Multi-Channel Uncertainty-Weighted Score Matching for Conditional Diffusion in Medical UDA

Chen Li*[1], Meilong Xu[1], Xiaoling Hu[2], Weimin Lyu[1], and Chao Chen[1]

[1] Stony Brook University, Stony Brook, NY, USA
[2] Massachusetts General Hospital and Harvard Medical School, MA, USA

Abstract. Robust medical image segmentation across modalities remains challenging due to severe domain shifts and the lack of target-domain labels. While diffusion models have been explored for cross-domain generation and augmentation, target-domain conditional diffusion training typically relies on highly noisy pseudo masks; naively conditioning on a single Arg-Max pseudo-label can corrupt diffusion training and downstream segmentation. We propose **UPDiff-UDA**, a unified UDA framework whose core is an uncertainty-guided training objective for target-domain conditional diffusion. Given an imperfect source-trained segmenter, we use its per-pixel softmax distribution to form ranked pseudo-label maps (Arg-Max, Arg-2nd, Arg-3rd, ...). Each map yields a conditional score estimate, and we aggregate them via pixel-wise confidence weighting to obtain an uncertainty-reweighted score for score matching, improving robustness to pseudo-label noise while leveraging alternative plausible labels in uncertain regions. We further provide a theoretical justification showing that confidence-weighted aggregation follows a minimum-MSE convex-combination principle under the segmenter-induced surrogate label distribution. To improve pseudo-condition quality, we also introduce a feature-guided, low-degree-of-freedom Bézier curve adaptation to reduce appearance gaps. Experiments on multiple public datasets and modality shifts show that UPDiff-UDA generates high-fidelity labeled target-style samples for augmentation and consistently outperforms strong UDA baselines. The code for this project is available at: `https://github.com/superlc1995/Multi-Channel-Uncertainty-Diffusion-UDA`

Keywords: Diffusion model · Unsupervised domain adaptation · Bézier curve

1 Introduction

Deep learning models can achieve excellent segmentation performance when trained with dense annotations, but their reliability often collapses under domain shift, e.g., when imaging protocols, scanners, or institutions change. In medical

* Email: Chen Li (li.chen.8@stonybrook.edu).

imaging, such shifts are very common and can degrade well-trained models from another domain. Unsupervised domain adaptation (UDA) [5, 17, 20, 30] aims to transfer knowledge from a labeled source domain to an unlabeled target domain, reducing the need for expensive target annotations. GAN-based style translation is a popular UDA approach [31,56], either mapping target images to source style for inference with a source-trained model or translating source images to target style for data augmentation. However, beyond GAN training instability, such translations often degrade or distort highly variable regions, especially small, rare lesions with heterogeneous appearance, because reliable cross-domain correspondences and consistent style mappings are difficult to learn [45].

Recently, diffusion models, particularly conditional diffusion models (CDMs), have been explored for UDA [10, 18, 32], mostly as generative augmentation modules. Masks are provided as conditional input for CDMs to generate corresponding syntheic images. These mask-image pairs are supposed to mimic target domain annotation-image pairs to adapt the model. The issue, however, is we do not have ground truth masks in the target domain in the first place. Some pipelines implicitly rely on stronger assumptions by training with (image, mask) pairs via style translation and reusing source masks as labels, treating source labels as directly transferable to target-style data [37]. Another realistic alternative is to use pseudo-labels, i.e., masks predicted by a source-domain-trained segmenter on target-domain images [10, 15, 39]. However, under domain shift, a source-domain-trained model's predictions on target images are often uncertain and error-prone; the noise will be amplified by CDM, yielding unrealistic synthetic mask-image pairs, and ultimately degrading model adaptation [32]. See Fig. 1 for illustrations.

We focus on this critical yet under-addressed challenge in UDA: **how to robustly train CDMs using noisy target-domain pseudo-labels**. An intuitive approach, borrowed from semi-supervised learning [26, 40], is to leverage the uncertainty of these labels by discarding high-uncertainty samples. However, given the significant domain gap, pseudo-labels are often excessively noisy. This causes vanilla uncertainty-based strategies to fail: it either discards too much data, leading to inefficient learning, or admits too much noise, ultimately derailing the training process.

In this paper, we propose a novel multi-channel uncertainty-guided CDM training for UDA. The key idea is to treat the segmentor output not as a single pseudo-label mask, but as a distribution over plausible labels. For each target domain image, the segmentation model produces a softmax probability vector at every pixel. Instead of only using the prediction (i.e., the Arg-Max mask), we argue that for uncertain prediction, model's confidence on all labels are valuable information and should be used. During CDM training, we generate ranked pseudo-label maps (Arg-Max, Arg-2nd, Arg-3rd, etc.) as separate channels.[3] We use all these different channels to train the CDMs. Each channel has its own conditional score estimation. Aggregating these multi-channel

[3] Here "map" denotes a full spatial label map (an arg-k label at every pixel/voxel), not a single one-hot class channel.

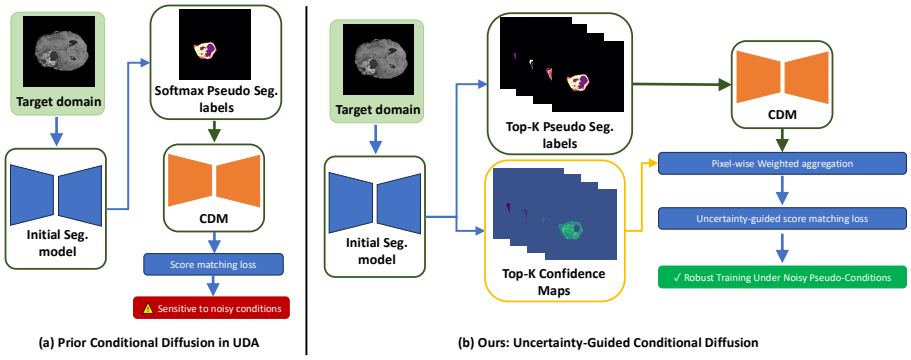

Fig. 1: The comparison between (a) Prior conditional diffusion in UDA and (b) our Uncertainty-guided conditional diffusion.

score estimates weighted by their confidence results in a much more effective uncertainty-reweighted score-matching objective. This design has two benefits: (1) it preserves useful semantic guidance from high-confidence regions, and (2) it reduces the negative impact of incorrect pseudo-labels by downweighting uncertain predictions while still exploiting alternative plausible labels. An illustration is shown in Fig. 1. In practice, one important strength is that our method does not change the diffusion architecture; it only changes the training objective, making the implementation easy and robust.

We further provide theoretical insight into the proposed aggregation: under a squared-error criterion, weighting multiple conditional score estimates by the segmenter's softmax probabilities yields the optimal convex combination with respect to a surrogate label distribution defined by the softmax outputs, highlighting why we should not collapse the softmax to an Arg-Max pseudo-label.

