BoostAdapter: Improving Vision-Language Test-Time Adaptation via Regional Bootstrapping

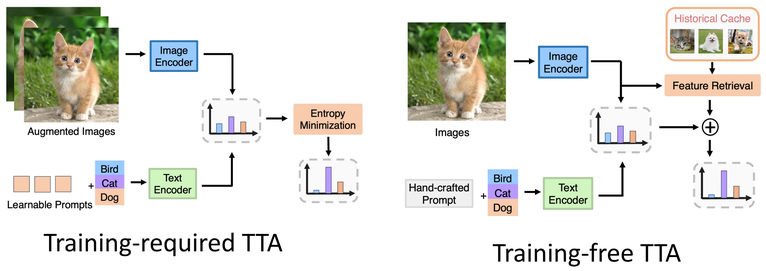
Taolin Zhang, Jinpeng Wang, Hang Guo, Tao Dai, Bin Chen, Shu-Tao Xia

Tsinghua University

Code Available

Motivation

Training-required v.s. Training-free TTA

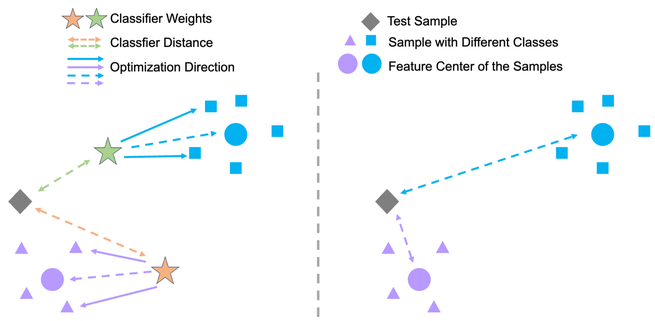


Training-required TTA

Training-free TTA

- Test-time adaptation (TTA) has been widely explored to mitigate misalignment issues in downstream tasks.
- Training-required methods utilize entropy minimization on augmented images, which is time-consuming.
- Training-free TTA methods perform feature retrieval on the historical samples, which do not effectively exploit the fine-grained information.
- Can we bridge the gap between them?

Cross-entropy Optimization v.s. Cache Classifier

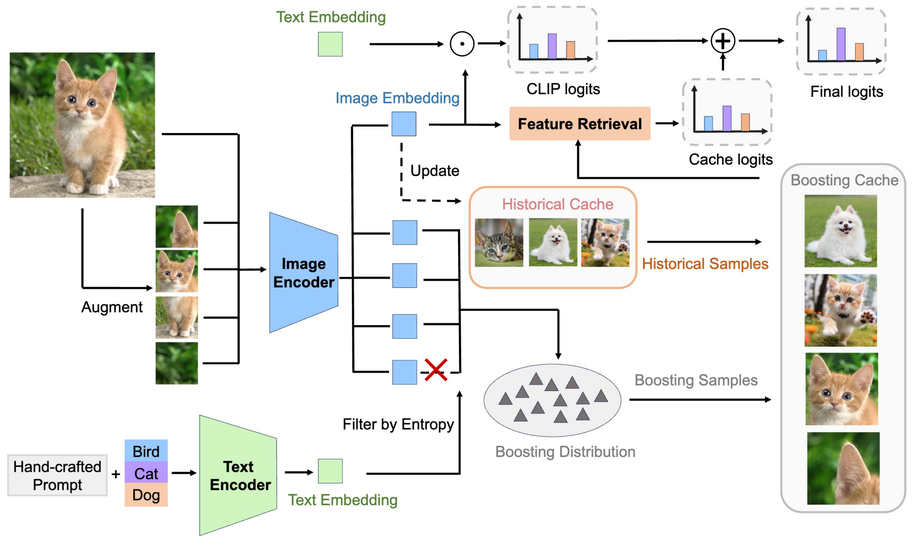


(a) Cross Entropy Optimization (b) Cache Classifier

- Samples pull the classifier weights closer of the same class while pushing them away from different class weights.
- With well-clustered samples, cross-entropy minimization will exhibit similar behavior with the cache classifier!

Method

Boosting your Training-free Adapters



How can we combine training-required TTA with training-free TTA?

Solution: Incorporate boosting samples into the historical cache!

Theoretical Analysis

- Historical Cache reduce Emperical Risk

Proposition 2. ***(Historical Cache reduce Emperical Risk)*** *Given f as the training-free classfier consisting of historical samples only defined by Eq.(4). Let n_t to be the number of confident previously predicted samples in the target domain and k_t as the number of historical samples in the cache, with assumptions 1-3, the following results hold with high-probability for large enough k_t and n_t.*

$$\mathcal{E}(f) \leq \mathcal{O}\left(\left(\frac{1}{k_t}\right)^{1/4} + \left(\frac{k_t}{c_t n_t}\right)^{1/d}\right)^{1+\beta} \tag{9}$$

