



Semi-supervised Domain Adaptive Medical Image Segmentation Through Consistency Regularized Disentangled Contrastive Learning

Hritam Basak^(✉) and Zhaozheng Yin

Department of Computer Science, Stony Brook University, Stony Brook, NY, USA
hbasak@cs.stonybrook.edu

Abstract. Although unsupervised domain adaptation (UDA) is a promising direction to alleviate domain shift, they fall short of their supervised counterparts. In this work, we investigate relatively less explored semi-supervised domain adaptation (SSDA) for medical image segmentation, where access to a few labeled target samples can improve the adaptation performance substantially. Specifically, we propose a two-stage training process. First, an encoder is pre-trained in a self-learning paradigm using a novel domain-content disentangled contrastive learning (CL) along with a pixel-level feature consistency constraint. The proposed CL enforces the encoder to learn discriminative content-specific but domain-invariant semantics on a global scale from the source and target images, whereas consistency regularization enforces the mining of local pixel-level information by maintaining spatial sensitivity. This pre-trained encoder, along with a decoder, is further fine-tuned for the downstream task, (i.e. pixel-level segmentation) using a semi-supervised setting. Furthermore, we experimentally validate that our proposed method can easily be extended for UDA settings, adding to the superiority of the proposed strategy. Upon evaluation on two domain adaptive image segmentation tasks, our proposed method outperforms the SoTA methods, both in SSDA and UDA settings. Code is available at [GitHub](#).

Keywords: Contrastive Learning · Style-content disentanglement · Consistency Regularization · Domain Adaptation · Segmentation

1 Introduction

Despite their remarkable success in numerous tasks, deep learning models trained on a source domain face the challenges to generalize to a new target domain, especially for segmentation which requires dense pixel-level prediction. This is

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43901-8_25.

attributed to a large semantic gap between these two domains. Unsupervised Domain Adaptation (UDA) has lately been investigated to bridge this semantic gap between labeled source domain, and unlabeled target domain [27], including adversarial learning for aligning latent representations [23], image translation networks [24], etc. However, these methods produce subpar performance because of the lack of supervision from the target domain and a large semantic gap in style and content information between the source and target domains. Moreover, when an image’s content-specific information is entangled with its domain-specific style information, traditional UDA approaches fail to learn the correct representation of the domain-agnostic content while being distracted by the domain-specific styles. So, they cannot be generalized for multi-domain segmentation tasks [4].

Compared to UDA, obtaining annotation for a few target samples is worthwhile if it can substantially improve the performance by providing crucial target domain knowledge. Driven by this speculation, and the recent success of semi-supervised learning (SemiSL), we investigate semi-supervised domain adaptation (SSDA) as a potential solution. Recently, Liu *et al.* [14] proposed an asymmetric co-training strategy between a SemiSL and UDA task, that complements each other for cross-domain knowledge distillation. Xia *et al.* [22] proposed a co-training strategy through pseudo-label refinement. Gu *et al.* [7] proposed a new SSDA paradigm using cross-domain contrastive learning (CL) and self-ensembling mean-teacher. However, these methods force the model to learn the low-level nuisance variability, which is insignificant to the task at hand, hence failing to generalize if similar variational semantics are absent in the training set. Fourier Domain Adaptation (FDA) [26] was proposed to address these challenges by an effective spectral transfer method. Following [26], we design a new Gaussian FDA to handle this cross-domain variability, without feature alignment.

Contrastive learning (CL) is another prospective direction where we enforce models to learn discriminative information from (dis)similarity learning in a latent subspace [2, 10]. Liu *et al.* [15] proposed a margin-preserving constraint along with a self-paced CL framework, gradually increasing the training data difficulty. Gomariz *et al.* [6] proposed a CL framework with an unconventional channel-wise aggregated projection head for inter-slice representation learning. However, traditional CL utilized for DA on images with entangled style and content leads to mixed representation learning, whereas ideally, it should learn discriminative content features invariant to style representation. Besides, the instance-level feature alignment of CL is subpar for segmentation, where dense pixel-wise predictions are indispensable [1].