To further strengthen the pseudo-conditions and uncertainty signals used by our diffusion training, we also introduce a supporting appearance-alignment component, Bézier adaptation. Rather than learning an unconstrained, free-form translation, Bézier adaptation applies a learnable Bézier-curve-based transformation with limited degrees of freedom and feature-guided optimization. This improves stability and generalization and, crucially for our main contribution, yields better initial segmentation predictions and more informative uncertainty maps in the target domain, improving uncertainty-guided diffusion training.

Putting these components together, we present UPDiff-UDA, a UDA framework that (i) enables robust target-domain CDM training under noisy pseudo-labels via uncertainty-guided score matching and (ii) produces stronger pseudo-conditions via stable, constrained appearance alignment. The trained CDM then generates labeled target-style image–mask pairs for augmentation, allowing us to train a high-quality target-domain segmentation model. Please see Fig. 2 for an illustration. In summary, our main contributions are:

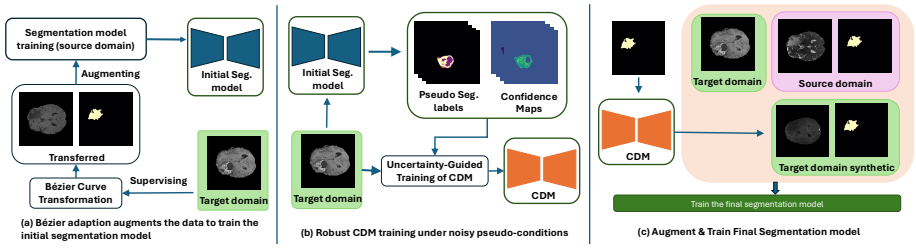

Fig. 2: The overview of our pipeline: Bézier adaptation, uncertainty guided CDM, and Plug-and-play integration with common UDA pipelines.

- **Multi-channel uncertainty-guided CDM training under noisy pseudo-label conditions**. We introduce a new score-matching objective that leverages the segmentation model's ranked softmax predictions and pixel-wise confidence weighting to robustly train conditional diffusion in the target domain without ground-truth target masks.
- **Theoretical justification and robustness analysis.** We provide theoretical justification for the proposed confidence-weighted aggregation by establishing a minimum-MSE convex aggregation principle under an assumed per-pixel label distribution.
- **Bézier adaptation as a supporting pillar**. We propose a constrained, feature-guided Bézier-curve transformation that reduces appearance gaps and improves the quality of pseudo-labels/uncertainty estimates feeding into uncertainty-guided diffusion training.

Using the robustly trained CDM, we synthesize labeled target-style image–mask pairs to strengthen target-domain training and improve segmentation performance. We demonstrate that our framework integrates with common UDA pipelines and consistently improves performance across datasets and modality shifts.

2 Related Work

Unsupervised domain adaptation (UDA) for segmentation has been extensively studied as a cost-effective solution to the challenge of acquiring high-quality segmentation annotations [13,16,24,29,42,53,55], particularly in the context of medical imaging, where manual annotations demand significant expertise and resources [6,9,17,20,30,54]. Self-training [4,14,38,49,50] is widely utilized in UDA due to the substantial performance gains achieved through pseudo-labeling. However, the gain of self-training is limited due to two aspects: First, the quality of pseudo-labels can be degraded due to the domain gap. Second, the available target domain images determine the performance upper bound of UDA methods. In contrast, our framework leverages the generative capacity of CDM, enabling us to generate an unlimited number of labeled target domain images, irrespective of the availability of target domain samples.

GAN-based style transfer [2, 11, 21, 33] mitigates domain gaps by adapting the image style of the source domain to match that of the target domain. This enables the segmentation model trained on style-transferred images to achieve improved performance in the target domain [13,20,56]. However, the effectiveness of these domain style transfer methods is limited. Achieving a thorough and precise style transformation is highly challenging. In tumor regions, style transfer can be particularly problematic and may even have a counterproductive effect due to the sparsity and location variability of tumor regions. Although methods like CyCADA [13] use semantic consistency loss to force the translated image to have the same segmentation map as before, the need to use multiple loss functions impairs the model's ability to depict tumor regions.

Diffusion models [8, 12, 36, 41] are widely applied to vision tasks due to their ability to produce high-quality samples. For the segmentation task, diffusion models are applied to both natural images [1, 7, 19, 37] and medical images [27, 46, 48, 51]. In terms of domain adaptation, [37] proposes a diffusion-based image translation framework guided by pixel-wise semantic labels for semantic segmentation. In Wang et al. [44], the diffusion model is utilized as an encoder to learn domain-invariant representations, facilitating UDA in medical image segmentation. To the best of our knowledge, we are among the first to leverage conditional diffusion models for directly augmenting the target domain with high-quality labeled synthetic images.

3 Method

Our *UPDiff-UDA* framework is built around an *uncertainty-guided conditional diffusion model (CDM)* for target-domain data augmentation (Sec. 3.2), with a lightweight Bézier-curve-based style alignment module (*Bézier adaptation*, Sec. 4.1) as a supporting component. The core challenge we address is that training a CDM in UDA requires conditional labels in the target domain, yet only *noisy pseudo-labels* are available due to domain shift. Naively conditioning a diffusion model on a single Arg-Max pseudo mask can propagate label errors into generation and undermine downstream segmentation. To overcome this bottleneck, we propose a *noise-robust training strategy* for conditional diffusion: instead of treating the segmenter output as one pseudo mask, we exploit the *full softmax distribution* by constructing multiple ranked pseudo-label maps (Arg-Max, Arg-2nd, Arg-3rd, etc.) and using their pixel-wise confidence to reweight conditional score estimates during diffusion training. This uncertainty-guided score matching objective enables the CDM to leverage useful semantic cues while suppressing the impact of incorrect pseudo conditions, resulting in more reliable target-style labeled synthesis.

As shown in Fig. 2(b), we train the CDM on target-domain images and then use it as a label-preserving generator. Under the standard UDA assumption that the semantic label distribution is shared across domains, the CDM takes a segmentation mask (e.g., from the source domain) as condition and generates a corresponding target-style image, producing labeled synthetic pairs for aug-

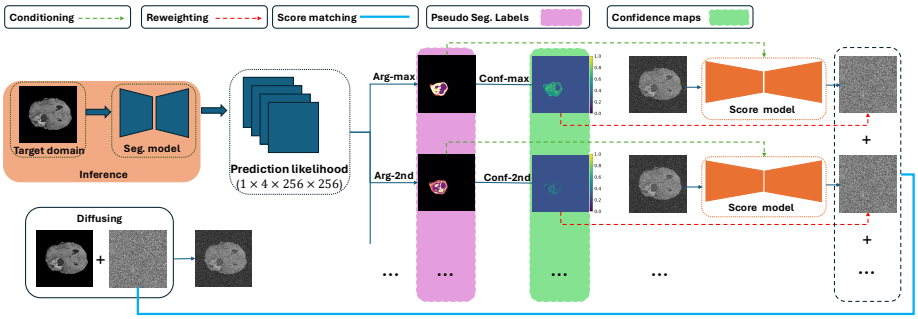

Fig. 3: An overview of the proposed uncertainty-guided CDM training framework.

mentation. These generated image–mask pairs can be combined with unlabeled target images and plugged into existing UDA pipelines to learn a strong target-domain segmentation model. To further improve the quality of pseudo conditions and the associated uncertainty signals used for CDM training, we introduce *Bézier adaptation* as a supporting alignment step: it applies a constrained, learnable Bézier-curve intensity transformation optimized via feature-space similarity, narrowing the appearance gap and yielding more reliable segmentation predictions on target images. The overall framework is illustrated in Fig. 2.

