TAAN: Task-Aware Attention Network for Few-shot Classification

Zhe Wang, Li Liu and FanZhang Li
Soochow University

Introduction

Metric-based few-shot learning

- classify unseen data instances into a set of new categories, given just a small number of labeled instances in each class

- represent image data in an appropriate feature space and use a distance metric to predict image labels

Challenge

- extract features from samples independently

- few pay enough attention to the discriminability of the extracted features

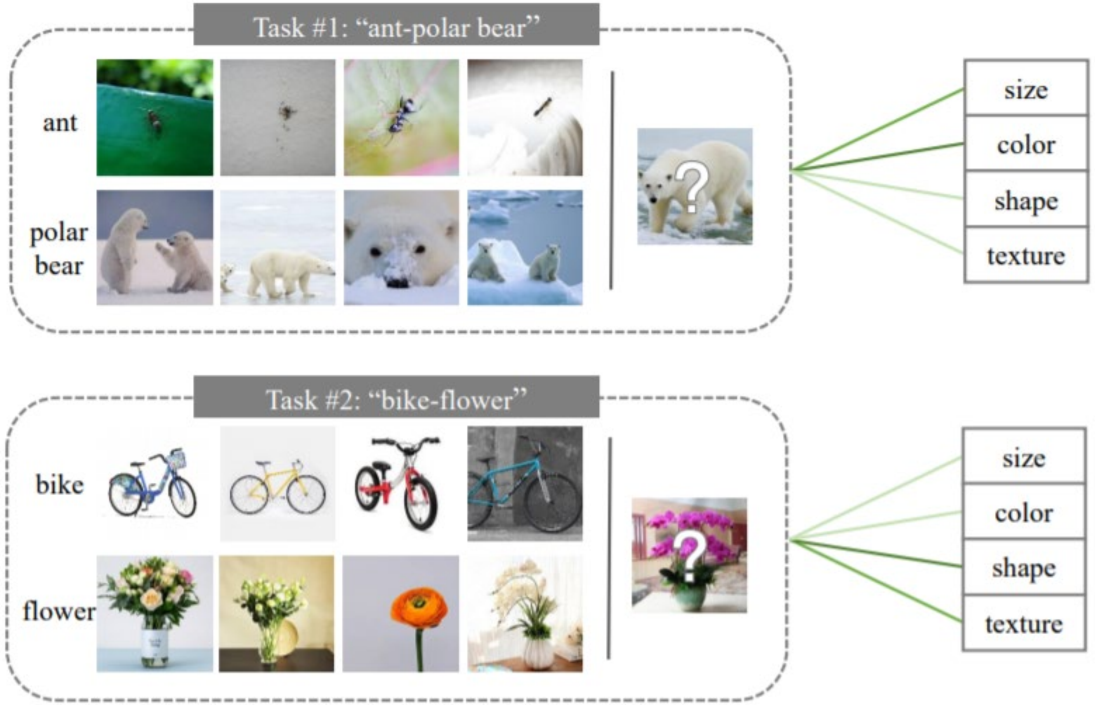

Fig. 1. A motivating example of two different 2-way 4-shot classification tasks, with each instances represented as four simple features. The task consists of a support set (left, 4 labeled examples per class) and a query example (right). Intuitively, we find size and color are key features for task (1), whereas shape and texture are key features for task (2). However, few of previous approaches consider the discriminability of extracted features.

Methodology

Task-Relevant channel Attention module

- enhance the feature discriminability
- The ***Squeeze Layer*** form a channel descriptor which represented as the instance-level channel context
- The ***Attention Transformer*** transform the channel descriptor into task-level channel attention by looking at the whole support set.