To alleviate these three underlined shortcomings, we propose a novel contrastive learning with pixel-level consistency constraint via disentangling the style and content information from the joint distribution of source and target domain. Precisely, our contributions are as follows: **(1)** We propose to disentangle the style and content information in their compact embedding space using a joint-learning framework; **(2)** We propose encoder pre-training with two CL strategies: *Style CL* and *Content CL* that learns the style and content information respectively from the embedding space; **(3)** The proposed CL is complemented

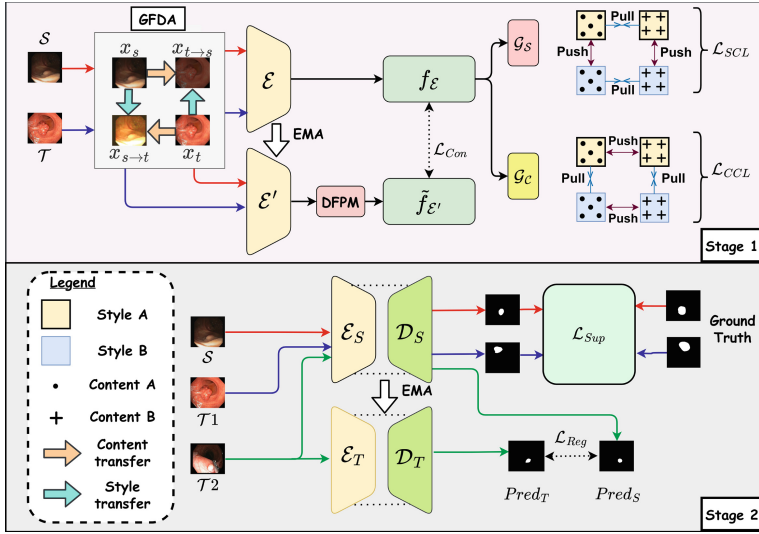


Fig. 1. Overall workflow of our proposed method. **Stage 1:** Encoder pre-training by GFDA and CL on disentangled style and content branches, and pixel-wise feature consistency module DFPM; **Stage 2:** Fine-tuning the encoder in a semi-supervised student-teacher setting.

with a pixel-level consistency constraint with dense feature propagation module, where the former provides better categorization competence whereas the later enforces effective spatial sensitivity; (4) We experimentally validate that our SSDA method can be extended in the UDA setting easily, achieving superior performance as compared to the SoTA methods on two widely-used domain adaptive segmentation tasks, both in SSDA and UDA settings.

2 Proposed Method

Given the source domain image-label pairs $\{(x_s^i, y_s^i)_{i=1}^{N_s} \in \mathcal{S}\}$, a few image-label pairs from target domain $\{(x_{t1}^i, y_{t1}^i)_{i=1}^{N_{t1}} \in \mathcal{T}1\}$, and a large number of unlabeled target images $\{(x_{t2}^i)_{i=1}^{N_{t2}} \in \mathcal{T}2\}$, our proposed pre-training stage learns from images in $\{\mathcal{S} \cup \mathcal{T}; \mathcal{T} = \mathcal{T}1 \cup \mathcal{T}2\}$ in a self-supervised way, without requiring any labels. The following fine-tuning in SSDA considers image-label pairs in $\{\mathcal{S} \cup \mathcal{T}1\}$ for supervised learning alongside unlabeled images $\mathcal{T}2$ in the target domain for unsupervised prediction consistency. Our workflow is shown in Fig. 1.

