



# A Survey on Cross-Domain Few-Shot Image Classification

Shisheng Deng<sup>1,2</sup>, Dongping Liao<sup>3</sup>, Xitong Gao<sup>1(✉)</sup>, Juanjuan Zhao<sup>1</sup>,  
and Kejiang Ye<sup>1</sup>

<sup>1</sup> Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences,  
Shenzhen 518000, China

{ss.deng, xt.gao, jj.zhao,kj.ye}@siat.ac.cn

<sup>2</sup> University of Chinese Academy of Sciences, Beijing 100049, China

<sup>3</sup> University of Macau, Macau SAR 999078, China  
yb97428@um.edu.mo

**Abstract.** Due to the limited availability of labelled data in many real-world scenarios, we have to resort to data from other domains to improve models' performance, which prompts the advancement of research regarding the cross-domain few-shot image classification task. In this paper, we systematically review existing cross-domain few-shot image classification algorithms published in recent years. We categorize these algorithms into data-augmentation and feature-alignment paradigms and present their recent progress. We summarize three commonly-used cross-domain datasets for benchmarking few-shot image classification tasks and relevant scenarios. Finally, we outline existing limitations and future perspectives.

**Keywords:** Deep Learning · Few-shot Image Classification · Cross-Domain

## 1 Introduction

Over the last decade, deep learning (DL) [30] has achieved excellent results on many application scenarios, including computer vision [20], natural language processing [14], *etc.* Traditional DL methods are not effective in tasks with limited training data. In contrast, humans can leverage their accumulated knowledge to quickly learn the characteristics of unfamiliar things with a limited amount of data. To address this issue, researchers have introduced the concept of Few-Shot Learning (FSL) [57]. FSL aims to mimic the human learning process and achieve better generalization performance by using a limited number of training samples in scenarios where data is scarce. Recently, Few-Shot Image Classification (FSIC) [57] algorithms have demonstrated better classification accuracy than humans in image classification. However, these remarkable outcomes are limited

---

S. Deng and D. Liao—Equal contribution.

to scenarios where there is only a slight difference between the distribution of the training data and the test data. For situations where there is a sizeable distributional difference between the training and test data, the model will suffer significant performance degradation due to the discrepancy between the different domains. Researchers have thus formalized the Cross-Domain Few-Shot Image Classification (CDFSIC) [7], along with its corresponding classification algorithms to investigate the challenges in cross-domain few-shot learning.

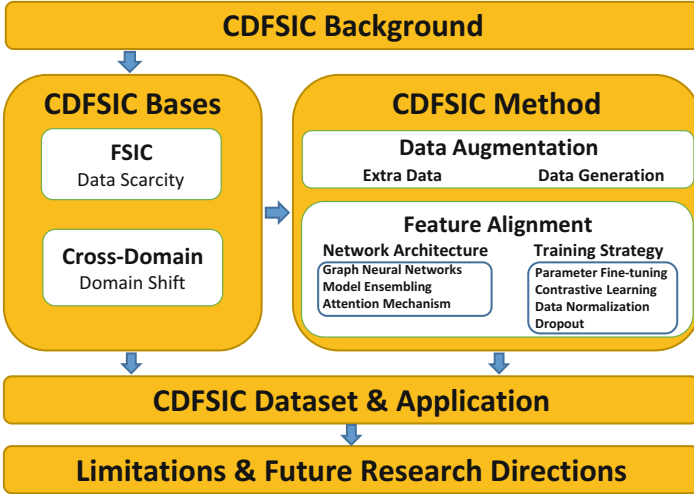


Fig. 1. The framework of survey.

This paper presents a thorough and systematic review of CDFSIC. As shown in Fig. 1, the survey is structured as follows. First, following the introduction of CDFSIC in this section, we present the preliminaries of CDFSIC in Sect. 2, which includes the definitions of FSIC and Cross-Domain problems. We then provide a summary of the current CDFSIC methods, including an introduction to standard datasets and applications. Finally, we discuss the limitations and challenges of CDFSIC that may present future research opportunities.

## 2 Preliminaries of CDFSIC

### 2.1 Few-Shot Image Classification

Few-Shot Learning (FSL) [57] is a machine learning technique that involves training a model to achieve strong generalization performance using only a limited number of training examples. One of the most widely-used benchmark for evaluating FSL algorithms is Few-Shot Image Classification (FSIC), which has numerous realistic applications [57].

A FSIC task can be defined as  $\mathcal{D}_{\text{FSIC}} = \{\mathcal{D}_{\text{train}}, \mathcal{D}_{\text{test}}\}$ , where  $\{y \mid (x, y) \in \mathcal{D}_{\text{train}}\} \cap \{y \mid (x, y) \in \mathcal{D}_{\text{test}}\} = \emptyset$ , *i.e.*, the test and train datasets do not contain common labels. Following [29], most recent works on FSIC employ the standard *N-way K-shot (M-query)* episodic task learning.