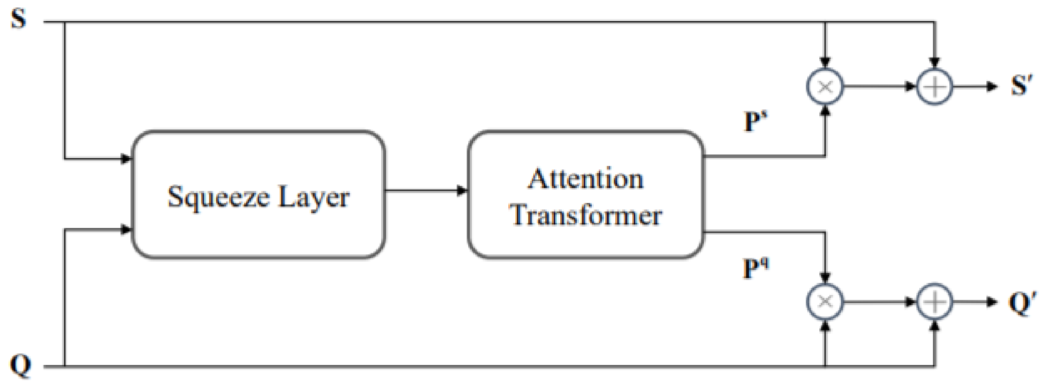

Fig. 2. An illustration of the proposed *Task-Relevant Channel Attention Module* (TRCAM). The *Squeeze Layer* firstly generates the channel representation for each instance, which is treated as instance-level channel attention. Then, the *Attention ansformer* helps to transform channel representation to task-level channel attention by looking at all support classes.

Methodology

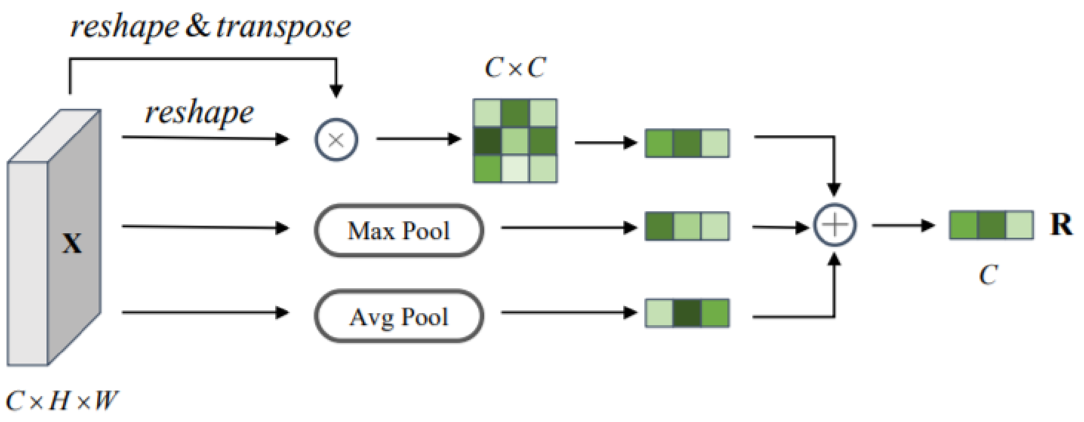

Fig. 3. An illustration of the proposed *Squeeze Layer*.

