This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Sequence-Independent Continual Test-Time Adaptation with Mixture of Incremental Experts for Cross-Domain Segmentation

Dunyuan Xu[1], Yuchen Yuan[1], Donghao Zhou[1], Xikai Yang[1], Jingyang Zhang[3], Jinpeng Li[1]✉, and Pheng-Ann Heng[1,2]

[1] Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China
jpli21@cse.cuhk.edu.hk
[2] Institute of Medical Intelligence and XR, The Chinese University of Hong Kong, Hong Kong, China
[3] The School of Computer Science and Engineering, Southeast University.

Abstract. Continual Test-Time Adaptation (CTTA) adapts a model pretrained on the source domain to sequentially arriving unlabeled target domains. However, existing approaches predominantly assume that model would complete adaptation to all samples within the same target domain before transitioning to the next domain, deviating from realistic clinical scenarios where samples from diverse domains arrive stochastically. Such gradual adaptation strategies suffer from performance drop under rapid domain shifts and limits their clinical applicability. To address this issue, we propose Mixture of Incremental Experts (MoIE), a lightweight network structure that maps new patterns to established knowledge. Specifically, MoIE incorporates two key innovations: 1) Progressive Expert Expansion (PEE), which dynamically adds experts when existing ones fail to effectively process the current sample, enabling stable and swift adaptation to target domains; 2) Knowledge-Transfer Initialization (KTI), which initializes new experts by combining existing ones through domain-similarity based weights, enabling fast adaptation to unseen domains while preserving learned knowledge to prevent immediate forgetting. Experiments on two CTTA tasks (prostate and fundus segmentations) indicate its superiority by achieving SOTA performance with minimal performance gaps across diverse inference sequences. (Code available at https://github.com/dyxu-cuhkcse/MoIE)

Keywords: Continual Test-Time Adaptation · Mixture of Experts · Dynamic Network.

1 Introduction

Deep learning has achieved remarkable success in medical image segmentation [3,1]. However, most approaches suffer from performance degradation when deployed on out-of-distribution domains [11], which poses a critical challenge in

Dunyuan Xu and Yuchen Yuan contributed equally.

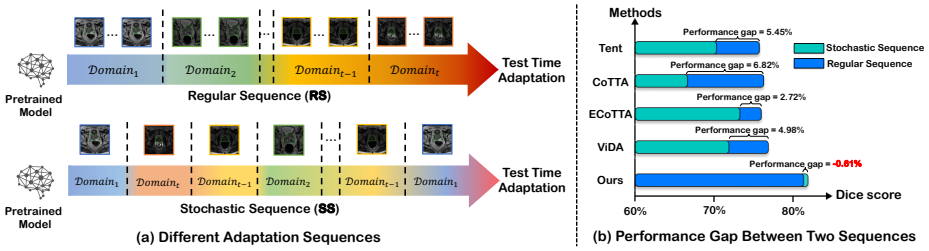

Fig. 1: **The problem setting and motivation**. (a) Different definitions of two adaptation sequences; (b) Performance degradation between two settings.

clinical scenarios where data is often collected from heterogeneous and isolated sources continuously (e.g. different hospitals or imaging devices) [13]. To address this issue, Continual Test-Time Adaptation (CTTA) [5,24] has emerged as a promising paradigm, enabling models to adapt to data stream from continually shifting target domains. Despite the advances achieved, current CTTA methods typically follow the Regular Sequence (**RS**) setting, where the model processes samples from one domain exhaustively before transitioning to the next domain.

