AI

MDPI

Review

Few-Shot Fine-Grained Image Classification: A Comprehensive Review

Jie Ren [1], Changmiao Li [1], Yaohui An [1], Weichuan Zhang [2,*] and Changming Sun [3]

[1] College of Electrical and Information, Xi'an Polytechnic University, Xi'an 710048, China; renjie@xpu.edu.cn (J.R.); changmiaoli@163.com (C.L.); yaohui_22@163.com (Y.A.)
[2] School of Electronic Information and Artificial Intelligence, Shaanxi University of Technology and Science, Xi'an 710021, China
[3] CSIRO Data61, P.O. Box 76, Epping, NSW 1710, Australia; changming.sun@csiro.au
* Correspondence: zwc2003@163.com

Abstract: Few-shot fine-grained image classification (FSFGIC) methods refer to the classification of images (e.g., birds, flowers, and airplanes) belonging to different subclasses of the same species by a small number of labeled samples. Through feature representation learning, FSFGIC methods can make better use of limited sample information, learn more discriminative feature representations, greatly improve the classification accuracy and generalization ability, and thus achieve better results in FSFGIC tasks. In this paper, starting from the definition of FSFGIC, a taxonomy of feature representation learning for FSFGIC is proposed. According to this taxonomy, we discuss key issues on FSFGIC (including data augmentation, local and/or global deep feature representation learning, class representation learning, and task-specific feature representation learning). In addition, the existing popular datasets, current challenges and future development trends of feature representation learning on FSFGIC are also described.

Keywords: few-shot fine-grained image classification; feature representation learning; meta-learning; metric-learning

Citation: Ren, J.; Li, C.; An, Y.; Zhang, W.; Sun, C. Few-Shot Fine-Grained Image Classification: A Comprehensive Review. *AI* **2024**, *5*, 405–425. https://doi.org/10.3390/ai5010020

Academic Editor: Arslan Munir

Received: 11 December 2023
Revised: 1 March 2024
Accepted: 4 March 2024
Published: 6 March 2024

1. Introduction

Few-shot fine-grained image classification (FSFGIC) methods [1] refer to the classification of images (e.g., birds [2], flowers [3], and airplanes [4]) belonging to different subclasses of the same species by a small number of labeled samples. As illustrated in Figure 1, image classification tasks can be divided into coarse-grained image classification (CGIC) and fine-grained image classification (FGIC) according to different classification granularity. CGIC is a task of cross-species classification, and these classes usually have obvious differences in appearance characteristics, with the characteristics of large inter-class differences and small intra-class differences. FGIC is a classification task of different subclasses of the same species, and the differences between these classes may be very small, with the characteristics of small inter-class differences and large intra-class differences.

The researchers found that two-year-old children can classify objects into different categories after viewing just a few images, but the child may be confused about fine-grained image classification with a limited number of samples [5,6], due to the following reasons: (1) Objects for FSFGIC are obtained from sub-categories of one category, making them visually very similar. Some images may differ only in subtle visual features, requiring experts in the field to distinguish between specific categories; (2) Samples are affected by factors such as background, pose, occlusion, light intensity, and shooting angle, the differences between different subclasses may be small, and the differences within the same subclass may be greater, resulting in a classification problem of small inter-class differences and large intra-class differences.

Fine-grained image datasets usually have a small number of samples and need domain experts to label the datasets. However, the traditional image classification algorithm requires a large amount of labeled data for model training, which is obviously not suitable for FGIC tasks. Therefore, how to use few-shot learning to complete FGIC tasks is a research hotspot in this field. Since the objects in different sub-categories of the same entry-level category are very similar to each other, a key consideration in FSFGIC is how to effectively learn discriminative features from extremely limited training samples, which makes FSFGIC a very challenging research problem.

Figure 1. Comparison of coarse-grained image classification (CGIC) and fine-grained image classification (FGIC).

Recently, with the growing attention on FSFGIC, various FSFGIC methods have been proposed. Many few-shot learning methods have also been applied to handle FSFGIC tasks with impressive results. Currently, there is no survey about FSFGIC. This paper aims to fill this gap. It is worth noting that the quality of feature representation learning directly affects the classification performance on FSFGIC. The reason is that the quality of feature representation learning determines whether the FSFGIC methods can make better use of the limited sample information and learn more discriminant feature representations, thus greatly improving the classification accuracy and generalization ability of the FSFGIC methods. In this way, a taxonomy of feature representation learning for FSFGIC is proposed. According to this classification, we discuss different types of FSFGIC methods in depth. It is worth to note that those few-shot image classification algorithms (e.g., [7,8]) that have achieved good classification performance in some FSFGIC datasets are also introduced in this survey.

The contributions of this survey comprise the following aspects. This is the first work to review FSFGIC under a taxonomy of feature representation learning. Subsequently, different types of feature representation learning techniques for FSFGIC are reviewed. Additionally, the relationships among different FSFGIC methods are presented. Furthermore, combining with representative existing FSFGIC techniques, the main unresolved issues on FSFGIC are discussed.

2. Problem, Datasets, and Categorization of FSFGIC Methods

In this section, the problem formulation of FSFGIC, categorization of FSFGIC methods, and representative benchmark datasets for FSFGIC are presented.

2.1. Problem Formulation

For an FSFGIC task, the dataset $\mathcal{D}$ is typically divided into a training set $\mathcal{D}_{train}$, a validation set $\mathcal{D}_{val}$, and a test set $\mathcal{D}_{test}$. The $\mathcal{D}_{train}$ is used to train the parameters of the model, the $\mathcal{D}_{val}$ is used to verify and tune the model, and the $\mathcal{D}_{test}$ is used to finally evaluate the accuracy of the FSFGIC method. That is, the three stages of training, validation, and testing of the model. Each stage consists of many epochs, each containing thousands of episodes.

$$\mathcal{D} = \{\mathcal{D}_{train} \cup \mathcal{D}_{val} \cup \mathcal{D}_{test}\}, \tag{1}$$

where $\mathcal{D}_{train} \cap \mathcal{D}_{val} = \emptyset$, $\mathcal{D}_{train} \cap \mathcal{D}_{test} = \emptyset$, and $\mathcal{D}_{val} \cap \mathcal{D}_{test} = \emptyset$.

The FSFGIC task is denoted as a C-way K-shot task, which means that C categories are selected in each episode, K samples in each category are selected as support samples, and part of the remaining samples in the C categories are selected as query samples. Each episode's dataset $\mathcal{D}_{episode}$ consists of a support set $\mathcal{S}$ consisting of $C \times K$ labeled support samples and a query set $\mathcal{Q}$ consisting of $C \times J$ unlabeled query samples.

$$\mathcal{D}_{episode} = \{\mathcal{S} = \{(x_i, y_i)_{i=1}^{C \times K}\} \cup \mathcal{Q} = \{(x_j)_{j=1}^{C \times J}\}\}, \tag{2}$$

where $x_i \cap x_j = \emptyset$, x_i and x_j denote fine-grained samples and $(x_i, x_j) \in C$, and $y_i \in C$ represents the ground truth label of x_i.

The purpose of the FSFGIC method is to successfully predict the category of x_j using x_i and y_i. The evaluation criterion of FSFGIC method is classification accuracy, which is calculated by dividing the number of successfully predicted query samples by the total number of query samples.

2.2. A Taxonomy of the Existing Feature Representation Learning for FSFGIC

According to the difference of contents and representations of learned features, the existing feature representation learning techniques for FSFGIC can be divided into three categories: local and/or global deep feature representation learning based FSFGIC methods [9,10], class representation learning based FSFGIC methods [11,12], and task-specific feature representation learning based FSFGIC methods [13,14]. According to different types of feature representation learning paradigms, a taxonomy of feature representation learning for FSFGIC methods is illustrated in Figure 2.

Local and/or global deep feature representation learning based FSFGIC methods utilize the degree of difference of the local and/or global deep feature representations between query and support samples for performing FSFGIC tasks. Class representation learning based FSFGIC methods utilize deep feature representations from all training samples in a class to construct a class feature representation (e.g., class-level graph [15] or class-level local deep feature representation [7]) for this class. And then class feature representation is used to perform FSFGIC tasks. Task-specific feature representation learning based FSFGIC methods utilize deep feature representations from all training images in a task (i.e., one training episode) to construct a task-specific feature representation (e.g., task-level graph relationship representation [16] or task-level local deep feature representation [8]) for this task and to perform FSFGIC tasks.

It is worth noting that after feature representation is learned, most meta-learning based techniques, which can be divided into two branches (i.e., optimization-based techniques and metric-based techniques), are utilized for performing FSFGIC tasks. Optimization-based techniques aim to converge the model to novel tasks, which learns how to update the parameters of a given initial model with only a few training samples for each category. Metric-based techniques aim to learn a transferable feature knowledge and obtain a distribution based on similarity metrics between different samples. In this way, for each type of feature representation learning, both optimization-based and metric-based techniques used for FSFGIC will be reviewed in detail.

