Bi-directional Feature Reconstruction Network for Fine-Grained Few-Shot Image Classification

Jijie Wu[1], Dongliang Chang[2], Aneeshan Sain[3], Xiaoxu Li[1], Zhanyu Ma[2], Jie Cao[1], Jun Guo[2], and Yi-Zhe Song[3]

[1]Lanzhou University of Technology, [2]Beijing University of Posts and Telecommunications, [3]SketchX, University of Surrey

Introduction

Problems:

1. A key consideration in few-shot fine-grained image classification is to learn feature representations with higher inter-class and lower intra-class variations, with a mere few labelled samples.
2. Prior works predominately use a support set to reconstruct the query image and then utilize metric learning to determine its category. Such unidirectional reconstruction methods only help to increase inter-class variations and are not effective in tackling intra-class variations.

Motivation:

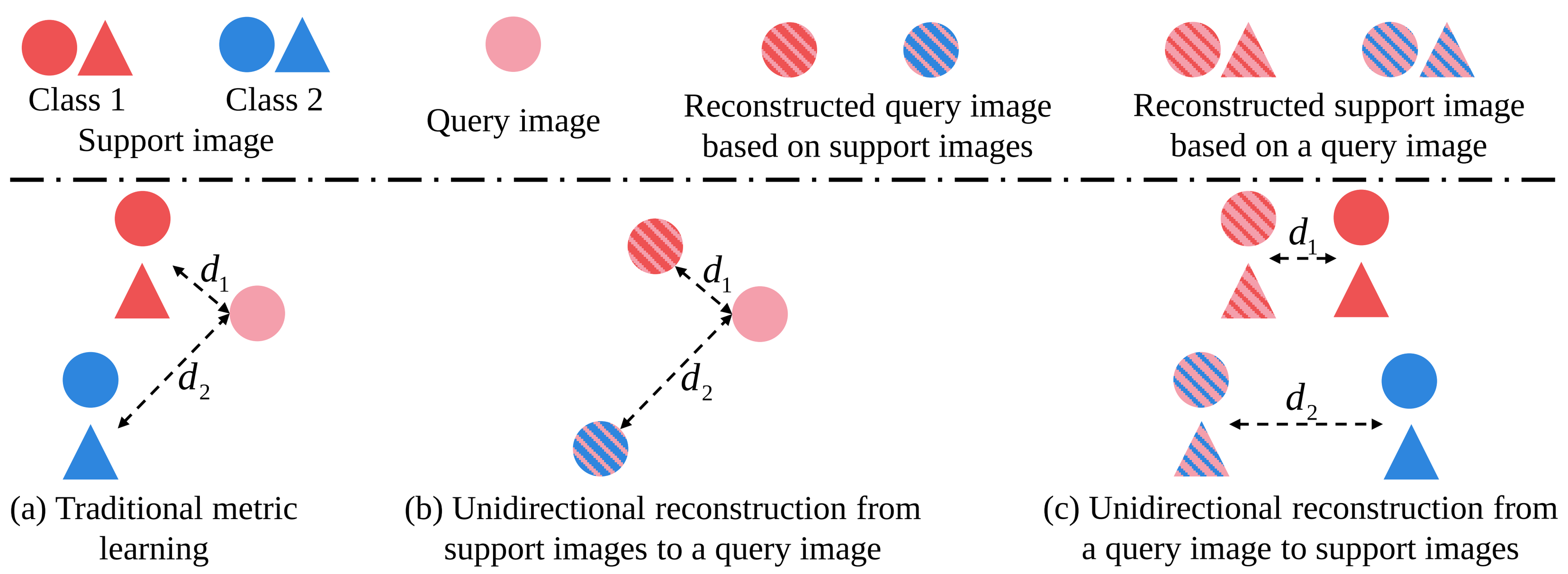

Fig. 1: Motivation of the proposed *Bi-directional Feature Reconstruction Network* (*Bi-FRN*). (a) is the traditional metric based method. (b) is the method proposed in FRN [1]. (b) + (c) is the method proposed in this paper. (b) can help the model increase the inter-class variations, and (c) can help the model decrease the intra-class variations.

