Unsupervised Domain Adaptation for Medical Image Segmentation with Dynamic Prototype-based Contrastive Learning

Qing En QINGEN@CUNET.CARLETON.CA
School of Computer Science, Carleton University, Ottawa, Canada

Yuhong Guo YUHONG.GUO@CARLETON.CA
School of Computer Science, Carleton University, Ottawa, Canada
Canada CIFAR AI Chair, Amii, Canada

Abstract

Medical image segmentation typically requires numerous dense annotations in the target domain to train models, which is time-consuming and labour-intensive. To alleviate this challenge, unsupervised domain adaptation (UDA) has emerged to enhance model generalization in the target domain by harnessing labeled data from the source domain along with unlabeled data from the target domain. In this paper, we introduce a novel Dynamic Prototype Contrastive Learning (DPCL) framework for UDA on medical image segmentation, which dynamically updates cross-domain global prototypes and excavates implicit discrepancy information in a contrastive manner. DPCL learns cross-domain global feature representations while enhancing the discriminative capability of the segmentation model. Specifically, we design a novel cross-domain prototype evolution module that generates evolved cross-domain prototypes by employing dynamic updating and evolutionary strategies. This module facilitates a gradual transition from the source to the target domain while acquiring cross-domain global guidance knowledge. Moreover, we devise a cross-domain embedding contrastive module that establishes contrastive relationships within the embedding space. This module captures both homogeneous and heterogeneous information within the same category and among different categories, thereby enhancing the discriminative capability of the segmentation model. Experimental results demonstrate that the proposed DPCL is effective and outperforms the state-of-the-art methods.

Data and Code Availability This paper uses the MultiModality Whole Heart Segmentation (MMWHS) challenge 2017 dataset (Zhuang and Shen, 2016)[1], which is pre-processed and available on the PnP-AdaNet repository (Dou et al., 2019)[2]. The source code is available on github[3].

Institutional Review Board (IRB) This work does not require IRB approval.

1. Introduction

Medical image analysis is crucial for medical diagnosis and paramedical treatment (Duncan and Ayache, 2000; Shen et al., 2017). One of the most critical and challenging medical image analysis tasks is medical image segmentation, which involves classifying each pixel within the input medical image (Ronneberger et al., 2015; Chen et al., 2021). While fully-supervised segmentation methods can yield satisfactory results, real-world medical applications frequently entail diverse data modalities (e.g., CT or MRI) from different hospitals, scanners, and protocols (Guan and Liu, 2021), leading to a significant domain shift phenomenon. One straightforward approach entails annotating data from the target domain and then fine-tuning the segmentation model (Ghafoorian et al., 2017; Swati et al., 2019). However, this process is both time-consuming and labour-intensive, presenting a significant obstacle for medical image analysis.

Recently, unsupervised domain adaptation (UDA) methods have been proposed to learn target domain segmentation models using labeled source domain data and unlabeled target domain data (Chen et al.,

1. https://zmiclab.github.io/zxh/0/mmwhs/
2. https://github.com/carrenD/Medical-Cross-Modality-Domain-Adaptation
3. https://github.com/EnQing626/DPCL

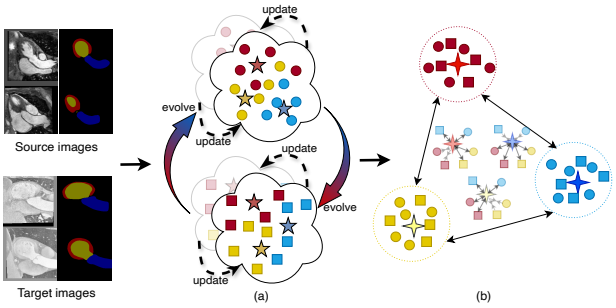

Figure 1: Illustration of the proposed idea. The proposed DPCL (a) dynamically updates cross-domain global prototypes and (b) excavates the implicit discrepancy information of data from both domains in a contrastive manner.

2019; Liu et al., 2022; Shin et al., 2023). In particular, several methods utilize image-to-image translation techniques to bridge the cross-domain gap. They generate source-to-target data or vice versa and employ adversarial training for distribution adaptation (Chen et al., 2020a; Zou et al., 2020; Zhao et al., 2022). However, distortions in data translation may adversely affect the model, potentially resulting in overconfident segmentation masks in the target domain. Some other methods focus on identifying reliable pixels within the target domain segmentation masks to impose constraints and attain semantic alignment (Sun et al., 2022; Feng et al., 2023). However, their effectiveness in segmenting target domain data is limited as they overlook discriminative information among different categories.

In general, UDA for medical image segmentation poses the following fundamental challenges: (1) Significant distribution variations arising from diverse data origins, such as different scanners and patient demographics, lead to a domain shift issue. This challenge is exacerbated by the difficulty in obtaining reliable cross-domain global guidance information. (2) The absence of ground-truth labels in the target domain, combined with the challenges of capturing discriminative information due to subtle differences in organ appearance, further compounds the inherent difficulty of the task. Humans possess a dynamic cognitive capacity to evolve and distinguish complex functions and behaviours as they acquire new knowledge (Luppi et al., 2022; Rensink, 2000). Inspired by this, we aim to address the challenges by integrating dynamically updating strategies within contrastive learning paradigms, which is desirable from both biological and practical aspects.

In this paper, we propose a novel Dynamic Prototype Contrastive Learning (DPCL) framework for UDA on medical image segmentation, as shown in Figure 1. The framework dynamically updates cross-domain global prototypes and exploits implicit discrepancy information in a contrastive manner. DPCL learns cross-domain global feature representations while enhancing the discriminative capability of the segmentation model. Specifically, we devise a cross-domain prototype evolution module (CDPE) to generate evolved cross-domain prototypes through dynamic updating and evolutionary strategies driven by the source domain ground-truth labels and the target domain predictions. This module enables the progressive evolution of prototypes from the source to the target domain, acquiring cross-domain global guidance knowledge. Moreover, we design a cross-domain embedding contrastive module (CDEC) by establishing contrastive relationships between evolved prototypes and object embeddings to capture both homogeneous and heterogeneous information within the same category and among different categories. This module aggregates pixel-level features for object classes, enhancing the discriminative capability of the segmentation models. Finally, we incorporate the two modules to train the DPCL framework end-to-end. The main contributions of our paper are summarized as follows:

- We propose a novel DPCL framework for UDA on medical image segmentation by incorporating dynamic updating strategies and contrastive learning paradigms to excavate cross-domain implicit information. It enhances the discriminative capability of the segmentation model while learning cross-domain global feature representations.

- We devise a cross-domain prototype evolution module to acquire cross-domain global guidance knowledge with dynamic updating and evolutionary strategies.

- We design a cross-domain embedding contrastive module to enhance discriminitive ability of segmentation models by establishing contrastive relationships between evolved prototypes and object embeddings.

- Experimental results demonstrate that the proposed DPCL framework achieves state-of-the-art performance on the MMWHS cross-domain medical image segmentation dataset.

