Soft Computing
https://doi.org/10.1007/s00500-023-07816-7

FOCUS

A novel method of data and feature enhancement for few-shot image classification

Yirui Wu[1,3] · Benze Wu[1] · Yunfei Zhang[1] · Shaohua Wan[2]

Accepted: 3 January 2023

Abstract
Deep learning has shown remarkable performance in quantity of vision tasks. However, its large network generally requires quantity of samples to support sufficient parameters learning during training process. Such high request greatly reduces efficiency when applying on a small dataset with few samples. To alleviate this problem, we propose a novel data enhancement method for few-shot learning via a cutout approach and feature enhancement. After enhancement, the generated network not only produces distinguish feature map without collecting more samples, but also achieves advantage of feature representation with high efficiency for computing. Specifically, cutout approach is simple yet highly effective for image regulation, which enhances input image matrix by adding a fixed mask to improve robustness and overall performance of network. Afterward, we perform feature enhancement by proposing a feature promotion module, which uses characteristics of dilated convolution and sequential processing to improve feature representation ability, thus improving efficiency of the whole network. We conduct comparative experiments on both miniImageNet and CUB datasets, where the proposed method is superior to comparative methods in both 1-shot and 5-shot cases.

Keywords Few−shot classification · Feature enhancement · Data enhancement · Feature promotion module

1 Introduction

In recent years, many deep learning methods have achieved significant results in computer vision domain. It is noted that these effective deep learning models largely rely on deep neural networks trained with thousands of labeled instances. However, these labels are time-consuming and annoying for manually collecting, thus in many cases not being sufficient for training. In the case of limited training data, most popular deep learning models will encounter the problem of overfitting. Essentially, it takes only a few samples for humans to understand a new concept, which inspires researchers to transfer knowledge from the known to the unknown. In the past few years, learning how to perform vision task with limited labeled examples, called few-shot learning (FSL), has attracted considerable attention, which has been studied in image classification, face recognition, action recognition, and so on.

In order to avoid overfitting, it is popular to apply regularization techniques in computer vision, such as data augmentation or the judicious addition of noise to activations, parameters, or data. Researchers generally name them as hallucination based methods in FSL, which directly deals with data deficiency by "learning to augment" (Chen et al. 2019). Their design patterns are defined as first learning a generator from labeled data and then using the learned generator to hallucinate new data for data augmentation. They aim to transfer knowledge or distribution variance or data feature from the labeled to new, thus aiding in generating

Communicated by Wei Wang.

✉ Shaohua Wan
shaohua.wan@uestc.edu.cn

Yirui Wu
wuyirui@hhu.edu.cn

Benze Wu
hhuwubenze@163.com

Yunfei Zhang
zhangyunfei@hhu.edu.cn

[1] College of Computer and Information, Hohai University, Focheng, Nanjing 211100, Jiangsu, China

[2] Key Laboratory of AI and Information Processing, Hechi University, Guangxi 546300, Yizhou, China

[3] Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Junxu, Changchun 130012, Jilin, China

Published online: 13 January 2023

more distinguished feature map for vision tasks. One complicated form of such generator is GAN models (Wu et al. 2018), which is capable to transfer semantical of images like drawing style from several paintings to generate a new one.

Meanwhile, training and testing sets may not belong to the same domain in real scenarios, which causes another famous problem in FSL, i.e., domain adaptation. For example, the learning network might deal with samples from testing set of CUB, while learning parameters on training dataset of mini-ImageNet. The inherent difference of data distributions or high-level semantical knowledge from both datasets would lead to low performance phenomenon known as overfitting problem, which could be even worse due to few samples. Therefore, domain adaptation techniques aim to reduce the domain shifts between source and target datasets. For example, Dong et al. (2018) addresses one-shot category domain adaptation problem, where both domain and category in the testing stage are verified to be different from those in training stage.

We start our work from the idea to combine power of data augmentation and domain adaption techniques. Essentially, we believe the former could alleviate data requirement of FSL methods by intuitively adding more samples, the latter offers enhancement on feature representation by digging semantical links between the training and testing samples. Specifically, we aim to apply a simple but effective regulation technique, i.e., cutout, rather than complicated GAN models, since GAN could significantly brings computation burden. The original motivation for cutout is object occlusion, where the generated occluded examples could provide sufficient information for model when encountering occlusions in the wild. Moreover, it enforces the model to take more context information into account, thus better solving the problem of occlusions with inferences. The initial form of cutout is maxdrop (Park and Kwak 2016), which would eliminate salient visual features, i.e., the maximally activated feature map, from the input of the image, thus encouraging the model to consider less prominent features for decision making. Later, Devries and Taylor (2017) find that randomly removing regions of a fixed size could achieve nearly the same performance as maxdrop, without the complicated preprocessing steps to locate salient feature map. In fact, cutout encourages the network to better utilize the full context of the image, rather than relying on the presence of a small set of specific visual features.

