# Automated segmentation of 3D anatomical structures on CT images by using a deep convolutional network based on end-to-end learning approach

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

[a] Department of Intelligent Image Information, Division of Regeneration and Advanced Medical Sciences, Graduate School of Medicine, Gifu University, Gifu-shi, 501-1194, Japan
[b] Department of Computer Science and Engineering, University of South Carolina, Columbia, SC 29208, USA
[c] School of Information Culture, Nagoya Bunri University, 365 Maeda, Inazawa-cho, Inazawa-shi, 492-8520, Japan
*zxr@fjt.info.gifu-u.ac.jp

## ABSTRACT

We have proposed an end-to-end learning approach that trained a deep convolutional neural network (CNN) for automatic CT image segmentation, which accomplished a voxel-wised multiple classification to directly map each voxel on 3D CT images to an anatomical label automatically. The novelties of our proposed method were (1) transforming the anatomical structures segmentation on 3D CT images into a majority voting of the results of 2D semantic image segmentation on a number of 2D-slices from different image orientations, and (2) using "convolution" and "de-convolution" networks to achieve the conventional "coarse recognition" and "fine extraction" functions which were integrated into a compact all-in-one deep CNN for CT image segmentation. The advantage comparing to previous works was its capability to accomplish real-time image segmentations on 2D slices of arbitrary CT-scan-range (e.g. body, chest, abdomen) and produced correspondingly-sized output. In this paper, we propose an improvement of our proposed approach by adding an organ localization module to limit CT image range for training and testing deep CNNs. A database consisting of 240 3D CT scans and a human annotated ground truth was used for training (228 cases) and testing (the remaining 12 cases). We applied the improved method to segment pancreas and left kidney regions, respectively. The preliminary results showed that the accuracies of the segmentation results were improved significantly (pancreas was 34% and kidney was 8% increased in Jaccard index from our previous results). The effectiveness and usefulness of proposed improvement for CT image segmentations were confirmed.

**Keywords:** 3D CT images, anatomical structures segmentation, deep learning, convolutional neural network

## 1. INTRODUCTION

Fully automatic image segmentation is a fundamental step for understanding anatomical structures on 3D CT scans by converting a physical image signal to useful abstractions [1]. Conventional approaches for CT image segmentation are usually constituted by transferring the human consideration and experience directly to a computer-based processing pipeline including a number of hand-crafted image processing algorithms and image features [2]. Although many mathematic models have been introduced recently for supporting image segmentation to reflect more human knowledge and learn the parameters from real images instead of human experience [3], it still attempts to emulate limited human's rules or operations that are insufficient for computer-based image segmentation [4-10]. In order to further improve the accuracy and robustness of the automatic CT image segmentation, we have to handle a large variance of ambiguous image appearances, shapes, and relations of anatomical structures. It is difficult to achieve this goal by explicitly defining and considering human experience and rules. Data-driven approaches, such as deep convolutional neural networks

Medical Imaging 2017: Image Processing, edited by Martin A. Styner, Elsa D. Angelini, Proc. of SPIE
Vol. 10133, 1013324 · © 2017 SPIE · CCC code: 1605-7422/17/$18 · doi: 10.1117/12.2254201

(CNNs), provide a new way to solve this problem by automatically learning image features and model parameters from a large dataset with a simple and unified layer structure.

Recently, several research works were reported to apply deep CNNs to medical image analysis. Many of them use the deep CNNs to lesion detection or classifications. Shin et al. evaluated different structures of deep CNNs for thoraco-abdominal lymph node detection and interstitial lung disease classification [11]. Ciompi et al. used deep CNNs to detect pulmonary peri-fissural nodules on CT images [12]. These works usually divide CT images into a number of small 2D/3D patches at different locations and classify these patches into a number of pre-defined categories. The deep CNNs are used to learn a set of optimized image features (sometimes combined with a classifier) to achieve the best classification rate for the image patches. Similarly, deep CNNs were also embedded into conventional image analysis pipelines such as lesion classification [13], lesion detection [14] and image segmentation processes to reduce the false positives in the results or predict the likelihoods of image patches [15-17]. However, anatomical segmentations on CT images over a wide range of human body is still challenging due to the similar image appearance between small image patches and difficulty to ensure global spatial consistency of patches on different CT cases.

