Performance evaluation of 2D and 3D deep learning approaches for automatic segmentation of multiple organs on CT images

Xiangrong Zhou*[a], Kazuma Yamada [a], Takuya Kojima [a], Ryosuke Takayama [a], Song Wang [b], Xinxin Zhou [c], Takeshi Hara [a], and Hiroshi Fujita [a]

[a] Department of Electrical, Electronic & Computer Engineering, Faculty of Engineering, Gifu University, Gifu-shi, 501-1193, Japan
[b] Department of Computer Science and Engineering, University of South Carolina, Columbia, SC 29208, USA
[c] Faculty of Information and Media Studies, Nagoya Bunri University, 365 Maeda, Inazawa-cho, Inazawa-shi, 492-8520, Japan
*zxr@fjt.info.gifu-u.ac.jp

ABSTRACT

The purpose of this study is to evaluate and compare the performance of modern deep learning techniques for automatically recognizing and segmenting multiple organ regions in 3D CT images. CT image segmentation is an important and challenging task in medical image analysis. Deep learning approaches have demonstrated their capability for scene recognition and semantic segmentation on images of nature and have been used to address segmentation problems of medical images. Although several works have showed promising results, there is no comprehensive evaluation of segmentation performance of deep learning approaches on segmenting multiple organs in different portions of CT scans. In this study, we evaluated and compared the segmentation performance of two different deep learning approaches that used 2D- and 3D deep convolutional neural networks (CNN) without- and with a pre-processing step. A conventional approach that represents the state-of-the-art approach to CT image segmentation without deep learning was also compared. A dataset containing 240 CT images scanned on different portions of human bodies was used. A maximum of 17 types of organ regions in each CT scan were segmented automatically and compared to the human annotations by using ratio of intersection over union (IU) as the criterion. The experimental results demonstrated that the IUs of the segmentation results had a mean value of 79% and 67% by averaging 17 types of organs that were segmented by a 3D- and 2D deep CNN, respectively. All results using the deep learning approaches showed better accuracy and robustness than the conventional segmentation method that used probabilistic atlas and graph-cut methods. Thus, the effectiveness and usefulness of deep learning approaches were demonstrated for solving multiple organs segmentation problem in 3D CT images.

Keywords: 3D CT images, anatomical structure segmentation, deep learning, 3D convolutional neural networks.

1. INTRODUCTION

Understanding the anatomical structures of different patients by using 3D high-resolution CT images is an important task in medical image analysis, which helps to support image diagnosis, surgery planning, and radiation therapy [1]. Fully automatic segmentation of the anatomical structures of multiple organ regions can reduce the burden of manual annotation and enhance efficiency while obtaining further diagnostic insight on CT images. Although many research works on medical image segmentation have been presented [2], fully automatic segmentation of 3D CT images is still a challenge because of the poor contrast in CT images, high variability on image appearance as well as individual differences between patients, and partial visibility of the anatomical structures present in different CT scans. For example, 3D CT images used in clinical medicine are generally scanned under different image resolutions and qualities, and they cover various portions (chest, abdominal, or pelvic) that contain different types of organs. Modeling such a wide variety of image appearances for predicting anatomical structures within unseen CT images from different patients is challenging.

Medical Imaging 2018: Computer-Aided Diagnosis, edited by Nicholas Petrick, Kensaku Mori, Proc. of SPIE Vol. 10575, 105752C · © 2018 SPIE · CCC code: 1605-7422/18/$18 · doi: 10.1117/12.2295178

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 4/1/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

The conventional approach for modeling CT image appearance aims to represent anatomical structures of organs based on several hand-crafted features on contour shape, image intensity, and spatial relations of different organs, by using a number of patient cases [3-8]. The image segmentation process was designed as a pipeline that uses the model information to guide or constrain the actions of a number of hand-crafted processing procedures [3-8]. To improve the accuracy and robustness of CT image segmentation, the model construction should be able to handle a larger variance within ambiguous CT image appearances, shapes, and relations of different organs by using more patient cases. It was difficult to achieve this goal by explicitly defining and considering the rules from human experience. Therefore, data-driven approaches, such as deep learning, can be used to solve this problem by combining model construction with image segmentation, and by learning image features and model parameters from a large CT image dataset directly.

