Label-constrained unsupervised domain adaptation for semantic segmentation with diffusion models

Alexandre Stenger, Étienne Baudrier, Benoît Naegel, Nicolas Passat

Label-constrained Unsupervised Domain Adaptation for Semantic Segmentation with Diffusion Models

1st Alexandre Stenger
Lab. ICube
University of Strasbourg
France

2nd Étienne Baudrier
Lab. ICube
University of Strasbourg
France

3rd Benoît Naegel
Lab. ICube
University of Strasbourg
France

4th Nicolas Passat
Lab. CRESTIC
University of Reims
France

***Abstract*—Unsupervised Domain Adaptation (UDA) methods have emerged as a promising solution to generalize a learning to close datasets (domains) without the need to produce new ground truth. Nonetheless in biomedical images, some high domain shifts between source and target images lead to poor adaptation. To address this issue, we propose a method relying on two main ideas. First, we learn the source posterior label distribution with diffusion models. Assuming the target label distribution is similar, this learning helps us to guide the diffusion process to generate relevant segmentation masks on target domain. Alongside this probabilistic constraint, we propose a reconstruction pretext task on both source and target domain to extract common images features. Our approach is compared to the state of the art on three highly shifted mitochondria segmentation datasets. Our method ranks among the best in moderately difficult adaptation cases and succeeds in difficult adaptation cases where all other tested methods fail. Code will be available.**

***Index Terms*—Unsupervised Domain Adaptation, Biomedical Image Segmentation, Diffusion Models**

I. Introduction

In electron microscopy, imaging methods produce images of mitochondria with varying contrast and texture, even when coming from the same cellular culture. Segmentation methods are robust and efficient in supervised settings, but they often fail when dealing with new images. To overcome this burden, commonly named *domain shift*, Unsupervised Domain Adaptation (UDA) has emerged as a promising solution. It consists of learning on a source domain with annotation labels, and then adapting to a target domain without any labels. In essence, this topic can be distilled into two distinct subproblems. Firstly, there is a shift between source images and target images. This implies that a network trained on source data, which has no prior exposure to target images, is unable to effectively process the unfamiliar features of the target domain. The underlying challenge is to make the network more familiar with target images (see Fig. 1). This issue has been addressed through various techniques [1]–[3].

In this article, we propose to rely on a pretext task, which involves a separated reconstruction decoder at the output of the segmentation latent space. This decoder will be trained to reconstruct both source and target images, thereby allowing a common representation in the latent space. Alongside this classical step, there is a second one, which is more challenging since it aims at exploiting the following prior on label distribution. If the shift in the image is undeniable, the binary mask distributions of both source and target domains are very close to each others (see Fig. 1). Indeed, the shape and area of mitochondria are similar across the different imaging methods that do not denature geometric aspects of such organelles. Note that it is also a valid hypothesis in most UDA cases, named the *covariate shift assumption* [4], [5]. This assumption allows us to add a probabilistic constraint on the produced segmentation. Regarding Deep Learning methods, enforcing this probabilistic constraint mainly consists of using Adversarial training. Unfortunately, these methods often fail due to high domain shifts and the instability of adversarial training [6]. On the contrary, we propose to take advantage of recent advances in Diffusion Models for semantic segmentation. Our idea is to use them for the segmentation task, but also to learn the source label distribution thanks to their ability to model data distributions.

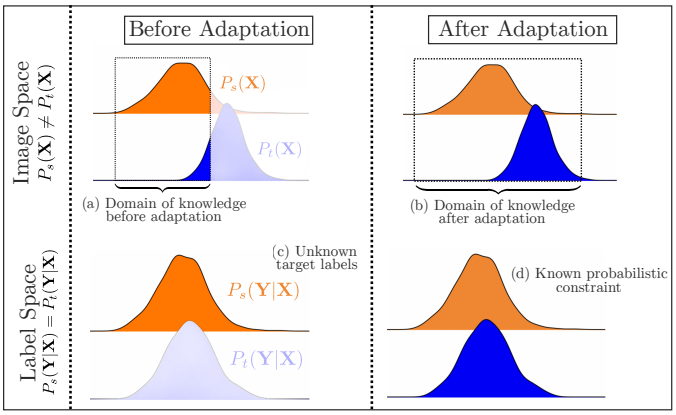

Fig. 1. Our approach for Unsupervised Domain Adaptation is based on the following ideas: (a) There is a shift in image space, which means that the segmentation network trained on the source has a narrow knowledge of the target. (b) This can be mitigated by using a pretext task to extract features from the target images so that the network can work on an extended domain. (c) On the other hand, one does not have access to the target labels. (d) Nevertheless, the covariate shift assumption $P_s(Y|X) = P_t(Y|X)$ allows us to use a probabilistic constraint to foster a meaningful adaptation.