2. Related Work

2.1. UDA for Medical Image Segmentation

Unsupervised domain adaptation aims to enhance segmentation models' performance on the target domain by leveraging labeled source domain data and unlabeled target domain data (Chen et al., 2019; Liu et al., 2022). Several methods utilize GANs (Goodfellow et al., 2014; Zhu et al., 2017), allowing the model to acquire target-specific information (Chen et al., 2020a; Zou et al., 2020; Zhao et al., 2022). Some methods enable end-to-end models for simultaneous cross-modal synthesis and semantic segmentation (Zhang et al., 2018; Huo et al., 2018). Additionally, various methods focus on image-level or feature-level alignment to acquire domain-invariant and domain-specific features (Chen et al., 2020a; Sun et al., 2022; Xie et al., 2022; Shin et al., 2021; You et al., 2020). Moreover, output-level adaptation is employed to ensure that predictions in both domains share similar structural and contextual information (Tsai et al., 2018; Vu et al., 2019; Luo et al., 2019). However, the aforementioned methods are susceptible to distortions during data transformation or lack the guidance of discriminative information based on cross-domain global knowledge. In contrast, the proposed DPCL framework employs dynamically updated and evolved cross-domain prototypes to explore discriminative information along with various classes of foreground features through a contrastive learning approach.

2.2. Contrastive and Prototypical Learning

Contrastive learning is a widely adopted technique in self-supervised learning for enhancing feature representations (Chen et al., 2020b; He et al., 2020; Chen et al., 2020c; Grill et al., 2020). The technology is versatile and applicable in various contexts, including supervised classification (Khosla et al., 2020), semi-supervised segmentation(Hu et al., 2021; Zhong et al., 2021), and domain adaptation for semantic segmentation (Liu et al., 2021a; Li et al., 2021). Prototypical learning (Snell et al., 2017) focuses on extracting representative information for each category and has found application in various scenarios (Li et al., 2020). It is widely utilized in few-shot segmentation (Dong and Xing, 2018; Liu et al., 2020a,c; Wang et al., 2019; Zhang et al., 2019b,a) and semi-supervised segmentation (Xu et al., 2022; Mai et al., 2023). These methods incorporate prototype alignment (Wang et al., 2019), employ graph attention (Zhang et al., 2019a), and utilize iterative refinement (Zhang et al., 2019b; Liu et al., 2020b) to harness support knowledge effectively.

UDA for medical image segmentation faces substantial challenges due to the lack of target domain ground-truth labels and the cross-domain distributional gaps. Our proposed DPCL, unlike existing methods, uses evolved cross-domain prototypes for contrastive learning, capturing invariant features within the same class across images and domains.

3. Method

In this section, we present the proposed DPCL framework for UDA on medical image segmentation. We begin by introducing the overall framework in Section 3.1. We then present the cross-domain prototype evolution module in Section 3.2 and the cross-domain embedding contrastive module in Section 3.3. Finally, we present the loss function in Section 3.4.

3.1. Overall Framework

In UDA for medical image segmentation, we aim to train an effective segmentation model for the target domain. We use a labeled source domain dataset $\mathcal{D}_S$ with $|\mathcal{D}_S|$ image-segmentation mask pairs denoted as $\{(I_s^n, Y_s^n)\}_{n=1}^{|\mathcal{D}_S|}$ and an unlabeled target domain dataset $\mathcal{D}_T$ with $|\mathcal{D}_T|$ images denoted as $\{(I_t^n)\}_{n=1}^{|\mathcal{D}_T|}$. Here $I \in \mathbb{R}^{1*H*W}$ denotes an input image, and $Y \in \{0,1\}^{K*H*W}$ represents a ground-truth segmentation mask, where K is the number of categories, and H and W are the height and width of the image.

The architecture of the proposed DPCL is shown in Figure 2. We first employ the segmentation network N_{seg} as a generator, accompanied by the discriminator N_{dis} for adversarial learning. This encourages consistent spatial semantic information in predictions across domains. Next, we introduce a cross-domain prototype evolution module (CDPE), designed to generate evolved cross-domain prototypes through dynamic updating and evolutionary strategies. This module enables the progressive transformation of prototypes from the source domain to the

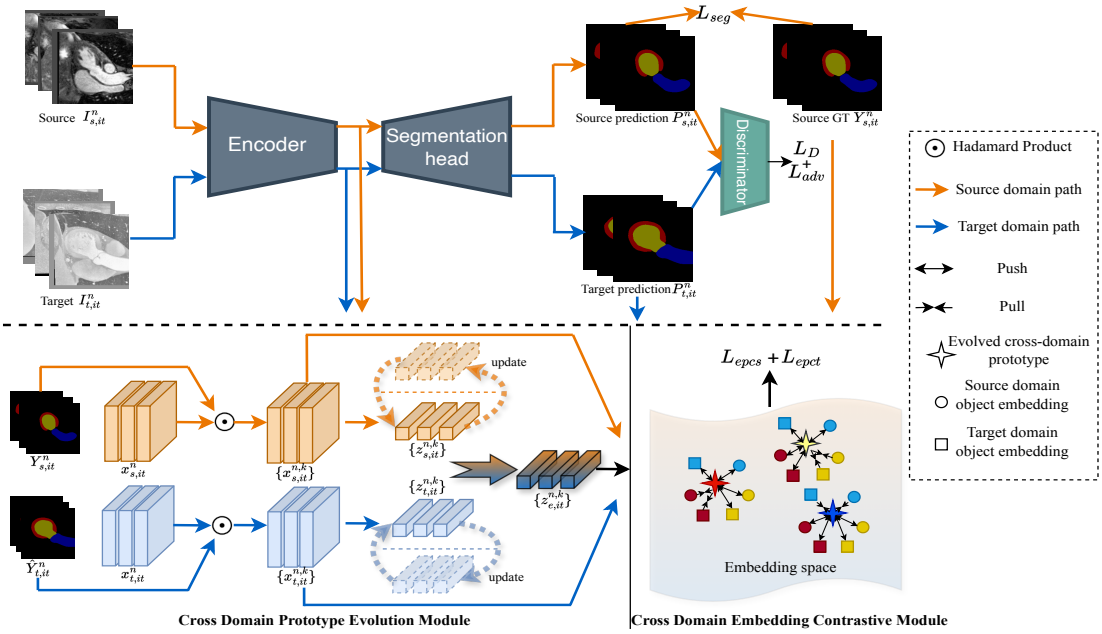

Figure 2: An overview of the proposed DPCL. We first predict segmentation masks for both domain images, which are used for calculating the discriminator loss L_D, the adversarial loss L_{adv} and the segmentation loss L_{seg}. Next, the CDPE module leverages the ground-truth of the source image $Y^n_{s,it}$, the prediction of the target image $\hat{Y}^n_{t,it}$, and the feature embeddings $x^n_{s,it}$ and $x^n_{t,it}$ to generate evolved cross-domain prototypes. The CDEC module is then deployed to calculate the evolved prototype contrastive loss for source domain L_{epcs} and target domain L_{epct}.

target domain, acquiring cross-domain global guidance knowledge. Furthermore, a cross-domain embedding contrastive module (CDEC) is devised to establish contrastive relationships between evolved cross-domain prototypes and object embeddings in both domains. This is achieved by utilizing our proposed cross-domain embedding contrastive loss function, which can aggregate pixel-level features for corresponding object classes and yield discriminative feature representations. Finally, the segmentation network is trained using a combination of adversarial learning loss, supervised segmentation loss, and the evolved prototype embedding contrastive loss. The segmentation network N_{seg} consists of an encoder $f : \mathbb{R}^{1\times H\times W} \rightarrow \mathbb{R}^{c\times h\times w}$ and a segmentation head $g : \mathbb{R}^{c\times h\times w} \rightarrow \mathbb{R}^{K\times H\times W}$, where c, h, and w denote the channels, height, and width of feature embeddings, respectively.

