

An Image Enhancement Method for Few-shot Classification

Yirui Wu

Hohai University

Benze Wu

Hohai University

Yunfei Zhang

Hohai University

Shaohua Wan (✉ shwanhust@gmail.com)

Zhongnan University of Economics and Law - Nanhu Campus: Zhongnan University of Economics and Law <https://orcid.org/0000-0001-7013-9081>

Research Article

Keywords: few-shot classification, feature enhancement, data enhancement, forgetting and updating

Posted Date: December 21st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1065912/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

An Image Enhancement Method for Few-shot Classification

Yirui Wu^{1,2}, Benze Wu^{1,2}, Yunfei Zhang^{1,2} and Shaohua Wan^{3*}

¹Key Laboratory of Water Big Data Technology of Ministry of Water Resources, Hohai University, Jiangning, Nanjing, 211100, Jiangsu, China.

²College of Computer and Information, Hohai University, Jiangning, Nanjing, 211100, Jiangsu, China.

³School of Information and Safety Engineering, Zhongnan University of Economics and Law, Hongshan, Wuhan, 430073, Hubei, China.

*Corresponding author(s). E-mail(s): shaohua.wan@ieee.org;
Contributing authors: wuyirui@hhu.edu.cn;
hhuwubenze@163.com; zhangyunfei@hhu.edu.cn;

Abstract

With the development of 5G/6G, IoT, and cloud systems, the amount of data generated, transmitted, and calculated is increasing, and fast and effective close-range image classification becomes more and more important. But many methods require a large number of samples to support in order to achieve sufficient functions. This allows the entire network to zoom in to meet a large number of effective feature extractions, which reduces the efficiency of small sample classification to a certain extent. In order to solve these problems, we propose an image enhancement method for the problems of few-shot classification. This method is an expanded convolutional network with data enhancement function. This network can not only meet the features required for image classification without increasing the number of samples, but also has the advantage of using a large number of effective features without sacrificing efficiency. structure. The cutout structure can enhance the matrix in the data image input process by adding a fixed area 0 mask. The structure of FAU uses dilated convolution and uses the characteristics

of the sequence to improve the efficiency of the network. We conduct a comparative experiment on the miniImageNet and CUB datasets, and the proposed method is superior to the comparative method in terms of effectiveness and efficiency measurement in the 1-shot and 5-shot cases.

Keywords: few-shot classification, feature enhancement, data enhancement, forgetting and updating

1 Introduction

In recent years, many deep learning methods have achieved good results in intelligent signal processing. These effective deep learning models largely rely on deep neural networks trained with thousands of tag instances. However, these tags are time-consuming and annoying, and in many cases there is not enough data to tag. In the case of limited training data, most popular deep learning models will encounter the problem of over fitting. In essence, 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 generalize new classes with limited labeled examples, called few-shot learning (FSL), has attracted considerable attention. And FSL has been studied in image classification, face recognition, action recognition and other fields.

Many FSL methods have been proposed to learn new classes based on limited samples. Some start from the limited sample itself, some start from the characteristics of the sample. But there are some problems, the former is simply to increase the number of samples to fit the existing popular network structure, the latter is to extract enough features as far as possible to meet the data needs of the model. All of these increase the burden of network to a certain extent.

In order to solve this problem, we propose a dilative convolutional neural network with data enhancement. This network can not only meet the needs of image classification features, but also has a structure of using a large number of effective features without sacrificing efficiency. In the process of image input, the clipping structure enhances the data by adding a fixed area 0 mask matrix, which achieves the effect of data enhancement without increasing the number of samples; The structure of FAU uses dilative convolution, which discards useless features and leaves useful features, so that the whole network can use a large number of effective features without adding extra burden.

The main contributions of this paper are as follows:

- We propose a feature enhancement method (FAU), which can not only forget useless information from support image and query image, but also enhance contribution information.
- In the case of not increasing the number of samples, the data is enhanced by Cutout.

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

2 Related Work

2.1 Few-shot Classification Methods

Machine learning is very successful in data-intensive applications, but it is often hindered when the data set is small. In order to solve this problem, few-shot learning (FSL) has been proposed in recent years. Using prior knowledge, FSL can quickly generalize to new tasks that contain only a small amount of supervised information samples.

