

An Image Enhancement Method for Few-shot Classification

Benze Wu.

College of Computer and Information
Hohai University
Nanjing, China
hhuwubenze@163.com

Yirui Wu*.

College of Computer and Information
Hohai University
Nanjing, China
wuyirui@hhu.edu.cn

Shaohua Wan.

School of Information and Safety Engineering
Zhongnan University of Economics and Law
Wuhan, China
shaohua.wan@ieee.org

Abstract—In order to predict the unknown image categories, few-shot image classification has recently become a very hot field. However, many methods need a large number of samples to support in order to achieve enough functions. This makes the whole network de amplification to meet a large number of effective feature extraction, and reduces the efficiency of few-shot classification to a certain extent. To solve these problems, we propose a dilate convolutional network with data enhancement. This network can not only meet the necessary features of image classification without increasing the number of samples, but also has a structure that utilizes a large number of effective features without sacrificing efficiency. The cutout structure can enhance the data by adding a fixed area 0 mask matrix in the process of image input. The structure of FAU uses dilate convolution and uses the characteristics of a sequence to improve the efficiency of the network.

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

I. INTRODUCTION

Recently, many deep learning methods have achieved excellent results in the field of multimedia. These efficient deep learning models largely rely on deep neural networks trained with thousands of label instances. However, these labels are very time-consuming and annoying, and there is not enough data to tag in many scenarios. In the case of limited training data, most of these 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.

II. RELATEDWORK

A. Few-shot Classification Methods

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]. 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. [3] 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 [4] and FEAT [5] 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. [6].

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. [7] 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. [8] proposed a new algorithm to replace the stochastic gradient descent optimizer.

B. Image Augmentation

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. [9] 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 Adams et al. [10]; Chen et al. [11]; Durand et al. [12]; Ding et al. [13]. And Gao et al. [14] 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. [15] proposed, achieves this by using bilateral filters on high-resolution guidance maps to generate piecewise smooth edgewise upsampling. Bilateral spatial optimization [Barron et al. [16]; Barron and Poole [17]] 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. [18], optical flow [ILG et al. [19]], super-resolution [Dong et al. [20]], and general image to image "translation" task [Isola et al. [21]]. Recent work has even explored learning deep networks in bilateral grids [Jam-

pani et al. [22]]. These methods focus on classification and semantic segmentation. Some architectures have been trained to approximate a general class of operators. Xu et al. [23] developed a three-layer network in gradient domain to accelerate the edge sensing smoothing filter. Liu et al. [24] proposed a learning recursive filter architecture for denoising, image smoothing, inpainting and color interpolation. Gao et al. [25] 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 extent, which not only increases the generalization ability of the model, but also speeds up the processing ability of the network.

III. 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.

A. 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. 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). 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.

B. Design of data enhancement module

The data enhancement module is simple, but the effect is very good. [26] 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

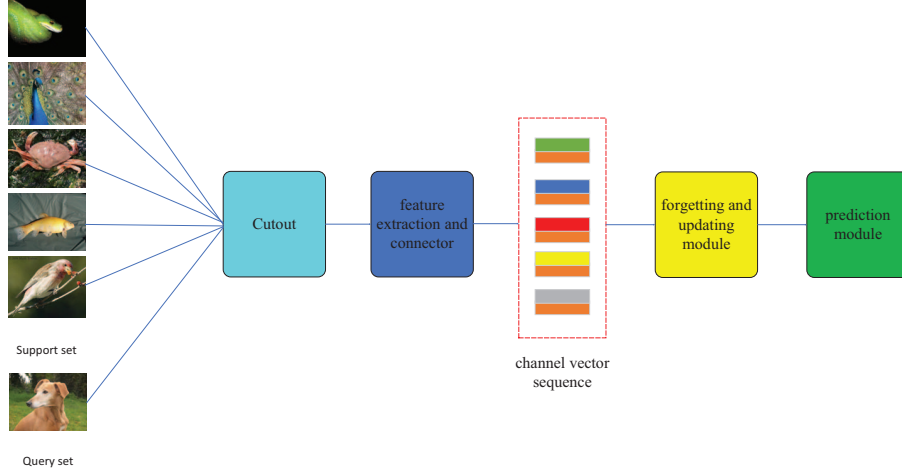


Figure 1. The overall framework of the proposed method.

