A GANs-based Modality Fusion and Data Augmentation for CrossMoDA Challenge

Jianghao Wu[1], Ran Gu[1], Shuwei Zhai[1], Wenhui Lei[1], and Guotai Wang[1](✉)

School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, China.
guotai.wang@uestc.edu.cn

Abstract. In this work, we implement a deep learning-based segmentation algorithm that can automatically segment the tumour and the cochlea for a better treatment planning of vestibular schwannoma (VS). The main challenge of the task is cross-modality unsupervised domain adaptation. Because the domain gap between the two modality is huge, we adapt cycleGAN to convert ceT1 images to fake hrT2 images. And we use the fake hrT2 images and the corresponding labels to train a segmentation model by PyMIC. After we get the probability maps from PYMIC, we use the simple CRF to refine the maps as pseudo labels. Finally, we use an algorithm to select the high-quality pseudo labels to train a new segmentation model based on fake hrT2 images. Our proposed framework achieves an average Dice score of 0.72 for the tumour and 0.71 for cochlea on the validation set and achieving an average Dice score of 0.72 on this challenge.

Keywords: Domain Adaption · CycleGAN · Vestibular Schwannoma.

1 Introduction

The incidence of vestibular schwannomas has increased significantly in recent years, and is now estimated to be 14 to 20 cases per million people in one year, and has increased significantly in recent years [1]. CT can cause missed tumor detection, but MRI has a good imaging effect on vestibular schwannomas, which is convenient to segment the tumor and measure the change in tumor volume. In the past, tumor segmentation was mostly performed on T1 images. Recent research shows that T2 images have better overall performance. Therefore, it is valuable to use labeled T1 images to segment tumour on T2 images. Many works are dedicated to proposing Domain Adaptation (DA) methods to diminish this domain shift [6–8].

This challenge aims to use T1 images and corresponding labels to automatically segment VS tumors and cochlea in unlabeled high-resolution T2 MRI. This will save a lot of labor cost, and can greatly speed up the clinical workflow and enable patients to monitor and image in a safer situation. However, the challenge is full of difficulties. First, there exists domain shift between T1 and T2 images, and the labeled T1 image has higher contrast than unlabeled T2 image, which brings more difficult for the tumor prediction on T2 image. DA has recently raised strong interests in the medical imaging analysis [10, 9]. In this challenge, it is urgent to solve the domain shift problem across the domain by means of the domain adaptation strategy.

In order to solve this challenge, we first transfer the image's style through CycleGAN and CUT, which are powerful and efficient image-to-image transformation networks used in image generation [14, 15]. Meanwhile, the organizer provides T1 labels for training. Therefore, we use CycleGAN and CUT to transfer the ceT1 image to the fake hrT2 image, and then convert the fake hrT2 images back to ceT1 through cycleGAN. Through this strategy, the number of training sample will be doubled. Then, we train a 2.5D segmentation network on the PyMIC platform. Finally, we transfer T2 images to fake T1 and feed into the trained segmentation network for prediction.Then we use simpleCRF to post-process the predicted mask to get a more accurate mask, and filter out the false positive mask through the algorithm, and finally get a high-quality mask. We then use these high-quality masks as labels to retrain a new segmentation model on the T2 modal.

The main procedure of our work is summarized as:

1). Use a unified algorithm to crop the ceT1 image and hrT2 image to a suitable size, which containing vestibular schwannoma and cochlea.

2). Transfer ceT1 to fake hrT2, and then transfer them back to ceT1 by cycleGAN and CUT. First, we extracted the 3D volumes into 2D slices and use cropped images to train cycleGAN and CUT to get the fake hrT2 images. Then, we transformed the fake hrT2 back to ceT1 style and obtained the fake ceT1. Finally, we concatenated the 2D fake ceT1 images as 3D volumes to increase the training samples.

3). We jointly train the segmentation model on PYMIC with original ceT1 images as well as the fake ceT1 images, which we use 2.5D unet as our backbone network.

4). Use cycleGAN to convert the hrT2 images of the training set into fake ceT1 images and input them into the segmentation model, in which we can get pseudo-labels with uneven quality.

