

# MS-MT: Multi-Scale Mean Teacher with Contrastive Unpaired Translation for Cross-Modality Vestibular Schwannoma and Cochlea Segmentation

Ziyuan Zhao<sup>1,2,3</sup>, Kaixin Xu<sup>1</sup>, Huai Zhe Yeo<sup>1,4,\*</sup>, Xulei Yang<sup>1,2</sup>, and Cuntai Guan<sup>3</sup>

<sup>1</sup> Institute for Infocomm Research (I<sup>2</sup>R), A\*STAR, Singapore

<sup>2</sup> Artificial Intelligence, Analytics And Informatics (AI<sup>3</sup>), A\*STAR, Singapore

<sup>3</sup> Nanyang Technological University, Singapore

<sup>4</sup> National University of Singapore, Singapore

**Abstract.** Domain shift has been a long-standing issue for medical image segmentation. Recently, unsupervised domain adaptation (UDA) methods have achieved promising cross-modality segmentation performance by distilling knowledge from a label-rich source domain to a target domain without labels. In this work, we propose a multi-scale self-ensembling based UDA framework for automatic segmentation of two key brain structures *i.e.*, Vestibular Schwannoma (VS) and Cochlea on high-resolution T2 images. First, a segmentation-enhanced contrastive unpaired image translation module is designed for image-level domain adaptation from source T1 to target T2. Next, multi-scale deep supervision and consistency regularization are introduced to a mean teacher network for self-ensemble learning to further close the domain gap. Furthermore, self-training and intensity augmentation techniques are utilized to mitigate label scarcity and boost cross-modality segmentation performance. Our method demonstrates promising segmentation performance with a mean Dice score of 83.8% and 81.4% and an average asymmetric surface distance (ASSD) of 0.55 mm and 0.26 mm for the VS and Cochlea, respectively in the validation phase of the crossMoDA 2022 challenge.

**Keywords:** Medical image segmentation · Unsupervised domain adaptation · Vestibular Schwannoma

## 1 Introduction

Medical image segmentation plays a vital role in the field of medical image analysis, delivering valuable information for diagnostic analysis and treatment planning [10]. Accurate segmentation and measurement of Vestibular Schwannoma (VS) and Cochlea from MRI can assist in VS treatment planning, improving

---

\* This work was done when Huai Zhe Yeo was an intern at I2R, A\*STAR.

clinical workflow [24]. To this end, researchers have turned to deep learning as a solution to perform autonomous segmentation of VS and Cochlea. Recent findings also suggest that high-resolution T2 (hrT2) MRI could be a safer and more cost-efficient alternative to contrast-enhanced T1 (ceT1) MRI. However, the large domain shift between MRI images with different contrasts coupled with the costly and laborious process of re-annotating medical image scans on another modality, makes it difficult for deep learning to generalize well across both domains. Therefore, we are encouraged to perform unsupervised domain adaptation (UDA) and conduct VS and Cochlea segmentation in the hrT2 domain by leveraging labeled ceT1 scans and unlabeled hrT2 scans. In this work, we propose an effective and intuitive UDA method based on image translation and self-ensembling learning. Firstly, we translate MRI scans from the ceT1 domain to the hrT2 domain using a modified CUT model [21] for image-level adaptation. Then, a nnU-Net [13] is trained on synthetic hrT2 images to generate pseudo annotations for unlabeled hrT2 images. Finally, we build a multi-scale mean-teacher (MS-MT) network [28, 15] to transfer multi-level knowledge from the teacher model to the student model for improving the cross-modality segmentation performance. The experimental results show that the proposed UDA network can greatly reduce the domain gap, achieving promising segmentation performance on hrT2 scans.

