Image classification based on few-shot learning algorithms: a review

Qiao Qi[1], Azlin Ahmad[1], Wang Ke[2]
[1]College of Computing, Informatics and Mathematics, Universiti Teknologi MARA (UiTM), Shah Alam, Malaysia
[2]School of Computer and Communication, Jiangsu Vocational College of Electronics and Information (JSEI), Huai an, China

Article Info

Article history:

Received Feb 1, 2024
Revised Apr 9, 2024
Accepted Apr 13, 2024

Keywords:

Data augmentation
Deep learning
Few-shot image classification
Machine learning
Meta-learning
Transfer learning

ABSTRACT

Image classification is a critical task in the field of computer vision, and its importance has significantly increased over the past few years. Machine learning and deep learning techniques have demonstrated immense potential in this field. However, traditional image classification models require a vast amount of training data, which can be challenging and expensive to obtain. To overcome this limitation, researchers are turning to few-shot learning, which aims to classify images with limited training samples. This paper presents a detailed analysis of the field of image classification using few-shot learning. First, it investigates the use of data augmentation, transfer learning, and meta-learning methods in this field. Then, it introduces several commonly used datasets and evaluation metrics in few-shot classification, compares several classical few-shot classification methods, and summarizes the experimental results obtained from public datasets. Finally, this paper analyzes the current challenges in few-shot image classification and suggests potential future directions.

Corresponding Author:

Azlin Ahmad
College of Computing, Informatics and Mathematics, Universiti Teknologi MARA (UiTM)
Shah Alam, Selangor, Malaysia
Email: azlin@fskm.uitm.edu.my

1. INTRODUCTION

Image is a general term that refers to various graphics and visual representations, and it serves as the most common carrier of information in human social activities. People receive a wealth of visual information through their eyes, and image data possesses rich colors and a vast amount of information, playing an indispensable role in daily learning and work. However, due to the swift progress of the latest information technology, data has exploded, and analyzing data quickly and effectively has become a challenging task [1]. Image classification algorithms play a crucial role in numerous applications in various fields, such as satellite image recognition and traffic system control. Over the past few decades, image classification has found extensive applications in object recognition, scene understanding, facial analysis, and other domains [2]. With the emergence and application of deep learning technology, the efficacy of image classification in various fields has significantly improved. Image classification is essentially pattern classification, which aims to classify images into different categories based on their features while minimizing classification errors. Features are crucial in distinguishing objects from one another. One of the research tasks in image classification is extracting image features. Image features are unique structures that can be identified within an image, highlighting the distinctive characteristics of a particular object and differentiating it from others. Image classification distinguishes images based on their unique features, including natural features such as brightness, edges, texture, and color that can be perceived intuitively [3]. In traditional image classification techniques, researchers mainly perform feature extraction on images, followed by feature combination and feature selection, and finally use discriminative classifiers. Traditional image classification algorithms traditionally rely on manual feature engineering to encode image features. Various feature extraction techniques have been proposed for traditional image classification, including histogram of gradients (HOG) [4], local binary pattern (LBP) [5], and scale-invariant feature transform (SIFT) [6].

Over the last few years, the field of image recognition has made significant strides with the help of sustained progress in artificial intelligence, the availability of large-scale datasets [7], and robust computing resources, particularly in deep neural network-based machine learning methods [8]. However, conventional machine learning approaches require a large dataset with annotations for model training [9], with the primary objective being to minimize errors on the verification set. In real-world scenarios, manual annotation of data requires a significant allocation of human and material resources and generally consumes substantial time. In many fields, it may be difficult to acquire significant quantities of data, such as medical images, remote sensing images, synthetic pore diameter radar images, and obtaining large amounts of data may be more challenging due to privacy, security, or moral concerns. When training data are limited, machine learning models used for training are prone to problems such as synthesis and generalization [10].

Few-shot learning addresses this issue by training models to learn from only a few examples, often as few as one or a few per class [11]. There are several problems in few-shot learning that urgently need to be solved. First is the problem of over-fitting. Because few-shot learning models have limited training data. They are vulnerable to overfitting, the model's performance is good on the training set, causing it to memorize the training examples, but it is poor on the test set. To address the issue of overfitting, regularization techniques such as weight decay and discarding can be used to prevent model memory training examples. Another aspect is the dataset bias problem. Few-shot learning models are typically trained on a limited number of examples, thus making them more prone to bias in comparison to traditional machine learning models that operate on abundant datasets. For example, if the training set is biased toward some types of examples, the model can learn to identify examples of these types well but may not perform well with other types of examples. To solve this problem, you can redesign data sets and ensure the diversity of data types.

The present work constitutes a comprehensive review article elucidating techniques pertinent to few-shot image classification. Moreover, various other studies have endeavored to explore and analyze strategies for classifying small samples from diverse perspectives, such as a survey of data augmentation-based method [12], a survey of transfer learning method [13], and a survey of meta-learning [14]. For the classification of lung images, it is recommended to refer to the classification method based on the convolutional neural network [15]. There is also a survey on few-shot learning methods for remote sensing images [16]. The main distinction of this article from the survey above lies in its comprehensive coverage of few-shot image classification algorithms, particularly addressing the challenges encountered in this domain. The paper delves further into the intricacies of different algorithmic approaches, such as those based on data augmentation, transfer learning, and meta-learning. Furthermore, it meticulously discusses the complexities of the datasets used in few-shot learning research, evaluates various evaluation metrics, and compares experimental results to gain a deeper insight into the performance of different algorithms. By thoroughly examining these aspects, this paper contributes to a more nuanced understanding of the current status and future research directions in few-shot image classification.

