Chinese Journal of Electronics
vol. 33, no. 3, pp. 766–775, May 2024
https://doi.org/10.23919/cje.2022.00.306

RESEARCH ARTICLE

The Squeeze & Excitation Normalization Based nnU-Net for Segmenting Head & Neck Tumors

Juanying XIE, Ying PENG, and Mingzhao WANG

School of Computer Science, Shaanxi Normal University, Xi'an 710119, China

Corresponding author: Juanying XIE, Email: xiejuany@snnu.edu.cn
Manuscript Received September 9, 2022; Accepted December 13, 2022

Abstract — Head and neck cancer is one of the most common malignancies in the world. We propose SE-nnU-Net by adapting SE (squeeze and excitation) normalization into nnU-Net, so as to segment head and neck tumors in PET/CT images by combining advantages of SE capturing features of interest regions and nnU-Net configuring itself for a specific task. The basic module referred to convolution-ReLU-SE is designed for SE-nnU-Net. In the encoder it is combined with residual structure while in the decoder without residual structure. The loss function combines Dice loss and Focal loss. The specific data preprocessing and augmentation techniques are developed, and specific network architecture is designed. Moreover, the deep supervised mechanism is introduced to calculate the loss function using the last four layers of the decoder of SE-nnU-Net. This SE-nnU-net is applied to HECKTOR 2020 and HECKTOR 2021 challenges, respectively, using different experimental design. The experimental results show that SE-nnU-Net for HECKTOR 2020 obtained 0.745, 0.821, and 0.725 in terms of Dice, Precision, and Recall, respectively, while the SE-nnU-Net for HECKTOR 2021 obtains 0.778 and 3.088 in terms of Dice and median HD95, respectively. This SE-nnU-Net for segmenting head and neck tumors can provide auxiliary opinions for doctors' diagnoses.

Keywords — Head and neck tumors, PET/CT images, Image segmentation, SE-nnU-Net, Squeeze and excitation normalization.

Citation — Juanying XIE, Ying PENG, and Mingzhao WANG, "The Squeeze & Excitation Normalization Based nnU-Net for Segmenting Head & Neck Tumors," *Chinese Journal of Electronics*, vol. 33, no. 3, pp. 766–775, 2024. doi: 10.23919/cje.2022.00.306.

I. Introduction

Head and neck cancer refers to a group of malignant tumors occurring in the head and neck, including oropharyngeal cancer, hypopharyngeal cancer, laryngeal cancer, nasopharyngeal cancer, and salivary gland cancer, et cetera. It is the seventh common cancer in the world, with 890000 diagnoses and more than 450000 deaths [1]. It is very challenging to diagnose and make appropriate treatment to the head and neck tumors, due to the complex anatomical structure and the heterogeneity of the head and neck tumors. The current available treatment methods are radiotherapy, surgery and chemotherapy. Detecting or marking the tumor location is prerequisite to take treatment using medical images. However, manual labeling the tumor location is not only time-consuming and laborious, but also depends on the subjective judgments and professional experiences of doctors, leading to unstable [2]. Therefore, the automatic segmentation method is expected urgently, due to its property of reducing the workload of doctors, reducing the differences between the labeling results from different doctors, and realizing the automatic segmentation process, and facilitating it possible for the routine clinical use [3].

PET/CT (Positron emission tomography/Computed tomography) is an important imaging tool for clinical diagnosis, staging and prognosis for head and neck cancers. This technology includes PET imaging and CT imaging. It can reflect the metabolic and structural changes of the lesion area. Wu *et al*. [4] detected the candidate lesion areas with regard to the features of PET and CT images and the prior knowledge of anatomical features of NPC (nasopharyngeal carcinoma) patients, and then classified them as benign/malignant tumors by SVM (support vector machines). The experimental results show that the system has obtained high accuracy and good clinical application prospect. Song *et al*. [5]

Associate Editor: Prof. Chenggang YAN, Hangzhou Dianzi University.

modeled the segmentation problem as a minimization problem of a Markov random field model. It coded the PET and CT images, respectively, then segmented them by constructing subgraphs. The optimization problem is solved through graph cutting method. In order to obtain the consistent results in both types of images, context arcs are added between the two sub-images, so as to achieve adaptive context cost. This method was tested on PET/CT images of 23 lung cancer patients, and it achieved better results than that using PET or CT images alone. Moe *et al*. [6] proposed an automatic segmentation method for primary tumors and metastatic lymph nodes in head and neck based on 2D U-Net. The model was trained on 142 cases and tested on 40 cases. The experimental results show that this method performed as similar as the clinician level. On the basis of this study, Andrearczyk *et al*. [7] studied the differences between four strategies for the automatic segmentation tasks of head and neck tumors based on 2D/3D V-Net, including only using PET or CT images, early fusion of PET and CT images, and late fusion of them. The experimental results show that the PET and CT image late fusion strategy can achieve better results than other three strategies. In order to better integrate information at different scales, Yuan *et al*. [8] proposed a dynamic scale attention module which incorporates low-level details with high-level semantics from feature maps at different scales. Each decoding layer merged the output features from all encoding layers, so as to capture both fine-grained details and coarse-grained semantic information at full scales. The weights of features on each scale are adjusted adaptively to highlight important scales and suppress irrelevant ones when the model learns. Ma *et al*. [9] proposed an automatic segmentation method for head and neck tumors based on 3D U-Net and hybrid active contour. A multi-channel 3D U-Net was used to segment tumors in PET/CT images. Then the uncertainty of segmentation was estimated by ensemble model, and a segmentation quality score was defined to select cases with high uncertainty. Finally, a hybrid active contour model was adopted to refine those segmentation results of cases with high uncertainty. Experimental results show that this method can achieve good segmentation results. Chen *et al*. [10] proposed a deep learning based framework to iteratively refine the segmentation results of head and neck tumors. There are multiple 3D U-Nets in this framework. The features and predictive results learned by the upstream model are used as the additional information for the next model to further refine and improve the segmentation results. Experimental results show that the iterative refinement strategy can improve the performance of the model.

Most of the existing methods are appropriate to specific task, such that they may not perform well when facing new tasks. The nnU-Net [11] proposed recently combines multiple U-Net methods [12], and it can automatically adjusts itself to use multiple U-Net architectures, such that it can model any given type of input images. It is reported that the nnU-Net outperforms most of the existing approaches without any human intervention [11]. Therefore, nnU-Net was investigated for segmenting early ischemic changes on non-contrast computed tomography in patients with acute ischemic stroke [13]. The conditional nnU-Net was proposed for multiorgan segmentation from partially labeled datasets [14].

