Domain Adaptation for Medical Image Segmentation Using Transformation-Invariant Self-training

Negin Ghamsarian[1(✉)], Javier Gamazo Tejero[1], Pablo Márquez-Neila[1], Sebastian Wolf[2], Martin Zinkernagel[2], Klaus Schoeffmann[3], and Raphael Sznitman[1]

[1] Center for AI in Medicine, Faculty of Medicine, University of Bern, Bern, Switzerland
negin.ghamsarian@unibe.ch
[2] Department of Ophthalmology, Inselspital, Bern, Switzerland
[3] Department of Information Technology, Klagenfurt University, Klagenfurt, Austria

Abstract. Models capable of leveraging unlabelled data are crucial in overcoming large distribution gaps between the acquired datasets across different imaging devices and configurations. In this regard, self-training techniques based on pseudo-labeling have been shown to be highly effective for semi-supervised domain adaptation. However, the unreliability of pseudo labels can hinder the capability of self-training techniques to induce abstract representation from the unlabeled target dataset, especially in the case of large distribution gaps. Since the neural network performance should be invariant to image transformations, we look to this fact to identify uncertain pseudo labels. Indeed, we argue that transformation invariant detections can provide more reasonable approximations of ground truth. Accordingly, we propose a semi-supervised learning strategy for domain adaptation termed transformation-invariant self-training (TI-ST). The proposed method assesses pixel-wise pseudo-labels' reliability and filters out unreliable detections during self-training. We perform comprehensive evaluations for domain adaptation using three different modalities of medical images, two different network architectures, and several alternative state-of-the-art domain adaptation methods. Experimental results confirm the superiority of our proposed method in mitigating the lack of target domain annotation and boosting segmentation performance in the target domain.

Keywords: Semi-Supervised Learning · Domain Adaptation · Semantic Segmentation · Self Training · Cataract Surgery · MRI · OCT

This work was funded by Haag-Streit Switzerland.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43907-0_32.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14220, pp. 331–341, 2023.
https://doi.org/10.1007/978-3-031-43907-0_32

1 Introduction

Semantic segmentation is a prerequisite for a broad range of medical imaging applications, including disease diagnosis and treatment [13], surgical workflow analysis [6,9], operation room planning, and surgical outcome prediction [7]. While supervised deep learning approaches have yielded satisfactory performance in semantic segmentation [8,10], their performance is heavily limited by the labeled training dataset distribution. Indeed, a network trained on a dataset acquired with a specific device or configuration can dramatically underperform when evaluated on a different device or conditions. Overcoming this entails new annotations per device, a demand that is hard to meet, especially for semantic segmentation, and even more so in the medical domain, where expert knowledge is essential.

Driven by the need to overcome this challenge, numerous semi-supervised learning paradigms have looked to alleviate annotation requirements in the target domain. Semi-supervised learning refers to methods that encourage learning abstract representations from an unlabeled dataset and extending the decision boundaries towards a more-generalized or target dataset distribution. These techniques can be categorized into (i) consistency regularization [4,15–17,19,22], (ii) contrastive learning [2,11], (iii) adversarial learning [22], and (iv) self-training [24–26]. Consistency regularization techniques aim to inject knowledge via penalizing inconsistencies for identical images that have undergone different distortions, such as transformations or dropouts, or fed into networks with different initializations [4]. Specifically, the Π model [15] penalizes differences between the predictions of two transformed versions of each input image to reinforce consistent and augmentation-invariant predictions. Temporal ensembling [15] is designed to alleviate the negative effect of noisy predictions by integrating predictions of consecutive training iterations. Cross-pseudo supervision regularizes the networks by enforcing similar predictions from differently initialized networks.

More recent deep self-training approaches based on pseudo labels have emerged as promising techniques for unsupervised domain adaptation. These techniques assume that a trained network can approximate the ground-truth labels for unlabeled images. Since no metric guarantees pseudo-label reliability, several methods have been developed to alleviate pseudo-label error back-propagation. To progressively improve pseudo-labeling performance, reciprocal learning [25] adopts a teacher-student framework where the student network performance on the source domain drives the teacher network weights updates. ST++ [24] proposes to evaluate the reliability of image-based pseudo labels based on the consistency of predictions in different network checkpoints. Subsequently, half of the more reliable images are utilized to re-train the network in the first step, and the trained network is used for pseudo-labeling the whole dataset for a second re-training step. Despite the effectiveness of state-of-the-art pseudo-labeling strategies, we argue that one important aspect has been underexplored: how can a trained network self-assess the reliability of its pixel-level predictions?