Few-shot learning algorithms are designed to learn new classes with a limited number of labeled samples. In order to overcome the problem of data efficiency, people have made a lot of efforts. They can be divided into three categories: classifier learning, metric learning and initialization based methods.

The first few-shot learning method is based on classifier learning. These FSL methods are implemented by learning classifiers, such as [1][2][3]. These methods train feature extractors and classifiers with a large number of sampling tasks in the training set in the training phase, and then determine the parameters of feature extractors and classifiers with a small number of labeled samples in the test phase. Finally, the feature extractor and the improved classifier are used to predict the unlabeled samples.

The second few-shot learning method is based on metric learning. The purpose of these methods is to make the samples from the same category closer in the embedding space, while the samples from different categories are farther apart. For example, Wei et al. [4] proposed a simple and interpretable general weighted framework to estimate the informativeness of heterogeneous features, which provides a tool for analyzing the interpretability of various loss functions. Protonet [5] and FEAT [6] are based on Euclidean distance measures, and use embedded mean values from the same category as prototypes of that category. The intuition is that if a model can determine the similarity of two images, it can classify an unseen input image with the labeled instances Koch et al. [7].

The last few-shot learning method is based on initialization. This kind of method solves the learning problem of FSL by "learning fine tuning", and its purpose is to learn the appropriate model initialization or prediction network parameters. For example, Lee et al. [8] 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. Ravi et al. [9] proposed a new algorithm to replace the stochastic gradient descent optimizer.

2.2 Image Augmentation

Image enhancement in computer vision artificially introduces prior knowledge for visual invariance (semantic invariance). Data enhancement has basically become the simplest and direct way to improve model performance. The data-enhanced network learns some invariance or some regularization. The first enhanced sample is strongly correlated with the original sample (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). That is, we force the network to learn some sample transformation methods, and if these transformation methods improve the performance of the network, then it can be simply considered that the network has not learned the relevant transformations before, or the learning is not comprehensive.

Although image enhancement algorithm has become the focus of a lot of research, most of the complex algorithms cost too much time and work too much. Because of this, previous work has identified specific key operations and developed new algorithms to accelerate them. For example, farbman et al.[10] introduced convolution pyramid to accelerate linear translation invariant filter. Similarly, due to the universality of edge sensing image processing, many methods have been proposed to accelerate bilateral filtering such as [11][12][13][14][15][16][17][18]. And Gao et al.[19] improves the difficulty of sample clustering.

One way to speed up the operator is to simply apply it at low resolution and upsample the results. Simple up sampling usually leads to unacceptable blur output, but this problem can be improved by using more complex up sampling techniques (respecting the edge of the original image). Joint bilateral upsampling ,Kopf et al.[20] proposed, achieves this by using bilateral filters on high-resolution guidance maps to generate piecewise smooth edgware upsampling. Bilateral spatial optimization [21][22] established this idea by solving a compact optimization problem in a two-sided grid, resulting in maximum smooth upsampling results.

Neural network for image processing. Recently, deep convolution network has made great progress in low-level vision and image processing tasks, such as depth estimation eigen et al.[23], optical flow [24], super-resolution [25], and general image to image "translation" task [26]. Recent work has even explored learning deep networks in bilateral grids [27]. These methods focus on classification and semantic segmentation. Some architectures have been trained to approximate a general class of operators. Xu et al.[28] developed a three-layer network in gradient domain to accelerate the edge sensing smoothing filter. Liu et al.[29] proposed a learning recursive filter architecture for denoising, image smoothing, inpainting and color interpolation. Gao et al.[30] proposed four feature enhancement evaluation criteria. They jointly train the set of recurrent network and convolutional network to predict the propagation weight related to the image. These methods have achieved image enhancement to a certain