This work of the Interdisciplinary Thematic Institute HealthTech, as part of the ITI 2021-2028 program of the University of Strasbourg, CNRS and Inserm, was supported by IdEx Unistra (ANR-10-IDEX-0002) and SFRI (STRAT'US project, ANR-20-SFRI-0012) under the framework of the French Investments for the Future Program. The authors would like to acknowledge the High Performance Computing Center of the University of Strasbourg for supporting this work by providing scientific support and access to computing resources.

Thus, because of the hypothesis on the label Y distributions $P_s(Y|X) = P_t(Y|X)$ (where X is the image domain, s and t stand for *source* and *target*, see Fig. 1), these models have the potential to produce high-standard masks in the target domain. Our contributions can be summarised as follows: (1) A diffusion model is used as a segmentation tool that is able to learn the source label distribution, and then segment consistent target masks (that actually resemble mitochondria masks). (2) This segmentation diffusion model is trained to be familiar with target features by aggregating a Decoder on it. This decoder reconstructs both target and source images from the latent space, allowing a common feature representation.

II. Related Works

Unsupervised Domain Adaptation (UDA) has many applications. The predominant use case is for autonomous driving, in a synthetic to real scenario on the GTA5-Synthia-Cityscapes datasets [6], [7]. However, it is pointed out in [8] that these methods do not perform well when applied to biomedical images. By contrast to the general scenario described above, UDA for biological images segmentation faces the challenge of higher domain shifts. Thus, we propose to review existing methods especially developed in UDA for biomedical images.

1) Unsupervised Domain Adaptation for biomedical image segmentation: Techniques designed for biomedical images can be divided in three categories. The methods of the first category are based on images: they use the information contained in the target image and try to expand the knowledge of the segmentation network on the whole image space (see Fig. 1). In the case of YNet [9], the reconstruction of both source and target images allows a common feature representation in latent space. An adaptation based on recomputing batch normalization layers is proposed in [10]. The second category is the label-based adaptation. The idea is to enforce target segmentation to be as close as possible to the source one by constraining the segmentation network. In [11], adversarial training is considered: a discriminator has to distinguish which segmentation comes from the source and which from the target domain, encouraging a cross-domain consistency. Finally, other methods make use of both image and label based adaptation. The method CellSegUDA [12] typically crosses Adversarial training with a reconstruction network, while in [2] the same basis as CellSegUDA is used in conjunction with Self-ensembling.

2) Diffusion Models for Semantic Segmentation: Over the last few years, Score-based models got significant interest. From DDPM [13]–[15] to DDIM [16] and lastly Consistency models [17], various sampling techniques have emerged to overcome challenges such as computation time and images generation quality. In this article, we will use a DDIM basis since it allows a greater flexibility than DDPM. A diffusion model is a purely generative process, nonetheless, the community showed that it can also be used in a decisional purpose. Both [18] and [19], alongside other papers [20], [21], propose a diffusion model trained to sample masks constrained by the image to segment in a supervised way, introducing the possibility to use Diffusion models as segmentation networks. When it comes to UDA, most existing methods using diffusion models consist of classical style transfer [22], or generation of image-mask pairs [23] which has already been done using GAN or Cycle-GAN [24]. On the contrary, we believe that it is relevant to build upon the segmentation diffusion framework presented above, because of the inherent ability of Diffusion model to learn a probabilistic distribution, which could help to consider the equality in the label space $P_s(Y|X) = P_t(Y|X)$. In the next section, we present our method built upon the former explanation.

