electronics

Article

Collaborative Self-Supervised Transductive Few-Shot Learning for Remote Sensing Scene Classification

Haiyan Han [1], Yangchao Huang [1] and Zhe Wang [2,*]

[1] Information and Navigation School, Air Force Engineering University, Xi'an 710038, China; haiyanhanAFEU@gmail.com (H.H.); gxyxhbwhyc@sohu.com (Y.H.)
[2] Air Defense and Antimissile School, Air Force Engineering University, Xi'an 710038, China
* Correspondence: zhewangafeu@gmail.com

Abstract: With the advent of deep learning and the accessibility of massive data, scene classification algorithms based on deep learning have been extensively researched and have achieved exciting developments. However, the success of deep models often relies on a large amount of annotated remote sensing data. Additionally, deep models are typically trained and tested on the same set of classes, leading to compromised generalization performance when encountering new classes. This is where few-shot learning aims to enable models to quickly generalize to new classes with only a few reference samples. In this paper, we propose a novel collaborative self-supervised transductive few-shot learning (CS^{2}TFSL) algorithm for remote sensing scene classification. In our approach, we construct two distinct self-supervised auxiliary tasks to jointly train the feature extractor, aiming to obtain a powerful representation. Subsequently, the feature extractor's parameters are frozen, requiring no further training, and transferred to the inference stage. During testing, we employ transductive inference to enhance the associative information between the support and query sets by leveraging additional sample information in the data. Extensive comparisons with state-of-the-art few-shot scene classification algorithms on the WHU-RS19 and NWPU-RESISC45 datasets demonstrate the effectiveness of the proposed CS^{2}TFSL. More specifically, CS^{2}TFSL ranks first in the settings of five-way one-shot and five-way five-shot. Additionally, detailed ablation experiments are conducted to analyze the CS^{2}TFSL. The experimental results reveal significant and promising performance improvements in few-shot scene classification through the combination of self-supervised learning and direct transductive inference.

Keywords: few-shot learning; remote sensing scene classification; transductive inference; self-supervised learning

Citation: Han, H.; Huang, Y.; Wang, Z. Collaborative Self-Supervised Transductive Few-Shot Learning for Remote Sensing Scene Classification. *Electronics* **2022**, *12*, 3846. https://doi.org/10.3390/electronics12183846

Academic Editor: Chiman Kwan

Received: 20 August 2023
Revised: 6 September 2023
Accepted: 6 September 2023
Published: 11 September 2023

1. Introduction

Remote sensing scene classification refers to the task of identifying and categorizing different objects or scenes in remote sensing imagery [1]. It plays a crucial role in geological exploration [2], environmental monitoring [3], urban planning [4], and disaster monitoring [5]. By performing scene classification, we can obtain vital geographic information to support decision making and resource management, thereby promoting sustainable development and the construction of smart cities [6].

With the advancement of deep learning and the widespread availability of large-scale remote sensing datasets, the accuracy of remote sensing scene classification has been significantly enhanced. Deep learning-based scene classification of remote sensing images is a technique that utilizes convolutional neural networks [7], attention mechanisms [8], recurrent neural networks [9], and other methods to automatically classify and identify objects and scenes in remote sensing images [10,11]. It effectively captures spatial and semantic information by automatically learning and extracting image features, resulting in more accurate and efficient scene classification [12–14]. Additionally, deep learning-based

methods possess powerful feature learning and complex relationship modeling capabilities, making them suitable for classifying different scales and complex scene requirements [15]. Through training on large-scale data and transfer learning, this approach can improve the performance of classification models and provide significant support for geographic information analysis and applications [16].

However, traditional deep learning models rely on a large amount of annotated data for training in order to accurately capture the features and relationships between different categories [17]. In practical applications, acquiring a large number of annotated samples can be challenging and costly [18,19]. Additionally, traditional deep learning models use the same categories for training and testing, which leads to poor generalization when encountering new categories. This is because the model has not been exposed to annotated samples of these new categories and cannot accurately classify and recognize them [20].

Therefore, few-shot learning has become crucial in remote sensing scene classification [21]. Few-shot learning is a task that aims to learn informative and discriminative feature representations from only a limited number of labeled examples [22]. Through techniques such as transfer learning and meta-learning, the knowledge acquired can be applied to new categories for accurate classification [23]. This approach helps overcome the problem of insufficient annotated data and improves the model's generalization ability on new categories [24]. Currently, there has been extensive research on few-shot remote sensing scene classification, which can generally be categorized into two different training paradigms: meta-learning-based [25–27] and transfer-learning-based approaches [28,29]. Among them, improving the metric space is a key focus of current research by scholars [30]. However, current research has found that a good feature extractor is the key to few-shot classification [30–32]. An effective feature extractor has a much more significant impact on the performance than a cleverly designed metric space.

In light of this indication, in this paper, we propose a novel collaborative self-supervised transductive few-shot learning (CS^{2}TFSL) algorithm for remote sensing scene classification. In CS^{2}TFSL, we train the feature extractor by leveraging two distinct self-supervised auxiliary tasks collaboratively to obtain a powerful feature representation. Unlike traditional few-shot learning methods that rely on labeled training data, self-supervised learning does not require any labels. Instead, it constructs pretext tasks and generates pseudo-labels for training. Afterwards, the trained feature extractor requires no further training, and its parameters are frozen, allowing for direct transfer to the inference stage. During the testing stage, we employ transductive inference, enhancing the associative information between the support and query sets by incorporating additional sample information from the dataset. Extensive comparisons with state-of-the-art (SOTA) few-shot scene classification algorithms demonstrate the effectiveness of the CS^{2}TFSL. Additionally, we conduct detailed ablation experiments to analyze the components of CS^{2}TFSL. The experimental results highlight significant and promising performance improvements in few-shot scene classification achieved through the combination of self-supervised learning and transductive inference.

Overall, our contributions can be summarized in the following two aspects:

(1) We propose a collaborative self-supervised training strategy for few-shot scene classification. Based on the transfer learning paradigm, our approach allows the model to directly freeze and perform inference after self-supervised training, eliminating the need for complex episode-based training procedures in metric learning.
(2) We introduce the simple soft k-means (SKM) classifier for the few-shot transductive inference. Compared to inductive inference, our model can benefit from more sample references for inference, and it is also simpler compared to the complex graph neural network of transductive inference.

The remaining sections of this article are structured as follows. Section 2 introduces some relevant studies on few-shot remote sensing classification and self-supervised learning. In Section 3, the proposed method is described in detail. The experimental results are reported and discussed in Section 4 and Section 5. The conclusion is presented in Section 6.

