HEAL: LEARNING-FREE SOURCE FREE UNSUPERVISED DOMAIN ADAPTATION FOR CROSS-MODALITY MEDICAL IMAGE SEGMENTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Growing demands for clinical data privacy and storage constraints have spurred advances in Source Free Unsupervised Domain Adaptation (SFUDA). SFUDA addresses the domain shift by adapting models from the source domain to the unseen target domain without accessing source data, even when target-domain labels are unavailable. However, SFUDA faces significant challenges: the absence of source domain data and label supervision in the target domain due to source free and unsupervised settings. To address these issues, we propose **HEAL**, a novel SFUDA framework that integrates **H**ierarchical denoising, **E**dge-guided selection, size-**A**ware fusion, and **L**earning-free characteristic. Large-scale cross-modality experiments demonstrate that our method outperforms existing SFUDA approaches, achieving state-of-the-art (SOTA) performance. The source code is publicly available at: https://anonymous.4open.science/r/HEAL-10C5.

1 INTRODUCTION

Medical image segmentation is crucial in diagnosis, treatment planning, and disease progression monitoring. Deep learning methods have shown remarkable success in this field Havaei et al. (2017)Isensee et al. (2021)He et al. (2025). However, these methods typically require large labeled datasets for training, which are often difficult and expensive to acquire due to the need for expert annotation. Furthermore, data collected from different institutions using various imaging modalities or acquired from different devices can exhibit a significant domain shift, which significantly degrades the cross-modality generalization capability of models Sandfort et al. (2019)Wachinger et al. (2018).

In response to the increasing challenges of domain shift, unsupervised domain adaptation (UDA) has emerged as a crucial research direction. UDA strategically utilizes labeled data from the source domain to enable model adaptation in the completely unlabeled target domain Oza et al. (2023). However, these methods require access to sensitive source domain data, which may raise privacy concerns or make collaboration between institutions difficult due to data sharing restrictions Abadi et al. (2016)Yang et al. (2019). Therefore, source free unsupervised domain adaptation (SFUDA) has emerged as a promising paradigm in which a pre-trained source model is adapted to unlabeled target domains without accessing data from the source domain Tian & Zhao (2024). The source free and unsupervised settings are particularly appealing in medical image analysis, where data privacy is paramount. However, SFUDA presents significant challenges due to the inaccessibility of source domain data and the absence of guidance in target domain adaptation. Several SFUDA approaches focus on generating pseudo-labels in the target domain and using them for self-training Dai et al. (2020); Zou et al. (2018). Other approaches leverage generative models to generate source-like data for domain gap mitigation Litrico et al. (2023), where diffusion models Ho et al. (2020) are promising in this paradigm Zeng et al. (2024). However, existing SFUDA methods often involve self-training or fine-tuning the pre-trained model on the target domain, which can still raise concerns about computational cost and potential privacy leakage.

In this work, we propose a novel SFUDA framework (HEAL) for cross-modality medical image segmentation. **HEAL** incorporates **H**ierarchical denoising (HD), **E**dge-guided selection (EGS), size-**A**ware fusion (SAF), and **L**earning-free characteristic.

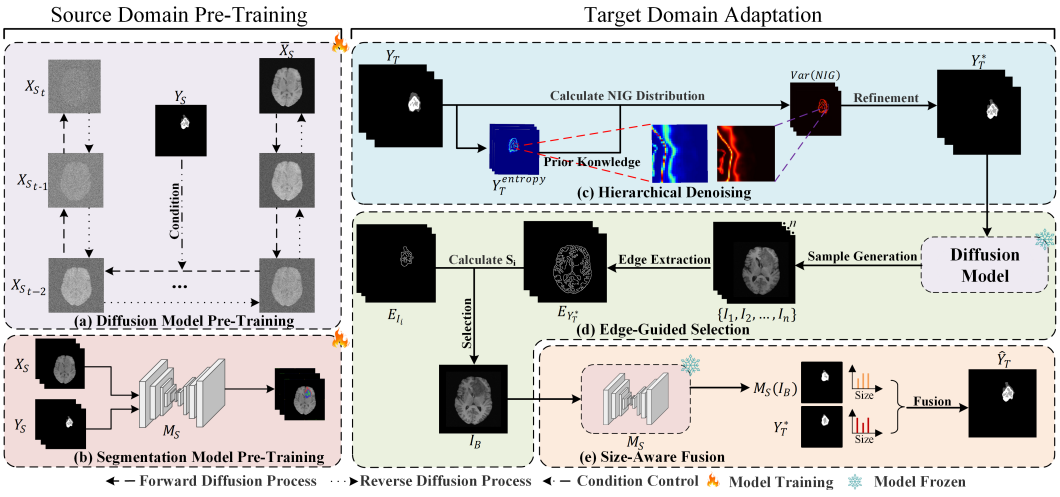

Figure 1: An overview of HEAL, illustrating its key components: hierarchical denoising, edge-guided selection, and size-aware fusion.