Specifically, for each FSIC task, we sample  $n$  episodic tasks  $\{T_1, \dots, T_n\}$  from  $\mathcal{D}_{\text{train}}$  as training episodes, and  $m$  episodic tasks  $\{T_1, \dots, T_m\}$  from  $\mathcal{D}_{\text{test}}$  as testing episodes. Each episodic task  $T_i$  consists of a support set  $T_i^S$  and a query set  $T_i^Q$ . From a dataset, each episodic tasks randomly samples  $N$  categories respectively, with each category sampling  $K$  image-label pairs  $(x, y)$ ,  $T_i^S = \{(x_k, y_k)\}_{k=1}^{N \times K}$  for support set, and each category sampling  $M$  image-label pairs  $(x, y)$ ,  $T_i^Q = \{(x_k, y_k)\}_{k=1}^{N \times M}$  for query set. Both  $\mathcal{D}_{\text{train}}$  and  $\mathcal{D}_{\text{test}}$  samples the support and query sets following the above configuration, except that the  $\mathcal{D}_{\text{test}}$  provides no labels for the query set, namely,  $T_i^Q = \{(x_k)\}_{k=1}^{N \times M}$ .

## 2.2 The Cross-Domain Problem

Blanchard *et al.* [3] formally presented the Cross-Domain (CD) problem in machine learning, while Torralba *et al.* [47] brought research attention to the cross-domain problem in computer vision tasks. They investigated the performance of classification models by thorough evaluation on six popular benchmark datasets. Their experiments showed that the intrinsic dataset bias introduced by the domain gap will lead to poor generalization performance.

A domain is defined as a joint distribution  $P(X, Y)$  [70] of the input (data) space  $X$  and output (label) space  $Y$ . For the Cross-Domain problem, the source-domain distribution  $P_S(X, Y)$  and the target-domain distribution  $P_T(X, Y)$  are notably different. Moreover, the data of target domain is not available during the model training process. Most of the research has focused on the multi-source scenario, which presupposes the availability of several distinct yet relevant domains. Specifically, given  $K$  similar but distinct source domains,  $S = \{S_k = \{(x^k, y^k)\}_{k=1}^K\}$ , each domain is represented by a joint distribution  $P_S^k(X, Y)$ . Note that  $P_S^k(X, Y)$  is dissimilar to  $P_S^{k'}(X, Y)$ , with  $k \neq k'$  for  $k, k' \in \{1, \dots, K\}$ . The joint distribution corresponding to the target domain is denoted as  $P_T(X, Y)$ . In addition,  $P_T(X, Y)$  is also dissimilar to  $P_S^k(X, Y)$ , where  $k \in \{1, \dots, K\}$ .

The *cross-domain few-shot image classification* (CDFSIC) problem, first introduced by Chen *et al.* [7], poses challenges of both Cross-Domain and Few-Shot Image Classification, including a scarce sample size and considerable differences between the training and testing data distributions. The models trained under CDFSIC would thus require stronger generalization capabilities than traditional FSIC models for better adaptation to novel target domains.

## 3 CDFSIC Algorithm

In general, CDFSIC faces two challenges: data scarcity and domain shift. Based on these challenges, the current approach of CDFSIC can be categorized into two camps: *data augmentation* and *feature alignment* methods.

### 3.1 Data Augmentation Methods

Data augmentation [45], commonly utilized in deep learning methods, can mitigate the possibility of overfitting, which may happen when the training dataset has a limited number of samples, while having low diversity. Recently, some researchers employ additional larger datasets (*e.g.*, ImageNet [12]) as training data to augment the FSIC task. This technique aims to learn valuable features from a varied dataset with higher diversity [18]. Additionally, data generation [45] is another popular data augmentation technique. Based on these approaches, we categorize data augmentation methods into two: *extra data* and *data generation*.

**Extra Data.** As part of their work, Chen *et al.* [7] introduced the first benchmark dataset for the CDFSIC task, namely MiniImageNet  $\rightarrow$  CUB. They employed MiniImagenet [52] as the source domain, which is relatively similar to the target domain, CUB [53].

Real-world CDFSIC scenarios involve domains that differ greatly in data volume and distribution. Addressing this issue, Guo *et al.* [18] proposed a broader CDFSIC baseline than previous work. Employing ImageNet as the source domain, they conduct experiments on four datasets with varying degrees of similarity to the natural image based on 3 orthogonal criteria: 1) existence of perspective distortion, 2) the semantic content, and 3) color depth. Experiments showed that the accuracy of CDFSIC methods is dependent on the degree of similarity between the source and target domain. While Chen *et al.* [7] proposed a 2-stage training approach (pretrain  $\rightarrow$  metatrain), Hu *et al.* [24] introduced a 3-stage training pipeline (pretrain  $\rightarrow$  metatrain  $\rightarrow$  finetune). Hu *et al.* also evaluated the effectiveness of various feature extraction networks and showed that Vision Transformer [27] performs better than standard convolutional networks [37] and residual networks (ResNets) [20].

Compared to traditional FSIC approaches, methods that leverage extra data are useful but computationally demanding. Therefore, data generation methods that are less computationally intensive have been introduced for the CDFSIC task.