Typically, given the highly similar semantic structural information shared between the source domain and target domain data (Tsai et al., 2018), adversarial learning (Vu et al., 2019) is used to align the entropy of the target domain predictions with that of the source domain. Specifically, in the it-th iteration, the adversarial learning is conducted on a batch of N labeled source images $\{(I^n_{s,it}, Y^n_{s,it})\}^N_{n=1}$ and N target images $\{(I^n_{t,it})\}^N_{n=1}$. We feed a source image $I^n_{s,it}$ and a target image $I^n_{t,it}$ into the segmentation network N_{seg} to generate predictions $P^n_{s,it} \in [0,1]^{K*H*W}$ and $P^n_{t,it} \in [0,1]^{K*H*W}$ as follows:

$$\begin{aligned} P^n_{s,it} &= \text{softmax}(N_{seg}(I^n_{s,it})), \\ P^n_{t,it} &= \text{softmax}(N_{seg}(I^n_{t,it})), \end{aligned} \tag{1}$$

where the softmax function is used to compute pixel-wise prediction probabilities over K categories. Next, we generate the entropy-information maps, denoted as $E^{n,(u,v)}_{s,it} = -P^{n,(u,v)}_{s,it} \cdot \log P^{n,(u,v)}_{s,it}$ and $E^{n,(u,v)}_{t,it} = -P^{n,(u,v)}_{t,it} \cdot \log P^{n,(u,v)}_{t,it}$, with (u, v) denoting the spatial location index. These maps serve as input for the

discriminator N_{dis}, distinguishing between the source and target domains. The generator, represented by our segmentation network N_{seg}, aims to fool the discriminator. Hence, the loss function for training the discriminator N_{dis} is defined as follows:

$$L_D = \sum_{n=1}^{N}(L_{bce}(N_{dis}(E^n_{s,it}), 1) + L_{bce}(N_{dis}(E^n_{t,it}), 0)), \tag{2}$$

where L_{bce} indicates the binary cross-entropy loss function; the adversarial loss for the segmentation network N_{seg} is defined as:

$$L_{adv} = \sum_{n=1}^{N} L_{bce}(N_{dis}(E^n_{t,it}), 1), \tag{3}$$

3.2. Cross-Domain Prototype Evolution Module

While adversarial learning aligns the output of the target domain with spatial structures and context information similar to those of the source domain, it may lack semantic consistency and feature-level discriminative information, limiting segmentation performance. For UDA on medical image segmentation, cross-domain global guidance knowledge is crucial in learning discriminative information. But acquiring this knowledge is challenging due to the domain shift issue and the absence of labels in the target domain. To tackle this problem, we introduce a cross-domain prototype evolution module (CDPE), as shown in Figure 2, to generate evolved cross-domain prototypes through dynamic updating and evolutionary strategies, capturing cross-domain global guidance knowledge.

In the it-th iteration, we start by integrating feature embeddings for each k-th category using global average pooling (Zhang et al., 2020) over the foreground, which serves as the prototype for the specific category organ. This process involves the ground-truth labels from the source domain $Y^n_{s,it}$ and the prediction from the target domain $\hat{Y}^n_{t,it} = \text{argmax}(P^n_{t,it})$ in the it-th iteration. Consequently, the prototypes for category k in the n-th source image and target image in the it-th iteration are determined as follows:

$$z^{n,k}_{s,it} = \frac{\sum_{(u,v)} x^{n,(u,v)}_{s,it} \mathbb{1}[y^{n,(k,u,v)}_{s,it} \neq 0]}{\sum_{(u,v)} \mathbb{1}[y^{n,(k,u,v)}_{s,it} \neq 0]}, \tag{4}$$

$$z^{n,k}_{t,it} = \frac{\sum_{(u,v)} x^{n,(u,v)}_{t,it} \mathbb{1}[\hat{y}^{n,(k,u,v)}_{t,it} \neq 0]}{\sum_{(u,v)} \mathbb{1}[\hat{y}^{n,(k,u,v)}_{t,it} \neq 0]}, \tag{5}$$

where (k, u, v) represents the category and the spatial location index; $\mathbb{1}[\cdot]$ denotes an indicator function; $x^n_{s,it} = f(I^n_{s,it})$ and $x^n_{t,it} = f(I^n_{t,it})$ are the feature embeddings produced by the encoder f in the it-th iteration; $y^n_{s,it} \in \mathbb{R}^{K\times h\times w}$ and $\hat{y}^n_{t,it} \in \mathbb{R}^{K\times h\times w}$ are downsampled $Y^n_{s,it}$ and $\hat{Y}^n_{t,it}$, yielding the same dimension as the feature embeddings. As the prototypes in the current iteration lack global historical category information and are influenced only by the current image embeddings, we employ a dynamic updating strategy (Xie et al., 2018). This strategy alleviates prototype fluctuations by incorporating information from both the current and previous iterations as follows:

$$z^{n,k}_{s,it} \longleftarrow \alpha * z^{n,k}_{s,it-1} + (1-\alpha) * z^{n,k}_{s,it}, \tag{6}$$

$$z^{n,k}_{t,it} \longleftarrow \alpha * z^{n,k}_{t,it-1} + (1-\alpha) * z^{n,k}_{t,it}. \tag{7}$$

Here, we obtain the updated source domain prototype $z^{n,k}_{s,it}$ and the updated target domain prototype $z^{n,k}_{t,it}$ for category k in the n-th source and target images during the it-th iteration by aggregating prototypes from the previous iteration. We use Equations (4) and (5) to generate initial prototypes when it=0.

Subsequently, we aim to combine prototypes from both domains to obtain cross-domain global guidance knowledge. While the target domain lacks ground-truth labels, the segmentation network can gradually generate better predictive segmentation masks as the training progresses. Thus, we gradually increase the weight of the updated target domain prototypes when combining with the ones from the source domain, resulting in a cross-domain evolved prototype $z^{n,k}_{e,it}$ for category k at the it-th iteration:

$$z^{n,k}_{e,it} = \beta * z^{n,k}_{s,it} + (1-\beta) * z^{n,k}_{t,it}. \tag{8}$$

Here $\beta = (\frac{1-it}{maxit})^{0.9}$ is a hyper-parameter that is dynamically updated during training using the poly strategy (Liu et al., 2015; Chen et al., 2017), determining the dynamic combination weights of the source and target prototypes, where $maxit$ represents the maximum number of iterations for training.

