our model to more accurately interpret and classify the intricate patterns inherent in PolSAR imagery.

3.3. Self-Supervised Learning with Knowledge Distillation

As we navigate through the challenges instigated by noisy labels and a scant quantity of labeled samples, we explore avant-garde techniques to bolster the discriminative prowess of our model. The presence of label noise and limited labeled samples present a dichotomy; while we require robustly discriminative features for label correction and subsequent classification, using these labels directly for learning might culminate in procuring misleading discriminative features. Enter contrastive learning, which offers a resolute solution by gleaning more illuminative supervised signals from raw unlabeled PolSAR data in an unsupervised fashion. To amplify the discriminative capacity of the model, we enlist knowledge distillation methodologies. At its core, knowledge distillation conceives a streamlined student model and hones it through the mentorship of a superior-performing

teacher model. The quintessence of this paradigm lies in transmitting knowledge from the teacher to the student, optimizing performance.

Our approach heralds a more kinetic interaction between teacher and student models. This synergy is materialized by gauging the disparity between the outcomes of the student and teacher models. This ushers in our feature extraction technique based on self-distillation contrastive learning. In the subsequent sections, we delve deep into aspects encompassing pretraining tasks, loss functions, and the architecture of the encoder and self-distillation module.

3.3.1. Pretext Task and Loss Function

In traditional supervised learning, models are honed to discern the intricate relationships between input data and their associated output labels, necessitating the availability of class information. Diverging from this paradigm, we propose an approach grounded in instance discrimination tasks. Within this framework, a neural network is self-supervised, training itself on two distinct data augmentation views. This methodology capacitates the network to concurrently project two variant views of an identical sample to a congruent representation space while projecting views from distinct samples to separate representation spaces. The inherent advantage is that the samples intrinsically act as their own supervisors, obviating the need for manual labeling. This strategy paves the way for harnessing vast repositories of unlabeled PolSAR images. Furthermore, by pretraining this network, we establish a deep feature network that is transferable. The network exhibits strong discriminative feature extraction capabilities, facilitating accurate label correction. Additionally, it adeptly addresses the small-sample challenges often encountered in classification tasks.

Figure 1 illustrates our proposed self-distillation contrastive learning model tailored for PolSAR data. This model is architecturally segmented into two networks: a student network, g_{θ_s}, and a teacher network, g_{θ_t}, visually discernible through orange and green modules, respectively. Both these networks, characterized by their respective parameters θ_s and θ_t, are intrinsically structured into three foundational components: an encoder, a projection head, and a predictor. Upon the sequential processing through these components, each network computes a probability distribution over Q dimensions, respectively denoted as P_s and P_t. Within the framework of our self-distillation contrastive learning approach, the designed loss function plays a pivotal role. It serves to nudge the neural networks into aligning similar instances in close proximity within the feature representation space, while simultaneously pushing apart dissimilar instances. This strategic configuration aids in fostering the extraction of robust discriminative features. A key element in this mechanism is the temperature parameter, denoted as $\tau_s > 0$, which dictates the acuteness of the distribution contour of P_s as

$$P_s(x)^{(i)} = \frac{\exp\left(\frac{g_{\theta_s}(x)^{(i)}}{\tau_s}\right)}{\sum_{k=1}^{Q} \exp\left(\frac{g_{\theta_n}(x)^{(k)}}{\tau_s}\right)} \tag{1}$$

In a parallel fashion, the temperature parameter τ_t governs the sharpness of P_t. To harmonize these distributions, we adopt a strategy of minimizing the cross-entropy loss concerning the parameters θ_s of the student network, all the while maintaining the teacher network g_{θ_t} in a static state. The objective function can be formally expressed as

$$\min_{\theta_s} H(P_t(x), P_s(x)) \tag{2}$$

where the relationship $H(a, b) = -a \log b$ holds true. We generate a set of views, V, from the PolSAR images, where views x_1 and x_2 are two randomly augmented views. Our primary pursuit is encapsulated in the minimization of the loss, articulated as

$$\min_{\theta_s} \sum_{x \in \{x_1, x_2\}} \sum_{x' \in V x' \neq x} H(P, P(x')) \tag{3}$$

To refine the parameters θ_s, we employ the stochastic gradient descent method, targeting the minimization of Equation (3).

3.3.2. Architecture of Encoder and Self-Distillation Module

In light of the aforementioned principles, we architected a network for self-distillation contrastive learning. The encoder in our model incorporates the VGGNet-8 structure, serving as a convolutional feature extractor designed for processing input images. It is composed of three convolutional blocks, each containing two layers that use 3×3 convolutional kernels, followed by a ReLU activation function and 2×2 max-pooling, effectively capturing and processing image features. In parallel, the projection head transforms the input feature vectors into a lower-dimensional space through dense layers, enabling the learning of more compact yet abstract data representations while preserving crucial feature information. Additionally, the predictor utilizes a fully connected layer to map these feature vectors into Q dimensions. This dimensionality reduction is achieved using a softmax activation function, which calculates the probability distribution across various classes, ensuring an effective and efficient classification process.

During the training regime, neither network updates its parameters based on labeled data. An input image, denoted as x, undergoes random augmentations to yield two distinct variants, x_1 and x_2. Subsequently, these variants are independently channeled into both the student and teacher networks. It is imperative to note that while these networks architecturally mirror each other, they possess unique parameters, thus fostering independent learning and nuanced data comprehension. To achieve consistent representations, the output of the teacher network is centralized by computing its mean over the entire batch, subsequently normalizing these features across individual samples. Both networks yield an M-dimensional feature vector, which undergoes further normalization via a temperature-regulated softmax operation across its dimensions. The congruence between the feature vectors from the student and teacher networks is ascertained using a cross-entropy loss. This loss function measures the discrepancy between the predicted probability distributions of the two networks. By striving to minimize this loss, we compel the networks to generate analogous representations for equivalent input samples, thus enhancing the knowledge transfer from the teacher to the student. It is paramount during training to restrict the flow of gradients solely to the student network. To achieve this, we deploy a stop-gradient operator on the teacher network, ensuring its immunity from external updates and guaranteeing that only the student network receives iterative refinements.

Our methodology presents a notable divergence from traditional knowledge distillation practices, especially in its approach to temperature scaling. Conventionally, the teacher temperature parameter is held invariant throughout the training, serving to temper the fluctuations in its output probabilities. In contrast, our approach harnesses a temperature scheduling mechanism that methodically diminishes the temperature of the teacher model as training advances. The initiation phase employs a heightened temperature to ensure a robust training foundation, which is progressively tapered to bolster the distillation impact. By refashioning the teacher model's knowledge, manifested as soft targets or feature representations, we aim to effectively shepherd the student's learning trajectory. Furthermore, we introduce a mechanism for updating the center vector based on the outputs of the teacher model. This innovation not only enhances knowledge distillation but also marks a distinction from traditional methodologies.

Unlike the conventional approach of initializing the teacher network by directly copying the student network's weights, our strategy crafts the teacher network based on antecedent iterations of the student network. This process is refined using the nuances of the exponential moving average, as demonstrated by the following rule: $\theta_t \leftarrow \lambda \theta_t + (1 - \lambda) \theta_s$.

As training ensues, we adopt a λ value that commences at 0.996 and ascends, tracing a cosine trajectory until it culminates at unity. Consequently, in the nascent stages, the teacher network's parameters gravitate swiftly toward their student counterparts. Yet, as the training journey evolves, this adaptation pace decelerates, culminating in a poised equilibrium. This meticulously crafted strategy strikes a harmonious balance between maintaining the stability of the teacher network and optimizing its directive potency on the student network's representations. To encapsulate, our proposed self-distillation contrastive learning method undergoes cyclical refinements, capitalizing on variances between views to adeptly mediate the knowledge transference between the student and teacher constructs. The outcome of this innovative methodology is the adept extraction of discerningly potent features, leading to a marked enhancement in model proficiency.

4. Enhancing Classification Accuracy

In this section, we address two pivotal aspects of classification accuracy: label correction and addressing class imbalance. The label correction module corrects mislabeled instances, while our class imbalance strategy ensures a fair representation of all classes. This dual approach is crucial for the precise categorization of PolSAR data, where both label quality and balanced class representation significantly impact the classifier's performance.