III. Method

A. Context

1) Notations: We consider a domain $\mathcal{D} = \{\mathcal{X}, P\}$, structured by a probability distribution P and a feature space $\mathcal{X}$ such that the image set is a subset of $\mathcal{X}$ sampled from P. In the context of supervised learning, we have a training set $\{(X_i, Y_i)\}_{i=1}^n$ with an image set $\mathbf{X} = \{X_i\}_{i=1}^n \subseteq \mathcal{X}$ and a label set $\mathbf{Y} = \{Y_i\}_{i=1}^n$, knowing that $X_i, Y_i \in \mathbb{R}^{h\times w}$. In domain adaptation, there are a source domain $\mathcal{D}_s = \{\mathcal{X}_s, P_s\}$ and a target domain $\mathcal{D}_t = \{\mathcal{X}_t, P_t\}$ associated to $(\mathbf{X}_s, \mathbf{Y}_s)$ and $(\mathbf{X}_t, \mathbf{Y}_t)$, respectively. When it comes to the UDA setting, one does not have access to the target labels $\mathbf{Y}_t$.

2) Generative Diffusion Models: Let a data distribution be of the form $x_0 \sim q(x_0)$. Diffusion models aim to learn this true data distribution $q(x_0)$ so it is possible to generate new samples following $q(x_0)$. We call $p_\theta(x_0)$ the learnt approximation of $q(x_0)$. Diffusion models learn p_θ by a two-step training stage. First, a sample $x_0 \sim q(x_0)$ is gradually noised with a noising scheme $\epsilon(t)$, from $t = 0$ to $t = T$. Then $x_T \sim q(x_T)$ can be approximated as a nearly isotropic Gaussian noise. The second step consists of denoising $\hat{x}_T$. Formally, we sample $\hat{x}_T \sim p_\theta(x_T) \sim \mathcal{N}(0, \mathbf{I})$ where $\mathbf{I}$ denotes the identity matrix, and gradually denoise it from $t = T$ to $t = 0$. This is done in practice by introducing a denoising U-Net ϵ_θ which predicts the noise to remove from $\hat{x}_t$ at time t. This is trained with the following objective:

$$\mathcal{L}_{\theta_{\text{generative}}} = E_{x_0\sim q(x_0),\epsilon\sim\mathcal{N}(0,\mathbf{I})}[\|\epsilon(x_t, t) - \epsilon_\theta(\hat{x}_t, t)\|^2] \quad (1)$$

with $\epsilon(x_t, t)$ being the noise added at time t during the noising process and $\epsilon_\theta(\hat{x}_t, t)$ the prediction of the noise to remove. In the following, we present how this training procedure was modified to propose segmentation diffusion-models [18].

3) Segmentation Diffusion Models: The goal is to predict a segmentation mask $\hat{Y}_i$ related to the image to segment X_i. In that way, early papers propose to noise a training label Y_i (from the label $Y_{i,0}$ to the nearly Gaussian noise $Y_{i,T}$) and then to denoise it conditioned by the image to segment X_i. It denoises from $\hat{Y}_{i,T}$ to $\hat{Y}_{i,0}$, such that $\hat{Y}_{i,0}$ is the final proposed segmentation. The loss used for training is then the following:

$$\mathcal{L}_{\theta_{\text{seg\_supervised}}} = E_{\epsilon\sim\mathcal{N}(0,\mathbf{I})}[\|\epsilon(Y_{i,t}, t) - \epsilon_\theta(\hat{Y}_{i,t}, X_i, t)\|^2] \quad (2)$$

In the next section, we see how to take advantage of this existing supervised-segmentation framework with diffusion models to perform Unsupervised Domain Adaptation.

B. Proposed method

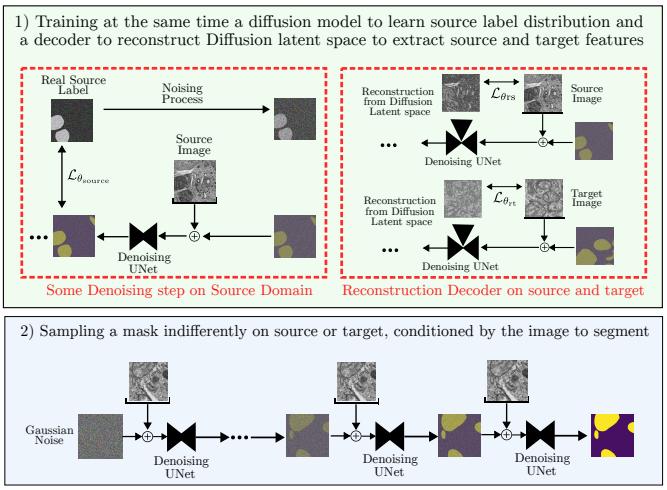

Fig. 2. Overview of our method. We base our segmentation network on the diffusion segmentation. By contrast to its supervised use [18], we use it to enforce relevant segmentation on target domain thanks to its generative learning of source label distribution. Along this probabilistic constraint, we propose a feature extraction module on both domains through a decoder plugged in the latent space of our segmentation diffusion-network.