2.1 Gaussian Fourier Domain Adaptation (GFDA)

Manipulating the low-level amplitude spectrum of the frequency domain is the easiest way for style transfer between domains [26], without notable alteration

in the visuals of high-level semantics. However, as observed in [26], the generated images consist of incoherent dark patches, caused by abrupt changes in amplitude around the rectangular mask. Instead, we propose a Gaussian mask for a smoother transition in frequency. Let, $\mathcal{F}_A(\cdot)$ and $\mathcal{F}_P(\cdot)$ be the amplitude and phase spectrum in frequency space of an RGB image, and \mathcal{F}^{-1} indicates inverse Fourier transform. We define a 2D Gaussian mask g_σ of the same size as \mathcal{F}_A , with σ being the standard deviation. Given two randomly sampled images $x_s \sim \mathcal{S}$ and $x_t \sim \mathcal{T}$, our proposed GFDA can be formulated as:

$$x_{s \rightarrow t} = \mathcal{F}^{-1}[\mathcal{F}_P(x_s), \mathcal{F}_A(x_t) \odot g_\sigma + \mathcal{F}_A(x_s) \odot (1 - g_\sigma)], \quad (1)$$

where \odot indicates element-wise multiplication. It generates an image preserving the semantic content from \mathcal{S} but preserving the style from \mathcal{T} . Reciprocal pair $x_{t \rightarrow s}$ is also formulated using the same drill. The source and target images, and the style-transferred versions $\{x_s, x_{s \rightarrow t}, x_t, x_{t \rightarrow s}\}$ are then used for contrastive pre-training below. Visualization of GFDA is shown in the supplementary file.

2.2 CL on Disentangled Domain and Content

We aim to learn discriminative content-specific features that are invariant of the style of the source or target domain, for a better pre-training of the network for the task at hand. Hence, we propose to disentangle the style and content information from the images and learn them jointly in a novel disentangled CL paradigm: Style CL (*SCL*) and Content CL (*CCL*). The proposed *SCL* imposes learning of domain-specific attributes, whereas *CCL* enforces the model to identify the ROI, irrespective of the spatial semantics and appearance. In joint learning, they complement each other to render the model to learn domain-agnostic and content-specific information, thereby mitigating the domain dilemma. The set of images $\{x_s, x_{s \rightarrow t}, x_t, x_{t \rightarrow s}\}$, along with their augmented versions are passed through encoder \mathcal{E} , followed by two parallel projection heads, namely style head (\mathcal{G}_S) and content head (\mathcal{G}_C) to obtain the corresponding embeddings. Two different losses: style contrastive loss \mathcal{L}_{SCL} and content contrastive loss \mathcal{L}_{CCL} , are derived below.

Assuming $\{x_s, x_{t \rightarrow s}\}$ (along with their augmentations) having source-style representation (style *A*), and $\{x_t, x_{s \rightarrow t}\}$ (and their augmentations) having target-style representation (style *B*), in style CL, embeddings from the same domain (style) are grouped together whereas embeddings from different domains are pushed apart in the latent space. Considering the i^{th} anchor point $x_t^i \in \mathcal{T}$ in a minibatch and its corresponding style embedding $s_t^i \leftarrow \mathcal{G}_S(\mathcal{E}(x_t^i))$ (with style *B*), we define the *positive* set consisting of the same target domain representations as $\Lambda^+ = \{s_t^{j+}, s_{s \rightarrow t}^{j+}\} \leftarrow \mathcal{G}_S(\mathcal{E}(\{x_t^j, x_{s \rightarrow t}^j\})), \forall j \in \text{minibatch}$, and *negative* set having unlike source domain representation as $\Lambda^- = \{s_s^{j-}, s_{t \rightarrow s}^{j-}\} \leftarrow \mathcal{G}_S(\mathcal{E}(\{x_s^j, x_{t \rightarrow s}^j\})), \forall j \in \text{minibatch}$. Following SimCLR [5] our style contrastive loss can be formulated as:

$$\mathcal{L}_{SCL} = \sum_{i,j} -\log \frac{\exp(\text{sim}(s^i, s^{j+})/\tau)}{\exp(\text{sim}(s^i, s^{j+})/\tau) + \sum_{j \in \Lambda^-} \exp(\text{sim}(s^i, s^{j-})/\tau)}, \quad (2)$$

