



Few-shot and meta-learning methods for image understanding: a survey

Kai He¹ · Nan Pu² · Mingrui Lao¹ · Michael S. Lew¹

Received: 21 April 2023 / Revised: 25 May 2023 / Accepted: 29 May 2023 / Published online: 29 June 2023
© The Author(s) 2023

Abstract

State-of-the-art deep learning systems (e.g., ImageNet image classification) typically require very large training sets to achieve high accuracies. Therefore, one of the grand challenges is called few-shot learning where only a few training samples are required for good performance. In this survey, we illuminate one of the key paradigms in few-shot learning called meta-learning. These meta-learning methods, by simulating the tasks which will be presented at inference through episodic training, can effectively employ previous prior knowledge to guide the learning of new tasks. In this paper, we provide a comprehensive overview and key insights into the meta-learning approaches and categorize them into three branches according to their technical characteristics, namely metric-based, model-based and optimization-based meta-learning. Due to the major importance of the evaluation process, we also present an overview of current widely used benchmarks, as well as performances of recent meta-learning methods on these datasets. Based on over 200 papers in this survey, we conclude with the major challenges and future directions of few-shot learning and meta-learning.

Keywords Few-shot learning · Meta-learning · Image classification · Deep neural networks

1 Introduction

Image classification [67, 142] is an important application in computer vision [4, 162] and machine learning [91, 193]. With the continuous development of deep learning [5, 79, 132], recent years have witnessed great breakthroughs in this area [48, 153]. However, such success relies on a huge amount of data [22, 136] (usually in the order of million), which is difficult and time-consuming in the real world. In order to reduce the data requirement, there has been growing interest in small-sample image classification [80, 140, 201],

such as few-shot classification [1, 18, 115], which learns a classification rule from few (1-5) labeled samples.

A core challenge in few-shot image classification is to alleviate the susceptibility of models to overfitting under few-data regime [27, 110, 168]. To address this problem, researchers have proposed several promising approaches, such as transfer learning [123, 203], meta-learning [38, 122, 145] and data augmentation [7, 16, 57]. In transfer learning, a model is first trained on a source domain where abundant source data is available. Then this trained model is fine-tuned [15, 137, 195] on another target domain with few labeled target samples. The learnt prior knowledge can be transferred from source tasks to target tasks during this process. Meta-learning, or learning to learn, has emerged as one of the prominent approaches for few-shot learning. It is proposed to train a meta-learner which can quickly generalize to new tasks with few examples [33, 45, 165, 178]. A meta-learning procedure also involves learning at two levels, within and across tasks. Meta-learning approaches simulate the tasks that will be presented at inference through episodic training [116, 170, 202], enabling the generalization ability of meta-learner within minor adaption steps. Data augmentation methods are often used as preprocessing in few-shot learning (FSL). In order to solve the problem of insufficient

✉ Kai He
k.he@liacs.leidenuniv.nl

Nan Pu
nan.pu@unitn.it

Mingrui Lao
m.lao@liacs.leidenuniv.nl

Michael S. Lew
m.s.lew@liacs.leidenuniv.nl

¹ LIACS Media Lab, Leiden University, Leiden, Netherlands

² The Department of Information Engineering and Computer Science, University of Trento, Trento, Italy

training data, they introduce various kinds of existing data variance for the model to capture. For image classification, one commonly used method is deformation [69, 119, 164, 185], including horizontal flipping, cropping and rotation. Besides these, more advanced methods, such as generating training samples and pseudolabels [28, 29, 192], are also an important part of data augmentation.

In this paper, we present a survey of recent meta-learning methods for few-shot image classification. Meta-learning focuses on learning prior knowledge from previous tasks which can bring efficient downstream learning to new tasks. This learning mechanism enables models can learn new concepts quickly where only few samples are available. Meta-learning deserves special attention as it is an essential part of few-shot image classification and it has also demonstrated outstanding performance on benchmark datasets [64, 144]. To be specific, in this survey we divide meta-learning into three categories according to the different mechanisms, namely metric-based, model-based and optimization-based methods [40, 58, 89, 166].

A number of surveys on FSL have been proposed. In 2018, Shu et al. [140] provided an early survey on small-sample learning, discussing approaches for different scenarios (zero-shot learning [124, 179, 180] and FSL) and tasks (image classification, visual question answering [6, 90, 139] and object detection [62, 114, 151]). Wang et al. [167] conducted a comprehensive review in 2021, which provides a formal definition of FSL and distinguishes it from other machine learning problems, exploring FSL from a fundamental viewpoint of error decomposition in supervised learning. Li et al. [74] published another comprehensive review on FSL in 2021, which is entirely focused on meta-learning and review literature [39, 43, 44, 156] over a long period in this area. There is another review on few-shot image classification [76] published in 2023, which is fully devoted to metric learning methods [103, 141, 188]. Compared with these surveys [74, 76, 140, 167], our review presents an up-to-date survey of meta-learning approaches for few-shot image classification and provides a thorough analysis of these different kinds of methods to better understand their individual strengths and limitations.

The remainder of this survey is organized as follows. In Sect. 2, we provide the preliminary concepts of meta-learning, including the definition of few-shot image classification, commonly used datasets and the evaluation procedure. In Sect. 3, we mainly introduce the category of meta-learning methods and review both classical and state-of-the-art meta-learning approaches. We also present other kinds of few-shot learning methods to do a comparison. In Sect. 4, we discuss the major challenges, along with future directions. Finally, we conclude this survey in Sect. 5.

2 The framework of few-shot image classification

2.1 Notation and definitions

In this section, we first present a brief introduction about few-shot learning and meta-learning, and then provide the notation and unified definitions of few-shot image classification [23, 56, 155].

Few-shot learning is a surprising research area that focuses on learning patterns from a set of data (base classes) and then adapting to a disjoint set (novel classes) with limited training samples. Few-shot image classification is the one with most attention and researches. As the most popular approach for few-shot learning, meta-learning organizes the learning process into two phases, called meta-training and meta-testing. During each phase, the meta-training set or meta-testing set is split into multiple episodes. Each episode samples from the task distribution and is further divided into a small training set and a testing set.

In the standard few-shot image classification setting, two distinct datasets are involved, namely base dataset $D_{base} = \{(x_i, y_i); x_i \in X_{base}, y_i \in Y_{base}\}_{i=1}^{N_{base}}$ and novel dataset $D_{novel} = \{(x_i, y_i); x_i \in X_{novel}, y_i \in Y_{novel}\}_{i=1}^{N_{novel}}$, where x_i represents the original feature vector of i -th image and y_i is the corresponding class label; N_{base} and N_{novel} denote the total numbers of instances in D_{base} and D_{novel} , respectively. The base dataset is an auxiliary dataset that is used to train the classifier to learn some prior or shared knowledge and the novel dataset is used for the classifier to perform new classification tasks. Note that D_{base} and D_{novel} are disjoint, which means $Y_{base} \cap Y_{novel} = \emptyset$. In order to train and test the classifier, the D_{novel} is usually split into the support set D_S and the query set D_Q and they share the same label space.

Definition 1 The few-shot image classification task aims to learn a classifier from D_{base} and D_S to correctly classify the samples in D_Q . It is generally termed as a N -way K -shot problem, where N and K denote the number of classes and instances in D_S , respectively. If $K = 1$, it becomes a one-shot image classification task; and if $K = 0$, then the task is called zero-shot classification.

Definition 2 A few-shot image classification task is called cross-domain few-shot image classification when the base dataset and the novel dataset are from two different domains, i.e., $X_{base} \neq X_{novel}$.

2.2 Datasets

In this section, we briefly introduce several well-known datasets for few-shot image classification. According to different data types, we categorize them into simple image

Fig. 1 Sample images of these benchmark datasets widely used for few-shot image classification



dataset (Omniglot [70]), complex image dataset (MiniImageNet [120, 161], TieredImageNet [122], CIFAR-FS [10] and FC100 [107]) and special image dataset (CUB-200 [163, 175]). Among these datasets, CIFAR-FS and FC100 are considered more difficult as the resolution of images from the two datasets is 32×32 . It is more challenging for models to extract useful information from low-resolution images. Statistics of these datasets and popular experimental settings are summarized below. We also present some sample images from these benchmark datasets in Fig. 1.

Omniglot is one of the most frequently used benchmarks for evaluating few-shot image classification algorithms. It contains 1623 handwritten characters collected from 50 different alphabets. Each character consists of 20 samples, drawn by different human subjects. This dataset is usually augmented by the rotations in multiples of 90 degrees, and 1200 characters are used for training and the rest for evaluation.

MiniImageNet and *TieredImageNet* are two mini versions of the large ImageNet dataset [129]. *MiniImageNet* is composed of 60,000 color images from 100 classes, with 600 images in each class. Following the widely used splitting protocol proposed by Revi and Larochelle [120], 64 classes are used for training, 16 classes for validation and 20 classes for evaluation. *TieredImageNet* is another larger subset of ImageNet with a hierarchical structure. It contains 779,165 images from 34 high-level categories (or 608 classes), which are further split into 20 base categories (351 classes), 6 validation categories (97 classes) and 8 novel categories (160 classes).

CIFAR-FS and *FC100* are two widely used datasets derived from CIFAR-100 [68]. *CIFAR-FS* is constructed

from 100 classes with 600 images per class. The 100 classes are split into 64, 16 and 20 classes for training, validation and evaluation, respectively. *FC100* also contains 100 classes, which are further divided into 20 super-categories, with five classes in each super-categories. *FC100* is split into 12 base, 4 validation and 4 novel super-categories.

CUB-200 is a fine-grained dataset consisting of 200 bird species. The *CUB-200* dataset has two versions, while the initial version was proposed in 2010 [175] which includes 6033 images and is extended to 11,788 images in 2011 [163]. The *CUB-200-2010* dataset is often split into 130 base, 20 validation and 50 novel classes [85], while the *CUB-200-2011* dataset is divided into 100 classes for training, 50 classes for validation and 50 classes for testing [18].

MiniImageNet \rightarrow **CUB** is a dataset designed for cross-domain few-shot image classification [93, 159, 169, 190]. *MiniImageNet* plays the role of the base dataset, while 50 classes of *CUB-200-2011* are used for validation, and the remaining 50 classes serve for evaluation.

