Boosting Few-Shot Classification with Lie Group Contrastive Learning

Feihong He and Fanzhang Li(✉)

School of Computer Science and Technology, Soochow University, 215006 Suzhou, China
fhheloafei@stu.suda.edu.cn, lfzh@suda.edu.cn

Abstract. Few-shot learning can alleviate the issue of sample scarcity, however, there remains a certain degree of overfitting. There have been solutions for this problem by combining contrastive learning with few-shot learning. In previous works, sample pairs are usually constructed with traditional data augmentation. The fitting of traditional data augmentation methods to real sample distributions poses difficulties. In this paper, our method employs Lie group transformations for data augmentation, resulting in the model learning more discriminative feature representations. Otherwise, we consider the congruence between contrastive learning and few-shot learning with respect to classification objectives. We also incorporate an attention mechanism into the model. Utilizing the attention module obtained through contrastive learning, the performance of few-shot learning can be improved. Inspired by the loss function of contrastive learning, we incorporate a penalty term into the loss function for few-shot classification. This penalty term serves to regulate the similarity between classes and non-classes. We conduct experiments with two different feature extraction networks on the standard few-shot image classification benchmark datasets, namely miniImageNet and tieredImageNet. The experimental results show that the proposed method effectively improves the performance of the few-shot classification.

Keywords: Few-shot learning · Contrative learning · Lie group

1 Introduction

In recent years, deep neural networks perform satisfactorily with the support of large amounts of data. However, acquiring large amounts of labeled data requires too many human and financial resources. And, in many sample-sparse domains, obtaining enough samples for deep neural network training is impossible. Under such circumstances, deep learning often fails to demonstrate its full efficacy. As a result of these challenges, there has been significant interest in the field of few-shot learning [5,7,12,22,24,25].

Few-shot learning allows the model to adapt to a task with a very small number of labeled samples. Meta-learning [5,7,22,24,25] is a popular class of methods

L. Iliadis et al. (Eds.): ICANN 2023, LNCS 14254, pp. 99–111, 2023.
https://doi.org/10.1007/978-3-031-44207-0_9

used in few-shot learning. We usually divide meta-learning into two general directions: optimization-based [7] and metric-based [22]. Specifically, metric learning is used to classify samples by learning transferable feature extraction capabilities on the training set. It learns the feature representation capabilities specific to that task from a small number of samples during the testing phase and constructs a feature space to classify the samples by the metric. In meta-learning, feature extraction networks also suffer from overfitting problems due to sample sparsity. Unsupervised learning is proposed to address the problem of labeled sample scarcity. Contrastive learning is a class of methods for unsupervised learning. Networks trained by contrastive learning exhibit strong generalizations and are commonly used in diverse downstream tasks.

Inspired by the generalization capability of contrastive learning across diverse tasks, we propose a method that combines contrastive learning and meta-learning, aiming to endow meta-learning with enhanced generalization ability. Specifically, we divided the model training into two phases, the contrastive training phase and the meta-training phase. In the contrastive learning phase, we improve the data augmentation method for constructing sample pairs. Typically, traditional image augmentation, such as cropping, flipping, and color distortion, is commonly employed in contrastive learning. Recent works combining contrastive learning and few-shot learning have shown exceptional performance but have relied on traditional image augmentation methods. More powerful image augmentation can facilitate the creation of more diverse sample pairs. More diverse sample pairs enable the model to learn more discriminative expressions. We introduce Lie group transformations in the comparative learning stage to construct more diverse sample pairs. Specifically, we utilize the SO3 group, which conforms to the structure of Lie groups, to implement an image augmentation module. We refer to this module as the Lie transformation. Meanwhile, we incorporated an attention module in the contrastive learning phase. In meta-training phase, we will transfer the attention module trained in the contrastive learning phase. This transfer will enable the sample features to exhibit diverse expressive abilities in the channel dimension. Moreover, we formulate a penalty term based on contrastive learning in the meta-training phase. This penalty term implements inter-class constraints on samples by constructing positive and negative sample pairs based on the support set. The contributions of this paper are as follows:

- · Using the Lie group transformation method, we improve the image augmentation module in contrastive learning. By integrating it with meta-learning, we enhance the sample representation capability of meta-learning.
- · We introduce an attention module and add a penalty term to the meta-learning loss function to correct the deviation of prototype points in the sample space.
- · The result of our experiments on two popular few-shot classification benchmark datasets – miniImagenet and tieredImagenet, demonstrate that our algorithm outperforms state-of-the-art methods significantly on both 1-shot and 5-shot tasks.