Let $X_i^{(\mathbf{s})} \in \mathbf{X}_s$ be images to segment on source domain alongside their corresponding binary masks $Y_i^{(\mathbf{s})} \in \mathbf{Y}_s$. Let $X_i^{(\mathbf{t})} \in \mathbf{X}_t$ be images to segment on target domain, without any available labels on target. The only information on target labels is that $P_s(\mathbf{Y}|\mathbf{X}) = P_t(\mathbf{Y}|\mathbf{X})$. By contrast to the supervised setting, diffusion models are here employed to learn the probability density function $P_s(\mathbf{Y}|\mathbf{X})$. Thus, at the inference time, the segmentation diffusion-model will follow $P_s(\mathbf{Y}|\mathbf{X}) = P_t(\mathbf{Y}|\mathbf{X})$ when segmenting a new image. This is the first component of our adaptation strategy: we encourage the segmentation framework to produce consistent labels across domain by learning this probabilistic constraint. The following objective is then a classical segmentation loss, but we also believe that this allows to learn the above constraint:

$$\mathcal{L}_{\theta_{\text{seg}}} = E_{\epsilon\sim\mathcal{N}(0,\mathbf{I})}[\|\epsilon(Y_{i,t}^{(\mathbf{s})}) - \epsilon_\theta(\hat{Y}_{i,t}^{(\mathbf{s})}, X_i^{(\mathbf{s})}, t)\|^2] \quad (3)$$

Along this segmentation loss that permits to work on the label-space, it is necessary for the segmentation network to be able to extract features from the target image domain (see Fig. 1). That is why we propose here to introduce a Decoder that will reconstruct the input of our denoising U-Net ϵ_θ from its latent space. We note this decoder D, and the encoder of the denoising U-Net ϵ_θ^E.

We propose to reconstruct from both domains diffusion latent spaces by introducing the same loss on source ($\mathcal{L}_{\theta_{\text{rs}}}$) and target ($\mathcal{L}_{\theta_{\text{rt}}}$):

$$\mathcal{L}_{\theta_{\text{r s/t}}} = E[\|X_i^{(\mathbf{s/t})} - D(\epsilon_\theta^E(\hat{Y}_{i,t}^{(\mathbf{s/t})}, X_i^{(\mathbf{s/t})}, t))\|^2] \quad (4)$$

This should allow the encoder of our diffusion model to create a common latent space between source and target image domains, and thus the segmentation decoder will be able to take advantage of this common representation. Finally, the global loss of our framework is a combination of the segmentation loss on source and both reconstruction losses weighted by λ:

$$\mathcal{L}_{\theta_{\text{global}}} = \mathcal{L}_{\theta_{\text{seg}}} + \lambda(\mathcal{L}_{\theta_{\text{rs}}} + \mathcal{L}_{\theta_{\text{rt}}}) \quad (5)$$

A graphical explanation of our method is proposed in Fig. 2. Finally, as this remains a generative process, segmentation mask could exhibit undesired artefacts. This is why we propose a classical post-processing strategy [18] by gathering 10 prediction and thresholding over this mean prediction, with a value of 0.5.

TABLE I
RESULTS FOR THE DIFFERENT METHODS. "NO ADA" MEANS THAT NO ADAPTATION IS PROCEEDED (SOURCE TRAINED)

Settings / Methods	W → FS1	FS1 → W	FS2 → 1	FS2 → W	FS1 → 2	W → FS2
No Ada. UNet	0.01	0.02	0.01	0.10	0.57	0.70
No Ada. SegFormer	0.01	0.10	0.01	0.20	0.45	0.50
No Ada. Att.UNet	0.02	0.06	0.15	0.11	0.18	0.40
BN [10]	0.14	0.08	0.14	0.30	**0.74**	0.67
Adv. [11]	0.17	0.08	0.19	0.46	0.43	0.62
CellSeg. [12]	0.13	0.05	0.13	0.18	0.36	0.17
YNet [9]	0.27	0.12	0.15	0.32	0.72	0.65
SelfEns. [2]	0.12	0.11	0.13	0.14	0.28	0.39
Ours	**0.29** ±.007	**0.17** ±.007	**0.25** ±.005	**0.49** ±.005	0.68 ±.011	**0.70** ±.007
Target Supervised	0.84	0.81	0.84	0.81	0.92	0.92

