SpMVNet: Spatial Multi-view Network for Head and Neck Organs at Risk Segmentation

Hongzhi Liu[1], Shiyu Zhu[2], Qianjin Feng[3(✉)], and Yang Chen[1,4]

[1] School of Computer Science and Engineering, Southeast University, Nanjing, China
chenyang.list@seu.edu.cn
[2] National Key Laboratory of Transient of Physics, Nanjing University of Science and Technology, Nanjing, China
[3] School of Biomedical Engineering, Southern Medical University, Guangzhou, China
fengqj99@fimmu.com
[4] Key Laboratory of Computer Network and Information Integration (Ministry of Education), Southeast University, Nanjing, China

Abstract. Head and neck (HaN) cancers are often treated with radiotherapy. Since radiation inevitably causes damage to human organs, it is necessary to control the dose of radiation in different areas during radiation therapy to protect organs at risk (OARs). To solve these incompatible problems, we proposed an end-to-end spatial multi-view network for head and neck organs at risk segmentation, named SpMVNet, to take advantage of both spatial continuous context and multi-view relevance in whole volume CT images. The proposed method includes a symmetric segmentation network (SymNet) and a continuous context network (CCNet), making full use of organs' structural symmetry in CT slices and spatial contextual information of volume data. Our proposed method is validated on the MICCAI 2015 Head and Neck Automatic Segmentation Challenge datasets. Extensive experiments show that it achieves lower error range for most organ segmentation with better evaluation metrics than state-of-the-art methods. This proposed method is helpful to improve the precision of organ segmentation in radiotherapy.

Keywords: Automated segmentation · Organs at risk · Head and neck CT images

1 Introduction

Cancer is a common disease in the world, with a high fatality rate threatening human life and health. More than millions of people die of cancer every year, among which head and neck (HaN) cancer is one of the most difficult cancers to treat because of its complex anatomical structure [15]. And for clinical treatment, the high precision radiotherapy is often the preferred treatment for head and neck cancer, but it is necessary to limit the radiation dose to avoid damage

X. Yang et al. (Eds.): ADMA 2023, LNAI 14177, pp. 257–271, 2023.
https://doi.org/10.1007/978-3-031-46664-9_18

to the organs at risk (OARs), as well as reduce sequelae and complications. It can be seen that accurately delineating the areas of organs at risk is particularly important for the design of radiotherapy schedules. Organs at risk are highly sensitive to radiation, such as the optic nerve and optic chiasm, which cannot tolerate excessive radiation. And the key step in radiation therapy planning is the identification of the boundaries of high-risk organs. Therefore, the automatic segmentation of high-risk organs helps reduce the workload of doctors in radiation therapy planning, resulting in a reduction in the overall cost of radiation therapy from both a time and economic perspective.

CT imaging overcomes the problem of human anatomical structure information overlapping in X-ray imaging, and has the characteristics of high acquisition speed, high spatial accuracy and resolution. Its three-dimensional (3D) data can clearly display the spatial density and accurate position information of human organs, and two-dimensional (2D) plain scans can be used to detect suspicious lesions. Therefore, computed tomography (CT)-based treatment planning remains to be the mainstream in current clinical treatment.

For the multi-target segmentation task in this paper, how to extract the representation of human organs from CT images is a thought-provoking problem due to the large sizes and shape differences of human organs and the complex spatial structure positions. For 2D neural network, it processes slice images layer by layer, which cannot learn the correlation between successive slices, resulting in the loss of spatial information. However, for the 3D framework of voxel-by-voxel image processing, patch training is usually used to counter the large increase in parameters caused by the network, and the maximum receiving range of the network will be limited by computing resources, thus it is easy to lose the global information of large organs.

In actual clinical practice, radiologists usually manually segment the OARs on the each layer of CT images, which is time-consuming and lies on rich experience. Even so, this process of segmentation could also lead to incorrect and misdiagnosis problems. Our proposed method can accurately delineate the organs at risk for radiotherapy schedules according to the prior knowledge of doctors, which can save time and labor cost while explaining the objectivity and interpretability of the method. The research in this paper is based on a publicly available dataset. The aim is to perform the aforementioned blade segmentation on head and neck computed tomography (CT) images. The example of CT images and labels are shown in Fig. 1.

