An Enhanced Coarse-to-Fine Framework for the Segmentation of Clinical Target Volume

Huai Chen\(^1\), Dahong Qian\(^2\), Weiping Liu\(^3\), Hui Li\(^4\), and Lisheng Wang\(^1(\text{✉})\)

\(^1\) Institute of Image Processing and Pattern Recognition, Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, People’s Republic of China
lswang@sjtu.edu.cn

\(^2\) School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200240, People’s Republic of China

\(^3\) Department of Algorithm and Research, Shanghai Aitrox Technology Co., Ltd., Shanghai, China

\(^4\) National Key Laboratory of Science and Technology on Nano/Micro Fabrication, Key Laboratory for Thin Film and Micro Fabrication of the Ministry of Education, Institute of Micro-Nano Science and Technology, Shanghai Jiao Tong University, Shanghai 200240, People’s Republic of China

Abstract. In radiation therapy, obtaining accurate boundary of the clinical target volume (CTV) is the vital step to decrease the risk of treatment failures. However, it is a time-consuming and laborious task to obtain the delineation by hand. Therefore, an automatic algorithm is urgently needed to realize accurate segmentation. In this paper, we propose an enhanced coarse-to-fine framework to automatically fuse the information of CT, T1 and T2 images to get the target region. This framework includes a coarse-segmentation stage to identify the region of interest (ROI) of targets and a fine-segmentation stage to iteratively refine the segmentation. In the coarse-segmentation stage, the F-loss is proposed to keep the high recall rate of the ROI. In the fine segmentation, the ROI of target will be first cropped according to the ROI obtained by coarse-segmentation and be fed into a 3D-Unet to get the initial results. Then, the prediction and medium features will be set as additional information for the next one network to refine the results. When evaluated on the validation dataset of challenge of Anatomical Brain Barriers to Cancer Spread (ABCs), our method won the 3\(^{th}\) place in the public leaderboard.

Keywords: Clinical target volume segmentation · Multi-modality medical images · Coarse-to-fine · Iterative refinement

1 Introduction

In radiation therapy, the delivery of a radiation dose to the treatment target is very conformal, with as low as 1 mm uncertainty [1]. Obtaining accurate boundary of the clinical target volume (CTV) is important for conformal treatments.
to reduce the risk of treatment failures. Recently, radiologists mark the CTV regions slice by slice by hand. It is time-consuming and deeply depends on the experiments of doctors. Therefore, automatic and accurate segmentation of CTV is urgently needed for alleviating the workload of clinicians and improving the efficiency of treatment planning.

The goal of ABCs is to identify the best methods of segmenting brain structures that serve as barriers to the spread of brain cancers and structures to be spared from irradiation, for use in computer assisted target definition for glioma and radiotherapy plan optimization. The segmentation of brain structures is a challenging task as different structures can be visually appreciated more or less favorably on different imaging modalities. Furthermore, as multi-modality images are usually acquired at different time points, they could present subtle anatomical differences even for brain imaging. This presents a unique technological challenge as information from multi-modality imaging is used to define the clinical target volume and the healthy organs for each individual patient’s disease [1].

There are totally 15 targets in this challenge, including 5 CTVs and 10 structures. And all of them have various shapes and size. Therefore, the first challenge is to propose an unified framework fitting to all of them. To address this challenge, the coarse-to-fine framework, which firstly adopt coarse segmentation to identify the ROI of target and then adopt fine segmentation to get accurate predictions, is a good choice for the base model. When utilizing the idea of coarse-to-fine, we can not only divide complex segmentation task of 15 targets into simple one-target task but also be successfully process targets with extreme small size.

For the fusion of multi-modal medical images, previous works can be divided into 3 categories according to the fusion style. They are input-level fusion [3, 5], medium-level fusion [7, 8] and decision-level fusion [4, 6]. Input-level fusion based methods directly concatenates the original images, while medium-level fusion based methods fuse features of medium layers and decision-level fusion based methods fuse final outputs of models. In this paper, we utilize the input-level fusion.

Based on the base idea of above analysis, which including utilizing coarse-to-fine framework and input-level fusion, we propose a enhanced coarse-to-fine method. In the coarse-segmentation stage, we propose F-loss to keep high recall rate of ROI to alleviate the missing of target region. And in the fine-segmentation, we propose iterative refinement to iteratively refine the results based on previous predictions and features. Finally, referring to our previous work of [2], we reuse features and results of models in fine-segmentation stage and fuse them to make final predictions. When evaluated on the validation set of challenge of ABCs, our method won the 3rd place with mean DSC of 0.883, mean SDSC of 0.980 for task1 and mean DSC of 0.775, mean SDSC of 0.942 for task2.
Fig. 1. An illustration of the coarse-to-fine framework.

Fig. 2. An illustration of the iterative refinement.