3.1 Preliminaries: score-based diffusion models

Diffusion models synthesize data by transforming Gaussian noise into samples via a reverse-time stochastic process. Let $x_t \in \mathbb{R}^d$ denote the data state at time $t \in [0, T]$. The forward diffusion process is described by the SDE

$$dx_t = f(x_t, t)\, dt + g(t)\, dw_t, \tag{1}$$

where $f(\cdot, \cdot)$ is the drift, $g(\cdot)$ the diffusion coefficient, and w_t is a standard Wiener process. Let $p_t(x)$ be the marginal density of x_t. Under mild regularity conditions, the corresponding reverse-time SDE is

$$dx_t = \Big[f(x_t, t) - g(t)^2 \nabla_{x_t} \log p_t(x_t)\Big] dt + g(t)\, d\bar{w}_t, \tag{2}$$

where $\bar{w}_t$ is a reverse-time Wiener process. Since the score $\nabla_{x_t} \log p_t(x_t)$ is intractable, it is approximated by a neural network s_θ trained via denoising score matching (DSM) [12]:

$$\mathcal{L}_{\mathrm{DSM}}(\theta) = \mathbb{E}_t \left[\lambda(t)\, \mathbb{E}_{(x_0,y)\sim p_0(X,Y)}\, \mathbb{E}_{x_t \sim p_{t|0}(\cdot|x_0)} \left\| s_\theta(x_t, y, t) - \nabla_{x_t} \log p_{t|0}(x_t \mid x_0) \right\|_2^2 \right], \tag{3}$$

where $p_{t|0}(x_t \mid x_0)$ is the Gaussian perturbation kernel implied by (1), and $\lambda(t)$ is a weighting schedule.

3.2 Uncertainty-guided conditional diffusion model on the target domain

We aim to train a target-domain conditional diffusion model (CDM) using only target images X_t and an imperfect segmentation model f_p (trained on the source domain and thus potentially unreliable on X_t). We adapt f_p to target images and use its predictions as conditional inputs for CDM training. Unlike standard pseudo-labeling that uses only the Arg-Max prediction, we jointly utilize multiple semantic predictions from f_p. For a given target image x_0, we generate multiple pseudo segmentation masks (Arg-Max, Arg-2nd, Arg-3rd, etc.), and use their corresponding confidence maps to aggregate multiple conditional score estimates in a pixel-wise manner. The overall training pipeline and an illustrative example of these ranked pseudo maps are shown in Fig. 3.

Pseudo-label distribution and confidence maps. Let $p_u(k) := \mathrm{softmax}(f_p(x_0))_{u,k}$ be the segmentation probability for class k at pixel/voxel u. For each pixel, we view the softmax vector $\{p_u(k)\}_{k=1}^{|\mathcal{K}|}$ as a surrogate distribution over plausible labels in the target domain. When the Arg-Max confidence is low, the Arg-2nd (or other ranked) class can have a substantial chance to be correct; thus incorporating the alternatives can mitigate pseudo-label errors in uncertain regions.

Uncertainty-reweighted score estimation. We feed pseudo-label maps derived from f_p to the conditional score network s_θ to obtain label-conditional score estimates. These estimates are aggregated using pixel-wise confidence weights to form a single uncertainty-reweighted score:

$$\hat{s}_\theta(x_t, t) = \sum_{k=1}^{|\mathcal{K}|} p(\cdot = k) \;\odot\; s_\theta(x_t, y = k, t), \tag{4}$$

where $p(\cdot = k)$ denotes the per-pixel weight map $u \mapsto p_u(k)$ and "$\odot$" denotes the Hadamard (pixel-wise) product.

Uncertainty-guided score matching objective. We train the CDM by replacing the score network in (3) with the uncertainty-reweighted score $\hat{s}_\theta$:

$$\mathcal{L}_{\mathrm{UGSM}}(\theta) = \mathbb{E}_t \left[\lambda(t)\, \mathbb{E}_{x_0 \sim p_t(X_t)}\, \mathbb{E}_{x_t \sim p_{t|0}(\cdot|x_0)} \left\| \hat{s}_\theta(x_t, t) - \nabla_{x_t} \log p_{t|0}(x_t \mid x_0) \right\|_2^2 \right]. \tag{5}$$

Optimizing (5) ensures that high-confidence predictions contribute strong semantic guidance to the score estimation, while lower-confidence predictions are still utilized but downweighted, thereby reducing the negative impact of erroneous pseudo-labels. Importantly, the reweighting is pixel-wise, reflecting the fact that confidence can vary across spatial locations.

4 Theoretical justification: why multi-channel

Our confidence-weighted aggregation is motivated by a minimum-MSE convex-combination principle: given a per-pixel label distribution, the squared-error optimal convex weights for aggregating label-conditional scores match that distribution. We instantiate this distribution using the segmenter softmax outputs, which clarifies why we need not collapse the softmax to an Arg-Max pseudo-label (Arg-Max is optimal only when predictions are nearly deterministic). We formalize this principle in Theorem 1.

Theorem 1 (Minimum-MSE convex aggregation under an assumed label distribution). *Fix a diffusion time t and pixel u. Let $Y(u) \in \{1, \dots, |\mathcal{K}|\}$ be a discrete label variable with an* assumed *per-pixel distribution $q_u(k) := \mathbb{P}(Y(u) = k \mid x_0)$ (e.g., provided by a segmentation model). For each label ℓ, denote the label-conditional score at u by $S_\ell(u) := \nabla_{x_t(u)} \log p_t(x_t \mid Y = \ell)$. Consider deterministic aggregated estimators of the form*

$$a_w(u) = \sum_{\ell=1}^{|\mathcal{K}|} w_u(\ell)\, S_\ell(u), \qquad w_u \in \Delta^{|\mathcal{K}|-1}.$$

Then the conditional mean squared error

$$\mathbb{E}\Big[\big\| a_w(u) - S_{Y(u)}(u) \big\|_2^2 \,\Big|\, x_0, x_t \Big]$$

is minimized by $w_u = q_u$.

Implication. Arg-Max conditioning corresponds to a one-hot q_u and is optimal only when the label distribution is (nearly) deterministic; otherwise, probability-weighted aggregation yields the minimum-MSE convex combination under q_u.