Recent works achieve promising performance under RS through unsupervised frameworks [15,2,19,20,25] or improving network architectures [21,22,6], such as teacher-student frameworks [18], lightweight meta-networks [16] and hybrid-rank adapters [12]. However, in most real clinical scenarios, samples from different domains typically appear randomly during the inference process. We refer to this setting as Stochastic Sequence (**SS**), and the differences between RS and SS are shown in Fig. 1(a). This scenario is becoming increasingly practical with the rise of telemedicine [7], where centrally deployed models must handle heterogeneous data streams from multiple institutions, often arriving in a random order. While practically important, the SS setting is rarely explored in the medical CTTA field. Therefore, in this paper, we aim to develop a method that can effectively handle the SS setting while maintaining stable performance in the conventional RS setting. The model should exhibit the following two key properties: 1) enabling stable and swift adaptation to handle rapid and recurring domain shifts, which is crucial under SS where data continuously arrives from unpredictable domains; and 2) preventing immediate forgetting of encountered domain patterns during the adaptation process, as the model is forced to make aggressive updates due to larger domain variations between consecutive samples introduced in the SS. Unfortunately, existing methods fail to handle the SS setting due to their gradual adaptation [10] to new domains and poor memory retention [9] during adaptation. These methods require a long-term process to slowly accumulate knowledge from abundant samples from a newly encountered domain, which makes them unsuitable for rapid domain shifts under the SS setting and leads to performance drop compared to RS, as demonstrated in Fig. 1(b).

To address this challenge, we introduce, to the best of our knowledge, *the first sequence-independent CTTA framework* for medical image segmentation,

which achieves stable and superior performance under both RS and SS settings. This is attributed to two key advancements: 1) a dynamic Mixture of Experts (MoE)-based network architecture that flexibly aligns target features with the source domain feature space by introducing only a few additional lightweight subnetworks (i.e., experts), enabling stable and swift adaptation by mapping newly encountered patterns to established feature representations; and 2) a domain-similarity based weighting system that guides both MoE feature ensemble and incremental expert initialization. Experts in MoE layers serve as dedicated memory units that store encountered knowledge through adaptation. For samples with fewer feature variations, the system enables self-adaptive knowledge retrieval by combining features across experts. For samples with large domain shifts, our weighted initialization integrates existing experts to maintain fast adaptation while preserving previous knowledge, effectively preventing forgetting. Specifically, we propose the Mixture of Incremental Experts (MoIE) which comprise multiple specialized experts, each optimized to handle specific patterns. To ensure model flexibility to mitigate the impact of rapid domain shifts, we design a self-adaptive Progressive Expert Expansion (PEE), which gradually adds new experts during adaptation when existing experts are insufficient to effectively process data from a new domain. Given the need for enhancing memorizability, we propose an efficient Knowledge-Transfer Initialization (KTI) module to initialize newly added experts by weighted integration of existing ones, which ensures fast adaptation while retaining previous knowledge. Moreover, a warm-up phase is designed to pretrain experts in MoIE layers before deploying for adaptation, enabling a more reliable start point. We evaluate our method on two CTTA tasks for prostate and fundus segmentation, showing its superior performance and robustness compared to other methods under both RS and SS settings.

2 Method

In the CTTA scenario, a model pretrained on the source domain performs continuous inference while adapting to stochastic sequential online samples from diverse target domains. Fig. 2(a) illustrates our proposed sequence-independent CTTA framework that only updates MoIE layers during both warm-up and adaptation phases by creating additional experts through PEE and initializing newly added experts via KTI. Our proposed MoIE module effectively provides the stable and swift adaptation while preserving previous knowledge under both RS and SS settings. A detailed theoretical visualization is shown in Fig. 2(b).

2.1 Progressive Expert Expansion

We propose a dynamic expert expansion mechanism that adds new experts when existing ones cannot adequately process the current sample. In contrast to traditional MoE methods with fixed numbers of experts [10], this adaptive expansion enables stable feature mapping from target domains to source knowledge.