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.

C. 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 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 [27], 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:

$$r = \text{Causal}(X, d, k) \quad (1)$$

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

$$X_{forgetting} = \text{sigmoid}(\text{Causal}(X^{(i)}, d, k)) \odot X^{(i)} \quad (2)$$

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:

$$X_{updating} = \text{tanh}(\text{Causal}(X^{(i)}, d, k)) \odot X^{(i)} \quad (3)$$

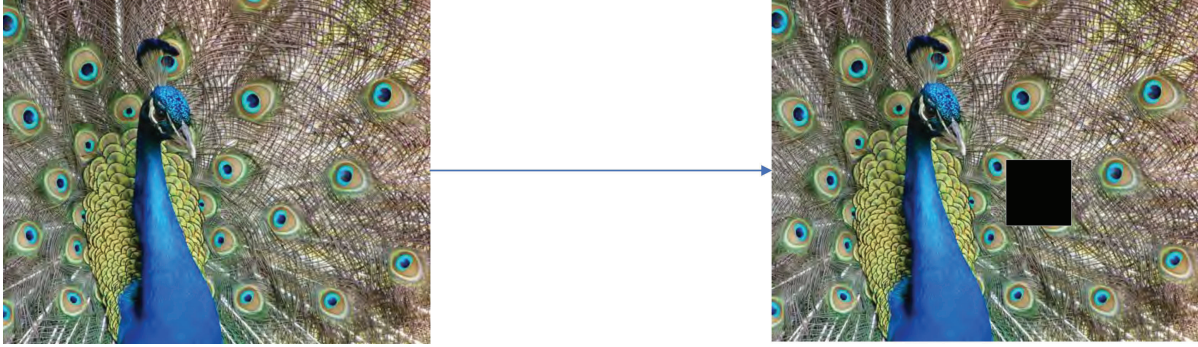


Figure 2. Visualization results of Cutout.

The whole process can be expressed:

$$X^{(i+1)} = X_{forgetting} \oplus X_{updating} \quad (4)$$

where $\text{Causal}(\cdot)$ is causal dilated convolutional function, $\text{sigmoid}(\cdot)$ 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, $\tanh(\cdot)$ is hyperbolic tangent activation function.

IV. 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.

A. Dataset

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

Among them, miniImageNet dataset is a subset of ImageNet, which is composed of 100 general object classes, and each class contains 600 images, and CUB dataset contains 11788 images of 200 bird species, which are usually used for fine-grained classification. There was no significant difference in the field of pups. According to common evaluation protocols, data sets are divided into 100, 50 and 50 categories for training, verification and testing.

B. Experimental Setups

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.

C. 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 [28]	56.07±0.94	73.28±0.69
MatchingNet [29]	60.51±0.89	72.88±0.67
ProtoNet [4]	49.39±0.88	66.21±0.72
RelationNet [30]	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 [5]	52.43±0.92	66.85±0.76
ours	62.54±0.79	74.82±0.71

D. Ablation Study

In order to verify that each module plays a positive role in classification, we conducted two groups of ablation experiments. The two groups of experiments deleted the cutout module and the forgetting and updating module

