Quest for Clone: Test-time Domain Adaptation for Medical Image Segmentation by Searching the Closest Clone in Latent Space

Hritam Basak[1], Zhaozheng Yin

Dept. of Computer Science, Stony Brook University, NY, USA
[1]hbasak@cs.stonybrook.edu

Abstract. Unsupervised Domain Adaptation (UDA) aims to align labeled source distribution and unlabeled target distribution by mining domain-agnostic feature representation. However, adapting the source-trained model for new target domains after the model is deployed to users poses a significant challenge. To address this, we propose a generative latent search paradigm to reconstruct the closest clone of every target image from the source latent space. This involves utilizing a test-time adaptation (TTA) strategy, wherein a latent optimization step finds the closest clone of each target image from the source representation space using variational sampling of source latent distribution. Thus, our method facilitates domain adaptation without requiring target-domain supervision during training. Moreover, we demonstrate that our approach can be further fine-tuned using a few labeled target data without the need for unlabeled target data, by leveraging global and local label guidance from available target annotations to enhance the downstream segmentation task. We empirically validate the efficacy of our proposed method, surpassing existing UDA, TTA, and SSDA methods in two domain adaptive image segmentation tasks. Code is available at GitHub.

Keywords: Domain Adaptation · Segmentation · Variational Inference.

1 Introduction

Supervised deep learning techniques have excelled with abundant annotated natural image datasets, but this success does not always translate to medical imagery. Clinician-grade annotations are the gold standard, but acquiring them is time-consuming and requires deep domain expertise. Hence, methods alleviating this requirement are highly expedient. Moreover, medical images often contain significant domain shifts, especially across institutions and device modalities, leading to variations in image intensities and appearance. Models trained solely on one dataset struggle to adapt to new domains with such profound distribution shifts. This emphasizes the need for solutions tailored to address domain shifts and alleviate the annotation effort on new domains.

Lately, Unsupervised Domain Adaptation (**UDA**) that uses labeled source ($\mathbf{S_L}$) and unlabeled target data ($\mathbf{T_U}$) for training has been instrumental [18]. Existing approaches include image reconstruction [25], self-training [1], divergence

minimization [12], etc. Recently, a dual-adaptation framework was proposed in [13] to align low-level edge information and high-level semantics separately for UDA. A Dual-Scheme-Fusion-Network (DSFN) [31] was also introduced for a balanced bidirectional domain adaptation. Another popular latent alignment was coined using an adversarial approach [6]. However, as exemplified by [20], these methods focus on mining domain-invariant representations, requiring unlabeled target images being available for training. While acquiring target images in new domains is doable, retraining the model for every new target domain is patently impractical, particularly after models are deployed to users [16].

To this end, test-time adaptation (**TTA**) and domain generalization (**DG**) are promising and practical solutions. In TTA, only labeled source data ($\mathbf{S_L}$) are available during training, whereas the model adapts to target data while inference. A seminal work by combining TTA and DG, using a dictionary-based consistency was proposed in [11]. Another promising work in a similar direction proposed prototypical TTA using feature distance [10]. SATTA [28] was introduced to address the existing challenges by utilizing uncertainty estimation for each semantic category and adaptive learning rate. While promising, they rely heavily on a high-quality pre-trained model and data alignment across domains, making them ineffective in addressing label shift [29] and correlation shift [19].

Recently, semi-supervised domain adaptation (**SSDA**) was introduced by proposing an asymmetric co-training for domain-agnostic information retrieval [14]. In SSDA, labeled source ($\mathbf{S_L}$), unlabeled target ($\mathbf{T_U}$), and a few labeled target data ($\mathbf{T_L}$) are available for training. DisCL [1] proposed another SSDA paradigm using contrastive learning to disentangle content and style. However, they tend to cause overfit and bias toward source data due to their abundance in the training set, leading to degraded adaptation ability while inference.

In this work, we introduce an innovative domain adaptation framework that eliminates the necessity for access to target data during training. Our approach involves training a segmentation model solely on labeled source data, followed by searching for a source-like clone in the source latent space for every target image. This clone serves as a proxy or surrogate of the target image to be processed by the model. Specifically, our contributions are: **(1) New TTA Paradigm**: We present a DA paradigm that circumvents the conventional reliance on target data, paving the way for more efficient and resource-conservative test-time adaptation. **(2) Enhanced Source-like Target Clone**: Our approach precisely maps every target image to its closest *clone* within the source latent space by a latent optimization strategy. By finding the clone of every target image from the source latent space, we introduce an effective way of adapting to new target domains, resulting in superior performance. **(3) Scalability to SSDA$^-$**: We further introduce global and local semantic guidance in a new SSDA$^-$ setting to explore beyond TTA's adaptation ability. Different from the traditional SSDA that uses labeled source ($\mathbf{S_L}$), both labeled ($\mathbf{T_L}$) and unlabeled ($\mathbf{T_U}$) target data, our SSDA$^-$ only uses $\mathbf{S_L}$ and $\mathbf{T_L}$, greatly reducing the training data acquisition effort in new target domains. Upon evaluation, our proposed

