Supplementary Material: Topology-Inspired Backward-Free Framework for Test-Time Adaptation in Medical Detection

Bin Pu[1], Xingguo Lv[1], Jiewen Yang[2], Kai Xu[1], Lei Zhao[1], Zuozhu Liu[3], Kenli Li[1*]

[1]Hunan University, Changsha, China
[2]The Hong Kong University of Science and Technology, HKSAR, China
[3]Zhejiang University, Hangzhou, China
{pubin, lvxg, zhaolei, lkl}@hnu.edu.cn, jyangcu@connect.ust.hk

Appendix

A1. Additional Ablation Studies

We conducted a series of additional ablation studies to evaluate the impact of various modules and augmentation strategies on detection performance across all datasets.

Effect of SPM. The core of Structural Perception Modeling (SPM) is Gaussian Kernel Fusion (GKF), which models the spatial distribution of organ centers across augmented views of a test sample using kernel density estimation. The results shown in Table 1 reflect the performance variations across all datasets, depending on whether augmentation was applied, whether GKF was used, and the choice of Box Fusion strategy. As observed, the TTA detection performance consistently improves with the use of augmentation followed by GKF. Additionally, when Box Regression Adaptation (BRA) fusion is applied, the model shows the greatest improvement in mAP. Specifically, a 12.64% improvement was observed on the Center 1→2 split, and a 15.03% improvement on the Center B→2 split. These results highlight the effectiveness of our design for TTA.

In contrast to current mainstream methods that rely on gradient backpropagation to update network parameters, our parameter-free, backward-free approach significantly enhances the network's adaptability to unknown domains more efficiently.

Effect of Different Augmentations. We conducted ablation experiments on different augmentation strategies across all datasets. We selected commonly used augmentation methods, including Flip, Rotate, Noise Injection, Crop, and Resize. As shown in Table 2, combining more augmentation strategies generally improves the model's generalization performance, helping it learn more domain-invariant features. However, simply stacking different augmentations does not lead to significant performance gains. As observed in rows 6-7, when both resized and transformed images are included, the model can better leverage multi-scale information, leading to more stable domain adaptation capabilities.

Parameters Sensitivity. Ablation studies were also conducted on the hyperparameters η, τ, and α. Specifically, α controls the bandwidth of the Gaussian kernel, τ defines the response threshold for local peak detection in Eq. (**??**), and η sets the minimum weight contribution for box refinement. Detailed results are provided in Table 3. Based on these experiments, we set the final hyperparameters to $\eta = 0.1$, $\tau = 0.05$, and $\alpha = 0.2$. The optimal hyperparameter values may vary depending on the dataset, and we selected values that yielded stable results across experiments.

Cross-Modality Evaluation on CT and MRI.

To further examine the generalizability of our framework beyond fetal ultrasound, we conduct comprehensive cross-modality experiments on abdominal CT and MRI. We construct an Abdominal Multi-organ Dataset consisting of 30 abdominal CT volumes from the MICCAI 2015 Multi-Atlas Abdomen Labeling Challenge (Landman et al. 2015) and 20 T2SPIR MRI volumes from the ISBI 2019 CHAOS Challenge (Kavur et al. 2021). Four abdominal organs are included for detection: liver, right kidney, left kidney, and spleen. Following the preprocessing protocol of (Chen et al. 2020), we normalize intensity ranges, resample the volumes to a consistent spatial resolution, and convert the provided segmentation masks into bounding-box annotations. Due to the substantial variation in slice thickness between CT and MRI scans, all 3D volumes are further decomposed into 2D slices for model inference and adaptation.

We evaluate our method under two challenging cross-modality adaptation settings: CT→MRI and MRI→CT, and benchmark against several state-of-the-art TTA approaches. As summarized in Table 4, T^{3}A consistently achieves the highest performance across both adaptation directions, demonstrating robust generalizability across modalities. In particular, T^{3}A surpasses the second-best baseline by 4.85% mAP in CT→MRI and 4.23% mAP in MRI→CT. These results highlight the strong cross-domain adaptability of our approach and its potential applicability to a broad range of medical imaging tasks beyond ultrasound.

*Corresponding authors.