In this survey, we present several noteworthy contributions:

- Novel classification scheme: our work categorizes methods for small-sample image classification into three distinct classes, delineated by the strategies and techniques employed to tackle the problem. These classes encompass methodologies centered around data augmentation, transfer learning, and meta-learning techniques.
- Thorough review: we have meticulously examined methods pertinent to small-sample image classification. Within each classification, we offer intricate descriptions of representative methodologies and elucidations of their interrelations, comparative analyses, and overarching summaries.
- Abundant resources: throughout our investigation, we have curated many resources, including cutting-edge methodologies, benchmark datasets, open-source code repositories, and exemplary applications.

The following segments of the paper are structured in the following manner: section 2 introduces methods for few-shot image classification, outlining several typical approaches. In section 3 presents benchmark datasets for few-shot image classification and evaluations of various model algorithms. It also delineates three challenges encountered in few-shot image classification. Lastly, section 4 offers a conclusion to the paper.

2. METHOD

Currently, there are three main approaches used for few-shot classification tasks: methods based on data augmentation, methods based on transfer learning, and meta-learning techniques. Data augmentation generates additional training data by applying various transformations to the existing data. Transfer learning leverages knowledge learned from a source task to aid in learning a target task with limited data. Meta-learning trains a model to adapt to new tasks with few examples quickly.

2.1. Methods based on data augmentation

The scarcity of available data makes few-shot learning a difficult task, which makes it difficult to optimize neural networks. One of the most promising approaches to dealing with these challenges is to incorporate data augmentation or enhanced feature engineering techniques. The objective of data augmentation is to expand the training dataset by generating new samples, either unlabeled or synthetic. This can help increase sample volume and diversity [17]. Feature enhancement, on the other hand, involves adding new features to the feature space to increase sample diversity. There are generally three categories of data augmentation methods: label-free data-based enhancement, data-based enhancement, and feature-based enhancement.

2.1.1. Methods based on label-free data

Based on label-free data, the original few-shot dataset can be expanded using different methods. Machine learning involves categorizing techniques into three categories: supervised learning, direct push learning, and semi-supervised learning. A classifier is employed in supervised learning to predict labels for data that has no labels (label-free data), and these predicted labels are appended to the original dataset. Semi-supervised learning incorporates labeled and unlabeled data to improve the accuracy of classification models. Direct push learning generates synthetic samples by pushing existing samples toward the decision boundary, expanding the dataset with new samples that are different from the original data.

The task of unsupervised few-shot learning is both difficult and full of potential. It involves extracting knowledge from training data that lacks explicit or labeled information, thereby reducing the cost of collecting and annotating data. Researchers have proposed various techniques to tackle this issue. Ji *et al.* [18] proposed a method that utilizes cluster analysis to construct pseudo labels for data in different clusters while employing meta-training optimization models. An approach for unsupervised few-shot learning was presented by Qin and colleagues in their study [19], it utilizes separation and enhancement techniques to generate diverse and representative samples for training the model. The framework pays careful attention to the differences in the distribution of the few-shot learning tasks, which helps alleviate overfitting and enhance the flexibility and generalization performance of the model. Wang *et al.* [20] proposed a self-training method to enhance the data by obtaining pseudo-labels without labeling samples and selecting highly confident samples using a new measure of pseudo-labels.

Zhang and Mortazavi [21] presented a semi-supervised learning model utilizing model agnostic meta-learning (MAML) [22], which utilizes unlabeled data and label sample adjustment-related parameters to adapt the model. Wang and Hebert [23] introduced an unsupervised training stage to enable the top-level unit to learn from a vast amount of unannotated data, thereby separating it from the original specific category dataset. Ren *et al.* [24] developed a few-shot learning approach with a semi-supervised methodology that builds on the prototype network [25], by incorporating unannotated data into the training dataset to enhance the accuracy of the classification boundary and the overall effectiveness of the classification process. Transductive learning directly transfers knowledge gained during the learning stage to the given data, which allows the model to have better generalization capabilities by observing both training and test data.

The direct push-type learning framework is utilized in a transductive propagation network that addresses a particular problem as presented by Liu *et al.* [26]. The labels assigned to the labeled samples are extended or applied to the unlabeled samples, after the forward pass through the neural network, the cross-entropy function is utilized to determine the loss. Subsequently, the relevant parameters are updated accordingly based on the outcome of the loss calculation. Hou *et al.* [27] introduced a cross-attention network that was developed using the direct push-type learning approach. When sampling the characteristics, the target object area is highlighted to make the distinctive features more prominent. Furthermore, they proposed a conversion reasoning algorithm that uses the available unlabeled data in an iterative manner to address the issue of data scarcity and enhance the feature optimization for each category.

2.1.2. Methods based on synthetic data

Data-driven methods can augment training data by generating synthetic labeled data. Goodfellow *et al.* [28] introduced a new method that involves simultaneously training two models. The first model is trained to generate samples, while the second model is trained to accurately distinguish between real samples and generated samples. Mehrotra and Dukkipati [29] introduced a novel approach for generating synthetic data by using an adversarial residual network based on generative adversarial network (GAN) architecture. However, the generated samples are still not real data. Hariharan and Girshick [30] proposed a two-stage learning approach consisting of a pre-training stage and a few-shot learning phase. During the pre-training stage, a general model is learned on a large set of samples, and the few-shot learning phase fine-tunes the model on new categories.

To tackle the problem of few-shot image classification, Xian *et al.* [31] presented a new network architecture known as F-VAEGAN-D2, which leverages the strengths of two popular deep learning techniques, the variational autoencoder (VAE) and GAN. Chen *et al.* [32] introduced a strategy for identifying and choosing high-probability samples from the training dataset and using neighboring classifiers to find multiple images that are likely to be in the same class, forming an expanded support set. By integrating the original features and transformed features of the samples, data augmentation is achieved. Finally, the expanded samples are used to calculate the classification loss and optimize the weights to generate a sub-network.