As illustrated in Figure 2, our proposed label correction strategy capitalizes on the inherent affinities among training samples to amend erroneously assigned labels. Within this strategy, the backbone network of a contrastive learning framework is employed to distill features and create a comprehensive distance matrix encompassing all training samples. For each pixel, we then identify its top-*K* nearest samples, based on the predefined distance metric. The label that exhibits the highest frequency among these nearest samples is then designated to the pixel under consideration. This approach adeptly harnesses representational affinities to ameliorate the classification of incorrectly labeled instances.
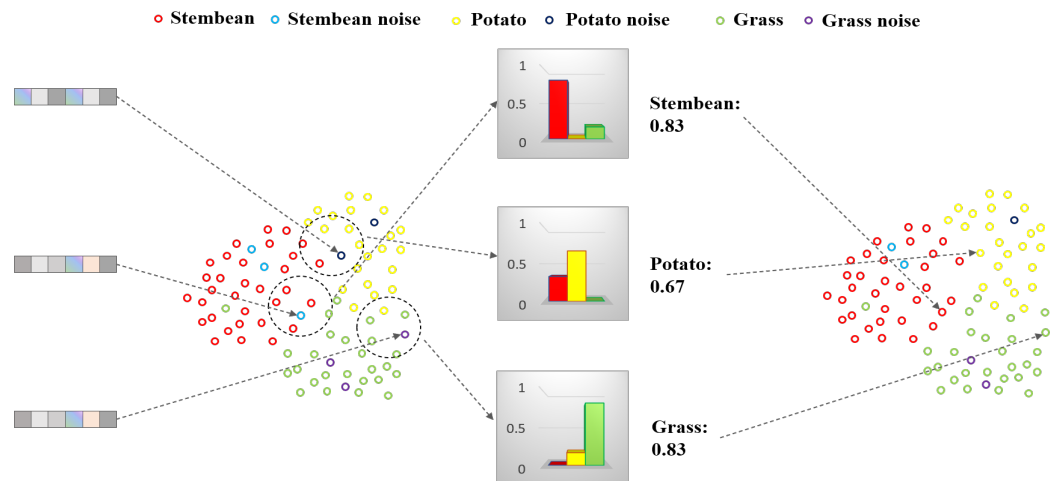

Figure 2. The label correction procedure can be delineated as depicted in this figure. Initially, a global distance matrix is constructed to discern pixels demonstrating the paramount similarity to individual samples. Subsequent to this step, the label for each pixel is determined based on the predominant label within its associated cluster of similar pixels.

Consider a scenario wherein a sample, erroneously labeled under a *non-Stembean* category, requires rectification, given that its ground truth designation is *Stembean*. Assuming a top-*K* threshold of 6, the six proximal samples in the feature space relative to this label are selected. The distribution among these reveals 1 *Grass* label, 5 *Stembean* labels, and no other categories. As a result, the *Stembean* category emerges with a probability of 0.83, surpassing the stipulated threshold for label correction. Consequently, the label is rectified to *Stembean*.

To better understand the pseudocode of our label correction algorithm presented in Algorithm 1, it is essential to define some key variables used within it. The total pixel count

is denoted as N, calculated as $H \times W$, where H and W represent the height and width, respectively. We denoted the set of labeled pixel pairs as $L = \{(x_1, t_1), (x_2, t_2), \dots, (x_n, t_n)\}$, where $\mathcal{X}$ and $\mathcal{Y}$ represent the data and label parts of L, respectively, and n is significantly smaller than N. Within this context, M denotes the total number of classes. The label of each sample x_i corresponding to the one-hot label vector y_i is expressed as $l_i = \arg_j[y_i(j) = 1] \in \{1, \dots, M\}$. The objective of our approach is to allocate a class label y_i to each pixel i, where $i \in \{1, 2, \dots, n\}$. Algorithm 1 delineates the pseudocode of our advanced label correction algorithm, which intakes both original features and augmented image labels. The primary objective of this module is to redress noisy labels. Its foundational architecture, denoted as f, is sculpted through self-distillation rooted in contrastive learning. For brevity, we define f_i as the representational feature of the sample x_i. p_{f_i} is the projection head derived from p_f, which exemplifies the encoder's prowess in capturing intricate, high-dimensional features. Additionally, p_f is employed to construct a *K*-Nearest Neighbors Classifier (KNNC) k_q, with $k_{q_i} \triangleq k_q\left(p_{f_i}\right)$ representing its predictive vector.

Algorithm 1: Label Correction Algorithm

---

1   **Input:**
     $(\mathcal{X}, \mathcal{Y})$
     n represents the size of the training set
     Sample relabelling threshold θ_s
     Max epochs E
     p_f represents the feature extractor
     $\mathcal{Y}_E$ is a list of elements denoted by $\mathcal{Y}_e$

2   **Output:**
     The clean label of $\mathcal{Y}$

    1: Data augmentation on small classes:
    2: **for** $i = 1$ **to** n **then**
    3:     Extract feature
    4: **end for**
    5: **for** $i = 1$ **to** n **then**
    6:     **for** $j = 1$ **to** n **then**
    7:        Calculate similarity between each representation: Equation (4)
    8: **end for**
    9: **for** $e = 1$ **to** E **then**
  10:     **for** $i = 1$ **to** n **then**
  11:        Measure of consistency c_i: Equation (6)
  12:        **if** $c_i < \theta_s$ **then**
  13:           l_i^r is likely to be wrong
  14:        **else**
  15:           $y_i^r \leftarrow l_i^r$
  16:        **end if**
  17:        $\mathcal{Y}_{Ei} = y_i^r$
  18: **end for**
  19: **for** $j = 1$ **to** E **then**
  20:     $\mathcal{Y}_j = Max_j \sum\limits_{i=1}^{n} \mathcal{Y}_{Ej}$
  21: **end for**

The affinity between the representations p_{f_i} and p_{f_j} of samples x_i and x_j is articulated as s_{ij}, where both i and j iterate from 1 to n. The cosine similarity is computed as

$$\mathbf{s}_{ij} = \frac{p_{f_i}^T p_{f_j}}{\left\| p_{f_i} \right\|_2 \left\| p_{f_j} \right\|_2} \tag{4}$$

This remains our measure of choice. The index set for the *S*-nearest neighbors of sample x_i in $\mathcal{X}$, predicated on this similarity, is denoted as N_i. For every sample x_i, the normalized label distribution is computed as

$$k'_{q_i} = \frac{1}{S} \sum_{n \in N_i} y_n^r \tag{5}$$

A subsequent balanced version, $k_{q_i} \in R^M$, adjusts for the label distribution $\pi = \sum\limits_{i=1}^{N} yi^r$ inherent to the dataset, with $kq_i = \pi^{-1}k'_{q_i}$, where π^{-1} comprises the inverse of π's entries, compensating for potential sample selection biases arising from class imbalances.

For each specific sample, we ascertain instances manifesting maximal similarity using their respective distance metrics, and based on these proximate samples, we proceed to refine the associated labels. For every pixel, the foremost top-*K* nearest samples delineated by the designated distance metric are identified. We introduce a consistency metric, represented as c_i, which gauges the congruence between sample label $l_i^r = \arg\max j y_i^r(j)$ and the prediction sourced from KNNC:

$$c_i = \frac{k_{q_i}(l_i^r)}{\max j k q_i(j)} \tag{6}$$

This metric is derived by dividing the value of the distribution k_{q_i} corresponding to the label l_i^r by its predominant peak $\max j k q_i(j)$. A pronounced c_i value for a given sample x_i insinuates a consensus among its neighboring samples in favor of its prevailing label l_i^r, suggesting its likely accuracy. Applying a threshold θ_s to c_i, a pristine subset $(\mathcal{X}_c, \mathcal{Y}_c^r)$ is derived. By default, we utilize $\theta s = 0.65$, implying that a sample x_i is deemed pristine when the consensus, as reflected in kq_i, among its neighbors corroborates its extant label y_i^r.

In light of limited labeling, we propose a data augmentation strategy that capitalizes on the original features and labeled image pairs. Specifically, our approach adopts an offline data augmentation technique tailored for underrepresented or minor-category samples, ensuring that transformations are conducted on the training data prior to their introduction into the label correction module. Historically, popular data augmentation methodologies have included translation, image flipping, rotation, and cropping, as corroborated by Hernandez et al. [58] and Wong et al. [59]. In alignment with these practices, we implement five cardinal data augmentation operations, represented as AUG-*i* (where $i \in 1, \ldots, 4$): AUG-1 denotes horizontal flipping, AUG-2 implies a 90° clockwise rotation, AUG-3 indicates a 180° clockwise rotation, and AUG-4 pertains to a 270° clockwise rotation. Subsequently, each training image patch pair, $(x_i, y_i) \in (\mathcal{X}, \mathcal{Y})$, where $i \in 1, 2, \ldots, n$, is extended into a series of eight image patch pairs. These include (x_i, y_i), (x_i^{R90}, y_i), (x_i^{R180}, y_i), (x_i^{R270}, y_i), (x_i^F, y_i), (x_i^{FR90}, y_i), (x_i^{FR180}, y_i), and (x_i^{FR270}, y_i). The subsequent seven pairs in this sequence correspond to transformations driven by the operations AUG-1 through AUG-4.

With the refined labels in place, the primary objective of the classification module is the categorization of PolSAR data. A projection head is utilized within this module, projecting the representations gleaned from the network onto a dimensionality defined by the class number. This is mathematically represented as

$$d_i(\vec{x}) = \frac{\exp\left(\vec{W}_i^T \vec{x} + \vec{b}_i\right)}{\sum\limits_{j=1}^{M} \exp\left(\vec{W}_j^T \vec{x} + \vec{b}_j\right)} \tag{7}$$

Here, $\vec{x}$ is the output of the projection head, with $\vec{W}_\bullet^T$ and $\vec{b}_\bullet$ representing the associated weight and bias, respectively. Furthermore, to effectively confront sample imbalance, we introduce a rebalancing loss, denoted as $\mathcal{L}_{CACE}$, encapsulated in Equation (5). The foundational loss function employed is the categorical cross-entropy [60], $\mathcal{L}_{CCE}$. The

derivation of $\mathcal{L}_{CACE}$ necessitates averaging two error magnitudes, both of which are scaled by the categorical weight W.

$$\mathcal{L}_{CACE} = -W \times [\mathcal{L}_{CCE}(\vec{y}, y)], \tag{8}$$

Here, W is formulated as $[\frac{1}{N_1}, \frac{1}{N_2}, \ldots, \frac{1}{N_M}]^T$. In this equation, $\vec{y}$ symbolizes the predicted label, y stands for the ground truth label, and N_k represents the count of labels in the kth class.

