



Left ventricle segmentation in the era of deep learning

Jelmer M. Wolterink, PhD^a

^a Image Sciences Institute, University Medical Center Utrecht, Utrecht, The Netherlands

Received Feb 20, 2019; accepted Feb 20, 2019
doi:10.1007/s12350-019-01674-3

See related article, pp. 976–987

Medical image analysis has recently been revolutionized through the widespread adoption of deep learning techniques.¹ This revolution has primarily been powered by supervised machine learning with convolutional neural networks (CNNs). CNNs typically operate on images, and provide one prediction per image sample (Figure 1), e.g., an image class label or quantitation of disease burden.² These networks contain a large number of parameters, which can be optimized or *trained* by repeatedly providing training samples and adjusting network parameters to minimize the discrepancy between predicted values and desired output values.

Deep learning has been used to analyze many types of medical images visualizing a wide range of anatomies,¹ including a large number of studies focusing on medical image segmentation. To this end, fully convolutional networks (FCNs) are often used.^{3,4} These networks are closely related to CNNs, but predict a value for each pixel or voxel, instead of a single prediction for the full image (Figure 1). Accurate segmentation models could allow fast and consistent quantitation of tissue volume and replace time-consuming manual annotation. An example application is preoperative planning in congenital heart disease patients, where deep learning-based segmentation of MR images could save hours of manual annotation.⁵ Head-to-head comparisons with conventional image analysis methods have established the superiority of

deep learning for medical image segmentation. For example, in the MR brain segmentation benchmark (MRBrainS),⁶ the first 16 ranked methods are all based on deep learning*. Similarly, all top ranking methods for CMR segmentation in the automatic cardiac diagnosis challenge (ACDC)⁷ used deep learning.

Successful deep learning applications in cardiac imaging include myocardial analysis in coronary CT angiography (CCTA) for identification of patients with functionally significant stenosis,⁸ and direct quantitation of left ventricular (LV) functional parameters in cardiac MR (CMR),⁹ among others.¹⁰ Nuclear cardiology has seen several applications of conventional machine learning,¹¹ but deep learning applications have thus far been scarce. A notable exception is the work of Betancur et al.¹² for identification of patients with obstructive disease based on myocardial perfusion SPECT (MPS) imaging. In this issue of the *Journal of Nuclear Cardiology*, Wang et al. present a feasibility study into deep learning-based segmentation of the LV myocardium in gated myocardial perfusion SPECT (MPS) images.¹³ An FCN is used to transform a 3D MPS image into a segmentation mask, labeling each voxel as part of the background, the region enclosed by the epicardial surface, or the region enclosed by the endocardial surface. The FCN is trained and evaluated using MPS images of 32 healthy subjects and 24 patients with mild, moderate, or severe myocardial ischemia. Experimental results show that in both groups, automatic segmentations of the LV myocardium overlap strongly with manual reference segmentations. The authors conclude that this deep learning-based method would allow quantitation of LV contractile functional indices within seconds and without human intervention.

The work by Wang et al. complements methods for deep learning-based LV segmentation in CCTA,⁸ CMR,⁷ and echocardiography.¹⁴ MPS images have several characteristics that facilitate fast and accurate segmentation: images are relatively small, they are

Reprint requests: Jelmer M. Wolterink, PhD, Image Sciences Institute, University Medical Center Utrecht, Utrecht, The Netherlands; J.M.Wolterink@umcutrecht.nl

J Nucl Cardiol 2020;27:988–91.
1071-3581/\$34.00

Copyright © 2019 American Society of Nuclear Cardiology.

* <http://mrbrains13.isi.uu.nl/results.php>.

