symmetry

MDPI

Article

TACDFSL: Task Adaptive Cross Domain Few-Shot Learning

Qi Zhang, Yingluo Jiang and Zhijie Wen *

Department of Mathematics, College of Sciences, Shanghai University, Shanghai 200444, China; zhangqi3552@163.com (Q.Z.); yingluojiang@shu.edu.cn (Y.J.)
* Correspondence: wenzhijie@shu.edu.cn

Abstract: Cross Domain Few-Shot Learning (CDFSL) has attracted the attention of many scholars since it is closer to reality. The domain shift between the source domain and the target domain is a crucial problem for CDFSL. The essence of domain shift is the marginal distribution difference between two domains which is implicit and unknown. So the empirical marginal distribution measurement is proposed, that is, WDMDS (Wasserstein Distance for Measuring Domain Shift) and MMDMDS (Maximum Mean Discrepancy for Measuring Domain Shift). Besides this, pre-training a feature extractor and fine-tuning a classifier are used in order to have a good generalization in CDFSL. Since the feature obtained by the feature extractor is high-dimensional and left-biased, the adaptive feature distribution transformation is proposed, to make the feature distribution of each sample be approximately Gaussian distribution. This approximate symmetric distribution improves image classification accuracy by 3% on average. In addition, the applicability of different classifiers for CDFSL is investigated, and the classification model should be selected based on the empirical marginal distribution difference between the two domains. The Task Adaptive Cross Domain Few-Shot Learning (TACDFSL) is proposed based on the above ideas. TACDFSL improves image classification accuracy by 3–9%.

Keywords: cross domain few-shot learning; domain shift; empirical marginal distribution; feature distribution transformation

check for updates

Citation: Zhang, Q.; Jiang, Y.; Wen, Z. TACDFSL: Task Adaptive Cross Domain Few-Shot Learning. *Symmetry* **2022**, *14*, 1097. https://doi.org/10.3390/sym14061097

Academic Editor: Jan Awrejcewicz

Received: 29 April 2022
Accepted: 24 May 2022
Published: 27 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

There are two main reasons for Few-Shot Learning (FSL) receiving increasing attention recently [1–5]. Firstly, though deep learning has achieved great success in visual recognition tasks, it greatly depends on plenty of labeled samples. However, collecting and labeling data always require high costs. Secondly, humans can learn quickly from a few samples, and even a child can identify pictures of a "zebra" after learning the concept of "zebra". The current mainstream FSL algorithms are based on meta-learning [6–12]. The main idea of meta-learning-based algorithms mainly focuses on the model's "learning to learn" ability so that the model can have a good generalization performance on the novel task. However, most FSL algorithms do not consider the domain shift between the source domain and the target domain. Due to the small number of labeled samples in the novel task, usually, FSL needs a source domain with a large number of labeled samples to train the model. Most FSL algorithms are evaluated on the same domains. For example, both the target domain and source domain are natural images. But many applications face a situation: collecting source datasets for a novel task in the same domain is not feasible or impossible. For instance, the novel task dataset is medical images, but a suitable dataset within the medical image area cannot be found. In this case, CDFSL is very valuable for practical application. Whether it is agricultural data or medical data, CDFSL can be applied [13–15].

In CDFSL, the model trained on the source domain cannot have a good generalization performance on the target domain for the large domain shift between the two domains. Tseng et al. propose to use affine transformations to simulate various feature distributions in different domains during the training phase [16]. However, the experimental datasets

used in their paper are not diverse enough and do not consider the problem of domain shift measurement. The measurement of domain shift is a significant problem in CDFSL.

Chen et al. evaluated the cross domain generalization ability of the FSL classification algorithm with datasets miniImagenet [6] and CUB [17] (both natural image datasets) and found that the current FSL algorithm did not perform well in cross domain situations, calling on more scholars to pay attention to CDFSL [18]. Guo et al. proposed that the CDFSL algorithm should consider the images with a diverse assortment of image acquisition methods, such as aviation or medical images and proposed A Broader Study of Cross-Domain Few-Shot Learning (BSCDFSL) benchmark [19]. BSCDFSL includes four datasets with different similarity to Imagenet [20]. As shown in Figure 1, the three criteria for similarity are perspective distortion, semantic content, and color depth. Following these three criteria, the similarity between these four target domain datasets (CropDisease [21], EuroSAT [22], ISIC [23], ChestX [24]) and the source domain is gradually decreasing. However, these three criteria are from the perspective of subjective understanding. A metric to measure domain shift from the data distribution perspective is proposed and a quantitative analysis of domain shift is made. In addition, the meta-learning-based CDFSL algorithm is not suitable for the problem with large domain shift. Hence, a pre-train and fine-tune method for CDFSL is proposed.

A simple but effective model is presented. Our contributions are shown as follows:

1. For CDFSL, the domain shift between the source domain and the target domain is a key problem and domain shift is essentially a marginal distribution difference. However, the marginal distribution of the domain is implicit and unknown, so from the perspective of sample features, an empirical marginal distribution which is suitable for CDFSL is proposed, that is, WDMDS (Wasserstein Distance for Measuring Domain Shift) and MMDMDS (Maximum Mean Discrepancy for Measuring Domain Shift).
2. Because of the large domain shift, the meta-learning-based method is not suitable for CDFSL. The rotation method to pre-train the feature extractor and fine-tune the novel task on the classifier is adopted, so that the model can fast and effectively generalize to novel tasks. The effectiveness of different classifiers on different datasets for CDFSL is compared. The research finds that selecting a classifier based on the basis of the empirical marginal distribution between domains is necessary. Hence, the Task Adaptive Cross Domain Few-Shot Learning (TACDFSL) is proposed.
3. An adaptive feature distribution transformation is used to correct the left-biased feature distribution obtained by the convolutional neural network, which effectively improves image classification accuracy.

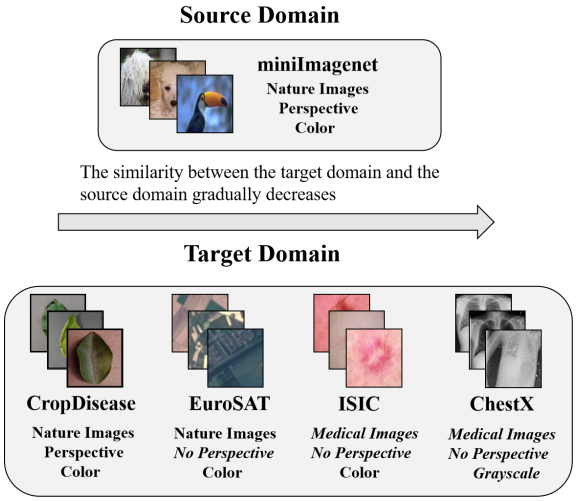


Figure 1. Schematic diagram of the similarity between the CDFSL benchmark datasets and miniImagenet. Here we use miniImagenet as the source domain.