5. Experimental Results

In this section, we provide a rigorous evaluation of our proposed method on four PolSAR datasets, both from quantitative and qualitative perspectives. We initially detail the experimental datasets and our chosen parameter settings in Section 5.1. In Section 5.2, an ablation study is presented to highlight the significance of the four pivotal components of our method: self-distillation backbone, noise label correction approach, sample rebalancing loss function, and augmented dataset.

For clarity, we elucidate that the classification average accuracy (AA) for a class is the proportion of accurately classified pixels for that class to the total pixels of the class, whereas the overall accuracy (OA) refers to the proportion of all correctly classified pixels in the entire image to the overall pixels in the image. Data with the highest accuracy are highlighted in bold for emphasis.

5.1. Experimental Data and Parameter Setting

Figure 3 offers visual insights, and Table 2 presents a summary of the PolSAR images employed in our experiments. The first dataset is composed of L-band four-look PolSAR data, captured by the NASA/JPL AIRSAR system over the Flevoland region, the Netherlands, in August 1989, whose PauliRGB image is portrayed in Figure 3(a1). Spanning an area of 750 × 1024 pixels, it offers a resolution of 6.6 m in the slant range and 12.1 m in the azimuth direction. The dataset delineates 15 distinct land cover classes, as illustrated in Figure 3(a2), with color codings that represent the legend of the ground truth map. The number of pixels for each class is listed as below: *Water* (12,671), *Barley* (7156), *Peas* (9111), *Stembean* (6103), *Beet* (10,050), *Forest* (14,822), *Bare soil* (3078), *Grass* (6269), *Rapeseed* (12,690), *Lucerne* (9477), *Wheat 1* (17,283), *Wheat 2* (10,591), *Wheat 3* (21,300), *Building* (476), and *Potato* (15,292). For model training, a random subset comprising 1% of the labeled samples is utilized. We then proceed to extract image patches of dimensions 12 × 12 × 6, where 12 × 12 signifies the window size and 6 represents the channel count.

Table 2. Summary of PolSAR datasets.

Dataset	Size	Spatial Resolution (m)	Bands	Classes
Flevoland (Dataset 1)	750 × 1024	6.6 × 12.1	L-band	15
Oberpfaffenhofen	1300 × 1200	1.5 × 1.8	L-band	3
Flevoland (Dataset 2)	1020 × 1024	6 × 12	L-band	14
San Francisco	1800 × 1380	3 to 100	C-band	5

Our second dataset, as illustrated in Figure 3(b1), comprises an E-SAR L-band image, which covers a 1300 × 1200 pixel area in the Oberpfaffenhofen region, Germany. This dataset includes several distinct land categories, with the number of pixels for each as follows: *Build-up* areas (333,955), *Wood Land* (265,516), and *Open Area* (760,769). Its diversity renders it apt for gauging the robustness of our method in varied landscapes. The ground truth class labels and their associated color legends for this area are delineated in Figure 3(b2), serving as a benchmark for our model's predictions and enabling classification accuracy quantification.
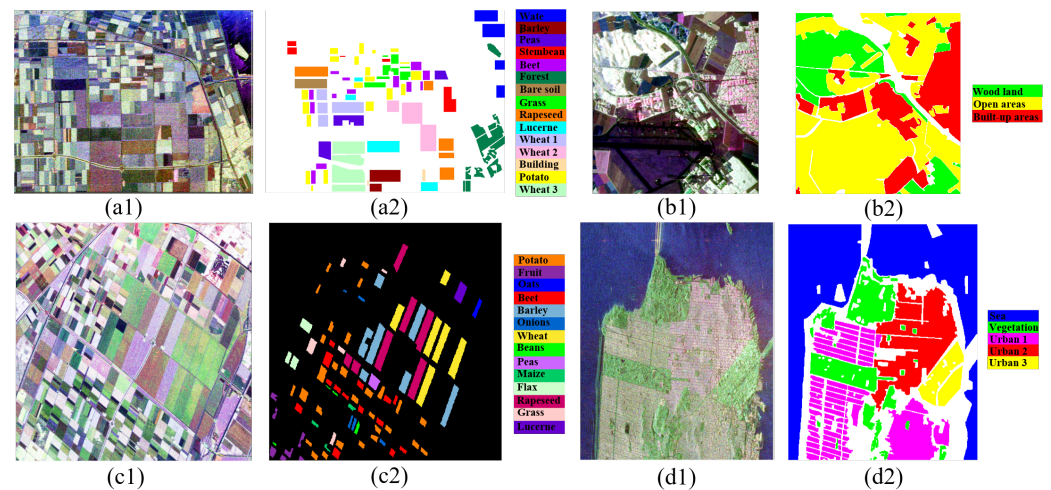
Figure 3. In this study, a series of experimental images were employed to rigorously assess the efficacy of our proposed method. The selected images encompass the following: Flevoland area dataset 1: (**a1**) A PauliRGB depiction of the region. (**a2**) The associated ground truth class labels, supplemented by their corresponding color codes. Oberpfaffenhofen Area Data Set: (**b1**) The PauliRGB representation of the aforementioned area. (**b2**) Ground truth class labels, paired with their relevant color codes. Flevoland area dataset 2: (**c1**) Another distinct PauliRGB portrayal from the Flevoland region. (**c2**) Its affiliated ground truth class labels, along with the matching color codes. San Francisco Area Data Set: (**d1**) The PauliRGB visualization of this iconic urban landscape. (**d2**) The ground truth class labels, harmonized with their specified color codes.

Figure 3(c1) showcases the third dataset: an L-band AIRSAR image captured over the Flevoland region in 1991. This dataset, spanning dimensions of 1020×1024 pixels, is indispensable for discerning the radar responses of different land cover types and augmenting our grasp of PolSAR data interpretation. Figure 3(c2) manifests the corresponding ground truth labels and color codings. This dataset, encapsulating 14 classes, is referred to as Flevoland area dataset 2 in Section 5. This dataset includes a diverse range of land types, with the number of pixels for each being *Potato* (21,539), *Fruit* (4062), *Oats* (1394), *Beet* (10,795), *Barley* (24,543), *Onions* (2130), *Wheat* (26,277), *Beans* (1082), *Peas* (2160), *Maize* (1290), *Flax* (4301), *Rapeseed* (28,235), *Grass* (4204), and *Bare Soil* (2952) pixels.

The fourth dataset entails a 25-look Radarsat-2 image of the San Francisco region from 2008, with a size of 1800×1380 pixels. This dataset features five classes, with the number of pixels for each being *Sea* (841,489), *Vegetation* (236,715), *Urban 1* (80,616), *Urban 2* (348,056), and *Urban 3* (282,975). Figure 3(d1) renders the PauliRGB image, while Figure 3(d2) displays the ground truth class labels. Notably, in Figure 3(d2), void regions are apparent, symbolizing unlabeled classes or interclass boundaries. These void zones are excluded from experimental consideration and analysis.

The optimization algorithm was parameterized with a learning rate (τ) set at 0.001, complemented by a momentum parameter of 0.9. During training, we utilized a batch size of 128. For all experiments, we initialized with a noisy label rate of 20%. All experiments were orchestrated within the TensorFlow framework, leveraging a Dell Z690 workstation equipped with a GeForce RTX 3090 GPU and a memory capacity of 64 GB.

5.2. Ablation Study

The proposed method, predicated on the robust self-distillation mechanism for correcting noisy labels, was rigorously tested on various prominent PolSAR images, as delineated in earlier sections. This study bifurcates into three critical experimental segments, each elucidating distinctive facets of the model's capabilities. Initially, the research accentuates the advantages of harnessing self-distillation for feature extraction, particularly when maneuvering high-dimensional vector distance computations in PolSAR imagery. For this

purpose, two contrasting experimentations were devised: one incorporating contrastive learning and the other omitting it. To testify the effectiveness of each component of our proposed SDBCS, we conducted four groups of experiments as follows: We start with VGGNet-8 as our baseline, which trains directly on noisy-labeled samples. We then examine the influence of our label correction module with the VGGNet-8+CS model. Advancing further, SDVGGNet-8+CS enriches the previous model by adding self-distillation-based contrastive learning, aiming for enhanced feature extraction. The penultimate step in our experimental series, SDVGGNet-8+CS+Aug, integrates data augmentation into the SDVGGNet-8+CS framework to further improve the model's resilience to noisy data and enhance generalization. The culmination of our experimental series, the SDBCS framework, incorporates data augmentation and balanced loss into SDVGGNet-8+CS, specifically designed to overcome class imbalance and enhance the model's classification efficacy.