In our previous work [18, 19], we have proposed an approach that naturally imitates brain mechanism of CT image interpretation by radiologists for image segmentation based on deep CNNs. This approach modeled the CT image segmentation as "multiple 2D proposals with a 3D integration" just like the radiologist to interpret a CT scan on many 2D slices and reconstruct 3D anatomical structures in the brain. The method used in this approach was a fully convolutional network without using any conventional image processing algorithms such as smoothing and filtering, and did not require any other pre- or post-processing. In addition, the proposed approach used one simple network to segment multiple organs simultaneously, adaptive to 3D or 2D images of arbitrary CT-scan-range (e.g. body, chest, abdomen) with a high computational efficiency by using parallel computations on GPU. Although, our previous works showed possibility of a deep CNN to segment multiple organs simultaneously on 3D CT images (the lasted result of our previous method has been submitted in [20]), there is still a room for improvement on the segmentation accuracy especially for some special organ types such as pancreas on CT images.

In this paper, we propose an improvement for our previous method that initially reported in [18, 19] by restricting the range of inputted image based on organ localizations [21]. We also compare the performance of our deep CNNs [19] before and after improvements by using the same CT dataset. In following sections, we firstly introduce the basic idea of our previous method that is initially reported in [18, 19], and then describe our improvement by adding the information of bounding box for a target organ for both training and testing. Finally, we show the experimental results for segmenting two kinds of organ types (pancreas and left kidney) with a conclusion.

## 2. METHODS

### 2.1 Outline

The process flow of the proposed scheme is shown in Fig.1. The input is a 3D CT case (it can also handle a 2D case by treating it as a special 3D case with only one slice) and the output is an annotated image of the same size and dimension on which a pre-defined anatomical label is associated to each pixel respectively. Our method firstly passed the CT image to an organ localization module to output multiple organ locations (a number of bounding boxes), and then accomplish the segmentation task by deciding the target organ regions within the bounding boxes. Both localization process and segmentation process are based on our idea, which conclude three steps: (1) sampling 2D sections on a 3D CT case from multiple directions, (2) detecting the organ location or segmenting organ regions on each 2D section individually, and (3) stacking the 2D results back to 3D with redundancy. Both localization and segmentation methods were based on our previous works [18-20], which use machine-learning approaches.

### 2.2 Organ localization [21]

The organ localization module includes two major processing steps: (1) individual organ localization that decides the bounding box of a target organ based on window sliding and pattern matching in Haar-like and LBP feature spaces; (2) group-wise calibration and correction based on general Hough voting of multiple organ locations [21]. The first step focuses on identifying local texture similarity of a target organ and output likelihood values of the target organ location on each pixel of CT images. The second step emphasizes the validation of the spatial relationships of the detected multiple organ locations constrained by human anatomy. The final locations of organs on CT images are determined by balancing the local texture similarity (likelihood value) and global spatial relationship of the multiple organ locations

(prior probability validated by anatomical knowledge) obtained in Step 1 and 2, respectively. Details of our organ localization method can be referred in [21, 22].

### 2.3 Organ segmentations [18-20]

The core part of our segmentation method is a deep CNN [19-20] that is used for the anatomical segmentation of 2D sections. This deep CNN is trained based on a set of CT cases with human annotations (label map) as ground truth. For the training images, all the 2D CT sections, together with their label maps, along the three body orientations are shuffled, and used to train a 2D deep CNN [22]. The training process repeats the feed-forward computation and back-propagation to minimize the loss function, which is defined as the sum of the pixel-wise losses between the network prediction and the label map annotated by the human experts. The gradients of the loss are propagated from the end to the start of the network, and the method for stochastic optimization is used to refine the parameters of each layer. In CT image segmentation, the trained CNN is applied to each 2D section independently, and each pixel is labeled automatically. The labels from each 2D section are then projected back to the original 3D locations for the final vote-based labeling [18-20].

For segmentation of an unseen CT case, the density resolution (CT number) of the CT images is first reduced from 12 to 8 bits by using linear interpolation and target organ region is cropped based on bounding box that is detected by localization module. The trained deep CNN is then applied to each 2D section within bounding box independently, and each pixel is labeled automatically. The labels from each 2D section are then projected back to its original 3D location for the final vote-based labeling, as described above.