Table 1: Ablation results (mAP) on augmentation, GKF, and fusion strategies on FUSH^2.

Augment	GKF	Box Fusion	Center1→2	Center2→1	Center A→B	Center B→A	Center 2→B	Center B→2
✗	✗	-	62.93	82.07	78.58	77.16	75.83	63.34
✓	✗	NMS	72.46	85.98	79.28	86.23	87.74	68.82
✓	✓	NMS	73.20	86.09	81.22	85.13	88.31	72.48
✓	✓	WBF	73.44	87.88	82.66	86.70	89.51	75.08
✓	✓	BRA (Ours)	**75.57**	**89.37**	**83.71**	**88.00**	**91.74**	**78.37**

Table 2: Effect of different augmentations on Center 1→2 of FUSH^2. Short side resized to $\{500, \ldots, 1200\}$ and "$\times 2$" denotes inclusion of both resized and transformed images.

Flip	Rotate	Noise	Crop	Resize	Center 1→2	Center 2→1	Center A→B	Center B→A	Center 2→B	Center B→2
✗	✗	✗	✗	$\{800\}$	62.93	78.23	74.44	80.93	82.40	62.18
✓	✗	✗	✗	$\{800\}$	63.62	78.80	73.84	82.34	82.82	63.68
✓	✗	✗	✗	$\{800\} \times 2$	63.98	79.03	74.71	82.59	83.43	64.09
✓	✓	✗	✗	$\{800, 1200\}$	66.22	79.48	74.00	83.08	85.07	64.70
✓	✓	✗	✗	$\{800, 1200\} \times 2$	67.05	78.93	74.24	83.26	84.79	65.21
✓	✓	✓	✗	$\{600, 800, 1000, 1200\}$	66.89	81.24	76.55	85.63	85.66	65.79
✓	✓	✓	✗	$\{600, 800, 1000, 1200\} \times 2$	68.14	80.78	77.31	86.13	85.80	66.12
✓	✓	✓	✗	$\{500, ..., 1200\} \times 2$	69.56	83.69	77.50	85.92	87.11	66.66
✓	✓	✓	✓	$\{500, ..., 1200\}$	70.71	84.07	78.81	**86.92**	86.25	67.51
✓	✓	✓	✓	$\{500, ..., 1200\} \times 2$	**72.46**	**85.98**	**79.28**	86.23	**87.74**	**68.82**

Table 3: Effect of hyperparameters η, τ, and α on Center 1→2 of FUSH^2.

Center 1→2					Center 2→1				
η	τ	mAP (%)	α	mAP (%)	η	τ	mAP (%)	α	mAP (%)
0.05	0.05	74.84	0.1	75.08	0.05	0.05	88.73	0.1	88.90
0.10	0.05	**75.08**	0.2	**75.57**	0.10	0.05	88.64	0.2	**89.37**
0.10	0.10	74.63	0.5	75.29	0.10	0.10	**89.01**	0.5	88.27
0.20	0.20	74.28	1.0	74.61	0.20	0.20	88.02	1.0	87.95
Center A→B					Center B→A				
η	τ	mAP (%)	α	mAP (%)	η	τ	mAP (%)	α	mAP (%)
0.05	0.05	82.99	0.1	83.10	0.05	0.05	87.27	0.1	87.29
0.10	0.05	**83.58**	0.2	**83.71**	0.10	0.05	**87.89**	0.2	**88.00**
0.10	0.10	83.33	0.5	83.19	0.10	0.10	87.46	0.5	87.40
0.20	0.20	82.89	1.0	83.41	0.20	0.20	87.03	1.0	87.25
Center 2→B					Center B→2				
η	τ	mAP (%)	α	mAP (%)	η	τ	mAP (%)	α	mAP (%)
0.05	0.05	90.59	0.1	90.67	0.05	0.05	78.00	0.1	78.03
0.10	0.05	**91.26**	0.2	**91.74**	0.10	0.05	**78.11**	0.2	**78.37**
0.10	0.10	91.30	0.5	91.33	0.10	0.10	77.93	0.5	77.98
0.20	0.20	91.09	1.0	91.13	0.20	0.20	77.57	1.0	77.64

Table 4: Performance comparison of various method in cross-modal adaptation.

