PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Automatic coronary artery lumen segmentation in computed tomography angiography using paired multi-scale 3D CNN

Fei Chen, Yu Li, Tian Tian, Feng Cao, Jimin Liang

Fei Chen, Yu Li, Tian Tian, Feng Cao, Jimin Liang, "Automatic coronary artery lumen segmentation in computed tomography angiography using paired multi-scale 3D CNN," Proc. SPIE 10578, Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging, 105782R (12 March 2018); doi: 10.1117/12.2293289

SPIE.

Event: SPIE Medical Imaging, 2018, Houston, Texas, United States

Automatic coronary artery lumen segmentation in computed tomography angiography using paired multi-scale 3D CNN

Fei Chen[a], Yu Li[a], Tian Tian[b], Feng Cao[b], Jimin Liang*[a]
[a]Engineering Research Center of Molecular and Neuro Imaging of Ministry of Education, School of Life Science and Technology, Xidian University, Xi'an, Shaanxi 710071, China;
[b]Department of Cardiology, Chinese PLA General Hospital, Beijing, 100089, China

ABSTRACT

Coronary artery disease (CAD) is one of the leading causes of death worldwide. The computed tomography angiography (CTA) is increasingly used to diagnose CAD due to its non-invasive nature and high-resolution three-dimensional (3D) imaging capability of the coronary artery anatomy. CTA allows for identification and grading of stenosis by evaluating the degree of narrowing of the blood-filled coronary artery lumen. Both identification and grading rely on the precise segmentation of the coronary arteries on CTA images. In this paper, a fully automatic segmentation framework is proposed to extract the coronary arteries from the whole cardiac CTA images. The framework adopts a paired multi-scale 3D deep convolutional neural networks (CNNs) to identify which voxels belong to the vessel lumen. Voxels that may belong to coronary artery lumen are recognized by the first CNN in the pair and both artery positives and artery-like negatives are distinguished by the second one. Each CNN is assigned to a different task. They share the same architecture in common but with different weights. In order to combine local and larger contextual information, we adopt a dual pathway architecture that can process the input image simultaneously on multiple scales. The experiments were performed on a CTA dataset from 44 patients. 35 CTA scans are used for training and the rests for testing. The proposed segmentation framework achieved a mean Dice similarity coefficient (DSC) of 0.8649 and mean surface distance (MSD) of 0.5571 with reference to manual annotations. Experimental results show that the proposed framework is capable of performing complete, accurate and robust segmentation of the coronary arteries.

Keywords: 3D convolutional neural network, deep learning, cardiac computed tomography angiography, coronary arteries segmentation

1. INTRODUCTION

Over the past decades, coronary artery disease (CAD) has been one of the most common causes of human deaths in the world.[1] Many factors may contribute to the development of CAD, of which stenosis caused by atherosclerosis is the most common factor.[2] Computed tomography angiography (CTA) has become the most commonly used technique for clinical plaque assessment and stenosis detection,[3] with the advantages of non-invasive nature and the high-resolution 3D imaging capability of the complete coronary tree.[4] Most plaque evaluation and stenosis detection rely on a (semi)-automatically segmentation of the coronary arteries.[5] Therefore, the coronary arteries are usually extracted first. Inaccurate segmentation of coronary arteries will result in fatal false diagnosis because a missing segment or mixed extraction of other structures may lead to the improper plaque evaluation and stenosis detection.

In recent years, numerous studies are committed to the extraction of vascular tree. The vast majority of these studies concentrate on the segmentation of coronary artery tree (CAT)[6] and fall broadly into two categories, data-based methods and model-based methods.[7] The data-based approaches typically focus on image enhancement techniques to identify blood vessels, the estimation of vessel diameter, the detection of bifurcations, centerline extraction and special structures recognition.[8] These studies treat the blood vessels as a group of interconnected tubular structures on which the bifurcation and stenosis presence, without the use of any anatomical information. Most recent published CAT segmentation studies based on the model-based approaches that improve the segmentation accuracy by involving anatomical models or atlases.[7] The model-based methods first build predefined models or atlases, then extract the relevant information from the existing dataset and perform a registration process. The pre-defined vascular model is the prerequisite for blood vessels and centerlines recognition, which is time consuming. In addition, the large differences in the anatomy of the coronary arteries between individuals also increase the difficulty of segmentation.

Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging,
edited by Barjor Gimi, Andrzej Krol, Proc. of SPIE Vol. 10578, 105782R · © 2018 SPIE
CCC code: 1605-7422/18/$18 · doi: 10.1117/12.2293289

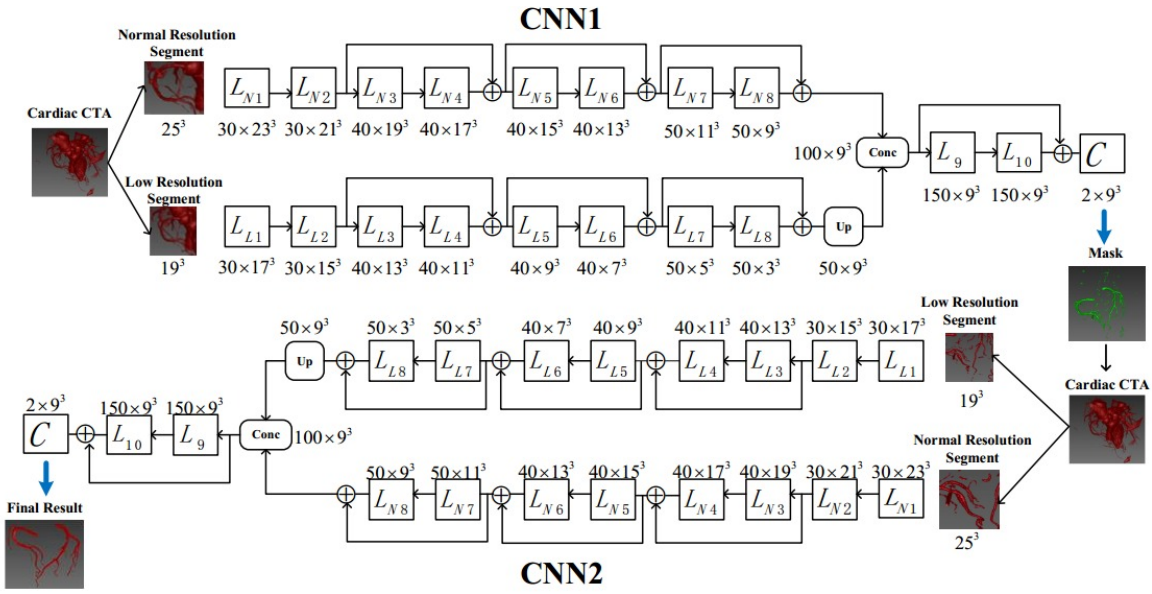


Figure 1. The proposed framework for fully automatic segmentation of the coronary arteries. The upper network is CNN1 and the lower one is CNN2. Each CNN has two convolutional pathways with residual connections. The input segments center of the two pathways is located at the same location of the image and the extracted segments from the down-sampled image are taken as input to the second pathway. *Up* represent the upsampling layer while *C* represent the classification layer. The number and size of feature maps (FMs) are recorded as (Number × Size).

Deep learning shows a trend of rapid growth in the field of machine learning, which currently drives the booming of artificial intelligence. In many applications of computer vision, the performance of convolutional neural networks (CNNs) has exceeded the state-of-the-art. Moreover, CNNs have also been successfully used in a variety of medical image processing tasks, such as the segmentation of knee cartilage,[9] lymph node detection,[10] brain tissue segmentation[11] and pulmonary nodule classification.[12]

Being inspired by the work concerned with the brain lesion segmentation in MRI images,[13] we propose a fully automatic framework that adopts a paired multi-scale 3D deep convolutional neural networks to extract the coronary arteries from the whole cardiac CTA images. Such a strategy has no need of any post-processing steps while can achieve satisfactory results.