Proof (compact). Condition on (x_0, x_t) and define $m_q(u) := \sum_\ell q_u(\ell)\, S_\ell(u)$. For any fixed (deterministic) $a_w(u)$,

$$\mathbb{E}\big[\|a_w(u) - S_{Y(u)}(u)\|_2^2 \mid x_0, x_t \big] = \|a_w(u) - m_q(u)\|_2^2 + C(u),$$

where

$$C(u) = \sum_\ell q_u(\ell) \|S_\ell(u)\|_2^2 - \|m_q(u)\|_2^2$$

is independent of w_u. Thus the minimizer satisfies $a_w(u) = m_q(u)$, which is achieved by $w_u = q_u$. □

Discussion. Theorem 1 provides a minimum-MSE convex aggregation principle: the optimal weights equal the underlying per-pixel label distribution q_u. In our method, we instantiate q_u using the segmenter softmax output $p_u(k) = \mathrm{softmax}(f_p(x_0))_{u,k}$, treating it as a surrogate estimate of target-domain label uncertainty under domain shift. Since this surrogate may be imperfect, the effectiveness of the aggregation depends on how well p_u reflects the underlying label uncertainty; we analyze the impact of probability mismatch on the estimation error in the supplementary material.

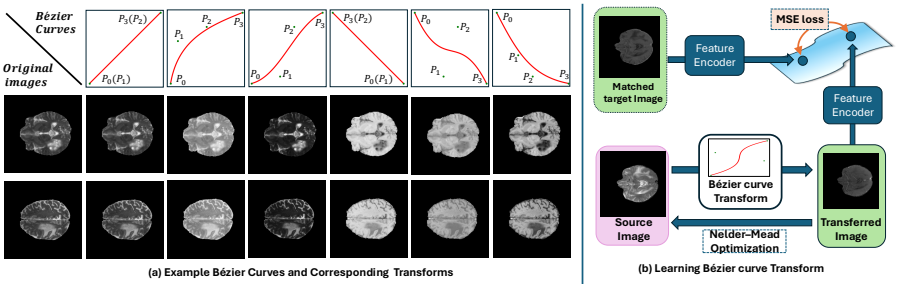

Fig. 4: The overview of Bézier-curve-based style transfer. (a) illustrates the impact of varying control points on Bézier curves and their effect on medical image styles. (b) is the control points optimization process of Bézier adaptation.

Top-K approximation (efficiency). Computing (4) over all $|\mathcal{K}|$ labels can be expensive. In practice, we restrict aggregation to the top-K classes per pixel and renormalize the weights. Let $\pi_u(1), \ldots, \pi_u(K)$ be the indices of the K largest values of $\{p_u(k)\}_{k=1}^{|\mathcal{K}|}$ and define $\tilde{y}^{(k)}(u) := \pi_u(k)$ and

$$\bar{c}^{(k)}(u) := \frac{p_u(\tilde{y}^{(k)}(u))}{\sum_{j=1}^{K} p_u(\tilde{y}^{(j)}(u)) + \epsilon}, \quad \sum_{k=1}^{K} \bar{c}^{(k)}(u) \approx 1. \tag{6}$$

We then approximate (4) as

$$\hat{s}_\theta(x_t, t) \approx \sum_{k=1}^{K} \bar{c}^{(k)} \odot s_\theta(x_t, \tilde{y}^{(k)}, t), \tag{7}$$

where K trades off computation and approximation fidelity (see supplementary material for further discussion).

High-confidence collapse with per-pixel renormalization. We observed that in pixels where the segmenter is extremely confident, higher-rank predictions (Arg-2nd/Arg-3rd/...) are not informative (their probabilities are near zero) and may unnecessarily complicate uncertainty-guided CDM training. To handle this case, we apply a pixel-wise collapse rule during CDM training. Let $c^{(1)}(i,j)$ denote the Arg-Max confidence at pixel (i,j) and let δ be a confidence threshold. For any $k > 1$, if $c^{(1)}(i,j) > \delta$, we overwrite the higher-rank pseudo-labels with the Arg-Max label,

$$\tilde{y}^{(k)}(i,j) \leftarrow \tilde{y}^{(1)}(i,j),$$

so that alternative labels are discarded at these highly confident locations. We then set the corresponding confidence values to a uniform constant before per-pixel renormalization,

$$c^{(k)}(i,j) \leftarrow \frac{1}{|C|}, \quad k > 1,$$

where $|C|$ is the number of classes. After renormalization, the aggregation remains equivalent to conditioning on Arg-Max only, because all channels share the same label at (i, j) (hence they produce the same conditional score estimate), while in pixels with $c^{(1)}(i, j) \leq \delta$ we keep the original ranked pseudo-labels so that plausible alternatives can contribute in uncertain regions.

4.1 Bézier adaptation

Bézier adaptation is a lightweight intensity alignment module used to reduce the appearance gap before training our uncertainty-guided CDM. We model the cross-domain intensity mapping with a cubic Bézier curve, which provides a smooth nonlinear transformation controlled by only a few parameters. Compared with GAN-based translation, this constrained parameterization is more stable and less likely to introduce structural artifacts; compared with histogram matching, it is flexible enough to fit a specific source–target domain pair.

To learn the transformation, we optimize the Bézier control points so that transformed source images become closer to target images. Because pixel-wise comparison is unreliable under large modality shifts, we measure similarity in a deep feature space extracted by a pretrained encoder. Concretely, we first select representative source prototypes via K-means in feature space, retrieve their nearest neighbors from the target set, and then fit Bézier parameters by minimizing the feature-space MSE between each matched pair. Since the optimization landscape can be noisy and the parameter space is low-dimensional, we use a derivative-free optimizer (Nelder–Mead) for stable fitting. The resulting Bézier transformation produces better-aligned images, which in turn improves the quality of pseudo-labels and uncertainty maps used by our uncertainty-guided diffusion training. The complete Bézier adaptation pipeline is in Fig. 4(b). More details about Bézier adaptation are in the supplementary material.

5 Experiment

In the experimental section, we mainly evaluate the effectiveness of our method on unsupervised domain adaptation (UDA) for medical image segmentation.

Datasets. We use three benchmark datasets (i.e., BraTS 2023 dataset [23], Multi-Modality Whole Heart Segmentation dataset [57], and Abdominal Multi-Organ datasets [22, 25]) to show the effectiveness of our method. Please refer to the supplementary material for more details about the datasets.

Dataset preprocessing For the **BraTS** and **Multi-Organ** datasets, each 3D image is normalized using min-max scaling, with voxel intensities rescaled to the range $[0, 1]$. For the MM-WHS dataset, we use the preprocessed images provided by Wang et al. [44].