A key distinguishing feature of HEAL is the ”learning-free” characteristic, which sets it apart from conventional SFUDA methods. In HEAL, the model is exclusively pre-trained on the source domain, and no further training, fine-tuning, or parameter updates are performed during domain adaptation to the target domain. Instead, the entire adaptation process is conducted purely through inference. This stands in stark contrast to many existing SFUDA approaches, which typically involve a target-domain learning phase, such as self-training with pseudo-labelsDai et al. (2020); Zou et al. (2018), iterative optimization to reduce domain shift, or training auxiliary networks to bridge domain gaps. By eliminating the need for such training, HEAL not only enhances computational efficiency and simplifies deployment but also preserves the integrity of the pre-trained source model and reduces the risks associated with learning from noisy or limited target data. The contributions of our study are summarized as follows:

1. We introduce HEAL, a novel SFUDA framework for cross-modality medical image segmentation. HEAL can achieve effective domain adaptation without requiring further training or parameter updates to the pre-trained model, which inherently ensures superior data privacy and exceptional computational efficiency.
2. We propose hierarchical denoising that refines pseudo-labels by sequentially leveraging entropy uncertainty and Normal-Inverse Gaussian (NIG) uncertainty, effectively mitigating error accumulation in the pseudo-labels.
3. We introduce edge-guided selection to identify the most reliable samples generated by the diffusion model. By leveraging the inherent stochasticity of the diffusion model, multiple candidate images are synthesized, and the one with the highest structural consistency metric is selected as the most trustworthy.
4. We develop size-aware fusion that dynamically integrates pseudo-labels based on the size of the segmentation targets. This approach stabilizes the overall segmentation structure while enhancing the performance of small targets.

2 METHOD

Figure 1 shows the overview architecture of HEAL. First, during the pre-training stage, we train a segmentation model M_S and a diffusion model using source domain data $\{X_S, Y_S\}$, where X_S are the source domain data and Y_S are the label. Then, the segmentation model M_S is employed to generate initial pseudo-labels Y_T from the target domain data X_T. To mitigate error accumulation in pseudo-labels, we introduce HD to refine pseudo-labels utilizing entropy and NIG uncertainty maps. Subsequently, EGS is used to select the most reliable source-like sample I_B by measuring

the structural consistency. Finally, we design SAF to dynamically fuse the segmentations of EGS-selected samples $M_S(I_B)$ and the HD-refined pseudo labels Y_T^*.

2.1 Hierarchical Denoising

2.1.1 Entropy Denoising:

Given pseudo-labels $Y_T = M_S(X_T)$, we first compute voxel-wise entropy from the probability distribution of the initial pseudo-labels Y_T. Voxel-wise entropy is assigned to each class in each voxel as a measure of uncertainty, and higher entropy values indicate greater uncertainty. The voxel-wise entropy is used to guide the refinement of coarse error regions by filtering out voxels with high uncertainty. For a given voxel v and a set of C classes, let $P(v|c)$ denote the predicted probability of pseudo-label that voxel v belongs to class c. The entropy $H(v)$ of the voxel v is computed as:

$$H(v) = -\sum_{c=1}^{C} P(c|v) \cdot \log_2 \left(P(c|v)\right) \tag{1}$$

For each voxel v in the target domain, the entropy-refined pseudo-labels $Y_T^{entropy}$ can be calculated:

$$Y_T^{entropy}(v) = Y_T(v) \cdot \mathbb{I}_{H(v)\leq\tau_1} = \begin{cases} Y_T(v), & \text{if } H(v) \leq \tau_1 \\ 0, & \text{if } H(v) > \tau_1 \end{cases} \tag{2}$$

Where $\mathbb{I}_{\{condition\}}$ is the indicator function, and τ_1 is the entropy threshold.

2.1.2 Normal-Inverse Gaussian Variance Denoising:

To further mitigate fine-grained pseudo-label errors, we introduce NIG denoising. Specifically, we model the class probability p of each voxel in Y_T as a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$, where the mean μ and variance σ^2 are governed by a NIG prior. This prior is derived from the entropy-refined pseudo-labels $Y_T^{entropy}$. The class probability p can be calculated:

$$p(\mu, \sigma^2 \mid \alpha, \beta, \gamma, \omega) = \frac{\beta^\alpha \sqrt{\omega}}{\Gamma(\alpha)\sqrt{2\pi\sigma^2}} \left(\frac{1}{\sigma^2}\right)^{\alpha+1} \exp\left\{-\frac{2\beta + \omega(\gamma-\mu)^2}{2\sigma^2}\right\} \tag{3}$$

