Automatic Whole Heart Segmentation Using Deep Learning and Shape Context

Chunliang Wang(✉) and Örjan Smedby

School for Technology and Health (STH), KTH Royal Institute of Technology, Hälsovägen 11C, 14152 Huddinge, Stockholm, Sweden
{chunliang.wang,orjan.smedby}@sth.kth.se
http://www.kth.se/sth

Abstract. To assist 3D cardiac image analysis, we propose an automatic whole heart segmentation using a deep learning framework combined with shape context information that is encoded in volumetric shape models. The proposed processing pipeline consists of three major steps: scout segmentation with orthogonal 2D U-nets, shape context estimation and refining segmentation with U-net and shape context. The proposed method was evaluated using the MMWHS challenge data. Two sets of networks were trained separately for contrast-enhanced CT and MRI. On the 20 training datasets, using 5-fold cross-validation, the average Dice coefficients for the left ventricle, the right ventricle, the left atrium, the right atrium and the myocardium of the left ventricle were 0.895, 0.795, 0.847, 0.821, 0.807 for MRI and 0.935, 0.825, 0.908, 0.881, 0.879 for CT, respectively. Further improvement may be possible given more training data or advanced data augmentation strategy.

Keywords: Deep learning · Fully convolutional network
Heart segmentation · Shape context · Statistic shape model

1 Introduction

Cardiovascular disease (CVD) is currently the leading cause of death worldwide. Approximately one in five deaths is currently related to cardiac disease in Europe and the US. Nearly 500,000 deaths caused by CVD are reported every year in the US, and over 600,000 in Europe. Approximately half of men and one third of women over 40 years old will develop CVD [1]. Both computed tomography (CT) and magnetic resonance imaging (MRI) are important diagnostic tools for CVD, allowing medical doctors to directly access morphological and functional changes of the heart. In addition to their clinical use, they are also widely used in clinical trials to study the remodeling of the heart due to various cardiac diseases [2]. Several large cardiac cohort studies include CT and/or MRI imaging in their data collection protocols. One recent example is the Swedish cardiopulmonary bioimage study (SCAPIS), where CT, MRI and ultrasound images as well as blood samples will be collected from 30,000 subjects [3].

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 242–249, 2018.
https://doi.org/10.1007/978-3-319-75541-0_26

Both in clinical diagnostic procedures and in the data analysis process in clinical trials, segmentation of the heart is often one of the primary steps required to generate any useful quantitative measurements. Heart segmentation in 3D is known to be time-consuming if performed manually. Considerable efforts have been made to automate the procedure and to reduce the involvement of the radiologist during the segmentation. Promising results have been reported in the literature, in particular with the statistical shape-model based methods [4] and atlas-based methods [5]. On the other hand, deep learning based methods are gaining more and more attention due to their superior performance in several image segmentation challenges [6,7].

In this study, we propose a hybrid method that attempts to integrate the statistical shape model into the deep learning process. This is done by feeding the estimated volumetric shape models of several cardiac structures as context layers to a fully convolutional network (FCN). In our preliminary experiments, the shape context layers seem to increase the segmentation accuracy of most structures compared to the plain FCN, when validated on a publicly available database of 20 contrast-enhanced CT scans and 20 contrast-enhanced MR scans with manually created ground truth.

2 Methods

As illustrated in Fig. 1, the proposed processing pipeline consists of three major steps: scout segmentation with orthogonal 2D U-nets, shape context estimation and refining segmentation with U-net and shape context. The individual steps are explained in the following sections.

2.1 2.5D Segmentation Using Orthogonal U-Nets

Convolutional neural networks (CNN) have been successfully used in many segmentation tasks [6]. In this study, we adopted a somewhat more sophisticated FCN architecture, called U-net, proposed by Ronneberger et al. [7]. In FCNs, the fully connected layers of classical CNNs are replaced by convolutional layers [6], which allows FCNs to be applied to images of any size and output label maps proportional to the input image. In combination with "skips" and up-sampling layers [6], the FCNs are often designed to produce the same size output image as the input image, thus eliminating the need for the time-consuming sliding window process used by classical CNN-based methods. To perform segmentation in 3D volumes, we used three U-nets that were trained independently to segment multiple structures in 2D slices acquired in three orthogonal projects, i.e., in axial, coronal and sagittal views. The final probability map of each structure is generated by averaging the outputs of these three U-nets.