where $\{s^i, s^{j+}\} \in \text{style } B$; $s^{j-} \in \text{style } A$, $\text{sim}(\cdot, \cdot)$ defines cosine similarity, τ is the temperature parameter [5]. Similarly, we define \mathcal{L}_{CCL} for content head as:

$$\mathcal{L}_{CCL} = \sum_{i,j} -\log \frac{\exp(\text{sim}(c^i, c^{j+})/\tau)}{\exp(\text{sim}(c^i, c^{j+})/\tau) + \sum_{j \in A^-} \exp(\text{sim}(c^i, c^{j-})/\tau)}, \quad (3)$$

where $\{c^i, c^j\} \leftarrow \mathcal{G}_C(\mathcal{E}(\{x^i, x^j\}))$. These contrastive losses, along with the consistency constraint below enforce the encoder to extract domain-invariant and content-specific feature embeddings.

2.3 Consistency Constraint

The disentangled CL aims to learn global image-level representation, which is useful for instance discrimination tasks. However, segmentation is attributed to learning dense pixel-level representations. Hence, we propose an additional Dense Feature Propagation Module (DFPM) along with a momentum encoder \mathcal{E}' with exponential moving average (EMA) of parameters from \mathcal{E} . Given any pixel m of an image x , we transform its feature $f_{\mathcal{E}'}^m$ obtained from \mathcal{E}' by propagating other pixel features from the same image:

$$\tilde{f}_{\mathcal{E}'}^m = \sum_{\forall n \in x} \mathcal{K}(f_{\mathcal{E}'}^m) \otimes \cos(f_{\mathcal{E}'}^m, f_{\mathcal{E}'}^n) \quad (4)$$

where \mathcal{K} is a linear transformation layer, \otimes denotes *matmul* operation. This spatial smoothing of learned representation is useful for structural sensitivity, which is fundamental for dense segmentation tasks. We enforce consistency between this smoothed feature $\tilde{f}_{\mathcal{E}'}$ from \mathcal{E}' and the regular feature $f_{\mathcal{E}}$ from \mathcal{E} as:

$$\mathcal{L}_{Con} = \sum_{[d(m,n) < Th]} -[\cos(\tilde{f}_{\mathcal{E}'}^m, f_{\mathcal{E}}^n) + \cos(f_{\mathcal{E}}^m, \tilde{f}_{\mathcal{E}'}^n)] \quad (5)$$

where $d(\cdot, \cdot)$ indicates the spatial distance, Th is a threshold. The overall pre-training objective can be summarized as:

$$\mathcal{L}_{Pre} = \lambda_1 \mathcal{L}_{SCL} + \lambda_2 \mathcal{L}_{CCL} + \mathcal{L}_{Con} \quad (6)$$

2.4 Semi-supervised Fine-Tuning

The pre-training stage is followed by semi-supervised fine-tuning using a student-teacher framework [18]. The pre-trained encoder \mathcal{E} , along with a decoder \mathcal{D} are used as a student branch, whereas an identical encoder-decoder network (but differently initialized) is used as a teacher network. We compute a supervised loss on the labeled set $\{\mathcal{S} \cup \mathcal{T}1\}$ along with a regularization loss between the prediction of the student and teacher branches on the unlabeled set $\{\mathcal{T}2\}$ as:

$$\mathcal{L}_{Sup} = \frac{1}{\mathbb{N}_s + \mathbb{N}_{t1}} \sum_{x^i \in \{\mathcal{S} \cup \mathcal{T}1\}} CE[\mathcal{D}_S(\mathcal{E}_S(x^i)), y^i] \quad (7)$$

$$\mathcal{L}_{Reg} = \frac{1}{\mathbb{N}_{t2}} \sum_{x^i \in \{T2\}} CE [\mathcal{D}_S (\mathcal{E}_S(x^i)), \mathcal{D}_T (\mathcal{E}_T(x^i))] \quad (8)$$

