Computers, Materials & Continua CMC, vol.65, no.2, pp.1771-1780, 2020

Tissue Segmentation in Nasopharyngeal CT Images Using Two-Stage Learning

Yong Luo[1], Xiaojie Li[2], Chao Luo[2], Feng Wang[1], Xi Wu[2], Imran Mumtaz[3] and Cheng Yi[1,*]

Abstract: Tissue segmentation is a fundamental and important task in nasopharyngeal images analysis. However, it is a challenging task to accurately and quickly segment various tissues in the nasopharynx region due to the small difference in gray value between tissues in the nasopharyngeal image and the complexity of the tissue structure. In this paper, we propose a novel tissue segmentation approach based on a two-stage learning framework and U-Net. In the proposed methodology, the network consists of two segmentation modules. The first module performs rough segmentation and the second module performs accurate segmentation. Considering the training time and the limitation of computing resources, the structure of the second module is simpler and the number of network layers is less. In addition, our segmentation module is based on U-Net and incorporates a skip structure, which can make full use of the original features of the data and avoid feature loss. We evaluated our proposed method on the nasopharyngeal dataset provided by West China Hospital of Sichuan University. The experimental results show that the proposed method is superior to many standard segmentation structures and the recently proposed nasopharyngeal tissue segmentation method, and can be easily generalized across different tissue types in various organs.

Keywords: Tissue segmentation, deep learning, two-stage network, convolutional neural network.

1 Introduction

Since nasopharyngeal image analysis can provide a wealth of information about tissue structure and morphology, they can be widely used in clinical practice, e.g., medical diagnosis, benign and malignant diagnosis and treatment effectiveness prediction [Chua, Wee, Hui et al. (2016); Lin, Dou, Jin et al. (2019); Li, Xu, Chen et al. (2018)]. However, the manual assessment of histopathological images by a nasopharyngologists is time-

[1] West China Hospital, Sichuan University, Chengdu, 610000, China.

[2] Chengdu University of Information Technology, Chengdu, 610000, China.

[3] University of Agriculture, Faisalabad, 38000, Pakistan.

[*]Corresponding Author: Cheng Yi. Email: yicheng6843@126.com.

Received: 10 February 2020; Accepted: 22 April 2020.

consuming and subjective [Men, Chen, Zhang et al. (2017); Ma, Wu and Zhou (2017)]. Digital nasopharyngeal image analysis is designed to automatically analyze nasopharyngeal images, which can significantly improve the repeatability and objectivity of the diagnosis [Wang, Zu, Hu et al. (2018); Mohammed, Ghani, Arunkumar et al. (2018)]. In particular, segmenting each tissue in a nasopharyngeal image is a fundamental and important task. Many techniques have been proposed, an extensive review is presented in Xing et al. [Xing and Yang (2016); Chen, Xiong, Xu et al. (2019)]. However, the task still faces some challenges, such as the difficulty of segmenting overlapping or contacted tissues, and the limited generalization of different organs and tissue types [Liang, Tang, Huang et al. (2019); Nejad and Shiri (2019)].

Recently, many advanced methods have been used for medical image segmentation. Inspired by the classic anti-generation network, Xue et al. [Xue, Xu, Zhang et al. (2018)] designed an end-to-end network (SegAN) for medical image segmentation tasks. Ma et al. [Ma, Zhou, Wu et al. (2019)] designed a multimodal convolutional neural network (M-CNN) for nasopharyngeal tumor segmentation to jointly learn multimodal similarity measures and segmentation of paired CT-MR images. Chen et al. [Chen, Xiong, Xu et al. (2019)] proposed a multimodal MRI fusion network (MMFNet) based on three MRI models (T1, T2 and contrast-enhanced T1) to complete the accurate segmentation of nasopharyngeal carcinoma (NPC). Daoud et al. [Daoud, Morooka, Kurazume et al. (2019)] proposed a second-order method for segmentation of CT images of nasopharyngeal carcinoma. Although these methods have lead to some performance improvements, they still have some disadvantages, such as the need for complex preprocessing and poor robustness.

