

2131

Automatic Cardiac Resting Phase Detection for Static Cardiac Imaging Using Deep Neural Networks

Seung Su Yoon¹, Elisabeth Hoppe¹, Michaela Schmidt², Christoph Forman², Puneet Sharma³, Christoph Tillmanns⁴, Andreas Maier¹, and Jens Wetzl²

¹Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, ²Magnetic Resonance, Siemens Healthcare GmbH, Erlangen, Germany, ³Siemens Medical Solutions USA, Inc., Princeton, NJ, United States, ⁴Diagnostikum Berlin, Berlin, Germany

Synopsis

To perform static cardiac imaging, manual inspection of CINE images is currently necessary to detect a quiescent window within the cardiac cycle. We propose an automated system using two chained Deep Neural Networks to determine localized end-systolic and end-diastolic resting phases. The first network finds a region of interest (e.g. RCA, right or left atrium) and the second determines a quantitative motion curve for this region. Training and evaluation was performed on data from volunteers and patients acquired on different scanners and field strengths and a comparison to manually annotated resting phases showed accuracy in the range of 35ms.

Introduction

Static cardiac imaging such as whole-heart coronary imaging requires the determination of a quiescent window within the cardiac cycle for data acquisition, either the end-systolic or end-diastolic resting phase (RP). In current practice, this is done by visual inspection of a CINE image series, e.g. the four-chamber-view (4CHV) and requires user experience. A further complicating factor for the detection of RPs is that different cardiac anatomies rest at different times of the cardiac cycle, so ideally a localized RP for the anatomy of interest should be determined. A previously published approach¹ has demonstrated automated RP detection, but this method computes RPs based on the entire field-of-view and does not provide localized RPs for specific anatomies. In this study, we investigate a new automated system which detects the RPs for several localized regions of the heart without manual user interaction.

Methods

System Overview: The proposed prototype system (Figure 1) consists of two chained Deep Neural Networks. While the first network identifies regions of interest (ROIs) in a 4CHV either by landmark detection or segmentation, the second performs motion detection within each ROI. As a preprocessing step, all images were interpolated to a fixed spatial and temporal resolution of 224x224x32 to be resolution-independent. The first network provided landmarks or segmentations from which rectangular ROIs were derived. The model for landmark detection is shown in Figure 2 and segmentation used a state-of-the-art U-net². For landmark-based ROIs, prior knowledge about the size of the anatomy was used to define a bounding box centered at the landmark position. For segmentation-based ROIs, the tightest fitting rectangular bounding box enclosing the contours was selected. In this abstract, we focused on landmark detection for the right coronary artery (RCA) and segmentation of the right atrium (RA) and left atrium (LA). The second network used the ROIs at each time point as input and computed a motion curve with one motion value per time point, see Figure 3. Ground truth motion values were derived from the ROIs by performing image registration³ and averaging the magnitudes of the motion vectors within the ROI.

Classification of RPs: Manual annotation of RPs was performed on 21 datasets. These were then used for the calculation of sensitivity and specificity in order to obtain a threshold for RP detection. Since the motion curve describes the average motion in millimeters, RPs could then be determined with an absolute threshold, e.g. < 0.2 mm average motion (Figure 4).

Experiments: The two networks were trained and evaluated on 4CHV CINE datasets from 960 volunteers/patients acquired on 1.5 T and 3 T clinical MRI scanners (MAGNETOM Avanto, Trio, Aera, Skyra; Siemens Healthcare, Erlangen, Germany). The full dataset was split into 70 % training, 15 % validation and 15 % testing data subsets. Statistics for each subset are given in Figure 5. For RCA detection, the 3D-DenseNet was validated by calculating the mean and standard deviation of the distance of ground truth and detected landmarks for each time point. For motion detection, the MDN was validated by calculating the root-mean-squared error (RMSE) between the predicted motion curves and the ground truth. Additionally, the RMSE for start and end points of manually annotated and automatically detected RPs for the RCA in milliseconds was computed.

Results

For RCA detection, the mean error was 7.2 ± 1.5 mm. The RMSE of the MDN motion curve was 0.12 ± 0.06 mm. The RMSE for start and end points was 35 ± 17 ms and 30 ± 18 ms (end-systolic RP) and 36 ± 17 ms, 37 ± 16 ms (end-diastolic RP). Sensitivity and specificity for the test sets were 78 % and 92 %. Qualitative results of the first and second network are shown in Figures 1 and 4, respectively.

Discussion

The automatic RCA detection was sufficiently accurate to find a bounding box for motion detection. In addition, the proposed MDN showed a good match between expert-annotated and automatically detected RPs. Compared to a previous automated RP detection approach¹, this method is an improvement both in terms of localization of the RPs, as well as providing a quantitative motion curve which enables the use of an interpretable threshold for RP classification. While we focused on RCA, LA and RA, this approach can be extended to detect other anatomies.

