Source free domain adaptation for medical image segmentation with fourier style mining

Chen Yang, Xiaoqing Guo, Zhen Chen, Yixuan Yuan*

Department of Electrical Engineering, City University of Hong Kong, Hong Kong SAR, China

ARTICLE INFO

ABSTRACT

Unsupervised domain adaptation (UDA) aims to exploit the knowledge learned from a labeled source dataset to solve similar tasks in a new unlabeled target domain. Existing UDA techniques typically assume that samples from source and target domains are freely accessible during the training. However, it may be impractical to access source images due to privacy concerns, especially in medical imaging scenarios with the patient information. To tackle this issue, we devise a novel source free domain adaptation framework with fourier style mining, where only a well-trained source segmentation model is available for the adaptation to the target domain. Our framework is composed of two stages: a generation stage and an adaptation stage. In the generation stage, we design a Fourier Style Mining (FSM) generator to inverse source-like images through statistic information of the pretrained source model and mutual Fourier Transform. These generated source-like images can provide source data distribution and benefit the domain alignment. In the adaptation stage, we design a Contrastive Domain Distillation (CDD) module to achieve feature-level adaptation, including a domain distillation loss to transfer relation knowledge and a domain contrastive loss to narrow down the domain gap by a self-supervised paradigm. Besides, a Compact-Aware Domain Consistency (CADC) module is proposed to enhance consistency learning by filtering out noisy pseudo labels with shape compactness metric, thus achieving output-level adaptation. Extensive experiments on cross-device and cross-centre datasets are conducted for polyp and prostate segmentation, and our method delivers impressive performance compared with state-of-the-art domain adaptation methods. The source code is available at https://github.com/CityU-AIM-Group/SFDA-FSM.

1. Introduction

Convolutional neural networks (CNNs) have achieved remarkable success for automatic medical image analysis in recent years (Ronneberger et al., 2015; Guo et al., 2020; Yuan et al., 2019). However, a large amount of labeled data is required for model training, which is time-consuming and error-prone especially in medical image segmentation task that needs pixel-wise annotations (Litjens et al., 2017). Moreover, a segmentation model trained on one clinical centre (i.e., source domain) often fails to generalize well when deployed in a new centre (i.e., target domain), due to the discrepancy in data distribution (Ghafoorian et al., 2017).

To tackle these problems, unsupervised domain adaptation (UDA) has been widely explored to transfer knowledge from sufficiently labeled source data to unlabeled target data (Tzeng et al., 2017), as illustrated in Fig. 1(a). Existing UDA methods

(Tzeng et al., 2017; Tsai et al., 2018; Hoffman et al., 2018; Chen et al., 2020; Zhang et al., 2019; Xing et al., 2019; Liu et al., 2021a) typically assume that the source dataset is available during the training, and utilize distribution alignment to diminish the domain discrepancy between source and target domains. But the requirement of accessing to the source domain data is sometimes impractical and restrictive in the real clinical applications due to the privacy or security concerns (Yang et al., 2021b). Moreover, some medical imaging datasets, such as videos or high-resolution images, are extremely large and may not be efficient for data storage and transmission. Hence, source free domain adaptation (SFDA) (Wang et al., 2021; Liang et al., 2020; Kim et al., 2020; Li et al., 2020; Kundu et al., 2020; Li et al., 2021; Qiu et al., 2021; Fleuret et al., 2021), which only requires a well-trained source model and an unlabeled target domain dataset for adaptation (Fig. 1(b)), has a high practical value in medical image analysis. For instance, with a well-trained model build on one hospital data, SFDA allows other clinical centres to adapt the model on their own data without exchanging sensitive health data, which largely saves time-consuming annotations. Compared with UDA, SFDA transfers

---

* Corresponding author.
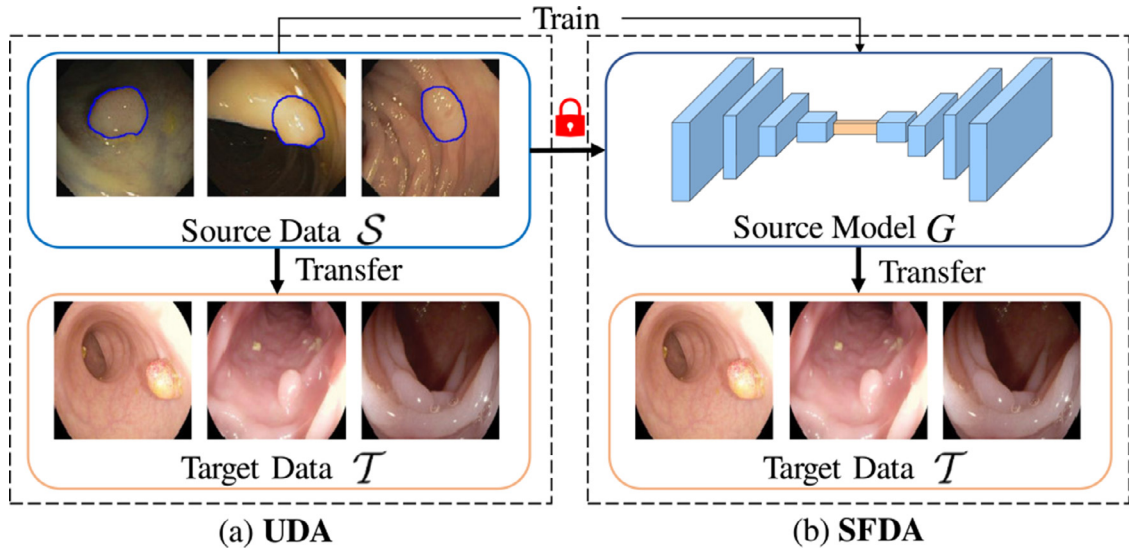 *E-mail address:* yxyuan.ee@cityu.edu.hk (Y. Yuan).

Fig. 1. Setting comparison between unsupervised domain adaptation (UDA) and source free domain adaptation (SFDA). Different from UDA that is able to access source data, SFDA merely utilizes the source model (i.e., a segmentation model) to achieve adaptation for unlabeled target domain.

the knowledge of the source model rather than the source data to unlabeled target domain. Although SFDA methods (Wang et al., 2021; Liang et al., 2020; Kim et al., 2020; Li et al., 2020; Kundu et al., 2020; Li et al., 2021; Qiu et al., 2021; Fleuret et al., 2021) have made great progress for natural image classification and object detection, there is still a blank in the source free domain adaptive medical image segmentation.

The major challenge of SFDA lies in the unavailability of source data, thus previous distribution alignment based UDA methods through adversarial learning (Tzeng et al., 2017; Tsai et al., 2018; Hoffman et al., 2018) and image translation (Chen et al., 2019; Zhu et al., 2017) cannot be applied directly. To access the source data distribution, we attempt to utilize the statistic information of source model to generate source-like images and resolve the problem of SFDA. Note that batch normalization (BN) layers store the moving average statistics of activations, which preserve distribution information of previously trained data (Yin et al., 2020). These model statistics are valuable to approximate data distribution and can be utilized to represent style information (Li et al., 2018). Considering that shallow features contain appearance information while deep features focus on semantic information (Pan et al., 2018), we apply the BN constraints on shallow features as the style loss and force deep features similar to the target ones as the content loss. Following traditional neural style transfer approaches that optimize noise image by feature discrepancy of different layers in a forzen VGG network (Gatys et al., 2016; Li and Wand, 2016; Zhang et al., 2018), we optimize input Gaussian noise image with the proposed style loss and content loss to generate source-like images that capture style information of source data and preserve content information of target data.

Even if these generated images have obvious source style information, the ambiguous problem still exists (Gatys et al., 2016). As the observation that low frequency part of the images can capture style information (Yang and Soatto, 2020; Yang et al., 2020), we apply mutual Fourier Transform to replace the style components of target image with the source-like ones, thus further improving the quality of generated images.

With generated source-like images, previous adversarial learning based methods (Tzeng et al., 2017; Hoffman et al., 2018; Tsai et al., 2018) can be utilized to reduce the domain gap and enhance the segmentation performance on the target domain. These adver-

sarial learning based methods (Tzeng et al., 2017; Hoffman et al., 2018; Tsai et al., 2018) can be divided into feature-level adaptation and output-level adaptation. Previous feature-level adaptation (Tzeng et al., 2017; Hoffman et al., 2018) aims at learning domain-invariant features by minimizing distribution distance between source and target samples, but ignores the corresponding structure relationship. Output-level adaptation (Tsai et al., 2018) is commonly based on the assumption that label space distribution of source and target domains is similar, but this assumption fails for medical image segmentation task where the spatial distribution of lesions or organs vary significantly. Moreover, the generated source-like images preserve semantic information and the corresponding content information share a large similarity with the target images, hence previous adversarial learning based methods may be hard to converge. To deal with these challenges, we introduce a Contrastive Domain Distillation (CDD) to transfer relation knowledge among the generated source-like images and target images in a self-supervised paradigm. The proposed CDD learns domain-invariant distribution by aligning cross domain structure-wise relationships, enabling the feature-level adaptation. As the generated source-like images and corresponding target images share same semantic information, they tend to produce same segmentation prediction. Hence, we propose Compact-Aware Domain Consistency (CADC) module to apply consistency learning (Kim and Byun, 2020) on these images to achieve output-level adaptation in SFDA. A shape compactness metric is further designed to filter out noisy pseudo labels, thus providing accurate supervision for the consistency learning.

In this work, we propose a SFDA framework with fourier style mining for medical image segmentation, composing of a generation stage and an adaptation stage. To the best of our knowledge, this represents the first effort to solve unsupervised domain adaptation with only the source model on medical image segmentation. In summary, our main contributions are listed as follows:

- In the generation stage, we propose a Fourier Style Mining (FSM) generator to produce high-quality source-like images from the well-trained source segmentation model through BN constraints and mutual Fourier Transform. FSM represents the

first attempt to generate source-like images from a segmentation model to deal with domain shift in SFDA.

- In the adaptation stage, CDD and CADC modules are proposed to achieve feature-level and output-level alignment based on paired source-like and target images. The CADC introduces shape compactness metric to filter out noisy pseudo labels for medical image segmentation.

- Comprehensive experiments on different domain scenarios (including modalities, vendors and clinical cites), demonstrate the effectiveness of our style mining guided SFDA framework, which outperforms state-of-the-art domain adaptation methods.