2.2 Shape Context Generation

In this study, we use the volumetric statistical shape models proposed by proposed by Leventon et al. [8]. As described in [8], the statistical model is created

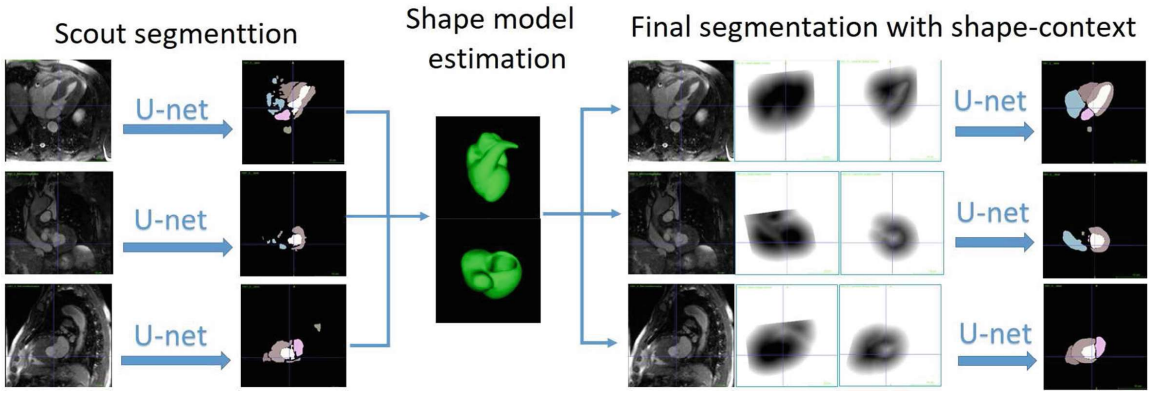

Fig. 1. An overview of the segmentation pipeline

by taking the mean of the signed distance functions of each segmented region and n prominent variations extracted via Principal Component Analysis (PCA). Then the model M that matches the current case is estimated by solving a level set function

$$\frac{\partial \phi}{\partial t} = \alpha F(x) + \beta M(T(x)) + \gamma \kappa(x) \left| \nabla \phi \right| \tag{1}$$

where F is the image force, which is simply a threshold function on the probability maps from U-nets in this case, M is the statistical model as a weighted sum of the mean shape and PCA components, T is the global transformation and κ is the mean curvature. The transformation T and the weighting factors of PCA components are updated iteratively by minimizing the squared distance between the model and the level set function, which is also a signed distance map. The weighting factors, α, β and γ are determined empirically. To speedup the process, a fast level set method using coherent propagation is used [9].

Since the heart is a complex structure with four chambers and several vessels attached to it, we use a hierarchical approach to model the whole heart, similar to the method reported in [10]. At the higher level, we create the overall shape model of the surface of the whole heart (Fig. 2A), and at a lower level the detailed structures are modeled separately with the relative position to the parent shape model preserved (e.g. Fig. 2C). To save computation time and memory consumption in this preliminary study, we merged the right ventricle with the myocardium of the left ventricle into a single structure (Fig. 2B) and ignored all other second level structures. For simplicity, we refer to this fused structure as "ventricle surface".

2.3 Shape-Context Guided U-Net

After estimating the shape models that fit to the probability map of the scout segmentation, the distance maps of the heart surface and the ventricle surface, together with the input images, are fed into another three U-nets that will perform the segmentation again in three orthogonal views. The architecture of these

