



## OPEN ACCESS

## EDITED BY

Alessandro Bria,  
University of Cassino, Italy

## REVIEWED BY

Xi Wang,  
The Chinese University of Hong Kong, China  
Maheshwari Prasad Singh,  
National Institute of Technology Patna, India

## \*CORRESPONDENCE

Zhicheng Zhang

✉ zhangzhicheng13@mails.ucas.edu.cn

Lu Zhou

✉ htcmu@126.com

†These authors have contributed equally to this work

RECEIVED 26 April 2024

ACCEPTED 26 July 2024

PUBLISHED 23 August 2024

## CITATION

Zhou L, Wu C, Chen Y and Zhang Z (2024)  
Multitask connected U-Net: automatic lung cancer segmentation from CT images using PET knowledge guidance.

*Front. Artif. Intell.* 7:1423535.

doi: 10.3389/frai.2024.1423535

## COPYRIGHT

© 2024 Zhou, Wu, Chen and Zhang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Multitask connected U-Net: automatic lung cancer segmentation from CT images using PET knowledge guidance

Lu Zhou<sup>1\*†</sup>, Chaoyong Wu<sup>2†</sup>, Yiheng Chen<sup>1</sup> and Zhicheng Zhang<sup>3\*</sup>

<sup>1</sup>Traditional Chinese Medicine (Zhong Jing) School, Henan University of Chinese Medicine, Zhengzhou, Henan, China, <sup>2</sup>Shenzhen Hospital, Beijing University of Chinese Medicine, Shenzhen, Guangdong, China, <sup>3</sup>Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

Lung cancer is a predominant cause of cancer-related mortality worldwide, necessitating precise tumor segmentation of medical images for accurate diagnosis and treatment. However, the intrinsic complexity and variability of tumor morphology pose substantial challenges to segmentation tasks. To address this issue, we propose a multitask connected U-Net model with a teacher-student framework to enhance the effectiveness of lung tumor segmentation. The proposed model and framework integrate PET knowledge into the segmentation process, leveraging complementary information from both CT and PET modalities to improve segmentation performance. Additionally, we implemented a tumor area detection method to enhance tumor segmentation performance. In extensive experiments on four datasets, the average Dice coefficient of 0.56, obtained using our model, surpassed those of existing methods such as Segformer (0.51), Transformer (0.50), and UctransNet (0.43). These findings validate the efficacy of the proposed method in lung tumor segmentation tasks.

## KEYWORDS

lung cancer, CT image, PET/CT, medical image segmentation, deep learning

## 1 Introduction

Despite significant advancements in diagnosing and treating lung cancer in recent decades, it remains the leading cause of cancer-related mortality globally, particularly among males (Sung et al., 2021). The introduction of low-dose computed tomography (CT) based lung cancer screening has notably reduced the mortality in clinical settings (Leiter et al., 2023). This is mainly due to the advantages of CT's high spatial resolution and acceptable economic burden, making it widely used in the clinical setting. However, the limited contrast between malignant and non-malignant lesions in lung tissue, makes timely detection and accurate segmentation of cancer boundaries on CT images a challenge. This difficulty is further compounded by variations in tumor locations, intensities, shapes, and attachments to adjacent structures, necessitating accurate lung tumor segmentation (Mercieca et al., 2021). The use of positron emission tomography/computed tomography (PET/CT) imaging has become essential in the diagnosis and staging of cancer, as it provides both functional information from PET images and anatomical localization from CT images (Baek et al., 2019). PET/CT has become integral in tumor management, aiding in diagnosis, staging, and follow-up (Bianconi et al., 2020). In particular, PET/CT is crucial for early differentiation of benign and malignant tumors as well as

disease severity assessment and progression. While CT images provide excellent spatial resolution, PET adds metabolic insights, allowing for superior tumor characterization. This further optimizes the screening and evaluation strategy for lung cancer; however, the high cost of PET and greater radioactive harm, limits its use compared to CT.

In the segmentation strategy for lung cancer, traditional manual contouring of boundaries on CT images is a time-consuming task prone to interobserver variability and misinterpretation. With the development and application of artificial intelligence (AI) technology, realizing the automatic segmentation of lung cancer based on CT images has become possible under the guidance of PET. Automatic tumor segmentation based on PET/CT images performs better than CT images alone (Ju et al., 2015). These automated segmentation methods largely focus on fusing the information extracted separately from the PET and CT modalities, under the assumption that each modality contains complementary information (Li et al., 2019; Bourigault et al., 2021; Cai et al., 2023). While acknowledging the advancements in PET/CT for lung cancer segmentation, addressing the associated drawbacks, such as the higher cost and increased radiation exposure to patients compared to CT alone has become crucial (Sheikhbahaei et al., 2017).

In this context, we explore the current status and challenges in lung cancer segmentation and highlight the evolving role of AI, particularly in the integration of PET and CT modalities for more accurate and comprehensive tumor segmentation. To mitigate the PET/CT concerns of higher cost and increased radiation exposure to patients compared with CT alone (Sheikhbahaei et al., 2017) and enhance the economic viability of lung cancer segmentation, we propose a novel deep learning-based connected U-Net model. This model aims to automatically fuse multimodal information by generating pseudo-PET images from CT images. The motivation behind this approach is to retain the benefits of PET-guided segmentation while minimizing the economic burden and radiation damage associated with PET examinations. Our contributions include:

- 1. Multitask modeling framework:** A connected U-Net model and teacher-student framework are presented. In this framework, the model can learn two tasks: PET generation and tumor segmentation. This modeling approach allows for lung tumor segmentation guided by learned PET knowledge, eliminating the requirement for actual PET images. This simplifies the process of tumor segmentation, making it more practical and effective in the delineation of tumor boundary.
- 2. Tumor area detection method:** To enhance the precision of tumor area delineation, we propose a tumor area detection method that enables more accurate segmentation by focusing on the area of the tumor.

This study mainly consists of six parts. **Part 1**, covers related work, presents a literature review of the field, and has been the inspiration for our research. **Part 2**, introduces the datasets and catalogs the public data sources utilized for training and validation. **Part 3** focuses on the methods, elucidating the architecture of multitask connected U-Net model and the semi-supervised learning based on our teacher-student framework and the proposed tumor area detection method. **Part 4**, provides the

experimental settings and details of the technical environment and hyperparameters. **Part 5**, summarizes the results and presents a discussion on the model's performance across various metrics and datasets. **Part 6**, is the conclusion, providing a synthesis of the findings and suggesting future directions.

## 2 Related work

In recent years, significant advancements have been made in the field of medical image segmentation. These advancements have been primarily driven by the development of novel architectures and methodologies, which have significantly improved the performance of medical image segmentation, providing valuable tools for screening, clinical diagnosis, and treatment planning of lung cancer.

### 2.1 U-Net architecture

The U-Net architecture, presented by Ronneberger et al. (2015), has been a cornerstone in medical image segmentation. The architecture, consisting of a contracting path to capture context and a symmetric expanding path to enable precise localization, can be trained end-to-end on very few images and has demonstrated superior performance on several benchmarks. Several extensions and improvements to the U-Net architecture have been proposed. For instance, nnU-Net, a deep learning-based segmentation method, automatically configures itself for any new task, reducing the need for manual hyperparameter tuning (Isensee et al., 2021; Ferrante et al., 2022). Unet 3+ extends the U-Net architecture by incorporating full-scale skip connections, allowing better feature representation and more accurate segmentation results (Huang et al., 2020). UNeXt, a novel medical image segmentation network based on a convolutional multilayer perceptron (MLP), significantly reduces the number of parameters and decreases computational complexity compared to existing methods (Valanarasu and Patel, 2022). Zhang G. et al. (2022) proposed an improved 3D dense connected UNet (I-3D DenseUNet) for lung cancer segmentation from CT images. The nested dense skip connection adopted in the I-3D DenseUNet aims to contribute similar feature maps to the encoder and decoder sub-networks, encouraging feature propagation and reuse. U-Net's advantages include efficient biomedical image processing, an encoder-decoder structure for capturing context and details, skip connections for enhanced spatial coherence, strong performance with limited data, and adaptability to various medical imaging tasks, achieving state-of-the-art results.

