Assessment of cardiac valve motion on time-resolved MRI images using deep learning

Master's Thesis in Medical Engineering

submitted
by

Maria Monzon Ronda

born 24.05.1995 in Bilbao

Written at

Lehrstuhl für Mustererkennung (Informatik 5)
Department Informatik
Friedrich-Alexander-Universität Erlangen-Nürnberg.

Advisor: Jens Wetzl, M.Sc., Daniel Giese, Ph.D., Michaela Schmidt (Siemens Healthineers GmbH) Seung Su Yoon, M.Sc. , Prof. Dr.-Ing. Andreas Maier (Pattern Recognition Lab) and Prof. Dr. rer. nat. Armin Nagel (MR Physics Group, Institute of Medical Physics)

Started: 27.11.2019

Finished: 01.07.2020

Übersicht Fortschritte in der kardiale Bildgebung helfen genaue und rechtzeitige Diagnosen zu stellen. Die Kardio-Magnetresonanztomographie (CMR) ist ein bevorzugtes, nicht-invasives Verfahren, welches sowohl anatomische als auch funktionelle Informationen liefern kann. Aufgrund der Bewegung des Herzens durch den Herzzyklus, stellt die Bildgebung der Klappenströmungen bei der zeitaufgelösten CMR-Bildgebung jedoch eine besondere Herausforderung dar. Für die Darstellung der Klappenbewegungen würde einen speziellen Prescanund die Lokalisierung der Herzklappen und der Klappenebene, die manuell von Experten durchgeführt werden benötigen. Um die zeitaufwändige Darstellung der Klappenbewegungen und den Arbeitsablauf zu verbessern, wird in dieser Arbeit eine eine vollständig automatisierte Methode für die Klappendetektion vorgestellt, welche auf Neuronalen-Netzwerken (Deep Learning) basiert. Das Ziel hierbei ist die Evaluierung von Deep-Learning basierter Landmarken-Tracking-Methoden, um die Bewegung der Mitralklappen während des gesamten Herzzyklus auf zeitaufgelösten MR-Bildern zu extrahieren. DieSchichtverfolgung wird eine präzisere Morphologie und Flussschätzung ermöglichen, was die Diagnose der diastolischen Dysfunktion möglicherweise verbessern wird. Anhand der annotierten Daten werden verschiedene Deep Learning Architekturen trainiert und ausgewertet. Schlieŝlich wird das optimale Netzwerk ausgewählt und präsentiert. Die Ergebnisse dieser Arbeit zeigen, dass das Deep Learning System eine hohe Genauigkeit der Landmarkenverfolgung in räumlicher und zeitlicher Dimension und der Bewegungsabschätzung aufweist.

Abstract Advances in cardiac imaging techniques help to achieve early and accurate diagnoses. Cardiac Magnetic Resonance (CMR) imaging is the gold standard non-invasive technique able to provide both anatomical and functional information. However, valvular flow imaging is specially challenging on time-resolved CMR imaging due to the motion of the heart through the cardiac cycle. In clinical setting, valve motion model would need a dedicated pre-scan as well as user manual invention for localizing the valvular landmarks and plane. In order to improve the time-consuming slice motion assessment, we proposed a fully slice-following method based on Deep Learning. The objective of the presented work is to evaluate deep learning-based landmarks tracking methods to extract the motion of the mitral valves throughout the cardiac cycle on time-resolved MR images. The imaging slice tracking will enable a more precise morphology and flow estimation, potentially improving the diagnosis of diastolic dysfunction. Different deep learning architectures are trained and evaluated using the annotated data. Finally, the optimal network is selected and presented. The outcomes of the thesis show that the deep learning system has high accuracy in terms of temporal landmark tracking and motion assessment.