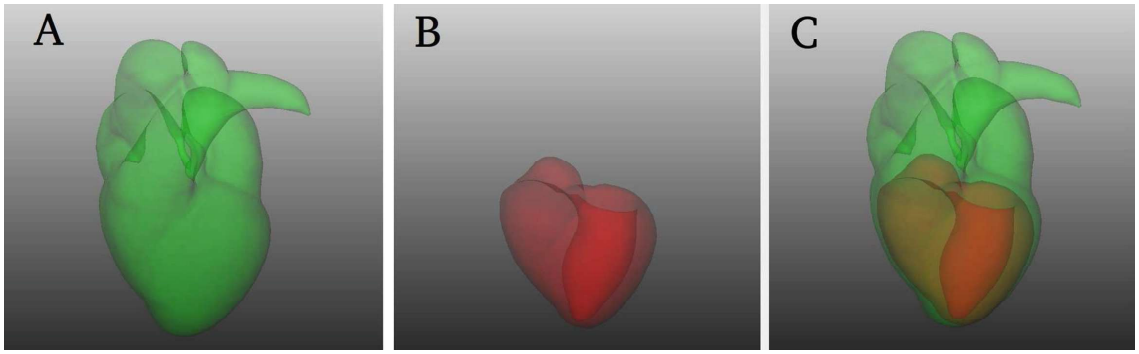

Fig. 2. Shape models used in this study. A, heart surface. B, ventricle surface. C, the relative position of two level models. (Volumetric shape models are used in this study; the illustrations above are created by taking the iso-surface of level 0)

U-nets is identical to the ones used in the scouting step, but retrained from scratch. The final probability map of each structure is again generated by taking the mean value of the outputs from these three U-nets.

2.4 Implementation Details

Our U-net implementation was based on the Keras framework with Theano backend (http://keras.io). The U-net architecture is identical to the one proposed in the original paper, with a little modification to the size of the input image. In our implementation, the segmentation is performed at 1 mm isotropic resolution, which means the CT images are down-sampled to about half the original resolution before being processed. This down-sampling step is introduced only to reduce the GPU memory consumption and reduce training time. The same resolution is used for both the scout segmentation and the refinement segmentation. For the MR scans, an additional landmark detection step using random forest [11] is used to first crop the image into a smaller region of interest ($256 \times 256 \times 256$) to reduce the size of the input image to the U-nets.

Different data normalization methods are used for CT and MRI images and the shape context images. For the CT image and the shape context channels, the voxel intensity is simply divided the group standard deviation (SD) without changing the reference point of 0. For MRI, the intensity of all MRI scans are normalized first individually and then together. During the individual normalization, the lower 5% cutting point of each subject's histogram is mapped to 0 and the upper 5% cutting point is mapped to 1.0. In the group normalization, all subjects are normalized together by subtracting the group mean and divided by group SD, i.e., the normalized images will have 0 mean and SD of 1.

The categorical cross-entropy is used as the loss function for multi-structure segmentation. Stochastic gradient descent (SGD) is used as the optimizer in all training process, and the number of epochs is fixed at 150 for all U-nets. The shape models were created using 10 randomly selected CT cases and used for both the CT and the MRI experiments.

3 Results

The proposed method was evaluated using the training data of the Multi-Modality Whole Heart Segmentation (MMWHS) challenge, which consists 20 contrast-enhanced CT scans and 20 contrast-enhanced T1-weighted MRI scans. In all cases, seven structures, namely the left ventricle blood cavity (LV), the right ventricle blood cavity (RV), the left atrium blood cavity (LA), the right atrium blood cavity (RA), the myocardium of the left ventricle (Myo), the ascending aorta (AA) and the pulmonary artery (PA) were manually delineated. The U-nets were trained separately for the CT images and MRI images. The evaluation was done using five-fold cross-validation, in each fold 16 cases were used for training and 4 cases were used for testing. Shape models, however, were not recreated in every fold.

The average accuracy (Dice coefficient) of the scout segmentation and the refined segmentation is compared in Tables 1 and 2 for MRI images and CT images, respectively. In most cases, the combined approach yielded somewhat higher accuracy than U-nets alone. Table 3 shows the segmentation accuracy of the scout segmentation on the testing datasets of the MMWHS challenge. The accuracy of the refined segmentation with shape context was not submitted by the submission deadline due to an implementation error. Because re-submission is not allowed after the deadline, the scores of the entire pipeline on the testing datasets are not available by the time this paper being accepted. Figure 3 shows the scout segmentation and refined segmentation results of three example cases. The overall processing time for both CT and MRI was about 5–7 min on a personal computer with an Nvidia GTX 1080 graphic card.

