Prediction of Major Torso Organs in Low-contrast Micro-CT Images of Mice using a Two-Stage Deeply Supervised Fully Convolutional Network

Hongkai Wang[1], Ye Han[1], Zhonghua Chen[2], Ruxue Hu[2], Arion F. Chatziioannou[3], Bin Zhang[1,4]

[1]School of Biomedical Engineering, Dalian University of Technology, Dalian, Liaoning 116024, China

[2]Faculty of Information Technology, University of Jyväskylä, Jyväskylä, Finland

[3]Crump Institute of Molecular Imaging, David Geffen School of Medicine, University of California, Los Angeles, CA 90066 USA

[4]Author to whom any correspondence should be addressed.

Abstract

Delineation of major torso organs is a key step of mouse micro-CT image analysis. This task is challenging due to low soft tissue contrast and high image noise, therefore anatomical *prior* knowledge is needed for accurate prediction of organ regions. In this work, we develop a deeply supervised fully convolutional network which uses the organ anatomy *prior* learned from independently acquired contrast-enhanced micro-CT images to assist the segmentation of non-enhanced images. The network is designed with a two-stage workflow which firstly predicts the rough regions of multiple organs and then refines the accuracy of each organ in local regions. The network is trained and evaluated with 40 mouse micro-CT images. The volumetric prediction accuracy (Dice score) varies from 0.57 for the spleen to 0.95 for the heart. Compared to a conventional atlas registration method, our method dramatically improves the Dice of the abdominal organs by 18~26%. Moreover, the incorporation of anatomical prior leads to more accurate results for small-sized low-contrast organs (e.g. the spleen and kidneys). We also find that the localized stage of the network has better accuracy than the global stage, indicating that localized single organ prediction is more accurate than global multiple organ prediction. With this work, the accuracy and efficiency of mouse micro-CT image analysis are greatly improved and the need for using contrast agent and high X-ray dose is potentially reduced.

Keywords: Fully convolutional network, micro-CT, mouse image, deeply supervised network, organ segmentation

Introduction

As a mammal model whose genome has 85 percent of protein-coding regions identical to the human genome, mice are commonly used in preclinical studies of drug development, pharmacokinetic analysis and tumor target identification (Kagadis *et al.*, 2010). *In vivo* mouse

imaging is of great importance for observing the phenotype and cancerous changes in the body. In the field of medical physics, micro-CT is a widely used preclinical imaging modality for observing internal organ structures (Badea *et al.*, 2011; Burk *et al.*, 2012; Guo *et al.*, 2011; Holbrook *et al.*, 2018) and for providing anatomical reference to functional signal (Zhang *et al.*, 2014). As a post-processing step, delineating important organ regions from mouse preclinical image is important for monitoring organ morphometry changes (Kockelkorn *et al.*, 2014), quantifying organ molecular tracer uptake (Maroy *et al.*, 2010; Cheng and Qi, 2010; Wang *et al.*, 2012a), assisting functional image reconstruction (Joshi *et al.*, 2010), measuring radiation dosimetry (Welch *et al.*, 2015), and etc.

To alleviate human operator burden, and to reduce subjectivity of image analysis in small animal imaging centers, automated mouse image analysis is preferred. However, soft tissues in non-enhanced micro-CT images always suffer from poor intensity contrast (as shown in Fig.1), making automated delineation a challenging problem. To enhance soft tissue contrast, X-ray attenuating agents can be used, leading to more complicated experiment procedures and higher experimental expenditures (Yan *et al.*, 2017). Therefore, most small animal micro-CT scans are still acquired without contrast agent to maintain high throughput and low cost. Using advanced software algorithms to predict organ regions from low-contrast micro-CT images offers a practical solution to this problem. Such an algorithm should be able to use prior anatomical knowledge to compensate for the lack of localizing features in the low-contrast image.

To this day, mainstream methods of mouse organ localization in low-contrast micro-CT images are based on atlas registration. By registering a digital mouse atlas with the target subject image, the pre-defined organ regions in the atlas are non-rigidly warped into the subject image, giving an estimation of the organ locations. Representative methods of this type include the articulated skeleton guided atlas registration by Baiker et al. (Baiker *et al.*, 2010; Khmelinskii *et al.*, 2011; Snoeks *et al.*, 2012; Baiker *et al.*, 2011; Kok *et al.*, 2010; Snoeks *et al.*, 2011), the statistical organ shape model approach by Wang et al. (Wang *et al.*, 2012b) and the multi-atlas registration method by Ren et al. (Ren *et al.*, 2016). Atlas-based approaches utilize the prior anatomical knowledge defined in the atlas to compensate for the lack of features of the low-contrast organs. In addition to atlas-based methods, there are also studies using multi-modality imaging strategy to compensate for the missing organ information in micro-CT images (Akselrod *et al.*, 2016), as well as using dedicated animal shuttles to improve segmentation accuracy by constraining the body position (Klose and Paragas, 2018). However, the use of additional imaging modalities or special animal holders further complicates the imaging procedure.

Deep learning methods have achieved remarkable success in the classification, diagnosis, segmentation and registration of clinical medical images (Litjens *et al.*, 2017). For organ segmentation, deep convolutional neural network (CNN) and its variants (Ronneberger *et al.*, 2015; Milletari *et al.*, 2016; Shelhamer *et al.*, 2017)) significantly outperformed conventional segmentation methods in terms of segmentation accuracy and computation speed. For example, Kearney *et al.* present a deeply supervised network (Kearney *et al.*, 2019), which combined attention gates (AGs) and channel boosting to improve the convergence of the model, increasing the numbers of learned features. Li *et al.* propose the H-DenseUNet to extract intra-slice and inter-slice features effectively in order to avoid heavy computational cost, instead of training 3D networks directly (Li *et al.*, 2018). Gibson *et al.* use Dense V-networks to segment multi-organ of abdominal CT automatically (Gibson *et al.*, 2018). Aided by modern graphics process units (GPU), a CNN

typically takes less than one minute to segment a volumetric image, while conventional atlas-based methods may take minutes or even hours. Considering the merits of deep learning approaches, it is worthy to apply the CNN-based methods to the segmentation of mouse micro-CT images. However, to the best of our knowledge, such applications are still rare, probably due to the following reasons.

i. Most existing CNN methods are developed to segment high-contrast organs with clear boundaries. These methods may fail to segment the low contrast organs which are almost indistinguishable from the surrounding tissues.
ii. Due to the low boundary contrast, even human operators cannot give accurate annotation. Therefore, the network training is hampered by the lack of label annotations.
iii. Considering the lack of enough image contrast, anatomical priors should be used to aid organ prediction. Although there have been studies introducing anatomical priors to CNN (Krizhevsky *et al.*, 2012), these studies were mostly focused on high-contrast organs rather than low-contrast organs.

