HHS Public Access

Author manuscript

*Med Phys*. Author manuscript; available in PMC 2020 February 01.

Published in final edited form as:

*Med Phys*. 2019 February ; 46(2): 619–633. doi:10.1002/mp.13331.

# Simultaneous cosegmentation of tumors in PET-CT images using deep fully convolutional networks

**Zisha Zhong**,
Department of Electrical and Computer Engineering, The University of Iowa, Iowa City, IA 52242, USA Department of Radiation Oncology, University of Iowa Hospitals and Clinics, Iowa City, IA 52242, USA

**Yusung Kim**,
Department of Radiation Oncology, University of Iowa Hospitals and Clinics, Iowa City, IA 52242, USA

**Kristin Plichta**,
Department of Radiation Oncology, University of Iowa Hospitals and Clinics, Iowa City, IA 52242, USA

**Bryan G. Allen**,
Department of Radiation Oncology, University of Iowa Hospitals and Clinics, Iowa City, IA 52242, USA

**Leixin Zhou**,
Department of Electrical and Computer Engineering, The University of Iowa, Iowa City, IA 52242, USA Department of Radiation Oncology, University of Iowa Hospitals and Clinics, Iowa City, IA 52242, USA

**John Buatti**, and
Department of Radiation Oncology, University of Iowa Hospitals and Clinics, Iowa City, IA 52242, USA

**Xiaodong Wu**[a)]
Department of Electrical and Computer Engineering, The University of Iowa, Iowa City, IA 52242, USA Department of Radiation Oncology, University of Iowa Hospitals and Clinics, Iowa City, IA 52242, USA

## Abstract

**Purpose—**To investigate the use and efficiency of 3-D deep learning, fully convolutional networks (DFCN) for simultaneous tumor cosegmentation on dual-modality nonsmall cell lung cancer (NSCLC) and positron emission tomography (PET)-computed tomography (CT) images.

**Methods—**We used DFCN cosegmentation for NSCLC tumors in PET-CT images, considering both the CT and PET information. The proposed DFCN-based cosegmentation method consists of

[a)] Author to whom correspondence should be addressed. xiaodong-wu@uiowa.edu; Telephone: (319) 335-6490; Fax: 319-335-6028.

CONFLICT OF INTEREST

The authors have no relevant conflicts of interest to disclose.

two coupled three-dimensional (3D)-UNets with an encoder-decoder architecture, which can communi- cate with the other in order to share complementary information between PET and CT. The weighted average sensitivity and positive predictive values denoted as Scores, dice similarity coefficients (DSCs), and the average symmetric surface distances were used to assess the performance of the pro- posed approach on 60 pairs of PET/CTs. A Simultaneous Truth and Performance Level Estimation Algorithm (STAPLE) of 3 expert physicians' delineations were used as a reference. The proposed DFCN framework was compared to 3 graph-based cosegmentation methods.

**Results—**Strong agreement was observed when using the STAPLE references for the proposed DFCN cosegmentation on the PET-CT images. The average DSCs on CT and PET are 0.861 ± 0.037 and 0.828 ± 0.087, respectively, using DFCN, compared to 0.638 ± 0.165 and 0.643 ± 0.141, respectively, when using the graph-based cosegmentation method. The proposed DFCN cosegmentation using both PET and CT also outperforms the deep learning method using either PET or CT alone.

**Conclusions—**The proposed DFCN cosegmentation is able to outperform existing graph-based segmentation methods. The proposed DFCN cosegmentation shows promise for further integration with quantitative multimodality imaging tools in clinical trials.

## Keywords

cosegmentation; deep learning; nonsmall cell lung cancer (NSCLC); tumor contouring

Positron emission tomography and computed tomography (PET-CT) have revolutionized modern cancer therapy. Several studies have demonstrated that the estimate of tumor extent and distribution is most accurate when functional and morphological image data are combined using PET- CT.[1–3] Outcomes are reported to improve with PET-CT guided radiotherapy.[4–7] To make full use of both PET and CT image modalities, accurate tumor delineation on PET- CT images is vital for tumor staging, response prediction, treatment planning, and prognostic assessment. The current standard of care for radiation therapy target determination relies on manually contouring the CT portion of combined PET-CT image dataset often combined with a threshold method for volume definition on the PET images. The manual contouring on CT is performed visually on a slice- by-slice basis by the radiation oncologist for tumor delineation. They have very limited support from automated segmentation tools and threshold determined volumes on PET often require significant manual editing. The development of standardized and highly reproducible PET-CT segmentation techniques would be immensely valuable for clinical care and for research.[8]

Although PET-CT images are routinely used in clinic, many clinically available PET-CT segmentation algorithms only work for a single modality or work for the fused PET- CT images. A major challenge in CT segmentation is that the pathological and physiological contrast uptake cannot be distinguished; meanwhile, the pathological and physiological changes are often more differentiable using molecularly based PET radiotracers. The American Association of Physics in Medicine (AAPM) Task Group (TG) 211[9] and the MICCAI challenge[10] have published recommendations for PET image segmentations. To take advantage of the dual modality nature of PET-CT imaging, cosegmentation aims to

simultaneously compute the tumor volume defined on the CT image as well as that defined on the PET image by combining physiological information from the PET image with the anatomical information from CT image.[11,12] Substantial progress has been made in automating the tumor definition extracted from PET-CT scans.[11–28] Although these methods showed promise, there are still limitations when adapting them for clinical use. Most previous methods depend on user-defined foreground seeds belonging to the tumor.[11,12] Secondly, features modeling components in graph-based cosegmentation algorithms were designed to complement human experience or more complicated clinical priors.[12,23–25] Therefore, finding optimal parameters and features is difficult especially with the presence of lesions. These challenges have restricted clinical application. Efforts to automate PET- CT tumor segmentation are consequently needed in modern radiotherapy.

Due to inherent differences in PET and CT imaging modalities, the tumor boundary defined in PET does not always match that in CT. Therefore, simultaneously segmenting tumors in both PET and CT while admitting the (subtle) difference of the boundaries defined in the two modalities is a more reasonable approach than using fused PET-CT images where identical tumor boundaries are assumed. Figure 1 shows the tumor boundary differences between CT and PET images in a lung tumor.

In this work, we attempt to address these challenges and seek data-driven deep learning solutions for automatically delineating features directly from PET-CT scans to develop a computer- aided automatic processing tool for tumor segmentation. Deep learning is able to outperform most conventional approaches[29] and is able to manage many medical tasks.[30–38]

Several publications detail the potential power of this approach.[39,40] In this paper, we focus on investigating 3D-UNet, deep-fully convolutional networks (DFCN) for tumor delineation in PET-CT scans. The 3D-UNet for semantic segmentation[31,38,41] performs voxel-wise classification and was adopted to label each voxel as lesion or background. To achieve PET-CT tumor cosegmentation, we propose a novel DFCN network which integrates two coupled 3D-UNets within an encoder-decoder architecture. One 3D-UNet performs the PET tumor segmentation and the other is used for performing CT tumor segmentation. The two U-Nets communicate with each other to allow the complementary features from both modalities to "flow" between the two U-Net networks to produce more consistent tumor contours. To demonstrate the applicability and performance of our method, we evaluate the proposed segmentation approach on PET-CT scans of nonsmall cell lung cancer (NSCLC) patients and compare its results to manual segmentation, which is the standard of care for NSCLC segmentation in PET-CT volumes.

## 1. METHODS AND MATERIALS

### 1.A. Image data

A total of 60 NSCLC patients who received stereotactic body radiation therapy (SBRT) were analyzed in this study following institutional review board (IRB) approval. All patients had PET-CT images for simulation and received follow-up CT images between 2 and 4 months after radiotherapy treatment. Fluorine 18-fluorodeoxyglucose (18F-FDG) PET and CT images were obtained using a dual PET/CT scanner (Siemens Biograph 40, Siemens

Medical Solutions, Erlangen, Germany). All patients were injected with 370 BMq ± 10% of 18F-FDG with an uptake time of 90 min ± 10%. In all cases, subjects fasted for more than 4 h and had a blood glucose of less than 200 mg/dl. The gross tumor volume (GTV) for each of the PET and CT image datasets was separately delineated by three radiation oncologists on both CT and 18F-FDG PET images, with the guidance of the corresponding images in the other modality. All contouring was completed using VelocityAI (Varian Medical System, Inc., Palo Alto, CA). In this study, while physicians referred to the other modality to define the tumor contours on either PET or CT, they did not visualize the corresponding PET and CT scans at the same time using the software's fusion feature. The reference standard for each scan was then generated by applying the STAPLE algorithm[42] to the three manual delineations.

## 1.B. Methods

**1.B.1. Data preprocessing—**In our experiments, we first resampled each pair of the registered PET-CT scans with an isotropic spacing in all 3D, and then cropped a fixed size of each 3D volume (that is, 96 9 96 9 48 voxels) centered at each lesion. In order to remove unrelated image details, we took a similar intensity thresholding strategy as that used by Zhong et al.[28] For the CT images, we truncated the intensity values of all scans to the range of [−500, 200], and for the PET images, we truncated the SUV of all scans to the range of [0.01, 20.0].

**1.B.2. DFCN-based cosegmentation—**The proposed DFCN-based cosegmentation framework (briefly called DFCN-CoSeg) consists of two coupled 3D- UNets with an encoder-decoder architecture, as illustrated in Fig. 2.

Each 3D-UNet is used to handle the tumor segmentation in PET or in CT, and both are communicating with each other to share the complementary features from the other modality.

In this work, we mainly employ the 3D version of the original two-dimensional (2D) U-Net with encoder-decoder architecture,[38] for single-modality segmentation, (CT-only or PET-only). This consists of a number of down-sampling (encoder) and up-sampling (decoder) modules,[38] as depicted in Fig. 2 and Table I. Given an input image cube with a size of 96 9 96 9 48, the first convolutional layer which produces 32 features maps is mainly adopted to extract the low level features. Based on these feature maps, a U-type network architecture is formed, in which the encoder module contains four convolutional and max-pooling (for down-sampling) layers with 64, 128, 256, and 512 feature maps, respectively; and the decoder module contains four de-convolutional (for up-sampling) and convolutional layers with 256, 128, 64, 32 feature maps respectively. For each convolutional layer, the size of all convolutional kernels is 3 9 3 9 3; while for all max-pooling layers, the pooling size is 2 9 2 9 2 with a stride of 2. In all deconvolutional layers, we up-sample the input features maps by a factor of 2. Using a technique similar to that described by Cicek et al.[31], the feature maps are concatenated after deconvolution with those corresponding features in the prior encoder module. More specifically, using the CT data as the input, the first convolutional layer produces 32 feature maps (denoted by F1), the encoder 1 produces 64 feature maps (denoted

as F2), and so on and so forth. Then, in the corresponding decoder module, we concatenate them to ensure maximizing information flow between layers, which helps improve the gradient flow computation and the network training.