2.1.3. Methods based on characteristic enhancement

The two methods mentioned above, label-free data and data-based methods, aim to increase the expansion of samples by leveraging additional data information. In addition, few-shot learning can use feature-based enhancement methods to increase the diversity of samples by enhancing their characteristics. Dixit *et al.* [33] suggested employing the attributed guided augmentation (AGA) model for producing a synthetic data map. This method facilitates the acquisition of synthetic data features, which are then mapped onto a designated space, the encoder and decoder functions are utilized to generate a wide variety of sample images.

The feature migration network (FATTEN) was introduced by Liu *et al.* [34]. This model comprises an encoder and decoder component that can track modifications in the motion trajectory triggered by alterations in an object’s pose. The role of the encoder is to map the characteristics of the target image as a pair of appearance and pose parameters. The role of the decoder is to convert these parameters into feature vectors. In their work, Chen *et al.* [35] introduced Trinet, a bidirectional network. Firstly, the semantic-image feature space mapping is facilitated by the label semantic space, which extracts multi-level depth features from the input image. Then, it maps the multi-level depth features to the semantic space. Feature extraction from images enhances the discriminative properties of the samples.

Jing *et al.* [36] proposed a feature enhancement network (FEN) for unconstrained palm pattern recognition in few-shot learning. It aims to eliminate the image variations caused by non-constraint collection and only enhance the characteristics of a few support samples. It is worth mentioning that data augmentation is not a panacea and should be used in conjunction with other techniques, such as regularization, feature engineering, and model selection. Additionally, the selection of augmentation techniques and the extent to which they are applied should be informed by the specific task at hand and the characteristics of the data.

2.2. Methods based on transfer learning

Transfer learning is a technique that involves utilizing knowledge and skills gained from previous learning experiences to improve the learning of new information [37]. First, a model is trained on a dataset with abundant data, and the model parameters are transferred to a new model to facilitate the training of the new model, thereby achieving knowledge transfer between different fields. Transfer learning methods are widely used in few-shot image classification research. Many techniques that utilize transfer learning rely on fine-tuning strategies for the model, updating its values, and obtaining a new classifier parameter [38]. Through ablation experiments, it has been demonstrated that the baseline method outperforms the initial SoftMax classifier. However, it should be noted that the choice of pre-trained model, transfer strategy, and fine-tuning approach can significantly impact the performance of the transfer learning method. Therefore, careful consideration should be given to these factors to achieve optimal results.

Dhillon *et al.* [39] proposed an improved Baseline method that is more effective. They added a loss function to the algorithm, which allows the model to learn more adaptable knowledge for the target data domain during the fine-tuning stage, even without labeled data. Meanwhile, Yu *et al.* [40] introduced a method for transfer learning that first utilizes a pre-trained feature extractor with data from the source domain. Then, the feature extractor is used to initialize the classification layer for the target domain category to further update the model. Fine-tuning the models that are trained on source-domain data enables the rebuilding of the classification layer, which facilitates the classification of target-domain data.

2.3. Methods based on meta-learning

Meta-learning is a concept that is also referred to as “learning to learn” [41], and has attracted significant interest due to its potential to enhance the efficiency of deep learning models. The primary objective of meta-learning is to train a model that can effectively adapt and generalize to novel and unseen tasks. Unlike traditional machine learning models that require a considerable amount of data for each new task, meta-learning models are designed as “fast learners” capable of quickly adapting to novel assignments with only a few examples. This is accomplished by training the model on a task distribution rather than a single task and by combining task-agnostic knowledge that can be easily adapted to new tasks. This section covers different meta-learning techniques that are used for few-shot image classification, meta-learning can be broadly classified into three categories: metric-based, optimization-based, and memory-based.

2.3.1. Metric-based

The metric-based meta-learning method is based on the concept of learning an embedded feature space using data from each K-shot category. This feature space allows the model to effectively measure the similarity between samples, where higher similarity is indicative of samples belonging to the same category. Samples are classified using non-parametric classification models. Van der Spoel *et al.* [42] utilized the Siamese-C3D network on two different datasets and achieved superior performance compared to other state-of-the-art methods. During the training phase, twin networks with the same convolutional neural network are trained to learn two similar input images. During the testing phase, the model inputs a query image to both twin networks, which return a condensed image representation of the query and each support image. The similarity between the query and each support image is calculated, and the query is classified based on the closest support image in the feature space.

Vinyals *et al.* [43] proposed a match network model that incorporates an attention mechanism and memory structure based on metric learning. This model maps a labeled support set and an unlabeled sample to a vector space, establishing a connection between the training embeddings and the test sample. In the training process, a matching training principle is used, allowing the model to quickly learn to match training embeddings with their corresponding categories. The match network model solves the problems existing in the Siamese-C3D network, as there is no need to input the test image and the training embeddings at the same time during the testing stage.

Jake *et al.* [44] suggested a network model prototype that entails allocating a central point to each category in a vector space, where it is exemplified as a prototype. By minimizing the loss function during model training, samples that belong to the same category are brought closer together in distance. The relation network (RN) model for end-to-end learning was introduced by Sung *et al.* [45] in their research study. It simplifies the structure of the measurement network based on the matching network, and its recognition performance is better than match network. This method learns a deep distance measurement, which uses the relationship between items in a dataset to predict the relationship scores between a test sample and a support set sample. These scores are then used to evaluate the similarity between different samples and classify the test sample.

2.3.2. Optimization-based

The optimization-based meta-learning approach stores priority knowledge in the parameters of the model. This approach can learn parameters that have strong generalization capabilities for network models as initialization parameters. To enhance the optimization of the model’s parameters, the training process and loss function should be designed accordingly. Mei *et al.* [46] proposed a bidirectional long short-term memory (Bi-LSTM) algorithm, which takes advantage of the characteristics of the LSTM structure and provides a better starting point than random parameters when training different data sets. The model continuously updates the values of the initialization parameters and the loss function obtained by predicting the sample to update the meta-learning device so that the meta-learning device parameter is optimized. Finally, the model predicts the results for the test data.