## 3. EXPERIMENT AND RESULTS

A CT image database that was produced and shared by a Japanese research project entitled "Computational Anatomy" [23] was used in this experiment. This database includes 640 3D volumetric CT scans from 200 patients at Tokushima

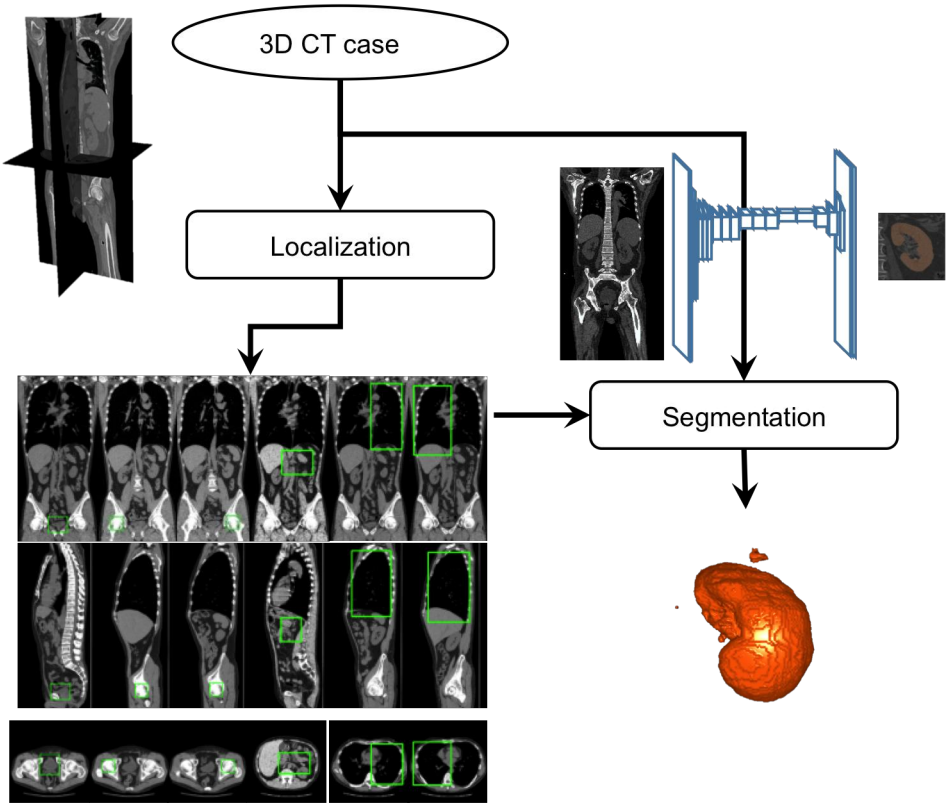


Fig.1. Process flow of our approach for an organ segmentation on a 3D CT case.

University Hospital. The anatomical ground truth (a maximum of 19 labels that include Heart, right/left Lung, Aorta, Esophagus, Liver, Gallbladder, Stomach and Duodenum (lumen and contents), Spleen, left/right Kidney, Inferior Vein Cava, region of Portal Vein, Splenic Vein, and Superior Mesenteric Vein, Pancreas, Uterus, Prostate, and Bladder) in 240 CT scans was also provided in the database [24]. Our experimental study used all of the 240 ground-truth CT scans, comprising 89 torsos, 17 chests, 114 abdomens, and 20 abdomen-with-pelvis scans.

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. We re-trained the segmentation module that was constructed by deep CNNs [18-20] based on the regions of interest (ROIs) on CT images indicated by bounding boxes of the target organs. Our deep CNNs were implemented by using Caffe package [26]. Twelve CT scans were picked at random as the test samples, and the remaining 228 CT scans were used for training the deep CNN. As previously mentioned, we took 2D sections along the axial, sagittal, and coronal body directions. We trained two deep CNNs based on the ground-truth labels of the pancreas and left kidney regions, respectively. A method for stochastic optimization (ADAM) [25] was used for the optimization. All the 2D images were used directly as the inputs for CNN training, without any patch sampling.

We tested the proposed scheme (Fig. 1) using 12 CT cases that were not used in the training stage. Two examples of the segmentation results for pancreases and left kidney on a CT case were shown in Fig. 2 and 3. We measured the intersection over union (IU) (also known as the Jaccard index) between the segmentation result and the ground truth, and voxel accuracy that shows true positive rate of the voxel-wised prediction. The evaluation results for the voxel accuracy and IU were 99 % and 88 % for left kidney, 84 % and 65 % for pancreas, when averaged over all the segmentation results of the test data. That means 99 % of the voxels within the left kidney (84% within pancreas) were labelled correctly, with a region-coincidence (3D Jaccard index) of 88 % (65% of pancreas) to human annotations for the test data.