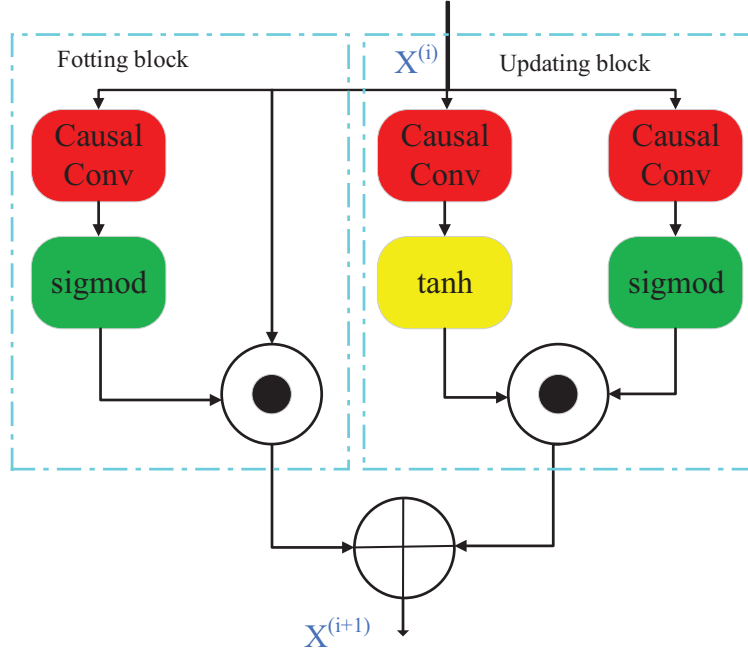


Figure 3. Forgetting and Updating model. The Forgetting and Updating model includes two parts. The dashed boxes from left to right represent the forgetting block and updating block, respectively. \odot indicates element-wise multiplication.

Table II

5-WAY 1-SHOT AND 5-WAY 5-SHOT CLASSIFICATION ACCURACIES ON MINIIMAGENET.

Model	1-shot	5-shot
MAML [28]	47.91 \pm 0.81	62.51 \pm 0.72
MatchingNet [29]	50.53 \pm 0.83	63.77 \pm 0.67
ProtoNet [4]	48.58 \pm 0.82	64.18 \pm 0.69
RelationNet [30]	50.43 \pm 0.78	66.30 \pm 0.70
Baseline [1]	36.43 \pm 0.61	55.41 \pm 0.66
Baseline++ [1]	38.26 \pm 0.55	55.86 \pm 0.65
FEAT [5]	46.11 \pm 0.74	62.76 \pm 0.67
ours	52.27 \pm 0.77	68.13 \pm 0.74

respectively, and the results are shown in Table 3 and Table 4 respectively. From the table, we can see that the two modules do have a positive effect for small sample classification and the latter has a higher contribution.

Table III

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

E. 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

Table IV

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
-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

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;

V. CONCLUSION

This paper proposes a feature enhancement method, which can not only forget useless information from support image and query image, but also enhance contribution information. In addition, we use random zero mask to enhance the data without increasing the number of samples.