U-Net is a structure based on a full convolutional neural network (FCN) [Ronneberger, Fischer and Brox (2015)]. The network has been widely used in image segmentation tasks due to its excellent segmentation performance. However, U-Net cannot be directly applied to the nasopharyngeal tissue segmentation task, and it has been clearly recognized that: 1) the receptive field and the segmentation accuracy cannot be simultaneously improved. When the receptive field selection is large, the segmentation accuracy is reduced. Similarly, in the case of a small receptive field, the accuracy of segmentation is not greatly improved. 2) There is a lot of noise in the nasopharyngeal CT image. The contrast between the target tissue and the surrounding tissue is very poor, which greatly reduces the accuracy of the segmentation. 3) The redundancy of the CT image segmentation task directly using U-Net is too large. Since each pixel has a gray value, the similarity of the gray values of two adjacent pixels is high, resulting in a large amount of redundancy, making the network training very slow [Xu, Luo, Zhang et al. (2018)].

Inspired by the advantages and disadvantages of U-Net mentioned above, we designed a two-stage training network based on U-Net [Yu, Wang, Shelhamer et al. (2018)]. The method solves the shortcomings of the above U-Net, and fully utilizes the advantages of U-Net, so that the nasopharyngeal tissue can be accurately segmented. Inspired by the advantages and disadvantages of U-Net mentioned above, we designed a network based on U-Net fusion two-stage method. The network structure is mainly composed of two modules: 1) Denoising module. Since the original image we input is large (512×512) and there is a large amount of noise area around each image, we design a roughly

segmented network consisting of 5 convolutional layers and deconvolutional layers based U-Net. The network can roughly segment the target area and reduce image noise. 2) Precise segmentation module. We design an accurate segmentation network consisting of 3 convolutional layers and deconvolutional layers. Considering the limited computing resources and training time, this module only constructs a 3-layer shallow network.

Consequently, this paper proposes a novel method for segmentation of nasopharyngeal tissue, which has two main contributions:

1. By constructing a two-stage segmentation network, the effect of noise on the segmentation results is effectively reduced, and the accuracy of segmentation is improved.

2. By constructing a noise reduction module, the influence of noise on segmentation accuracy is effectively reduced, and the robustness of the network is increased.

2 Methodology

In this section, we give more details of our proposed approach for nasopharyngeal tissue segmentation.