Squeeze Layer

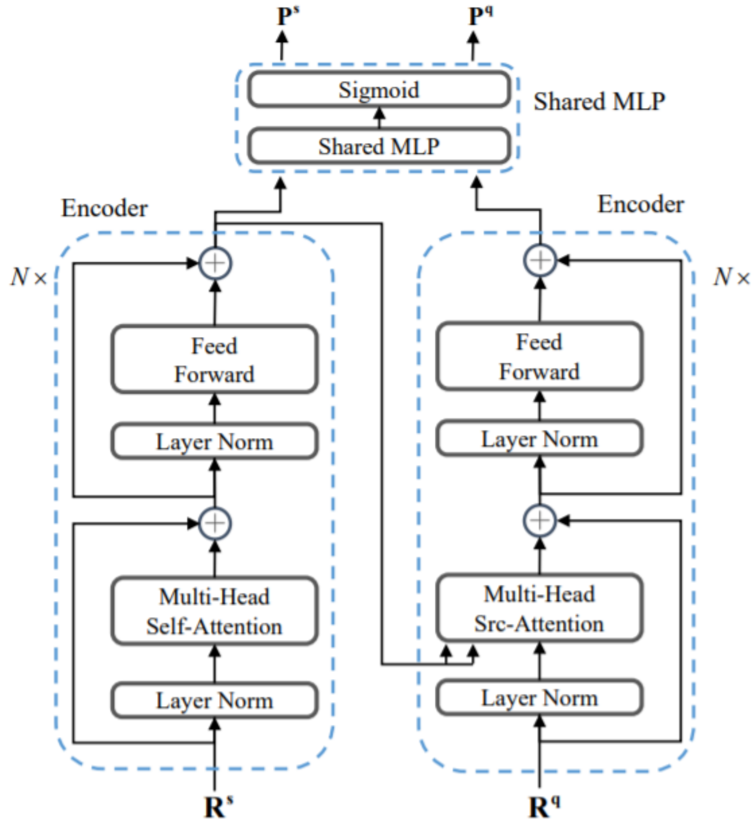

Fig. 4. An illustration of the proposed *Attention Transformer*.

