



Unsupervised Site Adaptation by Intra-site Variability Alignment

Shaya Goodman¹, Shira Kasten Serlin¹, Hayit Greenspan¹,
and Jacob Goldberger²(✉)

¹ Tel-Aviv University, Tel-Aviv, Israel
{yishayahug,kastenserlin}@mail.tau.ac.il, hayit@eng.tau.ac.il
² Bar-Ilan University, Ramat-Gan, Israel
jacob.goldberger@biu.ac.il

Abstract. A medical imaging network that was trained on a particular source domain usually suffers significant performance degradation when transferred to a different target domain. This is known as the domain-shift problem. In this study, we propose a general method for transfer knowledge from a source site with labeled data to a target site where only unlabeled data is available. We leverage the variability that is often present within each site, the *intra-site variability*, and propose an *unsupervised site adaptation* method that jointly aligns the intra-site data variability in the source and target sites while training the network on the labeled source site data. We applied our method to several medical MRI image segmentation tasks and show that it consistently outperforms state-of-the-art methods.

Keywords: Unsupervised domain adaptation · UDA · Intra-site variability · MRI segmentation

1 Introduction

Neural networks have been successfully applied to medical image analysis. Unfortunately, a model that is trained to achieve high performance on a certain dataset, often drops in performance when tested on medical images from different acquisition protocols or different clinical sites. This model robustness problem, known as domain shift, especially occurs in Magnetic Resonance Imaging (MRI) since different scanning protocols result in significant variations in slice thickness and overall image intensities. Site adaptation improves model generalization capabilities in the target site by mitigating the domain shift between the sites. Unsupervised Domain Adaptation (UDA) assumes the availability of data from the new site but without manual annotations. The goal of UDA is to train a network using both the labeled source site data and the unlabeled target site data to make accurate predictions about the target site data. In this study we

This research was supported by the Ministry of Science & Technology, Israel. The first two authors contributed equally.

concentrate on segmentation tasks (see an updated review on UDA for segmentation in [6]). Another setup is supervised domain adaptation where we also have labeled data from the target site (see e.g. [10]).

Recent UDA methods include feature alignment adversarial networks that are based on learning domain-invariant features using a domain discriminator which is co-trained with the network [23, 26, 28]. Image alignment adversarial networks (e.g. [4, 5, 8]) translate the appearance from one domain to another using multiple discriminators and a pixel-wise cycle consistency loss. Seg-JDOT [1] solves a site adaptation scenario using optimal transport theory by presenting a domain shift minimization in the feature space. Li et al. Another approach is transferring the trained model to a new domain by modulating the statistics in the Batch-Normalization layer [13, 16]. Some methods such as [2, 12] suggest test-time adaptation methods.

Intra-site variability can result from multiple reasons in the medical space, including slice variability across an imaged organ, varying scanning protocols and differences in the patient population being imaged. The intra-variability of the data collected from the source and target site is often based on similar factors. Importantly, this can be exploited in the site adaptation process. Recent studies on UDA for classification have used intra-site variability induced by different classes to divide the feature space into different subsets. [7, 9, 20, 27]. Pseudo labels, which are produced for samples in the target domain, are used for domain alignment. These methods cannot be applied to segmentation tasks, as mentioned in [1], since the number of possible segmentation maps is exponentially larger than the number of classes in a classification task.

This gap has motivated us to look for a different approach for solving the domain shift problem for segmentation tasks. We present a domain adaptation approach that tackles the inter-domain shift by aligning the intra-variability of the source and target sites. Our approach consistently out-performs the state-of-the-art site adaptation methods on several publicly available medical images segmentation tasks. The code to reproduce our experiments is available at <https://github.com/yishayahu/AIVA.git>.

