Coarse-To-Fine Segmentation of Organs at Risk in Nasopharyngeal Carcinoma Radiotherapy

Qiankun Ma[1], Chen Zu[2], Xi Wu[3], Jiliu Zhou[1,3], and Yan Wang[1](✉)

[1] School of Computer Science, Sichuan University, Chengdu, China
[2] Department of Risk Controlling Research, JD.com, Beijing, China
[3] School of Computer Science, Chengdu University of Information Technology, Chengdu, China

Abstract. Accurate segmentation of organs at risk (OARs) from medical images plays a crucial role in nasopharyngeal carcinoma (NPC) radiotherapy. For automatic OARs segmentation, several approaches based on deep learning have been proposed, however, most of them face the problem of unbalanced foreground and background in NPC medical images, leading to unsatisfactory segmentation performance, especially for the OARs with small size. In this paper, we propose a novel end-to-end two-stage segmentation network, including the first stage for coarse segmentation by an encoder-decoder architecture embedded with a target detection module (TDM) and the second stage for refinement by two elaborate strategies for large- and small-size OARs, respectively. Specifically, guided by TDM, the coarse segmentation network can generate preliminary results which are further divided into large- and small-size OARs groups according to a preset threshold with respect to the size of targets. For the large-size OARs, considering the boundary ambiguity problem of the targets, we design an edge-aware module (EAM) to preserve the boundary details and thus improve the segmentation performance. On the other hand, a point cloud module (PCM) is devised to refine the segmentation results for small-size OARs, since the point cloud data is sensitive to sparse structures and fits the characteristic of small-size OARs. We evaluate our method on the public Head&Neck dataset, and the experimental results demonstrate the superiority of our method compared with the state-of-the-art methods. Code is available at https://github.com/DeepMedLab/Coarse-to-fine-segmentation.

Keywords: Image segmentation · Organs at risk · Nasopharyngeal carcinoma · Coarse-to-fine

1 Introduction

Nasopharyngeal carcinoma (NPC) is a common but fatal malignant tumor arising from nasopharynx or upper throat [1]. For NPC patients, radiation therapy is one of the main treatments. During radiotherapy planning, delineating organs at risk (OARs) is a crucial step to avoid potential radiation risks to normal tissues. Currently, the OARs are always delineated by radiation oncologists manually based on computed tomography (CT) scans, which is extremely time-consuming and subjective [2, 3]. Thus, it is highly

M. de Bruijne et al. (Eds.): MICCAI 2021, LNCS 12901, pp. 358–368, 2021.
https://doi.org/10.1007/978-3-030-87193-2_34

desirable to develop an automatic OARs segmentation approach for NPC patients to deliver efficient and accurate radiotherapy planning.

With the rise of deep learning, a number of approaches have been proposed to delineate OARs from NPC CT images [4–12]. The current OARs segmentation methods can be divided into two categories: 1) organ-specific segmentation based on target region and 2) multi-organ segmentation based on entire CT image. Specifically, the organ-specific segmentation aims to design an exclusive segmentation model to achieve the best segmentation performance for each OAR [4, 5]. Nevertheless, such methods inevitably bring tremendous computational overhead due to the multiple specific models, thus limiting their applicability. In contrast, the multi-organ segmentation based on entire CT image targets at training a deep model which can segment multiple OARs simultaneously. For instance, Tong et al. [12] proposed a fully convolutional neural network (FCNN) with a shape representation model to learn the shape of segmentation targets and delineate the OARs of NPC. Zhu et al. [8] developed a 3D Squeeze-and-Excitation U-Net for OARs segmentation. Gao et al. [9] presented a model named FocusNet, which locates the center points of multiple OARs respectively to extract the corresponding 3D image patches and further segments them. Tang et al. [10] explored the 3D UNet to get the region of interests (ROIs) of each organ by embedding an OAR detection module. Liang et al. [11] proposed a multi-view spatial aggregation framework using 2D ROIs detection module to assist segmentation.