SE (squeeze-and-excitation) block was proposed for adaptively recalibrating channel-wise feature responses by explicitly modelling interdependencies between channels [15]. It was incorporates into U-Net for segmenting prostate zonal of multi-institutional MRI datasets [16]. It was also embedded into 3D U-Net to highlight tumor regions for segmenting lung tumors in CT images [17]. SE normalization [18] integrates instance normalization and SE attention module, i.e. SE block, such that it can capture the features of the interesting regions and has good generalization on multi-center data. Therefore, this paper proposed SE-nnU-Net by combining SE normalization into nnU-Net, along with other improvements, so as to advance the nnU-Net model for segmenting tumors in PET/CT images of the head and neck cancer patients from multi-centers.

The organizational structure of this paper is as follows: Section II introduces the data and methods, including the experimental data sets, SE normalization, and SE-nnU-Net network structure; Section III presents the experimental results and the analyses. Section IV comes the conclusions.

II. Data and Methods

1. Data

The data used in the experiments in this paper are from the HEad and neCK TumOR segmentation challenge (HECKTOR) [19]. The challenge presents a large dataset of PET/CT images from the patients with head and neck cancer. HECKTOR 2020 dataset contains 201 cases in training data set. They are from HGJ (Hôpital Général Juif), CHUS (Centre Hospitalier Universitaire de Sherbooke), HMR (Hôpital Maisonneuve-Rosemont), and CHUM (Centre Hospitalier de ll'Université de Montréal) from Canada, respectively, of 55, 72, 18 and 56 cases. There are 53 cases in testing data set. They are from CHUV (Centre Hospitalier Universitaire Vaudois) in Switzerland. Each case comprises PET/CT images in NIfTI format, the bounding box of the oropharyngeal region in CSV format, and patient gender and the age. The detail information of data collection and the examples from each data collection center in HECKTOR 2020 data set are shown in Table 1 and Figure 1, respectively.

As can be seen from Table 1, the manufacturers and models of scanning equipment are not same completely used by each data collection center in HECKTOR 2020 data set. The testing data from CHUV in Switzerland

Table 1 Details of each data collection center of the HECKTOR 2020 data

Center name	Center location	Equipment type	# Samples
HGJ	Canada	Discovery ST, GE Healthcare	55
CHUS	Canada	Gemini GXL 16, Philips	72
HMR	Canada	Discovery STE, GE Healthcare	18
CHUM	Canada	Discovery STE, GE Healthcare	56
CHUV	Switzerland	Discovery D690 TOF, GE Healthcare	53

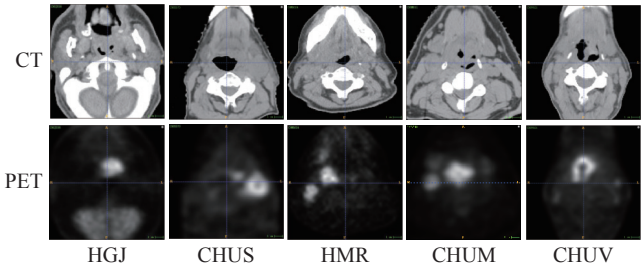

Figure 1 Examples of PET/CT images from different centers for HECKTOR 2020 data set.

are collected by using different scanning equipment compared to that used by centers in Canada for collecting training data. In addition, it can be seen from Figure 1 that there are differences in the quality of PET and CT images collected by each center.

HECKTOR 2021 data set are obtained by adding 71 cases from the CHUP (Centre Hospitalier Universitaire de Poitiers) to HECKTOR 2020 data set, where there are 23 cases being assigned to training data set, and 48 in testing data set. The scanning equipment used by the CHUP center is the Siemens Biograph mCT 40 ToF. Figure 2 shows an example of the PET/CT images from CHUP center for HECKTOR 2021 data set.

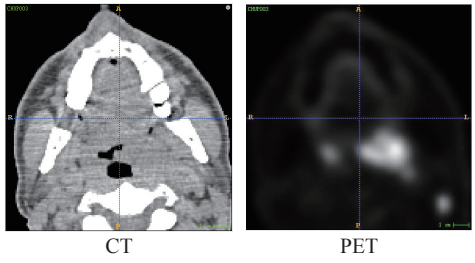

Figure 2 The example of PET/CT images from CHUP center for HECKTOR 2021 data set.

2. Data processing

In order to reduce the computing load, PET/CT images were cropped with regard to the bounding box of the oropharyngeal region. There are heterogeneous voxel spacing in the HECKTOR data, due to the different scanning equipment or the different protocols for collecting data. Hence, we resampled the training images and the corresponding labels to 1.0 mm × 1.0 mm × 1.0 mm, such that the size of the processed images became 144 × 144 × 144, so do the labels.

To highlight the related information of tumors while suppressing the background information, the gray values of CT images in the training data are constrained in $[-200, 200]$. Finally, Z-score method was used to normalize the PET and CT images.

3. Data augmentation

Data augmentation technology refers to the processing of the original images, so as to expand the data set. It can make up the lack of the training data, and avoid overfitting of the model to the training data while enhancing the generalization capability of the model. The following data augmentation methods are used in the experiments in this paper: such as rotation, scaling, mirroring, brightness transformation, contrast enhancement, gamma correction, Gaussian noises and Gaussian blurring. The rotation angle is in $[-30^{\circ}, +30^{\circ}]$. The scaling ratio is in [0.7, 1.4]. The Gaussian noise is in [0, 1]. The Gamma correction is in [0.7, 1.5].