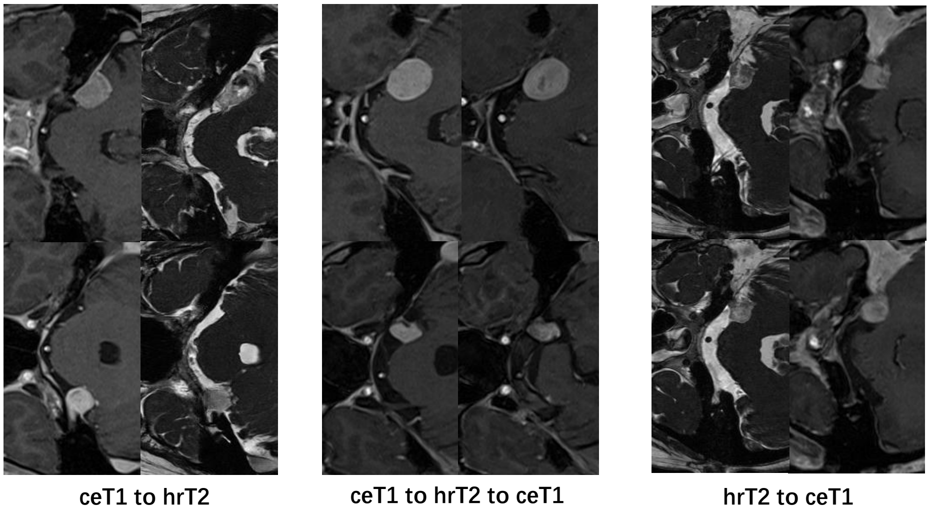
5). We use simpleCRF to post-process the predicted mask to get a more accurate mask, and filter out the false positive mask through the algorithm. Finally, we get a set of high-quality pseudo labels. We then use these pseudo labels to retrain a new segmentation model on the T2 modal.

2 Method

2.1 Modality Fusion

We set the ceT1 as the source domain and hrT2 as the target domain. Due to the big difference between the source domain and the target domain, the segmentation network trained directly using the source domain and performs extremely degrade on T2. First, we translate the source data to the target domain like, and then translate the fake target data back to source domain. Because the source domain is labeled, we convert the hrT2 image to the ceT1 image for final prediction. We first count the location information of the VS and the cochlea in the ceT1 label. These location information will help us to narrow the target area so as to eliminate the interference of other useless information. After getting the location information, we enlarge the boundary appropriately making it fit to the cycleGAN. Then, we crop the 3D images of ceT1 and hrT2 to appropriate size, in which most of the images are cropped as (40, 160, 272), and some images are cropped as (30, 160, 272) depending on the size of the spacing.

After finishing the cropping work, we perform image generation operations. CycleGAN and CUT are widely used for image generation, but the input objects of these two methods are 2D images. So we must slice a 3D image into multiple 2D images. We cut the 3D image along the depth direction. For example, for a 3D image with a size of (40,160,272), after slicing, we can get 40 images with a size of (160, 272) images. After slicing all the images in the training set, we put the T1 image and the T2 image in trainA and trainB respectively. And use the trainA and trainB data sets to train two different GAN networks, CycleGAN and CUT. We can convert the T1 image to T2 image and return the fake T2 to T1, which is beneficial to the training of the segmentation model because it will get more data for training.



ceT1 to hrT2

ceT1 to hrT2 to ceT1

hrT2 to ceT1

Fig. 1. Visual results of image translation using CycleGAN

2.2 Segmentation Model

After we train CycleGAN and CUT models, we convert the ceT1 image back to ceT1. Then, we train an excellent segmentation model on the PyMIC platform, and use the prediction mask of the T2 image as a pseudo-label to further train a segmentation model. The training process is shown in Figure 2. The training data set consists of the original ceT1, ceT1′ and ceT1″, where ceT1′ and ceT1″ are generated from CycleGAN and CUT respectively, and their labels are all annotations provided by the organizer. We train a 2.5D U-Net as a segmentation model, and consider the balance of computational performance and computational cost. In our training, we use data augmentation such as random cropping and random flipping.

We convert T2 images into T1 images through cycleGAN, and use the above model to predict the tumor and cochlea, but these predictions are inaccurate. We call these predictions pseudo-labels. But the quality of these pseudo-labels is very poor, there are insufficient predictions, as well as all-empty, and false-positive predictions. We perform a post-processing operation on the pseudo-labels, and use simpleCRF to perform CRF operations on the obtained pseudo-labels to improve the quality of the pseudo-labels. After that, we write an algorithm to filter out the pseudo-labels with higher precision from the pseudo-labels obtained in the above steps. The algorithm can choose a non-zero prediction mask as the pseudo-label of T2, and screen out the false-positive pseudo-labels, after which we will get a certain number of high-quality pseudo-labels. And use these high-quality pseudo-labels and their corresponding T2 images to further train the segmentation model, which we can call the T2 model. The training settings are the same as the training on ceT1. Finally, we use the second training segmentation model to directly predict the hrT2 image of the validation dataset.

