Chinese Conference on Pattern Recognition and Computer Vision (PRCV)
PRCV 2021: Pattern Recognition and Computer Vision pp 443-454 | Cite as

Local Mutual Metric Network for Few-Shot Image Classification

Authors | Authors and affiliations

Yaohui Li, Huaxiong Li, Haoxing Chen, Chunlin Chen

Conference paper
First Online: 22 October 2021

Part of the Lecture Notes in Computer Science book series (LNCS, volume 13019)

Abstract

Few-shot image classification aims to recognize unseen categories with only a few labeled training samples. Recent metric-based approaches tend to represent each sample with a high-level semantic representation and make decisions according to the similarities between the query sample and support categories. However, high-level concepts are identified to be poor at generalizing to novel concepts that differ from previous seen concepts due to domain shifts. Moreover, most existing methods conduct one-way instance-level metric without involving more discriminative local relations. In this paper, we propose a *Local Mutual Metric Network (LM2N)*, which combines low-level structural representations with high-level semantic representations by unifying all abstraction levels of the embedding network to achieve a balance between discrimination and generalization ability. We also propose a novel local mutual metric strategy to collect and reweight local relations in a bidirectional manner. Extensive experiments on five benchmark datasets (i.e. *mini*ImageNet, *tiered*ImageNet and three fine-grained datasets) show the superiority of our proposed method.

Keywords

Few-shot learning · Metric learning · Attention · Local representation

1 Introduction

Despite that deep learning [10, 18] has made tremendous advances in many machine learning tasks [37], its data-driven nature restricts its performance and efficiency in practical applications where collecting enough samples and correctly labeling them are expensive. Therefore, Few-shot Learning (FSL) [7, 8, 16] is widely studied recently, which aims to generalize to novel unseen categories with scarce data after training on seen categories with sufficient data. Few-shot Image Classification (FSIC) [12, 23] is one of the well-studied fields in FSL, which aims at making classification on unseen categories with a small number of training samples.

Recently, many meta-learning based approaches have been proposed to solve the FSIC problem, which can be divided into two main streams: optimization-based approaches [8, 31] and metric-based approaches [20, 29]. Specifically, approaches based on optimization target at learning a suitable initialization or updating strategy for the base model, which helps the base model to converge on novel tasks with a few steps of gradient descent. Metric-based approaches aim to learn a transferable feature space, in which the homogeneous samples are close to each other while the heterogeneous samples are far away and the query samples are categorized by measuring their similarities with the labeled support categories.

Metric-based methods [9, 21, 22, 30, 32, 34] have attracted a lot of attention due to their efficiency and simplicity on solving the FSIC problem. However, there are still some limitations remain to be solved.

First, most metric-based approaches adopt the high-level semantic representation [3, 30, 32] to describe each sample. Although high-level representations contain rich semantic information with high degree of discrimination, they are not shared between different categories compared with low-level features [19]. Meanwhile, recent metric-based approaches mainly follow the paradigm of meta-learning, which is inherently designed to enhance the generalization ability of the base model by learning inductive bias (meta knowledge) [11]. However, several recent works [5, 33] show that the classic pretraining combined with fine tuning beats the advanced meta-learning methods, which indicates that generalizing just by meta knowledge is far from enough. Second, recent methods mainly conduct instance-level metric [22, 30, 32] while ignore the local relations between two feature maps, which can not reveal the true similarities and is obviously inefficient in practice.

To solve the above limitations, in this paper, we first propose a local representation fusion strategy to combine low-level representations with high-level representations. We further propose a non-parametric Local Mutual Metric Module (LM3), which comprehensively compares two feature maps by collecting and reweighting their local relations.

Our main contributions are as follows:

- We propose a representation fusion strategy to make use of both low-level and high-level features by a learned fusion layer, which aims to achieve a balance between discrimination and generalization ability.
- We design a novel local mutual metric strategy for the FSIC problem, which explores the pairwise local relations between two feature maps in a bidirectional manner.
- We adopt a Convolutional Block Attention Module (CBAM) [36] to select and highlight discriminative semantic regions.