2.3 Evaluation process of few-shot image classification

In this section, we present a general procedure [30, 55, 148, 177, 200] to evaluate a classifier's performance on N -way K -shot image classification problems in Algorithm 1. The whole evaluation process is composed of lots of episodes. In each episode, we first randomly select N classes from the novel label space with K samples in each class to form a support set D_S and M examples from the rest samples of those N classes to compose a query set D_Q . A final classifier can be

obtained based on the base dataset and support set, which is used to predict labels of samples in D_Q . We use $acc^{(e)}$ to denote the classification accuracy in the e -th episode, and the performance of a learning algorithm can be measured by the averaged classification accuracy over all episodes.

Algorithm 1 Evaluation process of N -way K -shot image classification

Input: D_{base} ; D_{novel} ; the total number of episodes E and a classifier f .
for $e = 1, \dots, E$ **do**
 Randomly select N classes from D_{novel} .
 Randomly select K samples from each class to compose the support set $D_S^{(e)}$.
 Randomly select M samples from the remaining samples of N classes to compose the query set $D_Q^{(e)} = \{(X^{(e)}, Y^{(e)})\}$.
 Use the classifier f to predict labels of the samples from D_Q : $\hat{Y}^{(e)} = f(X^{(e)} | D_{base}, D_S^{(e)})$.
 Record accuracy $acc^{(e)} = \frac{1}{M} \sum_{j=1}^M \mathbb{1}[\hat{Y}^{(e)} = Y^{(e)}]$.
end for
 return mean accuracy $\frac{1}{E} \sum_{e=1}^E acc^{(e)}$.

3 Paradigms of meta-learning for few-shot image classification

The goal of meta-learning for few-shot image classification [24, 61, 82, 94, 111] is to enable models, especially deep neural networks, to perform well on new tasks when only few samples are available. With the rapid development of few-shot learning [50, 183, 205], a number of meta-learning approaches [19, 102, 184] have been proposed. In this section, we provide a comprehensive overview of recent meta-learning studies and their advances. In order to let beginners better understand, we follow the main trend and still categorize meta-learning into metric-based, model-based and optimization-based methods. Besides, we also present other few-shot learning methods to make a com-

parison. Figure 2 shows an overview of few-shot image classification.

3.1 Metric-based meta-learning

Metric-based meta-learning methods [49, 72, 75, 194] aim to learn a distance metric, which can effectively measure the similarity among samples, ensuring it is optimal for new learning tasks. For few-shot image classification problems, the learned metric should follow the rules that enable samples from the same (or different) class should a small (or large) distance.

Siamese network is one of the most widely used metric-based methods for one-shot image classification. The term ‘‘Siamese’’ was first proposed for signature verification [13] and the principal structure of Siamese network was introduced for the fingerprint similarity estimation problem [9]. In 2015, Koch et al. [65] adopted a pair of identical VGG-styled [142] convolutional layers with shared weights to extract high-level features from two input images and calculate the weighted L_1 distance between the two feature vectors. The network finally outputs a score, representing the probability that the two images belong to the same class. The architecture of Siamese network is shown in Fig. 3. Wang et al. [173] proposed an attention-based Siamese network, which exploits an attention kernel function to measure the similarity between two feature vectors. To bridge the gap between one-shot image recognition [17, 32, 160] and regular classification, Lungu et al. [88] proposed a multi-resolution Siamese network, which mixes different kernel size streams into one layer and adopts a hybrid training mechanism.

As another powerful metric-based meta-learning method, matching network [161] uses different networks to encode support and query images. For support images embedding, a bidirectional long-short-term memory (LSTM) [198] is used in the context of the support set D_S ; for query images embedding, an LSTM with an attention kernel is taken to enable the dependency on D_S , where the attention kernel [12, 105, 106] is used to compute cosine similarities between support and query images and then normalize the similarities through

Fig. 2 An overview of few-shot image classification approaches

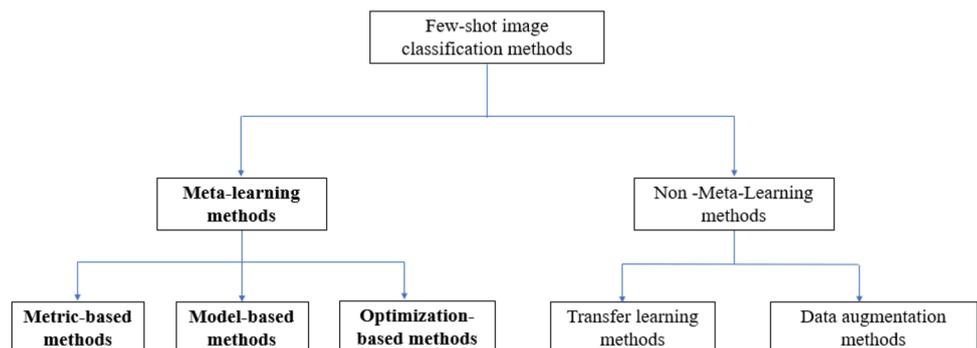
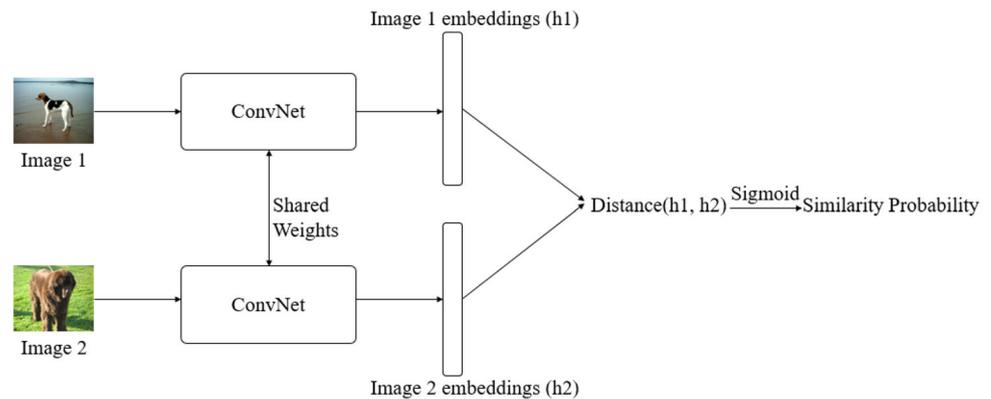
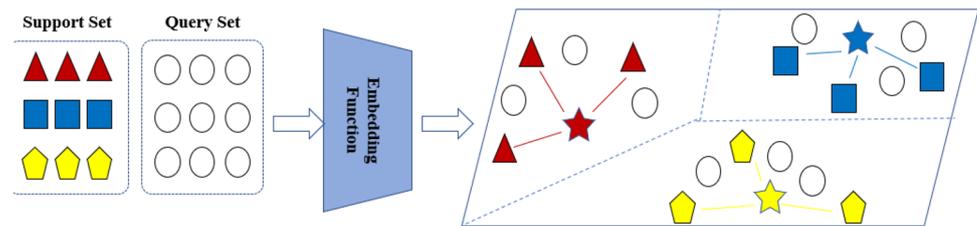


Fig. 3 Siamese network architecture**Fig. 4** Prototypical network architecture

a softmax function. Matching network's output is defined as a sum of the labels (one-hot encoded) of support images weighted by the attention kernel. In 2019, Mai et al. [92] proposed an attentive matching network (AMN), introducing a feature-level attention mechanism to pay more attention to the features that can better reflect the inter-class differences and a complementary cosine loss function for optimization.

The initial prototypical network was proposed by Snell et al. [145] based on the hypothesis that there exists an embedding space where each class can be represented by a unique prototype, and all samples are supposed to cluster around their corresponding prototypes. Figure 4 shows the architecture of prototypical network. A simple convolutional neural network with 4 layers is exploited to extract features, and the prototype of each class is defined as the mean value of feature embeddings from the support samples belonging to that class. The squared Euclidean distance is employed as a distance metric, calculating the distance between query embeddings and each class prototype. Build on this, Li et al. [86] proposed a covariance metric network (CovaMNet), using the covariance matrix of embedding vectors to represent the class prototype and also apply a covariance-based metric to measure the similarity between the query sample and the class prototype. Wang and Zhai [172] proposed a prototypical Siamese network (PSN), adding a prototype module in Siamese network to obtain high-quality prototype representations of each class.

Relation network [149] is the first study that employs a neural network to estimate a similarity score of feature embeddings rather than manual computation. This model consists of two main components: an embedding module

and a relation module. The embedding module is composed of convolutional blocks, mapping input images into an embedding space; and the relation module builds on two convolutional blocks and two fully connected layers, calculating a relation score between each query and support image (or a class prototype when the number of support samples is more than one). Note that the feature embeddings of support and query images need to be concatenated together before they are fed into the relation module. The architecture of relation network is presented in Fig. 5. In order to obtain discriminative features for fine-grained image classification [35, 59, 204], the subsequent work [73] proposed a bi-similarity network (BSNet), which combines an extra cosine module with the existing similarity measure as a new relation module, generating a more compact feature space by forcing features to adapt to the new relation module.