To this end, we propose a novel self-training framework with a self-assessment strategy for pseudo-label reliability. The proposed framework uses transformation-invariant highly-confident predictions in the target dataset for

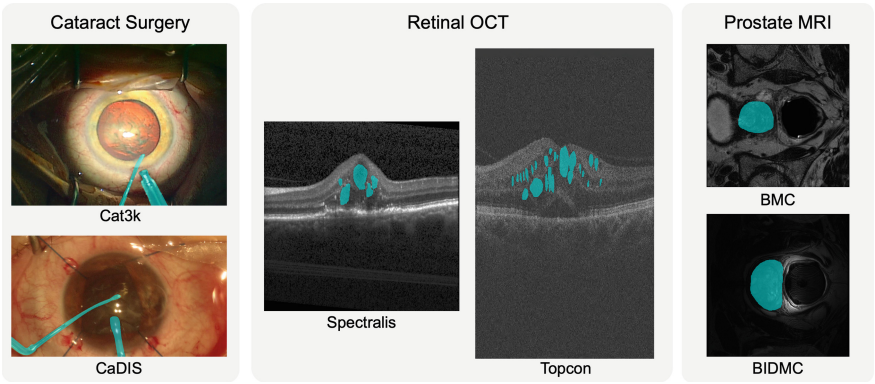

Fig. 1. Example images from the three adopted datasets: (1) cross-device-and-center instrument segmentation in cataract surgery videos (Cat101 vs. CaDIS), cross-device fluid segmentation in OCT (Spectralis vs. Topcon), and cross-institution prostate segmentation in MRI (BMC vs. BIDMC).

self-training. This objective is achieved by considering an ensemble of high-confidence predictions from transformed versions of identical inputs. To validate the effectiveness of our proposed framework on a variety of tasks, we evaluate our approach on three different semantic segmentation imaging modalities, including video (cataract surgery), optical coherence tomography (retina), and MRI (prostate), as shown in Fig. 1. We perform comprehensive experiments to validate the performance of the proposed framework, namely "Transformation-Invariant Self-Training"[1] (TI-ST). The experimental results indicate that TI-ST significantly improves segmentation performance for unlabeled target datasets compared to numerous state-of-the-art alternatives.

2 Methodology

Consider a labeled source dataset, $\mathcal{S}$, with training images $\mathcal{X}_\mathcal{S}$ and corresponding segmentation labels $\mathcal{Y}_\mathcal{S}$, while we denote a target dataset $\mathcal{T}$, containing only target images $\mathcal{X}_\mathcal{T}$. We aim to train a network using $\mathcal{X}_\mathcal{S}$, $\mathcal{Y}_\mathcal{S}$, and $\mathcal{X}_\mathcal{T}$ for semantic segmentation in the target dataset.

We propose to train the model using a self-supervised approach on the images $\mathcal{X}_\mathcal{T}$ by assigning pseudo labels during training. Typical pseudo labels are computed from independent predictions of unlabeled images. Instead, our proposed framework adopts a self-assessment strategy to determine the reliability of predictions in an unsupervised fashion. Specifically, we propose to target highly-reliable predictions generated by a network aiming for transformation-invariant confidence. Compared to self-ensembling strategies that penalize the distant predictions corresponding to the transformed versions of identical inputs, our goal is to filter out transformation-variant predictions. Indeed, our method reinforces

[1] The PyTorch implementation of TI-ST is publicly available at https://github.com/Negin-Ghamsarian/Transformation-Invariant-Self-Training-MICCAI23.