Recently, several research works reported success in organ segmentation from 3D CT images by using deep learning approaches [9-13]. However, most of these works only considered segmentation results for a few specific organ types (e.g. liver and brain) within a pre-defined portion of a CT image (e.g. chest and head). These kinds of organ segmentations had also been reported earlier, by using conventional methods. To our knowledge, there is still no comprehensive investigation of segmentation performance by using different deep learning approaches for all types of organs on different portions of CT scans. This issue has also not been addressed by conventional methods. The goal of this study is to evaluate and compare the segmentation performance by using 2D- and 3D deep learning approaches as well as a conventional method while segmenting a wide range of organ types based on the same dataset including 240 CT scans on different portions of human bodies.

2. METHODS

2.1 Outline

Two approaches (a) 2D deep learning with post-processing (3D label fusion based on majority voting), and (b) 3D deep learning with pre-processing (organ localization) were proposed for our CT image segmentation tasks. A convolutional neural network (CNN) was used as the core of the segmentation process. We used a convolutional layer as the basic unit and did not use fully connected layers. Two modern CNNs, a fully convolutional network (FCN) [14] and residual network (ResNet) [15] were used as the templates to construct our network structures. We used pixel-based labeling loss and region-based coincidence between the segmentation result and the human annotation as the loss function to train the 2D- and 3D deep CNN, respectively. Two optimization methods, stochastic gradient descent (SGD) and ADAM [16], were used for learning the network parameters. Details are provided in the following sections.

2.2 2D deep learning approach

A 2D deep learning method presented in our previous papers [17-20] was used in this work for multiple organ segmentation in 3D CT scans. This method modeled the CT image segmentation as “multiple 2D proposals with 3D integration” and took inspiration from the fact that radiologists interpret a CT scan from many 2D slices and reconstruct 3D anatomical structures mentally [20]. A 2D fully convolutional network (FCN) [14] extended from a VGG-16 network was used for generating the multiple 2D proposals of organ regions from three body directions on each CT scan independently, and an organ–label fusion was carried out on each voxel in 3D image space by using a majority voting method. In the training stage, we decompose the 3D CT scans with annotated organ regions into several 2D slices in three orthogonal body directions and treat each 2D slice with an annotation as a training sample. The sum of the pixel-wise classification error on 2D slices was used as the loss function during the training process. ADAM [16] was used as the optimization method. In the testing stage, we applied the trained FCN to segment partial organ regions on each 2D slice along three body directions independently, and then integrated the segmented 2D partial organ regions into 3D image space by using a voxel-wise label fusion to obtain the final 3D segmentation result. The implementation details of the processing steps and network structures can be found in [20].

2.3 3D deep learning approach

We proposed a novel 3D deep learning approach for multiple organ segmentation in this work. This approach accomplished organ segmentation through two steps, as shown in Fig. 1. We decoupled the organ detection and segmentation functions, and modeled the multiple organ segmentation processes as “detecting the bounding box of different organs first, and then segmenting the organ region within each 3D bounding box”. The organ detection module used in this work was presented in our previous work [21]. This method used a conventional machine-learning approach based on window sliding and pattern matching in Haar-like and LBP image feature spaces through the entire CT image

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 4/1/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

space to detect each organ region and return the coordinates of a bounding box as the organ location. Details of our organ detection method can be found in [21]. The organ segmentation module used a 3D deep CNN structure derived from VNet [11]. This 3D deep CNN was constructed using a number of 3D convolutional layers based on ResNet [15] modules with skip connections to combine the information in shallow and deep layers. The dice index between the segmented region and ground truth was used as the loss function, and the SGD algorithm was used for optimizing the network parameters. In the training stage, we trained organ localization and segmentation separately. We generated the ground truth of the bounding box for each organ type from the manual annotations in each CT scan and trained a set of organ-specific ad-hoc classifiers that detect the bounding box for each organ type in CT scans. Next, we cropped the volume-of-interest (VOI) regions for each organ type based on the bounding box and resized the VOIs to the same image size (128×128×128 voxels) as the input image of the segmentation model. The annotation of each organ region was also cropped and resized in the same manner to retain the correspondence of the spatial location of each voxel between the CT image and manual annotation data. The pairs of VOI and annotation of all organ regions in CT scans were used as the samples for training the 3D deep CNN. In the testing stage, our method detected the bounding boxes of all the organ regions within a CT scan and segmented target organ regions within the range of bounding boxes.