### 2.2 Based on transformer architecture

The transformer architecture, originally proposed for natural language processing tasks, has also been widely adapted for medical image segmentation. UCTransNet presents a new segmentation framework, which uses a CTrans module to replace the original U-Net skip connection and conducts multiscale channel-wise fusion using the Transformer (Wang et al., 2022). TransUNet

combines the strengths of the Transformer and U-Net, with the Transformer component encoding tokenized image patches from a convolutional neural network (CNN) feature map as the input sequence for extracting global contexts (Chen et al., 2021). SegFormer unifies the Transformer with the lightweight MLP decoder and includes a novel hierarchically structured Transformer encoder that outputs multiscale features (Xie et al., 2021). Recently, Tyagi et al. (2023) proposed an approach for lung cancer segmentation using an amalgamation of the vision transformer and CNN, demonstrating strong performance in lung cancer segmentation.

### 2.3 Cross-modality and teacher-student framework

Cross-modality learning, where information from one modality (e.g., PET) is used to enhance another modality (e.g., CT), has received increasing attention in recent years. The complementarity between PET and CT images allows the two modality images to be fused for automatic lung tumor segmentation. Zhang X. et al. (2022) proposed a network, based on two modality-specific encoders and two modality-specific decoders, that can fuse the complementary information and preserve modality-specific features of PET and CT images. In a similar vein, Bi et al. (2021) introduced a recurrent fusion network for multimodality PET/CT tumor segmentation, which iteratively fuses complementary image features from PET and CT images to refine segmentation results. These studies emphasize the simultaneous use of CT and PET medical images as input, to achieve information integration. However, there are some common concerns in PET scans, such as the high cost and usage of radioactive tracers among others, leading to a lack of PET data.

The teacher-student framework has been widely used in various fields and encompasses two design concepts, namely, knowledge distillation and semi-supervised learning. In the context of knowledge distillation, Hinton et al. (2015) demonstrated its effectiveness in compressing large models into smaller ones without significant loss of accuracy. For semi-supervised learning, Yu et al. (2019) presented a novel uncertainty-aware semi-supervised framework for left atrium segmentation from MR images. This framework consists of a student model and a teacher model, whereby the student model learns from the teacher model by minimizing segmentation and consistency losses with respect to the targets of the teacher model.

In this study, inspired by tumor segmentation using combined PET and CT images as well as the teacher-student framework in semi-supervised learning, we designed our teacher-student framework and proposed a connected U-Net model to integrate PET and CT knowledge. Utilizing the teacher-student framework, our model simultaneously learns the PET generation and tumor segmentation tasks, thereby leveraging the learned PET knowledge to generate pseudo-PET information, which is used in the segmentation process to eliminate the requirement for actual PET images and enhance segmentation performance.

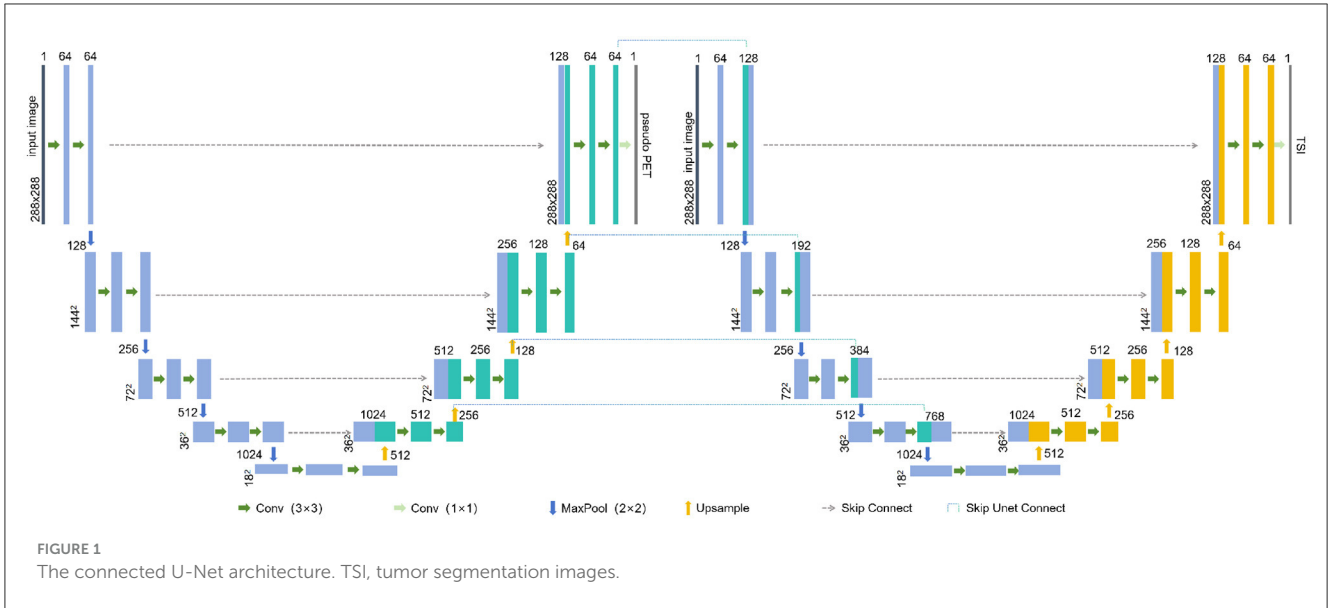
TABLE 1 Description of datasets.

Dataset	Total studies	Inclusion studies	Inclusion studies and used
NSCLC+Rad	211	126	Training (100)/ Validation (13)/ Test (13)
NSCLC-Rad	422	421	External Test (421)
MSD Task06	127	63	External Test (63)
NSCLC-Rad-Int	22	22	External Test (22)

## 3 Datasets

In this study, we used datasets from four public sources as shown in Table 1. These sources were: (1) NSCLC+Radiomics (NSCLC+Rad, updated 2021/06/01) (Bakr et al., 2018), (2) NSCLC-Radiomics (NSCLC-Rad, updated 2020/10/22), (3) NSCLC-Radiomics-Interobserver1 (NSCLC-Rad-Int, updated 2020/08/13), (4) Medical Segmentation Decathlon Task06 (MSD Task06) (Simpson et al., 2019; Antonelli et al., 2022). Among these four datasets, NSCLC + Rad has PET and CT images along with CT images with lung cancer segmentation labels; however, the segmentation labels do not correspond to PET images. The data of NSCLC-Rad comprise the largest number of CT images with lung cancer segmentation labels, but do not include PET images. The NSCLC-Rad-Int dataset also only includes CT images and segmentation labels, but each image contains multiple segmentation labels from different experts, whereby the regions that they collectively agree upon have been characterized as tumorous. The MSD Task06 is a well-known NSCLC segmentation competition dataset that includes CT images and tumor segmentation labels, specifically used to evaluate the performance of models in lung cancer segmentation. The production of these datasets involved the participation of domain experts and underwent rigorous proofreading, ensuring their high data quality.

In this study, only the NSCLC+Rad dataset was used for model training. We chose studies with CT images and corresponding PET images (Data A) or CT images with segmentation mask labels (Data B). After cleaning the data, 126 studies out of the original 211 in the NSCLC+Rad dataset were retained. These studies were divided as follows: 80% (100 studies) for training the model, 10% (13 studies) for validation, and the remaining 10% (13 studies) for testing the model. For external validation, we used three additional datasets with lung cancer mask labels, namely, NSCLC-Rad, NSCLC-Rad-Int, and MSD Task06, totaling 506 studies. To standardize the CT images across all datasets, we resized them to  $512 \times 512$  with a pixel spacing of 1 mm using trilinear interpolation. Each image was further cropped to a size of  $288 \times 288$  pixels to facilitate training on our devices.