**Data Generation.** Data generation refers to generating new labeled data through commonly-used data synthesis techniques, such as MixUp [63], geometric transformations [45], *etc.*

Fu *et al.* [16] propose a feature-wise domain adaptation module called Feature Distribution Matching (FDM) to guide the MixUp process. FDM measures the discrepancy between the feature distributions of the source and target domain and encourages the model to generate synthetic samples that are more similar to the target domain. Zhang *et al.* [64] and Deng *et al.* [13] apply rotation transformations to images and predict the rotation angle in the pretrain phase. Mazumder *et al.* [34] proposed the composite rotation auxiliary task as a data generation method for the CDFSIC task. This method involves two levels of rotation on the image: first, rotating patches within the image (inner rotation); and

then rotating the entire image (outer rotation) before assigning a rotation class to the transformed image for the model to learn to predict via self-supervision.

Although data generation methods require less computing effort and are easy to implement, they face limitations in significantly improving classification accuracy since the generated samples are derived from the original dataset. Therefore, while data generation methods may be used to boost accuracies of the CDFSIC task, their performance is relatively limited when compared to methods that utilize additional training data.

### 3.2 Feature Alignment Methods

To address data scarcity issue in CDFSIC, data augmentation based method essentially enhances the diversity of samples by expanding the sample space. To handle the problem of domain shift [56] in CDFSIC, feature alignment methods aims to align the features extracted from the source domain with those extracted from the target domain. We summarize the existing feature alignment based method by casting them into two categories: *network architecture design* and *training strategy improvement*.

**Network Architecture Design.** Network architecture design refers to designing or refining the model structure to enhance the ability of the model to generalize the source domain feature characteristics to the target domain. We summarize the existing network architecture design methods as follows:

- **Graph Neural Networks (GNN)** [44] are widely used in graph analysis due to their better scalability and interpretation comparing to traditional graph learning algorithms, such as, Graph Signal Processing, Random Walk and Matrix Factorization. In FSIC, researchers usually take an image as a node of the GNN, while the similarity of image pairs is considered as an edge of the GNN [43]. GNN-based methods parameterize the metric function in FSIC task, allowing a closer fit to the realistic metric function between image pairs. A number of excellent works have emerged in traditional FSIC tasks [28, 43, 59], and CDFSIC.

To alleviate the issue of information loss with the increasing number of the GNN layer and improve the graph-structured data features representation quality, Liu *et al.* [33] propose a geometric algebra graph neural network (GA-GNN) that maps graph nodes to a high-dimensional geometric algebraic space, allowing for a better measurement of the discrepancy between image pairs. Chen *et al.* [8] introduce a Flexible Graph Neural Network (FGNN) that adaptively selects the node feature dimensions to enhance the relevance between image pairs. Most current methods for domain alignment focus on utilizing local spatial information while neglecting the strong correspondence of non-local spatial information (non-local relationships). Accordingly, Zhang *et al.* [67] present a Dual Graph Cross-domain Few-shot Learning (DG-CFSL) framework to learn the domain distribution properties and mitigate

the domain shift, specifically, optimize the dual graph, feature graph and distribution graph simultaneously to achieve domain alignment.

The fundamental concept of the CDFSIC methods based on GNNs is to iteratively update the node features and deduce the relationships between nodes. It features strong interpretability [43] and exhibits great classification performance, but demands significant computational and memory resources. As every two images require the construction of an edge, the memory and computational cost will increase quadratically with the number of samples during inference. Therefore, in CDFSIC tasks, GNN-based method still suffer from the aforementioned limitations that merits further research and improvement.

- **Model Ensembling** [42] is considered as the state-of-the-art solution for many machine learning challenges, aiming to merge multiple models in some way (*e.g.*, voting, averaging, stacking, *etc.*) to extract their strengths and improve the generalization performance of the final model.

Liu *et al.* [31] have put forth a proposal for the CDFSIC task, which involves using an ensemble model with feature transformation. Specifically, they suggested constructing a prediction model by performing diverse feature transformations after extracting features using a network. While Liu *et al.* [31] ensemble the feature extractor, Adler *et al.* [1] integrate from the classifier perspective. In CDFSIC, domain shifts can cause a significant divergence in high-level concepts between the source and target domain. However, low-level concepts, such as image edges, may still retain relevance and applicability. To tackle the challenge, Adler *et al.* [1] introduce a novel approach called Cross-domain Hebbian Ensemble Few-shot learning (CHEF) that utilizes an ensemble of Hebbian learners, which operate on different layers of a deep neural network to merge representations. Through the fusion process, CHEF facilitates the transfer of useful low-level features while accommodating high-level concept shifts.

In CDFSIC tasks, ensemble of multiple models trained across different scenarios can equip algorithms with diverse knowledge of various scenes, effectively addressing the issue of limited generalization ability of models. However, it is important to note that the training of ensembles incurs significant computational and storage costs that increase linearly with the number of scenarios.

- The **Attention Mechanism** [5] in neural networks draws inspiration from the physiological perception of the environment by humans. For example, our visual system tends to selectively focus on certain parts of the visual field while disregarding irrelevant information. Similarly, in various natural language scenarios, some parts of the input to the model are more important than others. The attention mechanism allows for the selective processing of model features, enhancing the model’s generalization performance.