2 Site Adaptation Based on Intra-site Variability Alignment

We present an *unsupervised* site adaptation method that explicitly takes the intra-site variability into account. We concentrate on MRI image segmentation task. In this scenario, we are given a U-net network that was trained on the source site. We jointly align the feature space of the target site to the source site, so as when optimizing the model on the source site, we obtain a model that performs well on the target site as well. More specifically, our method minimizes the domain shift between the source and the target by aligning the intra-site variability of the target site with the intra-site variability of the source site. The intra-site variability is modeled by separately clustering the source and the target sites in a suitable embedded space. The centers of the clusters of the

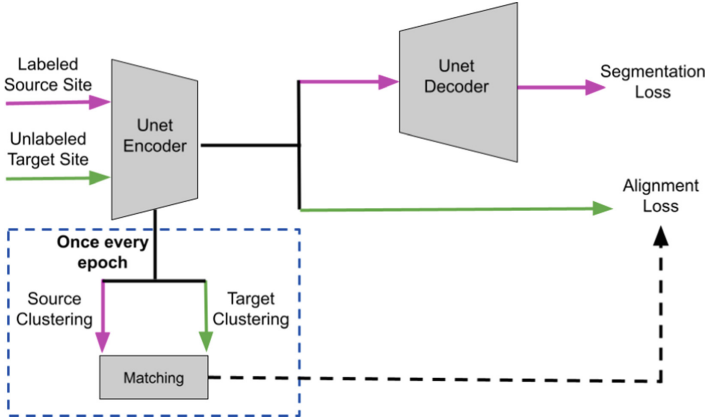


Fig. 1. A scheme of the fine-tuning loss assembly of the AIVA site adaptation method.

two sites are then matched, and each target cluster is pushed in towards its corresponding source cluster. In parallel, the segmentation loss is minimized on the source labeled data to maintain accurate semantic segmentation masks for the source site. Aligning the structure of the target site with the source site while maintaining good results on the source site, yields a good segmentation performance on the target site. In what follows, we provide a detailed description of each step of the proposed site adaptation algorithm.

Intra-site Variability Modeling. The intra-site variability is modeled by clustering the images of each site in a suitable embedded space. We compute an image embedding by considering the segmentation U-net bottleneck layer with its spatial dimensions and its convolutional filter dimension. We denote this image representation as the BottleNeck Space (BNS). Next, we apply the k -means algorithm to cluster the source site images in the BNS into k centers and in a similar manner we cluster the target site images into k centers. It is well known that applying k -means clustering to high-dimensional data does not work well because of the curse of dimensionality. Hence, in practice the actual clustering of the image representations is computed in a 2D embedding obtained by the PCA algorithm [11] followed by the t-SNE algorithm [19] that are applied jointly to the BNS representations of the source and target data points. We denote the 2D k clustering centers of the source site by $\{\mu_i^s\}_{i=1}^k$, and the 2D target site centers by $\{\mu_i^t\}_{i=1}^k$.

Clustering Matching. In this step, we align the intra-site variability structure of the target site to the source site by matching the two clusterings. We look for the optimal matching between the k source centers μ_1^s, \dots, μ_k^s and the k target centers μ_1^t, \dots, μ_k^t :

$$\hat{\pi} = \arg \min_{\pi} \sum_{i=1}^k \|\mu_i^t - \mu_{\pi(i)}^s\|^2 \quad (1)$$

where π goes over all the $k!$ permutations. The Kuhn-Munkers matching algorithm, also known as the Hungarian method [14,21] is an algorithm that can efficiently solve the minimization problem (1) in time complexity $O(k^3)$. The clustering of the source and target images and the matching between the clusterings’ centers are done once every epoch and are kept fixed throughout all the mini-batches of the epoch. This implies that the t-SNE procedure, the clustering and the matching algorithms do not need to be differentiable with respect to the model parameters since this process is separate from the backwards calculation of gradients and their impact on the total training running time is negligible (less than 2% addition to training time). Note that we can view the source (and target) site centers as the modes of a multi-modal distribution of the source (and target) data. Aligning the centers thus corresponds to aligning the source and target multi-modal distributions.

Alignment Loss. The assignment (1) found above is used to align the two sites by encouraging each target cluster center to be closer to the corresponding source center. Since in practice we work in mini-batches, we encourage the BNS representation of the average of target images in the current minibatch which were assigned to the same cluster, to be closer to the center of the corresponding source cluster. We define the following loss function in the BNS space:

$$L_{\text{alignment}} = \sum_{i=1}^k \|\bar{x}_i^t - \nu_{\pi(i)}^s\|^2 \quad (2)$$

such that \bar{x}_i^t is the average of all the target-site points in the minibatch that were assigned by the clustering procedure to the i -th cluster. The vector ν_i^s is the average of all source points that were assigned to the i -th cluster (μ_i^s is the average of the same set in the t -SNE embedded space). The domain shift between the source and target sites is thus minimized by aligning the data structure of the target site with the data structure of the source site. Note that in the alignment loss (2), while the source centers are kept fixed during an epoch, the target samples are obtained as a function of the model parameters, and the loss gradients with respect to the parameters are back propagated through them.