2 Related Work

2.1 Few-Shot Learning

We can divide few-shot learning into two categories: initialization-based method and metric-based method. The main idea of initialization-based few-shot learning methods is to find an optimal set of initialization parameters for the model through training on different tasks. These initialization parameters can be trained with a small amount of data and quickly adapt to new tasks to achieve good results. Chelsea Finn et al. proposed a classic model [7] in 2017, pioneering the field of initialization-based few-shot learning methods. The main idea of metric-based few-shot learning methods is to acquire prior knowledge through training the model with a large number of tasks, map the samples to a reasonable space using the prior knowledge, and classify the samples using a predetermined metric method. Prototypical Networks [22], Matching Network [25], and Siamese Network [5] are classic models in metric learning. Many subsequent works are based on the idea of these models and have made improvements. The current metric-based few-shot learning shows excellent performance.

2.2 Contrativate Learning

The two mainstream methods of unsupervised learning currently are contrastive learning and masked image modeling [10,13]. Contrastive learning is an unsupervised learning method that learns representations by contrasting positive and negative data pairs. The goal of contrastive learning is to make the representations of positive pairs similar while making the representations of negative pairs dissimilar. Contrastive learning recently gains a lot of attention in deep learning due to its impressive performance in various computer vision tasks, such as image recognition and object detection. Inst Disc [27] pushes the class discrimination task to the extreme and proposes for the first time an instance discrimination method that achieves remarkable performance in the unsupervised domain. In the unsupervised domain, a large number of contrastive learning works [2,9] emerge and make rapid progress. In our work, we exploit the powerful generalization of contrastive learning to improve the performance of few-shot learning.

2.3 Lie Group Machine Learning

Recent years, Lie groups plays an important role in driving the development of machine learning research. In [28], Lie algebra is used to perform unsupervised augmentation of unlabeled samples and improve the performance of the model using an expanded dataset. In [29], the intrinsic mean of Lie groups is introduced to describe remote sensing images, which better reflects the commonalities of objects and the relationship between feature expressions, thereby achieving better results. In order to preserve shallow features and enhance local features, Lie groups are introduced in [30] to achieve satisfactory results. In our work, we also apply Lie groups to contrastive learning to improve the performance of few-shot learning.

3 Method

In this section, we introduce two parts in detail. In the first part, we introduce the improvement of contrastive learning through Lie group transformations in the contrastive learning phase. And the second part, we present the combination of meta-learning and contrastive learning, which is integrated with attention mechanisms and loss penalty terms.

Fig. 1. Original Image means the image that has not been augmented. Traditional is the image augmented with traditional cropping, flipping, and color transformation. Lie Mean is the image augmented with Lie transformation module and blank filled with image mean. Lie Original is the image augmented with Lie transformation module and blank filled with the original image.

We adopt a traditional few-shot learning setup to evaluate our method. In meta-learning, we usually divide samples into a training set $D_t = \{(x_i, y_i); i = 1 \cdots N_t\}$ and a validation set $D_v = \{(x_i, y_i); i = 1 \cdots N_v\}$ $(D_t \cap D_v = \emptyset)$. Following the N-way K-shot few-shot learning task setting, we draw N categories from the dataset, with $K+Q$ samples per category. Of these, $N \times K$ samples are used as the support set $D_s = \{(x_{i,j}, y_i); i = 1 \cdots N, j = 1 \cdots K\}$, with their category labels are visible to the model. Where $N \times Q$ samples are used as the query set $D_q = \{(x_{i,j}, y_i); i = 1 \cdots N, j = 1 \cdots Q\}$ and their category labels are not visible to the model.