2. Related work

2.1. Unsupervised domain adaptation

Unsupervised domain adaptation (UDA) aims to adapt the knowledge learned in labeled source data to an unlabeled target domain. Recent UDA methods (Tzeng et al., 2017; Tran et al., 2019; Hoffman et al., 2018; Chen et al., 2019; Zhu et al., 2017; Tsai et al., 2018) can be categorized into two groups: image-level adaptation (Chen et al., 2019; Zhu et al., 2017) and feature-level adaptation (Tzeng et al., 2017; Tran et al., 2019). Image-level adaptation (Chen et al., 2019; Zhu et al., 2017) generates synthesized data to bridge the domain shift in appearance among source and target domains. While feature-level adaptation (Tzeng et al., 2017; Tran et al., 2019) learns domain-invariant representations by minimizing the specific distribution distance between the source and target domains. To further reduce the domain gap, Hoffman et al. (2018) jointly utilized image-level adaptation with feature-level adaptation and proposed a cycle-consistent framework.

Since automatic medical image analysis requires time-consuming labeling and professional knowledge, domain adaptation models attract great attentions recently (Chen et al., 2020; Yang et al., 2021a; Zhang et al., 2019; Xing et al., 2019; Guo et al., 2021). For instance, Chen et al. (2020) proposed synergistic image and feature alignments to apply CycleGAN (Zhu et al., 2017) to achieve adaptation between CT and MRI images. Zhang et al. (2019) learned domain-invariant feature representations by adversarial learning for WSI and MSI images. Xing et al. (2019) combined CycleGAN (Zhu et al., 2017) with self-training to achieve cell detection adaptation from bright-field microscopy to phase-contrast microscopy.

Despite the great progresses, these methods (Tzeng et al., 2017; Tran et al., 2019; Hoffman et al., 2018; Chen et al., 2019; Chen et al., 2020; Zhang et al., 2019; Xing et al., 2019) assume the availability of both source and target data. But this assumption fails in medical imaging scenarios where source images are not accessible due to the privacy concerns.

2.2. Source relaxed domain adaptation

To solve UDA problem under the privacy concern, source relaxed domain adaptation (SRDA) (Bateson et al., 2020; Stan and Rostami, 2021), is proposed recently. Different from previous UDA methods that have to access source data, SRDA only utilizes the well-trained source model and additional source information to achieve adaptation to the target domain. Bateson et al. (2020) integrated entropy minimization loss with a class-ratio constraint to achieve adaptation to the target domain with the source model and an additional class ratio predictor. Stan and Rostami (2021) first trained a prototypical distribution in the source domain, and then aligned target features to source prototypical distribution to achieve adaptation. However, this additional source information introduces extra computation and may also violate patient privacy information.

2.3. Source free domain adaptation

Due to data privacy concern, source free domain adaptation (SFDA) (Wang et al., 2021; Liang et al., 2020; Kim et al., 2020; Li et al., 2020; Kundu et al., 2020; Qiu et al., 2021; Fleuret et al., 2021) is recently proposed to achieve adaptation with only the source model rather than the source data. Wang et al. (2021) directly applied minimum entropy regularization on unlabeled target data to guide target domain training. Liang et al. (2020) exploited self-supervised pseudo-labeling and information maximization to constrain unlabeled target data and learn target-specific representations. Qiu et al. (2021) generated feature prototypes of each source class and target pseudo labels for domain alignment. Fleuret et al. (2021) applied entropy minimization to reduce the uncertainty of target predictions. Kim et al. (2020) utilized target-oriented pseudo labels obtained from the adaptive prototype memory to train the target model and prevented the target model from a self-biasing problem by source-oriented pseudo labels.

Even if these SFDA methods (Wang et al., 2021; Liang et al., 2020; Kim et al., 2020; Li et al., 2020; Kundu et al., 2020; Qiu et al., 2021; Fleuret et al., 2021) have made great progresses on object recognition and semantic segmentation, SFDA has not been well-investigated on medical image segmentation. Moreover, they simply apply the source model as a pre-trained model to generate pseudo labels for unlabeled target samples, ignoring distribution information stored in the source model. In contrast, we utilize style information stored in BatchNorm statistics of the source model to generate source-like images, thus making full use of the source model. Besides, instead of directly training target model with pseudo labels (Liang et al., 2020; Liu et al., 2021b), we consider the compact properties of lesions and organs in medical imaging scenarios and introduce a shape compactness metric to re-weight pseudo labels.

3. Methodology

In this paper, we design a fourier style mining guided SFDA framework for medical image segmentation, composing of a generation stage to generate source-like images and an adaptation stage to adapt source model to the target domain based on paired source-like and target images. The overall pipeline of the proposed approach is illustrated in Fig. 2.

Given unlabeled target images x_t and Gaussian noise image z, they are first fed into frozen source model G to produce coarse source-like images x_c by BN constraints. The input noise image is optimized through a style loss $\mathcal{L}_{style}$ and a content loss $\mathcal{L}_{content}$, which follows traditional neural style transfer approaches (Gatys et al., 2016; Li and Wand, 2016; Zhang et al., 2018). Then mutual Fourier Transform is applied to target images x_t and coarse generations x_c to generate high-quality fine generations x_g. We then augment x_t and the corresponding x_g to construct the augmented inputs $\{\hat{x}_t^1, \hat{x}_t^2, ..., \hat{x}_t^n\}$ and $\{\hat{x}_g^1, \hat{x}_g^2, ..., \hat{x}_g^n\}$. With the domain distillation loss $\mathcal{L}_{distill}$ and the domain contrastive loss $\mathcal{L}_{contrast}$, the CDD module optimizes the segmentation model to transfer relation knowledge and narrows down the domain gap by a self-supervised paradigm. For the segmentation predictions of x_g and x_t, the CADC module applies consistency learning to achieve output-level adaptation, and reweights pseudo labels by a shape compactness metric with self-training loss $\mathcal{L}_{consist}$. Through optimizing $\mathcal{L}_{contrast}$, $\mathcal{L}_{distill}$ and $\mathcal{L}_{consist}$, we can leverage the knowledge of the well-trained source model to improve segmentation performance on the target domain.
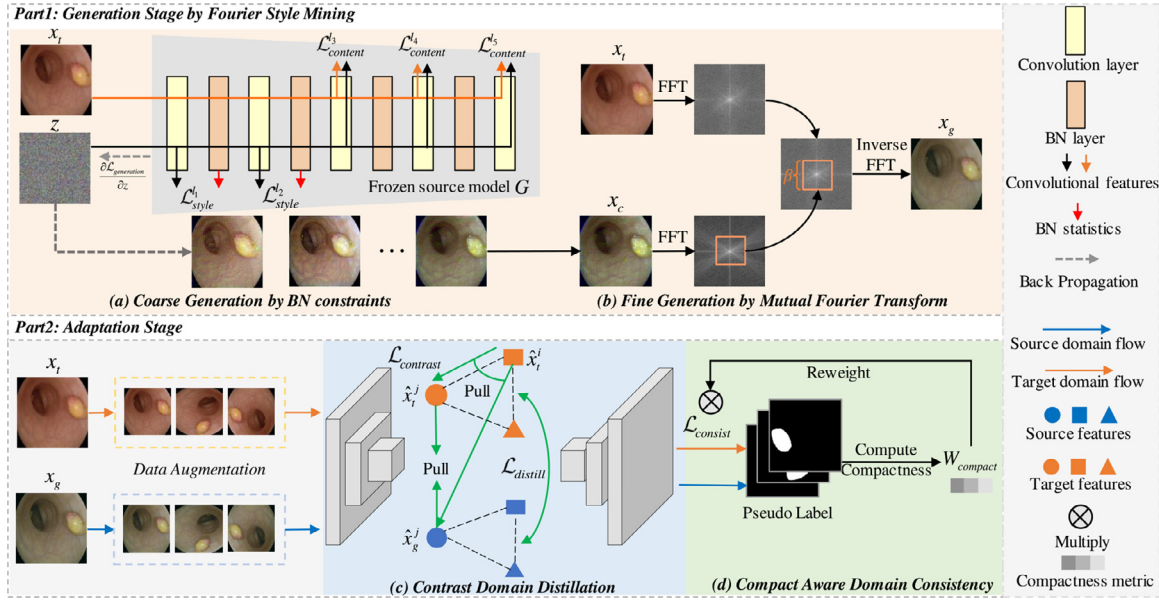
Fig. 2. The pipeline of the proposed style mining guided SFDA framework. *Part*1: the generation stage aims to optimize input noise image z by batch normalization constraints, and generate coarse source-like image x_c; then a mutual Fourier Transform is applied to obtain high-quality source-like image x_g. *Part*2: the adaptation stage optimizes the segmentation model G by contrast domain distillation (CDD) and compact-aware domain consistency (CADC).

3.1. Generation stage by fourier style mining

Previous SFDA methods (Wang et al., 2021; Liang et al., 2020; Kim et al., 2020; Li et al., 2020; Kundu et al., 2020; Bateson et al., 2020; Qiu et al., 2021; Fleuret et al., 2021) simply utilize the well-trained source model to generate pseudo labels and retrain a target model. But these methods ignore the distribution information stored in BN statistics of the source model, which could provide valuable information for the domain adaptation. To make full use of the pretrained model on source domain, we introduce a Fourier Style Mining (FSM) generator to produce high-quality source-like images through BN constraints and mutual Fourier Transform. The FSM generator is composed of two steps: coarse generation by BN constraints and fine generation by mutual fourier transform.

3.1.1. Coarse generation by BN constraints

In coarse generation stage, we utilize style loss and content loss to optimize noise image z, thus generating source-like images x_c. The generated source-like images can provide source data distribution to benefit domain alignment. This optimization process for input noise image is inspired by traditional neural style transfer methods (Gatys et al., 2016; Li and Wand, 2016; Zhang et al., 2018). The corresponding workflow is shown in Fig. 2(a).

