MS-MT++: Enhanced Multi-Scale Mean Teacher for Cross-Modality Vestibular Schwannoma and Cochlea Segmentation

Ziyuan Zhao[1,2], Ruikai Lin[1,3], Kaixin Xu[1], Xulei Yang[1], and Cuntai Guan[2]

[1] Institute for Infocomm Research (I^{2}R), A*STAR, Singapore
[2] Nanyang Technological University, Singapore
[3] National University of Singapore, Singapore

Abstract. Domain shift has been a long-standing issue for medical image segmentation. Unsupervised domain adaptation (UDA) methods have recently achieved promising cross-modality segmentation performance by distilling knowledge from a label-rich source domain to a target domain without labels. Different from CrossMoDA 2022, the challenge of 2023 includes highly heterogeneous MRI scans from more institutions and various scanners and subdivides the segmentation object into three key brain structures (intra/extra-vestibular schwannoma and cochlea), increasing the difficulty of domain adaptation. In this work, we improve our previous method and propose an enhanced multi-scale self-ensembling-based UDA framework for automatic segmentation of Vestibular Schwannoma and Cochlea on high-resolution T2 images. Our method demonstrates a mean Dice score of 0.669 and 0.766 for the extra/intra-VS joint area and Cochlea respectively, securing a top 5 finish in the CrossMoDA 2023 challenge.

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

1 Introduction

Medical image segmentation plays a pivotal role in medical image analysis, providing crucial information for diagnostic analysis and treatment planning [6]. Accurate segmentation and measurement of Vestibular Schwannoma (VS) and Cochlea from MRI are instrumental in assisting VS treatment planning and streamlining clinical workflows [25]. Recently, machine learning and deep learning [11] have demonstrated remarkable success across various domains [24, 32, 3, 33, 23, 19, 31, 9]. Consequently, researchers have increasingly employed deep learning techniques for autonomously segmenting VS and Cochlea [5, 14, 15]. In the case of Vestibular Schwannoma, the tumor area is anatomically classified into intra- and extra-meatal regions based on their location inside or outside the inner ear canal. Reporting guidelines underscore the importance of differentiating between intra- and extra-meatal regions when presenting results. Notably, the largest extra-meatal diameter serves as a key metric for indicating tumor size.

In addition, metrics such as size and volume derived from the extra-meatal region are crucial radiomic features for evaluating Vestibular Schwannoma (VS) growth. In this regard, the CrossMoDA 2023 dataset introduces more detailed annotations of intra- and extra-meatal VS to facilitate more accurate VS segmentation compared to the previous challenge [5]. While Contrast-enhanced T1 (ceT1) MR imaging is commonly used, there is a growing interest in non-contrast sequences, such as high-resolution T2 (hrT2) imaging. Recent findings suggest that high-resolution T2 (hrT2) MRI could be a safer and more cost-efficient alternative to contrast-enhanced T1 (ceT1) MRI. However, the significant domain shift between MRI images with different contrasts, along with the expensive and laborious process of re-annotating medical image scans in another modality, makes it challenging for deep learning models to generalize effectively across both domains. In this regard, unsupervised domain adaptation (UDA) [7, 1, 36] has emerged as a promising approach in medical imaging to enhance the robustness of deep learning algorithms on complex scenarios from multiple perspectives, such as image adaptation [41, 7, 18], feature adaptation [16, 29, 35] and their mixtures [2, 39, 40], enabling them to adapt to diverse input data domains and extend their utility in various clinical settings. Therefore, we are encouraged to perform unsupervised domain adaptation (UDA) and conduct intra-VS, extra-VS, and Cochlea segmentation in the hrT2 domain by leveraging both labeled ceT1 scans and unlabeled hrT2 scans.