2. Related Works

2.1. Few-Shot Remote Sensing Scene Classification

According to the categorization in [28,29], current research on few-shot remote sensing scene classification can be divided into two categories: methods based on meta-learning and transfer learning. The main difference between them lies in the training paradigm. Taking the N-way K-shot scenario as an example, algorithms based on meta-learning utilize the episodic training approach. In each episode, N remote sensing categories are selected, and K images from each category are chosen as the support set for training the network. On the other hand, algorithms based on transfer learning do not use explicit support and query sets during training, instead, all categories are used as input for training. During the testing phase, both methods follow the same procedure. The testing set is divided into multiple episodes according to the N-way K-shot scenario. It is worth noting that there is no overlap in categories between the training and test set [33,34].

In terms of meta-learning-based few-shot scene classification algorithms, Zhai et al. [27] explored the application of lifelong learning techniques for scene recognition in remote sensing images. Lifelong learning refers to the ability of a machine learning model to continuously learn and adapt from new data throughout its lifetime. Cheng et al. [25] leveraged a Siamese Prototypical Network (SPNet), where two branches share weights and learn to extract features from both support and query images, with prototype self-calibration and inter-calibration. Ji et al. [26] addressed the challenges of data scarcity and domain shift. The proposed method aims to improve the performance of few-shot scene classification models by leveraging auxiliary objectives and incorporating transductive inference. Li et al. [35] combined the benefits of deep feature extraction and metric learning to improve few-shot classification performance in remote sensing images. The key components of DLA-MatchNet [35] include a deep layer aggregation backbone network and a matching network. Zeng et al. [36] incorporated an iterative process that progressively refines the feature representations and classification boundaries. It consists of two main components: an embedding network and an iterative distribution learning strategy. Huang et al. [37] proposed a task-adaptive attention component, which combined the meta-learning training mechanism and graph neural network transductive inference in few-shot scene classification.

In terms of transfer learning based few-shot scene classification algorithms, Gong et al. [28] proposed the two-path aggregation attention network with quad-patch data augmentation, which improves the few-shot scene classification performance from both the perspectives of data and network structure. Li et al. [29] proposed a multiform ensemble enhancement strategy to explore the effects of different self-supervised auxiliary tasks on the feature extraction performance.

Self-Supervised Learning

Self-supervised learning is a method that utilizes the inherent information within data for unsupervised learning, with the goal of acquiring useful representations or features [38–40]. It can learn from large-scale unlabeled data, reducing reliance on annotated data, and is applicable in scenarios with limited data or challenging annotation conditions. Additionally, it can be used as a form of pretraining, where learned features are fine-tuned to enhance performance on other tasks [29,41].

In the field of remote sensing, self-supervised learning has been extensively applied to various specific tasks and has achieved promising results [42]. For example, self-supervised learning has been widely explored in various domains such as hyperspectral unmixing [43], change detection [44], SAR target recognition [45], and hyperspectral classification [46].

In terms of few-shot remote sensing scene classification, Zhai et al. [47] applied the Bootstrap Your Own Latent (BYOL) [48] contrastive learning algorithm to feature extraction in the context of few-shot scene classification, with the aim of obtaining better sample representations. Li et al. [29] explored the impact of different self-supervised auxiliary tasks combined in various forms on feature extraction for few-shot scene classification.

Gong et al. [43] employed the SimCLR [49] to explore the transfer relationship between the source domain and target domain in few-shot scene classification scenarios. In [50], the Meta-FSEO approach is proposed for few-shot remote sensing scene classification. Specifically, it is achieved through self-supervised embedding optimization. This method aims to train a model using a meta-learning framework that combines self-supervised learning and meta-learning concepts to improve the model's generalization performance and adaptability by optimizing embedding representations. In terms of the current research, the exploration of self-supervised few-shot scene classification is relatively insufficient, and most studies focus on inductive inference, with only a few studies focusing on transductive inference.

3. Methodology

3.1. Overview

Few-shot classification refers to the task of training a model with very limited samples per class, in order to accurately classify new unseen samples [13,19]. It is a demanding task that enables the model to have strong generalization ability, extracting representative features and patterns from a small number of samples and applying them to classify unseen samples. Generally, the dataset in few-shot learning is divided into a base dataset and a novel dataset, and it is important to note that these two datasets have no overlapping classes. We adopt a training paradigm based on transfer learning, which means that in the training phase, the base dataset is not divided into multiple episodes, but all classes are directly trained together like traditional classification algorithms. However, during testing, the novel dataset is divided into multiple episodes for inference. To explain our point in a more straightforward manner, we take the common scenario of an N-way K-shot as an example. First, in constructing the support set and query set, N classes are selected from the novel dataset, and K labeled samples are provided for each class as the support set. The query set consists of unlabeled samples from these N classes. The goal is to classify the samples in the query set into their respective classes with the assistance of a small number of reference samples. Few-shot scene classification presents a challenge due to the scarcity of training samples available for each class. This necessitates the model to learn and generalize from a small number of samples to accurately classify the unseen samples.

We propose the CS^{2}TFSL framework as illustrated in Figure 1. Initially, the feature extractor is trained by incorporating two collaborative self-supervised learning tasks, namely rotation and spatial contrastive learning (SCL), alongside the original semantic class prediction task. Once the training is complete, the feature extractor is fixed and directly applied during the inference phase. The features of both the support set and query set samples are first extracted, followed by using transductive inference to determine the class labels of the query samples.

3.2. Collaborative Self-Supervised Learning

In the training phase of the feature extractor, the total loss function L_T can include the following two components:

$$L_T = L_{CE} + L_{SSL}, \tag{1}$$

where L_{CE} and L_{SSL} represent the loss functions for semantic class prediction and self-supervised prediction, respectively.

The cross-entropy loss function, denoted as L_{CE}, is widely employed in classification tasks to quantify the disparity between the predicted class probabilities and the actual class labels. It serves as a measure of the model's accuracy in assigning the correct class. Mathematically, it can be defined as follows:

$$L_{CE} = -\sum(y * log(\hat{y})), \tag{2}$$

where y is the true class label, and p is the predicted class probabilities. The cross-entropy loss function penalizes significant disparities between the predicted probabilities and the

true label. By doing so, it motivates the model to reduce the differences and enhance its prediction accuracy. In essence, it encourages the model to better align its predictions with the true labels, leading to improved performance.