Deep learning methods represented by convolutional neural networks (CNN) in recent years, have made great achievements in the field of medical segmentation [1,9,13,14,22], and CNN has also been applied for OARs segmentation in head and neck CT images [10,18,20,23]. At present, some researchers have completed the related works on this task. The first [10] using deep learning methods proposed a 2D CNN for OARs segmentation from HaN CT images, but it only got a slight improvement in right submandibular gland and right optic nerve, and the performance for the other OARs was similar to that of the

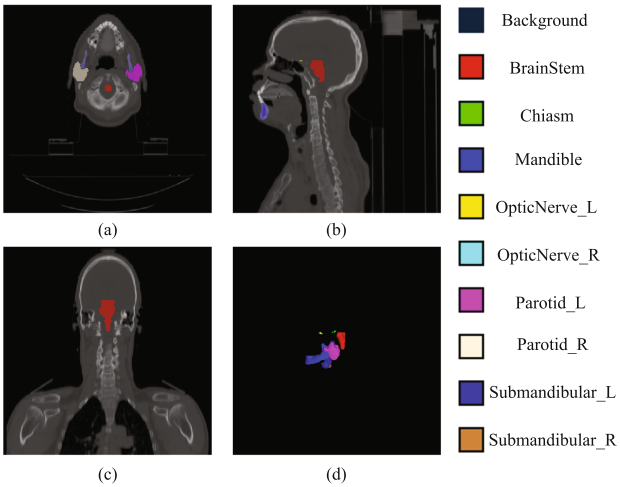

Fig. 1. Segmentation labels of nine organs at risk in head and neck CT images. (a)–(d) are axial, coronal and sagittal views as well as 3D masks, respectively. Different colors on the right represent relevant organs at risk.

traditional methods. Zhu *et at.* [23] proposed the end-to-end method AnatomyNet, a three dimensional squeeze-and-excitation U-Net (3D SE U-Net) based on the SE attention mechanism, combining dice loss and focal loss as optimization constraints. Tong *et at.* [20] designed a fully convolutional neural networks framework with stacked auto-encoder as a shape latent representation model for HaN radiotherapy. However, these existing deep learning-based methods usually produce accurate segmentation maps for large organs and ignore the characteristics of different views of CT data, which have influence on accuracy of small organs and may not be helpful for segmentation of symmetrical OARs.

In this paper, we proposed an end-to-end spatial multi-view network for OARs segmentation, named SpMVNet. The challenging head and neck organs segmentation problem is divided into three views as branches of processing. We first design a symmetric segmentation network (SymNet) to take advantage of the symmetric anatomical structure features of the axial and coronal views, and divide the input network into two parts to make it easier for the network to learn similar features of the symmetric structure. We raise a continuous context network (CCNet) to make full use of the spatially continuous structural information of CT images to make the segmentation masks to be continuous. And the proposed method shows great performance on MICCAI 2015 challenge datasets.

2 Method

In this section, we describe the method of OARs segmentation for head and neck CT images. Our strategy is to simulate the way experienced doctors observe, that is, to predict and locate OARs in different views of volume CT and then output

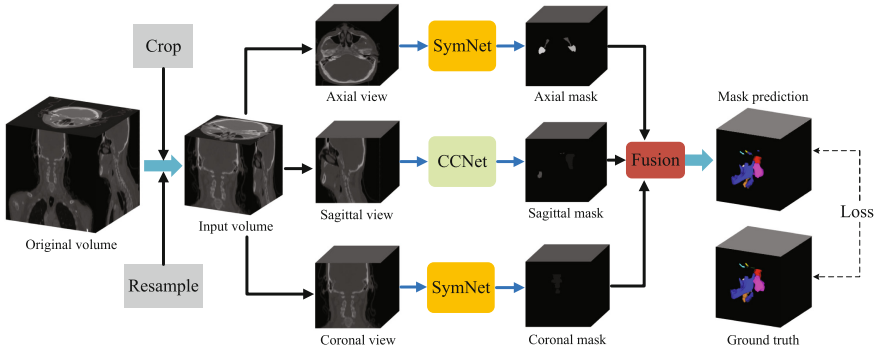

Fig. 2. Overall framework of our SpMVNet. The origin volume is preprocessed by cropping and resampling to obtain the input volume of the network. The data of axial, coronal and sagittal views are input into SymNet and CCNet module, then corresponding prediction masks are output. The masks of the three views are fused and used as the final mask prediction results.

the most probable segmentation results by fusing the masks of three branches at the same spatial position. The overall framework of the proposed SpMVNet has two main components, symmetric segmentation network (SymNet) and continuous context segmentation network (CCNet).

2.1 SpMVNet

We propose a novel end-to-end spatial multi-view network (SpMVNet) for HaN OARs segmentation and its structure is illustrated in Fig. 2. The input volumes are obtained from the origin volumes through image preprocessing, preserving the information of the key parts in HaN CT volumes. After our observation and consultation with hospital experts, we explore the segmentation network using the features of different views and divide the segmentation task into two main sub-networks, namely symmetric segmentation network (SymNet) and continuous context segmentation network (CCNet). We notice that OARs such as the parotid, optic nerve and submandibular have left-right symmetrical physiological structures, so that the CT volumes divided into left and right slices along the midline of the brain for feature learning in axial and coronal views.