In this work, we improve our previous method [37] and propose an effective cross-modality unsupervised domain adaptation (UDA) framework, called MS-MT++. In our framework, we first translate contrast-enhanced T1 (ceT1) scans to high-resolution T2 (hrT2) modality via a Segmentation-Enhanced Contrastive Unpaired Translation (SE-CUT) network [22]. Subsequently, three CycleGAN networks are employed for pixel-level intensity fine-tuning. We also perform intensity augmentation on the annotated regions of generated images. To generate pseudo labels for unlabeled real hrT2 scans, we apply a 3D full-resolution nnU-Net [8]. Finally, we build a multi-scale mean-teacher (MS-MT) network [27, 12] for improving the cross-modality segmentation performance. The experimental results demonstrate the effectiveness of the proposed UDA method in reducing the domain gap between different modalities, achieving promising segmentation performance on ceT2 scans.

2 Methods

Given an unpaired dataset of two modalities, *i.e.,* annotated contrast-enhanced T1 (ceT1) MRI images $\mathcal{D}_s = \{(\mathbf{x}_i^s, y_i^s)\}_{i=1}^N$ and non-annotated high-resolution T2 (hrT2) MRI scans $\mathcal{D}_t = \{(\mathbf{x}_i^t)\}_{i=1}^M$, both sharing the same classes (intra- and extra-meatal VS, and Cochlea), we aim to exploit $\mathcal{D}_s$ and $\mathcal{D}_t$ for unsupervised domain adaptation, aiming to enhance the cross-modality segmentation perfor-

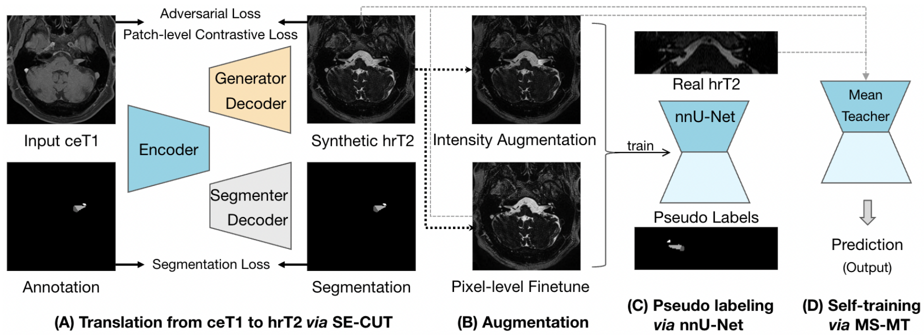

Fig. 1. The workflow of our proposed method. First, the SE-CUT network is proposed to translate ceT1 to hrT2. Then, the target areas are augmented with intensity, and multiple CycleGANs are used for pixel-level enhancement. After that, all synthetic and augmented scans are used for training a 3D nnU-Net, which can generate pseudo labels for all unlabeled real hrT2 images. Finally, an MS-MT network is employed for multi-scale self-ensemble learning.

mance of the VS and Cochlea on hrT2 MRI images. The overview of our UDA framework is illustrated in Fig. 1.

2.1 Segmentation-enhanced translation

To mitigate the domain gap across modalities, we conduct image-level domain adaptation to generate synthetic target samples. This involves training a model on synthetic target images for VS and Cochlea segmentation in real high-resolution T2 (hrT2) scans. For effective image-to-image translation, we leverage the Contrastive Unpaired Translation (CUT) [22] method for time efficiency. To preserve structural information during translation (see Fig. 1(A)), we enhance the 2D CUT with an additional segmentation decoder. The ResNet-based generator (generation decoder) transforms source domain images to the target domain, while a PatchGAN discriminator distinguishes real from generated images [22]. Inspired by the SIFA architecture design [2], we connect two layers of the encoder with the segmenter decoder to produce multi-level segmentation predictions. This segmentation loss guides the encoder to focus on relevant areas, thereby preserving the structure details of Vestibular Schwannoma (VS) and Cochlea in the translated images.