2. METHODS

This study aims to extract the coronary arteries from the CTA images with human-annotated masks. As shown in Fig.1, given CTA images as input data, we first construct a full convolutional network CNN1 (with a dual pathway architecture) model to generate the coronary artery candidate masks. Then CNN2 which has the same architecture as CNN1 is built to produce the final probability map, taking the samples extracted within the candidate mask generated by CNN1 as input. Finally, the probability map of coronary arteries is thresholded to get the final binary segmentation map. In the following sub-sections, components in the proposed framework for automatic segmentation of coronary arteries are introduced in details.

2.1 Data preparation and pre-processing

In this study, a total of 44 consecutive cardiac CTA exams of patients were included in the dataset, made on a dual-source CT scanner (Somatom Definition Flash, Siemens Healthcare, Forchheim, Germany). The CTA scans with ECG-triggering and contrast enhancement were acquired with 120 KVp tube voltage and 55 mAs tube current. The ground-truth manual annotations were obtained by two experienced cardiologists who were blinded to the patient's clinical information by using a commercially available software MIMICS.[14] The datasets were obtained with different voxel sizes between individuals and were resampled to the same voxel size of 0.5^3. Then we normalized the intensity of the images to a zero-mean, unit-variance space. The images were reflected along the sagittal axis to expand the training data.

2.2 Paired CNN architecture

In this paper, we use a pair of CNNs to classify voxels in the CTA images, which are purely convolutional, i.e. there are no pooling layers and fully connected layers are implemented as convolutions, thereby an efficient voxel classification can be achieved. Each CNN has a specific task. The first CNN (CNN1) selects voxels that may belong to coronary artery throughout the cardiac CTA images, focuses on detecting artery-like voxels and learning to discard most of the negative samples. The second CNN (CNN2) identifies artery voxels among all these suspected samples. The two CNNs are trained in order. For each training image, a coronary artery coarse mask is generated by CNN1, which contains artery-like voxels and excludes the negatives such as bones and adipose tissues. Then, CNN2 is trained using only the samples extracted from the coronary artery coarse mask obtained by CNN1. In order to take advantage of the already learned knowledge, CNN2 is initialized with the weights of the final version of CNN1 and fine-tuned with the samples from the coronary artery coarse mask. Therefore, CNN1 and CNN2 share their architecture but have different trainable parameters. The overview of the network architecture is shown in Figure.1.

2.3 Multi-scale 3D convolutional networks

Combining local and larger contextual information in the decision process has been shown beneficial in other works.[15] In order to efficiently achieve this merging, each CNN uses a dual pathway architecture that processes the input images at multiple scales simultaneously to obtain a larger receptive field for the final classification layer at low computational costs. The first pathway operates on the original images to capture the detailed local appearance of the structures, while the second pathway utilizes the down-sampled images to learn more advanced features such as the coronary arteries location information. In order to preserve the dense inference characteristic of the CNN, the feature maps (FMs) of the last convolutional layer of the second pathway are upsampled to match the dimensions of the last convolutional layer of the first pathway. Then, as shown in Fig.1, the two feature maps are connected together. Since the two pathways are separated in this way, context in any size can be handled using the second pathway by simply changing the down-sampling factor FD.

Unlike most popular used 2D CNNs, 3D CNNs adopted in our framework makes better use of the volumetric image data. Fully 3D CNNs brings about an increased number of parameters, with significant memory and computational requirements.[10] A few tricks are used to address these problems. First, when the input size is larger than CNN's receptive field, the fully convolutional networks can make dense-inference.[16] In order to enable dense-inference, the final classification layer is actually a convolutional layer with a kernel size of 1^3. When performing image segmentation, the receptive field is shifted over the input images, it is possible to predict multiple voxels simultaneously. This method avoids repeated convolution calculations on the same voxels over a large number of overlapping patches, which can greatly reduce the computational costs and memory loads. Second, as a large 3D networks are adopted in this paper, there are a plenty of FMs need to be cached. Due to the GPU memory limitations, the input of the networks uses segments composed of voxels instead of individual patches. The size of the segments is larger than the patches, but small enough for memory capacity. A training strategy that uses the dense inference on image segments is employed in this paper.[13] If the training batches are formed of N segments extracted from the training set, the cost function becomes