Hou *et al.* [22] propose a novel attention module to tackle the problem of generalization to novel classes, known as the Cross Attention Module (CAM). The CAM generates cross attention maps for each pair of class feature and query sample feature, with the aim of highlighting the relevant object regions and enhancing the discriminative power of the extracted features. The innovative method shows promising results in improving the performance of various

computer vision tasks, particularly in scenarios where generalization to new categories is required. Ye *et al.* [62] introduce an innovative attention method to customize instance embeddings for a given classification task using a set-to-set function. This approach generates task-specific embeddings that are also highly discriminative. To determine the most effective set-to-set functions, they conducted empirical investigations on several variations and discovered that the Transformer [27] was the best option. This is because the Transformer inherently satisfies the key properties required for the desired model. According to Liu *et al.* [32], model ensemble is an effective method for tackling the CDFSIC task. However, when combining models trained on different domains, it is important to take into account that the ratio of model parameter weights should not be equal in the final model. To address this issue, they propose a task-adaptive model weight method, which involves fixing the parameters of all feature extractors after training on the source domain, and subsequently training an attention structure. Sa *et al.* [41] present a simple and effective model for Attentive Fine-Grained Recognition (AFGR). They introduce a residual attention module (RAM) [54] that is integrated into the feature encoder of the residual network. This module enhances various semantic features linearly, enabling the metric function to locate fine-grained feature information better in an image.

Attention mechanism has been demonstrated effective to enhance the interpretability of CDFSIC algorithms and improve the semantic representation capabilities of models. As such, we believe that there is still considerable untapped potential for its application in this field. One potential future research direction is to explore the combination of attention mechanism with feature disentanglement [40] to propose more sophisticated and effective attention mechanisms. By doing so, we can further improve the accuracy and interpretability of CDFSIC methods.

**Training Strategy Improvement.** Training strategy improvement refers to improving the model performance during the model training process to align the source domain features with the target domain features. We summarize the existing training strategies as follows:

- **Parameter Fine-tuning** [23] is a machine learning technique that involves modifying the parameters of a pretrained model to adapt it to a new dataset while focusing on a specific task.

Chen *et al.* [7] propose two simple baselines, which provides the first evidence of the powerful capabilities of fine-tuning in CDFSIC. Similarly, Guo *et al.* [18] use a straightforward fine-tuning approach but differed from Chen *et al.* [7] by fixing the low-dimensional feature layer of the feature extractor during fine-tuning on the target domain on the last three layers. Meanwhile, Cai *et al.* [4] propose a meta fine-tuning mechanism, which utilizes a meta-learning [15] approach to initialize the weights that need to be fine-tuned, rather than directly fine-tuning an incompletely pretrained model. Reinitialization [65] has been widely explored in the natural language field, especially

in the BERT [14] model. Oh *et al.* [35] propose a method for CDFSIC that involves re-initializing the final residual block of the feature extractor before fine-tuning on the target domain. This is done after supervised training on the source domain. This approach reduces learning bias towards the source domain by simply re-initializing specific layers for a given domain, providing a fresh perspective for fine-tuning on CDFSIC.

Fine-tuning the parameters of a model can rapidly assist it in adapting to new scenarios and effectively align the features of both the source and target domains, making it a crucial technique for tackling cross-domain issues. In the case of CDFSIC tasks, there is still ample scope for further research in parameter fine-tuning.

- **Contrastive Learning.** In recent years, a new paradigm of Self-Supervised Learning (SSL) [26] called Contrastive Learning (CL) [36] has emerged as an effective tool for unsupervised learning. CL generates a similarity distribution of data by comparing pairs of samples, and adjusts the model parameters accordingly. By optimizing the contrastive loss [19], the model is encouraged to extract more similar features from pairs of samples in the same class, while features from pairs of samples in different classes are encouraged to be more disperse.

Zhang *et al.* [66] employ the AmdimNet [6] as backbone for training, which utilizes contrastive loss maximization on the mutual information between two new views generated from the same image. Das *et al.* [10] propose a Contrastive Learning and Feature Selection System (ConFeSS) for CDFSIC. ConFeSS optimizes in pretrain stage by contrastive loss and fine-tunes using sample pairs with masked relevant classification features to addresses the issue of overfitting and achieves improved performance. In order to mitigate overfitting, Das *et al.* [11] propose a new fine-tuning method that relies on contrastive loss. This approach utilizes unlabelled examples from the source domain as distractors, which serves to repurpose them and prevent overfitting. In the CDFSIC, the use of contrastive loss can enhance model’s ability to generalize by effectively leveraging the representation in unlabelled data to pull together intra-class samples and push apart inter-class ones. As a result, contrastive loss holds practical value in realistic scenarios where ample unlabelled data is available. However, due to the absence of explicit supervision, contrastive loss is susceptible to problems such as slow convergence and instability, necessitating further investigation.

