Automatic anatomy partitioning of the torso region on CT images by using a deep convolutional network with a majority voting

Xiangrong Zhou [a], Takuya Kojima [a], Song Wang [b], Xinxin Zhou [c], Takeshi Hara [a], Taiki Nozaki [d], Masaki Matsusako [d], and Hiroshi Fujita [a]

a: Department of Electrical, Electronic and 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: Department of Information and Media Studies, Nagoya Bunri University, 365 Maeda, Inazawa-cho, Inazawa-shi, 492-8520 Japan

d: Department of Radiology, St. Luke's International Hospital, Tokyo, 104-8560, Japan

Abstract (250 works)

We propose an automatic approach for anatomy partitioning on 3D CT images, which aims to divide the human torso region into several volumes-of-interest (VOIs) according to the anatomical definition. The proposed approach trains a deep convolutional neural network (CNN) to automatically detect the bounding boxes of major organs on two dimensional (2D) sections of CT images, and groups the coordinates of those boxes to vote for a 3D VOI (called localization) for each organ type respectively. We applied this approach to localize the VOIs of 17 types of major organs in the human torso and a four-fold cross-validation for performance evaluation on a dataset consisting of 240 3D CT scans with the human-annotated ground truth for each organ region. The preliminary results show that 84.3% VOIs of the 17 types of organs in 240 test CT images are localized with acceptable accuracies (mean of Jaccard indexes is 70.2%) against the human annotations. This performance is better than the state-of-the-art method reported recently. The experimental results demonstrate the improvements in efficiency and usefulness by using a deep CNN for anatomy partitioning on 3D CT images.

Description of purpose:

Fully automatic anatomy partitioning, which finds a bounding box for each organ on 3D CT images, is a fundamental pre-processing step in medical image analysis. Dividing the CT images into several volumes-of-interest (3D bounding boxes) has three benefits: (1) providing the diagnostic information about the spatial relations and volumes of different organs, (2) simplifying the complexity and difficulty of many challenging tasks such as organ segmentation, lesion detection, and image registration that are routinely used in computer-aided diagnosis, and (3) reducing the data size of a CT volume (a typical size is 512x512x1024 voxels) to several small-sized VOIs to increase the efficiency and speed of computation, which is especially useful in GPU based training and testing. The core part of anatomy partitioning is multi-organ localizations, which is one of the most important problems in the developments of computer-aided diagnosis systems.

A few prior works have been reported for multi-organ localizations on CT images [1-9]. The methods used in those previous works can be divided into two categories: atlas-based registration and machine-learning-based detection. The former tries to deform the CT images into a standard space by using image registration and localizes the multiple organs by referring to anatomical structures (called atlas) that are pre-defined in the standard space [1, 2]. The latter tries to train a detector that slides a window over the entire space of CT images to decide the target organ location. The detector is a classifier that is trained on a hand-craft feature space to judge the target organ within or outside the current window at each location of the CT images [3-8]. Thanks to the progress of the machine-learning technique that is proposed for face detection in computer vision, the machine-learning-based detection approach demonstrated a better robustness, generality, and efficiency than atlas-based methods and has become the mainstream solution to handle the multiple organs localization issue on CT images.

We proposed an approach [6] to localize the multiple organs on 3D CT images by using a machine-learning approach. The proposed approach detected a target organ location on 3D CT images by using ensemble-learning on 2D sections with a 3D majority voting method. The proposed scheme successfully solved the different kinds of organ detection problem generally and handled real clinical images generated by the different CT scanners, including non-contrast and contrast-enhanced, normal and abnormal CT cases [6]. This scheme was improved continually by expanding the feature pool to enhance the performance of the ensemble-learning [7], increasing the number of 2D sections sampled on more than three orientations for voting to enhance the robustness and accuracy of localization [8], and introducing the spatial relation between multiple organ locations to compensate the absence or outlier of the independent detections [8]. The performance of the proposed scheme was evaluated by detecting 35 types of organ regions on more than 3,000 CT scans including PET-CT images that were scanned in a low spatial resolution with a poor image contrast. While this scheme demonstrated many promising results [8], computational efficiency of this scheme needs to be improved on both training and testing because the detectors and features for each organ localization are different and need to be learned from the training dataset independently. This weakness was also discussed in [9].