## 2 Related Work

To bridge the domain gap across different modalities, many UDA methods have been developed in the medical imaging field to align the distributions between modalities from different perspectives, including image-level alignment [36, 12, 31], feature-level alignment [29, 18, 7, 30] and their combinations [11, 2]. Image alignment is a practical UDA approach that translates source images to appear as if they were sampled from the target domain or vice versa. CycleGAN [36] is mainly used to achieve unpaired image-to-image translation for UDA in medical image segmentation [12, 31]. For instance, Huo *et al.* [12] leveraged CycleGAN to perform MRI to CT synthesis for enabling CT splenomegaly segmentation without using target labels. Zhang *et al.* [31] proposed a task-driven generative adversarial network by introducing segmentation consistency into the DRR to X-ray translation process to achieve X-ray segmentation with the pre-trained DRR segmentation models. Recently, contrastive learning has demonstrated a powerful capacity for unsupervised visual representation learning in various computer vision tasks [14]. Park *et al.* [21] first proposed CUT to use contrastive learning for image synthesis, in which, the mutual information between original and generated patches was maximized to keep the content unchanged after translation. Several studies have applied CUT in cross-domain medical image analysis [3, 17]. Therefore, in this study, we use CUT for image-level adaptation. Another line of research in UDA is to align the feature distributions across domains from different aspects, including explicit discrepancy minimization [29, 18] and implicit adversarial learning [9, 7]. For example, Dou *et al.* [7] proposed to

fine-tune specific feature layers via adversarial learning for cross-modality cardiac segmentation. To achieve image and feature adaptation, CyCADA [11] was proposed to bridge the appearance gap using CycleGAN and align the feature spaces with adversarial learning separately. Chen *et al.* [2] further introduced additional discriminators in CycleGAN to achieve concurrent feature and image adaptation. These UDA methods have been used in cross-modality VS and Cochlea Segmentation for closing the domain gap [6]. More recently, the usage of weak labels, such as scribbles on the target domain, for weakly-supervised domain adaptation [5] has also shown much promise, receiving attention from the community.

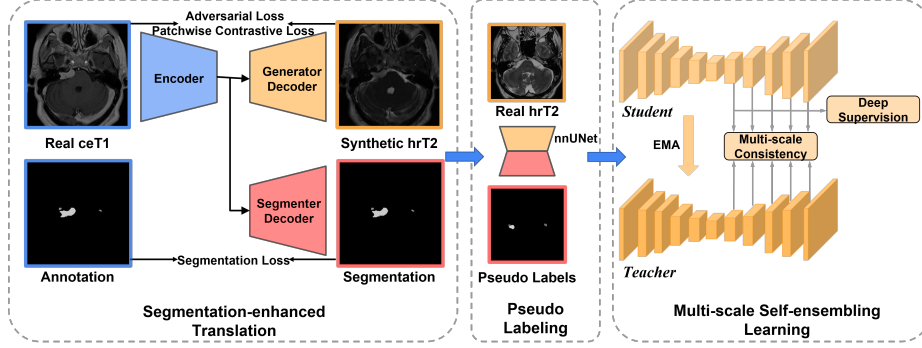
On the other hand, since the labels on the target domain are not accessible, not-so-supervised learning methods [27], such as semi-supervised learning (SSL) [1, 28] can be leveraged to relax the dependence on target labels for improving the adaptation performance [37, 33, 34]. Zou *et al.* [37] built a self-training pseudo-labeling framework for UDA, in which, pseudo-target labels are generated to retrain the model iteratively for improving the UDA performance. Zhao *et al.* [33, 34] investigated the source label scarcity problem in UDA, and leveraged self-ensembling models [28] to address annotation scarcity on both domains for label-efficient UDA. These works suggest that SSL techniques can also be leveraged in UDA to improve adaptation performance. In this regard, pseudo-labeling and self-ensembling learning methodologies were explored in our UDA framework.

### 3 Methods

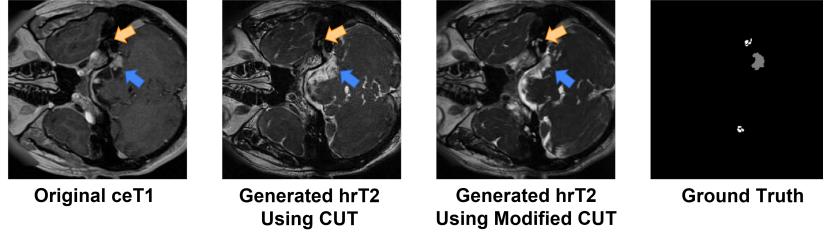
Given an unpaired dataset of two modalities, *i.e.*, annotated ceT1 MRI images  $\mathcal{D}_s = \{(\mathbf{x}_i^s, y_i^s)\}_{i=1}^N$  and non-annotated hrT2 MRI scans  $\mathcal{D}_t = \{(\mathbf{x}_i^t)\}_{i=1}^M$ , sharing the same classes (VS and Cochlea), we aim to exploit  $\mathcal{D}_s$  and  $\mathcal{D}_t$  for unsupervised domain adaptation to enhance the cross-modality segmentation performance of the VS and Cochlea on hrT2 MRI images. The overview of our UDA framework is shown in Fig. 1.