In addition to the alignment loss, we use a standard segmentation cross-entropy loss which is computed at the final output layer for the source samples and is designed to avoid degradation of the segmentation performances. Indirectly, it improves the segmentation of the target site data. The overall loss function is thus:

$$L = L_{\text{segmentation}} + \lambda L_{\text{alignment}}. \quad (3)$$

The regularization coefficient λ is a hyper-parameter that is usually tuned using cross-validation. Since there are no labels from the current target site, we cannot tune λ on a validation set. Instead, we use the following unsupervised tuning procedure: we average the values of $L_{\text{alignment}}$ in the first minibatches and define lambda as the reciprocal of the average. This makes the scaled alignment score close to 1 and makes it the same scale as our segmentation loss. The network is

Algorithm 1. The AIVA site adaptation algorithm

input: Source labeled images, target unlabeled images and a U-net that was trained on the source site.

for each epoch **do**

 Compute the BNS representation of all the source and the target samples.

 Cluster each site in a joint t-SNE embedding into k centers $\{\mu_i^s\}_{i=1}^k, \{\mu_i^t\}_{i=1}^k$.

 Find the optimal matching $\hat{\pi}$ between the source and the target centers:

$$\hat{\pi} = \arg \min_{\pi} \sum_{i=1}^k \|\mu_i^t - \mu_{\pi(i)}^s\|^2.$$

for each mini-batch **do**

 For each cluster i , let \bar{x}_i^t be the average of the target-site points (in the BNS domain) in the minibatch that were assigned to cluster i .

 Compute $L_{\text{alignment}} = \sum_{i=1}^k \|\bar{x}_i^t - \nu_{\hat{\pi}(i)}^s\|^2$ where ν_i^s is the average of all source-site points (in the BNS domain) that were assigned cluster i .

 Apply a gradient descent step to the loss: $L = L_{\text{segmentation}} + \lambda L_{\text{alignment}}$.

output: A model adapted to the target site.

pretrained on the source site, and then is adapted to the target site by minimizing the loss function (3). We dub the proposed method Adaptation by Intra-site Variability Alignment (AIVA). A scheme of the loss function of the AIVA algorithm is shown in Fig. 1. The AIVA algorithm is summarized in Algorithm Box 1.

3 Experiments

We evaluated the performance of our method and compared it with other unsupervised domain adaptation methods on two different medical image datasets for segmentation tasks. Our experiments were conducted on the following unsupervised domain adaptation setup: we have labeled data from a source site and unlabeled data from the target site and we are given a network that was trained on the source site data.

We chose a representative baseline from each of the three most dominant approaches today that deal with UDA (image statistics, domain shift minimization in feature space and feature alignment adversarial networks).

- *AdaBN*: recalculating the statistics of the batch normalization layers on the target site [16].
- *Seg-JDOT*: aligning the distributions of the source and the target sites using an optimal transport algorithm [1].
- *AdaptSegNet*: aligning feature space using adversarial learning [26].

We also directly trained a network on the target site using the labels of the training data of the target site, thereby setting an *upper bound* for UDA methods. In addition, we show the results on the pretrained model without any adaptation to set a lower bound.

MRI Skull Stripping: The publicly available dataset CC359 [25] consists of 359 MR images of heads where the task consists of skull stripping. The dataset

Table 1. Segmentation surface-Dice results on the brain MRI dataset CC359 [25].