Although current OARs segmentation methods have achieved promising progress, the performance is still somewhat unsatisfactory due to the following challenges. First, compared with other types of cancer which only have a small number of OARs, the number of OARs for NPC patients is up to more than ten, as shown in Fig. 1. Second, it is obvious in Fig. 1 that the size of OARs is highly variable, making the segmentation models prone to segment the large-size OARs (e.g., brain stem, mandible) but neglect the small-size ones (e.g., optical nerves, optical chiasm). Third, most of the OARs only occupy small volumes in CT images, causing the problem of target sparsity and regional imbalance between foreground and background. The ratio between the background and the smallest organ can even reach nearly 10^5:1 in some extreme cases [9]. Fourth, the boundary ambiguity of OARs in NPC CT images is also a sore point.

In this paper, to overcome the above-mentioned challenges, we propose a novel end-to-end coarse-to-fine segmentation model to automatically segment multiple OARs in CT images. Specifically, the entire framework consists of a coarse stage and a fine stage. The coarse stage tries to generate rough segmentation results using an encoder-decoder architecture embedded with a target detection module (TDM). According to the size of targets, the preliminary results from the coarse stage are further categorized into large- and small-size OARs groups using a preset threshold. In the fine stage, we design two exclusive refinement networks for the large- and small-size OARs, respectively. Particularly, to tackle the boundary ambiguity problem of the large-size targets, an edge-aware module (EMA) is devised for capturing the boundary details for performance refinement. Moreover, considering that the point cloud data is sensitive to sparse structures and fits the characteristic of small-size OARs, we explore a point cloud module (PCM) to refine the segmentation results of small-size OARs. We evaluate our method on the public Head&Neck dataset [14]. The experimental results demonstrate that our proposed method achieves better performance than other state-of-the-art methods in both qualitative and quantitative measures.

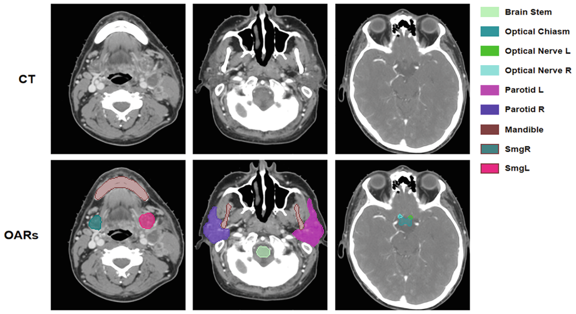

Fig. 1. Illustration of CT images and typical nine OARs to be delineated in NPC.

2 Methodology

The architecture of the proposed method is illustrated in Fig. 2, which consists of two stages. The first stage is the TDM guided coarse segmentation network which employs the encoder-decoder architecture as backbone while the second stage is responsible for the refinement by two elaborate modules for large-size and small-size OARs respectively. The details of our model and the objective function will be introduced in the following sub-sections.

2.1 Architecture

TDM Guided Coarse Segmentation Network: As shown in Fig. 2(a), the coarse segmentation network in the first stage is a U-Net-like network embedded with TDM. With the guidance of TDM, the network can locate the target areas of OARs and neglect irrelevant background regions. Specifically, taking a CT volume as input, the encoder is equipped with four down-sampling blocks to extract latent representative features, each with two residual sub-blocks based on 3D convolution and a max-pooling layer applied to halve the resolution. After four down-sampling operations, the final extracted feature maps are fed into the TDM for ROI extraction and cropping. Particularly, the TDM contains two separate heads, one for bounding box regression to indicate the location and size of ROI for each OAR, and the other for binary classification to judge whether the corresponding OAR class is detected correctly by the TDM. Then, we adopt the ROI Align layer [15] in order to get feature maps with fixed dimensions. At the end of TDM, two fully connected layers are subsequently applied to predict the class of each OAR proposal and further regress coordinates and size offsets of its bounding box, respectively. Then, these generated ROI proposals are down-sampled to the same size as the features from each encoder layer to crop them, aiming to reduce the interference of irrelevant background regions. In decoder, each up-sampling block initially adopts the trilinear up-sampling to double the size of inputs, and then applies a 3 × 3 × 3 convolution followed by local contrast normalization. After each up-sampling operation, the cropped feature maps from the encoder are concatenated with the feature maps of the decoder. At the last decoder layer, we apply a 1 × 1 × 1 convolution to the final feature maps to generate the coarse segmentation results. Finally, we calculate the size of the segmented multiple OARs and according to a threshold which is preset to 1000 [9] in this paper, the coarse segmentation results are divided into large- and small-size OAR groups and respectively sent to EAM and PCM of the second stage for refinement.