2 Related Works

Recent meta-learning based FSIC approaches can be broadly divided into two branches: optimization-based and metric-based approaches.

2.1 Optimization-Based Approaches

Optimization-based approaches [2, 8, 31] utilize the meta learner to learn optimal initialization parameters for the base learner, which is sensitive to novel samples so that the base model can fast adapt to unseen categories with only a few steps of gradient descent. MAML [8] is a typical optimization-based method for the FSIC, which concludes parameterized meta knowledge from a series of episodes (tasks) to boost model-agnostic generalization ability. Later optimization-based approaches mainly follow the ideology of MAML.

However, optimization-based approaches suffer from the computation of high-order gradient [25], which is out of question in our metric-based LM2N.

2.2 Metric-Based Approaches

Metric-based approaches [4, 9, 21, 22, 30, 32, 34] aim to learn a transferable embedding space (feature extractor), in which to map labeled support samples and unlabeled query samples. Final classification is completed by comparing the distances between query samples and support categories. Obviously, the key points of metric-based approaches are transferable feature representations and an effective metric strategy.

Specifically, Matching Network [34] utilizes a LSTM for feature embedding and compares the cosine similarities between feature maps for classification. Note that the Matching Network also proposes the widely used episodic training mechanism. Prototypical Network [30] utilizes the center of the support class as the class representation (prototype) and makes classification by computing the Euclidean distances between query samples and class prototypes. Relation Network [32] follows the representation strategy of [30] and proposes a novel MLP-based metric, which aims to learn a transferable and effective nonlinear metric function.

Unlike the above approaches, CovaMNet [22] represents images by local representations, which treats a feature map (a $h \times w \times d$ tensor) as $h \times w$ d-dimension local representation vectors. CovaMNet makes classification according to the distribution consistency between query samples and support categories. DN4 [21] and SAML [9] also follow the local representation strategy.

Based on the local representation mechanism, our LM2N unifies low-level and high-level features for a fused local representation. Unlike the existing instance-level metric, we propose a novel local mutual metric strategy, which explores the local relations between two feature maps bilaterally to reveal their local correlations comprehensively. Note that some methods have explored the local metric in FSIC, i.e. DN4 [21] and SAML [9]. DN4 utilizes a unidirectional Nearest Neighbor searching to acquire the local relations of two feature maps and makes classification according to the summation of the pairwise local similarities, which shows promising performance. SAML uses a relation matrix to collect and select useful pairwise local vectors, followed by a MLP-based nonlinear metric network [32].

3 Methodology

3.1 Problem Formulation

In this paper, we follow the classic episodic training [34] mechanism. Specifically, we utilize a series of episodes (tasks) on the training set to train the model and evaluate the model with a series of episodes on the test set.

In each episode, we randomly sample a support set with N categories (K samples per category) $\mathcal{S} = \{\{s_1^1, y_1\}, \{s_1^2, y_1\}, \cdots, \{s_N^K, y_N\}\}$ and an unlabeled query set with NM samples $\mathcal{Q} = \{q_1, q_2, \cdots, q_{NM}\}$. Notice that $\mathcal{S}$ and $\mathcal{Q}$ share the same label space. Finally, each episode can be viewed as a N-way K-shot task, which aims to categorize the NM query samples into the N categories with only K labeled samples per category.

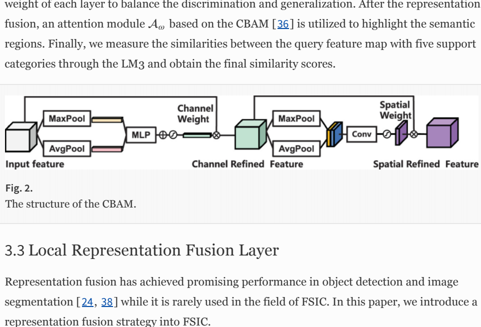

Fig. 1.
The overview of our proposed LM2N under the 5-way 1-shot setting.