SpMVNet for segmentation of HaN OARs can be interpreted as a mathematical theoretical model: a CT medical image I as input and a group of representation constraints C_i $(i = 1, 2, \cdots)$, and the segmentation of I is to acquire a delineation of it, which can be expressed by the following Eq. 1:

$$\bigcup_{x=1}^{N} R_x = I, \quad R_x \bigcap R_y = \varnothing, \\ \forall x \neq y, \quad x, y \in [1, N]. \tag{1}$$

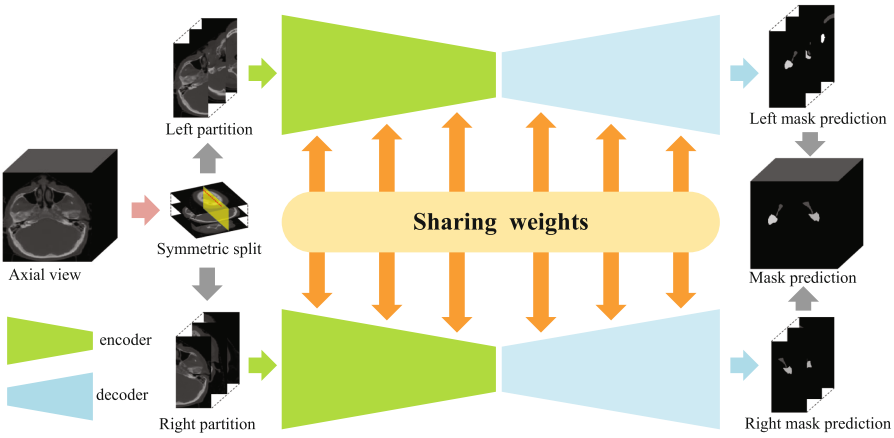

Fig. 3. The structure of our SymNet. The input volume is split into left and right sectors by the brain midline and fed into the siamese network to get the predicted masks of the left and right partitions respectively, and finally merged into the labels.

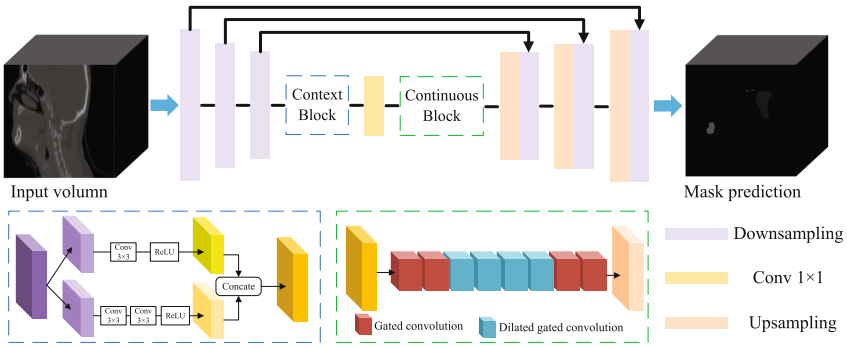

Fig. 4. Illustration of the proposed CCNet. It is on the basis of the segmentation network 3D U-Net with a context block and a continuous block. Input volumes are fed into a feature encoder module, where the ResNet-34 block pretrained from ImageNet [5] is used to replace the original U-Net encoder block.

Herein, R_x satisfies both sets of pixels of the HaN CT images I in the constraint C_i and so does R_y. There is no intersection between R_x and R_y. And x, y are used to distinguish the different regions. N indicates the number of classification including background and nine OARs.

2.2 SymNet

The head and neck CT images have structural symmetry in the axial and coronal views, so the images can be segmented along the midline of brains to obtain the left or right OARs structures, which inspires us to design a symmetrical network for organ feature extraction.

We first calculated the midlines of 2D slices from HaN CT volumes in axial and coronal views. The slices I_s are processed via automatic nonparametric and unsupervised threshold selection segmentation algorithm OTSU [16] to obtain regions of whole brain M_h. We then perform image inflation with a small kennel on the results of the last step and calculate the maximum connected regions R_c. The outer contour of regions C will be saved to the matrix m_c and filled, eliminating the holes inside. The end points P_{up} and P_{down} are searched up and down in the matrix m_c along the midpoint of the segmentation results M_h, and the boundary position is used as the search termination condition, so that the midlines l_m of the HaN slices can be obtained.

Our symmetric segmentation network (SymNet) is composed of the same shared weighted convolutional kernel of encoder and decoder based on the Siamese Network [12], with paired (I_1, I_2) as the network inputs, which is shown in Fig. 3. Siamese network uses shared weight convolution computation and maximum pooling procedures to calculate the similarity between the high-level features (F_1, F_2) of the input images.