3.1 Lie Contrative Learning

A Lie group is a mathematical object that simultaneously possesses a group structure and a smooth manifold structure. Firstly, we provide a formal definition for the structure of a Lie group. $(G, \bullet)$ is a group if it satisfies the following conditions:

1. $a \bullet b \in G, \forall a, b \in G$
2. $(a \bullet b) \bullet c = a \bullet (b \bullet c), \forall a, b, c \in G$

3. $\exists e \in G, \forall a \in G, e \bullet a = a \bullet e = a$
4. $\forall a \in G, \exists a^{-1}, a^{-1} \bullet a = a \bullet a^{-1} = e$

When a group structure satisfies the above conditions and it is also a differentiable manifold with the property that the group operations are compatible with the smooth structure, we call it a Lie group. It is commonly understood that matrix multiplication groups consisting of non-singular matrices can form Lie groups.

We define a new image augmentation operator as $r : R^3 \to R^3$. We demand that the operator satisfies the following conditions:

1. $\|r(v)\| = \sqrt{\langle r(v), r(v)\rangle} = \sqrt{\langle v, v\rangle} = \|v\|, \forall v \in R^3$
2. $\langle r(v), r(w)\rangle = \langle v, w\rangle = \|v\|\|w\| \cos\alpha, \forall v, w \in R^3$
3. $u \times v = w \longleftrightarrow r(u) \times r(v) = r(w)$

Based on the above properties, we can define:

$$SO(3) : \{r : R^3 \to R^3 \forall v, w \in R^3, \|r(v)\| = \|v\|, r(v) \times r(w) = r(v \times w)\}$$

Thus, we have obtained a transformation method, denoted by r, for an image in Euclidean space. Specifically, we can obtain a decomposed representation of the operator r by performing a decomposition on it:

$$r = R_x(\alpha) \bullet R_y(\beta) \bullet R_z(\gamma)$$

By decomposing its expression, we can construct a specific operator r based on three parameters α, β and γ:

$$R_x(\alpha) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix} R_y(\beta) = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} R_z(\gamma) = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

All possible operators that exist in r form a group structure known as SO3. For the sake of brevity in our exposition, we shall denote this process as $r(x)$. In Fig. 1, we compare the commonly used augmentation methods in contrastive learning and our two augmentation methods.

In contrastive learning phase, we put the samples in the training set through two traditional data augmentations and the random operator r to obtain the augmented samples $\{(r_l(x_i), r_r(x_i)); x_i \in D_t, i = 1 \cdots N_t\}$ after two different data augmentation methods. We treat two augmentations from the same sample as positive pairs, and one of the augmentations with two augmentations from the other sample as negative pairs. We expect more similarity between positive sample pairs and more variability between negative pairs, and have following loss function:

$$L = -\log \frac{\exp(r_l(x_i) \cdot r_r(x_i)/T)}{\sum_{i \neq j} \exp(r(x_i) \cdot r(x_j)/T)}$$

3.2 Attention and Penalty Items

It can be readily comprehended that the loss has a similar geometric meaning as the prototype network loss. In Fig. 2, it is evident that in contrastive learning, the positive sample pairs exhibit a closer distance in the corresponding metric space, whereas the negative sample pairs are farther apart. In prototypical networks, instances of the same class exhibit clustering, while instances of different classes demonstrate dispersion. Due to similar optimization objectives for the loss function, we can enhance the expressive ability of feature channels in meta-learning by training an attention module during the contrastive learning phase. This attention module assigns distinct weights to the embeddings of sample features in different channels. In the meta-learning phase, we transfer this attention module to the meta-learning model to improve the channel-wise representation capability of features in meta-learning.

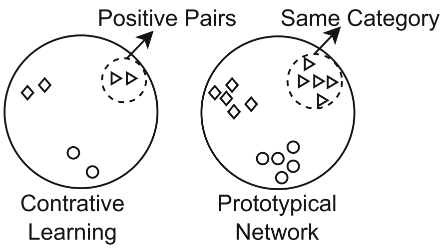

Fig. 2. The figure shows the spatial distribution of samples obtained from comparative learning and the spatial distribution characteristics of samples in the prototypical network (few-shot learning).

In the meta-training phase, we construct a penalty term by defining positive and negative pairs in the support set. Specifically, we consider samples within the support set belonging to the same class as positive pairs, and construct negative pairs from different classes. Therefore, our penalty term can be formulated as:

$$L_c = \frac{\sum_{i=1}^{N} d\left(x_{ip}, x_{iq}\right)}{\sum_{j,k=1}^{N} \sum_{m,n=1}^{K} d\left(x_{jm}, x_{kn}\right)}$$

The d function here represents the measurement method. After adding a penalty term, the meta-training loss can be uniformly expressed as: $L = L_{CE} + tL_c$. The t serves as a hyperparameter that balances the penalty term and cross-entropy loss function.