The meta-learner LSTM model has the capability to adapt swiftly to new tasks, even when presented with a limited number of samples. The MAML algorithm was introduced by Finn *et al.* [47] in their research paper, which is model-agnostic and can be used with other models trained by gradient descent. MAML’s main idea revolves around adjusting the model’s parameters to acquire the appropriate task-specific parameters. To address the limitations of MAML, Antoniou *et al.* [48] introduced the meta-learning adaptive model for few-shot learning (MAML++). MAML++ improves the flexibility of inner-layer optimization by learning internal optimization and direction and enhances the flexibility of outer-layer optimization through the strings of the meta-learning device. The application of MAML++, which is a novel model-agnostic meta-learning technique, has resulted in a remarkable enhancement in the performance of few-shot learning tasks.

2.3.3. Memory-based

The memory-based meta-learning approaches are designed to store image information in memory, which helps to make more efficient use of data. Faradonbeh and Safi-esfahani [49] proposed the neural turing machine (NTM), which uses additional memory storage modules to achieve the functions of machine learning systems. The NTM can store knowledge in the memory storage module and use this knowledge for prediction or classification. Santoro *et al.* [50] introduced a neural network model that utilizes memory augmentation to enhance its capabilities that include external memory modules to quickly store new data and use it for accurate prediction after training on a small number of samples. The MetNet method was presented by Munkhdalai and Yu in their paper [51], which is capable of learning across tasks. They divided the model into different learning modules to distinguish relevant knowledge and tasks learned by each module. The parameters and basic learning devices are then adjusted to adapt to new assignments. Jamal *et al.* [52] introduced a novel meta-learning algorithm called RN, which decouples representation learning from relationship learning to more effectively capture the similarities and differences between samples.

In addition, graph neural networks are favoured by some researchers. Garcia *et al.* [53] proposed to represent the relationships between categories by constructing graphs and utilizing graph neural networks (GNN) to learn these relationships. Then they use a me-ta-learner to infer the appropriate classifier, enabling it to quickly learn new tasks. Kipf and Welling [54] introduced the graph convolutional network (GCN) model for learning node representations by minimizing the classification loss function. They applied the GCN model to multiple graph datasets to solve the semi-supervised node classification task. Kim *et al.* [55] proposed the edge-labeling graph neural network (EGNN) model, which utilizes edge labels as additional input features to consider edge relationships during message passing, and interacts information between different layers to generate new node representations. The EGNN model achieves good performance on several datasets.

3. RESULTS AND DISCUSSION

Over the last few years, there has been a marked rise in interest in the field of few-shot image classification. This surge in attention coincides with the introduction of several datasets aimed at evaluating model performance across diverse scenarios and with various objects. These datasets serve as valuable benchmarks for assessing the effectiveness of algorithms in handling few-shot learning tasks.

3.1. Datasets

The purpose of few-shot image classification datasets is to evaluate the efficacy of few-shot learning algorithms in the domain of image classification through a pre-selected and organized collection of images. These datasets typically contain many image classes, but only a few examples per class, ranging from 1 to 20. The purpose of these datasets is to evaluate the generalization capacity of few-shot learning algorithms when confronted with novel image classes, which are provided with minimal training data. Over the last few years, a number of datasets have been put forward, each with its own challenges and characteristics. These datasets are utilized to assess the effectiveness of few-shot learning algorithms and to compare various approaches. In this section, we will investigate several datasets that are commonly used in applications of few-shot image classification. The basic information of datasets commonly used for few-shot learning is shown in Table 1.

Table 1. The basic information of datasets commonly used for few-shot learning

Dataset name	Image quantity	Category quantity	Sample quantity
MiniImageNet	60,000	100	600
Omniglot	32,460	1,623	20
CUB-200	11,788	120	171
TieredImageNet	779,165	608	1,282
CIFAR-100	60,000	100	600

- MiniImageNet: few-shot learning researchers frequently use the MiniImageNet dataset in their studies, making it a well-known and widely utilized dataset in this field [55]. It is a subset of the ImageNet dataset, a comprehensive labeled image dataset used in image recognition research. The MiniImageNet dataset is comprised of 100 categories, with each category having 600 full-color images that are 84×84 pixels in size. This dataset is widely used in few-shot learning research as it enables the evaluation of models that can be generalized to new classes with limited data. Few-shot learning refers to the scenario where a model is first trained on a limited set of base classes and then evaluated on a new set of novel classes that have very few examples available for each class. The MiniImageNet dataset is particularly useful in assessing few-shot learning models due to its large number of classes and relatively small number of examples per class, which makes it challenging for models to generalize. The high degree of variability within each class also adds to the dataset’s difficulty level. The ProtoNet and matching networks are some of the few-shot learning algorithms evaluated using this dataset.

- Omniglot: the Omniglot dataset is a collection of handwritten characters from 50 different writing systems [56], including Latin, Greek, Hebrew, Korean, Japanese, and others. Each alphabet has precisely 20 handwritten examples of each of its 1,623 characters, resulting in a total of 32,760 images. Each image is a black and white image of size 105×105 pixels, making it particularly challenging for models to generalize to new classes with limited data, a crucial challenge in few-shot learning. The Omniglot dataset has become a popular benchmark for evaluating machine learning models for few-shot learning due to its many alphabets and the similarity of many characters to each other. The dataset has led to significant advances in the field of few-shot learning, with various methods like convolutional neural networks (CNN) and recurrent neural networks (RNN) being utilized to tackle the problem. Meta-learning techniques are also employed to train these models.
- CUB-200: the CUB-200 dataset consists of 11,788 bird images that portray 200 distinct species [57], with around 60 images per category, and a size of 84*84 pixels. It was created by researchers at the California Institute of Technology, who collected images from various sources, including zoos, bird sanctuaries, and the internet. Due to the high similarity between many bird species, some images contain multiple birds or objects in the background, increasing the dataset's complexity. One common use of the CUB-200 dataset is for fine-grained image classification, such as the implementation of cross-attention networks.
- TieredImageNet: the TieredImageNet dataset is a recently introduced dataset designed to evaluate few-shot learning algorithms [58]. Within this dataset, there is a total of 779,165 images that have been categorized into 608 distinct classes, which are further divided into 34 high-level categories, such as animals and vehicles. Each class has 600 instances of 84×84 pixels size. The hierarchical structure and large number of classes make TieredImageNet a challenging dataset for few-shot learning algorithms, as it requires learning cross-modality mappings and generalization across different imaging modalities. TieredImageNet has been used to evaluate many few-shot learning algorithms, including MetaOptNet and TADAM.
- CIFAR-100: the CIFAR-100 dataset is a widely used image dataset in machine learning and computer vision [59]. With 60,000 colorful images that have been classified into 100 different categories, the CIFAR-100 dataset offers 600 images for each category. The dataset can be utilized for assessing convolutional neural network-based characteristic learning techniques, such as WA-CNN.