We then divide the two-dimensional slices along the dissection line into left and right partition as input and joint the two prediction masks in spatial position, which significantly reduces the amount of network parameters compared with other methods. The L1 distance is used to estimate the similarity of high level features, followed by weight multiplication and sigmoid function to map the value into [0, 1]. The similarity function is formulated as Eq. 2:

$$p = \sigma(W \cdot |f_1 - f_2|), \tag{2}$$

where σ is the sigmoid activation function, W is the weight parameters, is matrix product of two matrices, and f_i is the high level feature F.

The SymNet employs the U-Net [19] with long skip-connections as the baseline network. U-net consists of downsampling and upsampling processes to obtain the predicted segmentation masks. The skip-connection from the downsampling part to the upsampling part has several advantages in fusing local and global features for accurate segmentation with details and resolving the gradient vanishing problem in deep learning models. In our approach, In our approach, the network learns the similar anatomical shape of the left and right portions of the HaN organs, reducing the difference in segmentation results while transferring the convolutional features of the downsampling to the upsampling phase.

2.3 CCNet

The proposed CCNet consists of three major parts: Downsampling module, context block, continuous block and upsampling module. And its detailed illustration are shown in Fig. 4.

A challenge in OARs segmentation is the large variation of object sizes in HaN CT image. For example, a tumor in middle or late stage can be much larger than that in early stage. Motivated by the feature pyramids and multi-scale feature concatenation, we propose novel context block to encode the high-level semantic

feature maps. For segmentation task, the receptive field of the high-level network is relatively large, and the semantic information representation ability is strong, but the representation ability of geometric information is weak while low-level network is relatively small, and the geometric detail information representation ability is strong. In order to enable the network to fully learn features of different scales and improve the effectiveness of the features, our method extracts the feature maps output by stage3, and stage5 based on the Resnet network. For the input of 512×512 size, the output feature map size They are $64 \times 64 \times 512$ and $16 \times 16 \times 2048$, which correspond to the shallow texture features, intermediate transition features and deep semantic features of the image, and are input to the subsequent self-attention module for each layer features for further channel filtering. By combining the convolution of different rates, the context block is able to extract features for objects with various sizes.

Vanilla convolutions in a U-Net [19] have no significant effect for multi-organ segmentation. The inputs of skip connections are almost zeros thus cannot propagate detailed color or texture information to the decoder of that region. Therefore, We customize a continuous block with gated convolution and dilated gated convolution [21]. Gated convolution learns a dynamic feature selection mechanism for each channel and each spatial location and the mask feature output O_M can be formulated as Eq. 3.

$$\begin{aligned} Gating_M &= \sum\sum W_g \cdot V_{HaN}, \\ Feature_M &= \sum\sum W_f \cdot V_{HaN}, \\ O_M &= \phi(Feature_M) \odot \sigma(Gating_M), \end{aligned} \tag{3}$$

where $Gating_M$ and $Feature_M$ represent two type of features extracted from corresponding convolution filter W_g and W_f for the same input volume V_{HaN}. Besides, ϕ and σ mean sigmoid function and activation function.

For delineation boundaries of HaN OARs, our encoder-decoder architecture equipped with context block and continuous block is sufficient to obtain reasonably continuous segmentation results.

2.4 Loss Functions

As illustrated in Fig. 2, our approach needs to train the proposed network to predict each pixel in the CT images to be background or nine OARs, which is a pixel-wise classification problem. And a widely used loss function is cross entropy loss. However, the objects in this task such as chiasm and optic nerve often take up small regions in the CT images. In this paper, we use the dice coefficient loss function [2,4] to optimize network parameters, which helps to constrain the multi-organ masks from the ground truth. The comparison experiments and discussions are also conducted in the following section. The dice coefficient is a measure of overlap widely used to assess segmentation performance when ground truth is available, as in Eq. 4:

$$L_{dice} = 1 - \sum_{k}^{K} \frac{2\omega_k \sum_i^N p(k,i) g(k,i)}{\sum_i^N p_{(k,i)}^2 + \sum_i N g_{(k,i)}^2}, \tag{4}$$

where N is the pixel number, $p(k,i) \in [0,1]$ and $g(k,i) \in 0,1$ denote predicted probability and ground truth label for class k, respectively. K is the class number, and $\sum_k \omega_k = 1$are the class weights. In our paper, we set $\omega_k = \frac{1}{K}$ empirically.

We use shape-aware loss [7] to take shape of organs into account. In general, all the loss function values are calculated from the pixels in the image, but shape-aware loss calculates the average point to curve Euclidean distance D among points around curve of predicted segmentation $\hat{C}$ to the ground truth C_{GT} and use it as coefficient to cross-entropy loss function. It is defined as follows:

$$\begin{aligned} E_i &= D(\hat{C}, C_{GT}), \\ L_{shape} &= -\sum_{i}^{N} [CE(\hat{y}, y) - iE_i CE(\hat{y}, y)]. \end{aligned} \tag{5}$$

Using E_i the network learns to produce a prediction masks similar to the training shapes.