where CE indicates cross-entropy loss, \mathcal{E}_S , \mathcal{D}_S , \mathcal{E}_T , \mathcal{D}_T indicate the student and teacher encoder and decoder networks. The student branch is updated using a consolidated loss $\mathcal{L} = \mathcal{L}_{Sup} + \lambda_3 \mathcal{L}_{Reg}$, whereas the teacher parameters (θ_T) are updated using EMA from the student parameters (θ_S):

$$\theta_T(t) = \alpha \theta_T(t-1) + (1-\alpha) \theta_S(t) \quad (9)$$

where t tracks the step number, and α is the momentum coefficient [9].

In summary, the overall SSDA training process contains pre-training (Subsect. 2.1–Subsect. 2.3) and fine-tuning (Subsect. 2.4), whereas, we only use the student branch ($\mathcal{E}_S, \mathcal{D}_S$) for inference.

3 Experiments and Results

Datasets: We evaluate our work on two different DA tasks to evaluate its generalizability: (1) Polyp segmentation from colonoscopy images in Kvasir-SEG [11] and CVC-EndoScene Still [20], and (2) Brain tumor segmentation in MRI images from BraTS2018 [16]. Kvasir and CVC contain 1000 and 912 images respectively and were split into 4 : 1 training-testing sets following [10]. BraTS consists of brain MRIs from 285 patients with T1, T2, T1CE, and FLAIR scans. The data was split into 4 : 1 train-test ratio, following [14]. **Source→Target:** We perform experiments on $CVC \rightarrow Kvasir$ and $Kvasir \rightarrow CVC$ for polyp segmentation, and $T2 \rightarrow \{T1, T1CE, FLAIR\}$ for tumor segmentation. The SSDA accesses 10 – 50% and 1 – 5 labels from the target domain for the two tasks, respectively. For UDA, only \mathcal{S} is used for \mathcal{L}_{Sup} , whereas $\mathcal{T}1 \cup \mathcal{T}2$ is used for \mathcal{L}_{Reg} . **Implementation details:** Implementation is done in a PyTorch environment using a Tesla V100 GPU with 32GB RAM. We use U-Net [17] backbone for the encoder-decoder structure, and the projection heads \mathcal{G}_S and \mathcal{G}_C are shallow FC layers. The model is trained for 300 epochs for pre-training and 500 epochs for fine-tuning using an ADAM optimizer with a batch size of 4 and a learning rate of $1e-4$. $\lambda_1, \lambda_2, \lambda_3$, and Th are set to 0.75, 0.75, 0.5, 0.6, respectively by validation, τ, α are set to 0.07, 0.999 following [9]. Augmentations include random rotation and translation. **Metrics:** Segmentation performance is evaluated using Dice Similarity Score (DSC) and Hausdorff Distance (HD).

3.1 Performance on SSDA

Quantitative comparison of our proposed method with different SSDA methods [4, 14, 21, 24] for both tasks are shown in Table 1 and Table 2. ACT [14] simply ignores the domain gap and only learns content semantics, resulting in substandard performance on the BraTS dataset that has a significant domain gap. FSM [24], on the other hand, is adaptable to learning explicit domain information, but

Table 1. Comparison with state-of-the-art UDA and SSDA methods for polyp segmentation on KVASIR and CVC. SSDA results are shown for 10%-labeled (10%L) and 50%-labeled (50%L) data in the target domain. The results of cited methods are directly reported from the corresponding papers. **No DA**: the encoder-decoder model trained only using labeled data from the source domain is applied to the target domain without adaptation. **Supervised**: model is trained using all labeled data from source and target domains. The best and second-best results are highlighted in **RED** and **BLUE**, respectively.