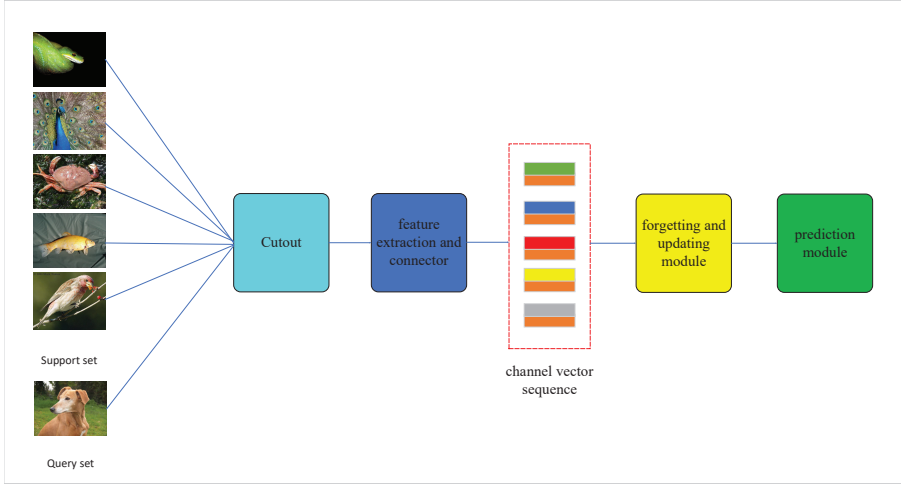


Fig. 1 The overall framework of the proposed method.

extent, which not only increases the generalization ability of the model, but also speeds up the processing ability of the network.

3 The Proposed Method

Because the few-shot learning problem is very suitable for the image enhancement, we solve the problem from the perspective of image enhancement. There are two directions in our proposed method. One is to consider the simple data enhancement from the sample to reduce the over fitting; The other is to enhance the generalization ability through feature enhancement.

3.1 Overall Architecture

The overall framework of the method is shown in Fig. 1, which is composed of data enhancement module, feature extraction and connector module, forgetting and updating module and prediction module.

Firstly, the data enhancement module(Cutout) adds a fixed region 0 mask to the random region of each image, and then inputs it to the feature extraction and connector module.

$$X = Cutout(X_0) \quad (1)$$

Then, feature extraction and connector transform it into a C -dimensional feature map, and then transform each feature map into a one-dimensional feature vector, then, channel connector is used to sew the feature vectors of query samples with the feature vectors of each class in the support set to obtain n channel vector sequences (n is the number of classes in the support set), Then these channel vector sequences(X) are put into the forgetting and updating module(FAU).

$$X_1 = FAU(X) \quad (2)$$

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. The prediction module infers the category of query samples from the relational embedding, and finally calculates the loss of mean square error and back propagation.

$$Y = FCN(X_1) \quad (3)$$

3.2 Design of data enhancement module

Convolutional neural networks can learn a powerful representation space, which is necessary for processing complex learning tasks. However, because of the model capacity required to capture such representations, they are often prone to overfitting, and therefore require proper regularization to generalize well. In this subsection, we show a simple regularization technique that randomly masks input square regions during training, we call it Cutout, which can be used to improve the robustness and overall performance of convolutional neural networks.

The data enhancement module is simple, but the effect is very good.[31] It randomly selects a fixed region of the image (some regions, even all regions, are allowed to be out of the range of the image), and applies a 0 mask to the region, as shown in Fig. 2. The data enhancement module is a simple convolutional neural network regularization technology, which removes the continuous part of the input image and effectively enhances the data network of partially occluded samples. In the specific operation, in the training process, when the image data is enhanced, the pixel coordinates in the image are randomly selected as the center point, and then a zero mask is placed around the position. This method allows all parts of the image not to contain a zero mask. The reason is that the model must receive some examples where most of the images can be seen during training. In this sense, Cutout is closer to data enhancement than dropout, because it does not produce noise, but begins to generate novel looking images on the network.

This 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. Simply put, CutOut allows CNN to use the global information of the entire image, rather than the local information composed of some small features. This idea is similar to the idea of most fine-grained papers.

3.3 Design of forgetting and updating module

Most of the latest FSL methods do not consider the differences in categories between training sets and testing sets. When the categories of training set and testing set are different, the effectiveness of these methods will be reduced. In this part, the forgetting update module is proposed to improve the discrimination ability in the domain transfer scenario. The forgetting update module is composed of stacked forgetting update blocks, and each forgetting update



Fig. 2 Visualization results of Cutout.