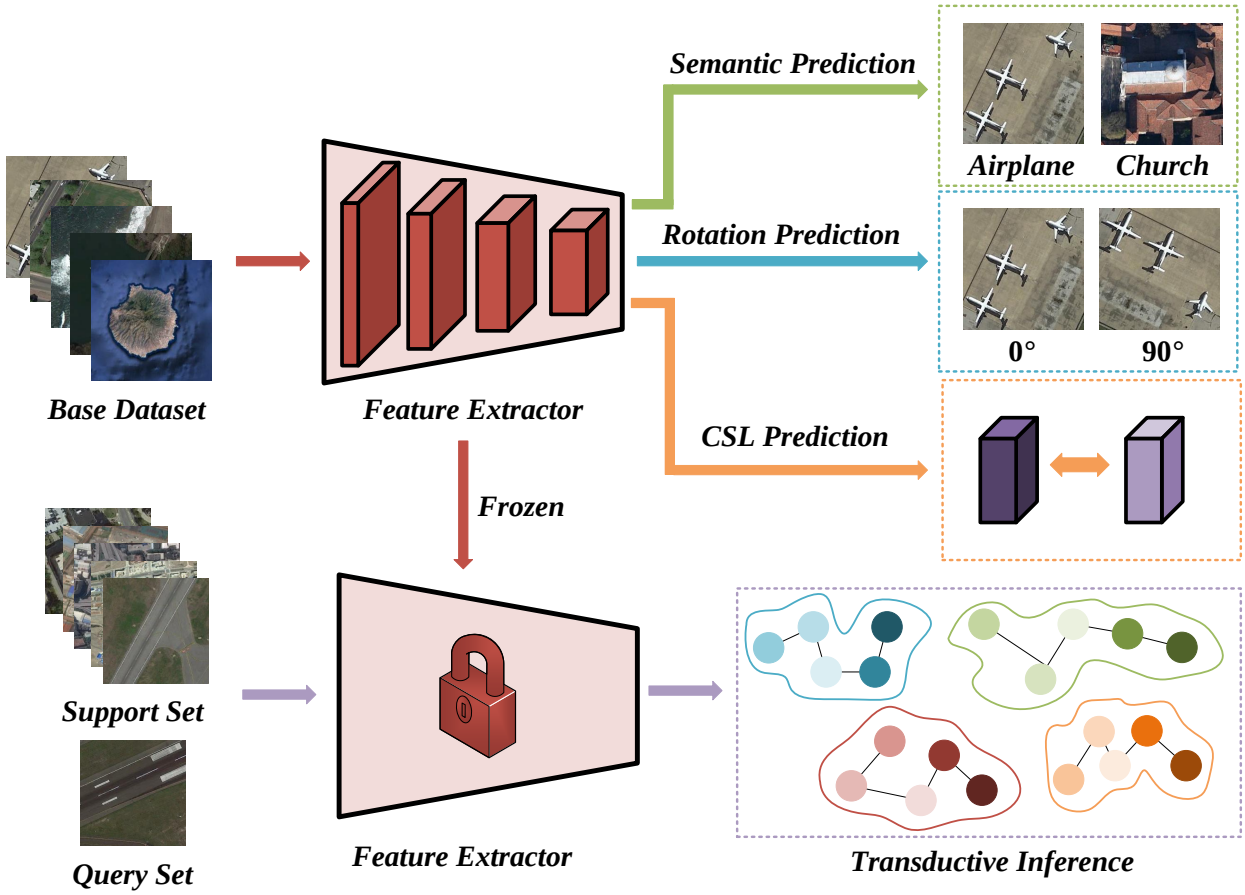


Figure 1. The framework of the proposed CS^{2}TFSL. Initially, the feature extractor is trained using three tasks: rotation, spatial contrastive learning, and semantic class prediction. These tasks are designed to improve the performance and generalization ability of the feature extractor through collaborative self-supervised learning. After the training phase, the feature extractor is frozen and directly deployed to the inference phase. During inference, both the support set and query set samples undergo feature extraction. The class labels of the query samples are then determined using transductive inference. We present a case study of five-way one-shot .

L_{SSL} is a composite loss function formed by two self-supervised auxiliary tasks. Its definition is as follows:

$$L_{SSL} = \lambda_R * L_R + \lambda_{SCL} * L_{SCL}, \tag{3}$$

where L_R is the loss function for the rotation pretext task. L_{SCL} is the loss function for the spatial contrastive learning pretext task. λ_R and λ_{SCL} are weight parameters that control the importance of each task in the overall loss. Each task aims to provide additional learning signals to improve the feature extractor's representation capabilities. The weights λ_R and λ_{SCL} allow for adjusting the relative impact of each task on the overall training process.

The purpose of the rotation prediction task is to determine the specific 2-D rotation transformation applied to an input image. In contrast to using semantic class labels, the supervision signal in this task is independent of any particular category, which facilitates information exchange across different classes. Additionally, using rotation as a supervisory signal assumes that the neural network has already acquired knowledge about object categories and learned about their constituent parts before being able to perform accurate rotation recognition. Through training the model to predict the precise rotation for an

image, the neural network implicitly learns to extract robust and discriminative features that capture relevant information about objects and their spatial relationships. This approach promotes a deeper understanding of objects and refines the model's ability to distinguish between different classes. We define a collection of rotation operators as $\Theta = \{\phi_\omega\}_{\omega=1}^{\Omega}$, where $x_\omega = \phi_\omega(x)$ represents the rotated image, and Ω is the total number of rotations. Given the parameters $\Delta_f = [\alpha_1, \alpha_2, ..., \alpha_R]$ of the rotation classifier, the probability for the input image x_j is given by:

$$p(\hat{y}_k^\phi = \omega | x_k) = \frac{\exp(\alpha_\omega^T f_\theta(\phi_\omega(x_k)))}{\sum_{\omega=1}^{\Omega} \exp(\alpha_\omega^T f_\theta(\phi_\omega(x_k)))}. \tag{4}$$

where f_θ represents the feature extractor.

Based on this, the L_R can be defined as follows:

$$L_R = -\sum_{k=1}^{B} \sum_{\omega}^{\Omega} \prod (y_k^\phi = \omega) \log(p(\hat{y}_k^\phi = \omega | x_k)). \tag{5}$$

where B represents the batchsize, and y_k and $\hat{y}_k$ denote the rotated ground-truth pseudo-labels and the predictions made by the model, respectively.

Spatial contrastive learning utilizes the concept of spatial similarity to evaluate the degree of similarity or agreement between two different feature representations. In simpler terms, when conducting spatial contrastive learning, we use a measure called "spatial similarity" to determine how similar or different two feature representations are from each other. This approach is integral in assessing the similarity and agreement between different features. This similarity captures the spatial relationship or alignment between the features extracted from different views or augmentations of the same image. The spatial similarity metric helps in determining the degree of correspondence or match between the extracted features, which is crucial for the contrastive loss function used in self-supervised learning. The definition of spatial similarity $sim(s_i, s_m)$ between two features is consistent with [51]. Following this lead, the spatial contrastive learning loss function can be defined as:

$$L_{SCL} = \sum_{i=1}^{2B} \frac{1}{2B_{y_i} - 1} \sum_{j=1}^{2B} \mathbb{1}_{i \neq j} \cdot \mathbb{1}_{y_i = y_j} \cdot \ell_{ij}, \tag{6}$$

$$\ell_{ij} = -\log \frac{\exp(sim(s_i, s_j)/\tau)}{\sum_{m=1}^{2B} \mathbb{1}_{i \neq m} \exp(sim(s_i, s_m)/\tau)}, \tag{7}$$

where $\mathbb{1}_{condition} \in \{0, 1\}$ is an indicator function, which takes a value of 1 only if the specified condition is satisfied. B_{y_i} represents the total number of images within a dataset that possess the same label y_i. Moreover, τ denotes the scalar temperature parameter.