In order to get optimal matching image regions, Zhang et al. [196] proposed a DeepEMD algorithm, which adopts the earth mover's distance (EMD) [112, 127, 191] as a distance metric to calculate the similarity. They introduce a cross-reference mechanism to produce the weights of elements in the EMD formulation and embed the EMD layer into the network for end-to-end training. Motivated by this, Xie et al. [181] proposed a deep Brownian distance covariance (DeepBDC) approach, which applies BDC metric for few-shot learning. To learn discriminative feature representations, Afrasiyabi et al. [2] proposed a mixture-based feature space learning (MixtFSL) approach, learning both the feature representations and the mixture model via an online manner. Different from those few-shot classification methods that extract a single feature vector from each image,

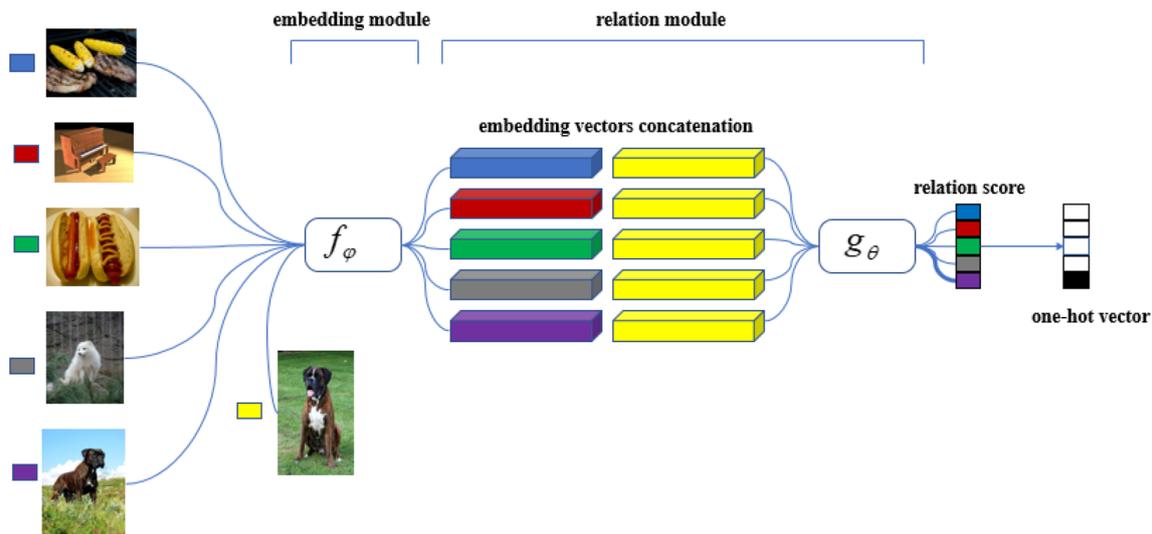


Fig. 5 Relation network architecture

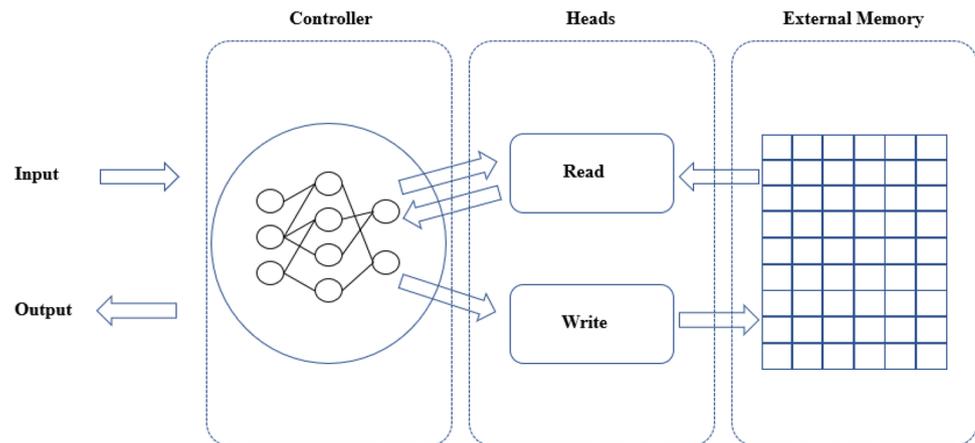
Table 1 A summary of presented metric-based meta-learning approaches

| Approach | Author | Year | Brief Description |
|---------------------------------------|------------------|------|---|
| Siamese network [65] | Koch et al | 2015 | Exploit a pair of identical networks with shared weights to extract features and calculate the distance between feature vectors |
| Attention-based Siamese network [173] | Wang et al | 2018 | Use an attention kernel function to measure the similarity |
| Multi-resolution Siamese network [88] | Lungu et al | 2020 | Mix different kernel size streams into one layer |
| Matching network [161] | Vinyals et al | 2016 | Use different networks to encode support and query images |
| AMN [92] | Mai et al | 2019 | Introduce a feature-level attention mechanism |
| Prototypical network [145] | Snell et al | 2017 | Propose class prototypes, and calculate the distances between query samples and class prototypes |
| CovaMNet [86] | Li et al | 2019 | Apply the covariance matrix of embedding vectors to represent the class prototype and use a covariance-based metric to measure the similarity |
| PSN [172] | Wang and Zhai | 2020 | Add a prototype module in Siamese network |
| Relation Network [149] | Sung et al | 2018 | Employ a neural network to estimate a similarity score of feature embeddings |
| BSNet [73] | Li et al | 2021 | Combine an extra cosine module with the existing similarity measure as a new relation module |
| DeepEMD [196] | Zhang et al | 2020 | Adopt the earth mover's distance (EMD) as a distance metric to calculate the similarity |
| DeepBDC [181] | Xie et al | 2022 | Apply the Brownian distance covariance (BDC) metric |
| MixtFSL [2] | Afrasiyabi et al | 2021 | Learn the feature representations and the mixture model jointly via an online manner |
| Matching feature sets [3] | Afrasiyabi et al | 2022 | Embed self-attention modules in between convolutional blocks and introduce set-to-set metrics |

Afrasiyabi et al. [3] held the view that a set-based representation can build a richer and more robust representation of images from base classes. To do so, they proposed a matching feature sets method which embeds self-attention modules in between convolutional blocks and introduces set-to-set metrics for evaluation. We summarize those introduced metric-based meta-learning approaches in Table 1.

3.2 Model-based meta-learning

With the goal of fast learning, model-based methods [63, 104] mainly focus on model architectures, adjusting model parameters based on presented tasks. There are several frequently used architectures in model-based methods, such as convolutional neural networks (CNNs) [71], recurrent

Fig. 6 Neural Turing machine scheme

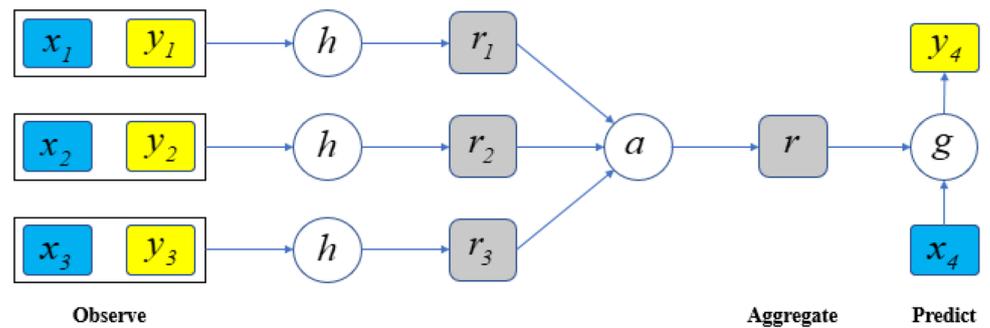
neural networks (RNNs) [128, 134] and long short-term memory (LSTM) [54]. According to the model architecture types, these model-based methods are further separated into memory-based, rapid adaptation-based and miscellaneous models.

Memory-augmented neural network (MANN) is a famous memory-based method proposed by Santoro et al. [131], which aims at improving task adaptation by utilizing the neural Turing machine (NTM) [21, 36, 47]. NTM is a neural network that integrates an external memory component during its learning process, enabling it has access to retrieve previously stored information. To be specific, NTM consists of a controller, interacting with an external memory module via a number of read and write heads. The NTM scheme is shown in Fig. 6. In MANN, a new addressing mechanism, namely least recently used access (LRUA) [131], is proposed, writing memories to either the least used memory location or the most recently used memory location. Through the stored information of a coupled representation-class label in the external memory, MANN can access them for later classification. Tran et al. [157] proposed a memory-augmented matching network (MAMN), which combines MANN and matching network. In MAMN, to reduce the biased on class prototypes caused by data distribution skew, weighted class prototypes are introduced by incorporating the distances of classwise samples. As another memory-based meta-learning method, memory matching network (MM-Net) [14] incorporates the memory module extracted from key-value memory network [96] into matching network. Different from traditional one-shot learning methods, MM-Net encodes and generalizes the whole support set into memory slots and can generate a unified model regardless of the number of shots and categories.

Meta-network (MetaNet) [100] is a model that designed with specific architecture and training process for rapid adap-

tion across tasks. Meta-network contains a base learner, a meta-learner and an external memory. It performs a generic knowledge acquisition in a meta-space and shifts its inductive biases via fast parameterization for rapid generalization. Conditional shifted neurons (CSNs) [101] is a generic neural mechanism designed for fast adaption, which is able to extract conditional information and generate conditional shifts for prediction during the meta-learning process. Compared with previous works [97, 100, 131], CSNs is more efficient computationally as the number of neurons is usually much smaller than that of weight parameters. Moreover, CSNs can be integrated into various neural architectures, including CNNs and RNNs. Similar to MetaNet, CSNs contains a base learner, a meta-learner and a memory module. During the description time, the meta-learner extracts and employs conditional information to generate memory values for samples within a task; at the prediction phase, the meta-learner generates query keys of query images by a key function for the purpose of getting the value of conditional shift.

Simple neural attentive learner (SNAIL) [98] is a general model-based meta-learning architecture that incorporates temporal convolution and soft attention mechanism. The temporal convolution acts as high-bandwidth memory access, and the soft attention enables access to specific pieces of information. This combination enables models to better leverage information from past experiences. Similar to SNAIL, Garnelo et al. [42] proposed conditional neural processes (CNPs) which consists of a meta-learner and task learner. The meta-learner generates a memory value by aggregating representations of the support set, and the task learner makes predictions by processing the aggregated representations. Figure 7 shows the CNPs scheme. We also make a short summary of those model-based meta-learning approaches and present it in Table 2.

Fig. 7 Conditional neural processes scheme**Table 2** A summary of presented model-based meta-learning approaches

| Approach | Author | Year | Brief Description |
|---------------|-------------------|------|--|
| MANN [131] | Santoro et al | 2016 | Employ a modified NTM to quickly assimilate new data into memory |
| MAMN [157] | Tran et al | 2019 | Combine MANN and matching network |
| MM-Net [14] | Cai et al | 2018 | Incorporate the memory module extracted from key-value memory network into matching network |
| MetaNet [100] | Munkhdalai and Yu | 2017 | Perform a generic knowledge acquisition in a meta-space and shift its inductive biases via fast parameterization |
| CSNs [101] | Munkhdalai et al | 2018 | Extract conditional information and generate conditional shifts for prediction during the meta-learning process |
| SNAIL [98] | Mishra et al | 2018 | Incorporate temporal convolution and soft attention mechanism |
| CNPs [42] | Garnelo et al | 2018 | Make predictions based on concise representations of seen classes |

3.3 Optimization-based meta-learning

Optimization-based meta-learning methods are an important vital branch in the field of few-shot image classification [11, 20, 37, 41, 121]. Basically, this kind of algorithm attempts to obtain a better initialization model or gradient descent direction by leveraging the meta-learning architecture and optimizes the initialization parameters through episodic training, enabling an optimization procedure to work on a small number of training samples. Optimization-based methods generally contain a task-specific learner trained for a given task and a meta-learner trained on distributions of tasks.

