BSNet: Bi-Similarity Network for Few-shot Fine-grained Image Classification (TIP 2021)

Xiaoxu Li*, Jijie Wu*, Zhuo Sun‡, Zhanyu Ma†, Jie Cao*, and Jing-Hao Xue‡

*Lanzhou University of Technology, †Beijing University of Posts and Telecommunications, ‡University College London

杭州 VALSE 2021
August 18 –20, 2021

Introduction

Problems:

- A key consideration in few-shot fine-grained image classification is how to learn discriminative features from few labeled images.
- Features obtained by adapting a single similarity metric are only discriminative in a single feature space. That is, using one single similarity measure may induce certain similarity bias that lowers the generalization ability of the model, in particular when the amount of training data is small.

Motivation:

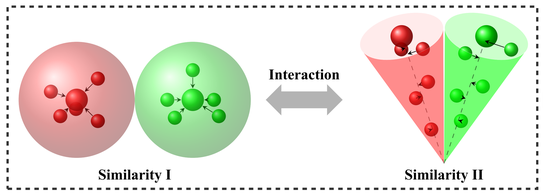

Fig. 1: Motivation of the proposed *Bi-Similarity Network* (*BSNet*). Here we use the Euclidean distance and the cosine distance as the similarity measures in feature spaces. The Euclidean distance and the cosine distance are for Similarity I and Similarity II, respectively. Different colors indicate different predicted labels; larger balls are class prototypes.

- If the obtained features can simultaneously adapt two similarity measures of diverse characteristics, the samples within one class can be mapped more compactly into a smaller feature space. This will result in a model embedding two diverse similarity measures, generating more discriminative features than using a single measure.

Contributions

- We propose a BSNet that leverages two similarity measures and significantly improves the performance on four fine-grained image datasets.
- We demonstrate that the model complexity of BSNet is less than the mean value of model complexities of two single-similarity networks, even though BSNet contains more model parameters.

The proposed method: BSNet

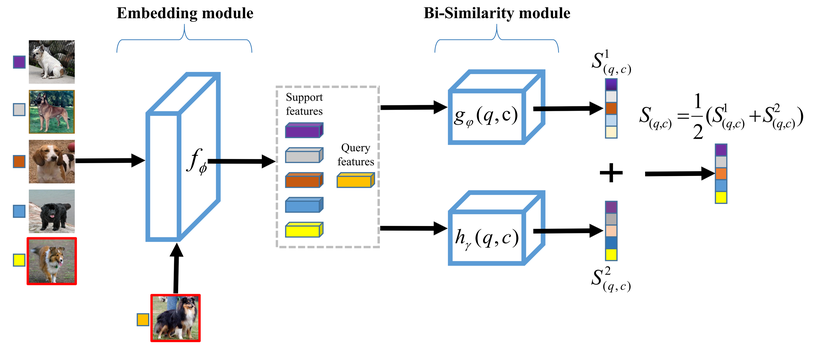

Fig. 2: Illustration of the proposed *Bi-Similarity Network* (*BSNet*). It consists of one embedding module f_ϕ, followed by a bi-similarity module which outputs two similarity scores between a query image and C class prototypes in C-way K-shot problems. $S^k_{(q,c)}$ denotes the kth similarity score between the qth query image x_q and the cth class in a task.

$$S^1_{(q,c)} = g_\varphi\left([\frac{1}{K}\sum_{s=1}^{K} f_\phi(x^{(i)}_{s,c}) || f_\phi(x^{(i)}_q)]\right),\ c = 1, \cdots C. \quad (1)$$

$$S^2_{(q,c)} = h^{cos}_\gamma\left(h^{em}_\gamma(\frac{1}{K}\sum_{s=1}^{K} f_\phi(x^{(i)}_{s,c})), h^{em}_\gamma(f_\phi(x^{(i)}_q))\right). \quad (2)$$

$$\hat{y}^{(i)}_q = [0, \cdots, \underbrace{1}_{\arg\max_c \frac{1}{2}\left(S^1_{(q,c)} + S^2_{(q,c)}\right), c \in \{1, \cdots, C\}}, \cdots, 0]. \quad (3)$$