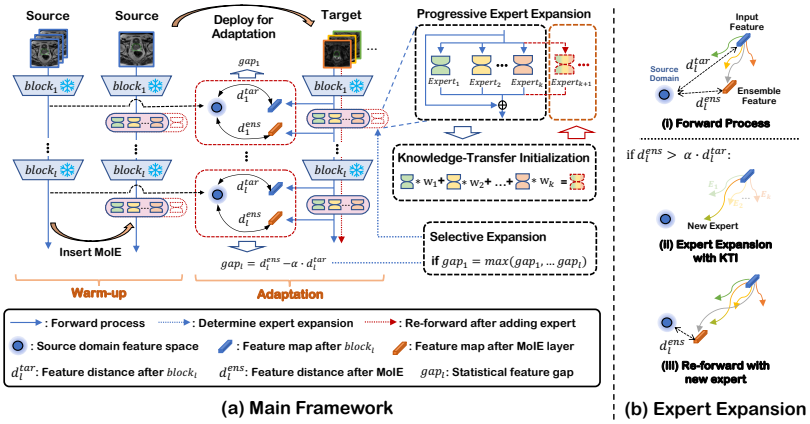

Fig. 2: (a) Our proposed Sequence-Independent CTTA framework with Mixture of Incremental Expert (MoIE) layers. Specifically, we design the Progressive Expert Expansion (PEE) to dynamically create new experts when existing ones become inadequate (Sec. 2.1) and utilize Knowledge-Transfer Initialization (KTI) mechanism to initiate newly added experts for rapid convergence (Sec. 2.2). During adaptation, we leverage a selective expansion strategy to minimizes network modifications thus reducing error accumulation (Sec. 2.3); (b) Our expert expansion strategy aims to map target features to the source domain feature space.

Expert Expansion. We first forward all samples in the source domain through the pretrained model to get block-wise feature statistics $F_l^{sour} = (\mu_l^{sour}, \sigma_l^{sour})$, where μ_l^{sour} and σ_l^{sour} denote the mean and standard deviation. Next, for each target sample, we obtain its feature statistics from all K experts in every MoIE layer, where each expert generates $F_l^k = (\mu_l^k, \sigma_l^k)$, forming a set of K feature representations for layer l. Then we can obtain the statistical domain-similarity ρ_l^k for each expert using the cosine-based similarity function δ_{cos}:

$$\rho_l^k := \delta_{cos}(F_l^k, F_l^{sour}) = \frac{\mu_l^k \cdot \mu_l^{sour}}{\|\mu_l^k\| \cdot \|\mu_l^{sour}\|} + \frac{\sigma_l^k \cdot \sigma_l^{sour}}{\|\sigma_l^k\| \cdot \|\sigma_l^{sour}\|}, \tag{1}$$

Since the MoIE layer aims to better align target features with the source domain knowledge, the new expert is introduced when existing experts fail to achieve adequate knowledge alignment. Specifically, we calculate F_l^{tar} and F_l^{ens} as the target sample's feature map outputted by the l-th block of the pretrained model and the ensembled feature after MoIE layer l, respectively. The target-to-source feature distances for F_l^{ens} and F_l^{tar} are computed by subtracting their domain-similarity with source domain knowledge as calculated in Eq. (1):

$$d_l^{ens} = 2 - \delta_{cos}(F_l^{ens}, F_l^{sour}), \quad d_l^{tar} = 2 - \delta_{cos}(F_l^{tar}, F_l^{sour}). \tag{2}$$

A new expert is created when $d_l^{ens} > \alpha \cdot d_l^{tar}$, where $\alpha \in (0, 1)$ is the threshold.

