Application of Multi-Scale Fusion Attention U-Net to Segment the Thyroid Gland on CT Localization Images for Radiotherapy

Xiaobo Wen  
the Third Affiliated Hospital of Kunming Medical University

Biao Zhao  
the Third Affiliated Hospital of Kunming Medical University

Meifang Yuan  
the Third Affiliated Hospital of Kunming Medical University

Jinzhi Li  
the Third Affiliated Hospital of Kunming Medical University

Mengzhen Sun  
the Third Affiliated Hospital of Kunming Medical University

Lishuang Ma  
the Third Affiliated Hospital of Kunming Medical University

Chaoxi Sun  
the Third Affiliated Hospital of Kunming Medical University

Yi Yang (yiyangrt@126.com)  
the Third Affiliated Hospital of Kunming Medical University

Research Article

Keywords: U-net model, Multi-scale Fusions, medical image segmentation, thyroid, radiotherapy, cSE

Posted Date: October 5th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-949323/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Application of Multi-scale Fusion Attention U-net to segment the thyroid gland on CT localization images for radiotherapy

Xiaobo Wen¹, Biao Zhao¹, Meifang Yuan¹, Jinzhi Li¹, Mengzhen Sun¹, Lishuang MA¹, Chaoxi Sun², Yi Yang¹

1. Department of Radiotherapy, Yunnan Cancer Hospital, the Third Affiliated Hospital of Kunming Medical University, Kunming, Yunnan, China.
2. Department of Neurosurgery, Yunnan Cancer Hospital, the Third Affiliated Hospital of Kunming Medical University, Kunming, Yunnan, China.

Corresponding Authors:
Yi Yang, Department of Radiotherapy, Yunnan Cancer Hospital, the Third Affiliated Hospital of Kunming Medical University, 519 Kunzhou Road, Xishan District, Kunming City 650118, Yunnan Province, China.
E-mail: yiyangrt@126.com
Application of Multi-scale Fusion Attention U-net to segment the thyroid gland on CT localization images for radiotherapy

Abstract

Objectives To explore the performance of Multi-scale Fusion Attention U-net (MSFA-U-net) in thyroid gland segmentation on CT localization images for radiotherapy.

Methods CT localization images for radiotherapy of 80 patients with breast cancer or head and neck tumors were selected; label images were manually delineated by experienced radiologists. The data set was randomly divided into the training set (n=60), the validation set (n=10), and the test set (n=10). Data expansion was performed in the training set, and the performance of the MSFA-U-net model was evaluated using the evaluation indicators Dice similarity coefficient (DSC), Jaccard similarity coefficient (JSC), positive predictive value (PPV), sensitivity (SE), and Hausdorff distance (HD).

Results With the MSFA-U-net model, the DSC, JSC, PPV, SE, and HD indexes of the segmented thyroid gland in the test set were 0.8967±0.0935, 0.8219±0.1115, 0.9065±0.0940, 0.8979±0.1104, and 2.3922±0.5423, respectively. Compared with U-net, HR-net, and Attention U-net, MSFA-U-net showed that DSC increased by 0.052, 0.0376, and 0.0346 respectively; JSC increased by 0.0569, 0.0805, and 0.0433, respectively; SE increased by 0.0361, 0.1091, and 0.0831, respectively; and HD increased by −0.208, −0.1952, and −0.0548, respectively. The test set image results showed that the thyroid edges segmented by the MSFA-U-net model were closer to the standard thyroid delineated by the experts, in comparison with those segmented by the other three models. Moreover, the edges were smoother, over-anti-noise interference was stronger, and oversegmentation and undersegmentation were reduced. Conclusion The MSFA-U-net model can meet basic clinical requirements and improve the efficiency of physicians' clinical work.
Keywords: U-net model; Multi-scale Fusions; medical image segmentation; thyroid; radiotherapy; cSE

Introduction

Head and neck tumors and breast cancer are currently the tumors with the highest morbidity and mortality worldwide(1). In 2020, there were 19.29 million new cancer cases worldwide, of which 4.57 million cases (23.7%) were in China. Radiotherapy is an effective and common method for the treatment of head and neck cancer and breast cancer (2-4). Accurately delineating organs at risk (OARs) when designing radiotherapy plans can effectively avoid radiation side effects. At present, the outline of OARs is mainly done by physicians; thus, the process is subjective, time-consuming, and labor-intensive.