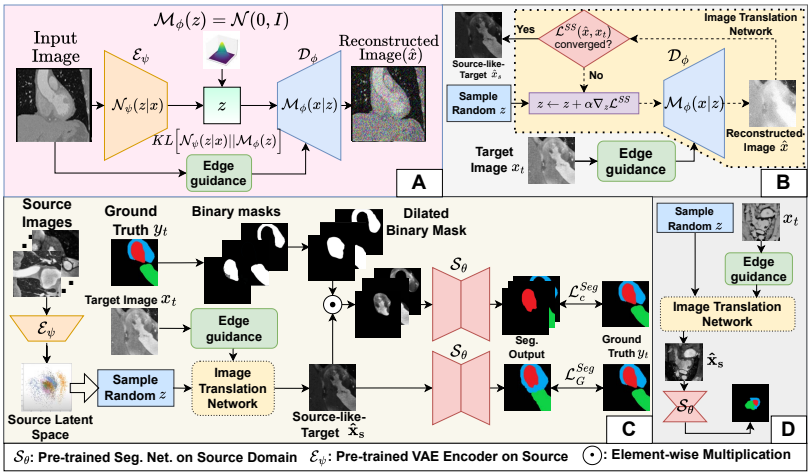

Fig. 1: **(A)** Training VAE via image reconstruction on source domain, obtaining latent feature representation through $\mathcal{E}_\psi$, **(B)** Clone a target image in the source latent space. A point z is sampled randomly from source latent space, which is then iteratively optimized through the proposed algorithm (subsection 2.2), resulting in a source-like clone $\hat{x}_s$ for a target image x_t, **(C)** For SSDA$^-$, we fine-tune $\mathcal{S}_\theta$ using a few target annotations. **(D)** The inference pipeline.

method performs superior on multi-organ and multi-modal benchmark datasets, compared to existing UDA, TTA, and SSDA works.

2 Proposed Method

Given source-domain image-label pairs $\{\mathcal{X}_s, \mathcal{Y}_s\}$, we first propose to leverage VAE-based reconstruction network to learn source distribution while concurrently normalizing the latent space $\mathscr{L}_s$ effectively (subsection 2.1). Then, we use VAE-decoder to find the clone of a target sample from $\mathscr{L}_s$ (subsection 2.2). We further explain the SSDA$^-$ setting in subsection 2.3.

2.1 Variational Sampling of Source Latent Space

In variational inference-based models using VAE [8], the underlying premise revolves around a dual-stage data generation: (i) a random sample from a distribution $\mathcal{M}_\phi(z)$ of a latent variable z is generated, (ii) given a specific latent variable, generate data x by sampling from $\mathcal{M}_\phi(x|z)$. However, $\mathcal{M}_\phi(x|z)$ and $\mathcal{M}_\phi(z)$ are governed by unknown parameters, whereas the posterior $\mathcal{M}_\phi(z|x)$ is computationally intractable. A new variational distribution $\mathcal{N}_\psi(z|x)$ is introduced to circumvent these obstacles. With this, we decompose the log-likelihood of observed data into an evidence lower bound (ELBO) based on KLD [9] term:

$$\mathcal{L}(\psi, \phi) = \mathbb{E}_{\mathcal{N}_\psi(z|x)}[\log(\mathcal{M}_\phi(x|z))] - KL\Big[\mathcal{N}_\psi(z|x)||\mathcal{M}_\phi(z)\Big] \quad (1)$$

Equation 1 serves as the lower bound of the log-likelihood and can be maximized in a VAE by parameterizing $\mathcal{N}_\psi(z|x)$ and $\mathcal{M}_\phi(x|z)$ using encoder-decoder network $\{\mathcal{E}_\psi, \mathcal{D}_\phi\}$, where ψ, ϕ denote the parameters of VAE encoder and decoder, respectively. With this, we train the VAE encoder-decoder $\{\mathcal{E}_\psi, \mathcal{D}_\phi\}$ on source domain image x to reconstruct $\hat{x}$, where we approximate the first term in Equation 1 to a pixel-wise reconstruction loss $\mathcal{L}^R$ on $\mathcal{D}_\phi$:

$$\mathbb{E}_{\mathcal{N}_\psi(z|x)}[\log(\mathcal{M}_\phi(x|z))] \approxeq \mathcal{L}^R = ||\hat{x} - x||^2 \tag{2}$$