Conclusion

Automated RP detection for multiple ROIs in 4CHV images was successfully performed with high accuracy by the proposed system. Future work will focus on evaluating further anatomical structures as well as clinical validation.

Acknowledgements

No acknowledgement found.

References

1. Piccini, Davide, et al. Automated Cardiac Resting Phase Detection in 2D cine MR images for Acquisition Window Selection in High-Resolution Coronary MRI. *Prod. Intl. Soc. Mag. Reson. Med.* 25 (2017) 2862
2. Ronneberger, Olaf, et al. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham, 2015.
3. Chefd'hotel C, et al. Flows of Diffeomorphisms for Multimodal Image Registration. In: 2000 IEEE International Symposium on Biomedical Imaging, July 7–10, 2002, Washington D.C., USA. pp 753–756.
4. Huang, Gao, et al. Densely Connected Convolutional Networks. *CVPR*. Vol. 1. No. 2. 2017.
5. Xie, Saining, et al. Rethinking spatiotemporal feature learning: Speed-accuracy trade-offs in video classification. *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.
6. Kingma, Diederik P., et al. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
7. Robbins, Herbert, et al. A stochastic approximation method. *Herbert Robbins Selected Papers*. Springer, New York, NY, 1985. 102-109.

Figures

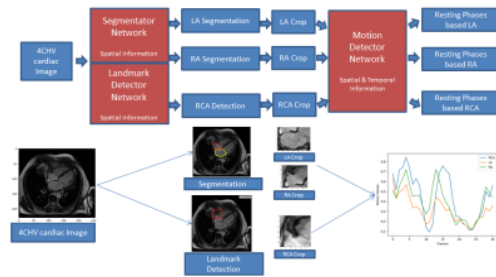


Figure 1: The structure of the proposed system for the detection of resting phases. An abstract representation of the system is given in the top row while the bottom row shows example inputs and outputs of the components.

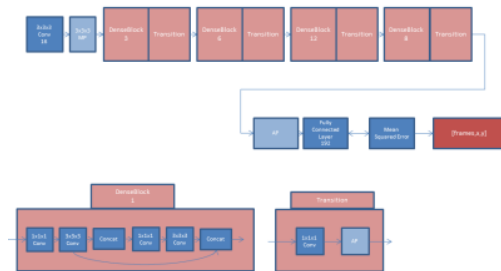


Figure 2: For landmark detection, a 3D-DenseNet based on previous work⁴ with the grow rate 12 was trained using Adam⁶ with $\lambda = 10^{-2}$. The number of dense layer sizes was configured as 3-6-12-8. In the bottom row, an example of a DenseBlock with one dense layer and a transition block is shown. Before and after 1x1x1 convolutional layer, a 3D-Batch Normalization and ReLU activation layer were added. For the downsampling at the beginning, a Max Pooling layer (MP) was used after the first convolutional layer. A global average pooling, fully connected layer and mean-squared error layer were used for landmark regression.

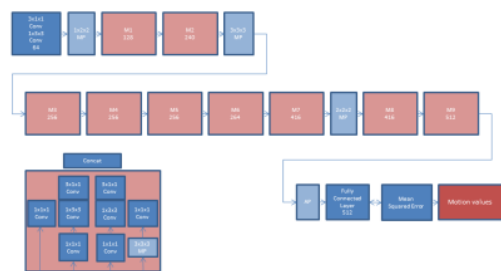


Figure 3: For motion detection, a network based on the separable Spatial-Temporal Network⁵ was trained using SGD⁷ with $\beta = 0.9$ and $\lambda = 10^{-5}$. Separable convolution blocks consisting of one spatial convolution (1x3x3) and one temporal convolution (3x1x1) were used to extract spatial and temporal features. Each convolution block was followed by Instance Normalization and ReLU activation layers. For downsampling, Max Pooling layers (MP) were used. The number of filters in each convolution block or module is listed. The average pooling (AP) and a fully connected layer were used for the regression of motion values at the end.

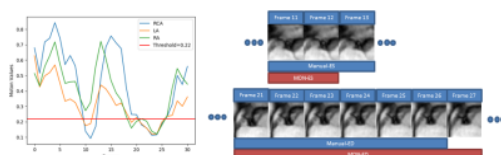


Figure 4: Left: A plot of motion curves for RCA, LA and RA generated by the MDN, as well as a threshold for RCA RP detection. Right: RCA end-systolic (ES) and end-diastolic (ED) RPs from manual annotation and threshold-based classification from the MDN output.

	Number of volunteers/patients	Age	Gender	Field Strength
Training Set	682	50.9±21.3	60% male, 40% female	28% 1.5 T, 72% 3 T
Validation Set	147	57.1±17.5	66% male, 34% female	14% 1.5 T, 86% 3 T
Testing Set	131	57.1±18.2	67% male, 33% female	13% 1.5 T, 87% 3T

Figure 5: Statistics of datasets used for training, validating and testing in both networks.

Proc. Intl. Soc. Mag. Reson. Med. 27 (2019)
2131