The purpose of this work is to improve the computational efficiency of the multi-organs localization developed in our previous work [6] by replacing the core part of the organ detection from ensemble-learning based on ad-hoc hand-draft features with a deep CNN [10] that learns feature (representation) from CT images automatically. We train and use a deep CNN to localize all the organ regions simultaneously on all 2D sections that sample along the different orientations from 3D CT images based on unified feature space. We describe the improvement of this scheme and details of the deep CNN for both training and testing and give a performance comparison with the state-of-the-art method [9] that was published recently.

<u>Methods:</u>

The process flow of our scheme is shown in Fig.1. The input is a 3D CT image and the output is a set of anatomical partitions (bounding boxes) annotated on CT image. In contrast to our previous work [6], we handle the localizations of different inner organs jointly by using a deep CNN. Our method still treats 3D organ localizations on a 3D CT image as detecting several independent 2D objects in a series of 2D image slices originally proposed in [6]. The idea of reducing the feature dimension (3D to 2D) for classification and increasing the size of training samples (one 3D training sample consists of many 2D training samples) improves the robustness and accuracy of organ localizations [6]. For an unseen 3D CT scan, our method applies the trained deep CNN to detect several 2D candidates for each organ type and votes all 2D candidates back to the

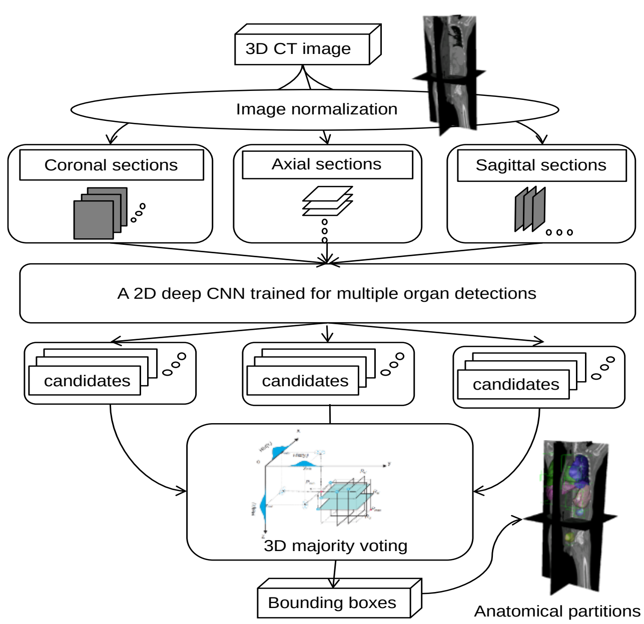

Fig.1. The process flow for anatomy partitioning on a 3D CT image.

3D space. Finally, our algorithm judges the existence of the target by examining the mutual consent of all 2D candidates and selecting the location-majority of the related 2D candidates in the 3D voting space as the final target location for each organ region.

Actually, our scheme repeats to sample 2D sections from a CT case, passes to a deep CNN for object detection, and stacks the detected 2D bounding boxes back to 3D. Finally, the 3D bounding box for each organ is decided based on the most accumulated coordinates belong to two corner points on the CT image. This deep CNN is trained based on a set of CT images with the human annotations. All the 2D CT sections with the annotations along the three body orientations are shuffled and used to train the CNN. The training process repeats feed-forward computation and back-propagation to minimize the loss function, which is defined as the weighted sum of a confidence associated with an organ type and an error of the boxes between the network prediction and the human annotation. The gradients of the loss are propagated from the end to the start of the network, and the method of stochastic gradient descent with momentum is used to refine the parameters of each layer. During the testing stage, the trained CNN is applied to each 2D section of a CT image independently, and each organ is localized automatically. The coordinates of the bounding box from each 2D section are then projected back to their original 3D locations for the final vote-based decision, as described above.

Experiment and Results:

A shared dataset [10] produced by a research project titled "Computational Anatomy" was used for performance evaluations. 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 and distributed within the dataset. In this work, we only consider the 17 types of organ regions as the localization targets for performance evaluations.