In summary, organ prediction in low-contrast images is a different problem from organ segmentation in high-contrast images. In many applications of preclinical image analysis, prediction of organ region is used for quantifying tracer uptake (Maroy *et al.*, 2008). In such cases, only a region of interest (ROI) is required to be drawn in the organ center, while complete segmentation of the entire organ is not necessary. If the algorithm can guarantee that the core part of the predicted region lies inside the target organ, a reliable ROI can be created to provide accurate tracer quantification. Therefore, the objective of organ prediction is to ensure the major part of the predicted region overlaps with the ground truth region, rather than to accurately delineate the organ boundary. The network meeting this objective ought to infer the organ regions from very week contrast appearance. To achieve such inference ability, the memory of organ locations should be deeply implanted into the network architecture. We've made the following efforts to address these issues.

i. To solve the problem of missing training labels, we created pseudo low-contrast micro-CT images from contrast-enhanced micro-CT images. The pseudo low-contrast images were used to train the network, and the training labels come from human annotations of the original contrast-enhanced images. The trained network was then evaluated on real low-contrast images and its effectiveness was validated.
ii. To implant the memory of organ locations into the network, we adopt the mechanism of deep supervision for network training (Dou *et al.*, 2017). The term 'deep supervision' means to use ground truth segmentation labels to supervise the learning process of the intermediate network layers. In contrast, the networks without deep supervision only use the ground truth labels to supervise the last output layer. In this paper, the labels coming from contrast-enhanced images were used as anatomical priors to supervise the training of intermediate network layers so that the network outputs anatomically plausible results even when the organ contrast is low.
iii. The labels coming from contrast-enhanced images were used as anatomical priors to supervise the training of intermediate network layers so that the network outputs anatomically plausible results even when the organ contrast is low.
iv. To ensure that the major part of the predicted region overlaps with the ground truth region, we design a two-stage network structure in which the first stage localizes rough organ regions from the whole-body image, and the second stage gives more accurate prediction of every organ region. We found that the localized prediction of each single organ is more accurate than the

globalized prediction of multiple organs.

Materials

Because the network is designed to process low contrast micro-CT images, the training images should present low soft tissue contrast. However, accurately annotating organ regions in low contrast images is difficult even for human experts. To tackle this problem, we use contrast-enhanced micro-CT images to provide accurate label annotations, and then simulate pseudo low contrast images using the high contrast images to generate the training data. Moreover, we use real low-contrast micro-CT images to further prove the effectiveness of our method on real-world data.

Forty contrast-enhanced whole-body mouse micro-CT images were randomly selected from the database of small animal images at the Crump Institute of UCLA (Stout *et al.*, 2005). These images were acquired for mouse subjects injected with the Fenestra LC contrast agent (ART, Montreal, QC, Canada) which enhances the intensities of the liver, spleen and kidneys. The reconstructed voxel size was 0.20 mm and the image array size was 256×256×496, the exposure settings were 70 kVp, 500 mAs with 2.0 mm aluminum filtration. The scanner was MicroCAT II small animal CT system (Siemens Preclinical Solutions, Knoxville, TN, USA). From the contrast-enhanced images, target organ regions were manually segmented by a human expert of mouse anatomy, and the label image of multiple organ regions were created from the expert labels.

Within the expert-segmented regions of the liver, spleen and kidneys, the voxel intensity was substracted with an intensity offset $\mu \in (200, 800)$ to create the pseudo low contrast image. Because the enhanced organ intensities were inconsistent between the different organs and different subjects, the values of μ were manually adjusted for each organ of each subject. Fig.1 shows a pseudo low-contrast image compared with the contrast-enhanced image and the real low-contrast image of the same mouse subject. It can be observed that the pseudo low-contrast (PLC) image visually resembles the real low-contrast image.

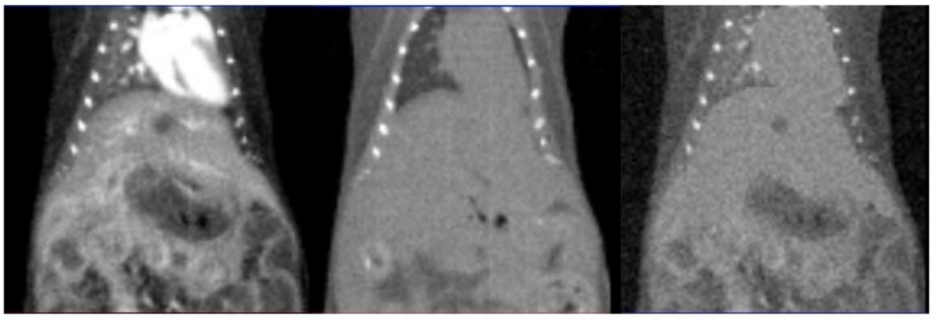

Fig.1. A contrast-enhanced mouse micro-CT image (left), the low-contrast CT image of the same mouse subject (middle) and the pseudo low-contrast (PLC) image generated from the contrast-enhanced image (right).

To reduce the computational overhead and the training time, torso regions were cropped from the pseudo low contrast images and their corresponding label images. The cropped images were resampled to the size of 96×96×96 and their grayscales were normalized to the range of [0,1].

Method

Fig.2 illustrates the training and inference workflow of the proposed two-stage deeply supervised network (TS-DSN). The network takes a volumetric micro-CT image of mouse torso as the input, it outputs a same-sized label image of multiple organs including the heart, lungs, liver, spleen, left kidney and right kidney. These organs are important for preclinical biodistribution and tracer metabolism studies. As shown by Fig.2(a), the training workflow includes two stages. The first stage generates pseudo low-contrast (PLC) images from high-contrast images and then use the generated images to train a rough prediction network. The second stage crops out the local image of each organ and trains multiple prediction networks for different organs. For inference, the first stage network is used to localize the rough organ regions, then the second stage network of each organ is used for more accurate region prediction (Fig.2(b)). Each step of the training and inference workflows is elaborated in detail as below.

A. The First-Stage Network Training

The first stage network is designed with a deep supervision architecture as illustrated in Fig.3. We adopt the idea of deep supervision (Dou *et al.*, 2017) to incorporate anatomical priors into the network, as well as to prevent the gradient vanishing problem of training deep 3D networks.

The network contains 12 layers and the convolution kernel size was set to $3\times3\times3$ for a good receptive field. The sampling stride was set to 2 to produce a double size down-scaling convolution map. Each convolution layer contains a batch normalization layer and uses the rectified linear unit (ReLU) as an activation function. The ReLU function was defined as F(x) = x for x>0 and 0 otherwise.