## 4. DISCUSSION AND CONCLUSION

We confirmed that the left kidney and pancreas were localized and segmented correctly in all test CT images. Our segmentation targets cover a wide range of shapes, volumes, and sizes, either with or without contrast enhancement, and at different locations in the human body. These experimental results showed the potential of our approach to recognize those organs that may appear in CT images. The IUs of the target organs (left kidney: 88 %, pancreas: 65 %) were comparable to the accuracies reported from the previous state-of-the-art methods. Comparing to our previous method (left kidney: 80 %, pancreas: 31 %), the IUs for segmenting left kidney and pancreas were 8 % and 34 % higher. From this result, we see that accuracy of the organ segmentation based on deep CNNs can be improved significantly by only focusing an ROI (bounding box of the target) instead of the whole image region on CT images.

The performance of the deep CNNs for multiple organ segmentations based on an end-to-end deep learning has been confirmed in our previous works [18-20]. However, for some organ types (e.g., pancreases, gallbladder, portal vein, splenic vein, and superior mesenteric vein), our previous method could not show particularly high IUs [18,19]. This experimental result demonstrated a solution by restricting the inputted image range for training and testing deep CNNs

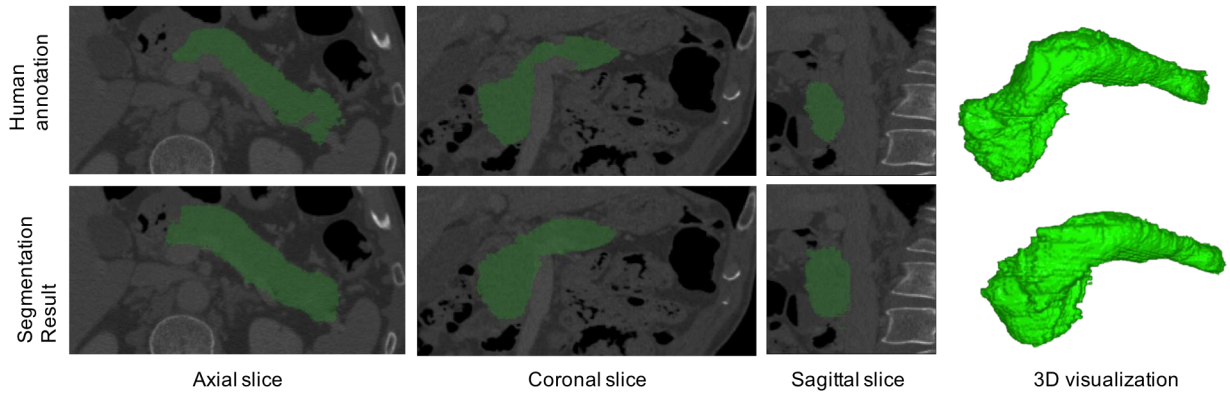


Fig.2. Example of pancreas segmentation on a CT scan by using the proposed method. Three typical slices of a 3D CT scan and a 3D view of pancreas region (green color) are shown from left to right. Human annotation and segmentation result are shown in upper and bottom, respectively.

that can improve the accuracy of segmenting those organ types. We found the organ localization technique on CT images established in our previous work [21, 27] is still useful as a pre-processing to improve the performance of deep CNN based image segmentation.

In conclusion, an improvement of a deep-learning approach for CT image segmentation was proposed. We applied our method to segmenting two types of organs based on a dataset of 240 3D CT cases with different scan ranges and protocols. The segmentation results demonstrated a significant improvement in accuracies than these of our previous works. The proposed method can be applied to improve accuracy of the other anatomical structure segmentations in the future.