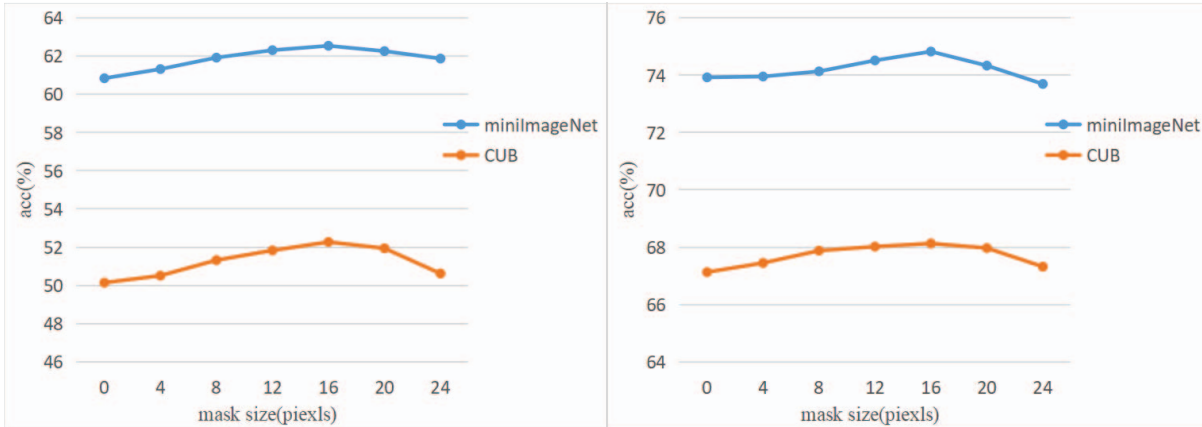


Figure 4. Sensitivity of mask size on 1-shot and 5-shot.

ACKNOWLEDGMENT

Throughout the writing of this dissertation I have received a great deal of support and assistance.

I would first like to thank my supervisor, Yirui Wu, whose expertise was invaluable in formulating the research questions and methodology. Your insightful feedback pushed me to sharpen my thinking and brought my work to a higher level.

I would also like to thank my tutors, Shaohua Wan, for their valuable guidance throughout my studies. You provided me with the tools that I needed to choose the right direction and successfully complete my dissertation.

REFERENCES

- [1] Wei-Yu Chen and Yen-Cheng Liu and Zsolt Kira and Yu-Chiang Frank Wang and Jia-Bin Huang, *A Closer Look at Few-shot Classification*. New Orleans, LA, USA: 7th International Conference on Learning Representations, 2019.
- [2] Siyuan Qiao and Chenxi Liu and Wei Shen and Alan L. Yuille, *Few-Shot Image Recognition by Predicting Parameters From Activations*. Salt Lake City, UT, USA: 2018 IEEE Conference on Computer Vision and Pattern Recognition, 2018.
- [3] Jiwei Wei and Xing Xu and Yang Yang and Yanli Ji and Zheng Wang and Heng Tao Shen, *Universal Weighting Metric Learning for Cross-Modal Matching*. Seattle, WA, USA: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.
- [4] Jake Snell and Kevin Swersky and Richard S. Zemel, *Prototypical Networks for Few-shot Learning*. Long Beach, CA: Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems, 2017.
- [5] Han-Jia Ye and Hexiang Hu and De-Chuan Zhan and Fei Sha, *Few-Shot Learning via Embedding Adaptation With Set-to-Set Functions*. Seattle, WA: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.
- [6] Koch, Gregory and Zemel, Richard and Salakhutdinov, Ruslan, *Siamese neural networks for one-shot image recognition*. Harlow, England: ICML deep learning workshop, 2015.
- [7] Lee, Kwonjoon and Maji, Subhansu and Ravichandran, Avinash and Soatto, Stefano, *Meta-learning with differentiable convex optimization*. Harlow, England: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.
- [8] achin Ravi and Hugo Larochelle, *Optimization as a Model for Few-Shot Learning*. Toulon, France: 5th International Conference on Learning Representations, 2017.
- [9] Farbman, Zeev and Fattal, Raanan and Lischinski, Dani, *Convolution pyramids*. Harlow, England: ACM Trans. Graph, 2011.
- [10] Adams, Andrew and Baek, Jongmin and Davis, Myers Abraham, *Fast high-dimensional filtering using the permutohedral lattice*. Harlow, England: Computer Graphics Forum, 2010.
- [11] Chen, Jiawen and Paris, Sylvain and Durand, Frédo, *Real-time edge-aware image processing with the bilateral grid*. Harlow, England: ACM Transactions on Graphics (TOG), 2007.
- [12] Paris, Sylvain and Durand, Frédo, *A fast approximation of the bilateral filter using a signal processing approach*. Harlow, England: European conference on computer vision, 2006.
- [13] Songtao Ding and Shiru Qu and Yuling Xi and Shaohua Wan, *Stimulus-driven and concept-driven analysis for image caption generation*. Salt Lake City, UT: Neurocomputing, 2020.
- [14] Zan Gao and Haixin Xue and Shaohua Wan, *Multiple Discrimination and Pairwise CNN for view-based 3D object retrieval*. Salt Lake City, UT: Neural Networks, 2020.
- [15] Kopf, Johannes and Cohen, Michael F and Lischinski, Dani and Uyttendaele, Matt, *Joint bilateral upsampling*. New York, USA: ACM Transactions on Graphics (ToG), 2007.

