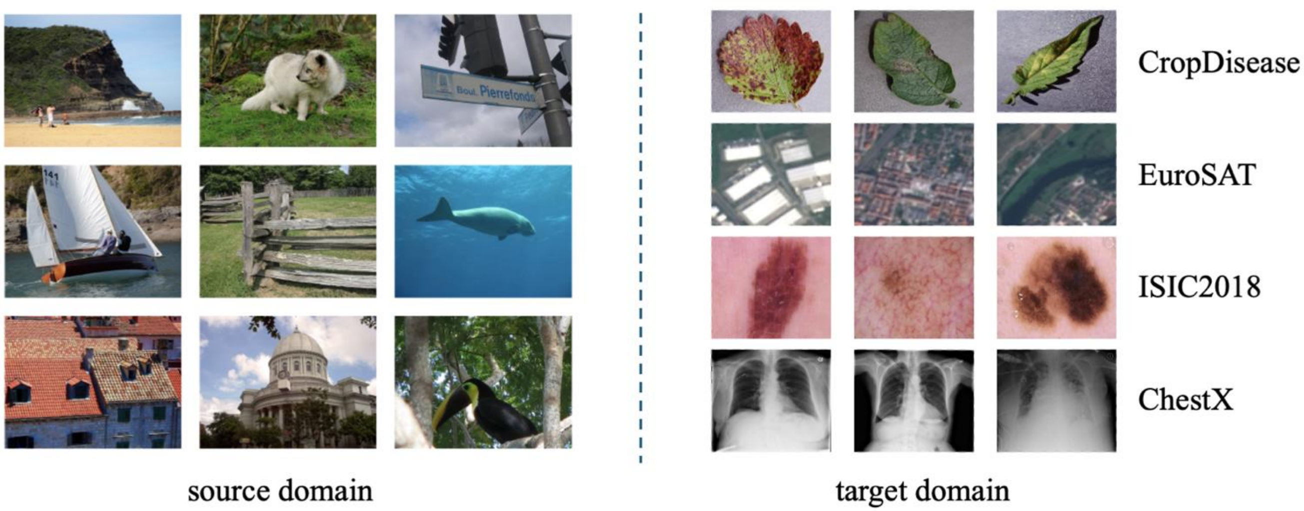
Attention Temperature Matters in ViT-Based Cross-Domain Few-Shot Learning

Yixiong Zou , Ran Ma , Yuhua Li and Ruixuan Li
School of Computer Science and Technology, Huazhong University of Science and Technology
{yixiongz, ranma, idcliyuhua, rxli}@hust.edu.cn

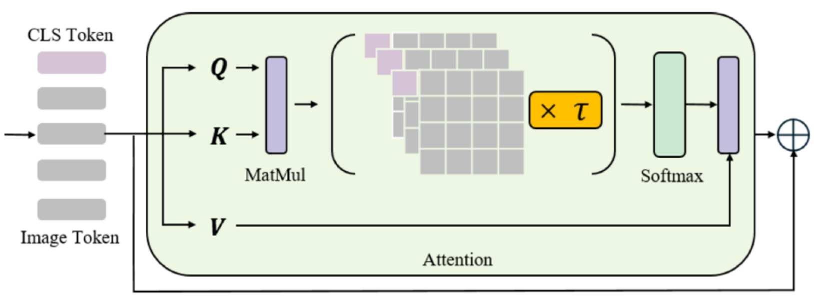
Cross-Domain Few-Shot Learning (CDFSL)