3.2 Overview

The overview of our proposed LM2N is shown in Fig. 1. First, we feed the support set (5-way 1-shot setting) and a query sample into the embedding network $\mathcal{F}_\theta$. Then, a fusion layer $\mathcal{I}_\phi$ is applied to fuse the outputs of all layers of the feature extractor, which automatically learns the weight of each layer to balance the discrimination and generalization. After the representation fusion, an attention module $\mathcal{A}_\omega$ based on the CBAM [36] is utilized to highlight the semantic regions. Finally, we measure the similarities between the query feature map with five support categories through the LM3 and obtain the final similarity scores.

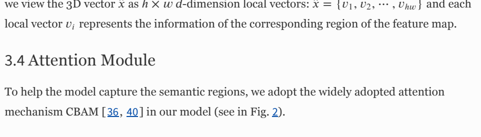

Fig. 2.
The structure of the CBAM.

3.3 Local Representation Fusion Layer

Representation fusion has achieved promising performance in object detection and image segmentation [24, 38] while it is rarely used in the field of FSIC. In this paper, we introduce a representation fusion strategy into FSIC.

Specifically, given a n-layer embedding network, we aim to unify the outputs of these n layers (n representation maps). In practice, there is no enough prior knowledge to give appropriate weights to these n feature maps. Therefore, a fusion layer $\mathcal{I}_\phi$ is utilized to learn the fusion weights under the meta-learning framework. In practice, we resize these posterior feature maps to the same size and combine them by concatenation before putting into the fusion layer. After the representation fusion, we can obtain a balanced feature map:

$$\hat{x} = \mathcal{I}_\phi(\mathcal{F}_\theta(x)) \in \mathbb{R}^{h \times w \times d}. \tag{1}$$

Where $\mathcal{F}_\theta(x)$ denotes the outputs of the meaningful layers of the backbone $\mathcal{F}_\theta$. Following [9], we view the 3D vector $\hat{x}$ as $h \times w$ d-dimension local vectors: $\hat{x} = \{v_1, v_2, \cdots, v_{hw}\}$ and each local vector v_i represents the information of the corresponding region of the feature map.

3.4 Attention Module

To help the model capture the semantic regions, we adopt the widely adopted attention mechanism CBAM [36, 40] in our model (see in Fig. 2).

Generally, CBAM consists of two parts, i.e. channel attention and spatial attention. Specifically, channel attention first squeezes the feature map along the spatial dimension by global average pooling and global max pooling to obtain two weight tensors. Then the two weight tensors are combined to generate the channel attention weight, which is further used to refine the feature map. Similar to the channel attention, spatial attention squeezes the feature map along the channel dimension and generates the spatial attention weight, which expands along the channel dimension to refine the feature.

Through the attention module $\mathcal{A}_\omega$, the semantic objects are effectively highlighted, which contributes to the further metric (see Table 3).

$$\bar{x} = \mathcal{A}_\omega(\hat{x}). \tag{2}$$

3.5 Local Mutual Metric Module (LM3)

In this paper, we propose a novel local mutual metric strategy to explore the local relations between two feature maps in a bidirectional manner. Given a support class $S_n = \{s_n^1, s_n^2, \cdots, s_n^K\}$ and a query instance q, following the above procedures, we adopt the central class strategy [30] here for concise expression:

$$\bar{s_n} = \frac{1}{|S_n|} \sum_{(s_n^i, y_n) \in S_n} \mathcal{A}_\omega(\mathcal{I}_\phi(\mathcal{F}_\theta(s_n^i))). \tag{3}$$

We rewrite them by local representations:

$$\bar{s_n} = \{t_1, t_2, \cdots, t_{hw}\}, \bar{q} = \{l_1, l_2, \cdots, l_{hw}\}. \tag{4}$$

