Masked Frequency Consistency for Domain-Adaptive Semantic Segmentation of Laparoscopic Images

Xinkai Zhao[1(✉)], Yuichiro Hayashi[1], Masahiro Oda[1,2], Takayuki Kitasaka[3], and Kensaku Mori[1,4,5(✉)]

[1] Graduate School of Informatics, Nagoya University, Nagoya, Japan
xkzhao@mori.m.is.nagoya-u.ac.jp, kensaku@is.nagoya-u.ac.jp
[2] Information Strategy Office, Information and Communications, Nagoya University, Nagoya, Japan
[3] Department of Information Science, Aichi Institute of Technology, Toyota, Japan
[4] Information Technology Center, Nagoya University, Nagoya, Japan
[5] Research Center for Medical Bigdata, National Institute of Informatics, Tokyo, Japan

Abstract. Semantic segmentation of laparoscopic images is an important issue for intraoperative guidance in laparoscopic surgery. However, acquiring and annotating laparoscopic datasets is labor-intensive, which limits the research on this topic. In this paper, we tackle the Domain-Adaptive Semantic Segmentation (DASS) task, which aims to train a segmentation network using only computer-generated simulated images and unlabeled real images. To bridge the large domain gap between generated and real images, we propose a Masked Frequency Consistency (MFC) module that encourages the network to learn frequency-related information of the target domain as additional cues for robust recognition. Specifically, MFC randomly masks some high-frequency information of the image to improve the consistency of the network's predictions for low-frequency images and real images. We conduct extensive experiments on existing DASS frameworks with our MFC module and show performance improvements. Our approach achieves comparable results to fully supervised learning method on the CholecSeg8K dataset without using any manual annotation. The code is available at github.com/MoriLabNU/MFC.

Keywords: Unsupervised Domain Adaptation · Laparoscopic Image · Semantic Segmentation

1 Introduction

Laparoscopic surgery is a minimally invasive surgical technique in which a camera and surgical instruments are inserted through a series of small skin punctures.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43907-0_63.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14220, pp. 663–673, 2023.
https://doi.org/10.1007/978-3-031-43907-0_63

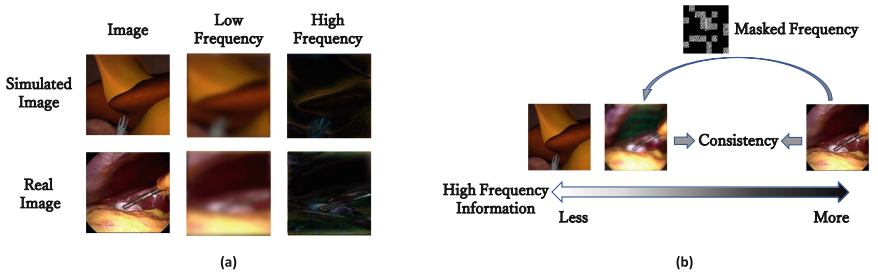

Fig. 1. Motivation for this work. (a) real images tend to contain more high frequency information than simulated images, and (b) our method aims to mitigate domain gaps by minimizing the discrepancy between high and low frequency information.

During this procedure, the surgeon relies heavily on a screen to visualize the surgical site, which can be a serious challenge. An inaccurate interpretation of abdominal anatomy can result in serious injury to the patient's bile ducts [25]. Therefore, deploying neural networks to accurately identify anatomical structures during laparoscopic surgery can markedly enhance both the quality and safety of the procedure [15]. Despite the marked achievements of deep neural networks in various medical computer vision tasks, the training of supervised models necessitates a substantial volume of precisely annotated images. Because the acquisition of large, high-quality datasets of laparoscopic images is labor-intensive and requires expert knowledge, the size and quality of publicly available datasets limit current research on semantic segmentation of laparoscopic images [21]. To overcome these limitations, several active learning [1,18] and domain generalization [14] methods have been developed to minimize the manual annotation required for network training. We take a step further and employ unsupervised domain adaptation (UDA) to eliminate the dependence on manual annotation.