2.2 Intensity augmentation and pseudo labeling

Considering the heterogeneous signal intensity of tumors [20] and the hyper-intense signal intensity of cochleas [13] in T2-weighted imaging, we perform intensity augmentation and generate augmented data for diversifying the training distributions. Using the generated hrT2 images and corresponding ground truth annotations, the signal intensities of annotated regions in each scan are

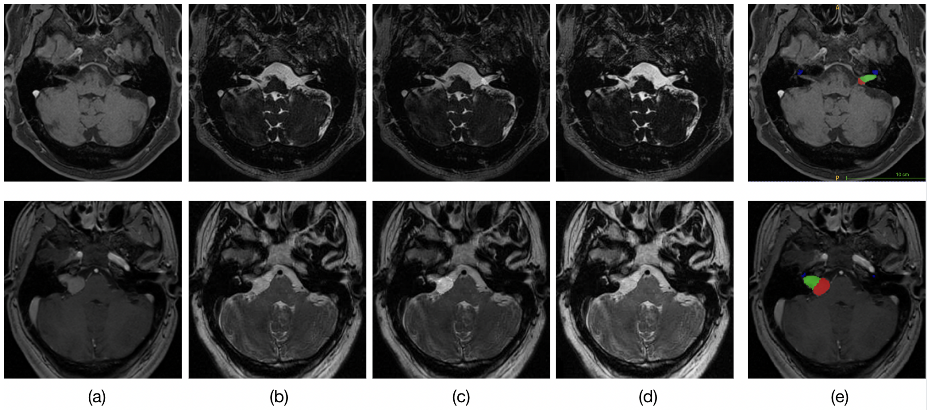

Fig. 2. Visual comparison for (a) Original ceT1, (b) Generated hrT2 using SE-CUT, (c) Generated hrT2 after intensity augmentation, (d) Generated hrT2 after pixel-level fine-tuning via CycleGAN, and (e) Ground truth of VS and cochlea.

randomly adjusted by a factor of 50% or 25% for mute and intensify respectively, effectively doubling the training data. Simultaneously, we employ multiple CycleGANs [41] on different sources of data for further fine-tuning at the pixel level of synthetic hrT2, improving the data utilization and training efficiency, which also augments the data by another copy. On the other hand, to enhance the segmentation performance on real hrT2 images, we utilize a Pseudo-Labeling (PL) strategy to leverage unlabeled hrT2 images by generating pseudo hrT2 annotations. We train a 3D full-resolution nnU-Net [8] using synthetic hrT2 images and augmented images. These trained models are then used to generate pseudo-labels for the unlabeled hrT2 images, further improving segmentation accuracy.

2.3 Multi-scale self-ensembling learning

To maximize the utilization of available data, we propose using the self-ensembling network, mean teacher (MT) [27], where a teacher model is constructed with the same architecture as the student model and updated using an exponential moving average (EMA) of the student's parameters during training. We ensure consistency between student and teacher outputs by minimizing their difference with the mean square error (MSE) loss. Additionally, we leverage the success of multi-scale learning in medical image analysis [38, 12, 34] and other fields [21, 28], and introduce a multi-scale mean teacher (MS-MT) network following [12]. This approach utilizes multi-scale predictions for deep supervision and consistency regularization. Both the teacher and student networks utilize the 3D full-resolution nnU-Net [8] as the backbone, with auxiliary layers connected to each block of the last five blocks to obtain multi-scale predictions. This combination allows us to make the most use of available data, leading to enhanced cross-modality segmentation performance.

3 Experiments and Results

3.1 Dataset

The CrossMoDA 2023 challenge provides a highly heterogeneous dataset, sourced from multiple centers, comprising 227 annotated contrast-enhanced T1 (ceT1) scans and 391 unlabeled high-resolution T2 (hrT2) scans (295 scans are used for training and 96 scans are used as the validation set) [5, 10, 30]. The London (LDN) and Tilburg SC-GK (ETZ) data are obtained using different scanners and imaging sequences, encompassing both T1-weighted and high-resolution T2-weighted imaging. The UK MC-RC (UKM) data are obtained from various scanners with different magnetic field strengths and slice thickness, in which, voxel volume and intensities vary significantly across all ceT1 weighted and T2 weighted imaging. Besides, 341 unpublished hrT2 scans are used as the external test set during the testing phase.