We name the encoding and decoding parts of the network as the main network. The hidden deconvolutional layers in the dotted-lined box of Fig.3 are named as the auxiliary network. The main network performs the image segmentation as the conventional fully convolutional network (FCN), while the auxiliary network compares the intermediate layer feature maps with the down-sampled ground truth label map. The loss function of the entire network incorporates both losses of the main network and the auxiliary network. In this way, the anatomical priors from the ground truth labels are used to deeply supervise the training of each intermediate layer. In the main network, the main loss function $\mathcal{L}$ is defined as

$$\mathcal{L}(\mathcal{X},\mathcal{Y};W) = \sum_{x_i\in\mathcal{X}} -logp(y_i|x_i;W) \quad (1)$$

where x_i means the voxel and y_i means the label of x_i. W is the weight of the main network. $p(y_i|x_i;W)$ means the probability of x_i belonging to the class of y_i. In the auxiliary network, the auxiliary loss $\mathcal{L}_u$ is defined as

$$\mathcal{L}_u(\mathcal{X},\mathcal{Y};W_u) = \sum_{x_i\in\mathcal{X}} -logp(y_i|x_i;W_u) \quad (2)$$

where W_u means the weights of the auxiliary network. Therefore, the loss function of the deeply supervised network is defined as

$$\mathcal{L} = \mathcal{L}(\mathcal{X},\mathcal{Y};W) + \sum_{u\in\mathcal{U}} \alpha_u\mathcal{L}_u(\mathcal{X},\mathcal{Y};W_u) + \mu(\|W\|^2 + \sum_{u\in\mathcal{U}}\|W_u\|^2) \quad (3)$$

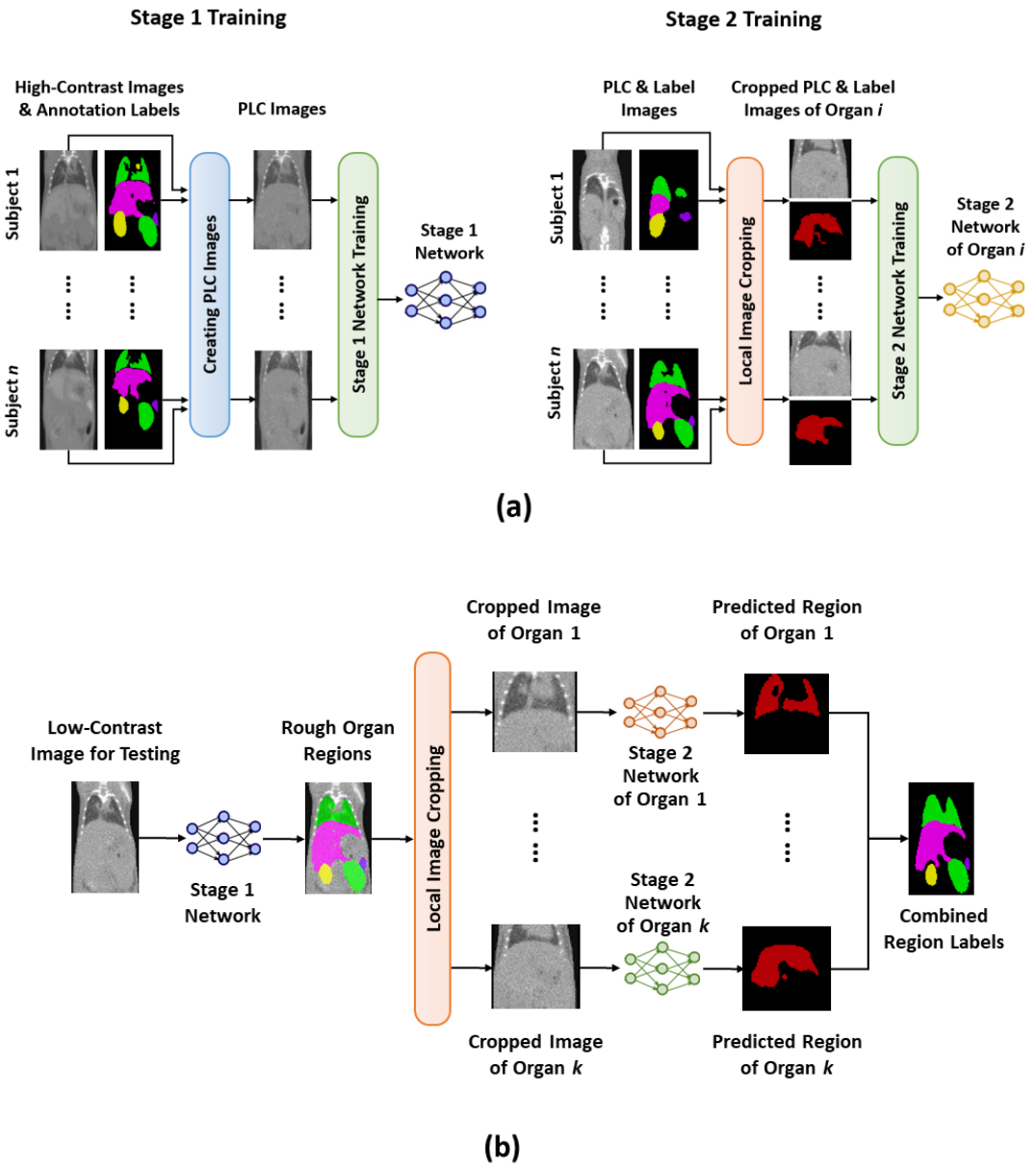

Fig.2. The workflows of network training and inference. (a). The workflow of the first training stage and the second training stage. (b). The inference workflow.

where $\mathcal{X}$ means the training set and $\mathcal{Y}$ means the label. $(\mathcal{X},\mathcal{Y})$ consists the training dataset pair, $\mathcal{W}$ is the weight of the main network, W_u is the weight of the auxiliary network, which is shown in Fig 3. μ is the hyperparameter of the L2 regularization, α_u is the weight of $\mathcal{L}_u$. The first term is the main loss, the second term comes from deeply supervised auxiliary loss, and the third term means L2 regularizations. The hidden layer deconvolves the intermediate feature maps until the input and output sizes are the same. The lower-level feature maps obtained via the deconvolution were then used to calculate the loss using the softmax layer.

