RESEARCH ARTICLE

MEDICAL PHYSICS

Automated segmentation of lung, liver, and liver tumors from Tc-99m MAA SPECT/CT images for Y-90 radioembolization using convolutional neural networks

Anucha Chaichana[1] | Eric C. Frey[2,3] | Ajalaya Teyateeti[4] | Kijja Rhoongsittichai[4] | Chiraporn Tocharoenchai[1] | Pawana Pusuwan[4] | Kulachart Jangpatarapongsa[5]

[1] Department of Radiological Technology, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand

[2] Johns Hopkins School of Medicine, Johns Hopkins University, Baltimore, Maryland 21218, USA

[3] Radiopharmaceutical Imaging and Dosimetry, LLC, Lutherville, Maryland 21093, USA

[4] Department of Radiology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand

[5] Center for Research and Innovation, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand

Correspondence
Kulachart Jangpatarapongsa, Center for Research and Innovation, Faculty of Medical Technology, Mahidol University, Bangkok, 10700, Thailand.
Email: kulachart.jan@mahidol.ac.th

Funding information
Mahidol University's Academic Development Scholarship; Faculty of Medical Technology, Mahidol University; Johns Hopkins Radiological Physics Division

Abstract

Purpose: ^{90}Y selective internal radiation therapy (SIRT) has become a safe and effective treatment option for liver cancer. However, segmentation of target and organ-at-risks is labor-intensive and time-consuming in ^{90}Y SIRT planning. In this study, we developed a convolutional neural network (CNN)-based method for automated lungs, liver, and tumor segmentation on ^{99m}Tc-MAA SPECT/CT images for ^{90}Y SIRT planning.

Methods: ^{99m}Tc-MAA SPECT/CT images and corresponding clinical segmentations were retrospectively collected from 56 patients who underwent ^{90}Y SIRT. The collected data were used to train three CNN-based segmentation algorithms for lungs, liver, and tumor segmentation. Segmentation performance was evaluated using the Dice similarity coefficient (DSC), surface DSC, and average symmetric surface distance (ASSD). Dosimetric parameters (volume, counts, and lung shunt fraction) were measured from the segmentation results and were compared with clinical reference segmentations.

Results: The evaluation results show that the method can accurately segment lungs, liver, and tumor with median [interquartile range] DSCs of 0.98 [0.97–0.98], 0.91 [0.83–0.93], and 0.85 [0.71–0.88]; surface DSCs of 0.99 [0.97–0.99], 0.86 [0.77–0.93], and 0.85 [0.62–0.93], and ASSDs of 0.91 [0.69–1.5], 4.8 [2.6-8.4], and 4.7 [3.5–9.2] mm, respectively. Dosimetric parameters from the three segmentation networks show relationship with those from the reference segmentations. The overall segmentation took about 1 min per patient on an NVIDIA RTX-2080Ti GPU.

Conclusion: This work presents CNN-based algorithms to segment lungs, liver, and tumor from ^{99m}Tc-MAA SPECT/CT images. The results demonstrated the potential of the proposed CNN-based segmentation method for assisting ^{90}Y SIRT planning while drastically reducing operator time.

KEYWORDS

90Y selective internal radiation therapy, 99mTc macro-aggregated albumin SPECT/CT, convolutional neural network, hepatocellular carcinoma, segmentation

---

1 | INTRODUCTION

According to the GLOBOCAN 2018 data, liver cancer is the sixth most commonly diagnosed cancer and the fourth most common cause of cancer-related death worldwide; hepatocellular carcinoma (HCC) accounts for more than 80% of primary liver cancers.[1] Median survival of untreated liver cancers range from less than one month to ten months.[2] Surgical resection is the first treatment option for solitary HCCs in patients with preserved liver function. However, the large majority of patients present in locally advanced or metastatic stages or are poor surgical candidates for resection.[3]

In the past decade, radioembolization (RE), also known as selective internal radiation therapy (SIRT), with yttrium-90 (^{90}Y) microspheres has been established as a safe and effective treatment option in the management of patients with primary and metastatic liver cancer.[4] Several studies have shown that, in localized disease, outcomes for ^{90}Y SIRT were similar to or better than those for other locoregional therapies such as transarterial chemoembolization or ablation.[5,6] In ^{90}Y SIRT, millions of microspheres containing ^{90}Y are delivered into tumor-feeding hepatic arteries during a femoral arterial catheterization. The goal is to deliver a tumoricidal absorbed dose via ^{90}Y microspheres to the tumor(s) while sparing the normal liver parenchyma. Before committing treatment, the absorbed dose to tumor(s) and normal liver parenchyma must be estimated to ensure delivery of an appropriate quantity of ^{90}Y microspheres. In ^{90}Y SIRT planning using the partition model,[7] macro-aggregated albumin (MAA) particles labeled with technetium-99 m (^{99m}Tc) are injected, and planar scintigraphy and single-photon emission computed tomography with X-ray computed tomography (SPECT/CT) are performed.[8] These images are used to quantify a possible liver-to-lung shunt, determine extrahepatic uptake, and predict the intrahepatic distribution of the ^{90}Y microspheres, enabling pre-therapeutic dosimetry.[7]

Accurate segmentation of target and organs-at-risk (OARs) is critical for precise predictive dosimetry because it directly affects the estimation of absorbed dose. This is especially important when planning ^{90}Y SIRT, where the estimated absorbed dose is used to calculate the ^{90}Y activity to achieve the best efficacy and lowest toxicity profile.[9] In current clinical practice, segmentation of lung, liver, and liver tumor volumes for ^{90}Y SIRT planning is generally performed manually with little machine assistance. This process is a labor-intensive and time-consuming task, requiring many hours of physician attention for a single subject. Therefore, an automated segmentation algorithm for segmenting the lungs, liver, and liver tumors from ^{99m}Tc-MAA SPECT/CT images is highly desirable for ^{90}Y SIRT planning.

In the past, numerous semi-automatic and automatic segmentation methods for lung and liver segmentation of CT images have been developed. For instance, segmentation methods based on region-growing,[10,11] active shape models,[12,13] supervised voxel classification,[14] and recently neural network[15] have been used for lung segmentation. For liver segmentation, segmentation methods using region-growing,[16] graph cut,[17] geometric deformable models,[18] and neural networks[19] have been proposed. Despite these efforts, lungs and liver segmentation remain challenging tasks, especially using the non-contrast and low-dose CT images typically acquired as part of ^{99m}Tc-MAA SPECT/CT.

In recent years, deep learning methods, in particular convolutional neural networks (CNN), have emerged as a powerful tool for image analysis, achieving state-of-art performance in numerous computer vision problems including image classification,[20] object detection,[21] and semantic segmentation.[22] In medical image segmentation, the application of CNNs has ranged from anatomical structure segmentation to pathological lesion segmentation, particularly for computed tomography (CT) images and magnetic resonance images (MRI).[23–25] Although numerous segmentation algorithms have been developed, there has been little work on SPECT segmentation in general and CNNs for SPECT in particular. This is also true of the non-contract, low-dose CT images that are acquired as part of SPECT/CT acquisitions for attenuation compensation.

This study aimed to develop an automated image segmentation method that can be used for ^{90}Y SIRT planning using pretreatment simulation based on ^{99m}Tc-MAA SPECT/CT images. For this purpose, we developed and evaluated three CNN-based segmentation algorithms for lung and liver segmentation using non-contrast, low-dose CT images and for liver tumor segmentation using ^{99m}Tc-MAA SPECT images.

2 | MATERIALS AND METHODS

We used three CNN-based segmentation algorithms to segment the lungs, liver, and tumor liver independently. The lung (LungNet) and liver (LiverNet) segmentation networks were trained to segment lungs and whole liver, respectively, from low-dose CT images acquired in the ^{99m}Tc-MAA study. The tumor segmentation network (TumorNet) was trained to segment liver tumors from ^{99m}Tc-MAA SPECT images. The details of the dataset used in this study are explained in Section 2.1. The three networks shared the same network architecture designed to perform segmentation of ^{99m}Tc-MAA SPECT/CT images. The details of network architecture and its implementation are explained in Sections 2.2. and 2.3., respectively.