4. SE normalization

There are many layers in deep neural networks. The parameter variations of each layer will lead to the distribution variation of the input data of the next layer. Hence, there is a dramatic variation in data distribution at the very deeper layer if there is a shift in the data distribution at the shallow layer. The deeper the layer, the more dramatic data distribution variation will happen, so that the model optimization becomes difficult or even impossible to realize. Therefore, normalization is a very important in deep neural networks. Iantsen *et al.* [18] put forward the SE normalization method by integrating the instance normalization and SE attention module, so as to capture the features of the interesting region. It was reported that the SE normalization has obtained good capability of generalization in several segmentation tasks such as the brain tumor segmentation [18] and the head and neck tumor segmentation [20]. The structure of SE normalization is shown in Figure 3. The left branch is the instance normalization, and the middle and right branches are both the SE attention modules with Sigmoid and Tanh activation functions at the last layer, respectively. The SE normalization is different from the regular batch normalization and the instance normalization. The output of the middle and right branches of the SE normalization are calculated by using equations (1) and (2), respectively. They are both the functions of input X. But the γ_i and β_i learned in the training process of the batch or instance normalization are fixed during inference process and independent of input X.

$$\gamma = f_{\gamma}(X) \tag{1}$$

$$\beta = f_{\beta}(X) \tag{2}$$

where, $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_N)$, $\beta = (\beta_1, \beta_2, \ldots, \beta_N)$, and f_{γ}

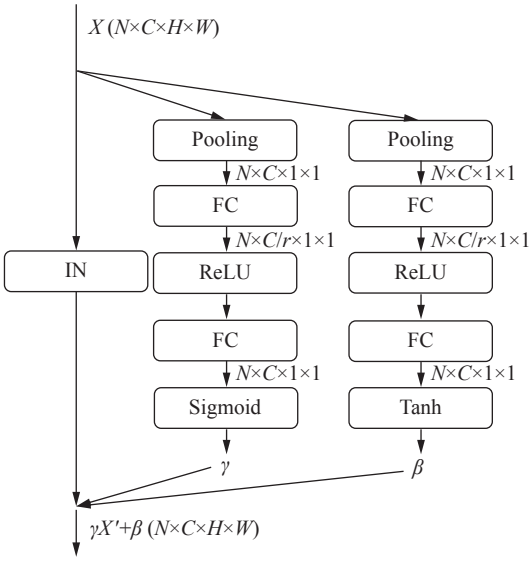

Figure 3 SE normalization.

is the output of SE attention module adopting the Sigmoid activation function at the last layer of the middle branch in Figure 3, and f_β is the output of the SE attention module with Tanh activation function of the last layer of the right branch of Figure 3.

5. Network structure of SE-nnU-Net

U-Net combines low and high resolution information, so that it can obtain high generalization capability without large amount of training data [12]. Therefore, it has been widely used in the medical image segmentation field. However, some available U-Net variations, such as those combining dense connectivity, attention mechanism, and et cetera, are over-suitable to specific tasks to some extent, or may be affected by imperfect verification, so that they are not appropriate to new tasks and cannot get good performance for the new tasks. To solve this problem, Isensee *et al*. [11] extensively studied the expert-driven configuration methods and the data-driven deep learning methods, and designed the nnU-Net method. This nnU-Net can automatically configure itself, including preprocessing, network architecture, training, and post-processing for any new task in the biomedical domain, so that the search space is greatly reduced for the empirical design choices when facing a new task. In addition, it is very fast for this nnU-Net to realize the automatic configuration with only executing the simple rules and a small number of empirical decisions without additional computational costs, so that it can be used as an out-of-the-box tool. SE normalization [18] can better capture the features of the interesting region. It can bring about significant improvements in segmentation quality, such as in brain tumor segmentation [18] and head and neck tumor segmentation [20].

With respect to the advantages of the nnU-Net and the SE normalization, and with the task in mind to segment the head and neck tumors in the PET/CT images, collecting by the multiple centers from different countries using different types of scanning devices for the head and neck cancer patients, the SE-nnU-Net method is proposed by integrating SE normalization into the nnU-Net while combining some specific improvements, such as designing Conv-ReLU-SE module and the network architecture, and combining Dice loss and Focal loss as loss function, and introducing the deep supervise mechanism, et cetera, so as to segment the head and tumors as accurate as possible. The structure of the SE-nnU-Net is shown in Figure 4.

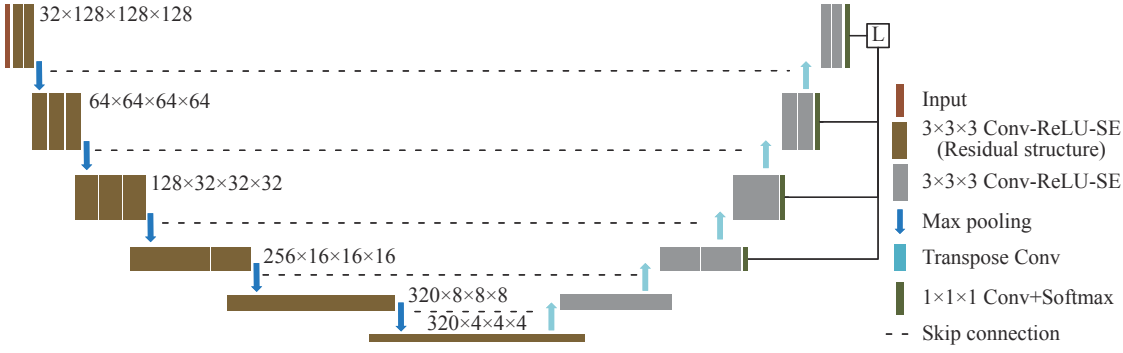

Figure 4 The architecture of our SE-nnU-Net network.

The basic module in our SE-nnU-Net is the 3 × 3 × 3 convolution-ReLU-SE module. It should be noted that there is residual structure in the 3 × 3 × 3 convolution-ReLU-SE module in the encoder of the SE-nnU-Net. There are five down samplings in the encoder of the SE-nnU-Net. These down samplings are achieved by max pooling. The decoder of the SE-nnU-Net is composed of the stacked 3 × 3 × 3 convolution-ReLU-SE modules without residual structure.

The output feature of each down sampling is skip connected as the input feature of the corresponding up sampling. There are 32 initial feature channels in the SE-nnU-Net. The number of channels is doubled with each down sampling, up to the maximum of 320 channels. The number of channels is halved with each up sampling. The output of each layer of decoder is up sampled using a transpose convolution with a stride of 2 × 2 × 2 .

The loss function L is calculated by using the last four layers of the decoding part of the SE-nnU-Net, due to the lots of noises in the feature maps of the shallow layers of the network. Specifically, labels are down sampled according to the ratios of {1, 2, 4, 8} during the process of online data enhancement process. Then, the last four-layer outputs of the decoding part are input into a convolution layer with two filters and the convolution kernel of the size of $1 \times 1 \times 1$. Finally, the Softmax activation function is used to obtain the probability map of segmentation, so as to calculate the error of the segmentation using the weights of $\frac{8}{15}$, $\frac{4}{15}$, $\frac{2}{15}$ and $\frac{1}{15}$ for the last four-layer outputs, respectively. The loss of the last output layer together with the auxiliary losses of the middle layers simultaneously stimulate the gradient back-propagation. The parameters can be efficiently updated at each iteration of the model.