- **Data Normalization** [46] is a crucial technique in data processing that involves mapping data into a common scale. It is especially important when dealing with data from different sources, as it allows for easier comparison and analysis. In the context of CDFSIC, images from the source and target domains usually exhibit significant differences in terms of style, color, and quality. These differences could have a negative impact on the model’s ability to generalize well to new data.

Wang *et al.* [55] and Xu *et al.* [58] both normalize the extracted image features before classification to reduce the discrepancy between samples from the source and target domains. However, they employ different normaliza-



tion techniques. Wang *et al.* [55] standardize the feature vectors using 1, 2, 3, and  $\infty$  p-norms, while Xu *et al.* [58] use two learnable parameters  $\gamma, \beta$  for Instance Normalization  $IN(F) = \gamma \frac{F - \mu(F)}{\sigma(F)} + \beta$ , where  $F$  refer to the image feature,  $\mu(\cdot)$  and  $\sigma(\cdot)$  denote the mean and standard deviation calculated at the channel level for each sample. Yazdanpanah *et al.* [60, 61] and Tseng *et al.* [49] make improvements to the Batch Normalization (BN) Layer in the feature extraction network. According to Yazdanpanah *et al.* [61], the use of trainable parameters in the BN layer of convolutional neural networks will lead to a shift in the distribution of batch data, while also improving the convergence rate during training on the source domain. However, it may not generalize well to the target domain, which can limit classification performance. To address the issue, Yazdanpanah *et al.* [61] replaced the BN layer in the convolutional network with a Feature Normalization (FN) layer,  $FN(h_c) = \frac{h_c - \mu_c}{\sqrt{\sigma_c^2 + \epsilon}}$ , Here,  $h_c$  denotes batch data feature,  $\mu_c$  and  $\sigma_c$  are the first and second moments [38] of  $h_c$ . In contrast to the BN layer, the FN layer discards the trainable parameters for shifting and scaling. In their subsequent work, Yazdanpanah *et al.* [60] propose that the parameters within the BN layer are trained using source domain data, leading to a potential mismatch between the internal BN parameters and the data distribution during inference caused by domain shift. To tackle the issue, they introduce a Visual Domain Bridge (VDB) that replaces the statistical mean and variance of the target domain data with those of the source domain, generating a transformed data feature, then fine-tune the model using the transformed feature to alleviate the mismatch between the BN layer’s internal parameters and the target domain’s data distribution. Tseng *et al.* [49] propose adding a Feature-Wise Transformation (FWT) layer after the BN layer in convolutional neural networks to simulate feature distributions in different domains, improving the generalization ability of the feature extractor.

Data normalization is crucial for improving image classification accuracy. It helps the model converge in cross-domain scenarios and aligns the feature distributions of the source and target domains by reducing distribution discrepancies. Therefore, data normalization is a practical method to enhance the generalization ability of the model in CDFSIC task.

- **Dropout** is a commonly-used technique in deep learning to regularize training. Hinton *et al.* [21] point out that over-parameterization of the model can easily lead to overfitting, while dropout can effectively alleviate overfitting and to some extent act as regularization, improving the performance of the network.

According to Huang *et al.* [25], dropout can be a useful technique in CDFSIC. By dropping out the activations of the most important features in the training data, the network is forced to activate the second most important features that are related to the labels. This approach can effectively unlock the potential of the network, leading to enhanced generalization performance. Tu *et al.* [50] propose a simple and effective dropout-style method to enhance model trained on low-complexity concepts from the source domain. The app-

roach involves sampling multiple sub-networks by dropping neurons or feature maps to create a diverse set of models with varied features for the target domain. The most suitable sub-networks are selected to form an ensemble for target domain learning. This method enables the model to generalize better to the target domain, where it may encounter novel and complex concepts. In conclusion, dropout can effectively alleviate overfitting on CDFSIC task without increasing computational or memory overhead.

## 4 CDFSIC Dataset and Application

### 4.1 Standard Datasets

Currently, in CDFSIC, the datasets used in different literature are not entirely consistent. Table 1 shows three commonly-used benchmark datasets.

**Table 1.** Standard Dataset of CDFSIC

Dataset	Published In	Code/Data Link
MiniImageNet $\rightarrow$ CUB [7]	ICLR 19	<a href="https://github.com/wyharveychen/CloserLookFewShot">https://github.com/wyharveychen/CloserLookFewShot</a>
BSCDFSL [18]	ECCV 20	<a href="https://github.com/IBM/cdfls-benchmark">https://github.com/IBM/cdfls-benchmark</a>
Meta-Dataset [48]	ICLR 20	<a href="https://github.com/google-research/meta-dataset">https://github.com/google-research/meta-dataset</a>

MiniImageNet  $\rightarrow$  CUB and BSCDFSL are widely-used datasets in recent works. Due to the late release of MetaDataset, there are only a few works evaluated on this dataset.

### 4.2 CDFSIC Application

CDFSIC algorithms have already found applications in various fields, including medical imaging such as X-ray images [9], skin disease images [17], and satellite remote sensing images [2] as well as hyperspectral images [68]. Moreover, we foresee that CDFSIC algorithms have immense potential in other domains, such as aerospace, cultural heritage preservation, and public safety.