To alleviate difficulties of insufficient data and variances in appearance, we propose feature promotion module (FPM) to perform feature enhancement for image classification. Since some distinguishing features in one dataset may be invalid in another parakeet due to domain shift, the proposed FPM is designed to perform update and forget steps in FSL. More specifically, FPM is in favor of two major characteristics. First, we improve the capability of image classification by specific network structure design on the basis of the convolutional neural network, which extracts distinguish and informative features to solve the problem of large variances in appearances for samples between training and testing dataset. In fact, FPM extracts the relationship embeddings of channel vector sequence of feature map containing the general relationship of two samples from training and testing dataset. Second, FPM successfully integrates and fuses features from multiple feature maps, thus forgetting useless and enhancing contributive information to guide image classification task even facing domain shifting. Essentially, FPM provides a new way to offer deep feature representation for effective learning. Owing to the FPM-based enhancement method, a small number of labeled images can be involved into learning process to guarantee desirable classification results without bringing extra computation burden.

The main contributions of this paper are as follows:

- We propose a novel data enhancement method for few-shot learning, which involves cutout approach for data generation and feature promotion module for feature enhancement, thus solving difficulties of image classification with few labeled samples.
- A learnable feature enhancement module, named as feature promotion module, is proposed to improve feature representation ability for image classification by forgetting useless and enhancing contributive information.
- Experiments were carried out on two public datasets with several comparative methods, where the proposed method achieves the best classification accuracy.

The rest of this article is organized as follows. Section 2 gives an overview of related work. In Sect. 3, the details of the proposed structure are discussed, including the overall network architecture, designs of cutout and FPM module . Section 4 shows our experimental results with several comparison methods. Finally, Sect. 5 concludes the article.

2 Related work

The existing methods related to our work can be categorized into the following two types: few-shot classification methods and image augmentation.

2.1 Few-shot classification methods

Using prior knowledge, FSL can quickly generalize to perform new tasks that contain only a small amount of supervised samples. In order to overcome the problem of data efficiency, researchers have proposed quantity of methods, where they can be divided into three categories: initialization, metric learning, hallucination-based methods, and domain

adaptation. It is noted that we have offered illustration on the third category of method in Introduction.

The first category is implemented by initialization (Chen et al. 2019; Qiao et al. 2018). These methods first train classifiers with a large number of samples from large training set in training phase and then determine the parameters of classifiers with a small number of labeled samples in testing phase. In short, such methods tackle the few-shot learning problem by "learning to fine-tune." They try to learn good model initialization (i.e., the parameters of a network) so that the classifiers for new classes can be learned with a limited number of labeled examples and a small number of gradient update steps. Ravi and Larochelle (2017) propose the LSTM-based meta-learner to replace the stochastic gradient decent optimizer. Later, Munkhdalai and Yu (2017) propose the weight-update mechanism with an external memory to upgrade gradient optimizer as well. Afterward, Lee et al. (2019) thought that training linear classifier in the low shot mode can provide better generalization performance, and they have successfully learned the feature embedding which can be generalized under the new classification rules.

The second category is based on metric learning, which separates samples by measuring how close in feature representation they are with each other. In short, they address the few-shot classification problem by "learning to compare," where metric learning learns a sophisticated comparison models, performing classification task conditioned on distance or metric to few labeled instances during the training process. For example, Wei et al. (2020) propose a simple and interpretable general weighted framework to estimate the informativeness of heterogeneous features, which provides a tool to analyze interpretability of various loss functions. FEAT (Ye et al. 2020) is designed on the basis of Euclidean distance measures, where they use embedded mean values from the same category as prototypes of different categories. To perform incremental few-shot semantic segmentation, Shi et al. (2022) propose an embedding adaptive-update strategy to avoid catastrophic forgetting, where hyper-class embeddings remain fixed to maintain old knowledge, and category embeddings are adaptively updated with a class-attention scheme.

The fourth category, i.e., domain adaptation, is a particular case of transfer learning. Domain adaptation could reduce domain shift from source domain to target domain. Many methods have been proposed to address this problem, such as divergence-based, adversarial-based, and reconstruction-based domain adaptation. The former (Rozantsev et al. 2019) reduce divergence criterion between the training and testing data distributions. The middle (Zhu et al. 2017) minimize the gap between distributions by using adversarial training. The last (Yuan et al. 2022) consider domain shifting as an auxiliary reconstruction task by creating a shared representation for both domains.

2.2 Image augmentation

Image enhancement introduces prior knowledge for visual invariance, which has become the simplest and direct way to improve model performance in deep learning methods (Ni et al. 2022; Ding et al. 2022). Basically, researchers propose different data-enhanced networks by learning invariance or regularization rules for enhancement. The most common enhanced samples are strongly correlated with the original samples by different spatial operations, like cropping, flipping, rotating, scaling, warping, and other geometric transformations, as well as pixel perturbation, adding noise, lighting adjustment, contrast adjustment, sample addition or interpolation, segmentation patch wait. By involving enhanced samples for training, we enforce the network to learn inherent transformation representations for such operators, thus greatly improving performance by dealing with these transformations existed in the wild.