The final loss function is defined as:

$$L_{loss} = L_{dice} + L_{shape} + L_{reg}. \tag{6}$$

Herein, L_{reg} represents the regularization loss (also called to weight decay) [8] used to avoid overfitting.

3 Experiments

In this section, we conduct evaluation experiments to evaluate the performance of the different methods on MICCAI 2015 Head and Neck Auto-Segmentation Challenge dataset [17]. Nine anatomical segmentation structures in the dataset are highly relevant OARs for radiation therapy treatment in the head and neck, including brainstem, mandible, chiasm, left and right optic nerves, left and right parotid glands, as well as left and right submandibular. And manual contouring data used are segmented by three different medical imaging experts. For fair comparison, all methods are trained and validated using the same data and condition settings. The predicted segmentation results are quantitatively evaluated by two widely used metrics. Furthermore, we demonstrates the outperformance of our proposed approach through segmentation visualization and ablation study.

3.1 Dataset Preprocessing

The dataset consists of 48 CT scan sequences, of which 38 cases are used as training set and 10 cases as testing set following [6]. In this work, nine anatomical structures are considered as segmentation targets, including brainstem, mandible, chiasm, bilateral optic nerves, bilateral parotid glands, and bilateral

Table 1. Dice score coefficient (%) ↑ of results by different compared methods on MICCAI 2015 dataset. The larger the value, the more accurate the segmentation.

Organ	3D U-Net	AnatomyNet	FocusNetv2	Ours
Brain Stem	0.814	0.863	0.879	**0.895**
chiasm	0.508	0.541	0.708	**0.719**
Mandible	0.801	0.921	0.940	**0.952**
Optic nerve left	0.613	0.721	0.788	**0.793**
Optic nerve right	0.608	0.691	**0.809**	0.805
Parotid glands left	0.836	0.878	0.887	**0.895**
Parotid glands right	0.802	0.872	0.892	**0.908**
Submandibular glands left	0.759	0.808	0.836	**0.842**
Submandibular glands right	0.771	0.807	0.829	**0.833**

Table 2. 95% HD score (mm) ↓ of results by different compared methods on MICCAI'15 dataset. The smaller the value, the more accurate the segmentation.

Organ	3D U-Net	AnatomyNet	FocusNetv2	Ours
Brain Stem	11.122	8.396	1.839	**0.574**
chiasm	4.418	1.741	1.144	**0.996**
Optic nerve left	3.539	2.549	2.980	**2.080**
Optic nerve right	1.157	2.827	1.909	**0.855**
Mandible	1.074	0.578	**0.511**	0.531
Parotid glands left	4.716	6.447	4.106	**3.715**
Parotid glands right	8.045	4.177	5.732	**4.108**
Submandibular glands left	5.479	2.938	1.819	**1.406**
Submandibular glands right	3.322	1.534	1.321	**0.908**

submandibular glands. We first convert the original imaging data to NIfTI format, keeping the same size 512×512 pixels with $110 - 190$ slices. And in-plane pixel spacing varied between 0.76×0.76 mm and 1.27×1.27 mm. We then normalized the data to satisfy a standard normal distribution with a mean of 0 and variance of 1. In addition, we normalized the grayscale values of the images.

3.2 Implementation Details

We implemented our model with PyTorch framework. Batch size was set to be 1 because of different sizes of whole-volume CT images. We first used SGD optimizer with momentum 0.9, learning rate 0.001 and the number of epochs being 50. Then, Adam optimizer [11] was used for training, with $\beta_1 = 0.5$ and $\beta_2 = 0.999$, and the number of epochs 600. During training process, we apply the following image augmentations to enhance the training set: random resize with