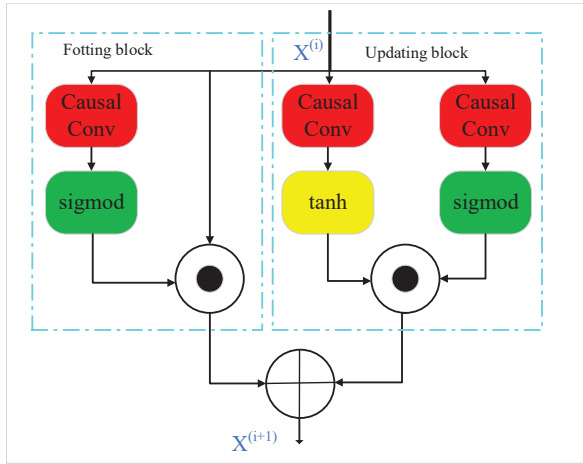


Fig. 3 Forgetting and Updating model. The Forgetting and Updating model includes two parts. The dashed boxes from left to right represent the forgetting block and updating block, respectively. \odot indicates element-wise multiplication.

block is composed of forgetting block and update block. Forgetting blocks learn forgetting rate according to context, and updating blocks generate new information according to context learning. Through training a large number of scenes, the forgetting update module learns how to forget the noise information that is not suitable for the context, and generates new information based on the context. The forgetting update module is used to extract the embedding relation from the channel vector sequence, so that the distribution of training set and test set is consistent. The details of forget update block are shown in Fig. 3.

Causal dilated convolution is the basis of forget-update block. Causal convolution is first applied as a special one-dimensional convolution in Wavenet[32], which 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, the network can produce outputs of the same length

as the inputs and can obtain features as data leakage free with few network layers. It can be formalized as in:

$$r = \text{Causal}(X, d, k) \quad (4)$$

Dilated convolution is adopted to improve the range of receptive field on the channel vector sequence.

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 $X_{\text{forgetting}}$ of the same size as the input, which can be formalized as in:

$$X_{\text{forgetting}} = \text{sigmoid}(\text{Causal}(X^{(i)}, d, k)) \odot X^{(i)} \quad (5)$$

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_{updating} with the same sequence length as the input, which can be formalized as in:

$$X_{\text{updating}} = \tanh(\text{Causal}(X^{(i)}, d, k)) \odot X^{(i)} \quad (6)$$

The whole process can be expressed:

$$X^{(i+1)} = X_{\text{forgetting}} \oplus X_{\text{updating}} \quad (7)$$

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

4 Experiments

In this section, the effectiveness of the proposed method is evaluated. For the fairness of the experiment, the unified experimental platform provided by reference [1] is used for comparative experiments.

4.1 Implementation Details

In the experiment of this paper, we use two kinds of data sets, one is miniImageNet, the other is CUB.

The miniImageNet dataset is an excerpt from the ImageNet dataset. ImageNet is a very famous large-scale visual data set, which was established to promote the research of visual recognition. Training the ImageNet dataset

requires a lot of computing resources. ImageNet annotated more than 14 million images and provided borders for at least 1 million images. ImageNet contains more than 20,000 categories, such as "balloon", "tire" and "dog". Each category of ImageNet has no less than 500 images. Training so many images requires a lot of resources, so in 2016, the google DeepMind team Oriol Vinyals and others extracted the miniImageNet dataset on the basis of ImageNet [33]. The DeepMind team used the miniImageNet dataset for small sample learning research for the first time. Since then, miniImageNet has become a benchmark dataset in the field of meta-learning and small samples. 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 experiment of this article, the categories of the training set and test set of this data set are divided into: 80:20, 80:20, and 80:20. Compared with the CIFAR10 data set, the miniImageNet data set is more complex, but it is more suitable for prototyping and experimental research.

The CUB data set is a fine-grained data set proposed by the California Institute of Technology in 2010, and it is also the current benchmark image data set for fine-grained classification and recognition research. The data set has 11788 bird images, including 200 bird sub-categories. The training data set has 5994 images and the test set has 5794 images. Each image provides image class tag information, and the bounding of birds in the image box, the key part information of the bird, and the attribute information of the bird. In the experiment of this article, the categories of the training set and test set of this data set are divided into: 80:20, 80:20, and 80:20. Compared with miniImagenet, the cub data set is used to test that the model has a better classification effect on fine-grained samples.