With the rapid development of artificial intelligence, Ronneberger (5) and others proposed the U-net neural network model in 2015. The segmentation and delineation method based on deep learning has gradually been developed and applied in clinical work(6-10). Ye et al.(7) used an improved model, a dense connectivity embedding U-net, to train and segment the T1 and T2 MRI images of 44 patients with nasopharyngeal carcinoma; the authors obtained DSC of 0.87 after 10-time cross-validation. However, in segmentation studies on OARs for head and neck tumors and breast cancer, the thyroid gland has often been ignored and has not been considered an OAR for automatic segmentation studies (8); therefore, the automatic delineation of the thyroid gland in CT localization images for radiotherapy has rarely been studied. It has been shown that side effects, such as thyroid function decline, occur when the radiation dose of the thyroid gland exceeds 26 Gy (11). Franco and others (12) conducted a retrospective study on 3-dimensional conformal radiation therapy (3D-CRT) for breast cancer. They
found that in about 45% of patients with lymph node-positive breast cancer, the thyroid gland was exposed to a radiation dose higher than 26 Gy. Other studies have shown (13, 14) that, at 5–10 years after receiving radiotherapy, the incidence of hypothyroidism in patients with nasopharyngeal cancer or breast cancer is 20% to 52%; moreover, the incidence of hypothyroidism increases with the increasing time of follow-up. Therefore, during radiotherapy planning, it is necessary to limit the radiation of the thyroid gland. Considering that CT localization for radiotherapy involves a simulated-positioning large-aperture CT (Somatom Sensation Open, 24 rows, Φ85 cm), which is limited by small size and poor image resolution, automatic segmentation of the thyroid gland based on deep learning model is difficult. It is necessary to further explore the performance of the deep learning model on CT localization images for radiotherapy. This study proposed a model that combines cSE attention mechanism and HR-net on the basis of U-net, and applied it to segment the thyroid gland on CT localization images so as to assist with the delineation of the thyroid gland as an OAR in radiotherapy.

1. Materials and methods

1.1 Data set acquisition

The experimental data set of this study was obtained from 80 patients with nasopharyngeal carcinoma or breast cancer admitted to the same attending group at the Department of Radiotherapy of Yunnan Cancer Hospital from June 2014 to April 2019. Siemens large-aperture CT (Sensation Open 24 CT) was used to perform the simulated localization for each patient. The slices were 5 mm or 3 mm thick, and the images of 512 × 512 were obtained in DICOM format. The mask images were drawn by senior radiotherapy physicians using the 3D slicer software. After drawing, the mask images in DICOM format were converted to PNG format; the result is shown in Figure 1.
Fig 1. CT localization image, label image, and 3D image (a. a standard image of the imported model (CT image), b. the corresponding label image (mask), c. the thyroid gland drawn in 3D)

The data set was divided (6:1:1) into the training set, the validation set, and the test set. Due to the small number of medical data sets and the high cost of drawing, it was difficult to collect a large number of data sets; however, if the training data are not large enough, there is a risk of overfitting the model. Therefore, in this study, data expansion was conducted by means of rotation, flipping, zooming,
and shearing, so as to increase the training sample size in the training set and avoid overfitting.

1.2 Data set preprocessing

To better highlight the region of interest (ROI), we first adopted HU value conversion to convert image pixels into HU values and then adjusted the window width and window level of the converted data to highlight the thyroid gland; finally, we used adaptive histogram equalization to further enhance the contrast (contrast) and normalize the images (normalization).