#### 3.1 Segmentation-enhanced translation

To close the domain gap across the modalities, we first conduct image-level domain adaptation to generate synthetic target samples. In this regard, the model trained on the synthetic target images will be used for VS and Cochlea segmentation on real hrT2 scans. For unpaired image-to-image translation, we adopt the Contrastive Unpaired Translation (CUT) [21] as our backbone since it is faster and less memory-intensive than CycleGAN [36]. Moreover, we enhance the 2D CUT with an additional segmentation decoder for maintaining the structural information of VS and Cochlea (see Fig. 1). Specifically, a ResNet-based generator is used to translate images from the source domain to the target domain, while a PatchGAN discriminator is employed to distinguish between the real and generated images [21]. We follow the SIFA architecture [2] and connect two



**Fig. 1.** The overview of our proposed method. First, synthetic hrT2 images are generated with the proposed segmentation-enhanced translation network. Then, we employ a nnU-Net for training on synthetic hrT2 images and generate pseudo labels for unlabeled hrT2 images. Finally, a multi-scale mean teacher network is employed to further close the domain gap.



**Fig. 2.** Visual comparison of the synthetic hrT2 images by different methods.

layers of the encoder, specifically at the last layer and the layer before the last downsampling, with the segmenter decoder to generate multi-level segmentation predictions. The segmentation loss can help the encoder focus more on areas related to the segmentation task and preserve the structure of VS and Cochlea in the translated images. In Fig. 2, we can observe that the modified CUT can better preserve the shape of the VS and Cochlea in comparison with the original CUT framework.

### 3.2 Intensity augmentation and pseudo-labeling

Considering that tumors exhibit T2 heterogeneous signal intensity [20] and Cochleas show T2 hyperintense signal intensity [16], we perform intensity augmentation (IA) and generate augmented data for diversifying the training distributions, thereby improving the model generalizability. Using the generated hrT2 images and the ground truth annotations, the signal intensities of both the VS and Cochlea were muted and intensified by 50% respectively, doubling the training data. On the other hand, to boost the segmentation performance on

real hrT2 images, we adopt a Pseudo-Labeling (PL) strategy to leverage unlabeled hrT2 images by generating pseudo hrT2 annotations. We employ a 3D full resolution nnU-Net [13] for training on synthetic hrT2 images and augmented images, which is then used on unlabeled hrT2 images to generate pseudo-labels.

### 3.3 Multi-scale self-ensembling learning

To further utilize all available data, we propose to take advantage of the self-ensembling network, mean teacher (MT) [28], in which, a teacher model is constructed with the same architecture as the student model, and updated with an exponential moving average (EMA) of the student parameters during training. In our training process, the outputs of the student and teacher models with different perturbations are optimized to be consistent by minimizing the difference using the mean square error (MSE) loss. Inspired by the great success of multi-scale learning in medical image analysis [8, 15, 32], we follow [15] to construct a multi-scale mean teacher (MS-MT) network to leverage multi-scale predictions for deep supervision and consistency regularization. We implement 3D full resolution nnU-Net [13] as the backbone for both teacher and student networks, in which, auxiliary layers are connected to each block of the last five blocks to obtain multi-scale predictions (see Fig. 1).