Table 1. Comparing the accuracy (Dice coefficient) of the scout segmentation with U-net and the final segmentation with U-net and shape context in MR images

Structures	U-net	U-net & Shape context
Myocardium	0.785 ± 0.091	0.807 ± 0.059
Left atrium	0.877 ± 0.026	0.847 ± 0.061
Left ventricle	0.873 ± 0.082	0.895 ± 0.057
Right atrium	0.749 ± 0.225	0.821 ± 0.087
Right ventricle	0.688 ± 0.205	0.795 ± 0.102
Ascending aorta	0.708 ± 0.257	0.679 ± 0.180
Pulmonary artery	0.622 ± 0.232	0.743 ± 0.204

4 Discussion and Conclusion

Overall, adding the shape context as additional input to the U-nets seems to increase the segmentation accuracy, especially in the case of MRI images where the Dice coefficients of the original U-nets are relatively low compared to the

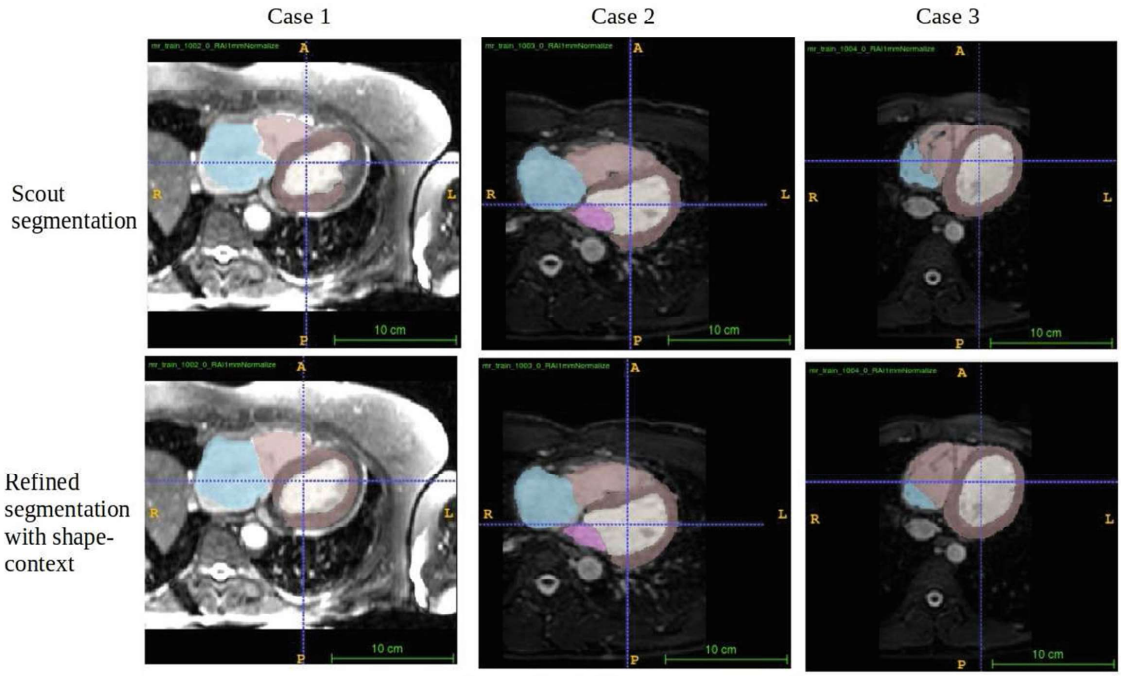

Fig. 3. Three example cases. The upper row shows the scout segmentation results, the lower row shows the refined results with shape context

Table 2. Comparing the accuracy (Dice coefficient) of the scout segmentation with U-net and the final segmentation with U-net and shape context in CT images

Structures	U-net	U-net & Shape context
Myocardium	0.872 ± 0.084	0.879 ± 0.068
Left atrium	0.903 ± 0.062	0.908 ± 0.067
Left ventricle	0.911 ± 0.035	0.935 ± 0.046
Right atrium	0.858 ± 0.105	0.881 ± 0.082
Right ventricle	0.884 ± 0.075	0.825 ± 0.082
Ascending aorta	0.956 ± 0.021	0.959 ± 0.023
Pulmonary artery	0.830 ± 0.125	0.815 ± 0.131