1.3 Model framework

The model in this study was improved on the basis of the U-net and HR-net model architectures. The main improvement in this study included the replacement of the two feature extraction convolutions of different resolutions in the U-net downsampling process with multiple convolution blocks in HR-net and feature fusion between different scales, and introduction of Spatial Squeeze and Channel Excitation Block (cSE) attention mechanism (15), called MSFA-U, into each convolution block, as shown in Figure 2. In the downsampling process of the model, we connected an cSE module after extracting two 3 × 3 convolutional features and fused the input features with the features after the scale operation by means of residual connection. The residual connection consisted of a 1 × 1 2d convolution and a normalization layer(16) (Batch Normalization, BN). In the cSE module, we used a global average pooling layer (GlobalAveragePooling, GAP) to convert a feature map from Channel × height × width to Channel × 1 × 1 and then used Dense to reduce the feature channel by half, which was achieved by activating the function Relu. After that, the feature channel was restored to normal size by using Dense, and the function Sigmoid was used to activate the channel. Finally, the calibrated
feature map was obtained through the channel-wise multiplication. The schematic diagram of the residual connection and cSE module structure is shown in Figure 3. The residual connection effectively prevents the model from disappearing and exploding with the deepening of the network (17). Moreover, the cSE module is able to effectively reflect the relationship between different channels and assign different weights so that the model can focus on important features that accurately segment the thyroid gland during the training process. The whole module is called an Attention Resblock (Figures 2 and 3).

The traditional U-net model uses the maximum pooling layer to perform downsampling and reduce the amount of parameters. This method may lead to the loss of information during the feature extraction process. Therefore, in this study, the stepped convolution was used to perform the downsampling. Stride convolution can remove redundant information, thereby reducing the size of the feature map. The model uses multiple branches of different resolutions to extract features in parallel during the training process, and it performs feature fusion between different scales after each attention residual block so as to achieve strong semantic information and precise location during the training process.

One or more transposed convolutions (3 × 3) are used in the conversion from low resolution to high resolution, while one or more stride convolutions (3 × 3) are used in the conversion from high resolution to low resolution, as shown in Fig. 3. In the upsampling part, the attention residual block replaces the two convolution operations in U-net to avoid excessive parameters. Meanwhile, dropout layer is added after each shortcut connection (the parameter set to 0.2) to avoid the decrease of generalization caused by the overfitting resulting from multiple feature fusions between different scales during the training process.
Fig. 2 MSFA-U-net structure

(a. Attention Resblock Module, the blue cuboid is the cSE module, and the red cuboid is the)

Fig. 3 Attention Resblock Module and Feature fusion of different scales

(b. High resolution to a low resolution
c. Low resolution to a High resolution

(a. Attention Resblock Module, the blue cuboid is the cSE module, and the red cuboid is the
Resblock module; b. one or more stride convolutions (3 × 3) are used in the conversion from high resolution to low resolution; c. one or more transposed convolutions (3 × 3) are used in the conversion from low to high resolution

1.4 Model operating environment and parameters

TensorFlow and Keras were used to build the model, and Python 3 was used to program the model. We used a Windows 10 64-bit operating system and the following hardware characteristics: CPU Intel (R) Core (TM) i9-10900KF CPU @ 3.70 GHz; the graphics card, NVIDIA GTX3090 24G; and 128 GB memory. The model hyperparameters were selected from the best results according to the experimental conditions, as shown in Table 1. Batch Size represents the number of input images for each iteration, Epoch represents the batch to be trained, Image Size represents the input size of the image, Learning Rate represents the initial learning rate using exponential decay, Decay steps indicate how many steps have been experienced for a learning rate decay, and Decay_Rate indicates the learning rate decay coefficient.

Table 1 Network training parameters

| Model          | Batch Size | Epoch | Image Size (Pixels) | Learning Rate | Decay Steps | Decay_Rate |
|----------------|------------|-------|---------------------|---------------|-------------|------------|
| U-net          | 2          | 120   | 512×512             | 1e-5          |             |            |
| HR-net         | 2          | 120   | 512×512             | 8e-5          | 300         | 0.96       |
| Attention U-net| 2          | 120   | 512×512             | 8e-4          | 300         | 0.96       |
| MSFA-U-net     | 2          | 120   | 512×512             | 2e-4          | 300         | 0.96       |

1.5 Loss function
Due to the small size, the thyroid gland occupies less space on a CT image. Therefore, the use of the traditional cross-entropy loss function would make the model more inclined to predict the background, resulting in the model not able to accurately identify the thyroid gland. Milletari et al. (18) proposed a loss function for sample imbalance in medical image segmentation in the research of V-net. Dice loss function, which is a loss function based on the DSC. It directly compares the overlap between the model prediction and the real segmentation, thereby effectively solving the problem of serious thyroid imbalance. The formula is shown in 1.1:

\[ DL = 1 - 2 \frac{|X \cap Y|}{|X| + |Y| + \varepsilon}, \]  

(1.1),

where X represents the label matrix of the real thyroid gland, Y is the prediction matrix of the model predicting the thyroid gland, and \( \varepsilon \) represents a constant included to avoid being divided by zero.

