TTT++: When Does Self-Supervised Test-Time Training Fail or Thrive?

Yuejiang Liu, Parth Kothari, Bastien Van Delft, Baptiste Bellot-Gurlet, Taylor Mordan, Alexandre Alahi

OVERVIEW

Background:

- Test-time training (TTT) through self-supervised learning (SSL) to tackle **distributional shifts**
- Despite encouraging results, not as competitive as other recent approaches

Contribution:

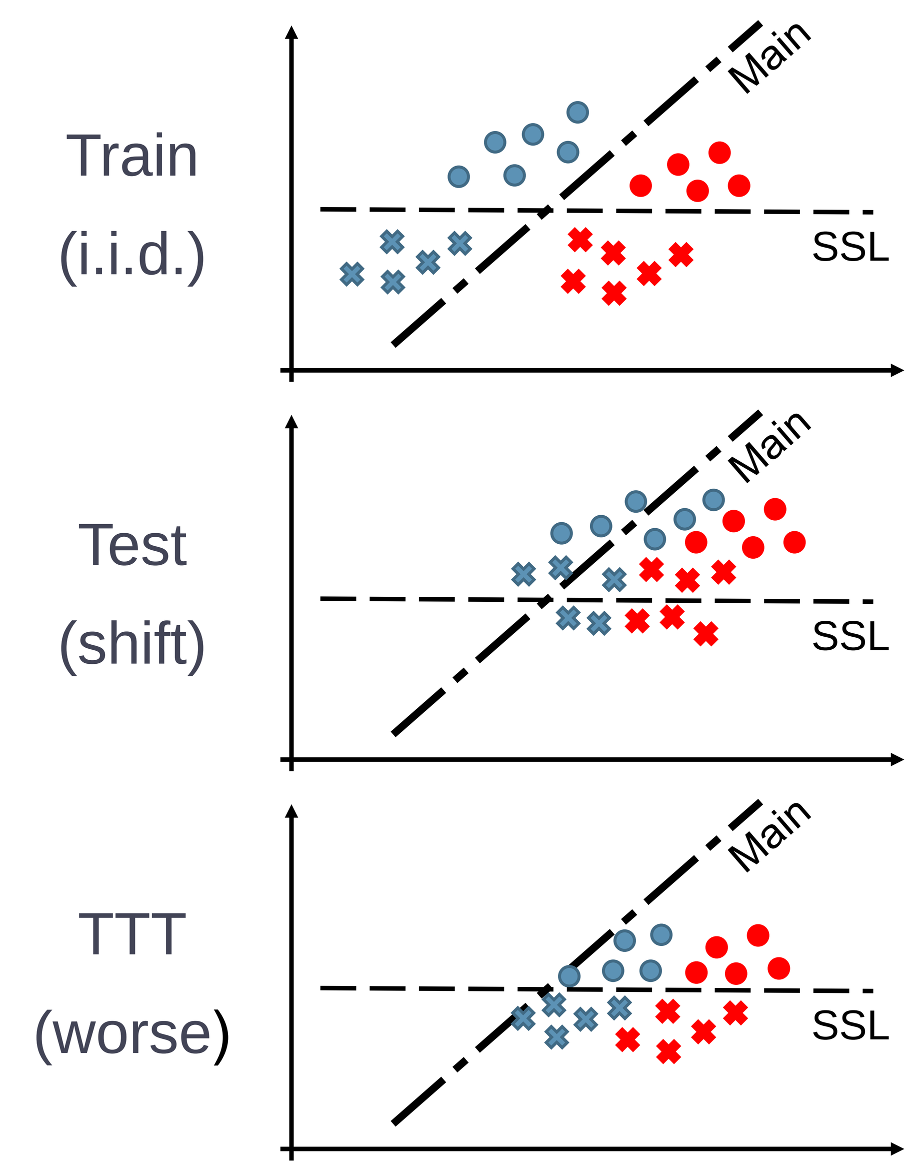
➢ P1: feature alignment remains necessary

✓ S1: offline summarization + online matching

➢ P2: relation btw the main and SSL tasks matters

✓ S2: contrastive method as a strong SSL learner

WHEN DOES TTT FAIL?



WHEN DOES TTT THRIVE?