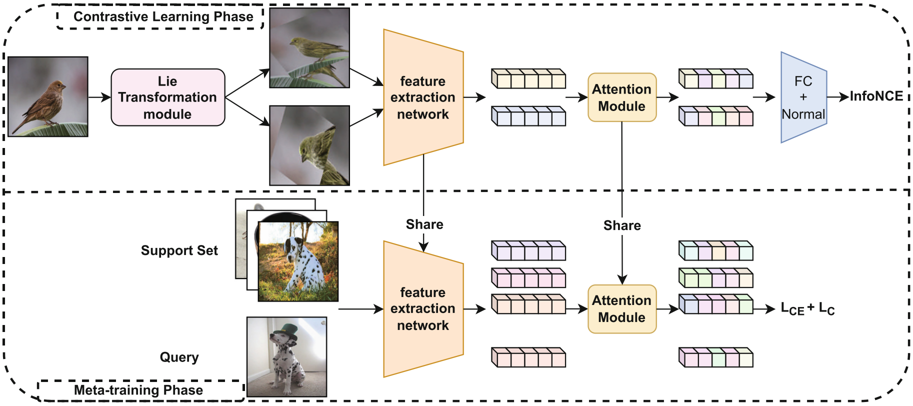

Fig. 3. The overall process framework of our method.

In Fig. 3, we present the overall workflow of the proposed method. Our method divides the training process into two stages: contrastive learning phase and meta-training phase. In the contrastive learning phase, we subject the input samples to two augmentations using a Lie transformation module, and obtain a pair of augmented samples. These augmented samples are first input into a feature extraction network. Then, the output is fed into an attention module, before being processed through two fully connected layers to obtain the sample feature representation. Following the conventional setup of instance discrimination tasks, InfoNCE is computed using sample feature representations to optimize the network. We incorporated attention modules following the feature extraction network in the meta-training phase. We shared the parameters of both the feature extraction network and attention modules trained in the contrastive learning phase, and then optimize the model by incorporating a meta-training loss function with a penalty term.

4 Experiments

In this section, we verify the method's performance through extensive experiments.

4.1 Datasets

We test our method on two public few-shot learning datasets with the 5way-5shot and 5way-1shot tasks, respectively.

MiniImageNet [25]: The miniImageNet dataset is selected from the sizeable visual dataset ImageNet. It contains 100 categories, 600 samples per category, and a total of 60,000 color images. Each image's resolution is set to 84×84. It is partitioned into a training set of 64 categories, a validation set of 16 categories, and a test set of 20 categories.

TieredImageNet [19]: The tieredImageNet dataset, as a subset of the ImageNet dataset, is richer in categories than miniImageNet. There are 608 categories, split into 351, 97 and 160 for the training, validation, and test sets, respectively.

4.2 Implementation Details

For a fair comparison, we used ResNet18 and ResNet12 as the backbones commonly used in few-shot learning.

Contrastive Learning Phase: In the contrastive learning phase, we used the adam [6] optimiser to optimise the model. We seted the initial learning rate to 0.001, the decay factor to 0.1 ,the weight decay was 0.00006 and the momentum to the default value of 0.9. Our batch size was set to 64 and trained through 200 epochs. In the image augmentation phase, we used cropping, flipping, colour transformation and Lie transformation to generate sample pairs. We seted the three randomly generated variables α, β, and γ in the Lie transformation to range between -0.5 and 0.5. We seted the temperature parameter to 0.5 in the loss function of the contrastive learning phase.