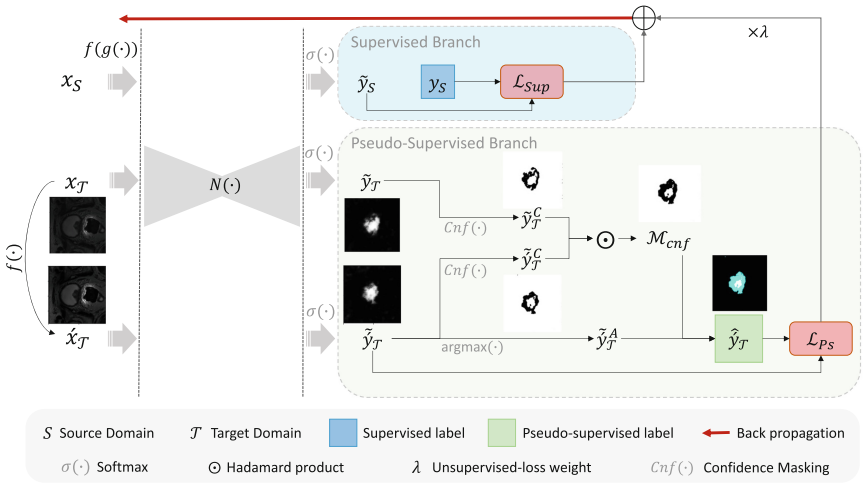

Fig. 2. Overview of the proposed semi-supervised domain adaptation framework based on transformation-invariant self-training (TI-ST). Ignored pseudo-labels during unsupervised loss computation are shown in turquoise.

the ensemble of high-confidence predictions from two versions of the same target sample. Our proposed TI-ST framework simultaneously trains on the source and target domains, so as to progressively bridge the intra-domain distribution gap. Figure 2 depicts our TI-ST framework, which we detail in the following sections.

2.1 Model

At training time, images from the source dataset are augmented using spatial $g(\cdot)$ and non-spatial $f(\cdot)$ transformations and passed through a segmentation network, $N(\cdot)$, by which the network is trained using a standard supervision loss. At the same time, images from the target dataset are also passed to the network. Specifically, we feed two versions of each target image to the network: (1) the original target image $x_\mathcal{T}$, and (2) its non-spatially transformed version, $\acute{x_\mathcal{T}} = f(x_\mathcal{T})$. Once fed through the network, the corresponding predictions can be defined as $\tilde{y_\mathcal{T}} = \sigma(N(x_\mathcal{T}))$ and $\tilde{\acute{y}_\mathcal{T}} = \sigma(N(\acute{x_\mathcal{T}}))$, where $\sigma(\cdot)$ is the Softmax operation. We then define a confidence-mask ensemble as

$$\mathcal{M}_{cnf} = Cnf(\tilde{y_\mathcal{T}}) \odot Cnf(\tilde{\acute{y}_\mathcal{T}}), \tag{1}$$

where $\odot$ refers to Hadamard product used for element-wise multiplication, and Cnf is the high confidence masking function,

$$Cnf_{\in (W \times H)}(y) = \begin{cases} 1, & \text{if } \max_C(y) > \tau \\ 0, & \text{else.} \end{cases} \tag{2}$$

where $\tau \in (0.5, 1)$ is the confidence threshold, and H, W, and C are the height, width, and number of classes in the output, respectively. Specifically, $\mathcal{M}_{cnf}$

encodes regions of confident predictions that are invariant to transformations. We can then compute the pseudo-ground-truth mask for each input from the target dataset as

$$\hat{y_{\mathcal{T}}} = \begin{cases} \text{argmax}_{\text{C}}(\tilde{y_{\mathcal{T}}}), & \text{if } \mathcal{M}_{cnf} = 1 \\ \text{ignore}, & \text{else.} \end{cases} \tag{3}$$

2.2 Training

To train our model, we simultaneously consider both the source and target samples by minimizing the following loss,

$$\mathcal{L}_{overall} = \mathcal{L}_{Sup}(\tilde{y_{\mathcal{S}}}, y_{\mathcal{S}}) + \lambda\Big(\mathcal{L}_{Ps}(\tilde{y_{\mathcal{T}}}, \hat{y_{\mathcal{T}}})\Big), \tag{4}$$

where $\mathcal{L}_{Sup}$ and $\mathcal{L}_{Ps}$ indicate the supervised and pseudo-supervised loss functions used, respectively. We set λ as a time-dependent weighing function that gradually increases the share of pseudo-supervised loss. Intuitively, our pseudo-supervised loss enforces predictions on transformation-invariant highly-confident regions for unlabeled images.