2.1 | Dataset

2.1.1 | Image acquisition

We retrospectively collected data from 56 HCC patients who underwent evaluation for ^{90}Y SIRT at Siriraj Hospital during 2016 and 2019. Each dataset includes ^{99m}Tc-MAA SPECT images, CT images, and corresponding manual lung, liver, and liver tumor segmentation performed by nuclear medicine physicians. Institutional review board approval was obtained, and no informed consent was required for this retrospective analysis.

For each patient, a SPECT/CT scan was performed about 1 h after injection of 4–5 mCi of ^{99m}Tc-MAA using either a GE Discovery 670 or 670 pro system (GE Healthcare, USA) with a low-energy high-resolution collimator. SPECT acquisitions were performed in 2 FOVs to encompass the lungs and liver. The acquisition energy windows were set at $140 \pm 20\%$ keV and $120 \pm 10\%$ keV for window-based scatter correction. Sixty projections in a 128×128 matrix size were acquired over a 360° angular range with a 20-s acquisition duration at each view. SPECT images were reconstructed on a Xeleris workstation (GE Healthcare, USA) using 5 iterations and 10 subsets of OSEM with CT-based attenuation correction, window-based scatter correction, and collimator-detector response modeling. The CT scan was acquired in low-dose mode (80 reference mAs, 120 kVp) during free breathing with a 3.75 mm slice thickness. Reconstructed SPECT and CT images were resampled to have the exact dimensions of $256 \times 256 \times 256$ voxels with an isotropic voxel size of 2.20 mm.

2.1.2 | Reference segmentation

Segmentations performed by nuclear medicine physicians served as the reference for comparison in this work. The segmentations were performed using the Dosimetry Toolkit on the Xeleris workstation. These reference segmentations include volume of interest (VOI) definitions for the lungs, liver, and liver tumor. For the lung segmentation, physicians used CT images and intensity-based semi-automatic tools provided in the software to define initial lung VOIs. Manual modification was usually performed to remove the trachea, primary bronchus, and lung base near hepatic dome from the lung VOI. For liver segmentation, the entire liver was generally segmented based on the non-contrast, low-dose CT from the MAA study. Regions of interest (ROIs) were manually drawn on CT slices and stacked to create liver VOIs. Tumor VOIs were initially created on SPECT images using region-growing and thresholding methods; manual modification was then performed to account for the potential overlooked or over-covered tumor volumes that appeared on the fused SPECT/CT, hepatic angiography, and contrast-enhanced CT. The final segmentation results were regarded as the reference segmentation for the subsequent training and evaluation processes.

2.2 | Network architecture

We achieved tumor, liver, and lung segmentation by using separate networks for (1) lung segmentation, (2) liver segmentation, and (3) tumor segmentation. All three networks shared the same network architecture (see Figure 1). Each network was comprised of an encoding path to extract features from input data and a decoding path to return to the original resolution. Between the encoding and decoding paths, skip connections were used to forward early-extracted features from encoding directly to the decoding path. The network was designed to receive a small 3D patch as input and produce a small 3D patch as segmentation output. The output patches need to be reassembled into the segmentation result of the original (full-size) image.

Typically, fully convolutional network (FCN) architectures (e.g., 3D U-Net,[26] V-Net,[23] DenseVNet[25]) use full images (or volumes) as inputs and then use pooling/unpooling (or strided convolution/deconvolution in V-Net and DenseVNet) layers to limit the number of parameters and resulting memory requirement. These layers can lead to the loss of spatial information from repeated downsampling operations. Instead, the proposed network uses sub-volumes (3D patches) as inputs. This strategy reduces the memory requirements of the network, thereby removing the need for a pooling layer. More importantly, sampling volumes into small 3D patches substantially increases the number of training examples without the need for data augmentation. Similar to DenseNet and DenseVNet structures, the encoding path contains dense connectivity to support gradient propagation and feature reuse. However, the proposed network does not use pooling layers or strided convolution for downsampling. The decoding path upsampled the feature map using only deconvolution layers (also known as transposed convolution) instead of bilinear upsampling in DenseVnet.

In the encoding path, there are 16 convolutional layers, where each layer was defined as a composite function consisting of 3D convolution, batch normalization (BN),[27] and a Parametric Rectified Linear Unit (PReLU) activation function.[28] The convolution was responsible for producing a feature map by the convolution of the input with the encoder kernel. The kernel contained the weights (also known as learnable parameters) to be learned during the network training. The kernel weights represent the structure or feature that the kernel will detect. The number of kernels in the convolutional layer defines the number of output feature maps. Batch normalization was used to normalize the convolution's feature map to the standard Gaussian distribution. Batch normalization helps in speed up training by allowing use
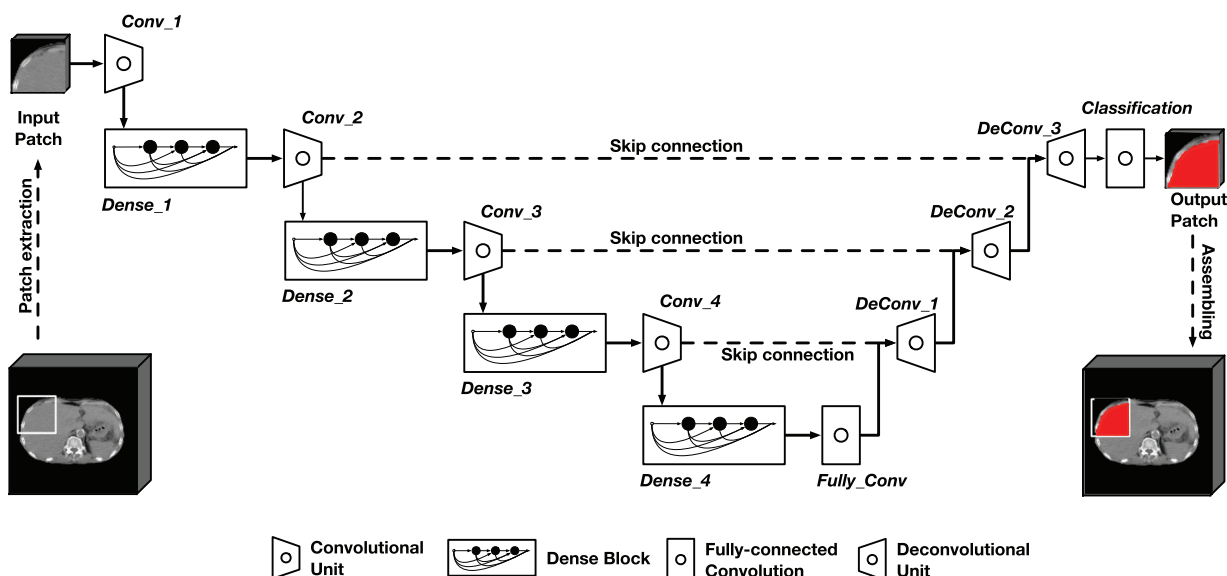
FIGURE 1 The schematic representation of the proposed convolutional neural network architecture

of high learning rates without overfitting.[27] The PReLU activation function was used to increase non-linearity in the output that has been computed using linear operations during the convolution.[28]

The encoding path can be divided into four stages. Each stage contains four convolutional layers: one in the convolutional unit and three in the dense block. The convolutional unit allows the network to increase the receptive field of the feature maps being used as input for subsequent layers. The output of the l-layer convolutional unit (C_l) can be expressed as:

$$C_l = H_l\,(C_{l-1}),\qquad(1)$$

where C_{l-1} is the output of the previous layer or the input volume in the first layer, and H_l is a composite function consisting of convolution, PReLU activation function, and batch normalization, as previously described. We designed each convolutional unit to perform convolution using a volumetric kernel of $3 \times 3 \times 3$ voxels with stride one and no padding applied to the input; stride refers to the amount of movement between consecutive applications of the kernel to the input volume. Each convolutional unit had an output size of two voxels smaller and a number of kernels two times larger than the previous stage. The architecture was designed not to include pooling operations to avoid the loss of spatial resolution of the feature map. However, due to memory constraints, we allowed the convolutional unit to gently reduce the feature map's size as the number of feature maps increased during passage through network. The numbers and sizes of kernels for each layer are shown in Table 1.