Then a LM3 is applied to calculate the similarity between $\bar{s_n}$ and $\bar{q}$. We first define a local compare function $g(,)$ [9] to calculate the similarity between two local vectors. Mathematically, the local mutual metric function is defined as:

$$h(\bar{s_n}, \bar{q}) = \sum_{i=1}^{hw} \sum_{j=1}^{hw} \alpha_{i,j} g(t_i, l_j) + \sum_{j=1}^{hw} \sum_{i=1}^{hw} \beta_{i,j} g(t_i, l_j), \tag{5}$$

$$\alpha_{i,j} = \frac{e^{[g(t_i, l_j)/T]}}{\sum_{k=1}^{hw} e^{[g(t_i, l_k)/T]}}, \beta_{i,j} = \frac{e^{[g(t_i, l_j)/T]}}{\sum_{k=1}^{hw} e^{[g(t_k, l_j)/T]}}. \tag{6}$$

Here, T denotes the influence coefficient of the similarity, which gives the similar local vector pairs much higher weights while gives the dissimilar local vector pairs much lower weights. The value of the local mutual metric function $h(,)$ represents the similarity score between the query sample q and the n-th support class S_n.

Given an episode, there are N categories in the support set $\{S_1, S_2, \cdots, S_N\}$ and NM unlabeled samples in the query set $\{q_1, q_2, \cdots, q_{NM}\}$. Therefore, for each query sample there are N similarity scores involving N support classes. We utilize the softmax function to compute the probabilities and the final loss function of this episode is defined as:

$$Loss = -\frac{1}{NM} \sum_{i=1}^{NM} \sum_{j=1}^{N} \varphi(\hat{y}_i = y_j) log \frac{e^{h(\bar{s_j}, \bar{q_i})}}{\sum_{k=1}^{N} e^{h(\bar{s_k}, \bar{q_i})}}, \tag{7}$$

in which $\hat{y}_i$ represents the predicted label for q_i and $\varphi()$ is a flag function that equals one if its argument is true and zero otherwise.

4 Experiments

4.1 Datasets

We evaluate our method on five benchmark datasets, and all images are resized to 84×84 in our model.

***mini* ImageNet.** *mini*ImageNet [34] is sampled from the ImageNet [6], which consists of 100 categories and 600 samples per category. We split it into 64, 16 and 20 categories for training, validation and test by convention, respectively [22].

***tiered* ImageNet.** *tiered*ImageNet [27] is also a subset of the ImageNet [6]. There are 608 categories and 779, 165 images in it. We split the dataset into 351, 97 and 160 categories for training, validation and test, respectively.

***Stanford Dogs* , *Stanford Cars* and *CUB-200-2011* .** We utilize three fine-grained datasets to evaluate the fine-grained FSIC performance of our method. For *Stanford Dogs* [14], we divide it into 70, 20 and 30 classes. For *Stanford Cars* [17], we split it into 130, 17 and 49 categories while for *CUB-200-2011* [35], we choose 100 classes for training and both 50 classes for validation and test [22].

4.2 Network Architecture

In our model, the main network architecture consists of the embedding network, the representation fusion layer and the attention module.

Specifically, the embedding network is a 4-layer CNN [21, 22] with four convolutional blocks and each convolutional block contains a convolutional layer with 64 3×3 filters, a batch normalization layer and a Leaky ReLU layer. In addition, we add 2×2 max pooling after the first two convolutional blocks. The representation fusion layer is actually a 1×1 convolutional layer.

The attention module consists of a 2-layer MLP to learn the channel attention weight and a 1×1 convolutional layer to learn the spatial attention weight [40].

4.3 Experimental Settings

We implement our model under the framework of Pytorch [26]. Specifically, under the setting of the meta-learning (episodic training), we train, validate and test our model on a series of N-way K-shot episodes (tasks) randomly sampled from the training set, validation set and test set respectively. Specifically, for the end-to-end training stage, we randomly construct 200, 000 episodes from the training set and 20, 000 episodes from the validation set. When $N = 5, K = 1$, a 5-class episode contains 1 support sample and 10 query samples per calss; when $N = 5, K = 5$, a 5-class episode contains 5 support samples and 10 query samples per class.