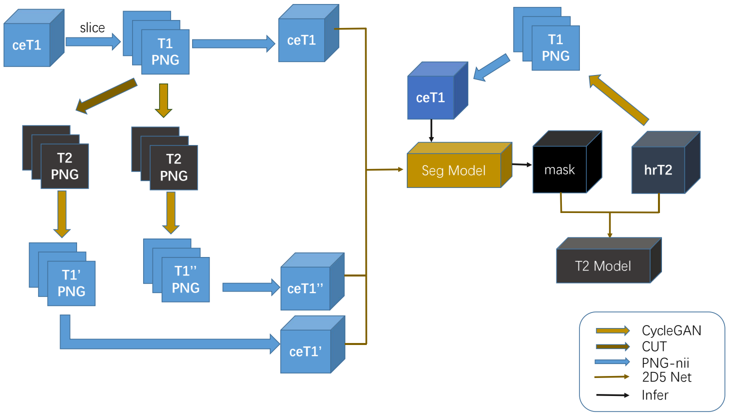


Fig. 2. The pipeline of training a segmentation model.

3 Experiments and Results

3.1 Datasets and Implementations

The organizer offered 105 ceT1 with their corresponding labels and 105 unlabeled hrT2 images for training, and additionally provided 32 hrT2 images for validation. While at the test, they will finally evaluate the model adaptation and ability of transferring knowledge by 100 hrT2 images. All images were obtained on a 32-channel Siemens Avanto 1.5T scanner using a Siemens single-channel head coil. The Contrast-enhanced T1-weighted imaging was performed with an MPRAGE sequence with in-plane resolution of 0.4×0.4 mm, in-plane matrix of 512×512, and slice thickness of 1.0 to 1.5 mm, and the High-resolution T2-weighted imaging was performed with a 3D CISS or FIESTA sequence in-plane resolution of 0.5×0.5 mm, in-plane matrix of 384×384 or 448×448, and slice thickness of 1.0 to 1.5 mm.

For our data prepossessing, because the size of data is different in depth wise, so we crop the hrT2 3D images to a new size. First, if the depth is under 40, We keep the pixel depth area from 5 to depth minus 5. And if the depth is more than 40, we take the depth information into considering. If the spacing is 1.0, we crop the depth in a range, and if the spacing is 1.5, we crop the depth in another range. The wide is cropped in the range of 120 to 392, and the height is in the range of 205 to 365.

3.2 Network training

We implemented the segmentation network by PyMIC, a PyTorch library provided for medical image segmentation[1] [16], and trained the model on Ubuntu with an NVIDIA GeForce RTX 2080 Ti with 11 GB RAM. For segmentation model training, the input size of data is (32,128,128) with the input channel of 1. The dropout rate at each level of the model is set to (0.0, 0.0, 0.3, 0.4, 0.5). Adam is select to optimize the model parameters. We use Dice loss for training with the learning rate of le-3, and it will decay by half in each iteration of 10000. We save the checkpoint models in each 10000 iterations. We train the segmentation model with 40000 iterations.

3.3 Experimental Results

We compare several datasets for the segmentation model training, adapting 1) only source T1 dataset, 2) the T1 and T1', 3) the T1, T1' and T1" datasets. The experimental results can be seen in the Table 1. When we only use the T1 data as the training dataset, the result of cochlea is the worst, only gets 0.1904 of the dice score. While the T1+T1'+T1" achieves the best Performance, and it's average dice score is 0.6561. After we further trained with the T2 pseudo labels, the average dice score improves dramatically, which get the highest score

[1] https://github.com/HiLab-git/PyMIC

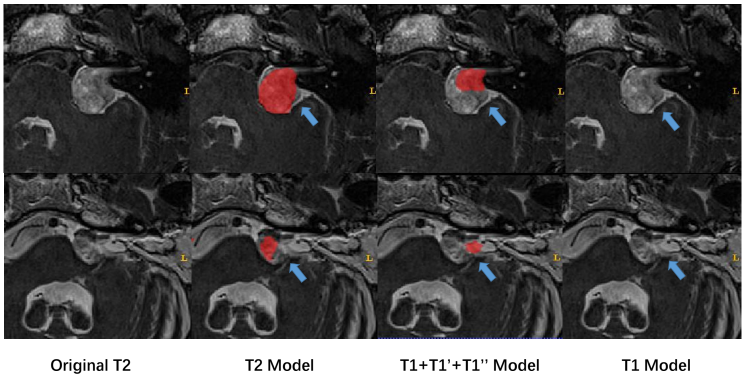


Fig. 3. The Result of Different Segmentation Model.

in both tumour and cochlea. The average dice score of T2 model gets at 0.7242 with the standard deviation at 0.1421. The visual segmentation results can be seen in the Fig. 3. We can see that the T2 model achieved the best segmentation results compared with the other two methods.