At each stage, following the convolutional unit, we used a dense block to extract further information. Motivated by Gibson et al.[25] and Huang et al.,[29] each dense block contained three convolutional layers; the input of each layer was the concatenated output of all preceding layers within the same block. The concatenated feature map helps increase variation in the input of subsequent layers by allowing the feature map learned by different layers to be accessed by subsequent layers. This encourages feature reuse throughout the dense block and leads to a more compact network that is easier to train and highly parameter efficient.[29] The output of the l-th layer within a dense block (D_l) can be expressed as:

$$D_l = H_l\,([D_0, D_1, \dots, D_{l-1}]),\qquad(2)$$

where $[D_0, D_1, \dots, D_{l-1}]$ refer to the concatenation of the feature maps produced in the preceding layers 0, 1, …, $l-1$, and H_l is a composite function: convolution, PReLU, and batch normalization. The convolution operation in a dense block was performed using a volumetric kernel of size $3 \times 3 \times 3$ voxels with stride one. However, in contrast to the convolutional unit, where padding was not applied to the input. Instead, each side of the input was zero-padded by one voxel to keep the size of the output feature map the same as its input. Each dense block had a number of kernels two times larger than the previous stage (see Table 1.).

Between the encoding path and the decoding path, we added a fully-connected convolutional layer (Fully_Conv) that performed convolution using 256 kernels of size $1 \times 1 \times 1$ voxel. This allowed the reducing number of feature maps to 256 and retained spatial information about the decoding path's input, thus reducing memory requirements. As we observed in experiments, adding the fully-connected convolutional layer also improved accuracy and reduced convergence time.

TABLE 1 Detailed parameters of the proposed networks

Layer	Kernel size	Sub-unit(#layers × #kernels)	Zero-padding	Input(size × #feature map)	Output(size × #feature map)
Conv_1	3^3	1×16	–	$24^3 \times 1$	$22^3 \times 16$
Dense_1	3^3	3×16	Yes	$22^3 \times 16$	$22^3 \times 64$
Conv_2	3^3	1×32	–	$22^3 \times 64$	$20^3 \times 32$
Dense_2	3^3	3×32	Yes	$20^3 \times 32$	$20^3 \times 128$
Conv_3	3^3	1×64	–	$20^3 \times 128$	$18^3 \times 64$
Dense_3	3^3	3×64	Yes	$18^3 \times 64$	$18^3 \times 256$
Conv_4	3^3	1×128	–	$18^3 \times 256$	$16^3 \times 128$
Dense_4	3^3	3×128	Yes	$16^3 \times 128$	$16^3 \times 512$
Fully_Conv	1^3	1×256	–	$16^3 \times 512$	$16^3 \times 256$
DeConv_1	3^3	1×128	–	$16^3 \times 384$	$18^3 \times 128$
DeConv_2	3^3	1×64	–	$18^3 \times 192$	$20^3 \times 64$
DeConv_3	3^3	1×64	–	$20^3 \times 96$	$22^3 \times 64$
Classification	1^3	1×2	–	$22^3 \times 80$	$22^3 \times 2$

While several pairs of convolutional units and dense blocks were used to increase feature richness for robust segmentation, the spatial resolution of the feature maps was simultaneously reduced. The loss of spatial resolution is typically not beneficial for image segmentation tasks, where the spatial position of features is critical for boundary delineation.[30] Therefore, it is necessary to restore the spatial resolution of the feature maps.

The decoding path is responsible for upsampling the low-resolution feature map(s) produced by the encoding path. In the proposed network, we used a series of 3D deconvolution operations to slightly increase the feature map's size up to the segmentation resolution. Unlike a fixed layer (e.g., bilinear or trilinear upsampling), kernel parameters in deconvolution can be learned while training the network. We observed that, with an equal number of learnable parameters, the network that used 3D deconvolution as an upsampling operator provided better training accuracy and computational efficiency than a network that used the trilinear upsampling combined with typical 3D convolution.

The decoding path contained three deconvolutional units (DeConv_1 to DeConv_3), which were a composite function consisting of 3D deconvolution, a PReLU activation function, and batch normalization. The deconvolution used a volumetric kernel of size $3 \times 3 \times 3$ voxels with stride one and no padding to produce an output feature map of size two voxels larger than the previous stage.

We added skip connections to forward the early-extracted feature maps from the encoding path directly to the decoding path on the same scale. Skip connection is represented in Figure 1 by the horizontal connection between convolutional unit and deconvolutional units. The skip connections allowed the network to combine shallow, fine, appearance information and deep, coarse, semantic information.[22]

At the end of the network, a fully-connected convolutional layer using 2 kernels of size $1 \times 1 \times 1$ voxels was used to encode semantic information and produce 2 feature maps. Softmax function[31] was then applied voxel-wise and return probabilities of each class (foreground, background) for each voxel with the target class (foreground) having the highest probability.

2.3 | Network training

We divided 56 images into training, validation, and test set containing 30, 6, and 20 images, respectively. We used all the training data without any data augmentation. All the networks were trained and validated on the training and validation set, respectively. The final model was chosen based on validation accuracy, and then the test set was used to measure the performance. Three networks for lung segmentation (LungNet), liver segmentation (LiverNet), and tumor segmentation (TumorNet) were trained independently. The goal of network training is to find the optimal learnable parameters (e.g., kernel weights, bias, PReLU parameters) that minimize the segmentation error determined by a loss function. In this work, we used weighted cross-entropy as loss function to solve the class imbalance problem inherent in the dataset.[32] The learnable parameters were updated using the Adam optimizer[33] with default parameter values ($\beta 1 = 0.9$, $\beta 2 = 0.999$, $\varepsilon = 10^{-8}$) and initial learning rate of 0.001. During training, the network was evaluated using a holdout validation set after every 2 epochs. If the performance (as measured by the Dice similarity coefficient) on the validation set did not improve for 10 epochs, then the training process

was stopped. The network with the highest validation performance of all the tested epochs was selected and used for testing (using the test set).

To reduce memory requirements, we trained the networks with small 3D patches instead of the entire SPECT or CT volumes. The 3D patches were sampled using a sliding window with overlap between neighboring patches. During training, the overlap was five voxels to increase number of patches available for training. Since the network has an output size of 1 voxel smaller than the input size in each direction, the overlap was 1 voxel during prediction to ensure that the output patch could be assembled into a continuous full-size image without overlap. Extracted patches can be categorized into foreground patches (at least one percent of the total voxels are foreground) and background patches. For every epoch, we randomly extracted 80% of foreground patches and 20% of background patches. The randomly extracted patches were used to train the network in a batch of size 20. This process of randomly sampling patches was repeated for each training epoch. During hyperparameter tuning, different patch size (20^3, 24^3, and 28^3) and the total number of patches per epoch (3,760 and 7,500) were used to train the proposed network, and the validation accuracy was monitored. We observed that increasing patch size from 24^3 to 28^3 yielded a minimal improvement in validation accuracy while convergence time increased. The network trained with 3,760 and 7,500 samples (patches) per epoch provided similar validation accuracy, but 7,500 samples required shorter convergence training time. Hence, we used the patch size of 24^3 and 7,500 training samples to train the proposed networks and their variants in the rest of the study.

We implemented the networks in the Python programming language using the PyTorch library based on the implementation published by Dolze et al.[34] Training and testing were performed on a computer equipped with 32 GB of memory, an Intel® Core™ i7-8700K CPU, and an Nvidia GeForce RTX 2080Ti GPU with 11 GB of video memory. Once training was completed, each network required less than 1 min to segment an input image of size $256 \times 256 \times 256$ voxels.

2.4 | Evaluation and experiment

2.4.1 | Network evaluation

After training, performance of each network was tested on the test set. Segmentation performance for each network was assessed relative to the reference segmentation using the Dice similarity coefficient (DSC),[35] surface Dice similarity coefficient (Surface DSC) with a tolerance parameter of three voxel widths,[36] and average symmetric surface distance (ASSD).[37] This study also compared the volume, counts, and lung shunt frac-

tion (LSF) calculated from the network segmentation results with those from reference segmentation. The LSF represents the degree of blood shunting between the liver and lungs, which is essential for calculating the prescribed activity of ^{90}Y microsphere. The LSF can be calculated from ^{99m}Tc-MAA SPECT images as the total counts in lung divided by total count in lung plus liver.

