TANGO: Learning Distribution-wise Foundation Prior Consistency and Instance-wise Style Calibration for Medical Image Generalization

Chuang Liu[1]; Yichao Cao[2]; Xiu Su[2*]; Haogang Zhu[1*]

[1]Beihang University, [2]Central South University

CVPR DENVER COLORADO JUNE 3-7, 2026

Abstract and Motivation

- **Background**: Test-time adaptation (TTA) has emerged as a promising solution to address real world domain shifts in medical image segmentation.
- **Limitations and Motivations**: Current approaches adapt by updating or regularizing a pre-trained source model. However, they face two major issues: (i) the source models on which they rely are prone to overfitting under domain shifts; (ii) in dynamic continual testing scenarios, error accumulation and class forgetting are further exacerbated.

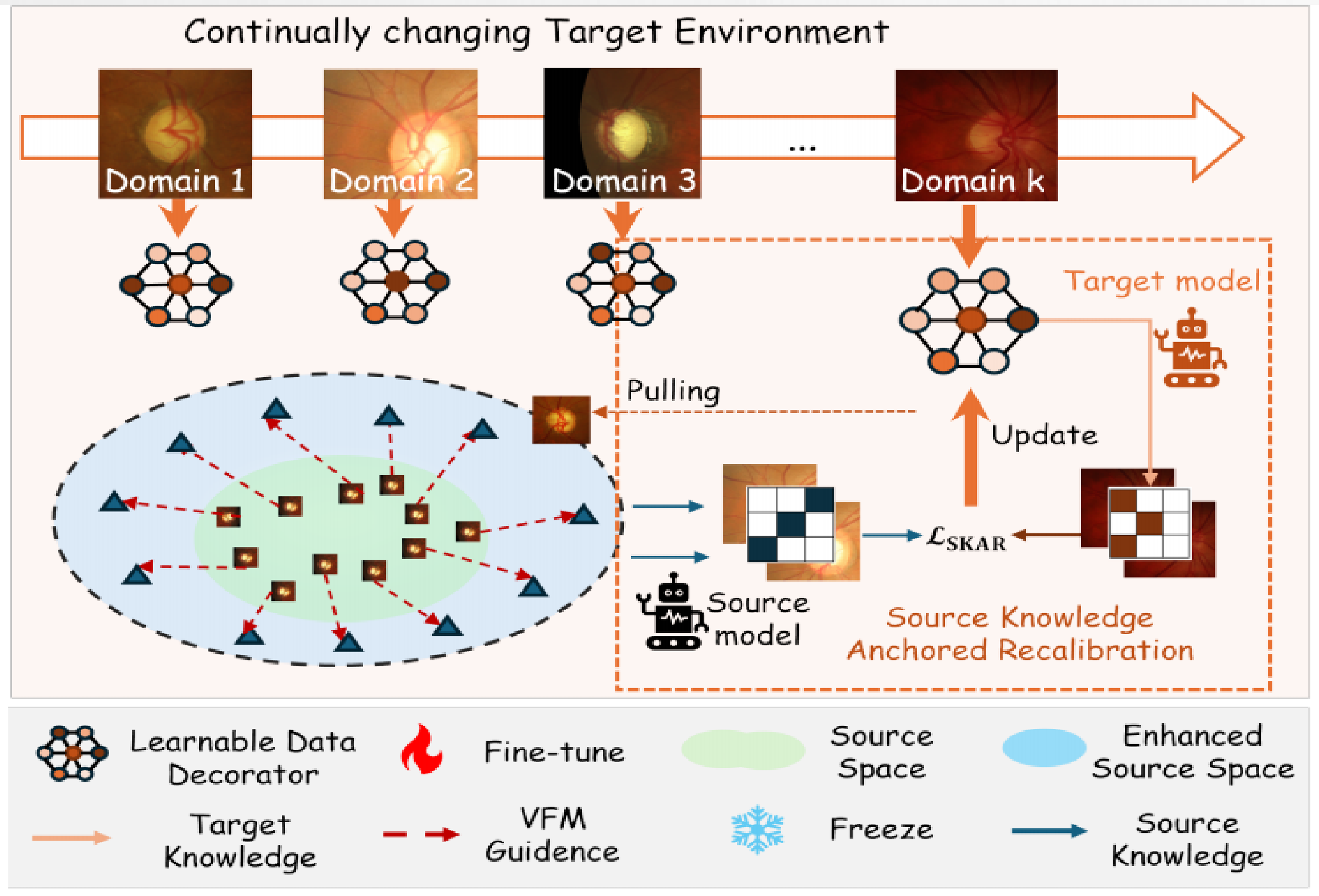
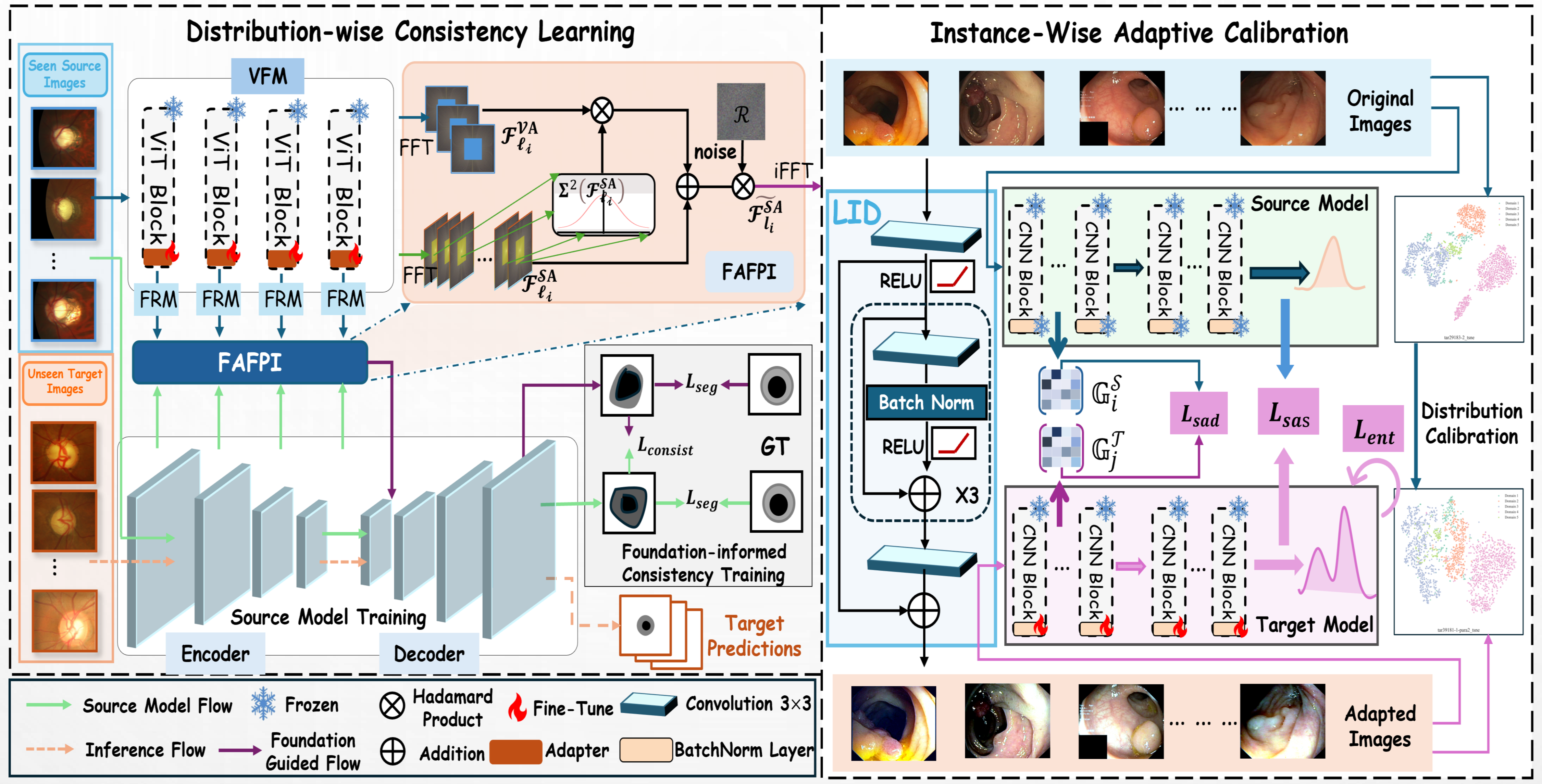


Figure 1. Our Training-to-Adapt CTTA paradigm. During training, we learn generalization priors consistency from a frozen foundation model to enhance the source model's feature robustness. At test time, a learnable per-sample data decorator, together with Source-Knowledge Anchored Recalibration, continuously pulls the evolving test distribution toward the enhanced source feature space, enabling stable and effective online adaptation.