For the style loss, BN layers store the moving average statistics: running means and running variances for batch-wise images, which implicitly capture the feature distribution of the training data. Moreover, shallow features contain appearance information while deep features focus on semantic information (Pan et al., 2018). In light of these, we utilize BN statistics of first P blocks that are stored in the source segmentation model as a representative of the source style information. Then the style loss $\mathcal{L}_{style}$ is defined as follows to optimize the noise image z to follow source data distribution.

$$\mathcal{L}_{style} = \frac{1}{P} \sum_{n=1}^{P} \mathcal{L}_{style}^{l_n}$$

$$= \frac{1}{P} \sum_{n=1}^{P} ||\mu_z^n - \mu_{stored}^n||_2 + ||\sigma_z^n - \sigma_{stored}^n||_2, \quad (1)$$

where μ_z^n and σ_z^n represent the mean and standard variation of the noise image in the l_n layer, respectively; μ_{stored}^n and σ_{stored}^n denote the running mean and the square root of running variance stored in the source model, respectively. $|| \cdot ||_2$ represents the Euclidean distance. Through minimizing the style loss, we can mine out the latent style information stored in the source segmentation model and generate the source-like images optimized from the input noise images.

To generate the source-like images that preserve the semantic information of original target images x_t, we regard target images x_t to provide content knowledge, and calculate the feature discrepancy of x_t and z in last Q layers as the content loss,

$$\mathcal{L}_{content} = \frac{1}{Q} \sum_{n=P+1}^{P+Q} \mathcal{L}_{content}^{l_n}$$

$$= \frac{1}{Q} \sum_{n=P+1}^{P+Q} ||f^n(z) - f^n(x_t)||_2, \quad (2)$$

where $f^n(z)$ and $f^n(x_t)$ denote feature maps of noise image and the target images at layer $n \in \{P+1, P+2, ..., P+Q\}$. We follow the setting of Zhang et al. (2018) to set P and Q as 2 and 3 with ResNet backbone (He et al., 2016).

Different from previous style transfer methods (Gatys et al., 2016; Huang and Belongie, 2017; Li and Wand, 2016; Zhang et al., 2018) for natural images that can access style images to extract style information from shallow features, our coarse generation does not require these source images under privacy concerns, and only utilizes BN statistics of the source segmentation model to mine out the source style information. Note that style loss and content loss are trade-off to the optimization of input noise image z, only optimizing by these two losses can achieve source free image translation.

3.1.2. Fine generation by mutual fourier transform

As shown in Fig. 2(a), the generated images by BN constraints have obvious source style but is ambiguous with undesirable texture information. As low-frequency components in Fourier Transform can represent the style information (Yang and Soatto, 2020), we introduce a mutual Fourier Transform to ex-

change low-frequency amplitude of the generated images with that of original target images to obtain high quality source-like images as shown in Fig. 2(b).

Specifically, for the generated coarse images x_c and the corresponding target images x_t, we apply Fourier Transform $\mathcal{F}$ and extract the amplitude and phase components $\mathcal{F}^A(x_c)$, $\mathcal{F}^P(x_c)$, $\mathcal{F}^A(x_t)$ and $\mathcal{F}^P(x_t)$. Then through exchanging low-frequency parts on a mask region $M_\beta = \mathbb{1}_{(h,w)\in[-\beta H:\beta H, -\beta W:\beta W]}$ and mapping them back to RGB space by inverse Fourier Transform, the fine generated source-like images are derived through:

$$x_g = \mathcal{F}^{-1}([M_\beta \circ \mathcal{F}^A(x_c) + (1 - M_\beta) \circ \mathcal{F}^A(x_t), \mathcal{F}^P(x_t)]), \tag{3}$$

where $\mathcal{F}^{-1}$ represents inverse Fourier Transform, $\circ$ is element-wise multiplication, and β controls size of exchanged mask region. This process is illustrated in Fig. 2(b), where the mask M_β is marked by the orange box. It is a binary mask to control the scale of low-frequency component within amplitude spectrum to be exchanged, whose value is 1 at the central region and 0 elsewhere. Through mutual Fourier Transform, we can obtain high quality source-like images, which provide distribution information of the source domain to explicitly reduce the domain gap.

3.2. Adaptation stage

With high-quality source-like images generated in the generation stage, we feed paired source-like images x_g and target samples x_t into the segmentation model G to learn domain-invariant representations by distribution alignment. Specifically, CDD and CADC modules are proposed to achieve feature-level and output-level alignment.

3.2.1. Contrastive domain distillation (CDD)

Different from previous UDA methods (Tsai et al., 2018; Tzeng et al., 2017; Hoffman et al., 2018) that reduce the distribution gap between unpaired source and target data by adversarial learning, the CDD module introduces a domain distillation loss $\mathcal{L}_{distill}$ to transfer relation knowledge among two domains and domain contrastive loss $\mathcal{L}_{contrast}$ to narrow down the domain gap by a self-supervised paradigm as illustrated in Fig. 2(c). The proposed CDD module, transferring relation-based knowledge, is more stable and robust than previous adversarial learning methods that transfer feature-based knowledge (Gou et al., 2021; Peng et al., 2019).

Specifically, we conduct data augmentation to pair-wise target images x_t and generated source-like images x_g to construct two domain set $\{\hat{x}_t^1, \hat{x}_t^2, ..., \hat{x}_t^n\}$ and $\{\hat{x}_g^1, \hat{x}_g^2, ..., \hat{x}_g^n\}$ and compute structure-wise distance within each domain:

$$\psi(\hat{x}^i, \hat{x}^j) = \| \hat{f}^i - \hat{f}^j \|_2, \tag{4}$$

where $\hat{x}^i$ and $\hat{x}^j$ denote augmented samples, $\hat{f}^i$ and $\hat{f}^j$ are deep features of $\hat{x}^i$ and $\hat{x}^j$ in last layer of ResNet-101. Aiming at learning structure-wise knowledge by mutual relations between these two domain sets, we calculate the distillation loss between different domain sets:

$$\mathcal{L}_{distill} = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \| \psi(\hat{x}_t^i, \hat{x}_t^j) - \psi(\hat{x}_g^i, \hat{x}_g^j) \|_2. \tag{5}$$

By constraining corresponding cross-domain structure-wise distances, we can narrow down the distribution gap between the source and target domains.

Moreover, a domain contrastive loss is designed to enable discriminative data representations from unlabeled data through the self-supervised paradigm. Different from previous contrasitive learning (Kang et al., 2019) that takes the class information into account, our domain contrastive loss focus on minimizing domain

discrepancy. Given an anchor sample $\hat{x}_t^i$, we define the corresponding positive sample $\hat{x}_t^j$ from same domain set and negative sample $\hat{x}_g^j$ from different domain sets. Since the difference among the positive sample and the negative sample lies in style information, the feature discrepancy $\psi(\hat{x}_t^j, \hat{x}_g^j)$ is small when the model is domain-invariant. Furthermore, feature discrepancy $\psi(\hat{x}_t^i, \hat{x}_t^j)$ between the anchor sample and the positive sample within same domain tends to be lower than feature discrepancy $\psi(\hat{x}_t^i, \hat{x}_g^j)$ between the anchor sample and the negative sample that come from different domains. With these sample relationships, the domain contrastive loss is calculated to pull features of samples with same semantics closer by:

$$\mathcal{L}_{contrast} = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \psi(\hat{x}_t^j, \hat{x}_g^j) + ||\psi(\hat{x}_t^i, \hat{x}_g^j) - \psi(\hat{x}_t^i, \hat{x}_t^j)||_1, \tag{6}$$

where the first term minimizes feature discrepancy between samples with same contents but different styles, which can explicitly narrow down the domain gap. The second term minimizes distance between intra-domain feature discrepancy and inter-domain feature discrepancy to implicitly eliminate domain shift. Combining these two terms together, the domain contrastive loss can learn the inherent relationship between samples by self-supervised learning, and thereby enhance the generalization of the proposed SFDA framework.

Combining the domain distillation loss with the domain contrastive loss, we can learn domain-invariant representations on structure-wise relationships. These two losses utilize the self-supervised paradigm to explore high-level semantic information with the absence of ground truth supervision.

3.2.2. Compact-aware domain consistency (CADC)

As the target images x_t and generated source-like images x_g share same semantics but with different style information, they should produce similar segmentation results. Hence, we introduce a CADC module to apply consistency learning on x_t and x_g to achieve output-level adaptation as shown in Fig. 2(d). We first forward fine generations x_g into the source model G to obtain pseudo labels $\hat{y}$. Then versatile consistency constraint is represented as below:

$$\mathcal{L}_{consist} = -\sum_{h,w} \sum_{c=1}^{C} (\hat{y}^{h,w,c} log P_t^{(h,w,c)} + \hat{y}^{h,w,c} log P_g^{(h,w,c)}), \tag{7}$$

where P_t and P_g represent prediction probability maps of target images x_t and source-like images x_g. This consistency loss forces the segmentation predictions of different style images (i.e., x_t and x_g) that share the same segmentation structure to be similar, thus enabling the segmentation model to filter out style information and learn domain-invariant features.