3.3. Transductive Inference

In few-shot transductive inference, a classifier that operates with limited training examples has access to both the support dataset (containing labeled examples) and the complete query dataset (unlabeled examples) during the prediction phase. In our approach, we utilize a straightforward soft k-means (SKM) classifier [30] for transductive inference. This classifier divides the query set into smaller subsets, each associated with a small number of labeled support sets. After the previous training is completed, we freeze the feature extractor and proceed directly to inference. Firstly, we obtain the sample feature representations $\mathcal{K}$ for the support set and query set. Next, we need to preprocess the vectors by centering them and projecting them onto a hypersphere, as shown below:

$$\mathcal{K}_C = \frac{\mathcal{K} - \widetilde{\mathcal{K}}}{||\mathcal{K} - \widetilde{\mathcal{K}}||_2}, \tag{8}$$

where the $\widetilde{\mathcal{K}}$ is the average feature of the support set.

First, given that the barycenters in transductive inference require recalculation on each occasion, where t is the index of the sequence, the initialization of $\widetilde{N}_t$ can be computed as follows:

$$\widetilde{N}_t^{u+1} = \sum_{\mathcal{K}_C \in \mathcal{S}_k \cup \mathcal{Q}} \frac{w(\mathcal{K}_C, \widetilde{N}_t^u)}{\sum_{\mathcal{K}_C' \in \mathcal{S}_k \cup \mathcal{Q}} w(\mathcal{K}_C', \widetilde{N}_t^u)} \mathcal{K}_C, \tag{9}$$

where $\mathcal{S}$ represents the support set, and $\mathcal{Q}$ denotes the query set. $\mathcal{S}_k$ is the feature extracted by the model in the k-th category. $w(\mathcal{K}_C, \widetilde{N}_t^u)$ is the weighting function on $\mathcal{K}_C$, which determines the probability associated with the $\widetilde{C}_k$, and the weight is calculated based on a decreasing function ℓ_2 of the distance between the data samples and the class centroids. Its definition is as follows:

$$w(\mathcal{K}_{CH}, \widetilde{C}_k^t) = \begin{cases} 1, & if\ \mathcal{K}_C \in \mathcal{S}_k \\ \frac{\exp(-\beta||\mathcal{K}_C - \widetilde{N}_t^u||_2^2)}{\sum_{u=1}^{N} \exp(-\beta||\mathcal{K}_C - \widetilde{N}_t^u||_2^2)}, & if\ \mathcal{K}_C \in \mathcal{Q} \end{cases},$$

where β represents the temperature value.

Based on this, all the unlabeled samples $\mathcal{K}_C$ in the query set can be classified as follows:

$$Class(\mathcal{K}_C, [\widetilde{N}_1^\infty ..., \widetilde{N}_N^\infty]) = \arg\min_t ||\mathcal{K}_C - \widetilde{N}_N^\infty||_2. \tag{10}$$

4. Experimental Results and Analysis

4.1. Datasets

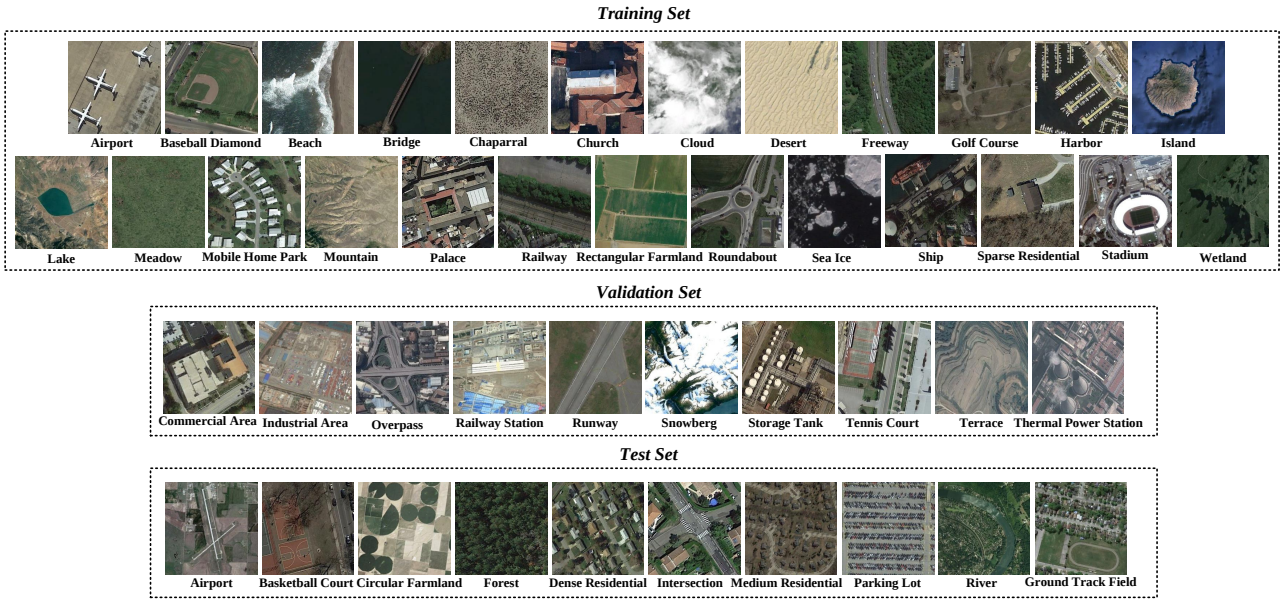
To assess the efficacy of our proposed CS^{2}TFSL, we conducted comprehensive comparisons between the CS^{2}TFSL and state-of-the-art methods on two widely recognized benchmark remote sensing few-shot scene classification datasets, namely the WHU-RS19 dataset [52] and the NWPU-RESISC45 dataset [53].

The WHU-RS19 dataset [52] is a widely used remote sensing dataset for scene classification. It was created by Wuhan University and consists of 19 classes of high-resolution remote sensing images. This dataset covers various land cover types, including urban areas, agricultural fields, forests, water bodies, and more. The images are captured by satellites with different sensors, such as SPOT-5 and GeoEye-1, providing a diverse range of spectral and spatial information. With a total of 2000 images, the dataset offers a balanced distribution of samples across different classes, ensuring equal representation for each category. Each image has a resolution of 600 $\times$ 600 pixels and is labeled with one of the 19 scene classes.

The NWPU-RESISC45 dataset [53] is created by Northwestern Polytechnical University and consists of 45 different scene categories. The images in the NWPU-RESISC45 dataset are captured from various regions using high-resolution satellite sensors, covering a wide range of scene types such as urban areas, farmland, forests, bodies of water, and industrial zones. Specifically, it is obtained using Google Earth satellite imagery and the Google Earth API. Each category in the dataset contains approximately 700 images, resulting in a total of around 31,500 images. This ensures an adequate number of samples and class balance, enabling effective evaluation of scene classification algorithms. Each image has a resolution of 256 $\times$ 256 pixels and is stored in JPEG format. The dataset also includes label files for each image, indicating their corresponding scene category.