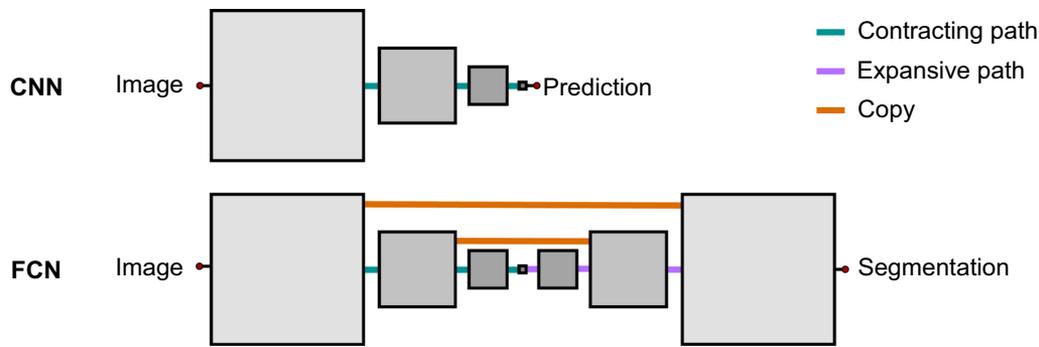


Figure 1. Schematic drawing of a convolutional neural network (CNN) and fully convolutional network (FCN). A CNN progressively reduces the representation size in a contracting path to provide one prediction per input image. In an FCN, this is followed by an expansive path that reuses information from the contracting path to provide one prediction per image pixel, e.g., a segmentation^{3,4}.

intrinsically 3D, and the contrast between the myocardium and the surrounding tissue is generally high. This enables the use of a 3D FCN architecture that considers a cropped 3D MPS volume with a fixed size of $32 \times 32 \times 16$ voxels and simultaneously predicts labels for all voxels in the image. The FCN architecture used in this study is based on the V-Net architecture proposed by Milletari et al.³ It contains a contracting path in which image information is extracted at multiple image scales, and an expansive path that combines this information into a segmentation. This allows the FCN to identify *what* is present *where* in the image. To quantitatively evaluate performance of the segmentation method, Wang et al. use a combination of criteria. First, the Dice similarity coefficient (DSC) for overlap and the Hausdorff distance for contour similarity are computed. Second, the agreement between automatic results and the reference standard is determined for LV myocardium volumes, and the LV ejection fraction (LVEF) is derived from the segmentation masks. To separate images that are used to optimize the FCN from images that are used for evaluation, a leave-one-out cross-validation setup is used. The FCN architecture used, the evaluation, and the experimental setup are generally in line with other works on image segmentation in other modalities.

Nevertheless, the study also has some limitations. The paper is positioned as a feasibility study, as the dataset is likely too small and homogeneous to evaluate generalizability to clinical practice. Although both normal subjects and patients with myocardial ischemia were included, no other pathologies were included, and the total number of 56 scans is small in comparison to the 1903 scans included in a previous study evaluating automatic LV segmentation in MPS.¹⁵ Moreover, previous experiences with deep learning-based systems have shown that performance may drop considerably

when transferring trained models from one center to another.¹⁶ In a potential future validation study, data from multiple centers could be included to assess generalizability to centers with different imaging protocols. Such a study could also include images acquired with stress, in addition to the images acquired at rest that were used in the current evaluation.

The FCN method was evaluated for both normal subjects and patients with myocardial ischemia. In each of these patient groups, a leave-one-out cross-validation experiment was performed. Although these experiments showed that the FCN architecture is capable of segmenting both kinds of scans, it is unclear whether a single trained model would be able to segment images of both groups of patients. Because cross-validation was performed separately in each group, models were either trained with only scans of healthy subjects, or only scans of patients with disease, which may have led to specialized models. In clinical practice, it will not be known beforehand whether patients are healthy or not, and a single trained model should be able to segment images from both patient groups. Such a model could be evaluated in a future study.

Performance metrics in the current study were determined based on agreement with manual reference segmentation in MPS. In addition, results were compared to commercially available software (Emory Cardiac Toolbox), which showed reasonable agreement regarding LVEF values ($r = 0.644$). This toolbox has previously been shown to overestimate LVEF compared to other software¹⁷ and CMR.¹⁸ To assess whether the proposed deep learning method mitigates or aggravates this overestimation, the comparison could be extended with additional software packages and an external reference standard in CMR. This might clarify whether the volumes determined by the model are correct, and

whether the method performs on par with or better than other automatic methods in MPS.