Baselines. 'No adaptation' denotes the baseline results obtained by evaluating a segmentation model trained solely on source domain data without applying any UDA techniques. CycleGAN [56], CyCADA [13], SIFA [6], and PSIGAN [20] are

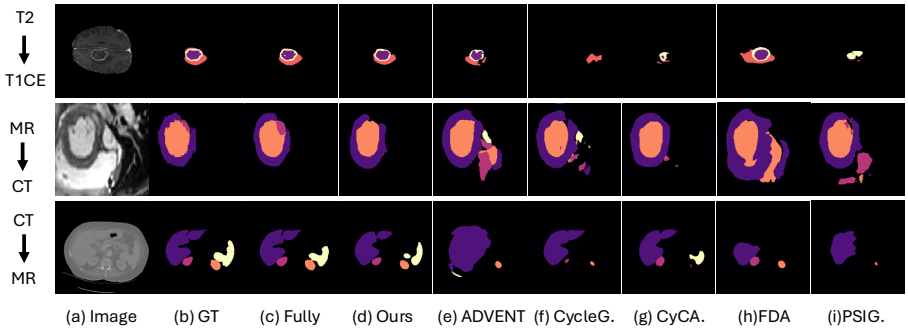

Fig. 5: Segmentation results of Ours-PL and other baselines. From top to bottom: (1) BraTS dataset – Purple, orange, and yellow represent NCR, ED, and ET classes, respectively; (2) MM-WHS dataset – Purple, pink, orange, and yellow represent LVM, LAB, LVB, and AA classes, respectively; (3) Multi-Organ dataset – Purple, pink, orange, and yellow represent Liver, R. Kidney, L. Kidney, and Spleen classes, respectively.

representative GAN-based UDA methods. ADVENT [43] enhances UDA performance by introducing entropy minimization losses, while FDA [52] leverages the Fourier Transform to reduce domain discrepancies. GenericSSL [44] uses diffusion models to extract domain-invariant representations. FPL+ [47] employs cross-domain data augmentation and dual-domain pseudo label generation to effectively mitigate domain shift. Diff-style [37] employs a classifier-guided diffusion model for image style transfer, effectively mitigating domain gaps.

Implementation details. We use a U-Net with ResNet34 as the backbone for the segmentation task. The input images are augmented with random rotation, random scale, and Bézier-curve-based style augmentation. For the selected UDA method, we enhance performance by adding our generated target domain images, paired with their corresponding segmentation conditions, into the labeled training set, while also applying Bézier adaptation as an additional augmentation during training. **Ours-AD**, **Ours-GS**, and **Ours-PL** denote the integration of our proposed generated data with ADVENT [43], GenericSSL [44], and pseudo-labeling (details in the supplementary material), respectively.

5.1 Results

We compare our UPDiff-UDA augmented UDA methods with various baselines on three benchmark datasets. We report Dice (Dice) as the primary metric for the main comparisons in the paper; 95% Hausdorff distance (95HD) and additional results are provided in the supplementary material. In the ablation study, we report Dice and 95HD to comprehensively analyze the impact of each component.

BraTS is a challenging UDA dataset due to the variability of tumors across different modalities. As shown in Tab. 1, our method outperforms other baselines under the UDA setting. When T1 is the target domain, our *UPDiff-UDA* approach significantly augments the ADVENT method, improving the Dice score

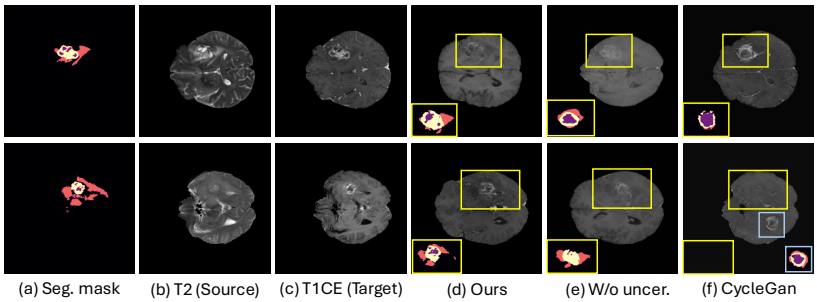

Fig. 6: Qualitative results for generative models. We also show the segmentation results of the tumor region. Our method (d) is much better than baselines: CDM trained without uncertainty (e), and CycleGAN (f). In the second row, CycleGAN even hallucinates the tumor at the wrong location.

Table 1: Comparison of different methods on the BraTS dataset (Dice % ↑).

Target domain	T1				T1CE			
Method	WT	TC	ET	Ave.	WT	TC	ET	Ave.
No Adaptation	8.30	8.81	5.34	7.48	9.96	20.35	14.57	14.96
Fully supervised	73.30	61.63	41.63	58.86	75.70	83.79	80.37	79.95
CycleGAN	7.21	5.95	3.88	5.68	12.55	23.93	12.49	16.32
CyCADA	13.00	18.25	12.08	14.44	22.26	35.71	16.79	24.92
ADVENT	54.24	49.49	32.91	45.55	40.67	48.28	34.81	41.25
FDA	32.52	36.21	22.87	30.53	39.91	53.17	35.77	42.95
SIFA	55.43	50.72	33.81	46.66	42.37	49.74	35.62	42.58
PSIGAN	47.62	40.92	25.57	38.04	21.68	32.83	17.98	24.16
GenericSSL	57.99	53.62	32.93	48.18	46.07	55.18	35.30	45.52
FPL+	58.97	54.83	34.68	49.49	48.72	56.17	36.63	47.17
Diff-style	46.11	46.91	31.25	41.42	37.69	45.12	19.18	34.00
Ours-AD	63.53	**57.71**	**38.69**	**53.31**	57.99	61.77	34.38	51.38
Ours-GS	**65.09**	55.55	38.03	52.89	46.34	67.35	**44.33**	52.67
Ours-PL	60.90	56.95	38.21	52.02	**62.47**	**67.89**	44.12	**58.16**

by 17.0%. When adapting T2 domain images to T1CE, **Ours-PL** outperforms other methods by a large margin. The generative-based baselines perform worse on this dataset because they rely on style transfer, which struggles in tumor regions due to the variability and sparsity of tumors.

MM-WHS serves as a strong benchmark for UDA methods due to the realistic domain gap between CT and MRI scans, particularly in contrast, noise characteristics, and anatomical appearance. Our method achieves impressive performance. As shown in Tab. 2, our *UPDiff-UDA* approach boosts ADVENT [43] substantially, achieving a 19.7% improvement in Dice score.

The results of **Multi-Organ** are shown in Tab. 3. Although ADVENT [43] achieves subpar results on this dataset, our *UPDiff-UDA* pipeline significantly improves its performance, surpassing other baselines. By combining our *UPDiff-UDA* with the pseudo-labeling framework, the **Ours-PL** surpasses other baselines by a large margin. GenericSSL [44] underperforms on this dataset, while the incorporation of our synthetic target domain images significantly improves its performance, highlighting the effectiveness of our pipeline.

Table 2: Comparison of different methods on the MM-WHS dataset (Dice % ↑).