The aim of UDA is to train a model on a labeled source and an unlabeled target domain for enhanced target domain performance. Various UDA methods exist, but we concentrate on two types that are relevant to our approach: self-training-based and Fourier transform-based. Self-training-based approaches [2,23,29] apply different data augmentations, multiple models or domain mixtures to the images and gauge the consistency regularization between them. On the other hand, Fourier transform-based approaches [12,27,28] exchange the low frequency components across domains to transform source domain images into target domain ones. While UDA has been extensively investigated in the medical field, the majority of existing research has focused only on domain migration between datasets from different sources [10,11] or segmentation of some distinct categories (e.g. instrument [11,19,20]). In contrast, we utilize computer-simulated images and unlabeled laparoscopic images to train a semantic segmentation network for laparoscopic images, which is more demanding and practical, as it deals with a more severe domain shift.

We propose a novel module for the UDA task in laparoscopic semantic segmentation, aiming to promote the network's exploration of consistency regularization between high-frequency and low-frequency images. Our approach is motivated by the observation in Fig. 1(a) that computer-generated images lack the high-frequency details present in real images. For example, the computer-generated abdominal wall appears smooth, whereas the real abdominal wall has rich textural information. Inspired by the effectiveness of masked image models [5,9], we randomly mask the high frequency information of the image in the frequency domain and train the network to predict the semantic segmentation result of the image lacking high frequency information. We use the pseudo-label generated by the exponential moving average (EMA) teacher network for supervision. During training, the masking of high frequency regions allows the network to discover a shared latent space for both high and low frequency images. Consequently, real laparoscopic images with high-frequency information and computer-generated images with low-frequency information can share the same feature space, facilitating the transfer of knowledge learned in the generated images to the real images, as illustrated in Fig. 1(b).

This paper's primary contributions are as follows: (1) We creatively address the domain-adaptive semantic segmentation task for laparoscopic images, which involves training the model with both unlabeled real images and computer-generated simulated images. (2) To bridge the severe domain shift between generated and real images, we propose a novel masking frequency consistency (MFC) module to reduce the domain gap. MFC encourages the network to learn shared features between high-frequency and low-frequency images. To our knowledge, MFC is the first UDA method that applies masking strategy on frequency domain. (3) We collect a vast number of image frames from public datasets and train them with computer-generated images. We evaluate our method, demonstrating that our MFC approach not only outperforms existing state-of-the-art UDA methods but also achieves performance comparable to fully supervised methods without the need for any manual annotation.

2 Method

This paper addresses the task of domain-adaptive semantic segmentation of laparoscopic images. Suppose we have a source domain of computer-generated simulated laparoscopic images, consisting of N images $\mathcal{X}^S = \{\boldsymbol{X}_i^S \mid i = 1, 2, \ldots, N\}$ with corresponding pixel-level annotations $\mathcal{Y}^S = \{\boldsymbol{Y}_i^S \mid i = 1, 2, \ldots, N\}$, and a target domain of M real laparoscopic images $\mathcal{X}^T = \{\boldsymbol{X}_j^T \mid j = 1, 2, \ldots, M\}$ without annotations. Our objective is to train a network f with robust semantic segmentation capability on the unlabeled target domain. To achieve this, we introduce a masked frequency consistency module for self-learning on the unlabeled target domain images $\boldsymbol{X}_i^T$, while supervised loss is

used for training on the labeled source domain images $\boldsymbol{X}_i^S$. Our approach can integrate with different networks, effectively bridging the domain gap that occurs when applying networks to the target domain.