2.1 Learning framework

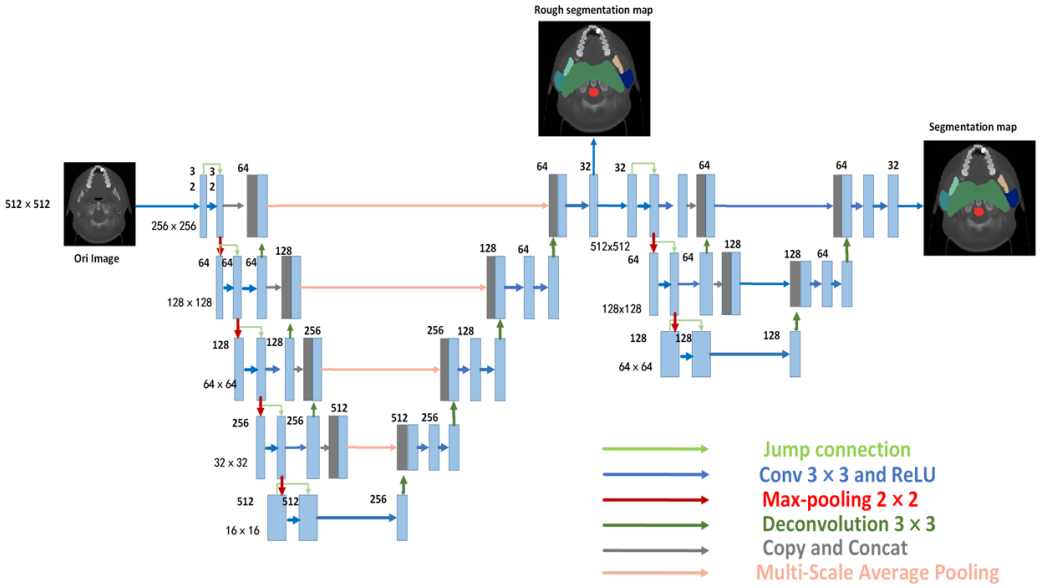

Figure 1: Our network

The architecture of our proposed network is illustrated in Fig. 1. In the first module, we constructed a noise reduction module consisting of five convolutional layer and deconvolutional layer. Each convolution unit is composed of a convolution unit, a batch normalization (BN) layer, and an activation function (ReLU) layer. For the second module, a shallow network is used to refine the nasopharyngeal contour segmentation map generated from the first stage for the final nasopharyngeal tissues segmentation map.

According to our experiments, we did not find performance differences between deep and shallow architectures. Therefore, considering the computational cost and efficiency, we chose a shallower architecture.

In addition, we also designed the MSP (Multi-scale Average Pooling) module, which consists of 4 average pooling layers of different kernel sizes. This module can fully learn the features of different sizes, effectively use the features of the image, and avoid feature loss.

2.1.1 Multi-scale block

The Multi-Scale block is a module composed of a fusion of average pooling layers with kernel sizes of 1×1, 3×3, 5×5, and 7×7. Kernel sizes of different sizes can not only increase the size of the receptive field, make full use of features of different sizes of the image, effectively avoid feature omissions, but also reduce the amount of network parameters, reduce the complexity of the network, and increase the training speed.

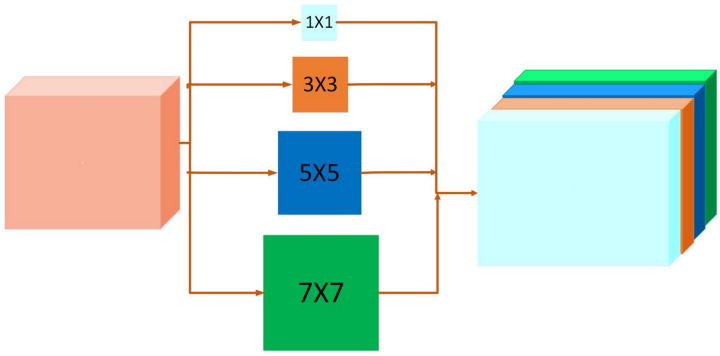

Figure 2: Multi-scale block

2.2 Loss function

Dice loss function and focal loss function are combined as the loss function of each step:

$$Dice = 1 - \frac{\sum_{n=1}^{N} p_n r_n + \epsilon}{\sum_{n=1}^{N} p_n + r_n + \epsilon} - \frac{\sum_{n=1}^{N} (1-p_n)(1-r_n) + \epsilon}{\sum_{n=1}^{N} (1-p_n) + (1-r_n) + \epsilon} \quad (1)$$

In Eq. (1), p_n is the predicted probability value and r_n is the true label value. ϵ is a constant in the range $(0, 1)$. The Dice loss function is a classic segmentation loss. Since the essence of segmentation is pixel-level classification, and the accumulated pixels are extremely imbalanced, the dice function can effectively avoid the problem of class imbalance, thereby improving the accuracy of segmentation.

$$Focal = \begin{cases} -\alpha(1-y')^{\gamma} log y' & , y = 1 \\ -(1-\alpha) y'^{\gamma} log(1-y'), & y = 0 \end{cases} \quad (2)$$

In Eq. (2), y' is the predicted probability value and y is the true label value. α is a constant in the range (0,1). In the segmentation task, there are often difficult samples that are difficult to segment and cause the performance of the network to decrease. The focal loss function can increase the weight of difficult samples and less the weight of easy samples, thereby making the network pay more attention to the samples and improve the

performance of the network.

$$Loss = \beta Dice + (1 - \beta)Focal \quad (3)$$

As shown in Eq. (3), in order to take full advantage of the dice loss function and the focal loss function, we fuse the two loss functions, where β is the weight parameter$\beta \in (0, 1)$. Through this loss function, the network segmentation performance is greatly improved.