3.2. Few-shot learning evaluation indicators

Few-shot learning evaluation indicators are crucial for assessing the performance of models trained with limited annotated data, tasked with accurately classifying novel categories using only a few labeled samples. Typically, this evaluation is conducted in the N-way K-shot setting, where each task comprises a support set and a verification set containing N categories (way) and K samples in the support set for each category. The algorithm’s accuracy is determined by testing it on the verification set, repeating the evaluation multiple times, and computing the average accuracy. More precisely, the accuracy of classifying a few-shot method is determined by the proportion of accurately classified instances compared to the total dataset size.

In the context of few-shot learning, another vital evaluation indicator is meta-learning efficiency. Meta-learning efficiency assesses the model’s ability to generalize knowledge across different tasks and adapt quickly to new tasks with minimal training data. This is particularly relevant in scenarios where the model must rapidly learn new concepts or categories with limited labeled examples. Efficient meta-learning algorithms can leverage prior knowledge effectively and adapt to new tasks, improving performance on few-shot classification tasks.

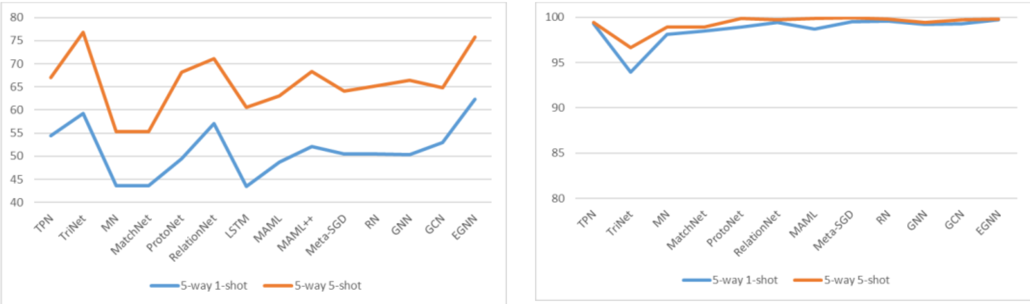
Furthermore, in addition to classification accuracy and meta-learning efficiency, it is essential to consider other performance metrics such as precision, recall, and F1-score. These metrics provide a more nuanced understanding of the model’s performance by considering false positives, false negatives, and the balance between precision and recall. By evaluating models using a combination of accuracy, meta-learning efficiency, and other relevant metrics, researchers can comprehensively assess their effectiveness in few-shot learning scenarios.

3.3. Comparison of experimental results

This section aims to evaluate the effectiveness of various few-shot image classification algorithms by selecting representative models on the MiniImageNet and Omniglot datasets [60]-[62], as they are frequently used in the field. The selected models are evaluated based on their classification accuracy in two different settings, namely 5-way 1-shot and 5-way 5-shot, which are widely acknowledged as standard measures for evaluating few-shot image classification techniques. We utilize 5-way 1-shot and 5-way 5-shot as baseline tasks. The results of our investigation are presented in Table 2 and Figure 1. Figure 1(a) displays classification accuracy for different methods on MiniImageNet, while Figure 1(b) shows accuracy on the Omniglot dataset.

Table 2. Few-shot image classification models are compared based on their classification accuracy

Method	MiniImageNet		Omniglot		Available code
	5-way 1-shot	5-way 5-shot	5-way 1-shot	5-way 5-shot	
TPN [26]	54.44	67.05	99.26	99.44	×
TriNet [35]	59.23	76.83	93.93	96.68	×
SiameseNet [42]	43.56	55.31	98.1	98.9	×
MatchNet [43]	43.56	55.31	98.51	98.92	√
ProtoNet [44]	49.42	68.20	98.90	99.85	√
RelationNet [45]	57.02	71.07	99.42	99.72	√
LSTM [46]	43.44	60.6	-	-	√
MAML [47]	48.70	63.11	98.7	99.9	√
MAML++ [48]	52.15	68.32	-	-	×
Meta-SGD [49]	50.47	64.03	99.53	99.93	×
RN [52]	50.44	65.32	99.6	99.8	×
GNN [53]	50.33	66.41	99.2	99.4	√
GCN [54]	53.03	64.78	99.26	99.72	√
EGNN [55]	62.34	75.77	99.75	99.77	√



(a) (b)

Figure 1. Classification accuracy on (a) MiniImageNet dataset and (b) Omniglot dataset

The results reveal notable variances in classification accuracy across different small-shot classification algorithms when applied to both the Omniglot and MiniImageNet datasets. This disparity can be attributed to the more straightforward image content present in the Omniglot dataset. Conversely, within the MiniImageNet dataset, a notable observation is made: the classification accuracy of the "5-shot" learning algorithm significantly surpasses that of the "1-shot" learning algorithm. This discrepancy stems from utilizing five instances per class in the 5-shot learning approach, enabling the algorithm to derive a more comprehensive and informative feature representation than the limited, single-instance representation offered by the 1-shot learning approach.