Although image enhancement algorithms have become focus of research, most of such algorithms cost too much time for computing. To emphasize effectively computing, several previous work has proposed specific operations for computation acceleration. For example, Farbman et al. (2011) introduce convolution pyramid operator to accelerate linear translation invariant filter. Similarly, many methods (Adams et al. 2010; Wang et al. 2021) have been proposed to accelerate bilateral filtering, which focuses on image processing application by performing edge sensing.

Another way to speed up computing is to first apply at low resolution and then up-sample the imaging results. However, simple up-sampling usually leads to undesirable blur output, which could be mitigated by designing special sampling techniques with respect to edge information of the original images. For example, Kopf et al. (2007) propose Joint bilateral up-sampling, which uses bilateral filters on high-resolution guidance maps to generate piecewise smooth edge-aware up-sampling. Bilateral spatial optimization (Barron and Poole 2016) solves a compact optimization problem in a two-sided grid, resulting in maximum smooth up-sampling results.

Recently, deep convolution network has made great progress in low-level vision and image processing tasks, such as depth estimation (Eigen et al. 2014), optical flow (Ilg et al. 2017), and image to image "translation" task (Isola et al. 2017). Some architectures have been trained to approximate a general class of operators. For example, Xu et al. (2015) develop a three-layer network in gradient domain to accelerate the edge sensing smoothing filter. Later, Liu et al. (2016) propose a learning recursive filter architecture for tasks of denoising, image smoothing, inpainting, and color interpolation. Recently, Zhang et al. (2022) propose local correlation ensemble (LCE) with graph convolution network based on attention features for cross-domain person re-identification

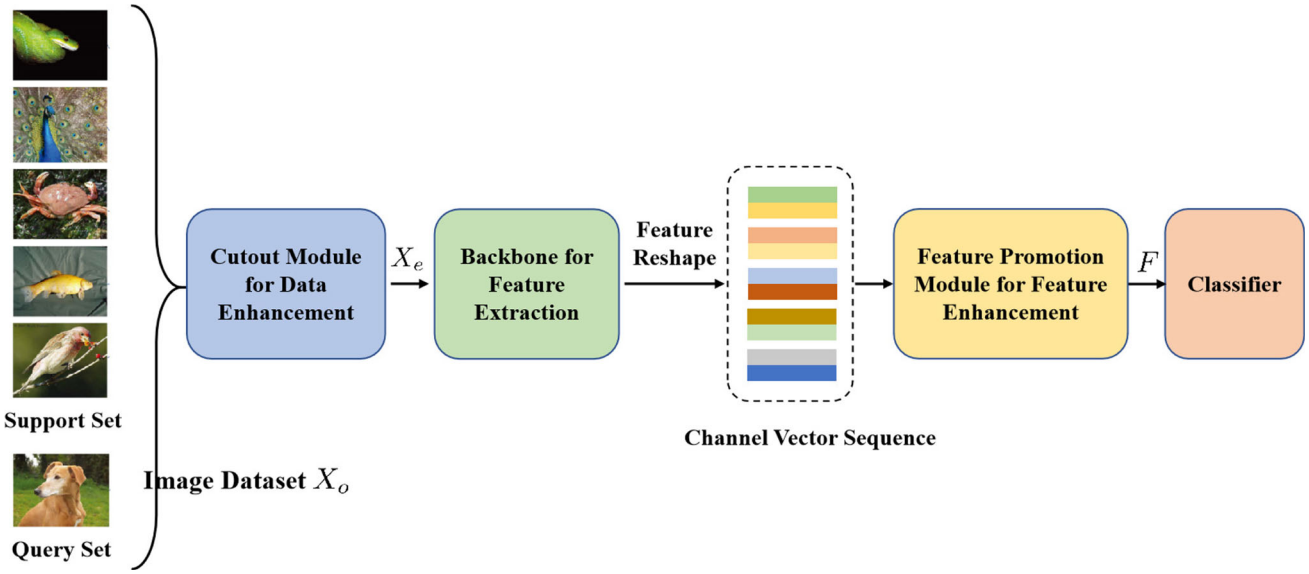

Fig. 1 Overall framework of the proposed method, which consists of cutout for data enhancement and feature promotion module for feature enhancement

and add a pedestrian attention module to obtain more robust person features.

3 The proposed method

3.1 Overall architecture

The overall framework is shown in Fig. 1, which is composed of cutout module, backbone for feature extraction, feature promotion module for feature enhancement, and classifier to assign category labels. It is noted that we perform enhancement from two aspects, where one is to consider the simple data enhancement by enlarging dataset to reduce the over fitting, and the other is to enhance the generalization ability of network through feature enhancement. After enhancing with two different modules from different aspects, the proposed method is capable to perform accurate image classification task with low computation cost, since the former module increases training time other than inference time, and the latter one is low in computation burden with only several layers.