2. Related Work

Few-Shot Learning Great progress has been made in FSL based on meta-learning. These algorithms are mainly divided into three categories: metric-based, optimization-based, and data augmentation-based. Metric-based FSL algorithm mainly measures the distance between sample features and category features and uses specific metric functions to classify samples [6–9]. MatchingNet uses cosine similarity in a recurrent network [6]. ProtoNet proposes euclidean distance for measuring distance between samples and classes [7]. RelationNet utilizes similarity score in Convolutional Neural Network (CNN) [8], while GNN applies the graph convolutional networks as the metric function [9]. Optimization-based FSL algorithm trains initialization parameter which can generalize to the novel task after few step iterations [10,25–27]. The idea of FSL algorithm based on data augmentation aims to extend prior knowledge by generating more diverse samples [28,29].

In addition to the meta-learning-based FSL algorithm, an algorithm based on transfer-learning is also popular [18,30–32]. Chen et al. proposed that a simple pre-train and fine-tune training strategy can achieve comparable results to complex meta-training [18]. The transfer-learning-based algorithm mainly focuses on feature extractor with good feature extraction ability and fine-tune on the novel task. Mangla et al. investigates the role of learning relevant feature manifold for few-shot tasks using self-supervision and regularization techniques [30]. An FSL task is called as N-way K-shot problem when target task contains N classes and each class contains K labels sample [6].

Cross Domain Few-Shot Learning CDFSL pays attention to the problem that the source domain and the target domain has large domain shift. CDFSL is in the early stage of development. Tseng et al. propose to use affine transformations to simulate various feature distributions in different domains during the training phase; therefore, the sample feature is enhanced [16]. Guo et al. propose a benchmark for cross domain few shot learning [19] and this work transfer the knowledge pre-trained on ImageNet [20] to four novel datasets: CropDiseases [21], EuroSAT [22], ISIC2018 [23], and ChestX [24]. Phoo et al. apply a self-training method to adapt representations to target domains. The unlabeled samples from the target domain and the labeled samples from the source domain are added to the self-training process, so as to obtain a feature extractor adapted to the target domain [33].

Domain Adaptation Domain Adaptation (DA) aims to reduce the domain shift between the source domain and the target domain. From the perspective of the feature alignment, a lot of work apply adversarial training to align the source and target feature distributions in feature space [34–37]. However, in DA, the source domain and the target domain have the same label space, so the features of these two domains can be aligned. Contrary to DA, CDFSL has different label space, thus the algorithm cannot blindly narrow the distance between them.

3. Methodology

As shown in Figure 2, TACDFSL mainly contains four parts: a feature extractor which is pre-trained on $\mathcal{D}_s$ with a self-supervised method—rotation [30] (introduced in Section 3.1); a domain shift measurement model (introduced in Section 3.2); a feature transformation module which adaptively modifies the distribution for each sample (introduced in Section 3.3); and an adaptive select classification module (introduced in Section 3.4).

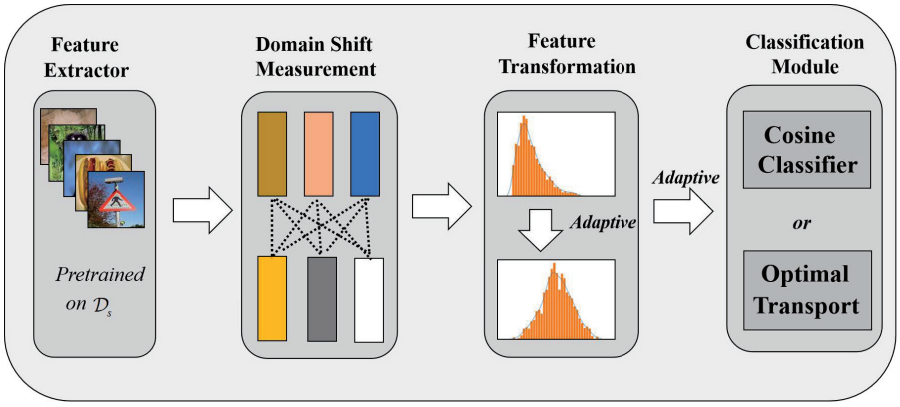


Figure 2. Schematic diagram of TACDFSL.

3.1. Feature Extractor

This paper uses a self-supervised method, namely, Rotation, to pre-train the feature extractor [30].

Let $f_\theta(\cdot)$ be the feature extractor with parameters θ, which extracts the feature of sample x. Let $m_{\theta_c}(\cdot)$ denote the classifier, which labels the sample x. Then the classification objective is,

$$\min_{\theta,\theta_c} \frac{1}{N_s} \sum_{(x_i,y_i)\sim\mathcal{D}_s} L_{CE}(m_{\theta_c}(f_\theta(x_i)), y_i), \tag{1}$$

where N_s is the number of samples in $\mathcal{D}_s$, y_i is the label of sample x_i, and L_{CE} is the cross entropy loss.

Adding rotation as an auxiliary task in the model training process can help the model learn better feature representation. Therefore, the input sample is rotated by four angles, i.e., $0°, 90°, 180°, 270°$. The target task of the model is not only to predict the image category correctly but also to predict the angle of image rotation. Let x^r denotes an image x rotated by r angles, where $r \in R = \{0°, 90°, 180°, 270°\}$. In this case, a classifier m_{θ_r} is used to predict the rotation angle. The rotation loss L_R and classification loss L_{CL} are

$$\begin{aligned} L_R &= \frac{1}{N_r} \sum_{x_i\in\mathcal{D}_s} \sum_{r\in R} L_{CE}(m_{\theta_r}(f_\theta(x^r)), r), \\ L_{CL} &= \frac{1}{N_s} \sum_{(x_i,y_i)\in\mathcal{D}_s} L_{CE}(m_{\theta_c}(f_\theta(x_i^r)), y_i), \end{aligned} \tag{2}$$

where $N_r = 4$ means the rotation is divided into four classes. The final model training loss L is

$$L = \frac{1}{2}L_R + \frac{1}{2}L_{CL}. \tag{3}$$

3.2. Empirical Marginal Distribution for CDFSL

The marginal distribution of the domain means the data generating distribution. As the actual marginal distribution $P(x)$ of the domain is implicit and unknown, so the empirical marginal distribution is used to approximate the domain marginal distribution from the perspective of sample features. Wasserstein distance can measure the minimum cost for converting one distribution to another [38], while MMD can measure the empirical marginal distribution difference of two domains in the feature space [39–41]. Both of two are used to measure the domain shift, respectively.