Table 2 shows that EGNN [55] and TriNet [35] achieved the highest accuracy in the 1-shot task and 5-shot task on the MiniImageNet dataset, respectively. TriNet conducts few-shot learning in high-dimensional concept space. It maps a sample instance to the concept space and generates new instances by interpolating between concepts to reduce learning difficulty. EGNN utilizes intra-cluster similarity and inter-cluster dissimilarity to update edge labels on the graph instead of node labels, thereby achieving explicit clustering evolution. It is also very suitable for inferring various new categories without the need for retraining. Concept learning and GNN play important roles in few-shot learning. As shown in Figure 1, the early works [42]-[44] use relatively simple network structures, which affect their accuracy. If their network structures are upgraded, their accuracy will be improved.

3.4. The challenges faced by few-shot image classification

Few-shot image classification is a challenging task that has seen proficient performance on datasets with relatively simple image content, such as the Omniglot character dataset. However, for more complex datasets, despite continuous improvements to algorithm models, classification accuracy remains unsatisfactory. This suggests that further research and innovation are necessary to address the complexities inherent in few-shot learning tasks involving intricate image content.

3.4.1. Explainability of deep learning

Due to the multitude of neural network structures and their significant parameters, researchers cannot always discern the specific role of each parameter. Therefore, improving model performance requires conducting multiple experiments continuously. Consequently, advancing research on deep learning mechanisms can assist researchers in enhancing their models and techniques.

3.4.2. Challenge of datasets

The MiniImageNet and Omniglot datasets are commonly used for few-shot image classification tasks, but they pose a challenge due to the unequal distribution of data across categories. With some categories having more training examples than others. As few-shot learning is becoming an increasingly crucial aspect of image classification, there is a growing need to develop datasets specifically for few-shot image classification, which is an active area of research in this field.

3.4.3. Universal few-shot learning method

Few-shot image classification methods present challenges due to their limited universality and high complexity. Developing classifiers that can adapt to different datasets is particularly difficult, as universal algorithms struggle with task and data variability. Additionally, few-shot classifiers require a deep understanding of image representations and relationships between different categories, and the limited availability of training data can further exacerbate these challenges. Ongoing research is focused on addressing these challenges and making few-shot image classification algorithms more robust and versatile. Efforts to develop new algorithms or enhance existing ones aim to improve their universality while maintaining high accuracy across different datasets and tasks.

4. CONCLUSION

In conclusion, this review thoroughly examines image classification techniques employing few-shot learning algorithms. The investigation delves into methodologies such as data augmentation, transfer learning, and meta-learning, elucidating their efficacy in addressing the challenges stemming from limited data availability in image classification tasks. Additionally, examining benchmark datasets and evaluation metrics sheds light on the performance of various model algorithms within few-shot learning contexts. Despite significant progress, persistent challenges include the interpretability of deep learning models, constraints inherent in existing datasets, and the pursuit of a universal few-shot learning approach. Advancing the field of few-shot image classification necessitates sustained research and development efforts to surmount these obstacles.

ACKNOWLEDGEMENTS

This research was partially supported by the Jiangsu Qing Lan Project under Grant (2022) 29, partially Jiangsu Province Industry-Academic-Research Cooperation Project (BY20231023), and partially supported by the Talent Project of Jiangsu Vocational College of Electronics and Information. We thank our colleagues from the Faculty of Computer and Mathematical Sciences of Universiti Teknologi MARA (UiTM), who provided insight and expertise that greatly assisted the research.

REFERENCES

[1] J. G. Lee and M. Kang, “Geospatial big data: challenges and opportunities,” *Big Data Research*, vol. 2, no. 2, pp. 74–81, Jun. 2015, doi: 10.1016/j.bdr.2015.01.003.

[2] Z. T. Abbood, M. N. H. Al-Turfi, and L. K. A. Almajmaie, “An abbreviated review of deep learning-based image classification models,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 30, no. 1, pp. 491–500, 2023, doi: 10.11591/ijeecs.v30.i1.pp491-500.

[3] D. Ciresan, U. Meier, and J. Schmidhuber, “Multi-column deep neural networks for image classification,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2012, pp. 3642–3649, doi: 10.1109/CVPR.2012.6248110.

[4] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, vol. 1, pp. 886–893, doi: 10.1109/CVPR.2005.177.

[5] T. Ahonen, A. Hadid, and M. Pietikäinen, “Face description with local binary patterns: application to face recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, 2006, doi: 10.1109/TPAMI.2006.244.

[6] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004, doi: 10.1023/B:VISI.0000029664.99615.94.

[7] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, “ImageNet: a large-scale hierarchical image database,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2010, pp. 248–255, doi: 10.1109/cvpr.2009.5206848.

[8] B. J. Khadhim, Q. K. Kadhim, W. K. Shams, S. T. Ahmed, and W. A. W. Alsiadi, “Diagnose COVID-19 by using hybrid CNN-RNN for chest X-ray,” *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 29, no. 2, pp. 852–860, Feb. 2023, doi: 10.11591/ijeecs.v29.i2.pp852-860.

[9] L. Yan, Y. Zheng, and J. Cao, “Few-shot learning for short text classification,” *Multimedia Tools and Applications*, vol. 77, no. 22, pp. 29799–29810, Nov. 2018, doi: 10.1007/s11042-018-5772-4.

[10] S. Rahman, S. Khan, and F. Porikli, “A unified approach for conventional zero-shot, generalized zero-shot, and few-shot learning,” *IEEE Transactions on Image Processing*, vol. 27, no. 11, pp. 5652–5667, Nov. 2018, doi: 10.1109/TIP.2018.2861573.

[11] A. Parnami and M. Lee, “Learning from few examples: a summary of approaches to few-shot learning,” *arXiv*, 2022, doi: 10.48550/arXiv.2203.04291.