Discussion: The quantity and distribution of supervised data are determining factors in neural networks' performance. With highly distributed large-scale supervisory data, neural networks converge to an optimal state efficiently. However, when only limited supervisory data with heterogeneous distribution from the inference dataset are available, using more sophisticated methods to leverage a priori knowledge is essential. Our proposed use of invariance of network predictions with respect to data augmentation is a strong form of knowledge that can be learned through dataset-dependent augmentations. The trained network is then expected to provide consistent predictions under diverse transformations. Hence, the transformation variance of the network predictions can indicate the network's prediction doubt and low confidence correspondingly. We take advantage of this characteristic to assess the reliability of predictions and filter out unreliable pseudo-labels.

3 Experimental Setup

Datasets: We validate our approach on three cross-device/site datasets for three different modalities:

- **Cataract:** instrument segmentation in cataract surgery videos [12,21]. We set the "Cat101" [21] as the source dataset and the "CaDIS" as the target domain dataset [12].
- **OCT:** IRF Fluid segmentation in retinal OCTs [1]. We use the high-quality "Spectralis" dataset as the source and the lower-quality "Topcon" dataset as the target domain.

– **MRI:** multi-site prostate segmentation [18]. We sample volumes from "BMC" and "BIDMC" as the source and target domain, respectively.

We follow a four-fold validation strategy for all three cases and report the average results over all folds. The average number of labeled training images (from the source domain), unlabeled training images (from the target domain), and test images per fold are equal to $(207, 3189, 58)$ for Cataract, $(391, 569, 115)$ for OCT, and $(273, 195, 65)$ for MRI dataset.

Baseline Methods: We compare the performance of our proposed transformation-invariant self-training (SI-ST) method against seven state-of-the-art semi-supervised learning methods: Π models [15], temporal ensembling [15], mean teacher [19], cross pseudo supervision (CSP) [4], reciprocal learning (RL) [25], self-training (ST) [24], and mutual correction framework (MCF) [23].

Networks and Training Settings: We evaluate our TI-ST framework using two different architectures: (1) DeepLabV3+ [3] with ResNet50 backbone [14] and (2) scSE [20] with VGG16 backbone. Both backbones are initialized with the ImageNet [5] pre-trained parameters. We use a batch size of four for the Cataract and MRI datasets and a batch size of two for the OCT dataset. For all training strategies, we set the number of epochs to 100. The initial learning rate is set to 0.001 and decayed by a factor of $\gamma = 0.8$ every two epochs. The input size of the networks is 512×512 for cataract and OCT and 384×384 for the MRI dataset. As spatial transformations $g(\cdot)$, we apply cropping and random rotation (up to 30 degrees). The non-spatial transformations, $f(\cdot)$, include color jittering (brightness = 0.7, contrast = 0.7, saturation = 0.7), Gaussian blurring, and random sharpening. The confidence threshold τ for the self-training framework and the proposed TI-ST framework is set to 0.85 except in the ablation studies (See the next section). In Eq. (4), the weighting function λ ramps up from the first epoch along a Gaussian curve equal to $\exp[-5(1 - \text{current-epoch}/\text{total-epochs})]$. The self-supervised loss is set to the cross-entropy loss, and the supervised loss is set to the *cross entropy log dice* loss, which is a weighted sum of cross-entropy and the logarithm of soft dice coefficient. For the TI-ST framework, we only use non-spatial transformations for the self-training branch for simplicity.

4 Results

Table 1 compares the performance of our transformation-invariant self-training (TI-ST) approach with alternative methods across three tasks and using two network architectures. According to the quantitative results, TI-ST, RL, ST, and CPS are the best-performing methods. Nevertheless, our proposed TI-ST achieves the highest average relative improvement in dice score compared to naive supervised learning (16.18% average improvement). Considering our main competitor (RL), we note that our proposed TI-ST method is a one-stage framework using one network. In contrast, RL is a two-stage framework (requiring a

Table 1. Quantitative comparisons in Dice score (%) among the proposed (TI-ST) and alternative methods for DeepLabV3+ [3] (DLV3+) and scSENet [20] and the three datasets. Relative Dice computed over the Supervised baseline. The best results are shown in green.