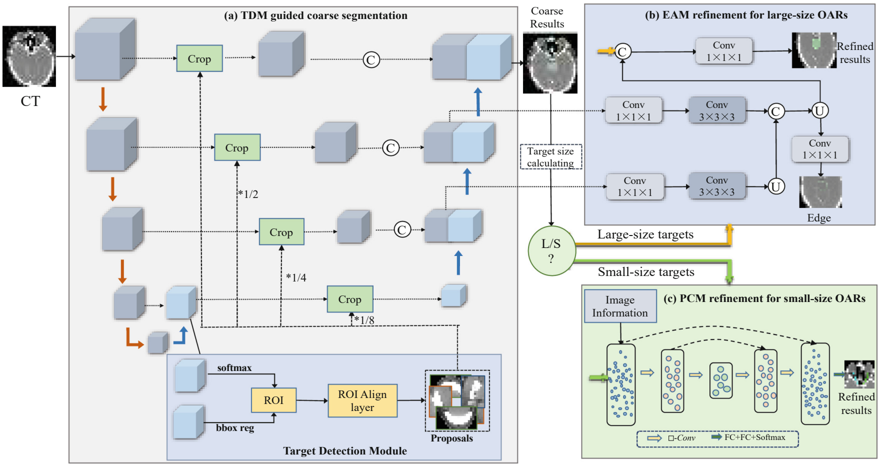

Fig. 2. Overview of our network architecture, including (a) TDM guided coarse segmentation (b) EAM refinement for large-size OARs and (c) PCM refinement for small-size OARs. ‘Conv’ denotes the convolutional layer, ‘C’ and ‘U’ denote concatenation and up-sampling.

Edge-Aware Module (EAM): Aiming to solve the boundary ambiguity problem of segmentation, we design an EAM to refine the coarse segmentation results of large-size OARs. Considering that the rich edge information is mainly contained in the low-level feature maps, we only employ the cropped features of the first two encoder blocks in the first stage as the input of the EAM, as shown in Fig. 2(b). Then, these two low-level feature maps undergo the processing of convolution, concatenation and up-sampling, to obtain the corresponding edge maps which represent the boundaries of OARs. Finally, the edge maps and the coarse segmentation results are fused by a 1 × 1 × 1 convolution to obtain refined segmentation results.

Point Cloud Module (PCM): In view of the advantages of point cloud networks in handling the sparse data, we create a PCM for small-size OARs refinement, as shown in Fig. 2(c). Our PCM is mainly based on the encoder-decoder architecture proposed by Balsiger et al. [13]. Specifically, for each small-size OAR, the volumetric coarse segmentation result is converted into a point cloud $P = [p_1, p_2, \ldots, p_K]$ with K (set to 2048) points $p_i \in R^3$. Moreover, we additionally extract the image information by following [13] and concatenate the extracted image information with the point cloud P as the input of our PCM. Finally, the outputs of PCM are utilized to replace the values of the corresponding points in the coarse segmentation result to obtain the refined segmentation result.

