

Table 1. Comparison with state-of-the-art works by 1-shot.

Methods	backbone	FT	Mark	Crop.	Euro.	ISIC.	Ches.	Ave.
MN+AFA	ResNet10	×	ECCV-22	80.07	69.63	39.88	23.18	53.19
GNN+FT	ResNet10	×	ICLR-20	87.07	78.02	40.87	24.28	57.06
GNN+ATA	ResNet10	×	IJCAI-21	90.59	83.75	44.91	24.32	60.39
LDP-net	ResNet10	×	CVPR-23	89.40	82.01	48.06	26.67	61.29
GNN+AFA	ResNet10	×	ECCV-22	88.06	85.58	46.01	25.02	61.67
SDT	ResNet10	×	NN-24	90.27	82.02	48.66	27.20	62.04
FLoR	ResNet10	×	CVPR-24	91.25	80.87	51.44	26.70	62.32
MEM-FS	ViT-S	×	TIP-23	93.74	86.49	47.38	26.67	63.57
StyleAdv	ViT-S	×	CVPR-23	94.85	88.57	47.73	26.97	64.53
MICM	ViT-S	×	MM-24	94.61	90.08	46.85	27.11	64.66
SDT	ViT-S	×	NN-24	95.00	89.60	47.64	26.72	64.75
FLoR	ViT-S	×	CVPR-24	95.28	90.41	49.52	26.71	65.48
Ours	ViT-S	×	Ours	95.68	91.04	50.46	27.23	66.10
Fine-tuning	ResNet10	✓	ECCV-20	89.84	81.59	49.51	25.37	61.58
FLoR	ResNet10	✓	CVPR-24	92.33	83.06	56.74	26.77	64.73
TACDFSL	WideResNet	✓	SB-22	93.42	85.19	45.39	25.32	62.33
PMF	ViT-S	✓	CVPR-22	92.96	85.98	50.12	27.27	64.08
StyleAdv	ViT-S	✓	CVPR-23	95.99	90.12	51.23	26.97	66.08
FLoR	ViT-S	✓	CVPR-24	96.47	90.75	53.06	27.02	66.83
Ours	ViT-S	✓	Ours	96.27	91.53	54.69	27.66	67.54
ConFess*	ResNet10	✓	ICLR-2022	88.88	84.65	48.85	27.09	62.37
LDP-net*	ResNet10	✓	CVPR-23	91.89	84.05	48.44	26.88	62.82
RDC*	ResNet10	✓	CVPR-22	93.30	84.29	49.91	25.07	63.14
TPN+ATA*	ResNet10	✓	IJCAI-21	93.56	85.47	49.83	24.74	63.40
FLoR*	ResNet10	✓	CVPR-24	93.60	83.76	57.54	26.89	65.45
MEM-FS+RDA*	ViT-S	✓	TIP-23	95.04	88.77	51.02	27.98	65.70
Ours*	ViT-S	✓	Ours	96.62	91.68	55.66	28.25	68.05

Table 2. Comparison with state-of-the-art works by 5-shot.

Method	backbone	FT	Mark	Crop.	Euro.	ISIC.	Ches.	Ave.
GNN+FT	ResNet10	×	ICLR-20	60.74	55.53	30.22	22.00	42.12
MN+AFA	ResNet10	×	ECCV-22	60.71	61.28	32.32	22.11	44.10
GNN+ATA	ResNet10	×	IJCAI-21	67.47	61.35	33.21	22.10	46.53
GNN+AFA	ResNet10	×	ECCV-22	67.61	63.12	33.21	22.92	46.97
LDP-net	ResNet10	×	CVPR-23	69.64	65.11	33.97	23.01	47.18
FLoR	ResNet10	×	CVPR-24	73.64	62.90	38.11	23.11	49.69
SDT	ResNet10	×	NN-24	73.92	65.87	36.45	23.22	49.97
MEM-FS	ViT-S	×	TIP-23	81.11	68.11	32.97	22.76	51.24
StyleAdv	ViT-S	×	CVPR-23	81.22	72.15	33.05	22.92	52.34
SDT	ViT-S	×	NN-24	81.03	72.71	33.40	22.79	52.48
FLoR	ViT-S	×	CVPR-24	81.81	72.39	34.20	22.78	52.80
Ours	ViT-S	×	Ours	83.51	74.08	34.21	22.93	53.68
Fine-tuning	ResNet10	✓	ECCV-20	73.43	66.17	34.60	22.13	49.08
FLoR	ResNet10	✓	CVPR-24	84.04	69.13	38.81	23.12	53.78
PMF	ViT-S	✓	CVPR-22	80.79	70.74	30.36	21.73	50.91
FLoR	ViT-S	✓	CVPR-24	83.55	73.09	35.49	23.26	53.85
StyleAdv	ViT-S	✓	CVPR-23	84.11	74.93	33.99	22.92	53.99
Ours	ViT-S	✓	Ours	84.54	74.97	35.56	23.39	54.62
LDP-net*	ResNet10	✓	CVPR-23	81.24	73.25	33.44	22.21	52.54
TPN+ATA*	ResNet10	✓	IJCAI-21	82.47	70.84	35.55	22.45	52.83
RDC*	ResNet10	✓	CVPR-22	85.79	70.51	36.28	22.32	53.73
FLoR*	ResNet10	✓	CVPR-24	86.30	71.38	41.67	23.12	55.62
MEM-FS+RDA*	ViT-S	✓	TIP-23	83.74	75.91	37.07	23.85	55.14
Ours*	ViT-S	✓	Ours	87.39	78.41	37.20	23.88	56.72

Table 3. Verification of re-initializing the CLS token.