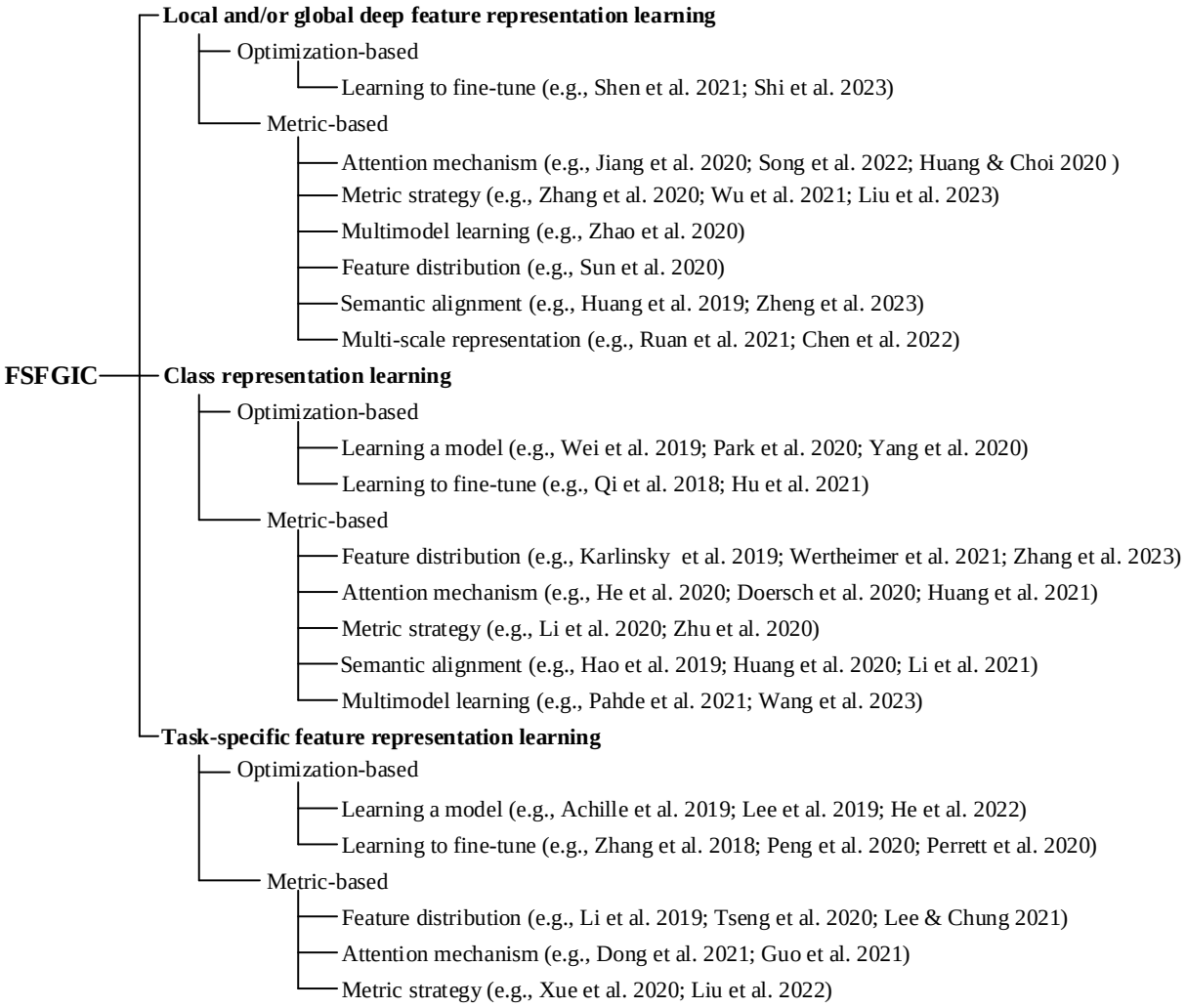

Figure 2. Classification of feature representation learning techniques in existing FSFGIC methods [1,6–8,16–62].

2.3. Benchmark Datasets

Datasets have become one of the most critical roles in the development of FSFGIC, not only as a means for evaluating the classification accuracy of different FSFGIC methods, but also for greatly promoting the development of the field of FSFGIC (e.g., solving more complex, practical, and challenging problems). Currently, the representative datasets for training and evaluation on FSFGIC are CUB-200-2010 [63], CUB-200-2011 [2], Stanford Dogs [64], Stanford Cars [65], FGVC-Aircraft [4], NABirds [66], SUN397 [67], and Oxford 102 Flowers [3]. The number of images and the number of categories corresponding to these datasets are shown in Table 1.

A detailed description of the datasets available on FSFGIC can be accessed at https://paperswithcode.com/task/fine-grained-image-classification (accessed on 1 March 2024). In addition, several ultra-fine-grained image datasets (such as Cottons and Soybeams [68]) exist in this field. Compared with the current widely used FSFGIC datasets (e.g., CUB-200-2011 [2]), the inter-class differences among ultra-fine-grained images are much smaller, which put forward greater requirements on the design of FSFGIC algorithms.

Table 1. Representative benchmark datasets for FSFGIC.

Dataset Name	Class	Images	Categories
CUB-200-2010 [63]	Birds	6033	200
CUB-200-2011 [2]	Birds	11,788	200
Stanford Dogs [64]	Dogs	20,580	120
Stanford Cars [65]	Cars	16,185	196
FGVC-Aircraft [4]	Aircrafts	10,000	100
NABirds [66]	Birds	48,562	555
SUN397 [67]	Scenes	108,754	397
Oxford 102 Flowers [3]	Flowers	8189	102

3. Methods on FSFGIC

In this section, we first review the data augmentation techniques for FSFGIC. Then, local and/or global deep feature representation learning based FSFGIC methods, class representation learning based FSFGIC methods, and task-specific feature representation learning based FSFGIC methods are introduced in detail. Furthermore, the relationships among the classification methods for different types of FSFGIC techniques are also introduced.

3.1. Data Augmentation Techniques for FSFGIC

Data augmentation techniques aim to enhance both the quantity and diversity of training data, thus alleviating overfitting and improving generalization ability. Currently, two types of data augmentation techniques are widely used on FSFGIC. The first type of data augmentation techniques (e.g., random horizontal flipping [50,69], jittering [39,44], scaling [1], random cropping [6], translation [70], zooming [70], and random rotation [56,71]) are used as a basic image manipulation in FSFGIC methods.

The second type of data augmentation techniques [33,72–74] are based on deep learning mechanisms which aim to mimic the characteristics of real data. For example, in [72], generative adversarial networks (GAN) were utilized to generate realistic samples from a given dataset. In [75], a feature encoder–decoder was used to augment the dataset by generating feature representations. In [76], a pre-trained GAN without discriminator was applied to generate subtle features of fine-grained images. And in [77], GAN was used to generate hallucination images. In [45], a self-training strategy was developed with unlabeled data for augmenting data, and in [78], they applied a self-taught learning strategy to measure the credibility of each pseudo-labeled instance. In [27], a fully annotated auxiliary dataset which has similar distribution with the target dataset was used to train a meta learner, which can transfer knowledge from an auxiliary dataset to a target dataset. In [79] a diversity transfer network (DTN) was proposed to learn to transfer latent diversities from training data to testing data. Xu et al. [74] first proposed a variational autoencoder (VAE)-based feature disentanglement method on FSL problems to generate images. Δ-encoder [80] utilized an autoencoder to find deformations between different samples of the same category, then generated new samples for the other categories. Ref. [81] proposed a method of foreground extraction and posture transformation, which can extract the foreground from base classes and generate additional samples for novel sub-classes to realize data expansion. Inspired by the hypothesis that language can help learn new visual objects [82], auxiliary semantic modalities (e.g., attribute annotations [50,83]) were applied for the support set while ignoring the query set. In addition, other data augmentation techniques will be described in detail in the following review of FSFGIC methods.

3.2. Local and/or Global Deep Feature Representation Learning Based FSFGIC Methods

In the field of FSFGIC, some scholars consider that local deep feature representations have the ability to recognize the discriminative regions for distinguishing subtle differences of fine-grained features. Some scholars argue that combining global and local deep feature representation learning can effectively improve the capability of deep feature representation. Currently, there are two main research directions (i.e., optimization-based techniques and

metric-based techniques) which utilize local and/or global deep feature representations for performing FSFGIC tasks as illustrated in the following.

3.2.1. Optimization-Based Local and/or Global Deep Feature Representation Learning

The existing optimization-based methods for local and/or global deep feature representation learning mainly focus on learning fine-tuning techniques. These methods aim to improve the model's performance with limited training samples by integrating the fine-tuning process during the meta-training stage.

Learning to fine-tune. By using multiple attention mechanisms, a multi-attention meta-learning (MattML) method [6] applied attention mechanisms to both the basic learner and the task learner to capture the feature information of subtle and local parts of an image. It was indicated in [17] that some knowledge in the base data may be biased against the new class, so transferring the entire knowledge in the base data to the new class may not obtain a good meta learner or classifier. An evolutionary search strategy was proposed for transferring partial knowledge by fine-tuning particular layers in the base model after obtaining deep feature representations through feature extractor. First, several fine-tuning strategies were randomly generated and their corresponding classification accuracies on the validation set are obtained. K strategies with the highest accuracy were selected as parents. Second, with the help of gene mutation and gene crossover as in an evolutionary algorithm, offspring vectors were obtained and their corresponding classification accuracies were calculated. By repeating this process in iterations, the best fine-tuning strategy can be obtained. This proposed evolutionary search strategy can be embedded into a metric-based method [84] and an optimization-based method [85] for performing FSFGIC tasks. By introducing enhancement methods that combine global and local perception features into the feature space and adding semantic orthogonality constraints, ref. [18] achieved a more comprehensive and accurate representation of image feature information.

3.2.2. Metric-Based Local and/or Global Deep Feature Representation Learning

Metric-based local and/or global deep feature representation learning methods can be classified into six categories: attention mechanism, metric strategy, multimodel learning, feature distribution, semantic alignment, and multi-scale representation.

Attention mechanism. Following the idea that a self-attention mechanism has the ability to indicate the discriminative regions in an image [46], a novel network architecture [86] that incorporated saliency information as input was designed. Local deep feature representations from training samples and their corresponding saliency maps obtained from [87] were combined for improving the classification performance on FSFGIC. Following the idea of object localization strategy [88], a meta-reweighting strategy [19] was designed to extract and exploit local deep feature representations of support samples. Furthermore, an adaptive attention mechanism based on the meta-reweighting model was designed to localize the region of interest in query samples. The aim of the designed adaptive attention mechanism was to match query images and support images to highlight relevant regions of interest for obtaining more discriminative local deep feature representations. A trilinear spatial-awareness network (S3Net) [23] was proposed to strengthen the spatial representation of each local descriptor by adding a global relationship feature with self-attention. They construct the multi-scale features to enhance rich representation in global features. Finally, a local loss and a global loss were combined to learn the discriminative features. In [29], they proposed an attention-based pyramid structure to weight the different areas of the feature maps and produce multi-scaled features. Ref. [20] proposed a fusion spatial attention method that performs spatial attention simultaneously in both the image and the embedded space. Ref. [21] proposed a self-attention based prototype enhancement network (SAPENet) to obtain a more representative prototype for each class. In [89], they proposed an automatic salient region selection network without the use of a bounding box or part annotation mechanism for locating salient regions from images.

Metric strategy. The DeepEMD method [22] formalized the problem of image classification as an optimal image matching problem. And then earth mover's distance (EMD) was applied to select local discriminative feature representations for finding optimal matching between query samples and support samples. In [90], a two-stage comparison strategy was proposed to mine hard examples which correspond to the top two relation scores outputted by the first relation network and then were inputted into a second relation network to distinguish similar classes. A subtle difference module [23] was proposed to classify confused or near-duplicated samples based on the cooperation of local and global similarities between query image and the prototype of each class. Ref. [24] used the Sinkhorn distance to find an optimal matching between images, mitigating the object mismatch caused by misaligned position. Meanwhile, they proposed the intraimage and interimage attentions as the bilateral normalization on the Sinkhorn distance to suppress the object mismatch caused by background clutter.