Method	Cardiac MRI → Cardiac CT					Cardiac CT → Cardiac MRI				
	Dice (%) ↑					Dice (%) ↑				
	LVM	LAB	LVB	AA	Ave.	LVM	LAB	LVB	AA	Ave.
No Adaptation	63.72	66.73	82.84	68.19	70.37	25.94	22.11	45.25	10.31	25.90
Fully supervised	88.33	89.69	94.70	83.10	88.95	90.94	94.47	92.98	96.13	93.63
CycleGAN	74.76	83.02	88.66	84.34	82.70	65.91	36.21	79.55	18.84	50.13
CyCADA	69.78	87.63	84.69	65.43	76.88	64.56	47.64	81.55	31.39	56.28
ADVENT	75.08	79.19	89.11	87.90	82.82	62.60	47.63	80.92	33.15	56.08
FDA	59.09	33.88	73.90	17.99	46.22	35.82	14.77	53.69	0.74	26.26
SIFA	65.72	78.36	81.59	71.82	74.37	51.57	67.15	78.23	**66.07**	65.75
PSIGAN	13.90	22.26	27.42	63.24	31.71	69.95	42.81	79.04	39.79	57.90
GenericSSL	**87.95**	91.88	91.02	**91.06**	**90.48**	**74.16**	55.07	86.13	54.55	67.48
FPL+	75.50	81.20	86.40	78.30	80.35	59.84	54.97	79.12	44.37	59.58
Diff-style	51.71	36.92	31.83	64.77	46.31	57.30	25.89	64.69	3.12	37.75
Ours-AD	79.04	85.23	89.45	85.92	84.91	68.31	63.97	**86.77**	56.59	68.91
Ours-GS	86.84	**92.04**	**92.63**	89.07	90.14	63.10	**70.56**	86.24	61.66	**70.39**
Ours-PL	79.18	84.50	86.78	89.23	84.92	66.95	63.37	83.83	50.00	66.04

Table 3: Comparing on the Abdominal Multi-Organ dataset (Dice % ↑).

Method	Abdominal MRI → Abdominal CT					Abdominal CT → Abdominal MRI				
	Liver	R. Kid	L. Kid	Spleen	Avg.	Liver	R. Kid	L. Kid	Spleen	Avg.
No Adaptation	9.66	70.32	61.45	17.05	39.62	22.12	57.94	59.52	56.35	48.98
Fully supervised	89.04	85.49	90.50	87.22	88.06	83.60	83.69	77.01	65.90	77.55
CycleGAN	74.09	67.88	62.03	76.24	70.06	57.15	51.62	55.38	56.59	55.18
CyCADA	63.31	69.60	72.06	74.09	69.77	53.59	60.18	60.74	58.70	58.30
ADVENT	21.18	67.98	72.19	14.95	44.08	20.77	57.94	59.52	56.35	48.64
FDA	71.55	81.27	71.02	80.06	75.97	51.16	64.20	63.90	67.96	61.81
SIFA	41.34	37.92	32.96	53.59	41.45	48.43	46.55	58.63	56.27	52.47
PSIGAN	63.53	76.94	77.09	71.01	72.14	34.20	51.65	53.99	58.79	49.66
GenericSSL	36.79	56.33	36.22	9.72	34.77	62.87	53.03	45.21	49.99	52.78
FPL+	74.20	79.15	71.60	78.45	75.85	55.90	66.85	64.30	68.10	63.79
Diff-style	41.52	73.38	76.85	76.23	67.00	29.15	57.94	58.84	57.44	50.84
Ours-AD	81.03	84.50	75.30	88.67	82.38	54.56	67.84	69.80	75.34	66.89
Ours-GS	75.05	**86.72**	77.03	84.24	80.76	**68.73**	72.61	**75.37**	73.25	**72.49**
Ours-PL	**84.89**	85.74	**82.33**	**90.47**	**85.86**	59.02	**74.83**	70.78	**79.44**	71.02

Overall, the performance gains achieved by our *UPDiff-UDA* demonstrate the effectiveness of the proposed method. Qualitative segmentation results are presented in Fig. 5. The Qualitative synthetic results in Fig. 6 highlight the ability of our uncertainty-guided CDM to produce high-fidelity target domain images with accurate alignment to the conditional masks.

Table 4: Ablation study of components.

Target domain	T1C							
Method	Dice (%) ↑				95HD ↓			
	WT	TC	ET	Average	WT	TC	ET	Average
w/o real	54.43	60.33	33.75	49.51	25.83	39.10	41.38	35.44
Mean-teacher	60.20	62.08	35.36	52.55	36.86	51.89	45.17	44.64
w/o threshold	57.30	58.13	30.54	48.66	32.42	37.81	44.37	38.20
w/o uncertainty	59.05	62.94	36.66	52.88	23.65	39.30	43.73	35.56
Ours-PL	**62.47**	**67.89**	**44.12**	**58.16**	**20.89**	**23.07**	**27.84**	**23.93**

Table 5: Ablation study of cold start.

Target domain	T1C						
Method	Accuracy ($\times 10^2$)		Uncertainty evaluation ($\times 10^3$)				
	Dice↑	95HD↓	AURC↓	E-AURC↓	ECE↓	NLL↓	Brier↓
No adaptation	14.96	61.23	0.86	0.78	7.92	41.27	7.64
Histogram matching	11.98	64.72	0.78	0.65	9.61	62.36	9.78
Fourier transform	14.72	58.23	0.39	0.34	6.23	46.11	6.39
Random Bézier	41.88	34.87	0.06	0.05	2.72	13.65	3.22
Bézier adapt	**48.45**	**30.25**	**0.05**	**0.04**	**2.34**	**11.50**	**2.79**

5.2 Ablation Study

In this section, we conduct extensive ablation studies to justify the effectiveness of different components. More ablation study results about the impact of varying hyperparameters are in the supplementary material.

Components. We conduct experiments on the BraTS 2023 dataset to validate the effectiveness of each component of our method, with a specific focus on domain adaptation from T2 to T1CE and the mean-teacher-like pseudo-labeling framework. The results are shown in Tab. 4. In "w/o uncertainty", the CDM is trained by pseudo-labels obtained by applying the argmax function to the initial segmentation model output. Without our proposed training strategy, a significant performance drop is observed in target domain segmentation. "mean-teacher" denotes the result when the pseudo-labeling framework is applied without augmentation from our high-quality target image generations. It reflects the effectiveness of our strategy in boosting the pseudo-labeling method. "w/o real" shows the result of the segmentation model trained with our proposed generations without real source domain images. We observe that even without the support of real images and their ground truth, the segmentation model trained on our labeled synthetic target images achieves respectable performance in target domain segmentation. This demonstrates the high quality of our generation.

Number of confidence maps. We conduct an ablation study on the number of confidence maps (k) employed for uncertainty-guided CDM training. As illustrated in Tab. 7, the better performance is achieved when $k = 2$. In practice, we also observe that, except for Arg-Max and Arg-2nd, other predictions contain less information (i.e., most of their confidence maps exhibit near-zero values). In the T2→T1CE setting, we evaluate voxel-level Top-K accuracy, defined as whether the ground-truth class is included among the K most probable predicted classes. The Top-1/2/3/4 accuracies are 0.9783, 0.9954, 0.9987, and 1.0000, respectively. These results indicate that the Top-2 predictions already contain sufficient useful information, whereas the Arg-3rd and Arg-4th maps contribute only marginal additional gains. In practice, larger values of K mainly lead to higher VRAM usage and smaller feasible batch sizes. On two NVIDIA H100 80GB GPUs [4], the maximum batch sizes for $K = 1$, $K = 2$, and $K = 3$ are 48, 24, and 16, respectively. As a result, increasing K substantially slows training convergence. We may safely drop these predictions in order to save GPU memory and maximize our batch size, given limited computational resources.