Chapter 7

Summary

Background Valvular flow imaging is specially challenging on time-resolved Cardiovascular Magnetic Resonance (CMR) imaging due to the motion of the heart through the cardiac cycle. The CMR fixed slice acquisition will not depict the same tissue in all cardiac phases, since mitral valve moves in and out of the acquisition slice. It is important to mention that quantified mitral valve flow assessment will be incorrect unless the acquisition slice position is re-configured during the acquisition process. To overcome these limitations, the slice position can be updated for each cardiac phase if the valve motion is known prior to the examination.
However, the motion assessment is clinically expensive, due to the need of a dedicated pre-scan, expert manual intervention and image based algorithm to update the valve-plane during the acquisition. Despite the clinical importance of an improved diagnosis regarding cardiac dysfunction, CMR temporal imaging is yet usually performed at fixed slice positions throughout the cardiac cycle. Therefore, a fully automated slice-tracking algorithm would overcome the time consuming and observer dependent limitations of the current valve-tracking techniques in clinical practice.

Objective The presented work aimed to develop a novel method for the task of valve motion assessment for prospective slice tracking in CMR temporal image (CINE) acquisition. The objective this thesis is to evaluate deep learning-based landmarks tracking methods to extract the motion of the mitral valves throughout the cardiac cycle on time-resolved CMR four-chamber-view (4CHV) images. The fully automated CNN based algorithm would improve the current slice following acquisition workflow, which needs from manual user intervention. The automatic imaging slice tracking will enable a more precise morphology and flow estimation, potentially improving the diagnosis of diastolic dysfunction.

Dataset The dataset used in this work is a composed of 1.5T 4CHV CINE CMR series from the Cardiac Atlas Project landmark detection challenge. All patients in the training set are accompanied by the mitral valve landmark coordinates annotated by an experienced analyst. As a preprocessing step, the CINE series were interpolated into a fixed and temporal resolution and horizontally flipped to have all the images oriented with the apex of the heart upwards. Finally,the input images intensity were normalized to 0-1 range values. After cleaning the data, the dataset contain 87 series that are split into 70% for training and the rest 30% equally between test and validation. To increase the variability of the dataset, online data augmentation was performed when training the network in forms of shift, center cropping, rotation, Guassian noise addition, contrast enhancement and Gaussian blurring.

System Overview The motion assesment system is composed of four main stages. The preprocessed CINE dataset are forwarded as input to a convolutional based neural network (CNN). The CNN is designed to regress the mitral valve-annulus landmarks for each time-frame points. The predicted temporal coordinates are translated into mm-space. Finally, the valvular plane motion can be derived from the two predicted landmark distance. For the CNN model , three different approaches were tested and evaluated.

Direct Coordinate Regression Network The CNN learns to output directly the output pixel given an as (x, y) coordinate. The fully connected output layer performs the actual regression to coordinates from feature maps. The direct landmark regression model is based on the state-of-the-art adapted 3D DenseNet-169. The DenseNet increases the non-linearity with the concatenation of dense concatenated blocks. The main computation of dense block is the $3 \times 3 \times 3$ convolution, Batch Normalization and ReLU activation. The training was performed by regressing of the Mean Square Error (MSE) loss between the predicted and ground truth coordinates for each frame.

Heatmap Regression Network The model is trained to produce output images that resemble the synthetic heatmaps. For regressing heatmaps a necessary pre-processing step is the grouns truth target generation. This is achieved by rendering a Gaussian 2D blob around the landmark position. The standard deviation of this Gaussian is empirically fixed to a value of $\sigma = 7$ pixels. The heatmap regression guides the fully convolutional architecture training to localize using probability maps. The basic proposed architecture for heatmap regression is the modified 3D UNet with $23\,889\,000$ trainable parameters. The feature extraction relies on the $3 \times 3 \times 3$ convolutions kernel were followed by LReLU activation and a batch normalization layer. The pooling layers had an spatial stride of 2×2 was applied, in order to only downsample the image

spatial dimensions. The configuration of the number of filters at the feature extraction phase is 32-64-128-256-512 until to reach the bottleneck. The MSE loss that matches the ground truth Gaussian heatmaps with the output heatmaps is penalized during training.
Finally the coordinates were inferred with the so-called soft-argmax layer, which is a differentiable maxima response retrieval implementation. This function applies a 2D spatial softmax activation and computes the coordinates as the average mean coordinate of the probability maps.