Multimodel learning. In [25], Zhao et al. argued that cross-modal external knowledge will help improve the classification performance on FSFGIC. In this way, a mirror mapping network (MMN) was designed to map multimodal features (i.e., external knowledge and global and local feature representations) into the same semantic space. The external knowledge which was extracted from textual descriptions and knowledge graph was utilized to generate global and local features for training samples. Finally, global and local feature representations from samples and external knowledge were combined for performing FSFGIC tasks.

Feature distribution. Sun et al. [26] proposed a domain-specific FSFGIC task of marine organisms. They designed a feature fusion model to focus on the key regions. Specifically, the framework consisted of a ConvNet-based feature extractor, a feature fusion model, and a classifier. As the key component, the feature fusion model utilized the focus-area location and high-order integration to generate feature representations which contained more identifiable information.

Semantic alignment. Huang et al. [27] proposed a novel pairwise bilinear pooling to recognize the subtle difference of fine-grained images. Specifically, they designed a fine-grained features extractor which contained an alignment loss regularization and a pair-wise bilinear pooling layer. The alignment loss aimed to match the features of the same position and the pair-wise bilinear pooling layer was able to capture comparative features from pairs of images. The bi-directional local alignment strategy [28] was proposed to encode image features using shared embedding networks, construct bi-directional distances to align similar semantic information, and optimize the network for FSFGIC tasks. Traditional feature generation networks failed to capture the subtle difference between fine-grained categories; to address this problem, a feature composition framework was proposed in [91] to generate fine-grained features for novel classes. In the training stage, they proposed a dense attribute-based attention to compute attention features for all attributes and then aligned them with attribute semantic vectors to obtain a similarity score. After that, they applied these attribute features to construct features of novel classes.

Multi-scale representation. Different from the single-scale representation, multi-scale enhances the representation of global features because the large-scale with larger receptive fields contains richer information [92–99]. In [23], a structural-pyramid descriptor was constructed by exploiting the pyramid pooling of the global feature with different scales. Then, multi-scale features were magnified to the same size and fused together by bilinear interpolations. Ruan et al. [29] proposed a spatial attentive comparison network (SACN) for the FSFGIC task. They constructed a selective-comparison similarity module (SCSM) based on pyramid structure and attention mechanism to assign different weights to the background and target, aiming to produce multi-scaled feature maps for classification. In [30], they were the first to attempt integrating the idea of multi-scale representation into the cross-domain few-shot classification problem by proposing a new hierarchical residual-like block applicable to lightweight ResNet structures such as ResNet-10. In [31], Zhang et al. proposed a multi-scale second-order relation network (MsSoSN), which equipped

second-order pooling and a scale selector to create multi-scale second-order representations. They proposed a scale and discrepancy discriminator to reweight multi-scale features, which were trained using a self-supervision method.

3.3. Class Representation Learning Based FSFGIC Methods

The authors of class representation learning based methods argue that local and/or global deep feature representations learned from extremely limited training samples cannot effectively represent a novel class, and class representations (e.g., class-level graph [15] or class-level local deep feature representation [7]) can be used to alleviate the phenomenon of overfitting and effectively represent a novel class.

3.3.1. Optimization-Based Class Representation Learning

The existing optimization-based class representation learning can be divided into two categories: (1) learning a model-based method, which aims to design network architectures to efficiently adapt to target tasks through only several gradient descent steps; (2) learning fine-tune-based methods.

Learning a model. In [32], an optimization-based FSFGIC method was proposed, which included a bilinear feature learning module and a classifier mapping module that encoded discriminative information and mapped features to decision boundaries using a "piecewise mappings" function. The meta variance transfer method [33] was proposed to transfer factors of variations between classes to improve classification performance on unseen examples, allowing deep learning models to generalize better with scarce data instances and enhance robustness against various factors of variations. In order to combine distribution-level and instance-level relation, Yang et al. [34] proposed a distribution propagation graph network (DPGN). The features of support images and query images were fed into a dual complete graph network, where a point-to-distribution aggregation strategy was applied to aggregate instance similarities to construct distribution representations. Additionally, a distribution-to-point aggregation strategy was applied to calculate similarity with both distribution-level and instance-level relations. Few-shot image classification methods faced challenges in capturing diverse context and intraclass variations with limited labeled images, leading to object and scale mismatch issues, which were addressed by the bilaterally normalized scale-consistent Sinkhorn distance (BSSD) method proposed by He et al. [100] for improved performance on few-shot benchmarks.

Learning to fine-tune. A weight imprinting strategy was proposed in [35], which aimed to set weights directly of a ConvNet classifier for new categories. They applied a normalization layer with a scaling factor in the classifier which aimed to transform the features of new category samples into activation vectors as the weights of the normalization layer. In [36], a transfer-based method was proposed to generate class representations. They applied a power transform mechanism to preprocess support features to make them closer to the Gaussian distribution. According to the Gaussian-like distribution, they applied maximum a posteriori probability to find the estimates of each class center, which is similar to the minimization of Wasserstein distance. Then an iterative algorithm based on a Wasserstein distance was used to estimate the optimal transport from the initial distribution of the features to the Gaussian distribution in order to update the center. In [37], they proposed an adaptive distribution calibration (ADC) method, which addressed distribution bias in few-shot learning by adaptively transferring and calibrating information from base classes to improve classification performance on novel classes.

3.3.2. Metric-Based Class Representation Learning

Many techniques have been put forward for effective metric-based class representations, which can be broadly divided into five categories: feature distribution, attention mechanism, metric strategy, semantic alignment, and multimodel learning.

Feature distribution. In [101], it was demonstrated that the GANs-based feature generator [102] suffered from the issue of mode collapse. To address this problem, varia-

tional autoencoder (VAE) [103] and GANs were combined together to form a conditional feature generation model [73], which aimed to learn the conditional distribution of image features on the labeled class data and the marginal distribution of image features on the unlabeled class data. Alternatively, a multi-mixed feature distribution could be learned to represent each category in RepMet [38] and perform FSFGIC tasks. Davis et al. [39] extended the DeepEMD method [22] by reconstructing each query sample as a weighted sum of components from the same class for obtaining class-level feature distribution. In [40], a re-abstraction and perturbing support pair network (RaPSPNet) was proposed to improve the performance of FSFGIC by enhancing feature discrimination through a feature re-abstraction embedding (FRaE) module and a novel perturbing support pair (PSP)-based similarity measure module.

Afrasiyabi [69] proposed two distribution alignment strategies to align the novel categories to the related base categories, aiming to obtain better class representations. A centroid alignment strategy and an adversarial alignment strategy based on Wasserstein distance were designed to enforce intra-class compactness. Das et al. proposed a non-parametric approach [104] to address the problem that only base-class prototypes were available. They considered that all class prototype distributions were arranged on a manifold. They first estimated the novel-class prototypes by calculating the mean of the prototypes, which were near the novel samples. A graph was structured with all the class prototypes, and an induced absorbing Markov chain was applied to complete the classification task. Ref. [105] proposed compositional prototypical networks (CPN) to learn transferable component prototypes for improved feature reusability, which could be adaptively fused with visual prototypes using a learnable weight generator for recognizing novel classes based on human-annotated attributes.

In order to learn fine-grained structure in the feature space, Luo et al. [106] proposed a two-path network to adaptively learn the views. One path was label-guided classification, where the support features belonging to the same class were aggregated into a prototype and the similarities were calculated between the prototypes and query images. Another path was instance-level classification, which aimed to produce different views for an image, then map them into feature space to construct a better fine-grained semantic structure. Ref. [107] proposed to combine the frequency features with routine features. In addition to a regular CNN module, a discrete cosine transformation was applied to generate frequency feature representations. Then, the two kinds of features were concatenated as the final features. Current approaches overlooked intra-class distribution details while focusing on learning a generalized class-level metric. Ref. [108] proposed improved prototypical networks (IPN) to address the issue by using an attention-analogous strategy with varied sample weights based on representativeness and a distance-scaling strategy to enhance class-distribution exploration and discriminative information across classes. To gain Gaussian-like distributions, ref. [109] proposed a transfer-based method to process features belonging to the same class. They introduced transforms to adjust the distribution of features, and a Wasserstein distance-based iterative algorithm to calculate the prototype for each class. Similarly, ref. [110] proposed an optimal-transport algorithm to transform features into Gaussian-like distributions and estimate the best class centers.

Attention mechanism. The attention strategy aims to select discriminative feature or region from the extracted feature space for effective class-level feature representation. In [46], an attention mechanism [111] was applied to locate and reweight semantically relevant local region pairs between query and support samples, which aimed to strengthen discriminative objects and suppress the background. He et al. [41] indicated that object localization (using local discriminative regions) could provide great help for FSFGIC. Then a self-attention-based complementary module, which utilized channel attention and spatial attention was designed for performing weakly supervised object localization and finding their corresponding discriminative regions. Ref. [48] utilized channel attention and spatial attention to find discriminative regions from query and support samples for improving the classification performance of FSFGIC. A novel transformer-based neural network

architecture called CrossTransformers [42] was designed which applied a cross-attention mechanism to find coarse spatial correspondence between the query and support labeled samples in a class. In [50], an attention mechanism was proposed to mix two modalities (i.e., semantic and visual modalities) and ensure that the representations of attributes were in the same space with visual representation. Single prototype-based methods might fail to capture the subtle information of a class. To address this problem, Huang et al. [43] proposed a descriptor-based multi-prototype network (LMPNet) to learn multi-prototype. They designed an attention mechanism to weight all channels in each spatial position of all samples adaptively to obtain local descriptors, and constructed multiple prototypes based on these descriptors which contained more complete information of a class.

Metric strategy. To obtain discriminative class representations for FSFGIC, image-to-class metric strategies were proposed. Deep nearest neighbor neural network (DN4) [7] aimed to learn optimal class-level local deep feature representation of a class space based on the designed image-to-class similarity measure strategy in the case of extremely limited training samples. A discriminative deep nearest neighbor neural network (D2N4) [112] extended the DN4 method [7] by adding a center loss function [113]. And then class-level local and global feature representations were learned for improving the quality discriminability features in the framework of the DN4 method [7]. The Bi-Similarity Network (BSNet) [44] was proposed to use two different similarity measures to create more discriminative feature maps from a small number of images, resulting in a significant boost in generalization performance. In [45], Zhu et al. argued that a large amount of unlabeled data had the high potential to improve the classification performance in FSFGIC tasks. A progressive point to set metric learning (PPSML) [45] was presented to improve few-shot classification accuracy by defining a distance metric and using a self-training strategy. To avoid overfitting and calculate a robust class representation under the condition of extremely limited training samples, a deep subspace network (DSN) [114] was introduced to transform class representation into an adaptive subspace and generate a corresponding classifier.