3.2.1. Wasserstein Distance for Measuring Domain Shift (WDMDS)

Measuring domain shift is a discrete Wasserstein distance and can be formalized as the following linear programming problem. Let $P = \{(p_1, \omega_{p_1}), \cdots, (p_m, \omega_{p_m})\}$ denote the signature of the source domain, where $p_i, i = 1, \cdots, m$ represents the ith eigenvalue of the

source domain and ω_{p_i} represents the weight of p_i. Let $Q = \{(q_1, \omega_{q_1}), \cdots, (q_n, \omega_{q_n})\}$ be the signature of the target domain. $q_i, i = 1, \cdots, n$ and ω_{q_i} represent the features of the target domain and corresponding weights. Let $D = [d_{ij}]$ represent the distance matrix, where d_{ij} is the distance between p_i and q_j, finding a transmission matrix $F = [f_{ij}]$ is expected, where f_{ij} represents the amount transferred from p_i to q_j, then $F = [f_{ij}]$ can minimize the cost

$$WD(F; P, Q, D) = \sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} f_{ij}, \tag{4}$$

which is subjected to

$$\begin{aligned} f_{ij} &\geq 0, \quad 1 \leq i \leq m, \ 1 \leq j \leq n, \\ \sum_{j=1}^{n} f_{ij} &\leq \omega_{p_i}, \ 1 \leq i \leq m, \\ \sum_{i=1}^{m} f_{ij} &\leq \omega_{q_i}, \ 1 \leq j \leq n, \\ \sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} &= \min\left(\sum_{i=1}^{m} \omega_{p_i}, \sum_{j=1}^{n} \omega_{q_j} \right). \end{aligned} \tag{5}$$

After obtaining the transmission matrix $F = [f_{ij}]$, the final Wasserstein distance is calculated as

$$WD(P, Q) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} f_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}}. \tag{6}$$

The above equations have an intuitive understanding: the minimum amount of energy required to convert a matrix composed of sample features in the target domain into a matrix composed of sample features in the source domain.

In this paper, p_i and q_i are the features obtained by the feature extractor, and each sample feature is equally important, so taking the weight $\omega_{p_i} = \omega_{q_j} = 1$.

3.2.2. Maximum Mean Discrepancy for Measuring Domain Shift (MMDMDS)

Let n_s, n_t represent the number of samples in the source domain $\mathcal{D}_s$ and the target domain $\mathcal{D}_t$ respectively, then the MMD distance between $\mathcal{D}_s$ and $\mathcal{D}_t$ is defined as

$$\mathrm{MMD}^2[\mathcal{D}_s, \mathcal{D}_t] = \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \phi(\mathbf{x}_i) - \frac{1}{n_t} \sum_{j=1}^{n_t} \phi(\tilde{\mathbf{x}}_j) \right\|_{\mathcal{H}}^2, \tag{7}$$

where $\mathcal{H}$ is Reproducing Kernel Hilbert Space (RKHS). $\mathbf{x}_i, \tilde{\mathbf{x}}_j$ are the features of source domain and target domain samples extracted by feature extractor $f(\cdot)$ respectively. ϕ is the mapping.

Let $k(x, x') := \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}$. The Gaussian kernel function is used, $k(u, v) = e^{\frac{-\|u-v\|^2}{\sigma}}$, then Equation (7) is written as

$$\begin{aligned} \mathrm{MMD}^2[\mathcal{D}_s, \mathcal{D}_t] &= \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \phi(\mathbf{x}_i) - \frac{1}{n_t} \sum_{j=1}^{n_t} \phi(\tilde{\mathbf{x}}_j) \right\|_{\mathcal{H}}^2 \\ &= \frac{1}{n_s^2} \left[\sum_{i=1}^{n_s} \phi(\mathbf{x}_i) \right]^2 + \frac{1}{n_t^2} \left[\sum_{j=1}^{n_t} \phi(\tilde{\mathbf{x}}_j) \right]^2 - \frac{2}{n_s n_t} \left[\sum_{i=1}^{n_s} \phi(\mathbf{x}_i) \sum_{j=1}^{n_t} \phi(\tilde{\mathbf{x}}_j) \right] \\ &= \frac{1}{n_s^2} \sum_{i=1}^{n_s} \sum_{i=1}^{n_s} k(\mathbf{x}_i, \mathbf{x}_i) + \frac{1}{n_t^2} \sum_{j=1}^{n_t} \sum_{j=1}^{n_t} k(\tilde{\mathbf{x}}_j, \tilde{\mathbf{x}}_j) - \frac{2}{n_s n_t} \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} k(\mathbf{x}_i, \tilde{\mathbf{x}}_j). \end{aligned} \tag{8}$$

The essence of MMDMDS is to calculate the difference between two domains with the mean values of the sample features in RKHS.

As shown in Figure 3, let $\mathcal{T}_n = \{(x_i, y_i)\}_{i=1}^{NK}$ be the novel task. Random selecting N-way K-shot samples in $\mathcal{D}_s$, denoted as $\mathcal{T}_s = \{(x_i, y_i)\}_{i=1}^{NK}$, both novel task and base task extract features through pre-trained backbone $f_{\theta^*}(\cdot)$. Then, calculate domain shift with WDMDS or MMDMDS. With the same novel task, random sampling E batches base task, then calculate 95% interval confidence of E domain shift.

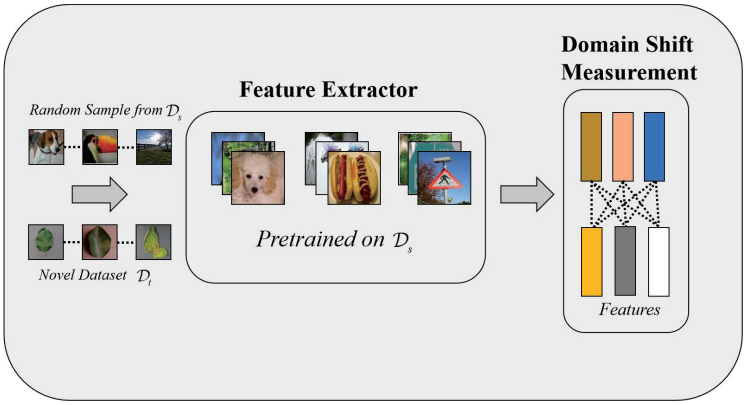


Figure 3. Schematic diagram of the domain shift measurement algorithm, the figure takes 3-way 1-shot as an example.

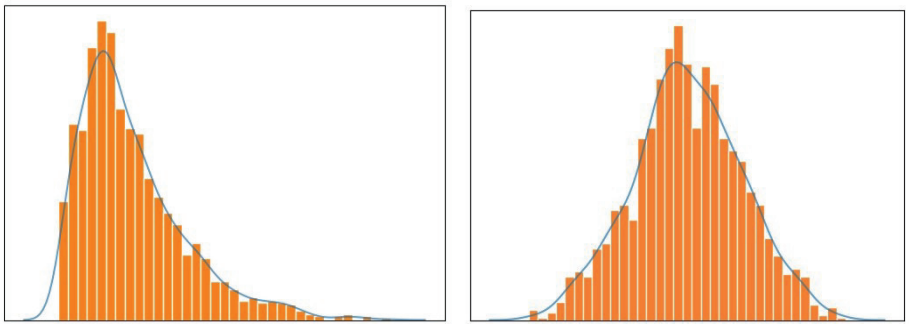
3.3. Adaptive Feature Distribution Transformation

The feature extracted by the feature extractor f_θ has the characteristics of highly-dimension and left-skewed. And each sample's feature distribution is different, so adaptively searching for suitable parameters for each sample is important so that the feature distribution can be approximately Gaussian distribution. The Box-Cox Transformation is used [42] to modify the feature distribution.

$$g(v) = \begin{cases} \dfrac{(v+c)^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log(v+c) & \lambda = 0 \end{cases}, \tag{9}$$

where $v = f_\theta(x)$ is the feature of sample x, and c is a constant. In order to ensure that $v + c$ is positive, $c = 1$ in the experiment. Each sample will obtain the most suitable λ for itself.