1.6 Evaluation indexes

The commonly used Dice similarity coefficient (DSC), Jaccard similarity coefficient (JSC), positive predictive value (PPV), sensitivity (SE), and Hausdorff distance (HD) were used to further evaluate the generalization ability and segmentation accuracy of the model.

DSC(19) and JSC(20) calculation is shown in formulas (1.2) and (1.3):

\[ DSC = 2 \frac{|X \cap Y|}{|X| + |Y|}, \]  

(1.2),

\[ JSC = \frac{|X \cap Y|}{|X \cup Y|}, \]  

(1.3),

where X represents the standard segmentation map drawn by a radiologist, Y is the prediction image segmented by the neural network model, and \( |X \cap Y| \) represents the overlap between the standard map drawn by the radiologist and the model predicted image. The value range of DSC and JSC is 0 to 1, and values closer to 1 indicate better prediction.
The calculation of PPV(21) and SE(22) is shown in formulas (1.4) and (1.5):

\[ PPV = \frac{TP}{TP + FP}, \quad (1.4), \]

\[ SE = \frac{TP}{TP + FN}, \quad (1.5), \]

where TP represents the foreground target value that is predicted correctly, and FP represents the foreground target value that is predicted incorrectly, and FN represents the background target value that is predicted incorrectly.

HD(23) calculation is shown in formula (1.6):

\[ H(X, Y) = \max_{x \in X} \min_{y \in Y} \| a - b \| \]

\[ h(Y, X) = \max_{y \in Y} \min_{x \in X} \| a - b \|. \]

Smaller values of HD indicate better prediction.

2. Experimental results

2.1 Comparison model design

In order to prove the validity of the proposed MSFA-U-net model, we selected three model architectures related to MSFA-U-net and conducted comparative experiments:

1) U-net(5): U-shaped symmetrical structure, composed of upsampling, downsampling, and skip connection. The skip connection effectively combines the feature information between different resolutions and makes up for the loss of the high-resolution features in the downsampling process, and it can output the feature map more accurately. The U-net model is one of the models widely used in the medical field.

2) HR-net(24): This model maintains high-resolution output during the feature extraction process.

It has multiple parallel subnets with different resolutions to compress and extract features, and perform
multiple different scale feature fusion to get richer high-resolution features. In the original study, the author used bilinear interpolation upsampling. In order to better extract features for fusion, in this study, we used transposed convolution to convert from low resolution to high resolution.

3) Attention U-net(25): This model introduces an attention-gating mechanism, so that the information in the jump connection of the U-net model has different weights and pays more attention to the ROI area.

2.2 Qualitative analysis of results

Figure 4 shows the results of the segmentation label map of the four models in the test set of the thyroid gland, and Figure 5 shows the coverage map of the four models on the CT image of the radiotherapy location. Given that there are a large number of blood vessels and soft tissues with similar gray levels around the thyroid gland(26), oversegmentation and undersegmentation are expected in the edge segmentation. As shown in the segmented label map in Figure 4, a part of the surrounding blood vessels and soft tissues was mistakenly segmented as the part of the left lobe of the thyroid gland when the U-net model segmented the left lobe of the thyroid gland. As for HR-net and Attention U-net, although oversegmentation of the surrounding soft tissues and blood vessels decreased, there were still some noise points and uneven edges. However, the MSFA-U-net architecture used in the study achieved smooth edges and decreased noise. Moreover, although we adjusted the window width and window level and adopted adaptive contrast enhancement, some lesions at some levels of the thyroid may have resulted in less obvious enhancement. In the thyroid segmentation at these levels, MSFA-U-net exhibited better robustness than the other three models. In summary, compared with the other three models, MSFA-U-net improved the performance of the thyroid gland segmentation on CT localization.
images for radiotherapy.