Meta-Training Phase: In the meta-training phase, we used the adam [6] optimiser to optimise the model. The optimiser parameters were the same as those used in the comparative learning phase. In the loss function, we seted the temperature parameter t of the penalty term to 0.5. In the 5way-5shot task, we randomly selected 5 categories in the training set. Each category had 5 samples to form the support set and 16 to form the query set. Each task consisted of 105 samples. In the 5way-1shot task, we randomly selected 5 categories in the training set, with 1 sample from each category formed the support set and 16 samples formed the query set. Each task consisted of 85 samples. In 1-shot tasks, the limited number of samples precludes the calculation of penalty terms. We employed Lie transformations to generate auxiliary samples for penalty term computation to address this issue. Each batch contained one task in both the 5shot and 1shot tasks, and there were 100 batches in each epoch, and 400 epochs were used for training.

Evaluation Metric: For the sake of fairness, we followed the assessment scheme unchanged. We evaluate our method with 1000 tasks and report the average accuracy with 95% confidence intervals.

4.3 Results

Following the standard setting, we conducted experiments using ResNet18 as the backbone, employing the original image and mean padding methods to fill the image's blank spaces. We conducted experiments on both miniImagenet and tieredImageNet, and the results are shown in Table 1. The state-of-the-art comparative methods were categorized into Baselines, Optimization-based and Metric-based. As our approach is metric-based, we selected more metric-based models for comparative analysis. We use the Prototypical Network [22] as the baseline, which we re-implemented using ResNet18 as the backbone, and test it

Table 1. FEW-SHOT LEARNING CLASSIFICATION OF RESNET-18 ACCURACIES ON MINI-IMAGENET AND TIERED-IMAGENET UNDER THE SETTING OF 5-WAY 1-SHOT AND 5-WAY 5-SHOT WITH 95% CONFIDENCE INTERVAL. ('-' NOT REPORTED)

Model	**Backbone**	**mini-ImageNet**		**tiered-ImageNet**	
		1-shot	*5-shot*	*1-shot*	*5-shot*
Optimization-based					
MAML [7]	Resnet-18	49.68 ± 0.84	65.73 ± 0.83	-	-
LEO [20]	WRN-28-10	61.76 ± 0.08	77.59 ± 0.12	66.33 ± 0.05	81.44 ± 0.09
Metrics-based					
Matching network [25]	Resnet-18	52.92 ± 0.81	68.93 ± 0.65	-	-
Relation network [24]	Resnet-18	52.19 ± 0.83	70.20 ± 0.66	54.48 ± 0.93	71.32 ± 0.78
SimpleShot [26]	Resnet-18	62.92 ± 0.83	79.07 ± 0.70	69.09 ± 0.22	84.58 ± 0.16
Neg-Cosine [15]	Resnet-18	62.31 ± 0.81	80.97 ± 0.55	-	-
TEAM [17]	Resnet-18	60.10 ± 0.24	75.94 ± 0.23	-	-
CTM [14]	Resnet-18	64.12 ± 0.28	80.51 ± 0.86	68.41 ± 0.39	84.28 ± 1.73
TADAM [16]	Resnet-18	58.50 ± 0.60	76.70 ± 0.45	-	-
PFA [18]	Resnet-18	59.60 ± 0.49	73.74 ± 0.36	-	-
CC+rot [8]	WRN-28-10	62.93 ± 0.45	79.87 ± 0.33	62.93 ± 0.45	79.87 ± 0.33
PSST [4]	WRN-28-10	64.16 ± 0.44	80.64 ± 0.32	-	-
Baseline	Resnet-18	61.18 ± 0.74	79.58 ± 0.64	66.82 ± 0.12	80.82 ± 0.53
Ours:LieOrigin	Resnet-18	62.68 ± 0.49	80.41 ± 0.54	67.22 ± 0.42	82.16 ± 0.61
Ours:LieMean	Resnet-18	$\mathbf{64.92} \pm 0.52$	$\mathbf{82.63} \pm 0.62$	$\mathbf{69.23} \pm 0.34$	$\mathbf{84.92} \pm 0.63$

using the same settings. By observation, our method shows excellent advantages compared to the baseline. Our method also shows better performance compared to optimization-based methods. Compared with the metric-based methods of the same category, [22,24,25] only focus on existing samples and do not solve the problem of sample scarcity, whereas our method expands the sample set and solves the problem to some extent. Our approach exploits the similarity between contrastive learning and metric learning by acquiring a channel attention module during training, enabling it to develop a more discriminative feature. Our method shows better performance in similar methods that exploit the attention mechanism [14,16]. In methods [4,8], which are similar to ours, we use the lie group approach to expand the image set and introduce channel attention to obtain more discriminative features to achieve a more competitive result.