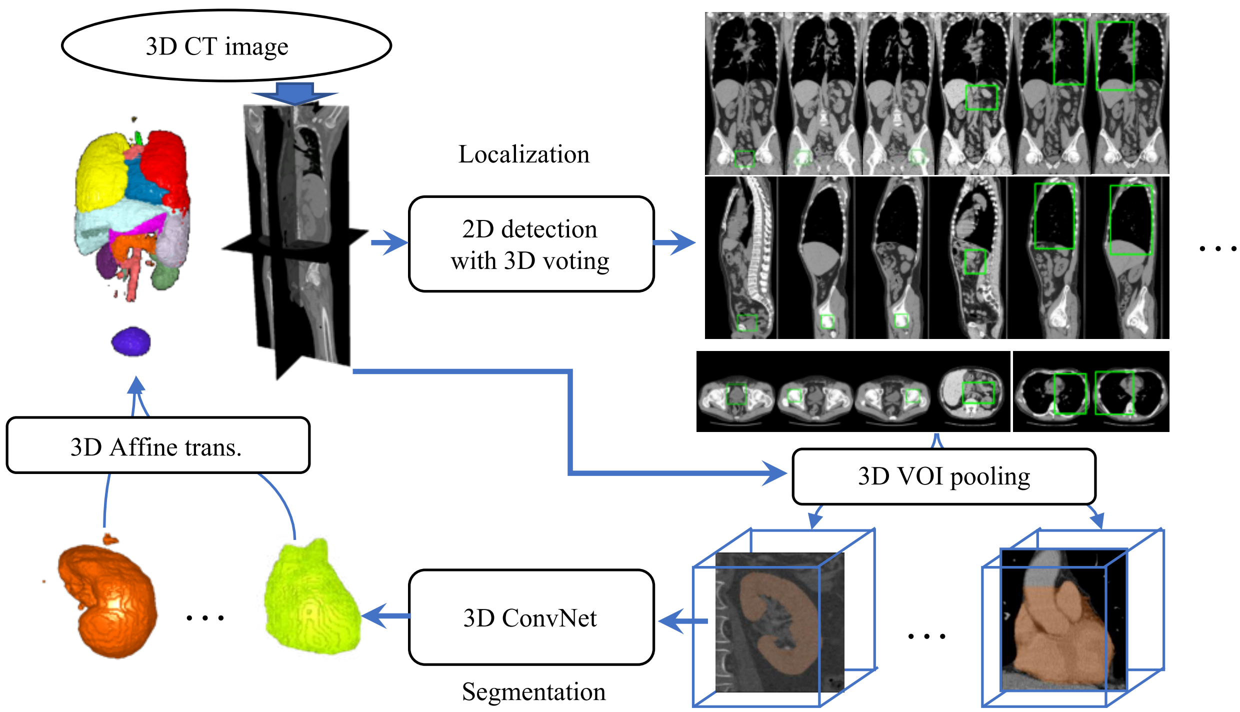

Fig.1. The process flow of multiple organ segmentations by combining organ localization and organ segmentation from CT images

2.4 Conventional image segmentation approach

A conventional method presented by Okada et al. [7] was used for evaluating the performance of the deep learning approaches. This conventional method modeled the intensity (CT number) and shape information of each target organ by using statistic shape modeling and probabilistic atlas, and combined intra-organ information with an inter-organ relationship to guide the segmentation process. This approach demonstrated state-of-the-art multiple organ segmentation performances among conventional methods in CT image segmentation and reported the detailed evaluation results based on a large CT image dataset [7].

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 4/1/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

3. EXPERIMENT AND RESULTS

The performance evaluations were based on a shared dataset produced by a research project titled "Computational Anatomy [22]". This dataset consists of 240 CT scans from 200 patients for the diagnosis of diverse types of lesions, obtained at Tokushima University Hospital. These CT images were scanned for different portions (89 torso, 17 chest, 114 abdomen, and 20 abdomen-with-pelvis scans) by a multi-slice CT scanner (Aquilion from Toshiba Medical Systems Corporation) and reconstructed by different protocols and kernel functions, leading to different image resolutions (distributed between 0.625–1.148 mm with 1.0 mm slice thickness) and different image quality (specialized for lung or abdominal observations). Contrast media enhancements were applied in 155 CT scans. The anatomical ground truth (a maximum of 19 labels that show 17 major organ types and 2 interesting regions inside the human body) of 240 CT scans was manually annotated [23] and distributed within the dataset. In this work, we only consider the 17 organ regions as the segmentation targets for performance evaluations.