Triantafillou et al. proposed a mean average precision (mAP) [115], which aimed to learn a similarity metric based on information retrieval. They extended the work that optimized for AP in order to account for all possible choices of query among the batch points. They then used the frameworks of SSVM (Structural Support Vector Machine) and DLM (Direct Loss Minimization) for optimization of mAP. Liu et al. [116] introduced a negative margin loss to reduce inter-class variance and generate more efficient decision boundaries. Hilliard et al. [70] proposed a metric-agnostic conditional embeddings (MACO) network. MACO contained four stages: the feature stage was used to obtain features, the relational stage produced a single vector as the class representation of each class. The conditioning stage connected the class representations to query image features which aimed to learn the class representation that was more relevant to the query image and the classifier made the final prediction.

Semantic alignment. It was indicated in [47] that people tended to compare similar objects thoroughly in a pairwise manner, e.g., comparing the heads of two birds first, then their wings and feet. In this manner, it was natural to enhance feature information during the comparison process. A low-rank pairwise bilinear pooling operation network [47] was designed for obtaining class-level deep feature representation between query and support samples in terms of the way that people compared similar objects. According to [46], the main object could be situated anywhere in the image, leading to potential ambiguity when directly computing the distance between query and support samples. To address this problem, semantic alignment metric learning (SAML) [46] was proposed to align the semantically related local regions on samples by a "collect and select" strategy. On the one hand, the similarities of all local region pairs from query samples and support class in a relation matrix were calculated and obtained. On the other hand, an attention mechanism [111] was applied to "select" the semantically relevant pairs. Li et al. [48] extended the method in [46], and a convolutional block attention module [117] was applied to capture discriminative regions. To eliminate the influence of noise and improve the

efficiency of a similarity measure, query-relevant regions from support samples were selected for semantic alignment. Then, multi-scale class-level feature representations were utilized to represent discriminative regions of the query, support samples in a class, and perform FSFGIC tasks. In [69], a centroid associative alignment strategy was proposed to enforce intra-class compactness and obtain better class representations.

Alternatively, an end-to-end graph-based approach called explicit class knowledge propagation network (ECKPN) [15] was proposed, which aimed to learn and propagate the class representations explicitly. First, a comparison module was used to explore the relationship between paired samples for learning sample representations in instance-level graphs. Secondly, a squeeze strategy was proposed to make the instance-level graph generate the class-level graph, which helped obtain class-level visual representation. Third, the class-level visual representations were combined with the instance-level sample representations for performing FSFGIC tasks.

Multimodel learning. Inspired by the prototypical network [85], a multimodal prototypical network [49] was designed for mapping text data into the visual feature space by using GANs. In [50], Huang et al. indicated that some methods, which applied auxiliary semantic modalities into a metric learning framework, only augmented the feature representations of samples with available semantics and ignored the query samples, which might lose the potential for the improvement of classification performance and could lead to a shift between the modalities combination and the pure-visual representation. To address this issue, an attributes-guided attention module (AGAM) was proposed, which aimed to make more effective use of human-annotated attributes and learn more discriminative class-level feature representations. An attention alignment mechanism was designed to distill knowledge from attribute guidance to the pure visual feature selection process, so that it could learn to pay attention to more semantic features without using the restriction of attribute annotation. To better align the visual and language feature distributions that described the same object class, a cross-modal distribution alignment module [51] was proposed, in which a vision-language prototype was introduced for each class to align the distributions, and the earth mover's distance (EMD) was adopted to optimize the prototypes.

Gu et al. [118] proposed a two-stream neural network (TSNN), which not only learned features from RGB images, but also focused on steganalysis features via a steganalysis rich model filter layer. The RGB stream aimed to distinguish the difference between support images and query images based on the global-level features and calculated the representations of each support class; the steganalysis stream extracted steganalysis features to locate critical regions. An extractor and fusion module was used to fuse the two-stream features by a general convolutional block. An image-to-class deep metric was applied to produce the similarity scores. Zhang et al. [119] introduced fine-grained attributes into the prototype network and proposed a prototype completion network (ProtoComNet). In the meta-training stage, ProtoComNet extracted representative attribute features as priors. They applied an attention-based aggregator to aggregate the attribute features and prototype to obtain the completed prototype. In addition, a Gaussian-based prototype fusion strategy was designed to learn mean-based prototypes from unlabeled samples, and applied Bayesian estimation to fuse the two kinds of prototypes, aiming to produce more representative prototypes.

3.4. Task-Specific Feature Representation Learning Based FSFGIC Methods

Task-specific feature representation learning based FSFGIC methods aim to overcome the problem of overfitting and poor generalization and utilize deep feature representation from all training samples in a task (i.e., one training episode) to construct a task-specific feature representation (e.g., task-level graph relationship representation [16] or task-level local deep feature representation [8]) for this task.

3.4.1. Optimization-Based Task-Specific Feature Representation Learning

The existing optimization-based task-specific feature representation learning methods can be divided into two categories: learning a model and learning to fine-tune.

Learning a model. In [52], a task embedding network was presented to learn task-specific feature representations via a Fisher information matrix [120] for exploring the nature of the target task and its relationship to other tasks. Meanwhile, the learned task-specific feature representations could also show the similarity between two different tasks. It was indicated in [53] that the existing optimization-based methods learned to equally utilize meta-knowledge in each task without considering the diversity of each task. To address this problem, they extended the model-agnostic meta-learning method [121] to deal with the imbalance of the number of samples in each task instance and out-of-distribution tasks, but the encoding of complex datasets and calculation of balance variables for each task increased the computational complexity of the algorithm.

A meta neural architecture search method called M-NAS [122] was proposed to effectively obtain a task-specific architecture for each new task. Specifically, an autoencoder was designed to generate a task-aware model architecture which had the ability to tailor the globally shared meta-parameters. It was indicated in [123] that meta-learning models were prone to overfitting in a new task with limited samples. In this way, a gradient dropout regularization was proposed to efficiently adapt to a new task. The key idea was to impose uncertainty on the meta-training stage by adding a noise gradient to parameters to improve the generalization of the model. In [54], new transformers called HCTransformers were introduced, which enhanced data efficiency for visual recognition by leveraging spectral token pooling and attribute surrogate learning. They addressed the limitations of vision transformers with limited data, providing better performance through improved parameter optimization and image structure utilization.

In order to improve the representation ability of meta-learning methods, a deep meta-learning (DEML) method [124] was proposed to generate high-level concepts for each image in a task. These concepts could guide the meta-learner to adapt quickly to new tasks. Moreover, a concept discriminator was designed to recognize different images. Tian et al. [125] proposed a new consistent meta-regularization (Con-MetaReg) to enhance the learning ability of meta-learning models. Specifically, a base learner trained on the support set, then another learner trained on a novel query set. Con-MetaReg was proposed to align the two learners by the Frobenius norm of the difference between parameters to eliminate the data discrepancy for better meta-knowledge. In [126], a label-free loss function called Self-Critique and Adapt (SCA) was proposed. SCA could be added to a base model to learn knowledge with an unsupervised loss from a critic loss network. The features learned from the base model were sent to the critic network to create a loss for the target task.

Learning to fine-tune. In order to overcome overfitting and the poor generalization ability caused by limited training samples, an effective scheme [1] for selecting samples from the auxiliary data was proposed. According to a given classifier with shared parameters, some samples with similar feature distributions to some given target samples were selected from an auxiliary dataset with rich samples. The selected samples from an auxiliary dataset and the given target samples were sent into the classifiers to pre-train a weight initialization. Finally, the remaining target samples were used to fine-tune the parameters corresponding to the classifiers for quickly adapting to target tasks.

In order to improve the generalization on the novel domain, ref. [55] proposed a combining domain-specific meta-learners (CosML) method. CosML pre-trained a set of meta-learners on different domains to learn domain-specific parameters. CosML generated task and domain prototypes to represent each task and domain in the feature space. For the novel domain, they initialized a subnetwork with the domain-specific meta-parameters, which were weighted by the similarity of these domains and the novel domain. In the optimizing phase, properties in an image that were not related to the target task interfered with the optimization results. A context-agnostic (CA) [56] method was proposed to abandon the additional properties in training data. In the training task, they applied a context-adversarial network to generate another object without extra information to the base network to initialize context-agnostic weights.

3.4.2. Metric-Based Task-Specific Feature Representation Learning

The existing metric-based task-specific feature representation learning methods can be classified into three categories: feature distribution, attention mechanism, and metric strategy.

Feature distribution. In [57], a covariance metric network (CovaMNet) was proposed, which aimed to obtain task-level covariance representations and a covariance metric between query and support samples. Furthermore, a novel deep covariance metric was designed to measure the consistency of distributions between query and support samples for performing FSFGIC tasks. The metric function might have failed to generalize due to the discrepancy between the feature distributions of the base and novel domains in a task. To address this problem, Tseng et al. [58] proposed a cross-domain approach which applied a feature-wise transformation layer to simulate the feature distributions of different domains. In the training stage, the feature-wise transformation layer was inserted into the feature encoder and optimized by two hyper-parameters via a learning-to-learn strategy. Ref. [59] proposed an unsupervised embedding adaptation mechanism called early-stage feature reconstruction (ESFR). ESFR contained a feature-level reconstruction training stage and a dimensionality-driven early stopping stage, which aimed to find out more generalizable features.

Attention mechanism. In [8], an adaptive episodic attention module was designed to select and weight key regions among the entire task. Alternatively, attention strategy was also used in graph neural networks (GNNs) for effectively obtaining task-level relation representations. Guo et al. indicated in [16] that existing GNN-based FSFGIC methods focused on the sample-to-sample relations while neglecting task-level relationships. Then, a GNN based sample-to-task FSFGIC method named attention-based task-level relation module (ATRM) was proposed to consider the specificity of different tasks. In ATRM, task-relation representations between the embedding features of a target sample and the embedding features of all samples in the task were obtained by calculating the absolute difference between the target sample and all samples in the task. Then, an attention mechanism was used to learn task-specific relation representations for each task.

Metric strategy. It was indicated in [8] that the existing image-to-image similarity measure [19] or image-to-class similarity measure [7] could not make full use of local deep feature representations. To address this problem, an adaptive task-aware local representations network (ATL-Net) was designed to select local descriptors with learned thresholds and assign selected local representations different weights based on episodic attention for improving the local deep feature representations. In [60], a region comparison network was proposed which aimed to reveal how FSFGIC worked in neural networks. In order to explore more fine-grained information and find the critical regions, each support sample was divided into several parts, and task-level local deep feature representations between each region in a support sample and each query sample were used to calculate their feature similarities and their corresponding region weights. Then, an explainable network was designed to find the critical regions related to the final classification results. A discriminative mutual nearest neighbor neural network (DMN4) [61] extended the DN4 method [7] and a mutual nearest neighbor mechanism [127] was applied to obtain task-level local-feature representations between query and support samples for performing FSFGIC tasks. Li et al. extended a triplet network [128] into a deep K-tuplet network [62] for learning a task-level local deep feature representation by utilizing the relationship among the input samples in a training episode.