## 4 Methods

### 4.1 Connected U-Net architecture

The connected U-Net architecture is illustrated in Figure 1. In this dual U-Net model, the first U-Net generates pseudo-PET images, whereas the second U-Net produces tumor segmentation images (TSI). For the segmentation process, the features that are used to generate pseudo-PET images are obtained by upsampling the first U-Net and connected with the features obtained by downsampling the second U-Net, thereby incorporating the PET information into the segmentation process. As illustrated in Figure 2, we employ the teacher-student framework, which enables the proposed model to simultaneously learn the PET generation and tumor segmentation tasks, whereby the model can use the learned PET knowledge to generate pseudo-PET information, using it in the segmentation process, as shown in Figure 1.

In the teacher-student framework, Data A's CT images were fed into the student U-Net, resulting in the production of pseudo-PET<sub>SA</sub> images. The quality of the pseudo-PET effects was assessed using mean-squared-error (MSE) loss. Concurrently, Data B's CT images were input into the student U-Net. This process yields pseudo-PET<sub>SB</sub> from the first U-Net, whereas the second U-Net outputs a TSI. Additionally, the teacher U-Net processes the randomly rotated the Data B's CT images to generate pseudo-PET<sub>TB</sub>. The similarity between pseudo-PET<sub>SB</sub> and pseudo-PET<sub>TB</sub> was measured using consistency loss to evaluate the consistency of the model output. The tumor segmentation model performance was evaluated through focal and dice losses, to assess tumor identification and segmentation efficiency.

**First U-Net:** Each U-Net consisted of an input layer, an architecture of four downsampling (encoder) layers and four upsampling (decoder) layers. Each encoder architecture consisted of two 3 × 3 convolutions (each followed by a leaky rectified linear unit *LeakyReLU* and instance normalization operation) with padding set to 1. We define this convolution-based operation as *Conv*. After the convolutions, a 2 × 2 max pooling operation with

a stride of 1 was employed for each downsampling step, which we define as *maxPool*. The decoder architecture consisted of a convolution operation and an upsampling operation followed by a 2 × 2 convolution, which we define as *UpSampling*. For the output of the first U-Net, a 1 × 1 convolution was used to map each feature vector to the pixel value of the PET, which we define as *outConv*. The down-sampling and up-sampling processes of the first U-Net can be represented as follows:

The input layer:

$$\hat{x}_1 = Conv(LeakyReLU(x_1)) \quad (1)$$

$$q_1 = Conv(LeakyReLU(\hat{x}_1)) \quad (2)$$

Each down-sampling (encoder) layer:

$$x_i = maxPool(q_i) \quad (3)$$

$$\hat{x}_i = Conv(LeakyReLU(x_i)) \quad (4)$$

$$q_{i+1} = Conv(LeakyReLU(\hat{x}_i)) \quad (5)$$

Each up-sampling (decoder) layer:

$$u_j = skipConnect(q_i, e_j) \quad (6)$$

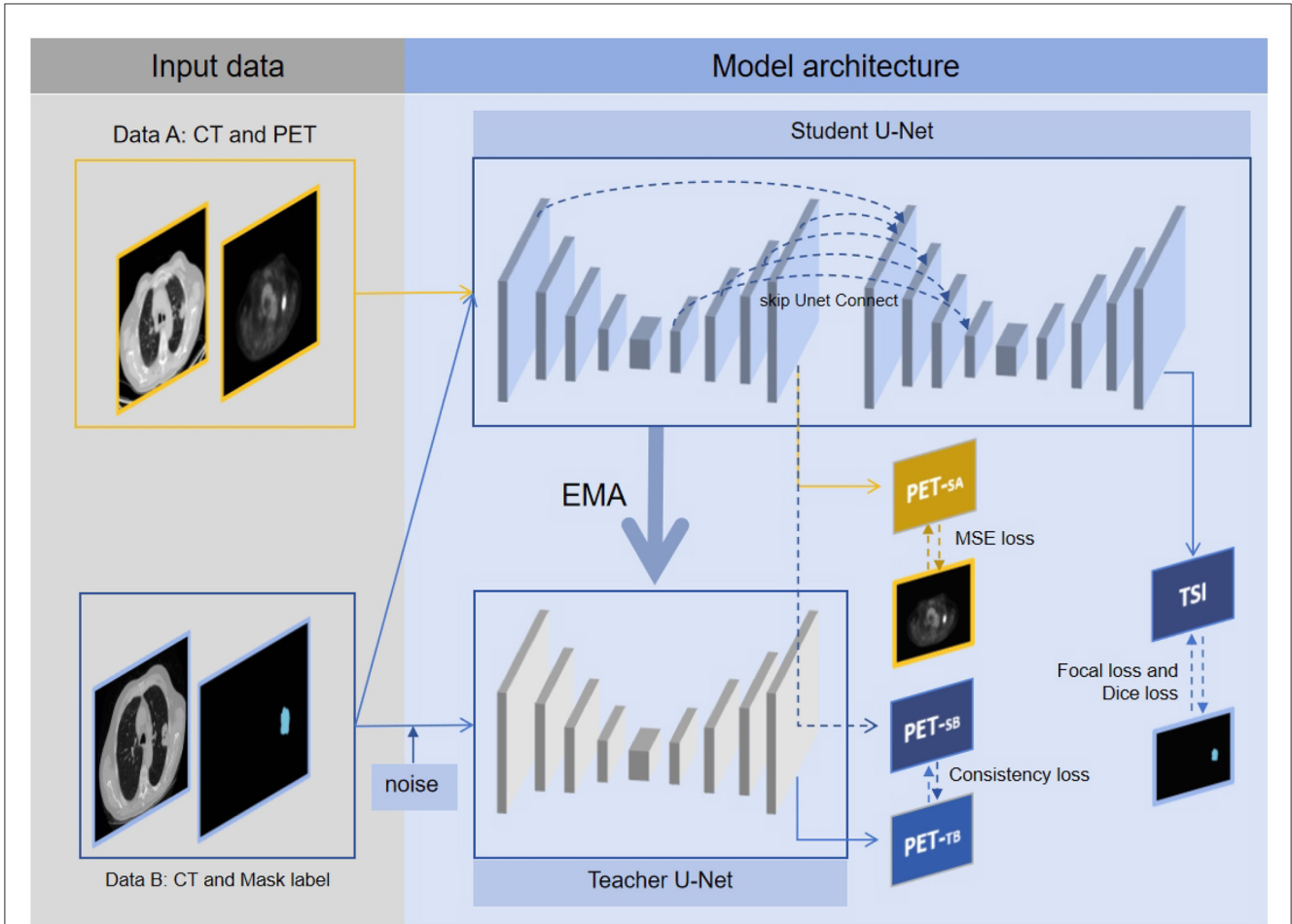
$$\hat{u}_j = UpSampling(u_j) \quad (7)$$

$$\hat{e}_j = Conv(LeakyReLU(\hat{u}_j)) \quad (8)$$

$$e_{j+1} = Conv(LeakyReLU(\hat{e}_j)) \quad (9)$$

The output layer:

$$\hat{e}_{last} = Conv(LeakyReLU(e_{last})) \quad (10)$$



**FIGURE 2**  
The multitask connected U-Net model with teacher-student framework. Data A are solely utilized for model construction, with the first U-Net in student U-Net generating pseudo  $PET_{SA}$ , optimizing parameters through MSE loss, and updating parameters of teacher U-Net via EMA. Data B are fed into both student U-Net and teacher U-Net, with the first U-Net of the student U-Net generating pseudo  $PET_{SB}$  and the pseudo  $PET_{TB}$  produced by the teacher U-Net. Based on pseudo  $PET_{SB}$  and  $PET_{TB}$ , consistency loss is obtained. Additionally, TSI is generated by the second U-Net of the student U-Net, with parameters optimized through focal, dice, and consistency losses. Through the above framework, the proposed segmentation model can simultaneously learn tumor segmentation and corresponding pseudo-PET generation, thereby integrating PET knowledge into the segmentation model.

$$y_{pet} = OutputConv(Conv(LeakyReLU(\hat{e}_{last}))) \quad (11)$$

In the Equations 1–11,  $x_i$  is the input of the  $i$ -th downsampling process.  $x_1$  is the CT image, and  $u_j$  is the input of the  $j$ -th upsampling process. The function  $skipConnect$  represents the concatenation of the corresponding feature map from upsampling to that from downsampling in the same U-Net.