Preliminary adaptation. In this paper, we employ Bézier adaptation to improve the quality of pseudo-labels and their associated confidence maps, ensuring they are sufficiently reliable to guide our strategy for training a conditional diffusion model (CDM) capable of generating high-quality target-domain images. To validate its effectiveness, we compare Bézier adaptation with alternative approaches using metrics from two perspectives: (i) the quality of confidence maps and (ii) the final UDA performance. For confidence map evaluation, we

[4] For all other results, the diffusion model is trained on three NVIDIA RTX 8000 GPUs. The two NVIDIA H100 80GB GPUs are used only in this analysis to estimate the feasible batch sizes for different values of K in the Top-K setting.

Table 6: Ablation study on the impact of different style transfer methods.

Target domain	T1C							
Style transfer	Dice (%) ↑				95HD ↓			
	WT	TC	ET	Average	WT	TC	ET	Average
Source only	26.88	25.71	13.88	22.16	56.21	98.27	105.10	86.53
CycleGAN	41.09	42.73	25.54	36.45	34.42	51.94	59.74	48.70
CyCADA	57.25	58.15	30.89	48.76	29.15	33.90	38.92	33.99
PSIGAN	34.76	41.97	19.26	32.00	31.30	74.24	75.92	60.49
FDA	41.81	45.24	29.12	38.72	39.01	61.39	66.79	55.73
Ours-Bézier	**62.47**	**67.89**	**44.12**	**58.16**	**20.89**	**23.07**	**27.84**	**23.93**

Table 7: Ablation study on the # of confidence maps used in UPDiff-UDA.

Target domain	T1C							
k	Dice (%) ↑				95HD ↓			
	WT	TC	ET	Average	WT	TC	ET	Average
1	59.05	62.94	36.66	52.88	23.65	39.30	43.73	35.56
2	**62.47**	**67.89**	**44.12**	**58.16**	**20.89**	**23.07**	**27.84**	**23.93**

adopt metrics from the uncertainty estimation literature. The area under the risk-coverage curve (AURC) quantifies failure prediction, where a lower value indicates that correctly predicted samples receive higher confidence scores, enabling more effective filtering. Excess-AURC (E-AURC) serves as the normalized variant of AURC [28,34]. Calibration assesses the alignment between model confidence and the true likelihood of correctness. To this end, we use expected calibration error (ECE) [35], the Brier score [3], and negative log-likelihood (NLL) to evaluate calibration quality.

As shown in Tab. 5, compared with other style transfer baselines, our Bézier adaptation consistently achieves better performance on both segmentation accuracy and uncertainty-aware metrics. These gains indicate that Bézier adaptation yields more trustworthy pseudo-labels and better-calibrated confidence maps from the initial segmenter, which directly strengthens the supervision signal for our uncertainty-guided CDM training. We further evaluate the impact of different style transfer strategies on end-to-end UDA performance. As reported in Tab. 6, Bézier adaptation again provides the strongest results among all alternatives, highlighting its effectiveness for mitigating appearance gaps and supporting it as a key component of our overall framework.

6 Conclusion

We propose *UPDiff-UDA*, a unified UDA framework for medical image segmentation that enables robust *target-domain conditional diffusion training* under noisy pseudo-labels. Our core contribution is an uncertainty-guided score matching objective that leverages ranked softmax pseudo-label maps and pixel-wise confidence weighting, with theoretical support from a minimum-MSE convex-aggregation principle under a surrogate label distribution. We further introduce a feature-guided Bézier-curve adaptation module for constrained appearance alignment to improve pseudo-conditions and uncertainty estimates. Together, our method delivers consistent improvements across UDA benchmarks, while plug-and-play with existing pipelines.

Acknowledgments

This work was partially supported by grants NSF CCF-2144901, NIH R01NS143143, and R01CA297843.

References

1. Amit, T., Shaharbany, T., Nachmani, E., Wolf, L.: Segdiff: Image segmentation with diffusion probabilistic models. arXiv preprint arXiv:2112.00390 (2021)
2. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: ICML (2017)
3. Brier, G.W.: Verification of forecasts expressed in terms of probability. Monthly weather review (1950)
4. Brüggemann, D., Sakaridis, C., Brödermann, T., Van Gool, L.: Contrastive model adaptation for cross-condition robustness in semantic segmentation. In: ICCV (2023)
5. Chen, C., Dou, Q., Chen, H., Qin, J., Heng, P.A.: Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation. In: AAAI (2019)
6. Chen, C., Dou, Q., Chen, H., Qin, J., Heng, P.A.: Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation. TMI (2020)
7. Chen, T., Li, L., Saxena, S., Hinton, G., Fleet, D.J.: A generalist framework for panoptic segmentation of images and videos. In: ICCV (2023)
8. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. In: NeurIPS (2021)
9. Dorent, R., Joutard, S., Shapey, J., Bisdas, S., Kitchen, N., Bradford, R., Saeed, S., Modat, M., Ourselin, S., Vercauteren, T.: Scribble-based domain adaptation via co-segmentation. In: MICCAI (2020)
10. Gong, H., Wang, Y., Wang, Y., Xiao, J., Wan, X., Li, H.: Diffuse-uda: Addressing unsupervised domain adaptation in medical image segmentation with appearance and structure aligned diffusion models. arXiv preprint arXiv:2408.05985 (2024)
11. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: NeurIPS (2014)
12. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: NeurIPS (2020)
13. Hoffman, J., Tzeng, E., Park, T., Zhu, J.Y., Isola, P., Saenko, K., Efros, A., Darrell, T.: Cycada: Cycle-consistent adversarial domain adaptation. In: ICML (2018)
14. Hoyer, L., Dai, D., Van Gool, L.: Daformer: Improving network architectures and training strategies for domain-adaptive semantic segmentation. In: CVPR (2022)
15. Hu, D., Li, H., Liu, H., Wang, J., Yao, X., Lu, D., Oguz, I.: Adaptdiff: Cross-modality domain adaptation via weak conditional semantic diffusion for retinal vessel segmentation. In: International Workshop on Simulation and Synthesis in Medical Imaging (2024)
16. Hu, X., Zeng, X., Puonti, O., Iglesias, J.E., Fischl, B., Balbastre, Y.: Learn2synth: Learning optimal data synthesis using hypergradients for brain image segmentation. In: ICCV (2025)
17. Huo, Y., Xu, Z., Moon, H., Bao, S., Assad, A., Moyo, T.K., Savona, M.R., Abramson, R.G., Landman, B.A.: Synseg-net: Synthetic segmentation without target modality ground truth. TMI (2018)
18. Ji, W., Chung, A.C.: Diffusion-based domain adaptation for medical image segmentation using stochastic step alignment. In: MICCAI (2024)
19. Ji, Y., Chen, Z., Xie, E., Hong, L., Liu, X., Liu, Z., Lu, T., Li, Z., Luo, P.: Ddp: Diffusion model for dense visual prediction. In: ICCV (2023)