6. Loss function

The loss function L is the sum of Dice loss (L_{Dice}) and Focal loss (L_{Focal}). It is calculated in equation (3). Where, L_{Dice} and L_{Focal} are calculated in equations (4) and (5), respectively. The N is the number of pixels in the segmentation map encoded using one-hot, and C is the number of classes. The y_i^c is a sign function. Its value is 1 when the true category of the ith pixel labeled y is c; otherwise, its value is 0. The p_i^c is the probability of the ith pixel of the segmentation map predicted by the model belonging to the category c. The L_{Focal} will become the cross-entropy loss when $\gamma = 0$. The L_{Focal} will focus on samples which are difficult to classify, especially for imbalanced data, it will pay attention to the minority class. This is why we introduce it into loss function of the SE-nnU-Net. The parameter α can make a tradeoff between the importance of positive and negative samples. The parameters α and γ are 0.25 and 2, respectively, in the Focal loss function in this study.

$$L = L_{\text{Dice}} + L_{\text{Focal}} \tag{3}$$

$$L_{\text{Dice}} = -\frac{2}{C}\sum_{c=1}^{C}\frac{\sum_{i=1}^{N} p_i^c y_i^c}{\sum_{i=1}^{N} p_i^c + \sum_{i=1}^{N} y_i^c + \epsilon} \tag{4}$$

$$L_{\text{Focal}}(y,p) = \begin{cases} -\dfrac{1}{N}\displaystyle\sum_{c=1}^{C}\sum_{i=1}^{N}\alpha(1-p_i^c)^{\gamma}\log(p_i^c), & y_i^c = 1 \\ -\dfrac{1}{N}\displaystyle\sum_{c=1}^{C}\sum_{i=1}^{N}(1-\alpha)p_i^{c\gamma}\log(1-p_i^c), & y_i^c = 0 \end{cases} \tag{5}$$

III. Experimental Results and Analyses

1. Evaluation metrics

To quantify the segmentation performance of the model, the Dice [21], Precision, Recall, and median HD95 (Huasdorff 95 Distance) [22] are adopted to value the segmentation results of each model. Dice (Dice similarity coefficient) is one of the most common metrics used to evaluate the performance of segmentation models, measuring volumetric overlap between segmentation results and annotations. It is a good evaluation metric of segmentation for imbalanced segmentation problems, i.e., the region to segment is small as compared to the image size. It is commonly used in the evaluation and ranking of segmentation algorithms and particularly tumor segmentation tasks [21]. Its domain is [0, 1]. The bigger the Dice value, the better is the segmentation result of models, and vice versa. Let P be the segmentation results predicted by the model, and GT the marked region by experts, then the Dice is calculated in equation (6).

$$\text{Dice} = \frac{2|P \cap \text{GT}|}{|P| + |\text{GT}|} \tag{6}$$

Assume TP represents the number of positive samples that are correctly predicted by model, and FP the number of negative samples incorrectly predicted as positive samples by the model, and FN the number of positive samples incorrectly predicted as negative samples by the model, then the Precision and Recall metrics can be defined in equations (7) and (8), respectively.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

Let $A = \{a_1, a_2, a_3, \ldots\}$, $B = \{b_1, b_2, b_3, \ldots\}$, then HD (Huasdorff Distance) metric is to measure the similarity between A and B. It is defined as follows.

$$H(A,B) = \max(h(A,\ B),\ h(B,\ A)) \tag{9}$$

$$h(A,\ B) = \max_{a\in A}(\min_{b\in B}\|a-b\|) \tag{10}$$

$$h(B,\ A) = \max_{b\in B}(\min_{a\in A}\|b-a\|) \tag{11}$$

where, $\|\cdot\|$ is a kind of distance, such as $L1$ norm, Euclidean distance, or any other distance. $H(A,\ B)$ is also called the bidirectional HD. It is the basic form of HD while $h(A,\ B)$ is called the single directional HD from set A to set B, and $h(B,\ A)$ is the one-way HD from set B to A.

The Euclidean distance is adopted in this paper to measure the distance between sets A and B. Here A and B will be P and GT, respectively. To exclude the unreasonable distance caused by some outliers, the median

HD95 metric is adopted in this paper to measure the segmentation results of models. HD95 is the 95th quantile of the distances between sets A and B. The smaller the HD95 value, the more similar between sets A and B, meaning the better segmentation results obtained by the model; otherwise the more dissimilar between sets A and B, and the worse segmentation results obtained by the model. The median HD95 is the median of the value of HD95 metric on all test samples.

2. Experimental design

The experiments are conducted on the following environments, including the Linux operating system, and the NVIDIA GeForce RTX 3090 GPU with 24GB video memory used to accelerate the training process. The codes are written based on that of the nnU-Net [11]. The repository of the nnU-Net is https://github.com/MIC-DKFZ/nnUNet. Table 2 shows the specific experimental environment configurations.

Table 2 Experimental environment settings

Names	Descriptions
Operating systems	Linux
CPU	Intel(R) Xeon(R) Silver 4214R CPU @ 2.40 GHz
Video card	NVIDIA GeForce RTX 3090
CUDA version	11.1
Programming language	Python 3.7
Deep learning framework	Pytorch 1.8.0

The stochastic gradient descent (SGD) optimizer with momentum is used to optimize the network. The momentum is set to 0.99. The initial learning rate is set to 0.01, and the weight decay coefficient is 3×10^{-5}. The image patch size is $128 \times 128 \times 128$. The batch size is 2. The total number of iterations is 1000. The learning rate is dynamically updated by using polyLR [23]. The updating equation is shown as

$$\alpha = \alpha_0 \times \left(1 - \frac{e}{T}\right)^{0.9} \tag{12}$$

where, α_0 indicates the initial learning rate, and e is the current iteration number, and T is the total iterations.

3. Experiments on HECKTOR 2020 data

This subsection will adopts the leaving one center out method to partition the HECKTOR 2020 training data into training subset and validation subset. The experiments are carried out on the training subset. The trained model is tested on the validation subset. The experimental results of our SE-nnU-Net on HECKTOR 2020 training data, that is, on the validation subset are displayed in Table 3. Figure 5 shows several examples of the segmentation results of our SE-nnU-Net on the validation subsets, where the ground truth (GT) label is shown in green color and the predictive result (PR) by our SE-nnU-Net is shown in red color. The CT in Figure 5 indicates the original CT images. The four trained SE-nnU-Net models were used jointly to do the prediction to the HECKTOR 2020 test data. In order to improve the prediction accuracy, the data augmentation technique TTA (test time augmentation) is adopted to mirror the x, y, z-axes in the prediction process. The results obtained by SE-nnU-Net on HECKTOR 2020 test data are compared to those obtained by other winning teams taking part in the HECKTOR 2020 challenge. The results of Top 10 teams are shown in Table 4.