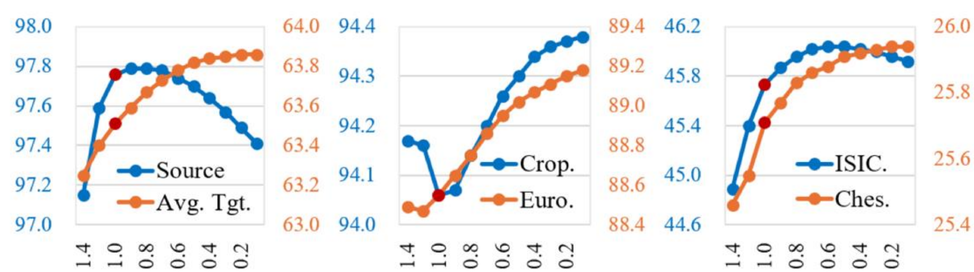
- Setting
  - Large-scale source-domain dataset
  - Few-shot target-domain datasets
- Task
  - Recognize target domain samples
- Key
  - Large domain gap + Sample scarcity



An Intriguing Phenomenon

- Target-domain performance consistently increases
  - By simply multiplying a temperature (even as small as 0)
- Contribution
  - Unveil the importance of the attention temperature in ViT-based CDFSL methods
  - Find the query-key attention mechanism shows limited transferability against large domain gaps
  - Propose a CDFSL method to boost ViT's transferability





Preliminaries

- Task definition
  - Source dataset:$D^S = \{x_i^S, y_i^S\}_{i=1}^N$; target dataset: $D^T = \{x_i^T, y_i^T\}_{i=1}^{N'}$
  - In target dataset, learn from a support set $\{x_{ij}, y_i\}_{i=1,j=1}^{k,n}$, evalute on a query set $\{x_q\}$
- ViT-based backbne

$$f(x_i^S) = M(A(M(\cdots A(E(x_i^S))\cdots)))$$

$$L = L_{cls}(\phi(f(x_i^S)), y_i^S)$$

- Baseline
  - Source Domain:
    - Cross entropy
  - Target Domain:
    - Fix backbone, prototype-based classification

Interpreting the phenomenon

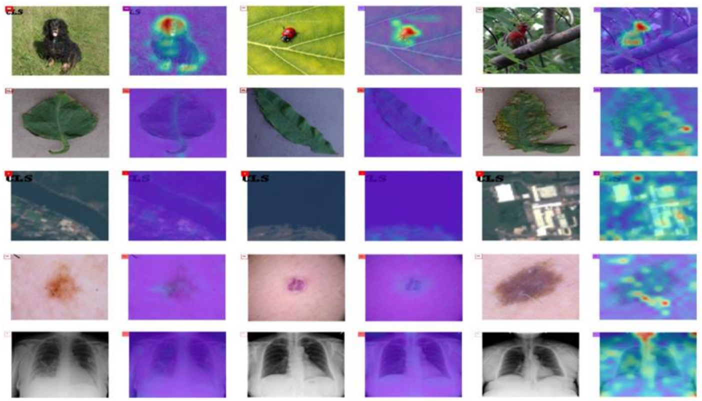
- Attention Temperature Remedies Target-Domain Attentions
  - Intuitive Observation of Ineffective Target-Domain Attentions
    - Source-domain-trained ViT focus on the CLS token and ignores all image tokens
    - Tends to focus on a large range of noisy regions instead of meaningful objects
  - Quantitative Verification of Target-Domain Attentions' Ineffectiveness
    - the attention value on the CLS token

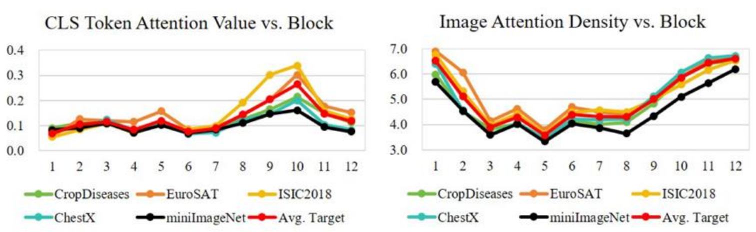
$$V(A) = \frac{1}{b}\frac{1}{n_h}\frac{1}{n_t}\sum_{i,j,k} r(A_{i,j,k})_{[0]}$$

    - the sparsity of the attention on image tokens

$$norm(A) = \frac{1}{b}\frac{1}{n_h}\frac{1}{n_t}\sum_{i,j,k} L_1(r(A_{i,j,k})_{[1:]})$$