[12] S. Y. Feng *et al.*, “A survey of data augmentation approaches for NLP,” *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pp. 968–988, 2021, doi: 10.18653/v1/2021.findings-acl.84.

[13] R. Ribani and M. Marengoni, “A survey of transfer learning for convolutional neural networks,” in *Proceedings - 32nd Conference on Graphics, Patterns and Images Tutorials, SIBGRAPI-T 2019*, 2019, doi: 10.1109/SIBGRAPI-T.2019.00010.

[14] X. Li, Z. Sun, J. H. Xue, and Z. Ma, "A concise review of recent few-shot meta-learning methods," *Neurocomputing*, vol. 456, pp. 463–468, Oct. 2021, doi: 10.1016/j.neucom.2020.05.114.
[15] L. N. Mahdy, A. I. B. El Seddawy, and K. A. Ezzat, "Automatic COVID-19 lung images classification system based on convolution neural network," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 5, pp. 5573–5579, Oct. 2022, doi: 10.11591/ijece.v12i5.pp5573-5579.
[16] X. Sun, B. Wang, Z. Wang, H. Li, H. Li, and K. Fu, "Research progress on few-shot learning for remote sensing image interpretation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 2387–2402, 2021, doi: 10.1109/JSTARS.2021.3052869.
[17] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "MixUp: beyond empirical risk minimization," *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, 2018, doi: 10.48550/arXiv.1710.09412.
[18] Z. Ji, X. Zou, T. Huang, and S. Wu, "Unsupervised few-shot learning via self-supervised training," *arXiv*, 2018, doi: 10.48550/arXiv.1912.12178.
[19] T. Qin, W. Li, Y. Shi, and Y. Gao, "Diversity helps: unsupervised few-shot learning via distribution shift-based data augmentation," *arXiv*, 2020, doi: 10.48550/arXiv.2004.05805.
[20] Y. Wang, C. Xu, C. Liu, L. Zhang, and Y. Fu, "Instance credibility inference for few-shot learning," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, doi: 10.1109/CVPR42600.2020.01285.
[21] L. Zhang and B. J. Mortazavi, "Semi-supervised meta-learning for multi-source heterogeneity in time-series data," *Proceedings of Machine Learning Research*, vol. 219, pp. 923–941, 2023.
[22] Y. Shao, W. Wu, X. You, C. Gao, and N. Sang, "Improving the generalization of MAML in few-shot classification via bi-level constraint," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 7, pp. 3284–3295, Jul. 2023, doi: 10.1109/TCSVT.2022.3232717.
[23] Y. X. Wang and M. Hebert, "Learning from small sample sets by combining unsupervised meta-training with CNNs," *Advances in Neural Information Processing Systems*, pp. 244–252, 2016, doi: 10.5555/3157096.3157124.
[24] M. Ren *et al.*, "Meta-learning for semi-supervised few-shot classification," *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, Mar. 2018, [Online]. Available: http://arxiv.org/abs/1803.00676.
[25] Z. Li, F. Ouyang, C. Zhou, Y. He, and L. Shen, "Few-shot relation classification research based on prototypical network and causal intervention," *IEEE Access*, vol. 10, pp. 36995–37002, 2022, doi: 10.1109/ACCESS.2022.3164688.
[26] Y. Liu *et al.*, "Learning to propagate labels: transductive propagation network for few-shot learning," *7th International Conference on Learning Representations, ICLR 2019*, May 2018, [Online]. Available: http://arxiv.org/abs/1805.10002.
[27] R. Hou, H. Chang, B. Ma, S. Shan, and X. Chen, "Cross attention network for few-shot classification," *Advances in Neural Information Processing Systems*, vol. 32, 2019.
[28] I. Goodfellow *et al.*, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020, doi: 10.1145/3422622.
[29] A. Mehrotra and A. Dukkipati, "Generative adversarial residual pairwise networks for one shot learning," *arXiv*, 2017, [Online]. Available: http://arxiv.org/abs/1703.08033.
[30] B. Hariharan and R. Girshick, "Low-shot visual recognition by shrinking and hallucinating features," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017, vol. 2017-Octob, pp. 3037–3046, doi: 10.1109/ICCV.2017.328.
[31] Y. Xian, S. Sharma, B. Schiele, and Z. Akata, "F-VAEGAN-D2: a feature generating framework for any-shot learning," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2019, vol. 2019-June, pp. 10267–10276, doi: 10.1109/CVPR.2019.01052.
[32] Z. Chen, Y. Fu, Y.-X. Wang, W. L. L. Ma, and M. Hebert, "Image deformation meta-networks for one-shot learning," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, doi: 10.48550/arXiv.1905.11641.
[33] M. Dixit, R. Kwitt, M. Niethammer, and N. Vasconcelos, "AGA: attribute guided augmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
[34] B. Liu, X. Wang, M. Dixit, R. Kwitt, and N. Vasconcelos, "Feature space transfer for data augmentation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9090–9098, doi: 10.1109/CVPR.2018.00947.
[35] Z. Chen, Y. Fu, Y. Zhang, Y. G. Jiang, X. Xue, and L. Sigal, "Multi-level semantic feature augmentation for one-shot learning," *IEEE Transactions on Image Processing*, vol. 28, no. 9, pp. 4594–4605, Sep. 2019, doi: 10.1109/TIP.2019.2910052.
[36] K. Jing, X. Zhang, Z. Yang, and B. Wen, "Feature augmentation learning for few-shot palmprint image recognition with unconstrained acquisition," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, May 2022, vol. 2022-May, pp. 3323–3327, doi: 10.1109/ICASSP43922.2022.9746999.
[37] A. S. Hatem, M. S. Altememe, and M. A. Fadhel, "Identifying corn leaves diseases by extensive use of transfer learning: a comparative study," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 29, no. 2, pp. 1030–1038, Feb. 2023, doi: 10.11591/ijeecs.v29.i2.pp1030-1038.
[38] W.-Y. Chen, Y.-C. Liu, Z. Kira, Y.-C. F. Wang, and J.-B. Huang, "A closer look at few-shot classification," *arXiv*, 2020.
[39] G. S. Dhillon, P. Chaudhari, A. Ravichandran, and S. Soatto, "A baseline for few-shot image classification," *8th International Conference on Learning Representations, ICLR 2020*, 2020.
[40] Z. Yu, L. Chen, Z. Cheng, and J. Luo, "Transmatch: a transfer-learning scheme for semi-supervised few-shot learning," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2020, pp. 12853–12861, doi: 10.1109/CVPR42600.2020.01287.
[41] S. Thrun and L. Pratt, "Learning to learn: introduction and overview," in *Learning to Learn*, Boston, MA: Springer US, 1998.
[42] E. Van der Spoel *et al.*, "Siamese neural networks for one-shot image recognition," *ICML - Deep Learning Workshop*, vol. 7, no. 11, pp. 956–963, 2015.
[43] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra, "Matching networks for one shot learning," *Advances in Neural Information Processing Systems*, pp. 3637–3645, 2016.
[44] S. Jake, S. Kevin, and Z. Richard, "Prototypical networks for few-shot learning," *Advances in Neural Information Processing Systems*, no. Nips, pp. 4077–4087, 2017.
[45] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. S. Torr, and T. M. Hospedales, "Learning to compare: relation network for few-shot learning," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 1199–1208, doi: 10.1109/CVPR.2018.00131.
[46] S. Mei, X. Li, X. Liu, H. Cai, and Q. Du, "Hyperspectral image classification using attention-based bidirectional long short-term memory network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, 2022, doi: 10.1109/TGRS.2021.3102034.