In the context of few-shot remote sensing scene classification, we follow the partitioning method as described in [26,28,29,35,47]. The detailed division of the training set, validation set, and testing set for the two benchmark datasets is illustrated in Figure 2. From the information above, it can be inferred that the WHU-RS19 and NWPU-RESISC45 datasets pose comprehensive challenges to the few-shot scene classification algorithm performance in terms of the scale, resolution, and data volume.

(a)



(b)

Figure 2. The detailed division of the training set, validation set, and testing set for the two benchmark datasets. (**a**) WHU-RS19 dataset. (**b**) NWPU-RESISC45 dataset.

4.2. Implementation Details

The backbone used in CS^{2}TFSL is ResNet-12, which is commonly employed in few-shot scene classification task. In line with [26,28,29,35,47], we report the results for both five-way one-shot and five-way five-shot scenarios, with data presented within a 95% confidence interval. We conducted all our experiments using the Pytorch framework on CUDA 11.3, with a single NVIDIA RTX 3090 graphics card equipped with 24 GB. Based on the above software and hardware conditions, we trained our network with 600 epochs and a batch size of 16. We selected the SGD (Stochastic Gradient Descent) optimizer with a momentum of 0.9. In terms of comparative algorithms, we selected 14 classical and state-of-the-art few-shot classification algorithms, namely MAML [54], LLSR [27], Ji et al. [26], *baseline* [33], S2M2 [55], Meta-SGD [56], MatchingNet [57], ProtoNet [58], RelationNet [59], DLA-MatchNet [35], TAE-Net [37], IDLN [36], CAN+T [60], SPNet [25], and DANet [28]. For fair comparison, all comparison methods were reported from their respective papers and taken from the final experimental results.

4.3. Comparison with Other Methods

4.3.1. Results on the WHU-RS19 Dataset

Table 1 presents a performance comparison between our proposed CS^{2}TFSL and the counterparts on the WHU-RS19 dataset. The best results are highlighted in bold. In the five-way one-shot scenario, the top three performers achieved accuracy rates of $86.03 \pm 0.13\%$ for CS^{2}TFSL, $85.41 \pm 0.35\%$ for Ji et al. [26], and $81.06 \pm 0.60\%$ for SPNet [25]. More specifically, CS^{2}TFSL not only outperformed the second-place method by a margin of 0.62% in terms of accuracy but also significantly surpassed the other competitors. The outstanding performance of the CS^{2}TFSL is also evident in the five-way five-shot scenario. In more detail, CS^{2}TFSL outperformed the second-best algorithm (DANet [28]) by 1.19% and surpassed the third-best algorithm (*baseline* [33]) by an additional 2.86%. In general, the proposed CS^{2}TFSL performed well on the small-scale WHU-RS19 dataset. This is attributed to the synergistic training of two self-supervised auxiliary tasks, which effectively explore the unlabeled data in the WHU-RS19 dataset. Additionally, the combination of information from other samples in the dataset and the use of transductive inference contribute to its performance.

Table 1. Comparing the overall performance (in %) of the proposed CS^{2}TFSL with other few-shot scene classification methods on the WHU-RS19 dataset; we highlight the best result in bold.

Method	**5-Way**	
	1-Shot	**5-Shot**
MAML [54]	59.92 ± 0.35	82.30 ± 0.23
LLSR [27]	57.10	70.65
Ji et al. [26]	85.41 ± 0.35	92.28 ± 0.13
baseline [33]	75.57 ± 0.36	88.65 ± 0.18
S2M2 [55]	69.00 ± 0.41	82.14 ± 0.21
Meta-SGD [56]	51.54 ± 2.31	61.74 ± 2.02
MatchingNet [57]	76.14 ± 0.35	84.00 ± 0.20
ProtoNet [58]	77.00 ± 0.36	91.70 ± 0.15
RelationNet [59]	77.76 ± 0.34	86.84 ± 0.15
DLA-MatchNet [35]	70.21 ± 0.32	81.86 ± 0.52
IDLN [36]	73.89 ± 0.88	83.12 ± 0.56
TAE-Net [37]	73.67 ± 0.74	88.95 ± 0.52
CAN+T [60]	69.79 ± 0.56	79.71 ± 0.22
SPNet [25]	81.06 ± 0.60	88.04 ± 0.28
DANet [28]	75.02 ± 0.16	89.21 ± 0.07
CS^{2}TFSL (Ours)	$\mathbf{86.03 \pm 0.13}$	$\mathbf{93.09 \pm 0.05}$

4.3.2. Results on the NWPU-RESISC45 Dataset

Table 2 presents a performance comparison between our proposed CS^{2}TFSL and the counterparts on the NWPU-RESISC45 dataset. The best results are highlighted in bold. From the experimental results, it can be observed that the CS^{2}TFSL also performed remarkably well on the large-scale NWPU-RESISC45 dataset, achieving the best results in both five-way one-shot and five-way five-shot scenarios. Specifically, in the five-way one-shot scenario, the CS^{2}TFSL outperformed the second-best algorithm (IDLN [36]) by a significant margin of 8.42%. However, in the five-way five-shot scenario, it only surpassed the second-best algorithm (DANet [28]) by a slight margin of 1.19%. This is because the NWPU-RESISC45 dataset is a very large-scale dataset, and with only one reference sample, the limitation of inter-class similarity becomes significantly magnified. However, as the number of reference samples increases, the model's requirement for small-sample classification decreases. This precisely demonstrates that the CS^{2}TFSL also exhibits very stable performance in extreme scenarios.

Table 2. Comparing the overall performance (in %) of the proposed CS^{2}TFSL with other few-shot scene classification methods on the NWPU-RESISC45 dataset; we highlight the best result in bold.

Method	5-Way	
	1-Shot	5-Shot
MAML [54]	58.99 ± 0.45	72.67 ± 0.38
LLSR [27]	51.43	72.90
Ji et al. [26]	69.80 ± 0.53	82.03 ± 0.30
baseline [33]	69.02 ± 0.46	85.62 ± 0.25
S2M2 [55]	63.24 ± 0.47	83.23 ± 0.28
Meta-SGD [56]	60.63 ± 0.90	75.75 ± 0.65
MatchingNet [57]	61.57 ± 0.49	76.02 ± 0.34
ProtoNet [58]	64.52 ± 0.48	81.95 ± 0.30
RelationNet [59]	65.52 ± 0.85	78.38 ± 0.31
DLA-MatchNet [35]	71.56 ± 0.30	83.77 ± 0.64
IDLN [36]	75.25 ± 0.75	84.67 ± 0.23
TAE-Net [37]	69.13 ± 0.83	82.37 ± 0.52
CAN+T [60]	69.89 ± 0.58	81.04 ± 0.33
SPNet [25]	67.84 ± 0.87	83.94 ± 0.50
DANet [28]	74.30 ± 0.20	87.29 ± 0.11
CS^{2}TFSL (Ours)	**83.67 ± 0.25**	**88.48 ± 0.17**