Notice that we initialize the learning rate to 0.005 and halve it per 40, 000 episodes. A widely adopted Adam [15] optimizer is applied to optimize our model during the training procedure. During the test stage, we randomly construct 600 episodes to evaluate the performance of our model on novel concepts.

Table 1.
Comparisons with the state-of-the-art methods on *mini*ImageNet and *tiered*ImageNet with 95% confidence intervals. The best and the second best results of each column are shown in red and blue respectively.

Model	Type	*mini*ImageNet		*tiered*ImageNet	
		1-shot(%)	5-shot(%)	1-shot(%)	5-shot(%)
MAML [8]	Optimization	48.70 ± 1.84	63.11 ± 0.92	51.67 ± 1.81	70.30 ± 1.75
MTL [31]	Optimization	45.60 ± 1.84	61.20 ± 0.90	–	–
MAML++ [2]	Optimization	52.15 ± 0.26	68.32 ± 0.44	–	–
MatchingNet [34]	Metric	43.56 ± 0.84	55.31 ± 0.73	–	–
IMP [1]	Metric	49.60 ± 0.80	68.10 ± 0.80	–	–
SAML [9]	Metric	52.64 ± 0.56	67.32 ± 0.75	–	–
DSN [29]	Metric	51.78 ± 0.96	68.99 ± 0.69	–	–
TNet [39]	Metric	52.39	67.89	–	–
GNN [28]	Metric	50.33 ± 0.36	66.41 ± 0.63	–	–
ProtoNet [30]	Metric	49.42 ± 0.78	68.20 ± 0.66	48.58 ± 0.87	69.57 ± 0.75
RelationNet [32]	Metric	50.44 ± 0.82	65.32 ± 0.70	54.48 ± 0.93	71.31 ± 0.78
CovaMNet [22]	Metric	51.19 ± 0.76	67.65 ± 0.63	54.98 ± 0.90	71.51 ± 0.75
DN4 [21]	Metric	51.24 ± 0.74	71.02 ± 0.64	53.37 ± 0.86	74.45 ± 0.70
LM2N(Ours)	Metric	53.22 ± 0.76	71.98 ± 0.91	54.66 ± 0.79	73.31 ± 0.85

Table 2.
Experiments on three fine-grained datasets. The best and the second best results of each column are shown in red and blue respectively.

Model	5-Way accuracy(%)					
	Stanford dogs		*Stanford Cars*		*CUB-200-2011*	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
MatchingNet [34]	35.80	47.50	34.80	44.70	61.16	72.86
ProtoNet [30]	37.59	48.19	40.90	52.93	51.31	70.77
MAML [8]	45.81	60.01	48.17	61.85	55.92	72.09
RelationNet [32]	44.49	56.35	48.59	60.98	62.45	76.11
CovaMNet [22]	49.10	63.04	56.65	71.33	60.58	74.24
DN4 [21]	45.41	63.51	59.84	88.65	52.79	81.45
LRPABN [13]	46.17	59.11	56.31	70.23	63.63	76.06
LM2N(Ours)	51.15	67.63	59.98	84.65	64.95	79.82

Table 3.
Ablation study on *mini*ImageNet.

Ablation models			5-way accuracy (%)	
Fusion	Attention	LM3	1-shot	5-shot
			49.24	68.70
	√		50.42	68.96
√			50.16	68.03
		√	51.63	70.13
√	√		50.91	69.65
	√	√	52.46	71.01
√		√	51.93	71.15
√	√	√	53.22	71.98

4.4 Comparisons with Other Methods

Experiments on the Routine FSIC Datasets. We compare our LM2N with both optimization-based and metric-based SOTA methods on *mini*ImageNet and *tiered*ImageNet. As is shown in Table 1, our LM2N achieves promising performance on both datasets under both settings. Specifically, on the *mini*ImageNet, our LM2N achieves the best results under both 1-shot and 5-shot settings with 0.58% and 0.96% improvements respectively. Compared with our base model Prototypical Network, our method achieves 3.80% and 3.78% improvements. On the *tiered*ImageNet, our LM2N also achieves competitive performance compared with the other methods.