Where $\Gamma(\cdot)$ is the gamma function and the NIG parameters $\gamma, \omega, \alpha, \beta$ are calculated from $Y_T^{entropy}$. Specifically, γ is directly governed by $Y_T^{entropy}$, which anchors the distribution center to high-confidence regions. In contrast, α is modulated by the discrepancy between Y_T and $Y_T^{entropy}$. Meanwhile, ω and β are dynamically tuned based on the regional entropy $E(v)$, ensuring finer adjustments in regions of higher uncertainty. The NIG parameters can be calculated as follows:

$$\alpha = \kappa \cdot \left|Y_T - Y_T^{entropy}\right| + \epsilon, \beta = \zeta_1(1 - E(v)) + \zeta_2, \omega = \frac{1}{\eta_1 E(v) + \eta_2} \tag{4}$$

Where $\kappa, \zeta_1, \eta_1 > 0$ and $\epsilon, \zeta_2, \eta_2$ are small positive offsets. Here, we clarify that $E(v)$ is distinct from the voxel-wise entropy $H(v)$ used in the previous entropy denoising stage. While $H(v)$ quantifies uncertainty at the individual voxel level, $E(v)$ captures regional uncertainty by averaging entropy values within a local 3×3 voxel neighborhood around v.

After we approximate the NIG distribution, the NIG uncertainty is quantified through the variance of NIG distribution to generate a binary mask for refining the Y_T. The HD-refined pseudo-label Y_T^* can be obtained by:

$$Var(\text{NIG}) = \frac{\omega}{\beta(\alpha-1)}, Y_T^*(v) = Y_T(v) \cdot \mathbb{I}_{\{Var(\text{NIG})\leq\tau_2\}} \tag{5}$$

2.2 Edge-Guided Selection

Due to the inherent stochasticity of diffusion models, the quality of generated samples can vary significantly Song & Ermon (2020). Therefore, selecting high-fidelity samples is crucial to ensure effective segmentation by the source pre-trained model Ho & Salimans (2022). To address this issue,

we introduce EGS to select the most reliable generated sample based on the structural consistency metric S_i. First, we use the pre-trained diffusion model to generate n samples conditioned on Y_T^*. Then, we apply n sets of Canny edge detectors, where each set consists of two independent detectors: one for processing the generated sample I_i (where $i = 1, 2, ..., n$) and the other for the corresponding diffusion model conditions Y_T^*:

$$\{\{E_{I_1}, E_{Y_T^*}\}, \{E_{I_2}, E_{Y_T^*}\}, ..., \{E_{I_n}, E_{Y_T^*}\} = \text{Canny}(\{I_1, I_2, ..., I_n\}, Y_T^*) \tag{6}$$

Where E_{I_i} represents the edge map of the i_{th} generated image I_i. $E_{Y_T^*}$ represents the edge map of the corresponding diffusion model conditions Y_T^*. $\text{Canny}(\cdot, \cdot)$ denotes the application of the Canny edge detector. To quantify the consistency between the generated image and the conditions, we calculate the structural consistency metric S_i for each generated image I_i:

$$S_i = \frac{|E_{I_i} \cap E_{Y_T^*}|}{|E_{Y_T^*}|} \tag{7}$$

Where S_i is the structural consistency metric for the i_{th} generated image I_i. $|E_{I_i} \cap E_{Y_T^*}|$ represents the number of aligned edge pixels between the generated image I_i and the diffusion condition Y_T^*. $|E_{Y_T^*}|$ represents the total number of edge pixels in the diffusion model conditions Y_T^*. A higher S_i indicates greater consistency between the generated sample and the conditions, suggesting a more reliable sample. After calculating the structural consistency S_i, we select I_B with the highest structural consistency S_i for subsequent segmentation.

2.3 Size-Aware Fusion

The EGS-selected sample I_B demonstrates superior structural preservation for large anatomical structures due to its global consistency modeling Karras et al. (2022), whereas the HD-refined pseudo-labels Y_T^* provide reliable guidance for small objects Xu et al. (2024). Therefore, we propose SAF that dynamically selects the most reliable source for each target according to the size. Specifically, we fuse the categories with the smallest proportion of voxels in Y_T^* with other categories in $M_S(I_B)$. The final segmentation result $\widehat{Y}_T$ can be obtained by:

$$\widehat{Y}_T = \sum_{k=1}^{C} \lambda_k \cdot [\mathbb{I}(v_k = v_{min}) \cdot Y_T^* + \mathbb{I}(v_k \neq v_{min}) \cdot M_S(I_B)] \tag{8}$$

Where $v_{min} = min\{v_1, v_2, ..., v_C\}$ denotes the fewest voxels of the target category, $M_S(I_B)$ is the segmentation of the EGS-selected sample I_B , and $\lambda_k = \frac{v_k^{-1}}{\Sigma v_i^{-1}}$ is the dynamic weight of the smallest target.