## 4 Experiments and Results

### 4.1 Dataset and pre-processing

The training dataset released by the MICCAI challenge crossMoDA 2022 [6, 22, 23, 4], includes 210 labeled ceT1 images and 210 unlabeled hrT2 images. Due to the varying voxel spacing in the training data, we resampled the images into a common spacing of  $0.6 \times 0.6 \times 1.0$  mm and normalized the intensity to  $[0, 1]$  using the Min-Max scaling. To remove the noise, the images were cropped into  $256 \times 256$  pixels in the xy-plane using the 75 percentile binary threshold [3], resulting in  $256 \times 256 \times N$  image volumes for 3D nnU-Net training. The processed 3D volumes were sliced along the z-axis to produce N number of 2D image training samples for the 2D CUT. The Dice Score (DSC [%]) [26] and the Average Symmetric Surface Distance (ASSD [voxel]) [19] were used to assess the model performance on VS and Cochlea segmentation.

### 4.2 Implementation details

We used single NVIDIA A40 GPU with 48GB of memory for model training. We followed [21] to optimize the proposed segmentation-enhanced CUT network, in which the weights for adversarial loss, contrastive loss, and segmentation loss were set the same. Following [2], the loss weights for additional segmentation in the modified CUT were set to 1 and 0.1 for the last layer and the second last downsampling layers respectively. For pseudo-labeling, we closely followed the

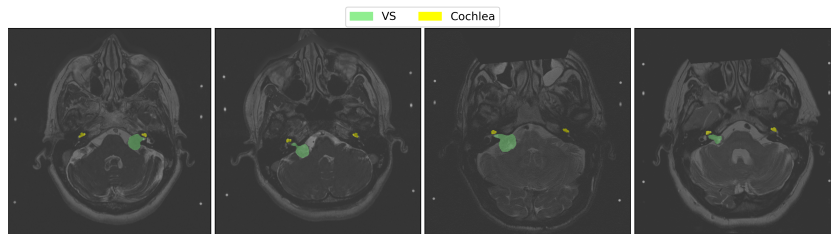
**Table 1.** Quantitative results on the validation dataset. The metrics are presented in the format of mean $\pm$ std.

VS Dice	VS ASSD	Cochlea Dice	Cochlea ASSD	Mean Dice
0.8380 $\pm$ 0.097	0.5555 $\pm$ 0.3647	0.8139 $\pm$ 0.0312	0.2652 $\pm$ 0.1553	0.8260 $\pm$ 0.0508

nnU-Net optimized settings [13] and trained for 200 epochs with the generated hrT2 images and augmented hrT2 images using cross-validation. The segmentation results from an ensemble of five-fold cross-validations on unlabeled hrT2 images were then used as pseudo labels for the following processes. For the multi-scale mean teacher, the EMA update  $\alpha$  was set to 0.9, and the loss weights for consistency regularization were set to  $\{0.05, 0.05, 0.05, 0.4, 0.5\}$ , assigned to each feature map according to its size in ascending order, and we followed the deep supervision scheme in nnU-Net [13]. The loss weights were ramped up across 160 epochs using the sigmoid function. We used a combination of Dice and cross-entropy losses as our objective function and trained the MS-MT model with an initial learning rate of 0.01 using stochastic gradient descent for 300 epochs. Following [28], the results from the five-fold teacher models were ensembled, and then post-processed by computing the largest connected component (LCC) to remove unconnected labels (The first LCC was preserved for the VS while the first and second LCCs were preserved for the Cochlea) for final submission.

### 4.3 Experimental results

The results of our method on the validation dataset are presented in Table. 1. With a mean Dice score of 0.826, our method secures a top-10 finish in the Cross-MoDA 2022 competition. Fig. 3 highlights some qualitative results produced by our method on the validation set. Finally, our method ranked 5th in the test phase of the crossMoDA 2022 challenge with a mean Dice score of 85.65%, and ranked in 3rd place for the Vestibular Schwannoma segmentation with a mean Dice score of 86.7%.

**Fig. 3.** Qualitative results produced by our method. The VS and Cochlea are indicated in green and yellow color, respectively.

#### 4.4 Ablation study

Table 2 shows the results of methods of different components, *e.g.*, intensity augmentation(IA), pseudo-labeling (PL), and mean teacher(MT) on the validation dataset. We start by training a 3D nnU-Net on generated images by CUT, *i.e.*, nnU-Net+CUT, which achieves a mean dice of 78.95%. By adding augmented images into the training process, we can observe an improvement in the segmentation performance of both Cochlea and VS. Then, we introduce pseudo-labeling into the training process, which leads to 1.06% increments in mean Dice, and a continued increase was observed with the inclusion of self-ensembling learning. We further build a multi-scale mean teacher, which achieved better performance. Finally, compared to CUT, the model trained on generated images with the proposed segmentation-enhanced CUT (SE-CUT) can achieve better performance.