Weighted Feature Ensemble. Having defined the expert expansion strategy in MoIE, another critical aspect in MoE-based methods is how to determine the

ensemble strategy for combining outputs from multiple experts. Most existing MoE strategies employ gated networks to automatically weight expert features [10,23], yet such mechanism becomes ineffective in CTTA where unlabeled data provides no supervision for effective model updates, particularly for newly added parameters when creating additional experts. Therefore, we propose to determine the ensemble weight w_l^k for the k-th expert in layer l (E_l^k) based on its adaptation effectiveness. Specifically, the similarity calculated in Eq. (1) actually define each expert's capability in aligning target features with source domain knowledge. We can obtain the domain-similarity based ensemble weights $\{w_l^k\}_{k=1}^K$ and calculate the ensembled feature outputted from this MoIE layer (F_l^{ens}) based on $\{\rho_l^k\}_{k=1}^K$:

$$F_l^{ens} = Z_l + \sum_{k=1}^{K} w_l^k \cdot E_l^k(Z_l), \quad \{w_l^k\}_{k=1}^K = \text{Softmax}(\rho_l^1, ..., \rho_l^K), \tag{3}$$

where Z_l is the feature from l-th block of pretrain model. Our progressive expert expansion strategy dynamically adjusts network architecture based on requirements, thereby ensuring stable and rapid adaptation to the target sample.

2.2 Knowledge-Transfer Initialization

Although progressive expert expansion adaptively aligns target features with source domain feature space, newly added experts without proper initialization are not able to leverage previous learned knowledge stored in existing experts.

Weighted Expert Initialization. To fully utilize previously learned knowledge, we propose to initialize incremental experts by domain-similarity based aggregation of existing experts, where each expert acts as a dedicated memory unit of encountered patterns. We formulate the initialization of the newly added expert as: $\Theta_l^{K+1} = \sum_{k=1}^{K} w_l^k \cdot \Theta_l^k$, where w_l^k represents the expert weights calculated based on domain-similarity as shown in Eq. (3) and Θ_l^k represents the network's parameters in the k-th expert from MoIE layer l. The ensemble weights indicate existing experts' effectiveness for processing current sample, thus experts with higher weights should transfer more knowledge to new experts.

Re-forward with New Experts. To immediately incorporate the effect of the newly added expert for the current target sample, we re-forward the sample through this expert to update the ensemble weights from $\{w_l^k\}_{k=1}^K$ to $\{w_l^k\}_{k=1}^{K+1}$ after expert expansion (similar as Eq. (3)). This re-forward operation is computationally efficient, requires only forward once through the newly added expert and updates the softmax and ensemble calculations, making its time consumption comparable to single forward strategies. To prevent infinite experts expansion, each MoIE layer adds experts until reaching its maximum capacity (K_{max}).

2.3 CTTA Learning Process

We achieve stable adaptation and memorizablity by utilizing the PEE and KTI. To further improve the effectiveness of our method, we perform a warm-up training for MoIE layers on the source domain to fully utilize its available labels.

Table 1: Performance comparison on prostate segmentation task under CTTA scenario (SS and RS). **Bold** font denotes the best performance.

	Stochastic Sequence (SS)						**Regular Sequence (RS)**						**Gap↓**
Site	B	C	D	E	F	**Avg**	B	C	D	E	F	**Avg**	
	Dice Similarity Coefficient (DSC) (%)												
Source	57.12	80.40	53.85	5.49	0.00	39.37	57.12	80.40	53.85	5.49	00.00	39.37	0.00
BN Adapt	73.43	75.53	74.95	71.62	80.13	75.13	73.07	75.66	75.99	70.79	80.39	75.18	0.05
Tent	51.84	77.63	71.63	74.38	76.26	70.35	71.87	78.25	78.21	73.55	77.13	75.80	5.45
CoTTA	72.80	73.25	73.90	64.99	63.52	69.69	73.82	75.83	78.05	77.56	77.30	76.51	6.82
ECoTTA	73.66	74.11	72.08	70.10	76.67	73.32	76.03	78.90	73.37	73.38	78.55	76.04	2.72
ViDA	63.85	79.04	70.86	69.11	76.86	71.94	75.55	77.90	77.92	77.83	75.38	76.92	4.98
Ours	**80.72**	79.91	**80.52**	**82.82**	**85.79**	**81.95**	**80.17**	77.90	**80.81**	**83.87**	**83.96**	**81.34**	**-0.61**

MoIE Initialization. We insert MoIE layers after every encoder block of the pretrained model. Each MoIE layer starts with two randomly initialized experts to provide diverse foundational expert bases, ensuring sufficient flexibility to generate various initialization patterns for newly constructed experts.