We leverage the Oberpfaffenhofen dataset to verify the efficacy of our method. Table 3 elucidates the foundational methodology, wherein a VGGNet-8 neural network was trained directly on the dataset, inclusive of the noise-labeled samples, sans any modification. This primary approach served as a litmus test for gauging model performance. Resultant accuracies across various classes were as follows: *Build-up* at 65.69%, *Wood Land* at 68.55%, and *Open Area* at 84.87%. Consequently, the OA was pegged at 76.98%, with an AA of 73.04%. The Precision, which indicates the accuracy of positive predictions, was recorded at 73.09%. The F1-Score, which balances precision and recall, was 73.05%, indicating a moderate balance in the model's ability to correctly identify classes and its robustness in terms of recall. The Kappa statistic, measuring agreement beyond chance, stood at 60.96%, suggesting a fair level of agreement. The Mean IoU, crucial for assessing the model's performance in segmenting classes, was 58.34%.

Table 3. OA values (%) of Oberpfaffenhofen area data for our proposed method.

Method	Build-up	Wood Land	Open Area	OA	AA
VGGNet-8	65.69	68.55	84.87	76.98	73.04
VGGNet-8+CS	63.89	76.28	87.02	79.25	75.73
SDVGGNet-8+CS	72.94	91.13	88.22	85.04	84.10
SDVGGNet-8+CS+Aug	81.07	92.19	87.48	86.82	86.91
SDBCS	79.08	89.12	92.38	88.48	86.86

Method	Precision	F1-Score	Kappa	Mean IoU	
VGGNet-8	73.09	73.05	60.96	58.34	
VGGNet-8+CS	75.15	75.37	64.85	61.30	
SDVGGNet-8+CS	81.57	82.58	75.08	70.91	
SDVGGNet-8+CS+Aug	84.02	85.25	78.23	74.84	
SDBCS	85.87	86.35	80.58	76.49	

5.2.1. Noisy Label Correction

To elevate the established baseline, we incorporated a correction mechanism into the VGGNet-8 model. For confident label determination, an intricate global distance matrix encompassing all pixels was constructed. The objective was to discern the most congruent pixels for each sample and subsequently adopt the predominant label within its pixel cohort. Each training sample was aligned to the label of the nearest k training data points. After computing the feature distance, the samples were sorted based on proximity. This strategy was devised to counteract the detriments of noisy labels and bolster classification accuracy. Despite inevitable trade-offs, the method showcased an upswing in performance. The achieved accuracies were *Build-up* at 63.89%, *Wood Land* at 76.28%, and *Open Area* at 87.02%. OA increased to 79.25%, with an AA of 75.73%. Additionally, precision improved to 75.15%, F1-Score to 75.37%, Kappa to 64.85%, and Mean IoU to 61.30%.

5.2.2. Self-Distillation Feature Extraction

The model's performance was further augmented by embedding a self-distillation technique, thereby enabling the model to introspectively refine from its own predictions. This adaptation yielded notable enhancements, with the accuracies for *Build-up*, *Wood Land*, and *Open Area* classes registering at 72.94%, 91.13%, and 88.22%, respectively. The OA marked an impressive 85.04%, culminating in an AA of 84.10%. Precision increased to 81.57%, F1-Score to 82.58%, Kappa to 75.08%, and Mean IoU to 70.91%.

In the realm of PolSAR data analysis, initial steps involve feature extraction from PolSAR data using VGGNet-8 and self-distillation methodologies. A supplementary set, termed VGGNet-8, was introduced for a comparative evaluation, which essentially trains without noise labels. These methodologies illuminate intricate relationships within the data, rendering high-dimensional features that encapsulate pivotal backscattering properties of PolSAR data. The subsequent phase emphasizes dimensionality mitigation.

Efficient dimensionality reduction is pivotal for interpreting high-dimensional data. One salient technique in this domain is t-distributed Stochastic Neighbor Embedding (t-SNE) [61], a sophisticated nonlinear algorithm grounded in neighborhood graphs, tailored to preserve the data's intrinsic local structure. This is achieved by t-SNE's transformation of interpoint distances into congruent probability distributions spanning various dimensions. Leveraging t-SNE, we embarked on visualizing both raw and quantized feature spaces. Given its design as an unsupervised algorithm tailored for dimensionality reduction and 3D data projection, t-SNE demonstrates exceptional prowess in rendering visualizations of intricate, high-dimensional datasets, thereby enhancing the interpretation of PolSAR data. The utility of t-SNE is further accentuated when amalgamated with visual aids like scatter plots and pseudocolor images, facilitating a lucid conveyance of intricate data relationships and patterns.

Figure 4 presents a detailed visual exposition of the spatial and polarimetric attributes across three preselected regions from the Oberpfaffenhofen dataset. The visualization unmistakably illustrates a clear delineation of three terrain typologies within the three-dimensional space charted by t-SNE. In particular, Figure 4a,d underscores the aptitude of the feature extraction network, shining light on its innate ability to capture and epitomize the quintessential characteristics of terrain surfaces. A deeper foray into Figure 4b,c provides a comparative purview against Figure 4a. Significantly, Figure 4c, harmonized with the self-distillation paradigm, exhibits heightened alignment with the ground truth, especially in the positionings pertaining to the three distinct categories.

In parallel with our assessment of Figure 4d, a detailed comparative analysis is presented in Figure 4e,f, bringing forth salient observations. Notably, Figure 4f, emblematic of the self-distillation-based method, highlights a pronounced aggregation in the central positions associated with various categories. This stands in stark contrast to the more scattered distribution observed within the VGGNet-8 influenced outcomes, as delineated in Figure 4e. Collectively, these observations underscore the superior discriminative capacity of the self-distillation approach, adeptly capturing inherent class distinctiveness and intricate intercategory dynamics. This fortifies the assertion of its pivotal role in elevating feature representation in the analyzed PolSAR dataset.

5.2.3. Data Augmentation and Balanced Loss

We further refined the SDVGGNet-8+CS model by integrating data augmentation, resulting in the SDVGGNet-8+CS+Aug configuration. This intermediate step was crucial in assessing the incremental benefits brought by data augmentation to the self-distillation process. The SDVGGNet-8+CS+Aug model demonstrated a significant improvement in dealing with noisy data and generalization capabilities, as evidenced by the following accuracies: *Build-up* at 81.07%, *Wood Land* at 92.19%, and *Open Area* at 87.48%. The OA and AA were recorded at 86.82% and 86.91%, respectively. Additionally, the model saw improvements in Precision (84.02%), F1-Score (85.25%), Kappa (78.23%), and Mean IoU

(74.84%). These advancements highlight the method's impact in not only improving accuracy but also precision, consistency, and segmentation effectiveness.
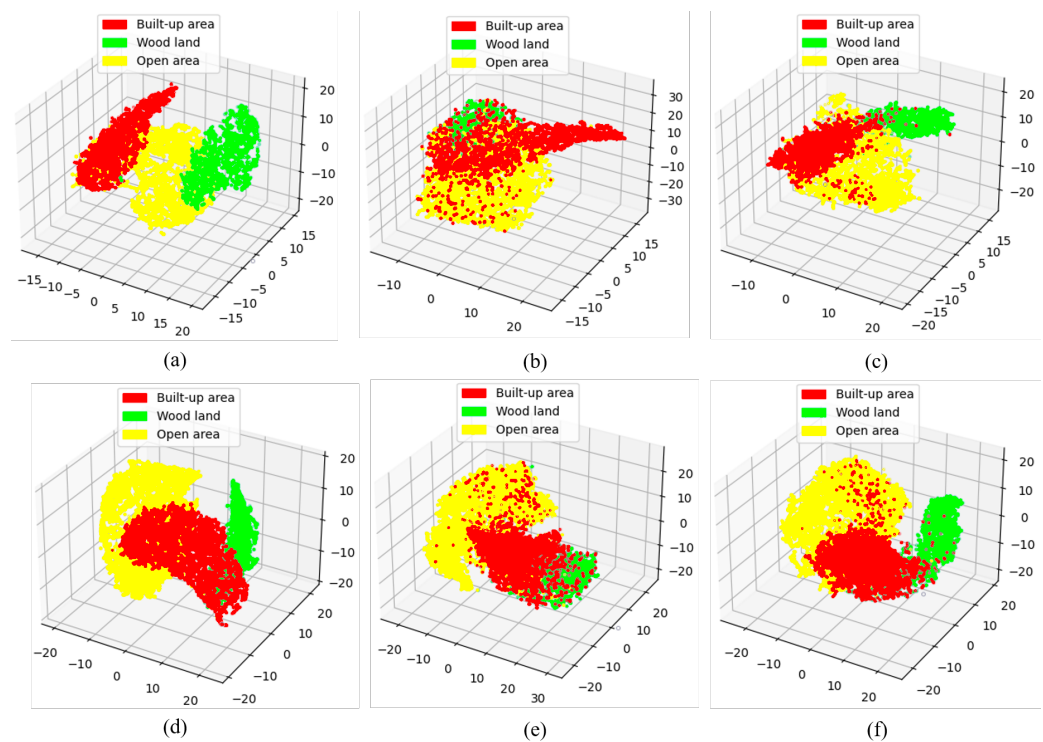

Figure 4. The presented figures showcase t-SNE plots derived from the Oberpfaffenhofen images. For enhanced visualization fidelity, the dataset is judiciously bifurcated into two subsets according to their sample proportions. Specifically, subsets (**a**–**c**) constitute 0.5% of the overarching samples, whereas subsets (**d**–**f**) account for 1%. In terms of methodological delineation, subsets (**a**,**d**) resonate with the features from the VGGNet-8 backbone trained devoid of noise labels, and subsets (**b**,**e**) are aligned with the VGGNet-8 training approach. Conclusively, subsets (**c**,**f**) are emblematic of the feature extraction facilitated through the self-distillation-based paradigm. Such a structured presentation aids in an in-depth comparison and assessment of the respective methodologies across varied sample sizes.