3 Experiments and Results

3.1 Experimental Setup

3.1.1 Dataset and Evaluation Metrics:

We validate our method on the BraTS 2021 dataset Menze et al. (2014), Kvasir-SEG Jha et al. (2020), and CVC-ClinicDB Bernal et al. (2015). **BraTS 2021** provides 1,251 fully annotated 3D MRI volumes with four modalities: T1, contrast-enhanced T1 (T1ce), T2, and FLAIR. The segmentation targets within this dataset encompass three tumor subregions: whole tumor (WT), tumor core (TC), and enhancing tumor (ET). **Kvasir-SEG** is a widely used benchmark dataset consisting of 1,000 endoscopic images with pixel-wise annotated segmentation masks. **CVC-ClinicDB** comprises 612 still frames extracted from multiple colonoscopy video sequences with corresponding binary masks delineating the polyp regions.

We evaluate HEAL on two medical image segmentation tasks across four domain adaptation directions: T1→T1ce and T2→FLAIR for brain tumor segmentation, and Kvasir-SEG→CVC-ClinicDB and CVC-ClinicDB→Kvasir-SEG for polyp segmentation. Segmentation performance is quantitatively assessed using two standard metrics: the Dice similarity coefficient (Dice) Dice (1945) and the average surface distance (ASD) Taha & Hanbury (2015). For the polyp segmentation tasks, due

Method	T1→T1ce								T2→FLAIR							
	Dice (%) ↑				ASD (mm) ↓				Dice (%) ↑				ASD (mm) ↓			
	WT	TC	ET	Mean	WT	TC	ET	Mean	WT	TC	ET	Mean	WT	TC	ET	Mean
No Adaptation	64.3	57.2	55.1	58.9	3.3	2.0	3.2	2.8	67.2	63.0	50.9	60.4	3.2	5.1	2.4	3.6
Supervised	86.2	93.1	73.2	84.2	1.3	0.6	1.4	1.1	93.0	84.3	81.1	86.1	0.8	1.4	1.0	1.1
ProtoContraYu et al. (2023)	46.1	47.9	33.4	42.5	6.0	5.5	4.0	5.2	63.3	56.7	37.6	52.5	4.3	5.8	4.4	4.8
DPLChen et al. (2021)	54.3	49.9	44.8	49.7	5.3	3.7	3.6	4.2	58.8	36.4	18.5	37.9	4.0	7.0	7.8	6.3
IAPCCao et al. (2024)	59.4	58.8	34.8	51.0	7.3	6.1	6.1	6.5	65.1	53.4	30.9	49.8	4.5	5.4	5.8	5.2
UPLWu et al. (2023)	66.1	67.7	60.0	64.6	4.9	3.6	2.8	3.7	54.0	48.4	27.7	43.4	9.9	7.4	6.5	7.9
HEAL	**80.7**	**82.9**	**68.1**	**77.3**	**2.5**	**1.5**	**2.0**	**2.0**	**82.9**	**74.5**	**63.0**	**73.5**	**2.8**	**2.6**	**2.4**	**2.6**

Table 1: Performance comparison of SFUDA methods for brain tumor segmentation. Best results are in **bold**, second best are underlined.

Method	Kvasir-SEG→CVC-ClinicDB		CVC-ClinicDB→Kvasir-SEG	
	Dice (%) ↑	ASD (mm) ↓	Dice (%) ↑	ASD (mm) ↓
No Adaptation	65.1±21.5	5.8±11.1	60.2±27.0	23.7±20.4
Supervised	99.8±0.1	0.36±0.1	97.4±1.3	0.87±1.1
ProtoContraYu et al. (2023)	78.0±12.1	**0.9±0.6**	**68.9±26.4**	9.5±15.0
DPLChen et al. (2021)	67.2±24.0	8.0±12.4	66.1±33.1	**7.8±17.7**
IAPCCao et al. (2024)	73.3±29.6	4.4±7.9	61.2±26.3	13.7±13.2
UPLWu et al. (2023)	67.1±14.6	2.2±0.9	60.2±31.3	11.4±17.2
HEAL	**81.8±22.6**	4.3±8.2	66.5±28.8	9.15±11.6

Table 2: Performance comparison of SFUDA methods for polyp segmentation. Best results are in **bold**, second best are underlined.

to their single-class nature, HEAL is applied without the size-aware fusion, employing only hierarchical denoising and edge-guided selection. As the target domain is not used during training, we do not partition it into training and testing subsets. Instead, the model is directly performing inference on the full target domain to assess its adaptation performance.

3.1.2 Implementation Details:

All experiments were implemented using the PyTorch framework and conducted on an NVIDIA RTX 4090 GPU with 24GB memory. We used Med-DDPM Dorjsembe et al. (2024) with a noise schedule of 250 time steps t, and the nnUNet framework Isensee et al. (2021), which employs a ResNet-based U-Net architecture, trained for 300 epochs. Both the entropy threshold τ_1 and the NIG distribution variance threshold τ_2 were set to 0.2. In EGS, the number of generated samples n is 6, and the threshold of the Canny edge detector is 0.1.