Although the consistency learning could narrow down the domain gap, it just utilizes pseudo labels generated through the source model for the model training, which may contain noises and lead to bias performance of the target domain. As clean pseudo labels are always complete and noisy pseudo labels are disorganized with random distributed pixels, we introduce a compactness reweighting mechanism to filter out noisy pseudo labels by shape compactness metric. The shape compactness metric is defined as follows to measure the reliability of pseudo labels:

$$\mathcal{W}_{compact} = \frac{A}{M^2} = \frac{\sum_{i\in\Omega} |\hat{y}_i|}{\sum_{i\in\Omega} \sqrt{(\nabla\hat{y}_{ui})^2 + (\nabla\hat{y}_{vi})^2}}, \tag{8}$$

where M represents the perimeter of the pseudo label boundary, and A is the area of the pseudo label $\hat{y}$. $\hat{y}_{ui}$ and $\hat{y}_{vi}$ are the probability gradients for each pixel i in horizontal and vertical directions.
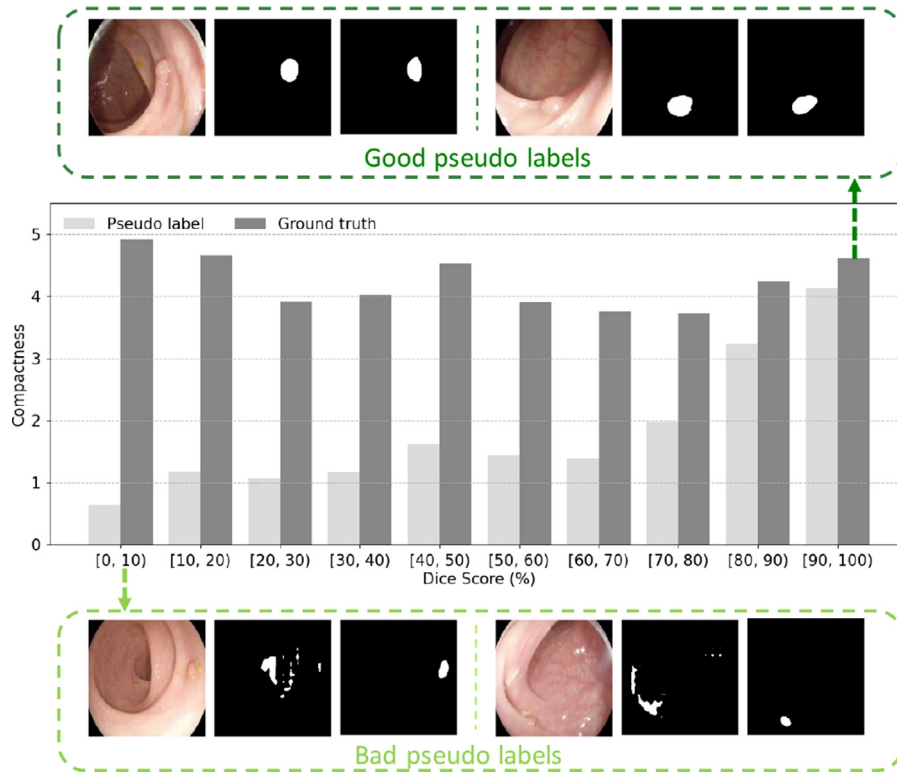
Fig. 3. Relationship between shape compactness (vertical axis) and reliability (horizontal axis) of pseudo labels. The reliability is measured via *Dice score* between pseudo label and ground truth label. Four typical target examples are illustrated to explain the relationship, and each example includes image x_t, pseudo label $\hat{y}$ and ground truth y from left to right. The upper part visualizes two typical examples with compact pseudo labels that are matched with ground truth. The lower part includes examples with fragmentary pseudo labels that are unreliable.

Ω is the pixel set of pseudo label $\hat{y}$. Noisy pseudo labels usually have larger perimeter with pixels distributed irregularly, leading to the small shape compactness metric.

We visualize the relationship among the shape compactness metric and reliability of pseudo labels (*Dice score* with ground truth) in Fig. 3. It is evident that the shape compactness metric can intuitively reflect the reliability of the pseudo label. With the shape compactness of pseudo label increasing, the corresponding reliability improves. Hence, large weights should be assigned to those samples whose pseudo labels are with large compactness metric. Given generated pseudo labels, we first calculate the shape compactness of these pseudo labels, and then reweight previous consistency loss in Eq. (7) through the shape compactness metric, as follows:

$$\mathcal{L}_{st} = \mathcal{W}_{compact}\mathcal{L}_{consist}. \tag{9}$$

With the CADC module, the proposed SFDA framework mitigates the influence of noisy pseudo labels and achieves output-level adaptation among different domains.

3.3. Optimization summary

The proposed style mining guided SFDA framework consists of two stages: a generation stage and an adaptation stage. In the generation stage, we input noise image z and target image x_t to the fixed source trained model G and optimize z by minimizing the generation loss:

$$\mathcal{L}_{generation} = \lambda_{style}\mathcal{L}_{style} + \lambda_{content}\mathcal{L}_{content}, \tag{10}$$

where λ_{style} and $\lambda_{content}$ are scalar weights to balance the two loss components. The gradient with respect to the pixel values $\frac{\partial \mathcal{L}_{generation}}{\partial z}$ can be used to optimize input Gaussian noise image z.

After step by step optimization on z, we can obtain coarse source-like image x_c. We then apply mutual Fourier Transform on x_c and x_t to get fine source-like image x_g.

In the adaptation stage, we first forward x_g into the source trained model G to generate pseudo label $\hat{y}$. Then we feed paired source-like and target images x_g and x_t into the source trained model G and optimize the model by minimizing the adaptation loss:

$$\mathcal{L}_{adaptation} = \mathcal{L}_{st} + \lambda_{contrast}\mathcal{L}_{contrast} + \lambda_{distill}\mathcal{L}_{distill}, \tag{11}$$

where $\lambda_{contrast}$ and $\lambda_{distill}$ are hyper-parameters to balance the optimization. The optimization procedure of our style mining guided SFDA framework is summarized in Algorithm 1.

4. Experiments

4.1. Datasets

Comprehensive experiments were conducted on both polyp and prostate segmentation to evaluate the proposed style mining guided SFDA framework, covering cross-device and cross-centre domain scenarios. Specifically, cross-device SFDA experiment was conducted on colonoscopy images for polyp segmentation, and cross-centre SFDA experiment was conducted on MRI images for prostate segmentation.

4.1.1. Polyp segmentation

The public available colonoscopy datasets EndoScene (Vázquez et al., 2017) and ETIS-Larib (Silva et al., 2014), were adopted for cross-device adaptation. The EndoScene dataset consists of 912 images acquired from 36 patients. It was collected by Olympus Q160AL and Q165L, Extra II video processor. We set the EndoScene dataset as the source domain and followed the

---

Algorithm 1 Optimization

Require: Source model G and unlabeled target data $D_t(x_t)$
Ensure: Fine generation x_g and target prediction p_t

1: Update input noise image z by Eq. (10) to obtain coarse generation x_c
2: Obtain fine generation x_g by Eq. (3) based on x_c and x_t
3: Generate pseudo labels $\hat{y}$ and corresponding shape compactness metric $\mathcal{W}_{compact}$ with fine generations x_g on frozen model G
4: **for** k = 1: *epoch* **do**
5:   Construct pair-wise domain sets $\{\hat{x}_t^1, \hat{x}_t^2, ..., \hat{x}_t^n\}$ and $\{\hat{x}_g^1, \hat{x}_g^2, ..., \hat{x}_g^n\}$ via data augmentations to x_t and x_g
6:   Feed two domain sets into G and generate corresponding features $\{\hat{f}_t^1, \hat{f}_t^2, ..., \hat{f}_t^n\}$ and $\{\hat{f}_g^1, \hat{f}_g^2, ..., \hat{f}_g^n\}$
7:   Optimize G with feature-level and output-level adaptation by Eq. (11)
8:   Update pseudo labels $\hat{y}^t$ and corresponding shape compactness metric $\mathcal{W}_{compact}$ with target images x_t on optimized model G
9: **end for**

---

standard setup for polyp segmentation as in Vázquez et al. (2017), leading to 547 training, 183 validation, and 182 testing colonoscopy images. The ETIS-Larib dataset (Silva et al., 2014) consists of 196 frames collected by Pentax 90i series, EPKi 7000 video processor. This dataset was used as the target domain and randomly split into 80% frames for training and 20% frames for testing. We resized all the images to 256 × 256 for training and testing.

4.1.2. Prostate segmentation
We collected prostate T2-weighted MRI images from two different clinical centres to evaluate the performance for cross-centre source free domain adaptation. Site A included 421 slices of 30 MRI samples by Radboud University Nijmegen Medical Centre from NCI-ISBI 2013 challenge (Nicholas Bloch, 2015). Site B included 261 slices of 12 MRI samples by Beth Israel Deaconess Medical Center from PROMISE12 challenge (Litjens et al., 2014). All the slices were sampled from axial plane and resized to 256 × 256 for prostate segmentation.

4.2. Implementation details

The segmentation model of polyp and prostate datasets was based on Deeplab-v2 (Chen et al., 2017) with ResNet-101 (He et al., 2016) backbone pre-trained on ImageNet (Deng et al., 2009). We ran the experiments with the PyTorch library on an NVIDIA 2080 Ti GPU. To obtain the source model, we first trained a segmentation network on labeled source data with a batch size of 8 for 150 epochs. SGD was chosen for optimization with learning rate of 1e-3. In the generation stage, we optimized the input noise image by Adam optimizer with a learning rate of 1e-2 for 100 epochs. In the adaptation stage, the segmentation model was trained on source-like and target images with a batch size of 4 for 150 epochs, and optimized by SGD with a learning rate of 5e-4. Considering the convolution layer is not rotation invariant but the segmentation model requires the property of rotation invariance, we applied the image rotation with random degrees as data augmentation strategy in the CCD module and regularized the segmentation model to be rotation consistent. The configuration of hyper-parameters in the framework included $\beta = 5e\text{-}3$, $\lambda_{style} = 1$, $\lambda_{content} = 0.5$, $n = 2$, $\lambda_{contrast} = 0.01$ and $\lambda_{distill} = 0.1$. Seven common criteria were applied to evaluate the performance on both polyp and prostate segmentation, including Dice similarity coefficient (*Dice*), Specificity (*Spe*), Sensitivity (*Sen*), Accuracy (*Acc*), IoU of polyp/prostate regions (IoU_P), IoU of backgrounds (IoU_B) and Mean IoU (*mIoU*).

4.3. Comparison with state-of-the-art domain adaptation methods

To evaluate the effectiveness of the proposed style mining guided SFDA framework, we compared our method with state-of-the-art unsupervised domain adaptation (UDA) (Yang and Soatto, 2020; Zhang et al., 2021), source relaxed domain adaptation (SRDA) (Bateson et al., 2020; Stan and Rostami, 2021) and source free domain adaptation (SFDA) (Liang et al., 2020; Liu et al., 2021b) methods in semantic segmentation. The oracle methods, including 'Source only' and 'Target only', represent segmentation models trained with only source data or target data, which can be seen as the lower bound and upper bound for domain adaptation problem. UDA methods (Yang and Soatto, 2020; Zhang et al., 2021) can access source images and corresponding annotations to transfer source domain knowledge to the target domain. SFDA methods (Liang et al., 2020; Liu et al., 2021b) provide a source model but without source data to achieve adaptation to the target domain. Besides the source model, SRDA methods (Bateson et al., 2020; Stan and Rostami, 2021) can access some additional information of the source domain, such as source prototypical distribution (Stan and Rostami, 2021) or source class ratio predictor(Bateson et al., 2020), to guide the training of segmentation model in the target domain. Since these methods are trained on different tasks with natural datasets, directly applying their models is not appropriate. To ensure fair comparison with these methods, we retrain these methods on our polyp and prostate datasets with the same backbone ResNet (He et al., 2016).