As illustrated in Table 4, we explored different loss functions in order to find a robust option. The studied loss functions include $\mathcal{L}_{CCE}$ [60], Label Smoothing Categorical Cross-Entropy Loss [62] ($\mathcal{L}_{SCCE}$), Focal Loss [63] ($\mathcal{L}_{focal}$), and our proposed $\mathcal{L}_{CACE}$.

It is evident that both $\mathcal{L}_{SCCE}$ and $\mathcal{L}_{focal}$ demonstrate promising results under certain parameter settings. However, it is crucial to note that slight changes in their parameters can lead to significant drops in classification performance. For instance, when the ϵ parameter in $\mathcal{L}_{SCCE}$ changes from 0.3 to 0.2, there is a notable decrease in OA by 2.11%. Similarly, in Focal Loss, a change in the γ parameter from 1.8 to 2.0 results in a reduction in OA by 2.17%. This sensitivity to parameter adjustments indicates that both $\mathcal{L}_{SCCE}$ and Focal Loss may not be robust across different categories or datasets, as their effectiveness heavily relies on fine-tuning specific parameters.

In contrast, our $\mathcal{L}_{CACE}$, meticulously designed to overcome the limitations of existing methods, demonstrated remarkable results. Significantly, $\mathcal{L}_{CACE}$ stands out due to its parameter-free design, eliminating the need for meticulous parameter tuning that plagues other loss functions. This unique feature enhances its robustness, making it exceptionally suitable for a wide range of PolSAR datasets. It achieved impressive classification accuracies and showcased enhanced Precision (85.87%), F1-Score (86.35%), Kappa (80.58%), and Mean IoU (76.49%). The absence of parameters in $\mathcal{L}_{CACE}$ not only simplifies its application but also ensures consistent performance across various scenarios in PolSAR datasets.

In conclusion, this investigative endeavor presents a holistic exploration of innovative methodologies tailored for optimizing neural-network-centric classifiers within the remote sensing land cover classification domain. The empirical findings highlight the paramount importance of bespoke strategies, especially when confronting challenges like label noise and constrained data availability. The integration of self-distillation, data augmentation, and balanced loss within the SDBCS framework emerges as a testament to this. Such revelations not only augment our contemporary understanding of effective strategies within this discipline but also establish an empirical benchmark, poised to guide and inspire subsequent research trajectories in analogous domains.

Table 4. Performance comparison of different loss functions with SDVGGNet-8+CS+Aug architecture on Oberpfaffenhofen area data, utilizing 0.05% of ground truth labels as the training set.

Loss Function	Build-up	Wood Land	Open Area	OA	AA
$\mathcal{L}_{CCE}$	81.07	92.19	87.48	86.82	86.91
$\mathcal{L}_{SCCE}\ \epsilon(0.2)$	81.42	**92.94**	87.58	87.11	87.31
$\mathcal{L}_{SCCE}\ \epsilon(0.3)$	80.97	91.15	92.16	**89.22**	**88.09**
$\mathcal{L}_{SCCE}\ \epsilon(0.5)$	**81.82**	92.23	87.75	87.17	87.27
$\mathcal{L}_{focal}\ \gamma(2.0)\ \alpha(0.37)$	81.10	90.82	88.02	86.87	86.65
$\mathcal{L}_{focal}\ \gamma(2.0)\ \alpha(0.50)$	81.10	91.26	87.97	86.93	86.78
$\mathcal{L}_{focal}\ \gamma(1.8)\ \alpha(0.50)$	75.87	90.93	**94.27**	89.10	87.02
$\mathcal{L}_{CACE}$	79.08	89.12	92.38	88.48	86.86

Loss Function	Precision	F1-Score	Kappa	Mean IoU
$\mathcal{L}_{CCE}$	84.02	85.25	78.23	74.84
$\mathcal{L}_{SCCE}\ \epsilon = 0.2$	84.17	85.50	78.73	75.17
$\mathcal{L}_{SCCE}\ \epsilon = 0.3$	86.39	**87.18**	**81.92**	**77.73**
$\mathcal{L}_{SCCE}\ \epsilon = 0.5$	84.24	85.52	78.82	75.21
$\mathcal{L}_{focal}\ \gamma(2.0)\ \alpha(0.37)$	84.13	85.21	78.25	74.78
$\mathcal{L}_{focal}\ \gamma(2.0)\ \alpha(0.50)$	84.09	85.23	78.39	74.84
$\mathcal{L}_{focal}\ \gamma(1.8)\ \alpha(0.50)$	**86.53**	86.69	81.51	77.06
$\mathcal{L}_{CACE}$	85.87	86.35	80.58	76.49

6. Discussion

In this section, we provide a rigorous evaluation of our proposed method on four PolSAR datasets, both from quantitative and qualitative perspectives. Section 6.1 delves into a sensitivity analysis, assessing the robustness of the proposed SDBCS framework on the Oberpfaffenhofen dataset. Section 6.2 furnishes a comparison between our proposed method and four contemporary state-of-the-art competitors employing deep learning techniques, namely Sel-CL [64], SSR [65], PASGS [22], and Auto-PASGS [22].

6.1. Sensitivity Analysis

To elucidate the robustness of the proposed method across varying proportions of training data, this section meticulously evaluates the SDBCS framework on the Oberpfaffenhofen dataset. This dataset served as a canvas for a rigorous appraisal of the land cover classification efficacy of SDBCS, with the Average Correction Rate (ACR) as the evaluation metric. Table 5 presents the corrected label rate for three principal land cover classes—*Build-up* area, *Wood Land*, and *Open Area*—across 0.05%, 0.1%, and 0.2% data proportions.

In juxtaposition with Sel-CL and SSR, the supremacy of SDBCS was consistently evident. It is noteworthy that, particularly in the *Build-up* area class, SDBCS was adept at maintaining commendable classification accuracy, even with limited data, underscoring its potent capacity for generalization relative to other methods. A salient aspect of the study was the discernible prowess of SDBCS in classifying the *Wood Land* segment, even when confronted with constrained data volumes. For the *Open Area* category, SDBCS's consistency in distinguishing between diverse land cover types was evident, signifying its resilience and robustness in comparison with alternative methodologies.

Table 5. Comparative analysis of the performance of SDBCS on varying proportions of the Oberpfaffenhofen dataset (Corrected label rates %).

	0.05%			
	Build-up	**WoodLand**	**OpenArea**	**ACR**
Initial	81.82	79.84	79.13	80.01
Sel-CL	83.96	81.45	84.01	83.53
SSR	83.95	82.26	83.74	83.53
SDBCS	85.56	93.55	88.35	88.53
	0.1%			
	Build-up	**WoodLand**	**OpenArea**	**ACR**
Initial	80.39	78.24	80.43	80.01
Sel-CL	85.36	82.05	83.83	83.90
SSR	85.64	82.06	81.11	82.50
SDBCS	83.15	96.18	84.78	86.54
	0.2%			
	Build-up	**WoodLand**	**OpenArea**	**ACR**
Initial	81.40	78.23	79.96	80.00
Sel-CL	82.98	90.37	80.96	83.27
SSR	83.55	88.63	80.89	83.05
SDBCS	83.69	97.69	84.29	86.69

SDBCS's consistently superior performance, relative to Sel-CL and SSR, across categories and proportions, accentuates the method's robustness and efficiency. Its capacity to sustain high accuracy, especially evident in the *Wood Land* category, underscores its potential for precise classification even in resource-constrained scenarios.

6.2. Results and Comparisons

Figure 5 provides a visual representation of the efficacy of each method across the Flevoland area dataset 1, Oberpfaffenhofen dataset, Flevoland area dataset 2, and San Francisco dataset. Following this visual exploration, an intricate analysis aligned with the associated tables is provided. Table 6 furnishes an exhaustive evaluation of the experimental outcomes from the Flevoland area dataset 1. Our SDBCS method is benchmarked against the prevalent state-of-the-art techniques: Sel-CL, SSR, PASGS, and Auto-PASGS. The core of this evaluation revolves around classification accuracy across diverse land cover categories, elucidating the subtle yet pivotal advantages proffered by SDBCS.

Table 6. Classification performances (%) of Flevoland area dataset 1 for the proposed method.