In 2017, Finn et al. [38] proposed model-agnostic meta-learning (MAML), the first algorithm for learning an initialization. The key idea of MAML is to enable a model's parameters can adapt fast to new unseen tasks through the gradient-based learning rule. During the meta-training phase, MAML attempt to update the task-specific parameters and the global initialization jointly in an iterative manner. The

MAML scheme is presented in Fig. 8. The main contribution of MAML is its compatibility in different application domains, not only in classification, but also in regression [133, 135, 199] and reinforcement learning [34, 51, 84]. To address the limitation of neural networks that are trained with gradient-based optimization on few-shot learning tasks [26, 143, 186], Ravi and Larochelle [120] proposed an LSTM-based meta-learner to learn both the exact task-specific optimization of a classifier, as well as good initialization values for the parameters of task-specific learner.

By taking ideas from prototypical network and MAML, Triantafillou et al. [158] proposed Proto-MAML, incorporating the advantages of both the former's simple inductive bias and the latter's flexible adaptation mechanism. As an extension to MAML, CAVIA [206] divides the model into parameters and task-specific context parameters which are shared across tasks. Compared with MAML, CAVIA is less prone to meta-overfitting and easier to parallelize. To address the issue that meta-learning models would be too

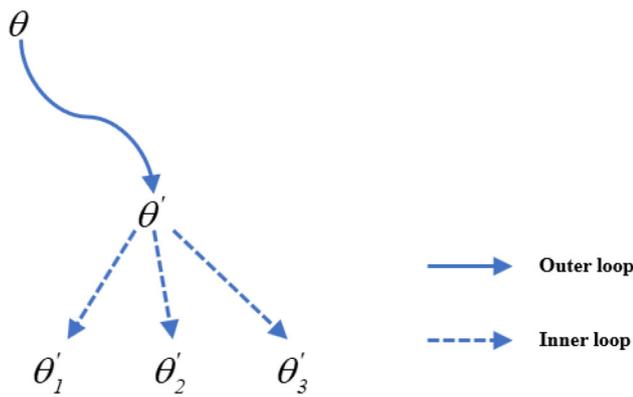


Fig. 8 Model-agnostic meta-learning scheme

biased toward existing tasks and lead to poor generalization, Jamal and Qi [60] proposed a task-agnostic meta-learning (TAML) algorithm, where two approaches are exploited to train a model unbiased over tasks. In order to improve generalization performance, BaiK et al. [8] proposed a novel framework called meta-learning with task-adaptive loss function (MeTAL). Particularly, MeTAL learns a task-adaptive loss function through two meta-learners and can be applied to different MAML variants.

Wang et al. [171] introduced a new approach called task-aware feature embeddings for low-shot learning (TAFE-Net) which mainly concentrates on tuning task-specific feature embedding through the generic embedding of a meta-learner. TAFE-Net is composed of a meta-learner and a prediction network, where the task-aware feature embedding is obtained by utilizing the meta-learner to develop task-specific feature layers of the prediction network. Sun et al. [152] introduced a meta-transfer learner (MTL) method, which focuses on generating task-specific feature extractors by leveraging both meta-learning and transfer learning. In MTL, scaling and shifting operations are introduced on pre-trained feature embeddings to freeze the feature extractor. Besides, similar fine-tuning steps are taken in MTL as those in previous work [18]. This work also proposed a novel hard task meta-batch

process that put more focus on hard tasks through sampling extra instances from the classes that the classifier failed.

Considering difficulties that exist in optimization on high-dimensional parameter spaces such as those faced by MAML [38], Rusu et al. [130] proposed an innovative algorithm called latent embedding optimization (LEO) that learns a low-dimensional latent representation of model parameters and performs optimization-based meta-learning in this space. Similar to MAML, LEO also consists of an inner loop training where the task-specific values are learned and an outer loop training where global shared initializations are updated. To instantiate low-dimensional latent embedding of model’s parameters, samples pass through a combination of an encoder and a relation network. The encoder is used to generate hidden codes from the support set. Then, these hidden codes are concatenated pairwise and fed into a relation network, leading to a probability distribution over latent codes in a lower dimension. Finally, the decoder produces task-specific initial parameters which are differentiable to backpropagate for adaptation. The LEO scheme is shown in Fig. 9. We present a short summary of optimization-based meta-learning approaches in Table 3.

3.4 Other methods

Transfer learning involves leveraging knowledge learned from a related task to enhance learning in a new task [52, 125, 126, 187, 189]. In the few-shot image classification scenario, transferring knowledge from another network is a viable option when original data is too limited to train a deep neural network from scratch. Compared with meta-learning, the learning experience involved in transfer learning is much narrower. To address few-shot hyperspectral image classification problems, Qu et al. [118] applied the transfer learning scheme to extract learned intrinsic representations from the same kind of objects in different domains. Tai et al. [154] proposed a novel few-shot transfer learning approach for synthetic aperture radar image classification, which uses a

Fig. 9 Latent embedding optimization scheme

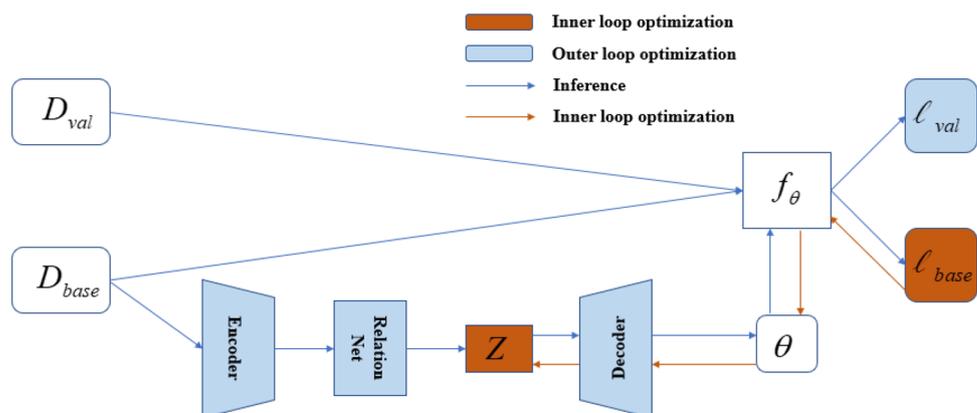


Table 3 A summary of presented optimization-based meta-learning approaches

| Approach | Author | Year | Brief Description |
|-------------------------------|---------------------|------|--|
| MAML [38] | Finn et al | 2017 | Enable a model's parameters can adapt fast to new unseen tasks through the gradient-based learning rule |
| LSTM-based meta-learner [120] | Ravi and Larochelle | 2017 | Learn both the exact task-specific optimization of a classifier and good initialization values for the parameters of task-specific learner |
| Proto-MAML [158] | Triantafillou et al | 2020 | Combine prototypical network and MAML |
| CAVIA [206] | Zintgraf et al | 2019 | Divide the model into parameters and task-specific context parameters which are shared across tasks |
| TAML [60] | Jamal and Qi | 2019 | Train a model unbiased over tasks |
| MeTAL [8] | Baik et al | 2021 | Learn a task-adaptive loss function through two meta-learners |
| TAFE-Net [171] | Wang et al | 2019 | Tune task-specific feature embedding through the generic embedding of a meta-learner |
| MTL [152] | Sun et al | 2019 | Generate task-specific feature extractors by leveraging both meta-learning and transfer learning |
| LEO [130] | Rusu et al | 2019 | Learn a low-dimensional latent representation of model parameters |

connection-free attention module to transfer features from a source network to a target network. Sun and Yang [147] proposed trans-transfer learning, a two-phase learning method for few-shot fine-grained visual categorization problems. In some cases, knowledge transfer may also fail when the source domain and target domain are not related to each other, even causing negative transfer. To address this problem, Liu et al. [83] proposed an analogical transfer learning (ATL), following the analogy strategy to effectively control the occurrence of negative transfer.

Considering the fundamental problem in few-shot image classification that models are prone to overfitting caused by few training samples, many researchers proposed a number of data augmentation approaches [108, 117, 174] to improve sample diversity and prevent overfitting during training. Goodfellow et al. [46] proposed the well-known Generative Adversarial Nets (GAN), which contains a generator for generating similar images and a discriminator for distinguishing. Based on GAN, Mehrotra and Dukkipati [95] proposed to generate samples for specific tasks, enabling these generated samples more suitable for few-shot learning. Zhang et al. [197] proposed MetaGAN. To help the classifier learn a clearer decision boundary, MetaGAN involves GAN and part of the classification network during the training process. Li et al. [87] proposed Adversarial Feature Hallucination Network (AFHN), using conditional Wasserstein Generative Adversarial Network (cWGAN) to generate samples.

We present experimental results of recent meta-learning methods in Tables 4 and 5. Table 4 shows performances

of different approaches on Omniglot. Omniglot is a handwritten dataset with multiple handwriting styles, languages and stroke types, this diversity makes Omniglot suitable for training deep learning algorithms. Table 4 shows that most meta-learning approaches obtain over 98% accuracies on Omniglot. Table 5 shows experimental results on Mini-ImageNet and TieredImageNet. These two datasets contain images with different objects, scenes and lighting conditions, which can improve the model's robustness. However, the limitations in dataset size and image quality may affect the model's performance. Table 5 shows that DeepBDC [181] and matching feature sets [3] achieved best results on both datasets.

4 Major challenges and future directions

Although meta-learning methods have achieved promising performance in few-shot image classification, there remain some vital challenges that ought to be dealt with in the future. These existing issues and suggested future research directions are outlined here.