3.5. Comparison of Experimental Results

In Table 2, we select experimental data of some research results for the above three feature representation learning methods to be shown. It is worth noting that the data in Table 2 are derived from the corresponding original papers. Different backbone networks and different feature representation learning methods make the final model performance different. At present, some researchers [7,46] combine two or more feature representation learning methods to make the model obtain better classification results. In this paper, we

classify the above models according to the feature representation learning method, which occupies the largest proportion in the original method.

Table 2. FSFGIC results of CUB-200-2010, CUB-200-2011, Stanford Dogs, and Stanford Cars.

Methods			Published in	Backbone	Accuracy							
					CUB_2010		CUB_2011		Dogs		Cars	
					1 shot	5 shot	1 shot	5 shot	1 shot	5 shot	1 shot	5 shot
LG [1]	O [4]	MattML [6]	IJCAL 2020	Conv-64F	-	-	66.29	80.34	54.84	71.34	66.11	82.80
		P-Transfer [17]	AAAI 2021	ResNet-12	-	-	73.88	87.81	-	-	-	-
		GLFA [18]	PR 2023	ResNet-12	-	-	76.52	90.27	-	-	-	-
	M [5]	PABN [27]	ICME 2019	Bilinear CNN	-	-	66.71	76.81	55.47	66.65	56.80	68.78
		DeepEMD [22]	CVPR 2020	ResNet-12	-	-	75.65	88.69	-	-	-	-
		Adaptive Attention [19]	Arxiv 2020	Conv-64F	64.51	78.62	-	-	61.74	77.37	70.73	87.72
		MMN [25]	ICME 2020	ResNet-18	-	-	72.5	86.1	-	-	-	-
		SACN [29]	KBS 2021	Conv-32F	-	-	71.50	79.77	64.30	71.65	68.23	78.70
		S3Net [23]	ICME 2021	Conv-64F	64.27	78.02	72.30	84.23	63.56	77.54	71.19	84.40
		LCCRN [129]	TCSVT 2023	ResNet-12	-	-	82.97	93.63	-	-	87.04	96.19
		SAPENet [21]	PR 2023	Conv-64F	-	-	70.38	84.47	-	-	-	-
CR [2]	O	PCM [32]	TIP 2019	Bilinear CNN	-	-	42.10	62.48	28.78	46.92	29.63	52.28
		DPGN [34]	CVPR 2020	ResNet-12	-	-	75.71	91.48	-	-	-	-
		ADC [37]	Information Sciences 2022	ResNet-12	-	-	80.2	91.42	-	-	-	-
	M	MACO [70]	Arxiv 2018	Conv-32F	-	-	60.76	74.96	-	-	-	-
		SAML [46]	ICCV 2019	Conv-64F	-	-	69.35	81.37	-	-	-	-
		DN4 [7]	CVPR 2019	Conv-64F	53.15	81.90	-	-	45.73	66.33	61.51	89.60
		LRPABN [47]	TMM 2020	Bilinear CNN	-	-	67.97	78.04	54.52	67.12	63.11	72.63
		TSNN [118]	ECAI 2020	Conv-64F	57.02	70.33	48.62	63.45	-	-	-	-
		Centroid [69]	ECCV 2020	ResNet-18	-	-	74.22	88.65	-	-	-	-
		BSNet [44]	TIP 2020	Conv-64F	-	-	62.84	85.39	43.42	71.90	40.89	86.88
		CTX [42]	NIPS 2020	ResNet-34	-	-	-	84.06	-	-	-	-
		D2N4 [112]	TGRS 2020	Conv-64F	56.85	77.78	-	-	47.74	70.76	59.46	86.76
		FRN [39]	Arxiv 2020	ResNet-12	-	-	83.55	92.92	-	-	-	-
		Neg-Cosine [116]	ECCV 2020	ResNet-18	-	-	72.66	89.40	-	-	-	-
		PPSML [45]	ICIP 2020	Conv-64F	63.43	78.76	-	-	52.16	72.00	71.71	90.02
		AGAM [50]	AAAI 2021	ResNet-12	-	-	79.58	87.17	-	-	-	-
		PN+VLCL [106]	ICME 2021	WRN	71.21	85.08	-	-	-	-	-	-
		ECKPN [15]	CVPR 2021	ResNet-12	-	-	77.43	92.21	-	-	-	-
		QPN [48]	Arxiv 2021	Conv-64F	-	-	66.04	82.85	53.69	70.98	63.91	89.27
		LMPNet [43]	PR 2021	ResNet-12	65.59	68.19	-	-	61.89	68.21	68.31	80.27
		ProtoComNet [119]	CVPR 2021	ResNet-12	-	-	93.20	94.90	-	-	-	-
		TOAN [130]	TCSVT 2021	ResNet-256	-	-	67.17	82.09	51.83	69.83	76.62	89.57
		EASE+SIAMESE [11]	CVPR 2022	WRN	-	-	91.68	94.12	-	-	-	-
		CPN [105]	Arxiv 2023	ResNet-12	-	-	87.29	92.54	-	-		
		RaPSPNet [40]	PR 2023	Conv-64F	67.54	83.73	73.53	91.21	55.77	73.58	71.39	92.60
TR [3]	O	DEML+Meta-SGD [124]	Arxiv 2018	ResNet-50	-	-	66.95	77.11	-	-	-	-
		CosML [55]	Arxiv 2020	Conv-64F	46.89	66.15	-	-	-	-	47.74	60.17
		ANIL+CM [125]	TNNLS 2021	ResNet-12	-	-	59.89	74.35	-	-	-	-
		CA-MAML++ [56]	ACCV 2020	ResNet-18	-	-	43.3	57.9	-	-	-	-
		M-NAS [122]	AAAI 2020	Conv-64F	-	-	58.76	72.22	-	-	-	-
		GNN [131]	Sensors 2022	GNN	-	-	61.1	78.6	49.8	65.3	-	-
	M	CovaMNet [57]	AAAI 2019	Conv-64F	52.42	63.76	-	-	49.10	63.04	56.65	71.33
		ATL-Net [8]	IJCAI 2020	Conv-64F	60.91	77.05	-	-	54.49	73.20	67.95	89.16
		DPGN+ATRM [16]	Arxiv 2021	ResNet-12	-	-	77.53	90.39	-	-	-	-
		DMN4 [61]	Arxiv 2021	Conv-64F	-	-	78.36	92.16	-	-	-	-
		TRSN-T [14]	TNNLS 2023	ResNet-12	-	-	93.58	95.09	-	-	-	-

[1] LG: Local and/or global deep feature representation learning. [2] CR: Class representation learning. [3] TR: Task-specific feature representation learning. [4] O: Optimization-based. [5] M: Metric-based.

4. Summary and Discussions

Our investigation indicates that the existing FSFGIC methods have made great process in FSFGIC tasks, but there are still some important challenges to FSFGIC that need to be dealt with in the future.

Trade-off between the problem of overfitting and the ability of image feature representation. Our investigation indicates that the existing FSFGIC algorithms are still at the stage of theoretical exploration and cannot be used in practical applications. Currently, data augmentation, regularization, and modeling of the feature extraction process can effectively alleviate the overfitting problem caused by extremely limited training samples and can also enhance the ability of feature representation, but there is

still a trade-off between overcoming the overfitting problem and enhancing the ability of image feature representation. On the one hand, image feature representation is used not only to represent train samples, but also to construct classifiers for performing FSFGIC tasks. In this manner, the quality of feature representation directly affects the classification performance on FSFGIC. On the other hand, due to the extremely limited number of training samples on FSFGIC, the existing FSFGIC methods utilize a relatively simple network as a backbone (e.g., Conv-64F [132]) for alleviating the overfitting problem. Our investigation indicates that the existing simple networks cannot effectively learn discriminative features from training samples compared with the existing large networks (e.g., ResNet50 [133]). Therefore, how to balance the problem of overfitting and the ability of image feature representation is one of the most important challenges on FSFGIC.

Generalization in FSFGIC. There exist two main challenges on generalization in FSFGIC methods. On the one hand, an ideal FSFGIC algorithm should have the ability to handle various learning tasks with different complexity and diversity of data. Our investigation indicates that, currently, the number of tasks and datasets available for FSFGIC training is very limited (much less than the number of instances available in few-shot learning). Most of the existing FSFGIC methods are over-designed for specific benchmark tasks and data sets which may weaken the applicability of the existing FSFGIC methods for dealing with more general FSFGIC tasks. On the other hand, our investigation indicates that most of the existing FSFGIC studies focus on common application scenarios with small-scale tasks and large-scale labeled auxiliary data. However, the actual FSFGIC tasks that need to be solved may be dynamic and the labeled auxiliary data is not available. Therefore, it is necessary to generalize the technique of feature representation learning to effectively perform cross-domain or multi-domain FSFGIC tasks.

Theoretical research. In essence, all FSFGIC solutions are designed by specific techniques to obtain feature representations that can be used to accurately represent samples and to perform FSFGIC tasks. Although the quality of feature representation directly affects the classification performance of FSFGIC, our investigation indicates that no one has considered how to establish a theoretical approach to measure whether the feature representation learned from training samples can correctly reflect the inherent characteristics of the training samples. Therefore, constructing a systematic theory for FSFGIC from the perspective of improving the accuracy of feature representations obtained from training samples can bring new inspiration to FSFGIC researchers.

Performance and efficiency. As shown in Figure 3, the FSFGIC methods still have some challenges in terms of performance and efficiency. Researchers still need to make breakthroughs in the following aspects: (1) how to obtain more discriminating local significant features; (2) how to achieve better classifier performance; (3) how to reduce the model complexity and time complexity, so as to avoid overfitting and strengthen the robustness of the model.