3.2 Data preprocessing

Due to variations in voxel spacing in the raw scans, all images were resampled into a common spacing of $0.6 \times 0.6 \times 1.0$ mm and the intensity was normalized to the range [0, 1] using Min-Max scaling. To delineate the regions of interest (ROI) and remove the noises, the images were cropped into 256×256 pixels in the xy-plane using a 75-percentile binary threshold [4], resulting in $256 \times 256 \times$ N image volumes for 3D nnU-Net training. These processed 3D volumes were also sliced along the z-axis to create N 2D images for SE-CUT and CycleGAN training. The Dice Score (DSC [%]) [26] and the Average Symmetric Surface Distance(ASSD [voxel]) [17] are used to assess the model performance on VS and Cochlea segmentation.

3.3 Implementation details

We used a single NVIDIA A40 GPU with 48GB of memory for model training. Fig. 2 shows the effects of using SE-CUT to translate (a) original ceT1 into (b) synthetic hrT2 after training for 120 epochs. In the SE-CUT, We followed [22] and [2] to keep the same weights for adversarial loss, contrastive loss, and segmentation loss; the loss weights for additional segmentation were set to 1 and 0.1 for the last layer and the second last downsampling layer, respectively. The segmentation-enhanced architecture effectively retained structural information for Vestibular Schwannoma (VS) and cochlea in the original ceT1 scans, as demonstrated in Figures 2 (c) and (d). These images represent synthetic hrT2 results after intensity augmentation and pixel-level fine-tuning, respectively. To improve pixel intensity similarity to real hrT2 scans, three CycleGANs were individually trained for 50 epochs on scans from different sources (ETZ, LDN, and UKM). The resulting datasets were collectively used to train a 3D full-resolution nnU-Net with generalization ability for hrT2. After 400 epochs of training and 5-fold cross-validation, the nnU-Net generated pseudo labels for all

Table 1. Performance of our model during validation and testing phases.

	Cochlea		VS		extra-VS		intra-VS	
	Dice	ASSD	Dice	ASSD	Dice	ASSD	Dice	ASSD
Validation Phase	0.729	19.836	0.568	32.409	0.620	40.384	0.457	40.364
Testing Phase	0.766	11.026	0.669	10.378	0.676	13.874	0.563	18.607

unlabeled real hrT2 scans. Subsequently, a self-training process using synthetic hrT2 with real labels and real hrT2 with pseudo labels was conducted. The multi-scale mean teacher (MS-MT) network, with a backbone of two 3D nnU-Nets, was trained using stochastic gradient descent for 300 epochs. The initial learning rate was set to 0.01, and the objective function employed a combination of Dice and cross-entropy losses. The deep supervision scheme in nnU-Net [8] was enabled. In the MS-MT network, the exponential moving average (EMA) update α was set to 0.9, and the loss weights for consistency regularization were assigned as $\{0.05, 0.05, 0.05, 0.4, 0.5\}$ in ascending order of feature map size.

3.4 Experimental results

Following the preprocessing step, the validation set was evaluated using the 5-fold ensemble MS-MT model, as depicted in Fig. 3 and summarized in Table 1. Notably, the Dice scores for Vestibular Schwannoma (VS) and cochlea in the validation phase leaderboard were 0.729 and 0.568, respectively. Finally, the largest connected component (LCC) was calculated, and the segmentation results were post-processed (*i.e.*, the first LCC was preserved for the connected intra and extra VS while the first and second LCCs were preserved for the cochlea) for the final submission during the testing phase. The segmentation performance of the final model on 341 cases of unpublished real high-resolution T2 (hrT2) scans during the testing phase is detailed in Table 1. In the region segmentation of Cochlea, extra-VS, and intra-VS, our method obtained dice scores of 0.766, 0.676, and 0.563 respectively, securing the fifth place. The superior performance during the testing phase further confirms the effectiveness of the proposed framework. Fig. 3 provides some qualitative results produced by our method.