2.2 Objective Function

The objective function consists of four parts: segmentation loss, TDM loss, EAM loss and PCM loss. Specifically, the segmentation loss function can be expressed as:

$$Loss_{seg} = \sum\nolimits_{c=1}^{C} I(c)\left(1 - \varphi\left(m^c, g^c\right)\right)$$

$$\varphi(m, g) = \frac{\sum_{i=1}^{N} m_i g_i}{\sum_{i=1}^{N} m_i g_i + \alpha \sum_{i=1}^{N} m_i(1 - g_i) + \beta \sum_{i=1}^{N} (1 - m_i) g_i + \varepsilon}, \tag{1}$$

where g^c and m^c respectively represent the ground truth and the final predicted mask of OAR c. C is the total number of OARs. $I(c)$ is set to 1 if the OAR class is detected correctly by the TDM, otherwise set to 0. $\varphi(m, g)$ computes a soft Dice score between the ground truth g and the predicted mask m, where i is a voxel index and N denotes the total number of voxels. α and β are hyper-parameters for controlling the weights of penalizing false negatives and false positives and ε is to ensure the numerical stability of the loss function.

We employ a multi-task loss function to train our TDM, including a classification loss L_c for OARs classification task and a regression loss L_r for bounding box regression task, as formulated in Eq. 2:

$$Loss_{TDM} = \frac{1}{M} \sum_{i} L_c\left(P_i, P_i^*\right) + \lambda \frac{1}{M} \sum_{i} L_r\left(t_i, t_i^*\right), \tag{2}$$

where L_c adopts the cross entropy (CE) loss and L_r uses the smooth L1 loss. λ is a hyper-parameter to balance these two terms.M is the total number of anchors participating in the calculation, while P_i^*, t_i^* denote the predicted class label and box parameter, respectively, and P_i, t_i are their ground truths.

Aiming to capture sufficient edge information, the EAM loss is introduced to constrain the edge map and can be expressed as:

$$Loss_{EAM} = \frac{1}{C} \sum_{c \in C} \overline{\Delta Jc}(m(c)),$$

$$m_i(c) = \begin{cases} 1 - p_i(c) \text{ if } c = y_i(c) \\ p_i(c) \text{ otherwise} \end{cases}, \tag{3}$$

where C denotes the total number of OARs. For the pixel i of OAR c, $y_i(c)$ denotes the ground truth binary mask while $p_i(c)$ denotes the predicted probability between 0 and 1. $\overline{\Delta Jc}$ represents the Lovasz extension of the Jaccard index [16].

For PCM, we employ the binary cross entropy (BCE) loss to constrain the output of the point cloud network. The loss of PCM is defined as:

$$Loss_{PCM} = L_{BCE}\left(S, S^{*}\right), \tag{4}$$

where the S and S^{*} are the output and ground truth of PCM.

The total loss is defined as:

$$Loss_{total} = Loss_{seg} + \mu_1 Loss_{TDM} + \mu_2 Loss_{EAM} + \mu_3 Loss_{PCM}, \tag{5}$$

where μ_1, μ_2 and μ_3 are balance terms.

2.3 Training Details

Our network is trained on PyTorch framework and an NVIDIA GeForce GTX 1080Ti with 11 GB memory. Specifically, we take Adaptive moment estimation (Adam) optimizer with a momentum of 0.9 to optimize the network. The proposed network is trained for 200 epochs and the batch size is set to 1. The initial learning rate is set to 0.001, and decays to 0.0001 and 0.00001 when the epoch respectively reaches 100 and 150. Based on our trial studies, α and β in Eq. 1 both equal to 0.5, λ in Eq. (2) is set to 1. μ_1, μ_2 and μ_3 in Eq. (5) are set to 1, 2 and 2, respectively.