Task	Method	Target label	CVC \rightarrow Kvasir		Kvasir \rightarrow CVC	
			DSC \uparrow	HD \downarrow	DSC \uparrow	HD \downarrow
No DA	Source only	0%L	62.2	5.6	53.9	6.2
UDA	PCEDA [25]	0%L	73.6	4.4	70.1	4.7
	ASN [19]	0%L	80.1	3.6	83.7	3.7
	BDL [12]	0%L	77.8	4.0	81.7	4.1
	CoFo [10]	0%L	82.8	3.6	81.1	3.5
	FDA [26]	0%L	80.4	3.9	75.1	4.2
	Ours	0%L	83.8	3.4	84.5	3.1
SSDA	DLD [21]	10%L	84.2	3.2	85.1	3.1
	ACT [14]	10%L	86.9	3.0	87.3	2.9
	SLA [4]	10%L	85.5	3.1	86.2	3.3
	FSM [24]	10%L	85.8	3.4	86.2	3.1
	Ours	10%L	87.7	2.9	86.9	2.7
	DLD [21]	50%L	87.6	2.8	87.9	2.6
	ACT [14]	50%L	89.4	2.6	90.3	2.4
	SLA [4]	50%L	88.6	2.7	89.3	2.8
	FSM [24]	50%L	89.1	2.6	89.8	2.5
	Ours	50%L	90.6	2.4	90.8	2.2
Supervised	Source+Target	100%L	92.1	2.1	93.8	2.0

lacks strong pixel-level regularization on its prediction, resulting in subpar performance. We address both of these shortcomings in our work, resulting in superior performance on both tasks. Other methods like [4, 21], which are originally designed for natural images, lack critical refining abilities even after fine-tuning for medical image segmentation and hence are far behind our performance in both tasks. The margins are even higher for less labeled data (1L) on the BraTS dataset, which is promising considering the difficulty of the task. Moreover, our method produces performance close to its fully-supervised counterpart (last row in Table 1 and Table 2), using only a few target labels.

3.2 Performance on UDA

Unlike SSDA methods, UDA fully relies on unlabeled data for domain-invariant representation learning. To analyze the effectiveness of DA, we extend our model to the UDA setting (explained in Sect. 3 [Source \rightarrow Target]) and compare it with SoTA methods [3, 8, 10, 12, 13, 25, 26, 28] in Table 1 and Table 2. Methods like [10, 19] rely on adversarial learning for aligning multi-level feature space, which is not effective for small-sized medical data. Other methods [12, 25] rely on an image-translation network but fail in effective style adaptation, resulting

Table 2. Comparison with state-of-the-art UDA and SSDA methods for whole tumor segmentation on BraTS2018, where source domain is T2. SSDA results are demonstrated for 1-labeled (1L) and 5-labeled (5L) data in the target domain.

Task	Method	Target Label	DSC \uparrow			HD \downarrow		
			T1	T1CE	FLAIR	T1	T1CE	FLAIR
No DA	Source only	0L	3.9	6.0	64.4	56.9	50.8	30.4
UDA	SSCA [13]	0L	59.3	63.5	82.9	12.5	11.2	7.9
	SIFA [3]	0L	51.7	58.2	68.0	19.6	15.0	16.9
	DSA [8]	0L	57.7	62.0	81.8	14.2	13.7	8.6
	DSFN [28]	0L	57.3	62.2	78.9	17.5	15.5	13.8
	Ours	0L	60.7	64.4	83.3	11.1	10.9	7.3
SSDA	DLD [21]	1L	65.8	66.5	81.5	12.0	10.3	7.1
	ACT [14]	1L	69.7	69.7	84.5	10.5	10.0	5.8
	ACT-EMD [14]	1L	67.4	69.0	83.9	10.9	10.3	6.4
	SLA [4]	1L	64.7	66.1	82.3	12.2	10.5	7.1
	Ours	1L	72.2	71.9	85.8	10.0	9.5	5.2
	DLD [21]	5L	67.8	68.3	83.3	11.2	9.9	6.6
	ACT [14]	5L	71.3	70.8	85.0	10.0	9.8	5.2
	ACT-EMD [14]	5L	70.3	69.8	84.4	10.4	10.2	5.7
	SLA [4]	5L	67.2	71.2	83.1	11.7	10.1	6.8
	Ours	5L	73.1	72.4	86.1	9.7	9.3	4.8
Supervised	Source+Target	all labeled	73.6	72.9	86.6	9.5	9.1	4.6

Table 3. Ablation experiment for polyp segmentation in SSDA(50%L) setting to identify the contribution of individual components. TCL: traditional CL [2], SCL: proposed style CL, CCL: proposed content CL. The last row, highlighted in **RED**, indicates our results.