3 Experiments and results

In this section, the data set, evaluation indicators, experimental details, and experimental results of this experiment are mainly introduced, and the experimental results are discussed.

3.1 Datasets

We evaluate our proposed tissue segmentation approach on West China Hospital of Sichuan University datasets, the datasets contain a total of 124 CT images of nasopharyngeal carcinoma with 512×512 resolution. It is a highly diverse data set since the shape of the segmentation targets contained in each image is inconsistent and the number of categories of segmentation targets is different. The data set contains a total of 14 segmentation target species (locref, CTV2, CTVnd, Mandible L, Mandible R, Parotid L, Parotid R, Brain Stem, Spinal Cord, Eyes L, Eyes R, Optic Nerves L, Optic Nerves R, Optic Chiasm), we converted the 3D CT image of each patient into a 2D CT image according to the slice, and removed the image without the label. A total of 8889 2D CT images were available. According to the five-fold cross-validation method, 7112 images were used for training and 1777 images were used for testing.

3.2 Evaluation metrics

We use of four indicators for the purpose of measure the performance of the network, which includes the dice similarity coefficient (DSC), area under the curve (AUC), Jaccard similarity coefficient (JSC), and F1-score for the assessment of the segmentation accuracy. The DSC was mostly employed for the calculation of the overlap metric between the results of segmentation and the ground truth. The DSC for bit vectors was defined as:

$$DSC = \frac{2||PG||_2}{||p||_2 + ||G||_2} \quad (4)$$

where $||PG||$ is the element-wise product of the prediction (P) and the ground truth (G), and $||x||_2$ is the L2-norm of x. The AUC is a probability value. The greater the AUC value, the better the performance. The AUC score was computed with a closed-form formula:

$$AUC = \frac{S_0 - n_0(n_0+1)/2}{n_0 n_1} \quad (5)$$

where n_0 is the number of pixel that belong to the ground truth, n_1 is the opposite and $S_0 = \sum_{i=1}^{n_0} r_i$, where r_i is the rank given by the predict model of the ground truth to the $i - th$

pixel in the CMR image. The F1 score is the harmonic average of precision and recall, wherein an F1 score reaches its best value at one (perfect precision and recall) and the worst at zero. These 3 metrics are defined as:

$$precision = \frac{TP}{TP+FP}, recall = \frac{TP}{FN+TP}, F1 = 2\frac{precision \cdot recall}{precision+recall} \quad (6)$$

where TP is the number of positive samples that are correctly predicted and FN is the number of negative samples with incorrect prediction results. FP is the number of positive samples with incorrect prediction results.

The JSC is put to use for the improvement of similarities and differences between finite sample sets. The larger the JSC value, the higher the sample similarity.

3.3 Implementation details

In order to ensure the stability and efficiency of the experiment, we conducted several experiments to explore the optimal settings of the parameters. Finally, the optimal parameter configuration scheme we adopted is as follows.

Learning rate strategy: We conducted several trials with different learning rates, and the results showed that the learning rate of 0.001 was the most appropriate. If the learning rate is too high, it will lead to over-fitting, and the network cannot learn the feature information correctly. If the learning rate is too low, the network fitting speed is very slow. Therefore we set the learning rate to 0.001, the initial learning rate is exponentially degraded every 10 iterations at a learning rate decay rate of 0.9.

Experiment configurations: To ensure the consistency of the experiment, we use accuracy as the quantization metric, 100 epochs are trained for each experiment. We use 5-fold cross-validation experimental method. All experiments are implemented in python 2.7 by using Tensorflow and Keras framework. We train the networks on a NVIDIA Tesla M40 GPU and the model that performs the best on test data set are saved for further analysis.