After the decoder, a soft- max classifier implemented by the fully convolutional operation is used to generate voxel-level probability maps and the final predictions. After that, the probabilities and predictions were given as inputs to the loss functions. In this work, two kinds of well-known loss functions were studied:

1. The cross-entropy loss function (denoted as CELoss), which is defined as:

$$L_{CE} = -\frac{1}{N}\sum_{i-1}^{N}\sum_{c=1}^{2} y_i^c \, log p_i^c$$

Where $p_i^c$ denotes the probability of each voxel $i$ belongs to class $c$ (i.e., nontumor or tumor), $y_i^c$ indicates the ground truth label for voxel $i$.

2. The dice coefficient loss function (denoted as DICE-Loss), which is defined as:

$$L_{DICE} = -\frac{1}{N}\sum_{i-1}^{N}\left(1 - 2\frac{|A \cap B|}{|A| + |B|}\right)$$

where A is the reference standard volume (ground truth), and B is the predicted volume.

To facilitate the complementary feature flow between the two 3D-UNets to achieve PET/CT tumor co segmentation, we propose a feature-level fusion scheme, which takes advantage of skipping connections between layers. Figure 2 shows the organization of the proposed feature fusion scheme. The network has two parallel 3D-UNets, one for CT and the other for PET, which share the same network architecture as previously described. For feature fusion, we concatenate the feature maps from the corresponding encoders from both CT and PET branches using either the CT or the PET decoder in each branch. In this way, the decoder module can incorporate the complementary features each modality extracted in their respective encoder modules, which maximizes the information flow between either short or long range connections while preserving the various low-mid-high semantic scales of the levels of the layers. In contrast with the single 3D-UNet, the DFCN-CoSeg network simultaneously generates two tumor label predictions: one for CT and the other for PET. During the process of net- work training, the DFCN-CoSeg loss function is the sum of the two separate losses for CT and PET.

The 3D-UNets were implemented using the open source TensorFlow-GPU[43] package. All networks ran on NVIDIA GeForce GTX 1080 Ti GPU with 11GB of memory. The 3D-UNets were trained by the Adam optimization method with a mini-batch size of 4 and for 21 epochs. For the proposed DFCN-CoSeg network, the mini-batch size is set to two due to the GPU memory limit. The learning rate was initialized as 10–4 and half-decreased according

to a piecewise linear scheme. With regard to weight initialization, we adopted the truncated normal distribution with zero mean and a standard deviation of 0.01. To avoid overfitting, the weight decay was adopted to obtain the best performance on the test set.

**1.B.3. DFCN-CoSeg network training—**Of the total 60 pairs of PET-CT scans, 38 pairs were used as the training set, and the remaining 22 pairs were used as the test set. The selection procedure began by sorting all scan pairs according to the size of the tumor volume. Then three scan pairs out of every five pairs in that order were selected to be in the training set. The test set consisted of the remaining 22 scans. This stratified strategy ensured that the training set is representative of the whole dataset in terms of tumor volume. All parameters were tuned on the training set. All reported results were obtained on the test set. Among the 22 cases of the test set, 10 cases were selected as the validation set to observe the learning curves for single- modality 3D-UNets and DFCN-CoSeg. Based on the learning curves, their best models on the validation set were determined and used to evaluate their performance on the whole test set (22 cases). None of the 22 cases in the test set were employed for training the networks.

Due to the extremely limited annotated data, several data augmentation methods were adopted to enhance the training set. For each pair of coregistered PET-CT scans in the training set, a number of rigid translation, rotation and flip operations were performed to obtain additional training datasets. To perform rigid translation, the corresponding region-of-interest (ROI) bounding boxes were cropped by shifting the gravity mass center in a fixed voxel range in (5, 10, 15, 20). This shift occurred along the eight combinations of the three axis directions. For each original ROI this resulted in 32 translated ROIs. The rotation and flip operations were extended to those ROIs to further enlarge the training set. For the rigid rotation operation, each ROI image was rotated 90°, 180°, and 270° counter-clockwise around the z axis (slice-axis) to generate new ROI images. In addition, each ROI image was flipped horizontally and vertically to generate new ones. In this augmentation process, the duplicated ROIs were removed.

See Fig. 3 for detailed some examples of these data augmentations.

### 1.C. Compared methods

We conducted quantitative comparisons for three semiautomatic graph-based cosegmentation approaches: (a) the graph-based PET-CT cosegmentation method,[12] (b) the random-walker-based cosegmentation method[23] and (c) the matting-based cosegmentation method.[28] For these semiautomatic methods, a few manual seeds should be defined on the tumor. In our experiment, the same initialization procedure was employed as described by Song et al.[12] The user needs to specify two concentric spheres with different radii to serve as object and background seeds. All voxels inside the smaller sphere were used as the object seeds. All voxels outside the larger sphere were used as background seeds. For the hyper-para- meters, a grid search strategy based on a training set was used. The proposed DFCN-CoSeg method was compared to the deep learning-based method, where 3D-UNet was directly applied to segment the tumor boundaries using either PET or CT.

According to AAPM TG2116 recommendations,[44] the segmentation performance was evaluated using three criteria:

1. Accuracy score (Score), which is defined as the weighted average sensitivity (SE) and positive predictive value (PPV). Following the characterization of Hatt et al.[20], we computed the accuracy Score as Score = 0.5 9 SE + 0.5 9 PPV.
2. Dice similarity coefficient (DSC), which measures the volume overlap of two segmentations, A and B. It is defined as 2|A ∩ B|/(|A| + |B|), with a range of [0,1]. The higher the DSC is, the better volume overlap the two segmentations have.
3. Average symmetric surface distance (ASSD), which is defined as:

$$ASSD = \frac{\sum_{a \in A} min_{b \in B} d(a,b) + \sum_{b \in B} min_{a \in A} d(a,b)}{|A| + |B|}$$

where $A$ denotes the boundary surface of the reference standard and $B$ denotes the computed surface; $a$ and $b$ are mesh points on the reference surface and the computed surface respectively. $d(a, b)$ Represents the distance between $a$ and $b$

$|A|$ and $|B|$ are the number of points on $A$ and $B$ respectively. The lower the ASSD is, the better volume overlap the two segmentations have.

Statistical significance of the observed differences was determined using a 2-tailed paired $t$-test for which a $P$ value of 0.05 was considered significant.

## 3. RESULTS

Table II shows the mean values and standard deviations for the three performance metrics (Score, DSC and ASSD) for the evaluated methods on the test scans (22 cases). In addition to the cosegmentation results, we include three reference methods (Song et al.[12], Ju et al.[23], and Zhong et al.[28]) using either CT or PET. Compared to the three reference methods, the pro- posed DFCN-CoSeg approach (with the best Scores of 0.865 ± 0.034 on CT and 0.853 ± 0.063 on PET) achieves significantly better results using either the DSCs or ASSDs as metrics. This demonstrates that the trained deep learning net- work can learn to be more descriptive and better discriminate between features than the traditional manual methods. When incorporating dual image modality information (PET-CT) into the segmentation process, segmentation accuracy using either deep learning based or other graph-cut based methods can consistently improve the results of both CT and PET scans than a single modality is able to achieve alone. This suggests that the dual image modality information (PET-CT) facilitates simultaneous cosegmentation. Considering the different loss functions for the deep learning based methods (3D-UNets or DFCN-CoSeg), the models trained with dice coefficient loss performed better than those using cross-entropy based models. For example, the deep learning based 3D-UNet method with the dice coefficient loss, achieved much higher DSCs on both CT and PET (0.811 ± 0.151 over 0.638 ± 0.165, and 0.794

± 0.111 over 0.643 ± 0.141, respectively), compared to either method used by Song et al.[12], Ju et al.[23], or Zhong et al.[28]

The first MICCAI challenge for PET tumor segmentation (Hatt et al.[10]) reported 0.642 to 0.810 DSC values from 15 different segmentation methods (see Table I, Hatt et al.[10]). The DSC values of the DFCN-CoSeg approach with results of over 80% are comparable to and even competitive with the results of the MICCAI challenge. The DFCN-CoSeg method achieved the best overall Score on both CT and PET (0.865 ± 0.034 and 0.853 ± 0.063, respectively) and revealed much smaller standard deviations than any other of the compared methods. Although the average Scores, DSCs or ASSDs from the DFCN-CoSeg are higher than those obtained when using the single-modality 3D-UNets, the statistical comparison did not show a significant improvement on either CT or PET. However, the DFCN-CoSeg standard deviation is much smaller than those from the single-modality 3D-UNets (0.037 over 0.151 on CT, 0.087 over 0.111 on PET). This suggests that while the 3D-UNets can effectively learn to discriminate features to recognize tumor voxels, the simple multimodality feature fusion can only provide limited improvement given the high segmentation accuracy on the PET image. We surmise that although the features learned by the network from PET images are able to localize the overall tumor position, they may not provide more useful information on the true tumor boundary. This provides an advantage over the previous single modality graph-cut based semiautomatic methods that promise to diminish physician work load and consequently facilitate the widespread use of PET-CT images.

Figure 4 illustrates the segmentation results of the com- pared methods on four PET-CT scan pairs. The proposed DFCN-CoSeg method can obtain more consistent results against the STAPLE-generated ground truth. The example results demonstrate that the proposed DFCN-CoSeg method is able to locate tumor boundaries on PET images more accurately than the other compared methods.

The intermodality consistency between the STAPLE based ground truth on each image modality and the manual con- tours of three physicians were tested and summarized in Table III. The DCS value of all manual contours on CT images over their ground truth of STAPLE was 0.867, while that on PET images was 0.875. No significant inter image modality variations were found between the STAPLE ground truth and all corresponding manual contours ($P = 0.423$). In Table III, the three physicians are denoted as P1, P2, and P3. Their respective manual contours on CT (PET) are denoted as P1_CT, P2_CT, and P3_CT (P1_PET, P2_PET, and P3_PET). The STAPLE_CT (STAPLE_PET) is obtained from the STAPLE algorithm with P1_CT (P1_PET), P2_CT (P2_PET) and P3_CT (P3_PET) as inputs.