Fig. 4 Thyroid gland segmentation of the four models on CT localization images for radiotherapy

Fig. 5 The thyroid coverage map of the four models on CT localization images for radiotherapy

2.3 Quantitative analysis of the results

Table 2 shows the comparison of the results of the four models in the test set of thyroid gland segmentation indexes on CT localization images for radiotherapy. Compared with other models, MSFA-U-net showed better results in terms of the following evaluation indicators: DSC, 0.8967; JSC,
0.8219; PPV, 0.9065; SE, 0.8979; and HD, 2.39. Compared with the other three mainstream medical image segmentation models, MSFA-U-net achieved great improvement in multiple evaluation indicators: DSC improvement range, [0.0346, 0.052]; JSC improvement range [0.0433, 0.0805]; SE improvement range [0.0361, 0.1091]; and HD improvement range [−0.208, −0.0548]. As for the PPV evaluation index, MSFA-U-net was better than the U-net model and worse than the HR-net and Attention U-net models; however, as for the other evaluation indexes, HR-net and Attention U-net models were worse than the MSFA-U-net.

Table 2 Assessment indexes of the test set (𝑥̅ ± s)

|               | U-net     | HR-net    | Attention U-net | MSFA-U-net |
|---------------|-----------|-----------|-----------------|------------|
| DSC           | 0.8591±0.1046 | 0.8447±0.0938 | 0.8621±0.1502   | 0.8967±0.0935 |
| JSC           | 0.7650±0.1337 | 0.7414±0.1273 | 0.7786±0.1633   | 0.8219±0.1115 |
| PPV           | 0.8775±0.1179 | 0.9257±0.0759 | 0.9523±0.0689   | 0.9065±0.0940 |
| SE            | 0.8618±0.1212 | 0.7888±0.1276 | 0.8148±0.1683   | 0.8979±0.1104 |
| HD            | 2.6002±0.5731 | 2.5874±0.5412 | 2.4470±0.6872   | 2.3922±0.5423 |

* Bold numbers mean optimal value.

2.4 Box plot of the results

To further evaluate the differences between the four models, we made box plots of the evaluation indicators (Figures 6). The box plot results showed that MSFA-U-net had a smaller distance between the upper quartile and the lower quartile than the other three models; it also had fewer outliers and they were closer to the median. These findings indicate that MSFA-U-net provides better segmentation of the thyroid gland on CT localization images for radiotherapy than the other three models; moreover, it
ensures more consistent segmentation results, and it has better generalization and robustness.

Fig. 6 Box plot diagrams in the test set (a. Box plot diagrams of DSC in the test set. b. Box plot diagrams of JSC in the test set. c. Box plot diagrams of PPV in the test set. d. Box plot diagrams of SE in the test set. e. Box plot diagrams of HD in the test set)
2.5 Summary and analysis of the results

Taken together, the Attention U-net (which introduces gated attention) and the HR-net did not show obvious advantages in thyroid gland segmentation on CT localization images for radiotherapy compared with the U-net. Instead, they were even worse in some of the evaluation indicators. The reason may be that although the gated attention mechanism can effectively segment the target category and location, it may also lead to an increase in model false-positive predictions for small-volume segmentation with variability in shape. The HR-net performs multiple simple feature fusions; although it can effectively fuse features and obtain rich high-resolution features, it is also more likely to cause overfitting in the event of relatively scarce training data. Thus, the phenomenon of integration leads to a decline in its generalization ability. Clearly, the increase in model parameters and the increase in resource consumption may not necessarily improve the corresponding results.

3. Discussion

Radiotherapy is an important part of the comprehensive treatment of head and neck tumors and breast cancer. For the design and implementation of radiotherapy plans, accurate regulation of the radiation dose within the target area and limiting the dose to the surrounding OARs are important parts of the evaluation of a treatment plan. Precise delineation of OARs is important to effectively limit the dose outside the target area and avoid side effects of radiotherapy (27). The thyroid gland is an OAR during treatment of head and neck tumors and breast cancer, so it needs to be protected during radiotherapy.

