Fast Single Direction Translation for Brain Image Domain Adaptation

Yanwu Xu[1], Mingming Gong[2], and Kayhan Batmanghelich[1]

[1] Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh , Pennsylvania , USA

[2] School of Mathematics and Statistics, Melbourne Centre for Data Science, The University of Melbourne, Australia

Abstract. In this paper, we apply a fast adaptation model for the Domain Adaptation of Segmentation on Medical Images. To be concrete, we separate the training model into two stages. 1. Fast translation stage: T1→T2 3D translation, 2. Training Segmentation Model. For the first stage, we apply one of the most popular model CUT [2] and extend the 2D translation to 3D translation. For the second stage, we apply the provided attention U-Net as our back bone model and then build a detection module upon it.

Keywords: 3D translation · segmentation · U-Net.

1 Data Prepossessing

To manually align the distribution between the source and the target, we apply some data prepossessing method at the low-level. 1. We restrict the intensity value to be in the range of [0,4100] for every patient scan. 2. The 3D volume for all the patients are re-sampled to have the same spacing. 3. We apply the intensity normalization to all volumes to be 0 mean and 0.5 variance. We also noticed that, the size the the patients varies. Thus we apply the mixed cropping size during the training time. And we test our method on [4,3]

2 Method

In this section, we apply two stage training. 1. the single direction translation of T1 → T2 is done by the 2D translation model. 2. we train our segmentation on both the original T1 volume and the translated T2 volume.

2.1 Model Design

For the translation model, we have three components, generator G, discriminator D and the feature extractor F. For the G, we simply apply the default 3D attention U-Net from VS Seg. Besides, we remove two down-sampling layers from VS Seg but keep the same structure. For the D network, we use a PatchGAN discriminator [1]. However, to fit the PatchGAN discriminator to our 3D structure, we extend it to the 3D form. The model structure of D is the 6 layers of Resnet, and when feeding images to the D, the images are divided into 16 equal-size patches, which is faster for model feed-forward without sacrificing any performance. The F is just a multi-layer full-connected layers as [2].

For the segmentation network, we keep the same settings as VS Seg.

2.2 3D Fast Image Translation

As we have observed that merging the training of image translation and the segmentation would cause the training collapse and no positive effect. Thus, we train the translation model first. At this stage, we apply the 3D attention U-Net as our translation network and the 3D ResNet as our Discriminator. For the implementation details, we follow the [2] and extract the intermediate feature maps from the Generator to anticipate the contrastive leaning. Rather than randomly sampling 256 features, we choose 1024 features for the 3D translation. Also, for the multi-scale purpose, we extract the features from 4 layers. For the training of the translation network, we keep the same settings as [2], except that we repalce all the 2D models with the modified 3D models. The optimizer is set to be ADAM for G, D, F, the learning rate is $2e-4$ and the training epoch is 70.

2.3 Segmentation

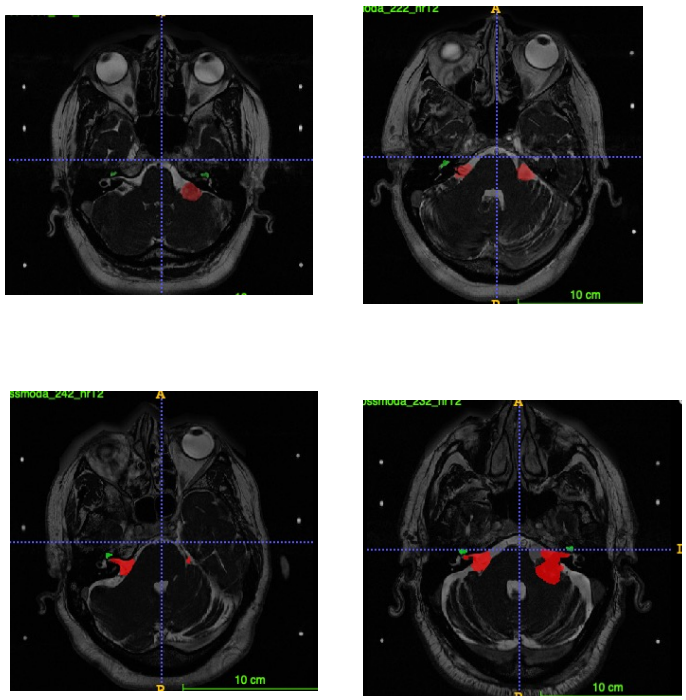
In this stage, we are trying to segment the vestibular schwannoma and the cochlea in the brain. The brain tumor is usually in larger size, however the cochlea is rather smaller than the brain tumor and hard to be well-segmented by the model. Thus we utilize a rough model to segment a coarse mask for the brain tumor and the cochlea. And then for the cochlea, we extract the smaller patch by the coarse mask and then zoom in by the scale factor of 2. We apply another attention U-Net to segment the cochlea. For the training settings, we only change the optimizer to be the SGD with initial learning rate of 2e-2 with linear decaying to the end of training.

3 Post processing

We follow the standard operation of morphological operations and components selection. For the morphological operations, we are trying to fill the small hole inside the predicted segmentation maps and for the components selection, we want to remove the noisy segmentation by the model.

4 Result

For our predicted model, it reaches the rank of 34 on the validation leader board for the 2D translation model. But we haven't submitted the 3D version, thus, in the testing phase, the results should have a improvement on this. Here we show several visual results from the validation dataset. The average dice score of our predicted model is 0.4638 ± 0.1862.



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