3.2 Comparison with SOTA Methods

In our experiments, **No Adaptation** indicates training on the source domain and directly testing on the target domain without adaptation. **Supervised** indicates training and testing on the target domain. Furthermore, we compared HEAL with four SOTA SFUDA methods: 1) ProtoContra Yu et al. (2023), 2) DPL Chen et al. (2021), 3) UPL Wu et al. (2023), and 4) IAPC Cao et al. (2024). We reproduce all the comparison methods according to their official codes and configurations.

The quantitative evaluation results for cross-modality brain tumor segmentation are presented in Table 1. A substantial performance gap is observed between the **Supervised** and **No Adaptation** baselines in both directions, highlighting the severe domain shifts between the source and target domains. In both directions, **HEAL** remarkably outperforms all other SFUDA approaches on WT, TC, and ET, achieving the highest mean Dice of 77.3%, 73.5%, and the lowest mean ASD of 2.0mm, 2.6mm in T1→T1ce and T2→FLAIR, respectively. Moreover, compared to **No Adaptation**, the competitive performance of **HEAL** on both Dice and ASD can be attributed to the synergistic effect of HD, EGS, and SAF.

Table 2 presents the quantitative results for cross-modality polyp segmentation. There is a clear performance gap between the **Supervised**, with Dice scores of 99.8% and 97.4%, and the **No Adaptation**, which achieves Dice scores of only 65.1% and 60.2% in the Kvasir-SEG→CVC-ClinicDB

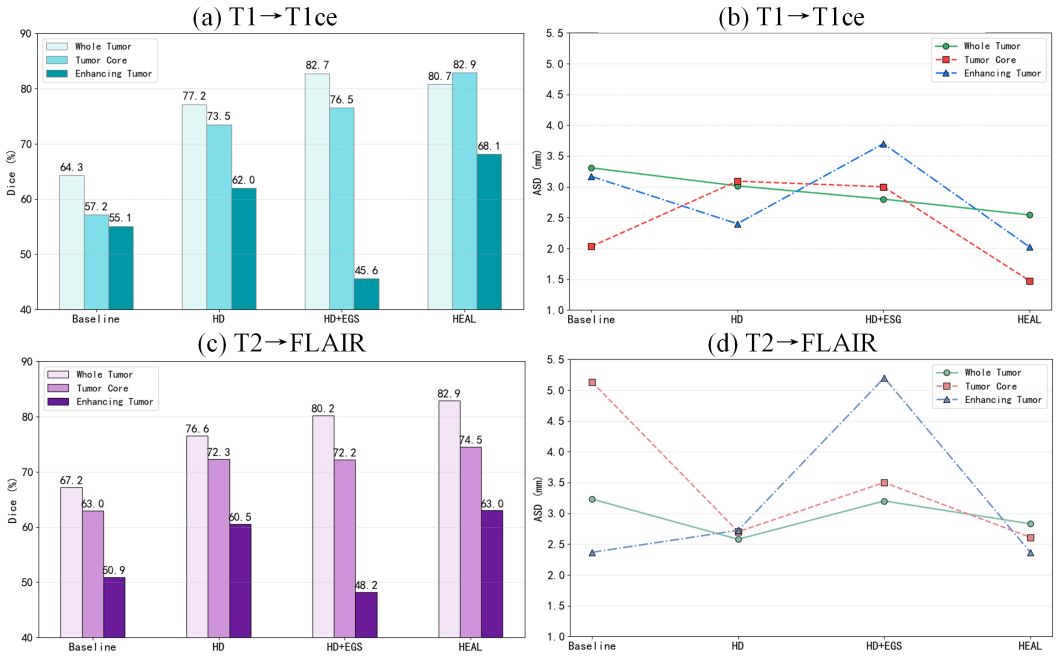

Figure 2: Ablation study of each component in HEAL on brain tumor segmentation. HD refers to the results obtained by hierarchical denoising, HD+EGS denotes the combination of hierarchical denoising and edge-guided selection, and HEAL represents the full method with hierarchical denoising, edge-guided selection, and size-aware fusion.