4.3.1. Results on polyp segmentation
The quantitative performance of different domain adaptation methods on the polyp segmentation task is reported in Table 1. It indicates that our approach attains the best *Sen* of 80.923% and a promising *Dice* of 74.123% compared with other baselines. Specifically, our method improves source only method with performance gain of 18.102% in *Dice* and 24.490% in *Sen*, demonstrating the successful adaptation to the target domain. In comparison to state-of-the-art SFDA techniques (Liang et al., 2020; Liu et al., 2021b), our method shows surpassing segmentation performance with increments of 5.910% and 10.042% in *Dice*. Our method outperforms other methods in all metrics except *Spe*. Due to domain gap, 'Source only' suffers from performance degradation on the target domain, and it fails to detect polyps and tends to recognize all pixels as majority class background as shown in Fig. 4(c). Since the definition of Specificity (*Spe*) is $\frac{TN}{FP+TN}$, which means the ability to recognize negative samples, it's rational that the *Spe* of 'Source only' is higher than others.

We further compared our approach with state-of-the-art UDA (Yang and Soatto, 2020; Zhang et al., 2021) and SRDA (Bateson et al., 2020; Stan and Rostami, 2021) techniques. It is evident that our method possesses the superior ability of polyp segmentation with increments of 10.460%, 1.478%, 10.937% and 14.254% in *Dice* compared with Yang and Soatto (2020); Zhang et al. (2021); Bateson et al. (2020) and Liu et al. (2021b), respectively. The significant improvements are owing to the proposed fourier style mining guided source-free domain adaptation framework leverages the well-optimized style information stored in the source model to distill useful information, which avoids the impurity information inevitably exists in all source images. On the contrary, these noisy information within source data may interfere the optimization of UDA models, and the additional information, such as source class ratio, involved in SRDA methods may be inconsistent with that of target data, leading to biased optimization. Moreover, the proposed CDD and CADC strategies advanced to facilitate the self-training also contribute to the performance gains against existing UDA and SRDA methods.
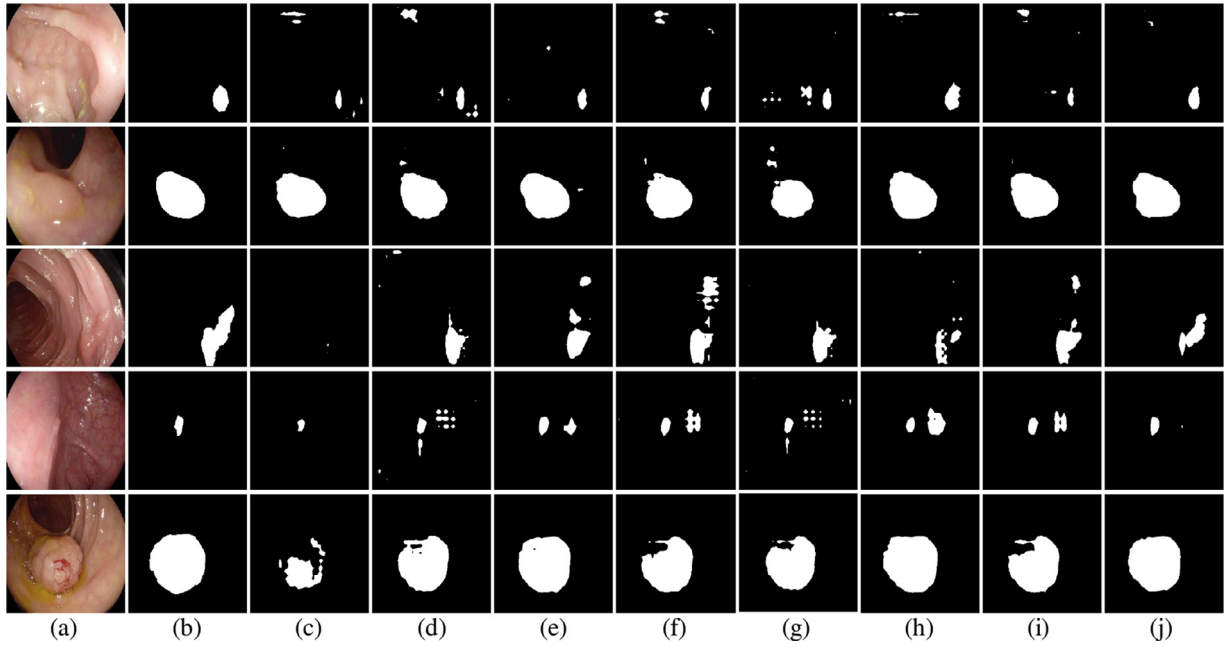
Fig. 4. Qualitative comparison of polyp segmentation results by different domain adaptation techniques. From left to right are (a) the target images, (b) ground truth, (c) results of source only method, results of state-of-the-art methods (d) Yang and Soatto (2020), (e) Zhang et al. (2021), (f) Bateson et al. (2020), (g) Stan and Rostami (2021), (h) Liang et al. (2020), (i) Liu et al. (2021b), and (j) our segmentation results.

Table 1
Comparison with state-of-the-art domain adaptation methods on the EndoScene to ETIS-Larib Adaptation. Best and second best results are highlighted and underlined.

Method	Dice (%)	Spe (%)	Sen (%)	Acc (%)	IoU_P (%)	IoU_B (%)	mIoU (%)
Source only	56.021	**98.994**	56.433	97.325	47.366	97.211	72.289
Target only	86.737	99.467	88.971	99.150	78.163	99.075	88.619
Yang and Soatto (2020)	63.663	<u>98.987</u>	68.715	97.612	52.695	97.480	75.087
Zhang et al. (2021)	<u>72.645</u>	98.884	<u>79.512</u>	<u>98.167</u>	<u>62.932</u>	<u>98.072</u>	<u>80.502</u>
Bateson et al. (2020)	63.186	98.822	69.271	97.694	53.154	97.563	75.359
Stan and Rostami (2021)	59.869	98.120	74.848	97.142	48.100	96.995	72.548
Liang et al. (2020)	68.213	98.819	77.012	98.033	58.459	97.937	78.198
Liu et al. (2021b)	64.081	98.321	76.918	97.697	54.266	97.585	75.925
Ours	**74.123**	98.854	**80.923**	**98.201**	**64.162**	**98.116**	**81.139**

Table 2
Comparison with state-of-the-art domain adaptation techniques on the NCI-ISBI 2013 to PROMISE12 Adaptation. Best and second best results are highlighted and underlined.

Method	Dice (%)	Spe (%)	Sen (%)	Acc (%)	IoU_P (%)	IoU_B (%)	mIoU (%)
Source only	61.041	**99.933**	49.207	98.276	46.466	98.246	72.356
Target only	85.140	99.510	90.548	99.215	75.445	99.192	87.319
Yang and Soatto (2020)	63.633	99.362	<u>60.790</u>	98.052	47.995	98.013	73.004
Zhang et al. (2021)	67.190	99.797	57.775	98.330	53.253	98.298	75.276
Bateson et al. (2020)	60.743	99.873	49.216	98.205	45.950	98.175	72.063
Stan and Rostami (2021)	62.238	<u>99.926</u>	50.762	98.296	48.036	98.268	73.152
Liang et al. (2020)	<u>67.210</u>	99.860	58.239	<u>98.523</u>	<u>53.289</u>	<u>98.494</u>	<u>75.892</u>
Liu et al. (2021b)	65.574	99.897	53.934	98.413	51.247	98.384	74.816
Ours	**71.090**	99.839	**61.077**	**98.595**	**57.026**	**98.566**	**77.796**

To qualitatively validate the effectiveness of the proposed method, we then visualize the segmentation predictions of the proposed method and comparison methods (Liang et al., 2020; Liu et al., 2021b; Yang and Soatto, 2020; Zhang et al., 2021; Bateson et al., 2020), as shown in Fig. 4. It is obvious that our method can generate clearer and more accurate polyp segmentation results.

4.3.2. Results on prostate segmentation

We further conducted experiments on the prostate segmentation task. As shown in Table 2, our SFDA framework attains a surpassing segmentation performance than other domain adaptation methods, with the *Dice* of 71.090% and *Sen* of 61.077%. Besides,

our approach surpasses the second best model (Li et al., 2021) with performance gain of 3.880% in *Dice* and 2.838% in *Sen*. Moreover, it is worth noting that the proposed framework achieves superior segmentation performance than state-of-the-art UDA and SRDA techniques (Yang and Soatto, 2020; Zhang et al., 2021; Bateson et al., 2020), which can access more source domain information than ours. We conjecture that it is owing to the proposed style mining strategy that makes full use of the source model to extract distribution knowledge from the source model. Furthermore, the qualitative results of the proposed approach and comparison methods (Liang et al., 2020; Liu et al., 2021b; Yang and Soatto, 2020; Zhang et al., 2021; Bateson et al., 2020) are shown in Fig. 5. It can be observed that the predictions of our method have much fewer
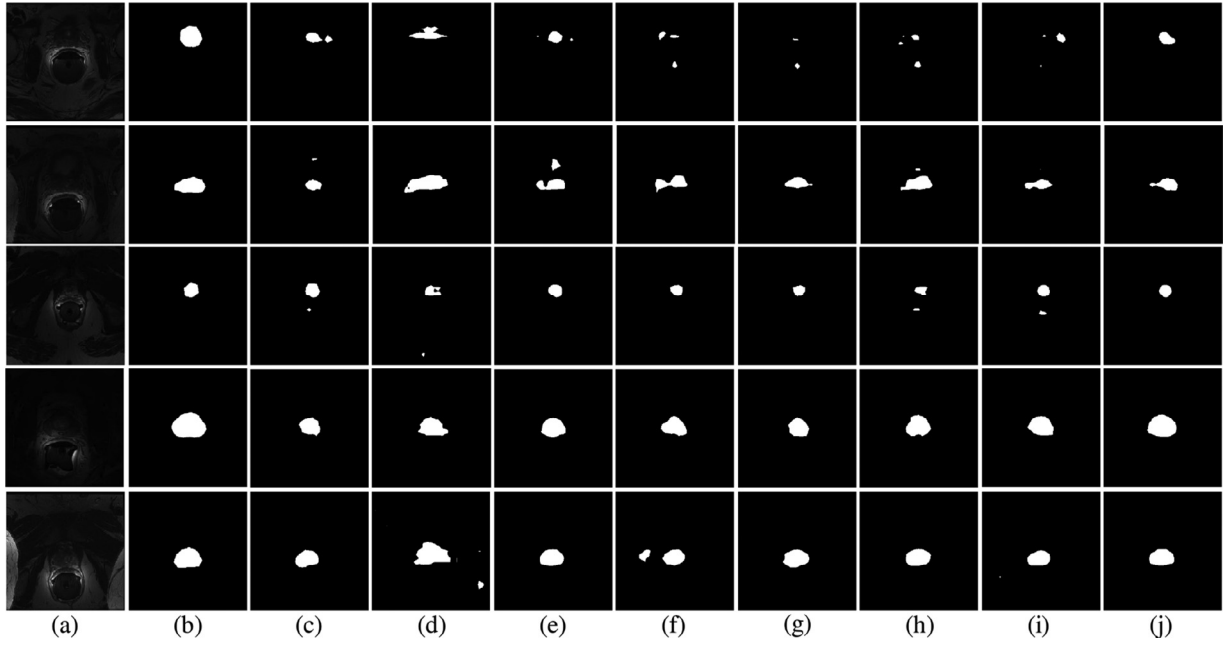
Fig. 5. Qualitative comparison of prostate segmentation results by different domain adaptation methods. From left to right are (a) the target images, (b) ground truth, (c) results of source only method, results of state-of-the-art methods (d) Yang and Soatto (2020), (e) Zhang et al. (2021), (g) Stan and Rostami (2021), (f) Bateson et al. (2020), (h) Liang et al. (2020), (i) Liu et al. (2021b), and (j) our segmentation results.