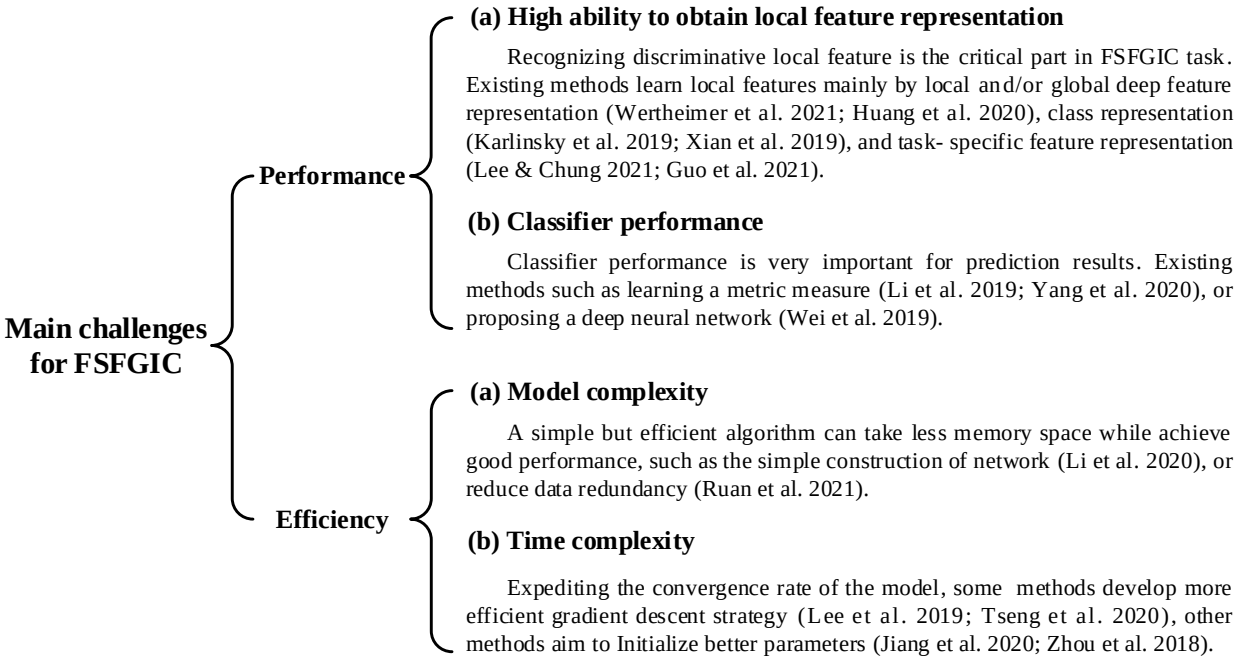

Figure 3. Main challenges for FSFGIC [7,16,19,29,32,34,38,39,44,47,53,59,73,112,123,124].

5. Conclusions

It is obvious that the fine-grained datasets are small in scale, and the samples between different subclasses often exist only in local subtle regions. Therefore, only the method that can extract the feature information of the local salient regions of the image without a large number of labeled samples for model training can achieve better classification performance of fine-grained datasets. The general few-shot algorithm is not designed for the fine-grained features of the image, so it cannot effectively extract the subtle differences in the image, resulting in poor performance [134]. Based on this, many FSFGIC methods have been proposed by researchers, and satisfactory results have been achieved. The excellent classification performance of these methods is mainly due to the following two reasons: (1) focusing on the feature information of the significant region of the image, it can obtain more distinctive and effective feature representation; (2) the inter-class distance between different subclasses is increased, and the intra-class distance within the same subclass is reduced.

In this paper, we presented a comprehensive review on feature representation learning for FSFGIC. A taxonomy for FSFGIC is proposed. In terms of this taxonomy, different issues of FSFGIC methods are discussed. The main unresolved problems related to feature representation learning for FSFGIC are summarized and discussed. We hope that this survey can help newcomers and practitioners position themselves in this growing field and work together to keep pushing the field forward.

Author Contributions: Conceptualization, methodology, and writing—review and editing, J.R., W.Z. and C.S.; investigation, writing—original draft, J.R., C.L. and Y.A.; supervision, W.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Shaanxi Natural Science Basic Research Project under Grant 2022JM-394 and the Scientific Research Program funded by Shaanxi Provincial Education Department under Grant 23JY029.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

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

References

1. Zhang, Y.; Tang, H.; Jia, K. Fine-grained visual categorization using meta-learning optimization with sample selection of auxiliary data. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 233–248.
2. Wah, C.; Branson, S.; Welinder, P.; Perona, P.; Belongie, S. *The Caltech-Ucsd Birds-200-2011 Dataset*; California Institute of Technology: Pasadena, CA, USA, 2011.
3. Nilsback, M.E.; Zisserman, A. Automated flower classification over a large number of classes. In Proceedings of the Indian Conference on Computer Vision, Graphics & Image Processing, Bhubaneswar, India, 16–19 December 2008; pp. 722–729.
4. Maji, S.; Rahtu, E.; Kannala, J.; Blaschko, M.; Vedaldi, A. Fine-grained visual classification of aircraft. *arXiv* **2013**, arXiv:1306.5151.
5. Smith, L.B.; Slone, L.K. A developmental approach to machine learning? *Front. Psychol.* **2017**, *8*, 2124. [CrossRef]
6. Zhu, Y.; Liu, C.; Jiang, S. Multi-attention Meta Learning for Few-shot Fine-grained Image Recognition. In Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence, Yokohama, Japan, 11–17 July 2020; pp. 1090–1096.
7. Li, W.; Wang, L.; Xu, J.; Huo, J.; Gao, Y.; Luo, J. Revisiting local descriptor based image-to-class measure for few-shot learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 7260–7268.
8. Dong, C.; Li, W.; Huo, J.; Gu, Z.; Gao, Y. Learning task-aware local representations for few-shot learning. In Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence, Yokohama, Japan, 11–17 July 2020; pp. 716–722.
9. Cao, S.; Wang, W.; Zhang, J.; Zheng, M.; Li, Q. A few-shot fine-grained image classification method leveraging global and local structures. *Int. J. Mach. Learn. Cybern.* **2022**, *13*, 2273–2281. [CrossRef]
10. Abdelaziz, M.; Zhang, Z. Learn to aggregate global and local representations for few-shot learning. *Multimed. Tools Appl.* **2023**, *82*, 32991–33014. [CrossRef]
11. Zhu, H.; Koniusz, P. EASE: Unsupervised discriminant subspace learning for transductive few-shot learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 9068–9078.
12. Li, Y.; Bian, C.; Chen, H. Generalized ridge regression-based channelwise feature map weighted reconstruction network for fine-grained few-shot ship classification. *IEEE Trans. Geosci. Remote. Sens.* **2023**, *61*, 1–10. [CrossRef]
13. Hu, Z.; Shen, L.; Lai, S.; Yuan, C. Task-adaptive Feature Disentanglement and Hallucination for Few-shot Classification. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *33*, 3638–3648. [CrossRef]
14. Zhou, Z.; Luo, L.; Zhou, S.; Li, W.; Yang, X.; Liu, X.; Zhu, E. Task-Related Saliency for Few-Shot Image Classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *early access*. [CrossRef] [PubMed]
15. Chen, C.; Yang, X.; Xu, C.; Huang, X.; Ma, Z. Eckpn: Explicit class knowledge propagation network for transductive few-shot learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 6596–6605.
16. Guo, Y.; Ma, Z.; Li, X.; Dong, Y. Atrm: Attention-based task-level relation module for gnn-based fewshot learning. *arXiv* **2021**, arXiv:2101.09840.
17. Shen, Z.; Liu, Z.; Qin, J.; Savvides, M.; Cheng, K.T. Partial is better than all: Revisiting fine-tuning strategy for few-shot learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtually, 2–9 February 2021; Volume 35, pp. 9594–9602.
18. Shi, B.; Li, W.; Huo, J.; Zhu, P.; Wang, L.; Gao, Y. Global-and local-aware feature augmentation with semantic orthogonality for few-shot image classification. *Pattern Recognit.* **2023**, *142*, 109702. [CrossRef]
19. Jiang, Z.; Kang, B.; Zhou, K.; Feng, J. Few-shot classification via adaptive attention. *arXiv* **2020**, arXiv:2008.02465.
20. Song, H.; Deng, B.; Pound, M.; Özcan, E.; Triguero, I. A fusion spatial attention approach for few-shot learning. *Inf. Fusion* **2022**, *81*, 187–202. [CrossRef]
21. Huang, X.; Choi, S.H. Sapenet: Self-attention based prototype enhancement network for few-shot learning. *Pattern Recognit.* **2023**, *135*, 109170. [CrossRef]
22. Zhang, C.; Cai, Y.; Lin, G.; Shen, C. Deepemd: Few-shot image classification with differentiable earth mover's distance and structured classifiers. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WS, USA, 13–19 June 2020; pp. 12200–12210.
23. Wu, H.; Zhao, Y.; Li, J. Selective, structural, subtle: Trilinear spatial-awareness for few-shot fine-grained visual recognition. In Proceedings of the IEEE International Conference on Multimedia and Expo, Shenzhen, China, 5–9 July 2021; pp. 1–6.
24. Liu, Y.; Zhu, L.; Wang, X.; Yamada, M.; Yang, Y. Bilaterally normalized scale-consistent sinkhorn distance for few-shot image classification. *IEEE Trans. Neural Netw. Learn. Syst.* 2023, *early access*. [CrossRef]
25. Zhao, J.; Lin, X.; Zhou, J.; Yang, J.; He, L.; Yang, Z. Knowledge-based fine-grained classification for few-shot learning. In Proceedings of the IEEE International Conference on Multimedia and Expo, London, UK, 6–10 July 2020; pp. 1–6.
26. Sun, X.; Xv, H.; Dong, J.; Zhou, H.; Chen, C.; Li, Q. Few-shot learning for domain-specific fine-grained image classification. *IEEE Trans. Ind. Electron.* 2020, *68*, 3588–3598. [CrossRef]