$$C(\Theta; B_n, l_n) = -\frac{1}{N \cdot M} \sum_{n=1}^{N} \sum_{m=1}^{M} \log(P_{l_n^m}(z^m)), \tag{1}$$

where B_n is the n-th segment in the batch and l_n is the corresponding true label. There are M predicted voxels in total. l_n^m represents the true label for the m-th voxel in the segment, z^m is the corresponding location information in the classification FMs and $P_{l_m^n}$ is the output of the softmax function. It's a mixed solution between the dense-inference training on a whole image and the commonly used training on individual patches.[17] Finally, inspired by VGG,[18] the adopted network makes use of small 3^3 convolutional kernels. This trick can be used to build deeper CNNs while keeping the number of training parameters at a lower level.

As shown in Fig.2, the number of positive samples is much less than the negative ones in the whole cardiac CTA images. If we generate the training segments using slide windows on the whole cardiac CTA images, this can lead to a serious class imbalance. It's a critical issue that will have a direct impact on the accuracy of the segmentation. In this study, we extract segments whose centers have a half probability of being on foreground or background voxels, alleviating the problem of class-imbalance.

Recent studies have shown that residual connections can facilitate preservation of the flowing signals and therefore capable of training very deep neural networks.[19] Although the networks used in biomedical field typically contains a relatively small number of layers compared to the modern architectures in computer vision, considering the trainable parameters in the 3D kernels, the problem of preserving the forward and backward propagating signals and the difficulty of optimization may be quite large in 3D CNNs.[13] In this study, for each pathway of CNNs, except for the first two layers, residual connections are added between the outputs of every two layers. Each layer block is ordered in batch-normalization, non-linearity and convolution. Previous studies have shown that this arrangement can achieve better performance.[20] In this paper, the element-wise addition $\oplus$ is performed between the output O_l of layer l and the input I_{l-1} of the previous layer with the operation of residual connection. Formula follows

$$I_{l+1}^{k} = \begin{cases} O_l^k \oplus \hat{I}_{l-1}^k, & \text{if } k \le K^{l-1} \\ O_l^k, & \text{otherwise} \end{cases} \tag{2}$$

where l represents any layer after adding a residual connection, the superscript k denotes the k-th channel and K^l is the number of FMs in layer l. In order to match the dimensions of O_l, $\hat{I}_l$ is padded in the (x, y, z) dimensions with reflection respectively.