Table 3
Experimental results comparison of our method with different experimental settings on polyp segmentation.

Methods	FSM		CDD		CADC	Dice (%)	Spe (%)	Sen (%)	Acc (%)	IoU_P (%)	IoU_B (%)	mIoU (%)
Baseline	Coarse	Fine	$\mathcal{L}_{distill}$	$\mathcal{L}_{contrast}$	$\mathcal{W}_{compact}$							
✓						56.021	98.994	56.433	97.325	47.366	97.211	72.289
✓	✓					46.402	98.651	48.444	97.232	40.317	97.138	68.727
✓	✓	✓				59.796	**99.375**	58.600	98.050	52.004	97.950	74.977
✓	✓	✓	✓			70.112	99.086	73.869	98.157	60.521	98.059	79.290
✓	✓	✓		✓		69.303	98.565	80.582	97.832	58.708	97.734	78.221
✓	✓	✓	✓	✓		71.074	98.489	**82.275**	97.952	61.207	97.871	79.539
✓	✓	✓	✓	✓	✓	**74.123**	98.854	80.923	**98.201**	**64.162**	**98.116**	**81.139**

false negatives, compared with other domain adaptation methods. This may be ascribed to the effect of CADC, which can filter out noisy pseudo labels and provide more accurate supervision for the target domain.

4.4. Ablation analysis of key components

4.4.1. Effectiveness of FSM

The proposed FSM aims to generate source-like images, and it includes coarse generation with BN constraints and fine generation through mutual Fourier Transform. We comprehensively evaluated the FSM module under two experimental settings. 1) Baseline + Coarse: predicting segmentation results with the source model for coarse generated images x_c; 2) Baseline + Coarse + Fine: predicting segmentation results with the source model for fine generated images x_g. The quantitative comparison of coarse generations and fine generations is shown in 1st-3rd rows of Table 3. Evaluating on the coarse generations instead of target images leads to 9.619% degradation in Dice due to blurry coarse generations, while evaluating on the fine generations shows superior segmentation performance with increments of 3.775% in Dice than original target images. This improvement illustrates that our FSM module is able to mine out source style information from the source model.

We also visualized coarse generations and fine generations in Fig. 6 to qualitatively verify the effectiveness of FSM module. It's obvious that source and target images vary significantly in appear-

ance, which demonstrates huge style gap. By applying BN constrains to the target polyp images, we can obtain coarse generations in Fig. 6(c). The coarse generations have similar appearance with source images but suffer from blurry information. By applying Mutual Fourier Transform on coarse generations and target images, we can obtain fine generation in Fig. 6(d). We further visualized coarse generations and fine generations of two samples in Fig. 7(c) and (d). As shown in enlarged red boxes of Fig. 7(c), coarse generation may produce artifacts and blurs, which destroy image content information. By using mutual fourier transform, fine generation can restore texture details of original target images as shown in Fig. 7(d). Quantitative and qualitative experiments demonstrate that mutual fourier transform (FFT) is necessary to generate high-quality source-like images.

4.4.2. Effectiveness of CDD

The CDD module includes a domain distillation loss $\mathcal{L}_{distill}$ and a domain contrastive loss $\mathcal{L}_{contrast}$ to achieve feature-level adaptation. To investigate the benefits of different constraints in the CDD module, we conducted ablation study using polyp segmentation dataset as in Table 3. We comprehensively evaluated the CDD module under three different experimental settings: 1) Baseline + FSM + $\mathcal{L}_{distill}$: optimizing the segmentation model with only a domain distillation loss; 2) Baseline + FSM + $\mathcal{L}_{contrast}$: optimizing the segmentation model with only domain contrast loss; 3) Baseline + FSM + CDD: optimizing the segmentation model
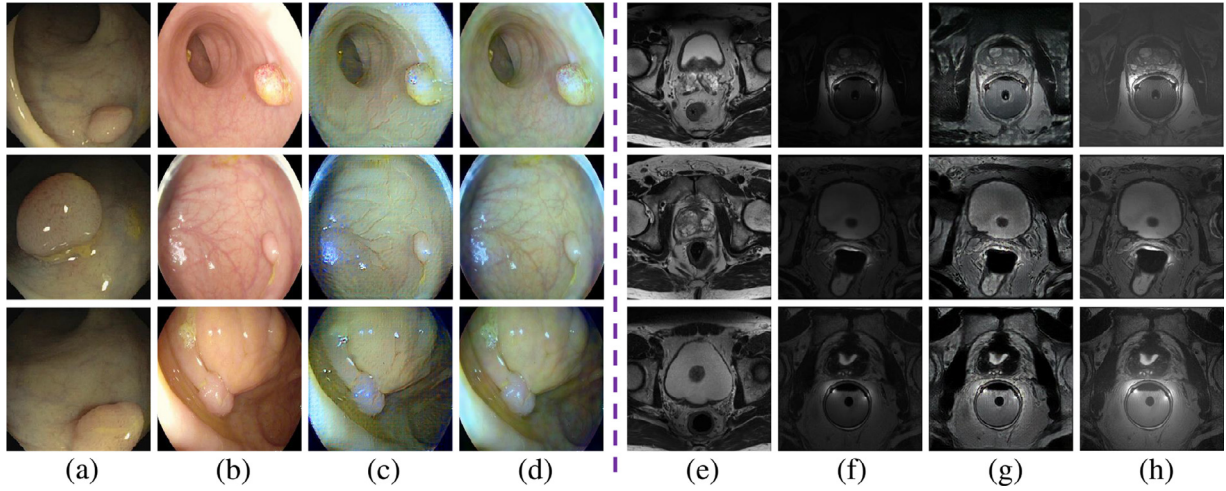
Fig. 6. Visualization of the FSM module. Polyp segmentation on the EndoScene to ETIS-Larib Adaptation: (a) source images. (b) target images. (c) coarse generations of the FSM. (d) fine generations of the FSM. Prostate Segmentation on the NCI-ISBI 2013 to PROMISE12 Adaptation: (e) source images. (f) target images. (g) coarse generations of the FSM. (h) fine generations of the FSM.
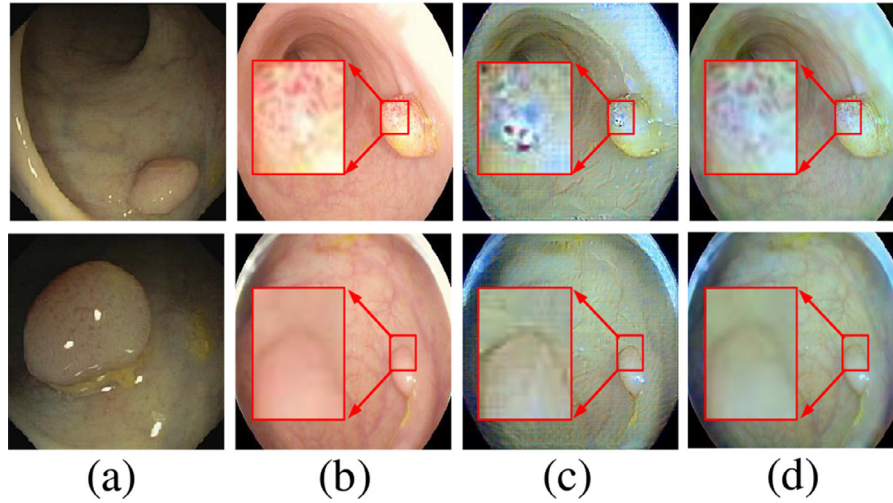

Fig. 7. Detailed visual comparison of coarse generations and fine generations. (a) source images. (b) target images. (c) coarse generations of the FSM. (d) fine generations of the FSM. Source images are inaccessible for our task. Here we show them for better visual comparison. Red boxes highlight blurs and artifacts of the coarse generations.

jointly with the domain distillation loss and the domain contrast loss. The results of $3^{rd} - 6^{th}$ row in Table 3 show that adding the domain distillation loss and the domain contrast loss lead to 10.316% and 9.507% performance gain in *Dice* than the model only applying the FSM module, demonstrating the effectiveness of these two constraints. Combining these two constraints together, we can achieve 11.278% performance gain in *Dice* than the model only applying the FSM module.