We compare using ResNet-12 as the backbone in the same experimental setup, as shown in Table 2. By observation, our method shows equally competitive experimental results under ResNet-12.

4.4 Ablation Study

This section verifies the effectiveness of the proposed Lie group image augmentation method and attention module through ablation experiments. We used only

Table 2. FEW-SHOT LEARNING CLASSIFICATION OF RESNET-12 ACCURACIES ON MINI-IMAGENET AND TIERED-IMAGENET UNDER THE SETTING OF 5-WAY 1-SHOT AND 5-WAY 5-SHOT WITH 95% CONFIDENCE INTERVAL.

Model	**Backbone**	**mini-ImageNet**		**tiered-ImageNet**	
		1-shot	*5-shot*	*1-shot*	*5-shot*
MAML [7]	ConvNet-4	47.78 ± 1.75	64.31 ± 1.1	52.07 ± 0.91	71.10 ± 1.67
Prototypical Network [22]	Resnet-12	60.76 ± 0.39	78.44 ± 0.21	66.25 ± 0.34	80.11 ± 0.91
Cosine Classifier [1]	Resnet-12	55.43 ± 0.81	77.18 ± 0.61	61.49 ± 0.91	82.37 ± 0.67
MTL [23]	Resnet-12	61.20 ± 1.80	75.50 ± 0.80	65.62 ± 1.80	80.61 ± 0.90
TapNet [31]	Resnet-12	61.65 ± 0.15	76.36 ± 0.10	63.08 ± 0.15	80.26 ± 0.12
Meta-Baseline [3]	Resnet-12	63.17 ± 0.23	79.26 ± 0.17	68.62 ± 0.27	83.29 ± 0.18
DSN-MR [21]	Resnet-12	64.60 ± 0.72	79.51 ± 0.50	67.39 ± 0.82	82.85 ± 0.56
MetaOptNet [11]	Resnet-12	62.64 ± 0.61	78.63 ± 0.46	65.99 ± 0.72	81.56 ± 0.63
Ours:LieMean	**Resnet-12**	$\mathbf{64.94} \pm 0.62$	$\mathbf{80.22 \pm 0.68}$	$\mathbf{68.78} \pm 0.64$	$\mathbf{83.48} \pm 0.58$

ResNet-18 as the feature extractor and the same experimental settings as in the comparison experiments section.

We conducted separate ablation experiments on the mean padding and original image padding Lie group augmentation methods and the attention module employed in the approach. Table 3 shows that the mean padding effect significantly outperforms the original image padding. This may be due to the fact that the positive pairs filled with the original image have a large number of identical features, and the network model found a classification shortcut. This method further improves the model effect and enhances the sample feature representation ability by adding an attention module. Figure 4 shows the Grad-CAM visualization results obtained by our method and prototypical network on the miniImageNet. In the Grad-CAM visualization, our proposed approach demonstrates a stronger capability to focus on the object of interest that requires classification in the image.

Table 3. ABLATION EXPERIMENTS ON MODULE. ('✓' WITH; '-' WITHOUT)

	Lie Group		AT	mini-ImageNet	
	Mean	Origin		1-shot	5-shot
(I)	-	-	-	61.18	79.58
(II)	✓	-	-	63.28	82.52
(III)	-	✓	-	62.32	80.21
(VII)	-	✓	✓	62.68	80.41
(IV)	✓	-	✓	64.92	82.63

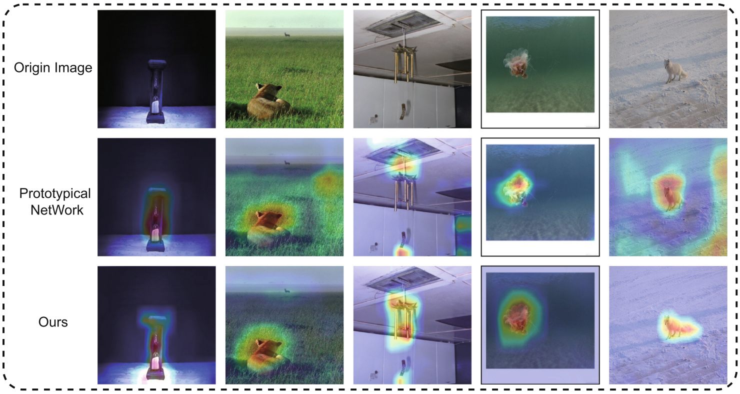

Fig. 4. Grad-CAM visualization of prototypical network and our method sampled randomly from mini-ImageNet.