Table 3 Leave-one-out experiment results of SE-nnU-Net on the HECKTOR 2020 training data

Center of validation subset	Dice	Median HD95	Precision	Recall
HGJ	**0.784**	5.997	0.753	0.847
CHUS	0.771	**5.423**	0.754	**0.851**
HMR	0.680	10.806	0.764	0.662
CHUM	0.703	7.498	**0.786**	0.734
Average	0.735	7.431	0.764	0.774

Note: bold fronts indicate best results.

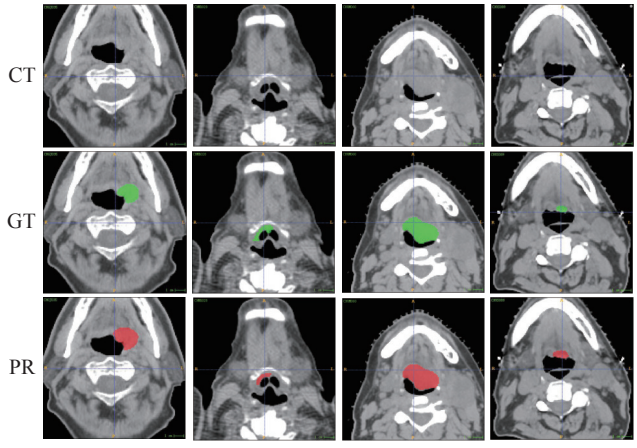

Figure 5 Examples of segmentation results of our SE-nnU-Net on the validation subset of HECKTOR 2020 challenge.

Table 4 Comparison of segmentation results between SE-nnU-Net and the Top 10 teams on HECKTOR 2020 test data

Teams/models	Dice	Precision	Recall
Iantsene *et al.* [20]	**0.759**	0.833	**0.740**
Ma *et al.* [9]	0.752	0.838	0.717
Xie *et al.* [24]	0.735	0.833	0.702
Yuan *et al.* [8]	0.732	0.785	0.732
Chen *et al.* [10]	0.724	**0.848**	0.670
Ghimire *et al.* [25]	0.691	0.753	0.693
Yousefirizi *et al.* [26]	0.668	0.729	0.716
Zhu *et al.* [27]	0.644	0.694	0.701
Naser *et al.* [28]	0.637	0.755	0.628
Rao *et al.* [29]	0.587	0.656	0.614
Our SE-nnU-Net	0.745	0.821	0.725

Note: bold fronts indicate best results.

It can be seen from Table 3 that the best results obtained by SE-nnU-Net on the validation subset are 0.784, 5.423, 0.786 and 0.851, in terms of Dice, median HD95, Precision and Recall, respectively. These four validation subsets are obtained by using the leave-one-center-out partitioning method. The average results were 0.735, 7.431, 0.764 and 0.774 in terms of Dice, median HD95, Precision and Recall, respectively. It can be noticed that the best result on each metric comes from different SE-nnU-Net model, which should be due to the partitioning method for the training data that makes the data from the validation subset comes from different data collection center compared to that from the training subset.

The results in Figure 5 show that our SE-nnU-Net network can identify the tumors in the CT images. Tough there are over-segmentation phenomena in the images in the 1st and 4th column results, while there are under-segmentation phenomena in the images in both the 2nd and 3rd column results. However, the segmentation results of our SE-nnU-Net on the validation subset are competitive, which can be seen from the quantity results in Table 3. Moreover, the segmentation results of Top 10 teams on HECKTOR 2020 test data shown in Table 4 will further prove the capability of the SE-nnU-Net.

The results in Table 4 show that the team of Isantsene *et al*. has obtained the best results of 0.759 and 0.740 in terms of Dice and Recall, respectively, on HECKTOR 2020 test data. The best Precision result of 0.848 has obtained by the team of Chen *et al*. on the HECKTOR 2020 test data. We obtained the segmentation results of 0.735, 0.833 and 0.702 in terms of Dice, Precision and Recall, respectively, ranked in the 3rd place. The predictive results of our SE-nnU-Net are 0.745, 0.821 and 0.725 in terms of Dice, Precision and Recall, respectively. Although SE-nnU-Net is slightly better than the method we designed and submitted for HECKTOR 2020 challenge [24] in terms of Dice and Recall, the rank is not changed in terms of Dice metric, the only one evaluation metric adopted by HEACKTOR 2020 challenge.

The results in Table 4 also show that the Dice values of Top 5 teams are over 0.7 while the other 5 teams within Top 10 are less than 0.7.

It should be noted that the differences between the SE-nnU-Net and the method we designed for HECKTOR 2020 [24] are as follows. The basic module we designed for the SE-nnU-Net adopts SE normalization and ReLU instead of the instance normalization and Leaky ReLU. The Focal loss instead of weighted cross-entropy loss combines Dice loss to design loss function for SE-nnU-Net, due to Focal loss is more appropriate for imbalanced data. Furthermore, SE-nnU-Net adopts deep supervision mechanism instead of integrating the scSE (spatial and channel SE) block at the end of each layer of the Decoder and Encoder. The image patch size and data preprocessing method for SE-nnU-Net are different to that in [24] for HECKTOR 2020 challenge. The last but not the least is that the methods for partitioning training data for the model in Reference [24] are different from that for SE-nnU-Net in this paper for HECKTOR 2020 challenge.

The above thorough analyses to the differences between the model in Reference [24] and the SE-nnU-Net in this paper disclose the reasons why the results of the SE-nnU-Net are superior to that of the method in reference [24] for HECKTOR 2020 challenge. This not only indicates that SE normalization significantly advances the performance of the nnU-net model, but also indicates that the hyper-parameters of a model will definitely influence its performance heavily.