4 Conclusion

In this study, we propose a comprehensive four-stage cross-modality unsupervised domain adaptation workflow, designed to seamlessly bridge the gap between annotated ceT1 and unlabeled hrT2 MRI scans. Our approach comprises unpaired image translation, intensity augmentation, and pixel-level fine-tuning, followed by pseudo labeling and multi-scale self-training. By leveraging only annotated ceT1 scans, we successfully trained a final model capable of segmenting intra-meatal Vestibular Schwannoma (VS), extra-meatal VS, and cochlea in hrT2 scans. Our proposed method demonstrates its efficacy by achieving a top-5

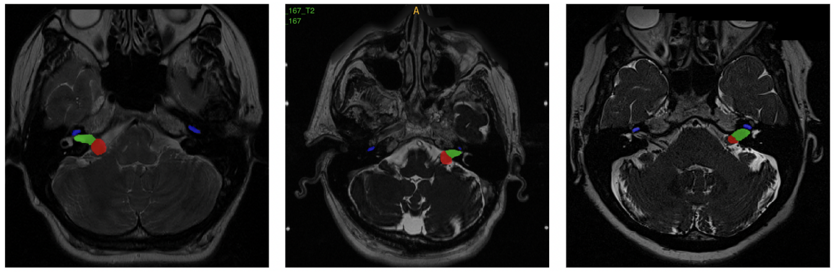

Fig. 3. Segmentation results of the validation set produced by our workflow. The intra-VS, extra-VS, and cochlea are indicated in red, green, and blue color, respectively.

finish in the testing phase of the CrossMoDA 2023 challenge, highlighting the model's ability to effectively bridge the gap between ceT1 and hrT2 MRI modalities and showcasing its potential in enhancing the segmentation of intra-meatal VS, extra-meatal VS, and cochlea.

References

1. Chen, C., Dou, Q., Chen, H., Qin, J., Heng, P.A.: Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation. In: Proceedings of the AAAI conference on artificial intelligence. vol. 33, pp. 865–872 (2019)
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. Chen, L., Lu, Y., Wu, C.T., Clarke, R., Yu, G., Van Eyk, J.E., Herrington, D.M., Wang, Y.: Data-driven detection of subtype-specific differentially expressed genes. Scientific reports **11**(1), 332 (2021)
4. Choi, J.W.: Using out-of-the-box frameworks for unpaired image translation and image segmentation for the crossmoda challenge. arXiv e-prints pp. arXiv–2110 (2021)
5. Dorent, R., Kujawa, A., Ivory, M., Bakas, S., Rieke, N., Joutard, S., Glocker, B., Cardoso, J., Modat, M., Batmanghelich, K., et al.: Crossmoda 2021 challenge: Benchmark of cross-modality domain adaptation techniques for vestibular schwnannoma and cochlea segmentation. arXiv preprint arXiv:2201.02831 (2022)
6. 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)
7. 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)
8. 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)