The outstanding performances indicate the effectiveness of our method, which explores the mutual local relations between the query samples and the support categories. Most previous methods compare the query samples and the support categories on the instance-level while DN4 conducts a local metric by searching out the local nearest neighbors. Unlike these methods, our L2MN takes a mutual view to collect local relations and takes a whole picture of the feature maps. Besides, the representation fusion strategy enhances the generalization ability to novel concepts and the attention mechanism highlights the discriminative features.

Experiments on the Fine-Grained FSIC Datasets. We also compare our method with the SOTA method on three fine-grained FSIC datasets. As is shown in Table 2, LM2N achieves competitive performances. Specifically, our LM2N achieve the best performances on four columns with 0.14%–4.12% improvement while on the other two columns LM2N both achieves the second best results.

Our LM2N is naturally suitable for fine-grained problems. For one thing, LM2N utilizes the local metric strategy to explore the local relations between two feature maps, for another, the attention mechanism helps to highlight the discriminative local regions, which is obviously effective on fine-grained FSIC.

5 Discussion

5.1 Ablation Study

In this paper we introduce the representation fusion strategy to FSIC tasks and propose a novel local mutual metric mechanism. Moreover, we adopt CBAM as the attention module to highlight the discriminative local regions [40]. In general, these three parts achieve promising performance together while we also want to know their respective contributions to the model. Therefore, we design the ablation study to reveal the contributions of each part. Specifically, we step-by-step remove one or several parts and make comparisons with the original model (see Table 3). Experimental results show that each component more or less contributes to the model.

5.2 Cross-Domain FSIC Analysis

To further evaluate the robustness of our LM2N when facing severe domain-shift, we conduct cross-domain experiments by training on the *mini*ImageNet and testing on the *CUB-200-2011* [5]. Experimental results are shown in Table 4. From the table we can see that our LM2N shows better generalization ability than most previous methods when facing large domain shift.

Table 4.
Cross-domain performance (***mini*ImageNet→ *CUB***). We adopt the results of the Resnet-18 reported by [5].

Model	Embedding	5-way 5-shot(%)
Baseline [5]	Resnet-18	65.57 ± 0.70
MatchingNet [34]	Resnet-18	53.07 ± 0.74
ProtoNet [30]	Resnet-18	62.02 ± 0.70
MAML [8]	Resnet-18	51.34 ± 0.72
RelationNet [32]	Resnet-18	57.71 ± 0.73
CovaMNet [22]	Conv-64F	63.21 ± 0.68
DN4 [21]	Conv-64F	63.42 ± 0.70
QPN (Ours)	Conv-64F	64.40 ± 0.79

Table 5.
The performance with different similarity functions on *mini*ImageNet.

Metric	5-way 1-shot	5-way 5-shot
Gaussian	51.88 ± 0.72	70.83 ± 0.85
Euclidean	51.08 ± 0.87	70.63 ± 0.82
Cosine	53.22 ± 0.76	71.98 ± 0.91

5.3 Influence of Different Local Compare Functions

In the local mutual metric module, we utilize the local compare function $g(,)$ to measure the similarity between two local vectors. Certainly, the local metric function has many choices, e.g. cosine similarity, Euclidean similarity and Gaussian similarity [9].

Cosine similarity:

$$g(a, b) = \frac{a^T b}{||a|| \cdot ||b||}. \tag{8}$$

Euclidean similarity:

$$g(a, b) = \frac{1}{e^{||a-b||^2}}. \tag{9}$$

Gaussian similarity:

$$g(a, b) = e^{a \cdot b}. \tag{10}$$

Obviously, as is shown in Table 5, cosine similarity achieves the best performance under both settings.

6 Conclusion

In this paper, we propose a Local Mutual Metric Network for the FSIC. We propose a novel metric strategy LM3 to explore the local relations between two feature maps in a bidirectional manner for a more effective metric. We also achieves a balance between discrimination and generalization ability by multi-level representation fusion. An attention module is adopted to highlight the semantic objects. Experimental results on five FSIC benchmark datasets show the superiority of our method. Ablation study and the experiments on the cross-domain FSIC task demonstrate the effectiveness and the robustness of our method.