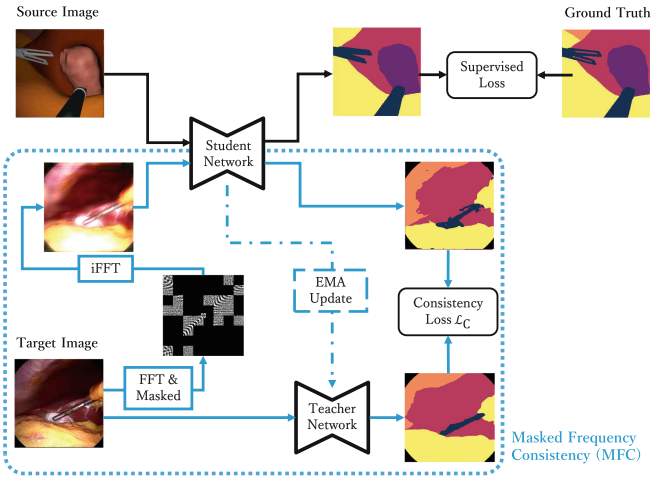

Fig. 2. The overview of our proposed Masked Frequency Consistency (MFC) module, which can be seamlessly integrated with different UDA methods and backbone networks. The MFC module works by augmenting the input image in the frequency domain using a mask, and then using a teacher-student structure to take both the original and the augmented image as inputs. A consistency loss is applied to facilitate the bridging of domains.

2.1 Image Frequency Representation

Considering an RGB image, $\boldsymbol{X} \in \mathbb{R}^{H \times W \times 3}$, we can generate its frequency representation map by applying the 2D Discrete Fourier Transform $\mathcal{F}$ for each channel $c \in \{0, 1, 2\}$, independently:

$$\mathcal{F}\left(\boldsymbol{X}\right)_{(u,v,c)} = \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} \boldsymbol{X}_{(h,w,c)} e^{-i2\pi\left(\frac{uh}{H} + \frac{vw}{W}\right)}, \text{with } i^2 = -1 \tag{1}$$

where (u, v) and (h, w) donate the coordinates in frequency map and image.

To facilitate subsequent operations, we rearrange the FFT data so that negative frequency terms precede positive ones, thereby centering the low frequency information. Furthermore, the inverse Fourier transform (iFFT) $\mathcal{F}^{-1}$ is utilized to transform the spectral signals back into the original image space:

$$\boldsymbol{X}_{(h,w,c)} = \frac{1}{HW} \sum_{u=0}^{H-1} \sum_{v=0}^{W-1} \mathcal{F}(\boldsymbol{X})_{(u,v,c)} e^{i2\pi\left(\frac{uh}{H} + \frac{vw}{W}\right)}, \text{with } i^2 = -1 \tag{2}$$

Computation of both the Fourier transform and its corresponding inverse is achieved through the Fast Fourier Transform (FFT) algorithm [16].

2.2 Masking Strategy

As illustrated in Fig. 2, MFC module perturbs the frequency information for target domain images. To do this, we define a mask $\mathcal{M} \in \{0,1\}^{H \times W}$ that randomly erases parts of the frequency map, thereby reducing the frequency data. Specifically, a patch mask $\mathcal{M}$ is randomly sampled as follows:

$$\mathcal{M}_{\substack{mb+1:(m+1)b, \\ nb+1:(n+1)b}} = [v > r], \text{with } v \sim \mathcal{U}(0,1) \tag{3}$$

where $[\cdot]$ is the Iverson bracket, $\mathcal{U}(0,1)$ the uniform distribution, b is the patch size, r represents the mask ratio, m and n are the patch indices. After this procedure, the patches in the mask are randomly masked.

However, using the random patch mask alone may result in the loss of all low frequency information, which would exacerbate the domain gap and lead to training instability. To avoid this, we set the central elements to 1, thus preserving the low frequency information from the images as:

$$\mathcal{M}_{\substack{H/2-h:H/2+h, \\ W/2-w:W/2+w}} = 1, \tag{4}$$

where h and w denote the size of the low-frequency information to be preserved.

We utilize the mask $\mathcal{M}$ to apply masking in the frequency domain and use the iFFT $\mathcal{F}^{-1}$ to transform the image back into the original spatial domain as the network input. The enhanced image can then be obtained as:

$$\boldsymbol{X}_m = \mathcal{F}^{-1}\left(\mathcal{F}\left(\boldsymbol{X}\right) \odot \mathcal{M}\right), \tag{5}$$

where $\odot$ is the Hadamard product between the matrices. Moreover, with conjugate symmetry's disruption inhibiting the imaginary component's cancellation, we employ complex number magnitudes as outputs.