The evolved cross-domain prototypes, obtained through a dynamic updating and evolved strategy, incorporate global category information for each domain and reasonable cross-domain knowledge. This can effectively mitigate the domain shift problem and reduce the fluctuating noise within each domain prototype. It is seamlessly integrated with the CDEC module for training segmentation models.

3.3. Cross-Domain Embedding Contrastive Module

It is challenging to capture discriminative information due to subtle differences in organ appearance and the lack of annotations in the target domain. In response, we introduce the CDEC module by leveraging the cross-domain evolved prototypes obtained in the CDPE module. This module constructs an evolved prototype embedding contrastive loss to enhance discriminative feature representations. The structure of the proposed CDEC module is illustrated in Figure 2. It aligns feature embeddings in each class with the corresponding cross-domain evolved prototypes in a contrastive way, thereby improving the discriminative capability of segmentation models.

The contrastive loss is constructed as follows. In the it-th iteration, for each target-domain image $I^n_{t,it}$ in the sampled batch, we first apply up-sampling on $z^{n,k}_{e,it}$ to match the size of the feature embeddings $x^n_{t,it}$, resulting in $Z^{n,k}_{e,it} \in \mathbb{R}^{c\times h\times w}$. Then we generate the masked feature embedding through entrywise product $x^{n,k}_{t,it} = x^n_{t,it} \circ \mathbb{1}[\hat{y}^{n,k}_{t,it} \neq 0]$, representing features specific to the k-th category on the n-th image. With such calculations across all the K categories and on all the target-domain images in the batch, the evolved prototype embedding contrastive loss in the target domain can be defined as follows, aiming to reduce intra-class variations and maximize inter-class variations:

$$L_{epct} = \sum_{n=1}^{N}\sum_{k=1}^{K} -\log\frac{pos}{pos+neg}, \tag{9}$$

where

$$pos = \exp(\cos(x^{n,k}_{t,it}, Z^{m,k}_{e,it})/\tau), \tag{10}$$

$$neg = \sum_{i}\sum_{j\neq k}\exp(\cos(x^{n,k}_{t,it}, Z^{i,j}_{e,it})/\tau), \tag{11}$$

where $\cos(\cdot,\cdot)$ denotes the cosine similarity function and τ is the temperature hyper-parameter. In this loss function, with $x^{n,k}_{t,it}$ representing the appearance of the k-th category in the n-th image, we establish the contrastive relation by randomly selecting an evolved cross-domain prototype, $Z^{m,k}_{e,it} \in \{Z^{n,k}_{e,it}\}^N_{n=1}$, for the same k-th category as a positive sample for $x^{n,k}_{t,it}$, while using all the evolved cross-domain prototypes for all the other categories $\{j : j \neq k\}$ as negative samples for $x^{n,k}_{t,it}$.

Similarly, for each source-domain image $I^n_{s,it}$ in the sampled batch of the it-th iteration, we utilize the down-sampled ground-truth labels $y^n_{s,it}$ to create the masked feature embedding $x^{n,k}_{s,it} = x^n_{s,it} \circ \mathbb{1}[y^{n,k}_{s,it} \neq 0]$ for each k-th category, and compute the evolved prototype embedding contrastive loss in the source domain as follows:

$$L_{epcs} = \sum_{n=1}^{N}\sum_{k=1}^{K} -\log\frac{pos}{pos+neg}, \tag{12}$$

where

$$pos = \exp(\cos(x^{n,k}_{s,it}, Z^{m,k}_{e,it})/\tau), \tag{13}$$

$$neg = \sum_{i}\sum_{j\neq k}\exp(\cos x^{n,k}_{s,it}, Z^{i,j}_{e,it})/\tau). \tag{14}$$

The designed cross-domain embedding contrastive loss function offers several advantages. Firstly, by leveraging the contrastive relationship established between the masked feature embeddings and the evolved cross-domain prototypes, it empowers the segmentation network to learn the connections between individual pixels and prototypes across images. This enables the network to capture richer information about the underlying structure of cross-domain data, and consequently facilitates the achievement of more precise segmentation outcomes. Secondly, through a contrastive exploration of the cross-domain global semantic relationship between pixels and prototypes, it maximizes the similarity of positive pairs

Algorithm 1: Training process of DPCL

Input: $\mathcal{D}_S = \{(I^n_s, Y^n_s)\}^{|\mathcal{D}_S|}_{n=1}$, $\mathcal{D}_T = \{(I^n_t)\}^{|\mathcal{D}_T|}_{n=1}$
Output: Trained segmentation network N_{seg}
for $it = 0$ **to** $maxit$ **do**
- Randomly sample a batch, $\{(I^n_{s,it}, Y^n_{s,it})\}^N_{n=1}$ and $\{(I^n_{t,it})\}^N_{n=1}$, from $\mathcal{D}_S$ and $\mathcal{D}_T$;
- Compute $\{P^n_{s,it}, P^n_{t,it}\}^N_{n=1}$ by Equation (1);
- Compute $\{\forall k : z^{n,k}_{s,it}, z^{n,k}_{t,it}\}^N_{n=1}$ by Equations (4), (5), (6) and (7));
- Compute $\{\forall k : z^{n,k}_{e,it}\}^N_{n=1}$, by Equation (8);
- Upsampling: $Z^{n,k}_{e,it} = Upsample(z^{n,k}_{e,it}), \forall n, k$;
- Create $\{x^{n,k}_{t,it}, x^{n,k}_{s,it}\}, \forall n, k$;
- Compute L_{epct}, L_{epcs} L_{adv} and L_{seg} by Equations (9), (12), (3) and (16);
- Update parameters of N_{seg} by minimizing L_{total} in Equation (15);
- Compute L_D by Equation (2);
- Update N_{dis} by minimizing L_D;

end

while minimizing the similarity of negative pairs. As a result, it can effectively constrain the embedding space, making similar categories more compact and distinct categories more separate.

3.4. Loss Function

The proposed DPCL can be trained end-to-end. In the it-th iteration, given the sampled batch of labeled source images $\{(I^n_{s,it}, Y^n_{s,it})\}^N_{n=1}$ and unlabeled target images $\{(I^n_{t,it})\}^N_{n=1}$, the overall loss function for training the segmentation network is as follows:

$$L_{total} = L_{seg} + \lambda_{adv} L_{adv} + \lambda_{epc}(L_{epcs} + L_{epct}), \quad (15)$$

where λ_{adv} and λ_{epc} are trade-off hyperparameters. It includes the adversarial loss L_{adv} from Equation (3), and the evolved prototype contrastive losses for the source and target domains, L_{epcs} and L_{epct}, from Equations (12) and (9), respectively, while the segmentation loss L_{seg} is defined as:

$$L_{seg} = \sum_{n=1}^{N} (L_{ce}(P^n_{s,it}, Y^n_{s,it}) + L_{dice}(P^n_{s,it}, Y^n_{s,it})) \quad (16)$$

Note the segmentation loss is only computed on the labeled source images, where L_{ce} denotes the cross-entropy loss function and L_{dice} denotes the Dice loss function. The training process of the proposed framework is described in Algorithm 1.