2.4.2 | Segmentation algorithm comparison

We compared segmentation performance using the proposed network to that from the V-Net architecture, which has been widely used for 3D image segmentation. We trained V-Net on the same dataset used for the proposed network. Three V-Net networks were trained independently for lung, liver, and liver tumor segmentation. Due to memory constraints, the size of the image was reduced to $128 \times 128 \times 128$. Random rotation ($\pm$ 10 degrees) and translation (5-10 voxels in each direction) were applied for data augmentation. For each epoch, 240 augmented training images (from 30 original training images) were processed with a batch size of 2. The training stop criteria were the same as for the proposed network. In post-processing, the segmentations were resampled to the original size. Performances of trained V-Net networks were evaluated on the test set, and the performances were compared with the proposed networks.

We further compared the proposed segmentation networks with other methods commonly used in medical image segmentation. For lung segmentation, we performed additional lung segmentation on the low-dose CT images using the seeded region-growing method (SRG).[38] For tumor segmentation, we perform additional tumor segmentation on ^{99m}Tc-MAA SPECT images using a thresholding-based method. A fixed-threshold value of 7% of the maximum voxel value was calculated from the training set and then applied to the test set. The SRG and thresholding-based segmentation results for 20 images were assessed relative to the reference segmentations. The Wilcoxon signed-rank test was used to evaluate the difference in the segmentation accuracy, in terms of DSC, surface DSC, and ASSD, among different segmentation algorithms.

2.4.3 | Architecture analysis

To quantify the contribution of each element of the proposed architecture, we conducted a series of experiments, where we altered elements underlying the architecture: multi-scale structure, dense connectivity, skip connection, and deconvolution for upsampling. Evaluating the network structure is challenging because network properties (i.e., number of channels, number of layers) interact with each other, and it is thus not feasible

to manipulate them independently while keeping all the others constant. Instead, we evaluated these properties together by comparing our four-stage network (four pairs of the convolutional unit and the dense block) to two alternatives with three and two pairs of the convolutional unit and the dense block (*ThreeStage* and *TwoStage*)

To evaluate the dense connectivity, we compare the proposed network to two networks without dense connection: *NoDenseL*, a network replacing the dense block with a standard convolutional unit with the same number of channels as the proposed network, but a smaller number of trainable parameters (due to connectivity); and *NoDenseH*, a network with standard convolutional unit having more channels to match the parameter count of the proposed network.

To evaluate the skip connection, intermediate fully convolution, and deconvolution, we compared our network to four alternative networks: *NoSkip*, a network without skip connections; *NoFC*, a network without a fully-connected convolutional layer between the encoding and decoding paths; *TriUp*, a network replacing the deconvolutional unit in the decoding path with trilinear upsampling; and *TriUp+Conv*, a network replacing the deconvolutional unit with trilinear upsampling and a standard convolutional unit to match the parameter count of the proposed network.

2.4.4 | Visual evaluation

Two nuclear medicine physicians (with seven and three years of ^{90}Y SIRT planning clinical experience) visually compared the VOIs of lung, liver, and tumor obtained from the proposed networks against the reference segmentation. The physicians were presented with corresponding VOIs from the two methods superimposed on SPECT, CT, and fused SPECT/CT images. The physicians were not revealed which method was used to generate the VOIs, and the VOIs were randomly ordered from one case to the next. The physicians were allowed to adjust slice and image orientation, as well as contrast and brightness of SPECT and CT images. For each pair of VOIs, the physicians selected their preferred VOIs among the two methods or selected both methods if the difference between methods was negligible.

3 | RESULTS

3.1 | Network evaluation

Figure 2 shows the segmentation results of the three segmentation networks from one subject in the test set. The medians of the segmentation metrics for each region evaluated on the test set (20 subjects) are

reported in Table 2. We further evaluated segmentation performance by comparing the differences in the volume, count, and LSF between the proposed and reference segmentation results over the 20 datasets. The median [interquartile range] of volume differences were 2.8% [2-4], 3.1% [-1-14], and 27% [10-78] for lung, liver, and tumor, respectively. The median [IQR] of count differences were 7.6% [2-15], 0.6% [-1-2], and 10% [4-27] for lung, liver, and tumor, respectively. We can see that the volumes and the counts determined by the three segmentation networks tended to be systematically higher than the volumes and the counts determined by the reference segmentation. For LSF comparison, the median [IQR] of relative difference (%) between the LSF quantified by the proposed method and the reference segmentation was 8% [1-16]. A low percentage difference between the estimated volumes, counts, and LSFs calculated from the two segmentation methods indicates the accuracy. It suggests that the liver and lung segmentation could be used clinically. Unfortunately, a high percentage difference and wide interquartile range indicate less precision in volume and count measured by the tumor segmentation network.

The segmentation results were visually reviewed, and we found the lowest accuracy for liver segmentation in patients where the CT images had uncommon features such as ascites, embolization coils, lipiodol deposition, or streak artifacts (Figure 3). Among the three tissues studied, tumor segmentation was the most challenging. The poorest tumor segmentations were found in patients with small tumors, low tumor-to-normal liver uptake (T/N) ratio, infiltrative/multiple tumors with heterogeneous uptake, residual tumor after transarterial chemoembolization (TACE) or radiofrequency ablation (RFA), or extra-hepatic activity adjacent to liver (Figure 4).

3.2 | Segmentation algorithm comparison

For algorithm comparison, the medians of the segmentation metrics for each region were evaluated on the test set (20 samples) and are reported in Table 2. For lung segmentation, the LungNet achieved better performance metrics (higher DSC, higher surface DSC, and lower ASSD) than V-Net and SRG methods. All performance metric differences were statistically significant. For liver segmentation, the LiverNet achieved better performance metric values than the V-Net network trained on the same training data. DSC and surface DSC between the LiverNet and the V-Net algorithms were statistically significant, but the ASSD difference was not statistically significant ($p = 0.2305$). For tumor segmentation, the TumorNet achieved better performance metric values than the V-Net, with all comparisons
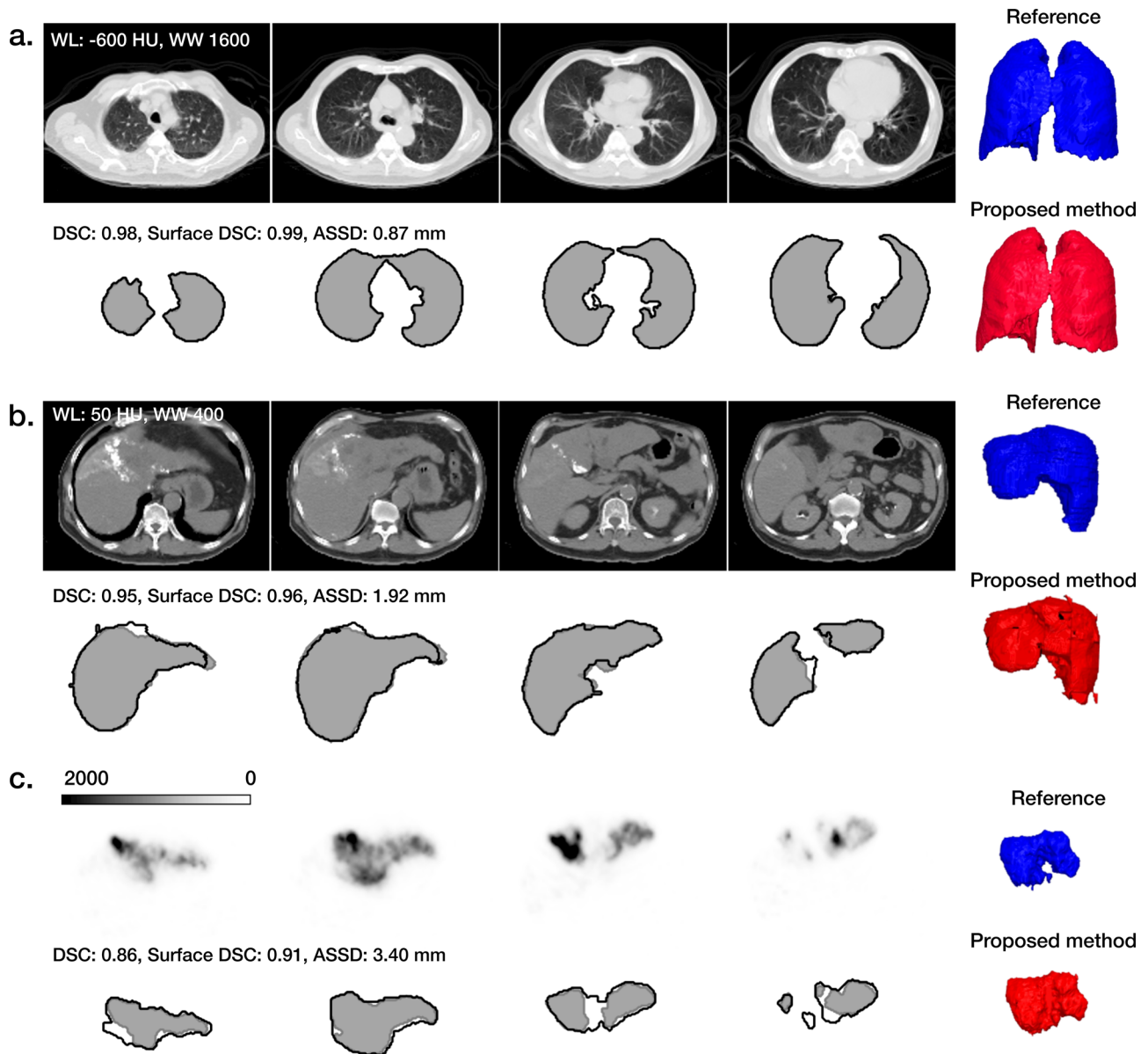
FIGURE 2 Example results from the (a) lung, (b) liver, and (c) liver tumor segmentation. The reference segmentations are shown on the second row of each sub-figure; the solid lines represent the results from the proposed segmentation algorithms, and shaded gray are voxels belonging to the reference segmentation. Columns 1 to 4 present four different slices of each label. Column 5 shows three-dimensional surface renderings of each label generated by the reference segmentation and the proposed segmentation algorithm