B. The Second-Stage Network Training

The purpose of the second stage network is to further improve the organ region prediction

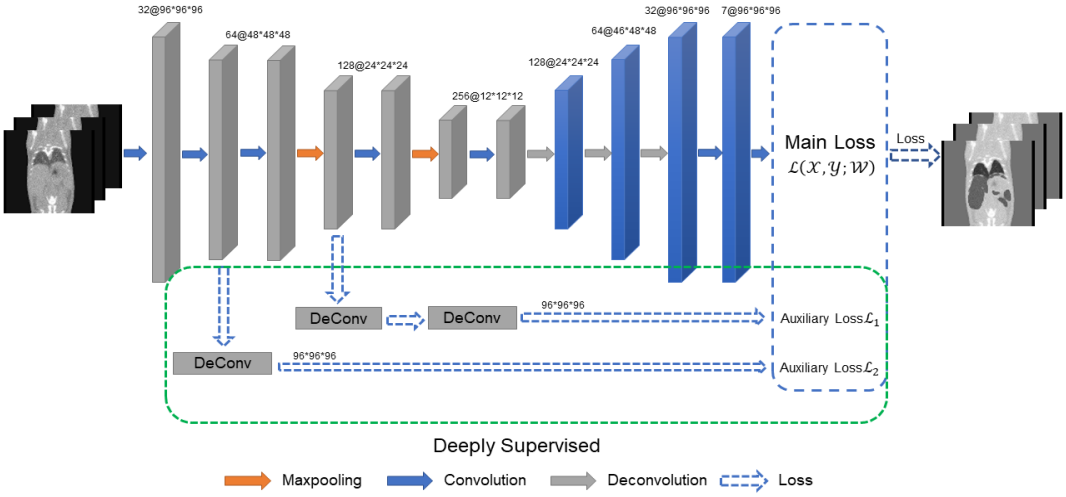

Fig.3. Structure of the deeply supervised network.

accuracy based on the first-stage result. The first-stage network takes the whole-torso image as input and predicts multiple organs simultaneously. It is difficult to train such a global network to produce accurate results for multiple organs, therefore we train separated local networks for individual organs in the second stage.

Based on the organ regions obtained from the first-stage result, the largest connected region of each organ is extracted, and the 3D bounding box of extracted region is obtained. Using the bounding box, the local image of each organ and its corresponding label image are cropped from the whole-torso micro-CT and label images. The cropped local images are used to train the specific networks for each organ. As a result, we trained six organ-specific networks for the six target organs. The network structure and training configurations of the second stage were the same as the first stage.

C. The Inference Workflow

After the two-stage training processes, one global network and six organ-specific local networks are obtained. For inference, the global network is used to generate rough predictions from the whole-torso micro-CT image, then the local organ images are cropped in the same manner as of the second-stage training. The cropped images are input to the organ-specific networks to generate the local label image of each organ. Finally, the local label images are combined into a whole-torso multi-label image as the final output of the two-stage workflow.

D. Implementation Details

The proposed network was implemented with 3D CAFFE (Dou *et al.*, 2017) on a server with three NVIDIA P40 GPUs and 128G RAM. Each session took about 30h to train. For inference, the two stages in Fig.2(b) totally took less than two seconds for one test image. The training process uses a stochastic gradient dependent (SGD) optimizer (Bottou, 2010) with cross-entropy loss. The initial learning rate, which influences the rate of the model converge, was set to 0.01. It is also decreased along a polynomial curve with power 0.9. The weight decay was set to 0.005 and the momentum was 0.9 to get rid of local optima. Due to the limitation of memory size of GPU, the batch size was set to 1. The deep supervision weights α_1 and α_2 were initialized as 0.3 and 0.4 and then decayed as training goes on. The network parameters were initiated with random Gaussian

distribution and trained to stop after 30000 iterations. The last layer of the network outputs the probability map of each organ. A threshold of 0.01 was used to generate the binary label maps. Due to limited training data and high-dimension network, data argumentation strategy was used to avoid overfitting. The augmented data was obtained by shifting the original image along three dimensions by 1, 3, 5, -1, -3, -5 and 0 voxels. It should be noted that all the above training parameters were empirically fine-tuned for the training images of this study therefore were specific for the low-contrast micro-CT images. In case the same methodology strategy is used for other small animal imaging modalities (e.g. MR and PET), the training parameters should be optimized otherwise. The network models of TS-DSN have been opened on GitHub website (https://github.com/DlutMedimgGroup/MouseCTSegmentation).

The evaluation of pseudo data was performed using a cross-validation strategy. Each time we used 5 of the 40 pseudo images as the testing data and used the remaining 35 images for network training. Such a test was repeated eight times and the results of all the eight folds were collected for accuracy validation.

E. Evaluation of the segmentation accuracy

We used four metrics to evaluate the prediction accuracy, including Dice coefficient (Dice), recovery coefficient (RC) of organ volume, precision and recall. The ground truth organ regions come from human expert labels based on the contrast-enhanced images. The metrics are defined as follows,

$$\text{Dice} = 2\frac{|R_p \cap R_g|}{|R_p| + |R_g|}, \tag{4}$$

$$\text{RC} = \frac{|R_p|}{|R_g|}, \tag{5}$$

$$\text{Precision} = \frac{|R_p \cap R_g|}{|R_p|}, \tag{6}$$

$$\text{Recall} = \frac{|R_p \cap R_g|}{|R_g|}, \tag{7}$$

where R_p and R_g represent the predicted and ground truth organ regions. $|\cdot|$ denotes the number of voxels, $\cap$ indicates overlapping between two regions; The Dice score reflects prediction accuracy; RC reflect the prediction accuracy of organ volume, it is important for the applications of preclinical biodistribution studies which requires a good estimation of organ volume. Precision reflects the ratio of true positive voxels versus all predicted voxels, while recall reflects the ratio of true positive voxels versus all ground truth voxels. Precision and recall are essential indices for molecular tracer uptake quantification which requires that most of the predicted voxels lie inside the ground truth region.

Results

A. Pseudo Low-Contrast Image Results

The proposed two-stage deeply supervised network (TS-DSN) was compared with two baseline methods, i.e. the conventional FCN without deep supervision mechanism (Shelhamer *et al.*,

2017) and the deeply-supervised FCN (DSN) without two-stage architecture (Dou *et al.*, 2017). The purpose of this comparison was to reveal the advantages of the deep supervision scheme and the two-stage strategy. Fig. 4 displays representative segmentation results of FCN, DSN and TS-DSN for different organs. The ground truth contours come from the human expert segmentation of the contrast-enhanced images used for generating the pseudo low contrast images. It can be visually inspected that all the three methods accurately segmented the high contrast organ boundaries like the lung boundary and the upper liver boundary. However, for the low contrast organ boundaries (e.g., the spleen and kidney boundaries and the lower liver boundary) which are almost indistinguishable in the image, the two deeply supervised networks (DSN and TS-DSN) yielded more precise contours than FCN, implying that the deep supervision scheme helps to generate anatomically plausible boundary using the prior shape knowledge. Moreover, as we further compare between the two deeply supervised networks, TS-DSN produces results closer to the ground truth than DSN (as indicated by the arrows in Fig. 4). For the small low-contrast organs like spleen and kidneys which are almost invisible in the image, TS-DSN produced less false-positive predictions outside the ground truth regions, proving the effectiveness of localized organ segmentation strategy (i.e., the second stage of the proposed workflow).