Assumption: aligned marginal distribution in the feature space, Z' ~ Z

Theorem 1. If the encoder only extracts information about the SSL task:

$$\mathbb{P}(\pi_m(Z') = Y'_m) = \sum_{y_s} \left[\mathbb{P}(Y_s = y_s) \sum_{y_m} \mathbb{P}(Y_m = y_m \mid Y_s = y_s)^2 \right].$$

Theorem 2. In the general case:

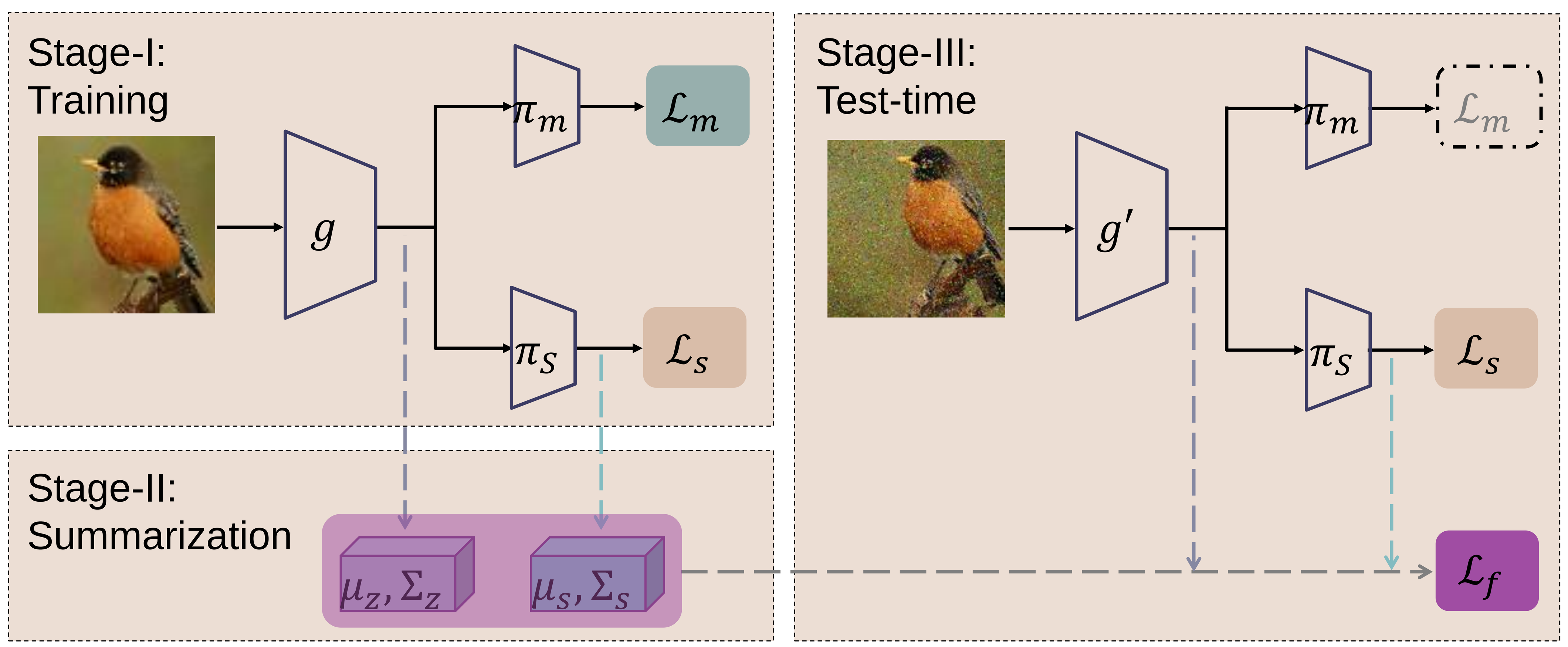
$$\mathbb{P}(\pi_m(Z') = Y'_m) \geq \sum_{y_s} \mathbb{P}(Y_s = y_s) \max \left\{ 0, 2 \left(\max_{y_m} \mathbb{P}(Y_m = y_m \mid Y_s = y_s) - \frac{1}{2} \right) \right\}.$$

Implication: the amount of shared information b/w the two tasks is crucial

$$\mathbb{P}(Y_m = y_m \mid Y_s = y_s)$$

PROPOSED METHOD

TTT++: adapt the encoder g from unlabeled test examples through (i) contrastive self-supervision $\mathcal{L}_s$ + (ii) online feature alignment $\mathcal{L}_f$