4. Experimental Results

4.1. Experimental Settings

Datasets and evaluation metrics The proposed DPCL is evaluated using the MultiModality Whole Heart Segmentation (MMWHS) challenge 2017 dataset (Zhuang and Shen, 2016). The dataset consists of unpaired volumes, including 20 CT and 20 MR scans, each accompanied by corresponding ground-truth label masks. We employ domain adaptation in two directions: MR → CT and CT → MR. Four organ categories are segmented: the ascending aorta (AA), the left atrium blood cavity (LAC), the left ventricle blood cavity (LVC) and the myocardium of the left ventricle (MYO). We utilize the pre-processed data provided by Chen et al. (2020a), where 80% of the data is used for training, and the remaining 20% is used for testing. We employ standard evaluation metrics, the Dice coefficient (Dice) and the average symmetric surface distance (ASD), to evaluate the proposed framework.

Implementation details DeepLabV2 (Chen et al., 2017) is employed as our segmentation network N_{seg}, initialized with ImageNet pre-trained parameters (Deng et al., 2009). The discriminator N_{dis} follows a PatchGAN setup (Isola et al., 2017), including four convolutional layers and the last classifier. The input image size is 256x256, with standard data augmentations (random scale, rotation, and intensity transformations) applied. We employ stochastic gradient descent (SGD) for segmentation network optimization with a weight decay of 5e-4, a momentum of 0.9, and a learning rate of 2.5e-4. We use the Adam optimizer for the discriminator with a learning rate of 1e-4. The batch size is set to 4, τ to 0.05, α to 0.2, λ_{adv} to 0.003, and λ_{epc} to 0.1 (CT → MR) and 0.01 (MR → CT). Following Tsai et al. (2018) and Feng et al. (2023), we implement multi-level output adaptation by utilizing features from $conv4$ and $conv5$ to generate segmentation masks. The features from $conv4$ are employed for computing L_{epct} and L_{epcs}. Our model is implemented in PyTorch and trained on an NVIDIA 3060Ti GPU.

4.2. Quantitative Evaluation Results

We conduct a comparative analysis over DPCL against ten state-of-the-art methods for UDA on medical image segmentation, including CycleGAN (Zhu et al., 2017), PnP-AdaNet (Dou et al., 2019), AdaOutput (Tsai et al., 2018), AdvEnt (Vu et al., 2019), CyCADA (Hoffman et al., 2018), SIFA V1 (Chen et al., 2019), SIFA V2 (Chen et al., 2020a), EBM (Liu et al., 2021b), CRST (Zou et al., 2019) and SECALA (Feng et al., 2023). We directly report results from SIFA V1 (Chen et al., 2019) and SIFA V2 (Chen et al., 2020a). Other compared methods use the same segmentation network and dataset to obtain results (Feng et al., 2023). We use the fully-supervised model trained on the annotated target domain dataset as the performance upper bound, while the model trained on the annotated source domain dataset serves as the performance lower bound.

Comparison results on MR → CT The test results of the proposed DPCL on the MR → CT task compared with the other state-of-the-art methods are shown in Table 1. The upper and lower bounds achieved Dice values of 90.4% and 23.3%, respectively. The results demonstrate that the proposed DPCL outperforms all the other comparison methods in terms of both Dice and ASD metrics.

Table 1: Quantitative comparison results on the MMWHS dataset for MR → CT adaptation. We report the class average results and the results for all individual classes in terms of Dice and ASD.

Method	Dice					ASD				
	AA	LAC	LVC	MYO	Dice.Avg↑	AA	LAC	LVC	MYO	ASD.Avg↓
Supervised training (upper bound)	89.3	91.4	92.8	88.0	90.4	2.3	2.9	1.5	3.2	2.5
Source model (lower bound)	30.8	36.8	18.3	7.2	23.3	20.2	8.9	33.6	27.8	22.6
CycleGAN (Zhu et al., 2017)	73.8	75.7	52.3	28.7	57.6	11.5	13.6	9.2	8.8	10.8
PnP-AdaNet (Dou et al., 2019)	74.0	68.9	61.9	50.8	63.9	12.8	6.3	17.4	14.7	12.8
AdaOutput (Tsai et al., 2018)	73.5	80.4	76.1	48.6	69.6	15.5	5.8	5.2	6.6	8.3
AdvEnt (Vu et al., 2019)	79.5	83.0	79.5	57.7	75.0	13.9	9.3	6.9	4.5	8.7
CyCADA (Hoffman et al., 2018)	72.9	77.0	62.4	45.3	64.4	9.6	8.0	9.6	10.5	9.4
SIFA V1 (Chen et al., 2019)	81.1	76.4	75.7	58.7	73.0	10.6	7.4	6.7	7.8	8.1
SIFA V2 (Chen et al., 2020a)	81.3	79.5	73.8	61.6	74.1	7.9	6.2	5.5	8.5	7.0
EBM (Liu et al., 2021b)	78.9	80.7	75.7	60.5	74.0	8.6	6.6	4.7	8.2	7.1
CRST (Zou et al., 2019)	79.6	80.5	78.3	63.7	75.5	8.8	6.4	4.5	7.5	6.8
SECALA (Feng et al., 2023)	83.8	85.2	82.9	71.7	80.9	9.6	4.2	**3.9**	3.9	5.4
DPCL (Ours)	**90.0**	**88.7**	**88.2**	**74.5**	**85.4**	**6.6**	**4.1**	4.4	**3.7**	**4.7**

PnP-AdaNet (Dou et al., 2019), AdaOutput (Tsai et al., 2018), and AdvEnt (Vu et al., 2019) emphasize feature and output-level alignment, resulting in Dice values of 63.9%, 69.6%, and 75.0%, respectively. Confidence-regularized techniques, namely EBM (Liu et al., 2021b), CRST (Zou et al., 2019), and SECALA (Feng et al., 2023), enhance performance by mitigating overfitting, yielding Dice values of 74.0%, 75.5%, and 80.9%, respectively. However, these methods lack cross-domain global guidance and discriminative information. In contrast, our proposed DPCL attains Dice and ASD values of 85.4% and 4.7, surpassing the second-best method, SECALA, by 4.5% in terms of Dice. These findings illustrate the effectiveness of our proposed DPCL.

Comparison results on CT → MR The test results of the proposed DPCL on the CT → MR task compared with the other state-of-the-art methods are shown in Table 2. The upper and lower bounds achieved Dice values of 85.1% and 20.4%, respectively. This highlights a substantial distribution gap between domains and underscores the increased difficulty when MR is the target domain. The experimental results demonstrate that the proposed DPCL framework achieved the best average Dice and ASD values of 71.0% and 3.9, respectively. Compared to SIFA V2 (Chen et al., 2020a), which utilizes CycleGAN for data synthesis and feature alignment, DPCL remains competitive. The second best method SECALA (Feng et al., 2023), which incorporates semantic alignment and entropy constraints, produces overall inferior performance in terms of Dice and ASD values, highlighting the effectiveness of DPCL.