Figure 4a is the left-skewed feature distribution. After transformation, it is approximately Gaussian distribution, as shown in Figure 4b.



(**a**) (**b**)

Figure 4. Schematic diagram of feature distribution before and after transformation: (**a**) Left-skewed distribution before transformation; (**b**) Approximate to Gaussian distribution after transformation.

3.4. Classifier Selection for CDFSL

This subsection introduces two representative classifiers used in this paper, Cosine Classifier (CC) and Optimal Transport (OT) classifier.

3.4.1. Cosine Classifier

Let f_{θ^*} denote the pre-trained model, i.e., feature extractor. Chen et al. propose to use the cosine classifier to adapt the novel task [18], that is

$$s_{ij} = \frac{f_{\theta^*}(x_i)^T \mathbf{w}_j}{\|f_{\theta^*}(x_i)\| \|\mathbf{w}_j\|}, \tag{10}$$

where s_{ij} is the similarity score between sample x_i and class j and $\mathbf{w}_j$ is weight parameter of class j.

3.4.2. Optimal Transport

According to [31], considering classification task as Optimal Transport (OT) problem and use sinkhorn algorithm to solve it. Hu et al. define mapping matrix $\mathbf{M}$ as follows

$$\mathbf{M}^* = Sinkhorn(\mathbf{L}, \mathbf{p}, \mathbf{q}, \alpha) = \arg \min_{\mathbf{M} \in \mathbb{U}(\mathbf{p},\mathbf{q})} \sum_{i,j} \mathbf{M}_{i,j} \mathbf{L}_{i,j} + \alpha H(\mathbf{M}), \tag{11}$$

and

$$\mathbb{U}(\mathbf{p}, \mathbf{q}) = \left\{ \mathbf{M} \in \mathbb{R}_+^{Nq \times N} \mid \mathbf{M}\mathbf{1}_N = \mathbf{p}, \mathbf{M}^T \mathbf{1}_{Nq} = \mathbf{q} \right\}, \tag{12}$$

where $\mathbf{p}$ is the number of class N, and $\mathbf{q}$ is $N \times \mathbf{q}$. $\mathbf{L}_{ij}$ is the euclidean distance between sample i and class center j. $\mathbf{M}_{ij}$ is mapping matrix, which denotes the probability of sample i belongs to class j. $\mathbf{M}_{ij}\mathbf{L}_{ij}$ is the element-wise product, which indicates the energy required to transport sample i to class j. $H(\mathbf{M}) = -\sum_{ij} \mathbf{M}_{ij} \log \mathbf{M}_{ij}$ is regularization term with hyper-parameter α.

4. Experiments

This section presents TACDFSL experiments results on image classification.

Datasets The experimental datasets are first introduced. **miniImagenet** contains 100 classes from ImageNet [20]. Each class has 600 images with size 84 × 84, and 100 classes are split randomly into 64 base, 16 validation, and 20 novel classes [6]. **CropDisease** is a crop disease dataset which is split into training set and test set. Total dataset consists of 38 classes and 54,305 images with size 256 × 256 [21]. **EuroSAT** is made up of satellite images, which consists of 10 classes, 27,000 images, where 5 classes have 3000 images each, 4 classes have 2500 images each, and 1 class has 2000 images. The size is 64 × 64 [22]. **ISIC** is a color medical dataset which consists of 7 classes, 10,015 images, with size 600 × 450 [23]. **ChestX** is a gray medical dataset, and it contains 12 classes, 112,120 images with the pixel 1024 × 1024 [24].

Experimental Details The code for this paper is implemented on PyTorch. The feature extractor is WideResNet [30], and the target domain image is resized to 80 × 80 before the feature extractor, which is consistent with the pre-train image size. For MMDMDS, five Gaussian kernels are used and sum the results up for stability. For adaptive feature transformation, $c = 1$.

4.1. Performance Evaluation with TACDFSL

The experimental results are shown in Table 1. The measure is accuracy, and the accuracy is calculated by averaging 600 randomly test episodes with 95% confidence intervals. In Table 1, † represents the results reported by Guo et al. [19]. CC represents the Cosine Classifier (CC), and CC + Box-Cox represents that the features are first subjected to Box-Cox transformation and then to the cosine classifier. OT + Box-Cox means that features are first subjected to Box-Cox transformation and then classified by OT.

Table 1 shows that for CC and OT, with different experimental datasets and different shot numbers, these two measurement methods have their own advantages and disadvantages. For CropDisease and EuroSAT, which are more similar to miniImagenet, OT is better when the number of shots is small, and CC is better for 50-shot cases. On the other hand, for ISIC and ChestX, which are less similar to miniImagenet, CC performs better than OT regardless of the number of shots. Figure 5 shows the accuracy of different methods under 5-way 5-shot, it is obviously that CC + Box-Cox has the best results for CropDisease and EuroSAT, while OT + Box-Cox has the best results for ISIC and ChestX. For more complex medical image datasets such as ISIC and ChestX, it is necessary to pay more attention to the intra-class gaping and inter-class gaping. For OT, a better class center cannot be learned on ISIC and ChestX. Besides this, for the 50-shot case, OT requires a larger computational cost.

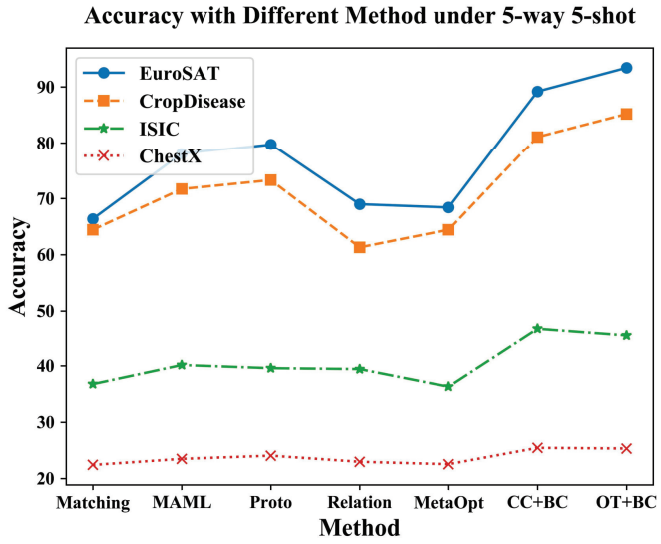


Figure 5. Schematic diagram of accuracy with different method under 5-way 5-shot. BC indicates Box-Cox.

In addition, in Table 1, the results with † are based on the meta-learning algorithm, while Backbone, CC + Box-Cox, and OT + Box-Cox are based on the transfer learning algorithm. Comparing all the results in the table, even a simple fine-tuning strategy like Backbone can give comparable results to complex meta-learning-based algorithms. Therefore, paying more attention to the generalization performance of CDFSL based on transfer learning will be better.