We further conducted experiments with different weights for these two introduced loss constraints to analyze the sensitivity of hyper parameters $\lambda_{distill}$ and $\lambda_{contrast}$ in the CDD module as shown in Fig. 8. Specifically, we evaluated the segmentation performance of our approach with $\lambda_{contrast}$ varies in {0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1} and $\lambda_{distill}$ ranges in {0.05, 0.1, 0.5, 1}. When $\lambda_{distill}$ and $\lambda_{contrast}$ equal 0.1 and 0.01, the proposed framework attains the best performance with 71.074% in *Dice* and 82.275% in *Sen*. Thus we choose $\lambda_{distill} = 0.1$ and $\lambda_{contrast} = 0.01$ in the proposed framework.

4.4.3. Effectiveness of CADC

The contribution of the CADC to our method is shown in $6^{th} - 7^{th}$ row of Table 3. After combining the CADC module with FSM and CDD modules, we can achieve 3.049% performance gain in

Dice. This improvement demonstrates that the shape compactness metric can largely filter out noises in the pseudo labels.

To further verify the effectiveness of the CADC module, we made comparison with three typical self-training methods: Threshold (Zou et al., 2018), MRENT(Zou et al., 2019) and Uncertainty (Zheng and Yang, 2021). For fair comparison with these approaches, FSM and CDD modules are kept same with our experimental setting. We only change the strategy to utilize pseudo labels: threshold pseudo labels or reweight them by some metrics (uncertainty or shape compactness metric). Results in Table 5 show that our self-training strategy attains the best performance with 74.123% in *Dice* and 98.201% in *Acc*. Specifically, our compact-aware strategy surpasses state-of-the-art self-training methods (Zou et al., 2018; Zheng and Yang, 2021; Zou et al., 2019) with increments of 3.049%, 1.843% and 2.500% in *Dice*. This confirms that the compact-aware mechanism can better filter out noisy pseudo labels than traditional methods. Previous self-training methods filter out noisy samples from the perspective of pixel-level information, but ignore the global information of pseudo labels. On the contrary, the proposed CADC method considers the compact properties of lesions and organs in medical imaging scenarios and introduces a shape compactness metric to reweight pseudo labels.
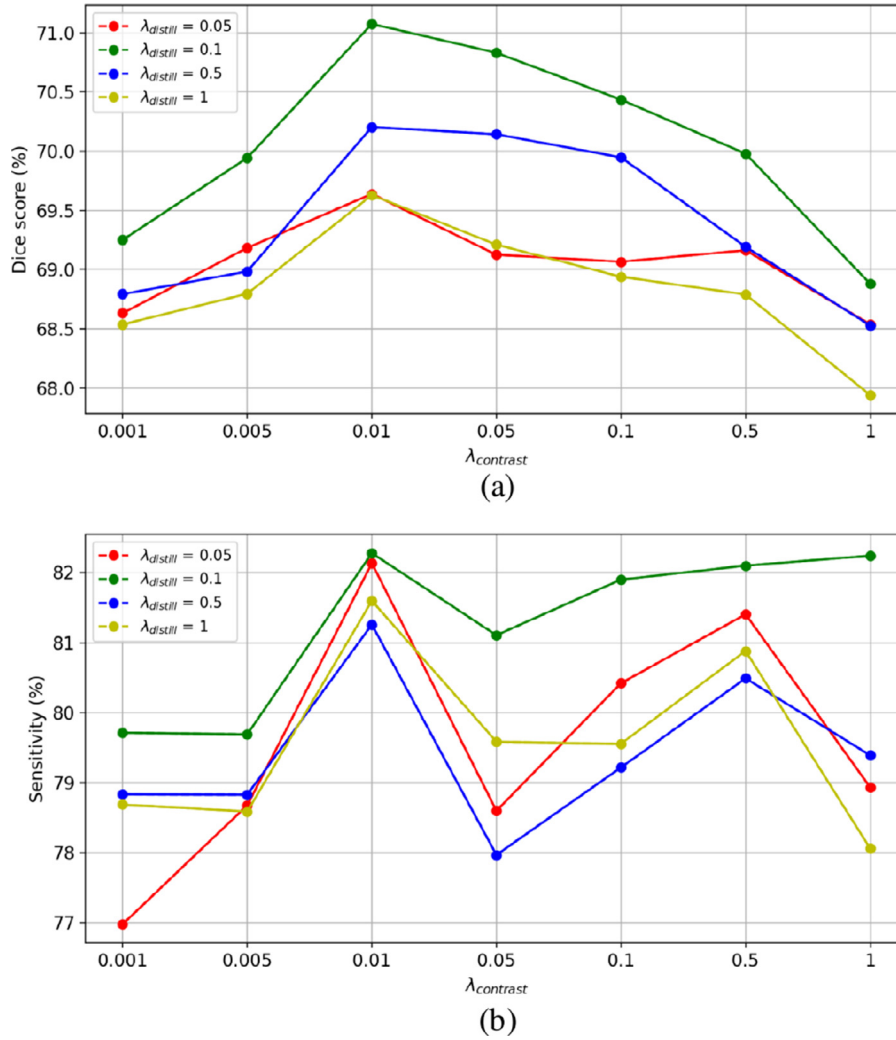
Fig. 8. Polyp segmentation performance with different weights of the CDD. (a) Dice score; (b) Sensitivity.

Table 4
Ablation study of mask size β on the EndoScene to ETIS-Larib Adaptation. Best and second best results are highlighted and underlined.

β	Dice (%)	Spe (%)	Sen (%)	Acc (%)	IoU$_P$ (%)	IoU$_B$ (%)	mIoU (%)
0	56.021	98.994	56.433	97.325	47.366	97.211	72.289
$1e-50$	58.003	99.364	57.215	97.921	50.041	97.813	73.927
0.0005	58.071	99.427	57.261	97.904	50.185	97.801	73.993
0.005	**59.796**	99.375	**58.600**	**98.050**	**52.004**	**97.950**	**74.977**
0.05	56.594	**99.502**	55.876	97.895	49.131	97.794	73.463
0.01	54.399	99.492	54.002	97.826	47.571	97.720	72.646
0.25	53.582	98.924	55.932	97.513	46.609	97.411	72.010
1	46.402	98.651	48.444	97.232	40.317	97.138	68.727

Table 5
Comparison with state-of-the-art self-training techniques on the EndoScene to ETIS-Larib Adaptation.

Method	Dice (%)	Spe (%)	Sen (%)	Acc (%)	mIoU (%)
Threshold	71.074	98.489	**82.275**	97.952	72.289
MRENT	72.280	**99.067**	76.933	98.147	80.097
Uncertainty	71.623	98.713	80.875	98.068	79.925
Ours	**74.123**	98.854	80.923	**98.201**	**81.139**

4.4.4. Ablation analysis of mask size β

To explore the style information with different mask sizes M_β, we conduct Mutual Fourier Transform between coarse generations and target images with different β values and visualize fine gen-erations as shown in Fig. 9. When mask size is large (e.g. 0.25, 0.1, 0.05), the generated images share much similar style informa-tion with the coarse generations, but suffer from visible artifacts as shown in 1st and 2nd rows of Fig. 9. With β decreasing, the fine generations gradually shift from the source style (see last row of Fig. 9). It is a trade-off between style information and image qual-ity.

To further evaluate the effect of various choices of β on the segmentation performance, we conduct ablation study of mask sizes β on the EndoScene to ETIS-Larib Adaptation as shown in Table 4. We input generated images with different mask sizes β to the source-only model and evaluate the segmentation results with seven common criteria. With β increasing from 0 to 0.005,
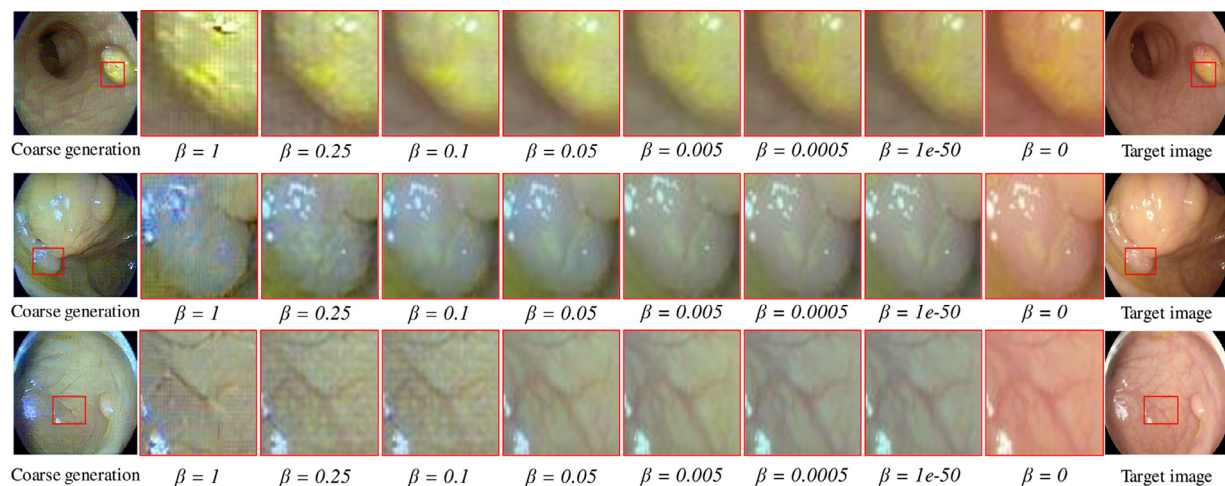
Fig. 9. Visualization of fine generations with different mask sizes β. When β equals to 0, fine generation is the same as target image. When β equals to 1, fine generation is similar to coarse generation.

the segmentation performance becomes better due to the more source style information extracted from coarse generations. However, when β becomes larger from 0.005 to 1, fine generations suffer from performance degradation due to image artifacts. Therefore, we choose β as 0.005 to preserve source style information and high image quality.

5. Conclusion

In this paper, we propose a novel SFDA framework with fourier style mining, which represents the first effort to solve SFDA for medical image segmentation. It intends to transfer knowledge of the source domain to the target domain with only a well-trained source model under the privacy concern. To obtain source domain distribution at the absence of the source data, a FSM generator is designed to generate source-like images from the source model. The FSM generator first optimizes noise image into coarse source-like images through BN constraints, and then applies mutual Fourier Transform to generate high-quality fine source-like images. With these source-like images, CDD and CADC modules are proposed to achieve feature-level and output-level distribution alignment. To filter out noisy pseudo labels, a shape compactness metric is proposed in the CADC to reweight pseudo labels. Extensive experiments were conducted on two different medical image segmentation tasks for cross-centre and cross-device adaptation, and the surpassing performance demonstrated the effectiveness of the proposed SFDA method. Our solution enables local clinical centers to access medical data distribution of other hospitals without privacy leakage, which has largely solved two major challenges in automatic medical image analysis: time-consuming annotations and privacy concerns.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests.