    - Curves of the source dataset are always located under those of target datasets
    - Temperature adjustment as a remedy for ineffective target-domain attention





Interpretion

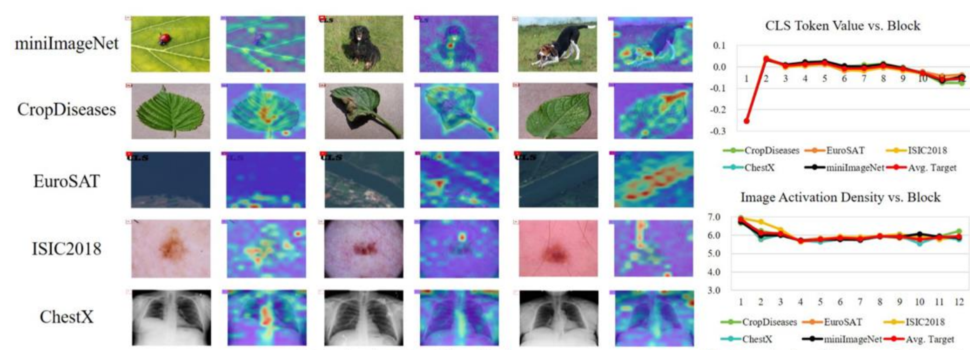
- Why do attention networks get ineffective on target domains?
  - Ineffective target-domain attention is majorly caused by the self-attention mechanism in the attention network
    - The SA mechanism is more on the side of discriminability than transferability
    - The default query-key relation contains the most domain information and discriminability
  - Handle the Ineffective Target-Domain Attention
    - Non-query-key features tend to be transferable but less discriminative
    - Encourage the learning of the non-query-key parameters in ViT and resist the learning of the query-key parts

Table 1: Ablation of the attention network from ViT's last block.

Method	*mini*ImageNet	CropDiseases	EuroSAT	ISIC2018	ChestX	Average
Input Tokens	90.17	79.63	73.12	32.81	22.41	51.99
Input Tokens + SA	92.59	79.10	73.17	32.54	22.47	51.82
Input Tokens + Identity SA	87.45	77.97	69.89	32.15	22.52	50.63
Input Tokens + Cosine SA	88.80	79.98	74.35	32.65	22.57	52.39
Input Tokens + Average SA	89.53	80.73	74.59	32.04	22.64	52.50

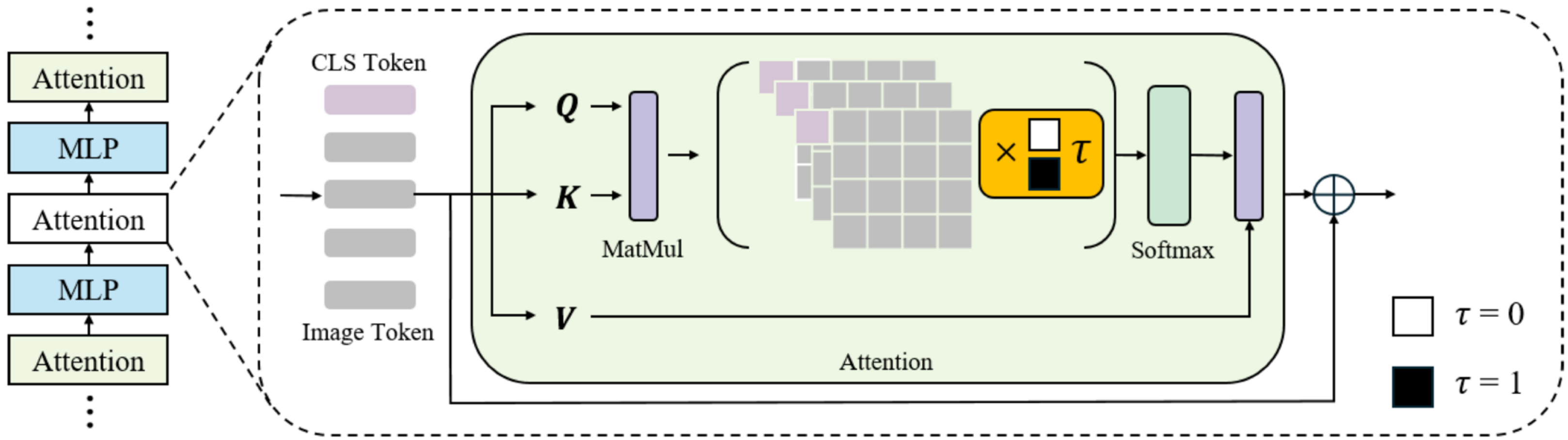
Table 2: Domain similarity w.r.t. ablated attention modules.