Additionally, we concatenate domain-invariant edge information (e.g., basic Sobel edge features) with intermediate features in VAE decoder. The overall VAE training is shown in Figure 1A. By initializing $\mathcal{M}_\phi(z) = \mathcal{N}(0, I)$, and due to inherent property of VAE, the source latent space $\mathscr{L}_s$ is approximated to a normal distribution using the KLD term. Post training, $\mathcal{D}_\phi$ operates as a pivotal sampler in two steps: (i) a sample z is drawn from a standard Gaussian distribution $\mathcal{N}(0, I)$, (ii) x is generated from the conditional distribution $\mathcal{M}_\phi(x|z)$. Given a target image x_t, first, we prove the existence of its source-like closet *clone* from $\mathscr{L}_s$, and then we present how to find the clone in the following subsection.

2.2 Source-like Clone of a Target Image

Lemma: Let's denote m i.i.d. points from source distribution $\mathscr{L}_s$ as $\mathcal{X}^m = \{x^1, x^2, .., x^m\}$ (m is large) and x_t be a point from target distribution. The fundamental postulate delineating the existence of the closest clone (in terms of distance metric $\mathfrak{d}$) is: *if* $\mathfrak{d}(\hat{x}, x_t) > \mathfrak{d}(\hat{x}_s, x_t)$, $\forall \hat{x} \in \mathcal{X}^m$, *then* x_t*'s closest clone* $\hat{x}_s$ *certainly exists in* $\mathscr{L}_s$*, i.e.,* $\hat{x}_s \in \mathcal{X}^m$.
Proof: We denote a p-radius sphere around point x_t as $\mathcal{S}_p(x_t) = \{x \in \mathcal{X}^m : \mathfrak{d}(x, x_t) < p\}$ using distance metric $\mathfrak{d}$. Using a separable metric space [4], the probability of non-empty sphere $\mathbb{P}\Big(\mathcal{S}_p(x_t)\Big)$ can be defined as:

$$\mathbb{P}\Big(\mathcal{S}_p(x_t)\Big) \triangleq \int_{\mathcal{S}_p(x_t)} \mathscr{L}_s(x)\, dx > 0; \tag{3}$$

The probability of no points $x_k \in \mathcal{X}^m$ are within $\mathcal{S}_\epsilon(x_t)$ with radius ϵ:

$$\mathbb{P}\Big(\min_k \mathfrak{d}(x_k, x_t) \geq \epsilon\Big) = \Big(1 - \mathbb{P}\Big(\mathcal{S}_\epsilon(x_t)\Big)\Big)^m, \forall x_k \in \mathcal{X}^m \tag{4}$$

Hence, the probability of $\hat{x}_s$, i.e. the closest point of x_t from source latent space $\mathscr{L}_s$ being within the sphere $\mathcal{S}_\epsilon(x_t)$ of radius ϵ is:

$$\mathbb{P}\Big(\hat{x}_s \in \mathcal{S}_\epsilon(x_t)\Big) = 1 - \Big(1 - \mathbb{P}\Big(\mathcal{S}_\epsilon(x_t)\Big)\Big)^m \tag{5}$$

$$= 1 \text{ for } m \text{ approaches } \infty \tag{6}$$

As $0 < \big[1 - \mathbb{P}\big(\mathcal{S}_\epsilon(x_t)\big)\big] < 1$, $m \to \infty$ (infinite points can be sampled from $\mathscr{L}_s$), $\exists\, \hat{x}_s \in \mathcal{X}^m$ i.e., a closest clone within small distance ϵ from x_t, with $\mathbb{P}$=1.

Finding the Closest Clone Given a target sample x_t, we perform latent optimization to find the target's source-like clone $\hat{x}_s$:

$$\hat{x}_s = \mathcal{D}_\phi(\hat{z}_s) : \mathfrak{d}(\hat{x}, x_t) > \mathfrak{d}(\hat{x}_s, x_t), \ \forall \hat{x} = \mathcal{D}_\phi(z) \sim \mathscr{L}_s(x), \ z \sim \mathcal{N}(0, I) \tag{7}$$

SSIM loss [24] $\mathcal{L}^{SS}$ is used for $\mathfrak{d}$ (refer to supplementary file for different distance metrics). As seen in Figure 1B, the clone search is an optimization step, i.e., gradient descent on z to find the closest clone $\hat{x}_s$ as proxy of target image x_t:

$$\hat{x}_s = \mathcal{D}_\phi(\hat{z}_s); \ \ \hat{z}_s = \underset{z}{argmin}\, \mathcal{L}^{SS}(x_t, \mathcal{D}_\phi(z)) \tag{8}$$

Here $\hat{z}_s$ is obtained by gradient descent: $z \leftarrow z + \alpha \nabla_z \mathcal{L}^{SS}$ with learning rate α.