4. Experiments on HECKTOR 2021 data

This subsection will show the experiments of our SE-nnU-Net on the HECKTOR 2021 challenge data. The 5-fold cross-validation experiments are adopted on HECKTOR 2021 training data, so the training data of HECKTOR 2021 are randomly partitioned into five subsets, and one of the subsets is selected as the validation subset while the rest four subsets together are used as the training subset to train the SE-nnU-Net model, till each subset is used as the validation subset. The 5-fold cross validation experimental results are shown in Table 5. Figure 6 shows several cases of the segmentation results of our SE-nnU-Net on validation subsets, where green color means the true label and red color indicates the predictive results of our SE-nnU-Net. The five trained SE-nnU-Net models were used jointly to predict the segmentation results for the HECKTOR 2021 test data. In

Table 5 Results of 5-fold cross-validation experiments on training data of HECKTOR 2021

Validation subset	Dice	Median HD95	Precision	Recall
Fold_1	0.731	9.324	0.709	0.859
Fold_2	0.745	8.787	0.741	0.789
Fold_3	0.764	**5.933**	0.776	0.832
Fold_4	**0.803**	6.733	**0.783**	**0.869**
Fold_5	0.775	6.558	0.755	0.846
Average	0.764	7.467	0.753	0.839

Note: bold fronts indicate best results.

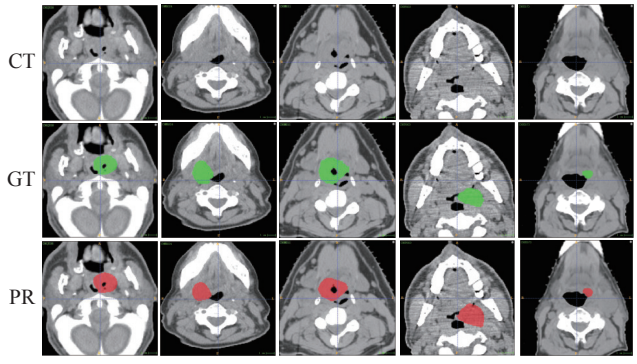

Figure 6 Example of the segmentation results of our SE-nnU-Net on validation subsets of HECKTOR 2021 challenge.

order to improve the segmentation accuracy, data augmentation technique TTA (test time augmentation) was used in the prediction process. The segmentation results of our SE-nnU-Net on HECKTOR 2021 test data are compared to the segmentation results obtained by other Top 9 teams participating in the HECKTOR 2021 challenge. The comparison of the results of Top 10 teams are shown Table 6.

It should be noted that the ranking was computed from the average Dice and median HD95 across all cases. The two metrics are ranked separately and the final rank is obtained by Borda counting. This ranking method was used first to determine the best submission of each participating team (ranking the 1 to 5 submissions), then to obtain the final ranking (across all participants). Each participating team had the opportunity to submit up to five (valid) runs [30].

Table 6 Comparison of segmentation results between the SE-nnU-Net and other Top 9 teams on HECKTOR 2021 test data

Teams/models	Dice	median HD95
An *et al.* [31]	0.773	**3.088**
Lu *et al.* [32]	0.774	**3.088**
Yousefirizi *et al.* [33]	0.771	**3.088**
Ren *et al.* [34]	**0.779**	3.155
Naser *et al.* [35]	0.770	3.143
De Biase *et al.* [36]	0.762	3.143
Wang *et al.* [37]	0.768	3.155
Cho *et al.* [38]	0.766	3.155
Meng *et al.* [39]	0.745	3.155
Our SE-nnU-Net [40]	**0.779**	**3.088**

Note: bold fronts indicate best results.

As can be seen from Table 5, the best results of the 5-fold cross-validation experiments of our SE-nnU-Net are 0.803, 5.933, 0.783 and 0.869, respectively, in terms of Dice, median HD95, Precision and Recall. The average results of the 5-fold cross-validation experiments of our SE-nnU-Net on HECKTOR 2021 training data are 0.764, 7.467, 0.753 and 0.839, respectively, in terms of Dice, median HD95, Precision and Recall.

It can be seen from Figure 6 that our SE-nnU-Net can segment the tumor in the images from the validation subset of HECKTOR 2021 challenge training data, but there are still over-segmentation or under-segmentation phenomena. However, the segmentation results of our SE-nnU-Net on the validation subsets of HECKTOR 2021 challenge are good, which can also be seen from the results in Table 5.

The results in Table 6 show that our SE-nnU-Net has obtained the best results of 0.779 (0.7785) and 3.088 (3.0882) in terms of Dice and median HD95, respectively. Our SE-nnU-Net ranked the first for HECKTOR 2021 challenge. The segmentation method proposed by An *et al.* is ranked the second due to the median HD95 slightly better than the third by Lu *et al.*, 3.088160269617 vs. 3.088161777508 [30], though the Dice of An *et al.* is slightly inferior to that of Lu *et al.*, 0.7733 vs. 0.7735 [30].

The results in Table 6 also show that there is large gap between the highest Dice of 0.779 and the lowest of 0.745. There is similar phenomenon in the median HD 95 with a large gap, the best is 3.088 (3.0882 [30]) and the worst is 3.155 (3.1549 [30]).

Moreover, the segmentation results of SE-nnU-Net on HECKTOR 2020 challenge and HECKTOR 2021 challenge demonstrate that the data partitioning method is very significant for training the model having good segmentation performance. The training data will affect the performance of the SE-nnU-Net model significantly. That is why the SE-nnU-Net performs best in HECKTOR 2021 challenge while ranking third in HECKTOR 2020 challenge. This further demonstrate the very common phenomenon in machine learning field that the performance of a model heavily depends on the training data.

IV. Conclusions

This paper integrated SE normalization into nnU-Net, such that the SE-nnU-Net was developed to segment the head and neck tumors in PET/CT images of cancer patients coming from multiple centers of different countries. The SE normalization combines instance normalization and SE attention module together, such that it can capture the features of interesting regions. The convolution-ReLU-SE module is designed, and combined with residual structure in encoder while without residual structure in decoder. The loss function combines the Dice loss and Focal loss together. The specific network architecture is designed, and the specific data preprocessing and augmentation techniques are developed, and deep supervision mechanism is introduced into Loss function using the last four layers of the decoder of the SE-nnU-Net.

This SE-nnU-Net was trained using the data coming from HECKTOR 2020 and HECKTOR 2021 challenges, respectively, under different experimental design. The SE-nnU-Net trained on HECKTOR 2021 training data by 5-fold cross validation experiments is superior to all other Top 9 teams and ranked first in HECKTOR 2021 challenge. Although the segmentation results on test data of the SE-nnU-Net trained using leave one center out method are better than that we submitted for HECKTOR 2020, the segmentation results is still ranked third compared to the segmentation results of the Top 10 teams participating in HECKTOR 2020 challenge.