## 4. DISCUSSION

### 4.A. Performance

Accurate tumor delineation in image-guided radiotherapy, is critically important yet efforts to automate the process radiotherapy treatment planning or delivery remain elusive.[19] In this work, our DFCN-CoSeg approach for PET-CT offers improved automation, requires no

direct interaction and enables efficient computer-aided segmentation which may facilitate eventual clinical use. Our framework takes advantage of the assumption that the combination of information derived from dual image modality (PET-CT) would vastly improve the capability of an automated, learning-based segmentation approach. Our goal is to automate the manual delineation done by radiation oncologists to define tumor by contouring PET-CT images so that the radiation oncologists can review and modify the automated tumor efficiently. Owing to the successful adoption of deep fully convolutional neural network architecture, the automated feature extraction is conducted on both dual-modal- ity PET-CT images and able to discriminate between tumors or nontumors. The proposed DFCN-CoSeg was evaluated on 60 NSCLC cases and the promising experimental results have demonstrated the efficiency over previous graph-based PET-CT cosegmentation methods.

Feature fusion between medical images acquired using distinct modalities is a challenging problem due to the inter- image variability inherent in each type. These include: different noise, various image resolutions, different contrast, or misregistration of the images to mention a few.[45,46] Traditional methods include morphology-based fusion,[47] wavelet- based fusion,[48] component analysis based fusion,[49] and hybrid fusion.[50] Different feature fusion approaches using deep neural networks are discussed:

1. The first and most basic method is to combine the input images/features and process them jointly in a single UNET. This describes most methods in the literature for handling multi-image modality datasets such as the fusion of multiparametric brain MR images of T1- and T2- weighted MRI for brain tumor segmentation.[34,37]
2. The second approach is to conduct feature fusion on images of different resolutions[34] through two steps of extracting different sizes of input patches in the input images and giving them as inputs of different networks to obtain the different feature levels to conduct the feature fusion.
3. The last method is to conduct feature fusion based on deep convolutional and recurrent neural networks (RNN), where the RNNs are responsible for exploiting the intraslice and interslice contexts respectively.[51]

The novelty of our approach lies in the investigation of the encoder-decoder based 3D-UNet for the cosegmentation in dual-modality PET-CT images. Our contribution is to con- sider the difference between each segmentation on either the PET or the CT image and to design a coupled feature fusion network based on the 3D-UNet architecture. This allows us to simultaneously produce high quality, voxel-wise segmentation for tumors in PET-CT images and specifically to cosegment tumors in PET-CT images. A novel DFCN network was proposed where two coupled 3D-UNets with an encoder-decoder architecture were integrated. One 3D-UNet is used to perform the PET image tumor segmentation and the other to perform the CT image segmentation. The two UNets communicate with each other to allow the complementary features from both modalities to “flow” between the two 3D-UNet networks to produce consistent tumor contours.

Figure 5 shows the DSC training curves on single-modality 3D-UNets trained on either CT- or PET- only data, and the training curves of DFCN-CoSeg on both PET and CT. From the

learning behaviors in those figures we can observe that the training set accuracy is much higher than that on the validation set on the single-modality 3D-UNet, for either CT or PET. The curves show a relatively large fluctuation com- pared to those from DFCN-CoSeg.

This may be due to the lack of complementary information from the other modality and the large variation in the tumor volume sizes. By contrast, the learning curves are much smoother than that of the single-modality 3D-UNet for the dual-modality DFCN-CoSeg network. This indicates that the complementary information exchanged between CT and PET images can help learn to better discriminate features, which in turn helps locate the local optimum for network training. Second, considering the loss functions, the DICELoss has achieved smoother learning curves compared to the CELoss, which indicates that the dice loss function is more robust on data variability in the training set compared to the simple cross-entropy loss. Third, as shown in these figures, we can see the mean DSCs on the training set are consistently increasing and become steady after about 15 epochs, while the mean DSCs on the validation set did not increase. We determined the best models for these networks based on their respective mean DSCs on the validation set, then evaluated them on the 22 cases that did not participate in the real net- work training.

### 4.B. Data augmentation improves general performance

In this subsection we study if data augmentation can improve general performance in PET-CT segmentation. We again ran the same experiments described above, but without the data augmentation on the training set, that is, the training set only includes the original 38 cases. All models were trained in the same way as described above. Table IV shows the segmentation performances based on Score, DSC and ASSD on our test set (22 cases) for these models trained either with or without data augmentation. Figure 6 illustrates the accuracy curves trained without data augmentation. As we can see in the table and figures, those network models trained with data augmentation can achieve significantly better performance than those without data augmentation. Compared to Fig. 5, the learning curves in Fig. 6 also get steady more quickly (about 10 000 steps), which indicates that these models were quickly fitted on the training set without data augmentation.

### 4.C. Testing on whole-body PET-CT images

The inputs for the proposed DFCN-CoSeg network are two fixed-size bounding boxes (i.e., cropped from the original whole-body PET/CT images respectively). In our experiments we first crop the bounding boxes containing lesions according to the ground truth segmentation as the input of the networks. However, it is very important in clinical practice to directly segment the whole-body PET-CT scans simultaneously. To this end, we conducted additional experiments to test the segmentation performance of the proposed method on resampled whole-body PET-CT scans (i.e., with voxel spacing of 1 9 1 9 1 mm). Specifically, we first resampled the whole-body PET/CT scans to isotropic voxel spacing (1 9 1 9 1 mm), then based on the sliding window technique, cropped the paired PET/CT sub-volume, gave them as inputs of the 3D-UNets or DFCN-CoSeg, and generated the final results on whole-body predictions.

Table IV shows the statistics of the segmentation results on the respective whole-body PET/CT images of the test set. As shown in this table, for most cases, the SEs and PPVs of the proposed DFCN-CoSeg are overall higher than those of the 3D-UNets. The high sensitivities indicate the proposed DFCN-CoSeg can correctly detect and segment the tumors in the whole-body images. However, in some cases (e.g., A- IA002114), the specificities of the proposed DFCN-CoSeg are lower than 50%.

The main reason for the relatively poor results for the tumor segmentation is that it is located in the boundary between two adjacent sliding windows. The tumor region located in the boundary presented relatively poor predictive results when compared to the tumor located in the center region. From the viewpoint of fully convolutional networks, as our DFCN-CoSeg models were trained based on tumor- centralized PET/CT bounding boxes, they may be biased in the centralized inputs.

The automatic object localization techniques could help improve this. It is worth noting that much research has been devoted to automatically obtaining these bounding boxes with a high degree of confidence that they will contain lesions within them (e.g., lung nodule object detection[52]), and results from these modules can be used as the inputs for the proposed DFCN-CoSeg method. It is also worth observing that recent advances in computer vision have demonstrated the efficiency of simultaneously conducting object detection and segmentation in a single deep network frame- work (e.g., Mask-RCNN).[53]

### 4.D. Limitations

Although the proposed DFCN-CoSeg method has achieved some improvement over the traditional graph-based method, the absolute assessment based on DSCs is still about 82% on PET and 86% on CT. There is still much improvement needed in terms of performance, robustness, and stability. For the single-modality 3D-UNets, we observed that the DSCs on two test datasets were below 70%. It seems the tumor patterns in the two datasets were not adequately represented in the training set. We thus plan to enlarge the training set in the future.

In terms of the network architecture design, our DFCN- CoSeg network was inspired by the encoder-decoder based 3D fully convolutional networks (3D-FCN)[32,37]and the 3D- UNets.[30,31,34] As a natural extension of the well-known 2D FCN proposed by Long et al.[41], 3D-FCNs have been success- fully applied to semantic segmentation tasks in medical imaging, such as liver segmentation,[32]brain tumor segmentation,[34] and pancreas segmentation.[54] As demonstrated in these studies, the skip connections designed in 3D-FCNs or 3D-UNets were very important to help recover the full spatial resolution at the network outputs, which is suitable for voxel-wise segmentation tasks. Various typical extensions include the extended U-Net based on the DenseNet,[55] or the short skip connection.[56] In this work, we utilized a coupled skip connection between the two 3D-UNets for CT or PET, taking advantage of both modalities to produce two separate segmentations. Although our proposed method achieved good results, designing a more efficient feature fusion architecture would be very beneficial.

The proposed DFCN-CoSeg method utilizes a pixel-wise cross-entropy loss or dice coefficient loss in the last layer of its network, which is insufficient to learn both local and global contextual relations between pixels. Although the use of UNet architecture may alleviate this problem, enabling it to implicitly learn some local dependencies between pixels; it is still limited by their pixel-wise loss function. This is because it lacks the ability to enforce the learning of multi-scale spatial constraints directly in the end-to-end training process. We propose the future integration of Conditional Random Fields (CRFs) into our DFCN-CoSeg framework with an end-to-end training process to enforce the pixel-wise labeling consistency to improve the segmentation accuracy.

AAPM Task Group 211[6] recommends the use of different types of ground truth datasets including phantom and clinical contouring. The proposed DFCN-CoSeg method has been thoughtfully validated on clinical contouring. It is natural to also validate our method on simulated PET-CT images and those obtained from phantoms. However, the performance of a model with supervised learning frequently deteriorates on data from a new deployment domain, which is known as a domain shift problem.[57] The domain shift from simulated images to real data can be particularly challenging.[58,59] Our DFCN-CoSeg model, as a supervised learning method trained on physician's manual contours, may not well work on the simulated and phantom data.