4.3. Ablation Study

Impact of different components To investigate the contributions of each component of DPCL to the performance, we conducted an ablation study on the MMWHS dataset, as shown in Table 3. All experiments use the base segmentation loss L_{seg}, which alone produces the performance lower bound. Our experiments initially incorporated the adversarial loss L_{adv} onto the segmentation loss L_{seg}, resulting in the Dice value of 77.3% and 66.6% for MR→CT and CT→MR, respectively. In the MR→CT adaptation task, further applying our evolved prototype contrastive loss to the source and target domains (L_{epcs} and L_{epct}) with the evolved cross-domain prototype leads to substantial enhancements, achieving Dice values of 83.1% and 84.6% and marking improvements of 5.8% and 7.3%, respectively. The best performance with a Dice value of 85.4% is achieved when the two evolved prototype contrastive loss terms, L_{epcs} and L_{epct}, are simultaneously applied to both domains. Without utilizing the evolved cross-domain prototype (i.e., dropping EP) and depending only on domain-specific prototypes from each source and target domain, the Dice value decreases to 83.8%. This shows the effectiveness of the proposed CDPE module. Similar findings are observed for CT→MR adap-

Table 2: Quantitative comparison results on the MMWHS dataset for CT → MR adaptation. We report the class average results and the results for all individual classes in terms of Dice and ASD.

Method	Dice					ASD				
	AA	LAC	LVC	MYO	Dice.Avg↑	AA	LAC	LVC	MYO	ASD.Avg↓
Supervised training (upper bound)	81.6	86.3	92.3	80.0	85.1	3.4	2.1	1.7	1.6	2.2
Source model (lower bound)	18.5	7.3	53.5	2.1	20.4	7.1	25.8	8.7	29.9	17.9
CycleGAN (Zhu et al., 2017)	64.3	30.7	65.0	43.0	50.7	5.8	9.8	6.0	5.0	6.6
PnP-AdaNet (Dou et al., 2019)	43.7	47.0	77.7	48.6	54.3	11.4	14.5	4.5	5.3	8.9
AdaOutput (Tsai et al., 2018)	52.3	71.7	79.5	49.2	63.2	9.0	3.5	5.1	5.4	5.8
AdvEnt (Vu et al., 2019)	54.4	72.0	77.5	51.8	63.9	6.8	3.2	3.9	4.0	4.5
CyCADA (Hoffman et al., 2018)	60.5	44.0	77.6	47.9	57.5	7.7	13.9	4.8	5.2	7.9
SIFA V1 (Chen et al., 2019)	67.0	60.7	75.1	45.8	62.1	6.2	9.8	4.4	4.4	6.2
SIFA V2 (Chen et al., 2020a)	65.3	62.3	78.9	47.3	63.4	7.3	7.4	3.8	4.4	5.7
EBM (Liu et al., 2021b)	65.9	64.2	76.9	49.1	64.1	6.9	7.5	5.6	3.8	6.0
CRST (Zou et al., 2019)	65.1	66.9	77.2	50.0	64.8	6.4	6.3	5.5	4.0	5.6
SECALA(Feng et al., 2023)	68.3	74.6	81.0	**55.9**	69.9	**4.9**	3.6	5.4	**3.2**	4.3
DPCL (Ours)	**70.3**	**77.0**	**82.9**	53.6	**71.0**	5.0	**2.9**	**3.2**	4.6	**3.9**

Table 3: Ablation study of the proposed components on the MMWHS dataset. We report the class average results and the results for all individual classes in terms of Dice. L_{adv}: using the adversarial loss. L_{epcs}: using the volved prototype contrastive loss for source domain. L_{epct}: using the volved prototype contrastive loss for target domain. EP: using the evolved cross-domain prototype.

				MR→CT					CT→MR				
L_{adv}	EP	L_{epcs}	L_{epct}	AA	LAC	LVC	MYO	Dice.Avg↑	AA	LAC	LVC	MYO	Dice.Avg↑
√	-	-	-	86.5	88.2	83.0	51.7	77.3	61.1	75.1	78.7	51.3	66.6
√	√	√	-	89.7	87.8	86.1	68.6	83.1	65.5	73.1	79.3	54.4	68.1
√	√	-	√	89.6	88.4	87.6	72.9	84.6	66.9	76.9	81.7	54.1	69.9
√	-	√	√	88.8	88.2	87.0	71.0	83.8	69.5	77.3	79.5	53.4	69.9
√	√	√	√	**90.0**	**88.7**	**88.2**	**74.5**	**85.4**	**70.3**	**77.0**	**82.9**	**53.6**	**71.0**

Table 4: Ablation study of using different layers of features for computing L_{epct} and L_{epcs} on the MR→CT adaptation task. We report the class average Dice and ASD results and the Dice results for all individual classes.

Layer	AA	LAC	LVC	MYO	Dice.Avg↑	ASD.Avg↓
*conv*4	90.0	88.7	88.2	74.5	**85.4**	**4.7**
*conv*5	89.8	88.4	87.9	72.7	84.7	4.9

tation. In summary, our experimental results validate the effectiveness of each component in the DPCL.

Impact of different layers of features We summarize the results of using different layers of features for computing L_{epct} and L_{epcs} on the MR → CT adaptation task of the MMWHS dataset in Table 4. In our proposed DPCL, the features from *conv*4 are employed for computing L_{epct} and L_{epcs}. With information from different resolutions, utilizing different layers of intermediate features (from *conv*4 or *conv*5) to compute the proposed L_{epct} and L_{epcs} influences the results. The results in Table 4 demonstrate that using features from *conv*4 to compute L_{epct} and L_{epcs} outperforms using features from *conv*5 across all categories, with a 0.7% improvement in the average Dice value. We attribute this to the fact that features from *conv*4 contain richer low-level information, enabling the generation of more useful feature representations