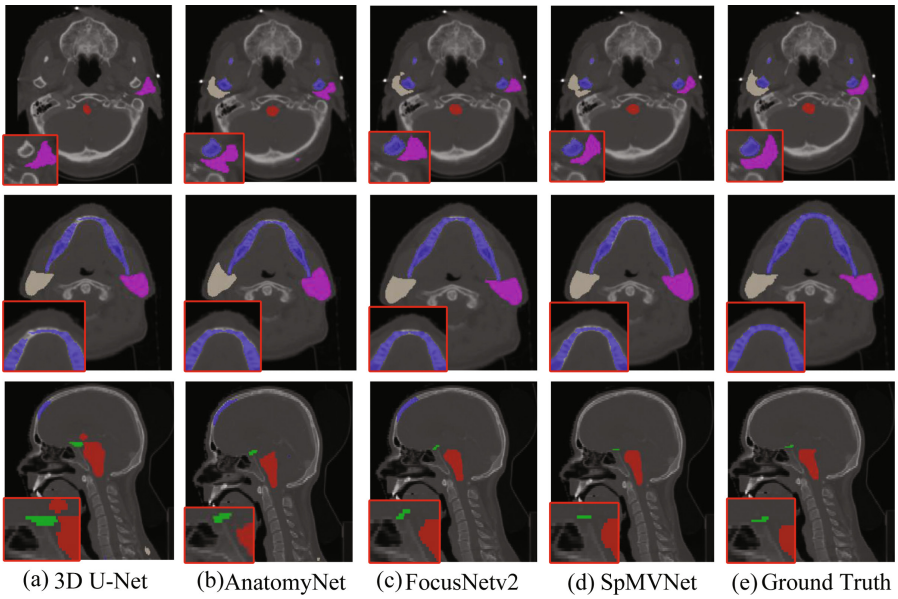

Fig. 5. Comparison of different methods for visualization on miccai dataset. (a)–(e) are the mask predicition of 3D U-Net, AnatomyNet, FocusNetv2, the proposed SpMVNet and ground truth, respectively.

Table 3. Dice score coefficient (%) ↑ of results by baseline and improved methods on MICCAI 2015 dataset. The larger the value, the more accurate the segmentation.

Organ	baseline	baseline w CCNet	baseline w SymNet	Ours
Brain Stem	0.751	0.830	0.864	**0.895**
chiasm	0.424	0.575	0.617	**0.719**
Mandible	0.742	0.847	0.881	**0.952**
Optic nerve left	0.569	0.694	0.769	**0.793**
Optic nerve right	0.608	0.627	0.696	**0.805**
Parotid glands left	0.771	0.776	0.842	**0.895**
Parotid glands right	0.726	0.799	0.852	**0.908**
Submandibular glands left	0.637	0.706	0.788	**0.842**
Submandibular glands right	0.688	0.690	0.775	**0.833**

scale range [0.5, 2.0], crop, and horizontal flipping with probability 0.5. The label images should do the same transformation as CT images. All the experiments were performed on a standard desktop with Ubuntu 16.04, using one NVIDIA GeForce RTX 3090 GPU with 24 GB memory.

Table 4. 95% HD score (mm) ↓ of results by baseline and improved methods on MICCAI 2015 dataset. The smaller the value, the more accurate the segmentation.

Organ	baseline	baseline w SymNet	baseline w CCNet	Ours
Brain Stem	14.227	5.733	3.357	**0.574**
chiasm	1.349	1.239	1.048	**0.996**
Optic nerve left	3.386	3.539	2.973	**2.080**
Optic nerve right	1.225	0.886	0.891	**0.855**
Mandible	1.234	0.974	0.612	**0.531**
Parotid glands left	6.852	5.496	3.909	**3.715**
Parotid glands right	9.645	6.594	5.394	**4.108**
Submandibular glands left	8.190	7.218	3.521	**1.406**
Submandibular glands right	1.839	1.241	1.091	**0.908**

3.3 Evaluation Metrics

In order to accurately evaluate the segmentation results, this article uses two evaluation indexes, Dice Similarity Coefficient and 95% Hausdorff Distance, to evaluate the segmentation results. They are the most common used metrics for evaluating 3D medical image segmentations and include volumeand overlap-based metric types. Multiple metrics are used because different metrics reflect different types of errors. For example, when segmentations are small, distance-based metrics such as HD are recommended over overlap-based metrics such as Dice coefficient. Overlap-based metrics are recommended if volume-based statistics are important. In the following, the metrics used are described in more detail:

The Dice coefficient measures the volumetric overlap between the automatic and manual segmentation. It is defined as:

$$Dice = \frac{2|A \cap B|}{|A| + |B|}, \tag{7}$$

where A and B are the labeled regions that are compared and |.| is the volume of a region. The Dice coefficient can have values between 0 (no overlap) and 1 (complete overlap).

The maximum HD measures the maximum distance of a point in a set A to the nearest point in a second set B. Commonly it is defined as:

$$\begin{aligned} H(A,B) &= \max(h(A,B), h(B,A)), \\ h(A,B) &= \max_{a \in A} \min_{b \in B} \parallel a - b \parallel, \end{aligned} \tag{8}$$

where ||.|| is the Euclidean distance, a and b are points on the boundary of A and B, and h (A, B) is often called the directed HD. It should be mentioned that maximum HD is sensitive to outliers but appropriate for nonsolid segmentations.