Method	Kvasir-SEG→CVC-ClinicDB		CVC-ClinicDB→Kvasir-SEG	
	Dice (%) ↑	ASD (mm) ↓	Dice (%) ↑	ASD (mm) ↓
Baseline	65.1±21.5	5.8±11.1	60.2±27.0	23.7±20.4
Entropy-refined	72.6±24.5	5.8±12.2	63.1±33.0	17.9±39.6
HD	<u>76.4±21.3</u>	<u>5.1±10.9</u>	<u>63.8±35.4</u>	<u>16.2±20.8</u>
HEAL (HD+EGS)	**81.8±22.6**	**4.3±8.2**	**66.5±28.8**	**9.15±11.6**

Table 3: Ablation study of HEAL on polyp segmentation. Entropy-refined refers to the pseudo-labels refined by voxel-wise entropy. Best results are in **bold**, second best are <u>underlined</u>.

and CVC-ClinicDB→Kvasir-SEG directions, respectively. This substantial gap highlights the challenges posed by domain shift, thereby underscoring the necessity for effective adaptation methods such as HEAL, which demonstrates strong domain adaptation capability. Specifically, in the Kvasir-SEG→CVC-ClinicDB direction, **HEAL** achieves the best Dice of 81.8%, a remarkable improvement over the **No Adaptation** and surpassing other SFUDA methods. In the reverse direction, CVC-ClinicDB→Kvasir-SEG, **HEAL** also delivers highly competitive performance, obtaining a Dice of 66.5% and an ASD of 9.15mm, which are marked as second-best results among the compared adaptation techniques. These results demonstrate **HEAL**'s effectiveness in significantly mitigating domain shift and enhancing segmentation accuracy across different polyp datasets and modalities without relying on source data.

3.3 Ablation Study

To assess the individual contributions of each component in HEAL, we perform ablation studies on brain tumor segmentation and polyp segmentation. For brain tumor segmentation, we evaluate four configurations, **Baseline** (No Adaptation), **HD**, **HD+EGS**, and **HEAL** (HD+EGS+SAF), under two adaptation adaptations: T1→T1ce and T2→FLAIR. For polyp segmentation, we conduct ablation under two adaptation directions: Kvasir-SEG→CVC-ClinicDB and CVC-ClinicDB→Kvasir-SEG, using four configurations: **Baseline**, **Entropy-refined**, **HD**, and **HEAL** (HD+EGS).

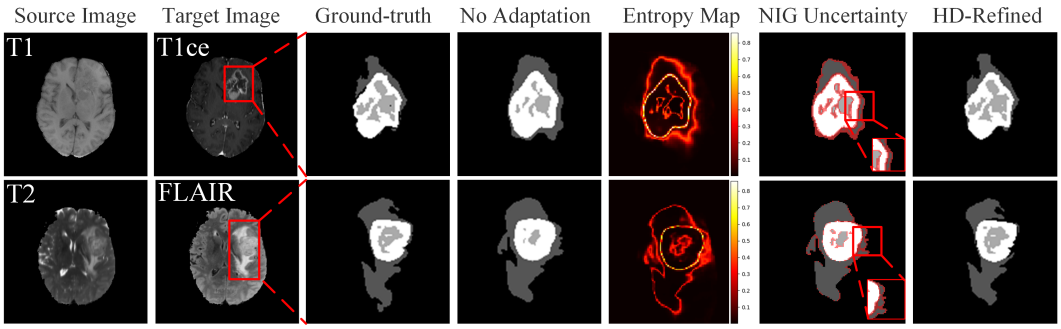

Figure 3: Qualitative ablation study of hierarchical denoising on brain tumor segmentation. The first row corresponds to patient #0000, and the second row to patient #0051.

As shown in Figure 2 (a), in the T1→T1ce direction, **HEAL** demonstrably enhances segmentation performance, yielding significant Dice improvements over the **Baseline** of 16.4%, 25.7%, and 13.0% for WT, TC, and ET, respectively. Figure 2 (b) reveals that **HD** reduces ASD for WT relative to the **Baseline**. Furthermore, the integration of EGS and SAF in **HEAL** further diminishes ASD to 2.8mm and 2.6mm, respectively. Similarly, Figure 2 (c) demonstrates that **HEAL** achieves Dice gains of 15.7%, 11.5%, and 12.1% compared to the **Baseline** in the FLAIR→T2 direction. As visualized in Figure 2 (d), **HEAL** effectively minimizes ASD across all tumor sub-regions in the FLAIR→T2 direction. These results collectively underscore the significant and consistent improvements in segmentation performance achieved by **HEAL** across different modalities.

Table 3 demonstrates that **HEAL** consistently enhances polyp segmentation performance across both adaptation directions. Beginning with the **Baseline**, the **Entropy-refined** led to a substantial Dice improvement of 7.5%, particularly in the Kvasir-SEG→CVC-ClinicDB direction. As the entropy-refined pseudo-labels provide reliable prior knowledge for subsequent NIG denoising, **HD** further enhances segmentation performance, leading to consistent improvements in Dice and reductions in ASD across both directions. Finally, integrating **EGS** with **HD**, **HEAL** achieves the best overall results, with Dice improvements of 16.7% and 6.3% over the baseline in Kvasir-SEG→CVC-ClinicDB and CVC-ClinicDB→Kvasir-SEG, respectively. Additionally, ASD is substantially reduced, particularly in the CVC-ClinicDB→Kvasir-SEG. These results highlight the effectiveness and complementary contributions of hierarchical denoising and edge-guided selection within **HEAL**.