All methods are trained from the beginning. The input image is normalized. 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 data sets. Unless otherwise specified, all results were averaged over 1000 episodes from the test set with a 95% confidence interval.

4.2 Comparison with State-of-the-arts

Table 1 and table 2 show the 1-shot and 5-shot results on the miniImageNet and CUB datasets, respectively. From these two tables, we can see that our network is superior to the previous methods and realizes the most advanced new performance.

As shown in Table 1, 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.

As shown in Table 2, our method is superior to the previous methods in single shot and five shot classification 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.

Table 1 5-way 1-shot and 5-way 5-shot classification accuracies on CUB.

Model	1-shot	5-shot
MAML[34]	56.07±0.94	73.28±0.69
MatchingNet[35]	60.51±0.89	72.88±0.67
ProtoNet[5]	49.39±0.88	66.21±0.72
RelationNet[36]	60.83±0.92	73.82±0.67
Baseline[1]	31.95±0.58	52.90±0.67
Baseline++[1]	43.58±0.76	60.82±0.76
FEAT[6]	52.43±0.92	66.85±0.76
ours	62.54±0.79	74.82±0.71

Table 2 5-way 1-shot and 5-way 5-shot classification accuracies on miniImageNet.

Model	1-shot	5-shot
MAML[34]	47.91±0.81	62.51±0.72
MatchingNet[35]	50.53±0.83	63.77±0.67
ProtoNet[5]	48.58±0.82	64.18±0.69
RelationNet[36]	50.43±0.78	66.30±0.70
Baseline[1]	36.43±0.61	55.41±0.66
Baseline++[1]	38.26±0.55	55.86±0.65
FEAT[6]	46.11±0.74	62.76±0.67
ours	52.27±0.77	68.13±0.74

4.3 Ablation Study

In order to verify that each module plays an active role in classification, we conducted two sets of ablation experiments. The cutout module and the forgetting update module were deleted in the two sets of experiments, and the results are shown in Table 3 and Table 4, respectively. In Table 3, we can find that the network performance after the Cutout module is added is even better. The network can make better use of 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. In Table 4, we can also find that the network performance after adding the FAU module is even better. 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 Table 3 and Table 4, we can find that the effect of the FAU module is better than that of the Cutout module. The reason we believe is that, compared with context information, extracting image category features is more conducive to Few-shot classification.

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

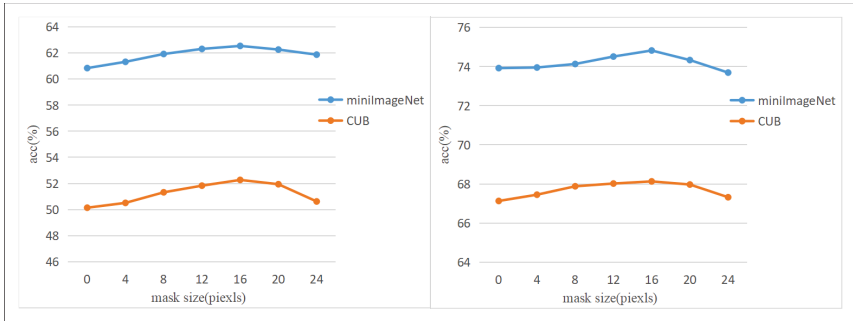
dataset	CUB		miniImageNet	
	1-shot	5-shot	1-shot	5-shot
-Cutout	60.84 \pm 0.93	73.92 \pm 0.82	50.14 \pm 0.67	67.13 \pm 0.84
ours	62.54 \pm 0.79	74.82 \pm 0.71	52.27 \pm 0.77	68.13 \pm 0.74

Table 4 5-way 1-shot and 5-way 5-shot classification without forgetting and updating module.

dataset	CUB		miniImageNet	
	1-shot	5-shot	1-shot	5-shot
Model				
-FAU	60.64 \pm 0.85	73.11 \pm 0.64	50.42 \pm 0.53	66.56 \pm 0.82
ours	62.54 \pm 0.79	74.82 \pm 0.71	52.27 \pm 0.77	68.13 \pm 0.74