Specifically, input training images set X_{o} are sent into cutout module for data enhancement, which adds a fixed region 0 mask to the random region of each image to generate new samples, and outputs the new set of training images X_{e}

$$X_{\mathrm{e}} = \mathrm{Cutout}(X_{\mathrm{o}}) \tag{1}$$

Then, the enhanced training image set are sent to the feature extraction module to extract feature maps. It is noted that we apply ResNet to encode sufficient information from images. After reshaping as channel feature vector sequences, the generated features are sent to promotion feature module for feature enhancement. These steps can be represented as:

$$F = \mathrm{FPM}(Reshape(\mathrm{Extractor}(X_{\mathrm{e}})) \tag{2}$$

More precisely, we transform each generated feature map into a one-dimensional feature vector in function $Reshape()$. Then, channel connector is used to sew feature vectors of query samples with the feature vectors of each class in the support set, thus obtaining channel feature vector sequences with size n. It is noted that n is the number of classes in the support set. The forgetting and updating module is composed of forgetting and updating blocks, which is used to extract the relational embedding of each channel vector sequence.

Finally, the classifier infers the category label Y of query samples based on enhanced feature map, which can be represented as

$$Y = \mathrm{Classifier}(F) \tag{3}$$

It is noted we use the loss of mean square error as loss function for training.

3.2 Design of cutout module for data enhancement

Due to the requirement of constructing large and complex representation space, most neural networks are often prone to be overfitting, thus requiring proper regularization to boost generalization. In this subsection, we show cutout for our pipeline, which is simple yet effective regularization technique that randomly offer input square regions zero masks during training. It is proved that cutout could significantly

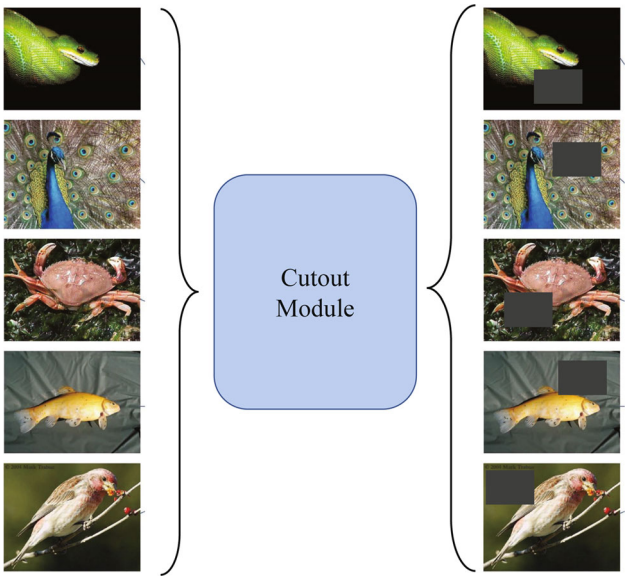

Fig. 2 Visualization results of Cutout on samples from training dataset

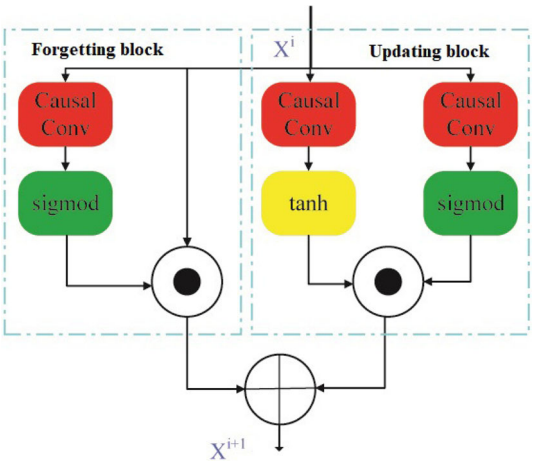

Fig. 3 Structure of Feature Promotion Module. The dashed boxes from left to right represent the forgetting block and updating block, respectively. $\odot$ and $\oplus$ indicates element-wise multiplication and addition, respectively

improve the robustness and overall performance of neural networks.

As shown in Fig. 2, the operations in cutout module randomly select a fixed region of the image, and then applies a total 0 mask to the region. It is noted that regions are allowed to be out of the range of the image. Essentially, cutout module is a simple neural network regularization technology, which removes continuous property of input images and effectively enhances data distribution by partially occluding samples. Specifically in the training process, pixel coordinates in the image are randomly selected as center points, where zero masks are placed around positions.

The proposed data enhancement module encourages the network to better utilize the full context of the image, rather than relying on the presence of a small set of specific visual features. In other words, cutout module allows neural network to use global information of the entire image, rather than the local information composed of some relatively low informative features. In fact, such idea is similar with the idea of most fine-grained papers, which emphasizes to receive some examples with large part being seen and some other being unseen during training. In this sense, cutout is closer to data enhancement than dropout, since it does not produce noise but generates novel views at images for network generalization.