4.2. Discussion

4.2.1. Measuring Domain Shift

This section shows the domain shift results obtained by WDMDS and MMDMDS in CDFSL. Table 2 shows the WDMDS and MMDMDS for the source domain and each target domain (CropDisease, EuroSAT, ISIC, ChestX) under the 5-way N-shot $N = \{1, 5, 20, 50\}$ setting. "-" in the WDMDS indicates that the amount of calculation is too large and WDMDS is not suitable for 20-shot and 50-shot situations, and "-" in MMDMDS indicates that MMDMDS values are unstable and are not suitable for 1-shot situations.

As shown in Table 2, the larger the value, the farther the target domain is from the source domain, and both methods can measure the gap between the source domain and the target domain except CropDisease. For CropDisease, since miniImagenet contains more natural image category information and CropDisease is one of the categories in miniImagenet, there will be a higher numerical value when sampling N-way K-shot samples on miniImagenet for domain shift experiments. In addition, for 5-way 1-shot and 5-way 5-shot, the WDMDS mentioned in this paper is more suitable for measuring the difference

between the source domain and the target domain. At the same time, MMDMDS is more appropriate to measure the multiple sample situation.

Table 1. Image classification accuracy (%) obtained by each method.

Methods	CropDisease 5-Way 5-Shot	CropDisease 5-Way 20-Shot	CropDisease 5-Way 50-Shot	EuroSAT 5-Way 5-Shot	EuroSAT 5-Way 20-Shot	EuroSAT 5-Way 50-Shot
MatchingNet †	66.39 ± 0.78	76.38 ± 0.67	58.53 ± 0.73	64.45 ± 0.63	77.10 ± 0.57	54.44 ± 0.67
MatchingNet+FWT †	62.74 ± 0.90	74.90 ± 0.71	75.68 ± 0.78	56.04 ± 0.65	63.38 ± 0.69	62.75 ± 0.76
MAML †	78.05 ± 0.68	89.75 ± 0.42	-	71.70 ± 0.72	81.95 ± 0.55	-
ProtoNet †	79.72 ± 0.67	88.15 ± 0.51	90.81 ± 0.43	73.29 ± 0.71	82.27 ± 0.57	80.48 ± 0.57
ProtoNet+FWT †	72.72 ± 0.70	85.82 ± 0.51	87.17 ± 0.50	67.34 ± 0.76	75.74 ± 0.70	78.64 ± 0.57
RelationNet †	68.99 ± 0.75	80.45 ± 0.64	85.08 ± 0.53	61.31 ± 0.72	74.43 ± 0.66	74.91 ± 0.58
RelationNet+FWT †	64.91 ± 0.79	78.43 ± 0.59	81.14 ± 0.56	61.16 ± 0.70	69.40 ± 0.64	73.84 ± 0.60
MetaOpt †	68.41 ± 0.73	82.89 ± 0.54	91.76 ± 0.38	64.44 ± 0.73	79.19 ± 0.62	83.62 ± 0.58
Backbone	80.61 ± 0.70	89.61 ± 0.48	91.97 ± 0.43	76.74 ± 0.66	83.74 ± 0.52	86.49 ± 0.47
CC + Box-Cox	89.20 ± 0.54	**95.24 ± 0.33**	**96.78 ± 0.26**	81.08 ± 0.61	**88.72 ± 0.40**	**91.40 ± 0.36**
OT + Box-Cox	**93.42 ± 0.55**	**95.49 ± 0.39**	95.88 ± 0.35	**85.19 ± 0.67**	**87.87 ± 0.49**	89.07 ± 0.43
Methods	**ISIC 5-Way 5-Shot**	**ISIC 5-Way 20-Shot**	**ISIC 5-Way 50-Shot**	**ChestX 5-Way 5-Shot**	**ChestX 5-Way 20-Shot**	**ChestX 5-Way 50-Shot**
MatchingNet †	36.74 ± 0.53	45.72 ± 0.53	54.58 ± 0.65	22.40 ± 0.70	23.61 ± 0.86	22.12 ± 0.88
MatchingNet+FWT †	30.40 ± 0.48	32.01 ± 0.48	33.17 ± 0.43	21.26 ± 0.31	23.23 ± 0.37	23.01 ± 0.34
MAML †	40.13 ± 0.58	52.36 ± 0.57	-	23.48 ± 0.96	27.53 ± 0.43	-
ProtoNet †	39.57 ± 0.57	49.50 ± 0.55	51.99 ± 0.52	24.05 ± 1.01	28.21 ± 1.15	29.32 ± 1.12
ProtoNet+FWT †	38.87 ± 0.52	43.78 ± 0.47	49.84 ± 0.51	23.77 ± 0.42	26.87 ± 0.43	30.12 ± 0.46
RelationNet †	39.41 ± 0.58	41.77 ± 0.49	49.32 ± 0.51	22.96 ± 0.88	26.63 ± 0.92	28.45 ± 1.20
RelationNet+FWT †	35.54 ± 0.55	43.31 ± 0.51	46.38 ± 0.53	22.74 ± 0.40	26.75 ± 0.41	27.56 ± 0.40
MetaOpt †	36.28 ± 0.50	49.42 ± 0.60	54.80 ± 0.54	22.53 ± 0.91	25.53 ± 1.02	29.35 ± 0.99
Backbone	37.42 ± 0.54	44.67 ± 0.53	49.79 ± 0.55	23.88 ± 0.41	27.04 ± 0.45	30.78 ± 0.50
CC + Box-Cox	**46.57 ± 0.61**	**56.18 ± 0.59**	**61.59 ± 0.56**	**25.44 ± 0.46**	**30.08 ± 0.52**	**32.92 ± 0.52**
OT + Box-Cox	45.39 ± 0.67	53.15 ± 0.59	56.68 ± 0.58	25.32 ± 0.48	29.17 ± 0.52	31.75 ± 0.51

† represents the results reported by Guo et al. [19].

Table 2. Domain shift between each target domain and miniImagenet.

Methods	CropDisease 5-Way 1-Shot	CropDisease 5-Way 5-Shot	CropDisease 5-Way 20-Shot	CropDisease 5-Way 50-Shot	EuroSAT 5-Way 1-Shot	EuroSAT 5-Way 5-Shot	EuroSAT 5-Way 20-Shot	EuroSAT 5-Way 50-Shot
WDMDS	1.10 ± 0.01	1.00 ± 0.00	-	-	0.89 ± 0.01	0.87 ± 0.01	-	-
MMDMDS	-	1.33 ± 0.01	1.21 ± 0.01	1.20 ± 0.01	-	1.18 ± 0.01	1.04 ± 0.01	1.13 ± 0.01
Methods	**ISIC 5-Way 1-Shot**	**ISIC 5-Way 5-Shot**	**ISIC 5-Way 20-Shot**	**ISIC 5-Way 50-Shot**	**ChestX 5-Way 1-Shot**	**ChestX 5-Way 5-Shot**	**ChestX 5-Way 20-Shot**	**ChestX 5-Way 50-Shot**
WDMDS	0.97 ± 0.01	0.95 ± 0.00	-	-	1.14 ± 0.01	1.07 ± 0.00	-	-
MMDMDS	-	1.40 ± 0.01	1.41 ± 0.01	1.43 ± 0.01	-	1.77 ± 0.00	1.76 ± 0.00	1.72 ± 0.00

Some experiments on MMDMDS for different shot numbers ($K = 5, 10, 20, 30, 40, 50$) show that for different K numbers, MMDMDS has better stability and can effectively measure the gap between the source domain and the target domain. The experimental results are shown in Table 3. Figure 6 shows the visualization of MMDMDS between each target domain and miniImagenet under different shot numbers. Figure 6a contains MMDMDS of four datasets, it can be found that for the increase of the number of shots, MMDMDS can still maintain its numerical stability, and as shown in Figure 6b, as analyzed above, except for CropDisease, the similarity between the other three target domain datasets and the source domain is significantly different, and MMDMDS can be used to measure the similarity between the source domain and the target domain.