A four-fold cross-validation was used for performance evaluation. We iteratively use 75% CT scans of the dataset for training parameters of the deep CNN and tested the trained network on the remaining 25% CT scans. The organ regions were divided equally per-type into each fold and the CT scans from the same patient were only used in the same training or testing stage. The accuracy of the localization was evaluated for each organ type and each CT scan. The intersection over union (IU) (also known as the Jaccard similarity coefficient) between the detected bounding box and ground truth was used as the evaluation metric. We used a condition of $IU > 0.5$ to determine the success of an organ localization.

We evaluated the localization results on all (17) types of the (810) organ regions on all (240) CT images, which are obtained in the test stage, and confirmed that 84.3% organ regions were localized successfully with a mean IU of 70.2%. This performance is better than the state-of-the-art method [9], which showed that 16.5% organ regions were localized successfully with a mean IU of 38.1%, based on the same validation process (both training and testing) on the same CT dataset. The worst performance was achieved on the uterus localization (only 50% uterus in total 8 CT cases were localized successfully with a mean IU of 59.3%). How to localize small-size organ types on non-contrast CT images by learning from a small number of training samples is still challenging and we plan to study it further in our future work.

New or breakthrough work to be presented

This work is the major revision of our previous work [6] that can successfully localize multiple organs on CT images. Deep learning model was embedded as a module into the existing system and demonstrated a better performance than the state-of-the-art method [9] published recently.

Conclusion:

We proposed a deep-learning-based multi-organs localization scheme that is a major revision of our previous works [6-8]. The organ detection in our previous works, which used an ensemble-learning approach based on organ-specific classifiers and ad-hoc feature-sets, was replaced by a deep learning approach, which used one classifier and one feature-set shared by all type of organ. The revision changed the organ localization from a sequential process (one organ per image scan) to a parallel process (all-organs in one image scan) and significantly improved computational efficiency for multiple organs localization on CT images which is an important pre-processing step in many computer-aided diagnosis systems.

This paper included new methodologies and results that had not been submitted for publication or presentation before.

References:

1. R. Wolz, C. Chu, K. Misawa, M. Fujiwara, K. Mori, D. Rueckert: Automated abdominal multi-organ segmentation with subject-specific atlas generation, IEEE Trans. Medical Imaging, 32, 1723-1730, 2013.
2. K. Sun, JK. Udupa, D. Odhner, Y. Tong, L. Zhao, DA. Torigian: Automatic thoracic anatomy segmentation on CT images using hierarchical fuzzy models and registration, Medical Physics, 43, 1882-1896, 2016.
3. Y. Zheng, B. Georgescu, D. Comaniciu: Marginal space learning for efficient detection of 2D/3D anatomical structures in medical images, in Information Processing in Medical Imaging. Berlin, Germany: Springer, 411-422, 2009.
4. F. C. Ghesu et al.: Marginal space deep learning: Efficient architecture for volumetric image parsing, IEEE Trans. Medical Imaging, 35(5), 1217-1228, 2016.
5. A. Criminisi et al.: Regression forests for efficient anatomy detection and localization in computed tomography scans, Medical Image Analysis, 17, 1293-1303, 2013.
6. 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.
7. X. Zhou, A. Yamaguti, Xin. Zhou, T. Hara, R. Yokoyam, M. Kanematsu, H. Fujita: Automatic organ localizations on 3D CT images by using majority-voting of multiple 2D detections based on local binary patterns and Haar-like features, Proc. of SPIE Medical Imaging 2013: Computer-Aided Diagnosis, 8670, 86703A-1 - 86703A-7, 2013.
8. X. Zhou, S. Morita, Xin. Zhou, H. Chen, T. Hara, R. Yokoyama, M. Kanematsu, H. Hoshi, 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, 9414, 94143K-1 - 94143K-6, 2015.
9. B. D. de Vos, J. M. Wolterink, P. A. de Jong, T. Leiner, M. A. Viergever, I. Isˇgum: ConvNet-based localization of anatomical structures in 3-D medical images, IEEE Trans. Medical Imaging, 36 (7), 1470-1481, 2017.
10. X. Zhou, R. Takayama, S. Wang, T. Hara, 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.