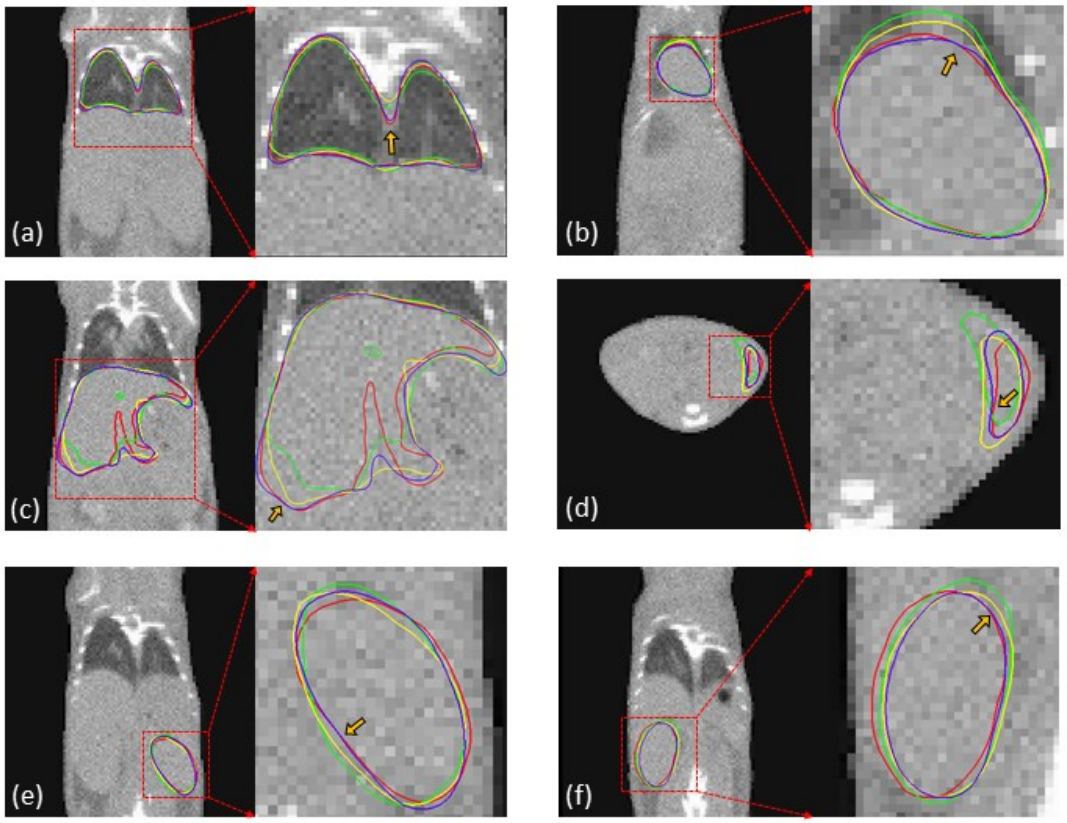

Fig.4. Representative segmentation results of the compared methods. (a)-(f) show the segmentation results of the lungs, heart, liver, spleen, left kidney, and right kidney, respectively. The red, green, yellow and blue contours indicate the ground truth and FCN, DSN, TS-DSN results, respectively. In each sub-figure, the right part shows a zoom-in view of the left part. The arrows points to locations where TS-DSN yielded more accurate results than DSN.

To give a quantitative comparison between different methods, Fig.5 shows the box plots of Dice, RC, Precision and Recall between FCN, DSN and TS-DSN. TS-DSN yielded increased median Dice compared to FCN and DSN. TS-DSN also has narrower Dice range than FCN and

DSN, meaning that the two-stage strategy lead to more stable results across different test data. The RC of TS-DSN is closer to 1 than FCN and DSN, revealing better volume recovering ability of TS-DSN. From Fig.5(c) and (d), it seems that FCN and DSN have higher recall but much lower precision than TS-DSN, implying that FCN and DSN make more false positive predictions than TS-DSN. This finding is also supported by the RC result (Fig.5(b)) where FCN and DSN tends to overestimate (i.e., RC>1) the organ volume. In contrast, TS-DSN is more conservative in positive prediction. TS-DSN has RC close to 1, it also has high precision and moderate recall values, meaning that its segmentation results mostly lie inside the ground truth organ region.

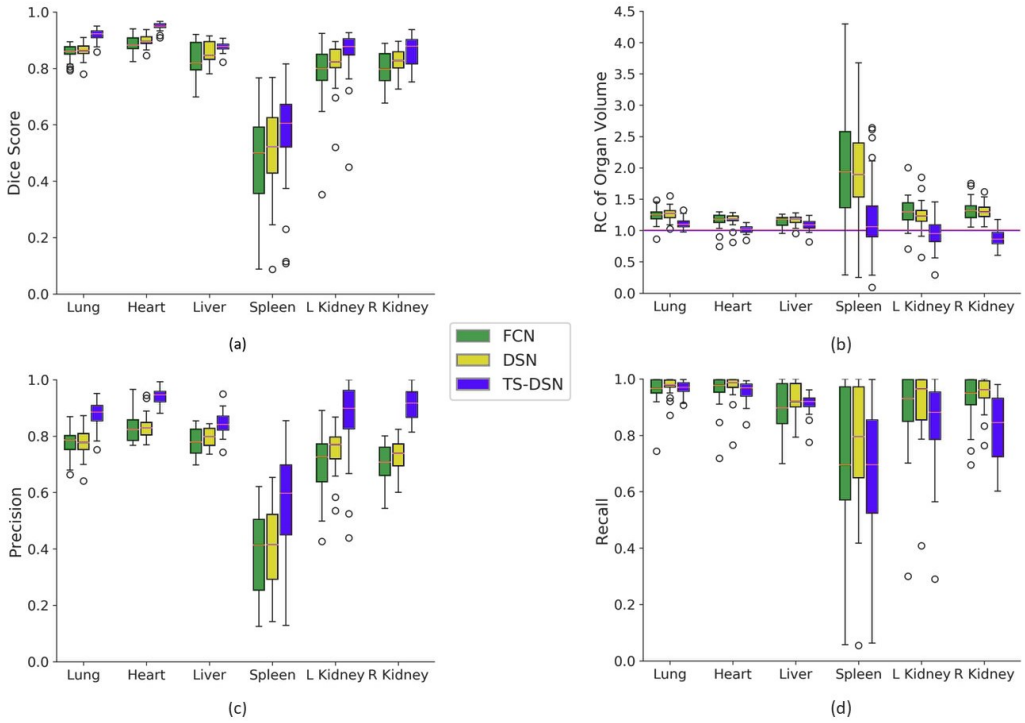

Fig.5. Box-plots of Dice, RC of organ volume, precision and recall of FCN (green), DSN (yellow) and TS-DSN (blue).