Warm-Up Stage. Before deploying the pretrained model for inference on target domains, we perform a warm-up phase. We freeze all parameters in the pretrained blocks and train the MoIE layers on the source domain, using a combination loss of Binary Cross-Entropy loss and Dice loss, formulated as:

$$L_{warm-up} = L_{BCE}(\Theta) + L_{Dice}(\Theta), \tag{4}$$

where Θ means all learnable parameters. During the warm-up phase, we expand the experts in all MoIE layers for each sample, following the criteria in Eq. (2).

Adaptation Stage. Given that ground truth becomes unavailable during the adaptation process, we reformulate our learning objective by minimizing the unsupervised entropy loss and the target-to-source feature distance:

$$L_{adapt} = -\sum p(\hat{y}) \ \log \ p(\hat{y}) - \sum_{l} gap_l, \tag{5}$$

where $gap_l = |d_l^{ens} - \alpha \cdot d_l^{tar}|$ is the feature distance gap calculated by the absolute value between two target-to-source feature distances as formulated in Eq. (2) and $\hat{y}$ represents the prediction result of the inference sample. Moreover, to prevent error accumulation during CTTA [2], we selectively expand only one MoIE layer for each target sample (i.e. the layer with the largest feature distance gap among all layers). This selective expansion strategy maximizes the effectiveness of newly incorporated experts while minimizes structural modifications to the model, mitigating unwanted error introduced during the CTTA process.

3 Experiments

Dataset. We evaluated our method on two CTTA tasks: 1) prostate MRI segmentation [13] with 116 MRI subjects from six diverse sites: RUNMC (Site A),

Table 2: Performance comparison on Fundus segmentation task under CTTA scenario (SS and RS). **Bold** font denotes the best performance.

	Stochastic Sequence (SS)					**Regular Sequence (RS)**					**Gap↓**
Site	B	C	D	E	**Avg**	B	C	D	E	**Avg**	
	Dice Similarity Coefficient (DSC) (%)										
Source	65.00	62.50	59.69	60.60	61.95	65.00	62.50	59.69	60.60	61.95	0.00
BN Adapt	67.53	63.24	61.30	65.02	64.27	67.51	63.61	61.32	64.78	64.31	0.04
Tent	65.20	64.24	62.91	62.10	63.61	68.57	**65.74**	66.50	67.39	67.05	3.44
CoTTA	64.58	63.50	60.80	**63.77**	63.16	67.68	64.03	65.34	70.17	66.80	3.64
ECoTTA	68.07	60.64	64.43	59.90	63.26	67.59	60.68	68.34	64.64	65.31	2.05
ViDA	76.52	45.84	**77.59**	56.92	64.22	**78.55**	49.25	**81.11**	61.71	67.65	3.43
Ours	**71.14**	**63.56**	66.59	63.10	**66.10**	69.93	64.87	66.57	**71.14**	**68.13**	**2.03**

BMC (Site B), HCRUDB (Site C), UCL (Site D), BIDMC (Site E) and HK (Site F); 2) fundus segmentation [4] with 1,441 cases from five different public datasets for optic disc and optic cup segmentation: BinRushed(Site A), Drishti_GS (Site B), Magrabia (Site C), ORIGA (Site D), REFUGE (Site E). For data pre-processing, images are resized to 256×256 and normalized to [-1, 1].