Experimental results

Model	5-Way 5-shot Accuracy (%)		5-Way 1-shot Accuracy (%)	
	Dogs	*CUB*	*Dogs*	*CUB*
Matching	59.79 ± 0.72	74.57 ± 0.73	**46.10 ± 0.86**	60.06 ± 0.88
BSNet (M&C)	**61.61 ± 0.69**	**74.68 ± 0.71**	45.91 ± 0.81	**60.73 ± 0.94**
Prototype	61.58 ± 0.71	75.06 ± 0.67	40.81 ± 0.83	50.67 ± 0.88
BSNet (P&C)	**62.61 ± 0.73**	**76.34 ± 0.65**	**43.13 ± 0.85**	**55.81 ± 0.97**
Relation	66.20 ± 0.74	77.87 ± 0.64	47.35 ± 0.88	63.94 ± 0.92
BSNet (R&C)	**68.60 ± 0.73**	**80.99 ± 0.63**	**51.06 ± 0.94**	**65.89 ± 1.00**
DN4	69.81 ± 0.69	84.41 ± 0.58	39.08 ± 0.76	57.45 ± 0.89
BSNet (D&C)	**71.90 ± 0.68**	**85.39 ± 0.56**	**43.42 ± 0.86**	**62.84 ± 0.95**

Table 1: Five-way few-shot classification performance.

Ablation study

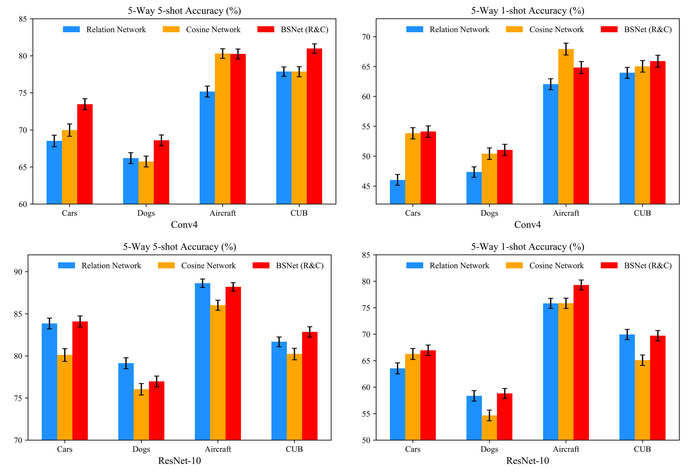

Fig. 3: Ablation study: effectiveness of the two-branch similarity.

Feature visualization

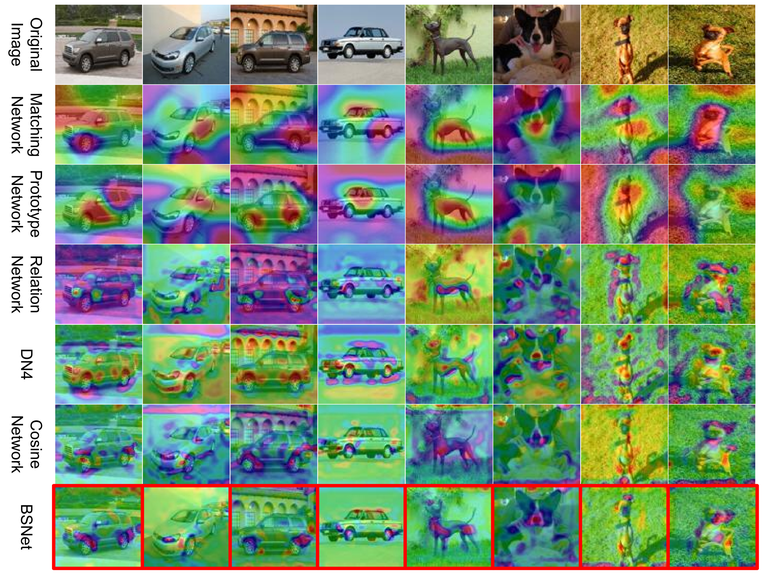

Fig. 4: Feature visualization of different methods. The redder the region, the more class-discriminative it is.

Open sources

- **Paper:** https://ieeexplore.ieee.org/document/9293172
- **Code:** https://github.com/PRIS-CV/BSNet