## 5 Limitations and Future Research Directions

In recent years, there are some advancements in addressing the problem of CDFSIC, particularly on challenges related to data scarcity and domain shift between source and target domain. However, despite these developments, there are still other limitations that need to be overcome in this field.

## 5.1 Limitations of the Current FSIC Settings

Currently, FSIC tasks generally follow  $N$ -way  $K$ -shot ( $M$ -query) setting, where  $N$  refers to the number of image categories in a sub-task, and  $K$  refers to the number of samples in each category contained in the support set.  $N$ -way  $K$ -shot setting is reasonable for real-world scenarios because the number of samples for each category in the support set can be artificially set when creating the dataset. However, in testing phase, the number of samples for each category in the query set may not be the same, denoted by  $M$ . Furthermore, we cannot predict the distribution of the query data easily, nor can we assume that it is evenly distributed among each category.

Veilleux *et al.* [51] propose to use Dirichlet Distribution to simulate imbalanced sample distribution for each category in the query set of a sub-task, making it closer to real-world scenarios. We believe that addressing imbalanced FSIC is an important area of future research.

## 5.2 Theoretical Insights

In the field of CDFSIC, current state-of-the-art algorithms are usually developed through empirical exploration, without sufficient theoretical guidance. For traditional FSIC tasks, various theoretical derivations have been proposed [15, 39]. However, for CDFSIC, current research merely combines traditional FSIC naively with cross-domain techniques. Therefore, there is an urgent need for future research that provides theoretical support for CDFSIC.

## 5.3 Cross-Hardware CDFSIC

In addition to the CDFSIC issues mentioned above, Zhao *et al.* [69] further explore the cross-hardware scenario of FSIC, optimizing the inference latency of the model on hardware devices such as GPUs, ASICs, and IoT platforms. As cross-domain scenarios do not require training and testing data to have consistent distributions, we anticipate that it is even more necessary for CDFSIC algorithms to optimize performance for hardware in order to meet its wider application prospects.

## 6 Conclusion

In the field of image classification, research on FSIC has recently extended to CDFSIC. This paper provides a detailed overview of the current state of research on CDFSIC, while analyzing the challenges faced by such research and providing a perspective on its future prospects.

**Acknowledgments.** This work is supported in part by National Key R&D Program of China (No. 2019YFB2102100), Key-Area Research and Development Program of Guangdong Province (No. 2020B010164003), and Shenzhen Science and Technology Innovation Commission (No. JCYJ20190812160003719).

## References

1. Adler, T., et al.: Cross-domain few-shot learning by representation fusion. arXiv preprint [arXiv:2010.06498](https://arxiv.org/abs/2010.06498) (2020)
2. Ammour, N., Bashmal, L., Bazi, Y., Al Rahhal, M.M., Zuair, M.: Asymmetric adaptation of deep features for cross-domain classification in remote sensing imagery. *IEEE Geosci. Remote Sens. Lett.* **15**(4), 597–601 (2018)
3. Blanchard, G., Lee, G., Scott, C.: Generalizing from several related classification tasks to a new unlabeled sample. In: *Advances in Neural Information Processing Systems*, vol. 24 (2011)
4. Cai, J., Shen, S.M.: Cross-domain few-shot learning with meta fine-tuning. arXiv preprint [arXiv:2005.10544](https://arxiv.org/abs/2005.10544) (2020)
5. Chaudhari, S., Mithal, V., Polatkan, G., Ramanath, R.: An attentive survey of attention models. *ACM Trans. Intell. Syst. Technol. (TIST)* **12**(5), 1–32 (2021)
6. Chen, D., Chen, Y., Li, Y., Mao, F., He, Y., Xue, H.: Self-supervised learning for few-shot image classification. In: *ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1745–1749. IEEE (2021)
7. Chen, W.Y., Liu, Y.C., Kira, Z., Wang, Y.C.F., Huang, J.B.: A closer look at few-shot classification. In: *International Conference on Learning Representations* (2018)
8. Chen, Y., et al.: Cross-domain few-shot classification based on lightweight res2net and flexible GNN. *Knowl.-Based Syst.* **247**, 108623 (2022)
9. Cohen, J.P., Hashir, M., Brooks, R., Bertrand, H.: On the limits of cross-domain generalization in automated x-ray prediction. In: *Medical Imaging with Deep Learning*, pp. 136–155. PMLR (2020)
10. Das, D., Yun, S., Porikli, F.: Confess: a framework for single source cross-domain few-shot learning. In: *International Conference on Learning Representations* (2022)
11. Das, R., Wang, Y.X., Moura, J.M.: On the importance of distractors for few-shot classification. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9030–9040 (2021)
12. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255. IEEE (2009)
13. Deng, S., Liao, D., Gao, X., Zhao, J., Ye, K.: Improving few-shot image classification with self-supervised learning. In: Ye, K., Zhang, L.J. (eds.) *CLOUD 2022*. LNCS, vol. 13731, pp. 54–68. Springer, Cham (2022). [https://doi.org/10.1007/978-3-031-23498-9\\_5](https://doi.org/10.1007/978-3-031-23498-9_5)
14. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. arXiv preprint [arXiv:1810.04805](https://arxiv.org/abs/1810.04805) (2018)
15. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: *International Conference on Machine Learning*, pp. 1126–1135. PMLR (2017)
16. Fu, Y., Fu, Y., Jiang, Y.G.: Meta-fdmixup: cross-domain few-shot learning guided by labeled target data. In: *Proceedings of the 29th ACM International Conference on Multimedia*, pp. 5326–5334 (2021)
17. Gu, Y., Ge, Z., Bonnington, C.P., Zhou, J.: Progressive transfer learning and adversarial domain adaptation for cross-domain skin disease classification. *IEEE J. Biomed. Health Inform.* **24**(5), 1379–1393 (2019)