5. Discussion

To fully demonstrate the effectiveness of the components proposed by the CS^{2}TFSL, we conducted ablation experiments in this section, including collaborative self-supervised learning and transductive inference. Figure 3 displays the heatmap visualization of the ablation experiments of the CS^{2}TFSL under different self-supervised auxiliary tasks. In the figure, "baseline" refers to the pretraining model of the feature extractor without any self-supervised auxiliary tasks. "rotation" and "SCL" indicate the inclusion of the rotation pretext task and spatial contrastive learning pretext task, respectively. It can be observed from the figure that the feature extractor in the baseline model lacks the ability to capture crucial objects in the scene. The network's attention does not focus on the key information in the scene. By incorporating self-supervised auxiliary tasks during pretraining, the models show improved ability to capture the information about crucial objects in the scene. However, these two pretext tasks, rotation and SCL, have their own strengths and weaknesses, and they only perform well in a few specific scenes. For example, in the basketball court scene, both tasks do not perform well. On the other hand, in the river scene, the rotation task performs worse than the SCL task. Conversely, in the ground track field scene, the opposite effect is observed.

In addition, the quantitative analysis of different self-supervised auxiliary task ablation experiments in two benchmark datasets is shown in Table 3. From the table, it can be observed that the design of the self-supervised auxiliary tasks improves the performance on few-shot remote sensing scene classification. The improvement in spatial contrastive learning is more significant compared to the effect of rotation tasks. Moreover, combining both approaches in the CS^{2}TFSL leads to an incremental improvement in accuracy for few-shot scene classification. Furthermore, Table 3 also presents the comparison results between inductive inference and transductive inference on two benchmark few-shot classification datasets. Compared to the traditional approach of using inductive inference in most few-shot scene classification algorithms, transductive inference, which incorporates additional information from other samples, has shown better performance in terms of classification accuracy. Specifically, transductive inference has shown a significant improvement in overall accuracy, with at least a 1% increase. This improvement is quite substantial in the context of the few-shot scene classification task.

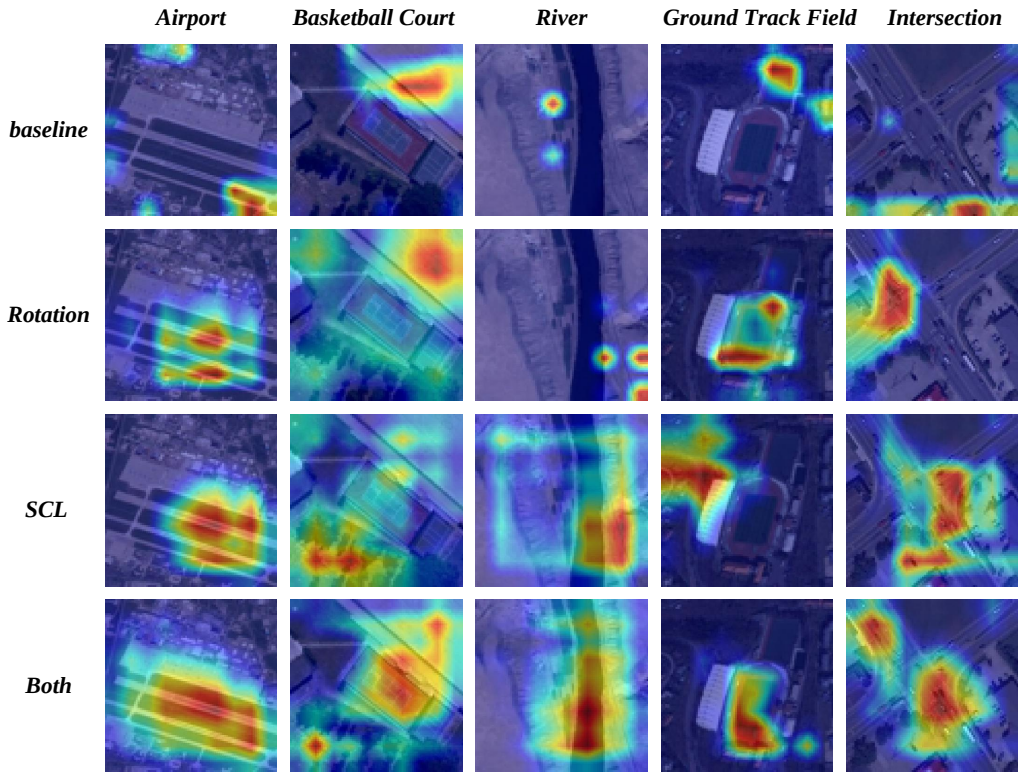


Figure 3. The heatmaps comparing the visualization of different self-supervised components in the CS^{2}TFSL, including the baseline (no self-supervised auxiliary task), rotation self-supervised auxiliary task, spatial contrastive learning (SCL) self-supervised auxiliary task, and the CS^{2}TFSL containing the collaboration of both.

Table 3. The ablation experimental results of the CS^{2}TFSL self-supervised auxiliary task, and the comparison between the inductive inference and transductive inference experiments, where "†" represents the inductive inference. We highlight the best result and the second-best result in bold and underline, respectively.

Dataset	Method	5-Way	
		1-Shot	5-Shot
WHU-RS19	baseline †	68.86 ± 0.22	76.74 ± 0.10
	+Rotation †	75.18 ± 0.18	80.57 ± 0.09
	+SCL †	79.69 ± 0.15	85.29 ± 0.09
	CS^{2}TFSL †	$\underline{85.26 \pm 0.13}$	$\underline{91.42 \pm 0.07}$
	baseline	70.16 ± 0.22	80.39 ± 0.10
	+Rotation	77.14 ± 0.21	83.18 ± 0.09
	+SCL	83.64 ± 0.17	88.83 ± 0.06
	CS^{2}TFSL	$\mathbf{86.03 \pm 0.13}$	$\mathbf{93.09 \pm 0.05}$
NWPU-RESISC45	baseline †	70.39 ± 0.29	80.65 ± 0.18
	+Rotation †	74.65 ± 0.26	82.38 ± 0.18
	+SCL †	78.19 ± 0.26	85.66 ± 0.18
	CS^{2}TFSL †	$\underline{80.37 \pm 0.26}$	$\underline{87.58 \pm 0.17}$
	baseline	73.07 ± 0.29	81.90 ± 0.18
	+Rotation	76.99 ± 0.25	83.67 ± 0.18
	+SCL	80.15 ± 0.25	84.10 ± 0.17
	CS^{2}TFSL	$\mathbf{83.67 \pm 0.25}$	$\mathbf{88.48 \pm 0.17}$

6. Conclusions

In this paper, we introduce a novel cooperative self-supervised transductive few-shot learning algorithm for remote sensing scene classification, called CS^{2}TFSL. We address the challenge of limited labeled data by leveraging two distinct self-supervised auxiliary tasks to collaboratively train the feature extractor and obtain a powerful representation. Subsequently, the trained feature extractor requires no further training, and its parameters are frozen, enabling seamless transfer to the inference stage. During the inference stage, we employ transductive inference, combining additional sample information in the data to enhance the associative information between the support and query sets. Extensive comparisons with state-of-the-art (SOTA) few-shot scene classification algorithms demonstrate the effectiveness of the CS^{2}TFSL. Furthermore, we conduct detailed ablation experiments to analyze the components of the CS^{2}TFSL. The experimental results highlight significant and promising performance improvements in few-shot scene classification achieved through the combination of self-supervised learning and transductive inference.