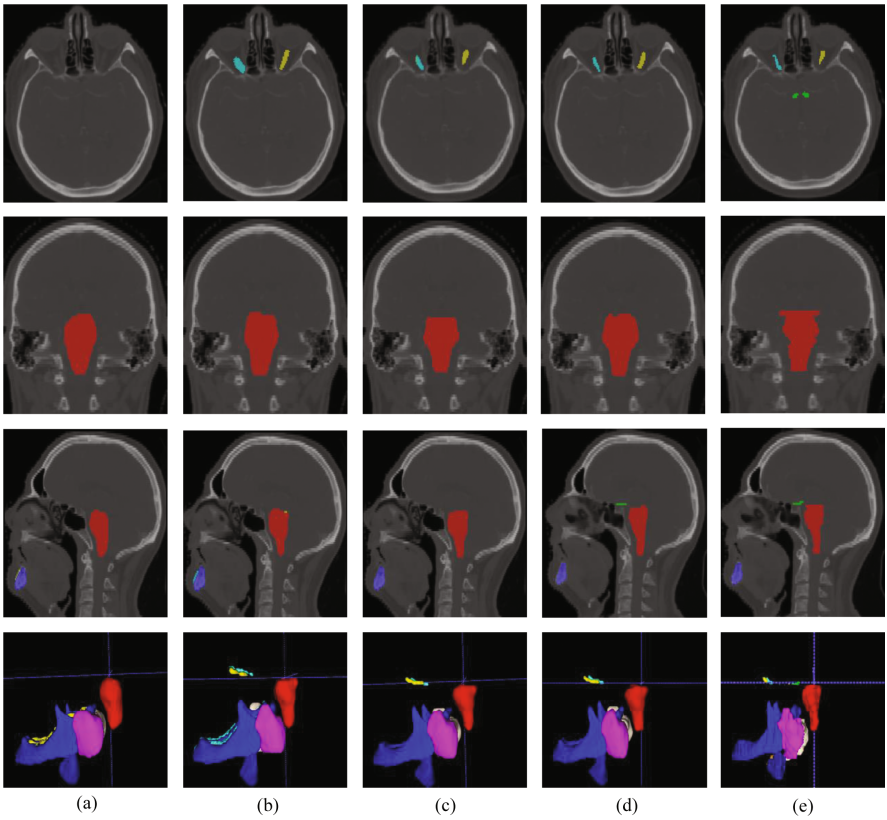

Fig. 6. Results of ablation experiments of our method. From top to bottom are the axial, coronal and sagittal views as well as 3D mask results. (a)–(e) are the mask prediction of baseline, baseline with CCNet, baseline with SymNet, our proposed method and ground truth, respectively.

The 95% HD is similar to maximum HD. However, in contrast to maximum HD, 95% HD is based on the calculation of the 95th percentile of the distances between boundary points in A and B. The purpose for using this metric is to eliminate the impact of a very small subset of inaccurate segmentations on the evaluation of the overall segmentation quality.

3.4 Quantitative Comparison

We compared our framework with three head and neck relevant segmentation methods, including 3D U-Net [3], AnatomyNet [23] and FocusNetv2 [6]. Note that we used the official code and results of 3D U-Net [3], AnatomyNet [23] as well as FocusNetv2 [6]. Compared with current state-of-the-art methods, our approach achieves effective improvements in the quantitative metrics of most OARs.

Table 1 and Table 2 shows the quantitative comparison of these methods. Most of the compared algorithms achieved above 0.7 on the Dice score coefficient of organs at risk except for chiasm. 3D U-Net [3] treats small and large organs equally, which will affect the segmentation results on small objects, and even organs with symmetrical structures are far inferior to other methods. Compared with FocusNetv2 [6], our framework achieves better performance on most organs at risk without using a complex multiple network architecture, corroborating that our strategy has the full capability to draw out the rich information from the CT data.

3.5 Qualitative Comparison

As shown in Fig. 5, our method shows the best visualized on parotid gland and optic nerve. As can be seen from the axial views in the first two rows, 3D U-Net [3] cannot identify this OARs, thus losing the information of OARs segmentation, and the segmentation on the mandible is discontinuous, so the OARs cannot be segmented completely. The problem of discontinuous segmentation also exists in AnatomyNet [23] and FocusNetv2 [6], and the segmentation information is incorrectly labeled at the position of the crania, which affects the segmentation results. Compared other methods, our method exploits the feature information of symmetrical OARs in the head and neck to help train network better to approximate the reference labels on left and right parotid. Our method extracts spatial context structure information and obtains convincing continuous OARs segmentation masks, which can achieve better segmentation results. Therefore, Our method is able to produce higher-quality OARs segmentation masks compared with other methods.