Figure 3 visually demonstrates the effectiveness of HD in the refinement of pseudo-labels. In the **No Adaptation** column, we observe that the segmentation results exhibit noticeable inaccuracies compared to the **Ground-truth**. The **Entropy Map** is a heatmap that highlights the regions of entropy uncertainty, primarily concentrated around the coarse errors. Subsequently, the **NIG Uncertainty** further refines these uncertainty regions, pinpointing areas requiring more precise error correction, as highlighted by the red bounding boxes. This qualitative assessment underscores the critical role of HD in enhancing the reliability of pseudo-labels and improving the overall segmentation performance of HEAL.

To further investigate the effectiveness of HEAL in aligning feature distributions across domains, we visualize the feature distribution using t-SNE, as shown in Figure 4. In both adaptation directions, the feature representations of samples from the source domain, the target domain, and source-like samples generated by the diffusion model are projected into a 2D space. The visualizations demonstrate that the source-like features lie closer to the source domain than the target domain, indicating that the diffusion model effectively re-renders the target domain's structural information into a new representation that is stylistically aligned with the source domain. By leveraging the source knowledge captured during diffusion model pretraining, HEAL mitigates the risks of over-segmentation and information loss that may arise from hierarchical denoising, ultimately facilitating more robust and accurate domain adaptation.

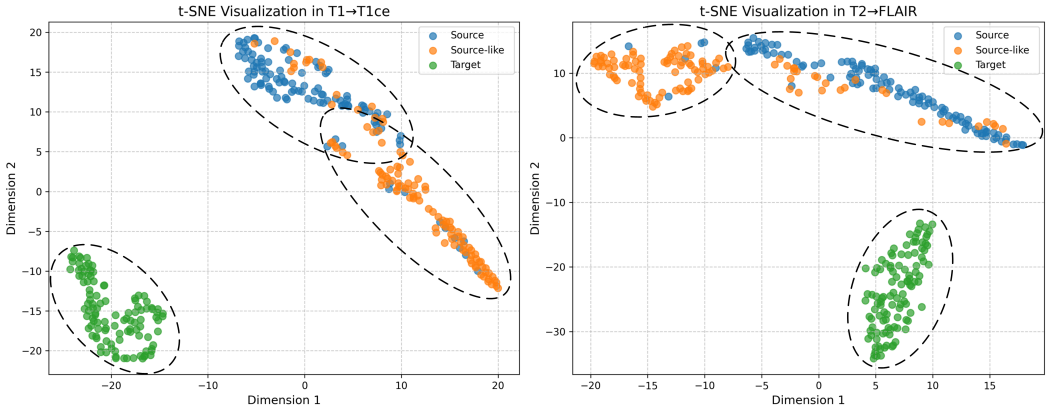

Figure 4: t-SNE visualizations of feature distribution in T1→T1ce and T2→FLAIR on BraTS2021.

4 Conclusion

In this work, we proposed HEAL, a novel SFUDA framework for cross-modality medical image segmentation without any training when performing target domain adaptation. The hierarchical denoising effectively mitigates error accumulation through entropy and NIG denoising, then the edge-guided selection ensures diffusion-generated samples preserve the critical anatomical structure and transfer the target domain knowledge to the source domain. Finally, the size-aware fusion effectively fuses the reliable classes in the HD-refined pseudo-labels and the segmentation of the EGS-selected sample. Experimental results on brain tumor and polyp segmentation tasks validate the superior performance of our method compared to existing SFUDA approaches.

A key aspect of HEAL's mechanism is that the framework does not attempt to make the source model directly interpret the target image. Instead, it leverages the structural information encapsulated in the HD-refined pseudo-label. The diffusion model, conditioned on the target image, is tasked with generating a source-like sample that is stylistically consistent with the source domain but structurally aligned with the target's anatomy. Therefore, the segmentation model is not processing the target image it has never seen; it is segmenting a synthesized source-like image where the tumor structures have been rendered according to the guidance from. This process effectively translates the adaptation challenge from a difficult cross-modality recognition problem into a more manageable in-domain segmentation task for the frozen source model.

limitations: The effectiveness of HEAL is inherently linked to the generalization capability of the pre-trained segmentation model. When the domain shift is excessively large, the initial pseudo-labels may be of low quality, which propagate errors through the subsequent process and ultimately degrade HEAL's performance. Furthermore, the learning-free nature of our approach may limit its ability to capture nuanced or novel target-specific features that could otherwise be addressed by methods involving target-domain parameter update.

References

Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pp. 308–318, 2016.