In this study, all PET and CT datasets were obtained from a single PET-CT scanner (Siemens Medical Solutions, Inc.). Different PET-CT scanners have unique acquisition and reconstruction properties, for PET datasets especially: spatial reconstruction, noise properties, and voxel size. The sensitivity study of the proposed DFCN-CoSeg method for each specific PET-CT scanner needs to be investigated. In fact, this is another domain shift problem in transfer learning, which has been widely recognized in machine learning or computer vision. Pre- trained convolutional neural networks have been designed and trained on the ImageNet dataset such as AlexNet,[60] Google InceptionNet,[61] and ResNet-50.[62] These pretrained models or weights have been widely adopted and transferred into other tasks such as RGBD segmentation[63] or remote sensing image analysis[64]to initialize the target networks

### 4.E. Impact

For one unique tumor, one segmentation represents patho- logically and morphologically true tumor boundaries. How- ever, the tumor boundaries are defined according to whether they are macroscopically or microscopically identified. If two image modalities are the same type of morphological or anatomical image such as CT and MRI, then the macroscopic tumor boundaries can be similar depending on tumor sites. For instance, lung cancer tumor boundaries that presented considerable image intensity changes on both MRI and CT are similar due to the significant electron density difference between the tumor and the adjacent lung tissue. The segmentations based upon their image intensities are similar even though the physics underlying the image generation for these images is different. In the case of cervical cancer, CT and MRI tumor boundaries vary considerably due to the poor soft-tissue contrast on CT. MRI is critical for tumors requiring high soft-tis- sue contrast such as gynecologic cancer, breast cancer, or prostate cancer. Even though each tumor is unique its boundary identification is different due to the physics underlying the generation of each

medical image. This is especially true between molecular images such as PET or functional MRI (e.g., DCE (dynamic contrast enhanced) or DWI (diffusion weighted image) MRI) and anatomical images such as CT. PET images visualize the uptake region of radiotracer such as F18-FDG (Fludeoxyglucose) as a marker for the tissue uptake of glucose, indicating a close correlation with tumor metabolism.[65] The high glucose uptake in tumor cells is used as a surrogate to identify an active tumor region using PET images, thus the highly active regions identified by F18-FDG PET images do not necessarily represent every single tumor cell or anatomically defined tumor. When tumor cells are not active or highly hypoxic, they are not included in the F18-FDG PET positive regions. In addition, when other radiotracers for PET images such as F18-FMISO (fluoromisonidazole) or F18-FLT are used, then specific subtumor regions are identified based on molecularly defined criteria. F18-FMISO PET images pre- sent hypoxic tumor regions[66] while F18-FLT PET images efficiently identify proliferating tumor regions.[67] In general, the identified tumor regions using molecular imaging are considerably different from anatomical CT imaging which identifies high electron density regions.

Besides the fundamental technical differences required to generate functional or anatomical images, PET images take a longer time (typically 10–20 min for acquisition vs less than 2 min in CT acquisition) to acquire. For a tumor site affected by respiratory motion like lung cancer, the acquisition time will include the full range of motion for PET images, while the tumor region in a CT scan (often fully acquired in a breath hold) is less affected by breathing motion. As a result, the tumor boundary varies between PET and CT images. In addition, PET images have a relatively poor spatial resolution ( “‘‘ 3–5 mm), while the spatial resolution of CT images can be submillimeter (e.g., 0.6 mm). These inherent imaging features all cause challenges in identifying tumor boundaries between PET and CT images.

Simultaneously segmenting tumors from both PET and CT while admitting the difference of the boundaries defined in the imaging modalities is a more reasonable method than those previously applied to fused PET-CT images, where identical tumor boundaries were assumed.

Advances in radiomics and machine learning in PET-CT images involve the accurate, robust, and reproducible segmentation of the tumor volume in each imaging modality to extract numerous unique features from each imaging modality. These radiomics features include 3D shape descriptors, intensity- and histogram-based metrics and 2nd or higher order textural features (Hatt et al.[10]). By obtaining different radiomics features from each imaging modality, including different segmentations from each image modality, we expect that the use of radiomics and CNN would improve the efficiency of diagnosis, prognosis determination and clinical decision-making critical for therapeutic interventions such as surgery, chemotherapy, or radiotherapy.

In our study, we presented two PET-CT cosegmentation approaches; Segmentation of PET images while interacting with CT images and vice versa. Two different cosegmentations could be variously adapted in current clinical applications. The PET-based cosegmentations can be further investigated to improve current PET-based prognostic assay studies in nuclear medicine in which the metrics of PET images, such as standardized uptake value (SUV) and

metabolic tumor volume (MTV) have been extensively studied.[68] In addition, the CT-based cosegmentation potentially improves contouring accuracies and robustness for radiation therapy, especially for determining high dose regions used in approaches such as SBRT in which the target (tumor) con- touring accuracy is critical. To our knowledge, the proposed DFCN-CoSeg network is the first CNN-based, simultaneous segmentation approach for both PET and CT.

Our proposed deep learning DFCN-CoSeg cosegmentation framework can be successfully applied to the simultaneous cosegmentation of GTVs in the PET-CT image dataset. The cosegmented GTV is expected to improve the efficacy of radiotherapy. The proposed DFCN-CoSeg approach will not fully replace input and review by physicians. However, we expect this supplementary segmentation tool would act as an aid and would improve the robustness and consistency of contouring. Beichel at al.[69] compared a semiautomated segmentation vs experts contouring method, and presented the intra- and interoperator standard deviations that were significantly lower for the semi-automated segmentation. Also, con- touring capability is expected to improve the efficiency of molecular image guided radiotherapy when used in concert with a PET-CT Linac system[70] in which fast, accurate con- touring is critical for adaptive replanning on each fraction. The developed deep fully convolutional network based segmentation can also be applied to current kV-cone-beam, CT based, adaptive replanning approaches. The improved accuracy of using deep learning based cosegmentation is expected to improve its prognostic power for the therapy selection for lung cancer patients and improve clinical outcomes. In addition, the current state-of-the-art in tumor diagnosis and characterization does not perform 3D contouring due to its extended acquisition time but instead performs three, 1D tumor measurements (length, width, and height) from which tumor size and stage is determined. Before and after chemotherapy, 3D contouring (segmentation) is not clinically performed for response assessment. No contouring is per- formed following radiation therapy in order to assess response to therapy. DFCN-CoSeg could provide a metric following treatment if correctly applied. The prognostic power, of DFCN-CoSeg applied to GTV is being studied and its results will be assessed. Given that integrated PET-CT scanners are widely available, the developed deep learning based cosegmentation technique in this work is readily avail- able as a tool to be utilized in nonsmall cell lung carcinoma clinical trials. In addition, the proposed computer-aided, auto- mated tumor volume identification method using deep learning is expected to advance the accuracy of target delineation for radiation therapy planning. This will be especially important for SBRT in which high-precision tumor identification is crucial in the process of defining the spatial extent of the therapeutic radiation dose distribution. It is expected that the improved accuracy and robustness of using deep learning based cosegmentation on both PET and CT images would improve outcomes and reduce the toxicity of radiation therapy for nonsmall cell lung cancer.

## 5. CONCLUSIONS

In this work, we investigated the deep 3D fully convolutional neural networks for tumor cosegmentation on dual- modality PET-CT images. Experimental results demonstrated the effectiveness of our proposed method and improved accuracy compared to several common existing methods. Our method would benefit clinics as it does not require unique knowledge

to implement effectively. Our dual 3D UNet cosegmentation framework could be further applied to other multi-modality data, for example, PET/MR. Validation of the PET/CT segmentation on other datasets including multi-institutional trials and more different neural network architectures will also be investigated in the future.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

This work was partially supported by the grants 1R21CA209874, UL1TR002537, and U01CA140206 from the National Cancer Institute.

## REFERENCES

1. Ahn PH, Garg MK. Positron emission tomography/computed tomogra- phy for target delineation in head and neck cancers. Semin Nucl Med. 2008;38:141–148. [PubMed: 18243850]
2. Caldwell CB, Mah K, Skinner M, Danjoux CE. Can PET provide the 3D extent of tumor motion for individualized internal target volumes? A phantom study of the limitations of CT and the promise of PET. Int J Radiat Oncol Biol Phys. 2003;55:1381–1393. [PubMed: 12654451]
3. Fox JL, Rengan R, O'Meara W, et al. Does registration of PET and planning CT images decrease interobserver and intraobserver variation in delineating tumor volumes for non-small-cell lung cancer? Int J Radiat Oncol Biol Phys. 2005;62:70–75. [PubMed: 15850904]
4. Bagci U, Udupa JK, Mendhiratta N, et al. Joint segmentation of anatomical and functional images: applications in quantification of lesions from PET, PET-CT, MRI-PET, and MRI-PET-CT images. Med Image Anal. 2013;17:929–945. [PubMed: 23837967]
5. Foster B, Bagci U, Mansoor A, Xu Z, Mollura DJ. A review on segmen- tation of positron emission tomography images. Comput Biol Med. 2014;50:76–96. [PubMed: 24845019]
6. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D. Characterization of PET/CT images using texture analysis: the past, the pre- sent… any future? Eur J Nucl Med Mol Imaging. 2017;44:151–165. [PubMed: 27271051]
7. Zaidi H, El Naqa I. PET-guided delineation of radiation therapy treat- ment volumes: a survey of image segmentation techniques. Eur J Nucl Med Mol Imaging. 2010;37:2165–2187. [PubMed: 20336455]
8. Markel D, Zaidi H, El Naqa I. Novel multimodality segmentation using level sets and Jensen-R'enyi divergence. Med Phys. 2016;40:121908.
9. Hatt M, Lee JA, Schmidtlein CR, et al. Classification and evaluation strategies of auto-segmentation approaches for PET: Report of AAPM task group No. 211. Med Phys. 2017;44:e1–e42. [PubMed: 28120467]
10. Hatt M, Laurent B, Ouahabi A, et al. The first MICCAI challenge on PET tumor segmentation. Med Image Anal. 2018;44:177–195. [PubMed: 29268169]
11. Han D, Bayouth J, Song Q, et al. Globally optimal tumor segmentation in PET-CT images: a graph-based co-segmentation method. Inf Process Med Imaging. 2011;22:245–256. [PubMed: 21761661]
12. Song Q, Bai J, Han D, et al. Optimal co-segmentation of tumor in pet-ct images with context information. IEEE Trans Med Imaging. 2013;32: 1685–1697. [PubMed: 23693127]
13. Bagci U, Udupa JK, Yao J, Mollura DJ. Co-segmentation of functional and anatomical images. Med Image Comput Comput Assist Interv. 2012;7512:459–467.
14. Drever L, Roa W, McEwan A, Robinson D. Comparison of three image segmentation techniques for target volume delineation in positron emission tomography. J Appl Clin Med Phys. 2007;8:93–109. [PubMed: 17592458]