TABLE 2 Median [interquartile range] segmentation metrics for comparison of segmentation algorithm

Region	Algorithm	DSC	Surface DSC	ASSD (mm)
Lung	Proposed	0.98 [0.97-0.98]	0.99 [0.97-0.99]	0.9 [0.7-1.5]
	V-Net	0.87 [0.85-0.91]*	0.81[0.76-0.87]*	7.5 [5.2-9.8]*
	SRG	0.96 [0.80-0.97]*	0.94 [0.78-0.97]*	1.6 [1.2-9.5]*
Liver	Proposed	0.91 [0.83-0.93]	0.86 [0.77-0.93]	4.8 [2.6-8.4]
	V-Net	0.84 [0.81-0.85]*	0.71 [0.66-0.74]*	6.5 [5.8-7.2]
Tumor	Proposed	0.85 [0.71-0.88]	0.85 [0.62-0.93]	4.7 [3.5-9.2]
	V-Net	0.77 [0.61-0.84]*	0.56 [0.37-0.80]*	7.7 [5.3-10]*
	Thresholding	0.77 [0.65-0.87]	0.74 [0.68-0.91]	6.5 [3.4-9.3]

Abbreviations: DSC, dice similarity coefficient; ASSD, average symmetric surface distance; SRG, seeded region growing.
*Statistically significant differences in median of the metric compared to the proposed algorithm with a p-value < 0.05.
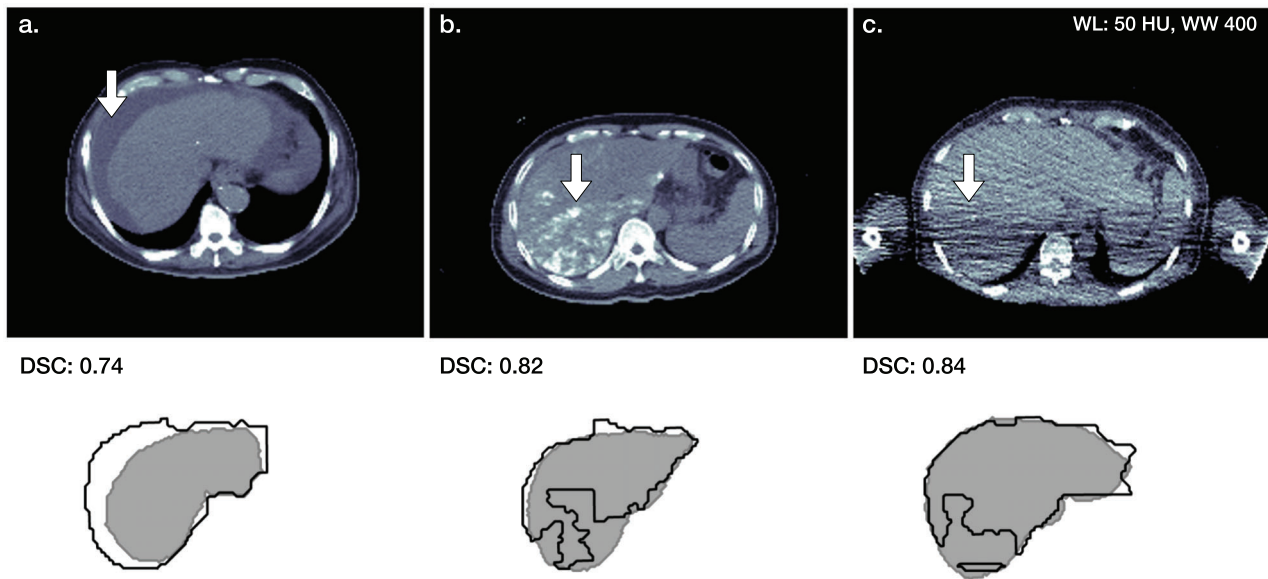
FIGURE 3 Examples of the cases where the liver segmentation by the network was less successful. The reference segmentations are shown on the second row of each sub-figure; the solid lines represent the results from the proposed segmentation algorithms, and shaded gray are voxels belonging to the reference segmentation. The arrows indicate the sites of segmentation errors due to uncommon features in the CT images, including (a) ascites, (b) lipiodol deposition, and (c) a streak artifact likely due to count starvation
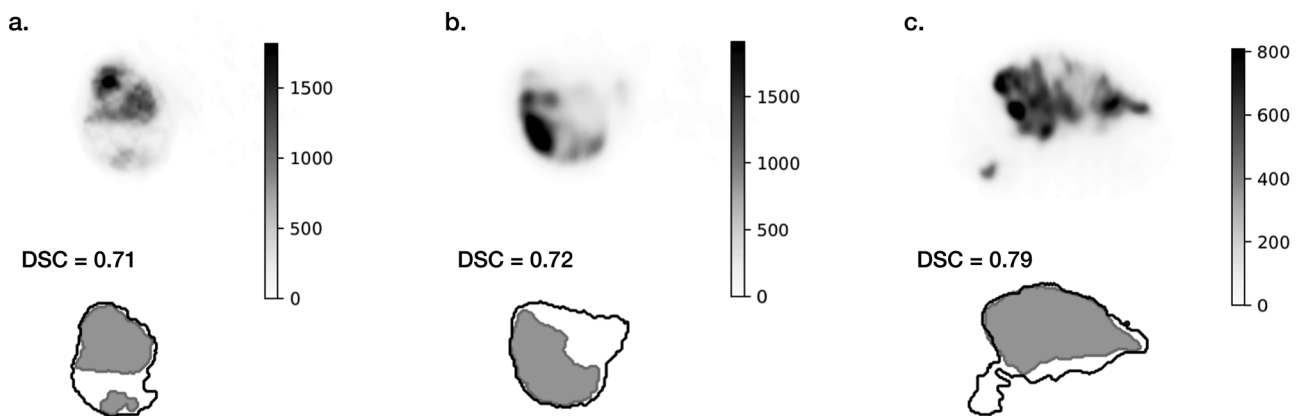

FIGURE 4 Examples of the cases where the tumor segmentation by the network was less successful, for example, (a) infiltrative/multiple tumors with heterogeneous uptake, and (b, c) tumors with extra-hepatic activity. The reference segmentations are shown on the bottom row of each sub-figure; the solid lines represent the results from the proposed segmentation algorithms, and shaded gray are voxels belonging to the reference segmentation

statistically significant. Although the TumorNet yielded better performance metric values than the optimized thresholding method in all metrics, the differences were not statistically significant.