5 Conclusion

In this paper, we propose a method of few-shot learning based on Lie group contrastive method. Specifically, we are inspired by contrastive learning's strong generalization and use Lie group to improve it. We apply it to few-shot learning to enhance its generalization capabilities. In addition, we use an attention mechanism and a loss penalty term in our approach. They optimize the model regarding sample channels and sample space distribution, respectively. Experimental results show that our method performs significantly on popular few-shot classification benchmark datasets.

Acknowledgments. This work is Supported by the National Key Research and Development Program of China (No. 2018YFA0701700; No. 2018YFA0701701) and National Natural Science Foundation of China (62002253, 62176172, 61672364).

References

1. Chen, W., Liu, Y., Kira, Z., et al.: A closer look at few-shot classification. In: 7th International Conference on Learning Representations, New Orleans, LA, USA, May 6–9, 2019 (2019)
2. Chen, X., Fan, H., Girshick, R., He, K.: Improved baselines with momentum contrastive learning. arXiv preprint arXiv:2003.04297 (2020)
3. Chen, Y., Wang, X., Liu, Z., Xu, H., Darrell, T., et al.: A new meta-baseline for few-shot learning. arXiv preprint arXiv:2003.04390 2(3), 5 (2020)
4. Chen, Z., Ge, J., Zhan, H., et al.: Pareto self-supervised training for few-shot learning. In: IEEE Conference on Computer Vision and Pattern Recognition, virtual, June 19–25, 2021, pp. 13663–13672. Comput. Vision Found. / IEEE (2021). https://doi.org/10.1109/CVPR46437.2021.01345

5. Chopra, S., Hadsell, R., LeCun, Y.: Learning a similarity metric discriminatively, with application to face verification. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 20–26 June 2005, San Diego, CA, USA, pp. 539–546. IEEE Computer Society (2005). https://doi.org/10.1109/CVPR.2005.202
6. Diederik, P., Kingma, E.A.: Adam: a method for stochastic optimization. In: 3rd International Conference on Learning Representations, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings (2015)
7. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: Proceedings of the 34th International Conference on Machine Learning, Sydney, NSW, Australia, 6–11 August 2017. Proceedings of Machine Learning Research, vol. 70, pp. 1126–1135. PMLR (2017)
8. Gidaris, S., Bursuc, A., Komadakis, N., et al.: Boosting few-shot visual learning with self-supervision. In: 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Korea (South), October 27 - November 2, 2019, pp. 8058–8067. IEEE (2019). https://doi.org/10.1109/ICCV.2019.00815
9. Grill, J., et al.: Bootstrap your own latent - a new approach to self-supervised learning. In: Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, December 6–12, 2020, virtual (2020)
10. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked autoencoders are scalable vision learners. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 16000–16009 (2022)
11. Lee, K., Maji, S., Ravichandran, A., Soatto, S.: Meta-learning with differentiable convex optimization. In: IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, June 16–20, 2019, pp. 10657–10665. Computer Vision Foundation / IEEE (2019). https://doi.org/10.1109/CVPR.2019.01091
12. Li, G., Zheng, C., Su, B.: Transductive distribution calibration for few-shot learning. Neurocomputing **500**, 604–615 (2022)
13. Li, G., Zheng, H., Liu, D., Wang, C., Su, B., Zheng, C.: Semmae: Semantic-guided masking for learning masked autoencoders. arXiv preprint arXiv:2206.10207 (2022)
14. Li, H., Eigen, D., Dodge, S., Zeiler, M., Wang, X.: Finding task-relevant features for few-shot learning by category traversal. In: IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, June 16–20, 2019, pp. 1–10. Computer Vision Foundation / IEEE (2019). https://doi.org/10.1109/CVPR.2019.00009
15. Liu, B., et al.: Negative margin matters: Understanding margin in few-shot classification. In: Computer Vision - ECCV 2020–16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IV. Lecture Notes in Computer Science, vol. 12349, pp. 438–455. Springer (2020). https://doi.org/10.1007/978-3-030-58548-8_26
16. Oreshkin, B.N., López, P.R., Lacoste, A.: TADAM: task dependent adaptive metric for improved few-shot learning. In: Advances in Neural Information Processing Systems 31: December 3–8, 2018, Montréal, Canada, pp. 719–729 (2018)
17. Qiao, L., Shi, Y., Li, J., Wang, Y., Huang, T., Tian, Y.: Transductive episodic-wise adaptive metric for few-shot learning. In: 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Korea (South), October 27 - November 2, 2019, pp. 3602–3611. IEEE (2019). https://doi.org/10.1109/ICCV.2019.00370