3.3 Design of feature promotion module for feature enhancement

When categories of training and testing dataset are different, the effectiveness of neural network will be greatly reduced. In this subsection, feature promotion module is proposed to improve the discrimination ability in the domain transfer scenario. We design it to extract informative and salient feature for classification, rather than adopting all features for processing. In other words, the proposed PFM aims to alleviate the ineffective features in input images during training, which could be represented as procedures of learning, updating and forgetting. Essentially, PFM not only integrates feature vectors with multiple feature levels extracted from input images, but also builds more inherent feature representations by forgetting useless content and enhancing contribution information.

PFM is composed of serially connected multiple forget and enhance block (FEB), where we show the structure design of FEB in Fig. 3. FEB consists of forget block and enhance block, where the former forget block forgets category-level extrinsic features, and the latter enhance block strengthens the representation of contributive and intrinsic features. Improve the recognition rate by learning to retain and generate new features. The forgotten part calculates the retention rate of the feature through the forgetting block and then connects the partially retained feature with the newly extracted feature obtained through the update section. Afterward, the cascading result is used as input for the next "FEB" step. The cyclic process realizes the forgetting and updating of the channel vector sequence in the feature extraction process.

Causal dilated convolution is the basic operation of forget-update block, which is first applied as a special one-dimensional convolution in Wavenet (van den Oord et al. 2016), and can be implemented by shifting the output of a normal convolution by a few steps. For two-dimensional data, the equivalent of causal convolution is a masked convolution. When combine the casual convolution with dilated convolution, network can produce outputs of the same length as the inputs and can obtain features as data leakage free with few network layers, since dilated convolution can improve the range of receptive field on the channel vector sequence.

Specifically, the proposed forgetting block learns how to forget low-recognition features based on the context. The initial context is channel vector sequence, and all subsequent contexts are the output of the previous forget-update block. Forgetting block implements the forgetting mechanism by calculating the forgetting rate of the input sequence. The forgetting block generates data X_{f} of the same size as the input, which can be formalized as:

$$X_{\mathrm{f}} = Sigmod(Causal(X^i, d, k)) \odot X^i \tag{4}$$

where function $Causal()$ is causal dilated convolutional function, $Sigmod()$ is a sigmoid function, d is dilated rate, k is kernel size, X^{i+1} is the input to the ith forget-update block in PFM module, $\odot$ indicates element-wise multiplication.

The proposed updating block is designed to generates new features based on context. Specifically, the channel vector sequence e x is used as the initial context of the first forget-update block, and the rest of the contextual information is the output from the previous layer. Updating block generates data X_u with the same sequence length as the input, which can be formalized as:

$$X_u = tanh(\mathrm{Causal}(X^i, d, k)) \odot X^i \tag{5}$$

where $tanh()$ is tangent activation function. The whole process of FPM can thus be expressed as:

$$X^{i+1} = X_f \oplus X_u \tag{6}$$

where $\oplus$ is element-wise addition operation.

4 Experiments

In this section, we show the effectiveness and efficiency of the proposed enhancement method for image classification task. We first introduce dataset. Then, we describe implementation details for readers' convenience. Afterward, we conduct comparison and ablation experiments on two public datasets to demonstrate the effectiveness of the proposed method. Finally, we perform parameter setting experiment to analyze sensitivity of cutout module size. For fairness, experimental platform provided by Chen et al. (2019) is used for comparative experiments.

4.1 Datasets

In the experiment of this paper, we use two kinds of datasets, i.e., miniImageNet and CUB. MiniImageNet dataset is an excerpt from the ImageNet dataset, which is a famous large-scale visual dataset to promote visual recognition. ImageNet annotates more than 14 million images and provide at least 1 million images. ImageNet contains more than 20,000 categories, and each category of ImageNet has no less than 500 images. Since training so many images requires a lot of resources, Google DeepMind team extracted the miniImageNet dataset on the basis of ImageNet (Vinyals et al. 2016), which is used for small-size dataset learning research. MiniImageNet thus has become a benchmark dataset in the field of meta-learning. MiniImageNet contains a total of 60,000 color images in 100 categories, of which there are 600 samples in each category, and the size of each image is 84×84. In the experiments, the categories of training set and test set are divided into 80:20. Compared with CIFAR10 dataset, miniImageNet dataset is more complex, being more suitable for FSL experiments.

CUB dataset is a fine-grained dataset proposed by the California Institute of Technology, which is the current benchmark image dataset for fine-grained classification and recognition. It has 11788 bird images corresponding to 200 bird sub-categories, where training set has 5994 images and testing set has 5794 images. Each image provides image class labels, bounding boxes for bids, key parts information of birds, and attribute information of birds. Compared with miniImageNet, CUB dataset is used to test whether the proposed model has a convinced classification performance on fine-grained samples.

4.2 Implementation details

All methods are trained from the scratch. All input image are normalized for training and testing. Adam is used as an optimizer. The initial learning rate of the optimization algorithm is 0.001. When the test accuracy stagnates in seven consecutive training steps, the learning rate decreases by 10%. The most common FSL classification settings, 5-way 1-shot and 5-way 5-shot, have been tested on all datasets. Unless otherwise specified, all results were averaged over 1000 episodes from the test set with a 95% confidence interval.