All FCN models were trained and evaluated using manually drawn contours. A potential limitation in the current study is that these contours were drawn by a single observer, which may have led to a bias. Supervised machine learning models are incentivized to replicate whatever is in the training set, and thus the model might learn to mimic the annotation style of the observer, including potential systematic errors made by this observer. Therefore, automatic results on the test set could be excellent when comparing with reference annotations by this observer, but agreement with other observers could be poorer. This effect has been found in subjective tasks like vessel segmentation in retinal fundus images,¹⁹ but may also have been present in the current study, as agreement with the reference standard was slightly higher for the automatic method than for a second observer. Thus, while the use of an automatic model may reduce interoperator variability, the model is still affected by and biased toward the observer setting the reference standard. In future work, this risk could be mitigated with a reference standard set by multiple observers in a consensus reading.

The FCNs were trained to perform a multiclass segmentation task, where each image voxel is assigned one label. In a typical multiclass segmentation task, reference labels are mutually exclusive: a voxel is expected to have one and only one label. For example, in the ACDC dataset, LV voxels are labeled as either myocardium or cavity, but never both.⁷ To encourage a deep learning model to assign a single label to each voxel, a softmax activation function is generally used, which imposes the sum of predicted probabilities for all

classes to be 1. However, classes in the current study were defined as follows: region within endocardial surface, region within epicardial surface, and background region. Hence, a voxel within the endocardial surface could have two equally correct reference labels: it is within the endocardial surface but also within the epicardial surface. The FCN architecture included a softmax output layer and was thus forced to choose between these two classes, which may have complicated optimization (Figure 2). In addition, the combination of a multiclass softmax activation function with a binary cross-entropy loss term is uncommon, as multiclass softmax outputs are more commonly used in combination with a categorical cross-entropy loss term. While a binary cross-entropy loss term only considers correct classification into the target class, a categorical cross-entropy term also considers misclassification between classes. In potential future development of the method, these methodological choices could be reconsidered to facilitate easier FCN optimization.

Despite these limitations, it is promising to see applications of deep learning permeate fields like nuclear cardiology to potentially reduce the workload of clinicians. Wang et al. have presented a feasibility study showing how deep learning could be used to segment MPS images. Results on a small dataset are promising, but several questions about the generalizability of the trained models remain to be answered in a larger evaluation study. This would most likely also include retraining of the FCN with a large and diverse training dataset.

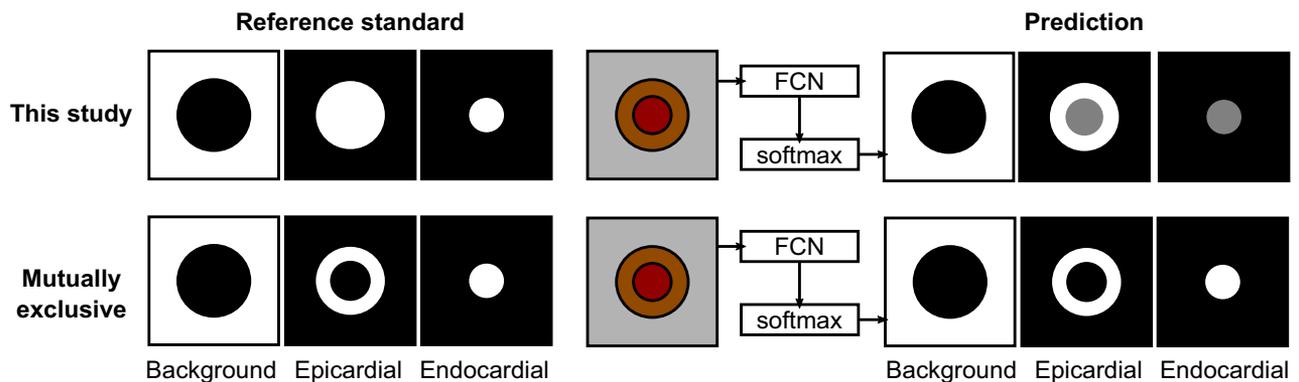


Figure 2. Schematic drawing of overlapping class definitions used in this study and mutually exclusive class definitions used in a typical multiclass setting. The definition used in this study requires the softmax activation function to choose between two equally likely classes for voxels within the endocardial surface.