Method	Stembeans	Peas	Forest	Lucerne	Wheat	Beet
Sel-CL	85.94	85.35	89.88	93.76	90.08	83.23
SSR	87.87	87.25	89.01	93.66	93.31	80.74
PASGS	93.77	94.66	99.26	93.01	93.82	86.85
Auto-PASGS	96.03	92.32	97.39	95.03	96.38	89.27
SDBCS	93.48	85.62	97.23	91.53	96.88	79.91
Method	**Potatoes**	**Bare Soil**	**Grass**	**Rapeseed**	**Barley**	**Wheat 2**
Sel-CL	81.48	82.09	71.01	60.42	94.49	69.82
SSR	84.33	84.73	66.56	57.22	92.27	64.69
PASGS	92.52	81.03	70.87	70.55	95.08	81.72
Auto-PASGS	95.57	34.19	80.85	75.31	95.40	78.01
SDBCS	94.29	99.94	88.53	84.00	99.68	92.03

Table 6. *Cont.*

Method	Wheat 3	Water	Building	OA	AA	Precision
Sel-CL	95.01	87.59	72.68	84.81	82.86	79.94
SSR	94.23	76.51	71.63	83.37	81.61	79.04
PASGS	95.65	90.56	40.76	89.74	85.34	84.21
Auto-PASGS	83.92	95.58	59.24	88.88	84.30	89.93
SDBCS	97.52	83.14	84.24	91.87	91.21	88.89

Method	F1-Score	Kappa	Mean IoU
Sel-CL	80.43	83.36	69.46
SSR	79.34	81.67	67.57
PASGS	84.52	88.83	75.78
Auto-PASGS	85.61	87.86	76.50
SDBCS	89.47	91.13	81.64

(a1) Sel-CL   (a2) SSR   (a3) PASGS   (a4) Auto-PASGS   (a5) SDBCS

(b1) Sel-CL   (b2) SSR   (b3) PASGS   (b4) Auto-PASGS   (b5) SDBCS

(c1) Sel-CL   (c2) SSR   (c3) PASGS   (c4) Auto-PASGS   (c5) SDBCS

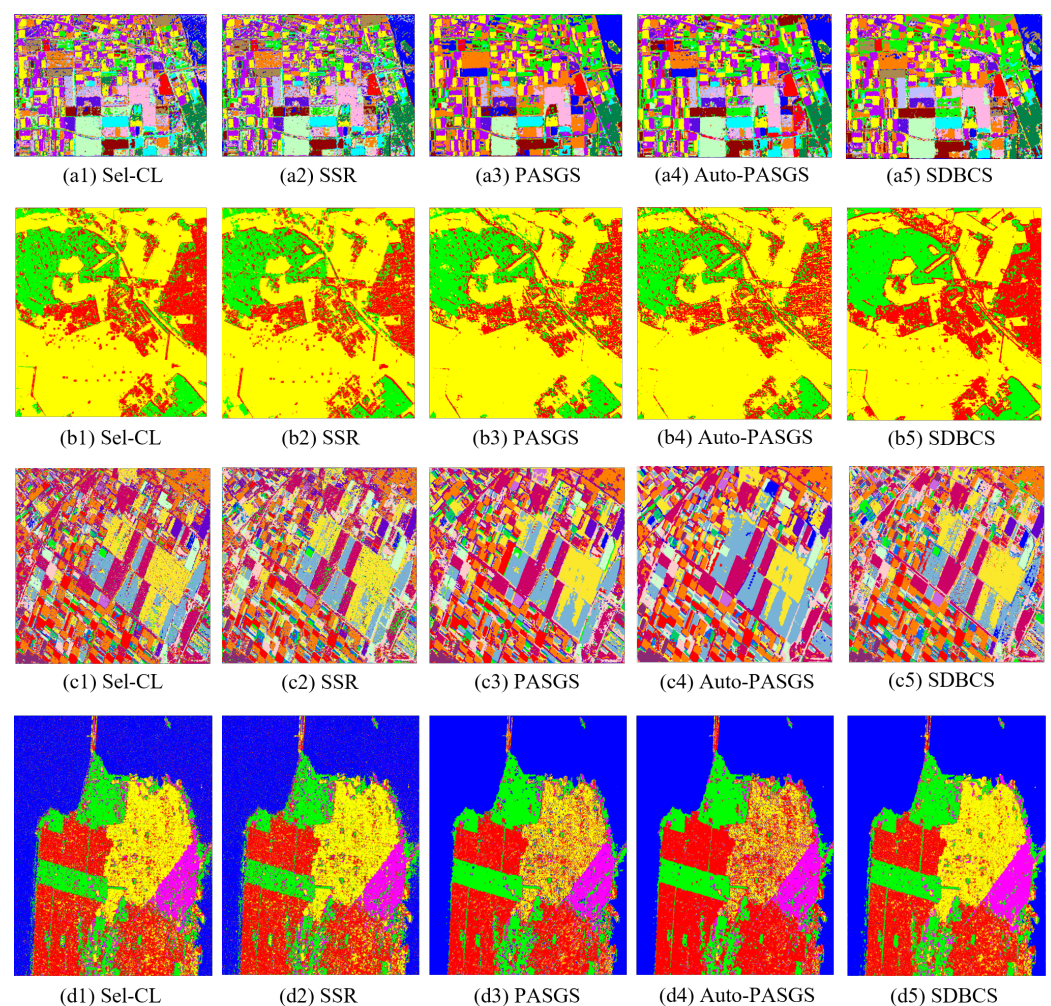(d1) Sel-CL   (d2) SSR   (d3) PASGS   (d4) Auto-PASGS   (d5) SDBCS

Figure 5. The figures presented offer a comprehensive visualization of the following: (**a1–a5**) class label predictions for the Flevoland area dataset 1, as forecasted by Sel-CL [64], SSR [65], PASGS [22], Auto-PASGS [22], and SDBCS. Subsequently, (**b1–b5**) showcases outcomes from the Oberpfaffenhofen dataset, (**c1–c5**) presents findings associated with the Flevoland area dataset 2, and, lastly, the San Francisco area dataset is elucidated in (**d1–d5**). Such systematic representation facilitates an insightful comparison and evaluation across the diverse methodologies and datasets.

Dissecting individual land cover classes reveals the consistent preeminence of SD-BCS. As a case in point, within the *Stembeans* category, SDBCS registers a commendable accuracy of 93.48%, surpassing Sel-CL and SSR, which have accuracies of 85.94% and

87.87%, respectively. Further, SDBCS achieves accuracies of 99.68% for *Barley* and 84.00% for *Rapeseed*, outperforming its competitors. This performance accentuates the capability of SDBCS to address intricate and multifaceted land cover types. Aggregating results across all classes, SDBCS achieves a commendable OA of 91.87%, overshadowing Sel-CL (84.81%), SSR (83.37%), PASGS (89.74%), and Auto-PASGS (88.88%). SDBCS not only excels in overall accuracy but also demonstrates superior performance in other metrics. It attains the highest Precision (88.89%), F1-Score (89.47%), Kappa (91.13%), and Mean IoU (81.64%).

These empirical findings highlight SDBCS's paramount stance in land cover classification, particularly amidst noise-induced challenges. Its unwavering performance across a range of land cover categories substantiates its potential to enhance the accuracy of land cover classification in remote sensing.

Table 7 illustrates the performance metrics of various agricultural land cover classification methodologies applied to the 1% Flevoland area dataset 2. SDBCS emerges as the superior method, outstripping competitors across several categories. It achieves stellar accuracy rates, exemplified by *Potatoes* (98.33%) and *Beet* (94.95%), underscoring its finesse in discerning pivotal agricultural variants. Its proficiency further extends to nuanced categories like *Oats* (92.47%) and *Barley* (81.80%). When compared with methods like Sel-CL, SSR, PASGS, and Auto-PASGS, SDBCS's superiority in accuracy remains evident. This exemplary performance, even in formidable land cover classes like *Bare Soil* (94.17%) and *Rapeseed* (96.70%), reinforces SDBCS's promise in remote sensing agricultural land cover classification. Additionally, SDBCS demonstrates robust performance in Precision (83.33%), F1-Score (85.88%), Kappa (90.47%), and Mean IoU (76.64%).

Table 8 sheds light on the performance assessment of multiple methodologies, including Sel-CL, SSR, PASGS, Auto-PASGS, and our proposed SDBCS, applied to the 0.05% San Francisco area dataset. This dataset focuses on diverse land cover classifications, including *Sea*, *Vegetation*, and three *urban* categories. The results in the table accentuate SDBCS's commendable adaptability, especially under sample-limited circumstances. While methods like Sel-CL and SSR display varying accuracies, Auto-PASGS manifests an intriguing trend, exhibiting a high accuracy for one category but faltering in others. SDBCS leads with the highest Precision (87.24%), F1-Score (88.09%), Kappa (88.50%), and Mean IoU (79.22%) SDBCS, however, consistently exhibits robustness across distinct land cover types, further cementing its efficacy in challenging classification scenarios.

Table 7. Classification performances (%) of the Flevoland area dataset 2 for the proposed method.

Method	Potatoes	Fruit	Oats	Beet	Barley	Onions
Sel-CL	87.93	90.10	73.60	91.56	82.98	42.39
SSR	85.10	89.96	64.56	92.02	84.97	59.81
PASGS	97.33	89.19	85.08	93.09	93.04	29.53
Auto-PASGS	99.57	98.01	88.95	91.18	96.33	40.47
SDBCS	98.33	89.36	92.47	94.95	81.80	75.77
Method	**Wheat**	**Beans**	**Peas**	**Maize**	**Flax**	**Rapeseed**
Sel-CL	86.57	70.89	91.20	72.25	94.12	89.87
SSR	86.45	71.90	81.99	66.59	88.17	85.19
PASGS	92.14	78.28	97.36	63.88	91.86	94.05
Auto-PASGS	80.48	22.09	100	90.47	93.75	96.15
SDBCS	91.89	72.83	99.95	83.80	92.84	96.70
Method	**Grass**	**Bare Soil**	**OA**	**AA**	**Precision**	**F1-Score**
Sel-CL	85.01	95.05	86.69	82.39	71.05	75.17
SSR	79.76	91.73	85.19	80.59	69.65	73.53
PASGS	75.95	91.67	91.64	83.75	83.84	82.50
Auto-PASGS	74.43	94.68	91.01	83.33	85.75	82.58
SDBCS	88.25	94.17	91.87	89.51	83.33	85.88

Table 7. *Cont.*

Method	Kappa	Mean IoU
Sel-CL	84.49	62.85
SSR	82.76	60.75
PASGS	90.16	73.32
Auto-PASGS	89.42	74.02
SDBCS	90.47	76.64

Table 8. Classification performances (%) of San Francisco area data for the proposed method.