B. Comparison with the state-of-the-art method

Among the published methods of low-contrast mouse micro-CT organ segmentation, our previous approach based on statistical shape model (SSM) (Wang *et al.*, 2012b) achieved higher accuracy than most existing studies. Besides, the recently developed Dense V-net (Gibson *et al.*, 2018) also demonstrated good performance on human CT multi-organ segmentation but was not evaluated on mouse CT images. Therefore, we compare our TS-DSN with the SSM approach and the Dense V-net method. Table 1 reports the Dice scores of the three methods based on the same test dataset. It can be seen that both TS-DSN and SSM have similar Dice scores for the lungs, heart and right kidney. TS-DSN has higher scores for all the organs than SSM. Especially, the mean Dice improvements for the left kidney and spleen are as high as 18% and 26%, respectively. TS-DSN has higher scores for the lungs, heart, liver and kidneys than Dense V-net, the improvements vary from 1% (right kidney) to 9% (liver). The advantage of TS-DSN over Dense V-net should be attributed to the mechanism of deep supervision. As for the computation efficiency, the SSM method took 4~5min to process one image, which is much slower than the two deep learning methods (TS-DSN

and Dense V-net) which both took less than two seconds per 3D image.

TABLE 1. DICE SCORES OF THE PROPOSED METHOD AND THE STATE-OF-THE-ART METHODS.

	Lungs	Heart	Liver	Spleen	Left Kidney	Right Kidney
TS-DSN	**0.92±0.02**	**0.95±0.01**	**0.88±0.02**	0.57±0.18	**0.85±0.09**	**0.86±0.06**
SSM	0.90±0.02	0.90±0.02	0.79±0.03	0.45±0.18	0.70±0.12	0.80±0.07
Dense V-net	0.89±0.02	0.91±0.02	0.81±0.02	**0.60±0.08**	0.80±0.05	0.85±0.03

C. Real Low-Contrast Image Results

Since the proposed TS-DSN was trained using pseudo low-contrast data, it is necessary to test its performance on real low-contrast datasets. We randomly selected 10 low contrast images from the database of small animal images at the Crump Institute of UCLA. These images were acquired with the same device settings as the contrast-enhanced images mentioned in the method section, except that no contrast agent was used. Fig.6 displays the typical results of different organs. It can be seen that the TS-DSN trained on pseudo data yield visually acceptable results on real-world data. The predicted organ regions overlap with the core regions of the target organs, ensuring reliable ROI delineation for preclinical molecular uptake quantifications. It is worth noting that Fig.6(b) presents a mouse subject with a large tumor xenograft on the shoulder. Our method yielded reasonable segmentation results even when this tumor is presented.

To further illustrate the segmentation accuracy on real low contrast image, we took pre- and post-contrast scans for the same subject in the same posture. We use the TS-DSN method to segment the pre-contrast image (i.e. low contrast image) and superimpose the results on both the pre- and post-contrast images, as shown in Fig.6(d) and (e), respectively. Moreover, the images of fig.6(d) and (e) were acquired with a different CT scan configuration (0.13 mm voxel size, 50 kVp voltage, 20.2 mAs exposure, 1.0 mm aluminum filtration with a homemade micro-CT system) from the configurations of Fig.6(a)-(c) (0.20 mm voxel size, 70 kVp voltage, 500 mAs exposure, 2.0 mm aluminum filtration with Siemens MicroCAT II CT system). It can be observed that our method yielded reasonable prediction of the organ regions for the test images with different scanner configurations, the predicted regions using the pre-contrast image (Fig.6(d)) agree well with the organ regions in the post-contrast image (Fig.6(e)).

Discussion

Due to the trade-off between the radiation dose and image quality for CT acquisition, low-contrast images are widely used in preclinical research. We use a deeply supervised network to cope with the problem of imperfect organ contrasts. Unlike most deep segmentation networks whose objective functions penalize the differences between the network output and the annotation labels, the deeply supervised network use the labels to supervise the training of intermediate network layers. With this approach, the intermediate layers are forced to memorize the correct organ locations and shapes. When the input image presents poor organ contrast, the implanted memories will help the network to output plausible organ locations and shapes. As revealed by Fig.4 and Fig.5, the deeply supervised networks yielded more accurate results than the FCN without deep supervision, especially for the low-contrast soft-tissue organs like the spleen and kidneys.

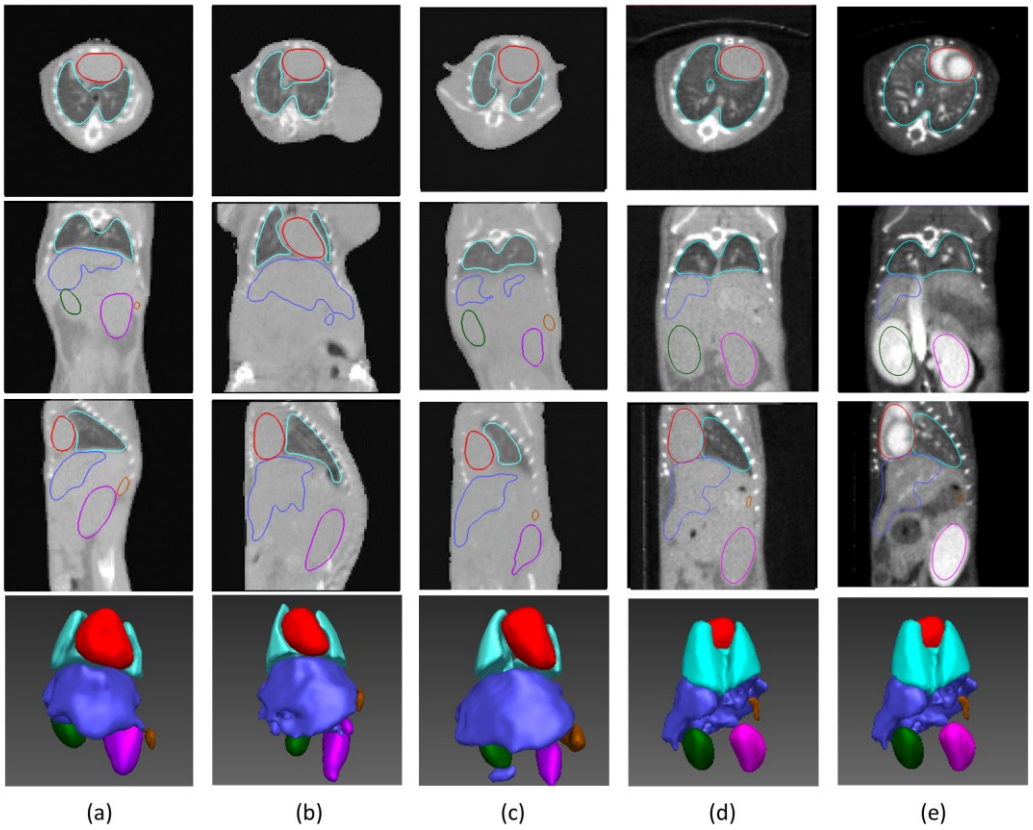

Fig.6 Organ prediction results of typical real low-contrast micro-CT images. The red, light blue, blue, brown, purple and green contours indicate the heart, lung, liver, spleen, left and right kidney, respectively. The first to the fourth row presents the results in transverse, sagittal, coronal and surface rendering views. Each column presents a different test subject. Columns (a)-(c) show the segmentation results of low-contrast images with the same acquisition configurations as the training data. Column (d) shows the segmentation result of the a low-contrast image acquired with different scanner configurations from (a)-(c). Column (e) superimposes the segmentation result of (d) on the post-contrast image of (d) to give a visual evaluation of the segmentation accuracy.