Disclosure

The author has nothing to disclose.

References

1. Litjens G, et al. A survey on deep learning in medical image analysis. *Med Image Anal* 2017;42:60-88.
2. de Vos BD, Wolterink JM, Leiner T, de Jong PA, Lessmann N, Isgum I. Direct automatic coronary calcium scoring in cardiac and chest CT. *IEEE Trans. Med. Imaging* 2019;1.
3. Milletari F, Navab N, Ahmadi SA. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In 2016 Fourth International Conference on 3D Vision (3DV), 2016, pp. 565-571.
4. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Cham: Springer; 2015. p. 234-41.
5. Dou Q, et al. 3D deeply supervised network for automated segmentation of volumetric medical images. *Med Image Anal* 2017;41:40-54.
6. Mendrik AM, et al. MRBrainS challenge: Online evaluation framework for brain image segmentation in 3T MRI scans. *Comput Intell Neurosci* 2015;2015:1-16.
7. Bernard O, et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved? *IEEE Trans Med Imaging* 2018;37:2514-25.
8. Zreik M, et al. Deep learning analysis of the myocardium in coronary CT angiography for identification of patients with functionally significant coronary artery stenosis. *Med Image Anal* 2018;44:72-85.
9. Luo G, Dong S, Wang K, Zuo W, Cao S, Zhang H. Multi-views fusion CNN for left ventricular volumes estimation on cardiac MR images. *IEEE Trans Biomed Eng* 2018;65:1924-34.
10. Slomka PJ, Dey D, Sitek A, Motwani M, Berman DS, Germano G. Cardiac imaging: Working towards fully-automated machine analysis & interpretation. *Expert Rev Med Devices* 2017;14:197-212.
11. Shrestha S, Sengupta PP. Machine learning for nuclear cardiology: The way forward. *J Nucl Cardiol* 2018: 1-4.
12. Betancur J, et al. Deep learning for prediction of obstructive disease from fast myocardial perfusion SPECT: A multicenter study. *JACC Cardiovasc Imaging* 2018;11:1654-63.
13. Wang T et al. A learning-based automatic segmentation and quantification method on left ventricle in gated myocardial perfusion SPECT imaging: A feasibility study. *J Nucl Cardiol* 2019: 1-12.
14. Carneiro G, Nascimento JC. Combining multiple dynamic models and deep learning architectures for tracking the left ventricle endocardium in ultrasound data. *IEEE Trans Pattern Anal Mach Intell* 2013;35:2592-607.
15. Xu Y, et al. Automated quality control for segmentation of myocardial perfusion SPECT. *J Nucl Med* 2009;50:1418-26.
16. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLoS Med* 2018;15:e1002683.
17. Hambye A-S, Vervaeke A, Dobbeleir A. Variability of left ventricular ejection fraction and volumes with quantitative gated SPECT: Influence of algorithm, pixel size and reconstruction parameters in small and normal-sized hearts. *Eur J Nucl Med Mol Imaging* 2004;31:1606-13.
18. Soneson H, et al. Development and validation of a new automatic algorithm for quantification of left ventricular volumes and function in gated myocardial perfusion SPECT using cardiac magnetic resonance as reference standard. *J Nucl Cardiol* 2011;18:874-85.
19. Maninis K-K, Pont-Tuset J, Arbeláez P, Van Gool L. Deep retinal image understanding. Cham: Springer; 2016. p. 140-8.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.