We directly used the organ localization module that was generated based on another CT image dataset in our previous work [21] without any fine-tuning to adapt to above CT image database. Because the segmentation approaches used in this work were based on machine learning, we used a four-fold cross-validation for performance evaluation. We used 75% CT scans of the dataset for training parameters of the different network structures and tested the trained networks on the remaining 25% CT scans. CT scans from the same patient were only used in the same training or testing stage. The accuracy of the segmentation was evaluated for each organ type and each CT scan. The intersection over union (IU) (also known as the Jaccard similarity coefficient) between the segmentation result and ground truth was used as the evaluation measures. Because a CT scan may contain different organ regions, we used a comprehensive evaluation of multiple organ segmentation results for all CT scans by considering the variance of the organ number and volume. The measures (mean voxel accuracy, mean IU, and frequency-weighted IU) that are commonly used in semantic segmentation and scene parsing [14] were employed in this study for the evaluations. Let n_{ij} be the number of pixels in target i classified as target j, n_{cl} be the total number of different targets in a CT case, and $t_i = \sum_j n_{ij}$ be the total number of pixels in target i. These measures are defined as:

- Mean voxel accuracy: $(\sum_i n_{ii}/t_i)/\, n_{cl}$ (1)
- Mean IU: $(\sum_i n_{ii}/\left(t_i + \sum_j n_{ji} - n_{ii}\right))/\, n_{cl}$ (2)
- Frequency-weighted IU: $(\sum_k t_k)^{-1} \sum_i t_i n_{ii}/\left(t_i + \sum_j n_{ji} - n_{ii}\right)$ (3)

An example of the result of multiple organ segmentation in a testing CT scan is shown in Fig. 2 by using a 3D volume rendering method. The evaluation results of deep CNNs demonstrated that the average segmentation accuracy of 17 types of organ over the 240 (4×60) test CT scans was 67% (2D CNN) and 78.8 % (3D CNN) in terms of the mean IUs, 84.9% (2D CNN) and 89 % (3D CNN) in terms of the frequency-weighted IUs, and 86.1% (2D CNN) and 88.4% (3D CNN) in terms of the mean voxel accuracy.

We tested the conventional method presented by Okada et al. [7] based on the same training and testing CT images by using their system for both model construction and organ segmentation. Because this conventional method was only applied to abdominal CT scans, our experiment was limited to seven organs types including liver, gallbladder, left and right kidneys, spleen, pancreas, and veins. The experimental results showed that the average segmentation accuracy of seven types of organs was 59% in terms of the mean IUs, which was lower than the results of the 2D and 3D CNN approaches.

The deep CNNs used in this work were implemented based on Caffe deep learning framework [24]. The computing time for training an 3D deep CNN was about three days. This was longer than the time taken for training a 2D deep CNN network, which is approximately one and a half days based on a GPU (NVIDIA Quadro P6000 with 24 GB of memory). The organ segmentation for one testing 3D CT scan with a 512×512×221 matrix took approximately 3 min by using a trained deep CNN. The efficiency in terms of system development and improvement of 2D- and 3D deep CNN approaches was better than that of the conventional method that used a CPU based approach.

4. DISCUSSION AND CONCLUSION

We evaluated and compared the performances of CT image segmentation by using 2D- and 3D deep learning approaches as well as a conventional method. The experimental results for 17 organ types on 240 CT scans showed that (1) Deep

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 4/1/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