4.3 Comparison with state of the arts

We conduct our experiments on the miniImageNet and CUB datasets with other comparative methods. The results are

Table 1 Accuracy for 5-way 1-shot and 5-way 5-shot classification on CUB dataset

Model	1-shot	5-shot
MAML Finn et al. (2017)	56.07±0.94	73.28±0.69
MatchingNet Vinyals et al. (2016)	60.51±0.89	72.88±0.67
ProtoNet Snell et al. (2017)	49.39±0.88	66.21±0.72
RelationNet Sung et al. (2018)	60.83±0.92	73.82±0.67
Baseline Chen et al. (2019)	31.95±0.58	52.90±0.67
Baseline++ Chen et al. (2019)	43.58±0.76	60.82±0.76
FEAT Ye et al. (2020)	52.43±0.92	66.85±0.76
ours	62.54±0.79	74.82±0.71

Table 2 Accuracy for 5-way 1-shot and 5-way 5-shot classification on miniImageNet dataset

Model	1-shot	5-shot
MAML Finn et al. (2017)	47.91±0.81	62.51±0.72
MatchingNet Vinyals et al. (2016)	50.53±0.83	63.77±0.67
ProtoNet Snell et al. (2017)	48.58±0.82	64.18±0.69
RelationNet Sung et al. (2018)	50.43±0.78	66.30±0.70
Baseline Chen et al. (2019)	36.43±0.61	55.41±0.66
Baseline++ Chen et al. (2019)	38.26±0.55	55.86±0.65
FEAT Ye et al. (2020)	46.11±0.74	62.76±0.67
ours	52.27±0.77	68.13±0.74

shown in Tables 1 and 2, respectively. It is noted that we follow (Chen et al. 2019) to choose the proper comparative studies to show the effectiveness of the proposed method, where Chen et al. (2019) is a recently published review paper with quantity of latest methods describing and analyzing.

Table 1 shows that our method is superior to the previous methods in single shot and five shot classification in CUB datasets. Through the Cutout data enhancement and forgetting and updating module, the classification performance is 1.71% higher than relationNet in 1-shot and 1.00% in 5-shot. Similarly, Table 2 shows our superior performance in miniImageNet datasets. Through the Cutout data enhancement and forgetting and updating module, the classification performance is 1.84% higher than relationNet in 1-shot and 1.83% in 5-shot.

Our model performs surprisingly on these two datasets. The reason is that the robustness of the model plays a huge role in the performance of few shot learning. The proposed data enhancement module and forgetting and updating module can both weaken the interference of irrelevant information. The former significantly reduce the impact of irrelevant context information such as background. The latter shields more useless information in the middle and high level features. Different resolution would not greatly affect the results of the proposed method. In fact, the proposed method focus on enhancement from multiple views. We think we can achieve progress in efficacy no matter what resolution images are. This analysis has been added in the revised version.

4.4 Ablation study

In order to verify that each module plays an active role in classification, we conduct two sets of ablation experiments. Cutout and feature promotion module are deleted in the two sets of experiments, and the results are shown in Tables 3 and 4, respectively.

In Table 3, we can find that the network performance after adding Cutout module is better. The reason it that the network fully utilize the complete context of the image instead of relying on the existence of a small set of specific visual features, thereby improving the accuracy of image classification. Similarly, Table 4 shows the contribution of the proposed FAU module. The network can make better use of the useful feature information in the image and discard the information that hinders the classification task, thereby improving the accuracy of image classification.

Combining Tables 3 and 4, we can find that the effect of the FAU module is better than that of the Cutout module. The reason is that in few shot learning, the display of context information and category information is not obviously related due to too few samples. It does not contribute as much as image category features.

4.5 Sensitivity analysis of cutout module size

In order to further study the influence of cutout on data enhancement, we do sensitivity experiments on the size of its mask. Experiments on 1-shot and 5-shot are carried out on CUB dataset and miniImageNet dataset, respectively, as shown in Fig. 4. We can see from Fig. 4 that when the size of 0 mask is 16 pixels, data enhancement effect is the best. And in the process of approaching 16 pixels, the increased income presents a diminishing benefit. When the size is greater than or less than 16 pixels, the effect is reduced. When the mask size is small, it is equivalent to noise and cannot enhance the data. When the mask size is too large, main features are blocked, and the influence network extracts the effective data.