**Table 2.** Quantitative results (Dice Score [%]) by different methods on the validation dataset.

Method	Cochlea	VS	Mean
nnU-Net+CUT	79.67 $\pm$ 3.72	78.22 $\pm$ 21.04	78.95 $\pm$ 11.13
nnU-Net+CUT+IA	80.48 $\pm$ 4.14	78.31 $\pm$ 21.00	79.39 $\pm$ 11.28
nnU-Net+CUT+IA+PL	81.06 $\pm$ 3.42	79.85 $\pm$ 12.88	80.45 $\pm$ 6.60
nnU-Net+CUT+IA+PL+MT	81.06 $\pm$ 3.28	82.67 $\pm$ 11.89	81.86 $\pm$ 6.18
nnU-Net+CUT+IA+PL+MS-MT	81.24 $\pm$ 3.31	82.98 $\pm$ 10.25	82.11 $\pm$ 5.39
nnU-Net+SE-CUT+IA+PL+MS-MT	81.39 $\pm$ 3.12	83.80 $\pm$ 9.72	82.60 $\pm$ 5.08

## 5 Discussion

In this work, we propose to explore image-level domain adaptation and semi-supervised learning to address the unsupervised domain adaptation problem in cross-modality vestibular schwannoma (VS) and cochlea segmentation. Additionally, intensity augmentation is performed to generate additional training data with different VS and cochlea intensities for diversifying the data distributions. Our final submission achieved a mean Dice score of 82.6% and 85.65% in the validation phase and the testing phase, respectively. In particular, our method achieved 5th place for overall segmentation performance and ranked 3rd for Vestibular Schwannoma segmentation. From our experiments, we observe that image adaptation contributes the most to the adaptation performance. We can regard the 3D nnU-Net trained on generated target images with CUT (*i.e.*, nnU-Net+CUT) as a strong baseline, achieving dice scores of approximately 80%, which signifies a competitive result within the leaderboards. However, synthetic images may include noises and artifacts, which would influence the follow-up segmentation performance. Many works [25, 21] have been proposed to address the limitations of GANs. In our work, we design a segmentation-enhanced translation network for improved adaptation performance. For data augmentation, we adjust the intensities of VS and cochlea based on clinical characteristics

and medical knowledge and achieved better performance. Besides, spatial augmentation methods [35] can be further explored in this UDA task. For SSL, we adopted pseudo-labeling and self-ensembling learning methodologies to improve the adaptation performance. According to our experimental results, both SSL methods are beneficial to improved segmentation performance. However, one limitation of pseudo-labeling is that low-quality pseudo-labels would adversely influence the training process. Since different perturbations are introduced to the self-ensembling models, the influence of noise from synthetic images and pseudo-labels could be weakened to some extent. We further advance multi-level consistency and deep supervision into self-ensembling learning, achieving better UDA performance.

## 6 Conclusion

In this work, we proposed a three-stage UDA method based on image translation, pseudo-labeling and multi-scale self-ensembling learning to close the gap between two domains and conduct effective image segmentation on the target domain. Specifically, we worked with annotated ceT1 and unannotated hrT2 data to perform segmentation on the target hrT2 domain. In addition, intensity augmentation was implemented to improve the generalizability of the model. With a mean Dice of 0.8216, we achieved a top 10 finish in the CrossMoDA 2022 validation phase and eventually, placed 5th overall in the final competition ranking, demonstrating its effectiveness in bridging the gap between the ceT1 and hrT2 MRI modalities.