Method	Sea	Vegetation	Urban 2	Urban 3	Urban 1	OA	AA
Sel-CL	91.13	78.43	75.30	82.28	81.38	84.53	81.70
SSR	91.22	75.89	75.49	81.83	81.86	84.23	81.26
PASGS	99.89	88.14	82.57	66.61	84.5	89.01	84.34
Auto-PASGS	99.70	86.98	90.54	53.40	88.42	88.41	83.81
SDBCS	99.75	86.53	80.96	86.57	93.84	92.00	89.53

Method	Precision	F1-Score	Kappa	Mean IoU
Sel-CL	76.63	78.60	78.19	65.74
SSR	76.14	78.14	77.76	65.14
PASGS	84.18	83.87	84.07	73.13
Auto-PASGS	83.66	82.29	83.25	71.35
SDBCS	87.24	88.09	88.50	79.22

Table 9 provides an analytical performance overview of diverse methodologies on the 0.05% Oberpfaffenhofen area dataset. Our model, SDBCS, consistently excels, outpacing counterparts in pivotal categories. Illustratively, in the *Build-up* category, SDBCS achieves an accuracy of 79.08%, superseding Sel-CL's 69.32%. In the *Wood Land* category, SDBCS's accuracy peaks at 89.12%, transcending SSR's 78.47%. Notably, in the OA metric, SDBCS's performance at 88.48% distinctly eclipses Sel-CL's 84.55% and SSR's 82.63%. SDBCS's prowess becomes manifest in the AA metric, with an accuracy of 86.86%, surpassing both SSR's 77.99% and PASGS's 78.70%. Furthermore, the superiority of SDBCS is underlined by its leading Precision of 85.88%, F1-Score of 86.35%, Kappa of 80.58%, and Mean IoU of 76.49%.

In summation, these empirical outcomes robustly underscore the innate capability of SDBCS to adeptly navigate challenges engendered by noisy labels, restricted sample sizes, and a gamut of land cover classifications. The intrinsic proficiency of SDBCS in rectifying label inaccuracies and capitalizing on limited annotations underscores its pivotal role in the evolutionary trajectory of research within remote sensing, with a particular emphasis on land cover classification endeavors.

6.3. Limitations and Enhancements

In Table 3, when the VGGNet-8 model was enhanced with our label correction module, there was a drop in prediction accuracy for the built-up land type. This decline can be linked to the unique properties of built-up areas in the Oberpfaffenhofen dataset, which are characterized by complex spatial structures and varied spectral signatures. The label correction process involves aligning each training sample with the label of its nearest k training data points based on feature distance. However, due to the spectral resemblance of some built-up areas to other land types, mislabeling may occur. This challenge is intrinsic to handling complex urban environments in remote sensing imagery.

While our method effectively addresses noisy labels, its performance might be influenced by the quality of the feature representations. If the feature extraction process fails to adequately distinguish between different classes, the label correction might not be as effective.

Table 9. Classification performances (%) of Oberpfaffenhofen area data for the proposed method.

Method	Build-up	Wood Land	Open Area	OA	AA	Precision
Sel-CL	69.32	79.14	93.13	84.55	80.53	81.46
SSR	62.66	78.47	92.84	82.63	77.99	79.42
PASGS	61.20	79.55	95.36	83.89	78.70	80.42
Auto-PASGS	59.24	75.99	95.87	82.99	77.03	79.73
SDBCS	79.08	89.12	92.38	88.48	86.86	85.88
Method	**F1-Score**	**Kappa**	**Mean IoU**			
Sel-CL	80.96	73.51	68.83			
SSR	78.57	70.00	65.73			
PASGS	79.29	72.07	66.88			
Auto-PASGS	78.10	70.27	65.40			
SDBCS	86.35	80.58	76.49			

To improve the performance in boundary areas and in general, we could consider integrating additional context-aware mechanisms. For instance, incorporating attention mechanisms could enable the model to focus on more relevant features, thereby improving the accuracy of the label correction, especially in complex regions. Another potential enhancement is to use multiscale feature representations. This approach could help in capturing both fine-grained details and broader contextual information, thereby improving the model's ability to handle diverse and challenging scenarios, including boundary regions.

7. Conclusions

Confronting the complexities of PolSAR image classification, our study introduces a novel label correction approach, designed for managing noisy labels, and leverages unsupervised contrastive learning to enhance polarimetric representation ability and further classification accuracy in label scarcity scenarios. The innovative label correction technique we developed employs similarities among training samples with a feature distance matrix derived from contrastive learning, which identifies and rectifies mislabeled samples, thereby addressing the noisy label issue. In addition, by adopting self-supervised representation learning, we significantly enhance the model's robustness and accuracy, especially in the context of limited labels in PolSAR image classification. Our method also includes strategic rebalancing and data augmentation techniques to tackle the class imbalance problem, improving the classification accuracy of minority classes. Extensive evaluations on four benchmark datasets have proven the effectiveness and superiority of the proposed method. To sum up, our approach effectively improves the accuracy and robustness of DNN-based PolSAR image classification methods in noisy and sparse label scenarios, addressing the initial challenges we set out to overcome.

Author Contributions: Conceptualization, N.W. and H.B.; Methodology, N.W. and H.B.; Software, N.W.; Validation, N.W., H.B. and F.L.; Formal analysis, H.B.; Investigation, N.W., H.B. and C.X.; Resources, H.B., F.L., C.X. and J.G.; Data curation, H.B., C.X. and J.G.; Writing—original draft, N.W.; Writing—review & editing, H.B.; Visualization, H.B., F.L. and C.X.; Supervision, H.B., F.L., C.X. and J.G.; Project administration, H.B. and J.G.; Funding acquisition, H.B. and J.G. All authors have read and agreed to the published version of the manuscript.

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 conflict of interest.

References

1. 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]
2. Lee, J.S.; Grunes, M.R.; Ainsworth, T.L.; Du, L.J.; Schuler, D.L.; Cloude, S.R. Unsupervised classification using polarimetric decomposition and the complex Wishart classifier. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 2249–2258.
3. Van Zyl, J.J. Unsupervised classification of scattering behavior using radar polarimetry data. *IEEE Trans. Geosci. Remote Sens.* **1989**, *27*, 36–45. [CrossRef]
4. Bi, H.; Sun, J.; Xu, Z. A graph-based semisupervised deep learning model for PolSAR image classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 2116–2132. [CrossRef]
5. Yu, P.; Qin, A.K.; Clausi, D.A. Unsupervised polarimetric SAR image segmentation and classification using region growing with edge penalty. *IEEE Trans. Geosci. Remote Sens.* **2011**, *50*, 1302–1317. [CrossRef]
6. Chen, Q.; Kuang, G.; Li, J.; Sui, L.; Li, D. Unsupervised land cover/land use classification using PolSAR imagery based on scattering similarity. *IEEE Trans. Geosci. Remote Sens.* **2012**, *51*, 1817–1825. [CrossRef]
7. Tu, S.T.; Chen, J.Y.; Yang, W.; Sun, H. Laplacian eigenmaps-based polarimetric dimensionality reduction for SAR image classification. *IEEE Trans. Geosci. Remote Sens.* **2011**, *50*, 170–179. [CrossRef]
8. Ersahin, K.; Scheuchl, B.; Cumming, I. Incorporating texture information into polarimetric radar classification using neural networks. In Proceedings of the IGARSS 2004, 2004 IEEE International Geoscience and Remote Sensing Symposium, Anchorage, AK, USA, 20–24 September 2004; Volume 1.
9. Wu, Y.; Ji, K.; Yu, W.; Su, Y. Region-based classification of polarimetric SAR images using Wishart MRF. *IEEE Geosci. Remote Sens. Lett.* **2008**, *5*, 668–672. [CrossRef]
10. Bi, H.; Yao, J.; Wei, Z.; Hong, D.; Chanussot, J. PolSAR Image Classification Based on Robust Low-Rank Feature Extraction and Markov Random Field. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 4005205. [CrossRef]
11. Bi, H.; Xu, F.; Wei, Z.; Xue, Y.; Xu, Z. An active deep learning approach for minimally supervised PolSAR image classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9378–9395. [CrossRef]
12. Xu, Y.; Li, Z.; Li, W.; Du, Q.; Liu, C.; Fang, Z.; Zhai, L. Dual-channel residual network for hyperspectral image classification with noisy labels. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5502511. [CrossRef]
13. Xiao, T.; Xia, T.; Yang, Y.; Huang, C.; Wang, X. Learning from massive noisy labeled data for image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 2691–2699.
14. Lee, K.H.; He, X.; Zhang, L.; Yang, L. Cleannet: Transfer learning for scalable image classifier training with label noise. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5447–5456.
15. Goldberger, J.; Ben-Reuven, E. Training deep neural-networks using a noise adaptation layer. In Proceedings of the International Conference on Learning Representations, San Juan, Puerto Rico, 2–4 May 2016.
16. Han, J.; Luo, P.; Wang, X. Deep self-learning from noisy labels. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 5138–5147.
17. Kim, Y.; Yim, J.; Yun, J.; Kim, J. Nlnl: Negative learning for noisy labels. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 101–110.
18. Ma, X.; Huang, H.; Wang, Y.; Romano, S.; Erfani, S.; Bailey, J. Normalized loss functions for deep learning with noisy labels. In Proceedings of the International Conference on Machine Learning, PMLR, Virtual, 13–18 July 2020; pp. 6543–6553.
19. Grill, J.B.; Strub, F.; Altché, F.; Tallec, C.; Richemond, P.; Buchatskaya, E.; Doersch, C.; Avila Pires, B.; Guo, Z.; Gheshlaghi Azar, M.; et al. Bootstrap your own latent-a new approach to self-supervised learning. In Proceedings of the Advances in Neural Information Processing Systems, Virtual, 6–12 December 2020; Volume 33, pp. 21271–21284.
20. Li, Y.; Xing, R.; Jiao, L.; Chen, Y.; Chai, Y.; Marturi, N.; Shang, R. Semi-supervised PolSAR image classification based on self-training and superpixels. *Remote Sens.* **2019**, *11*, 1933. [CrossRef]
21. Han, B.; Yao, Q.; Yu, X.; Niu, G.; Xu, M.; Hu, W.; Tsang, I.; Sugiyama, M. Co-teaching: Robust training of deep neural networks with extremely noisy labels. In Proceedings of the Advances in Neural Information Processing Systems, Montréal, QC, Canada, 3–8 December 2018; Volume 31.
22. 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]
23. Hou, B.; Wu, Q.; Wen, Z.; Jiao, L. Robust semisupervised classification for PolSAR image with noisy labels. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 6440–6455. [CrossRef]
24. 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]
25. 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*, 4512605. [CrossRef]
26. 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*, 5100316. [CrossRef]
27. 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; pp. 2529–2533.