For TTA, $\hat{x}_s$ (the closest clone of target image x_t) is directly fed into $\mathcal{S}_\theta$ to obtain x_t's segmentation mask (Figure 1D), where $\mathcal{S}_\theta$ is a segmentation network trained only using labeled source data. For SSDA$^-$, we can further fine-tune $\mathcal{S}_\theta$ as described below.

2.3 Semi-supervised Fine-tuning

In TTA, the segmentation model $\mathcal{S}_\theta$ is trained solely on the source domain, without incorporating any data from the target domain. However, if a few labeled data becomes available from the target domain, we can enhance $\mathcal{S}_\theta$ further in the SSDA$^-$ paradigm (defined in section 1).

As shown in Figure 1C, given a target label y_t, we extract a dilated version (m'_c) of binary mask m_c indicating the c^{th} class in target image x_t ($c \in \mathcal{C}$; $\mathcal{C}$ indicating total number of classes). m'_c is element-wisely multiplied ($\odot$) with the source-like clone $\hat{x}_s$ of the target image and is fed into $\mathcal{S}_\theta$. Subsequently, the output is employed to compute the local class-wise CE-loss $\mathcal{L}_c^{Seg}$:

$$\mathcal{L}_c^{Seg} = \mathcal{L}_{CE}\big(\mathcal{S}_\theta(\hat{x}_s \odot m'_c), y_t \odot m_c\big) \tag{9}$$

The class-wise segmentation can guide $\mathcal{S}_\theta$ to preserve local semantics for pixels in each class of source-like cloned target image $\hat{x}_s$ and provides strong supervision towards accurate segmentation in the class-boundary regions. We also propose a global image-level semantic supervision to complement this per-class guidance to enhance global harmony in the segmentation task. Specifically, we utilize a global CE loss for image-level supervision: $\mathcal{L}_G^{Seg} = \mathcal{L}_{CE}(\mathcal{S}_\theta(\hat{x}_s), \hat{y})$. Finally, $\mathcal{S}_\theta$ is fine-tuned using the gradient of combined local and global segmentation losses using hyperparameters γ and δ:

$$\theta' \leftarrow \theta + \gamma \nabla_\theta \sum_c \mathcal{L}_c^{Seg} + \delta \nabla_\theta \mathcal{L}_G^{Seg} \tag{10}$$

3 Experiments and Results

Datasets: The evaluation is performed on two widely-used domain-adaptive medical image segmentation benchmarks consisting of multiple structures/tumors and various modalities: **(1)** Cardiac structure segmentation from Multi-Modality Whole Heart Segmentation (MMWHS) Challenge dataset [30], and **(2)** Brain

tumor segmentation in MRI images from BraTS2018 dataset [15]. MMWHS contains 20 unpaired CT and 20 MRI volumes collected from multiple sources. Following [21], we choose four classes for domain adaptive segmentation: Left Ventricle (LV), Left Atrium (LA), Myocardium (MYO), and Ascending Aorta (AA). BraTS contains MRI volumes of 285 patients with four modalities: T1CE, T1, T2, and FLAIR. The rest of the experimental settings are consistent with previous works [1,13,14] for a fair comparison: **(a) Source→Target:** CT→MRI and MRI→CT in MMWHS. T2→{T1CE, T1, FLAIR} in BraTS. **(b) SSDA:** 1 or 5 labeled samples from the target domain are available.

Implementation Details: We implement our work in Python environment using a Nvidia TeslaV100 GPU with 32GB RAM. Training the model requires $2-3$ hours, whereas inference typically takes $5-8$ seconds for every test image in the new target domain, due to the clone-search in test time adaptation. We use U-Net [17] as the segmentation network $\mathcal{S}_\theta$. VAE was trained using RMSProp with a learning rate of 0.0001 whereas Adam optimizer was employed for training $\mathcal{S}_\theta$. α, γ, δ are set to $0.01, 0.5, 0.5$, respectively by validation.

Evaluation Metrics: Dice Similarity Coefficient (DSC), Hausdorff Distance (HD), and Average Surface Distance (ASD).