3 Experiment and Analysis

3.1 Dataset and Evaluation

Our proposed method is evaluated on the MICCAI 2015 Head&Neck Auto Segmentation Challenge dataset, which contains a set of CT volumes for 48 NPC patients with the image size varying from 512 × 512 × 39 to 512 × 512 × 181. Each patient involves nine OARs, including brain stem, mandible, optic chiasm, optic nerve (both left and right), parotid (both left and right), submandibular gland (both left and right). The dataset was split by the Challenge, where 33 subjects are used as training set and the remaining 15 subjects are used as the test set. All samples are preprocessed to fit the maximum input size of our model, i.e., 240 × 240 × 112. To evaluate and analyze the experimental results, we adopt three common metrics, including dice similarity coefficient (DSC), 95th percentile Hausdorff distance (95% HD), and average surface distance (ASD).

Table 1. Quantitative comparisons terms of DSC with state-of-the-art methods. † denotes $p < 0.05$ through paired t-test.

OAR	AnatomyNet[8]	FocusNet[9]	U_aNet[10]	Multi-view[11]	Proposed
Brain Stem	86.7 ± 2†	87.5 ± 2.6	87.5 ± 2.5†	**92.3 ± 1.0**	87.9 ± 2.4
Mandible	92.5 ± 2†	93.5 ± 1.9	**95.0 ± 0.8**	94.1 ± 0.7	94.5 ± 0.7
Optical Chiasm	53.2 ± 15†	59.6 ± 18.1	61.5 ± 10.2†	**71.3 ± 8.3**	65.9 ± 9.6

(*continued*)

Table 1. (*continued*)

OAR	AnatomyNet[8]	FocusNet[9]	U$_a$Net[10]	Multi-view[11]	Proposed
Optical Nerve L	72.1 ± 6†	73.5 ± 9.6	74.8 ± 7.1†	73.8 ± 4.6	**75.3 ± 6.8**
Optical Nerve R	70.6 ± 10†	74.4 ± 7.2	72.3 ± 5.9†	73.4 ± 5.1	**74.7 ± 5.3**
Parotid L	88.1 ± 2†	86.3 ± 3.6	88.7 ± 1.9†	88.2 ± 1.3	**89.2 ± 1.5**
Parotid R	87.4 ± 4†	87.9 ± 3.1	87.5 ± 5.0†	87.0 ± 1.5	**88.4 ± 4.9**
SMG L	81.4 ± 4†	79.8 ± 8.1	82.3 ± 5.2†	81.5 ± 2.9	**82.9 ± 4.8**
SMG R	81.3 ± 4†	80.1 ± 6.1	81.5 ± 4.5	80.0 ± 3.4	**81.5 ± 3.7**

3.2 Comparison with State-Of-The-Art Methods

To validate the advancement of our method, we compare it with four state-of-the-art (SOTA) OARs segmentation methods, including AnatomyNet [8], FocusNet [9], UaNet [10], Multi-view [11]. Table 1 gives the DSC results of the nine OARs segmented by different methods. As observed, our method achieves the best performance in six OARs. To study the statistical significance of our proposed method, we also perform paired t-tests (only for the methods with public available code) to compare the SOTAs against our method, through which we can find that almost all p-values are less than 0.05, demonstrating the statistical significance of the achieved improvement. Moreover, the quantitative results regarding 95% HD and ASD are shown in Fig. 3, from where we can see that our method obtains the competitive performance with the SOTAs.

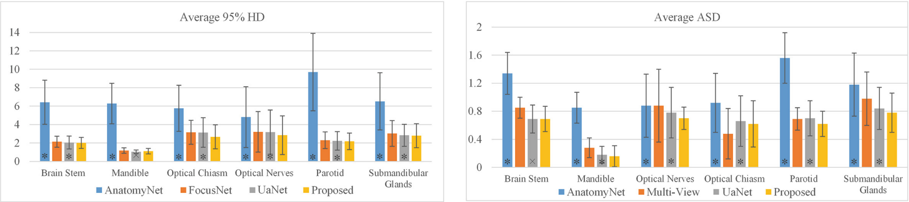

Fig. 3. Quantitative comparisons with the state-of-the-art methods in terms of 95% HD and ASD.