5 Conclusion

We propose a data enhancement method for few-shot classification. It not only forgets the useless information in the support image and query image, but also enhances the effective information and category features. Moreover, the proposed data enhancement method uses the random zero mask to enhance the experimental data without increasing the number of samples, improving the accuracy of few shot

Table 3 Accuracy for 5-way 1-shot and 5-way 5-shot classification without Cutout module

Dataset	CUB		MiniImageNet	
Model	1-shot	5-shot	1-shot	5-shot
Without cutout	60.84±0.93	73.92±0.82	50.14±0.67	67.13±0.84
ours	62.54±0.79	74.82±0.71	52.27±0.77	68.13±0.74

Table 4 Accuracy for 5-way 1-shot and 5-way 5-shot classification without feature promotion module

Dataset	CUB		MiniImageNet	
Model	1-shot	5-shot	1-shot	5-shot
Without FPU	60.64±0.85	73.11±0.64	50.42±0.53	66.56±0.82
Ours	62.54±0.79	74.82±0.71	52.27±0.77	68.13±0.74

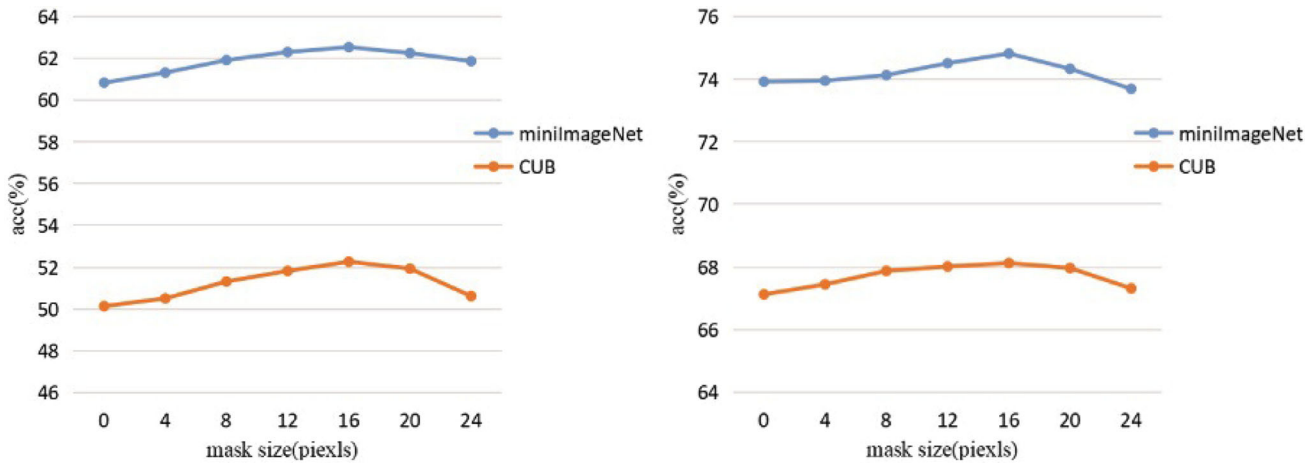

Fig. 4 Sensitivity analysis of mask size with 1-shot (left) and 5-shot tests (right)

classification from the perspective of feature enhancement and data enhancement. Experiments show that performance of the proposed method on public datasets is better than several of the latest methods. Cutout module can only be applied in only image processing domain. Meanwhile, feature promotion module is proper to be applied in other domains for feature map enhancement, which is potential to be adopted in multiple artificial intelligence domains. Considering the existing over-fitting problem, our future plan is to explore the idea of unsupervised learning to alleviate this concern. To make our model more generalized, we also plan to explore cross-domain image classification tasks.

Funding This work was supported in part by a grant from National Key R&D Program of China under Grant No. 2021YFB3900601, the Fundamental Research Funds for the Central Universities under Grant B220202074, Joint Fundation of the Ministry of Education (No.8091B022123), the Fundamental Research Funds for the Central Universities, JLU, and the Natural Science Foundation of China under Grant 61702160, Key Laboratory of AI and Information Processing (Hechi University), Education Department of Guangxi Zhuang Autonomous Region under Grant 2022GXZDSY014.

Data availability Enquiries about data availability should be directed to the authors.

Declarations

Conflict of interest The authors have not disclosed any competing interests.

References

Adams A, Baek J, Davis MA (2010) Fast high-dimensional filtering using the permutohedral lattice. Comput Graph Forum 29(2):753–762

Barron JT, Poole B (2016) The fast bilateral solver. In: Proceedings of European conference on computer vision, pp. 617–632

Chen W-Y, Liu Y-C, Kira Z, Wang Y-C, Huang J-B (2019) A closer look at few-shot classification. In: Proceedings of international conference on learning representations

Devries T, Taylor GW (2017) Improved regularization of convolutional neural networks with cutout. CoRR arXiv:1708.04552

Ding S, Wang H, Lu H, Nappi M, Wan S (2022) Two path gland segmentation algorithm of colon pathological image based on local semantic guidance. IEEE J Biomed Health Inform

Dong N, Xing EP (2018) Domain adaption in one-shot learning. Proc Eur Conf Mach Learn Knowledge Discovery Databases 11051:573–588

Eigen D, Puhrsch C, Fergus R (2014) Depth map prediction from a single image using a multi-scale deep network. In: Proceedings of neural information processing systems, pp. 2366–2374