4.1 Limitations and challenges

- *Data availability and computational complexity.* In image classification, a large dataset typically has a thousand (or more) categories. Meta-learning approaches also require a large amount of data and computational resources, but

Table 4 Accuracy results on Omniglot dataset reported in original papers, with mean accuracy (%) and 95% confidence interval. i: metric-based; ii: model-based; iii: optimization-based

| Model | Category | Omniglot five-way | |
|---------------------------------------|----------|-------------------|--------------|
| | | One-shot | Five-shot |
| Siamese network [65] | i | 97.30 | 98.40 |
| Attention-based Siamese network [173] | i | 99.60 | 99.80 |
| Multi-resolution Siamese network [88] | i | – | – |
| Matching network [161] | i | 98.10 | 98.90 |
| AMN [92] | i | 99.44 ± 0.09 | 99.86 ± 0.06 |
| Prototypical network [145] | i | 98.80 | 99.70 |
| CovaMNet [86] | i | – | – |
| PSN [172] | i | – | – |
| Relation Network [149] | i | 99.60 ± 0.20 | 99.80 ± 0.10 |
| BSNet [73] | i | – | – |
| DeepEMD [196] | i | – | – |
| DeepBDC [181] | i | – | – |
| MixtFSL [2] | i | – | – |
| Matching feature sets [3] | i | – | – |
| MANN [131] | ii | 82.80 | 94.90 |
| MAMN [157] | ii | 98.90 | 99.70 |
| MM-Net [14] | ii | 99.28 ± 0.08 | 99.77 ± 0.04 |
| MetaNet [100] | ii | 98.45 | – |
| CSNs [101] | ii | 98.42 ± 0.21 | 99.37 ± 0.28 |
| SNAIL [98] | ii | 99.07 ± 0.16 | 99.78 ± 0.09 |
| CNPs [42] | ii | 95.30 | 98.50 |
| MAML [38] | iii | 98.70 ± 0.40 | 99.90 ± 0.10 |
| LSTM-based meta-learner [120] | iii | – | – |
| Proto-MAML [158] | iii | – | – |
| CAVIA [206] | iii | – | – |
| TAML [60] | iii | 99.47 ± 0.25 | 99.83 ± 0.09 |
| MeTAL [8] | iii | – | – |
| TAFE-Net [171] | iii | – | – |
| MTL [152] | iii | – | – |
| LEO [130] | iii | – | – |

in few-shot scenarios, it is quite challenging to collect sufficient data. For deep testing of meta-learning we may need thousands of large datasets! This may also be very difficult and slow to process.

- **Model selection** There is not a one-size-fits-all so selecting an appropriate model is important. Model selection is more crucial in few-shot image classification scenarios as the model is prone to overfitting the training data. The model may perform well on the base set and lacks generalization on new tasks.
- **Transferability** Meta-learning models can transfer learned knowledge between various tasks. The success of transferability depends on the similarity between the tasks. Sometimes new tasks may have significant differences from old ones, making it difficult to transfer learned knowledge effectively, such as cross-domain tasks.

- **Task dependence** Most meta-learning approaches are designed to work for a specific set of tasks or domains. They may not perform well on new tasks or domains that are significantly different from the ones used during training. Improving meta-learning's generalization ability can be a hard task.
- **Interpretability** Interpretability is a critical aspect of neural approaches that refers to the ability to understand how a model works. Unfortunately, all neural approaches can be extremely challenging to interpret and thus difficult to understand how it learns to learn and make predictions or decisions. This issue can make it arduous to debug, diagnose and improve models' performances.

Table 5 Accuracy results on MiniImageNet and TieredImageNet datasets reported in original papers, with mean accuracy (%) and 95% confidence interval. i: metric-based; ii: model-based; iii: optimization-based

| Model | Category | MiniImageNet five-way | | TieredImageNet five-way | |
|---------------------------------------|----------|-----------------------|--------------|-------------------------|--------------|
| | | One-shot | Five-shot | One-shot | Five-shot |
| Siamese network [65] | i | – | – | – | – |
| Attention-based Siamese network [173] | i | 51.20 | 69.70 | – | – |
| Multi-resolution Siamese network [88] | i | – | – | – | – |
| Matching network [161] | i | 46.60 | 60.00 | – | – |
| AMN [92] | i | 54.97 ± 0.77 | 71.84 ± 0.67 | – | – |
| Prototypical network [145] | i | 49.42 ± 0.78 | 68.20 ± 0.66 | – | – |
| CovaMNet [86] | i | 51.19 ± 0.76 | 67.65 ± 0.63 | – | – |
| PSN [172] | i | 48.70 | 69.40 | – | – |
| Relation Network [149] | i | 50.44 ± 0.82 | 65.32 ± 0.70 | – | – |
| BSNet [73] | i | – | – | – | – |
| DeepEMD [196] | i | 65.91 ± 0.82 | 82.41 ± 0.56 | 71.16 ± 0.87 | 86.03 ± 0.58 |
| DeepBDC [181] | i | 67.83 ± 0.43 | 85.45 ± 0.29 | 73.82 ± 0.47 | 89.00 ± 0.30 |
| MixtFSL [2] | i | 64.31 ± 0.79 | 81.66 ± 0.60 | 70.97 ± 1.03 | 86.16 ± 0.67 |
| Matching feature sets [3] | i | 68.32 ± 0.62 | 82.71 ± 0.46 | 73.63 ± 0.88 | 87.59 ± 0.57 |
| MANN [131] | ii | – | – | – | – |
| MAMN [157] | ii | 49.80 | 66.50 | – | – |
| MM-Net [14] | ii | 53.37 ± 0.48 | 66.97 ± 0.35 | – | – |
| MetaNet [100] | ii | 49.21 ± 0.96 | – | – | – |
| CSNs [101] | ii | 56.88 ± 0.62 | 71.94 ± 0.57 | – | – |
| SNAIL [98] | ii | 55.71 ± 0.99 | 68.88 ± 0.92 | – | – |
| CNPs [42] | ii | – | – | – | – |
| MAML [38] | iii | 48.70 ± 1.84 | 63.11 ± 0.92 | – | – |
| LSTM-based meta-learner [120] | iii | 43.44 ± 0.77 | 60.60 ± 0.71 | – | – |
| Proto-MAML [158] | iii | – | – | – | – |
| CAVIA [206] | iii | 51.82 ± 0.65 | 65.85 ± 0.55 | – | – |
| TAML [60] | iii | 51.73 ± 1.88 | 66.05 ± 0.85 | – | – |
| MeTAL [8] | iii | 66.61 ± 0.28 | 81.43 ± 0.25 | 70.29 ± 0.40 | 86.17 ± 0.35 |
| TAFE-Net [171] | iii | – | – | – | – |
| MTL [152] | iii | 61.20 ± 1.80 | 75.50 ± 0.80 | – | – |
| LEO [130] | iii | 61.76 ± 0.08 | 77.59 ± 0.12 | 66.33 ± 0.05 | 81.44 ± 0.09 |

4.2 Future directions

- Enhancing generalized feature learning** To address the main challenge in few-shot learning that learn from a handful of samples [81, 146, 182], meta-learning employs shared knowledge from previously experienced tasks for unseen tasks. However, in most existing meta-learning methods, researchers attempt to learn discriminative features via attention mechanism, multitask learning, data augmentation and so on. One major research direction is developing new approaches for learning features that generalize better to new domains; and evaluation measures for assessment and selection of the learned features.
- Practice of episodic training strategy** In order to realize fast adaption to new tasks with limited samples, episodic training requires that each training episode should have the same number of classes and examples as the evaluation episode. But, this setting is prone to catastrophic forgetting [31, 138, 176] and leads to model underfitting in base classes. A number of approaches have been proposed to address this issue, and improving model performance on both base and novel classes remains a vital direction for future work.

- **Improving stability** Despite the continuous improvement of meta-learning in few-shot image classification, one existing issue is that some meta-learning methods obtain state-of-the-art performance on special datasets, but perform not well on other benchmarks. For example, a metric-based meta-learning method named global class representation (GCR) [78] achieved great performance on Omniglot, but cannot compete with other non-metric-based methods on miniImageNet. Further exploration of stable models [25, 66] will be very valuable.
- **Cross-domain and multimodal meta-learning** In principle, the base dataset D_{base} and novel dataset D_{novel} in few-shot learning can be from different domains [77, 150]. However, most model performances will decline when the difference between D_{base} and D_{novel} . Developing meta-learning methods on cross-domain performance can be one future research direction. Multimodal deep learning has also brought great opportunities to few-shot learning [53, 99, 109]. For example, Peng et al. [113] proposed a Knowledge Transfer Network (KTN), which combines semantic features and image features for few-shot image classification tasks. Therefore, how to design a more appropriate multimodal fusion method is a research trend in few-shot image classification.

5 Conclusions

This paper presents a survey comprised of over 200 papers on recent few-shot learning and meta-learning research for image understanding. Based on the research literature, we introduce the general approaches for few-shot learning and then turn to one of the key approaches called meta-learning. We separate existing meta-learning methods into three important categories: metric-based, model-based and optimization-based methods. We introduce both classical and state-of-the-art approaches in each category and summarize the state of the art. We also present the state-of-the-art performance of the literature approaches on well-known datasets. According to our study, we conclude with limitations, challenges and weaknesses for meta-learning and present promising directions of meta-learning from the perspectives of generalization, effectiveness and applicability.

Acknowledgements This work was supported by LIACS MediaLab at Leiden University and China Scholarship Council (CSC No.201703170183).

Author Contributions KH wrote the main manuscript text and prepared figures. All authors reviewed the manuscript and gave comments.

Funding This study was funded by China Scholarship Council (CSC No.201703170183).

Research Data Policy and Data Availability Statements The data that support the findings of this study are openly available at (<https://github.com/brendenlake/omniglot>), (<https://few-shot.yyliu.net/miniimagenet.html>), (<https://few-shot.yyliu.net/tieredimagenet.html>), (<https://few-shot.yyliu.net/fc100.html>), (<https://few-shot.yyliu.net/cifarfs.html>), (http://www.vision.caltech.edu/datasets/cub_200_2011/).