Method	Crop.	Euro.	ISIC.	Ches.	Ave.
Baseline	95.62	88.62	46.08	26.25	63.89
Ours	95.55	90.48	49.58	27.03	65.66
Ours + Re-Init	95.20	89.60	50.23	26.79	65.31

Table 4. Implementing our method with meta-learning baseline.

Method	Crop.	Euro.	ISIC.	Ches.	Ave.
ProtoNet	93.59	86.92	46.15	25.68	63.09
ProtoNet + Ours	95.03	89.42	48.67	27.15	65.07

Table 5. Verification of low-frequency images.

Maintained Frequency	75%	50%	25%	12.5%	6.25%	3.13%	2%	1%
Random Token (0.1)	9.29	9.29	9.29	9.29	8.67	4.69	3.67	1.63
CLS Token (0.1)	9.45	9.71	12.51	14.17	14.29	14.85	15.07	15.14
Random Token (0.2)	19.08	18.37	17.96	17.76	17.14	14.69	11.53	9.90
CLS Token (0.2)	19.68	20.36	24.71	27.43	28.78	29.73	29.93	30.17

Table 6. Verification of other pretraining method.

Method	Crop.	Euro.	ISIC.	Ches.	Ave.
DINO	81.19	72.06	32.85	22.83	52.23
DINO not load CLS	82.14	73.01	32.79	22.91	52.71
DINO + Ours	83.51	74.08	34.21	22.93	53.68
iBot	81.17	72.71	31.44	22.56	51.97
iBot not load CLS	82.20	73.22	31.46	22.78	52.41
iBot + Ours	82.47	73.83	32.87	22.88	53.01
CLIP	77.33	63.30	29.84	21.11	47.89
CLIP not load CLS	77.40	64.51	29.96	21.45	48.08
CLIP + Ours	78.55	64.88	30.95	22.25	49.16
ImageNet Supervision	52.71	47.59	27.55	21.04	37.22
ImageNet not load CLS	52.75	47.73	28.45	21.07	37.50

Table 7. Training with datasets of 5 constructed domains.

Method	Crop.	Euro.	ISIC.	Ches.	Ave.
Baseline	93.85	89.72	49.74	26.07	64.77
320 classes as domain token	95.03	90.49	49.10	26.81	65.35
64 classes as domain token	95.16	90.61	47.86	26.77	65.10
5 domains as domain token	95.52	90.62	50.77	27.00	65.98

Table 8. Add confidence intervals to Tab.1.

Method	Crop.	Euro.	ISIC.	Ches.	Ave.
(a.1)	81.22 ± 0.06	72.08 ± 0.05	32.75 ± 0.04	22.84 ± 0.05	52.22 ± 0.03
(a.2)	82.50 ± 0.05	72.58 ± 0.05	32.69 ± 0.06	22.94 ± 0.05	52.68 ± 0.03
(a.3)	82.54 ± 0.05	73.02 ± 0.06	32.76 ± 0.04	22.90 ± 0.04	52.81 ± 0.03
(b.1)	83.14 ± 0.06	72.81 ± 0.03	33.82 ± 0.05	23.09 ± 0.02	53.22 ± 0.04
(b.2)	83.32 ± 0.06	73.80 ± 0.03	33.37 ± 0.05	23.19 ± 0.03	53.42 ± 0.04
(c.1)	83.15 ± 0.04	73.26 ± 0.04	33.84 ± 0.03	23.07 ± 0.04	53.33 ± 0.03
(c.2)	83.37 ± 0.04	73.63 ± 0.05	34.26 ± 0.03	23.13 ± 0.03	53.60 ± 0.04

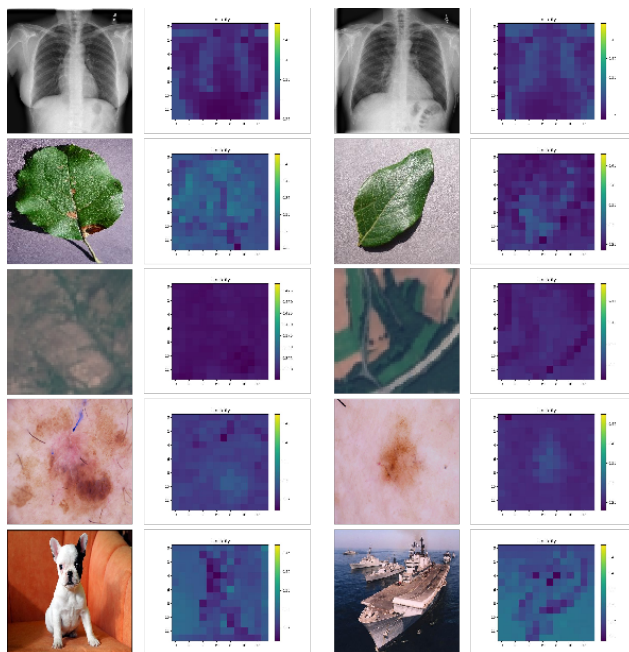


Figure 1. The random CLS token shows much lower similarity with patch tokens, but can capture a coarse contour at each domain.