3.1 Performance Evaluation on MMWHS

We compare our proposed method with state-of-the-art on UDA, TTA, and SSDA tasks, as shown in Table 1. Following previous works [5,2,7], we report DSC/ASD on MRI→CT and CT→MRI. SECASA [5] selectively maximizes low-entropy or high-confidence regions in segmentation but provides no substantial effort to bridge the domain gap. On the other hand, CPCL [2] uses prototypical contrastive learning, which is useful for image-level classification but lacks pixel-level supervision for segmentation. A similar drawback is present in T-MAtt [7], which mines long-range dependencies using meta-attention but lacks a pixel-level supervision module. We address these shortcomings by finding the target's clone from the source latent space, along with retaining its original structural integrity by edge guidance. When compared to TTA methods [22,27,26], they produce subpar performance (on average $\sim 5-6\%$ drop than ours in DSC) due to inefficient adaptation capabilities because of lack of target domain data. Surprisingly, even without using any target domain data for training, our test time adaptation beats the UDA methods [5,2,7] that use unlabeled target data for training by $\sim 1-3\%$ on average in DSC. For SSDA, SLS [3] and DLDM [23] demonstrate limited refinement capabilities for medical image analysis. Our method in SSDA$^-$ (only 1 or 5 labeled target data for training without using unlabeled target data) surpasses all the existing SSDA techniques that use both 1 or 5 labeled and unlabeled target data, justifying its superiority.

3.2 Performance Evaluation on BraTS2018

Similar observations were noted in tumor segmentation using the BraTS2018 dataset, shown in Table 2. DSAN [6] and DFN [31] rely on image translation for style adaptation, yielding biased outcomes toward the source domain. Although centered on target-to-source transfer, our approach leverages variational

Setting	Method	Data Access S_L	T_L	T_U	MRI→CT LV	LA	MYO	AA	avg.	CT→MRI LV	LA	MYO	AA	avg.
No DA	-	✓	✗	✗	53.5/18.7	7.3/25.8	2.1/29.9	18.5/27.3	20.4/25.4	18.3/33.6	36.8/18.9	7.2/27.8	30.8/20.2	23.3/25.1
TTA	DLTTA [26]	✓	✗	✗	74.5/7.7	88.2/9.2	60.2/4.9	85.5/12.1	77.1/8.5	72.0/11.6	59.4/13.8	57.1/7.8	75.2/5.7	65.9/9.5
	CoTTA [22]	✓	✗	✗	79.5/4.9	80.3/5.4	63.1/7.3	85.4/7.1	77.1/6.2	79.6/3.8	63.1/6.4	48.4/4.0	68.2/5.5	64.8/4.9
	DomAda [27]	✓	✗	✗	82.1/5.8	86.8/5.5	64.2/6.1	87.9/7.1	80.2/6.1	81.1/4.4	69.3/6.3	48.3/4.2	75.0/5.5	66.4/5.1
UDA	SECASA [5]	✓	✗	✓	82.9/5.4	85.2/3.6	71.7/6.2	83.8/4.9	80.9/5.1	81.0/ 3.9	74.6/4.2	55.9/3.9	68.3/9.6	69.9/5.4
	CPCL [2]	✓	✗	✓	85.6/4.5	88.6/3.9	61.5/6.2	89.6/6.1	82.1/5.2	81.2/4.6	70.1/4.1	60.8/3.5	62.5/9.9	70.1/5.5
	T-MAtt [7]	✓	✗	✓	77.6/7.6	88.7/3.8	67.4/5.2	90.8/5.7	81.1/5.6	77.5/5.6	67.4/6.9	59.1/4.2	71.0/5.4	68.7/5.3
TTA	**Ours**	✓	✗	✗	**85.6/4.6**	**89.2/3.6**	**71.9/5.1**	**91.2/4.8**	**84.5/4.6**	**81.4/3.7**	**70.5/4.0**	**60.9/3.5**	**75.2/5.4**	**71.3/4.4**
SSDA:1	SLS [3]	✓	1	✓	84.7/5.2	83.3/4.9	72.3/5.8	85.4/6.3	81.4/5.6	82.0/5.4	71.9/5.3	62.9/4.4	75.8/4.0	73.2/4.8
	DLDM [23]	✓	1	✓	85.0/4.7	83.1/5.6	71.7/5.8	85.9/6.2	81.4/5.6	82.7/5.3	70.4/6.7	63.1/4.5	76.0/3.7	73.3/5.1
	ACT [14]	✓	1	✓	87.9/3.9	88.5/3.6	75.5/4.8	88.3/5.1	85.1/4.4	85.1/4.2	74.8/4.2	65.9/3.7	74.7/3.7	75.1/4.0
	DisCL [1]	✓	1	✓	88.2/4.6	88.6/4.0	76.9/4.9	89.6/5.8	85.8/4.8	85.8/3.9	73.3/5.1	64.9/3.9	75.9/3.8	75.0/4.2
SSDA⁻:1	**Ours**	✓	1	✗	**88.7/3.8**	**89.8/2.8**	**77.4/4.7**	**91.7/4.0**	**86.8/3.8**	**87.7/2.8**	**76.8/3.0**	**66.9/3.3**	**76.3/4.1**	**76.9/3.3**
SSDA:5	SLS [3]	✓	5	✓	86.4/4.2	88.0/4.1	77.2/5.0	89.0/5.2	85.2/4.6	86.7/4.0	77.8/4.2	70.3/3.9	78.4/3.4	78.3/3.9
	DLDM [23]	✓	5	✓	87.5/3.9	88.1/4.1	78.0/4.7	89.3/5.2	85.7/4.5	87.9/3.7	77.7/4.5	72.2/3.6	78.6/3.3	79.1/3.8
	ACT [14]	✓	5	✓	89.0/3.1	89.6/3.3	78.4/4.1	91.6/4.4	87.2/3.7	88.6/3.1	78.3/3.1	73.9/3.3	79.1/3.2	80.0/3.2
	DisCL [1]	✓	5	✓	89.5/3.3	89.2/3.5	80.8/3.6	91.7/4.6	87.8/3.8	89.9/2.1	79.1/2.8	75.8/3.2	79.1/3.6	81.0/2.9
SSDA⁻:5	**Ours**	✓	5	✗	**90.2/2.8**	**90.5/2.0**	**83.3/2.1**	**92.4/3.2**	**89.1/2.5**	**90.1/1.8**	**79.7/2.5**	**77.0/3.0**	**80.5/2.7**	**81.8/2.5**
Supervised Joint Train		✓	✓	✓	91.9/1.3	91.1/2.3	87.8/1.2	92.7/1.5	90.9/1.6	92.4/1.4	80.5/1.7	78.8/2.5	82.8/1.1	83.6/1.7