[47] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," *34th International Conference on Machine Learning, ICML 2017*, vol. 3, pp. 1856–1868, 2017.
[48] A. Antoniou, A. Storkey, and H. Edwards, "How to train your MAML," *7th International Conference on Learning Representations, ICLR 2019*, 2019.
[49] S. M. Faradonbeh and F. Safi-esfahani, "A review on neural turing machine," *arXiv*, pp. 1–27, 2014.
[50] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. Lillicrap, "One-shot learning with memory-augmented neural networks," *Global Biogeochemical Cycles*, vol. 29, no. 6, pp. 775–792, 2016.
[51] T. Munkhdalai and H. Yu, "Meta networks," *34th International Conference on Machine Learning, ICML 2017*, vol. 5, pp. 3933–3943, 2017.
[52] M. A. Jamal, G.-J. Qi, and M. Shah, "Task-agnostic meta-learning for few-shot learning," arXiv, 2018.
[53] V. Garcia and J. Bruna, "Few-shot learning with graph neural networks," *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, 2018.
[54] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, 2017.
[55] J. Kim, T. Kim, S. Kim, and C. D. Yoo, "Edge-labeling graph neural network for few-shot learning," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2019, pp. 11–20, doi: 10.1109/CVPR.2019.00010.
[56] L. Zhang, J. Liu, M. Luo, X. Chang, Q. Zheng, and A. G. Hauptmann, "Scheduled sampling for one-shot learning via matching network," *Pattern Recognition*, vol. 96, p. 106962, Dec. 2019, doi: 10.1016/j.patcog.2019.07.007.
[57] S. Branson *et al.*, "Visual recognition with humans in the loop," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6314 LNCS, no. PART 4, 2010, pp. 438–451.
[58] E. Bennequin, M. Tami, A. Toubhans, and C. Hudelot, "Few-shot image classification benchmarks are too far from reality: build back better with semantic task sampling," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2022, vol. 2022-June, pp. 4766–4775, doi: 10.1109/CVPRW56347.2022.00523.
[59] L. Bertinetto, P. H. S. Torr, J. Henriques, and A. Vedaldi, "Meta-learning with differentiable closed-form solvers," *7th International Conference on Learning Representations, ICLR 2019*, 2019.
[60] Y. Liu, H. Zhang, W. Zhang, G. Lu, Q. Tian, and N. Ling, "Few-shot image classification: current status and research trends," *Electronics*, vol. 11, no. 11, p. 1752, May 2022, doi: 10.3390/electronics11111752.
[61] C. Zhang, Y. Cai, G. Lin, and C. Shen, "DeepEMD: differentiable earth mover's distance for few-shot learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 5, 2022, doi: 10.1109/TPAMI.2022.3217373.
[62] B. N. Oreshkin, P. Rodriguez, and A. Lacoste, "Tadam: task dependent adaptive metric for improved few-shot learning," *Advances in Neural Information Processing Systems*, vol. 2018-December, pp. 721–731, 2018.

BIOGRAPHIES OF AUTHORS

Qiao Qi is pursuing a Ph.D. in Computer Science at Universiti Teknologi MARA. He received his master's degree from the Jiangnan University, China. His research interests include image classification and machine learning. He has won many awards, such as the first prize of the teaching achievement award, the first prize of the communications industry association teaching achievement award, the first prize of the micro course teaching competition in colleges and universities, the third prize of the national industrial and information technology skills competition. He can be contacted at email: qiaoqi@jsei.edu.cn.

Azlin Ahmad received her Ph.D. from Universiti Teknologi Malaysia (UTM) Kuala Lumpur in 2016. Currently, she is a senior lecturer at the Center for Information System Studies, FSKM, Universiti Teknologi MARA (UiTM). Her current research interests are in data analytics, machine learning artificial intelligence, and its applications. She can be contacted at email: excersiseuni@gmail.com.

Wang Ke received his Ph.D. degree in Electronic and Information Engineering from Nanjing University of Aeronautics and Astronautics. He is an Associate Professor at Jiangsu Vocational College of Electronics and Information. His research interests include SAR image detection, SAR image recognition, and machine learning. He can be contacted at email: 072112@jsei.edu.cn.