2.3 Consistency Regularization

Consistency regularization is employed to extract common representations between high and low frequency images, thereby enhancing the generality of the network. Specifically, during training, the student segmentation network f_S takes the enhanced image $\boldsymbol{X}_m$ as input, whereas the original image $\boldsymbol{X}$ serves as the input for the teacher network f_T. The weight $\boldsymbol{\theta}_T$ of the teacher network undergoes updates using the exponential moving average (EMA) [22] of the weight $\boldsymbol{\theta}_S$ belonging to the student network:

$$\boldsymbol{\theta}_T = \alpha \boldsymbol{\theta}_S + (1-\alpha)\boldsymbol{\theta}'_T\ , \tag{6}$$

where $\boldsymbol{\theta}'_T$ representing the weight from the previous training step. The EMA teacher generates a series of stable pseudo-labels over time, a tactic frequently utilized in both semi-supervised learning [22] and UDA [7–9,23].

To evaluate the prediction results, we employ the mean squared error (MSE) as as our loss function, which quantifies the divergence between the predictions:

$$\mathcal{L}_{\mathrm{C}} = \boldsymbol{q}_T \odot MSE(f_T(\boldsymbol{X}), f_S(\boldsymbol{X}_m)), \tag{7}$$

where $\boldsymbol{q}_T$ denotes the quality weight. Due to potential inaccuracies in pseudo-labeling, like prior works [7,8,23], we only use confident pixels surpassing the maximum probability threshold τ, defined as:

$$\boldsymbol{q}_T = \frac{\sum_{h=0}^{H-1}\sum_{w=0}^{W-1}[\max_c f_T(\boldsymbol{X})_{(hwc)} > \tau]}{H \cdot W}. \tag{8}$$

3 Experiments

3.1 Datasets and Implementation

Datasets. Our experiments were conducted on three laparoscopic datasets, described as follows: (1) Simulation dataset [17] consists of 20,000 labeled images of 3D scenes assembled from CT data, with 6 categories. In addition, I2I [17] used generative adversarial networks (GAN) to translate these images into five different styles of realistic laparoscopic images, resulting in a total of 100,000 images. We used these two types of images (simulated and translated) as the source domain datasets with annotations. (2) CholecSeg8k [6] is a semantic segmentation dataset containing laparoscopic cholecystectomy images from 17 video clips of the Cholec80 dataset [24], labeled with 13 categories. We used these 17 video clips as a test set. (3) The Cholec80 dataset [24] consists of 80 videos of cholecystectomy procedures with annotations for phase and instrument presence, but no annotations related to segmentation. To train the UDA model, we selected 6819 images from 63 surgical videos, excluding the 17 videos used in CholecSeg8k [6]. Specifically, for the video of the preparation phase of surgery, we extracted one frame per second as an unlabeled target domain image. A more detailed description is available in supplementary material.

Dataset Partitioning. Our experiments were performed with two different settings: (1) simulated images to real images and (2) translated images to real images. Considering the 6 categories present in the simulated dataset and the 13 categories in the CholecSeg8k dataset, we performed semantic segmentation on the following 6 categories: Background (BG), Abdominal Wall (AW), Liver, Fat, Gallbladder (GB), and Instruments (INST).

Implementation. We used the mmsegmentation [4] codebase and trained each model on a single NVIDIA Tesla V100 GPU. We evaluated the Segformer [26] and DeepLabV2 [3] backbone networks, based on HRDA strategy [8], and initialized all backbone networks with pre-training on ImageNet. Training was performed using AdamW [13], with hyper-parameters taken from previous works [8,9,26]. In all experiments, we trained the models on a batch of randomly cropped $256px \times 256px$ images for 40k iterations, and the batch size is set to 4. As indicated in the ablation studies presented in the supplementary material, the optimal hyper-parameters for the MFC method vary according to the type of input datasets. Therefore, in all subsequent experiments, we set r to 0.7, b to 32, and h and w to 8, without further optimization of these hyper-parameters.