3.4 Results and discussions

To verify the advancement of our approach, we compare our approach to the classic standard segmentation architectures such as FCN-8 [Long, Shelhamer and Darrell (2015)], U-Net, Mask R-CNN [Kumar, Verma, Sharma et al. (2017)] and the results are shown in Tab. 1. These experimental results indicate that our model with two improvements performs significantly better and achieved the highest overall DSC, AUC, F1 score and JSC compared with other segmentation methods on nasopharyngeal datasets. Even the performance of our model with just one improvement is better than the majority of the other methods. In addition, we can clearly see from our experimental results that our method is more robust because there are 14 different types of tissue in each image.

Table 1: Experimental results of 4 different networks. The table shows the average of the four indicators

Model	DSC	AUC	F1	JSC
U-Net	0.7528	0.7326	0.7975	0.7832
FCN-8	0.6808	0.6051	0.7361	0.7029
Mask R-CNN	0.7331	0.6771	0.7829	0.7483
Ours	**0.8251**	**0.8592**	**0.9564**	**0.8732**

It is quite evident to observe from Fig. 3 that the proposed network is capable of segmenting the nasopharyngeal tissue efficiently. Through the comparison of the segmentation results with U-Net, FCN-8 and Mask R-CNN we can figure it out that the segmentation result of our network is not only less noisy but also closer to the ground truth. It can be clearly seen from the Fig. 3 that we can correctly segment the target area, and U-Net, FCN-8 and Mask R-CNN not only have poor segmentation effect but also have large deviations in the position of the segmented region. In addition, although our results also have noise areas, overall our results are far superior to the other three methods.

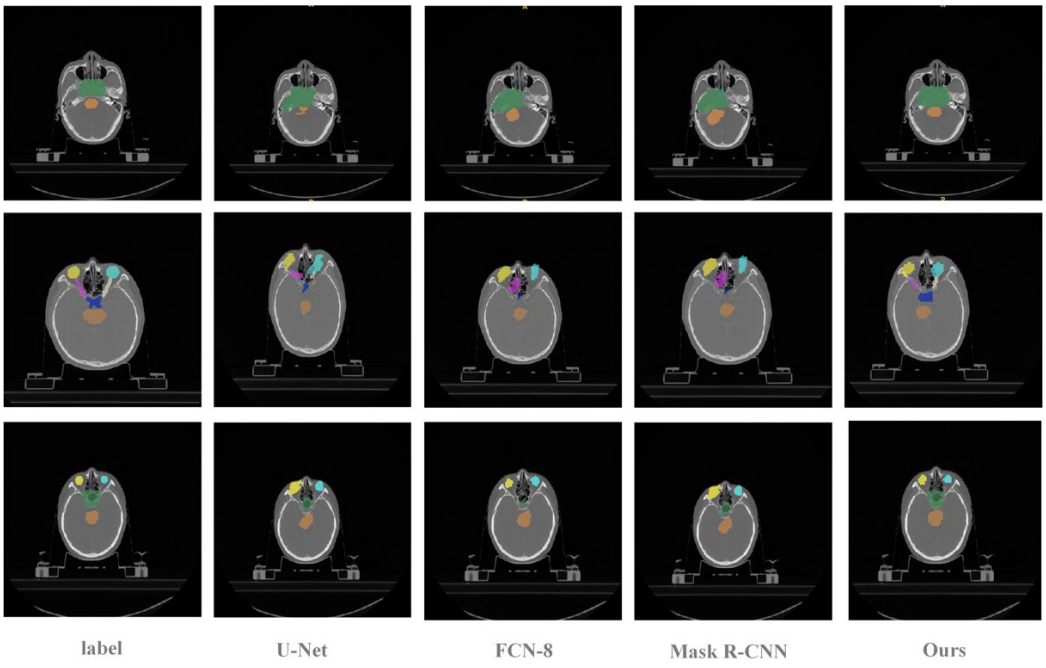

Figure 3: Segmentation results of three different samples in different networks

Fig. 4 is a graph of the results of each experiment. It is clear from the figure that the DSC value of each experiment of our method is the highest, and the range of change is the smallest. Therefore, the segmentation performance of our method is much stronger than the other three classical segmentation networks, and it is more robust.