18. Guo, Y., et al.: A broader study of cross-domain few-shot learning. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020, Part XXVII. LNCS, vol. 12372, pp. 124–141. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-58583-9\\_8](https://doi.org/10.1007/978-3-030-58583-9_8)
19. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9729–9738 (2020)
20. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
21. Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R.: Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint [arXiv:1207.0580](https://arxiv.org/abs/1207.0580) (2012)
22. Hou, R., Chang, H., Ma, B., Shan, S., Chen, X.: Cross attention network for few-shot classification. In: Advances in Neural Information Processing Systems, vol. 32 (2019)
23. Howard, J., Ruder, S.: Universal language model fine-tuning for text classification. arXiv preprint [arXiv:1801.06146](https://arxiv.org/abs/1801.06146) (2018)
24. Hu, S.X., Li, D., Stühmer, J., Kim, M., Hospedales, T.M.: Pushing the limits of simple pipelines for few-shot learning: external data and fine-tuning make a difference. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9068–9077 (2022)
25. Huang, Z., Wang, H., Xing, E.P., Huang, D.: Self-challenging improves cross-domain generalization. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020, Part II. LNCS, vol. 12347, pp. 124–140. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-58536-5\\_8](https://doi.org/10.1007/978-3-030-58536-5_8)
26. Jing, L., Tian, Y.: Self-supervised visual feature learning with deep neural networks: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **43**(11), 4037–4058 (2020)
27. Khan, S., Naseer, M., Hayat, M., Zamir, S.W., Khan, F.S., Shah, M.: Transformers in vision: a survey. *ACM Comput. Surv. (CSUR)* **54**(10s), 1–41 (2022)
28. Kim, J., Kim, T., Kim, S., Yoo, C.D.: Edge-labeling graph neural network for few-shot learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11–20 (2019)
29. Lake, B.M., Salakhutdinov, R., Tenenbaum, J.B.: Human-level concept learning through probabilistic program induction. *Science* **350**(6266), 1332–1338 (2015)
30. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2015)
31. Liu, B., Zhao, Z., Li, Z., Jiang, J., Guo, Y., Ye, J.: Feature transformation ensemble model with batch spectral regularization for cross-domain few-shot classification. arXiv preprint [arXiv:2005.08463](https://arxiv.org/abs/2005.08463) (2020)
32. Liu, L., Hamilton, W., Long, G., Jiang, J., Larochelle, H.: A universal representation transformer layer for few-shot image classification. arXiv preprint [arXiv:2006.11702](https://arxiv.org/abs/2006.11702) (2020)
33. Liu, Q., Cao, W.: Geometric algebra graph neural network for cross-domain few-shot classification. *Appl. Intell.* **52**(11), 12422–12435 (2022)
34. Mazumder, P., Singh, P., Namboodiri, V.P.: Few-shot image classification with composite rotation based self-supervised auxiliary task. *Neurocomputing* **489**, 179–195 (2022)