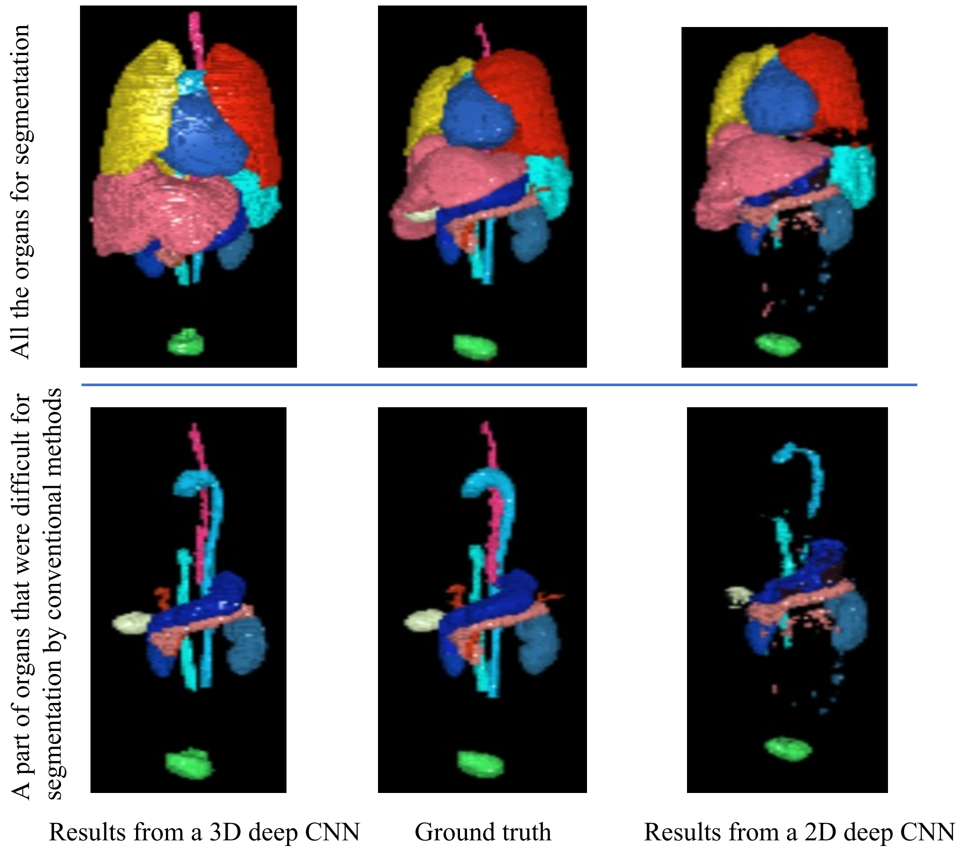

Fig.2. 3D views of the anatomical structure segmentation results in a CT scan by using 3D (left side) and 2D (right side) deep CNN approaches with the human annotation (middle column) as the ground truth.

learning approaches demonstrate better performance (especially in terms of robustness and generality) than the conventional method on CT image segmentation tasks. (2) A 2D convolutional network can achieve multiple organ segmentation on 3D CT images successfully. (3) A 3D convolutional network showed the best accuracy of segmentation in this experiment, especially for small volume organ types having a tube- or line-shape. In conclusion, the deep learning approach can be expected to solve 3D CT image segmentation problems with better accuracy and efficiency than the conventional approach.

ACKNOWLEDGEMENTS

The authors would like to thank all the members of the Fujita Laboratory in Gifu University, for their collaboration. We especially thank Dr. Okada of Tsukuba University for providing binary code and Dr. Roth of Nagoya University for giving advice on 3D deep CNN. We would like to thank all the members of the “Computational Anatomy” research project [22], especially Dr. Ueno of Tokushima University, for providing the CT image database. This research was supported in part by a Grant-in-Aid for Scientific Research on Innovative Areas (Grant No. 26108005), and in part by a Grant-in-Aid for Scientific Research (C26330134), MEXT, Japan.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 4/1/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