(<https://few-shot.yyliu.net/miniimagenet.html>), (<https://few-shot.yyliu.net/tieredimagenet.html>), (<https://few-shot.yyliu.net/fc100.html>), (<https://few-shot.yyliu.net/cifarfs.html>), (http://www.vision.caltech.edu/datasets/cub_200_2011/).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Afrasiyabi A, Lalonde J, Gagné C (2020) Associative alignment for few-shot image classification. In: ECCV, pp 18–35
2. Afrasiyabi A, Lalonde J, Gagné C (2021) Mixture-based feature space learning for few-shot image classification. In: ICCV, pp 9021–9031
3. Afrasiyabi A, Larochelle H, Lalonde J et al (2022) Matching feature sets for few-shot image classification. In: CVPR, pp 9004–9014
4. Akata Z, Geiger A, Sattler T (2021) Computer vision and pattern recognition 2020. *Int J Comput Vis* 129(12):3169–3170
5. Alzubaidi L, Zhang J, Humaidi AJ et al (2021) Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data* 8(1):53
6. Antol S, Agrawal A, Lu J et al (2015) VQA: visual question answering. In: ICCV, pp 2425–2433
7. Antoniou A, Storkey A (2019) Assume, augment and learn: Unsupervised few-shot meta-learning via random labels and data augmentation. *arXiv preprint arXiv:1902.09884*
8. Baik S, Choi J, Kim H et al (2021) Meta-learning with task-adaptive loss function for few-shot learning. In: ICCV, pp 9445–9454
9. Baldi P, Chauvin Y (1993) Neural networks for fingerprint recognition. *Neural Comput* 5(3):402–418
10. Bertinetto L, Henriques JF, Torr PHS et al (2019) Meta-learning with differentiable closed-form solvers. In: ICLR
11. Bian W, Chen Y, Ye X et al (2021) An optimization-based meta-learning model for MRI reconstruction with diverse dataset. *J Imaging* 7(11):231
12. Brauwers G, Frasincar F (2023) A general survey on attention mechanisms in deep learning. *IEEE Trans Knowl Data Eng* 35(4):3279–3298
13. Bromley J, Guyon I, LeCun Y et al (1993) Signature verification using a siamese time delay neural network. In: *NeurIPS*, pp 737–744
14. Cai Q, Pan Y, Yao T et al (2018) Memory matching networks for one-shot image recognition. In: CVPR, pp 4080–4088

15. Cai J, Shen SM (2020) Cross-domain few-shot learning with meta fine-tuning. arXiv preprint [arXiv:2005.10544](https://arxiv.org/abs/2005.10544)
16. Chao X, Zhang L (2021) Few-shot imbalanced classification based on data augmentation. *Multimed Syst*, pp 1–9
17. Chen Z, Fu Y, Wang Y et al (2019b) Image deformation meta-networks for one-shot learning. In: *CVPR*, pp 8680–8689
18. Chen W, Liu Y, Kira Z et al (2019a) A closer look at few-shot classification. In: *ICLR*
19. Chen Y, Liu Z, Xu H et al (2021) Meta-baseline: exploring simple meta-learning for few-shot learning. In: *ICCV*, pp 9042–9051
20. Cho H, Cho Y, Yu J et al (2021) Camera distortion-aware 3d human pose estimation in video with optimization-based meta-learning. In: *ICCV*, pp 11,149–11,158
21. Collier M, Beel J (2018) Implementing neural turing machines. In: *ICANN*, pp 94–104
22. Deng J, Dong W, Socher R et al (2009) Imagenet: a large-scale hierarchical image database. In: *CVPR*, pp 248–255
23. Deng S, Liao D, Gao X et al (2022) Improving few-shot image classification with self-supervised learning. In: *Cloud Computing*, pp 54–68
24. Dhillon GS, Chaudhari P, Ravichandran A et al (2020) A baseline for few-shot image classification. In: *ICLR*
25. Ding G, Han X, Wang S et al (2022a) Attribute group editing for reliable few-shot image generation. In: *CVPR*, pp 11,184–11,193
26. Ding L, Liu P, Shen W et al (2022b) Gradient-based meta-learning using uncertainty to weigh loss for few-shot learning. arXiv preprint [arXiv:2208.08135](https://arxiv.org/abs/2208.08135)
27. Dong J, Wang Y, Lai J et al (2022) Improving adversarially robust few-shot image classification with generalizable representations. In: *CVPR*, pp 9015–9024
28. dos Santos FP, Thumé GS, Ponti MA (2021) Data augmentation guidelines for cross-dataset transfer learning and pseudo labeling. In: *SIBGRAPI*, pp 207–214
29. Do J, Yoo M, Kim S (2022) A semi-supervised sar image classification with data augmentation and pseudo labeling. In: *ICCE-Asia*, pp 1–4
30. Dumoulin V, Houlsby N, Evci U et al (2021) Comparing transfer and meta learning approaches on a unified few-shot classification benchmark. arXiv preprint [arXiv:2104.02638](https://arxiv.org/abs/2104.02638)
31. Ebrahimi S, Petryk S, Gokul A et al (2021) Remembering for the right reasons: explanations reduce catastrophic forgetting. In: *ICLR*
32. Eloff R, Engelbrecht HA, Kamper H (2019) Multimodal one-shot learning of speech and images. In: *ICASSP*, pp 8623–8627
33. Elsken T, Staffler B, Metzen JH et al (2020) Meta-learning of neural architectures for few-shot learning. In: *CVPR*, pp 12,362–12,372
34. Fallah A, Mokhtari A, Ozdaglar A (2020) Provably convergent policy gradient methods for model-agnostic meta-reinforcement learning. arXiv preprint [arXiv:2002.05135](https://arxiv.org/abs/2002.05135)
35. Fan M, Bai Y, Sun M et al (2019) Large margin prototypical network for few-shot relation classification with fine-grained features. In: *CIKM*, pp 2353–2356
36. Faradonbe SM, Safi-Esfahani F, Karimian-kelishadrokh M (2020) A review on neural turing machine (NTM). *SN Comput Sci* 1(6):333
37. Feurer M, Springenberg JT, Hutter F (2015) Initializing Bayesian hyperparameter optimization via meta-learning. In: *AAAI*, pp 1128–1135
38. Finn C, Abbeel P, Levine S (2017) Model-agnostic meta-learning for fast adaptation of deep networks. In: Precup D, Teh YW (eds) *ICML*, pp 1126–1135
39. Finn C, Xu K, Levine S (2018) Probabilistic model-agnostic meta-learning. In: *NeurIPS*, pp 9537–9548
40. Gaikwad M, Doke A (2022) Survey on meta learning algorithms for few shot learning. In: *ICICCS*, pp 1876–1879
41. Gao K, Sener O (2020) Modeling and optimization trade-off in meta-learning. In: *NeurIPS*, pp 11,154–11,165
42. Garnelo M, Rosenbaum D, Maddison C et al (2018) Conditional neural processes. In: *ICML*, pp 1690–1699
43. Gidaris S, Komodakis N (2018) Dynamic few-shot visual learning without forgetting. In: *CVPR*, pp 4367–4375
44. Gidaris S, Komodakis N (2019) Generating classification weights with GNN denoising autoencoders for few-shot learning. In: *CVPR*, pp 21–30
45. Goldblum M, Fowl L, Goldstein T (2020) Adversarially robust few-shot learning: a meta-learning approach. In: *NeurIPS*, pp 17,886–17,895
46. Goodfellow IJ, Pouget-Abadie J, Mirza M et al (2014) Generative adversarial nets. In: *NeurIPS*, pp 2672–2680
47. Graves A, Wayne G, Danihelka I (2014) Neural turing machines. arXiv preprint [arXiv:1410.5401](https://arxiv.org/abs/1410.5401)
48. Gu J, Wang Z, Kuen J et al (2018) Recent advances in convolutional neural networks. *Pattern Recognit* 77:354–377
49. Guo N, Di K, Liu H et al (2021) A metric-based meta-learning approach combined attention mechanism and ensemble learning for few-shot learning. *Displays* 70(102):065
50. Guo Y, Codella N, Karlinsky L et al (2020) A broader study of cross-domain few-shot learning. In: *ECCV*, pp 124–141
51. Gupta A, Mendonca R, Liu Y et al (2018) Meta-reinforcement learning of structured exploration strategies. In: *NeurIPS*, pp 5307–5316
52. Gupta A, Thadani K, O’Hare N (2020) Effective few-shot classification with transfer learning. In: *COLING*, pp 1061–1066
53. Han G, Ma J, Huang S et al (2022) Multimodal few-shot object detection with meta-learning based cross-modal prompting. arXiv preprint [arXiv:2204.07841](https://arxiv.org/abs/2204.07841)
54. Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780
55. Hou R, Chang H, Ma B et al (2019) Cross attention network for few-shot classification. In: *NeurIPS*, pp 4005–4016
56. Hou M, Sato I (2022) A closer look at prototype classifier for few-shot image classification. In: *NeurIPS*, pp 25,767–25,778
57. Hu T, Tang T, Lin R et al (2020) A simple data augmentation algorithm and a self-adaptive convolutional architecture for few-shot fault diagnosis under different working conditions. *Measurement* 156(107):539
58. Huang W, He M, Wang Y (2021) A survey on meta-learning based few-shot classification. In: *MLICOM*, pp 243–253
59. Huang H, Zhang J, Zhang J et al (2019) Compare more nuanced: pairwise alignment bilinear network for few-shot fine-grained learning. In: *ICME*, pp 91–96
60. Jamal MA, Qi G (2019) Task agnostic meta-learning for few-shot learning. In: *CVPR*, pp 11,719–11,727
61. Kang D, Kwon H, Min J et al (2021) Relational embedding for few-shot classification. In: *ICCV*, pp 8802–8813
62. Kang B, Liu Z, Wang X et al (2019) Few-shot object detection via feature reweighting. In: *ICCV*, pp 8419–8428
63. Karunaratne G, Schmuck M, Le Gallo M et al (2021) Robust high-dimensional memory-augmented neural networks. *Nat Commun* 12(1):2468
64. Khodadadeh S, Bölöni L, Shah M (2019) Unsupervised meta-learning for few-shot image classification. In: *NeurIPS*, pp 10,132–10,142
65. Koch G, Zemel R, Salakhutdinov R et al (2015) Siamese neural networks for one-shot image recognition. In: *ICML deep learning workshop*
66. Köksal A, Schick T, Schütze H (2022) Meal: stable and active learning for few-shot prompting. arXiv preprint [arXiv:2211.08358](https://arxiv.org/abs/2211.08358)
67. Krizhevsky A, Sutskever I, Hinton GE (2017) Imagenet classification with deep convolutional neural networks. *ACM* 60(6):84–90