**Second U-Net:** The second U-Net introduces a skip connection  $skipU_{NetConnect}$  from the first U-Net’s upsampling to its downsampling. This connection is defined as:

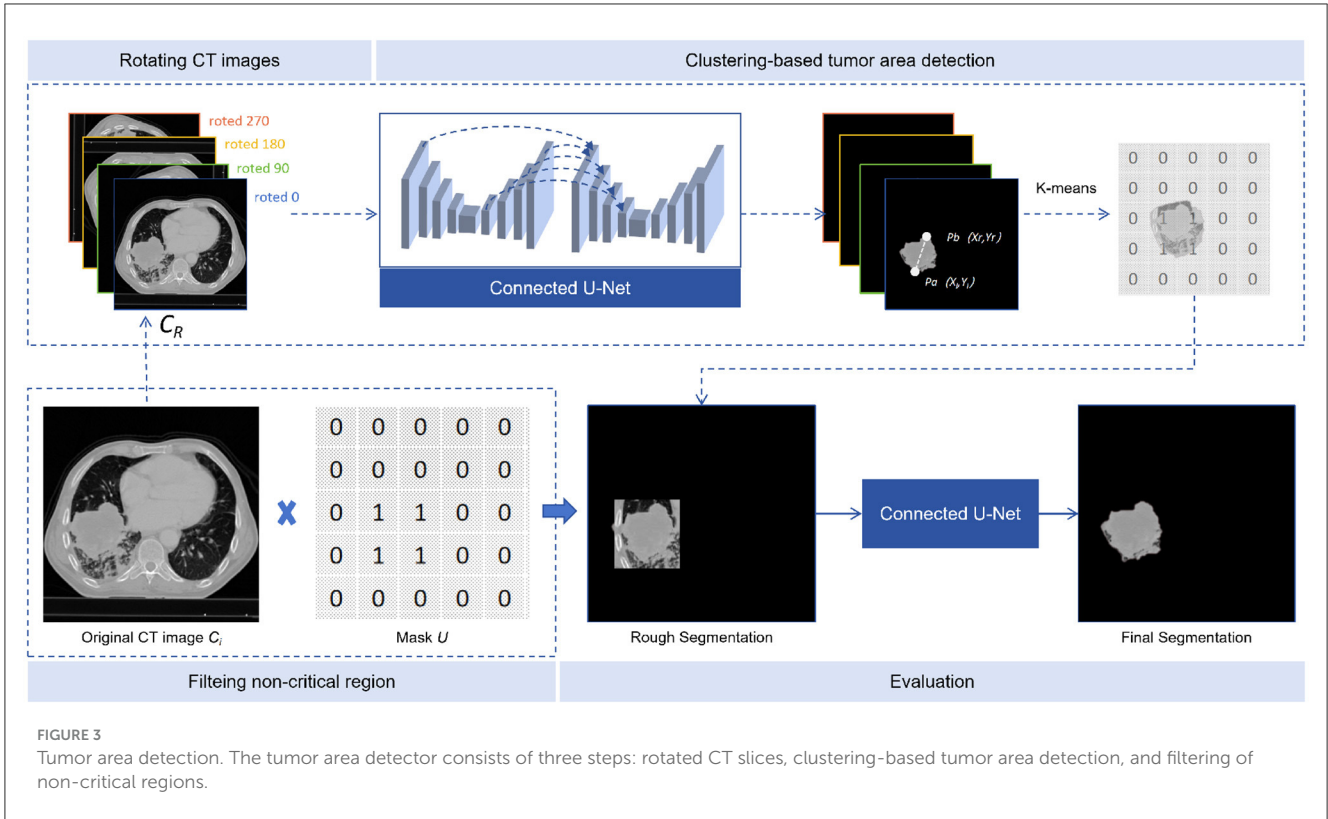
$$x_i = skipU_{NetConnect}(e_j, q_i) \quad (12)$$

In the Equation 12,  $skipU_{NetConnect}$  represents the concatenation of the corresponding feature map from the upsampling of the first U-Net to the downsampling of the second U-Net.  $q_i$  is the  $i$ -th downsampling feature of the second U-Net.  $e_j$  is the  $j$ -th upsampling feature of the first U-Net.

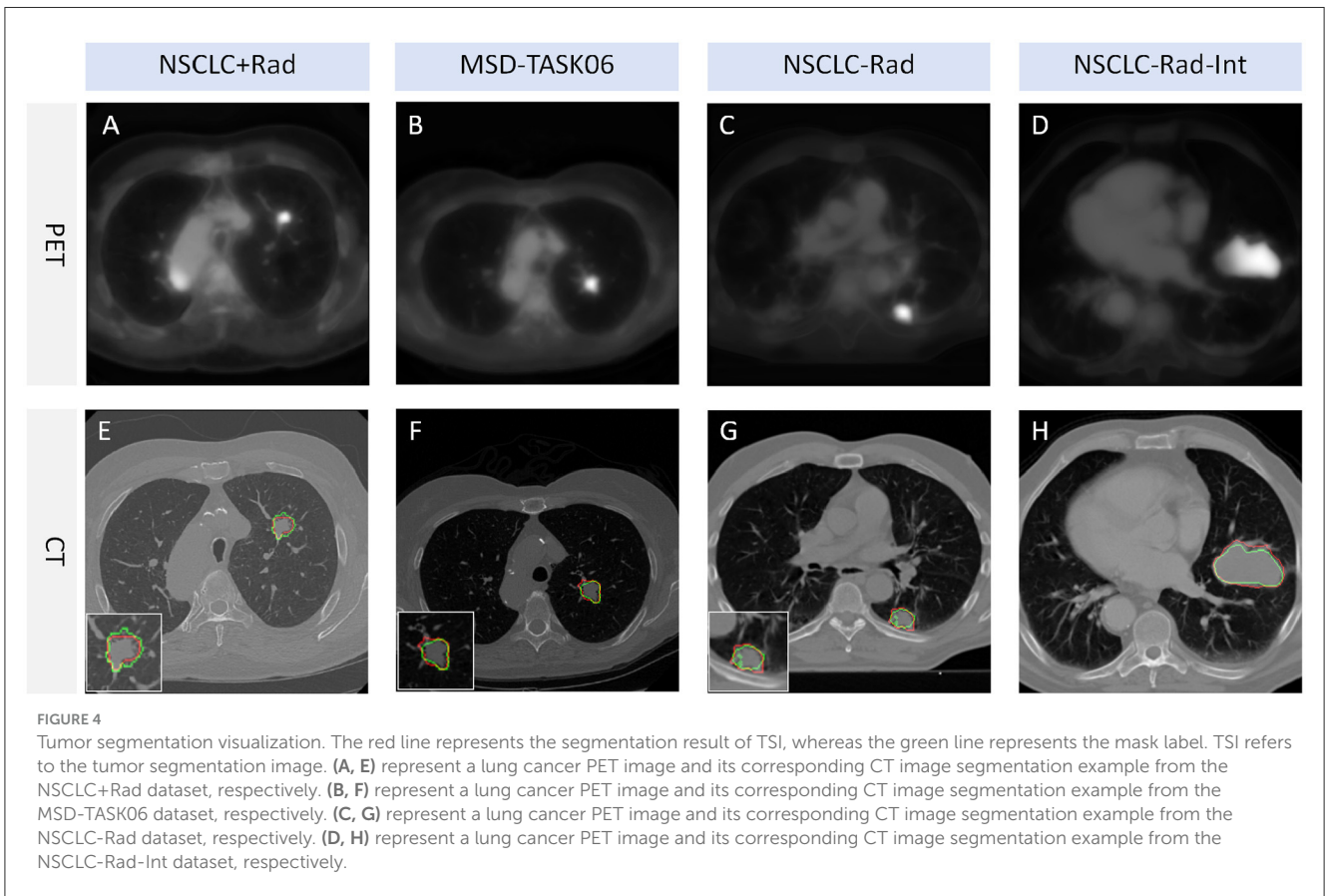
## 4.2 Semi-supervised multitask learning with teacher-student framework

PET images are more difficult to obtain than CT images. In our collected datasets, there were CT scans with segmentation labels but no PET images, and there were CT scans with PET images. To address this, we constructed a teacher-student framework to learn tumor segmentation and simultaneously learn pseudo-PET generation by semi-supervised learning, as illustrated in Figure 2. We denote Data A as  $D_l$  and Data B as  $D_u$ . The training loss  $L$  is defined as follows:

$$L = w \frac{1}{|D_l|} \sum_{x \in D_l} C(f_{\theta}(x), f_{\theta'}(x)) + \frac{1}{|D_u|} \sum_{x, y_{pet} \in D_u} H(y_{pet}, f_{\theta}(x)) + \frac{1}{|D_l|} \sum_{x, y_{seg} \in D_l} F(y_{seg}, f_{\theta}(x)) \quad (13)$$



**FIGURE 3** Tumor area detection. The tumor area detector consists of three steps: rotated CT slices, clustering-based tumor area detection, and filtering of non-critical regions.