3.3 | Architecture analysis

The median values for each of the segmentation evaluation metrics for each region are reported in Table 3 for evaluation of architecture features. Eliminating the dense connectivity (*NoDenseL* and *NoDenseH*)

yielded a statistically significant reduction in accuracy for liver segmentation. The accuracy difference between the proposed network and *NoDenseH* suggests that this improvement is not due to the number of trainable parameters, but rather the dense connectivity. The accuracy difference when eliminating skip connections and fully-connected convolutional layer (between the encoding and decoding paths) were not statistically significant for any metrics, except for liver segmentation where removing the skip connection yielded a statistically significant decrease in surface DSC.

TABLE 3 Median [interquartile range] segmentation metrics for evaluation of architecture features

Network structure	Trainableparameter	DSC	Surface DSC	ASSD (mm)
		Lung		
ThreeStage	1.7 M	0.98 [0.97-0.98]	0.99 [0.98-0.99]	0.80[0.65-1.3]
TwoStage	0.5 M	0.98 [0.97-0.98]	0.99 [0.98-0.99]	0.78 [0.60-1.2]
NoDenseL	3.5 M	0.98 [0.97-0.98]	0.99 [0.98-0.99]	0.78 [0.68-1.4]
NoDenseH	6.5 M	0.98 [0.97-0.98]	0.99 [0.98-0.99]	0.66 [0.59-1.3]
NoFC	7.4 M	0.98 [0.97-0.98]	0.99 [0.97-0.99]	0.87 [0.67-1.4]
NoSkip	6.0 M	0.98 [0.97-0.98]	0.99 [0.98-0.99]	0.87 [0.67-1.4]
TriUp	4.8 M	0.97 [0.95-0.98]*	0.97 [0.90-0.99]*	1.40 [0.93-5.2]*
TriUp+Conv	6.6 M	0.98 [0.97-0.98]	0.99 [0.98-0.99]	0.78 [0.66-1.1]
Proposed	6.6 M	0.98 [0.97-0.98]	0.99 [0.97-0.99]	0.91 [0.69-1.5]
		Liver		
ThreeStage	1.7 M	0.90 [0.83-0.92]*	0.83 [0.77-0.89]	5.6 [3.4-8.0]
TwoStage	0.5 M	0.83 [0.80-0.89]*	0.71 [0.65-0.81]*	7.9 [5.6-11]*
NoDenseL	3.5 M	0.86 [0.81-0.89]*	0.75 [0.69-0.79]*	8.5 [6.4-11]*
NoDenseH	6.5 M	0.89 [0.82-0.92]*	0.79 [0.73-0.86]*	7.2 [4.9-8.7]*
NoFC	7.4 M	0.90 [0.87-0.92]	0.87 [0.79-0.92]	4.0 [2.9-5.7]
NoSkip	6.0 M	0.89 [0.84-0.92]	0.80 [0.79-0.89]*	5.9 [4.0-7.8]
TriUp	4.8 M	0.86 [0.81-0.90]*	0.78 [0.72-0.85]*	6.5 [5.4-10]*
TriUp+Conv	6.6 M	0.88 [0.84-0.91]*	0.80 [0.77-0.87]*	6.2 [4.0-7.8]*
Proposed	6.6 M	0.91 [0.83-0.93]	0.86 [0.77-0.93]	4.8 [2.6-8.3]
		Tumor		
ThreeStage	1.7 M	0.84 [0.71-0.89]	0.84 [0.71-0.93]	4.6 [3.5-7.3]
TwoStage	0.5 M	0.83 [0.96-0.88]	0.81 [0.60-0.91]	4.5 [3.7-9.3]
NoDenseL	3.5 M	0.86 [0.72-0.88]	0.84 [0.60-0.93]	4.6 [3.2-8.1]
NoDenseH	6.5 M	0.84 [0.69-0.87]	0.85 [0.64-0.91]	4.6 [3.8-8.1]
NoFC	7.4 M	0.83 [0.68-0.86]	0.81 [0.59-0.91]	4.7 [3.7-9.8]
NoSkip	6.0 M	0.82 [0.72-0.89]	0.85 [0.66-0.94]	4.9 [3.0-7.0]
TriUp	4.8 M	0.82 [0.67-0.87]	0.72 [0.56-0.89]	6.4 [3.8-8.9]
TriUp+Conv	6.6 M	0.84 [0.69-0.89]	0.82 [0.62-0.91]	4.6 [3.8-9.9]
Proposed	6.6 M	0.85 [0.71-0.88]	0.85 [0.62-0.93]	4.7 [3.5-9.2]

Abbreviations: DSC, dice similarity coefficient; ASSD, average symmetric surface distance.

*Statistically significant differences in median of the metric for the method compared to the results from the proposed algorithm with a p-value < 0.05.

Altering the multi-scale structure by reducing operating stage (*ThreeStage* and *TwoStage*) yielded a reduction in accuracy for liver segmentation. However, differences for the lung and tumor were not significant. Replacing a deconvolutional unit in the decoding path with trilinear upsampling (*TriUp*) yielded a statistically significant reduction in accuracy for lung and liver segmentation. Replacing a deconvolutional unit with trilinear upsampling followed by a convolutional unit (*TriUp+Conv*) also yielded a statistically significant reduction in accuracy for liver segmentation.

3.4 | Visual evaluation

The selection frequency for each of segmentation methods is shown for each region in Figure 5. In the case of liver segmentation, both physicians prefer the reference over the proposed network. This was explained later by the physicians that, apart from uncommon features described in Figure 3, the proposed network usually fails to segment small areas in the left lobe when visually compared with the reference. On the other hand, the two physicians have different preferences in lung and tumor segmentation. Physician 1 prefers the reference for lung segmentation while physician 2 thinks that the difference between the two methods was negligible in more than 50% of the selection. Physician 1 prefers the reference as it excluded lung region near diaphragms. In practice, this region is removed to avoid overestimation of LSF due to ^{99m}Tc-MAA activity from liver dome. For tumor segmentation, physician 1 prefers the reference over the proposed network, whereas physician 2
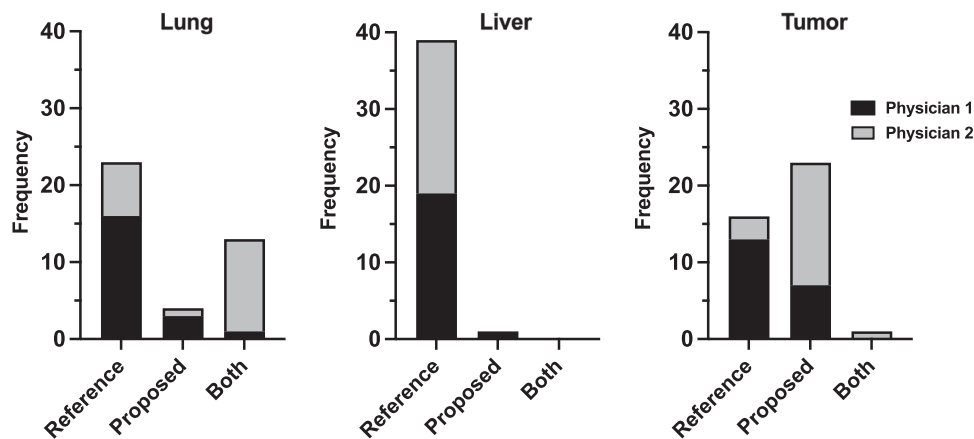
FIGURE 5 The frequency of preference for each of the segmentation methods for each region by two physicians. For each region, each physician visually compared and selected their preferred VOIs. "Reference" indicates the manually drawn VOI, whereas "Proposed" indicates the VOI defined by the proposed network, and "Both" means that the difference between the methods was negligible

prefers the proposed network over the reference. As per physician's comments, lack of clinical information (e.g., contrast-enhanced CT, angiography, or treatment approach) affected their judgements to some extent, particularly in patients with infiltrative/multiple tumors with heterogeneous uptake, or scattering post-treatment residual tumors.