Experimental Setting. For each task, we pretrained the model on the source domain (Site A for both prostate and fundus tasks), with the remainder serving as target domains for adaptation. During inference, we organized the CTTA in two scenarios: 1) RS, target sites arrive in an Alphabetical order (B→C→D→E→F for the prostate dataset, B→C→D→E for the fundus dataset); 2) SS, all target samples from diverse domains were randomly shuffled for inference. We evaluated the model performance using the Dice Score Coefficient (DSC), computed volume-wise for the prostate task and slice-wise for the fundus task.

Implementation. We used 2D UNet as our segmentation backbone for its scalability [8], each expert in the MoIE layer consist of two fully-connected layers. All experiments are conducted on an NVIDIA TITAN Xp GPU using Adam optimizer. We empirically set the threshold α as 0.65. The learning rates were set to 5×10^{-4} (batch size 8) for pretraining, 1×10^{-3} (batch size 4) for warm-up, and 5×10^{-3} (batch size 1) for adaptation. Each MoIE layer has the capacity of 4 during warm-up and a maximum capacity of 8 experts during adaptation.

Comparison with SOTA. We compared our Sequence-Independent CTTA framework with state-of-the-art methods: 1) **Source** directly inferences the pretrained model; 2) **BN Adapt** [14] only adapts the statistics parameters in Batch Normalization (BN) layers; 3) **TENT** [17] updates all parameters in BN layers; 4) **CoTTA** [18] utilizes the teacher-student framework to generate pseudo-labels; 5) **ECoTTA** [16] integrates the meta-networks for regularizing features; and 6) **ViDA** [12] leverages low-rank and high-rank features for adaptation.

Results on Prostate Task. As presented in Table 1, Source achieves only 39.37% average DSC due to severe domain shifts. BN Adapt maintains robust performance above 75% DSC under both RS and SS, serving as a strong baseline. Other approaches (TENT, CoTTA, ECoTTA and ViDA) effectively im-

prove average DSC scores under RS but degrade significantly under SS with clear sequence-specific gaps (decreases of 5.45%, 6.82%, 2.72% and 4.98% DSC respectively), showing limited sequence-related robustness. In contrast, our strategy demonstrates superior and consistent performance, achieving 81.34% average DSC under RS and 81.95% under SS, surpassing the second-best approach by 4.42% and 6.82% DSC. Notably, our method exhibits robust sequence invariance, evidenced by a negative performance gap between SS and RS (-0.61% DSC).

Results on Fundus Task. As shown in Table 2, Source achieves higher average performance than prostate segmentation, indicating smaller domain shifts. Current adaptation algorithms improve performance under RS but drop significant under SS, achieving even lower average DSC than the naive BN Adapt. This highlights their sensitivity to adaptation sequences. Meanwhile, our MoIE framework maintains the highest average DSC (68.13% under RS and 66.10% under SS) while achieving the smallest performance gap (2.03% DSC) between two scenarios. Notably, while demonstrating a comparable performance gap to ECoTTA, our approach outperforms it with higher average DSC, showing improvements of 2.82% under RS and 2.84% under SS. These results proof that our approach has superior robustness and sequence invariance across tasks.

Ablation Study for Each Component. To validate the effectiveness of each proposed component, we conducted separate ablation experiments as shown in Fig. 3(a). We evaluated five configurations: 1) fixed expert count with warm-up; 2) Progressively Expanding Experts (PEE) without Knowledge-Transfer Initialization (KTI) and warm-up; 3) warm-up and PEE but without KTI; 4) PEE and KTI without warm-up; and 5) the complete integration of all components. All components yield performance gains in average DSC and maintain sequence invariance, as evidenced by small performance gaps between SS and RS.

Explanation of Experts Expansion. Each expert within a MoIE layer specializes in distinct domains to promot diversity and preventing ensemble weight collapse. In Fig. 3(b), points in the radar graph indicate the ensemble weight rank of this expert receives across domains. The diverse shapes demonstrate unique expert characteristics. Fig. 3(c) shows the incremental pattern of new experts creation, where experts are added only when existing ones cannot effectively process current samples, maintaining appropriate expansion pace.