Modality	Cataract Surgery		OCT		MRI		Avg. Rel.
Network	DLV3+	scSENet	DLV3+	scSENet	DLV3+	scSENet	
Supervised	15.42	37.67	22.87	24.08	52.39	65.93	N/A
Π Model [15]	27.55	35.56	1.12	0.00	10.00	6.87	-22.88
TE [15]	33.10	42.32	42.13	39.86	63.41	67.25	11.62
Mean Teacher [19]	11.06	39.54	19.11	4.70	64.82	66.87	-2.04
RL [25]	34.40	45.13	48.73	47.70	60.79	70.20	14.77
CPS [4]	36.24	39.40	47.31	14.71	76.00	68.80	10.68
ST [24]	34.34	41.10	36.84	33.01	68.63	71.97	11.26
MCF [23]	26.97	40.19	40.12	36.52	54.17	50.23	7.46
TI-ST	37.69 (+22.27)	45.31 (+7.46)	50.93 (+28.06)	40.87 (+16.79)	66.56 (+14.17)	74.07 (+8.14)	16.18

pre-training stage) and uses a teacher-student network. Hence, TI-ST is also more efficient than RL in terms of time and computation. Furthermore, the proposed strategy demonstrates the most consistent results when evaluated on different tasks, regardless of the utilized neural network architecture.

Figure 3-(a–b) demonstrates the effect of the pseudo-labeling threshold on TI-ST performance compared with regular ST. We observe that filtering out unreliable pseudo-labels based on transformation variance can remarkably boost pseudo-supervision performance regardless of the threshold. Figure 3-(c) compares the performance of the supervised baseline, ST, and TI-ST with respect to the number of source-domain labeled training images. While ST performance converges when the number of labeled images increases, our TI-ST pushes deci-

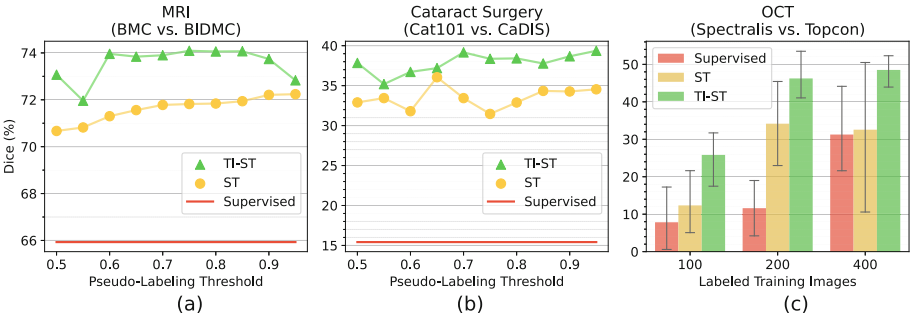

Fig. 3. Ablation studies on the pseudo-labeling threshold and size of the labeled dataset.

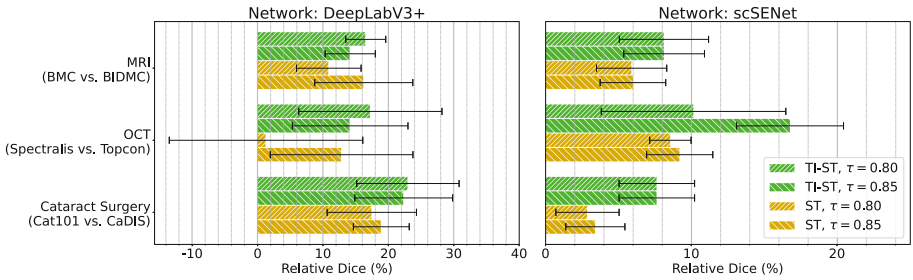

Fig. 4. Ablation study on the performance stability of TI-ST vs. ST across the different experimental segmentation tasks.