4.4 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 data set and miniImageNet data set respectively, as shown in Fig. 4. We can see from the Fig. 4 that when the size of 0 mask is 16 pixels, the data enhancement effect is the best. When the size is greater than or less than 16 pixels, the effect is reduced. Our conjecture about this phenomenon is: when the mask size is small, it is equivalent to noise and can not enhance the data; When the mask size is too large, the main features are blocked, and the influence network extracts the effective data;

**Fig. 4** Sensitivity of mask size on 1-shot and 5-shot.

5 Conclusion

This paper proposes a data enhancement method for few-shot classification. This method can not only forget the useless information in the support image and query image, but also enhance the effective information, so as to achieve the enhancement of the category features. At the same time, we also use the random zero mask to enhance the experimental data without increasing the number of samples. The data enhancement method we proposed improves

the accuracy of small sample classification from the perspective of feature enhancement and data enhancement. Experiments show that the performance of this method on public data sets is better than several of the latest methods.

Our future plan is to explore the idea of unsupervised learning to alleviate concerns about over-fitting performance of classification. At the same time, we also plan to explore cross-domain image classification tasks, hoping to make our method more generalized.

Acknowledgments. This work was supported by National Key R&D Program of China under Grant 2018YFC0407901, National Nature Science Foundation of China under Grant No. 62172438, the Fundamental Research Funds for the Central Universities under Grant B200202177.

Compliance with Ethical Standards

Conflicts of interest: The authors declare that there is no conflict of interests regarding the publication of this article.

Ethical approval: This article does not contain any studies with animals performed by any of the authors.

Consent to participate: All authors participated in this study with their consent.

References

- [1] Chen, W.-Y., Liu, Y.-C., Kira, Z., Wang, Y.-C., Huang, J.-B.: A closer look at few-shot classification. In: International Conference on Learning Representations (2019)
- [2] Qiao, S., Liu, C., Shen, W., Yuille, A.L.: Few-shot image recognition by predicting parameters from activations. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7229–7238 (2018)
- [3] Guo, T., Yu, K., Aloqaily, M., Wan, S.: Constructing a prior-dependent graph for data clustering and dimension reduction in the edge of aiot. Future Generation Computer Systems (2021)
- [4] Wei, J., Xu, X., Yang, Y., Ji, Y., Wang, Z., Shen, H.T.: Universal weighting metric learning for cross-modal matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13005–13014 (2020)
- [5] Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 4080–4090 (2017)

- [6] Ye, H.-J., Hu, H., Zhan, D.-C., Sha, F.: Few-shot learning via embedding adaptation with set-to-set functions. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8808–8817 (2020)
- [7] Koch, G., Zemel, R., Salakhutdinov, R., *et al.*: Siamese neural networks for one-shot image recognition. In: *ICML Deep Learning Workshop*, vol. 2 (2015). Lille
- [8] Lee, K., Maji, S., Ravichandran, A., Soatto, S.: Meta-learning with differentiable convex optimization. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10657–10665 (2019)
- [9] Ravi, S., Larochelle, H.: Optimization as a model for few-shot learning. In: *5th International Conference on Learning Representations, ICLR 2017, April 24–26, 2017, Conference Track Proceedings*, pp. 175–186 (2017)
- [10] Farbman, Z., Fattal, R., Lischinski, D.: Convolution pyramids. *ACM Trans. Graph.* **30**(6), 175 (2011)
- [11] Adams, A., Baek, J., Davis, M.A.: Fast high-dimensional filtering using the permutohedral lattice. *Comput. Graph. Forum* **29**(2), 753–762 (2010)
- [12] Ding, F., Yu, K., Gu, Z., Li, X., Shi, Y.: Perceptual enhancement for autonomous vehicles: Restoring visually degraded images for context prediction via adversarial training. *IEEE Transactions on Intelligent Transportation Systems* (2021)
- [13] Wu, L., Quan, C., Li, C., Wang, Q., Zheng, B., Luo, X.: A context-aware user-item representation learning for item recommendation. *ACM Transactions on Information Systems (TOIS)* **37**(2), 1–29 (2019)
- [14] Zhao, L., Liu, Y., Al-Dubai, A.Y., Zomaya, A.Y., Min, G., Hawbani, A.: A novel generation-adversarial-network-based vehicle trajectory prediction method for intelligent vehicular networks. *IEEE Internet of Things Journal* **8**(3), 2066–2077 (2020)
- [15] Cheng, X., Guo, Z., Shen, Y., Yu, K., Gao, X.: Knowledge and data-driven hybrid system for modeling fuzzy wastewater treatment process. *Neural Computing and Applications*, 1–22 (2021)
- [16] Chen, J., Paris, S., Durand, F.: Real-time edge-aware image processing with the bilateral grid. *ACM Trans. Graph.* **26**(3), 103 (2007)
- [17] Paris, S., Durand, F.: A fast approximation of the bilateral filter using a signal processing approach. *Int. J. Comput. Vis.* **81**(1), 24–52 (2009)