In addition, Fig. 5 is a box plot of 5 experiments for each method. As can be seen from the figure, the DSC fluctuation of our method is the smallest, with better performance and robustness.

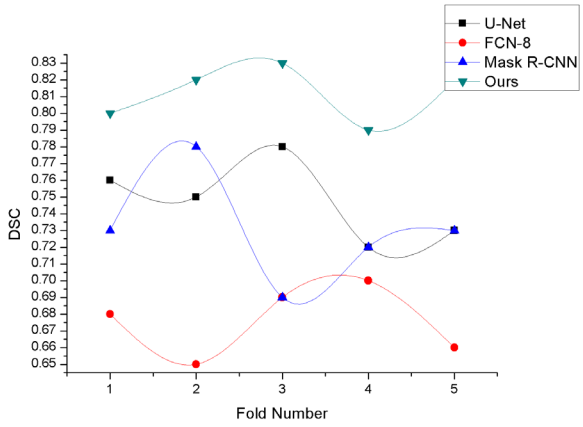

Figure 4: Curve of each experimental result

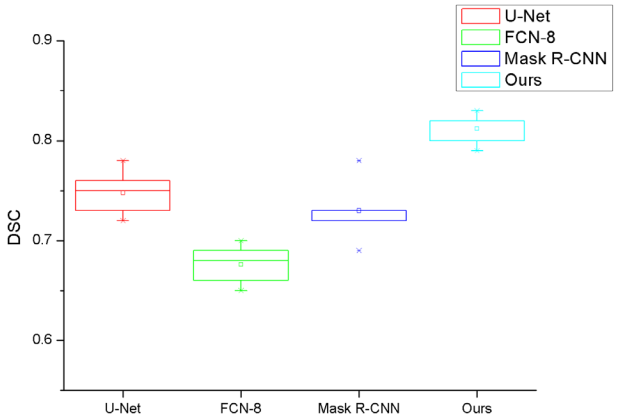

Figure 5: Box plots of experimental results for 4 methods

4 Conclusion

In this paper, we propose a two-stage network based on U-Net for segmentation of nasopharyngeal tissue. We constructed a denoising module that effectively filters the noise region of the image without manually eliminating noise in the data preprocessing section. Furthermore, we have designed a multi-scale average pooling module, which effectively improves the overall segmentation performance of the network. The experimental results show that compared with the standard segmentation method, the proposed method can segment different nasopharyngeal tissues better and has better robustness. However, the method proposed in this paper does not achieve very good results because the difference of the gray value of the CT image itself is small, the number of segmentation target area categories is large, and some segmentation regions are the target regions imagined by doctors. In the future work, we will continue to study the segmentation of nasopharyngeal tissue.

Funding statement: This work was supported by the National Natural Science Foundation of China (Grant No. 61602066) and the Scientific Research Foundation (KYTZ201608) of CUIT and the major Project of Education Department in Sichuan (17ZA0063 and 2017JQ0030), and partially supported by the Sichuan international science and technology cooperation and exchange research program (2016HH0018).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

Chen, Y. T.; Xiong, J.; Xu, W. H.; Zuo, J. W. (2019): A novel online incremental and decremental learning algorithm based on variable support vector machine. *Cluster Computing*, vol. 22, no. 3, pp. 7435-7445.

Chua, M. K.; Wee, J. T.; Hui, E. P.; Chan, A. C. (2016): Nasopharyngeal carcinoma. *The Lancet*, vol. 387, no. 10022, pp. 1012-1024.

Daoud, B.; Morooka, K.; Kurazume, R.; Leila, F.; Mnejja, W. et al. (2019): 3D segmentation of nasopharyngeal carcinoma from CT images using cascade deep learning. *Computerized Medical Imaging and Graphics*, vol. 77, pp. 101644.