Methodology



Overview of our TanGo framework

- **Distribution-wise Consistency Learning**: During training, the proposed Distribution-wise Style Consistency Learning transfers frequency-domain generalization priors from the foundation model into the source model to form a unified, domain-harmonized feature space.
- **Instance-wise Style Adaptive Calibration**: We develop an instance-wise style adaptive calibration method that jointly optimizes distributional, semantic, and uncertainty objectives, ensuring stable adaptation and semantic consistency under continuously testing conditions.

- FAFPI (Frequency Aware Foundation Prior Injection)
- LID (Learnable Instance Aware Decorator)
- SKAR (Source Knowledge Anchored Recalibration)

Experiments

Comparisons with the SOTAs

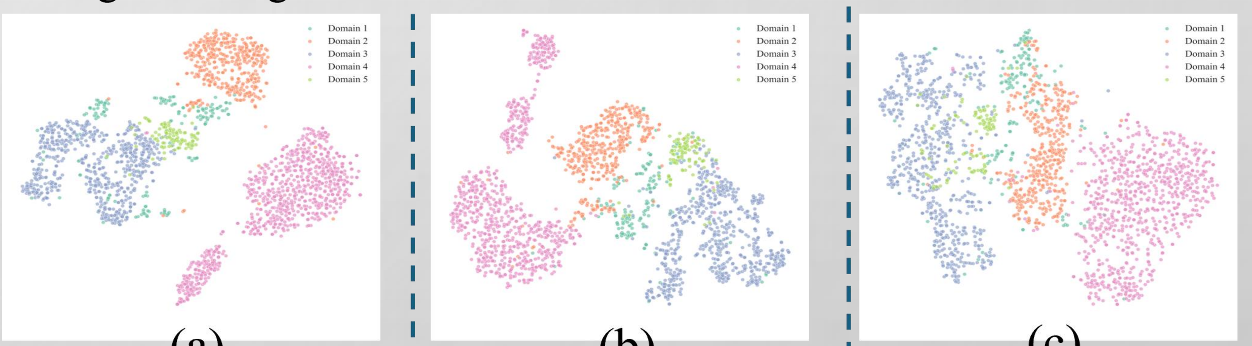
Methods	Domain A DSC	Domain B DSC	Domain C DSC	Domain D DSC	Domain E DSC	Average ↑ DSC
No Adapt [ResUNet-34]	64.53	76.06	71.18	52.67	64.87	65.86
TENT [53] (ICLR 2021)	73.07	78.66	71.94	46.81	70.20	68.13
CoTTA [54] (CVPR 2022)	75.39	75.98	69.14	53.99	70.40	68.98
DLTTA [57] (TMI 2022)	75.11	78.85	73.89	51.64	69.71	69.84
DUA [35] (CVPR 2022)	72.28	76.59	70.13	56.17	71.38	69.31
SAR [39] (ICLR 2023)	74.55	77.71	70.78	55.40	71.72	70.03
DomainAdaptor [59] (CVPR 2023)	74.50	76.39	71.81	56.78	70.55	70.01
VPTTA [6] (CVPR 2024)	73.91	79.36	74.51	56.51	75.35	71.93
CertainTTA [10] (IF 2025)	74.06	79.63	73.78	56.47	75.28	71.84
GraTa [7] (AAAI 2025)	76.58	78.72	76.27	67.15	72.88	74.32
TanGo (Ours)	87.76	85.96	84.49	84.41	83.52	85.23

s are highlighted as best, second and third, respectively.