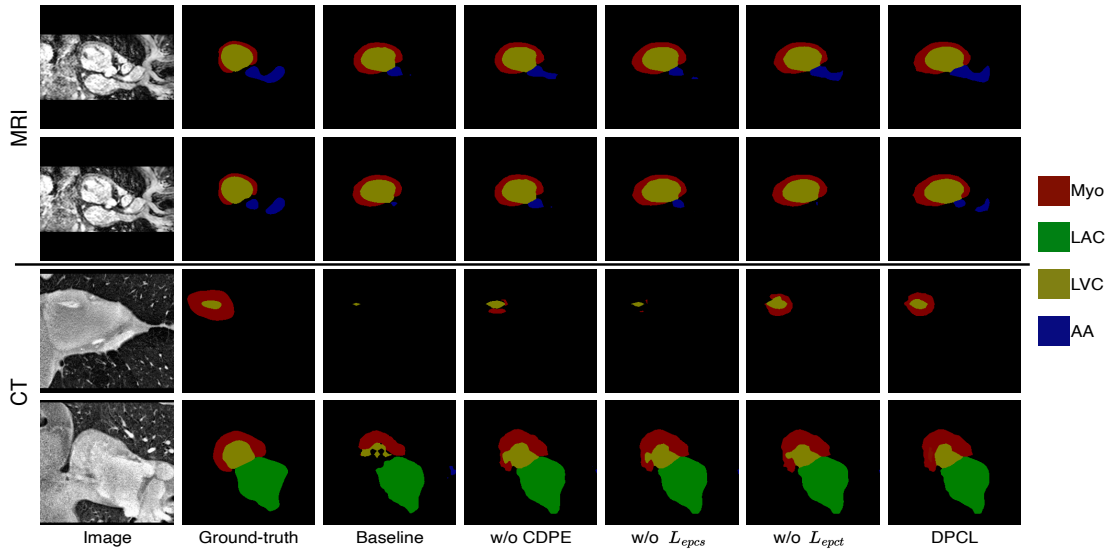

Figure 3: Visual examples of the segmentation results obtained by the proposed modules. Baseline: using L_{seg} and L_{adv}. w/o CDPE: only using the original prototypes without the updating and evolution strategies. w/o L_{cpes} or w/o L_{cpet}: without using L_{cpes} or L_{cpet}. DPCL: the full model.

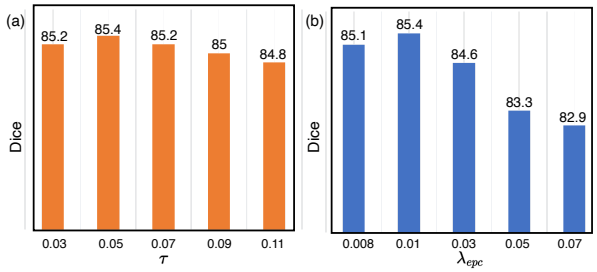

Figure 4: Impact of (a) the temperature τ and (b) the hyper-parameter λ_{eps}.

through contrastive loss. Importantly, DPCL produces satisfactory results with features from either one of the two layers, emphasizing the effectiveness of our approach without heavy dependence on a specific feature layer.

4.4. Further Analysis

Impact of the temperature τ We summarize the impact of the temperature hyperparameter τ on the MR $\rightarrow$ CT adaptation task in terms of Dice value in Figure 4 (a). In contrastive learning, the temperature parameter τ crucially regulates the severity of penalties imposed on challenging negative samples. Experimental results demonstrate superior performance when τ is set to smaller values. Conversely, increasing τ leads to a decrease in performance. The best results are obtained with a τ value of 0.05, resulting in a Dice value of 85.4%.

Impact of the weight of the evolved prototype embedding contrastive loss We summarize the impact of the weight value of L_{epcs} and L_{epct}, λ_{eps}, on the MR $\rightarrow$ CT adaptation task in terms of Dice value in Figure 4 (b). The results reveal that the optimal result of 85.4% is achieved when λ_{epc} is set to 0.01. Reducing or increasing the λ_{epc} value leads to degraded performance. This is because excessively small weights limit the effectiveness of contrastive learning, while larger weights hinder the establishment of useful contrast relations for segmentation due to imperfect prototypes.

4.5. Qualitative Evaluation Results

To illustrate the effectiveness of the proposed DPCL and its individual modules, we present visualized comparison results in Figure 3. The visualization results illustrate the effectiveness of the individual

modules, with the full DPCL providing the best segmentation results. The phenomenon of poorly identified edges and the lack of foreground contrast in medical images make semantic segmentation difficult. Through effective extraction of cross-domain global guidance and discriminative information, DPCL achieves favorable segmentation results.

5. Conclusion

In this paper, we introduce a novel DPCL framework for cross-domain medical image segmentation with unlabeled target domains. DPCL dynamically updates cross-domain global prototypes and excavates implicit discrepancy information in a contrastive manner. It utilizes a cross-domain prototype evolution module (CDPE) to facilitate the gradual transformation of prototypes from the source domain to the target domain with dynamic updating and evolutionary strategies. Moreover, a cross-domain embedding contrastive module (CDEC) is developed to enhance the discriminative capability of the segmentation model by establishing contrastive relationships between evolved prototypes and object embeddings. Experimental results demonstrate that DPCL is effective and outperforms many existing state-of-the-art methods.

References

Cheng Chen, Qi Dou, Hao Chen, Jing Qin, and Pheng-Ann Heng. Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2019.

Cheng Chen, Qi Dou, Hao Chen, Jing Qin, and Pheng Ann Heng. Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation. *IEEE transactions on medical imaging*, 39(7):2494–2505, 2020a.

Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.

Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning (ICML)*, 2020b.

Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey E Hinton. Big self-supervised models are strong semi-supervised learners. *Advances in Neural Information Processing Systems (NeurIPS)*, 2020c.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

Nanqing Dong and Eric P Xing. Few-shot semantic segmentation with prototype learning. In *The British Machine Vision Conference (BMVC)*, 2018.

Qi Dou, Cheng Ouyang, Cheng Chen, Hao Chen, Ben Glocker, Xiahai Zhuang, and Pheng-Ann Heng. Pnp-adanet: Plug-and-play adversarial domain adaptation network at unpaired cross-modality cardiac segmentation. *IEEE Access*, 7: 99065–99076, 2019.

James S Duncan and Nicholas Ayache. Medical image analysis: Progress over two decades and the challenges ahead. *IEEE transactions on pattern analysis and machine intelligence*, 22(1):85–106, 2000.

Wei Feng, Lie Ju, Lin Wang, Kaimin Song, Xin Zhao, and Zongyuan Ge. Unsupervised domain adaptation for medical image segmentation by selective entropy constraints and adaptive semantic alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2023.

Mohsen Ghafoorian, Alireza Mehrtash, Tina Kapur, Nico Karssemeijer, Elena Marchiori, Mehran Pesteie, Charles RG Guttmann, Frank-Erik de Leeuw, Clare M Tempany, Bram Van Ginneken, et al. Transfer learning for domain adaptation in mri: Application in brain lesion segmentation. In

International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), 2017.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in Neural Information Processing Systems (NeurIPS)*, 2014.

Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Hao Guan and Mingxia Liu. Domain adaptation for medical image analysis: a survey. *IEEE Transactions on Biomedical Engineering*, 69(3):1173–1185, 2021.

Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei Efros, and Trevor Darrell. Cycada: Cycle-consistent adversarial domain adaptation. In *International Conference on Machine Learning (ICML)*, 2018.

Xinrong Hu, Dewen Zeng, Xiaowei Xu, and Yiyu Shi. Semi-supervised contrastive learning for label-efficient medical image segmentation. In *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2021.

Yuankai Huo, Zhoubing Xu, Hyeonsoo Moon, Shunxing Bao, Albert Assad, Tamara K Moyo, Michael R Savona, Richard G Abramson, and Bennett A Landman. Synseg-net: Synthetic segmentation without target modality ground truth. *IEEE transactions on medical imaging*, 38(4):1016–1025, 2018.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Junnan Li, Pan Zhou, Caiming Xiong, and Steven Hoi. Prototypical contrastive learning of unsupervised representations. In *International Conference on Learning Representations (ICLR)*, 2020.