35. Oh, J., Kim, S., Ho, N., Kim, J.H., Song, H., Yun, S.Y.: Refine: re-randomization before fine-tuning for cross-domain few-shot learning. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management, pp. 4359–4363 (2022)
36. Oord, A.V.D., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint [arXiv:1807.03748](https://arxiv.org/abs/1807.03748) (2018)
37. O’Shea, K., Nash, R.: An introduction to convolutional neural networks. arXiv preprint [arXiv:1511.08458](https://arxiv.org/abs/1511.08458) (2015)
38. Papoulis, A., Unnikrishna Pillai, S.: Probability, random variables and stochastic processes (2002)
39. Rajeswaran, A., Finn, C., Kakade, S.M., Levine, S.: Meta-learning with implicit gradients. In: Advances in Neural Information Processing Systems, vol. 32 (2019)
40. Ren, J., Li, M., Liu, Z., Zhang, Q.: Disentanglement, visualization and analysis of complex features in DNNs (2020)
41. Sa, L., Yu, C., Ma, X., Zhao, X., Xie, T.: Attentive fine-grained recognition for cross-domain few-shot classification. *Neural Comput. Appl.* **34**(6), 4733–4746 (2022)
42. Sagi, O., Rokach, L.: Ensemble learning: a survey. *Wiley Interdiscipl. Rev. Data Min. Knowl. Discov.* **8**(4), e1249 (2018)
43. Satorras, V.G., Estrach, J.B.: Few-shot learning with graph neural networks. In: International Conference on Learning Representations (2018)
44. Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M., Monfardini, G.: The graph neural network model. *IEEE Trans. Neural Networks* **20**(1), 61–80 (2008)
45. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. *J. Big Data* **6**(1), 1–48 (2019)
46. Sun, J., Cao, X., Liang, H., Huang, W., Chen, Z., Li, Z.: New interpretations of normalization methods in deep learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 5875–5882 (2020)
47. Torralba, A., Efros, A.A.: Unbiased look at dataset bias. In: CVPR 2011, pp. 1521–1528. IEEE (2011)
48. Triantafillou, E., et al.: Meta-dataset: a dataset of datasets for learning to learn from few examples. arXiv preprint [arXiv:1903.03096](https://arxiv.org/abs/1903.03096) (2019)
49. Tseng, H.Y., Lee, H.Y., Huang, J.B., Yang, M.H.: Cross-domain few-shot classification via learned feature-wise transformation. arXiv preprint [arXiv:2001.08735](https://arxiv.org/abs/2001.08735) (2020)
50. Tu, P.C., Pao, H.K.: A dropout style model augmentation for cross domain few-shot learning. In: 2021 IEEE International Conference on Big Data (Big Data), pp. 1138–1147. IEEE (2021)
51. Veilleux, O., Boudiaf, M., Piantanida, P., Ben Ayed, I.: Realistic evaluation of transductive few-shot learning. *Adv. Neural. Inf. Process. Syst.* **34**, 9290–9302 (2021)
52. Vinyals, O., Blundell, C., Lillicrap, T., Wierstra, D., et al.: Matching networks for one shot learning. In: Advances in Neural Information Processing Systems, vol. 29 (2016)
53. Wah, C., Branson, S., Welinder, P., Perona, P., Belongie, S.: The caltech-ucsd birds-200-2011 dataset (2011)
54. Wang, F., et al.: Residual attention network for image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3156–3164 (2017)

55. Wang, H., et al.: Experiments in cross-domain few-shot learning for image classification. In: ECMLPKDD Workshop on Meta-Knowledge Transfer, pp. 81–83. PMLR (2022)
56. Wang, M., Deng, W.: Deep visual domain adaptation: a survey. *Neurocomputing* **312**, 135–153 (2018)
57. Wang, Y., Yao, Q., Kwok, J.T., Ni, L.M.: Generalizing from a few examples: a survey on few-shot learning. *ACM Comput. Surv. (CSUR)* **53**(3), 1–34 (2020)
58. Xu, Y., Wang, L., Wang, Y., Qin, C., Zhang, Y., Fu, Y.: Memrein: rein the domain shift for cross-domain few-shot learning (2021)
59. Yang, L., Li, L., Zhang, Z., Zhou, X., Zhou, E., Liu, Y.: DPGN: distribution propagation graph network for few-shot learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13390–13399 (2020)
60. Yazdanpanah, M., Moradi, P.: Visual domain bridge: a source-free domain adaptation for cross-domain few-shot learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2868–2877 (2022)
61. Yazdanpanah, M., Rahman, A.A., Desrosiers, C., Havaei, M., Belilovsky, E., Kahou, S.E.: Shift and scale is detrimental to few-shot transfer. In: NeurIPS 2021 Workshop on Distribution Shifts: Connecting Methods and Applications (2021)
62. 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 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8808–8817 (2020)
63. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: beyond empirical risk minimization. arXiv preprint [arXiv:1710.09412](https://arxiv.org/abs/1710.09412) (2017)
64. Zhang, Q., Jiang, Y., Wen, Z.: TACDFSL: task adaptive cross domain few-shot learning. *Symmetry* **14**(6), 1097 (2022)
65. Zhang, T., Wu, F., Katiyar, A., Weinberger, K.Q., Artzi, Y.: Revisiting few-sample BERT fine-tuning. arXiv preprint [arXiv:2006.05987](https://arxiv.org/abs/2006.05987) (2020)
66. Zhang, Y., Zheng, Y., Xu, X., Wang, J.: How well do self-supervised methods perform in cross-domain few-shot learning? arXiv preprint [arXiv:2202.09014](https://arxiv.org/abs/2202.09014) (2022)
67. Zhang, Y., Li, W., Zhang, M., Tao, R.: Dual graph cross-domain few-shot learning for hyperspectral image classification. In: ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3573–3577. IEEE (2022)
68. Zhang, Y., Li, W., Zhang, M., Wang, S., Tao, R., Du, Q.: Graph information aggregation cross-domain few-shot learning for hyperspectral image classification. *IEEE Trans. Neural Networks Learn. Syst.* (2022)
69. Zhao, Y., Gao, X., Shumailov, I., Fusi, N., Mullins, R.: Rapid model architecture adaption for meta-learning. *Adv. Neural. Inf. Process. Syst.* **35**, 18721–18732 (2022)
70. Zhou, K., Liu, Z., Qiao, Y., Xiang, T., Loy, C.C.: Domain generalization: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**, 4396–4415 (2022)