IV. EXPERIMENTS

A. Experimental settings and Results

We utilize three distinct publicly available datasets for mitochondria segmentation, denoted as FS1 [25], FS2 [26] and WeiH [27]. These datasets enable us to explore six distinct adaptation scenarios. In each scenario, one dataset serves as the source domain, while another is considered as a target domain. We use the notation *Source dataset* → *Target dataset* to indicate an adaptation scenario (for instance W → FS1 denotes WeiH as source and FS1 as target). The core of diffusion model relies on the Denoising U-Net. Its input image size was chosen 128^2 with a batch size of 10. The number of channels at the first layer is 128 with 4 layers in the down size and 4 layers in the upper side. This U-Net contains attention layers. The learning rate is set to 10^{-4} with an Adam optimiser. The weight λ in the loss was experimentally found to be optimal at 10^{-3}. For training, we proceed to 50 000 iterations. Finally, we use a DDIM sampling scheme trained on 1 000 noising steps, and for the upcoming experiment, the number of sampling steps is 100.

We compared our method with state-of-the-art methods in UDA for biological segmentation, that we abbreviated the following way: YNet [9], BN [10], Adv. [11], CellSeg [12], SelfEns [2]. To evaluate performance without adaptation, we also provide results from three state-of-the-art supervised segmentation methods trained on the source domain but not adapted to the target domain: U-Net [28], Attention-U-Net

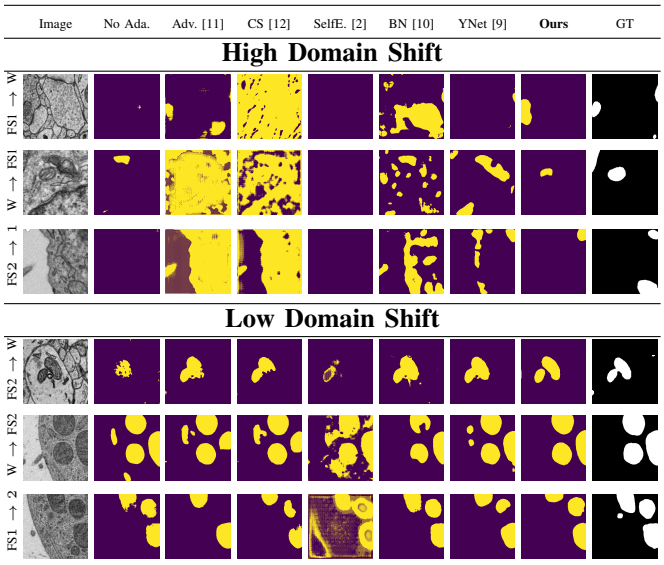

Fig. 3. Visual comparison of UDA methods. The column "Image" stands for the target domain image to segment. "No Ada(pt)" means the target domain U-Net mask prediction, U-Net being trained on the source domain (without adaptation). Then, column 3 to 8 depict different UDA methods, ours included. "GT" stands for Ground Truth segmentation mask. We divide the observed results in two categories: high and low domain shift (detailed in Section IV-A).

[29], and SegFormer [30]. In Tab. I, quantitative results verify that, for all metrics and cases taken together, our method is better or equivalent in 5 cases over 6, showing its robustness to a great variety of domain shifts. We precise that we computed the standard deviation only for our method because of its non-deterministic aspect as inference time. We run 10 inferences and set different seeds in the code to verify its robustness. The very low standard deviation in each case indicates the that results do not depend on random aspects. More generally, we depict two behaviours of the methods regarding the type of shift they are facing: a first case when without adaptation the segmentation network generalises well (Low Domain Shift), and another one where the network without adaptation is unable to segment anything (High Domain Shift).