State-of-the-Art Methods. We benchmarked our method against contemporary leading approaches, which include UDA methods (DAFormer [7], HRDA [8], and MIC [9]), image translation method [17], and fully supervised network on the CholecSeg8k dataset using cross-validation. For equitable comparison, all methods employed the same segmentation network, initialization, and optimizer.

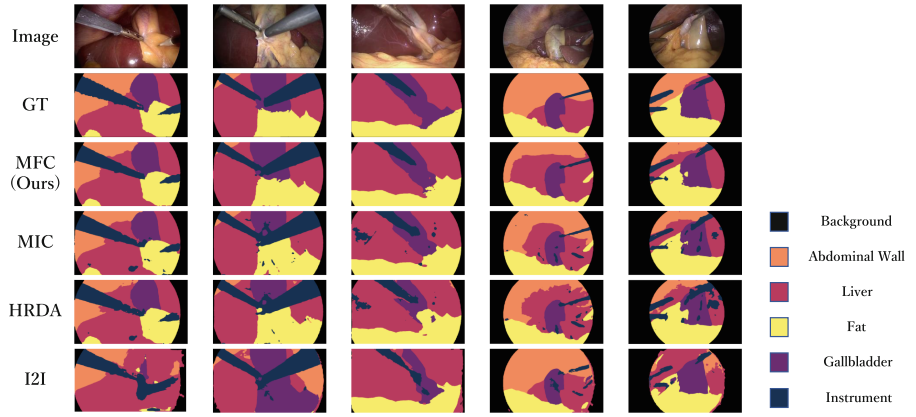

Fig. 3. Qualitative results of different methods applied to the CholecSeg8k [6] dataset.

3.2 Qualitative Evaluation

In Fig. 3, we present visualizations of the proposed MFC method and other compared methods on the CholecSeg8k dataset. All methods used only the simulated dataset for training, without any manual annotation. Our proposed method has two major advantages: (1) it effectively performs surgical instrument segmentation with minimal interference from reflections and shadows, and (2) it achieves a more accurate distinction between the boundaries of gallbladder and fat. However, we have found that our method exhibits imprecision in distinguishing the liver from the abdominal wall in certain cases.

Table 1. The results, alongside a comparison with other SOTA UDA methods utilizing simulated images as the source domain. The highest scores are emphasized in bold black text.

Network	UDA Method	BG	AW	Liver	Fat	GB	INST	mIoU
SegFormer [26]	Baseline	60.78	1.38	36.56	38.86	69.55	28.9	39.34
	Supervised	98.01	81.44	84.24	84.63	74.78	79.62	83.78
	I2I [17]	79.11	46.40	57.31	59.55	51.62	52.92	57.82
	DAFormer [7]	96.55	25.35	55.45	78.59	52.83	45.27	59.01
	HRDA [8]	97.80	44.04	63.23	78.90	55.39	55.28	65.77
	MIC [9]	98.19	**56.54**	**66.56**	85.09	54.49	59.26	70.02
	MFC(Ours)	**98.30**	43.48	60.40	**85.87**	**62.21**	**82.12**	**72.06**
	MFC w/o Eq.(3)	97.83	38.02	58.29	83.48	54.90	72.81	67.55
	MFC w/o Eq.(4)	98.11	34.35	56.53	85.81	61.10	83.37	69.88

Table 2. The results, alongside a comparison with other SOTA UDA methods utilizing translated images as the source domain. The highest scores are emphasized in bold black text.

Network	UDA Method	BG	AW	Liver	Fat	GB	INST	mIoU
DeepLabV2 [3]	Baseline [17]	62.25	38.20	52.29	57.48	51.08	43.61	50.82
	Supervised	97.04	72.45	85.64	68.79	58.72	82.42	77.51
	HRDA [8]	96.71	76.12	75.70	74.99	65.97	69.78	76.54
	MIC [9]	**97.63**	78.09	78.06	**75.36**	**67.91**	66.18	77.21
	MFC(Ours)	97.02	**80.46**	**78.51**	75.13	67.11	**70.98**	**78.20**
SegFormer [26]	Baseline [17]	79.11	46.40	57.31	59.55	51.62	52.92	57.82
	Supervised	98.01	81.44	84.24	84.63	74.78	79.62	83.78
	HRDA [8]	97.96	80.49	82.59	79.25	71.99	64.29	79.43
	MIC [9]	**98.66**	84.47	84.85	78.08	**75.94**	69.54	81.92
	MFC(Ours)	98.32	**86.77**	**87.21**	**80.10**	72.74	**70.71**	**82.64**