- [16] Barron, Jonathan T and Adams, Andrew and Shih, YiChang and Hernández, Carlos, *Fast bilateral-space stereo for synthetic defocus*. New York, USA: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.
- [17] Barron, Jonathan T and Poole, Ben, *The fast bilateral solver*. Harlow, England: European Conference on Computer Vision, 2016.
- [18] Eigen, David and Puhrsch, Christian and Fergus, Rob, *Depth map prediction from a single image using a multi-scale deep network*. Harlow, England: arXiv preprint arXiv:1406.2283, 2014.
- [19] Ig, Eddy and Mayer, Nikolaus and Saikia, Tonmoy and Keuper, Margret and Dosovitskiy, Alexey and Brox, Thomas, *FlowNet 2.0: Evolution of optical flow estimation with deep networks*. Harlow, England: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017.
- [20] Dong, Chao and Loy, Chen Change and He, Kaiming and Tang, Xiaoou, *Learning a deep convolutional network for image super-resolution*. Harlow, England: European conference on computer vision, 2014.
- [21] Isola, Phillip and Zhu, Jun-Yan and Zhou, Tinghui and Efros, Alexei A, *Image-to-image translation with conditional adversarial networks*. Harlow, England: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017.
- [22] Jampani, Varun and Kiefel, Martin and Gehler, Peter V, *Learning sparse high dimensional filters: Image filtering, dense crfs and bilateral neural networks*. Harlow, England: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.
- [23] Xu, Li and Ren, Jimmy and Yan, Qiong and Liao, Renjie and Jia, Jiaya, *Deep edge-aware filters*. Harlow, England: International Conference on Machine Learning, 2015.
- [24] Liu, Sifei and Pan, Jinshan and Yang, Ming-Hsuan, *Learning recursive filters for low-level vision via a hybrid neural network*. Harlow, England: European Conference on Computer Vision, 2016.
- [25] Zan Gao and Yinming Li and Shaohua Wan, *Exploring Deep Learning for View-Based 3D Model Retrieval*. Salt Lake City, UT: ACM Trans. Multim. Comput. Commun. Appl, 2020.
- [26] Linbo Wang and Hui Zhen and Xianyong Fang and Shaohua Wan and Weiping Ding and Yanwen Guo, *A unified two-parallel-branch deep neural network for joint gland contour and segmentation learning*. Salt Lake City, UT: Future Gener. Comput. Syst, 2019.
- [27] Aäron van den Oord and Sander Dieleman and Heiga Zen and Karen Simonyan and Oriol Vinyals and Alex Graves and Nal Kalchbrenner and Andrew W. Senior and Koray Kavukcuoglu, *WaveNet: A Generative Model for Raw Audio*. Harlow, England: arXiv preprint arXiv:1609.03499, 2016.
- [28] Chelsea Finn and Pieter Abbeel and Sergey Levine, *Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks*. Sydney, Australia: Proceedings of the 34th International Conference on Machine Learning, 2017.
- [29] Oriol Vinyals and Charles Blundell and Tim Lillicrap and Koray Kavukcuoglu and Daan Wierstra, *Matching Networks for One Shot Learning*. Harlow, England: Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, 2016.
- [30] Flood Sung and Yongxin Yang and Li Zhang and Tao Xiang and Philip H. S. Torr and Timothy M. Hospedales, *Learning to Compare: Relation Network for Few-Shot Learning*. Salt Lake City, UT: 2018 IEEE Conference on Computer Vision and Pattern Recognition, 2018.