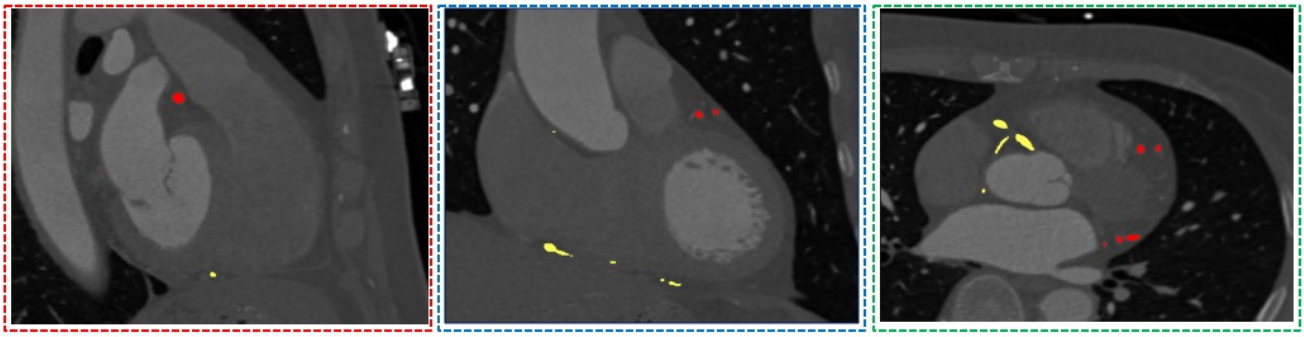
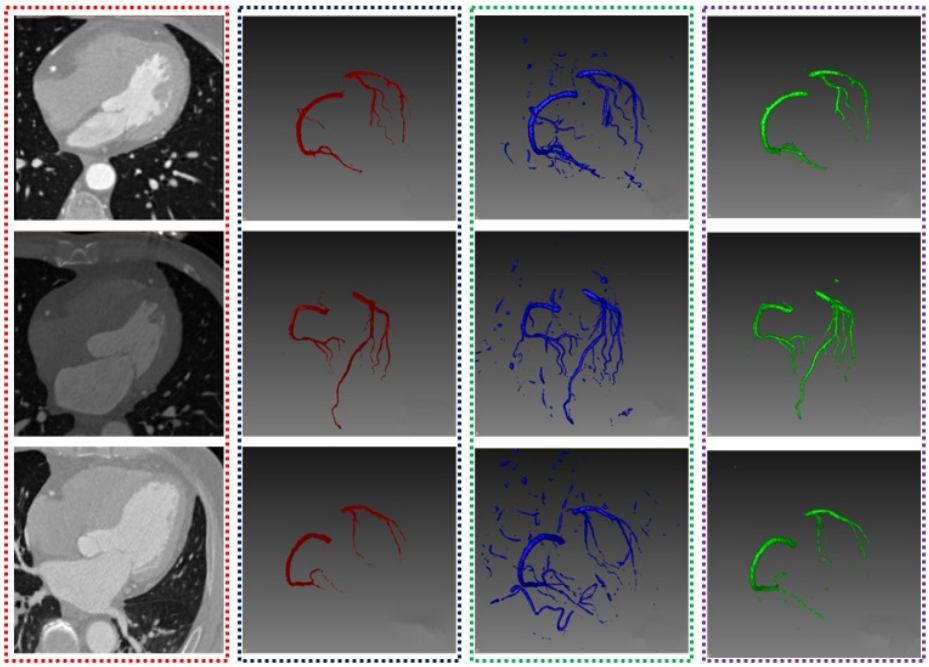


Figure 2. The distribution of positive and negative samples in sagittal, coronal and axial view, respectively. The yellow mask indicates the right coronary artery, the red stands for the left coronary artery and the remaining background for negative samples. As we can see that the number of negative samples is much larger than the positive ones.

3. RESULTS

During the training stage, the images are randomly divided into two datasets, i.e., the training set with 27 images and the validation set with 8 images respectively. At regular intervals, we extract the equal number of segments from each of the validation images for a total of 10k to monitor the changes of segmentation accuracy during training. Full segmentation of the validation datasets and the calculation of the mean Dice similarity coefficient (DSC) are performed every five epochs. The low resolution images were downsampled from the original images (normal resolution) by a factor of three. The deep learning library Theano[21] is used to realize the CNNS. The PReLu non-linearity is used in this paper. The RMSProp optimizer and Nesterov momentum with value 0.6 are applied to train the networks. L1 and L2 regularization are also used in our works with values of 10^{-6} and 10^{-4} respectively. The networks are trained with dense-training using batches of 10 segments. Each segment has a size of 25^3. One thing to note, the weights of CNN1 are initialized by sampling from the normal distribution $N(0, 0.01)$, while the weights of CNN2 are initialized with the final weights of CNN1. Throughout the training process, the initial learning rate is set to $lr = 10^{-4}$ and halved when the plateau is converged.

To evaluate the segmentation performance of the proposed framework, DSC and MSD are used to compare the thresholded probability map with the manually annotated result. The index of quantitative assessment is shown in Table 1. The segmentation results are compared with the results using only a single CNN in the paired CNN architecture. We then further show the qualitative segmentation results for three patients in Fig. 3. We can observe that satisfactory performance is achieved on our proposed framework, while the results on the single CNN are not reasonable.



(a) (b) (c) (d)

Figure 3. Qualitative evaluation of the coronary arteries segmentation framework for three patients. (a): CTA images, (b): manual annotations, (c): segmentation masks acquired by the single CNN, (d): final segmentation results.