In the Equation 13,  $f_{\theta}$  represents the connected U-Net model, and the  $\theta$  represent the model weights.  $C(f_{\theta}(x); f_{\theta'}(x))$  denotes the consistency loss, with  $w$  as the loss coefficient.  $H(y_{pet}, f_{\theta}(x))$

represents the consistency loss of the actual PET and pseudo PET, with the consistency loss computed by MSE.  $F(y_{seg}; f_{\theta}(x))$  denotes the sum of the focal and dice losses of the segmentation mask label

TABLE 2 DSC for different models on different datasets (mean ± SD).

Model	Param	Flops	NSCLC + Rad	NSCLC-Rad-Int	MSD Task06
Unet3+; Huang et al. (2020)	26.974M	199.856G	0.38 ± 0.31	0.29 ± 0.23	0.32 ± 0.30
UNeXt; Valanarasu and Patel (2022)	1.472M	0.554G	0.44 ± 0.30	0.25 ± 0.24	0.30 ± 0.30
UctransNet; Wang et al. (2022)	66.241M	54.386G	0.63 ± 0.29	0.32 ± 0.32	0.57 ± 0.32
TransUnet; Chen et al. (2021)	93.232M	40.812G	0.59 ± 0.30	0.43 ± 0.33	0.60 ± 0.31
Segformer; Xie et al. (2021)	84.595M	31.563G	<b>0.64 ± 0.29</b>	0.46 ± 0.32	0.64 ± 0.29
Ours	36.605M	50.716G	<b>0.64 ± 0.27</b>	<b>0.57 ± 0.29</b>	<b>0.66 ± 0.27</b>

DSC, Dice Similarity Coefficient. The bolded values represent the optimal evaluation metrics for different modeling strategies within the same dataset.

and TSI;  $f_{\theta'}$  represents the teacher U-Net. After training for  $t$  steps,  $\theta'$  is updated by exponential moving average (EMA):

$$\theta'_t = \alpha\theta(t - 1)' + (1 - \alpha)\theta_t \tag{14}$$

In the Equations 13, 14, both  $w$  and  $\alpha$  are dynamically adjusted according to:

$$\alpha = \min(1 - \frac{1}{step + 1}, 0.99) \tag{15}$$

The  $w$  is specified in Laine and Aila (2017) and expressed as:

$$w = 0.1 \times \exp(-5 \times (1 - T)^2) \tag{16}$$

In the Equations 15, 16,  $T = 1 - step/80$  (Laine and Aila, 2017), where  $step$  refers to the number of training steps.

### 4.3 Tumor area detection preprocessing

Tumor area detection aims to filter out areas without tumors from CT images, enabling our model to concentrate on the segmentation of tumor areas. As shown in Figure 3, the process consists of three stages: rotating CT slices, clustering-based tumor area detection, and filtering non-critical regions.

#### 4.3.1 CT slices rotation

Defining  $C_T$  as the sequence of CT slices, where  $C_i$  is the  $i$ -th slice, each slice was rotated by  $\alpha$  degrees (90, 180, and 270) to obtain three additional rotated slices. These were then inserted after the original slice, to form a new sequence  $C_R$ , as shown in Equation 17:

$$C_R = (C_0, C_0^{90}, C_0^{180}, C_0^{270}, C_1, C_1^{90}, C_1^{180}, C_1^{270}, \dots, C_i^\alpha) \tag{17}$$

#### 4.3.2 Clustering-based tumor area detection

The input of our model was  $C_R$  and the output TSI. In the multiple continuous TSI depicting the tumor boundary, the middle slice was selected as  $S_N$  and the 10 TSIs before and after it, together with the middle one, add up to 21 TSIs. On each selected TSI, points  $p_i$  and  $p_j$  can be represented as  $p_i(x_i, y_i)$  and  $p_j(x_j, y_j)$ . The Euclidean distance between pixels is calculated as Equation 18:

TABLE 3 IOU and HD95% of different models on different datasets.

Dataset	Model	IOU (mean ± SD)	HD95% (mean ± SD)
NSCLC+Rad	Unet3+	0.29 ± 0.26	116.21 ± 67.90
	UNeXt	0.33 ± 0.26	112.36 ± 74.89
	UctransNet	0.52 ± 0.27	57.60 ± 74.29
	TransUnet	0.48 ± 0.27	50.31 ± 59.95
	Segformer	<b>0.53 ± 0.27</b>	36.14 ± 54.57
	Ours	0.52 ± 0.25	<b>18.23 ± 32.25</b>
NSCLC-Rad-Int	Unet3+	0.19 ± 0.17	116.62 ± 49.44
	UNeXt	0.16 ± 0.18	112.29 ± 54.79
	UctransNet	0.24 ± 0.26	63.34 ± 54.06
	TransUnet	0.33 ± 0.28	63.79 ± 60.28
	Segformer	0.36 ± 0.28	37.94 ± 46.23
	Ours	<b>0.45 ± 0.26</b>	<b>21.46 ± 27.81</b>
MSD Task06	Unet3+	0.23 ± 0.25	133.64 ± 68.43
	UNeXt	0.22 ± 0.24	133.59 ± 73.64
	UctransNet	0.46 ± 0.29	38.83 ± 59.06
	TransUnet	0.50 ± 0.28	41.74 ± 58.51
	Segformer	0.53 ± 0.27	23.88 ± 45.08
	Ours	<b>0.55 ± 0.26</b>	<b>13.05 ± 23.27</b>

IOU, intersection over union; HD, Hausdorff distance. The bolded values represent the optimal evaluation metrics for different modeling strategies within the same dataset.

$$d(p_i, p_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{18}$$

We then find the two farthest points on each TSI:  $P_a(x_i, y_i)$  and  $P_b(x_r, y_r)$ , and form a quadruple, as shown in Equation 19:

$$quadruple = (x_i, y_i, x_r, y_r) \tag{19}$$

All quadruples from TSIs are input into the K-means algorithm to obtain two cluster centers:  $K_1$  and  $K_2$ .

#### 4.3.3 Non-critical area filtering

To compare the number of quadruples in  $K_1$  and  $K_2$ , we selected the group with more quadruples, calculating the mean values of each column in the quadruplets to obtain two points  $P_0$  and  $P'_0$ , thereby extending  $\lambda$  pixels ( $\lambda$  was set to 50 in our study)

TABLE 4 Performance of different models on NSCLC-Rad.