Jorge Bernal, F Javier Sánchez, Gloria Fernández-Esparrach, Debora Gil, Cristina Rodríguez, and Fernando Vilariño. Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized medical imaging and graphics*, 43:99–111, 2015.

Yihong Cao, Hui Zhang, Xiao Lu, Zheng Xiao, Kailun Yang, and Yaonan Wang. Towards source-free domain adaptive semantic segmentation via importance-aware and prototype-contrast learning. *IEEE Transactions on Intelligent Vehicles*, 2024.

Cheng Chen, Quande Liu, Yueming Jin, Qi Dou, and Pheng-Ann Heng. Source-free domain adaptive fundus image segmentation with denoised pseudo-labeling. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24*, pp. 225–235. Springer, 2021.

Dengxin Dai, Christos Sakaridis, Simon Hecker, and Luc Van Gool. Curriculum model adaptation with synthetic and real data for semantic foggy scene understanding. *International Journal of Computer Vision*, 128(5):1182–1204, 2020.

Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3): 297–302, 1945.

Zolnamar Dorjsembe, Hsing-Kuo Pao, Sodtavilan Odonchimed, and Furen Xiao. Conditional diffusion models for semantic 3d brain mri synthesis. *IEEE Journal of Biomedical and Health Informatics*, 28(7):4084–4093, 2024.

Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, and Hugo Larochelle. Brain tumor segmentation with deep neural networks. *Medical image analysis*, 35:18–31, 2017.

Weiheng He, Feifei Du, and Shilong Zhang. Non-separation method-based finite-time synchronization of discrete-time fractional-order complex-valued neural networks with time-varying delays. *Neurocomputing*, 648:130732, 2025. ISSN 0925-2312.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance, 2022.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.

Debesh Jha, Pia H Smedsrud, Michael A Riegler, Pål Halvorsen, Thomas de Lange, Dag Johansen, and Håvard D Johansen. Kvasir-seg: A segmented polyp dataset. In *International Conference on Multimedia Modeling*, pp. 451–462. Springer, 2020.

Tero Karras, Miika Aittala, Samuli Laine, and Timo Aila. Elucidating the design space of diffusion-based generative models. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2022. Curran Associates Inc. ISBN 9781713871088.

Mattia Litrico, Alessio Del Bue, and Pietro Morerio. Guiding pseudo-labels with uncertainty estimation for source-free unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7640–7650, 2023.

Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34 (10):1993–2024, 2014.

Poojan Oza, Vishwanath A Sindagi, Vibashan Vishnukumar Sharmini, and Vishal M Patel. Unsupervised domain adaptation of object detectors: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

Veit Sandfort, Ke Yan, Perry J Pickhardt, and Ronald M Summers. Data augmentation using generative adversarial networks (cyclegan) to improve generalizability in ct segmentation tasks. *Scientific reports*, 9(1):16884, 2019.

Yang Song and Stefano Ermon. Improved techniques for training score-based generative models. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.

Abdel Aziz Taha and Allan Hanbury. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC medical imaging*, 15:1–28, 2015.

Qing Tian and Mengna Zhao. Generation, division and training: A promising method for source-free unsupervised domain adaptation. *Neural Networks*, 172:106142, 2024.

Christian Wachinger, Martin Reuter, and Tassilo Klein. Deepnat: Deep convolutional neural network for segmenting neuroanatomy. *NeuroImage*, 170:434–445, 2018.

Jianghao Wu, Guotai Wang, Ran Gu, Tao Lu, Yinan Chen, Wentao Zhu, Tom Vercauteren, Sébastien Ourselin, and Shaoting Zhang. Upl-sfda: Uncertainty-aware pseudo label guided source-free domain adaptation for medical image segmentation. *IEEE Transactions on Medical Imaging*, 42 (12):3932–3943, December 2023. ISSN 1558-254X. doi: 10.1109/tmi.2023.3318364.

Chenchu Xu, Tong Zhang, Dong Zhang, Dingwen Zhang, and Junwei Han. Deep generative adversarial reinforcement learning for semi-supervised segmentation of low-contrast and small objects in medical images. *IEEE Transactions on Medical Imaging*, 2024.

Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2):1–19, 2019.

Qinji Yu, Nan Xi, Junsong Yuan, Ziyu Zhou, Kang Dang, and Xiaowei Ding. Source-free domain adaptation for medical image segmentation via prototype-anchored feature alignment and contrastive learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 3–12. Springer, 2023.

Hongye Zeng, Ke Zou, Zhihao Chen, Rui Zheng, and Huazhu Fu. Reliable source approximation: Source-free unsupervised domain adaptation for vestibular schwannoma mri segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 622–632. Springer, 2024.

Yang Zou, Zhiding Yu, BVK Kumar, and Jinsong Wang. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *Proceedings of the European conference on computer vision (ECCV)*, pp. 289–305, 2018.