## References

1. Bai, W., Oktay, O., Sinclair, M., Suzuki, H., Rajchl, M., Tarroni, G., Glocker, B., King, A., Matthews, P.M., Rueckert, D.: Semi-supervised learning for network-based cardiac mr image segmentation. In: MICCAI. pp. 253–260. Springer (2017)
2. 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 transactions on medical imaging* **39**(7), 2494–2505 (2020)
3. Choi, J.W.: Using out-of-the-box frameworks for contrastive unpaired image translation for vestibular schwannoma and cochlea segmentation: An approach for the crossmoda challenge. In: International MICCAI Brainlesion Workshop. pp. 509–517. Springer (2022)
4. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., et al.: The cancer imaging archive (tcia): maintaining and operating a public information repository. *Journal of digital imaging* **26**(6), 1045–1057 (2013)
5. Dorent, R., Joutard, S., Shapey, J., Bisdas, S., Kitchen, N., Bradford, R., Saeed, S., Modat, M., Ourselin, S., Vercauteren, T.: Scribble-based domain adaptation via co-segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 479–489. Springer (2020)



6. Dorent, R., Kujawa, A., Ivory, M., Bakas, S., Rieke, N., Joutard, S., Glocker, B., Cardoso, J., Modat, M., Batmanghelich, K., Belkov, A., Calisto, M.B., Choi, J.W., Dawant, B.M., Dong, H., Escalera, S., Fan, Y., Hansen, L., Heinrich, M.P., Joshi, S., Kashtanova, V., Kim, H.G., Kondo, S., Kruse, C.N., Lai-Yuen, S.K., Li, H., Liu, H., Ly, B., Oguz, I., Shin, H., Shirokikh, B., Su, Z., Wang, G., Wu, J., Xu, Y., Yao, K., Zhang, L., Ourselin, S., Shapey, J., Vercauteren, T.: Crossmoda 2021 challenge: Benchmark of cross-modality domain adaptation techniques for vestibular schwannoma and cochlea segmentation. *Medical Image Analysis* p. 102628 (2022). <https://doi.org/https://doi.org/10.1016/j.media.2022.102628>
7. Dou, Q., Ouyang, C., Chen, C., Chen, H., Glocker, B., Zhuang, X., Heng, P.A.: Pnp-adanet: Plug-and-play adversarial domain adaptation network at unpaired cross-modality cardiac segmentation. *IEEE Access* **7**, 99065–99076 (2019)
8. Dou, Q., Yu, L., Chen, H., Jin, Y., Yang, X., Qin, J., Heng, P.A.: 3d deeply supervised network for automated segmentation of volumetric medical images. *Medical image analysis* **41**, 40–54 (2017)
9. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V.: Domain-adversarial training of neural networks. *The journal of machine learning research* **17**(1), 2096–2030 (2016)
10. Hesamian, M.H., Jia, W., He, X., Kennedy, P.: Deep learning techniques for medical image segmentation: achievements and challenges. *Journal of digital imaging* **32**(4), 582–596 (2019)
11. Hoffman, J., Tzeng, E., Park, T., Zhu, J.Y., Isola, P., Saenko, K., Efros, A., Darrell, T.: Cycada: Cycle-consistent adversarial domain adaptation. In: *International conference on machine learning*. pp. 1989–1998. PMLR (2018)
12. Huo, Y., Xu, Z., Bao, S., Assad, A., Abramson, R.G., Landman, B.A.: Adversarial synthesis learning enables segmentation without target modality ground truth. In: *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*. pp. 1217–1220. IEEE (2018)
13. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods* **18**(2), 203–211 (2021)
14. Jaiswal, A., Babu, A.R., Zadeh, M.Z., Banerjee, D., Makedon, F.: A survey on contrastive self-supervised learning. *Technologies* **9**(1), 2 (2021)
15. Li, S., Zhao, Z., Xu, K., Zeng, Z., Guan, C.: Hierarchical consistency regularized mean teacher for semi-supervised 3d left atrium segmentation. In: *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. pp. 3395–3398. IEEE (2021)
16. Lin, E., Crane, B.: The management and imaging of vestibular schwannomas. *American Journal of Neuroradiology* **38**(11), 2034–2043 (2017)
17. Liu, H., Fan, Y., Cui, C., Su, D., McNeil, A., Dawant, B.M.: Unsupervised domain adaptation for vestibular schwannoma and cochlea segmentation via semi-supervised learning and label fusion. In: *International MICCAI Brainlesion Workshop*. pp. 529–539. Springer (2022)
18. Long, M., Cao, Y., Wang, J., Jordan, M.: Learning transferable features with deep adaptation networks. In: *International conference on machine learning*. pp. 97–105. PMLR (2015)
19. Lu, F., Wu, F., Hu, P., Peng, Z., Kong, D.: Automatic 3d liver location and segmentation via convolutional neural network and graph cut. *International journal of computer assisted radiology and surgery* **12**(2), 171–182 (2017)
20. Nguyen, D., de Kantzow, L.: Vestibular schwannomas: a review. *Appl Radiol* **48**(3), 22–27 (2019)