9. Jiang, C., Hui, B., Liu, B., Yan, D.: Successfully applying lottery ticket hypothesis to diffusion model. arXiv preprint arXiv:2310.18823 (2023)
10. Kujawa, A., Dorent, R., Connor, S., Thomson, S., Ivory, M., Vahedi, A., Guilhem, E., Bradford, R., Kitchen, N.D., Bisdas, S., Ourselin, S., Vercauteren, T.K.M., Shapey, J.: Deep learning for automatic segmentation of vestibular schwannoma: A retrospective study from multi-centre routine mri. In: medRxiv (2022)
11. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. nature **521**(7553), 436–444 (2015)
12. 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)
13. Lin, E., Crane, B.: The management and imaging of vestibular schwannomas. American Journal of Neuroradiology **38**(11), 2034–2043 (2017)
14. 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 (2021)
15. Liu, H., Fan, Y., Oguz, I., Dawant, B.M.: Enhancing data diversity for self-training based unsupervised cross-modality vestibular schwannoma and cochlea segmentation. arXiv preprint arXiv:2209.11879 (2022)
16. 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)
17. 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)
18. Lu, Y., Wang, H., Wei, W.: Machine learning for synthetic data generation: a review. arXiv preprint arXiv:2302.04062 (2023)
19. Murungi, N.K., Pham, M.V., Dai, X., Qu, X.: Trends in machine learning and electroencephalogram (eeg): A review for undergraduate researchers. arXiv preprint arXiv:2307.02819 (2023)
20. Nguyen, D., de Kanztow, L.: Vestibular schwannomas: a review. Appl Radiol **48**(3), 22–27 (2019)
21. Pahwa, R.S., Chang, R., Jie, W., Ziyuan, Z., Lile, C., Xun, X., Sheng, F.C., Choong, C.S., Rao, V.S.: 3d defect detection and metrology of hbms using semi-supervised deep learning. In: 2023 IEEE 73rd Electronic Components and Technology Conference (ECTC). pp. 943–950. IEEE (2023)
22. 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)
23. Qiu, Y., Zhao, Z., Yao, H., Chen, D., Wang, Z.: Modal-aware visual prompting for incomplete multi-modal brain tumor segmentation. In: Proceedings of the 31st ACM International Conference on Multimedia. pp. 3228–3239 (2023)
24. Qu, X., Liu, P., Li, Z., Hickey, T.: Multi-class time continuity voting for eeg classification. In: Brain Function Assessment in Learning: Second International Conference, BFAL 2020, Heraklion, Crete, Greece, October 9–11, 2020, Proceedings 2. pp. 24–33. Springer (2020)
25. 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)
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. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. arXiv preprint arXiv:1703.01780 (2017)
28. Wang, J., Chang, R., Zhao, Z., Pahwa, R.S.: Robust detection, segmentation, and metrology of high bandwidth memory 3d scans using an improved semi-supervised deep learning approach. Sensors **23**(12), 5470 (2023)
29. Wang, L., Wang, M., Zhang, D., Fu, H.: Unsupervised domain adaptation via style-aware self-intermediate domain. arXiv preprint arXiv:2209.01870 (2022)
30. Wijethilake, N., Kujawa, A., Dorent, R., Asad, M.H., Oviedova, A., Vercauteren, T., Shapey, J.: Boundary distance loss for intra-/extra-meatal segmentation of vestibular schwannoma. In: MLCN@MICCAI (2022)
31. Wu, J., Ye, X., Mou, C., Dai, W.: Fineehr: Refine clinical note representations to improve mortality prediction. In: 2023 11th International Symposium on Digital Forensics and Security (ISDFS). pp. 1–6. IEEE (2023)
32. Zeng, Z., Zhao, W., Qian, P., Zhou, Y., Zhao, Z., Chen, C., Guan, C.: Robust traffic prediction from spatial–temporal data based on conditional distribution learning. IEEE Transactions on Cybernetics **52**(12), 13458–13471 (2021)
33. Zhang, Z., Tian, R., Ding, Z.: Trep: Transformer-based evidential prediction for pedestrian intention with uncertainty. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37 (2023)
34. 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. arXiv preprint arXiv:2207.01883 (2022)
35. Zhao, Z., Qian, P., Yang, X., Zeng, Z., Guan, C., Tam, W.L., Li, X.: Semignn-ppi: Self-ensembling multi-graph neural network for efficient and generalizable protein–protein interaction prediction. In: Elkind, E. (ed.) Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23. pp. 4984–4992. International Joint Conferences on Artificial Intelligence Organization (8 2023). https://doi.org/10.24963/ijcai.2023/554, https://doi.org/10.24963/ijcai.2023/554, main Track
36. Zhao, Z., Xu, K., Li, S., Zeng, Z., Guan, C.: Mt-uda: Towards unsupervised cross-modality medical image segmentation with limited source labels. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24. pp. 293–303. Springer (2021)
37. Zhao, Z., Xu, K., Yeo, H.Z., Yang, X., Guan, C.: Ms-mt: Multi-scale mean teacher with contrastive unpaired translation for cross-modality vestibular schwannoma and cochlea segmentation. arXiv preprint arXiv:2303.15826 (2023)
38. Zhao, Z., Zeng, Z., Xu, K., Chen, C., Guan, C.: Dsal: Deeply supervised active learning from strong and weak labelers for biomedical image segmentation. IEEE Journal of Biomedical and Health Informatics **25**(10), 3744–3751 (2021)
39. Zhao, Z., Zhou, F., Xu, K., Zeng, Z., Guan, C., Zhou, S.K.: Le-uda: Label-efficient unsupervised domain adaptation for medical image segmentation. IEEE Transactions on Medical Imaging **42**(3), 633–646 (2022)

40. 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)
41. 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)