18. Qiao, S., Liu, C., et al.: Few-shot image recognition by predicting parameters from activations. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, June 18–22, 2018, pp. 7229–7238. Computer Vision Foundation / IEEE Computer Society (2018). https://doi.org/10.1109/CVPR.2018.00755
19. Ren, M., et al.: Meta-learning for semi-supervised few-shot classification. In: 6th International Conference on Learning Representations, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings. OpenReview.net (2018)
20. Rusu, A.A., et al.: Meta-learning with latent embedding optimization. In: 7th International Conference on Learning Representations, New Orleans, LA, USA, May 6–9, 2019. OpenReview.net (2019)
21. Simon, C., Koniusz, P., Nock, R., Harandi, M.: Adaptive subspaces for few-shot learning. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, June 13–19, 2020, pp. 4135–4144. Computer Vision Foundation / IEEE (2020). https://doi.org/10.1109/CVPR42600.2020.00419
22. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. In: Advances in Neural Information Processing Systems 30: December 4–9, 2017, CA, USA, pp. 4077–4087 (2017)
23. Sun, Q., et al.: Meta-transfer learning for few-shot learning. In: IEEE Conference on Computer Vision and Pattern Recognition, USA, June 16–20, 2019, pp. 403–412. Computer Vision Foundation / IEEE (2019). https://doi.org/10.1109/CVPR.2019.00049
24. Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P.H., Hospedales, T.M.: Learning to compare: Relation network for few-shot learning. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, June 18–22, 2018, pp. 1199–1208. Computer Vision Foundation / IEEE Computer Society (2018). https://doi.org/10.1109/CVPR.2018.00131
25. Vinyals, O., Blundell, C., Lillicrap, T., Wierstra, D.: Matching networks for one shot learning. In: Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5–10, 2016, Barcelona, Spain, pp. 3630–3638 (2016)
26. Wang, Y., Chao, W.L., Weinberger, K.Q., van der Maaten, L.: Simpleshot: Revisiting nearest-neighbor classification for few-shot learning. arXiv preprint arXiv:1911.04623 (2019)
27. Wu, Z., Xiong, Y., Yu, S.X., Lin, D.: Unsupervised feature learning via non-parametric instance discrimination. In: Proceedings of the IEEE Conference On Computer Vision and Pattern Recognition, pp. 3733–3742 (2018)
28. Xu, C., Zhu, G.: Semi-supervised learning algorithm based on linear lie group for imbalanced multi-class classification. Neural Process. Lett. **52**(1), 869–889 (2020). https://doi.org/10.1007/s11063-020-10287-8
29. Xu, C., Zhu, G., Shu, J.: Robust joint representation of intrinsic mean and kernel function of lie group for remote sensing scene classification. IEEE Geosci. Remote Sens. Lett. **18**(5), 796–800 (2021). https://doi.org/10.1109/LGRS.2020.2986779
30. Xu, C., Zhu, G., Shu, J.: A combination of lie group machine learning and deep learning for remote sensing scene classification using multi-layer heterogeneous feature extraction and fusion. Remote. Sens. **14**(6), 1445 (2022)
31. Yoon, S.W., Seo, J., Moon, J.: Tapnet: Neural network augmented with task-adaptive projection for few-shot learning. In: Proceedings of the 36th International Conference on Machine Learning, 9–15 June 2019, Long Beach, California, USA. Proceedings of Machine Learning Research, vol. 97, pp. 7115–7123. PMLR (2019)