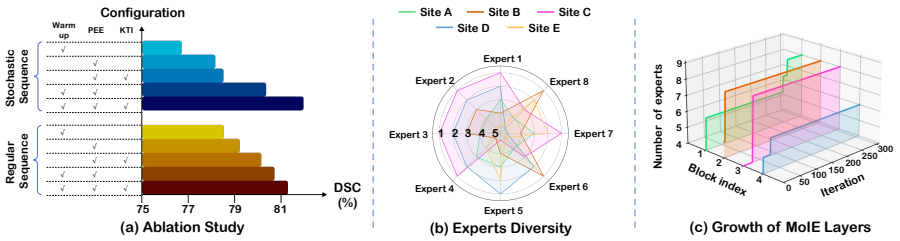

Fig. 3: **Illustrative experiments on prostate task**. (a) Ablation study; (b) Diversity of experts in a specific MoIE layer; (c) Growth of experts under SS.

4 Conclusion

This paper identifies that variation on inference sequences in CTTA degrades the model performance due to rapid domain shifts. To achieve effective and stable online CTTA, we propose a sequence-independent framework with Mixture of Incremental Experts (MoIE) layers, that incorporates Progressive Expert Expansion (PEE) with Knowledge-Transfer Initialization (KTI), enabling stable adaptation while maintaining good memory during adaptation. The broader impact of our work lies in establishing a stable inference platform for multiple clinical institutes while preserving data privacy and ensuring timely diagnosis.

Acknowledgments The work described in this paper was supported in part by the Research Grants Council of the Hong Kong Special Administrative Region, China, under Project T45-401/22-N; and by the Hong Kong Innovation and Technology Fund, under Project GHP/060/21GD.

Disclosure of Interests The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Azad, R., Aghdam, E.K., Rauland, A., Jia, Y., Avval, A.H., Bozorgpour, A., Karimijafarbigloo, S., Cohen, J.P., Adeli, E., Merhof, D.: Medical image segmentation review: The success of u-net. IEEE Transactions on Pattern Analysis and Machine Intelligence (2024) 1
2. Cao, H., Xu, Y., Yang, J., Yin, P., Yuan, S., Xie, L.: Multi-modal continual test-time adaptation for 3d semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 18809–18819 (2023) 2, 6
3. Chen, X., Wang, X., Zhang, K., Fung, K.M., Thai, T.C., Moore, K., Mannel, R.S., Liu, H., Zheng, B., Qiu, Y.: Recent advances and clinical applications of deep learning in medical image analysis. Medical image analysis **79**, 102444 (2022) 1
4. Chen, Z., Pan, Y., Ye, Y., Cui, H., Xia, Y.: Treasure in distribution: a domain randomization based multi-source domain generalization for 2d medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 89–99. Springer (2023) 7
5. Chen, Z., Pan, Y., Ye, Y., Lu, M., Xia, Y.: Each test image deserves a specific prompt: Continual test-time adaptation for 2d medical image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11184–11193 (2024) 2
6. Cho, Y., Kim, Y., Lee, D.: Beyond entropy: Style transfer guided single image continual test-time adaptation. arXiv preprint arXiv:2311.18270 (2023) 2
7. Haleem, A., Javaid, M., Singh, R.P., Suman, R.: Telemedicine for healthcare: Capabilities, features, barriers, and applications. Sensors international **2**, 100117 (2021) 2
8. Huang, Z., Wang, H., Deng, Z., Ye, J., Su, Y., Sun, H., He, J., Gu, Y., Gu, L., Zhang, S., et al.: Stu-net: Scalable and transferable medical image segmentation models empowered by large-scale supervised pre-training. arXiv preprint arXiv:2304.06716 (2023) 7