4 | DISCUSSION

In this work, we proposed 3D CNN-based segmentation algorithms to segment lungs, liver, and tumor from CT and SPECT images. A comparison of the proposed network with a V-Net based network demonstrated improved accuracy for all regions. It essential to note that, in this study, both networks were trained on a limited number of training samples. The proposed network used image patches, which is a strategy for using a smaller training set. Thus, one potential explanation for the better performance of the proposed network is that it was better able to handle the smaller training set. This, in fact, highlights one benefit of patch-based strategy used in the proposed network. In the architectural analysis, we found that the multi-scale structure, dense connectivity, and deconvolution layers were all critical contributors to the segmentation accuracy of this network, especially for a more complex structure like liver. The difference in segmentation accuracy without the fully-connected convolutional layer was not statistically significant in all regions. This suggests that it could be removed entirely from the proposed network with minimal change in segmentation performance. The minor performance degradation without the skip connection suggests that higher resolution information (from the encoding path) may not be necessary (in the decoding path) to achieve the observed accuracy, possibly due to

the use of the relatively small loss in the spatial resolution of the feature maps (24^3 to 16^3) as they pass through the network.

For lung segmentation, the results show a high degree of similarity between the proposed method and the semi-automatic segmentation performed by physicians used in clinical routine. Compared to the V-Net based network and the SRG method, the CNN method demonstrated improved segmentation performance in all evaluation metrics. We also compared the lung segmentation results obtained here with results reported in the literature for threshold-based region-growing segmentation methods. The DSC (mean = 0.97) in this work was comparable with the DSCs reported by Lassen et al. (mean = 0.973),[10] Rikxoort et al. (mean = 0.962),[11] and Weinheimer et al. (mean = 0.964).[39] It is important to note that lung segmentation in this work was performed on non-contrast, low-dose CT images obtained as part of the MAA SPECT-CT study. In contrast, other lung segmentation algorithms were applied to diagnostic chest CT images from the LOLA11 challenge. Compared with another CNN-based segmentation algorithm, Xu et al.[15] reported that they used SegNet[30] architecture to segment lungs from CT image and achieved an average DSC of 0.968, similar to the proposed method (0.97).

A comparison between the results from the proposed liver segmentation and liver segmentation algorithms reviewed by Moghbel et al.[40] also shows that the DSC (mean = 0.89) from this work is comparable with those from previously published segmentation algorithms (range [0.89, 0.97]). Again, the results here were obtained using non-contrast, low-dose CT images, while other results were obtained using diagnostic non-contrast or contrast-enhanced CT images. As in this study, Rangraz et al.[41] achieved an average DSC of 0.92 on liver segmentation for ^{90}Y SIRT planning using a joint region-growing method using information from

three co-registered images (CT images from ^{99m}Tc-MAA study, CT images from ^{18}F-FDG PET study, and the ^{18}F-FDG PET). Although their method achieved higher DSC than the proposed method, it depends on the accuracy of co-registration and requires three input images instead of the single low-dose CT used in this work. For CNN-based liver segmentation, Nanda et al.[19] recently reported an average DSC of 0.9557 on liver segmentation using SegNet, which is higher than the value of 0.89 from this study. The lower DSC in this study can be attributed to the use of non-contrast, low-dose CT compared to the diagnostic quality, contrast-enhanced abdominal CT used in their study.

These results show that the proposed lung and liver segmentation algorithms were comparable with reference segmentations by physicians and previously published segmentation algorithms.[10,11,39–41] The relatively low difference between the segmentation results from the proposed CNNs and reference segmentations, in terms of volume, counts, and lung shunt fraction, also show the clinical potential for assisting ^{90}Y SIRT planning.

Despite being inferior to the lung and liver segmentation algorithms, a comparison between the liver tumor segmentation results and other liver tumor segmentation algorithms shows that the DSC (mean = 0.79) was comparable with the DSCs reported in previously published segmentation algorithms (range [0.74, 0.83]).[40] The percentage difference in VOI volume (mean = 11.88%) is also comparable with the results reported in other work (range [4.02%, 30.65%]).[40] Nanda et al.[19] reported using CNN-based algorithms to segment liver tumors on contrast-enhanced CT and achieved an average DSC of 0.6976, which is lower than 0.79 in this study. It should be noted that the proposed liver tumor segmentation network was performed on ^{99m}Tc-MAA SPECT images, while other segmentations were performed on either diagnostic CT or contrast-enhanced CT images. The proposed tumor segmentation network also achieved slightly better segmentation performance than the thresholding-based method, which used the pre-calculated threshold value based on preliminary information from the reference segmentation.

Although the liver tumor segmentation network offered comparable accuracy in terms of DSC, the high percentage difference in volume and sizeable interquartile range in volume difference suggests that there is probably large degree of variability in tumor volume estimated with the proposed network. It should be noted that the proposed liver tumor segmentation was performed on SPECT images without additional information (e.g., medical history, angiography, contrast-enhanced CT, or MR images) required for segmenting liver tumor in current clinical practice. Nevertheless, the proposed network might be beneficial in the patient with a high T/N ratio and no post-treatment residual tumor. In general,

it could provide better initial tumor segmentation than the currently-used threshold, thus reducing the time and effort needed to refine the tumor segmentation manually.

The proposed method takes approximately 1 min to segment the lung, liver, and tumor for a single ^{99m}Tc-MAA SPECT/CT dataset with an image size of 256 × 256 × 256. Since the three segmentation networks are independent of each other, they could be applied to other clinical studies that require lung or liver segmentation from low-dose CT images.

5 | CONCLUSIONS

We developed and evaluated three CNN-based algorithms for automated segmentation of lungs, whole liver, and liver tumor for ^{90}Y SIRT planning. The CNNs for lung, liver, and liver tumor segmentation shared the same architecture, and were trained on ^{99m}Tc-MAA SPECT/CT datasets. The results showed that the three segmentation networks provided promising segmentation results compared to reference segmentations from expert human observers and performed similarly or better compared to currently used methods and published results of other methods despite operating on non-contrast, low-dose CT images. The proposed algorithms have the potential to support automated segmentation in ^{90}Y SIRT treatment planning.

CONFLICT OF INTEREST
The authors have no relevant conflicts of interest to disclose.

DATA AVAILABILITY STATEMENT
Research data are not shared.

REFERENCES
1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018; 68(6): 394-424.
2. Salem R, Thurston KG, Carr BI, Goin JE, Geschwind JFH. Yttrium-90 microspheres: radiation therapy for unresectable liver cancer. *J Vasc Interv Radiol.* 2002; 13(9 pt 2): 223-229.

3. El-Serag HB, Rudolph KL. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology*. 2007; 132(7): 2557-2576.

4. Tong AKT, Kao YH, Too CW, Chin KFW, Ng DCE, Chow PKH. Yttrium-90 hepatic radioembolization: clinical review and current techniques in interventional radiology and personalized dosimetry. *Br J Radiol*. 2016; 89(1062): 20150943.

5. Biederman DM, Titano JJ, Bishay VL, et al. Radiation segmentectomy versus TACE combined with microwave ablation for unresectable solitary hepatocellular carcinoma up to 3 cm: a propensity score matching study. *Radiology*. 2017; 283(3): 895-905.

6. Biederman DM, Titano JJ, Korff RA, et al. Radiation segmentectomy versus selective chemoembolization in the treatment of early-stage hepatocellular carcinoma. *J Vasc Interv Radiol*. 2018; 29(1): 30-37. e2.

7. Ho S, Lau WY, Leung TWT, et al. Partition model for estimating radiation doses from yttrium-90 microspheres in treating hepatic tumours. *Eur J Nucl Med*. 1996; 23(8): 947-952.

8. Lau WY, Kennedy AS, Kim YH, et al. Patient selection and activity planning guide for selective internal radiotherapy with yttrium-90 resin microspheres. *Int J Radiat Oncol Biol Phys*. 2012; 82(1): 401-407.

9. Kao YH, Hock Tan AE, Burgmans MC, et al. Image-guided personalized predictive dosimetry by artery-specific SPECT/CT partition modeling for safe and effective 90Y radioembolization. *J Nucl Med*. 2012; 53(4): 59-566.