Attention Transformer

Methodology

Task-Aware Attention Network

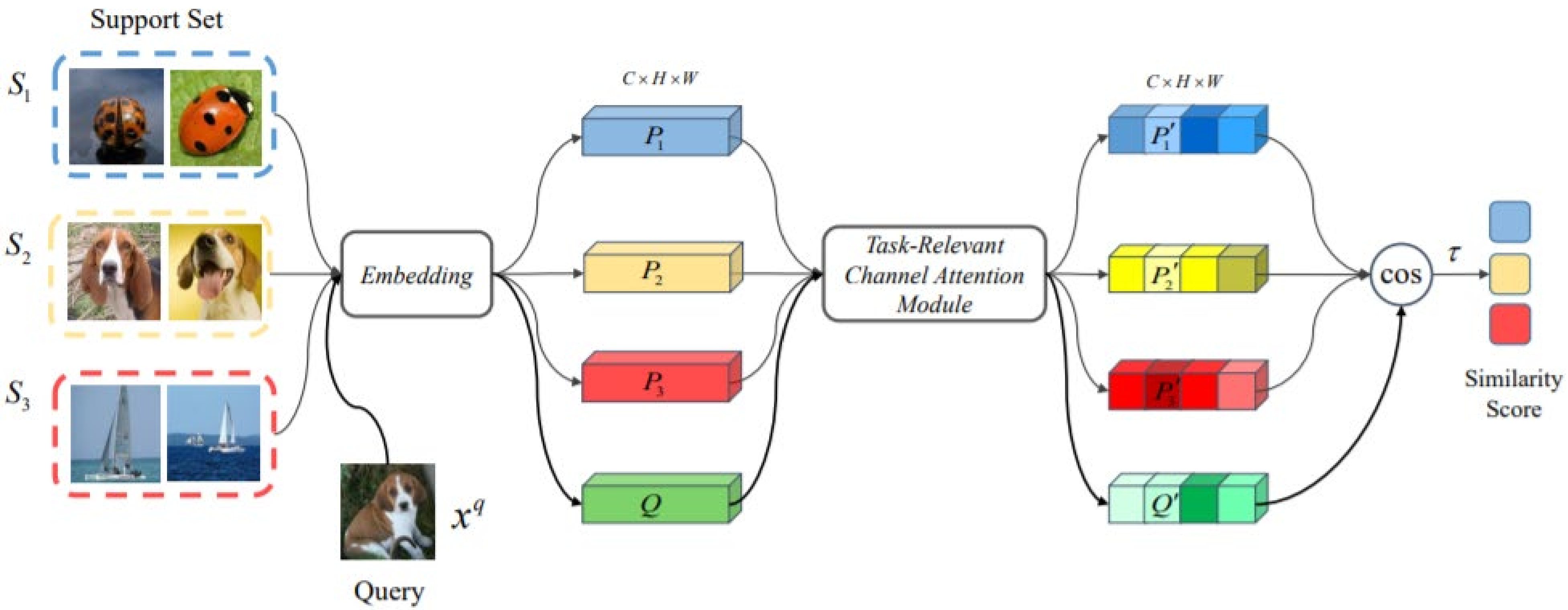

Fig. 5. The framework of the proposed TAAN approach.

Experiments

Results on two benchmark datasets

TABLE I
FEW-SHOT CLASSIFICATION ACCURACIES ON MINI-IMAGENET AND TIERED-IMAGENET. AVERAGE 5-WAY ACCURACY (%) IS REPORTED WITH 95% CONFIDENCE INTERVAL. ('-' NOT REPORTED)

Model	Backbone	mini-ImageNet		tierd-ImageNet	
		1-shot	5-shot	1-shot	5-shot
Model-based					
Meta-learner LSTM [2]	ConvNet-4	43.44 ± 0.77	60.60 ± 0.71	-	-
Meta-SGD [27], reported by [19]	ConvNet-4	54.24 ± 0.03	70.86 ± 0.04	62.95 ± 0.03	79.34 ± 0.06
SNAIL [15]	ResNet-12	55.71 ± 0.99	68.88 ± 0.92	-	-
Optimization-based					
MAML [16]	ConvNet-4	48.70 ± 1.75	63.15 ± 0.91	51.67 ± 1.81	70.30 ± 1.75
Cosine Classifier [4]	ResNet-12	55.43 ± 0.81	77.18 ± 0.61	61.49 ± 0.91	82.37 ± 0.67
MTL [28]	ResNet-12	61.20 ± 1.80	75.50 ± 0.80	-	-
LEO [19]	WRN-28-10	61.76 ± 0.08	77.59 ± 0.12	66.33 ± 0.05	81.44 ± 0.09
MetaOptNet [20]	ResNet-12	62.64 ± 0.61	78.63 ± 0.46	65.99 ± 0.72	81.56 ± 0.53
Metrics-based					
Matching network [8]	ConvNet-4	43.56 ± 0.84	55.31 ± 0.73	-	-
Relation network [10], reported by [20]	ConvNet-4	50.44 ± 0.82	65.32 ± 0.70	54.48 ± 0.93	71.32 ± 0.78
Prototypical network [9], reported by [20]	ConvNet-4	49.42 ± 0.78	68.20 ± 0.66	53.31 ± 0.89	72.69 ± 0.74
TADAM [21]	ResNet-12	58.5 ± 0.3	76.7 ± 0.3	-	-
Cross Attention network [29]	ResNet-12	63.85 ± 0.48	79.44 ± 0.34	69.89 ± 0.51	84.23 ± 0.37
CTM [30]	ResNet-18	62.05 ± 0.55	78.63 ± 0.06	64.78 ± 0.11	81.05 ± 0.52
CTM, data augment [30]	ResNet-18	64.12 ± 0.82	80.51 ± 0.13	68.41 ± 0.39	84.28 ± 1.73
Baselines					
Classifier-Baseline [31]	ResNet-12	58.91 ± 0.23	77.76 ± 0.17	68.07 ± 0.26	83.74 ± 0.18
Meta-Baseline [31]	ResNet-12	63.17 ± 0.23	79.26 ± 0.17	68.62 ± 0.27	83.29 ± 0.18
Prototypical Network (re-implement)	ResNet-12	59.88 ± 0.45	78.50 ± 0.33	65.65 ± 0.33	83.40 ± 0.13
Ours: TAAN	ResNet-12	**66.48 ± 0.45**	**81.40 ± 0.31**	**70.07 ± 0.51**	**84.56 ± 0.35**