15. El Naqa I, Yang D, Apte A, et al. Concurrent multimodality image segmentation by active contours for radiotherapy treatment planning. Med Phys. 2007;34:4738–4749. [PubMed: 18196801]
16. Foster B, Bagci U, Luna B, et al. Robust segmentation and accurate tar- get definition for positron emission tomography images using affinity propagation. IEEE 10th International Symposium on Biomedical Imaging. 2013:1461–1464.
17. Geets X, Lee JA, Bol A, Lonneux M, Gr'egoire V. A gradient-based method for segmenting FDG-PET images: methodology and validation. Eur J Nucl Med Mol Imaging. 2007;34:1427–1438. [PubMed: 17431616]
18. Gribben H, Miller P, Hanna GG, Carson KJ, Hounsell AR. MAP-MRF segmentation of lung tumors in PET/CT images. IEEE International Symposium on Biomedical Imaging: From Nano to Macro. 2009:290–293.
19. Hanzouli-Ben Salah H, Lapuyade-Lahorgue J, Bert J, et al. A frame-work based on hidden Markov trees for multimodal PET/CT image co- segmentation. Med Phys. 2017;44:5835–5848. [PubMed: 28837224]
20. Hatt M, Cheze le Rest C, Descourt P, et al. Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications. Int J Radiat Oncol Biol Phys. 2010;77: 301–308. [PubMed: 20116934]
21. Hong R, Halama J, Bova D, Sethi A, Emami B. Correlation of PET standard uptake value and CT window-level thresholds for target delineation in CT-based radiation treatment planning. Int J Radiat Oncol Biol Phys. 2007;67:720–726. [PubMed: 17293230]
22. Jentzen W, Freudenberg L, Eising EG, Heinze M, Brandau W, Bockisch a. Segmentation of PET volumes by iterative image thresholding. J Nucl Med. 2007;48:108–114. [PubMed: 17204706]
23. Ju W, Xiang D, Zhang B, Wang L, Kopriva I, Chen X. Random walk and graph cut for co-segmentation of lung tumor on PET-CT images. IEEE Trans Image Process. 2015;24:5854–5867. [PubMed: 26462198]
24. Lartizien C, Rogez M, Niaf E, Ricard F. Computer-aided staging of lymphoma patients with FDG PET/CT imaging based on textural information. IEEE J Biomed Health Inform. 2014;18:946–955. [PubMed: 24081876]
25. Li H, Bai J, Abu Hejle J, Wu X, Bhatia S, Kim Y. Automated cosegmentation of tumor volume and metabolic activity using PET-CT in non- small cell lung cancer (NSCLC). Int J Radiat Oncol Biol Phys. 2013;87: S528.
26. Li H, Thorstad WL, Biehl KJ, et al. A novel PET tumor delineation method based on adaptive region-growing and dual-front active con- tours. Med Phys. 2008;35:3711–3721. [PubMed: 18777930]
27. Nehmeh SA, El-Zeftawy H, Greco C, et al. An iterative technique to segment PET lesions using a Monte Carlo based mathematical model. Med Phys. 2009;36:4803–4809. [PubMed: 19928110]
28. Zhong Z, Kim Y, Buatti J, Wu X. 3D alpha matting based co-segmenta- tion of tumors on PET-CT images In: Molecular Imaging, Reconstruc- tion and Analysis of Moving Body Organs, and Stroke Imaging and Treatment. Vol. 10555 Cham: Springer Nature; 2017:31–42.
29. Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge, MA: MIT Press; 2016.
30. Chen H, Dou Q, Yu L, Qin J, Heng P-A. VoxResNet: deep voxelwise residual networks for brain segmentation from 3D MR images. NeuroI- mage. 2018;170:446–455.
31. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-N e t: Learning dense volumetric segmentation from sparse annotation In: Medical Image Computing and Computer-Assisted Intervention. Vol. 9901 Cham: Springer Nature; 2016:424–432.
32. Dou Q, Chen H, Jin Y, Yu L, Qin J, Heng P-A. 3D Deeply supervised network for automatic liver segmentation from CT volumes In: Medical Image Computing and Computer-Assisted Intervention. Vol. 9901 Cham: Springer Nature; 2016:149–157.
33. Jifara W, Jiang F, Rho S, Cheng M, Liu S. Medical image denoising using convolutional neural network: a residual learning approach. J Supercomput. 2017;1–15
34. Kamnitsas K, Bai W, Ferrante E, et al. Ensembles of multiple models and architectures for robust brain tumour segmentation In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Vol. 10670 Cham: Springer Nature; 2018:450–462.

35. Miao S, Wang ZJ, Liao R. A CNN regression approach for real-time 2D/3D registration. IEEE Trans Med Imaging. 2016;35:1352–1363. [PubMed: 26829785]
36. Nie D, Trullo R, Lian J, et al. Medical image synthesis with deep convolutional adversarial networks In: IEEE Transactions on Biomedical Engineering. Vol. 65 New York, NY: IEEE; 2018:2720–2730. [PubMed: 29993445]
37. Nie D, Wang L, Adeli E, Lao C, Lin W, Shen D. 3-D fully convolutional networks for multimodal isointense infant brain image segmentation. IEEE Trans Cybernet. 2018:1–14.
38. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation In: Medical Image Computing and Computer-Assisted Intervention. Vol. 9351 Cham: Springer Nature; 2015:234–241.
39. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60–88. [PubMed: 28778026]
40. Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. Ann Rev Biomed Eng. 2017;19:221–248. [PubMed: 28301734]
41. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2015:3431–3440.
42. Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging. 2004;23:903–921. [PubMed: 15250643]
43. Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems; 2015.
44. Berthon B, Spezi E, Galavis P, et al. Toward a standard for the evaluation of PET-auto-segmentation methods following the recommendations of AAPM task group No. 211: requirements and implementation. Med Phys. 2017;44:4098–4111. [PubMed: 28474819]
45. James AP, Dasarathy BV. Medical image fusion: a survey of the state of the art. Inf Fus. 2014;19:4–19.
46. James AP, Dasarathy BV. A Review of feature and data fusion with medi- cal images: multisensor data fusion: from algorithm and architecture design to applications. Boca Raton, FL: CRC Press/ Balkema; 2015.
47. Marshall S, Matsopoulos GK. Morphological data fusion in medical imaging In: IEEE Winter Workshop on Nonlinear Digital Signal Processing. New York, NY: IEEE; 1993:6.1_5.1–6.1_5.6.
48. Sayed S, Jangale S. Multimodal medical image fusion using wavelet transform In: Thinkquest~2010. New Delhi: Springer; 2011:301–304.
49. He C, Liu Q, Li H, Wang H. Multimodal medical image fusion based on IHS and PCA. Proc Eng. 2010;7:280–285.
50. Krishn A, Bhateja V, Himanshi, Sahu A. PCA based medical image fusion in ridgelet domain. In: Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014:475–482.
51. Chen J, Yang L, Zhang Y, Alber MS, Chen DZ. Combining fully convolutional and recurrent neural networks for 3D biomedical image segmentation. In: Proceedings of the 30th International Conference on Neural Information Processing Systems 2016:3044–3052.
52. Ding J, Li A, Hu Z, Wang L. Accurate pulmonary nodule detection incomputed tomography images using deep convolutional neural net- works In: Medical Image Computing and Computer-Assisted Intervention. Vol. 10435 Cham: Springer Nature; 2017:559–567.
53. He K, Gkioxari G, Doll'ar P, Girshick R. Mask R-CNN. In: Proceedings of the International Conference on Computer Vision (ICCV) New York, NY: IEEE; 2017:2980–2988.
54. Roth HR, Lu L, Lay N, et al. Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation. Med Image Anal. 2018;45:94–107. [PubMed: 29427897]
55. Huang G, Liu Z, dervan Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017:2261–2269.
56. Drozdzal M, Vorontsov E, Chartrand G, Kadoury S, Pal C. The Impor- tance of Skip Connections in Biomedical Image Segmentation; 2016: 179–187.

57. Quionero-Candela J, Sugiyama M, Schwaighofer A, Lawrence ND. Dataset Shift in Machine Learning. Cambridge, MA: The MIT Press; 2009.
58. Ganin Y, Ustinova E, Ajakan H, et al. Domain-adversarial training of neural networks. J Mach Learn Res. 2016;17:2096–2030.
59. Peng X, Usman B, Saito K, Kaushik N, Hoffman J, Saenko K. Syn2- Real: A New Benchmark for Synthetic-to-Real Visual Domain Adapta- tion. arXiv:180609755; 2018.
60. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks In: Pereira F, Burges CJC, Bottou L, Weinberger KQ, eds. Advances in Neural Information Processing Systems 25. Red Hook, NY: Curran Associates, Inc.; 2012:1097–1105.
61. Szegedy C, Wei L, Yangqing J, et al. Going deeper with convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) New York, NY: IEEE; 2015:1–9.
62. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) New York, NY: IEEE; 2016:770–778.
63. Gupta S, Girshick R, Arbelaez P, Malik J. Learning rich features from RGB-D images for object detection and segmentation In: European Conference on Computer Vision. Vol. 8695 Cham: Springer Nature; 2014:345–360.
64. Marmanis D, Datcu M, Esch T, Stilla U. Deep learning earth observation classification using imagenet pretrained networks. IEEE Geosci Rem Sens Lett. 2016;13:105–109.
65. Som P, Atkins HL, Bandoypadhyay D, et al. A fluorinated glucose ana- log, 2-fluoro-2-deoxy-D-glucose (F-18): Nontoxic tracer for rapid tumor detection. J Nucl Med. 1980;21:670–675. [PubMed: 7391842]
66. Tong X, Srivatsan A, Jacobson O, et al. Monitoring tumor hypoxia using 18F-FMISO PET and pharmacokinetics modeling after photody- namic therapy. Sci Rep. 2016;6:31551. [PubMed: 27546160]
67. Chen W, Cloughesy T, Kamdar N, et al. Imaging proliferation in brain tumors with 18F-FLT PET: comparison with 18F-FDG. J Nucl Med. 2005;46:945–952. [PubMed: 15937304]
68. Kim JW, Oh JS, Roh J-L, et al. Prognostic significance of standardized uptake value and metabolic tumour volume on 18F-FDG PET/CT in oropharyngeal squamous cell carcinoma. Eur J Nucl Med Mol Imaging. 2015;42:1353–1361. [PubMed: 26067088]
69. Beichel RR, Van Tol M, Ulrich EJ, et al. Semiautomated segmentation of head and neck cancers in 18F-FDG PET scans: a just-enough-interaction approach. Med Phys. 2016;43:2948–2964. [PubMed: 27277044]
70. Ishikawa M, Yamaguchi S, Tanabe S, et al. Conceptual design of PET- linac system for molecular-guided radiotherapy. Int J Radiat Oncol Biol Phys. 2010;78:S674.