Author Contributions: Conceptualization, H.H. and Y.H.; methodology, H.H., Y.H. and Z.W.; validation, H.H. and Z.W.; investigation, H.H. and Y.H.; writing—original draft preparation, H.H., Y.H. and Z.W.; writing—review and editing, H.H. and Y.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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. Cheng, G.; Xie, X.; Han, J.; Guo, L.; Xia, G.S. Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 3735–3756. [CrossRef]
2. Zhu, Q.; Guo, X.; Li, Z.; Li, D. A review of multi-class change detection for satellite remote sensing imagery. *Geo-Spat. Inf. Sci.* **2022**, 1–15. [CrossRef]
3. Wang, Z.; Li, J.; Liu, Y.; Xie, F.; Li, P. An adaptive surrogate-assisted endmember extraction framework based on intelligent optimization algorithms for hyperspectral remote sensing images. *Remote Sens.* **2022**, *14*, 892. [CrossRef]
4. Li, H.; Li, J.; Zhao, Y.; Gong, M.; Zhang, Y.; Liu, T. Cost-sensitive self-paced learning with adaptive regularization for classification of image time series. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 11713–11727. [CrossRef]
5. Zhu, Q.; Guo, X.; Deng, W.; Shi, S.; Guan, Q.; Zhong, Y.; Zhang, L.; Li, D. Land-use/land-cover change detection based on a Siamese global learning framework for high spatial resolution remote sensing imagery. *ISPRS J. Photogramm. Remote Sens.* **2022**, *184*, 63–78. [CrossRef]
6. Xia, G.S.; Hu, J.; Hu, F.; Shi, B.; Bai, X.; Zhong, Y.; Zhang, L.; Lu, X. AID: A benchmark data set for performance evaluation of aerial scene classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3965–3981. [CrossRef]
7. Cheng, G.; Han, J.; Guo, L.; Qian, X.; Zhou, P.; Yao, X.; Hu, X. Object detection in remote sensing imagery using a discriminatively trained mixture model. *ISPRS J. Photogramm. Remote Sens.* **2013**, *85*, 32–43. [CrossRef]
8. Wang, Q.; Liu, S.; Chanussot, J.; Li, X. Scene classification with recurrent attention of VHR remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 1155–1167. [CrossRef]
9. Wang, Q.; Gao, J.; Lin, W.; Yuan, Y. Learning from synthetic data for crowd counting in the wild. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 8198–8207.
10. Zhang, L.; Zhang, L.; Du, B. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 22–40. [CrossRef]
11. Zhang, L.; Zhang, L.; Tao, D.; Huang, X. On combining multiple features for hyperspectral remote sensing image classification. *IEEE Trans. Geosci. Remote Sens.* **2011**, *50*, 879–893. [CrossRef]
12. Qian, X.; Lin, S.; Cheng, G.; Yao, X.; Ren, H.; Wang, W. Object detection in remote sensing images based on improved bounding box regression and multi-level features fusion. *Remote Sens.* **2020**, *12*, 143. [CrossRef]
13. Zhang, Y.; Gong, M.; Zhang, M.; Li, J. Self-Supervised Monocular Depth Estimation With Self-Perceptual Anomaly Handling. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, 1–15. [CrossRef] [PubMed]
14. Gong, M.; Zhao, Y.; Li, H.; Qin, A.K.; Xing, L.; Li, J.; Liu, Y.; Liu, Y. Deep Fuzzy Variable C-Means Clustering Incorporated with Curriculum Learning. *IEEE Trans. Fuzzy Syst.* **2023**, 1–15. [CrossRef]
15. Li, J.; Li, H.; Liu, Y.; Gong, M. Multi-fidelity evolutionary multitasking optimization for hyperspectral endmember extraction. *Appl. Soft Comput.* **2021**, *111*, 107713. [CrossRef]