In this paper, we for the first time introduce a bi-directional reconstruction mechanism for few-shot fine-grained classification. Instead of using the support set to reconstruct the query set to increase inter-class variations only (as shown in Fig.1 (b)), we additionally use the query set to reconstruct the support set to simultaneously reduce intra-class variations (as shown in Fig. 1 (c)). This modification might sound overly simple at first sight, it however importantly fulfills both desired learning outcomes for the fine-grained setting – support to query to encourage large inter-class variations, and query to support to encourage small intra-class variations.

Ref: [1] Wertheimer, D.; Tang, L.; and Hariharan, B. 2021. Few-shot Classification With Feature Map Reconstruction Networks. In CVPR.

The Proposed Method

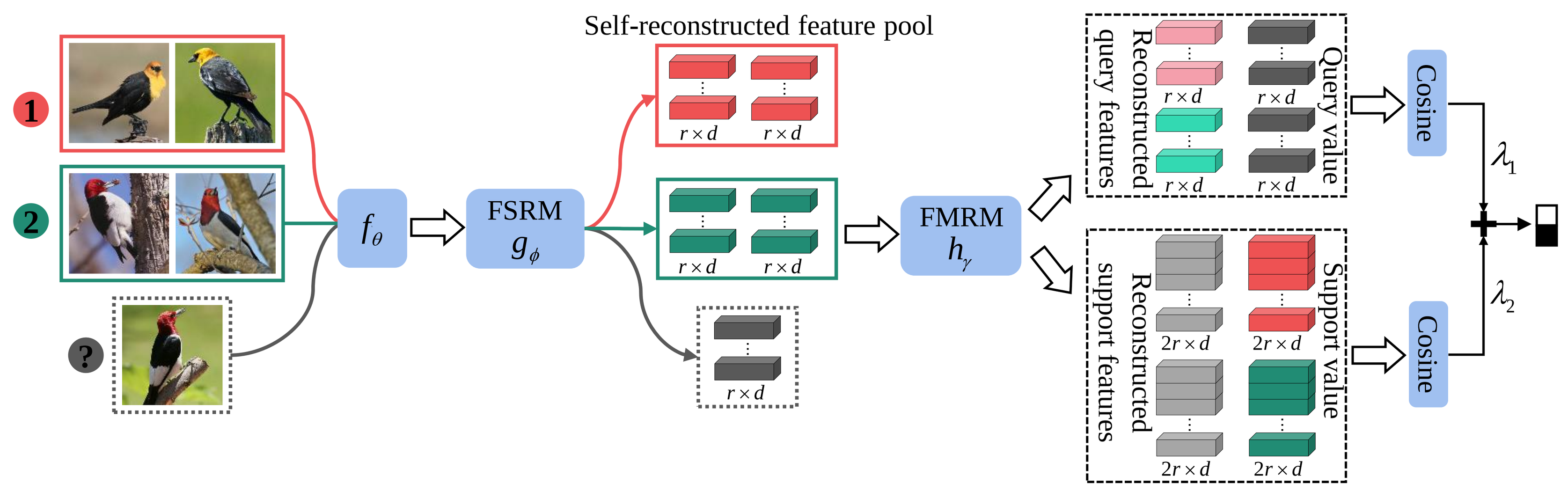

Fig. 2: The proposed Bi-Directional feature reconstruction network. FSRM refers to Feature Self-reconstruction Module and FMRM refers to Feature Mutual Reconstruction Module.

As shown in Fig.2, our model consists of four modules:

1. **Embedding module** f_θ for extracting deep convolutional image features.
2. **Feature self-reconstruction module** g_ϕ, in which the convolutional features of each image are reconstructed by themselves based on a self-attention mechanism.. This module can make the similar local features become more similar, while dissimilar ones even more dissimilar.
3. **Feature mutual reconstruction module** h_γ, which reconstructs sample features in a bidirectional form. This module not only uses the support sample to reconstruct the query sample but also reconstructs the support sample from the query sample.
4. **Euclidean metric module**, which is in charge of calculating the distance between origin sample and reconstructed sample.