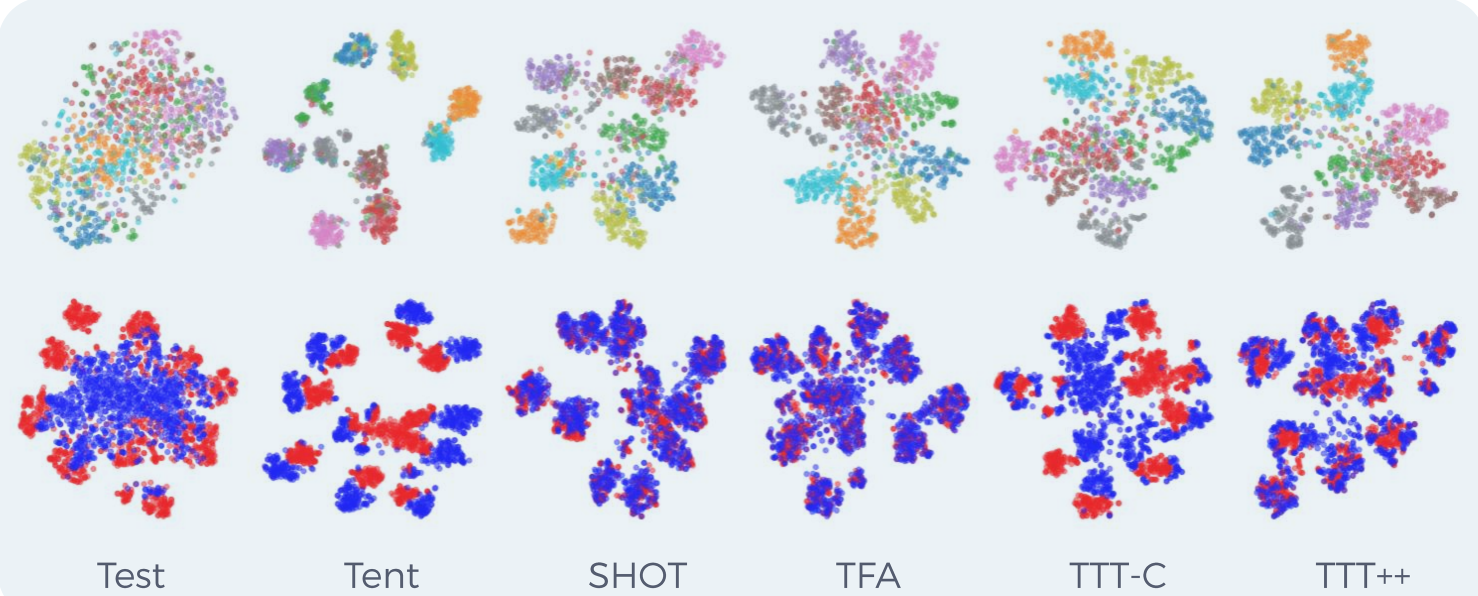


$$\mathcal{L}_{f,z} = \|\mu_z - \mu'_z\|_2^2 + \|\Sigma_z - \Sigma'_z\|_F^2 \qquad \mathcal{L}_{test} = \mathcal{L}_s + \lambda_z \mathcal{L}_{f,z} + \lambda_s \mathcal{L}_{f,s}$$

SOTA RESULTS

❑ CIFAR10

Method	C10-C	C100-C	C10.1
Test (No Adaptation)	29.1	61.2	12.1
BN (Schneider et al., 2020)	15.7	43.3	14.1
TTT-R (Sun et al., 2020a)	14.3	40.4	11.0
SHOT (Liang et al., 2020)	14.7	38.1	11.1
TENT (Wang et al., 2021)	12.6	36.3	13.4
TFA (Ours)	11.9	42.5	12.1
TTT-C (Ours)	10.7	36.9	9.7
TTT++ (Ours)	**9.6**	**36.1**	**9.5**



❑ VisDA-C

Method	plane	bcycl	bus	car	horse	knife	mcycl	person	plant	sktbrd	train	truck	Per-class
Test	56.52	88.71	62.77	**30.56**	81.88	99.03	17.53	95.85	51.66	77.86	20.44	99.51	58.72
BN	44.38	56.98	33.24	55.28	37.45	66.60	16.55	59.02	43.55	60.72	31.07	82.98	48.12
TENT	13.43	77.98	**20.17**	48.15	21.72	82.45	12.37	35.78	21.06	76.41	34.11	98.93	42.73
SHOT	5.73	**13.64**	23.33	42.69	7.93	86.99	19.17	**19.97**	11.63	11.09	15.06	**43.26**	25.04
TFA	28.25	32.03	33.67	64.77	20.49	**56.63**	22.52	36.30	24.84	35.20	25.31	64.24	39.58
TTT-C	5.46	32.23	25.42	37.03	7.84	85.20	9.14	23.80	11.72	11.00	7.74	56.87	25.72
TTT++	**4.13**	26.20	21.60	31.70	**7.43**	83.30	**7.83**	21.10	**7.03**	**7.73**	**6.91**	51.40	**22.46**

❑ Ablation

TFA	C10-C
w/o $\mathcal{L}_{f,s}$	13.01
w/o $\mathcal{L}_{f,z}$	12.68
w/o Σ	14.51
w/o μ	12.03
Full	**11.87**

Main References: Sun et al. TTT, ICML, 2020. | Liang et al. SHOT, ICML, 2020. | Schneider et al. BN, NeurIPS, 2020. | Wang et al. TENT, ICLR, 2021. | Chen et al. SimCLR, ICML, 2020.