Kumar, N.; Verma, R.; Sharma, S.; Bhargava, S.; Vahadane, A. et al. (2017): A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Transactions on Medical Imaging*, vol. 36, no. 7, pp. 1550-1560.

Li, Q.; Xu, Y.; Chen, Z.; Liu, D.; Feng, S. T. et al. (2018): Tumor segmentation in contrast-enhanced magnetic resonance imaging for nasopharyngeal carcinoma: deep learning with convolutional neural network. *BioMed Research International*, vol. 2018.

Liang, S.; Tang, F.; Huang, X.; Yang, K.; Zhong, T. et al. (2019): Deep-learning-based detection and segmentation of organs at risk in nasopharyngeal carcinoma computed tomographic images for radiotherapy planning. *European Radiology*, vol. 29, no. 4, pp. 1961-1967.

Lin, L.; Dou, Q.; Jin, Y. M.; Zhou, G. Q.; Tang, Y. Q. et al. (2019): Deep learning for automated contouring of primary tumor volumes by MRI for nasopharyngeal carcinoma. *Radiology*, vol. 291, no. 291, pp. 677-686.

Long, J.; Shelhamer, E.; Darrell, T. (2015): Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431-3440.

Ma, Z.; Wu, X.; Zhou, J. (2017): Automatic nasopharyngeal carcinoma segmentation in MR images with convolutional neural networks. *International Conference on the Frontiers and Advances in Data Science*, pp. 147-150.

Ma, Z.; Zhou, S.; Wu, X.; Zhang, H.; Yan, W. et al. (2019): Nasopharyngeal carcinoma segmentation based on enhanced convolutional neural networks using multi-modal metric learning. *Physics in Medicine and Biology*, vol. 64, no. 2, pp. 5-25.

Men, K.; Chen, X.; Zhang, Y.; Zhang, T.; Dai, J. et al. (2017): Deep deconvolutional neural network for target segmentation of nasopharyngeal cancer in planning computed tomography images. *Frontiers in Oncology*, vol. 7, pp. 315.

Mohammed, M. A.; Ghani, M. K. A.; Arunkumar, N.; Hamed, R. I.; Abdullah, M. K. et al. (2018): A real time computer aided object detection of nasopharyngeal carcinoma using genetic algorithm and artificial neural network based on Haar feature fear. *Future Generation Computer Systems*, vol. 89, pp. 539-547.

Nejad, M. B.; Shiri, M. E. (2019): A new enhanced learning approach to automatic image classification based on SALP Swarm Algorithm. *Computer System Science and Engineering*, vol. 34, no. 2, pp. 91-100.

Ronneberger, O.; Fischer, P.; Brox, T. (2015): U-net: convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-assisted Intervention*, pp. 234-241.

Wang, Y.; Zu, C.; Hu, G.; Luo, Y.; Ma, Z. et al. (2018): Automatic tumor segmentation with deep convolutional neural networks for radiotherapy applications. *Neural Processing Letters*, vol. 48, no. 3, pp. 1323-1334.

Wu, X.; Luo, C.; Zhang, Q.; Zhou, J.; Yang, H. et al. (2018): Text detection and recognition for natural scene images using deep convolutional neural networks. *Computers, Materials & Continua*, vol. 61, no. 1, pp. 289-300.

Xing, F.; Yang, L. (2016): A real time computer aided object detection of nasopharyngeal carcinoma using genetic algorithm and artificial neural network based on Haar feature fear. *Future Generation Computer Systems*, vol. 89, pp. 539-547.

Xue, Y.; Xu, T.; Zhang, H.; Long, L. R.; Huang, X. (2018): SEGAN: adversarial network with multi-scale l1 loss for medical image segmentation. *Neuroinformatics*, vol. 16, no. 4, pp. 383-392.

Yu, F.; Wang, D.; Shelhamer, E.; Darrell, T. (2018): Deep layer aggregation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2403-2412.