- Historical Cache benefits from Boosting Samples

Proposition 3. ***(Historical Cache benefits from Boosting Samples)*** *Let n_t to be all confident previously predicted samples in the target domain and n_b be the number of boosting samples that are drawn from the boosting distribution. Given k_t and k_b to be the number of historical samples and the number of boosting samples to be selected as the nearest neighbors stored in the cache, respectively. Let w_{ti} and w_{bi} be the weights defined in Eq.(5) of the historical samples and boosting samples. We have the following bound for the empirical risk of the cache classfier defined in 7.*

$$\mathcal{E}(f) \leq \mathcal{O}\left(\left(\frac{1}{k_t + k_b}\right)^{1/4} + \sum_{i=1}^{k_t} w_{ti}\left(\frac{k_t}{c_t n_t}\right)^{1/d} + \sum_{i=1}^{k_b} w_{bi}\left(\frac{k_b}{c_b n_b}\right)^{1/d}\right)^{1+\beta}. \tag{10}$$

State-of-the-art Comparison

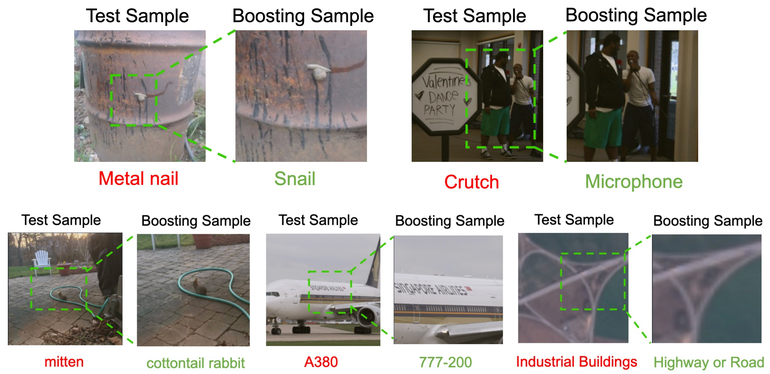
OOD Benchmark

	Imagenet-V2	Imagenet-Sketch	Imagenet-A	Imagenet-R	Average
CLIP [36]	60.86	46.09	47.87	73.98	57.20
CLIP+TPT [41]	64.35	47.94	54.77	77.06	60.81
CoOp [55]	64.20	47.99	49.71	75.21	59.28
CoOp+TPT [41]	**66.83**	49.29	57.95	77.27	62.84
Co-CoOp [54]	64.07	48.75	50.63	76.18	59.91
Co-CoOp+TPT [41]	64.85	48.27	58.47	78.65	62.61
Maple [18]	64.07	49.15	50.90	76.98	60.28
Maple + TPT [41]	64.87	48.16	58.08	78.12	62.31
PromptAlign [39]	65.29	50.23	59.37	79.33	63.55
DiffTPT [6]	65.10	46.80	55.68	75.00	60.52
TDA [17]	64.67	50.54	60.11	80.24	63.89
BoostAdapter	65.51	**51.28**	**64.53**	**80.95**	**65.57**

Cross-Domain Benchmark

	Caltech	Pets	Cars	Flowers	Food101	Aircraft	SUN397	DTD	EuroSAT	UCF101	*Average*
CLIP [36]	93.35	88.25	65.48	67.44	83.65	23.67	62.59	44.27	42.01	65.13	63.58
CLIP+TPT [41]	94.16	87.79	66.87	68.98	84.67	24.78	65.50	**47.75**	42.44	68.04	65.10
CoOp [55]	93.70	89.14	64.51	68.71	85.30	18.47	64.15	41.92	46.39	66.55	63.88
CoCoOp [54]	93.79	90.46	64.90	70.85	83.97	22.29	66.89	45.45	39.23	68.44	64.63
MaPLe [18]	93.53	90.49	65.57	72.23	86.20	24.74	67.01	46.49	48.06	68.69	66.30
MaPLe+TPT [41]	93.59	90.72	66.50	72.37	86.64	24.70	67.54	45.87	47.80	69.19	66.50
DiffTPT [6]	92.49	88.22	67.01	70.10	87.23	25.60	65.74	47.00	43.13	62.67	65.47
PromptAlign [39]	94.01	**90.76**	68.50	**72.39**	86.65	24.80	67.54	47.24	47.86	69.47	66.92
TDA [17]	94.24	88.63	67.28	71.42	86.14	23.91	67.62	47.40	58.00	70.66	67.53
BoostAdapter	**94.77**	89.51	**69.30**	71.66	**87.17**	**27.45**	**68.09**	45.69	**61.22**	**71.93**	**68.68**

Fine-grained Information Mining



Efficiency Analysis

	Augmentation	Views	Inference Speed (fps)	Memory (GB)	OOD Results	Cross-Domain Results
CLIP	-	-	82.3	0.7	57.20	63.58
TPT	Augmix	64	0.29	4.5	60.81	65.10
DiffTPT	Diffusion	64	0.10	14.4	60.52	66.92
TDA	Augmix	64	11.89	1.2	63.89	67.53
BoostAdapter	Rand. Crop & Rand. Horiz. Flip	64	11.23	1.2	65.57	68.68