Unsupervised Cross-Modality Domain Adaptation for Vestibular Schwannoma and Cochlea Segmentation

Chenkai Zhao[1], Yanwu Xu[2], Ming Gao[1] and Kayhan Batmanghelich[2]

[1] Department of Informatics and Networked Systems, University of Pittsburgh, Pittsburgh, USA

[2] Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, USA

Abstract. The purpose of this study is to achieve single-direction translation for segmentation on Magnetic Resonance Imaging (MRI) images from the cross-MoDA challenge. The training mode is consisted of two parts. First, we use the contrastive-unpaired-translation (CUT) model for domain adaptation from contrast-enhanced T1 to high-resolution T2. Second, for the segmentation part, we use the U-net with supervised attention module as our back bone model and train with additional images generated by changing the intensity of vestibular schwannoma.

Keywords: domain adaptation, segmentation, CUT, U-Net.

1 Introduction

Nowadays, most diagnosis and surveillance in patients with Vestibular Schwannoma are performed with contrast-enhanced T1 (ceT1) MR imaging. In order to achieve the diagnosis based on the high-resolution T2 (hrT2) MR imaging, another kind of the MR imaging method with less radiation risk and cost, we use domain adaptation model to translate the contrast-enhanced T1 images into high-resolution T2 images and segment the vestibular schwannoma and cochlea on the translated T2 images.

2 Method

2.1 Data Preprocessing

First, due to the difference between the cases in training data, we resample all images, including the labels, to the same voxel shaping referring to one specific image. Second, in order to reduce the distribution of the noise from the useless part of the images and focus on the brain, we crop both ceT1 and hrT2 images of training dataset along the z-axis with the center of all voxels with intensity higher than the 75th percentile of the whole volume as the size of 384*384*N. Then, we slice the 3D images along the z-axis into N 2D images with the size of 384*384 pixels because the CUT model only accepts 2D images. Before being used to train the translation model, all of these 2D images are rescaled to [0.0, 1.0] separately according to the original 3D images because of the huge difference between the intensity of different images.

2.2 Domain Adaptation

For domain adaptation, we use CUT to translate the contrast-enhanced T1 images into high-resolution T2 images. There are three components in the translation model, generator G, discriminator D and the feature extractor F. For the G, we simply use the ResNet-9block. For the D network, we use a PatchGAN discriminator [1]. The F is just a multi-layer fully-connected layers as [2]. The optimizer for G, D and F are all set to ADAM. The learning rate is 2e-4 and training epoch is 50.

2.3 Segmentation

In order to let the segmentation model handle the heterogeneous signal intensity of vestibular schwannoma in the training images, we apply data augmentation before training the segmentation model. We reduce the intensity of the area of vestibular schwannoma in half of the hrT2 images generated by the translation model by 45%. On the contrast, for the rest of the generated hrT2 images, we increase the intensity of the area of the tumor by 0.75x.

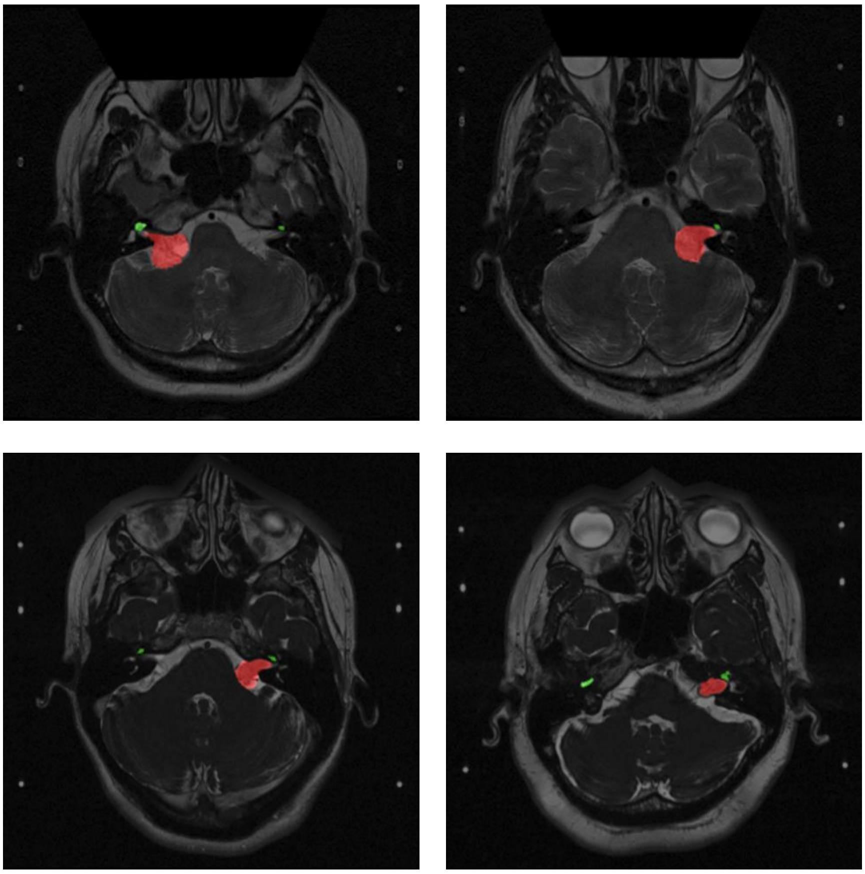
For the segmentation model, we use the U-net2d5 network as our backbone. Because of the small size and difficulty of finding the correct position of cochlea, we use the basic U-net2d5 network to segment the tumor and cochlea as a whole. Then we zoom in to the part of segmenting result and apply another U-net2d5 network as the attention module to segment the cochlea from the zoom-in image. The optimizers for both U-ne2d5 and attention U-net2d5 are ADAM with the initial learning rate of 1e-4 and sigmoid decaying until the end of training.

2.4 Post Processing

For the segmentation results, there are many small holes inside the areas labelled as tumor or cochlea. In order to figure this out, we try to fill these holes and choose the largest components for the tumor and cochlea and remove the rest which can be regarded as the noise.

3 Result

The validation dataset is from the crossMoDA2022 challenge. The main metric is the Mean Dice, which is used to rank in the leaderboard. Our predicted model is ranked 25th on the leaderboard. Here are some of our predicted results.



References

1. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5967–5976 (2017). https://doi.org/10.1109/CVPR.2017.632 2
2. Park, T., Efros, A.A., Zhang, R., Zhu, J.Y.: Contrastive learning for unpaired image-to-image translation. In: European Conference on Computer Vision (2020) 1, 2
3. Shapey, J., Kujawa, A., Dorent, R., Wang, G., Bisdas, S., Dimitriadis, A., Grishchuck, D., Paddick, I., Kitchen, N., Bradford, R., Saeed, S., Ourselin, S., Vercauteren, T.: Segmentation of Vestibular Schwannoma from Magnetic Resonance Imaging: An Open Annotated Dataset and Baseline Algorithm (2021). https://doi.org/10.7937/TCIA.9YTJ-5Q73, https://wiki.cancerimagingarchive. net/x/PZwvB, type: dataset 1
4. Shapey, J., Kujawa, A., Dorent, R., Wang, G., Dimitriadis, A., Grishchuk, D., Paddick I., Kitchen, N., Bradford, R., Saeed, S.R., Bisdas, S., Ourselin, S., Vercauteren, T.: Segmentation of vestibular schwannoma from mri — an open annotated dataset and baseline algorithm. medRxiv (2021). https://doi.org/10.1101/2021.08.04.21261588, https://www.medrxiv.org/content/early/2021/08/08/2021.08.04.21261588 1