Table 1. Comparison of the segmentation results.

Network	DSC [%]	MSD [mm]
Single CNN	0.7270 ± 0.0859	6.3757 ± 5.2899
Proposed framework	0.8649 ± 0.0329	0.5571 ± 0.4196

4. CONCLUSION AND DISCUSSION

In this paper, we propose a fully automatic framework to segment the coronary artery from the whole cardiac CTA images. This framework adopts a paired multi-scale 3D deep convolutional neural network to identify voxels that belong to the vessel lumen. CNN1 identifies voxels that may belong to the coronary artery, thereby excluding the vast majority negative voxels such as bones and adipose tissues. Subsequently, the identified artery-like voxels are classified by CNN2, which distinguishes between artery and artery-like negatives. The coronary arteries are segmented from the CTA images in a coarse-to-fine manner. This architecture can significantly improve the segmentation results. No manual interaction and post-processing are involved in the entire segmentation process. Our framework does not need to limit the volume of interesting (VOI), such as the bounding box around the heart. The coronary segmentation results are obtained directly from the whole cardiac CTA images, avoiding the trouble of determining the VOI. The experimental results show that the proposed framework has a good performance in the segmentation of coronary arteries from CTA images.

For the further study, we intend to combine coronary artery lumen segmentation with landmark detection. Coupled with landmark information may promote vascular segmentation. In addition, as far as we know, currently there is a lack of clinical software that automatically divides coronary arteries into 17 modified AHA segments.[22] Cardiologists need to artificially determine which coronary segment the lesion belongs to when performing surgical reports. This is a time-consuming procedure. Only after a period of training can cardiologists achieve reliable results. Recent study has successfully completed the task of craniomaxillofacial bone segmentation with the assistance of the landmark.[23] Due to

the limited medical image training data at hand, we may take the Two-Stage Task-Oriented Deep Neural Networks[24] into consideration to achieve better result.

ACKNOWLEDGMENTS

This work was supported by the Program of the National Key Research and Development Program of China [Grant No. 2016YFC1300300], National Natural Science Foundation of China [Grant No. 61571353 and 81530058], the Natural Science Basic Research Plan in Shaanxi Province of China [Grant No. 2015JZ019].

REFERENCES

[1] Roger, V. L., Go, A. S., Lloydjones, D. M., Benjamin, E. J., Berry, J. D., Borden, W. B., Bravata, D. M., Dai, S., Ford, E. S., Fox, C. S., Fullerton, H. J., Gillespie, C., Hailpern, S. M., Heit, J. A., Howard, V. J., Kissela, B. M., Kittner, S. J., Lackland, D. T., Lichtman, J. H., Lisabeth, L. D., Makuc, D. M., Marcus, G. M., Marelli, A. J., Matchar, D. B., Moy, C. S., Mozaffarian, D., Mussolino, M. E., Nichol, G., Paynter, N. P., Soliman, E. Z., Sorlie, P. D., Sotoodehnia, N., Turan, T. N., Virani, S. S., Wong, N. D., Woo, D. and Turner, M. B., "Heart disease and stroke statistics-2012 update: A report from the American Heart Association," Circulation 125 (1), 12-230 (2012).

[2] Tian, Y., Pan, Y., Duan, F., Zhao, S., Wang, Q. and Wang, W., "Automated segmentation of coronary arteries based on statistical region growing and heuristic decision method," BioMed Research International (2016).

[3] Benedek, T., Mester, A., Benedek, A., Rat, N., Opincariu, D. and Chitu, M., "Assessment of coronary plaque vulnerability in acute coronary syndromes using optical coherence tomography and intravascular ultrasound. A system review," Journal of Cardiovascular Emergencies 2(4), 173-184 (2016).

[4] Rodriguez-Granillo, G. A., Carrascpsa. P., Bruining, N., Waksman, R. and Garcia-Garcia, H. M., "Defining the non-vulnerable and vulnerable patients with computed tomography coronary angiography: evaluation of atherosclerotic plaque burden and composition," European Heart Journal-Cardiovascular Imaging 17(5), 481-491 (2016).