Notes

Acknowledgement

This work was partially supported by National Natural Science Foundation of China (Nos. 62176116, 71732003, 62073160, and 71671086) and the National Key Research and Development Program of China (Nos. 2018YFB1402600).

References

1. Allen, K., Shelhamer, E., Shin, H., Tenenbaum, J.: Infinite mixture prototypes for few-shot learning. In: ICML, pp. 232–241 (2019)
   Google Scholar

2. Antoniou, A., Edwards, H., Storkey, A.: How to train your maml. In: ICLR (2018)
   Google Scholar

3. Chen, H., Li, H., Li, Y., Chen, C.: Multi-scale adaptive task attention network for few-shot learning. arXiv preprint arXiv:2011.14479 (2020)

4. Chen, H., Li, H., Li, Y., Chen, C.: Multi-level metric learning for few-shot image recognition. arXiv preprint arXiv:2103.11383 (2021)

5. Chen, W.Y., Liu, Y.C., Kira, Z., Wang, Y.C.F., Huang, J.B.: A closer look at few-shot classification. In: ICLR (2019)
   Google Scholar

6. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: a large-scale hierarchical image database. In: CVPR, pp. 248–255 (2009)
   Google Scholar

7. Fe-Fei, L., et al.: A bayesian approach to unsupervised one-shot learning of object categories. In: ICCV, pp. 1134–1141 (2003)
   Google Scholar

8. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: ICML, pp. 1126–1135 (2017)
   Google Scholar

9. Hao, F., He, F., Cheng, J., Wang, L., Cao, J., Tao, D.: Collect and select: semantic alignment metric learning for few-shot learning. In: ICCV, pp. 8460–8469 (2019)
   Google Scholar

10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR, pp. 770–778 (2016)
    Google Scholar

11. Hochreiter, S., Younger, A.S., Conwell, P.R.: Learning to learn using gradient descent. In: International Conference on Artificial Neural Networks, pp. 87–94 (2001)
    Google Scholar

12. Hou, R., Chang, H., Ma, B., Shan, S., Chen, X.: Cross attention network for few-shot classification. In: NeurIPS, pp. 4003–4014 (2019)
    Google Scholar

13. Huang, H., Zhang, J., Zhang, J., Xu, J., Wu, Q.: Low-rank pairwise alignment bilinear network for few-shot fine-grained image classification. IEEE Trans. Multimedia **23**, 1666–1680 (2020)
    CrossRef Google Scholar

14. Khosla, A., Jayadevaprakash, N., Yao, B., Li, F.F.: Novel dataset for fine-grained image categorization: stanford dogs. In: CVPR Workshop on FGVC, vol. 2 (2011)
    Google Scholar

15. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: ICLR (2015)
    Google Scholar

16. Koch, G., Zemel, R., Salakhutdinov, R., et al.: Siamese neural networks for one-shot image recognition. In: ICML Deep Learning Workshop, vol. 2 (2015)
    Google Scholar

17. Krause, J., Stark, M., Deng, J., Fei-Fei, L.: 3D object representations for fine-grained categorization. In: ICCV Workshop, pp. 554–561 (2013)
    Google Scholar

18. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NeurIPS, vol. 25, pp. 1097–1105 (2012)
    Google Scholar

19. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015)
    CrossRef Google Scholar

20. Li, W., Wang, L., Huo, J., Shi, Y., Gao, Y., Luo, J.: Asymmetric distribution measure for few-shot learning, pp. 2957–2963 (2020)
    Google Scholar

21. Li, W., Wang, L., Xu, J., Huo, J., Gao, Y., Luo, J.: Revisiting local descriptor based image-to-class measure for few-shot learning. In: CVPR, pp. 7260–7268 (2019)
    Google Scholar