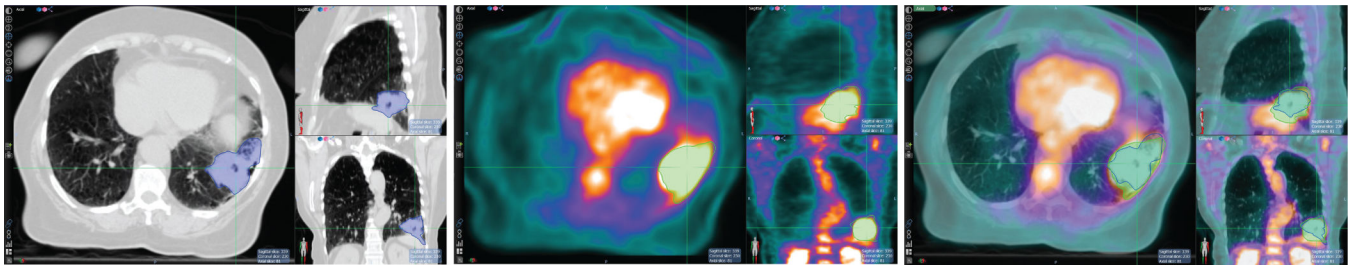

**Figure 1.**

Tumor contours are different in computed tomography (CT) (left), positron emission tomography (PET) (middle), and the fused PET and CT images (right). Due to the inherent differences between molecular imaging (PET) and morphological images (CT), the tumor boundary as defined between PET and CT images may differ.

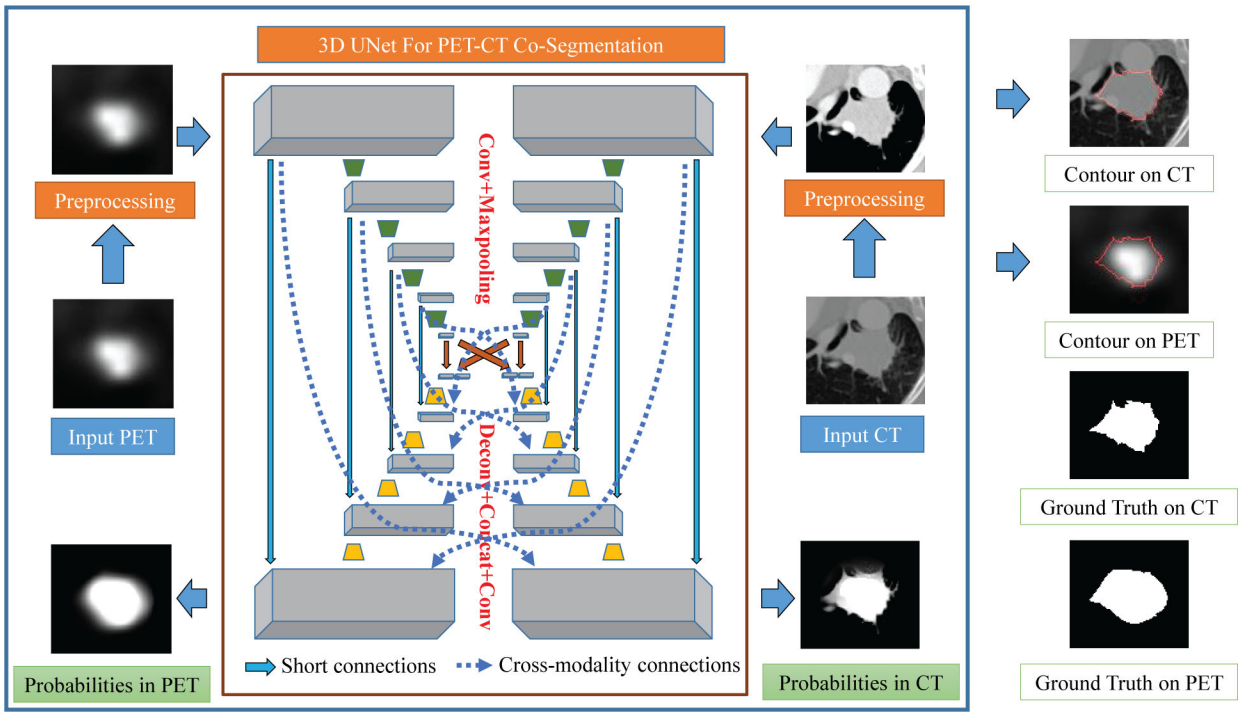


**Figure 2.**

A schematic illustration of our proposed deep learning, fully convolutional networks (DFCN)-CoSeg network with feature fusion for positron emission tomography (PET)-computed tomography (CT) cosegmentation. Two parallel 3D-UNets are built for CT and PET respectively. In DFCN-CoSeg, all feature maps produced by all the encoders of either the CT or PET branch are concatenated in the corresponding decoders, as depicted by the dotted arrow lines.

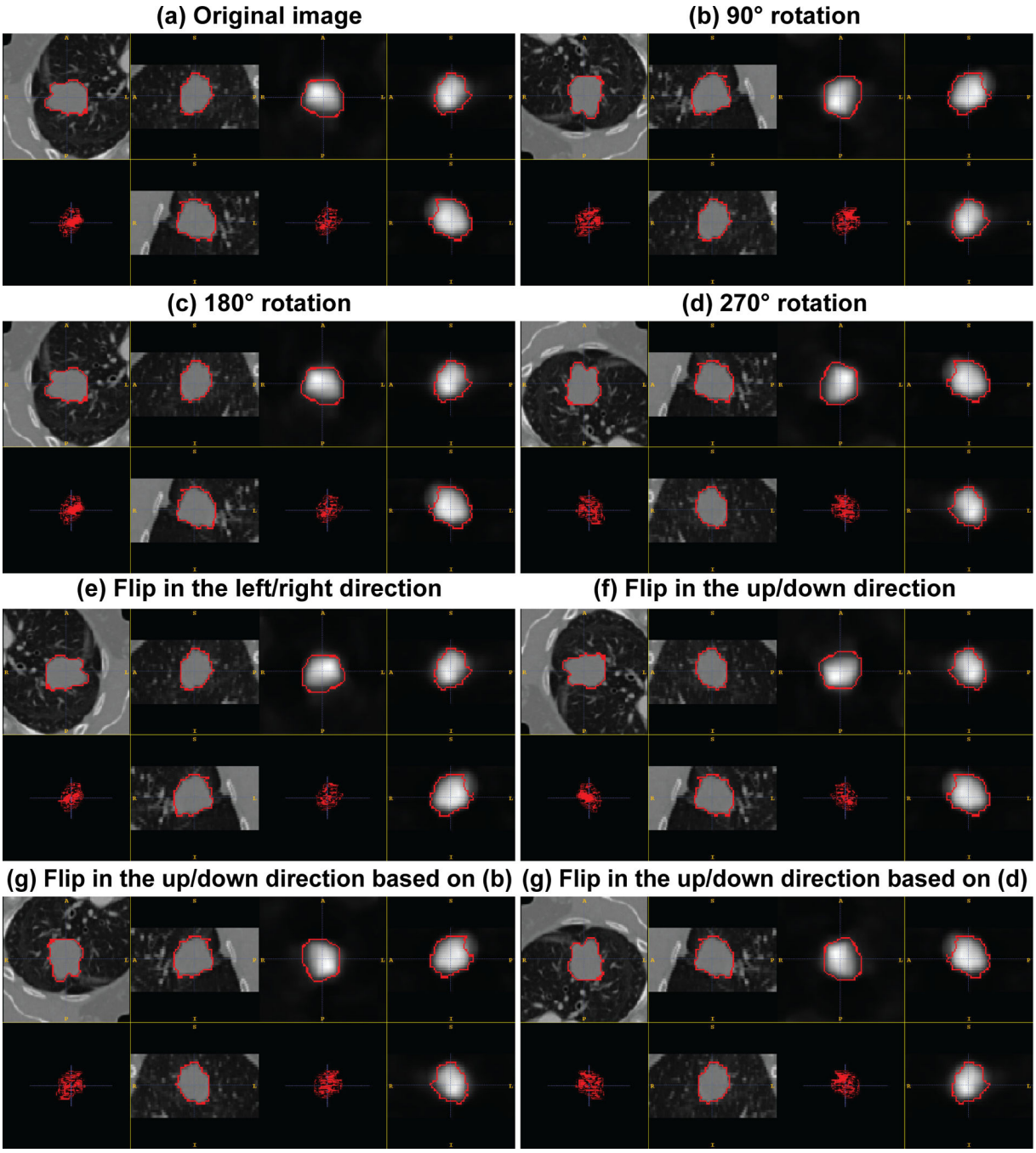


**Figure 3.**
Illustration of some examples using data augmentation. And these augmentation operations were conducted in the X–Y plane.