Contributions

1. We reveal the key problem in current reconstruction-based few-shot fine-grained classification lies with its inability in minimising intra-class variations.
2. We for the first time propose a bi-directional reconstruction network that simultaneously increase inter-class variations while reducing intra-class variations.
3. Experimental results and ablative analyses on three fine-grained few-shot image datasets consistently demonstrate the superiority of the proposed method and reveal insights on why the bi-directional approach is effective.

Experimental Results

Method	*CUB* 1-shot	*CUB* 5-shot	*Dogs* 1-shot	*Dogs* 5-shot	*Cars* 1-shot	*Cars* 5-shot
ProtoNet† (NeurIPS 2017)	64.82±0.23	85.74±0.14	46.66±0.21	70.77±0.16	50.88±0.23	74.89±0.18
Relation (CVPR 2018)	63.94±0.92	77.87±0.64	47.35±0.88	66.20±0.74	46.04±0.91	68.52±0.78
DN4 (CVPR 2019)	57.45±0.89	84.41±0.58	39.08±0.76	69.81±0.69	34.12±0.68	87.47±0.47
PARN† (ICCV 2019)	74.43±0.95	83.11±0.67	55.86±0.97	68.06±0.72	66.01±0.94	73.74±0.70
SAML (ICCV 2019)	65.35±0.65	78.47±0.41	45.46±0.36	59.65±0.51	61.07±0.47	88.73±0.49
DeepEMD (CVPR 2020)	64.08±0.50	80.55±0.71	46.73±0.49	65.74±0.63	61.63±0.27	72.95±0.38
LRPABN (TMM 2021)	63.63±0.77	76.06±0.58	45.72±0.75	60.94±0.66	60.28±0.76	73.29±0.58
BSNet(D&C) (TIP 2021)	62.84±0.95	85.39±0.56	43.42±0.86	71.90±0.68	40.89±0.77	86.88±0.50
CTX† (NeurrIPS 2020)	72.61±0.21	86.23±0.14	57.86±0.21	73.59±0.16	66.35±0.21	82.25±0.14
FRN† (CVPR 2021)	74.90±0.21	89.39±0.12	60.41±0.21	79.26±0.15	67.48±0.22	87.97±0.11
FRN+TDM† (CVPR 2022)	72.01±0.22	89.05±0.12	51.57±0.23	75.25±0.16	65.67±0.22	86.44±0.12
Ours	**79.08±0.20**	**92.22±0.10**	**64.74±0.22**	**81.29±0.14**	**75.74±0.20**	**91.58±0.09**
ProtoNet† (NeurIPS 2017)	81.02±0.20	91.93±0.11	73.81±0.21	87.39±0.12	85.46±0.19	95.08±0.08
CTX† (NeurIPS 2020)	80.39±0.20	91.01±0.11	73.22±0.22	85.90±0.13	85.03±0.19	92.63±0.11
DeepEMD† (CVPR 2020)	75.59±0.30	88.23±0.18	70.38±0.30	85.24±0.18	80.62±0.26	92.63±0.13
FRN† (CVPR 2021)	84.30±0.18	93.34±0.10	76.76±0.21	88.74±0.12	88.01±0.17	95.75±0.07
FRN+TDM† (CVPR 2022)	85.15±0.18	93.99±0.09	**78.02±0.20**	**89.85±0.11**	88.92±0.16	96.88±0.06
Ours	**85.44±0.18**	**94.73±0.09**	76.89±0.21	88.27±0.12	**90.44±0.15**	**97.49±0.05**

Fig. 3: 5-way few-shot classification performance on the *CUB*, *Dogs* and *Cars* datasets. The top block uses Conv-4 backbone and the bottom block uses ResNet-12 backbone.

Ablation on Reconstruction Designs of FMRM.