For qualitative comparison, we display the visual comparison results in Fig. 4. Note that, the visual results of the FocusNet and Multi-view are not given here since their code has not yet been released. Here, we also display the result of 3D Unet for comparison, due to its widely application in medical image segmentation tasks. As observed, the 3D Unet presents the worst segmentation results as the gap between the prediction and the ground truth is the largest. Compared with AnatomyNet and UaNet, our method gives more precise segmentation results with less false positive predictions.

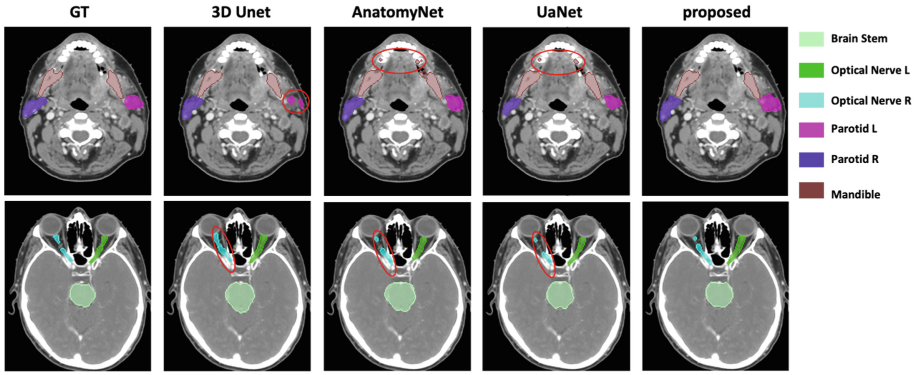

Fig. 4. Qualitative comparisons with the state-of-the-art methods. The areas with obvious improvement achieved by our method are circled by the red ellipsoids. (Color figure online)

3.3 Ablation Study

To evaluate the contributions of the components in the proposed method, we perform the ablation study in a progressive way using the 3D U-net (i.e., the first-stage without TDM) as the baseline. To be specific, our experimental settings include: (1) 3D U-net, (2) 3D U-net + TDM, (3) 3D U-net + TDM + EAM, (4) 3D U-net + TDM + PCM, (5) 3D U-net + TDM + EAM + PCM (proposed).

Table 2. Quantitative ablation study results. ✓ stands for the addition of corresponding module.

3D Unet	TDM	EAM	PCM	DSC [%]	95% HD	ASD
✓				75.42 ± 9.6	6.25 ± 3.7	1.64 ± 0.8
✓	✓			80.51 ± 5.3	2.76 ± 1.7	0.80 ± 0.4
✓	✓	✓		81.21 ± 4.9	2.47 ± 1.2	0.65 ± 0.3
✓	✓		✓	81.42 ± 4.8	2.54 ± 1.3	0.71 ± 0.3
✓	✓	✓	✓	**82.25 ± 4.4**	**2.38 ± 1.2**	**0.57 ± 0.2**

The quantitative ablation results are shown in Table 2. By comparing the first and second row, we find that DSC significantly improves from 75.42% to 80.51%, the 95% HD drops from 6.25 to 2.76 and the ASD drops from 2.46 to 0.89, respectively, demonstrating the necessity of TDM of the proposed method. Similarly, by further incorporating the EAM or PCM, the performance becomes better in terms of all three metrics. Undoubtedly, the completed model with TDM, EAM and PCM achieves the best performance, with the highest DSC and lowest 95%HD and ASD. Table 3 shows the DSC results of nine OARs based on different experimental settings. Specifically, by incorporating the TDM, the segmentation results were improved for all organs. With the introduction of the EAM and PCM, the 3D Unet + TDM + EAM achieves the best dice results on three large organs, (i.e., brainstem, parotid, and SMG R) and the 3D Unet + TDM + PCM