Cascaded CNN Network system The final proposed system comprises two chained CNNs based on heatmap regression approach: Localization Network + Landmark Detection Network. The task of the localization CNN model is to detect the landmarks only in the first temporal frame of the full 4CHV CINE series. The implemented architecture is a Res-Unet 2D model 49627 tunable parameters. The feature extraction was performed with dilated convolutions on the bottleneck and residual convolutional block. The block was formed by a 3×3 convolution - Batch Normalization - ReLU activation - 3×3 convolution - Batch Normalization -ReLU.
As an intermediate step to between the two networks a further image processing step was performed to mitigate the variance of the reduced dataset. Based on the heatmap localization region, the image were rotated to be horizontally aligned. After they were interpolated to a common 0.5 mm-resolution space and cropped to a 256x256 fixed window.
These cropped patches are the input for the 3D landmark detection network which will be predict the landmarks through the complete cardiac cycle. As the modified 3D Unet is the backbone architecture for heatmap regression network, analogous to the one employed in the heatmap regression approach. For this network, many hyper-parameter configurations implemented and evaluated. We further tested different convolutional blocks, loss functions and Gaussian standard deviation. The final proposed method was trained with an decaying standard deviation. First the ground truth heatmaps standard deviation was fixed to 32 pixels and it decreases exponentially along the training epochs.
In order to further refine these predictions, we designed a processing pipeline. For inference, the coordinates are computed applying the soft-argmax activation to the heatmaps, which constitute the initial guess for the non-least square Gaussian fitting. An iterative surface fitting algorithm refines the centroid sub-pixel landmark coordinate. This additional step enables not only a more accurate detection of the valves, but also quantifies smoother valve motion.

Experimental results The evaluation on the approaches was done in the test data composed of 13 CINE series. Although the training evaluation was computed on heatmaps, in testing only the landmark coordinates values were validated. The total accuracy was measured by means of the root mean square distance. We also defined further the evaluation methods for the temporal smoothness quantification. Based on temporal derivations, we proposed the metrics roughness and total variation of the slope. Several experiments were done to evaluate which of the proposed systems was more suitable for temporal landmark tracking. The choice of the best model for each model is done by evaluating several network configuration. The results can be summarized as the following:

- It is proven that heatmap based regression approach for landmark detection yields to a superior performance that direct coordinate regression. The output of the network represent the probability of the landmark motion over the cardiac cycle.
- In addition, we also compared spatio-temporal convolutions. In all the approaches and architectures, the 3D convolution showed showed outstanding performance compared only to spatial or separable 3D convolutions. Especially 3D networks predicted temporal smoother landmarks.
- The motion is better modeled by the two-stages architectures. The final proposed has a superior accuracy and a lower values for roughness. The complete system has an accuracy of 1.66 ± 0.75 mm. The roughness metric has a similar value as the ground truth annotations $R = 0.085 \pm 0.045$ and the total slope variation value is $TV_{slope} = 0.53 \pm 0.28$.
- Qualitative results in an small dataset indicates that the network is generalizable to other datasets despite the limited sized dataset employed for training and testing.

We found that training the two-stages CNN regressing the heatmap loss and transforming in inference into numerical coordinates using a the soft-argmax layer is very effective for the heart landmark detection.

For further test, we evaluated our approach in a really few dataset for qualitative assessment as we do not have the expert-annotated labels. Although the assessment is only qualitative, the final proposed system gives accurate results. The proposed 3D CNN model is capable of learning the spatio-temporal correlation of the cine series.

Conclusion and outlook In conclusion, the proposed system enables the valve tracking detection over time and therefore smooth valve motion assessment. Future work would focus on extension on the algorithm to 2CHV valve landmark detection. Evaluation with more test data and further refinement of the network. An integration of a single CNN could be convenient for faster inference time. Finally, when the validity of the method is proved , the scanner integration would be the next step.