Another advantage of the deep supervision architecture is that it helps the training process to avoid gradient vanishing. Three-dimensional networks contain more layers and parameters than 2D networks. As the numbers of layers and parameters increase, the gradient disappearance becomes more serious(Glorot and Bengio, 2010). Moreover, the absence of high contrast image features makes it more difficult to find the optimal gradient direction. The deep supervision mechanism introduces anatomical priors to guide the training of intermediate layers. By performing loss calculation on multiple intermediate layers, the problem of gradient disappearance can be reduced effectively. This mechanism is especially beneficial to the training of shallow layers which is prone to the gradient vanishing issue. The deeply supervised network produces more accurate segmentation than the network without deep supervision. This result means that deep supervision leads to better optimization of the shallow network layers which is related to the local features of the resultant image.

During the methodology development, we found that the multi-organ prediction from a whole-torso image is less accurate than the single-organ prediction from a local image region. This finding

is reflected from the inaccurate organ locations and noisy false positive predictions (see the 3D-DSN results in Fig.4). Interestingly, a similar finding was also presented in a recent study for multi-vertebrae segmentation (Lessmann *et al.*, 2019). Therefore, we designed the second stage network for single organ prediction to improve accuracy. As reflected from Fig.5, The TS-DSN has larger median Dice and Precision values of Dice for most organs. The organ volume RC of TS-DSN is also better than the compared methods. The improvement on RC of organ volume is important for preclinical biodistribution studies which use the organ volumes for organ exposure dose calculation. The increase of precision is also good for the quantifications of molecular tracer uptake since these applications require that most of the organ ROI lies inside the ground truth region, i.e., the ratio of true positive voxels over all predicted voxels should be high.

As compared to FCN, DSN and the traditional atlas registration method, the most remarkable accuracy improvements happen to the spleen and kidneys. These organs have small sizes and unstable positions. The small organ sizes cause imbalances of positive and negative voxel samples, posing challenges to the neural network training. The unstable position makes it difficult for the atlas registration method to get a good match. With the deep supervision and two-stage architecture, our TS-DSN method significantly improves the accuracy of these problematic organs. Moreover, the computation speed of the proposed method is much faster than the SSM method (a few seconds vs. a few minutes), the improvement on computation efficiency is quite significant.

One limitation of this study is reflected from the real world low-contrast image results. Although the prediction results in Fig.6 look reasonably good, it is still not as accurate as the pseudo CT results in Fig.4. Fortunately, the predicted regions mostly lie inside the target organ region, therefore the applications such as molecular tracer quantification are not handicapped. Another concern may arise from the generalizability of the method. In real-world situations, the CT scanners and configurations often vary. In the experiment, we use different CT devices and different scanner configurations to acquire real low-contrast images. As Fig.6 shows, our method performed well on different acquisition parameters or scanners. However, our pseudo CT training images could not resemble all the unknown cases of test images acquired with various protocols, animal sizes and/or body postures. This generalization concern is also a major limitation of current deep learning approaches. For the next step, we will enlarge our training set to enhance the generalizability of our method. We will consider simulating different pseudo training images with our previously developed deformable mouse phantom (Wang *et al.*, 2012b) or the MOBY phantom (Segars *et al.*, 2004) using different scanner configurations. Alternatively, we may also acquire both non-contrast and contrast-enhanced images of the same mouse subject to avoid the generation of pseudo low-contrast images. To do so, we will need to take care of the inter-scan motion of the mice and use non-rigid image registration to establish the organ label correspondences between the non-contrast and contrast-enhanced images.

Conclusion

In this study, we developed a deep learning method to estimate the regions of soft-tissue organs in low-contrast mouse micro-CT images. The deep supervision mechanism is used to impose anatomical priors to the network training, and a rough-to-fine two-stage workflow is designed to improve the prediction accuracy. Our method outperforms the state-of-the-art atlas registration method in both accuracy and efficiency, it enhances the role of deep learning in preclinical image

analysis. For future work, we will investigate more advanced methods of using anatomical priors to further improve the organ prediction accuracy in low-contrast micro-CT images.

Acknowledgements

This study was funded by the general program of the National Natural Science Fund of China (No. 61571076), the youth program of the National Natural Science Fund of China (No. 81401475), the general program of Liaoning Science & Technology Project (No. 2015020040), the Science and Technology Innovation Fund of Dalian City (2018J12GX042), the Fundamental Research Funds for the Central Universities (DUT16RC(3)099), and the Xinghai Scholar Cultivating Funding of Dalian University of Technology (No. DUT15LN02).

Reference

Akselrod B A, Dafni H, Addadi Y, Biton I, Avni R, Brenner Y and Neeman M 2016 Multimodal Correlative Preclinical Whole Body Imaging and Segmentation *Sci. Rep.* **6** 27940

Badea C T, Johnston S M, Qi Y and Johnson G A 2011 4D micro-CT for cardiac and perfusion applications with view under sampling *Phys. Med. Biol.* **56** 3351-69

Baiker M, Milles J, Dijkstra J, Henning T D, Weber A W, Que I, Kaijzel E L, Löwik C W G M, Reiber J H C and Lelieveldt B P F 2010 Atlas-based whole-body segmentation of mice from low-contrast Micro-CT data *Med. Image Anal.* **14** 723-37

Baiker M, Staring M, Löwik C W G M, Reiber J H C and Lelieveldt B P F 2011 Automated Registration of Whole-Body Follow-Up MicroCT Data of Mice *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention* pp 516-23

Bottou L 2010 Large-Scale Machine Learning with Stochastic Gradient Descent *Int. Conf. on Proceedings of COMPSTAT'2010* pp 177-86

Burk L M, Lee Y Z, Wait J M, Lu J and Zhou O Z 2012 Non-contact respiration monitoring for in-vivo murine micro computed tomography: characterization and imaging applications *Phys. Med. Biol.* **57** 5749-63

Cheng L and Qi J 2010 Segmentation of mouse dynamic PET images using a multiphase level set method *Phys. Med. Biol.* **55** 6549-69

Dou Q, Yu L, Chen H, Jin Y, Yang X, Qin J and Heng P 2017 3D deeply supervised network for automated segmentation of volumetric medical images *Med. Image Anal.* **41** 40-54