Experiment#	Stage 1				Stage 2	CVC \rightarrow Kvasir		Kvasir \rightarrow CVC	
	TCL	SCL	CCL	DFPM	SemiSL	DSC \uparrow	HD \downarrow	DSC \uparrow	HD \downarrow
(a)	✓	×	×	×	✓	81.7	4.4	82.1	4.2
(b)	×	✓	×	×	✓	83.2	3.9	84.7	3.5
(c)	×	×	✓	×	✓	84.5	3.8	85.4	3.1
(d)	×	✓	✓	×	✓	89.5	2.8	89.1	2.4
(e)	×	✓	✓	✓	✓	90.6	2.4	90.8	2.2

in source domain-biased subpar performance. Our method, although relies on FDA [26], outperforms it with a large margin of upto 12.5% DSC for polyp segmentation, owing to its superior learning ability of disentangled style and content semantics. Similar results are observed for the BraTS dataset in Table 2, where our work achieved a margin of upto 2.4% DSC than its closest performer.

3.3 Ablation Experiments

We perform a detailed ablation experiment, as shown in Table 3. The effectiveness of disentangling and joint-learning of style and content information is evident from the experiment (b)&(c) as compared to (a), where the introduction of *SCL* and *CCL* boosts overall performance significantly. Moreover, when combined together (experiment (d)), they provide a massive 9.54% and

8.52% DSC gain over traditional CL (experiment (a)) for $CVC \rightarrow Kvasir$ and $Kvasir \rightarrow CVC$, respectively. This also points out a potential shortfall of traditional CL: its inability to adapt to a complex domain in DA. The proposed DFPM (experiment (e)) provides local pixel-level regularization, complementary to the global disentangled CL, resulting in a further boost in performance ($\sim 1.5\%$). We have similar ablation study observations on the BraTS2018 dataset, which is provided in the supplementary file, along with some qualitative examples along with available ground truth.

4 Conclusion

We propose a novel style-content disentangled contrastive learning, guided by a pixel-level feature consistency constraint for semi-supervised domain adaptive medical image segmentation. To the best of our knowledge, this is the first attempt for SSDA in medical image segmentation using CL, which is further extended to the UDA setting. Our proposed work, upon evaluation on two different domain adaptive segmentation tasks in SSDA and UDA settings, outperforms the existing SoTA methods, justifying its effectiveness and generalizability.

References

1. Basak, H., Chattopadhyay, S., Kundu, R., Nag, S., Mallipeddi, R.: Ideal: improved dense local contrastive learning for semi-supervised medical image segmentation. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2023)
2. Chaitanya, K., Erdil, E., Karani, N., Konukoglu, E.: Contrastive learning of global and local features for medical image segmentation with limited annotations. In: Advances in Neural Information Processing Systems, vol. 33, pp. 12546–12558 (2020)
3. Chen, C., Dou, Q., Chen, H., Qin, J., Heng, P.A.: Synergistic image and feature adaptation: towards cross-modality domain adaptation for medical image segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 865–872 (2019)
4. Chen, S., Jia, X., He, J., Shi, Y., Liu, J.: Semi-supervised domain adaptation based on dual-level domain mixing for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11018–11027 (2021)
5. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International Conference on Machine Learning, pp. 1597–1607. PMLR (2020)
6. Gomariz, A., et al.: Unsupervised domain adaptation with contrastive learning for OCT segmentation. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) Medical Image Computing and Computer Assisted Intervention-MICCAI 2022: 25th International Conference, Part VIII, pp. 351–361. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16452-1_34
7. Gu, R., et al.: Contrastive semi-supervised learning for domain adaptive segmentation across similar anatomical structures. IEEE Trans. Med. Imaging **42**(1), 245–256 (2022)