Shuang Li, Binhui Xie, Bin Zang, Chi Harold Liu, Xinjing Cheng, Ruigang Yang, and Guoren Wang. Semantic distribution-aware contrastive adaptation for semantic segmentation. *arXiv preprint arXiv:2105.05013*, 2021.

Jinlu Liu, Liang Song, and Yongqiang Qin. Prototype rectification for few-shot learning. In *European conference on computer vision (ECCV)*, 2020a.

Wei Liu, Andrew Rabinovich, and Alexander C Berg. Parsenet: Looking wider to see better. *arXiv preprint arXiv:1506.04579*, 2015.

Weide Liu, Chi Zhang, Guosheng Lin, and Fayao Liu. Crnet: Cross-reference networks for few-shot segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020b.

Weizhe Liu, David Ferstl, Samuel Schulter, Lukas Zebedin, Pascal Fua, and Christian Leistner. Domain adaptation for semantic segmentation via patch-wise contrastive learning. *arXiv preprint arXiv:2104.11056*, 2021a.

Xiaofeng Liu, Bo Hu, Xiongchang Liu, Jun Lu, Jane You, and Lingsheng Kong. Energy-constrained self-training for unsupervised domain adaptation. In *International Conference on Pattern Recognition (ICPR)*, 2021b.

Yongfei Liu, Xiangyi Zhang, Songyang Zhang, and Xuming He. Part-aware prototype network for few-shot semantic segmentation. In *European conference on computer vision (ECCV)*, 2020c.

Zhizhe Liu, Zhenfeng Zhu, Shuai Zheng, Yang Liu, Jiayu Zhou, and Yao Zhao. Margin preserving self-paced contrastive learning towards domain adaptation for medical image segmentation. *IEEE Journal of Biomedical and Health Informatics*, 26(2): 638–647, 2022.

Yawei Luo, Liang Zheng, Tao Guan, Junqing Yu, and Yi Yang. Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

Andrea I Luppi, Pedro AM Mediano, Fernando E Rosas, Negin Holland, Tim D Fryer, John T O'Brien, James B Rowe, David K Menon, Daniel Bor, and Emmanuel A Stamatakis. A synergistic core for human brain evolution and cognition. *Nature Neuroscience*, 25(6):771–782, 2022.

Huayu Mai, Rui Sun, Tianzhu Zhang, Zhiwei Xiong, and Feng Wu. Dualrel: Semi-supervised mitochondria segmentation from a prototype perspective. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

Ronald A Rensink. The dynamic representation of scenes. *Visual cognition*, 7(1-3):17–42, 2000.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2015.

Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19:221–248, 2017.

Hyungseob Shin, Hyeongyu Kim, Sewon Kim, Yohan Jun, Taejoon Eo, and Dosik Hwang. 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 (CVPR)*, pages 7412–7421, 2023.

Seung Yeon Shin, Sungwon Lee, and Ronald M Summers. Unsupervised domain adaptation for small bowel segmentation using disentangled representation. In *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2021.

Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

Xiaoyi Sun, Zhizhe Liu, Shuai Zheng, Chen Lin, Zhenfeng Zhu, and Yao Zhao. Attention-enhanced disentangled representation learning for unsupervised domain adaptation in cardiac segmentation. In *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2022.

Zar Nawab Khan Swati, Qinghua Zhao, Muhammad Kabir, Farman Ali, Zakir Ali, Saeed Ahmed, and Jianfeng Lu. Brain tumor classification for mr images using transfer learning and fine-tuning. *Computerized Medical Imaging and Graphics*, 75:34–46, 2019.

Yi-Hsuan Tsai, Wei-Chih Hung, Samuel Schulter, Kihyuk Sohn, Ming-Hsuan Yang, and Manmohan Chandraker. Learning to adapt structured output space for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Pérez. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

Kaixin Wang, Jun Hao Liew, Yingtian Zou, Daquan Zhou, and Jiashi Feng. Panet: Few-shot image semantic segmentation with prototype alignment. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.

Qingsong Xie, Yuexiang Li, Nanjun He, Munan Ning, Kai Ma, Guoxing Wang, Yong Lian, and Yefeng Zheng. Unsupervised domain adaptation for medical image segmentation by disentanglement learning and self-training. *IEEE Transactions on Medical Imaging*, 2022.

Shaoan Xie, Zibin Zheng, Liang Chen, and Chuan Chen. Learning semantic representations for unsupervised domain adaptation. In *International Conference on Machine Learning (ICML)*, 2018.

Haiming Xu, Lingqiao Liu, Qiuchen Bian, and Zhen Yang. Semi-supervised semantic segmentation with prototype-based consistency regularization. *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Chenyu You, Junlin Yang, Julius Chapiro, and James S Duncan. Unsupervised wasserstein distance guided domain adaptation for 3d multi-domain liver segmentation. In *Interpretable and Annotation-Efficient Learning for Medical Image Computing*, 2020.

Chi Zhang, Guosheng Lin, Fayao Liu, Jiushuang Guo, Qingyao Wu, and Rui Yao. Pyramid graph networks with connection attentions for region-based one-shot semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019a.

Chi Zhang, Guosheng Lin, Fayao Liu, Rui Yao, and Chunhua Shen. Canet: Class-agnostic segmentation networks with iterative refinement and attentive few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019b.

Xiaolin Zhang, Yunchao Wei, Yi Yang, and Thomas S Huang. Sg-one: Similarity guidance network for one-shot semantic segmentation. *IEEE transactions on cybernetics*, 50(9):3855–3865, 2020.

Zizhao Zhang, Lin Yang, and Yefeng Zheng. Translating and segmenting multimodal medical volumes with cycle-and shape-consistency generative adversarial network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

Ziyuan Zhao, Fangcheng Zhou, Kaixin Xu, Zeng Zeng, Cuntai Guan, and S Kevin Zhou. Le-uda: Label-efficient unsupervised domain adaptation for medical image segmentation. *IEEE Transactions on Medical Imaging*, 42(3):633–646, 2022.

Yuanyi Zhong, Bodi Yuan, Hong Wu, Zhiqiang Yuan, Jian Peng, and Yu-Xiong Wang. Pixel contrastive-consistent semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2017.

Xiahai Zhuang and Juan Shen. Multi-scale patch and multi-modality atlases for whole heart segmentation of mri. *Medical image analysis*, 31:77–87, 2016.

Danbing Zou, Qikui Zhu, and Pingkun Yan. Unsupervised domain adaptation with dual-scheme fusion network for medical image segmentation. In *Proceedings of the International Joint Conferences on Artificial Intelligence (IJCAI)*, pages 3291–3298, 2020.

Yang Zou, Zhiding Yu, Xiaofeng Liu, BVK Kumar, and Jinsong Wang. Confidence regularized self-training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.