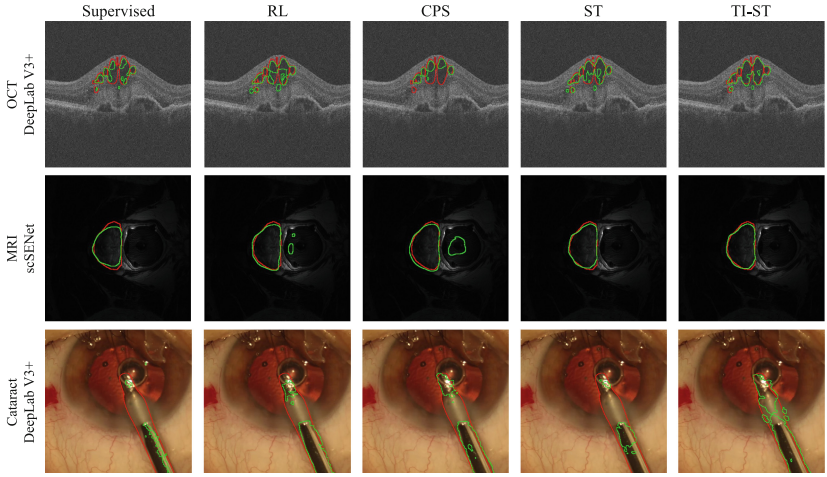

Fig. 5. Qualitative comparisons between the performance of TI-ST and four existing methods.

sion boundaries toward the target domain dataset by avoiding training with transformation variant pseudo-labels. We validates the stability of TI-ST vs. ST with different labeling thresholds (0.80 and 0.85) over four training folds in Fig. 4, where TI-ST achieves a higher average improvement relative to supervised learning for different tasks and network architectures. This analysis also shows that the performance of ST is sensitive to the pseudo-labeling threshold and generally degrades by reducing the threshold due to resulting in wrong pseudo labels. However, TI-ST can effectively ignore false predictions in lower thresholds and take advantage of a higher amount of correct pseudo labels. This superior performance is depicted qualitatively in Fig. 5.

5 Conclusion

We proposed a novel self-training framework with a self-assessment strategy for pseudo-label reliability, namely "Transformation-Invariant Self-Training" (TI-ST). This method uses transformation-invariant highly-confident predictions in the target dataset by considering an ensemble of high-confidence predictions from transformed versions of identical inputs. We experimentally show the effectiveness of our approach against numerous existing methods across three different source-to-target segmentation tasks, and when using different model architectures. Beyond this, we show that our approach is resilient to changes in the methods hyperparameter, making it well-suited for different applications.

References

1. Bogunović, H., et al.: RETOUCH: the retinal oct fluid detection and segmentation benchmark and challenge. IEEE Trans. Med. Imaging **38**(8), 1858–1874 (2019)
2. Chaitanya, K., Erdil, E., Karani, N., Konukoglu, E.: Contrastive learning of global and local features for medical image segmentation with limited annotations. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (eds.) Advances in Neural Information Processing Systems, vol. 33, pp. 12546–12558. Curran Associates, Inc. (2020)
3. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with Atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV), pp. 801–818 (2018)
4. Chen, X., Yuan, Y., Zeng, G., Wang, J.: Semi-supervised semantic segmentation with cross pseudo supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2613–2622 (2021)
5. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. IEEE (2009)
6. Ghamsarian, N.: Enabling relevance-based exploration of cataract videos. In: Proceedings of the 2020 International Conference on Multimedia Retrieval, ICMR '20, pp. 378–382 (2020). https://doi.org/10.1145/3372278.3391937
7. Ghamsarian, N., Taschwer, M., Putzgruber-Adamitsch, D., Sarny, S., El-Shabrawi, Y., Schoeffmann, K.: LensID: a CNN-RNN-based framework towards lens irregularity detection in cataract surgery videos. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12908, pp. 76–86. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87237-3_8
8. Ghamsarian, N., Taschwer, M., Putzgruber-Adamitsch, D., Sarny, S., El-Shabrawi, Y., Schöffmann, K.: ReCal-Net: joint region-channel-wise calibrated network for semantic segmentation in cataract surgery videos. In: Mantoro, T., Lee, M., Ayu, M.A., Wong, K.W., Hidayanto, A.N. (eds.) ICONIP 2021. LNCS, vol. 13110, pp. 391–402. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-92238-2_33
9. Ghamsarian, N., Taschwer, M., Putzgruber-Adamitsch, D., Sarny, S., Schoeffmann, K.: Relevance detection in cataract surgery videos by Spatio- temporal action localization. In: 2020 25th International Conference on Pattern Recognition (ICPR), pp. 10720–10727 (2021)