Table 3. Quantitative ablation results of each organ on DSC

OAR	3D Unet	3D Unet + TDM	3D Unet + TDM + EAM	3D Unet + TDM + PCM	Proposed
Brain Stem	84.5 ± 5.6	87 ± 2.6	**88.6 ± 2.0**	86.7 ± 2.9	87.9 ± 2.4
Mandible	89.3 ± 3.8	93.9 ± 1.1	94.4 ± 0.7	93.6 ± 1.8	**94.5 ± 0.7**
Optical Chiasm	48.4 ± 21.4	60.9 ± 11.2	59.8 ± 12.6	**67.3 ± 8.0**	65.9 ± 9.6
Optical Nerve L	68.1 ± 9.7	73.8 ± 7.3	73.5 ± 7.9	**75.8 ± 5.4**	75.3 ± 6.8
Optical Nerve R	64.9 ± 10.8	72.6 ± 6.8	72.8 ± 5.4	74.5 ± 5.2	**74.7 ± 5.3**
Parotid L	84.4 ± 8.5	87.9 ± 2.9	**89.5 ± 1.4**	87.2 ± 3.1	89.2 ± 1.5
Parotid R	82.8 ± 9.1	86.6 ± 5.6	88.2 ± 4.8	85.8 ± 6.7	**88.4 ± 4.9**
SMG L	77.5 ± 8.9	81.5 ± 5.5	82.6 ± 5.3	81.8 ± 5.2	**82.9 ± 4.8**
SMG R	78.9 ± 8.6	80.4 ± 4.8	**81.5 ± 3.6**	80.1 ± 4.8	81.5 ± 3.7
Average	75.42 ± 9.6	80.51 ± 5.3	81.21 ± 4.9	81.42 ± 4.8	**82.25 ± 4.4**

achieves the best dice results on two small organs, (i.e., Optical Chiasm and Optical Nerve L). These experimental results show that the introduction of EAM and PCM is helpful for the improvement of large- and small-size OARs.

The qualitative ablation results shown in Fig. 5 also demonstrate the effectiveness of each component we proposed. Particularly, from the results of (2) and (3), we can see that the EAM indeed could produce the segmentation results with more accurate boundaries, as indicated by the red arrows. By comparing (2) with (4), we can conclude that the PCM could refine the segmentation performance for the small-size OARs, as shown in the red squares. The visual effect in (5) yielded by our complete model has the smallest difference from the ground truth for both large- and small-size targets, supporting the findings in the statistical data in Table 2.

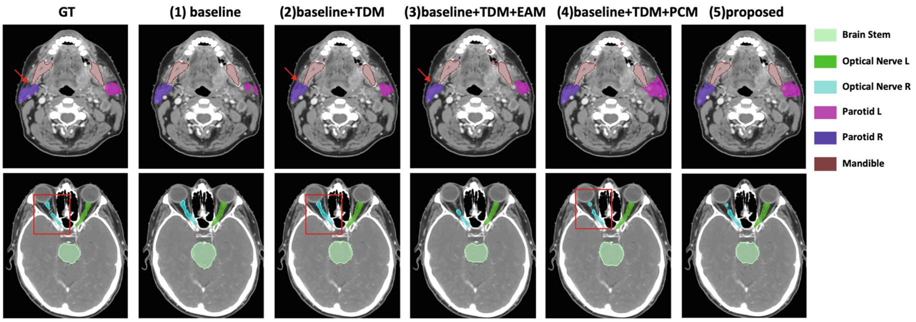

Fig. 5. Qualitative ablation study results. ‘Optical Nerve L’ and ‘Optical Nerve R’ are small-size OAR, the other four OARs appearing in the images are all large-size OAR.

4 Conclusion

In this paper, we propose a novel end-to-end two-stage segmentation network to automatically segment multiple OARs in NPC, including the first stage for coarse segmentation and the second stage for refinement by two elaborate modules. Concretely, in the first stage, we construct a well-performed target detection module (TDM) to locate and crop the general area for each OAR, thus eliminating the interference of large background area and making the network pay more attention to the OARs. In the second stage, an edge-aware module (EAM) is established to focus on the segmentation boundary of large-size targets and alleviate the boundary ambiguity problem. For small-size targets, since the point cloud data is sensitive to sparse structures, a point cloud module (PCM) is employed to further refine the segmentation performance. Experiments on the public Head&Neck dataset show that our method achieves competitive results compared with the state-of-the-art methods.