Ablation studies

Effectiveness of our TRACM

TABLE II
ABLATION EXPERIMENTS ON TRCAM. ('✓'WITH; '-' WITHOUT)

	SL			AT	1-shot	5-shot
	self-attention	avg-pool	max-pool			
(i)	-	-	-	-	62.67	78.20
(ii)	✓	✓	✓	-	63.56	79.56
(iii)	✓	-	-	✓	65.79	81.05
	-	✓	-	✓	65.83	81.02
	-	-	✓	✓	65.38	81.11
(iv)	✓	✓	✓	✓	66.59	81.54

Ablation studies

Visualization with Class Activation Mapping

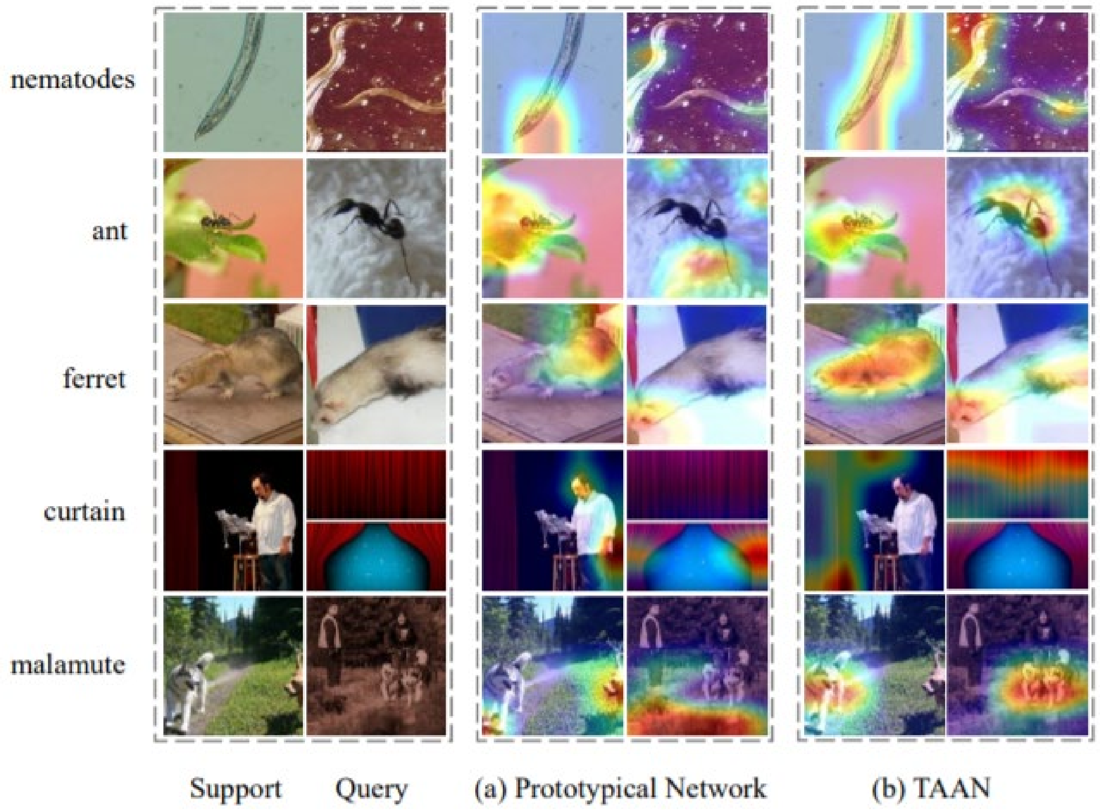

Fig. 6. Grad-CAM visualization on a 5-way 1-shot task with one query sample per class. Our method covers the target object regions better than Prototypical Network.

Thank you