3.3 Quantitative Evaluation

Simulated Images → Real Images. For quantitative evaluation, we employed the intersection over union (IoU) and its overall mean of 6 categories. A performance comparison of our method with other SOTA method is presented in Table 1. MFC, trained on the simulated data as the source domain, outperform the existing SOTA methods in all categories except for the abdominal wall and liver. Notably, our method significantly improves surgical instrument segmentation. This improvement can be attributed to the fact that our method excludes the disturbing high frequency noise such as reflections and shadows, which are absent in the source domain dataset. Such results indicate that our approach effectively bridges the domain gap between the generated and real images by

randomly masking high frequency information. Furthermore, DeepLabV2-based methods underperform SegFormer-based methods in this setting.

To verify the effectiveness of the patch mask outlined in Eq. (3) and the low frequency mask in Eq. (4), we also conduct ablation experiments. The results of variants of our method are presented in Table 1. It is observed that the adoption of either of the two masking strategies enhances the segmentation performance.

Translated Images → Real Images. Furthermore, we assessed the performance of various methods using translated images as the source domain, with results summarized in Table 2. These translated images reduced the domain disparity with real images, thereby boosting the performance of UDA methods on the target domain. To gauge the efficacy of our proposed module, we incorporated MFC into two different network backbones. The results show that our approach efficiently improves the mIoU performance of segmentation across different source domain datasets and network backbones.

4 Conclusion

This paper tackles the crucial issue of laparoscopic image segmentation, which is essential for surgical guidance and navigation. We propose a novel UDA module, called MFC, that leverages the consistency between high and low-frequency information in latent space. This consistency facilitates knowledge transfer from computer-simulated to real laparoscopic datasets for segmentation. Experimentally, MFC not only bolsters existing UDA models' performance but also outperforms leading methods, including fully supervised models that rely on annotated data. Our work unveils the potential of using computer-generated image data and UDA techniques for laparoscopic image segmentation. However, a limitation of our approach is that it does not account for the long-tail category distribution prevalent in real-world scenarios, such as venous vessels. Therefore, a future direction of our research is to extend our MFC module to handle rare category segmentation, thereby improving UDA models' generalization capabilities.

Acknowledgments. This work was supported in part by the JSPS KAKENHI Grant Numbers 17H00867, 21K19898, 26108006; in part by the JST CREST Grant Number JPMJCR20D5; and in part by the fellowship of the Nagoya University TMI WISE program from MEXT.

References

1. Aklilu, J., Yeung, S.: ALGES: active learning with gradient embeddings for semantic segmentation of laparoscopic surgical images. In: Proceedings of Machine Learning for Healthcare, pp. 892–911. PMLR (2022)
2. Araslanov, N., Roth, S.: Self-supervised augmentation consistency for adapting semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15384–15394. IEEE (2021)

3. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell. **40**(04), 834–848 (2018)
4. Contributors, M.: MMSegmentation: openmmlab semantic segmentation toolbox and benchmark (2020). https://github.com/open-mmlab/mmsegmentation
5. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked autoencoders are scalable vision learners. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 16000–16009. IEEE (2022)
6. Hong, W.Y., Kao, C.L., Kuo, Y.H., Wang, J.R., Chang, W.L., Shih, C.S.: CholecSeg8k: a semantic segmentation dataset for laparoscopic cholecystectomy based on Cholec80. arXiv preprint arXiv:2012.12453 (2020)
7. Hoyer, L., Dai, D., Van Gool, L.: DaFormer: improving network architectures and training strategies for domain-adaptive semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9924–9935. IEEE (2022)
8. Hoyer, L., Dai, D., Van Gool, L.: HRDA: context-aware high-resolution domain-adaptive semantic segmentation. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds.) ECCV 2022. LNCS, vol. 13690, pp. 372–391. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-20056-4_22
9. Hoyer, L., Dai, D., Wang, H., Van Gool, L.: MIC: masked image consistency for context-enhanced domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11721–11732. IEEE (2023)
10. Hu, S., Liao, Z., Xia, Y.: Domain specific convolution and high frequency reconstruction based unsupervised domain adaptation for medical image segmentation. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13437, pp. 650–659. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16449-1_62
11. Liu, J., Guo, X., Yuan, Y.: Prototypical interaction graph for unsupervised domain adaptation in surgical instrument segmentation. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12903, pp. 272–281. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87199-4_26
12. Liu, Q., Chen, C., Qin, J., Dou, Q., Heng, P.A.: FedDG: federated domain generalization on medical image segmentation via episodic learning in continuous frequency space. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1013–1023. IEEE (2021)
13. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
14. Lyu, J., Zhang, Y., Huang, Y., Lin, L., Cheng, P., Tang, X.: AADG: automatic augmentation for domain generalization on retinal image segmentation. IEEE Trans. Med. Imaging **41**(12), 3699–3711 (2022)
15. Madani, A., et al.: Artificial intelligence for intraoperative guidance: using semantic segmentation to identify surgical anatomy during laparoscopic cholecystectomy. Ann. Surg. **276**(2), 363–369 (2022)
16. Nussbaumer, H.J., Nussbaumer, H.J.: The fast fourier transform. Fast Fourier Transform and Convolution Algorithms, pp. 80–111 (1982)
17. Pfeiffer, M., et al.: Generating large labeled data sets for laparoscopic image processing tasks using unpaired image-to-image translation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11768, pp. 119–127. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32254-0_14

18. Qiu, J., Hayashi, Y., Oda, M., Kitasaka, T., Mori, K.: Class-wise confidence-aware active learning for laparoscopic images segmentation. Inter. J. Comput. Assisted Radiol. Surgery, 1–10 (2022)
19. Sahu, M., Mukhopadhyay, A., Zachow, S.: Simulation-to-real domain adaptation with teacher-student learning for endoscopic instrument segmentation. Int. J. Comput. Assist. Radiol. Surg. **16**(5), 849–859 (2021)
20. Sahu, M., Strömsdörfer, R., Mukhopadhyay, A., Zachow, S.: Endo-Sim2Real: consistency learning-based domain adaptation for instrument segmentation. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12263, pp. 784–794. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59716-0_75
21. Silva, B., et al.: Analysis of current deep learning networks for semantic segmentation of anatomical structures in laparoscopic surgery. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 3502–3505. IEEE (2022)
22. 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 30 (2017)
23. Tranheden, W., Olsson, V., Pinto, J., Svensson, L.: DACS: domain adaptation via cross-domain mixed sampling. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 1379–1389 (2021)
24. Twinanda, A.P., Shehata, S., Mutter, D., Marescaux, J., De Mathelin, M., Padoy, N.: Endonet: a deep architecture for recognition tasks on laparoscopic videos. IEEE Trans. Med. Imaging **36**(1), 86–97 (2016)
25. Way, L.W., et al.: Causes and prevention of laparoscopic bile duct injuries: analysis of 252 cases from a human factors and cognitive psychology perspective. Ann. Surg. **237**(4), 460–469 (2003)
26. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: SegFormer: simple and efficient design for semantic segmentation with transformers. Adv. Neural. Inf. Process. Syst. **34**, 12077–12090 (2021)
27. 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. IEEE (2020)
28. Zakazov, I., Shaposhnikov, V., Bespalov, I., Dylov, D.V.: Feather-light fourier domain adaptation in magnetic resonance imaging. In: Kamnitsas, K., et al. (eds.) DART 2022. LNCS, vol. 13542, pp. 88–97. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16852-9_9
29. Zhou, Q., et al.: Context-aware mixup for domain adaptive semantic segmentation. IEEE Trans. Circuits Syst. Video Technol. **33**, 804–817 (2021)