Target site	Target site model	Source site model	Unsupervised adaptation			
			AdaBN	Seg-JDOT	AdaptSegNet	AIVA
Siemens, 1.5T	80.13	58.44	62.20	64.27	63.41	67.43
Siemens, 3T	80.19	59.65	58.82	61.83	57.57	66.31
GE, 1.5T	81.58	38.96	58.04	50.06	57.71	59.95
GE, 3T	84.16	56.27	54.23	59.82	55.94	65.04
Philips, 1.5T	84.00	56.38	73.01	69.59	68.22	75.68
Philips, 3T	82.19	41.93	51.18	50.30	50.51	54.34
Average	82.04	51.94	59.58	59.31	58.89	64.79

was collected from six sites which exhibit domain shift resulting in a severe score deterioration [24]. For preprocessing we interpolated to $1 \times 1 \times 1$ mm voxel spacing and scaled the intensities to a range of 0 to 1. To evaluate the different approaches, we used the surface Dice score [22] at a tolerance of 1 mm. While preserving consistency with the methodology in [24], we also found that surface Dice score to be a more suitable metric for the brain segmentation task than the standard Dice Score (similar to [29]). We used a U-net network that processes each 2D image slice separately. All the models were pretrained on a single source data for 5K steps starting with a learning rate of 10^{-3} that polynomially decays with an exponential power of 0.9 and a batch size 16. All compared models were finetuned using 6.5K steps. For AIVA we used 12 clusters. We ensured that all the models reached the loss plateau. Each target site was split into a training set and a test set. Since the assumption here was that we only has unlabeled images from the target site we chose the checkpoint using the performance on the source test set. We used 25 pairs of source and target sites and averaged the results of each target site. The remaining five pairs were used to examine the robustness of the method to different amount of clusters. The surface-Dice results are shown at Table 1. It highlights the significant deterioration between the supervised and the no-adaptation. Furthermore, we observe that our model consistently outperformed the baselines for each new site.

We visualize the alignment process in the AVIA algorithm. Intuitively we expect the intra-variability to be represented by the different clusters and the matching to align them across the source and the target. This is demonstrated in Fig. 2 (clusters 1–4) by examples from each cluster. Figure 3 shows the clustering of the source and target slices and the matching between the clusters. The two clusterings are similar, but not perfectly aligned due to the domain shift. Figure 4 shows that after the adaptation process the two sites are better aligned as a result of minimization of the alignment loss. Finally, Fig. 5 shows the sDice score as a function of the number of clusters (averaged over 5 source-target pairs). We can see that AIVA is robust to the amount of clusters when it is at least 9.



Fig. 2. Matching clusters images examples from source (top) and target (bottom) from CC359 brain dataset (1–4) and the prostate dataset (5–7).

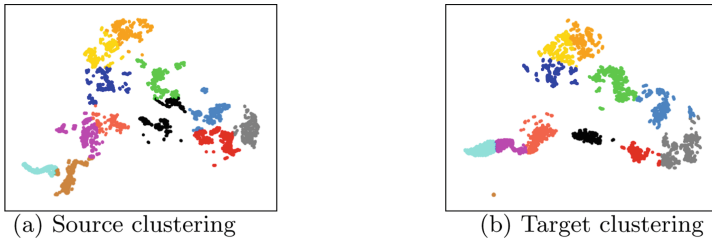


Fig. 3. Clustering of the slice samples of source (a) and target (b), at the beginning of the fine-tuning phase, in 2D space. Matched clusters - same color.

Prostate MRI Segmentation: To show the robustness of our method we evaluated it on a multi-source single-target setup as well. We used a publicly available multi-site dataset for prostate MRI segmentation which contains prostate T2-weighted MRI data (with segmentation masks) collected from different data sources with a distribution shift. Details of data and imaging protocols from the six different sites appear in [18]. Samples of sites A and B were taken from the NCI-ISBI13 dataset [3], samples of site C were from the I2CVB dataset [15], and samples of sites D, E and F were from the PROMISE12 dataset [17].

For pre-processing, we normalized each sample to have a zero mean and a unit variance in intensity value before inputting to the network. For each target site we used the other five sites as the source. The results were calculated on six possible targets. To evaluate different approaches, we used the Dice Score. We used the same network architecture and learning rate as in the experiment described above. We pretrained the network for 3.5K steps and finetuned the model for every method for another 3.5K steps. We ensured that all the models reached the loss plateau. Each site was split into a training set and a test set. We chose the checkpoint to evaluate using the source test set. We showed in the previous experiment that the AIVA algorithm is robust to the number of clusters. We fixed the number of clusters here to twelve as before.