Methods	**Abdominal CT ⟶ Abdominal MRI**					**Abdominal MRI ⟶ Abdominal CT**				
	Liver	R. kidney	L. kidney	Spleen	mAP (%)	Liver	R. kidney	L. kidney	Spleen	mAP (%)
No Adapt	63.07	26.79	33.51	51.30	43.67	49.65	19.76	30.52	55.96	38.97
TENT (ICLR'21) (Wang et al. 2021)	67.90	35.13	37.74	50.03	47.70	42.65	**54.86**	28.09	60.93	46.63
DLTTA (TMI'22) (Yang et al. 2022)	69.21	42.09	40.01	57.72	52.25	60.22	35.06	44.73	**67.24**	51.81
DomainAdaptor (ICCV'23) (Zhang et al. 2023)	67.78	40.66	45.83	51.22	51.37	57.47	38.50	42.24	64.50	50.67
MonoTTA (ECCV'24) (Lin et al. 2024)	61.01	33.99	40.13	56.73	47.96	52.58	25.94	33.59	59.83	42.98
VPTTA (CVPR'24) (Chen et al. 2024)	70.17	45.38	47.49	61.64	56.17	56.67	34.29	47.03	59.21	49.30
TTDG-MGM (CVPR'25) (Lv et al. 2025)	74.21	50.08	45.90	**67.25**	59.36	60.41	48.74	49.36	64.65	55.79
GraTa (AAAI'25) (Chen et al. 2025)	70.45	44.28	49.30	60.77	56.20	64.08	45.40	50.11	45.89	51.37
T^{3}A (Ours)	**76.82**	**58.33**	**54.69**	67.00	**64.21**	**71.03**	51.45	**58.90**	58.70	**60.02**

References

Chen, C.; Dou, Q.; Chen, H.; Qin, J.; and Heng, P. A. 2020. Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation. *IEEE Trans. Med. Imaging*, 39(7): 2494–2505.

Chen, Z.; Pan, Y.; Ye, Y.; Lu, M.; and Xia, Y. 2024. Each test image deserves a specific prompt: Continual test-time adaptation for 2d medical image segmentation. In *Proc. IEEE/CVF Comput. Vis. Pattern Recog.*, 11184–11193.

Chen, Z.; Ye, Y.; Pan, Y.; and Xia, Y. 2025. Gradient alignment improves test-time adaptation for medical image segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 2429–2437.

Kavur, A. E.; Gezer, N. S.; Barış, M.; Aslan, S.; Conze, P.-H.; Groza, V.; Pham, D. D.; Chatterjee, S.; Ernst, P.; Özkan, S.; et al. 2021. CHAOS challenge-combined (CT-MR) healthy abdominal organ segmentation. *Medical Image Analysis*, 69: 101950.

Landman, B.; Xu, Z.; Igelsias, J.; Styner, M.; Langerak, T.; and Klein, A. 2015. Miccai multi-atlas labeling beyond the cranial vault–workshop and challenge. In *Proc. MICCAI Multi-Atlas Labeling Beyond Cranial Vault—Workshop Challenge*, volume 5, 12.

Lin, H.; Zhang, Y.; Niu, S.; Cui, S.; and Li, Z. 2024. Monotta: Fully test-time adaptation for monocular 3d object detection. In *European Conference on Computer Vision*, 96–114. Springer.

Lv, X.; Dong, X.; Wang, L.; Yang, J.; Zhao, L.; Pu, B.; Jin, Z.; and Li, X. 2025. 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*, 15621–15631.

Wang, D.; Shelhamer, E.; Liu, S.; Olshausen, B.; and Darrell, T. 2021. Tent: Fully Test-Time Adaptation by Entropy Minimization. In *ICLR*.

Yang, H.; Chen, C.; Jiang, M.; Liu, Q.; Cao, J.; Heng, P. A.; and Dou, Q. 2022. Dltta: Dynamic learning rate for test-time adaptation on cross-domain medical images. *IEEE Transactions on Medical Imaging*, 41(12): 3575–3586.

Zhang, J.; Qi, L.; Shi, Y.; and Gao, Y. 2023. Domainadaptor: A novel approach to test-time adaptation. In *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 18971–18981.