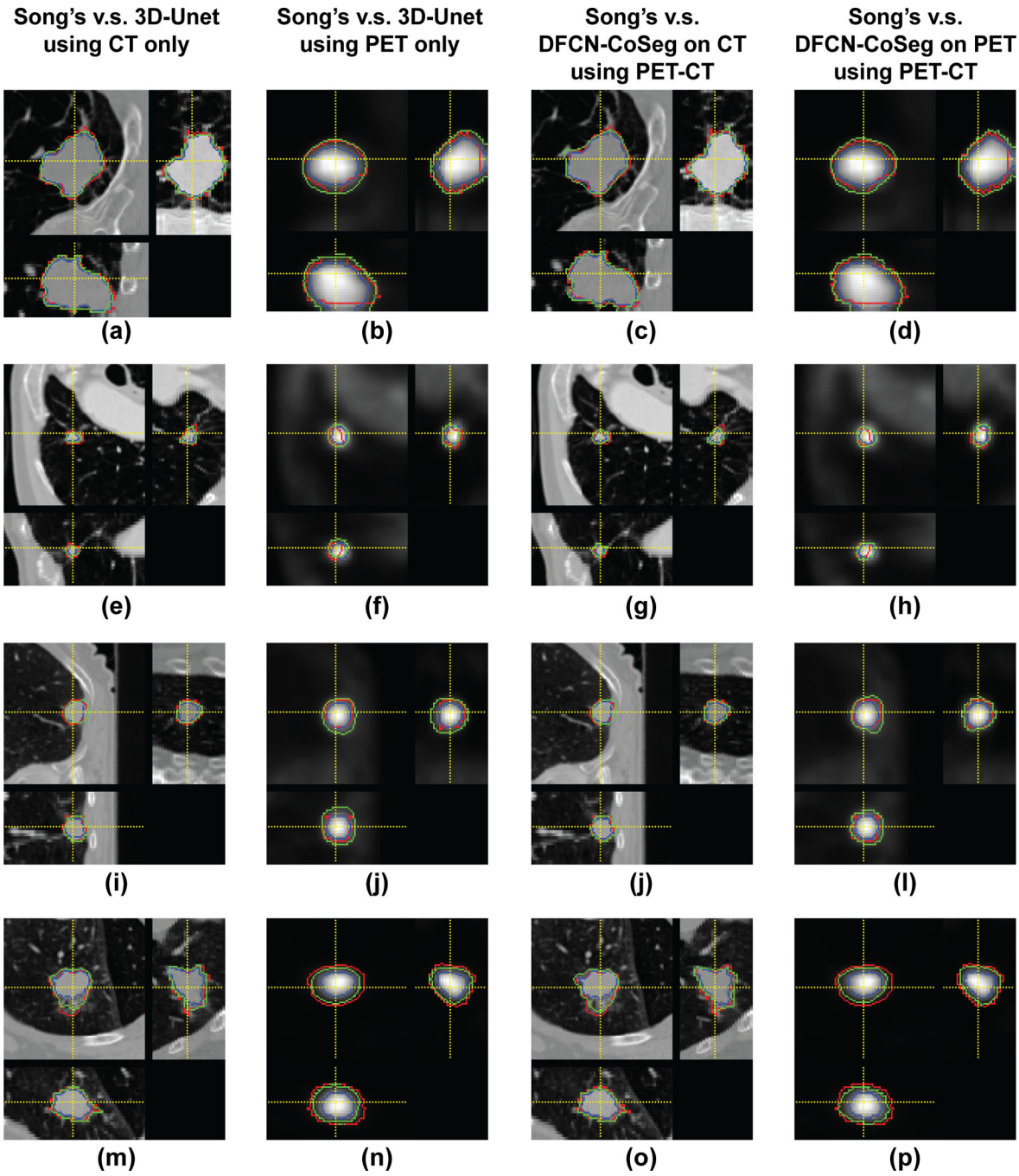


**Figure 4.**

Segmentation results of compared methods on four positron emission tomography-computed tomography scan pairs. Red: ground truth generated by Simultaneous Truth and Performance Level Estimation Algorithm, Blue: Song et al's.12 method, Green: the 3D-UNet method (first two columns), Green: Proposed deep learning, fully convolutional networks-CoSeg method (last two columns).

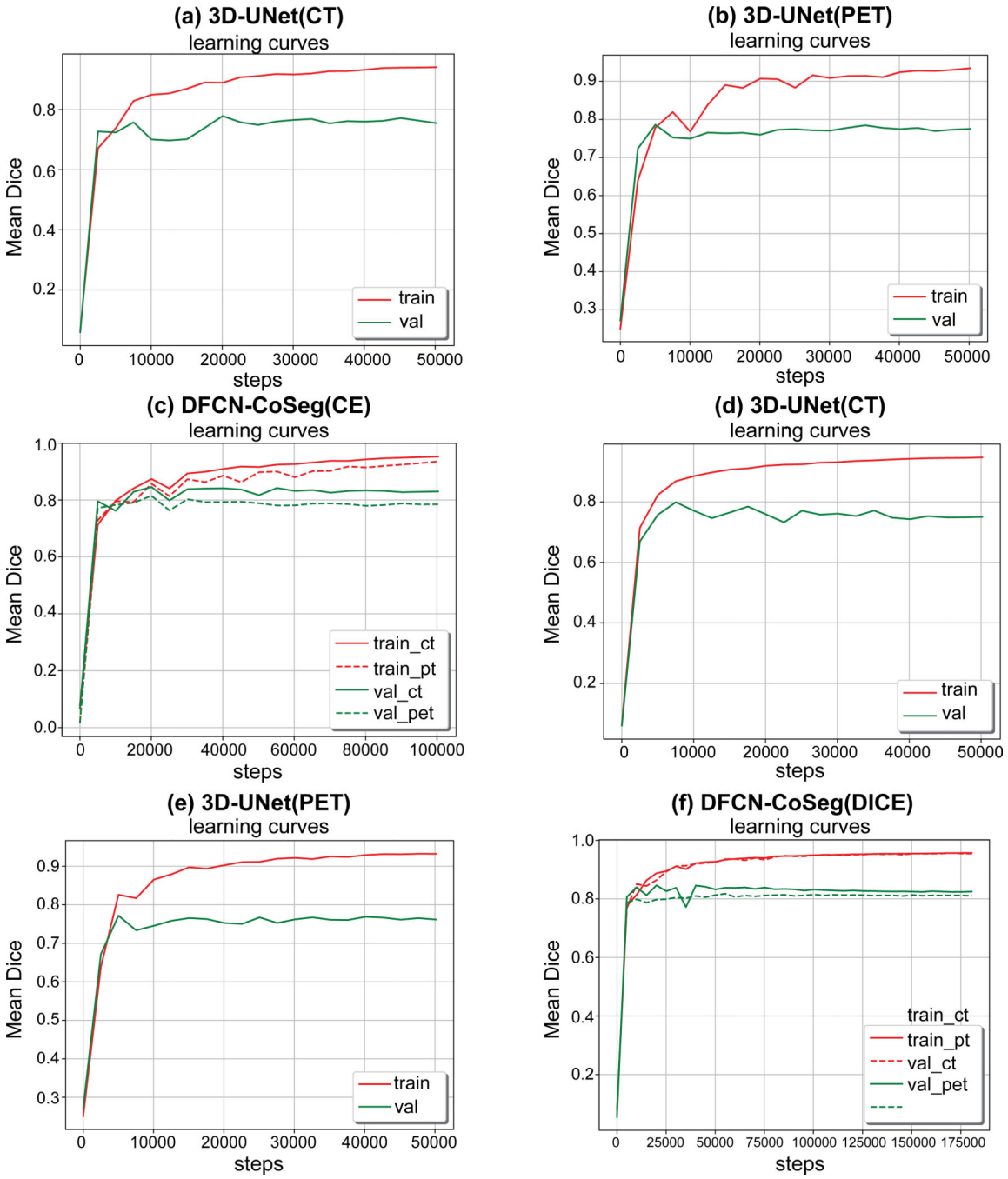


**Figure 5.**

Mean dice similarity coefficient (DSC)s on training and validation sets during the network training with CELoss (a,b,c) and DICELoss (d,e,f). 3D-UNet computed tomography (CT)/ positron emission tomography (PET) means 3D-UNet trained on CT or PET only data. Deep learning fully convolutional networks (DFCN)-CoSeg means the feature fusion DFCN-CoSeg network architecture for PET-CT cosegmentation.

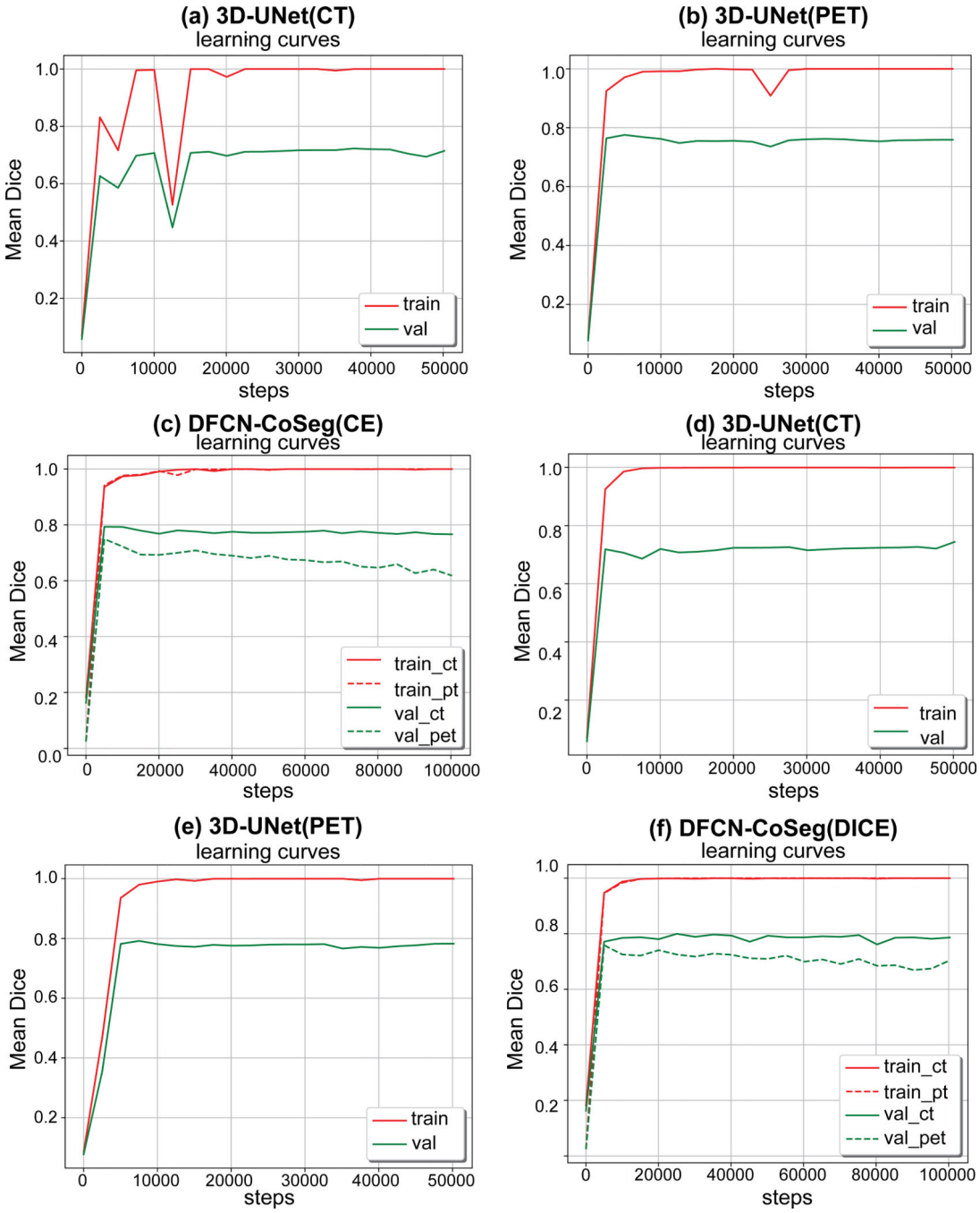


**Figure 6.**

Mean dice similarity coefficients (DSC)s on training and validation sets during the network training with CELoss(a,b,c) and DICELoss(d,e,f). These networks were trained with only the original 38 cases as the training set, i.e., without data augmentation

**TABLE I.**

Architecture of a basic three-dimensional (3D)-UNet.