Model	Dice (mean $\pm$ SD)	IOU (mean $\pm$ SD)	HD95% (mean $\pm$ SD)
Unet3+	0.23 $\pm$ 0.26	0.16 $\pm$ 0.20	105.02 $\pm$ 57.40
UNeXt	0.21 $\pm$ 0.27	0.15 $\pm$ 0.21	82.30 $\pm$ 61.12
UctransNet	0.23 $\pm$ 0.29	0.17 $\pm$ 0.23	75.52 $\pm$ 58.58
TransUnet	<b>0.38 <math>\pm</math> 0.33</b>	<b>0.29 <math>\pm</math> 0.27</b>	59.17 $\pm$ 55.07
Segformer	0.31 $\pm$ 0.32	0.23 $\pm$ 0.26	51.62 $\pm$ 49.76
Ours	<b>0.38 <math>\pm</math> 0.32</b>	<b>0.29 <math>\pm</math> 0.27</b>	<b>42.26 <math>\pm</math> 41.20</b>

DSC, Dice similarity coefficient; IOU, intersection over union; HD, Hausdorff distance. The bolded values represent the optimal evaluation metrics for different modeling strategies within the same dataset.

from the points obtained to create a rectangular region. A mask  $U$  was created, with the region set to 1 and the rest to 0. The critical region of each CT slice was found by applying the mask  $U$  to  $C_i$ , and the masked image was re-predicted using the connected U-Net model, as shown in Equation 20:

$$y_{seg} = f_{\theta}(C_i \times U) \quad (20)$$

## 5 Experimental settings

The connected U-Net model was constructed using Pytorch 1.13 and trained on an NVIDIA GeForce RTX 3080 Ti GPU 12 GB, Intel(R) Core(TM) i7-12700KF, RAM 16 GB. Throughout training, data augmentation techniques such as random rotation, random flipping, and random cropping were applied. The main hyperparameters were batch size, learning rate, and optimizer. Considering our hardware capability, the hyperparameter of batch size was set to 8. The learning rate of the hyperparameter was set to  $1e-4$ , utilizing the Adam optimizer and employing a cosine learning rate control scheme (the change in learning rate for each epoch was from  $1e-4$  to  $1e-5$ ), to avoid the problem of falling into saddle points through the periodic change in the learning rate. The model that yielded the best evaluation results was selected for testing.

The model was evaluated using the dice similarity coefficient (DSC), intersection over union (IOU), and Hausdorff 95% (HD 95%) as test metrics. These metrics were calculated using the Medpy package. Additionally, we analyzed the number of parameters and the speed of our model using the thop package, comparing these values with those of other models such as Unet3+ (Huang et al., 2020), UNeXt (Valanarasu and Patel, 2022), UctransNet (Wang et al., 2022), TransUnet (Chen et al., 2021), and SegFormer (Xie et al., 2021).

## 6 Results and discussion

The integration of PET and CT modalities has revolutionized lung cancer segmentation, by providing more anatomical information for superior tumor location. Some studies emphasize the significance of leveraging PET and CT images as inputs for tumor segmentation, highlighting that integration of multimodal imaging can enhance performance outcomes (Alshmrani et al., 2023; Marinov et al., 2023; Zhou et al., 2023). However, drawbacks such as higher costs and increased radiation exposure compared to

CT alone preclude the need for innovative solutions (Sheikhabaie et al., 2017; Edelman Saul et al., 2020). In this context, our proposed connected U-Net model, combined with teacher-student semi-supervised multitask framework, emerges as a promising method, aiming to fuse pseudo-PET features for segmentation processing without the need for actual PET images.

### 6.1 Examples of model performance on different datasets

Our model underwent testing on diverse datasets, demonstrating effective generalization capabilities. Guided by pseudo-PET and tumor area detection, the model excels in accurately pinpointing tumor growth areas. The model delivers impressive performance across various datasets, as illustrated in Figure 4. This result shows that the proposed model and framework do not mandate a correspondence between PET images and segmentation labels during training. Furthermore, the prediction stage also does not require PET images, suggesting promising avenues for practical applications and flexibility in real-world scenarios.

### 6.2 DSC of different models on different datasets

This study includes a series of experiments conducted on three public datasets, demonstrating the effectiveness of the proposed methods. Our model stands out among the evaluated models, showcasing remarkable adaptability across datasets. Notably, on the challenging MSD Task06 dataset, our model achieved a high DSC score of 0.66, surpassing competitor models, as detailed in Table 2. On the NSCLC+Rad dataset, our model achieved a DSC of 0.64, highlighting its effectiveness in the context of lung cancer segmentation. Moreover, on the NSCLC+Rad and NSCLC-Rad-Int datasets, our model achieved DSC of 0.64 and 0.57, respectively, highlighting its effectiveness in the context of lung cancer segmentation. While Segformer excels on the NSCLC+Rad dataset, our model matches and even surpasses its performance on the MSD Task06 dataset. These results indicate the proposed model's ability to adapt to diverse datasets, positioning it as a promising solution for a wide array of medical image segmentation tasks. Overall, our model not only competes effectively with existing models but also showcases reliability across varied medical imaging scenarios. The efficiency of the proposed model is evident in its moderate parameter count (36.605M) and FLOPs (50.716G), striking a balance between model complexity and segmentation accuracy. This underscores the practical applicability of our model in real-world medical image segmentation applications, offering a compelling blend of high performance and computational efficiency.

### 6.3 IOU and HD 95% of different models on different datasets

The proposed model stands out with outstanding segmentation performance across different datasets, particularly on the MSD



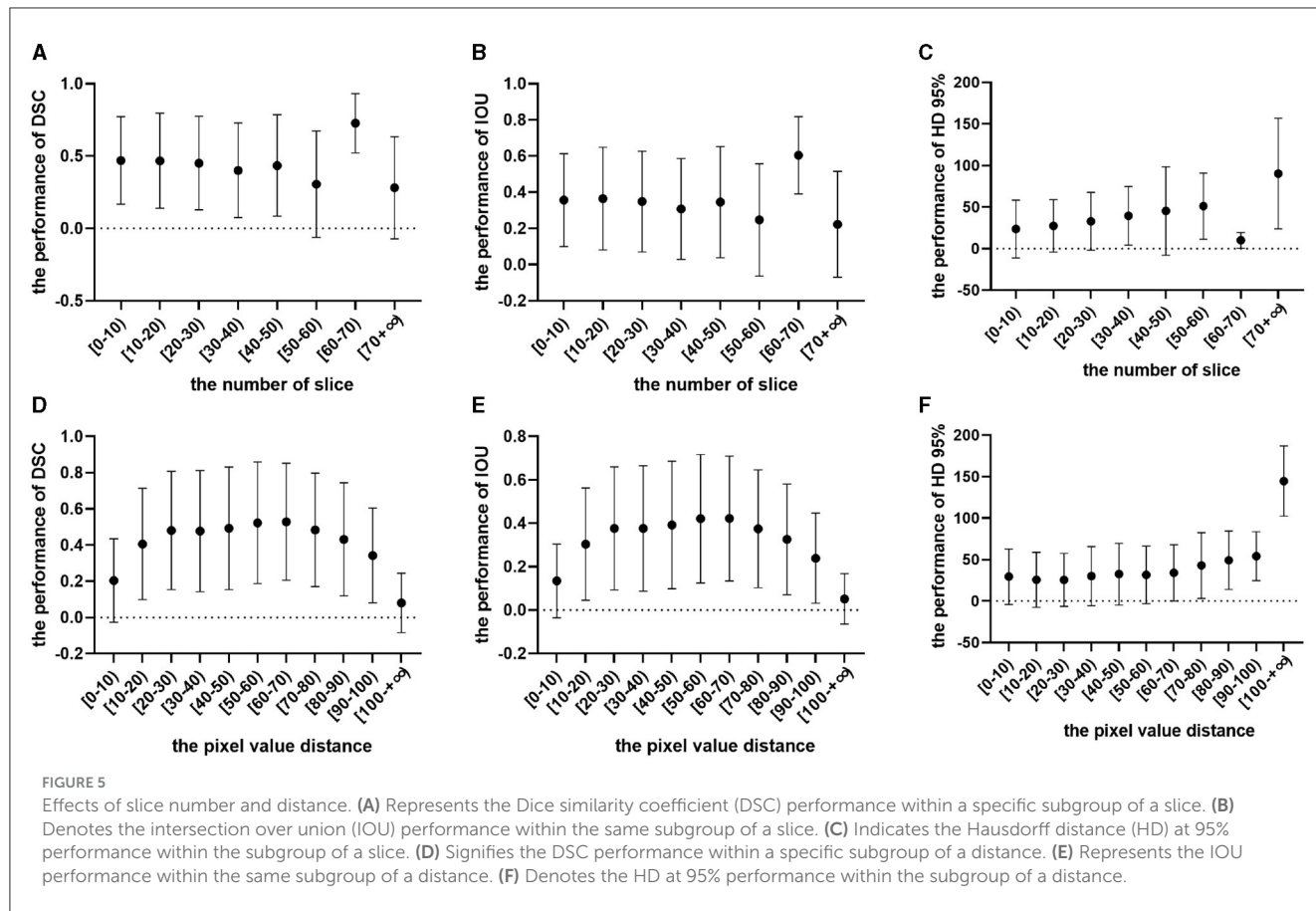


TABLE 5 Performance of multitask connected U-Net based on different components.