Backbone	*Method*	*CUB* 1-shot	*CUB* 5-shot	*Dogs* 1-shot	*Dogs* 5-shot	*Cars* 1-shot	*Cars* 5-shot
Conv-4	Baseline (ProtoNet)	64.82±0.23	85.74±0.14	46.66±0.21	70.77±0.16	50.88±0.23	74.89±0.18
	Ours (Q→S)	**79.88±0.20**	91.76±0.11	**65.26±0.22**	80.81±0.14	75.61±0.20	90.49±0.10
	Ours (S→Q)	76.54±0.21	88.03±0.14	64.39±0.22	78.36±0.15	72.71±0.22	85.11±0.14
	Ours (Mutual)	79.08±0.20	**92.22±0.10**	64.74±0.22	**81.29±0.14**	**75.74±0.20**	**91.58±0.09**
ResNet-12	Baseline (ProtoNet)	81.02±0.20	91.93±0.11	73.81±0.21	87.39±0.12	85.46±0.19	95.08±0.08
	Ours (Q→S)	83.72±0.19	93.31±0.09	76.50±0.21	87.95±0.12	87.37±0.17	95.10±0.08
	Ours (S→Q)	81.72±0.19	90.83±0.11	75.62±0.22	86.47±0.13	85.90±0.18	93.17±0.10
	Ours (Mutual)	**85.44±0.18**	**94.73±0.09**	**76.89±0.21**	**88.27±0.12**	**90.44±0.15**	**97.49±0.05**

Fig. 4: Ablation on reconstruction designs of FMRM.

Feature Visualization

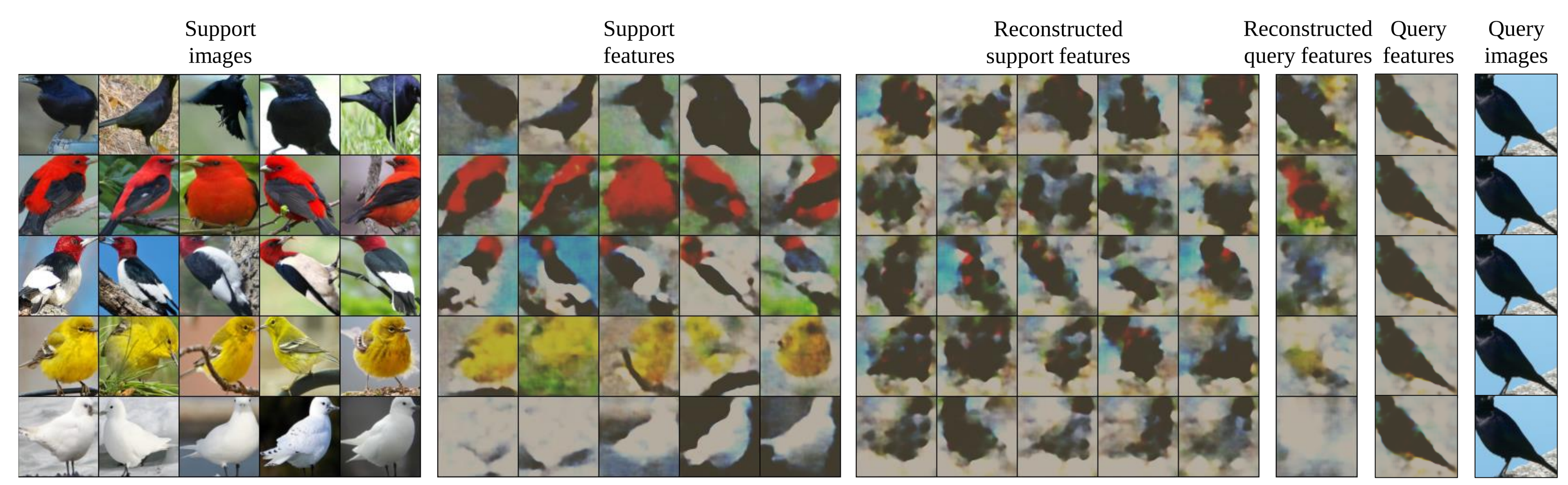

Fig. 5: Recovered images of different features by our method for the CUB dataset.

Open Sources

- **Code:** `https://github.com/PRIS-CV/Bi-FRN`
- **Email:** `{jijie,lixiaoxu}@lut.edu.cn`