28. 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; pp. 1061–1064.

29. Hu, J.; Hong, D.; Zhu, X.X. MIMA: MAPPER-induced manifold alignment for semi-supervised fusion of optical image and polarimetric SAR data. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9025–9040. [CrossRef]

30. 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*, 5235716. [CrossRef]

31. Wei, B.; Yu, J.; Wang, C.; Wu, H.; Li, J. PolSAR image classification using a semi-supervised classifier based on hypergraph learning. *Remote Sens. Lett.* **2014**, *5*, 386–395. [CrossRef]

32. Liu, W.; Yang, J.; Li, P.; Han, Y.; Zhao, J.; Shi, H. A novel object-based supervised classification method with active learning and random forest for PolSAR imagery. *Remote Sens.* **2018**, *10*, 1092. [CrossRef]

33. Qin, X.; Yang, J.; Zhao, L.; Li, P.; Sun, K. A Novel Deep Forest-Based Active Transfer Learning Method for PolSAR Images. *Remote Sens.* **2020**, *12*, 2755. [CrossRef]

34. Doz, C.; Ren, C.; Ovarlez, J.P.; Couillet, R. Large Dimensional Analysis of LS-SVM Transfer Learning: Application to Polsar Classification. In Proceedings of the ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4–10 June 2023; pp. 1–5.

35. Nie, W.; Huang, K.; Yang, J.; Li, P. A deep reinforcement learning-based framework for PolSAR imagery classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 4403615. [CrossRef]

36. 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]

37. Cui, Y.; Liu, F.; Liu, X.; Li, L.; Qian, X. TCSPANET: Two-staged contrastive learning and sub-patch attention based network for polsar image classification. *Remote Sens.* **2022**, *14*, 2451. [CrossRef]

38. Wu, Z.; Xiong, Y.; Yu, S.X.; Lin, D. Unsupervised feature learning via non-parametric instance discrimination. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3733–3742.

39. Caron, M.; Touvron, H.; Misra, I.; Jégou, H.; Mairal, J.; Bojanowski, P.; Joulin, A. Emerging properties in self-supervised vision transformers. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 9650–9660.

40. Ghosh, A.; Kumar, H.; Sastry, P.S. Robust loss functions under label noise for deep neural networks. In Proceedings of the AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; Volume 31.

41. Wang, Y.; Ma, X.; Chen, Z.; Luo, Y.; Yi, J.; Bailey, J. Symmetric cross entropy for robust learning with noisy labels. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 322–330.

42. Jiang, L.; Zhou, Z.; Leung, T.; Li, L.J.; Fei-Fei, L. Mentornet: Learning data-driven curriculum for very deep neural networks on corrupted labels. In Proceedings of the International Conference on Machine Learning, PMLR, Stockholm, Sweden, 10–15 July 2018; pp. 2304–2313.

43. Yu, X.; Han, B.; Yao, J.; Niu, G.; Tsang, I.; Sugiyama, M. How does disagreement help generalization against label corruption? In Proceedings of the International Conference on Machine Learning, PMLR, Long Beach, CA, USA, 9–15 June 2019; pp. 7164–7173.

44. Berthelot, D.; Carlini, N.; Goodfellow, I.; Papernot, N.; Oliver, A.; Raffel, C.A. Mixmatch: A holistic approach to semi-supervised learning. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; Volume 32.

45. Edwards, H.; Storkey, A. Towards a Neural Statistician. In Proceedings of the International Conference on Learning Representations, Toulon, France, 24–26 April 2017.

46. Kaiser, Ł.; Nachum, O.; Roy, A.; Bengio, S. Learning to remember rare events. *arXiv* **2017**, arXiv:1703.03129.

47. Ravi, S.; Larochelle, H. Optimization as a model for few-shot learning. In Proceedings of the International Conference on Learning Representations, San Juan, Puerto Rico, 2–4 May 2016.

48. Bertinetto, L.; Henriques, J.F.; Valmadre, J.; Torr, P.; Vedaldi, A. Learning feed-forward one-shot learners. In Proceedings of the Advances in Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; Volume 29.

49. Santoro, A.; Bartunov, S.; Botvinick, M.; Wierstra, D.; Lillicrap, T. Meta-learning with memory-augmented neural networks. In Proceedings of the International Conference on Machine Learning, PMLR, New York, NY, USA, 20–22 June 2016; pp. 1842–1850.

50. Snell, J.; Swersky, K.; Zemel, R. Prototypical networks for few-shot learning. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.

51. Finn, C.; Abbeel, P.; Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. In Proceedings of the International Conference on Machine Learning, PMLR, Sydney, Australia, 6–11 August 2017; pp. 1126–1135.

52. Lake, B.; Salakhutdinov, R.; Gross, J.; Tenenbaum, J. One shot learning of simple visual concepts. In Proceedings of the Annual Meeting of the Cognitive Science Society, Boston, MA, USA, 20–23 July 2011; Volume 33.

53. Oord, A.v.d.; Li, Y.; Vinyals, O. Representation learning with contrastive predictive coding. *arXiv* **2018**, arXiv:1807.03748.

54. Tian, Y.; Krishnan, D.; Isola, P. Contrastive multiview coding. In *Computer Vision–ECCV 2020, Proceedings of the 16th European Conference, Glasgow, UK, 23–28 August 2020*; Proceedings, Part XI 16; Springer: Berlin/Heidelberg, Germany, 2020; pp. 776–794.

55. He, K.; Fan, H.; Wu, Y.; Xie, S.; Girshick, R. Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 9729–9738.

56. Cloude, S.R.; Pottier, E. A review of target decomposition theorems in radar polarimetry. *IEEE Trans. Geosci. Remote Sens.* **1996**, *34*, 498–518. [CrossRef]

57. Zhou, Y.; Wang, H.; Xu, F.; Jin, Y.Q. Polarimetric SAR image classification using deep convolutional neural networks. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1935–1939. [CrossRef]

58. Hernández-García, A.; König, P. Do deep nets really need weight decay and dropout? *arXiv* **2018**, arXiv:1802.07042.

59. Wong, S.C.; Gatt, A.; Stamatescu, V.; McDonnell, M.D. Understanding data augmentation for classification: When to warp? In Proceedings of the 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Gold Coast, Australia, 30 November–2 December 2016; pp. 1–6.

60. De Boer, P.T.; Kroese, D.P.; Mannor, S.; Rubinstein, R.Y. A tutorial on the cross-entropy method. *Ann. Oper. Res.* **2005**, *134*, 19–67. [CrossRef]

61. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.

62. Müller, R.; Kornblith, S.; Hinton, G.E. When does label smoothing help? In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; Volume 32.

63. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.

64. Li, S.; Xia, X.; Ge, S.; Liu, T. Selective-supervised contrastive learning with noisy labels. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 316–325.

65. Wang, Y.; Sun, X.; Fu, Y. Scalable penalized regression for noise detection in learning with noisy labels. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 346–355.