Table 3. MDMDS between each target domain and miniImagenet under different shots.

Dataset	5-Shot	10-Shot	20-Shot	30-Shot	40-Shot	50-Shot
CropDisease	1.33 ± 0.01	1.33 ± 0.01	1.21 ± 0.01	1.21 ± 0.01	1.38 ± 0.01	1.20 ± 0.01
EuroSAT	1.18 ± 0.01	1.09 ± 0.01	1.04 ± 0.01	1.15 ± 0.01	1.08 ± 0.01	1.13 ± 0.01
ISIC	1.40 ± 0.01	1.47 ± 0.01	1.41 ± 0.01	1.47 ± 0.01	1.41 ± 0.01	1.43 ± 0.01
ChestX	1.77 ± 0.00	1.74 ± 0.00	1.76 ± 0.00	1.70 ± 0.00	1.70 ± 0.00	1.72 ± 0.00

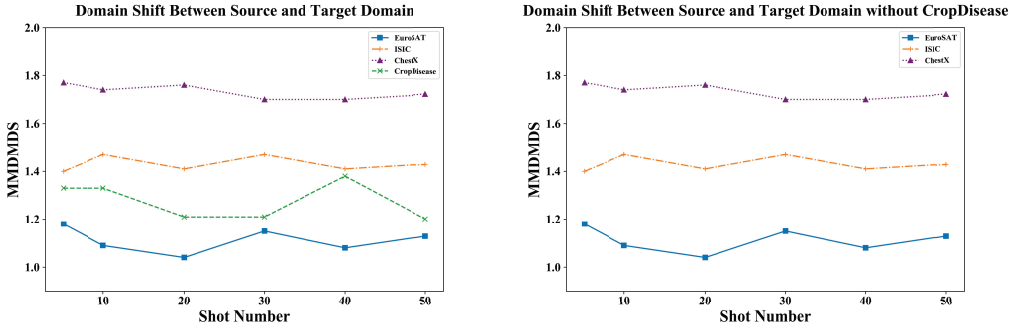


Figure 6. Schematic diagram of MMDMDS between each target domain and miniImagenet under different shot numbers: (**a**) MMDMDS between four target domain datasets and miniImagenet. (**b**) MMDMDS between target domain and miniImagenet except CropDisease

4.2.2. Adaptive Feature Distribution Transformation Ablation Study

This subsection performs ablation experiments on the adaptive feature distribution transformation and show the results. As shown in Table 4, the data in the table are 5-way N-shot for each target domain dataset, $N = \{5, 20, 50\}$. The classification accuracy of 95% confidence interval of 600 experiments. CC and OT indicate using cosine classifier or optimal transport, respectively, and no adaptive feature distribution transformation, CC + Box-Cox and OT + Box-Cox represent an adaptive feature distribution transformation. From Table 4, it can be found that, after the feature transformation distribution, whether it is the CC or OT, there is an improvement of 1% to 5%. It is important to find a suitable λ for each sample to perform feature distribution transformation. The t-SNE visualization on the samples before and after feature distribution transformation is shown in Figure 7. The left side of each subgraph is untransformed, and the right side is transformed. Each color in the subfigure represents a class, a good classification effect is the same color close to each other, but different colors are far away. It can be found that for CropDisease and EuroSAT, after the transformation, the aggregation effect of the same class is more obvious.

Some ablation experiments on the hyper-parameter c in the transformation are performed. The results are shown in Table 5. The data in the table are the classification accuracy (%) obtained by different c values under 5-way 5-shot, and its visual bar chart shown in Figure 8. From the data in the table and the figure, it can be found that the best result are obtained when $c = 1$. The ablation experiments for the case of $c > 1$, because the classification accuracy obtained when $c > 1$ is 20%, indicating that the model does not have the ability to learn. This is because the feature obtained by the convolutional neural network is a left-skewed distribution that is extremely close to 0 in this problem, and taking $c = 1$ can effectively alleviate this extreme left-skewed distribution.

Table 4. Classification accuracy (%) of different datasets before and after adaptive distribution transformation.

Methods	CropDisease 5-Way 5-Shot	CropDisease 5-Way 20-Shot	CropDisease 5-Way 50-Shot	EuroSAT 5-Way 5-Shot	EuroSAT 5-Way 20-Shot	EuroSAT 5-Way 50-Shot
CC	86.31 ± 0.63	93.67 ± 0.36	95.84 ± 0.26	80.49 ± 0.66	88.12 ± 0.46	90.76 ± 0.37
CC + Box-Cox	**89.20 ± 0.54**	**95.24 ± 0.33**	**96.78 ± 0.26**	**81.62 ± 0.60**	**88.95 ± 0.40**	**91.40 ± 0.36**
OT	88.06 ± 0.74	90.80 ± 0.55	92.50 ± 0.44	80.68 ± 0.78	84.53 ± 0.60	86.39 ± 0.54
OT + Box-Cox	**93.42 ± 0.55**	**95.49 ± 0.39**	**95.88 ± 0.35**	**85.19 ± 0.67**	**87.87 ± 0.49**	**89.07 ± 0.43**
Methods	**ISIC 5-Way 5-Shot**	**ISIC 5-Way 20-Shot**	**ISIC 5-Way 50-Shot**	**ChestX 5-Way 5-Shot**	**ChestX 5-Way 20-Shot**	**ChestX 5-Way 50-Shot**
CC	42.11 ± 0.59	52.88 ± 0.59	57.86 ± 0.56	24.60 ± 0.40	29.61 ± 0.50	32.30 ± 0.55
CC + Box-Cox	**46.57 ± 0.61**	**56.18 ± 0.59**	**61.59 ± 0.56**	**25.44 ± 0.46**	**30.08 ± 0.52**	**32.92 ± 0.52**
OT	38.60 ± 0.65	44.36 ± 0.60	47.98 ± 0.59	23.56 ± 0.44	26.81 ± 0.46	28.96 ± 0.48
OT + Box-Cox	**45.39 ± 0.67**	**53.15 ± 0.59**	**56.68 ± 0.58**	**25.32 ± 0.48**	**29.17 ± 0.52**	**31.75 ± 0.51**

