Learning Task-aware Local Representations for Few-shot Learning

Chuanqi Dong, Wenbin Li, Jing Huo, Zheng Gu, Yang Gao   Nanjing University, China

Introduction

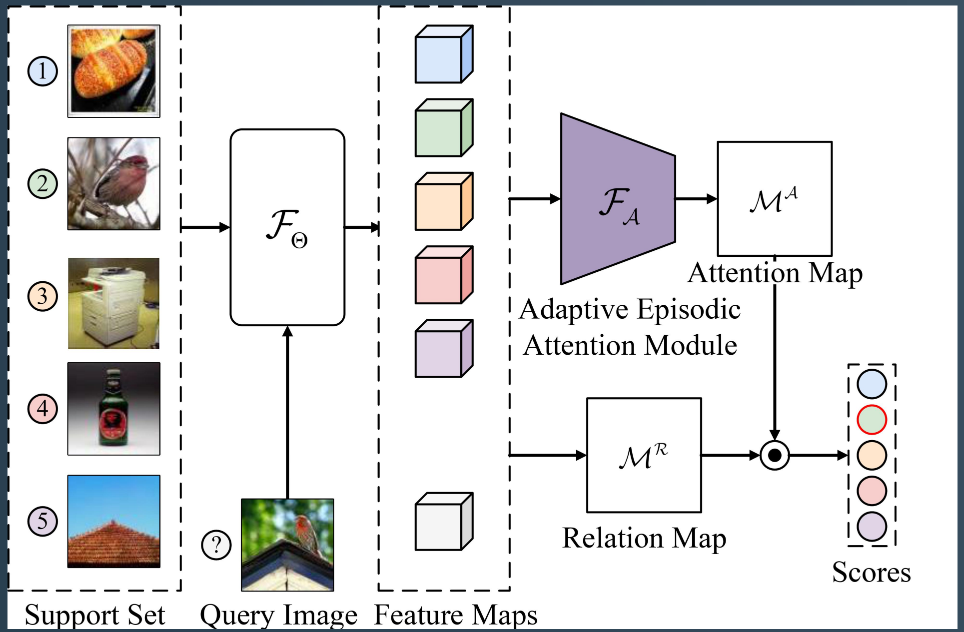
Few-shot Learning:

Extreme few training samples (e.g., 1 or 5) for each category

Key Problems:

- How to transfer information or knowledge?
- How to select important information among the task?
- How to measure the similarity between the query image and classes?

Proposed Method



Our Contributions:

- Propose a novel attention mechanism to explore discriminative semantic patches.
- Develop a trainable module to give the model the adaptive capacity.
- Achieve SOTA results on the *mini*Imagenet and three fine-grained datasets.

ATL-Net

Task-aware Local Representations:

$$\mathcal{M}^\mathcal{A}_{i,j} = \frac{I(\mathcal{M}^\Psi_{i,j})}{\sum_j I(\mathcal{M}^\Psi_{i,j})}$$

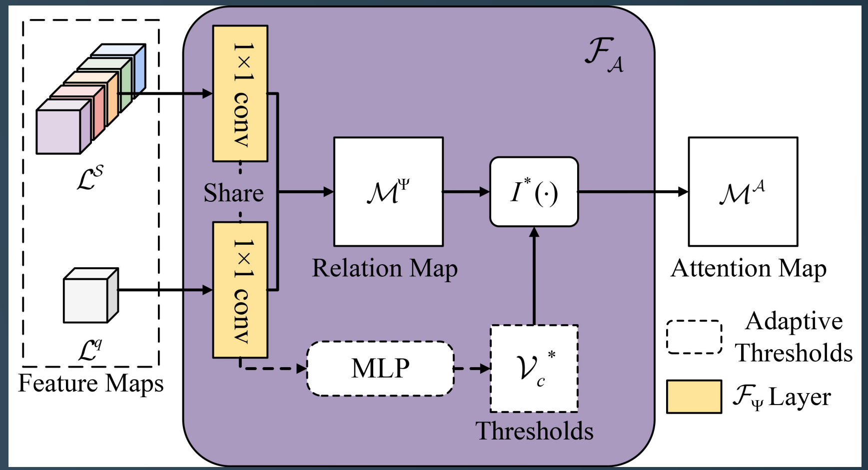
I is the step function, $\mathcal{M}_{i,j}$ is the local representations,

Adaptive Threshold for Episodic Attention

$$\mathcal{V}^*_c = \sigma(\mathcal{F}_\Gamma(\mathcal{L}^q_i))$$

$$I^*(x) = 1/(1+\exp^{-k(x-\mathcal{V}^*_c)})$$

σ is a sigmoid function



Results

Experiment Results:

Model	Backbone	Additional Stage	5-way 1-shot	5-way 5-shot
Matching Net [Vinyals *et al.*, 2016]	Conv-64F	N	43.56 ± 0.84	55.31 ± 0.73
MAML [Finn *et al.*, 2017]	Conv-32F	Y	48.70 ± 1.84	63.11 ± 0.92
Prototypical Net [Snell *et al.*, 2017]	Conv-64F	N	49.42 ± 0.78	68.20 ± 0.66
GNN [Satorras and Estrach, 2018]	Conv-256F	N	50.33 ± 0.36	66.41 ± 0.63
Relation Net [Sung *et al.*, 2018]	Conv-64F	N	50.44 ± 0.82	65.32 ± 0.70
MetaGAN [Zhang *et al.*, 2018]	Conv-64F	N	52.71 ± 0.64	68.63 ± 0.67
MM-Net [Cai *et al.*, 2018]	Conv-64F	N	$\underline{53.37 \pm 0.48}$	66.97 ± 0.35
MEPS [Chu *et al.*, 2019]	Conv-64F	N	51.03 ± 0.78	67.96 ± 0.71
CovaMNet [Li *et al.*, 2019c]	Conv-64F	N	51.19 ± 0.76	67.65 ± 0.63
DN4 [Li *et al.*, 2019b]	Conv-64F	N	51.24 ± 0.74	71.02 ± 0.64
GCR [Li *et al.*, 2019a]	Conv-64F	Y	53.21 ± 0.40	$\underline{72.34 \pm 0.32}$
ATL-Net (Ours)	Conv-64F	N	$\mathbf{54.30 \pm 0.76}$	$\mathbf{73.22 \pm 0.63}$

Table 2: Comparisons with other methods on *mini*Imagenet. The second column shows which kind of embedding module is employed. The third column denotes whether the model contains additional training stage, *e.g.* pretrain stage or fine-tune stage. We use the officially provided results for all the other methods. For each setting, the best and the second best results are highlighted.

Model	Stanford Dogs		Stanford Cars		CUB-200	
	5-way 1-shot	5-way 5-shot	5-way 1-shot	5-way 5-shot	5-way 1-shot	5-way 5-shot
Matching Net	35.80 ± 0.99	47.50 ± 1.03	34.80 ± 0.98	44.70 ± 1.03	45.30 ± 1.03	59.50 ± 1.01
Prototypical Net	37.59 ± 1.00	48.19 ± 1.03	40.90 ± 1.01	52.93 ± 1.03	37.36 ± 1.00	45.28 ± 1.03
GNN	46.98 ± 0.98	62.27 ± 0.95	55.85 ± 0.97	71.25 ± 0.89	51.83 ± 0.98	63.69 ± 0.94
DN4	45.41 ± 0.76	$\underline{63.51 \pm 0.62}$	$\underline{59.84 \pm 0.80}$	$\underline{88.65 \pm 0.44}$	46.84 ± 0.81	$\underline{74.92 \pm 0.64}$
CovaMNet	$\underline{49.10 \pm 0.76}$	63.04 ± 0.65	56.65 ± 0.86	71.33 ± 0.62	$\underline{52.42 \pm 0.76}$	63.76 ± 0.64
PABN+$_{cpt}$	45.65 ± 0.71	61.24 ± 0.62	54.44 ± 0.71	67.36 ± 0.61	-	-
LRPABN$_{cpt}$	45.72 ± 0.75	60.94 ± 0.66	60.28 ± 0.76	73.29 ± 0.58	-	-
ATL-Net (Ours)	$\mathbf{54.49 \pm 0.92}$	$\mathbf{73.20 \pm 0.69}$	$\mathbf{67.95 \pm 0.84}$	$\mathbf{89.16 \pm 0.48}$	$\mathbf{60.91 \pm 0.91}$	$\mathbf{77.05 \pm 0.67}$

Table 3: Comparisons with other methods on three fine-grained datasets. We adopt the results from [Li *et al.*, 2019c] for the first three methods and the officially provided results for the other methods. For each setting, the best and the second best results are highlighted.

Ablation Study:

Factor	5-way 1-shot	5-way 5-shot
(i) baseline	50.94 ± 0.79	65.16 ± 0.72
(ii) w/o $\mathcal{F}_\Gamma$ (TL-Net)	53.24 ± 0.80	71.87 ± 0.65
(iii) w/o $\mathcal{F}_\Psi$	$\underline{53.80 \pm 0.81}$	$\underline{72.95 \pm 0.64}$
ATL-Net (Ours)	$\mathbf{54.30 \pm 0.76}$	$\mathbf{73.22 \pm 0.63}$

Table 4: Ablation study on *mini*Imagenet for the proposed ATL-Net.

Model	Params	5-way 5-shot
Prototypical Net	**0.113M**	68.20
Relation Net	0.229M	65.32
GNN	1.619M	66.41
DN4	**0.113M**	71.02
GCR w/o Hallucinator	1.755M	72.34
ATL-Net (Ours)	0.117M	**73.22**

Table 5: The number of trainable parameters along with 5-way 5-shot performance of different models.

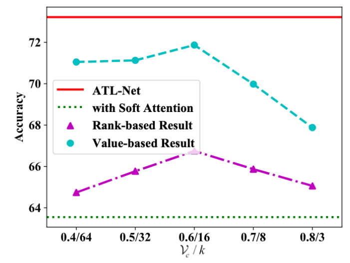


Figure 3: The results of the rank-based and value-based selections under the 5-way 5-shot setting.