CRediT authorship contribution statement

Chen Yang: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – review & editing. **Xiaoqing Guo:** Conceptualization, Methodology, Writing – review & editing. **Zhen Chen:** Conceptualization, Methodology, Writing – review & editing. **Yixuan Yuan:** Conceptualization, Methodology, Writing – review & editing, Supervision, Funding acquisition.

Acknowledgments

References

Bateson, M., Kervadec, H., Dolz, J., Lombaert, H., Ayed, I.B., 2020. Source-relaxed domain adaptation for image segmentation. In: MICCAI. Springer, pp. 490–499.

Chen, C., Dou, Q., Chen, H., Qin, J., Heng, P.A., 2020. Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation. IEEE Trans. Med. Imag. 39 (7), 2494–2505. doi:10.1109/TMI.2020.2972701.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans. Pattern Anal. Mach. Intell. 40 (4), 834–848.

Chen, Y., Li, W., Chen, X., Gool, L.V., 2019. Learning semantic segmentation from synthetic data: A geometrically guided input-output adaptation approach. In: CVPR, pp. 1841–1850.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. In: CVPR. IEEE, pp. 248–255.

Fleuret, F., et al., 2021. Uncertainty reduction for model adaptation in semantic segmentation. In: CVPR, pp. 9613–9623.

Gatys, L.A., Ecker, A.S., Bethge, M., 2016. Image style transfer using convolutional neural networks. In: CVPR, pp. 2414–2423.

Ghafoorian, M., Mehrtash, A., Kapur, T., Karssemeijer, N., Marchiori, E., Pesteie, M., Guttmann, C.R., de Leeuw, F.-E., Tempany, C.M., Van Ginneken, B., et al., 2017. Transfer learning for domain adaptation in mri: Application in brain lesion segmentation. In: MICCAI. Springer, pp. 516–524.

Gou, J., Yu, B., Maybank, S.J., Tao, D., 2021. Knowledge distillation: A survey. Int. J. Comput. Vis. 1–31.

Guo, X., Yang, C., Li, B., Yuan, Y., 2021. Metacorrection: Domain-aware meta loss correction for unsupervised domain adaptation in semantic segmentation. CVPR.

Guo, X., Yang, C., Liu, Y., Yuan, Y., 2020. Learn to threshold: Thresholdnet with confidence-guided manifold mixup for polyp segmentation. IEEE Trans. Med. Imag..

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: CVPR, pp. 770–778.

Hoffman, J., Tzeng, E., Park, T., Zhu, J.-Y., Isola, P., Saenko, K., Efros, A., Darrell, T., 2018. Cycada: Cycle-consistent adversarial domain adaptation. In: ICML. PMLR, pp. 1989–1998.

Huang, X., Belongie, S., 2017. Arbitrary style transfer in real-time with adaptive instance normalization. In: CVPR, pp. 1501–1510.

Kang, G., Jiang, L., Yang, Y., Hauptmann, A.G., 2019. Contrastive adaptation network for unsupervised domain adaptation. In: CVPR, pp. 4893–4902.

Kim, M., Byun, H., 2020. Learning texture invariant representation for domain adaptation of semantic segmentation. In: CVPR, pp. 12975–12984.

Kim, Y., Hong, S., Cho, D., Park, H., Panda, P., 2020. Domain adaptation without source data. arXiv preprint arXiv:2007.01524.

Kundu, J.N., Venkat, N., Babu, R.V., et al., 2020. Universal source-free domain adaptation. In: CVPR, pp. 4544–4553.

Li, C., Wand, M., 2016. Combining markov random fields and convolutional neural networks for image synthesis. In: CVPR, pp. 2479–2486.

Li, R., Jiao, Q., Cao, W., Wong, H.-S., Wu, S., 2020. Model adaptation: Unsupervised domain adaptation without source data. In: CVPR, pp. 9641–9650.

Li, X., Chen, W., Xie, D., Yang, S., Yuan, P., Pu, S., Zhuang, Y., 2021. A free lunch for unsupervised domain adaptive object detection without source data. AAAI.

Li, Y., Wang, N., Shi, J., Hou, X., Liu, J., 2018. Adaptive batch normalization for practical domain adaptation. Pattern Recognit. 80, 109–117.

Liang, J., Hu, D., Feng, J., 2020. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In: ICML. PMLR, pp. 6028–6039.

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88.

Litjens, G., Toth, R., van de Ven, W., Hoeks, C., Kerkstra, S., van Ginneken, B., Vincent, G., Guillard, G., Birbeck, N., Zhang, J., et al., 2014. Evaluation of prostate segmentation algorithms for mri: the promise12 challenge. Med. Image Anal. 18 (2), 359–373.

Liu, X., Guo, X., Liu, Y., Yuan, Y., 2021. Consolidated domain adaptive detection and localization framework for cross-device colonoscopic images. Med. Image Anal. 71, 102052.

Liu, Y., Zhang, W., Wang, J., 2021. Source-free domain adaptation for semantic segmentation. CVPR.

Nicholas Bloch, A.M., 2015. NCI-ISBI 2013 Challenge: Automated Segmentation of Prostate Structures. The Cancer Imaging Archive doi:10.7937/K9/TCIA.2015.ZF0VLOPV.

Pan, X., Luo, P., Shi, J., Tang, X., 2018. Two at once: Enhancing learning and generalization capacities via ibn-net. In: ECCV, pp. 464–479.

Peng, B., Jin, X., Liu, J., Li, D., Wu, Y., Liu, Y., Zhou, S., Zhang, Z., 2019. Correlation congruence for knowledge distillation. In: ICCV, pp. 5007–5016.

Qiu, Z., Zhang, Y., Lin, H., Niu, S., Liu, Y., Du, Q., Tan, M., 2021. Source-free domain adaptation via avatar prototype generation and adaptation. IJCAI.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. Springer, pp. 234–241.

Silva, J., Histace, A., Romain, O., Dray, X., Granado, B., 2014. Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. Int. J. Comput. Assist. Radiol. Surg. 9 (2), 283–293.

Stan, S., Rostami, M., 2021. Privacy preserving domain adaptation for semantic segmentation of medical images. arXiv preprint arXiv:2101.00522.

Tran, L., Sohn, K., Yu, X., Liu, X., Chandraker, M., 2019. Gotta adapt'em all: Joint pixel and feature-level domain adaptation for recognition in the wild. In: CVPR, pp. 2672–2681.

Tsai, Y.-H., Hung, W.-C., Schulter, S., Sohn, K., Yang, M.-H., Chandraker, M., 2018. Learning to adapt structured output space for semantic segmentation. In: CVPR, pp. 7472–7481.

Tzeng, E., Hoffman, J., Saenko, K., Darrell, T., 2017. Adversarial discriminative domain adaptation. In: CVPR, pp. 7167–7176.

Vázquez, D., Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., López, A.M., Romero, A., Drozdzal, M., Courville, A., 2017. A benchmark for endoluminal scene segmentation of colonoscopy images. J. Healthcare Eng..

Wang, D., Shelhamer, E., Liu, S., Olshausen, B., Darrell, T., 2021. Tent: Fully test-time adaptation by entropy minimization. ICLR.

Xing, F., Bennett, T., Ghosh, D., 2019. Adversarial domain adaptation and pseudo-labeling for cross-modality microscopy image quantification. In: MICCAI. Springer, pp. 740–749.

Yang, C., Guo, X., Zhu, M., Ibragimov, B., Yuan, Y., 2021. Mutual-prototype adaptation for cross-domain polyp segmentation. IEEE J. Biomed. Health. Inf..

Yang, D., Xu, Z., Li, W., Myronenko, A., Roth, H.R., Harmon, S., Xu, S., Turkbey, B., Turkbey, E., Wang, X., et al., 2021. Federated semi-supervised learning for covid region segmentation in chest ct using multi-national data from china, italy, japan. Med. Image Anal. 70, 101992.

Yang, Y., Lao, D., Sundaramoorthi, G., Soatto, S., 2020. Phase consistent ecological domain adaptation. In: CVPR, pp. 9011–9020.

Yang, Y., Soatto, S., 2020. Fda: Fourier domain adaptation for semantic segmentation. In: CVPR, pp. 4085–4095.

Yin, H., Molchanov, P., Alvarez, J.M., Li, Z., Mallya, A., Hoiem, D., Jha, N.K., Kautz, J., 2020. Dreaming to distill: Data-free knowledge transfer via deepinversion. CVPR.

Yuan, Y., Qin, W., Guo, X., Buyyounouski, M., Hancock, S., Han, B., Xing, L., 2019. Prostate segmentation with encoder-decoder densely connected convolutional network (ed-densenet). In: ISBI. IEEE, pp. 434–437.

Zhang, P., Zhang, B., Zhang, T., Chen, D., Wang, Y., Wen, F., 2021. Prototypical pseudo label denoising and target structure learning for domain adaptive semantic segmentation. CVPR.

Zhang, Y., Chen, H., Wei, Y., Zhao, P., Cao, J., Fan, X., Lou, X., Liu, H., Hou, J., Han, X., et al., 2019. From whole slide imaging to microscopy: Deep microscopy adaptation network for histopathology cancer image classification. In: MICCAI. Springer, pp. 360–368.

Zhang, Y., Qiu, Z., Yao, T., Liu, D., Mei, T., 2018. Fully convolutional adaptation networks for semantic segmentation. In: CVPR, pp. 6810–6818.

Zheng, Z., Yang, Y., 2021. Rectifying pseudo label learning via uncertainty estimation for domain adaptive semantic segmentation. Int. J. Comput. Vis. 129 (4), 1106–1120.

Zhu, J.-Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: ICCV, pp. 2223–2232.

Zou, Y., Yu, Z., Kumar, B., Wang, J., 2018. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In: ECCV, pp. 289–305.

Zou, Y., Yu, Z., Liu, X., Kumar, B., Wang, J., 2019. Confidence regularized self-training. In: ICCV, pp. 5982–5991.