The performance of deep learning-based segmentation models heavily depend on the training data, i.e. data partitioning methods for training the models. Furthermore, the segmentation results are simultaneously determined by the network architecture, loss function, data preprocessing and data augmentation techniques, basic operating module, and other hyper-parameters.

It is very challenging to obtain the excellent model for segmenting the head and neck tumors of PET/CT

images of cancer patients. The network structure, experimental design, data preprocessing, data augmentation, and specific detail of other hyper-parameters must be considered simultaneously. How to use task-specific empirical optimization, domain knowledge and data inherit properties to design the specific methods for obtaining models having the potential excellent performance may need further studying, especially in the medical image segmentation field.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant Nos. 62076159, 120 31010, and 61673251) and the Fundamental Research Funds for the Central Universities (Grant Nos. GK2021 05003 and GK202207017). We acknowledge those who organized the HECKTOR 2020 & HECKTOR 2021 challenges and published the training data and the related information for us to use in this study.

References

[1] F. Bray, J. Ferlay, I. Soerjomataram, *et al.*, "Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA:A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[2] S. L. Breen, J. Publicover, S. De Silva, *et al.*, "Intraobserver and interobserver variability in GTV delineation on FDG-PET-CT images of head and neck cancers," *International Journal of Radiation Oncology, Biology, Physics*, vol. 68, no. 3, pp. 763–770, 2007.

[3] D. Kawahara, M. Tsuneda, S. Ozawa, *et al.*, "Stepwise deep neural network (stepwise-net) for head and neck auto-segmentation on CT images," *Computers in Biology and Medicine*, vol. 143, article no. 105295, 2022.

[4] B. X. Wu, P. L. Khong, and T. Chan, "Automatic detection and classification of nasopharyngeal carcinoma on PET/CT with support vector machine," *International Journal of Computer Assisted Radiology and Surgery*, vol. 7, no. 4, pp. 635–646, 2012.

[5] Q. Song, J. J. Bai, D. F. Han, *et al.*, "Optimal co-segmentation of tumor in PET-CT images with context information," *IEEE Transactions on Medical Imaging*, vol. 32, no. 9, pp. 1685–1697, 2013.

[6] Y. M. Moe, A. R. Groendahl, M. Mulstad, *et al.*, Deep learning for automatic tumour segmentation in PET/CT images of patients with head and neck cancers, arXiv preprint arXiv: 1908.00841, 2019, doi: 10.48550/arXiv.1908.00841.

[7] V. Andrearczyk, V. Oreiller, M. Vallières, *et al.*, "Automatic segmentation of head and neck tumors and nodal metastases in PET-CT scans," in *Proceedings of the International Conference on Medical Imaging with Deep Learning*, Montréal, Canada, pp.33–43, 2020.

[8] Y. D. Yuan, "Automatic head and neck tumor segmentation in PET/CT with scale attention network," in *Proceedings of First 3D Head and Neck Tumor Segmentation in PET/CT Challenge*, Lima, Peru, pp.44–52, 2021.

[9] J. Ma and X. P. Yang, "Combining CNN and hybrid active contours for head and neck tumor segmentation in CT and pet images," in *Proceedings of First 3D Head and Neck Tumor Segmentation in PET/CT Challenge*, Lima, Peru, pp.59–64, 2021.

[10] H. Chen, H. B. Chen, and L. S. Wang, "Iteratively refine the segmentation of head and neck tumor in FDG-PET and CT images," in *Proceedings of First 3D Head and Neck Tumor Segmentation in PET/CT Challenge*, Lima, Peru, pp.53–58, 2021.

[11] F. Isensee, P. F. Jaeger, S. A. A. Kohl, *et al.*, "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, 2021.

[12] T. Falk, D. Mai, R. Bensch, *et al.*, "U-net: Deep learning for cell counting, detection, and morphometry," *Nature Methods*, vol. 16, no. 1, pp. 67–70, 2019.

[13] H. El-Hariri, L. A. S. M. Neto, P. Cimflova, *et al.*, "Evaluating nnU-Net for early ischemic change segmentation on non-contrast computed tomography in patients with acute ischemic stroke," *Computers in Biology and Medicine*, vol. 141, article no. 105033, 2022.

[14] G. B. Zhang, Z. Y. Yang, B. Huo, *et al.*, "Multiorgan segmentation from partially labeled datasets with conditional nnU-Net," *Computers in Biology and Medicine*, vol. 136, article no. 104658, 2021.

[15] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp.7132–7141, 2018.

[16] L. Rundo, C. Han, Y. Nagano, *et al.*, "Use-net: Incorporating squeeze-and-excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets," *Neurocomputing*, vol. 365, pp. 31–43, 2019.

[17] B. H. Zhang, S. L. Qi, Y. N. Wu, *et al.*, "Multi-scale segmentation squeeze-and-excitation UNet with conditional random field for segmenting lung tumor from CT images," *Computer Methods and Programs in Biomedicine*, vol. 222, article no. 106946, 2022.

[18] A. Iantsen, V. Jaouen, D. Visvikis, *et al.*, "Squeeze-and-excitation normalization for brain tumor segmentation," in *Proceedings of 6th International MICCAI Brainlesion Workshop*, Lima, Peru, pp.366–373, 2021.

[19] V. Andrearczyk, V. Oreiller, M. Jreige, *et al.*, "Overview of the HECKTOR challenge at MICCAI 2020: Automatic head and neck tumor segmentation in PET/CT," in *Proceedings of First 3D Head and Neck Tumor Segmentation in PET/CT Challenge*, Lima, Peru, pp.1–21, 2021.

[20] A. Iantsen, D. Visvikis, and M. Hatt, "Squeeze-and-excitation normalization for automated delineation of head and neck primary tumors in combined PET and CT images," in *Proceedings of First 3D Head and Neck Tumor Segmentation in PET/CT Challenge*, Lima, Peru, pp.37–43, 2021.

[21] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[22] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850–863, 1993.

[23] L. C. Chen, G. Papandreou, I. Kokkinos, *et al.*, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2018.

[24] J. Y. Xie and Y. Peng, "The head and neck tumor segmentation using nnu-net with spatial and channel 'squeeze & excitation' blocks," in *Proceedings of First 3D Head and Neck Tumor Segmentation in PET/CT Challenge*, Lima, Peru, pp.28–36, 2021.

[25] K. Ghimire, Q. Chen, and X. Feng, "Patch-based 3D UNet for head and neck tumor segmentation with an ensemble of conventional and dilated convolutions," in *Proceedings of First 3D Head and Neck Tumor Segmentation in PET/CT Challenge*, Lima, Peru, pp.78–84, 2021.