REFERENCES

1. K. Doi, "Computer-aided diagnosis in medical imaging: Historical review, current status, and future potential," Comput. Med. Imaging Graph., 31, 198–211 (2007).
2. D. L. Pham, C. Xu, J. L. Prince, “Current methods in medical image segmentation,” Biomed. Eng., 2, 315–33 (2000).
3. T. Heimann, H. P. Meinzer, “Statistical shape models for 3D medical image segmentation: A review,” Med. Image Anal., 13 (4), 543–563 (2009).
4. Y. Xu, C. Xu, X. Kuang, H. Wang, E. I.C. Chang, W. Huang, Y. Fan, “3D-SIFT-Flow for atlas-based CT liver image segmentation,” Med. Phys. 43 (5), 2229–2241 (2016).
5. N. Lay, N. Birkbeck, J. Zhang, S. K. Zhou, “Rapid multi-organ segmentation using context integration and discriminative models,” Proc. IPMI, 7917, 450–462 (2013).
6. R. Wolz, C. Chu, K. Misawa, M. Fujiwara, K. Mori, D. Rueckert, “Automated abdominal multi-organ segmentation with subject-specific atlas generation,” IEEE Trans. Med. Imaging, 32 (9), 1723–1730 (2013).
7. T. Okada, M.G. Linguraru, M. Hori, R.M. Summers, N. Tomiyama, Y. Sato, “Abdominal multi-organ segmentation from CT images using conditional shape-location and unsupervised intensity priors,” Med. Image Anal., 26 (1), 1–18 (2015).
8. K. Sun, J. K. Udupa, D. Odhner, Y. Tong, L. Zhao, D. A. Torigian, “Automatic thoracic anatomy segmentation on CT images using hierarchical fuzzy models and registration,” Med. Phys., 43 (4), 1882–1896 (2016).
9. H.C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R. M. Summers, “Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning,” IEEE Tran. Med. Imaging, 35 (5), 1285–98 (2016).
10. O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI), 234-241 (2015).
11. F. Milletari, N. Navab, S.-A. Ahmadi, “V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation,” arXiv:1606.04797 (2016).
12. H. R. Roth, A. Farag, L. Lu, E. B. Turkbey, R. M. Summers, “Deep convolutional networks for pancreas segmentation in CT imaging,” Proc. SPIE, 9413, 94131G-1-94131G-8 (2015).
13. K. H. Cha, L. Hadjiiski, R. K. Samala, H.P. Chan, E. M. Caoili, R. H. Cohan, “Urinary bladder segmentation in CT urography using deep-learning convolutional neural network and level sets,” Med. Phys., 43 (4), 1882–1896 (2016).
14. J. Long, E. Shelhamer, T. Darrell, “Fully convolutional networks for semantic segmentation,” Proc. CVPR, 3431-3440 (2015).
15. K.He, X.Zhang, S.Ren, and J.Sun, “Deep Residual Learning for Image Recognition,”, Proc. CVPR, 770-778 (2016).
16. D. P. Kingma, J. L. Ba, “ADAM: A method for stochastic optimization,” Proc. ICLR 2015 (2015).
17. X.Zhou, T.Ito, R.Takayama, S.Wang, T.Hara, and H.Fujita, “Three-dimensional CT image segmentation by combining 2D fully convolutional network with 3D majority voting,” Proc. of Workshop on the 2nd Deep Learning in Medical Image Analysis (DLMIA) in MICCAI 2016, LNCS 10008, 111-120 (2016).
18. X.Zhou, R.Takayama, S.Wang, X.Zhou, T.Hara, and H.Fujita, “Automated segmentation of 3D anatomical structures on CT images by using a deep convolutional network based on end-to-end learning approach,” Proc. SPIE Medical Imaging 2017: Image Processing, vol. 10133, 1013324 (2017).
19. X.Zhou, T.Ito, R.Takayama, S.Wang, T.Hara, and H.Fujita, “First trial and evaluation of anatomical structure segmentations in 3D CT images based only on deep learning,” Brief Article, Medical Image and Information Sciences, 33 (3), 69-74 (2016).
20. X.Zhou, R.Takayama, S.Wang, T.Hara, and H.Fujita, “Deep learning of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method, Medical Physics,” 44 (10), 5221-5233 (2017).
21. X.Zhou, S.Morita, X.X.Zhou, H.Chen, T.Hara, R.Yokoyama, M.Kanematsu, H.Hoshi, and H.Fujita, “Automatic anatomy partitioning of the torso region on CT images by using multiple organ localizations with a group-wise calibration technique,” Proc. of SPIE Medical Imaging 2015: Computer-Aided Diagnosis, edited by L.M.Hadjiiski and G.D.Tourassi, Vol.9414, 94143K-1 – 94143K-6 (2015).
22. http://www.comp-anatomy.org/wiki/
23. H. Watanabe, A. Shimizu, J. Ueno, S. Umetsu, S. Nawano, H. Kobatake, “Semi-automated organ segmentation using 3-dimensional medical imagery through sparse representation,” Trans. of Jap. Society for Med. and Bio. Eng., 51(5), 300–312 (2013).
24. http://caffe.berkeleyvision.org

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 4/1/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use