Table 1. Average Dice scores on test data set.

Dateset	Dice		
	Tumour	Cochlea	Average
T1	0.5433	0.1904	0.3652
T1+T1'	0.4123	0.5604	0.4863
T1+T1'+T1"	0.6573	0.6548	0.6561
T2	**0.7342**	**0.7142**	**0.7242**

4 Conclusion

In this paper, we propose a multiple generation data for CrossMoDA framework to segment two key brain structures involved in the follow-up and treatment planning of vestibular schwannoma (VS): the tumour and the cochlea. We use tow GANs to generate T1 and T2 images, and train a segmentation model by all of them.

After we trained the segmentation model, we infer the mask of T2. And we use simpleCRF to post-process the predicted mask to get a more accurate mask, and filter out the false positive mask through the algorithm, and finally get a high-quality mask. We then use these high-quality masks as labels to retrain a new segmentation model on the T2 modality. It makes great progress in the result of valid dataset.

References

1. Shapey, Jonathan, et al. ”Segmentation of vestibular schwannoma from MRI—An open annotated dataset and baseline algorithm.” medRxiv (2021).
2. Shapey, Jonathan, 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 (2019): 171-179.
3. Wang, Guotai, et al. ”Automatic segmentation of vestibular schwannoma from T2-weighted MRI by deep spatial attention with hardness-weighted loss.” International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019.
4. Dorent, Reuben, et al. ”Scribble-based Domain Adaptation via Co-segmentation.” International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2020.
5. Stangerup, Sven-Eric, and Per Caye-Thomasen. ”Epidemiology and natural history of vestibular schwannomas.” Otolaryngologic Clinics of North America 45.2 (2012): 257-268.
6. Ganin, Yaroslav, et al. ”Domain-adversarial training of neural networks.” The journal of machine learning research 17.1 (2016): 2096-2030.
7. Zou, Yang, et al. ”Unsupervised domain adaptation for semantic segmentation via class-balanced self-training.” Proceedings of the European conference on computer vision (ECCV). 2018.
8. Wang, Mei, and Weihong Deng. ”Deep visual domain adaptation: A survey.” Neurocomputing 312 (2018): 135-153.
9. Guan, Hao, and Mingxia Liu. ”Domain adaptation for medical image analysis: a survey.” arXiv preprint arXiv:2102.09508 (2021).
10. Mahmood, Faisal, Richard Chen, and Nicholas J. Durr. ”Unsupervised reverse domain adaptation for synthetic medical images via adversarial training.” IEEE transactions on medical imaging 37.12 (2018): 2572-2581.
11. Ghafoorian, Mohsen, et al. ”Transfer learning for domain adaptation in mri: Application in brain lesion segmentation.” International conference on medical image computing and computer-assisted intervention. Springer, Cham, 2017.
12. Perone, Christian S., et al. ”Unsupervised domain adaptation for medical imaging segmentation with self-ensembling.” NeuroImage 194 (2019): 1-11.
13. Dou, Qi, et al. ”Unsupervised cross-modality domain adaptation of convnets for biomedical image segmentations with adversarial loss.” arXiv preprint arXiv:1804.10916 (2018).
14. Zhu, Jun-Yan, et al. ”Unpaired image-to-image translation using cycle-consistent adversarial networks.” Proceedings of the IEEE international conference on computer vision. 2017.
15. Park, Taesung, et al. ”Contrastive learning for unpaired image-to-image translation.” European Conference on Computer Vision. Springer, Cham, 2020.
16. Wang, Guotai, et al. ”A noise-robust framework for automatic segmentation of COVID-19 pneumonia lesions from CT images.” IEEE Transactions on Medical Imaging 39.8 (2020): 2653-2663.
17. Park, Taesung, et al. ”Contrastive learning for unpaired image-to-image translation.” European Conference on Computer Vision. Springer, Cham, 2020.
18. Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. ”U-net: Convolutional networks for biomedical image segmentation.” International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.