20. Jiang, J., Hu, Y.C., Tyagi, N., Rimner, A., Lee, N., Deasy, J.O., Berry, S., Veeraraghavan, H.: Psigan: Joint probabilistic segmentation and image distribution matching for unpaired cross-modality adaptation-based mri segmentation. TMI (2020)
21. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: CVPR (2019)
22. Kavur, A.E., Gezer, N.S., Barış, M., Aslan, S., Conze, P.H., Groza, V., Pham, D.D., Chatterjee, S., Ernst, P., Özkan, S., et al.: Chaos challenge-combined (ct-mr) healthy abdominal organ segmentation. MedIA (2021)
23. Kazerooni, A.F., Khalili, N., Liu, X., Haldar, D., Jiang, Z., Anwar, S.M., Albrecht, J., Adewole, M., Anazodo, U., Anderson, H., et al.: The brain tumor segmentation (brats) challenge 2023: focus on pediatrics (cbtn-connect-dipgr-asnr-miccai brats-peds). ArXiv (2024)
24. Kim, M., Byun, H.: Learning texture invariant representation for domain adaptation of semantic segmentation. In: CVPR (2020)
25. Landman, B., Xu, Z., Igelsias, J., Styner, M., Langerak, T., Klein, A.: Miccai multi-atlas labeling beyond the cranial vault–workshop and challenge. In: Proc. MICCAI Multi-Atlas Labeling Beyond Cranial Vault—Workshop Challenge (2015)
26. LI, C., Hu, X., Abousamra, S., Chen, C.: Calibrating uncertainty for semi-supervised crowd counting. In: ICCV (2023)
27. Li, C., Hu, X., Abousamra, S., Xu, M., Chen, C.: Spatial diffusion for cell layout generation. In: MICCAI (2024)
28. Li, C., Hu, X., Chen, C.: Confidence estimation using unlabeled data. In: ICLR (2023)
29. Li, Y., Yuan, L., Vasconcelos, N.: Bidirectional learning for domain adaptation of semantic segmentation. In: CVPR (2019)
30. Liu, F.: Susan: segment unannotated image structure using adversarial network. Magnetic resonance in medicine (2019)
31. Liu, M.Y., Breuel, T., Kautz, J.: Unsupervised image-to-image translation networks. In: NeurIPS (2017)
32. Luo, L., Hu, S., Chen, L.: Noise optimized conditional diffusion for domain adaptation. In: IJCAI (2025)
33. Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z., Paul Smolley, S.: Least squares generative adversarial networks. In: ICCV (2017)
34. Moon, J., Kim, J., Shin, Y., Hwang, S.: Confidence-aware learning for deep neural networks. In: ICML (2020)
35. Naeini, M.P., Cooper, G., Hauskrecht, M.: Obtaining well calibrated probabilities using bayesian binning. In: AAAI (2015)
36. Nichol, A.Q., Dhariwal, P.: Improved denoising diffusion probabilistic models. In: ICML (2021)
37. Peng, D., Hu, P., Ke, Q., Liu, J.: Diffusion-based image translation with label guidance for domain adaptive semantic segmentation. In: ICCV (2023)
38. Shen, F., Pataki, Z., Gurram, A., Liu, Z., Wang, H., Knoll, A.: Loopda: Constructing self-loops to adapt nighttime semantic segmentation. In: WACV (2023)
39. Shen, F., Zhou, L., Kucukaytekin, K., Liu, Z., Wang, H., Knoll, A.: Controluda: Controllable diffusion-assisted unsupervised domain adaptation for cross-weather semantic segmentation. arXiv preprint arXiv:2402.06446 (2024)
40. Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.L.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. In: NeurIPS (2020)

41. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. In: ICLR (2021)
42. Tsai, Y.H., Hung, W.C., Schulter, S., Sohn, K., Yang, M.H., Chandraker, M.: Learning to adapt structured output space for semantic segmentation. In: CVPR (2018)
43. Vu, T.H., Jain, H., Bucher, M., Cord, M., Pérez, P.: Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In: CVPR (2019)
44. Wang, H., Li, X.: Towards generic semi-supervised framework for volumetric medical image segmentation. In: NeurIPS (2023)
45. Wang, R., Heimann, A.F., Tannast, M., Zheng, G.: Cyclesgan: A cycle-consistent and semantics-preserving generative adversarial network for unpaired mr-to-ct image synthesis. Computerized Medical Imaging and Graphics (2024)
46. Wolleb, J., Sandkühler, R., Bieder, F., Valmaggia, P., Cattin, P.C.: Diffusion models for implicit image segmentation ensembles. In: MIDL (2022)
47. Wu, J., Guo, D., Wang, G., Yue, Q., Yu, H., Li, K., Zhang, S.: Fpl+: Filtered pseudo label-based unsupervised cross-modality adaptation for 3d medical image segmentation. TMI (2024)
48. Wu, J., Fu, R., Fang, H., Zhang, Y., Yang, Y., Xiong, H., Liu, H., Xu, Y.: Medsegdiff: Medical image segmentation with diffusion probabilistic model. In: MIDL (2024)
49. Wu, X., Wu, Z., Ju, L., Wang, S.: A one-stage domain adaptation network with image alignment for unsupervised nighttime semantic segmentation. PAMI (2021)
50. Xie, B., Li, S., Li, M., Liu, C.H., Huang, G., Wang, G.: Sepico: Semantic-guided pixel contrast for domain adaptive semantic segmentation. PAMI (2023)
51. Xu, M., Gupta, S., Hu, X., Li, C., Abousamra, S., Samaras, D., Prasanna, P., Chen, C.: Topocellgen: Generating histopathology cell topology with a diffusion model. In: CVPR (2025)
52. Yang, Y., Soatto, S.: Fda: Fourier domain adaptation for semantic segmentation. In: CVPR (2020)
53. Zhang, P., Zhang, B., Zhang, T., Chen, D., Wang, Y., Wen, F.: Prototypical pseudo label denoising and target structure learning for domain adaptive semantic segmentation. In: CVPR (2021)
54. Zhang, Z., Yang, L., Zheng, Y.: Translating and segmenting multimodal medical volumes with cycle-and shape-consistency generative adversarial network. In: CVPR (2018)
55. Zhao, X., Mithun, N.C., Rajvanshi, A., Chiu, H.P., Samarasekera, S.: Unsupervised domain adaptation for semantic segmentation with pseudo label self-refinement. In: WACV (2024)
56. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: ICCV (2017)
57. Zhuang, X., Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of mri. MedIA (2016)