High domain shift Over the three scenarios with the higher domain shifts (FS1 $\rightarrow$ W, W $\rightarrow$ FS1, FS2 $\rightarrow$ 1), our method performs better than any other. The main visual observation is that our method, in each of those cases, is the only one to provide a segmentation which is really resembling a mitochondria mask, which validates our hypothesis regarding the ability of Diffusion Models to apply the desired probabilistic constraint $P_s(\mathbf{Y}|\mathbf{X}) = P_t(\mathbf{Y}|\mathbf{X})$. Especially, in those scenarios, methods based on Adversarial training (Adv., CellSeg. and SelfEns.) fail due to the Discriminator that is impossible to be fooled such that it rapidly collapses during training (this is comprehensible with the visual aspects of the results without adaptation, which are mainly empty due to the too high domain shift). For BN and YNet, they are not powerful enough to guide the U-Net with the additional target image information they add, which is pictured in visual results by an adaptation that segment objects not related to mitochondria, with shape

TABLE II
ABLATION STUDY (SEE SECTION IV-B).

Configuration	Modules		W → FS1	FS1 → W	FS2 → 1	FS2 → W	FS1 → 2	W → FS2
	Diff.	Recons.	IoU	IoU	IoU	IoU	IoU	IoU
Diffusion	✓	×	0.17	0.16	0.11	0.16	0.45	0.36
YNet	×	✓	0.27	0.12	0.15	0.32	**0.72**	0.65
Ours	✓	✓	**0.29**	**0.17**	**0.25**	**0.49**	0.68	**0.70**

very far from a mitochondria one. Finally, quantitative results in Tab. I confirm those visual considerations.

Low domain shift: On less shifted scenarios, each method has a different behaviour. For instance, Adversarial Methods (Adv., CellSeg. and SelfEns.) that failed previously are now able to stabilize the Discriminator so it can play its adaptability role. Our method is still working very well, but as the Analytical findings in Tab. I point out, BN and YNet can perform better in those cases (especially FS1 $\rightarrow$ 2). This discussion led us to the fact that our method is way better on high domain shift thanks to its probabilistic constraint that forces to generate meaningful segmentation masks. For low domain shift, we still rank among the top, even if it could be more suitable to use other methods.

B. Ablation study

1) The role of the diffusion process: Instead of having a diffusion model as a segmentation backbone, we propose a classical U-Net. This setting corresponds to the YNet method [9] discussed in previous sections. Results on Tab. II confirm already former observations: the diffusion backbone allows better segmentation in highly shifted cases. Although its non-deterministic behaviour is detrimental for less shifted case, it still achieves convenient results. Those observations lead to the conclusion that the diffusion process effectively aids in bridging the domain gap by leveraging its ability to generate domain-agnostic representations.

2) The role of the reconstruction module: The other question is to know if the Diffusion process alone would not be powerful enough to overcome the domain shift. In Tab. II, we observe clearly its inability to significantly improve performance when applied in isolation. This suggests that while the diffusion process is a crucial component, it requires the synergy of a pretext task such as the reconstruction we proposed in order to proceed a quality adaptation.

V. CONCLUSION

We propose a method that utilises diffusion models to learn source posterior label distributions, alongside a more classical reconstruction task to extract both source and target features. Ablation studies show the usefulness of both tasks to provide accurate segmentations on the target images. Even though quantitative results are modest in high-shift cases, its segmentation results are visually consistent with ground truth where other methods fail. Our results demonstrate that generative processes can successfully be applied to unsupervised segmentation tasks to overcome the shift-related challenges.

REFERENCES

[1] D. Franco-Barranco, J. Pastor-Tronch, A. González-Marfil, A. Muñoz-Barrutia, and I. Arganda-Carreras, "Deep learning based domain adaptation for mitochondria segmentation on EM volumes," *Computer Methods and Programs in Biomedicine*, vol. 222, p. 106949, 2022.

[2] C. Li, Y. Zhou, T. Shi, Y. Wu, M. Yang, and Z. Li, "Unsupervised domain adaptation for the histopathological cell segmentation through self-ensembling," in *COMPAY@MICCAI, Procs.*, 2021, pp. 151–158.

[3] M. Klingner, J.-A. Termöhlen, J. Ritterbach, and T. Fingscheidt, "Unsupervised batchnorm adaptation (UBNA): A domain adaptation method for semantic segmentation without using source domain representations," in *WACV, Procs.*, 2022, pp. 210–220.

[4] K. P. Murphy, *Probabilistic Machine Learning: Advanced Topics*. MIT Press, 2023.

[5] J. G. Moreno-Torres, T. Raeder, R. Alaiz-Rodríguez, N. V. Chawla, and F. Herrera, "A unifying view on dataset shift in classification," *Pattern recognition*, vol. 45, pp. 521–530, 2012.