8. Han, X., et al.: Deep symmetric adaptation network for cross-modality medical image segmentation. *IEEE Trans. Med. Imaging* **41**(1), 121–132 (2022)
9. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9729–9738 (2020)
10. Huy, T.D., Huyen, H.C., Nguyen, C.D., Duong, S.T., Bui, T., Truong, S.Q.: Adversarial contrastive Fourier domain adaptation for polyp segmentation. In: *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pp. 1–5. IEEE (2022)
11. Jha, D., et al.: Kvasir-SEG: a segmented polyp dataset. In: Ro, Y.M., et al. (eds.) *MMM 2020. LNCS*, vol. 11962, pp. 451–462. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-37734-2_37
12. Li, Y., Yuan, L., Vasconcelos, N.: Bidirectional learning for domain adaptation of semantic segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6936–6945 (2020)
13. Liu, X., Xing, F., El Fakhri, G., Woo, J.: Self-semantic contour adaptation for cross modality brain tumor segmentation. In: *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pp. 1–5. IEEE (2022)
14. Liu, X., et al.: ACT: semi-supervised domain-adaptive medical image segmentation with asymmetric co-training. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) *Medical Image Computing and Computer Assisted Intervention-MICCAI 2022: 25th International Conference, Proceedings, Part V*, pp. 66–76. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16443-9_7
15. Liu, Z., Zhu, Z., Zheng, S., Liu, Y., Zhou, J., Zhao, Y.: Margin preserving self-paced contrastive learning towards domain adaptation for medical image segmentation. *IEEE J. Biomed. Health Inform.* **26**(2), 638–647 (2022)
16. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* **34**(10), 1993–2024 (2014)
17. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015. LNCS*, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
18. Tarvainen, A., Valpola, H.: Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results. In: *Advances in Neural Information Processing Systems*, vol. 30 (2017)
19. Tsai, Y.H., Hung, W.C., Schuster, S., Sohn, K., Yang, M.H., Chandraker, M.: Learning to adapt structured output space for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7472–7481 (2019)
20. Vázquez, D., et al.: A benchmark for endoluminal scene segmentation of colonoscopy images. *J. Healthcare Eng.* **2017** (2017)
21. Wang, Z., et al.: Alleviating semantic-level shift: a semi-supervised domain adaptation method for semantic segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 936–937 (2020)
22. Xia, Y., et al.: Uncertainty-aware multi-view co-training for semi-supervised medical image segmentation and domain adaptation. *Med. Image Anal.* **65**, 101766 (2020)

23. Xing, F., Cornish, T.C.: Low-resource adversarial domain adaptation for cross-modality nucleus detection. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) *Medical Image Computing and Computer Assisted Intervention-MICCAI 2022: 25th International Conference, Proceedings, Part VII*, pp. 639–649. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16449-1_61
24. Yang, C., Guo, X., Chen, Z., Yuan, Y.: Source free domain adaptation for medical image segmentation with Fourier style mining. *Med. Image Anal.* **79**, 102457 (2022)
25. Yang, Y., Lao, D., Sundaramoorthi, G., Soatto, S.: Phase consistent ecological domain adaptation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9011–9020 (2020)
26. Yang, Y., Soatto, S.: FDA: Fourier domain adaptation for semantic segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4085–4095 (2020)
27. Yao, K., et al.: A novel 3D unsupervised domain adaptation framework for cross-modality medical image segmentation. *IEEE J. Biomed. Health Inform.* **26**(10), 4976–4986 (2022)
28. Zou, D., Zhu, Q., Yan, P.: Unsupervised domain adaptation with dual-scheme fusion network for medical image segmentation. In: *IJCAI*, pp. 3291–3298 (2022)