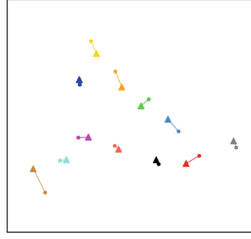
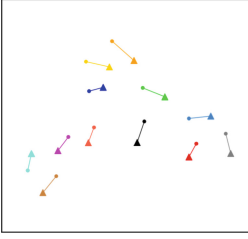


Fig. 4. Clusters’ centers of the source (circles) and the target (triangles) in the 2D space before (left) and after (right) the adaptation phase.

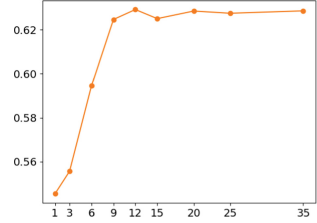


Fig. 5. The AIVA sDice-score as a function of the of number of clusters.

Results. Figure 2 (5–7) shows the matching clusters in the training process: whereas in the Brain data the clusters focused on morphological variations, here we see a focus on the image contrast variability. In Table 2 we present comparative performances for each target site. We could not get a convergence for seg-JDOT [1] on this dataset, probably due to lack of data. Therefore, we omitted it from the result report. We note that AIVA yielded the overall best Dice score. In some sites, the difference between the supervised training and the source model is relatively small. For these cases, relatively weak results were seen for some of the UDA methods. AIVA showed stability by consistently yielding improved results. Examples of segmentation results are shown in Fig. 6.

Table 2. Segmentation Dice results on the prostate MRI dataset [18].

Target site	Target site model	Source site model	Unsupervised adaptation		
			AdaBN	AdaptSegNet	AIVA
Site A	87.48	80.63	79.03	76.63	79.94
Site B	85.62	59.84	73.32	64.45	78.50
Site C	78.86	58.84	68.08	67.64	69.67
Site D	85.30	61.78	66.99	77.39	78.88
Site E	80.55	77.72	80.48	79.13	79.65
Site F	86.10	79.98	53.36	80.88	85.42
Average	83.98	69.80	70.21	74.35	78.67

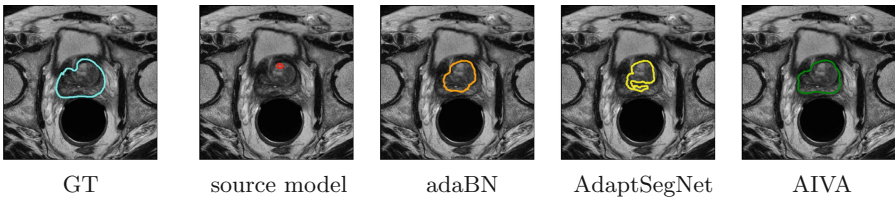


Fig. 6. Qualitative segmentation results from the prostate MRI dataset.

4 Conclusion

To conclude, in this study we presented AIVA, a general scheme for unsupervised site adaptation. The intra-site variability of the data collected from the source and target sites is often based on similar factors. AIVA uses this observation to align the two sites. Our experiments showed that AIVA is robust to the variations exhibited and consistently improves results over previous site adaptation methods. We concentrated here on two applications. The proposed method, however, is general and is especially suitable for segmentation tasks where we cannot align the source and target site using the labels.