Table 3. Dice coefficient of the scout segmentation on the testing datasets

Structures	MRI	CT
Myocardium	0.728 ± 0.142	0.874 ± 0.051
Left atrium	0.832 ± 0.093	0.908 ± 0.044
Left ventricle	0.855 ± 0.136	0.908 ± 0.059
Right atrium	0.782 ± 0.131	0.855 ± 0.064
Right ventricle	0.760 ± 0.174	0.806 ± 0.082
Ascending aorta	0.771 ± 0.219	0.835 ± 0.231
Pulmonary artery	0.578 ± 0.246	0.677 ± 0.240
Whole heart	0.792 ± 0.246	0.866 ± 0.048

scores for CT images. For CT images, small improvements were also observed for a majority of the structures; however, the segmentation accuracy of RV declined considerably when the shape context channels were added. This may be due to an over-fitting problem, as the U-nets rely too much on the shape context channels. Adding dropout layers may be helpful to overcome this problem. Also, including more augmented training samples that are designed to model the uncertainty of the shape context layer could be even more important for the network to learn to cope with the cases where shape context is not very precise.

In previous studies, researchers have also proposed the auto-context method [12], which uses the output of the first classifier to train a second classifier. The advantage of using the shape context is that the statistical shape could help to eliminate some of the false positive regions that are produced by the first classifier but do not fit to the overall shape of the heart or the ventricles. However, direct comparison between auto-context approaches and the proposed shape context was not performed due to time constraints.

In conclusion, we have proposed a hybrid image segmentation methods based on deep neural network and statistical shape modeling. In our preliminary experiments, the proposed method delivered promising results on cardiac structure segmentation in both CT and MRI.

Acknowledgments. This research has been partially funded by the Swedish Research Council (VR), grant no. 2014-6153, and the Swedish Heart-Lung Foundation (HLF), grant no. 2016-0609.

References

1. Lloyd-Jones, D.M., Larson, M.G., Beiser, A., Levy, D.: Lifetime risk of developing coronary heart disease. Lancet **353**(9147), 89–92 (1999)
2. Zhang, X., Cowan, B.R., Bluemke, D.A., Finn, J.P., Fonseca, C.G., Kadish, A.H., Lee, D.C., Lima, J.A.C., Suinesiaputra, A., Young, A.A., et al.: Atlas-based quantification of cardiac remodeling due to myocardial infarction. PLoS ONE **9**(10), e110243 (2014)
3. Bergström, G., Berglund, G., Anders Blomberg, J., Brandberg, G.E., Jan Engvall, M., Eriksson, U.F., Flinck, A., Hansson, M.G., et al.: The swedish cardiopulmonary bioimage study: objectives and design. J. Intern. Med. **278**(6), 645–659 (2015)
4. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. IEEE Trans. Med. Imaging **27**(11), 1668–1681 (2008)
5. Zhuang, X., Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. Med. Image Anal. **31**, 77–87 (2016)
6. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431–3440 (2015)
7. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

8. Leventon, M.E., Grimson, W.E.L., Faugeras, O.: Statistical shape influence in geodesic active contours. In: 2000 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 316–323. IEEE (2000)
9. Wang, C., Frimmel, H., Smedby, Ö.: Fast level-set based image segmentation using coherent propagation. Med. Phys. **41**(7), 073501 (2014)
10. Wang, C., Smedby, Ö.: Automatic multi-organ segmentation in non-enhanced CT datasets using hierarchical shape priors. In: Proceedings of the 22nd International Conference on Pattern Recognition (ICPR). IEEE (2014)
11. Wang, C., Wang, Q., Smedby, Ö.: Automatic heart and vessel segmentation using random forests and a local phase guided level set method. In: Zuluaga, M.A., Bhatia, K., Kainz, B., Moghari, M.H., Pace, D.F. (eds.) RAMBO/HVSMR -2016. LNCS, vol. 10129, pp. 159–164. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-52280-7_16
12. Tu, Z., Bai, X.: Auto-context and its application to high-level vision tasks and 3D brain image segmentation. IEEE Trans. Patt. Anal. Mach. Intell. **32**(10), 1744–1757 (2010)