Dataset	Model	Dice (mean ± SD)	IOU (mean ± SD)	HD95% (mean ± SD)
NSCLC+Rad	Only based on CT images	0.59 ± 0.29	0.48 ± 0.27	44.79 ± 57.82
	PET guide	<b>0.64 ± 0.29</b>	<b>0.53 ± 0.26</b>	28.57 ± 49.11
	PET guide and area detection	<b>0.64 ± 0.27</b>	0.52 ± 0.25	<b>18.23 ± 32.25</b>
NSCLC-Rad	Only based on CT images	0.35 ± 0.33	0.27 ± 0.28	54.74 ± 54.15
	PET guide	0.24 ± 0.28	0.17 ± 0.22	62.21 ± 57.18
	PET guide and area detection	<b>0.38 ± 0.32</b>	<b>0.29 ± 0.27</b>	<b>42.26 ± 41.20</b>
NSCLC-Rad-Int	Only based on CT images	0.51 ± 0.33	0.40 ± 0.29	44.40 ± 52.61
	PET guide	0.51 ± 0.32	0.40 ± 0.28	32.99 ± 47.15
	PET guide and area detection	<b>0.57 ± 0.29</b>	<b>0.45 ± 0.27</b>	<b>21.46 ± 27.81</b>
MSD Task06	Only based on CT images	0.65 ± 0.30	0.54 ± 0.28	30.28 ± 53.86
	PET guide	0.63 ± 0.30	0.52 ± 0.28	26.83 ± 49.49
	PET guide and area detection	<b>0.66 ± 0.27</b>	<b>0.55 ± 0.26</b>	<b>13.05 ± 23.27</b>
Average	Only based on CT images	0.53 ± 0.31	0.42 ± 0.28	43.55 ± 54.61
	PET guide	0.51 ± 0.30	0.41 ± 0.26	37.65 ± 50.73
	PET guide and area detection	<b>0.56 ± 0.29</b>	<b>0.45 ± 0.26</b>	<b>23.75 ± 31.13</b>

DSC, Dice similarity coefficient; IOU, intersection over union; HD, Hausdorff distance. The bolded values represent the optimal evaluation metrics for modeling models based on different.

Task06 dataset, as shown in Table 3. It achieves the highest IOU score of 0.55, showcasing superior accuracy in delineating segmented structures compared to other models. This result highlights the effectiveness of our model in capturing the

intricate details of medical images. Furthermore, in terms of 95% HD, our model also excels with a notable score of 13.05, demonstrating its ability to precisely capture the boundaries of segmented regions. The performance of the proposed model

surpasses or closely rivals other state-of-the-art models such as Segformer.

## 6.4 External validation on the challenge dataset

The NSCLC-Rad dataset has a larger sample size. On this dataset, all the models we used performed poorly. The performance of the model depends on various factors such as the quality of the images as well as the size and shape of the tumors (Tyagi et al., 2023). The poor performance is likely the result of distributional differences between this dataset and the training data. Nevertheless, our segmentation model, based on connected U-Net and tumor area detection, exhibits notable performance compared to other models, as shown in Table 4.

## 6.5 Subgroup analysis by slice and distance

Identifying small lesions in their early stages remains a challenging task. The size of the tumor significantly affects the performance of segmentation models. Coronal CT images allow the measurement of the long and short diameters of tumors and evaluation of their area. Additionally, continuous CT scans enable the calculation of tumor height in the sagittal plane based on layer thickness and the number of scanned layers, which can estimate tumor volume.

In our investigation, subgroup analyses were conducted using slice thickness and distance to explore the impact of tumor size on model segmentation across various datasets, as shown in Figure 5. Unexpectedly, the slice number was found to have no significant impact on tumor segmentation. In further analysis, the intermediate CT section yielded the best segmentation results when dealing with multiple slices containing tumor CT. Notably, patient chest CT scans from different sources may exhibit varying slice thicknesses, leading to potential errors in estimating tumor height based on slice number.

Furthermore, the long diameter of the tumor can be represented by the pixel value distance. As the tumor pixel value distance increases, the model's segmentation performance improves significantly. However, beyond a certain threshold, further expansion of the tumor pixel value may lead to a decline in segmentation effectiveness. This phenomenon could be attributed to limited modeling data and a lack of training samples for giant tumors.

## 6.6 Ablation experiment

The teacher-student framework allows our proposed model to concurrently learn PET generation and tumor segmentation tasks. Through this approach, our proposed model can use the learned PET knowledge to generate pseudo-PET information, subsequently integrating the information into segmentation process. To verify the effectiveness of integrating pseudo-PET information, we tested it on four datasets through ablation experiments, as outlined in Table 5. The results showed that our model, when

integrating pseudo-PET images, achieved better Dice values on the NSCLC+Rad dataset (0.64 vs 0.59) than the model based solely on CT images. In addition, the average of the HD 95% results on the four datasets (37.65 vs 43.55) also indicates that the model based on guidance from learned PET knowledge exhibits better performance in the processing of edge details for tumor boundaries.

Tumor area detection aims to filter out areas without tumors from CT images, enabling the established model to further focus on segmenting only the areas where tumors are present. The results showed that this approach further enhanced the model's performance on the NSCLC-Rad and NSCLC-Rad-Int datasets, reaffirming the generalizability and effectiveness of the proposed enhancements. The MSD Task06 dataset similarly reflects consistent improvement, culminating in the best overall performance when our model utilizes learned PET knowledge and applies tumor area detection. These shared trends underscore the efficacy of the proposed methods across various datasets, reinforcing their potential for enhancing medical image segmentation tasks.

## 7 Limitation

Although, our model aimed to alleviate the economic burden and radiation risks associated with PET examinations, extensive research is required to evaluate the long-term cost-effectiveness and safety implications of this approach. Continued investigations will be crucial in assessing the viability and sustainability of implementing our proposed methodology. Moreover, the tumor area detection in this study is constrained to single tumors, warranting validation for its effectiveness in localizing multiple tumors. This aspect requires further verification to ensure the model's applicability to scenarios involving multiple tumor instances.

## 8 Conclusion

In this study, we proposed a multitask connected U-Net model and a teacher-student framework. The framework can make the model learn the PET knowledge, whereby the model performs tumor segmentation using the learned PET knowledge without the need for real PET images. This method facilitates a more detailed delineation of the tumor boundaries. In addition to incorporating PET knowledge, our tumor area detection method is also beneficial in enhancing overall performance. Future work will focus on further refining the model and validating its performance on larger and more diverse datasets.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: NSCLC+Radiomics (NSCLC+Rad, updated 2021/06/01) <https://www.cancerimagingarchive.net/collection/nsclc-radiogenomics/>, NSCLC-Radiomics (NSCLC-Rad, updated 2020/10/22) <https://www.cancerimagingarchive.net/collection/nsclc-radiomics/>, NSCLC-Radiomics-Interobserver1 (NSCLCRad-Int, updated 2020/08/13), <https://www.cancerimagingarchive.net/collection/nsclc-radiomics-interobserver1/>

[net/collection/nslc-radiomics-interobserver1/](https://collection/nslc-radiomics-interobserver1/), and Medical Segmentation Decathlon Task06 (MSD Task06) <http://medicaldecathlon.com/dataaws/>.

## Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

## Author contributions

LZ: Funding acquisition, Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Writing – original draft, Writing – review & editing. CW: Data curation, Formal analysis, Validation, Visualization, Writing – original draft, Writing – review & editing. YC: Formal analysis, Methodology, Validation, Writing – review & editing. ZZ: Conceptualization, Resources, Supervision, Validation, Writing – review & editing.