[26] F. Yousefirizi and A. Rahmim, "GAN-based bi-modal segmentation using mumford-shah loss: Application to head and

neck tumors in PET-CT images," in *Proceedings of First 3D Head and Neck Tumor Segmentation in PET/CT Challenge*, Lima, Peru, pp.99–108, 2021.

[27] S. M. Zhu, Z. Z. Dai, and N. Wen, "Two-stage approach for segmenting gross tumor volume in head and neck cancer with CT and PET imaging," in *Proceedings of First 3D Head and Neck Tumor Segmentation in PET/CT Challenge*, Lima, Peru, pp.22–27, 2021.

[28] M. A. Naser, L. V. Van Dijk, R. J. He, *et al.*, "Tumor segmentation in patients with head and neck cancers using deep learning based-on multi-modality PET/CT images," in *Proceedings of First 3D Head and Neck Tumor Segmentation in PET/CT Challenge*, Lima, Peru, pp.85–98, 2021.

[29] C. Rao, S. Pai, I. Hadzic, *et al.*, "Oropharyngeal tumour segmentation using ensemble 3D PET-CT fusion networks for the HECKTOR challenge," in *Proceedings of First 3D Head and Neck Tumor Segmentation in PET/CT Challenge*, Lima, Peru, pp.65–77, 2021.

[30] V. Andrearczyk, V. Oreiller, S. Boughdad, *et al.*, "Overview of the HECKTOR challenge at MICCAI 2021: Automatic head and neck tumor segmentation and outcome prediction in PET/CT images," in *Proceedings of Second 3D Head and Neck Tumor Segmentation in PET/CT Challenge*, Strasbourg, France, pp.1–37, 2022.

[31] C. Y. An, H. Chen, and L. S. Wang, "A coarse-to-fine framework for head and neck tumor segmentation in CT and PET images," in *Proceedings of Second 3D Head and Neck Tumor Segmentation in PET/CT Challenge*, Strasbourg, France, pp.50–57, 2022.

[32] J. S. Lu, W. H. Lei, R. Gu, *et al.*, "Priori and posteriori attention for generalizing head and neck tumors segmentation," in *Proceedings of Second 3D Head and Neck Tumor Segmentation in PET/CT Challenge*, Strasbourg, France, pp.134–140, 2022.

[33] F. Yousefirizi, I. Janzen, N. Dubljevic, *et al.*, "Segmentation and risk score prediction of head and neck cancers in PET/CT volumes with 3D U-Net and cox proportional hazard neural networks," in *Proceedings of Second 3D Head and Neck Tumor Segmentation in PET/CT Challenge*, Strasbourg, France, pp.236–247, 2022.

[34] J. T. Ren, B. N. Huynh, A. R. Groendahl, *et al.*, "Pet normalizations to improve deep learning auto-segmentation of head and neck tumors in 3D PET/CT," in *Proceedings of Second 3D Head and Neck Tumor Segmentation in PET/CT Challenge*, Strasbourg, France, pp.83–91, 2022.

[35] M. A. Naser, K. A. Wahid, L. V. Van Dijk, *et al.*, "Head and neck cancer primary tumor auto segmentation using model ensembling of deep learning in PET/CT images," in *Proceedings of Second 3D Head and Neck Tumor Segmentation in PET/CT Challenge*, Strasbourg, France, pp.121–133, 2022.

[36] A. De Biase, W. Tang, N. Sourlos, *et al.*, "Skip-SCSE multi-scale attention and co-learning method for oropharyngeal tumor segmentation on multi-modal PET-CT images," in *Proceedings of Second 3D Head and Neck Tumor Segmentation in PET/CT Challenge*, Strasbourg, France, pp.109–120, 2022.

[37] G. S. Wang, Z. Y. Huang, H. Shen, *et al.*, "The head and neck tumor segmentation in PET/CT based on multi-channel attention network," in *Proceedings of Second 3D Head and Neck Tumor Segmentation in PET/CT Challenge*, Strasbourg, France, pp.68–74, 2022.

[38] M. Cho, Y. Choi, D. Hwang, *et al.*, "Multimodal spatial attention network for automatic head and neck tumor segmentation in FDG-PET and CT images," in *Proceedings of Second 3D Head and Neck Tumor Segmentation in PET/CT Challenge*, Strasbourg, France, pp.75–82, 2022.

[39] M. Y. Meng, Y. G. Peng, L. Bi, *et al.*, "Multi-task deep learning for joint tumor segmentation and outcome prediction in head and neck cancer," in *Proceedings of Second 3D Head and Neck Tumor Segmentation in PET/CT Challenge*, Strasbourg, France, pp.160–167, 2022.

[40] J. Y. Xie and Y. Peng, "The head and neck tumor segmentation based on 3D U-Net," in *Proceedings of Second 3D Head and Neck Tumor Segmentation in PET/CT Challenge*, Strasbourg, France, pp.92–98, 2022.

Juanying XIE is a Professor and a Ph.D. student Supervisor of the School of Computer Science of Shaanxi Normal University, Xi'an, China. She received the Ph.D. and M.S. degree from Xidian University in 2012 and 2004, respectively. She received the B.S. degree from Shaanxi Normal University in 1993. Her research interests include machine learning, data mining, and biomedical data analysis. Her research is highly cited, with one article in the top 1% of ESI and one is the top 3 hotspot article of *SCIENTIA SINICA Informationis* and 3 articles was included in F5000. She is an Associate Editor of *Health Information Science and Systems*, and an editor board member of *the journal of Shaanxi Normal University (Natural Science Edition)*.
(Email: xiejuany@snnu.edu.cn)

Ying PENG received the M.S. degree in the application technology of computer science and B.S. degree in computer science from Shaanxi Normal University, Xi'an, China, in 2022 and 2019, respectively. Her research interests include deep learning and biomedical image segmentation.
(Email: py183248@snnu.edu.cn)

Mingzhao WANG is a Post Doctor supervised by Professor Juanying Xie at the School of Computer Science of Shaanxi Normal University, Xi'an, China. He received the Ph.D. degree in bioinformatics and M.S. degree in the application technology of computer science from Shaanxi Normal University in 2021 and 2017, respectively. He received the B.S. degree in computer science from Shanxi Normal University, Linfen, China, in 2014. His research interests include machine learning and bioinformatics.
(Email: wangmz2017@snnu.edu.cn)