27. Huang, H.; Zhang, J.; Zhang, J.; Wu, Q.; Xu, J. Compare more nuanced: Pairwise alignment bilinear network for few-shot fine-grained learning. In Proceedings of the IEEE International Conference on Multimedia and Expo, Shanghai, China, 8–12 July 2019; pp. 91–96.
28. Zheng, Z.; Feng, X.; Yu, H.; Li, X.; Gao, M. BDLA: Bi-directional local alignment for few-shot learning. *Appl. Intell.* **2023**, *53*, 769–785. [CrossRef]
29. Ruan, X.; Lin, G.; Long, C.; Lu, S. Few-shot fine-grained classification with spatial attentive comparison. *Knowl.-Based Syst.* **2021**, *218*, 106840. [CrossRef]
30. Chen, Y.; Zheng, Y.; Xu, Z.; Tang, T.; Tang, Z.; Chen, J.; Liu, Y. Cross-domain few-shot classification based on lightweight Res2Net and flexible GNN. *Knowl.-Based Syst.* **2022**, *247*, 108623. [CrossRef]
31. Zhang, H.; Torr, P.; Koniusz, P. Few-shot learning with multi-scale self-supervision. *arXiv* **2020**, arXiv:2001.01600.
32. Wei, X.S.; Wang, P.; Liu, L.; Shen, C.; Wu, J. Piecewise classifier mappings: Learning fine-grained learners for novel categories with few examples. *IEEE Trans. Image Process.* **2019**, *28*, 6116–6125. [CrossRef] [PubMed]
33. Park, S.J.; Han, S.; Baek, J.W.; Kim, I.; Song, J.; Lee, H.B.; Han, J.J.; Hwang, S.J. Meta variance transfer: Learning to augment from the others. In Proceedings of the International Conference on Machine Learning, Virtually, 12–18 July 2020; pp. 7510–7520.
34. 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 Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 13390–13399.
35. Qi, H.; Brown, M.; Lowe, D.G. Low-shot learning with imprinted weights. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5822–5830.
36. Hu, Y.; Gripon, V.; Pateux, S. Leveraging the feature distribution in transfer-based few-shot learning. In Proceedings of the International Conference on Artificial Neural Networks, Bratislava, Slovakia, 14–17 September 2021; pp. 487–499.
37. Liu, X.; Zhou, K.; Yang, P.; Jing, L.; Yu, J. Adaptive distribution calibration for few-shot learning via optimal transport. *Inf. Sci.* **2022**, *611*, 1–17. [CrossRef]
38. Karlinsky, L.; Shtok, J.; Harary, S.; Schwartz, E.; Aides, A.; Feris, R.; Giryes, R.; Bronstein, A.M. Repmet: Representative-based metric learning for classification and few-shot object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 5197–5206.
39. Wertheimer, D.; Tang, L.; Hariharan, B. Few-shot classification with feature map reconstruction networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 8012–8021.
40. Zhang, W.; Zhao, Y.; Gao, Y.; Sun, C. Re-abstraction and perturbing support pair network for few-shot fine-grained image classification. *Pattern Recognit.* **2023**, *148*, 110158. [CrossRef]
41. He, X.; Lin, J.; Shen, J. Weakly-supervised Object Localization for Few-shot Learning and Fine-grained Few-shot Learning. *arXiv* **2020**, arXiv:2003.00874.
42. Doersch, C.; Gupta, A.; Zisserman, A. Crosstransformers: Spatially-aware few-shot transfer. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 21981–21993.
43. Huang, H.; Wu, Z.; Li, W.; Huo, J.; Gao, Y. Local descriptor-based multi-prototype network for few-shot learning. *Pattern Recognit.* **2021**, *116*, 107935. [CrossRef]
44. Li, X.; Wu, J.; Sun, Z.; Ma, Z.; Cao, J.; Xue, J.H. BSNet: Bi-similarity network for few-shot fine-grained image classification. *IEEE Trans. Image Process.* **2020**, *30*, 1318–1331. [CrossRef] [PubMed]
45. Zhu, P.; Gu, M.; Li, W.; Zhang, C.; Hu, Q. Progressive point to set metric learning for semi-supervised few-shot classification. In Proceedings of the IEEE International Conference on Image Processing, Abu Dhabi, United Arab Emirates, 25–28 October 2020; pp. 196–200.
46. Hao, F.; He, F.; Cheng, J.; Wang, L.; Cao, J.; Tao, D. Collect and select: Semantic alignment metric learning for few-shot learning. In Proceedings of the IEEE international Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 8460–8469.
47. Huang, H.; Zhang, J.; Zhang, J.; Xu, J.; Wu, Q. Low-rank pairwise alignment bilinear network for few-shot fine-grained image classification. *IEEE Trans. Multimed.* **2020**, *23*, 1666–1680. [CrossRef]
48. Li, Y.; Li, H.; Chen, H.; Chen, C. Hierarchical representation based query-specific prototypical network for few-shot image classification. *arXiv* **2021**, arXiv:2103.11384.
49. Pahde, F.; Puscas, M.; Klein, T.; Nabi, M. Multimodal prototypical networks for few-shot learning. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Virtually, 5–9 January 2021; pp. 2644–2653.
50. Huang, S.; Zhang, M.; Kang, Y.; Wang, D. Attributes-guided and pure-visual attention alignment for few-shot recognition. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtually, 2–9 February 2021; Volume 35, pp. 7840–7847.
51. Wang, R.; Zheng, H.; Duan, X.; Liu, J.; Lu, Y.; Wang, T.; Xu, S.; Zhang, B. Few-Shot Learning with Visual Distribution Calibration and Cross-Modal Distribution Alignment. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 23445–23454.
52. Achille, A.; Lam, M.; Tewari, R.; Ravichandran, A.; Maji, S.; Fowlkes, C.C.; Soatto, S.; Perona, P. Task2vec: Task embedding for meta-learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 6430–6439.

53. Lee, H.B.; Lee, H.; Na, D.; Kim, S.; Park, M.; Yang, E.; Hwang, S.J. Learning to balance: Bayesian meta-learning for imbalanced and out-of-distribution tasks. *arXiv* **2019**, arXiv:1905.12917.
54. He, Y.; Liang, W.; Zhao, D.; Zhou, H.Y.; Ge, W.; Yu, Y.; Zhang, W. Attribute surrogates learning and spectral tokens pooling in transformers for few-shot learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 9119–9129.
55. Peng, S.; Song, W.; Ester, M. Combining domain-specific meta-learners in the parameter space for cross-domain few-shot classification. *arXiv* **2020**, arXiv:2011.00179.
56. Perrett, T.; Masullo, A.; Burghardt, T.; Mirmehdi, M.; Damen, D. Meta-learning with context-agnostic initialisations. In Proceedings of the Asian Conference on Computer Vision, Virtually, 30 November–4 December 2020; pp. 70–86.
57. Li, W.; Xu, J.; Huo, J.; Wang, L.; Gao, Y.; Luo, J. Distribution consistency based covariance metric networks for few-shot learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 8642–8649.
58. Tseng, H.Y.; Lee, H.Y.; Huang, J.B.; Yang, M.H. Cross-domain few-shot classification via learned feature-wise transformation. *arXiv* **2020**, arXiv:2001.08735.
59. Lee, D.H.; Chung, S.Y. Unsupervised embedding adaptation via early-stage feature reconstruction for few-shot classification. In Proceedings of the International Conference on Machine Learning, Virtually, 18–24 July 2021; pp. 6098–6108.
60. Xue, Z.; Duan, L.; Li, W.; Chen, L.; Luo, J. Region comparison network for interpretable few-shot image classification. *arXiv* **2020**, arXiv:2009.03558.
61. Liu, Y.; Zheng, T.; Song, J.; Cai, D.; He, X. Dmn4: Few-shot learning via discriminative mutual nearest neighbor neural network. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtually, 22 February–1 March 2022; Volume 36, pp. 1828–1836.
62. Li, X.; Yu, L.; Fu, C.W.; Fang, M.; Heng, P.A. Revisiting metric learning for few-shot image classification. *Neurocomputing* **2020**, *406*, 49–58. [CrossRef]
63. Welinder, P.; Branson, S.; Mita, T.; Wah, C.; Schroff, F.; Belongie, S.; Perona, P. *Caltech-UCSD Birds 200*; California Institute of Technology: Pasadena, CA, USA, 2010.
64. Khosla, A.; Jayadevaprakash, N.; Yao, B.; Li, F.F. Novel dataset for fine-grained image categorization: Stanford dogs. In Proceedings of the CVPR Workshop on Fine-Grained Visual Categorization, Colorado Springs, CO, USA, 20–25 June 2011; Citeseer: State College, PA, USA, 2011; Volume 2.
65. Krause, J.; Stark, M.; Deng, J.; Fei-Fei, L. 3D object representations for fine-grained categorization. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Sydney, NSW, Australia, 1–8 December 2013; pp. 554–561.
66. Van Horn, G.; Branson, S.; Farrell, R.; Haber, S.; Barry, J.; Ipeirotis, P.; Perona, P.; Belongie, S. Building a bird recognition APP and large scale dataset with citizen scientists: The fine print in fine-grained dataset collection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 595–604.
67. Xiao, J.; Hays, J.; Ehinger, K.A.; Oliva, A.; Torralba, A. Sun database: Large-scale scene recognition from abbey to zoo. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 3485–3492.
68. Yu, X.; Zhao, Y.; Gao, Y.; Xiong, S.; Yuan, X. Patchy image structure classification using multi-orientation region transform. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 12741–12748.
69. Afrasiyabi, A.; Lalonde, J.F.; Gagné, C. Associative alignment for few-shot image classification. In Proceedings of the European Conference on Computer Vision, Virtually, 23–28 August 2020; pp. 18–35.
70. Hilliard, N.; Phillips, L.; Howland, S.; Yankov, A.; Corley, C.D.; Hodas, N.O. Few-shot learning with metric-agnostic conditional embeddings. *arXiv* **2018**, arXiv:1802.04376.
71. Zhang, M.; Wang, D.; Gai, S. Knowledge distillation for model-agnostic meta-learning. In Proceedings of the 24th European Conference on Artificial Intelligence, Virtually, 29 August–8 September 2020; pp. 1355–1362.
72. Pahde, F.; Nabi, M.; Klein, T.; Jahnichen, P. Discriminative hallucination for multi-modal few-shot learning. In Proceedings of the IEEE International Conference on Image Processing, Athens, Greece, 7–10 October 2018; pp. 156–160.
73. Xian, Y.; Sharma, S.; Schiele, B.; Akata, Z. f-vaegan-d2: A feature generating framework for any-shot learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 10275–10284.
74. Xu, J.; Le, H.; Huang, M.; Athar, S.; Samaras, D. Variational feature disentangling for fine-grained few-shot classification. In Proceedings of the IEEE International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 8812–8821.
75. Luo, Q.; Wang, L.; Lv, J.; Xiang, S.; Pan, C. Few-shot learning via feature hallucination with variational inference. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Virtually, 5–9 January 2021; pp. 3963–3972.
76. Tsutsui, S.; Fu, Y.; Crandall, D. Meta-reinforced synthetic data for one-shot fine-grained visual recognition. *arXiv* **2019**, arXiv:1911.07164.
77. Pahde, F.; Jähnichen, P.; Klein, T.; Nabi, M. Cross-modal hallucination for few-shot fine-grained recognition. *arXiv* **2018**, arXiv:1806.05147.
78. Wang, Y.; Xu, C.; Liu, C.; Zhang, L.; Fu, Y. Instance credibility inference for few-shot learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 12836–12845.