| | **Feature size** | **3DUNet** |
|---|---|---|
| Input | 1 9 96 9 96 9 48 | |
| Conv 1(F1) | 32 9 96 9 96 9 48 | Conv, 3 9 3 9 3, 32, stride 1 |
| Encoder 1(F2) | 64 9 48 9 48 9 24 | Conv, 3 9 3 9 3, 64, |
| | | maxpool, stride 2 |
| Encoder 2(F3) | 128 9 24 9 24 9 12 | Conv, 3 9 3 9 3, 128, |
| | | maxpool, stride 2 |
| Encoder 3(F4) | 256 9 12 9 12 9 6 | Conv, 3 9 3 9 3, 256, |
| | | maxpool, stride 2 |
| Encoder 4(F5) | 512 9 6 9 6 9 3 | Conv, 3 9 3 9 3, 512, |
| | | maxpool, stride 2 |
| Decoder 4 | 256 9 12 9 12 9 6 | Deconv, concat(F4), 3 9 3 9 3, 256, conv |
| Decoder 3 | 128 9 24 9 24 9 12 | Deconv, concat(F3), 3 9 3 9 3, 128, conv |
| Decoder 2 | 64 9 48 9 48 9 24 | Deconv, concat(F2), 3 9 3 9 3, 64, conv |
| Decoder 1 | 32 9 96 9 96 9 48 | Deconv, concat(F1), 3 9 3 9 3, 32, conv |
| Output | 2 9 96 9 96 9 48 | Conv, 1 9 1 9 1, 2, conv, stride 1 |

**TABLE II.**

Statistics of the compared methods on the test set (22 cases) based on the contours generated by Simultaneous Truth and Performance Level Estimation Algorithm (STAPLE). Average values and their standard deviations were reported.

| | | Score | | DSC | | ASSD | |
|---|---|---|---|---|---|---|---|
| **Methods** | **Modalities** | **Single** | **Dual** | **Single** | **Dual** | **Single** | **Dual** |
| Song et al. | CT | 0.712 ± 0.140 | 0.734 ± 0.089 | 0.597 ± 0.257 | 0.638 ± 0.165 | 2.574 ± 2.447 | 1.938 ± 1.132 |
| | PET | 0.735 ± 0.084 | 0.740 ± 0.074 | 0.629 ± 0.158 | 0.643 ± 0.141 | 2.901 ± 2.607 | 2.656 ± 2.131 |
| Ju et al. | CT | 0.778 ± 0.082 | 0.765 ± 0.098 | 0.765 ± 0.093 | 0.759 ± 0.100 | 1.484 ± 0.745 | 1.650 ± 0.823 |
| | PET | 0.817 ± 0.064 | 0.820 ± 0.061 | 0.776 ± 0.106 | 0.782 ± 0.099 | 1.526 ± 1.068 | 1.473 ± 0.970 |
| Zhong et al. | CT | 0.798 ± 0.058 | 0.809 ± 0.058 | 0.766 ± 0.095 | 0.783 ± 0.095 | 1.186 ± 0.828 | 1.080 ± 0.714 |
| | PET | 0.774 ± 0.065 | 0.816 ± 0.054 | 0.711 ± 0.123 | 0.778 ± 0.086 | 2.102 ± 2.002 | 1.502 ± 1.083 |
| 3D-UNet (CELoss) | CT | 0.812 ± 0.115 | — | 0.780 ± 0.185 | — | 1.667 ± 1.863 | — |
| | PET | 0.846 ± 0.084 | — | 0.811 ± 0.133 | — | 1.127 ± 0.718 | — |
| DFCN-CoSeg (CELoss) | CT | — | 0.850 ± 0.061 | — | 0.836 ± 0.095 | — | 0.895 ± 0.661 |
| | PET | — | 0.848 ± 0.064 | — | 0.823 ± 0.086 | — | 1.066 ± 0.660 |
| 3D-UNet (DICELoss) | CT | 0.839 ± 0.085 | — | 0.811 ± 0.151 | — | 1.291 ± 1.313 | — |
| | PET | 0.832 ± 0.075 | — | 0.794 ± 0.111 | — | 1.229 ± 0.587 | — |
| DFCN-CoSeg (DICELoss) | CT | — | 0.865 ± 0.034 | — | 0.861 ± 0.037 | — | 0.806 ± 0.605 |
| | PET | — | 0.853 ± 0.063 | — | 0.828 ± 0.087 | — | 1.079 ± 0.761 |

**TABLE III.**

Dice similarity coefficients (DSC) values of each physician expert.

| **Dice similarity coefficients (DSC) STAPLE_CT** | | |
|---|---|---|
| P1_CT | P2_CT | P3_CT |
| 0.846 | 0.819 | 0.934 |
| STAPLE_PET | | |
| P1_PET | P2_PET | P3_PET |
| 0.901 | 0.901 | 0.825 |

**Table 4**

| | | With Data-Augmentation | | | Without Data-Augmentation | | |
|---|---|---|---|---|---|---|---|
| **Methods** | **Modalities** | **Score** | **DSC** | **ASSD** | **Score** | **DSC** | **ASSD** |
| 3D-UNet(CELoss) | CT | 0.812 ± 0.115 | 0.780 ± 0.185 | 1.667 ± 1.863 | 0.752 ± 0.157 | 0.719 ± 0.207 | 3.154 ± 4.336 |
| | PET | 0.846 ± 0.084 | 0.811 ± 0.133 | 1.127 ± 0.718 | 0.819 ± 0.078 | 0.779 ± 0.127 | 2.491 ± 2.809 |
| DFCN-CoSeg (CELoss) | CT | 0.850 ± 0.061 | 0.836 ± 0.095 | 0.895 ± 0.661 | 0.818 ± 0.084 | 0.806 ± 0.108 | 1.226 ± 0.987 |
| | PET | 0.848 ± 0.064 | 0.823 ± 0.086 | 1.066 ± 0.660 | 0.793 ± 0.101 | 0.771 ± 0.116 | 1.993 ± 2.346 |
| 3D-UNet (DICELoss) | CT | 0.839 ± 0.085 | 0.811 ± 0.151 | 1.291 ± 1.313 | 0.759 ± 0.156 | 0.728 ± 0.204 | 3.431 ± 5.176 |
| | PET | 0.832 ± 0.075 | 0.794 ± 0.111 | 1.229 ± 0.587 | 0.832 ± 0.076 | 0.808 ± 0.105 | 1.308 ± 0.750 |
| DFCN-CoSeg (DICELoss) | CT | 0.865 ± 0.034 | 0.861 ± 0.037 | 0.806 ± 0.605 | 0.811 ± 0.098 | 0.785 ± 0.152 | 2.632 ± 4.657 |
| | PET | 0.853 ± 0.063 | 0.828 ± 0.087 | 1.079 ± 0.761 | 0.803 ± 0.087 | 0.778 ± 0.107 | 1.856 ± 1.976 |

**TABLE V.**

Statistics of SEs, PPVs on the whole-body PET/CT images of the test set. The "Voxels" means the number of tumor voxels in the resampled data with volume spacing of 1 9 1 9 1 mm.

| | | | 3D-Unets(CT)CoSeg(PET) | | 3D-Unets(PET) | | DFCN-CoSeg(CT) | | DFCN- | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Name** | **Voxels-ct** | **Voxels-pt** | **SE** | **PPV** | **SE** | **PPV** | **SE** | **PPV** | **SE** | **PPV** |
| A-IA002126 | 3430 | 6496 | 0.903 | 0.105 | 0.645 | 0.093 | 0.871 | 0.171 | 0.590 | 0.135 |
| A-IA002122 | 16806 | 31164 | 0.678 | 0.402 | 0.964 | 0.278 | 0.864 | 0.421 | 0.903 | 0.407 |
| A-IA0002096 | 12424 | 22990 | 0.502 | 0.143 | 0.166 | 0.430 | 0.655 | 0.748 | 0.074 | 1.000 |
| A-IA002119-M | 10522 | 7888 | 0.854 | 0.168 | 0.567 | 0.485 | 0.861 | 0.821 | 0.721 | 0.758 |
| A-IA0002111 | 2034 | 2476 | 0.755 | 0.015 | 0.999 | 0.141 | 0.873 | 0.249 | 0.924 | 0.323 |
| A-IA0002114 | 37628 | 23224 | 0.044 | 0.026 | 0.745 | 0.020 | 0.370 | 0.062 | 0.440 | 0.027 |
| A-IA0002097 | 2786 | 7036 | 0.293 | 0.025 | 0.891 | 0.114 | 0.857 | 0.087 | 0.867 | 0.260 |
| A-IA000991-M | 3116 | 12318 | 0.587 | 0.090 | 0.821 | 0.035 | 0.934 | 0.014 | 0.930 | 0.042 |
| A-IA0001254 | 65604 | 104091 | 0.365 | 0.023 | 0.599 | 0.610 | 0.564 | 0.540 | 0.558 | 0.793 |
| A-IA002135-M | 4494 | 5240 | 0.852 | 0.081 | 0.927 | 0.090 | 0.900 | 0.089 | 0.961 | 0.140 |
| A-IA002133 | 6910 | 9458 | 0.837 | 0.071 | 0.739 | 0.480 | 0.916 | 0.516 | 0.558 | 0.671 |
| A-IA0002094 | 1026 | 2040 | 0.175 | 0.006 | 0.647 | 0.032 | 0.666 | 0.030 | 0.578 | 0.025 |
| A-IA002130 | 5018 | 6156 | 0.886 | 0.254 | 0.633 | 0.946 | 0.938 | 0.817 | 0.927 | 0.852 |
| A-IA002134 | 21210 | 41666 | 0.425 | 0.083 | 0.870 | 0.208 | 0.869 | 0.234 | 0.896 | 0.257 |
| A-IA002131 | 11156 | 6510 | 0.767 | 0.258 | 0.012 | 0.000 | 0.768 | 0.029 | 0.736 | 0.010 |
| A-IA0002108 | 7832 | 2258 | 0.797 | 0.111 | 0.752 | 0.024 | 0.789 | 0.427 | 0.882 | 0.070 |
| A-IA001491 | 2068 | 1206 | 0.517 | 0.016 | 0.912 | 0.021 | 0.613 | 0.046 | 0.949 | 0.034 |
| A-IA0001345 | 39432 | 44552 | 0.844 | 0.590 | 0.949 | 0.787 | 0.869 | 0.839 | 0.885 | 0.883 |
| A-IA0002109 | 2908 | 6730 | 0.863 | 0.055 | 0.995 | 0.037 | 0.879 | 0.024 | 0.989 | 0.044 |
| A-IA0002117 | 12292 | 17586 | 0.699 | 0.108 | 0.787 | 0.671 | 0.870 | 0.906 | 0.838 | 0.983 |
| Mean | | | 0.632 | 0.132 | 0.731 | 0.275 | 0.796 | 0.354 | 0.760 | 0.386 |