Method	*mini*ImageNet	CropDiseases	EuroSAT	ISIC2018	ChestX	Average
Input Tokens	1.0	0.4569	0.4381	0.3608	0.3900	0.4115
Input Tokens + SA	1.0	0.1853	0.1829	0.1344	0.1998	0.1756
Input Tokens + Identity SA	1.0	0.5857	0.5873	0.5376	0.4836	0.5486
Input Tokens + Cosine SA	1.0	0.2692	0.2252	0.1616	0.2295	0.2214
Input Tokens + Average SA	1.0	0.2235	0.2226	0.1580	0.2002	0.2011



(a) Activation Map Visualization (b) Quantitative Measure of Activation

Method



- Boost ViT's transferability
  - Source-Domain Attention Abandonment
    - Stochastically abandon the query-key attention by multiplying a temperature of 0
    - Resist the learning of the query-key attention parameters
  - Target-Domain Attention Adjustment
    - Multipling a pre-difined hyper-parameter
    - Alleviate the influence of ineffective attention maps

Experiments

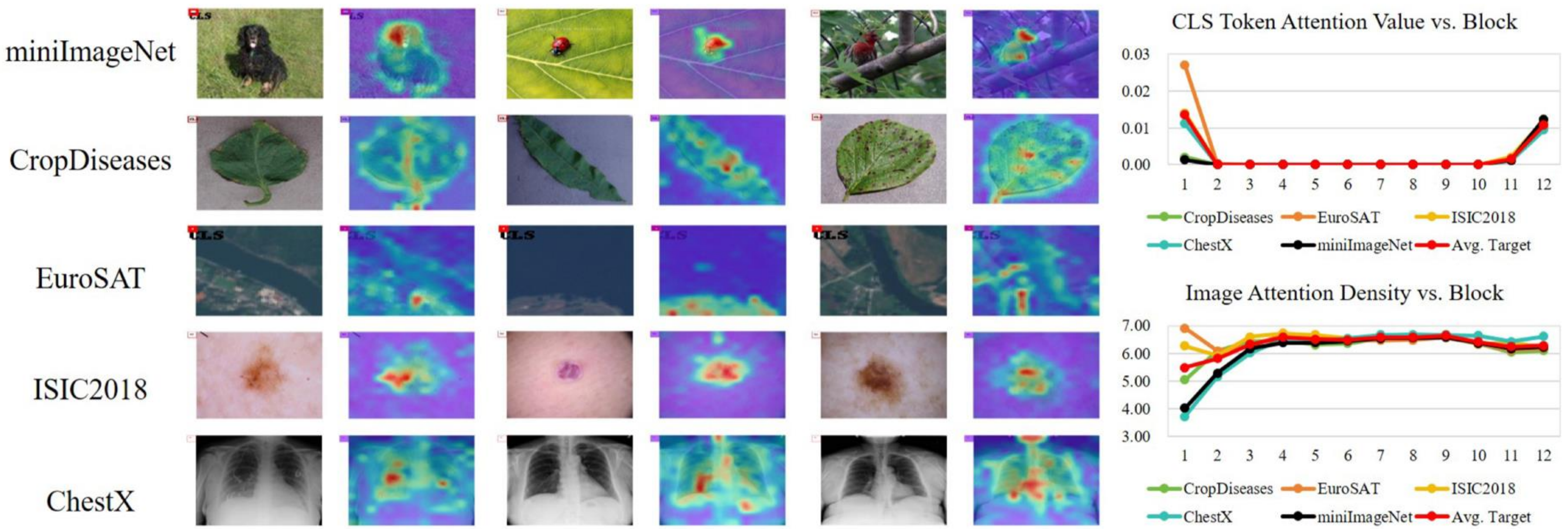
- State-of-the-art performance

Table 4: Comparison with state-of-the-art works by the 5-way 5-shot classification.

Methods	backbone	FT	Mark	Crop.	Euro.	ISIC.	Ches.	Ave.
LDP-net [50]	ResNet10	×	CVPR-23	89.40	82.01	48.06	26.67	61.29
GNN+AFA [15]	ResNet10	×	ECCV-22	88.06	85.58	46.01	25.02	61.67
SDT [29]	ResNet10	×	NN-24	90.27	82.02	48.66	27.20	62.04
FLoR [53]	ResNet10	×	CVPR-24	91.25	80.87	51.44	26.70	62.32
MEM-FS [42]	ViT-S	×	TIP-23	93.74	86.49	47.38	26.67	63.57
StyleAdv [10]	ViT-S	×	CVPR-23	94.85	88.57	47.73	26.97	64.53
MICM [49]	ViT-S	×	MM-24	94.61	90.08	46.85	27.11	64.66