Table 1: Comparison with SoTA UDA, TTA, and SSDA methods for cardiac structures segmentation on **MMWHS** dataset in **MRI→CT** and **CT→MRI** settings. Data used for training: $\mathbf{S}_L$: labeled source, $\mathbf{T}_L$: labeled target (1 and 5 labels in SSDA settings), $\mathbf{T}_U$: unlabeled target; Metrics used: **DSC↑/ASD↓**. The best and second-best values are highlighted in **red** and **blue**.

Setting	Method	Data Access S_L	T_L	T_U	DSC↑ T1CE	T1	FLAIR	HD↓ T1CE	T1	FLAIR
No DA	-	✓	✗	✗	10.9	6.7	55.1	60.3	50.4	28.1
TTA	DLTTA [26]	✓	✗	✗	42.1	38.1	63.3	17.2	25.4	23.2
	CoTTA [22]	✓	✗	✗	63.5	59.3	82.9	11.2	12.5	7.9
	DomAda [27]	✓	✗	✗	58.2	51.7	68.0	15.0	19.6	16.9
UDA	DFN [31]	✓	✗	✓	62.2	57.3	78.9	15.5	17.5	13.8
	DSAN [6]	✓	✗	✓	62.0	57.7	81.8	13.7	14.2	8.6
	DisCL [1]	✓	✗	✓	64.4	60.7	83.3	10.9	11.1	7.3
TTA	**Ours**	✓	✗	✗	**65.1**	**61.2**	**83.6**	**10.3**	**10.6**	**6.9**
SSDA:1	SLS [3]	✓	1	✓	66.1	64.7	82.3	10.5	12.2	7.1
	DLDM [23]	✓	1	✓	66.5	65.8	81.5	10.3	12.0	7.1
	ACT [14]	✓	1	✓	69.7	69.7	84.5	10.0	10.5	5.8
	ACT-Exp [14]	✓	1	✓	69.0	67.4	83.9	10.3	10.9	6.4
	DisCL [1]	✓	1	✓	71.9	72.2	85.8	9.5	10.0	5.2
SSDA⁻:1	**Ours**	✓	1	✗	**72.8**	**73.3**	**87.1**	**9.0**	**9.6**	**4.8**
SSDA:5	SLS [3]	✓	5	✓	71.2	67.2	83.1	10.1	11.7	6.8
	DLDM [23]	✓	5	✓	68.3	67.8	83.3	9.9	11.2	6.6
	ACT [14]	✓	5	✓	70.8	71.3	85.0	9.8	10.0	5.2
	ACT-Exp [14]	✓	5	✓	69.8	70.3	84.4	10.2	10.4	5.7
	DisCL [1]	✓	5	✓	72.4	73.1	86.1	9.3	9.7	4.8
SSDA⁻:5	**Ours**	✓	5	✗	**73.5**	**74.1**	**87.9**	**8.7**	**9.3**	**4.4**
Supervised Joint Train		✓	✓	✓	**73.8**	**74.9**	**88.3**	**8.4**	**9.0**	**4.1**