10. Lassen B, Kuhnigk J-M, Schmidt M, Krass S, Peitgen H-O. Lung and lung lobe segmentation methods at Fraunhofer MEVIS. In: *The Fourth International Workshop on Pulmonary Image Analysis*, 2011 September 18; Toronto, Canada: CreateSpace; 2011. Accessed April 2, 2020. 185-199. http://www.lungworkshop.org/2011/resources/Lung-and-Lung-Lobe-Segmentation-Methods-at-Fraunhofer-MEVIS.pdf

11. Van Rikxoort E, Van Ginneken B. Automatic segmentation of the lungs and lobes from thoracic CT scans. In: *The Fourth International Workshop on Pulmonary Image Analysis*. 2011 September 18; Toronto, Canada: CreateSpace; 2011: 261-268. Accessed April 2, 2020. http://www.lungworkshop.org/2011/resources/Automatic-Segmentation-of-the-Lungs-and-Lobes-from-Thoracic-CT-Scans.pdf

12. Sun S, Mclennan G, Hoffman EA, Beichel R. Model-based segmentation of pathological lungs in volumetric CT data. In: *The Third International Workshop on Pulmonary Image Analysis*. 2010 September 20; Beijing, China: CreateSpace; 2010: 31-40. Accessed April 2, 2020. http://www.lungworkshop.org/2010/proc2010/sun.pdf

13. Sun S, Bauer C, Beichel RR. Robust active shape model based lung segmentation in CT scans. In: *The Fourth International Workshop on Pulmonary Image Analysis*. 2011 September 18; Toronto, Canada: CreateSpace; 2011: 213-223. Accessed April 2, 2020. http://www.lungworkshop.org/2011/resources/Robust-Active-Shape-Model-Based-Lung-Segmentation-in-CT-Scans.pdf

14. Montillo A. Context selective decision forests and their application to lung segmentation in CT images. In: *The Fourth International Workshop on Pulmonary Image Analysis*. 2011 September 18; Toronto, Canada: CreateSpace; 2011: 201-212. Accessed April 5, 2020. https://www.aamontillo.net/publications/Montillo-MICCAIw-2011-ContextSelectiveDecisionForests.pdf

15. Xu M, Qi S, Yue Y, et al. Segmentation of lung parenchyma in CT images using CNN trained with the clustering algorithm generated dataset. *Biomed Eng Online*. 2019; 18(1): 1-21.

16. Goryawala M, Gulec S, Bhatt R, McGoron AJ, Adjouadi M. A low-interaction automatic 3D liver segmentation method using computed tomography for selective internal radiation therapy. *Biomed Res Int*. 2014; 2014: 198015.

17. Moghbel M, Mashohor S, Mahmud R, Iqbal Bin Saripan M. Automatic liver segmentation on computed tomography using random walkers for treatment planning. *EXCLI J*. 2016; 15: 500-517.

18. Huang C, Li X, Jia F, Automatic liver segmentation using multiple prior knowledge models and free-form deformation. In: *Proceedings of the VISCERAL Challenge at ISBI*. 2014 May 1; Beijing, China: CEUR-WS.org; 2014:22-24. Accessed August 7, 2020. http://ceur-ws.org/Vol-1194/visceralISBI14-2.pdf

19. Nanda N, Kakkar P, Nagpal S. Computer-aided segmentation of liver lesions in CT scans using cascaded convolutional networks and genetically Optimised Classifier. *Arab J Sci Eng*. 2019; 44(4): 4049-4062.

20. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst*. 2012; 25: 1097-1105.

21. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell*. 2017; 39(6): 1137-1149.

22. Long J, Shelhamer E, Darrell T, Fully convolutional networks for semantic segmentation. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, 7-12 June 2015. IEEE; 2015:3431-3440. https://doi.org/10.1109/CVPR.2015.7298965

23. Milletari F, Navab N, Ahmadi S-AA, V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: *2016 Fourth International Conference on 3D Vision (3DV)*, Stanford, CA, 25-28 October 2016. IEEE; 2016:565-571. https://doi.org/10.1109/3DV.2016.79

24. Dolz J, Desrosiers C. Ben Ayed I. 3D fully convolutional networks for subcortical segmentation in MRI: a large-scale study. *Neuroimage*. 2018; 170: 456-470.

25. Gibson E, Giganti F, Hu Y, et al. Automatic multi-Organ segmentation on abdominal CT with Dense V-Networks. *IEEE Trans Med Imaging*. 2018; 37(8): 1822-1834.

26. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-net: learning dense volumetric segmentation from sparse annotation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016: 19th International Conference*. 2016 October 17-21; Athens, Greece: Springer; 2016: 9901. 424-432.

27. Ioffe S, Szegedy C, Batch normalization: accelerating deep network training by reducing internal covariate shift. In: *Proceedings of 32nd International Conference on Machine Learning, ICML 2015*. Vol 37. 2015 July 6-11; Lille, France: JMLR.org; 2015:448-456. Accessed October 7, 2020. https://arxiv.org/pdf/1502.03167.pdf

28. He K, Zhang X, Ren S & Sun J. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: 2015 IEEE International Conference on Computer Vision (ICCV), 2015 December 7-13; Santiago, Chile: IEEE; 2015:1026-1034. https://doi.org/10.1109/ICCV.2015.123

29. Huang G, Liu Z, Van Der Maaten L & Weinberger KQ. Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017 July 21-26; Honolulu, HI: IEEE; 2017:2261-2269. https://doi.org/10.1109/CVPR.2017.243

30. Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell*. 2017; 39(12): 2481-2495.

31. Goodfellow I, Bengio Y, Courville A. 6.2.2.3 Softmax Units for Multinoulli Output Distributions. *Deep Learning*. MIT Press; 2016:180-184.

32. Fidon L, Li W, Garcia-Peraza-Herrera LC, et al. Generalised Wasserstein Dice score for imbalanced multi-class segmentation using holistic convolutional networks. *In: Crimi A., Bakas S., Kuijf H., Menze B., Reyes M. (eds) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2017*. 2017 September 14; Quebec City, Canada: Springer, Charm; 2018; 10670: 64-76.

33. Kingma DP, Ba JL. Adam: a method for stochastic optimization. In: *3rd International Conference on Learning Representations, ICLR 2015*. 2015 May 7-9; San Diego, CA: Conference Track Proceedings; 2015: 1-15. Accessed October 7, 2020. https://arxiv.org/pdf/1412.6980v9.pdf

34. Dolz J, Gopinath K, Yuan J, On HLI transactions, 2018. HyperDense-Net: a hyper-densely connected CNN for multimodal image segmentation. *IEEE Trans Med Imaging*. 2018; 38(5): 1116-1126.

35. Dice LR. Measures of the amount of ecologic association between species. *Ecology*. 1945; 26(3): 297-302.

36. Nikolov S, Blackwell S, Zverovitch A, et al. *Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy*. 2018:1-44.

37. Heimann T, Van Ginneken B, Styner MA, et al. Comparison and evaluation of methods for liver segmentation from CT datasets. *IEEE Trans Med Imaging*. 2009; 28(8): 1251-1265.

38. Adams R, Bischof L. Seeded region growing. *IEEE Trans Pattern Anal Mach Intell*. 1994; 16(6): 641-647.

39. Weinheimer O, Achenbach T, Heussel CP, Düber C. Automatic lung segmentation in MDCT images. In: *Fourth International Workshop on Pulmonary Image Analysis*. 2011 September 18; Toronto, Canada: CreateSpace; 2011: 241-255. Accessed April 5, 2020. http://www.lungworkshop.org/2011/resources/Automatic-Lung-Segmentation-in-MDCT-Images.pdf

40. Moghbel M, Mashohor S, Mahmud R, Bin Saripan MI. Review of liver segmentation and computer assisted detection/diagnosis methods in computed tomography. *Artif Intell Rev*. 2018; 50(4): 497-537.

41. Jafargholi Rangraz E, Coudyzer W, Maleux G, Baete K, Deroose CM, Nuyts J. Multi-modal image analysis for semi-automatic segmentation of the total liver and liver arterial perfusion territories for radioembolization. *EJNMMI Res*. 2019; 9(1): 1-21.