SDT [29]	ViT-S	×	NN-24	95.00	89.60	47.64	26.72	64.75
FLoR [53]	ViT-S	×	CVPR-24	95.28	**90.41**	49.52	26.71	65.48
AttnTemp	ViT-S	×	**Ours**	**95.53**	<u>90.13</u>	**53.09**	**27.72**	**66.62**
FLoR [53]	ResNet10	✓	CVPR-24	92.33	83.06	**56.74**	26.77	64.73
PMF [14]	ViT-S	✓	CVPR-22	92.96	85.98	50.12	27.27	64.08
StyleAdv [10]	ViT-S	✓	CVPR-23	95.99	90.12	51.23	26.97	66.08
FLoR [53]	ViT-S	✓	CVPR-24	96.47	90.75	53.06	27.02	66.83
AttnTemp	ViT-S	✓	**Ours**	**96.66**	**90.82**	<u>54.91</u>	**28.03**	**67.61**
LDP-net* [50]	ResNet10	✓	CVPR-23	91.89	84.05	48.44	26.88	62.82
RDC* [22]	ResNet10	✓	CVPR-22	93.30	84.29	49.91	25.07	63.14
FLoR* [53]	ResNet10	✓	CVPR-24	93.60	83.76	**57.54**	26.89	65.45
MEM-FS+RDA* [42]	ViT-S	✓	TIP-23	95.04	88.77	51.02	27.98	65.70
AttnTemp*	ViT-S	✓	**Ours**	**96.74**	**91.34**	<u>55.22</u>	**28.41**	**67.93**

Experiments

Verification of Improved Attention



(a) Our Attention Map Visualization (b) Quantitative Measure of Our Attention

Table 6: Verification of improved self-attention w.r.t. domain similarity and target-domain accuracy.

Metric.	1	2	3	4	5	6	7	8	9	10	11	12
BL CKA	0.9805	0.9500	0.9667	0.9654	0.9455	0.9146	0.8940	0.8406	0.7446	0.6337	0.2063	0.1756
Ours CKA	0.9857	0.9590	0.9659	0.9704	0.9547	0.9347	0.9148	0.8763	0.7903	0.6655	0.2955	0.1886
BL Acc.	34.67	39.88	42.19	44.73	47.20	48.93	50.05	50.98	52.60	53.02	52.03	51.82
Ours Acc.	34.91	40.47	43.01	45.19	47.28	48.93	50.40	51.47	53.26	54.34	53.98	53.70

Thanks!