68. Krizhevsky A, Hinton G et al (2009) Learning multiple layers of features from tiny images
69. Kulkarni TD, Whitney WF, Kohli P et al (2015) Deep convolutional inverse graphics network. In: NeurIPS, pp 2539–2547
70. Lake BM, Salakhutdinov R, Gross J et al (2011) One shot learning of simple visual concepts. In: Proceedings of the 33th annual meeting of the cognitive science society
71. LeCun Y, Bottou L, Bengio Y et al (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86(11):2278–2324
72. Li X, Yu L, Fu C et al (2020) Revisiting metric learning for few-shot image classification. *Neurocomputing* 406:49–58
73. Li X, Wu J, Sun Z et al (2021) Bsnet: Bi-similarity network for few-shot fine-grained image classification. *IEEE Trans Image Process* 30:1318–1331
74. Li X, Sun Z, Xue J et al (2021) A concise review of recent few-shot meta-learning methods. *Neurocomputing* 456:463–468
75. Li P, Zhao G, Xu X (2022) Coarse-to-fine few-shot classification with deep metric learning. *Inf Sci* 610:592–604
76. Li X, Yang X, Ma Z et al (2023) Deep metric learning for few-shot image classification: a review of recent developments. *Pattern Recognit* 138(109):381
77. Li W, Liu X, Bilen H (2022b) Cross-domain few-shot learning with task-specific adapters. In: CVPR, pp 7151–7160
78. Li A, Luo T, Xiang T et al (2019a) Few-shot learning with global class representations. In: ICCV, pp 9714–9723
79. Liu W, Wang Z, Liu X et al (2017) A survey of deep neural network architectures and their applications. *Neurocomputing* 234:11–26
80. Liu B, Guo W, Chen X et al (2020) Morphological attribute profile cube and deep random forest for small sample classification of hyperspectral image. *IEEE Access* 8:117:096–117:108
81. Liu Y, Zhang H, Zhang W et al (2022) Few-shot image classification: current status and research trends. *Electronics* 11(11):1752
82. Liu B, Cao Y, Lin Y et al (2020a) Negative margin matters: understanding margin in few-shot classification. In: ECCV, pp 438–455
83. Liu W, Chang X, Yan Y et al (2018) Few-shot text and image classification via analogical transfer learning. *ACM* 9(6):71:1–71:20
84. Liu H, Socher R, Xiong C (2019) Taming MAML: efficient unbiased meta-reinforcement learning. In: ICML, pp 4061–4071
85. Li W, Wang L, Xu J et al (2019b) Revisiting local descriptor based image-to-class measure for few-shot learning. In: CVPR, pp 7260–7268
86. Li W, Xu J, Huo J et al (2019c) Distribution consistency based covariance metric networks for few-shot learning. In: AAAI Conference on Artificial Intelligence, pp 8642–8649
87. Li K, Zhang Y, Li K et al (2020a) Adversarial feature hallucination networks for few-shot learning. In: CVPR, pp 13,467–13,476
88. Lungu I, Hu Y, Liu S (2020) Multi-resolution siamese networks for one-shot learning. In: AICAS, pp 183–187
89. Luo S, Li Y, Gao P et al (2022) Meta-seg: a survey of meta-learning for image segmentation. *Pattern Recognit* 126(108):586
90. Lu J, Yang J, Batra D et al (2016) Hierarchical question-image co-attention for visual question answering. In: NIPS, pp 289–297
91. Mahesh B (2020) Machine learning algorithms-a review. *IJSR* 9:381–386
92. Mai S, Hu H, Xu J (2019) Attentive matching network for few-shot learning. *Comput Vis Image Underst* 187(102):781
93. Mangla P, Singh M, Sinha A et al (2020) Charting the right manifold: manifold mixup for few-shot learning. In: WACV, pp 2207–2216
94. Ma J, Xie H, Han G et al (2021) Partner-assisted learning for few-shot image classification. In: ICCV, pp 10,553–10,562
95. Mehrotra A, Dukkipati A (2017) Generative adversarial residual pairwise networks for one shot learning. *arXiv preprint arXiv:1703.08033*
96. Miller AH, Fisch A, Dodge J et al (2016) Key-value memory networks for directly reading documents. In: EMNLP, pp 1400–1409
97. Mishra N, Rohaninejad M, Chen X et al (2017) Meta-learning with temporal convolutions. *arXiv preprint arXiv:1707.03141*
98. Mishra N, Rohaninejad M, Chen X et al (2018) A simple neural attentive meta-learner. In: ICLR
99. Moon J, Le NA, Minaya NH et al (2020) Multimodal few-shot learning for gait recognition. *Appl Sci* 10(21):7619
100. Munkhdalai T, Yu H (2017) Meta networks. In: ICML, pp 2554–2563
101. Munkhdalai T, Yuan X, Mehri S et al (2018) Rapid adaptation with conditionally shifted neurons. In: ICML, pp 3661–3670
102. Najdenkoska I, Zhen X, Worring M (2023) Meta learning to bridge vision and language models for multimodal few-shot learning. *arXiv preprint arXiv:2302.14794*
103. Nguyen VN, Løkse S, Wickstrøm K et al (2020) SEN: a novel feature normalization dissimilarity measure for prototypical few-shot learning networks. In: ECCV, pp 118–134
104. Nichol A, Schulman J (2018) Reptile: a scalable metalearning algorithm. *arXiv preprint arXiv:1803.02999*
105. Niu Z, Zhong G, Yu H (2021) A review on the attention mechanism of deep learning. *Neurocomputing* 452:48–62
106. Niu Z, Zhong G, Yu H (2021) A review on the attention mechanism of deep learning. *Neurocomputing* 452:48–62
107. Oreshkin BN, López PR, Lacoste A (2018) TADAM: task dependent adaptive metric for improved few-shot learning. In: NeurIPS, pp 719–729
108. Osahor U, Nasrabadi NM (2022) Ortho-shot: low displacement rank regularization with data augmentation for few-shot learning. In: WACV, pp 2040–2049
109. Pahde F, Puscas MM, Klein T et al (2021) Multimodal prototypical networks for few-shot learning. In: WACV, pp 2643–2652
110. Park S, Mello SD, Molchanov P et al (2019) Few-shot adaptive gaze estimation. In: ICCV, pp 9367–9376
111. Parnami A, Lee M (2022) Learning from few examples: a summary of approaches to few-shot learning. *arXiv preprint arXiv:2203.04291*
112. Pele O, Werman M (2009) Fast and robust earth mover’s distances. In: ICCV, pp 460–467
113. Peng Z, Li Z, Zhang J et al (2019) Few-shot image recognition with knowledge transfer. In: ICCV, pp 441–449
114. Pérez-Rúa J, Zhu X, Hospedales TM et al (2020) Incremental few-shot object detection. In: CVPR, pp 13,843–13,852
115. Qiao S, Liu C, Shen W et al (2018) Few-shot image recognition by predicting parameters from activations. In: CVPR, pp 7229–7238
116. Qiao L, Shi Y, Li J et al (2019) Transductive episodic-wise adaptive metric for few-shot learning. In: ICCV, pp 3602–3611
117. Qin T, Li W, Shi Y et al (2020) Diversity helps: unsupervised few-shot learning via distribution shift-based data augmentation. *arXiv preprint arXiv:2004.05805*
118. Qu Y, Baghbaderani RK, Qi H (2019) Few-shot hyperspectral image classification through multitask transfer learning. In: WHISPERS, pp 1–5
119. Ratner AJ, Ehrenberg HR, Hussain Z et al (2017) Learning to compose domain-specific transformations for data augmentation. In: NeurIPS, pp 3236–3246
120. Ravi S, Larochelle H (2017) Optimization as a model for few-shot learning. In: ICLR
121. Reif M, Shafait F, Dengel A (2012) Meta-learning for evolutionary parameter optimization of classifiers. *Mach Learn* 87(3):357–380
122. Ren M, Triantafyllou E, Ravi S et al (2018) Meta-learning for semi-supervised few-shot classification. In: ICLR
123. Rohrbach M, Ebert S, Schiele B (2013) Transfer learning in a transductive setting. In: NeurIPS, pp 46–54