3.6 Ablation Study

We design ablation experiments to verify the effectiveness of each part of the proposed method, as shown in Fig. 6. Firstly, the baseline method is to replace SymNet and CCNet with U-Net [19] and 3D U-Net [3] for segmentation of three views. However, the visualization results show that the baseline network does not have sufficient ability to recognize some obvious OARs structures leading to poor segmentation results.

Then we add our proposed SymNet and CCNet to the baseline model to analyze the continuity and symmetry of the segmentation results. Figure 6(b) and (c) show that CCNet and SymNet can make up for the lack of spatial context information and the inability to identify symmetric organs in the baseline method. It can be seen from the Fig. 6(d) that our proposed method is closer to the ground truth and the brainstem segmentation is more complete, but there is still the problem that small OARs cannot be identified.

Similar to the comparison method, we also calculated dice score coefficient and 95% HD scores in the ablation experiments, as shown in Table 3 and Table 4. Our method can take advantage of SymNet and CCNet to achieve promising results.

4 Conclusion

In this paper, we propose an end-to-end spatial multi-view network segmentation framework SpMVNet. Focusing on head and neck CT images, we explore a SymNet to combine multi-view probabilistic symmetry maps for mask predicition of specific organ volumes symmetrically distributed along the midline of the brain. The method innovatively improves the siamese network for OARs segmentation and takes the 2D slices on the left and right sides as input, and then synthesizes the 3D segmentation prediction results. We also solve the problem of lack of continuity in the segmentation of some OARs and achieve higher segmentation metrics through CCNet. We also reduce the segmentation errors of existing methods for OARs, and achieve a certain improvement in the accuracy of symmetric OARs segmentation. The evaluation results demonstrate the effectiveness of the proposed method in our paper. In this paper, an effective method is proposed to solve the difficulties of organ endangerment in radiotherapy, which will be helpful to the analysis and processing of biological information.

Acknowledgment. The work was supported in part by the State Key Project of Research and Development Plan under Grants 2022YFC2401600 and 2022YFC2408500, in part by the National Natural Science Foundation of China under Grant T2225025, in part by the Key Research and Development Programs in Jiangsu Province of China under Grant BE2021703 and BE2022768.

References

1. Cai, J., Lu, L., Zhang, Z., Xing, F., Yang, L., Yin, Q.: Pancreas segmentation in MRI using graph-based decision fusion on convolutional neural networks. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 442–450. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_51
2. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell. **40**(4), 834–848 (2017)
3. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 424–432. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_49
4. Crum, W.R., Camara, O., Hill, D.L.: Generalized overlap measures for evaluation and validation in medical image analysis. IEEE Trans. Med. Imaging **25**(11), 1451–1461 (2006)
5. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. IEEE (2009)
6. Gao, Y., et al.: Focusnetv 2: imbalanced large and small organ segmentation with adversarial shape constraint for head and neck CT images. Med. Image Anal. **67**, 101831 (2021)

7. Hayder, Z., He, X., Salzmann, M.: Shape-aware instance segmentation. arXiv preprint arXiv:1412.6980 (2016)
8. Hoerl, A.E., Kennard, R.W.: Ridge regression: biased estimation for nonorthogonal problems. Technometrics **12**(1), 55–67 (1970)
9. Hu, P., Wu, F., Peng, J., Bao, Y., Chen, F., Kong, D.: Automatic abdominal multi-organ segmentation using deep convolutional neural network and time-implicit level sets. Int. J. Comput. Assist. Radiol. Surg. **12**(3), 399–411 (2017)
10. 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)
11. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
12. Koch, G., Zemel, R., Salakhutdinov, R., et al.: Siamese neural networks for one-shot image recognition. In: ICML Deep Learning Workshop, Lille, vol. 2 (2015)
13. Milletari, F., et al.: Hough-CNN: deep learning for segmentation of deep brain regions in MRI and ultrasound. Comput. Vis. Image Underst. **164**, 92–102 (2017)
14. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571. IEEE (2016)
15. World Health Organization: World health statistics 2019: monitoring health for the SDGs, sustainable development goals. World Health Organization (2019)
16. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. **9**(1), 62–66 (1979)
17. 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)
18. Ren, X., et al.: Interleaved 3D-CNNs for joint segmentation of small-volume structures in head and neck CT images. Med. Phys. **45**(5), 2063–2075 (2018)
19. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
20. Tong, N., Gou, S., Yang, S., Ruan, D., Sheng, K.: 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)
21. Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T.S.: Free-form image inpainting with gated convolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4471–4480 (2019)
22. Zhu, Q., Du, B., Turkbey, B., Choyke, P.L., Yan, P.: Deeply-supervised CNN for prostate segmentation. In: 2017 International Joint Conference on Neural Networks (IJCNN), pp. 178–184. IEEE (2017)
23. Zhu, W., et al.: Anatomynet: deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. Med. Phys. **46**(2), 576–589 (2019)