16. Qian, X.; Zeng, Y.; Wang, W.; Zhang, Q. Co-saliency detection guided by group weakly supervised learning. *IEEE Trans. Multimedia* **2022**, *25*, 1810–1818. [CrossRef]
17. Lang, C.; Cheng, G.; Tu, B.; Han, J. Global Rectification and Decoupled Registration for Few-Shot Segmentation in Remote Sensing Imagery. *IEEE Trans. Geosci. Remote. Sens.* **2023**, *61*, 5617211. [CrossRef]
18. Gao, K.; Liu, B.; Yu, X.; Qin, J.; Zhang, P.; Tan, X. Deep relation network for hyperspectral image few-shot classification. *Remote Sens.* **2020**, *12*, 923. [CrossRef]
19. Zheng, W.; Tian, X.; Yang, B.; Liu, S.; Ding, Y.; Tian, J.; Yin, L. A few shot classification methods based on multiscale relational networks. *Appl. Sci.* **2022**, *12*, 4059. [CrossRef]
20. Lang, C.; Wang, J.; Cheng, G.; Tu, B.; Han, J. Progressive Parsing and Commonality Distillation for Few-shot Remote Sensing Segmentation. *IEEE Trans. Geosci. Remote. Sens.* **2023**, *61*, 5613610. [CrossRef]
21. Shuai, W.; Li, J. Few-shot learning with collateral location coding and single-key global spatial attention for medical image classification. *Electronics* **2022**, *11*, 1510. [CrossRef]
22. Oreshkin, B.; Rodríguez López, P.; Lacoste, A. Tadam: Task dependent adaptive metric for improved few-shot learning. In Proceedings of the NeurIPS 2018, Montreal, QC, Canada, 3–8 December 2018; Volume 31.
23. Lang, C.; Cheng, G.; Tu, B.; Han, J. Learning what not to segment: A new perspective on few-shot segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 8057–8067.
24. Gidaris, S.; Komodakis, N. Dynamic Few-Shot Visual Learning Without Forgetting. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4367–4375. [CrossRef]
25. Cheng, G.; Cai, L.; Lang, C.; Yao, X.; Chen, J.; Guo, L.; Han, J. SPNet: Siamese-prototype network for few-shot remote sensing image scene classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–11. [CrossRef]
26. Ji, H.; Yang, H.; Gao, Z.; Li, C.; Wan, Y.; Cui, J. Few-shot scene classification using auxiliary objectives and transductive inference. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
27. Zhai, M.; Liu, H.; Sun, F. Lifelong learning for scene recognition in remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1472–1476. [CrossRef]
28. Gong, M.; Li, J.; Zhang, Y.; Wu, Y.; Zhang, M. Two-path aggregation attention network with quad-patch data augmentation for few-shot scene classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–16. [CrossRef]
29. Li, J.; Gong, M.; Liu, H.; Zhang, Y.; Zhang, M.; Wu, Y. Multiform ensemble self-supervised learning for few-shot remote sensing scene classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–16. [CrossRef]
30. Bendou, Y.; Hu, Y.; Lafargue, R.; Lioi, G.; Pasdeloup, B.; Pateux, S.; Gripon, V. Easy—Ensemble augmented-shot-y-shaped learning: State-of-the-art few-shot classification with simple components. *J. Imaging* **2022**, *8*, 179. [CrossRef]
31. Schonfeld, E.; Ebrahimi, S.; Sinha, S.; Darrell, T.; Akata, Z. Generalized Zero- and Few-Shot Learning via Aligned Variational Autoencoders. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 8239–8247. [CrossRef]
32. Ye, H.J.; Hu, H.; Zhan, D.C.; Sha, F. Few-Shot Learning via Embedding Adaptation With Set-to-Set Functions. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 8805–8814. [CrossRef]
33. Chen, W.Y.; Liu, Y.C.; Kira, Z.; Wang, Y.C.F.; Huang, J.B. A closer look at few-shot classification. *arXiv* **2019**, arXiv:1904.04232.
34. Zhang, Y.; Gong, M.; Li, J.; Feng, K.; Zhang, M. Autonomous perception and adaptive standardization for few-shot learning. *Knowl.-Based Syst.* **2023**, *277*, 110746. [CrossRef]
35. Li, L.; Han, J.; Yao, X.; Cheng, G.; Guo, L. DLA-MatchNet for few-shot remote sensing image scene classification. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 7844–7853. [CrossRef]
36. Zeng, Q.; Geng, J.; Jiang, W.; Huang, K.; Wang, Z. IDLN: Iterative distribution learning network for few-shot remote sensing image scene classification. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [CrossRef]
37. Huang, W.; Yuan, Z.; Yang, A.; Tang, C.; Luo, X. TAE-net: Task-adaptive embedding network for few-shot remote sensing scene classification. *Remote Sens.* **2021**, *14*, 111. [CrossRef]
38. Jaiswal, A.; Babu, A.R.; Zadeh, M.Z.; Banerjee, D.; Makedon, F. A survey on contrastive self-supervised learning. *Technologies* **2020**, *9*, 2. [CrossRef]
39. Zhai, X.; Oliver, A.; Kolesnikov, A.; Beyer, L. S4l: Self-supervised semi-supervised learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27–28 October 2019; pp. 1476–1485.
40. Liu, X.; Zhang, F.; Hou, Z.; Mian, L.; Wang, Z.; Zhang, J.; Tang, J. Self-supervised learning: Generative or contrastive. *IEEE Trans. Knowl. Data Eng.* **2021**, *35*, 857–876. [CrossRef]
41. Zhang, Y.; Gong, M.; Li, J.; Zhang, M.; Jiang, F.; Zhao, H. Self-supervised monocular depth estimation with multiscale perception. *IEEE Trans. Image Process.* **2022**, *31*, 3251–3266. [CrossRef]
42. Wang, Y.; Albrecht, C.M.; Braham, N.A.A.; Mou, L.; Zhu, X.X. Self-supervised learning in remote sensing: A review. *arXiv* **2022**, arXiv:2206.13188.
43. Hong, D.; Gao, L.; Yao, J.; Yokoya, N.; Chanussot, J.; Heiden, U.; Zhang, B. Endmember-guided unmixing network (EGU-Net): A general deep learning framework for self-supervised hyperspectral unmixing. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 6518–6531. [CrossRef] [PubMed]

44. Hu, M.; Wu, C.; Zhang, L. HyperNet: Self-supervised hyperspectral spatial–spectral feature understanding network for hyperspectral change detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–17. [CrossRef]
45. Wen, Z.; Liu, Z.; Zhang, S.; Pan, Q. Rotation awareness based self-supervised learning for SAR target recognition with limited training samples. *IEEE Trans. Image Process.* **2021**, *30*, 7266–7279. [CrossRef]
46. Yue, J.; Fang, L.; Rahmani, H.; Ghamisi, P. Self-supervised learning with adaptive distillation for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–13. [CrossRef]
47. Li, X.; Shi, D.; Diao, X.; Xu, H. SCL-MLNet: Boosting few-shot remote sensing scene classification via self-supervised contrastive learning. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–12. [CrossRef]
48. 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. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 21271–21284.
49. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A simple framework for contrastive learning of visual representations. In Proceedings of the International Conference on Machine Learning (PMLR), Virtual Event, 13–18 July 2020; pp. 1597–1607.
50. Li, Y.; Shao, Z.; Huang, X.; Cai, B.; Peng, S. Meta-FSEO: A meta-learning fast adaptation with self-supervised embedding optimization for few-shot remote sensing scene classification. *Remote Sens.* **2021**, *13*, 2776. [CrossRef]
51. Ouali, Y.; Hudelot, C.; Tami, M. Spatial contrastive learning for few-shot classification. In Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD 2021), Bilbao, Spain, 13–17 September 2021; pp. 671–686.
52. Sheng, G.; Yang, W.; Xu, T.; Sun, H. High-resolution satellite scene classification using a sparse coding based multiple feature combination. *Int. J. Remote Sens.* **2012**, *33*, 2395–2412. [CrossRef]
53. Cheng, G.; Han, J.; Lu, X. Remote sensing image scene classification: Benchmark and state of the art. *Proc. IEEE* **2017**, *105*, 1865–1883. [CrossRef]
54. 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.
55. Mangla, P.; Kumari, N.; Sinha, A.; Singh, M.; Krishnamurthy, B.; Balasubramanian, V.N. Charting the right manifold: Manifold mixup for few-shot learning. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 2218–2227.
56. Li, Z.; Zhou, F.; Chen, F.; Li, H. Meta-sgd: Learning to learn quickly for few-shot learning. *arXiv* **2017**, arXiv:1707.09835.
57. Vinyals, O.; Blundell, C.; Lillicrap, T.; Wierstra, D. Matching networks for one shot learning. In Proceedings of the 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain, 5–10 December 2016; Volume 29.
58. Snell, J.; Swersky, K.; Zemel, R. Prototypical networks for few-shot learning. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; Volume 30.
59. Sung, F.; Yang, Y.; Zhang, L.; Xiang, T.; Torr, P.H.; Hospedales, T.M. Learning to compare: Relation network for few-shot learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1199–1208.
60. Hou, R.; Chang, H.; Ma, B.; Shan, S.; Chen, X. Cross attention network for few-shot classification. In Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, BC, Canada, 8–14 December 2019; Volume 32.