[5] Wolterink, J. M., Leiner, T., de Vos, B. D., van Hamersvelt, R. W., Viergever, M. A. and Išgum, I., "Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks," Medical Image Analysis 34, 123-136 (2016).

[6] Zhou, C., Chan, H. P., Chughtai, A., Patel, S., Hadjiiski, L. M., Wei, J. and Kazerooni, E. A., "Automated coronary artery tree extraction in coronary CT angiography using a multiscale enhancement and dynamic balloon tracking (MSCAR-DBT) method," Computerized Medical Imaging and Graphics 36(1), 1-10 (2012).

[7] Szilágyi, S. M., Popovici, M. M. and Szilágyi, L., "Automatic Segmentation Techniques of the Coronary Artery Using CT Images in Acute Coronary Syndromes," Journal of Cardiovascular Emergencies 3(1), 9-17 (2017).

[8] Wang, Y. and Liatsis, P., "Automatic segmentation of coronary arteries in CT imaging in the presence of kissing vessel artifacts," IEEE Transactions on Information Technology in Biomedicine 16(4), 782-788 (2012).

[9] Prasoon, A., Petersen, K., Igel, C., Lauze, F., Dam, E. and Nielsen, M., "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," Medical Image Computing and Computer-assisted Intervention, 246-253 (2013).

[10] Roth, H. R., Lu, L., Seff, A., Cherry, K. M., Hoffman, J., Wang, S., Liu, J., Turkbey, E. and Summers, R. M., "A new 2.5 D representation for lymph node detection using random sets of deep convolutional neural network observations," Medical Image Computing and Computer-Assisted Intervention, 520-527 (2014).

[11] Stollenga, M. F., Byeon, W., Liwicki, M. and Schmidhuber, J., "Parallel multi-dimensional lstm, with application to fast biomedical volumetric image segmentation," Neural Information Processing Systems, 2998-3006 (2015).

[12] Ciompi, F., de Hoop, B., van Riel, S. J., Chung, K., Scholten, E. T., Oudkerk, M., de Jong, P. A., Prokop, M. and van Ginneken, B., "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," Medical Image Analysis 26(1), 195-202 (2015).

[13] Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D. and Glocker, B., "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," Medical Image Analysis 36, 61-78 (2017).
[14] http://www.materialise.com/en/medical/software/mimics.
[15] Ronneberger, O., Fischer, P. and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," Medical Image Computing and Computer-Assisted Intervention, 234-241 (2015).
[16] Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R. and LeCun, Y., "Overfeat: Integrated recognition, localization and detection using convolutional networks," International Conference on Learning Representations, (2013).
[17] Long, J., Shelhamer, E. and Darrell, T., "Fully convolutional networks for semantic segmentation," Proc. CVPR, 3431-3440 (2015).
[18] Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition," International Conference on Learning Representations, (2014).
[19] He, K., Zhang, X., Ren, S. and Sun, J., "Deep residual learning for image recognition," Proc. CVPR, 770-778 (2016).
[20] He, K., Zhang, X., Ren, S. and Sun, J., "Identity mappings in deep residual networks," ECCV, 630-645 (2016).
[21] Bastien, F., Lamblin, P., Pascanu, R., Bergstra, J., Goodfellow, I., Bergeron, A. and Bengio, Y., "Theano: new features and speed improvements," arXiv: Symbolic Computation, (2012).
[22] American Heart Association, "A reporting system on patients evaluated for coronary artery disease," Circulation 51, 7-39 (1975).
[23] Zhang, J., Liu, M., Chen, S., Yuan, P., Li, J. and Shen, D., "Joint craniomaxillofacial bone segmentation and landmark digitization by context-guided fully convolutional networks," Medical Image Computing and Computer-Assisted Intervention, 720-728 (2017).
[24] Zhang, J., Liu, M. and Shen, D., "Detecting anatomical landmarks from limited medical imaging data using two-stage task-oriented deep neural networks," IEEE Transactions on Image Processing 26(10), 4753-4764 (2017).