This study proposes a multi-scale fusion attention U-net model to address the problem of thyroid
gland segmentation on CT localization images for radiotherapy. The innovation of this algorithm lies in the addition of multiple parallel channels on the basis of the traditional U-net model. It fully integrates the feature information between different resolutions, thereby avoiding the single information in the U-net downsampling process. In addition, our study also introduced the cSE attention mechanism, which inclined the model to the ROI during the training process. The experimental results showed that, compared with other similar representative segmentation algorithms, the proposed model improved both qualitative and quantitative results to a certain extent, and had better robustness and generalization. The image segmentation graphs revealed that MSFA-U-net effectively reduced oversegmentation and undersegmentation, and it achieved smoother edges. According to Zijdenbo et al. (28), DSC >0.7000 indicates that the segmentation meets the basic standard. While all of the models in this study reached this threshold, the DSC value of the MSFA-U-net reached 0.8967, indicating that this model can effectively segment the thyroid gland on CT localization images for radiotherapy. From the box plot diagrams, it is evident that the MSFA-U-net yields good quantitative results, where the upper and lower quartile gaps and outliers of most evaluation indicators were reduced, indicating that the model achieves a consistent segmentation on the different layers of the thyroid gland and can effectively segment the thyroid gland on CT localization images for radiotherapy. However, there are some limitations in the algorithm proposed in this study, which needs to be further improved. Some of the evaluation indicators of MSFA-U-net have not yet reached the optimal results. Moreover, the introduction of a large number of feature fusions between different scales resulted in an increase in the amount of model parameters. In addition, although the Dice loss function can effectively solve the problem of class imbalance, its gradient characteristics may cause the model to oscillate during the training process. In future research, we will explore how to reduce the parameter count of the model.
while using different loss functions to further optimize the model.

4. Conclusions

In summary, the MSFA-U-net model makes it possible for radiologists to automatically delineate the thyroid gland on CT localization images for radiotherapy, and the results show that the model can be well applied to clinical work. Compared with the three commonly used models in the medical field, the MSFA-U-net model can offer more help to radiologists by delineating the thyroid gland more accurately and helping save time in delineation.

List of abbreviations

CT: computed tomography; MSFA-U-net: Multi-scale Fusion Attention U-net; DSC: Dice similarity coefficient; JSC: Jaccard similarity coefficient; PPV: positive predictive value; SE: sensitivity; HD: Hausdorff distance; HR-net: High-Resolution net; OARs: organs at risk; MRI: magnetic resonance imaging; 3D-CRT: three-dimensional conformal radiotherapy; 3D: three-dimensional; ROI: region of interest; cSE: spatial squeeze and channel excitation block; BN: batch normalization; GAP: global average pooling.

Reference

1. Khazaei Z, Sohrabivafa M, Momenabadi V, Moayed L, Goodarzi E. Global Cancer Statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide prostate cancers and their relationship with the human development index. Advances in Human Biology. 2019;9.

2. Lin Z, Yang Z, He B, Wang D, Gao X, Tam SY, et al. Pattern of radiation-induced thyroid gland
changes in nasopharyngeal carcinoma patients in 48 months after radiotherapy. PLOS ONE. 2018;13:e0200310.

3. Buchholz T. Radiation after breast conservative surgery or after mastectomy reduces local recurrence and improves overall survival in breast cancer. Early Breast Cancer Trialists' Collaborative Group. Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomized trials. Lancet 2005;366:2087–2106. The Women's Oncology Review. 2011;6.

4. Forastiere A, Zhang Q, Weber R, Maor M, Goepfert H, Pajak T, et al. Long-Term Results of RTOG 91-11: A Comparison of Three Nonsurgical Treatment Strategies to Preserve the Larynx in Patients With Locally Advanced Larynx Cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2012;31.

5. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation 2015. 234-41 p.

6. Hesamian MH, Jia W, He X, Kennedy P. Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. Journal of Digital Imaging. 2019;32.

7. Ye Y, Cai Z, Huang B, He Y, Zeng P, Zou G, et al. Fully-Automated Segmentation of Nasopharyngeal Carcinoma on Dual-Sequence MRI Using Convolutional Neural Networks. Frontiers in Oncology. 2020;10:166.

8. Zhu W, Huang Y, Zeng L, Chen X, Liu Y, Qian Z, et al. AnatomyNet: Deep Learning for Fast and Fully Automated Whole-volume Segmentation of Head and Neck Anatomy. Medical Physics. 2018;46.

9. Alom MZ, Yakopcic C, Hasan M, Taha T, Asari V. Recurrent residual U-Net for medical image segmentation. Journal of Medical Imaging. 2019;6.
10. Zhang Z, Wu C, Coleman S, Kerr D. DENSE-INception U-net for medical image segmentation. Computer Methods and Programs in Biomedicine. 2020;192:105395.