Fig. 3. An illustration of the ensemble refinement.
2 Method

Our framework can be divided into three stages. The first stage is coarse-segmentation, aiming to identify the ROI of targets. The F-loss is proposed to alleviate the missing of important regions by keeping high recall rate. The second stage is fine-segmentation stage, in which, ROI will be firstly cropped and be segmented. In this stage, several 3D-Unets are built, and we propose iterative refinement to repeatedly refine the results by reusing predictions and medium features of previous models in the new one. The third stage is ensemble refinement, where all of the final feature maps and predictions of networks in fine-segmentation stage will be fused and fed into a fusion block to make final decisions. The illustration of these stage are respectively shown in Figs. 1, 2 and 3.

2.1 F-Loss to Keep the High Recall Rate of Coarse Segmentation

In the coarse segmentation stage, keeping high recall rate to alleviate the missing of important regions is the key concern. We define a F-loss loss function to increase recall rates and thereby to alleviate incomplete ROI definitions. The F-loss is inspired by F-score ($F_{score} = \frac{(1 + \beta^2) \times \text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}}$), where setting $\beta > 1$ shows preference for Recall. F-loss function is defined as below:

$$\text{Loss}_F(\beta) = - (1 + \beta^2) \times \frac{\sum_{i=1}^{N} p_i g_i}{\sum_{i=1}^{N} p_i} \times \frac{\sum_{i=1}^{N} p_i g_i}{\sum_{i=1}^{N} g_i} \times \frac{\sum_{i=1}^{N} p_i g_i}{\sum_{i=1}^{N} g_i}$$

$$= - \frac{(1 + \beta^2)}{\beta^2} \times \frac{\sum_{i=1}^{N} p_i g_i}{\sum_{i=1}^{N} g_i + \sum_{i=1}^{N} p_i} \quad (1)$$

We can set $\beta > 1$ ($\beta = 4$ in this paper) to keep high recall rate. And it worth noting that it will be the dice loss if $\beta = 1$.

2.2 Iterative Refinement to Iteratively Refine the Results

Iterative refinement is proposed to reuse predictions and features of previous models as additional information to enhance the new model. As shown in Fig. 1 and Fig. 2, the fine segmentation stage of base coarse-to-fine framework, named as step2 in Fig. 1, will firstly obtain the initial prediction and medium features. And then these features and predictions will be as additional information for the next network (named as step3), where predictions will be directly merged into the original images and features will be added into the corresponding features. Similarly, previous features of step2 and step3 will be added into step4 and results of step3 will be merged into the input of step4 for step4 iterative refinement.
2.3 Ensemble Refinement to Fuse Multiple Information to Get Finer Results

The final stage is ensemble refinement, in which all previous predictions and features are merged with original images to get finer results. As shown in Fig. 3, all features and predictions from step2, step3 and step4 are merged with original images and fed into a fusion block consisted with 3 convolution layers.

3 Experiments

3.1 DataSet

The challenge of ABCs2020 provides 45 multi-modal images (CT, T1 and T2) with their ground truth annotations as the training data, 15 instances as the validation data and 15 instances as the final test data.

3.2 Training Details

Adam is set as the optimizer for all models and the learning rate is set as $10^{-3}$. For the training of coarse-segmentation model, the loss function is F-loss with $\beta = 4$, while for the training of other models, the loss function is F-loss with $\beta = 1$ i.e. dice loss. The total epoch for each model is 50 and we split 5 cases from the training data as the validation data. Therefore, we can set learning rate to half when the validation loss does not decrease in 5 epochs.

4 Results

When evaluated on the validation set, our method obtain mean DSC with 0.883, mean SDSC with 0.980 for task 1, and mean DSC with 0.775, mean SDSC with 0.942 for task 2. The final mean score is 0.888 and is ranked 3$^{th}$ place in the public leaderboard.

References

1. https://abcs.mgh.harvard.edu/index.php
2. Chen, H., Wang, X., Huang, Y., Wu, X., Yu, Y., Wang, L.: Harnessing 2D networks and 3D features for automated pancreas segmentation from volumetric CT images. MICCAI 6, 339–347 (2019)
3. Havaei, M., et al.: Brain tumor segmentation with deep neural networks. Med. Image Anal. 35, 18–31 (2017)
4. Kamnitsas, K., et al.: Ensembles of multiple models and architectures for robust brain tumour segmentation. In: Crimi, A., Bakas, S., Kujić, H., Menze, B., Reyes, M. (eds.) BrainLes 2017. LNCS, vol. 10670, pp. 450–462. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75238-9_38
5. Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med. Image Anal. 36, 61–78 (2017)
6. Nie, D., Wang, L., Gao, Y., Shen, D.: Fully convolutional networks for multimodality isointense infant brain image segmentation. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 1342–1345. IEEE (2016)
7. Tseng, K.L., Lin, Y.L., Hsu, W., Huang, C.Y.: Joint sequence learning and cross-modality convolution for 3D biomedical segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6393–6400 (2017)
8. Valada, A., Mohan, R., Burgard, W.: Self-supervised model adaptation for multimodal semantic segmentation. arXiv preprint arXiv:1808.03833 (2018)