22. Li, W., Xu, J., Huo, J., Wang, L., Gao, Y., Luo, J.: Distribution consistency based covariance metric networks for few-shot learning. In: AAAI, pp. 8642–8649 (2019)
    Google Scholar

23. Li, Y., Li, H., Chen, H., Chen, C.: Hierarchical representation based query-specific prototypical network for few-shot image classification. arXiv preprint arXiv:2103.11384 (2021)

24. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: CVPR, pp. 2117–2125 (2017)
    Google Scholar

25. Mishra, N., Rohaninejad, M., Chen, X., Abbeel, P.: A simple neural attentive meta-learner. In: ICLR (2018)
    Google Scholar

26. Paszke, A., et al.: Pytorch: an imperative style, high-performance deep learning library. In: NeurIPS, vol. 32, pp. 8026–8037 (2019)
    Google Scholar

27. Ren, M., et al.: Meta-learning for semi-supervised few-shot classification. In: ICLR (2018)
    Google Scholar

28. Satorras, V.G., Estrach, J.B.: Few-shot learning with graph neural networks. In: ICLR (2018)
    Google Scholar

29. Simon, C., Koniusz, P., Nock, R., Harandi, M.: Adaptive subspaces for few-shot learning. In: CVPR, pp. 4136–4145 (2020)
    Google Scholar

30. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. In: NeurIPS, pp. 4080–4090 (2017)
    Google Scholar

31. Sun, Q., Liu, Y., Chua, T.S., Schiele, B.: Meta-transfer learning for few-shot learning. In: CVPR, pp. 403–412 (2019)
    Google Scholar

32. Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P.H., Hospedales, T.M.: Learning to compare: relation network for few-shot learning. In: CVPR, pp. 1199–1208 (2018)
    Google Scholar

33. Tian, Y., Wang, Y., Krishnan, D., Tenenbaum, J.B., Isola, P.: Rethinking few-shot image classification: a good embedding is all you need? In: ECCV, pp. 266–282 (2020)
    Google Scholar

34. Vinyals, O., Blundell, C., Lillicrap, T., Kavukcuoglu, K., Wierstra, D.: Matching networks for one shot learning. In: NeurIPS, pp. 3637–3645 (2016)
    Google Scholar

35. Wah, C., Branson, S., Welinder, P., Perona, P., Belongie, S.: The caltech-ucsd birds-200-2011 dataset (2011)
    Google Scholar

36. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: Cbam: convolutional block attention module. In: ECCV, pp. 3–19 (2018)
    Google Scholar

37. Zhang, C., Li, H., Chen, C., Qian, Y., Zhou, X.: Enhanced group sparse regularized nonconvex regression for face recognition. IEEE Trans. Pattern Anal. Mach. Intell. (2020). https://doi.org/10.1109/TPAMI.2020.3033994
    CrossRef Google Scholar

38. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: CVPR, pp. 2881–2890 (2017)
    Google Scholar

39. Zhu, W., Li, W., Liao, H., Luo, J.: Temperature network for few-shot learning with distribution-aware large-margin metric. Pattern Recogn. **112**, 107797 (2021)
    CrossRef Google Scholar

40. Zhu, Y., Liu, C., Jiang, S.: Multi-attention meta learning for few-shot fine-grained image recognition. In: IJCAI, pp. 1090–1096 (2020)
    Google Scholar

Copyright information

About this paper

Check for updates

Cite this paper as:
Li Y., Li H., Chen H., Chen C. (2021) Local Mutual Metric Network for Few-Shot Image Classification. In: Ma H. et al. (eds) Pattern Recognition and Computer Vision. PRCV 2021. Lecture Notes in Computer Science, vol 13019. Springer, Cham. https://doi.org/10.1007/978-3-030-88004-0_36

First Online
22 October 2021

DOI
https://doi.org/10.1007/978-3-030-88004-0_36

Publisher Name
Springer, Cham

Print ISBN
978-3-030-88003-3

Online ISBN
978-3-030-88004-0

eBook Packages
Computer Science
Computer Science (R0)

Reprints and Permissions