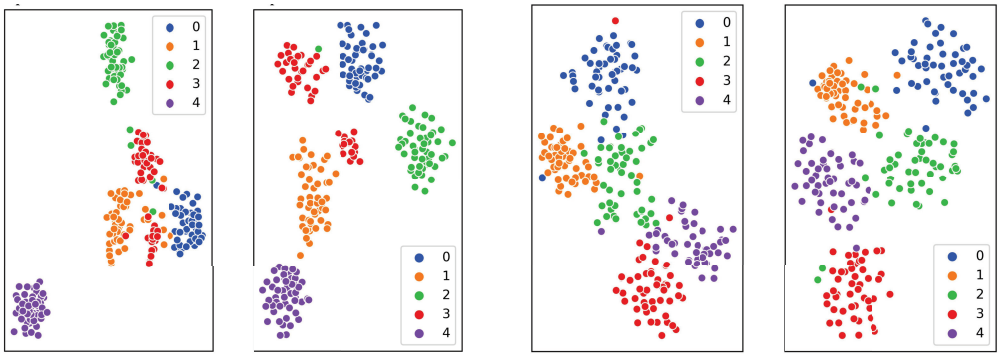


Figure 7. t-SNE visualization of CropDisease and EuroSAT before and after feature distribution transformation. The left side of each subfigure is untransformed, and the right side is transformed. (**a**) CropDisease. (**b**) EuroSAT.

Table 5. Classification accuracy (%) with different c values under 5-way 5-shot.

	5-Shot	10-Shot	20-Shot	30-Shot	40-Shot	50-Shot
Dataset	$c = 0.1$	$c = 0.3$	$c = 0.5$	$c = 0.6$	$c = 0.8$	$c = 1.0$
CropDisease	84.18 ± 0.68	82.75 ± 0.73	83.25 ± 0.73	84.20 ± 0.70	84.45 ± 0.74	$\mathbf{89.20 \pm 0.54}$
EuroSAT	53.14 ± 0.89	44.24 ± 0.88	46.78 ± 0.87	50.81 ± 0.88	62.26 ± 0.87	$\mathbf{81.08 \pm 0.61}$
ISIC	25.04 ± 0.42	21.46 ± 0.30	22.47 ± 0.32	23.14 ± 0.39	38.33 ± 0.57	$\mathbf{46.57 \pm 0.61}$
ChestX	24.54 ± 0.40	24.40 ± 0.41	24.45 ± 0.41	24.18 ± 0.40	24.62 ± 0.41	$\mathbf{25.44 \pm 0.46}$

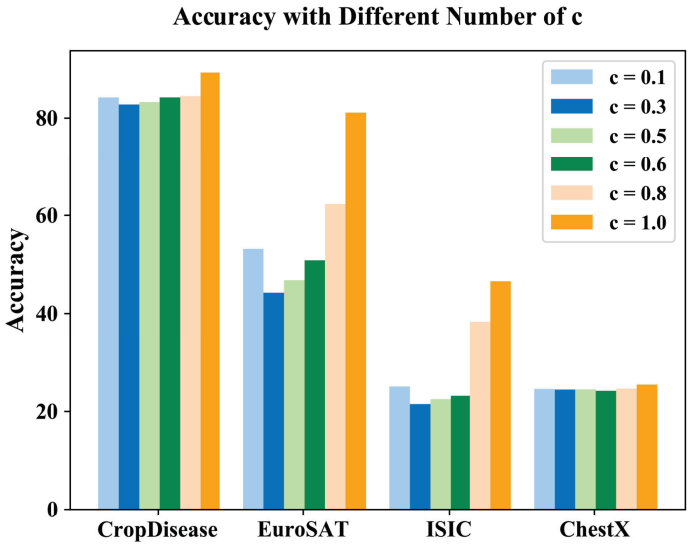


Figure 8. Schematic diagram of classification accuracy obtained with different c values under 5-way 5-shot.

5. Conclusions

This paper proposes a Task Adaptive Cross Domain Few-Shot Learning (TACDFSL) based on the empirical marginal distribution. The empirical marginal distribution measurement experiments show that WDMDS is more suitable for novel tasks with a relatively small sample size (such as 5-way 1-shot, 5-way 5-shot, etc.), while MMDMDS is more ideal for multiple samples (such as 5-way 10-shot, 5-way 20-shot, etc.). Besides this, adaptive feature distribution transformation in image classification tasks of CDFSL is necessary, and the ablation experiments indicate that the approximate symmetric distribution can effectively improve the image classification accuracy. In addition, considering the large domain shift between the two domains, meta-learning-based algorithms do not have a

good generalization of the novel task. Therefore, the method based on transfer-learning will be a better choice for CDFSL. Therefore, the rotation to pre-train the feature extractor and fine-tune the classifier on the novel task is adopted. In this training strategy, it is essential to select the classifier. The applicability of CC and OT for the CDFSL is investigated. For the novel task, the classification module should be chosen based on the difference of the empirical marginal distribution. The analysis of the experimental results shows that it is crucial to find a better class center for medical image datasets, so this research will continue to explore how to find a better class center for such medical datasets with large domain shift. In addition, in image classification, it is important to reduce the intra-class gap and increase the inter-class gap. Since CC and OT have their own superiority, fusing the cosine distance into OT will be considered in the future work.

Author Contributions: Methodology, Q.Z. and Z.W.; validation, Z.W. and Y.J.; writing, Q.Z. and Y.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the National Natural Science Foundation of China (11971296).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used in this paper are all publicly available. Here the download URLs for the four target domain datasets are given. The source domain dataset miniImagenet https://lyy.mpi-inf.mpg.de/mtl/download/ (accessed on 26 April 2016) is a classic dataset commonly used for FSL. EuroSAT: http://madm.dfki.de/downloads (accessed on 6 August 2019), ISIC: http://challenge2018.isic-archive.com (accessed on 27 July 2018), CropDisease: https://www.kaggle.com/saroz014/plant-disease/ (accessed on 26 May 2016), ChestX: https://www.kaggle.com/nih-chest-xrays/data (accessed on 1 January 2017).

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

Abbreviations

The following abbreviations are used in this paper:

FSL	Few-Shot Learning
CDFSL	Cross Domain Few-Shot Learning
TACDFSL	Task Adaptive Cross Domain Few-Shot Learning
WDMDS	Wasserstein Distance for Measuring Domain Shift
MMDMDS	Maximum Mean Discrepancy for Measuring Domain Shift
CC	Cosine Classifier
OT	Optimal Transport

References

1. Fan, Z.; Yu, J.G.; Liang, Z. Fgn: Fully guided network for few-shot instance segmentation. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 9172–9181.
2. Ganea, D.A.; Boom, B.; Poppe, R. Incremental few-shot instance segmentation. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 1185–1194.
3. Simon, C.; Koniusz, P.; Nock, R. Adaptive subspaces for few-shot learning. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 4136–4145.
4. Jeong, T.; Kim, H. OOD-MAML: Meta-learning for few-shot out-of-distribution detection and classification. In Proceedings of the 2020 Conference on Neural Information Processing Systems, Online, 6–12 December 2020; Volume 33, pp. 3907–3916.
5. Zhao, Y.; Wang, L.; Tian, Y. Few-shot neural architecture search. In Proceedings of the 2021 International Conference on Machine Learning, Online, 18–24 July 2021; pp. 12707–12718.
6. Vinyals, O.; Blundell, C.; Lillicrap, T. Matching networks for one shot learning. In Proceedings of the Annual Conference on Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; Volume 29.
7. Snell, J.; Swersky, K.; Zemel, R. Prototypical networks for few-shot learning. In Proceedings of the Annual Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.
8. Sung, F.; Yang, Y.; Zhang, L. Learning to compare: Relation network for few-shot learning. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1199–1208.