Farbman Z, Fattal R, Lischinski D (2011) Convolution pyramids. ACM Trans Graph 30(6):175

Finn C, Abbeel P, Levine S (2017) Model-agnostic meta-learning for fast adaptation of deep networks. In: Proceedings of international conference on machine learning, pp. 1126–1135

Ilg E, Mayer N, Saikia T, Keuper M, Dosovitskiy A, Brox T (2017) Flownet 2.0: Evolution of optical flow estimation with deep networks. In: Proceedings of CVF/IEEE conference on computer vision and pattern recognition, pp. 2462–2470

Isola P, Zhu J-Y, Zhou T, Efros AA (2017) Image-to-image translation with conditional adversarial networks. In: Proceedings of CVF/IEEE conference on computer vision and pattern recognition, pp. 1125–1134

Kopf J, Cohen MF, Lischinski D, Uyttendaele M (2007) Joint bilateral upsampling. ACM Trans Graph 26(3):96

Lee K, Maji S, Ravichandran A, Soatto S (2019) Meta-learning with differentiable convex optimization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 10657–10665

Liu S, Pan J, Yang M-H (2016) Learning recursive filters for low-level vision via a hybrid neural network. In: Proceedings of European conference on computer vision, pp. 560–576

Munkhdalai T, Yu H (2017) Meta networks. Proc Int Conf Mach Learn 70:2554–2563

Ni B, Liu Z, Cai X, Nappi M, Wan S (2022) Segmentation of ultrasound image sequences by combing a novel deep siamese network with a deformable contour model. Neural Comput Appl, 1–15

Park S, Kwak N (2016) Analysis on the dropout effect in convolutional neural networks. Proc Asian Conf Comput Vision 10112:189–204

Qiao S, Liu C, Shen W, Yuille AL (2018) Few-shot image recognition by predicting parameters from activations. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp. 7229–7238

Ravi S, Larochelle H (2017) Optimization as a model for few-shot learning. In: Proceedings of international conference on learning representations, pp. 175–186

Rozantsev A, Salzmann M, Fua P (2019) Beyond sharing weights for deep domain adaptation. IEEE Trans Pattern Anal Mach Intell 41(4):801–814

Shi G, Wu Y, Liu J, Wan S, Wang W, Lu T (2022) Incremental few-shot semantic segmentation via embedding adaptive-update and hyper-class representation. In: Proceedings of ACM international conference on multimedia, pp. 5547–5556

Snell J, Swersky K, Zemel R (2017) Prototypical networks for few-shot learning. In: Proceedings of neural information processing systems, pp. 4080–4090

Sung F, Yang Y, Zhang L, Xiang T, Torr PH, Hospedales TM (2018) Learning to compare: relation network for few-shot learning. In: Proceedings of CVF/IEEE conference on computer vision and pattern recognition, pp. 1199–1208

van den Oord A, Dieleman S, Zen H, Simonyan K, Vinyals O, Graves A, Kalchbrenner N, Senior AW, Kavukcuoglu K (2016) Wavenet: a generative model for raw audio. In: Proceedings of the 9th ISCA speech synthesis workshop, p. 125

Vinyals O, Blundell C, Lillicrap T, Wierstra D et al (2016) Matching networks for one shot learning. Proc Neural Inf Proc Sys 29:3630–3638

Vinyals O, Blundell C, Lillicrap T, Kavukcuoglu K, Wierstra D (2016) Matching networks for one shot learning. In: Proceedings of neural information processing systems, pp. 3630–3638

Wang H, Zhang D, Ding S, Gao Z, Feng J, Wan S (2021) Rib segmentation algorithm for x-ray image based on unpaired sample augmentation and multi-scale network. Neural Comput Appl, 1–15

Wei J, Xu X, Yang Y, Ji Y, Wang Z, Shen HT (2020) Universal weighting metric learning for cross-modal matching. In: Proceedings of IEEE/CVF conference on computer vision and pattern recognition, pp. 13005–13014

Wu Y, Yue Y, Tan X, Wang W, Lu T (2018) End-to-end chromosome karyotyping with data augmentation using GAN. In: Proceedings of IEEE international conference on image processing, pp. 2456–2460

Xu L, Ren J, Yan Q, Liao R, Jia J (2015) Deep edge-aware filters. In: Proceedings of international conference on machine learning, pp. 1669–1678

Ye H-J, Hu H, Zhan D-C, Sha F (2020) Few-shot learning via embedding adaptation with set-to-set functions. In: Proceedings of IEEE/CVF conference on computer vision and pattern recognition, pp. 8808–8817

Yuan M, Cai C, Lu T, Wu Y, Xu Q, Zhou S (2022) A novel forget-update module for few-shot domain generalization. Pattern Recognit 129:108704

Zhang Y, Zhang F, Jin Y, Cen Y, Voronin V, Wan S (2022) Local correlation ensemble with gcn based on attention features for cross-domain person re-id. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)

Zhu J, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: proceedings of IEEE international conference on computer vision, pp. 2242–2251