## References

- Alshmrani, G. M., Ni, Q., Jiang, R., and Muhammed, N. (2023). Hyperdense\_lung\_seg: Multimodal-fusion-based modified u-net for lung tumour segmentation using multimodality of ct-pet scans. *Diagnostics* 13:3481. doi: 10.3390/diagnostics13223481
- Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B. A., et al. (2022). The medical segmentation decathlon. *Nat. Commun.* 13:4128. doi: 10.1038/s41467-022-30695-9
- Baek, S., He, Y., Allen, B. G., Buatti, J. M., Smith, B. J., Tong, L., et al. (2019). Deep segmentation networks predict survival of non-small cell lung cancer. *Sci. Rep.* 9:17286. doi: 10.1038/s41598-019-53461-2
- Bakr, S., Gevaert, O., Echegaray, S., Ayers, K., Zhou, M., Shafiq, M., et al. (2018). A radiogenomic dataset of non-small cell lung cancer. *Scient. Data* 5, 1–9. doi: 10.1038/sdata.2018.202
- Bi, L., Fulham, M., Li, N., Liu, Q., Song, S., Feng, D. D., et al. (2021). Recurrent feature fusion learning for multi-modality pet-ct tumor segmentation. *Comput. Methods Programs Biomed.* 203:106043. doi: 10.1016/j.cmpb.2021.106043
- Bianconi, F., Palumbo, I., Spanu, A., Nuvoli, S., Fravolini, M. L., and Palumbo, B. (2020). Pet/ct radiomics in lung cancer: an overview. *Appl. Sci.* 10:1718. doi: 10.3390/app10051718
- Bourigault, E., McGowan, D. R., Mehranian, A., and Papież, B. W. (2021). “Multimodal pet/ct tumour segmentation and prediction of progression-free survival using a full-scale unet with attention,” in *3D Head and Neck Tumor Segmentation in PET/CT Challenge* (Cham: Springer), 189–201.
- Cai, L., Huang, J., Zhu, Z., Lu, J., and Zhang, Y. (2023). A localization-to-segmentation framework for automatic tumor segmentation in whole-body pet/ct images. *arXiv [Preprint]*. arXiv:2309.05446. doi: 10.48550/arXiv.2309.05446
- Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., et al. (2021). Transunet: Transformers make strong encoders for medical image segmentation. *arXiv [Preprint]*. arXiv:2102.04306. doi: 10.48550/arXiv.2102.04306
- Edelman Saul, E., Guerra, R. B., Edelman Saul, M., Lopes da Silva, L., Aleixo, G. F., Matuda, R. M., et al. (2020). The challenges of implementing low-dose computed tomography for lung cancer screening in low-and middle-income countries. *Nat. Cancer* 1, 1140–1152. doi: 10.1038/s43018-020-00142-z
- Ferrante, M., Rinaldi, L., Botta, F., Hu, X., Dolp, A., Minotti, M., et al. (2022). Application of nnu-net for automatic segmentation of lung lesions on ct images and its implication for radiomic models. *J. Clin. Med.* 11:7334. doi: 10.3390/jcm11247334
- Hinton, G., Vinyals, O., and Dean, J. (2015). Distilling the knowledge in a neural network. *arXiv [Preprint]*. arXiv:1503.02531.
- Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., et al. (2020). “Unet 3+: A full-scale connected unet for medical image segmentation,” in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Barcelona: IEEE), 1055–1059.
- Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., and Maier-Hein, K. H. (2021). nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* 18, 203–211. doi: 10.1038/s41592-020-01008-z
- Ju, W., Xiang, D., Zhang, B., Wang, L., Kopriva, I., and Chen, X. (2015). Random walk and graph cut for co-segmentation of lung tumor on pet-ct images. *IEEE Trans. Image Proc.* 24, 5854–5867. doi: 10.1109/TIP.2015.2488902
- Laine, S., and Aila, T. (2017). Temporal ensembling for semi-supervised learning. *arXiv [Preprint]*. doi: 10.48550/arXiv.1610.02242
- Leiter, A., Veluswamy, R. R., and Wisnivesky, J. P. (2023). The global burden of lung cancer: current status and future trends. *Nat. Rev. Clin. Oncol.* 20, 624–639. doi: 10.1038/s41571-023-00798-3
- Li, L., Lu, W., Tan, Y., and Tan, S. (2019). Variational pet/ct tumor co-segmentation integrated with pet restoration. *IEEE Trans. Radiat. Plasma Med. Sci.* 4, 37–49. doi: 10.1109/TRPMS.2019.2911597
- Marinov, Z., Reiß, S., Kersting, D., Kleesiek, J., and Stiefelhagen, R. (2023). Mirror u-net: Marrying multimodal fission with multi-task learning for semantic segmentation in medical imaging. *arXiv [Preprint]*. arXiv:2303.07126. doi: 10.1109/ICCVW60793.2023.00242
- Mercieca, S., Belderbos, J. S., and van Herk, M. (2021). Challenges in the target volume definition of lung cancer radiotherapy. *Transl. Lung Cancer Res.* 10:1983. doi: 10.21037/tlcr-20-627
- Ronneberger, O., Fischer, P., and Brox, T. (2015). “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18* (Munich: Springer), 234–241.
- Sheikhbahaee, S., Mena, E., Yanamadala, A., Reddy, S., Solnes, L. B., Wachsmann, J., et al. (2017). The value of fdg pet/ct in treatment response assessment, follow-up, and surveillance of lung cancer. *Am. J. Roentgenol.* 208, 420–433. doi: 10.2214/AJR.16.16532

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by The Academic Mentor Program for Backbone Teachers of Traditional Chinese Medicine (Zhongjing) School at Henan University of Chinese Medicine (No. 15102138-2024).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Simpson, A. L., Antonelli, M., Bakas, S., Bilello, M., Farahani, K., Van Ginneken, B., et al. (2019). A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv [Preprint]*. arXiv:1902.09063.
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 71, 209–249. doi: 10.3322/caac.21660
- Tyagi, S., Kushnure, D. T., and Talbar, S. N. (2023). An amalgamation of vision transformer with convolutional neural network for automatic lung tumor segmentation. *Comput. Med. Imag. Graph.* 108:102258. doi: 10.1016/j.compmedimag.2023.102258
- Valanarasu, J., and Patel, V. (2022). Unext: Mlp-based rapid medical image segmentation network. *arXiv [Preprint]*. arXiv:2203.04967. doi: 10.1007/978-3-031-16443-9\_3
- Wang, H., Cao, P., Wang, J., and Zaiane, O. R. (2022). Uctransnet: rethinking the skip connections in u-net from a channel-wise perspective with transformer. *Proc. Int. AAAI Conf. Weblogs. volume 36*, 2441–2449. doi: 10.1609/aaai.v36i3.20144
- Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., and Luo, P. (2021). Segformer: Simple and efficient design for semantic segmentation with transformers. *Adv. Neural Inf. Process. Syst.* 34, 12077–12090. doi: 10.48550/arXiv.2105.15203
- Yu, L., Wang, S., Li, X., Fu, C.-W., and Heng, P.-A. (2019). “Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation,” in *Medical image computing and computer assisted intervention-MICCAI 2019: 22nd international conference, Shenzhen, China, October 13-17, 2019, Proceedings, Part II 22* (Cham: Springer), 605–613.
- Zhang, G., Yang, Z., and Jiang, S. (2022). Automatic lung tumor segmentation from ct images using improved 3d densely connected unet. *Med. Biol. Eng. Comput.* 60, 3311–3323. doi: 10.1007/s11517-022-02667-0
- Zhang, X., Zhang, B., Deng, S., Meng, Q., Chen, X., and Xiang, D. (2022). Cross modality fusion for modality-specific lung tumor segmentation in pet-ct images. *Phys. Med. Biol.* 67:225006. doi: 10.1088/1361-6560/ac994e
- Zhou, T., Zhang, X., Lu, H., Li, Q., Liu, L., and Zhou, H. (2023). Gmre-iunet: isomorphic unet fusion model for pet and ct lung tumor images. *Comput. Biol. Med.* 166:107514. doi: 10.1016/j.compbiomed.2023.107514