11. Zhao N, Yang R, Jiang Y, Tian S, Guo F, Wang J. A hybrid IMRT/VMAT technique for the treatment of nasopharyngeal cancer. BioMed research international. 2015;2015:940102.

12. Franco A, Jolly LA, Russell S, Goldstein L, DeHart J. Abstract P5-06-01: The role of thyroid hormones in breast tumorigenesis: A translational study utilizing mouse models, cell culture and patient data. Cancer Research. 2016;76:P5-06.

13. Luo R, Wu V, He B, Gao X, Xu Z, Wang D, et al. Development of a normal tissue complication probability (NTCP) model for radiation-induced hypothyroidism in nasopharyngeal carcinoma patients. Bmc Cancer. 2018;18(1).

14. Rønjom M, Brink C, Bentzen S, Hegedüs L, Overgaard J, Petersen J, et al. External validation of a normal tissue complication probability (NTCP) model for radiation-induced hypothyroidism in an independent cohort. Acta oncologica (Stockholm, Sweden). 2015;54:1-9.

15. Hu J, Shen L, Sun G, Albanie S. Squeeze-and-Excitation Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017;PP.

16. Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. 2015.

17. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2015;7.

18. Milletari F, Navab N, Ahmadi S-A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. 2016.

19. Hermsen M, Bel T, Boer M, Steenbergen E, Kers J, Florquin S, et al. Deep Learning–Based Histopathologic Assessment of Kidney Tissue. Journal of the American Society of Nephrology.
20. Eelbode T, Bertels J, Berman M, Vandermeulen D, Maes F, Bisschops R, et al. Optimization for Medical Image Segmentation: Theory and Practice When Evaluating With Dice Score or Jaccard Index. IEEE Transactions on Medical Imaging. 2020;PP:1-.

21. Talluri S. Positive predictive value. BMJ. 2009;339:b3835-b.

22. Skaik Y. Understanding and using sensitivity, specificity and predictive values. Indian journal of ophthalmology. 2008;56:341; author reply

23. Kreveld M, Miltzow T, Ophelders T, Sonke W, Vermeulen J. Between shapes, using the Hausdorff distance. Computational Geometry. 2022;100:101817.

24. Ke S, Xiao B, Liu D, Wang J. Deep High-Resolution Representation Learning for Human Pose Estimation 2019.

25. Oktay O, Schlemper J, Folgoc L, Lee M, Heinrich M, Misawa K, et al. Attention U-Net: Learning Where to Look for the Pancreas. 2018.

26. Tajran J, Gosman AA. Anatomy, Head and Neck, Scalp. StatPearls. Treasure Island (FL): StatPearls Publishing

27. Peng Y-L, Chen I, Shen G-Z, Li Y-N, Yao J-J, Xiao W-W, et al. Interobserver variations in the delineation of target volumes and organs at risk and their impact on dose distribution in intensity-modulated radiation therapy for nasopharyngeal carcinoma. Oral oncology. 2018;82:1-7.

28. Zijdenbos A, Dawant B, Margolin R, Palmer A. Morphometric analysis of white matter lesions in MR images. IEEE transactions on medical imaging. 1994;13:716-24.

Declarations
Ethics approval and consent to participate

Institutional Review Board approval was not required because in order to protect the privacy of the patients, all the CT images in this study, when they were copied from the image library, were modified to delete the private information of the patients such as name, age, diagnosis, etc. This is a scientific study and isn't aimed at diagnosis and treatment, causing no harm to patients.

Consent for publication

Not applicable

Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available due to patient confidentiality, but are available from the corresponding author on reasonable request.

Competing interests

The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Funding

This study has received funding by Climbing Foundation of National Cancer Center (NCC201925B03).

Authors' contributions

All authors were involved in the conception of this study, design, and implementation. XBW was a major contributor in designing the models, drawing the figures and writing the manuscript and was also involved in delineating some label images. YY performed the checking and proofing of the manuscript and the data apart from guidance to the writing of the manuscript. BZ, MFY, JZL, MZS, LSM, and CXS were major contributors in data collection, delineation of most of the label images and output of data. All authors read and approved the final manuscript.
Acknowledgements

We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.