124. Romera-Paredes B, Torr PHS (2015) An embarrassingly simple approach to zero-shot learning. In: ICML, pp 2152–2161
125. Rostami M, Kolouri S, Eaton E et al (2019) Deep transfer learning for few-shot SAR image classification. *Remote Sens* 11(11):1374
126. Rostami M, Kolouri S, Eaton E et al (2019b) SAR image classification using few-shot cross-domain transfer learning. In: CVPR, pp 907–915
127. Rubner Y, Tomasi C, Guibas LJ (2000) The earth mover's distance as a metric for image retrieval. *Int J Comput Vis* 40(2):99–121
128. Rumelhart DE, McClelland JL (1986) On learning the past tenses of English verbs
129. Russakovsky O, Deng J, Su H et al (2015) Imagenet large scale visual recognition challenge. *Int J Comput Vis* 115(3):211–252
130. Rusu AA, Rao D, Sygnowski J et al (2019) Meta-learning with latent embedding optimization. In: ICLR
131. Santoro A, Bartunov S, Botvinick MM et al (2016) Meta-learning with memory-augmented neural networks. In: ICML, pp 1842–1850
132. Sarker IH (2021) Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN Comput Sci* 2(6):420
133. Satrya WF, Yun J (2023) Combining model-agnostic meta-learning and transfer learning for regression. *Sensors* 23(2):583
134. Schuster M, Paliwal KK (1997) Bidirectional recurrent neural networks. *IEEE Trans Signal Process* 45(11):2673–2681
135. Sendera M, Tabor J, Nowak A et al (2021) Non-gaussian gaussian processes for few-shot regression. In: NeurIP, pp 10,285–10,298
136. Shahroudy A, Liu J, Ng T et al (2016) NTU RGB+D: a large scale dataset for 3d human activity analysis. In: CVPR, pp 1010–1019
137. Shen Z, Liu Z, Qin J et al (2021) Partial is better than all: revisiting fine-tuning strategy for few-shot learning. In: AAAI, pp 9594–9602
138. Shi G, Chen J, Zhang W et al (2021) Overcoming catastrophic forgetting in incremental few-shot learning by finding flat minima. In: NeurIPS, pp 6747–6761
139. Shih KJ, Singh S, Hoiem D (2016) Where to look: focus regions for visual question answering. In: CVPR, pp 4613–4621
140. Shu J, Xu Z, Meng D (2018) Small sample learning in big data era. arXiv preprint [arXiv:1808.04572](https://arxiv.org/abs/1808.04572)
141. Simon C, Koniusz P, Nock R et al (2020) Adaptive subspaces for few-shot learning. In: CVPR, pp 4135–4144
142. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: ICLR
143. Singh R, Bharti V, Purohit V et al (2021) Metamed: few-shot medical image classification using gradient-based meta-learning. *Pattern Recognit* 120(108):111
144. Singh R, Bharti V, Purohit V et al (2021) Metamed: few-shot medical image classification using gradient-based meta-learning. *Pattern Recognit* 120(108):111
145. Snell J, Swersky K, Zemel RS (2017) Prototypical networks for few-shot learning. In: NeurIPS, pp 4077–4087
146. Song Y, Wang T, Mondal SK et al (2022) A comprehensive survey of few-shot learning: evolution, applications, challenges, and opportunities. arXiv preprint [arXiv:2205.06743](https://arxiv.org/abs/2205.06743)
147. Sun N, Yang P (2023) T2L: trans-transfer learning for few-shot fine-grained visual categorization with extended adaptation. *Knowl Based Syst* 264(110):329
148. Sun X, Xv H, Dong J et al (2021) Few-shot learning for domain-specific fine-grained image classification. *IEEE Trans Ind Electron* 68(4):3588–3598
149. Sung F, Yang Y, Zhang L et al (2018) Learning to compare: relation network for few-shot learning. In: CVPR, pp 1199–1208
150. Sun J, Lapuschkin S, Samek W et al (2020) Explanation-guided training for cross-domain few-shot classification. In: ICPR, pp 7609–7616
151. Sun B, Li B, Cai S et al (2021a) FSCE: few-shot object detection via contrastive proposal encoding. In: CVPR, pp 7352–7362
152. Sun Q, Liu Y, Chua T et al (2019) Meta-transfer learning for few-shot learning. In: CVPR, pp 403–412
153. Szegedy C, Liu W, Jia Y et al (2015) Going deeper with convolutions. In: CVPR, pp 1–9
154. Tai Y, Tan Y, Xiong S et al (2022) Few-shot transfer learning for sar image classification without extra sar samples. *IEEE J Sel Top Appl Earth Obs Remote Sens* 15:2240–2253
155. Tian Y, Wang Y, Krishnan D et al (2020) Rethinking few-shot image classification: A good embedding is all you need? In: ECCV, pp 266–282
156. Tokmakov P, Wang Y, Hebert M (2019) Learning compositional representations for few-shot recognition. In: ICCV, pp 6371–6380
157. Tran K, Sato H, Kubo M (2019) Memory augmented matching networks for few-shot learnings. *Int J Mach Learn Comput* 9(6)
158. Triantafyllou E, Zhu T, Dumoulin V et al (2020) Meta-dataset: a dataset of datasets for learning to learn from few examples. In: ICLR
159. Tseng H, Lee H, Huang J et al (2020) Cross-domain few-shot classification via learned feature-wise transformation. In: ICLR
160. Tsutsui S, Fu Y, Crandall DJ (2019) Meta-reinforced synthetic data for one-shot fine-grained visual recognition. In: NeurIPS, pp 3057–3066
161. Vinyals O, Blundell C, Lillicrap T et al (2016) Matching networks for one shot learning. In: NeurIPS, pp 3630–3638
162. Voulodimos A, Doulamis N, Doulamis AD et al (2018) Deep learning for computer vision: a brief review. *Comput Intell Neurosci* 2018(7068,349:13):7086,349:1
163. Wah C, Branson S, Welinder P et al (2011) The caltech-UCSD birds-200-2011 dataset
164. Wang J, Perez L et al (2017) The effectiveness of data augmentation in image classification using deep learning. *Convolutional Neural Netw Vis Recognit* 11(2017):1–8
165. Wang D, Cheng Y, Yu M et al (2019) A hybrid approach with optimization-based and metric-based meta-learner for few-shot learning. *Neurocomputing* 349:202–211
166. Wang S, Wang D, Kong D et al (2020) Few-shot rolling bearing fault diagnosis with metric-based meta learning. *Sensors* 20(22):6437
167. Wang Y, Yao Q, Kwok JT et al (2021) Generalizing from a few examples: a survey on few-shot learning. *ACM* 53(3):63:1–63:34
168. Wang R, Zhu F, Zhang X et al (2023) Training with scaled logits to alleviate class-level over-fitting in few-shot learning. *Neurocomputing* 522:142–151
169. Wang H, Deng Z (2021) Cross-domain few-shot classification via adversarial task augmentation. In: IJCAI, pp 1075–1081
170. Wang K, Liu X, Bagdanov A et al (2022) Incremental meta-learning via episodic replay distillation for few-shot image recognition. In: CVPR, pp 3728–3738
171. Wang X, Yu F, Wang R et al (2019b) Tafe-net: task-aware feature embeddings for low shot learning. In: CVPR, pp 1831–1840
172. Wang J, Zhai Y (2020) Prototypical siamese networks for few-shot learning. In: ICEIEC, pp 178–181
173. Wang J, Zhu Z, Li J et al (2018) Attention based siamese networks for few-shot learning. In: ICSESS, pp 551–554
174. Wei J, Huang C, Vosoughi S et al (2021) Few-shot text classification with triplet networks, data augmentation, and curriculum learning. In: NAACL-HLT, pp 5493–5500
175. Welinder P, Branson S, Mita T et al (2010) Caltech-UCSD birds 200
176. Wen J, Cao Y, Huang R (2018) Few-shot self reminder to overcome catastrophic forgetting. arXiv preprint [arXiv:1812.00543](https://arxiv.org/abs/1812.00543)
177. Wertheimer D, Tang L, Hariharan B (2021) Few-shot classification with feature map reconstruction networks. In: CVPR, pp 8012–8021

178. Widhianingsih TDA, Kang D (2022) Augmented domain agreement for adaptable meta-learner on few-shot classification. *Appl Intell* 52(7):7037–7053
179. Xian Y, Lorenz T, Schiele B et al (2018) Feature generating networks for zero-shot learning. In: *CVPR*, pp 5542–5551
180. Xian Y, Schiele B, Akata Z (2017) Zero-shot learning—the good, the bad and the ugly. In: *CVPR*, pp 3077–3086
181. Xie J, Long F, Lv J et al (2022) Joint distribution matters: deep brownian distance covariance for few-shot classification. In: *CVPR*, pp 7962–7971
182. Yang J, Guo X, Li Y et al (2022) A survey of few-shot learning in smart agriculture: developments, applications, and challenges. *Plant Methods* 18(1):1–12
183. Yang S, Liu L, Xu M (2021) Free lunch for few-shot learning: Distribution calibration. In: *ICLR*
184. Yang P, Ren S, Zhao Y et al (2022b) Calibrating cnns for few-shot meta learning. In: *WACV*, pp 408–417
185. Yang S, Xiao W, Zhang M et al (2022c) Image data augmentation for deep learning: a survey. *arXiv preprint arXiv:2204.08610*
186. Yap PC, Ritter H, Barber D (2021) Addressing catastrophic forgetting in few-shot problems. In: *ICML*, pp 11,909–11,919
187. Yazdanpanah M, Rahman AA, Chaudhary M et al (2022) Revisiting learnable affines for batch norm in few-shot transfer learning. In: *CVPR*, pp 9099–9108
188. Yoon SW, Seo J, Moon J (2019) Tapnet: neural network augmented with task-adaptive projection for few-shot learning. In: *ICML*, pp 7115–7123
189. Yu Z, Chen L, Cheng Z et al (2020) Transmatch: a transfer-learning scheme for semi-supervised few-shot learning. In: *CVPR*, pp 12,853–12,861
190. Yue Z, Zhang H, Sun Q et al (2020) Interventional few-shot learning. In: *NeurIPS*, pp 2734–2746
191. Yu Z, Herman G (2005) On the earth mover’s distance as a histogram similarity metric for image retrieval. In: *ICME*, pp 686–689
192. Yu J, Zhang L, Du S et al (2022) Pseudo-label generation and various data augmentation for semi-supervised hyperspectral object detection. In: *CVPR*, pp 304–311
193. Zhang Z, Sejdic E (2019) Radiological images and machine learning: trends, perspectives, and prospects. *Comput Biol Med* 108:354–370
194. Zhang P, Bai Y, Wang D et al (2021) Few-shot classification of aerial scene images via meta-learning. *Remote Sens* 13(1):108
195. Zhang J, Bui T, Yoon S et al (2021a) Few-shot intent detection via contrastive pre-training and fine-tuning. In: *EMNLP*, pp 1906–1912
196. Zhang C, Cai Y, Lin G et al (2020) Deepemd: Few-shot image classification with differentiable earth mover’s distance and structured classifiers. In: *CVPR*, pp 12,200–12,210
197. Zhang R, Che T, Ghahramani Z et al (2018) Metagan: an adversarial approach to few-shot learning. In: *NeurIPS*, pp 2371–2380
198. Zhang S, Zheng D, Hu X et al (2015) Bidirectional long short-term memory networks for relation classification. In: *PACLIC*
199. Zhao C, Chen F (2020) Unfairness discovery and prevention for few-shot regression. In: *ICKG*, pp 137–144
200. Zheng W, Tian X, Yang B et al (2022) A few shot classification methods based on multiscale relational networks. *Appl Sci* 12(8):4059
201. Zhu F, Ma Z, Li X et al (2019) Image-text dual neural network with decision strategy for small-sample image classification. *Neurocomputing* 328:182–188
202. Zhu P, Zhu Z, Wang Y et al (2022) Multi-granularity episodic contrastive learning for few-shot learning. *Pattern Recognit* 131(108):820
203. Zhuang F, Qi Z, Duan K et al (2021) A comprehensive survey on transfer learning. *Proc IEEE* 109(1):43–76
204. Zhu Y, Liu C, Jiang S (2020) Multi-attention meta learning for few-shot fine-grained image recognition. In: *IJCAI*, pp 1090–1096
205. Ziko IM, Dolz J, Granger E et al (2020) Laplacian regularized few-shot learning. In: *ICML*, pp 11,660–11,670
206. Zintgraf LM, Shiarlis K, Kurin V et al (2019) Fast context adaptation via meta-learning. In: *ICML*, pp 7693–7702

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.