[6] L. Hoyer, D. Dai, and L. Van Gool, "HRDA: Context-aware high-resolution domain-adaptive semantic segmentation," in *ECCV, Procs.*, 2022, pp. 372–391.

[7] ——, "DAFormer: Improving network architectures and training strategies for domain-adaptive semantic segmentation," in *CVPR, Procs.*, 2022, pp. 9924–9935.

[8] H. Shin, H. Kim, S. Kim, Y. Jun, T. Eo, and D. Hwang, "SDC-UDA: Volumetric unsupervised domain adaptation framework for slice-direction continuous cross-modality medical image segmentation," in *CVPR, Procs.*, 2023, pp. 7412–7421.

[9] J. Roels, J. Hennies, Y. Saeys, W. Philips, and A. Kreshuk, "Domain adaptive segmentation in volume electron microscopy imaging," in *ISBI, Procs.*, 2019, pp. 1519–1522.

[10] A. Stenger, L. Vedrenne, P. Schultz, S. Faisan, É. Baudrier, and B. Naegel, "Fast and interpretable unsupervised domain adaptation for FIB-SEM cell segmentation," in *ISBI, Procs.*, 2023.

[11] M. Javanmardi and T. Tasdizen, "Domain adaptation for biomedical image segmentation using adversarial training," in *ISBI, Procs.*, 2018, pp. 554–558.

[12] M. M. Haq and J. Huang, "Adversarial domain adaptation for cell segmentation," in *MIDL, Procs.*, 2020, pp. 277–287.

[13] A. Q. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," in *ICLR, Procs.*, 2021, pp. 8162–8171.

[14] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *NeurIPS, Procs.*, 2020.

[15] P. Dhariwal and A. Nichol, "Diffusion models beat GANs on image synthesis," *NeurIPS, Procs.*, pp. 8780–8794, 2021.

[16] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," *arXiv:2010.02502*, 2020.

[17] Y. Song, P. Dhariwal, M. Chen, and I. Sutskever, "Consistency models," *arXiv:2303.01469*, 2023.

[18] J. Wolleb, R. Sandkühler, F. Bieder, P. Valmaggia, and P. C. Cattin, "Diffusion models for implicit image segmentation ensembles," in *MIDL, Procs.*, 2022, pp. 1336–1348.

[19] T. Amit, T. Shaharbany, E. Nachmani, and L. Wolf, "SegDiff: Image segmentation with diffusion probabilistic models," *arXiv:2112.00390*, 2021.

[20] A. Rahman, J. M. J. Valanarasu, I. Hacihaliloglu, and V. M. Patel, "Ambiguous medical image segmentation using diffusion models," in *CVPR, Procs.*, 2023, pp. 11 536–11 546.

[21] L. Zbinden, L. Doorenbos, T. Pissas, A. T. Huber, R. Sznitman, and P. Márquez-Neila, "Stochastic segmentation with conditional categorical diffusion models," in *ICCV, Procs.*, 2023, pp. 1119–1129.

[22] D. Peng, P. Hu, Q. Ke, and J. Liu, "Diffusion-based image translation with label guidance for domain adaptive semantic segmentation," in *ICCV, Procs.*, 2023, pp. 808–820.

[23] Y. Benigmim, S. Roy, S. Essid, V. Kalogeiton, and S. Lathuilière, "One-shot unsupervised domain adaptation with personalized diffusion models," in *CVPR, Procs.*, 2023, pp. 698–708.

[24] R. Gong, W. Li, Y. Chen, and L. Van Gool, "DLOW: Domain flow for adaptation and generalization," in *CVPR, Procs.*, 2019, pp. 2477–2486.

[25] Zenodo, "FS1 dataset." [Online]. Available: https://zenodo.org/records/8341172

[26] ——, "FS2 dataset." [Online]. Available: https://zenodo.org/records/8344292

[27] EMPIAR, "WeiH dataset." [Online]. Available: https://www.ebi.ac.uk/empiar/EMPIAR-11037

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *MICCAI, Procs.*, 2015, pp. 234–241.

[29] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, "Attention U-Net: Learning where to look for the pancreas," *arXiv:1804.03999*, 2018.

[30] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and efficient design for semantic segmentation with transformers," *Advances in neural information processing systems*, vol. 34, pp. 12 077–12 090, 2021.