Acknowledgments. This work is supported by National Natural Science Foundation of China (NSFC 62071314) and Sichuan Science and Technology Program (2021YFG0326, 2020YFG0079).

References

1. Zhong, T., Huang, X., Tang, F., et al.: Boosting-based cascaded convolutional neural networks for the segmentation of CT organs-at-risk in nasopharyngeal carcinoma. Med. Phys. **46**(12), 5602–5611 (2019)
2. Nelms, B.E., Tomé, W.A., Robinson, G., et al.: Variations in the contouring of organs at risk: test case from a patient with oropharyngeal cancer. Int. J. Radiat. Oncol. Biol. Phys. **82**(1), 368–378 (2012)
3. Brouwer, C.L., Steenbakkers, R.J.H.M., van den Heuvel, E., et al.: 3D variation in delineation of head and neck organs at risk. Radiat. Oncol. **7**(1), 1–10 (2012)
4. Ren, X., Xiang, L., Nie, D., et al.: Interleaved 3D-CNN s for joint segmentation of small-volume structures in head and neck CT images. Med. Phys. **45**(5), 2063–2075 (2018)
5. Ibragimov, B., Xing, L.: Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. Med. Phys. **44**(2), 547–557 (2017)
6. Raudaschl, P.F., et al.: Evaluation of segmentation methods on head and neck CT: auto-segmentation challenge 2015. Med. Phys. **44**(5), 2020–2036 (2017)
7. Wang, Z., Wei, L., Wang, L., et al.: Hierarchical vertex regression-based segmentation of head and neck CT images for radiotherapy planning. IEEE Trans. Image Process. **27**(2), 923–937 (2017)
8. Zhu, W., Huang, Y., Tang, H., et al.: Anatomynet: deep 3d squeeze-and-excitation u-nets for fast and fully automated whole-volume anatomical segmentation. Med. Phys. **46**(2), 576–589 (2019)
9. Gao, Y., et al.: Focusnet: imbalanced large and small organ segmentation with an end-to-end deep neural network for head and neck ct images. In: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.-T., Khan, A. (eds.) MICCAI 2019. LNCS, vol. 11766, pp. 829–838. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32248-9_92
10. Tang, H., Chen, X., Liu, Y., et al.: Clinically applicable deep learning framework for organs at risk delineation in CT images. Nat. Mach. Intell. **1**(10), 480–491 (2019)

11. Liang, S., Thung, K.H., Nie, D., et al.: Multi-view spatial aggregation framework for joint localization and segmentation of organs at risk in head and neck CT images. IEEE Trans. Med. Imaging **39**(9), 2794–2805 (2020)
12. Tong, N., Gou, S., Yang, S., et al.: Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks. Med. Phys. **45**(10), 4558–4567 (2018)
13. Balsiger, F., Soom, Y., Scheidegger, O., Reyes, M.: Learning shape representation on sparse point clouds for volumetric image segmentation. In: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.-T., Khan, A. (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 273–281. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_31
14. Raudaschl, P.F., Zaffino, P., Sharp, G.C., et al.: Evaluation of segmentation methods on head and neck CT: auto-segmentation challenge 2015. Med. Phys. **44**(5), 2020–2036 (2017)
15. He, K., Gkioxari, G., Dollár, P., et al.: Mask R-CNN. In: IEEE International Conference on Computer Vision (ICCV), pp. 2980–2988 (2017)
16. Berman, M., Triki, A.R., Blaschko, M.B.: The lovász-softmax loss: a tractable surrogate for the optimization of the intersection-over-union measure in neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4413–4421 (2018)