79. Chen, M.; Fang, Y.; Wang, X.; Luo, H.; Geng, Y.; Zhang, X.; Huang, C.; Liu, W.; Wang, B. Diversity transfer network for few-shot learning. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 10559–10566.
80. Schwartz, E.; Karlinsky, L.; Shtok, J.; Harary, S.; Marder, M.; Kumar, A.; Feris, R.; Giryes, R.; Bronstein, A. Delta-encoder: An effective sample synthesis method for few-shot object recognition. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 2850–2860.
81. Wang, C.; Song, S.; Yang, Q.; Li, X.; Huang, G. Fine-grained few shot learning with foreground object transformation. *Neurocomputing* **2021**, *466*, 16–26. [CrossRef]
82. Lupyan, G.; Ward, E.J. Language can boost otherwise unseen objects into visual awareness. *Natl. Acad. Sci.* **2013**, *110*, 14196–14201. [CrossRef] [PubMed]
83. Tokmakov, P.; Wang, Y.X.; Hebert, M. Learning compositional representations for few-shot recognition. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6372–6381.
84. Chen, W.Y.; Liu, Y.C.; Kira, Z.; Wang, Y.C.F.; Huang, J.B. A closer look at few-shot classification. *arXiv* **2019**, arXiv:1904.04232.
85. Snell, J.; Swersky, K.; Zemel, R. Prototypical networks for few-shot learning. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4080–4090.
86. Flores, C.F.; Gonzalez-Garcia, A.; van de Weijer, J.; Raducanu, B. Saliency for fine-grained object recognition in domains with scarce training data. *Pattern Recognit.* **2019**, *94*, 62–73. [CrossRef]
87. Tavakoli, H.R.; Borji, A.; Laaksonen, J.; Rahtu, E. Exploiting inter-image similarity and ensemble of extreme learners for fixation prediction using deep features. *Neurocomputing* **2017**, *244*, 10–18. [CrossRef]
88. Zhang, X.; Wei, Y.; Feng, J.; Yang, Y.; Huang, T.S. Adversarial complementary learning for weakly supervised object localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1325–1334.
89. Liao, Y.; Zhang, W.; Gao, Y.; Sun, C.; Yu, X. ASRSNet: Automatic Salient Region Selection Network for Few-Shot Fine-Grained Image Classification. In Proceedings of the International Conference on Pattern Recognition and Artificial Intelligence, Paris, France, 1–3 June 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 627–638.
90. Chen, Q.; Yang, R. Learning to distinguish: A general method to improve compare-based one-shot learning frameworks for similar classes. In Proceedings of the IEEE International Conference on Multimedia and Expo, Shanghai, China, 8–12 July 2019; pp. 952–957.
91. Huynh, D.; Elhamifar, E. Compositional fine-grained low-shot learning. *arXiv* **2021**, arXiv:2105.10438.
92. Zhang, W.; Sun, C. Corner detection using second-order generalized Gaussian directional derivative representations. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 1213–1224. [CrossRef] [PubMed]
93. Zhang, W.; Sun, C.; Gao, Y. Image intensity variation information for interest point detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 9883–9894. [CrossRef] [PubMed]
94. Jing, J.; Liu, S.; Wang, G.; Zhang, W.; Sun, C. Recent advances on image edge detection: A comprehensive review. *Neurocomputing* **2022**, *503*, 259–271. [CrossRef]
95. Zhang, W.; Zhao, Y.; Breckon, T.P.; Chen, L. Noise robust image edge detection based upon the automatic anisotropic Gaussian kernels. *Pattern Recognit.* **2017**, *63*, 193–205. [CrossRef]
96. Jing, J.; Gao, T.; Zhang, W.; Gao, Y.; Sun, C. Image feature information extraction for interest point detection: A comprehensive review. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 4694–4712. [CrossRef]
97. Zhang, W.; Sun, C.; Breckon, T.; Alshammari, N. Discrete curvature representations for noise robust image corner detection. *IEEE Trans. Image Process.* **2019**, *28*, 4444–4459. [CrossRef]
98. Zhang, W.; Sun, C. Corner detection using multi-directional structure tensor with multiple scales. *Int. J. Comput. Vis.* **2020**, *128*, 438–459. [CrossRef]
99. Shui, P.L.; Zhang, W.C. Corner detection and classification using anisotropic directional derivative representations. *IEEE Trans. Image Process.* **2013**, *22*, 3204–3218. [CrossRef] [PubMed]
100. He, J.; Kortylewski, A.; Yuille, A. CORL: Compositional representation learning for few-shot classification. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 2–7 January 2023; pp. 3890–3899.
101. Arjovsky, M.; Bottou, L. Towards principled methods for training generative adversarial networks. *arXiv* **2017**, arXiv:1701.04862.
102. Xian, Y.; Lorenz, T.; Schiele, B.; Akata, Z. Feature generating networks for zero-shot learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5542–5551.
103. Verma, V.K.; Arora, G.; Mishra, A.; Rai, P. Generalized zero-shot learning via synthesized examples. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4281–4289.
104. Das, D.; Moon, J.; George Lee, C. Few-shot image recognition with manifolds. In Proceedings of the Advances in Visual Computing: International Symposium, San Diego, CA, USA, 5–7 October 2020; pp. 3–14.
105. Lyu, Q.; Wang, W. Compositional Prototypical Networks for Few-Shot Classification. *arXiv* **2023**, arXiv:2306.06584.
106. Luo, X.; Chen, Y.; Wen, L.; Pan, L.; Xu, Z. Boosting few-shot classification with view-learnable contrastive learning. In Proceedings of the IEEE International Conference on Multimedia and Expo, Shenzhen, China, 5–9 July 2021; pp. 1–6.
107. Chen, X.; Wang, G. Few-shot learning by integrating spatial and frequency representation. In Proceedings of the Conference on Robots and Vision, Burnaby, BC, Canada, 26–28 May 2021; pp. 49–56.
108. Ji, Z.; Chai, X.; Yu, Y.; Pang, Y.; Zhang, Z. Improved prototypical networks for few-shot learning. *Pattern Recognit. Lett.* **2020**, *140*, 81–87. [CrossRef]

109. Hu, Y.; Pateux, S.; Gripon, V. Squeezing backbone feature distributions to the max for efficient few-shot learning. *Algorithms* **2022**, *15*, 147. [CrossRef]
110. Chobola, T.; Vašata, D.; Kordík, P. Transfer learning based few-shot classification using optimal transport mapping from preprocessed latent space of backbone neural network. In Proceedings of the AAAI Workshop on Meta-Learning and MetaDL Challenge, Virtually, 9 February 2021; pp. 29–37.
111. Zagoruyko, S.; Komodakis, N. Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer. *arXiv* **2016**, arXiv:1612.03928.
112. Yang, X.; Nan, X.; Song, B. D2N4: A discriminative deep nearest neighbor neural network for few-shot space target recognition. *IEEE Trans. Geosci. Remote. Sens.* **2020**, *58*, 3667–3676. [CrossRef]
113. Wen, Y.; Zhang, K.; Li, Z.; Qiao, Y. A discriminative feature learning approach for deep face recognition. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 499–515.
114. Simon, C.; Koniusz, P.; Nock, R.; Harandi, M. Adaptive subspaces for few-shot learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 4136–4145.
115. Triantafillou, E.; Zemel, R.; Urtasun, R. Few-shot learning through an information retrieval lens. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 2252–2262.
116. Liu, B.; Cao, Y.; Lin, Y.; Li, Q.; Zhang, Z.; Long, M.; Hu, H. Negative margin matters: Understanding margin in few-shot classification. In Proceedings of the European Conference on Computer Vision, Virtually, 23–28 August 2020; pp. 438–455.
117. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 3–19.
118. Gu, Q.; Luo, Z.; Zhu, Y. A Two-Stream Network with Image-to-Class Deep Metric for Few-Shot Classification. In Proceedings of the ECAI 2020, Santiago de Compostela, Spain, 29 August–8 September 2020; pp. 2704–2711.
119. Zhang, B.; Li, X.; Ye, Y.; Huang, Z.; Zhang, L. Prototype completion with primitive knowledge for few-shot learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 3754–3762.
120. Jaakkola, T.; Haussler, D. Exploiting generative models in discriminative classifiers. *Adv. Neural Inf. Process. Syst.* **1998**, *11*, 487–493.
121. Finn, C.; Abbeel, P.; Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. In Proceedings of the International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 1126–1135.
122. Wang, J.; Wu, J.; Bai, H.; Cheng, J. M-nas: Meta neural architecture search. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 6186–6193.
123. Tseng, H.Y.; Chen, Y.W.; Tsai, Y.H.; Liu, S.; Lin, Y.Y.; Yang, M.H. Regularizing meta-learning via gradient dropout. In Proceedings of the Asian Conference on Computer Vision, Virtually, 30 November–4 December 2020.
124. Zhou, F.; Wu, B.; Li, Z. Deep meta-learning: Learning to learn in the concept space. *arXiv* **2018**, arXiv:1802.03596.
125. Tian, P.; Li, W.; Gao, Y. Consistent meta-regularization for better meta-knowledge in few-shot learning. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 7277–7288. [CrossRef] [PubMed]
126. Antoniou, A.; Storkey, A.J. Learning to learn by self-critique. *Adv. Neural Inf. Process. Syst.* **2019**, *32*.
127. Gowda, K.; Krishna, G. The condensed nearest neighbor rule using the concept of mutual nearest neighborhood. *IEEE Trans. Inf. Theory* **1979**, *25*, 488–490. [CrossRef]
128. Ye, M.; Guo, Y. Deep triplet ranking networks for one-shot recognition. *arXiv* **2018**, arXiv:1804.07275.
129. Li, X.; Song, Q.; Wu, J.; Zhu, R.; Ma, Z.; Xue, J.H. Locally-Enriched Cross-Reconstruction for Few-Shot Fine-Grained Image Classification. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *33*, 7530–7540. [CrossRef]
130. Huang, H.; Zhang, J.; Yu, L.; Zhang, J.; Wu, Q.; Xu, C. TOAN: Target-oriented alignment network for fine-grained image categorization with few labeled samples. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *32*, 853–866. [CrossRef]
131. Zhou, X.; Zhang, Y.; Wei, Q. Few-Shot Fine-Grained Image Classification via GNN. *Sensors* **2022**, *22*, 7640. [CrossRef]
132. Vinyals, O.; Blundell, C.; Lillicrap, T.; Wierstra, D. Matching networks for one shot learning. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 3637–3645.
133. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
134. Liu, Y.; Bai, Y.; Che, X.; He, J. Few-Shot Fine-Grained Image Classification: A Survey. In Proceedings of the 2022 4th International Conference on Natural Language Processing (ICNLP), Xi'an, China, 25–27 March 2022; pp. 201–211.