21. Park, T., Efros, A.A., Zhang, R., Zhu, J.Y.: Contrastive learning for unpaired image-to-image translation. In: European conference on computer vision. pp. 319–345. Springer (2020)
22. Shapey, J., Kujawa, A., Dorent, R., Wang, G., Bisdas, S., Dimitriadis, A., Grishchuk, D., Paddick, I., Kitchen, N., Bradford, R., et al.: Segmentation of vestibular schwannoma from magnetic resonance imaging: An open annotated dataset and baseline algorithm. The Cancer Imaging Archive (2021)
23. Shapey, J., Kujawa, A., Dorent, R., Wang, G., Dimitriadis, A., Grishchuk, D., Paddick, I., Kitchen, N., Bradford, R., Saeed, S.R., et al.: Segmentation of vestibular schwannoma from mri, an open annotated dataset and baseline algorithm. *Scientific Data* **8**(1), 1–6 (2021)
24. Shapey, J., Wang, G., Dorent, R., Dimitriadis, A., Li, W., Paddick, I., Kitchen, N., Bisdas, S., Saeed, S.R., Ourselin, S., et al.: An artificial intelligence framework for automatic segmentation and volumetry of vestibular schwannomas from contrast-enhanced t1-weighted and high-resolution t2-weighted mri. *Journal of neurosurgery* **134**(1), 171–179 (2019)
25. Shin, H., Kim, H., Kim, S., Jun, Y., Eo, T., Hwang, D.: Cosmos: Cross-modality unsupervised domain adaptation for 3d medical image segmentation based on target-aware domain translation and iterative self-training. *arXiv preprint arXiv:2203.16557* (2022)
26. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Cardoso, M.J.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Deep learning in medical image analysis and multimodal learning for clinical decision support, pp. 240–248. Springer (2017)
27. Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J.N., Wu, Z., Ding, X.: Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Medical Image Analysis* **63**, 101693 (2020)
28. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems* **30** (2017)
29. Tzeng, E., Hoffman, J., Zhang, N., Saenko, K., Darrell, T.: Deep domain confusion: Maximizing for domain invariance. *arXiv preprint arXiv:1412.3474* (2014)
30. Wang, L., Wang, M., Zhang, D., Fu, H.: Unsupervised domain adaptation via style-aware self-intermediate domain. *arXiv preprint arXiv:2209.01870* (2022)
31. Zhang, Y., Miao, S., Mansi, T., Liao, R.: Task driven generative modeling for unsupervised domain adaptation: Application to x-ray image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 599–607. Springer (2018)
32. Zhao, Z., Hu, J., Zeng, Z., Yang, X., Qian, P., Veeravalli, B., Guan, C.: Mmgl: Multi-scale multi-view global-local contrastive learning for semi-supervised cardiac image segmentation. In: 2022 IEEE International Conference on Image Processing (ICIP). pp. 401–405 (2022)
33. Zhao, Z., Xu, K., Li, S., Zeng, Z., Guan, C.: Mt-uda: Towards unsupervised cross-modality medical image segmentation with limited source labels. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 293–303. Springer (2021)
34. Zhao, Z., Zhou, F., Xu, K., Zeng, Z., Guan, C., Kevin Zhou, S.: Le-uda: Label-efficient unsupervised domain adaptation for medical image segmentation. *IEEE Transactions on Medical Imaging* (2022)

35. Zhao, Z., Zhou, F., Zeng, Z., Guan, C., Zhou, S.K.: Meta-hallucinator: Towards few-shot cross-modality cardiac image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 128–139. Springer (2022)
36. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)
37. Zou, Y., Yu, Z., Kumar, B., Wang, J.: Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In: Proceedings of the European conference on computer vision (ECCV). pp. 289–305 (2018)