9. Jiang, J., Zhou, Q., Li, Y., Zhao, X., Wang, M., Ma, L., Chang, J., Zhang, J., Lu, X., et al.: Pcotta: Continual test-time adaptation for multi-task point cloud understanding. Advances in Neural Information Processing Systems **37**, 96229–96253 (2025) 2
10. Lee, D., Yoon, J., Hwang, S.J.: Becotta: Input-dependent online blending of experts for continual test-time adaptation. arXiv preprint arXiv:2402.08712 (2024) 2, 3, 5
11. Liu, C., Wang, L., Lyu, L., Sun, C., Wang, X., Zhu, Q.: Deja vu: Continual model generalization for unseen domains. arXiv preprint arXiv:2301.10418 (2023) 1
12. Liu, J., Yang, S., Jia, P., Zhang, R., Lu, M., Guo, Y., Xue, W., Zhang, S.: Vida: Homeostatic visual domain adapter for continual test time adaptation. arXiv preprint arXiv:2306.04344 (2023) 2, 7
13. Liu, Q., Dou, Q., Yu, L., Heng, P.A.: Ms-net: multi-site network for improving prostate segmentation with heterogeneous mri data. IEEE transactions on medical imaging **39**(9), 2713–2724 (2020) 2, 6
14. Nado, Z., Padhy, S., Sculley, D., D'Amour, A., Lakshminarayanan, B., Snoek, J.: Evaluating prediction-time batch normalization for robustness under covariate shift. arXiv preprint arXiv:2006.10963 (2020) 7
15. Shi, Z., Lyu, F., Liu, Y., Shang, F., Hu, F., Feng, W., Zhang, Z., Wang, L.: Controllable continual test-time adaptation. arXiv preprint arXiv:2405.14602 (2024) 2
16. Song, J., Lee, J., Kweon, I.S., Choi, S.: Ecotta: Memory-efficient continual test-time adaptation via self-distilled regularization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11920–11929 (2023) 2, 7
17. Wang, D., Shelhamer, E., Liu, S., Olshausen, B., Darrell, T.: Tent: Fully test-time adaptation by entropy minimization. arXiv preprint arXiv:2006.10726 (2020) 7
18. Wang, Q., Fink, O., Van Gool, L., Dai, D.: Continual test-time domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7201–7211 (2022) 2, 7
19. Wang, Y., Hong, J., Cheraghian, A., Rahman, S., Ahmedt-Aristizabal, D., Petersson, L., Harandi, M.: Continual test-time domain adaptation via dynamic sample selection. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 1701–1710 (2024) 2
20. Yang, X., Chen, X., Li, M., Wei, K., Deng, C.: A versatile framework for continual test-time domain adaptation: Balancing discriminability and generalizability. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 23731–23740 (2024) 2
21. Yang, X., Li, M., Yin, J., Wei, K., Deng, C.: Navigating continual test-time adaptation with symbiosis knowledge. In: Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence. pp. 5326–5334 (2024) 2
22. Zhang, R., Cheng, A., Luo, Y., Dai, G., Yang, H., Liu, J., Xu, R., Du, L., Du, Y., Jiang, Y., et al.: Decomposing the neurons: Activation sparsity via mixture of experts for continual test time adaptation. arXiv preprint arXiv:2405.16486 (2024) 2
23. Zhou, D., Gu, C., Xu, J., Liu, F., Wang, Q., Chen, G., Heng, P.A.: Repmode: learning to re-parameterize diverse experts for subcellular structure prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3312–3322 (2023) 5
24. Zhu, J., Bolsterlee, B., Chow, B.V., Song, Y., Meijering, E.: Uncertainty and shape-aware continual test-time adaptation for cross-domain segmentation of medical

images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 659–669. Springer (2023) 2
25. Zhu, J., Bolsterlee, B., Song, Y., Meijering, E.: Improving cross-domain generalizability of medical image segmentation using uncertainty and shape-aware continual test-time domain adaptation. Medical Image Analysis **101**, 103422 (2025) 2