Methods	Domain A DSC	Domain A E_ϕ^{max}	Domain A S_α	Domain B DSC	Domain B E_ϕ^{max}	Domain B S_α	Domain C DSC	Domain C E_ϕ^{max}	Domain C S_α	Domain D DSC	Domain D E_ϕ^{max}	Domain D S_α	Average ↑ DSC	Average ↑ E_ϕ^{max}	Average ↑ S_α
No Adapt (PraNet)	79.90	87.97	84.66	66.33	78.51	76.72	73.89	84.64	81.28	82.95	90.84	88.08	75.77	85.49	82.69
TENT [53] (ICLR 2021)	74.86	84.58	80.52	67.51	78.66	78.05	17.79	40.04	53.30	73.55	83.38	82.41	58.43	71.67	73.57
CoTTA [54] (CVPR 2022)	76.46	85.37	82.56	66.77	76.75	79.17	71.39	83.42	80.18	70.71	79.81	82.54	71.33	81.34	81.11
DLTTA [57] (TMI 2022)	76.27	85.23	82.41	66.58	77.00	79.24	63.72	78.23	75.56	71.20	81.32	83.47	69.44	80.45	80.17
DUA [35] (CVPR 2022)	78.93	87.37	83.96	66.84	78.52	77.51	76.53	86.45	83.05	86.24	93.23	89.82	77.13	86.39	83.58
SAR [39] (ICLR 2023)	76.48	85.89	81.49	66.45	77.35	78.05	71.46	83.23	79.40	70.41	80.11	81.07	71.20	81.65	80.00
ainAdaptor [59] (CVPR 2023)	77.48	86.31	82.40	70.82	81.76	80.88	71.96	83.06	79.97	76.89	85.89	84.45	74.29	84.26	81.93
VPTTA [6] (CVPR 2024)	81.00	88.91	84.91	76.87	87.31	84.08	77.58	87.48	83.64	86.39	93.47	89.87	80.46	89.29	85.62
CertainTTA [10] (IF 2025)	81.55	88.97	84.96	75.73	86.86	83.46	73.94	86.76	82.92	87.11	93.88	90.27	79.58	89.12	85.40
TanGo (Ours)	83.95	91.33	86.55	82.44	91.45	86.42	81.11	89.81	85.76	87.29	94.80	89.97	83.70	91.85	87.14

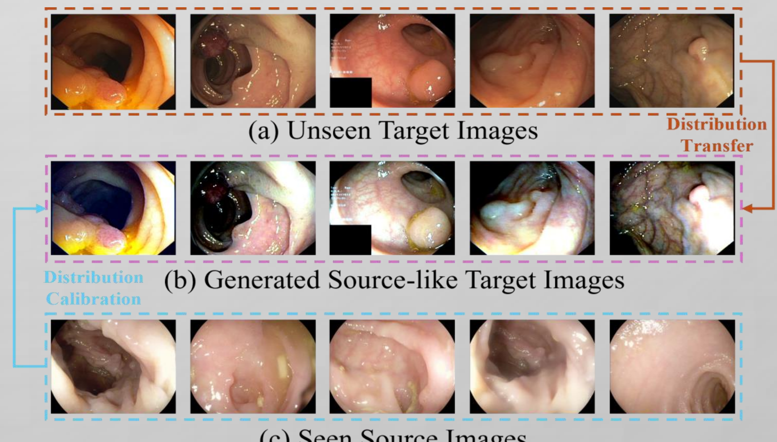
Our method achieves competitive or superior performance across most evaluation metrics and datasets.

Original Images VPTTA Ours

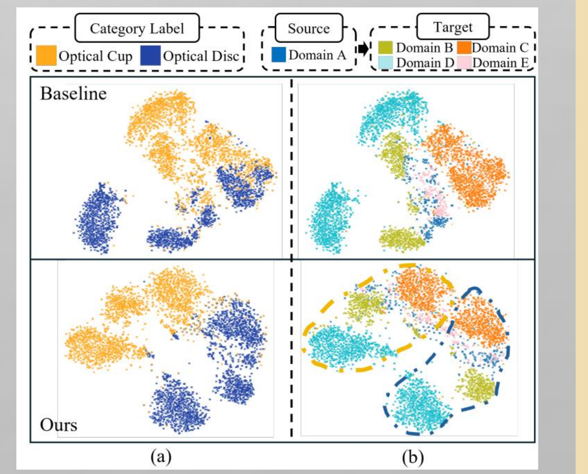


(a) (b) (c)

Visualizations of LID test images



t-SNE visualizations of the learned distributions



Contact

Chuang Liu
Beihang University
cliu_trans@buaa.edu.cn

My Wechat

References

Xingbo Dong, Liwen Wang, Xingguo Lv, Xiaoyan Zhang,Hui Zhang, Bin Pu, Zhan Gao, Iman Yi Liao, and Zhe Jin.Certaintta: Estimating uncertainty for test-time adaptation on medical image segmentation. Information Fusion, page 103300, 2025.
Xingguo Lv, Xingbo Dong, Liwen Wang, Jiewen Yang, Lei Zhao, Bin Pu, Zhe Jin, and Xuejun Li. Test-time domain generalization via universe learning: A multi-graph matching approach for medical image segmentation. In Proceedings of the Computer Vision and Pattern Recognition Conference, pages 15621–15631, 2025