- [18] Ding, S., Qu, S., Xi, Y., Wan, S.: Stimulus-driven and concept-driven analysis for image caption generation. *Neurocomputing* **398**, 520–530 (2020)
- [19] Gao, Z., Xue, H., Wan, S.: Multiple discrimination and pairwise CNN for view-based 3d object retrieval. *Neural Networks* **125**, 290–302 (2020)
- [20] Kopf, J., Cohen, M.F., Lischinski, D., Uyttendaele, M.: Joint bilateral upsampling. *ACM Trans. Graph.* **26**(3), 96 (2007)
- [21] Barron, J.T., Adams, A., Shih, Y., Hernández, C.: Fast bilateral-space stereo for synthetic defocus. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4466–4474 (2015)
- [22] Barron, J.T., Poole, B.: The fast bilateral solver. In: *European Conference on Computer Vision*, pp. 617–632 (2016). Springer
- [23] Eigen, D., Puhrsch, C., Fergus, R.: Depth map prediction from a single image using a multi-scale deep network. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (eds.) *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8–13 2014, Montreal, Quebec, Canada*, pp. 2366–2374 (2014)
- [24] Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., Brox, T.: Flownet 2.0: Evolution of optical flow estimation with deep networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2462–2470 (2017)
- [25] Chao, D., Chen, C.L., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. *ECCV* (2014)
- [26] Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1125–1134 (2017)
- [27] Jampani, V., Kiefel, M., Gehler, P.V.: Learning sparse high dimensional filters: Image filtering, dense crfs and bilateral neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4452–4461 (2016)
- [28] Xu, L., Ren, J., Yan, Q., Liao, R., Jia, J.: Deep edge-aware filters. In: *International Conference on Machine Learning*, pp. 1669–1678 (2015). PMLR
- [29] Liu, S., Pan, J., Yang, M.-H.: Learning recursive filters for low-level

- vision via a hybrid neural network. In: European Conference on Computer Vision, pp. 560–576 (2016). Springer
- [30] Gao, Z., Li, Y., Wan, S.: Exploring deep learning for view-based 3d model retrieval. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* **16**(1), 1–21 (2020)
 - [31] Wang, L., Zhen, H., Fang, X., Wan, S., Ding, W., Guo, Y.: A unified two-parallel-branch deep neural network for joint gland contour and segmentation learning. *Future Generation Computer Systems* **100**, 316–324 (2019)
 - [32] Oord, A.v.d., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., Kavukcuoglu, K.: Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499* (2016)
 - [33] Vinyals, O., Blundell, C., Lillicrap, T., Kavukcuoglu, K., Wierstra, D.: Matching networks for one shot learning. In: Lee, D.D., Sugiyama, M., von Luxburg, U., Guyon, I., Garnett, R. (eds.) *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016*, pp. 3630–3638 (2016)
 - [34] Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: *International Conference on Machine Learning*, pp. 1126–1135 (2017). PMLR
 - [35] Vinyals, O., Blundell, C., Lillicrap, T., Wierstra, D., *et al.*: Matching networks for one shot learning. *Advances in neural information processing systems* **29**, 3630–3638 (2016)
 - [36] Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P.H., Hospedales, T.M.: Learning to compare: Relation network for few-shot learning. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1199–1208 (2018)