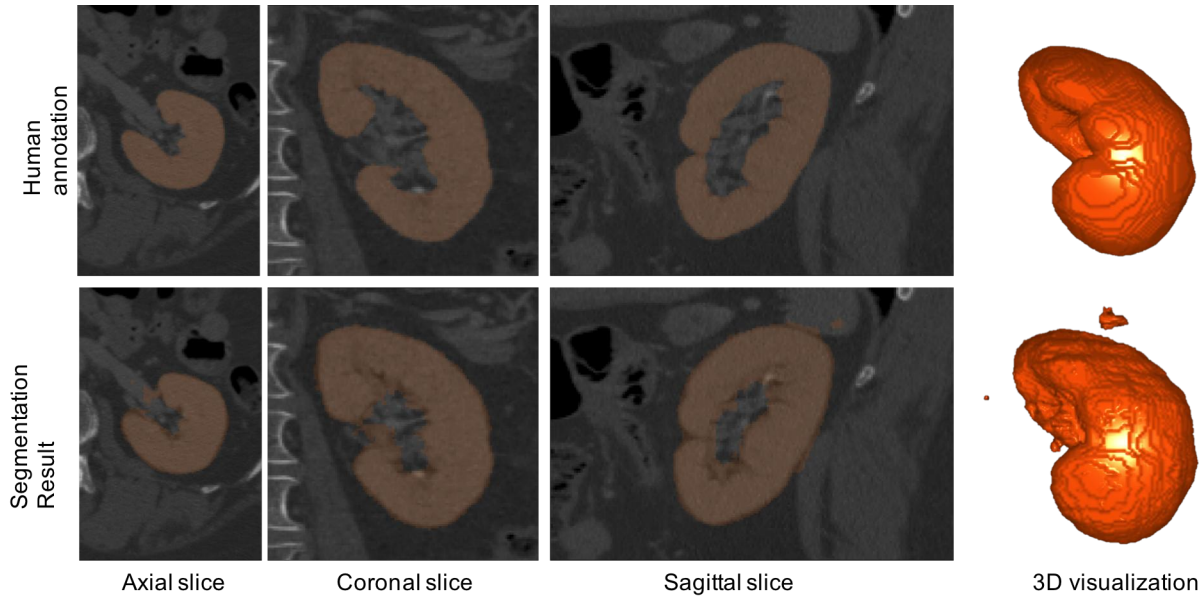


Fig.3. Example of left kidney segmentation on a CT scan by using the proposed method. Three typical slices of a 3D CT scan and a 3D view of left kidney region (red color) are shown from left to right. Human annotation and segmentation result are shown in upper and bottom, respectively.

## ACKNOWLEDGEMENTS

Authors thank the members of Fujita Laboratory. This research work was funded in part by a Grant-in-Aid for Scientific Research on Innovative Areas (26108005), and in part by Grant-in-Aid for Scientific Research (C26330134), MEXT, Japan.

## 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. A. Shimizu, R. Ohno, T. Ikegami, H. Kobatake, S. Nawano, D. Smutek, "Segmentation of multiple organs in non-contrast 3D abdominal CT images," Int. J. Compt. Assisted Radiol. Surg., 2, 135–142 (2007).
7. 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).
8. 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).
9. U. Bagci, J. K. Udupa, N. Mendhiratta, B. Foster, Z. Xu, J. Yao, X. Chen, D. J. Mollura, "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., 17 (8), 929–945 (2013).
10. 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).
11. 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).
12. F. Ciompi, B. de Hoop, S. J. van Riel, K. Chung, E. Scholten, M. Oudkerk, P. de Jong, M. Prokop, B. van Ginneken, "Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2d views and a convolutional neural network out-of-the-box," Med. Image Anal., 26 (1), 195–202 (2015).
13. A. Teramoto, H. Fujita, O. Yamamuro, T. Tamaki, "Automated detection of pulmonary nodules in PET/CT images: Ensemble false-positive reduction using a convolutional neural network technique," Med. Phys., 43 (6), 2821–2827 (2016).
14. J. J. N¨appi, T. Hironaka, D. Regge, H. Yoshida, "Deep transfer learning of virtual endoluminal views for the detection of polyps in CT colonography," Proc. SPIE, 9785, 97852B-1-97852B-8 (2016).
15. A. d. Brebisson, G. Montana, "Deep neural networks for anatomical brain segmentation," Proc. CVPR workshops, 20-28 (2015).
16. 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).
17. 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).
18. 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).
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, T.Ito, R.Takayama, S.Wang, T.Hara, and H.Fujita, "Automatic segmentation of anatomical structures in 3D CT images by voxel-wise voting from a 2D fully convolutional network based on end-to-end deep learning," Submitted and under the review.
21. X.Zhou, S.Morita, 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. J. Long, E. Shelhamer, T. Darrell, "Fully convolutional networks for semantic segmentation," Proc. CVPR, 3431-3440 (2015).
23. http://www.comp-anatomy.org/wiki/
24. 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).
25. D. P. Kingma, J. L. Ba, "ADAM: A method for stochastic optimization," Proc. ICLR 2015, (2015).
26. http://caffe.berkeleyvision.org
27. X.Zhou, S.Wang, H.Chen, T.Hara, R.Yokoyama, M.Kanematsu, and H.Fujita, "Automatic localization of solid organs on 3D CT images by a collaborative majority voting decision based on ensemble learning," Computerized Medical Imaging and Graphics, 36 (4), 304-313, (2012).