References

1. Ackaouy, A., Courty, N., Vallée, E., Commowick, O., Barillot, C., Galassi, F.: Unsupervised domain adaptation with optimal transport in multi-site segmentation of multiple sclerosis lesions from MRI data. *Frontiers Comput. Neurosci.* **14**, 19 (2020)
2. Bateson, M., Kervadec, H., Dolz, J., Lombaert, H., Ayed, I.B.: Source-relaxed domain adaptation for image segmentation (2020)
3. Bloch, N., et al.: NCI-ISBI 2013 challenge: automated segmentation of prostate structures. *Cancer Imaging Arch.* **370** (2015)
4. 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* (2019)
5. Chen, C., Dou, Q., Chen, H., Qin, J., Heng, P.A.: Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation. *IEEE Trans. Med. Imaging* **39**(7), 2494–2505 (2020)
6. Csurka, G., Volpi, R., Chidlovskii, B.: Unsupervised domain adaptation for semantic image segmentation: a comprehensive survey. *arXiv preprint [arXiv:2112.03241](https://arxiv.org/abs/2112.03241)* (2021)
7. Deng, Z., Luo, Y., Zhu, J.: Cluster alignment with a teacher for unsupervised domain adaptation. In: *International Conference on Computer Vision (ICCV)* (2019)
8. Dou, Q., et al.: Pnp-adanet: Plug-and-play adversarial domain adaptation network with a benchmark at cross-modality cardiac segmentation. *arXiv preprint [arXiv:1812.07907](https://arxiv.org/abs/1812.07907)* (2018)
9. Gao, B., Yang, Y., Gouk, H., Hospedales, T.M.: Deep clustering for domain adaptation. In: *International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2020)
10. Goodman, S., Greenspan, H., Goldberger, J.: Supervised domain adaptation using gradients transfer for improved medical image analysis. In: *MICCAI Workshop on Domain Adaptation and Representation Transfer (DART)* (2022)
11. Jolliffe, I.: *Principal Component Analysis* (1986)
12. Karani N, Erdil E, C.K.K.E.: Test-time adaptable neural networks for robust medical image segmentation. *MedIA* **68**, 101907 (2021)
13. Kasten-Serlin, S., Goldberger, J., Greenspan, H.: Adaptation of a multisite network to a new clinical site via batch-normalization similarity. In: *The IEEE International Symposium on Biomedical Imaging (ISBI)* (2022)

14. Kuhn, H.W.: The Hungarian method for the assignment problem. *Naval Res. Logistics Q.* **2**, 83–97 (1955)
15. Lemaître, G., Martí, R., Freixenet, J., Vilanova, J.C., Walker, P.M., Meriaudeau, F.: Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: a review. *CBM* **60**, 8–31 (2015)
16. Li, Y., Wang, N., Shi, J., Hou, X., Liu, J.: Adaptive batch normalization for practical domain adaptation. *Pattern Recogn.* **80**, 109–117 (2018)
17. Litjens, G., et al.: Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge. *MIA* **18**(2), 359–373 (2014)
18. Liu, Q., Dou, Q., Heng, P.-A.: Shape-aware meta-learning for generalizing prostate MRI segmentation to unseen domains. In: Martel, A.L., et al. (eds.) *MICCAI 2020*. LNCS, vol. 12262, pp. 475–485. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59713-9_46
19. Van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**(11), 2579–2605 (2008)
20. Menapace, W., Lathuilière, S., Ricci, E.: Learning to cluster under domain shift. In: *European Conference on Computer Vision* (2020)
21. Munkers, J.: Algorithms for the assignment and transportation problem. *J. Soc. Ind. Appl. Math.* **5**, 32–38 (1957)
22. Nikolov, S., et al.: Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy. *CoRR* abs/1809.04430 (2018)
23. Panfilov, E., Tiulpin, A., Klein, S., Nieminen, M.T., Saarakkala, S.: Improving robustness of deep learning based knee mri segmentation: Mixup and adversarial domain adaptation. In: *Proceedings of the IEEE International Conference on Computer Vision Workshops* (2019)
24. Shirokikh, B., Zakazov, I., Chernyavskiy, A., Fedulova, I., Belyaev, M.: First U-net layers contain more domain specific information than the last ones. In: Albarqouni, S., et al. (eds.) *DART/DCL -2020*. LNCS, vol. 12444, pp. 117–126. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-60548-3_12
25. Souza, R., et al.: An open, multi-vendor, multi-field-strength brain MR dataset and analysis of publicly available skull stripping methods agreement. *Neuroimage* **170**, 482–494 (2018)
26. 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* (2018)
27. Xu, T., Chen, W., Wang, P., Wang, F., Li, H., Jin, R.: Cdtrans: Cross-domain transformer for unsupervised domain adaptation. In: *International Conference on Learning Representations (ICLR)* (2022)
28. Yan, W., Wang, Y., Xia, M., Tao, Q.: Edge-guided output adaptor: highly efficient adaptation module for cross-vendor medical image segmentation. *IEEE Signal Process. Lett.* **26**(11), 1593–1597 (2019)
29. Zakazov, I., Shirokikh, B., Chernyavskiy, A., Belyaev, M.: Anatomy of domain shift impact on U-net layers in MRI segmentation. In: de Bruijne, M., et al. (eds.) *MICCAI 2021*. LNCS, vol. 12903, pp. 211–220. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87199-4_20