10. Ghamsarian, N., Taschwer, M., Sznitman, R., Schoeffmann, K.: DeepPyramid: enabling pyramid view and deformable pyramid reception for semantic segmentation in cataract surgery videos. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. Lecture Notes in Computer Science, vol. 13435, pp. 276–286. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16443-9_27
11. 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.) MICCAI 2022. Lecture Notes in Computer Science, vol. 13438, pp. 351–361. Springer Nature Switzerland, Cham (2022). https://doi.org/10.1007/978-3-031-16452-1_34
12. Grammatikopoulou, M., et al.: CaDIS: cataract dataset for surgical RGB-image segmentation. Med. Image Anal. **71**, 102053 (2021)
13. Guo, R., et al.: Using domain knowledge for robust and generalizable deep learning-based CT-free PET attenuation and scatter correction. Nat. Commun. **13**(1), 5882 (2022)
14. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
15. Laine, S., Aila, T.: Temporal ensembling for semi-supervised learning. CoRR abs/1610.02242 (2016). http://arxiv.org/abs/1610.02242
16. Li, C., Zhou, Y., Shi, T., Wu, Y., Yang, M., Li, Z.: Unsupervised domain adaptation for the histopathological cell segmentation through self-ensembling. In: Atzori, M., et al. (eds.) Proceedings of the MICCAI Workshop on Computational Pathology. Proceedings of Machine Learning Research, vol. 156, pp. 151–158. PMLR (2021)
17. Li, X., Yu, L., Chen, H., Fu, C.W., Xing, L., Heng, P.A.: Transformation-consistent self-ensembling model for semisupervised medical image segmentation. IEEE Trans. Neural Netw. Learn. Syst. **32**(2), 523–534 (2021)
18. Liu, Q., Dou, Q., Yu, L., Heng, P.A.: MS-Net: multi-site network for improving prostate segmentation with heterogeneous MRI data. IEEE Trans. Med. Imaging **39**, 2713–2724 (2020)
19. Perone, C.S., Ballester, P., Barros, R.C., Cohen-Adad, J.: Unsupervised domain adaptation for medical imaging segmentation with self-ensembling. Neuroimage **194**, 1–11 (2019)
20. Roy, A.G., Navab, N., Wachinger, C.: Recalibrating fully convolutional networks with spatial and channel "squeeze and excitation" blocks. IEEE Trans. Med. Imaging **38**(2), 540–549 (2019)
21. Schoeffmann, K., Taschwer, M., Sarny, S., Münzer, B., Primus, M.J., Putzgruber, D.: Cataract-101: video dataset of 101 cataract surgeries. In: Proceedings of the 9th ACM Multimedia Systems Conference, pp. 421–425 (2018)
22. Varsavsky, T., Orbes-Arteaga, M., Sudre, C.H., Graham, M.S., Nachev, P., Cardoso, M.J.: Test-time unsupervised domain adaptation. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12261, pp. 428–436. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59710-8_42
23. Wang, Y., Xiao, B., Bi, X., Li, W., Gao, X.: MCF: mutual correction framework for semi-supervised medical image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15651–15660 (2023)
24. Yang, L., Zhuo, W., Qi, L., Shi, Y., Gao, Y.: ST++: make self-training work better for semi-supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4268–4277 (2022)

25. Zeng, X., et al.: Reciprocal learning for semi-supervised segmentation. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12902, pp. 352–361. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87196-3_33
26. Zou, Y., Yu, Z., Vijaya Kumar, B.V.K., Wang, J.: Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11207, pp. 297–313. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01219-9_18