Table 2: Comparison with SoTA UDA, TTA, and SSDA methods for whole tumor segmentation on **BraTS2018** dataset in **T2** →{**T1CE**, **T1**, **FLAIR**} settings. Data used for training: $\mathbf{S}_L$: labeled source, $\mathbf{T}_L$: labeled target (1 and 5 labels in SSDA settings), $\mathbf{T}_U$: unlabeled target; Metrics used: **DSC↑/HD↓**. The best and second-best values are highlighted in **red** and **blue**.

Exp.#	Components			MMWHS		BraTS2018 (T2 Source)		
	Edge Guidance	Global $\mathcal{L}_G^{Seg}$	Local $\mathcal{L}_c^{Seg}$	MRI→CT	CT→MRI	T1CE	T1	FLAIR
(a)	✗	✗	✗	83.1 / 5.6	69.8 / 5.5	64.5 / 11.4	60.6 / 10.9	83.1 / 7.2
(b)	✓	✗	✗	84.5 / 4.6	71.3 / 4.4	65.1 / 10.3	61.2 / 10.6	83.6 / 6.9
(c)	✓	✓	✗	88.8 / 2.8	81.6 / 2.7	73.0 / 9.1	73.9 / 9.5	87.3 / 4.6
(d)	✓	✓	✓	**89.1 / 2.5**	**81.8 / 2.5**	**73.5 / 8.7**	**74.1 / 9.3**	**87.9 / 4.4**

Table 3: Ablation experiment for two segmentation tasks in SSDA$^-$ setting (with 5 target labels) to understand the contribution of individual components. Metrics used: **DSC↑/ASD↓** for **MMWHS** and **DSC↑/HD↓** for **BraTS**. The mean of four segmentation tasks (LV, LA, MYO, AA) is reported for **MMWHS**.

inference-based source sampling and generative latent optimization for efficient target-clone generation, demonstrating effectiveness in diverse and challenging domain adaptation scenarios like BraTS. Although quite effective, DCL [1] falls short in fully leveraging domain-agnostic information, resulting in a second-best performance for both UDA and SSDA. TTA approaches [26,27,22] lag behind their UDA counterpart due to lack of discriminative feature learning from the target domain. Interestingly, our TTA without accessing any target data for training beats all existing UDA and TTA methods in Table 2. Our SSDA$^-$ method that only accesses 1 or 5 labeled target data beats all other SSDA methods and even performs close to fully supervised ones (last row in Table 2).

3.3 Ablation Experiments

We conduct a detailed ablation experiment, as demonstrated in Table 3. The effectiveness of edge guidance in constructing the source latent space and finding a target's source-like clone is evident in experiment (b) as compared to (a), where the inclusion of domain-agnostic edge features leads to important spatial and structural guidance for reconstruction, resulting in superior segmentation performance. Please refer to supplementary material for qualitative performance. In exp.(c), global semantic guidance using $\mathcal{L}_G^{Seg}$ brings a significant boost as compared to exp. (a) (up to $\sim 13\%$ in MMWHS and $\sim 16\%$ in BraTS avg. DSC improvement), justifying the significance of semi-supervised fine-tuning. Furthermore, local semantic guidance in exp.(d) complements its global counterpart by bringing significant boost (up to $\sim 17\%$ and $\sim 23\%$ DSC gain in MMWHS and BraTS, respectively). Additional results are in **supplementary material**.

4 Conclusion

We present a novel latent search-based optimization strategy for generating the closest clone of each target image within the source latent space during inference. This approach eliminates the need for continually refining a deployed model with the inception of new domains. Additionally, we extend our method to semi-supervised settings by incorporating label guidance at both global and local scales. Through comprehensive experimental analysis on two domain adaptive segmentation tasks, we demonstrate the superior performance of our proposed approach over state-of-the-art UDA, TTA, and SSDA methods. Though currently tested on modality shifts only, we will extend this paradigm across datasets with different organs and anatomical variations.