9. Garcia, V.; Bruna, J. Few-shot learning with graph neural networks. In Proceedings of the 6th International Conference on Learning Representation, Vancouver, BC, Canada, 30 April–3 May 2018.
10. Finn, C.; Abbeel, P.; Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. In Proceedings of the 2017 International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 1126–1135.
11. Zhong, X.; Gu, C.; Huang, W. Complementing representation deficiency in few-shot image classification: A meta-learning approach. In Proceedings of the 25th International Conference on Pattern Recognition, Milan, Italy, 10–15 January 2021; pp. 2677–2684.
12. Rajasegaran, J.; Khan, S.; Hayat, M. An incremental task-agnostic meta-learning approach. In Proceedings of the 2020 Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 13588–13597.
13. Sugandh, U.; Khari, M.; Nigam, S. The integration of blockchain and IoT edge devices for smart agriculture: The challenges and use cases. *Adv. Comput.* **2022**. [CrossRef]
14. Sugandh, U.; Khari, M.; Nigam, S. How Blockchain Technology Can Transfigure the Indian Agriculture Sector: A Review. In *Handbook of Green Computing and Blockchain Technologies*; CRC Press: Boca Raton, FL, USA, 2022; pp. 69–88.
15. Srivastava, S.; Khari, M.; Crespo, R.G.; Chaudhary, G.; Arora, P. *Concepts and Real-Time Applications of Deep Learning*; Springer International Publishing: Cham, Switzerland, 2021.
16. Tseng, H.Y.; Lee, H.Y. Cross-Domain Few-Shot Classification via Learned Feature-Wise Transformation. In Proceedings of the 8th International Conference on Learning Representations, Addis Ababa, Ethiopia, 26–30 April 2020.
17. Wah, C.; Branson, S.; Welinder, P. *The Caltech-UCSD Birds-200-2011 Dataset*; California Institute of Technology: Pasadena, CA, USA, 2011.
18. Chen, W.Y.; Liu, Y.C.; Kira, Z. A closer look at few-shot classification. In Proceedings of the 7th International Conference on Learning Representations, New Orleans, LA, USA, 6–9 May 2019.
19. Guo, Y.; Codella, N.C.; Karlinsky, L. A broader study of cross-domain few-shot learning. In Proceedings of the 2020 European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 124–141.
20. Deng, J.; Dong, W.; Socher, R. A large-scale hierarchical image database. In Proceedings of the 2009 Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
21. Mohanty, S.P.; Hughes, D.P.; Salathé, M. Using deep learning for image-based plant disease detection. *Front. Plant Sci.* **2016**, *7*, 1419. [CrossRef] [PubMed]
22. Helber, P.; Bischke, B.; Dengel, A. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 2217–2226. [CrossRef]
23. Tschandl, P.; Rrsendahl, C.; Kittler, H. The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data* **2018**, *5*, 1–9. [CrossRef] [PubMed]
24. Wang, X.; Peng, Y.; LU, L. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In Proceedings of the 2017 Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2097–2106.
25. Rusu, A.A.; Rao, D.; Sygnowski, J. Meta-learning with latent embedding optimization. In Proceedings of the 7th International Conference on Learning Representations, New Orleans, LA, USA, 6–9 May 2019.
26. Vuorio, R.; Sun, S.H.; Hu, H. Multimodal model-agnostic meta-learning via task-aware modulation. In Proceedings of the 2019 Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; Volume *32*.
27. Tseng, H.Y.; De, M.S.; Tremblay, J. Few-shot viewpoint estimation. In Proceedings of the 31st British Machine Vision Virtual Conference, Online, 7–10 September 2020.
28. Wang, Y.X.; Girshick, R.; Hebert, M. Low-shot learning from imaginary data. In Proceedings of the 2018 Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7278–7286.
29. Alfassy A.; Karlinsky, L.; Aides, A. Laso: Label-set operations networks for multi-label few-shot learning. In Proceedings of the 2019 Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 6548–6557.
30. Mangla, P.; Kumari, N.; Sinha, A. Charting the right manifold: Manifold mixup for few-shot learning. In Proceedings of the 2022 Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 4–8 January 2022; pp. 2218–2227.
31. Hu, Y.; Gripon, V.; Pateux, S. Leveraging the feature distribution in transfer-based few-shot learning. In Proceedings of the 2021 International Conference on Learning Representations, Vienna, Austria, 3–7 May 2021; pp. 487–499.
32. Yang, S.; Liu, L.; Xu, M. Free lunch for few-shot learning: Distribution calibration. In Proceedings of the 2021 International Conference on Learning Representations, Vienna, Austria, 3–7 May 2021.
33. Phoo, C.P.; Hariharan, B. Self-training for few-shot transfer across extreme task differences. In Proceedings of the 2021 International Conference on Learning Representations, Vienna, Austria, 3–7 May 2021.
34. Wang, A.; Liu, C.; Xue, D. Hyperspectral Image Classification Based on Cross-Scene Adaptive Learning. *Symmetry* **2021**, *13*, 1817. [CrossRef]
35. Hsu, H.K.; Yao, C.H.; Tsai, Y.H. Progressive domain adaptation for object detection. In Proceedings of the 2020 Winter Conference on Applications of Computer Vision, Snowmass Village, CO, USA, 2–5 March 2020; pp. 749–757.
36. Wang, F.; Chai, G.; Li, Q. An Efficient Deep Unsupervised Domain Adaptation for Unknown Malware Detection. *Symmetry* **2022**, *14*, 296. [CrossRef]

37. Fickinger, A.; Cohen, S.; Russell, S. Cross-Domain Imitation Learning via Optimal Transport. In Proceedings of the 2022 International Conference on Learning Representations, Online, 25–29 April 2022.
38. Rubner, Y.; Tomasi, C.; Guibas, L.J. The earth mover's distance as a metric for image retrieval. *Int. J. Comput. Vis.* **2000**, *40*, 99–121. [CrossRef]
39. Borgwardt, K.M.; Gretton, A.; Rasch, M.J. Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics* **2006**, *22*, e49–e57. [CrossRef] [PubMed]
40. Gretton, A.; Sejdinovic, D.; Strathmann, H. Optimal kernel choice for large-scale two-sample tests. In Proceedings of the 2012 Annual Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–8 December 2021; Volume 25.
41. Duan, L.; Tsang, I.W.; Xu, D. Domain transfer multiple kernel learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 465–479. [CrossRef] [PubMed]
42. Box, G.E.P.; Cox, D.R. An analysis of transformations. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1964**, *26*, 211–243. [CrossRef]