Gibson E, Giganti F, Hu Y, Bonmati E, Bandula S, Gurusamy K, Davidson B, Pereira S P, Clarkson M J and Barratt D C 2018 Automatic multi-organ segmentation on abdominal CT with dense v-networks *IEEE Trans. Med. Imag.* **37** 1822-34

Glorot X and Bengio Y 2010 Understanding the difficulty of training deep feedforward neural networks *Int. Conf. on Proceedings of the thirteenth international conference on artificial intelligence and statistics* pp 249-56

Guo X, Johnston S M, Qi Y, Johnson G A and Badea C T 2011 4D micro-CT using fast prospective gating *Phys. Med. Biol.* **57** 257-71

Holbrook M, Clark D P and Badea C T 2018 Low-dose 4D cardiac imaging in small animals using

dual source micro-CT *Phys. Med. Biol.* **63** 025009

Joshi A A, Chaudhari A J, Li C, Dutta J, Cherry S R, Shattuck D W, Toga A W and Leahy R M 2010 DigiWarp: a method for deformable mouse atlas warping to surface topographic data *Phys. Med. Biol.* **55** 6197

Kagadis G C, Loudos G, Katsanos K, Langer S G and Nikiforidis G C 2010 In vivo small animal imaging: current status and future prospects *Med. Phys.* **37** 6421-42

Kearney V P, Chan J W, Wang T, Perry A, Yom S S and Solberg T D 2019 Attention-enabled 3D boosted convolutional neural networks for semantic CT segmentation using deep supervision *Phys. Med. Biol.* **64** 5001

Khmelinskii A, Baiker M, Kaijzel E L, Chen J, Reiber J H C and Lelieveldt B P F 2011 Articulated Whole-Body Atlases for Small Animal Image Analysis: Construction and Applications *Mol. Imag. Biol.* **13** 898-910

Klose A D and Paragas N 2018 Automated quantification of bioluminescence images *Nat. Commun.* **9** 4262

Kockelkorn T T J P, Schaefer P C M, Bozovic G, Muñoz B A, van Rikxoort E M, Brown M S, de Jong P A, Viergever M A and van Ginneken B 2014 Interactive lung segmentation in abnormal human and animal chest CT scans *Med. Phys.* **41** 081915

Kok P, Baiker M, Hendriks E A, Post F H, Dijkstra J, Lowik C W G M, Lelieveldt B P F and Botha C P 2010 Articulated Planar Reformation for Change Visualization in Small Animal Imaging *IEEE Trans. Vis. Comput. Gr.* **16** 1396-404

Krizhevsky A, Sutskever I and Hinton G 2012 ImageNet Classification with Deep Convolutional Neural Networks *Int. Conf. on Neural Information Processing Systems* pp 1097-1105

Lessmann N, van Ginneken B, de Jong P A and Išgum I 2019 Iterative fully convolutional neural networks for automatic vertebra segmentation and identification *Med. Image Anal.* **53** 142-55

Li X, Chen H, Qi X, Dou Q, Fu C and Heng P 2018 H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes *IEEE Trans. Med. Imag.* **37** 2663-74

Litjens G, Kooi T, Bejnordi B E, Setio A A A, Ciompi F, Ghafoorian M, van der Laak J A W M, van Ginneken B and Sánchez C I 2017 A survey on deep learning in medical image analysis *Med. Image Anal.* **42** 60-88

Maroy R, Boisgard R, Comtat C, Frouin V, Cathier P, Duchesnay E, Dolle F, Nielsen P E, Trebossen R and Tavitian B 2008 Segmentation of Rodent Whole-Body Dynamic PET Images: An Unsupervised Method Based on Voxel Dynamics *IEEE Trans. Med. Imag.* **27** 342-54

Maroy R, Boisgard R, Comtat C, Jego B, Fontyn Y, Jan S, Dubois A, Trébossen R and Tavitian B 2010 Quantitative organ time activity curve extraction from rodent PET images without anatomical prior *Med. Phys.* **37** 1507-17

Milletari F, Navab N and Ahmadi S 2016 V-net: Fully convolutional neural networks for volumetric medical image segmentation *Int. Conf. on 2016 Fourth International Conference on 3D Vision (3DV)* pp 565-71

Ren S, Hu H, Li G, Cao X, Zhu S, Chen X and Liang J 2016 Multi-atlas registration and adaptive hexahedral voxel discretization for fast bioluminescence tomography *Biomed. Opt. Express* **7** 1549-60

Ronneberger O, Fischer P and Brox T 2015 U-net: Convolutional networks for biomedical image segmentation *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*

pp 234-41

Segars W P, Tsui B M, Frey E C, Johnson G A and Berr S S 2004 Development of a 4-D digital mouse phantom for molecular imaging research *Mol. Imag. Biol.* **6** 149-59

Shelhamer E, Long J and Darrell T 2017 Fully Convolutional Networks for Semantic Segmentation *IEEE Trans. Pattern. Anal.* **39** 640-51

Snoeks T J A, Baiker M, Kaijzel E L, Lelieveldt B P F and Löwik C W G M 2012 CT-based handling and analysis of preclinical multimodality imaging data of bone metastases *BoneKEy reports* **1** 79

Snoeks T J A, Khmelinskii A, Lelieveldt B P F, Kaijzel E L and Löwik C W G M 2011 Optical advances in skeletal imaging applied to bone metastases *Bone* **48** 106-14

Stout D B, Chatziioannou A F, Lawson T P, Silverman R W, Gambhir S S and Phelps M E 2005 Small Animal Imaging Center Design: The Facility at the UCLA Crump Institute for Molecular Imaging *Mol. Imag. Biol.* **7** 393-402

Wang H, Stout D B, Taschereau R, Gu Z, Vu N T, Prout D L and Chatziioannou A F 2012a MARS: a mouse atlas registration system based on a planar x-ray projector and an optical camera *Phys. Med. Biol* **57** 6063-77

Wang H, Stout D B and Chatziioannou A F 2012b Estimation of Mouse Organ Locations Through Registration of a Statistical Mouse Atlas With Micro-CT Images *IEEE Trans. Med. Imag.* **31** 88-102

Welch D, Harken A, Randers-Pehrson G and Brenner D 2015 Construction of mouse phantoms from segmented CT scan data for radiation dosimetry studies *Phys. Med. Biol.* **60** 3589

Yan D, Zhang Z, Luo Q and Yang X 2017 A Novel Mouse Segmentation Method Based on Dynamic Contrast Enhanced Micro-CT Images *Plos. One.* **12** e0169424

Zhang B, Gao F, Wang M, Cao X, Liu F, Wang X, Luo J, Wang G and Bai J 2014 In vivo tomographic imaging of lung colonization of tumour in mouse with simultaneous fluorescence and X-ray CT *J. Biophotonics.* **7** 110-6