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

References

1. Basak, H., Yin, Z.: Semi-supervised domain adaptive medical image segmentation through consistency regularized disentangled contrastive learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 260–270. Springer (2023)
2. Cai, Z., Xin, J., Dong, S., You, C., Shi, P., Zeng, T., Zhang, J., Onofrey, J.A., Zheng, N., Duncan, J.S.: Unsupervised domain adaptation by cross-prototype contrastive learning for medical image segmentation. In: 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 819–824. IEEE (2023)
3. Chen, S., Jia, X., He, J., Shi, Y., Liu, J.: Semi-supervised domain adaptation based on dual-level domain mixing for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11018–11027 (2021)
4. Cover, T., Hart, P.: Nearest neighbor pattern classification. IEEE transactions on information theory **13**(1), 21–27 (1967)
5. Feng, W., Ju, L., Wang, L., Song, K., Zhao, X., Ge, Z.: Unsupervised domain adaptation for medical image segmentation by selective entropy constraints and adaptive semantic alignment. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 623–631 (2023)
6. Han, X., Qi, L., Yu, Q., Zhou, Z., Zheng, Y., Shi, Y., Gao, Y.: Deep symmetric adaptation network for cross-modality medical image segmentation. IEEE transactions on medical imaging **41**(1), 121–132 (2021)
7. Ji, W., Chung, A.C.: Unsupervised domain adaptation for medical image segmentation using transformer with meta attention. IEEE Transactions on Medical Imaging (2023)
8. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
9. Kullback, S., Leibler, R.A.: On information and sufficiency. The annals of mathematical statistics **22**(1), 79–86 (1951)
10. Li, B., Gao, Z., He, X.: Gradient-map-guided adaptive domain generalization for cross modality mri segmentation. In: Machine Learning for Health (ML4H). pp. 292–306. PMLR (2023)
11. Liu, Q., Chen, C., Dou, Q., Heng, P.A.: Single-domain generalization in medical image segmentation via test-time adaptation from shape dictionary. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 1756–1764 (2022)
12. Liu, S., Yin, S., Qu, L., Wang, M.: Reducing domain gap in frequency and spatial domain for cross-modality domain adaptation on medical image segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 1719–1727 (2023)
13. Liu, X., Xing, F., El Fakhri, G., Woo, J.: Self-semantic contour adaptation for cross modality brain tumor segmentation. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2022)
14. Liu, X., Xing, F., Shusharina, N., Lim, R., Jay Kuo, C.C., El Fakhri, G., Woo, J.: Act: Semi-supervised domain-adaptive medical image segmentation with asymmetric co-training. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 66–76. Springer (2022)

15. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging **34**(10), 1993–2024 (2014)
16. Pandey, P., Kyatham, V., Mishra, D., Dastidar, T.R., et al.: Target-independent domain adaptation for wbc classification using generative latent search. IEEE Transactions on Medical Imaging **39**(12), 3979–3991 (2020)
17. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015)
18. Shin, H., Kim, H., Kim, S., Jun, Y., Eo, T., Hwang, D.: Sdc-uda: Volumetric unsupervised domain adaptation framework for slice-direction continuous cross-modality medical image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7412–7421 (2023)
19. Sun, Q., Murphy, K., Ebrahimi, S., D'Amour, A.: Beyond invariance: Test-time label-shift adaptation for distributions with "spurious" correlations (2022)
20. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial discriminative domain adaptation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7167–7176 (2017)
21. Wang, H., Li, X.: Towards generic semi-supervised framework for volumetric medical image segmentation. Advances in Neural Information Processing Systems **36** (2024)
22. 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)
23. Wang, Z., Wei, Y., Feris, R., Xiong, J., Hwu, W.M., Huang, T.S., Shi, H.: Alleviating semantic-level shift: A semi-supervised domain adaptation method for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 936–937 (2022)
24. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing **13**(4), 600–612 (2004)
25. Yang, C., Guo, X., Chen, Z., Yuan, Y.: Source free domain adaptation for medical image segmentation with fourier style mining. Medical Image Analysis **79**, 102457 (2022)
26. Yang, H., Chen, C., Jiang, M., Liu, Q., Cao, J., Heng, P.A., Dou, Q.: Dltta: Dynamic learning rate for test-time adaptation on cross-domain medical images. IEEE Transactions on Medical Imaging **41**(12), 3575–3586 (2022)
27. Zhang, J., Qi, L., Shi, Y., Gao, Y.: Domainadaptor: A novel approach to test-time adaptation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 18971–18981 (2023)
28. Zhang, Y., Huang, K., Chen, C., Chen, Q., Heng, P.A.: Satta: Semantic-aware test-time adaptation for cross-domain medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 148–158. Springer (2023)
29. Zhao, H., Liu, Y., Alahi, A., Lin, T.: On pitfalls of test-time adaptation. In: International Conference on Machine Learning (ICML) (2023)
30. Zhuang, X., Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of mri. Medical image analysis **31**, 77–87 (2016)

31. Zou, D., Zhu, Q., Yan, P.: Unsupervised domain adaptation with dual-scheme fusion network for medical image segmentation. In: IJCAI. pp. 3291–3298 (2022)