MANET: MITRAL ANNULUS POINT TRACKING NETWORK IN CARDIAC MAGNETIC RESONANCE

Jianguo Chen[1*], *Xulei Yang*[1*], *Shuang Leng*[2,3], *Ru-San Tan*[2,3], *Zeng Zeng*[1], *Liang Zhong*[2,3†]

[1] Institute for Infocomm Research, Agency for Science, Technology and Research, 138632, Singapore.
[2] National Heart Research Institute Singapore, National Heart Centre Singapore, 169609, Singapore.
[3] Duke-NUS Medical School, 169857, Singapore.

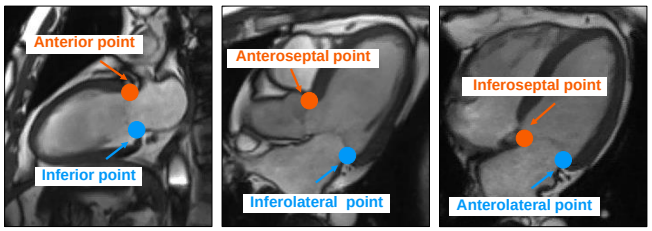
ABSTRACT

Cardiac magnetic resonance (CMR) imaging is frequently recommended for patients at intermediate risk of cardiovascular disease to triage them for medication or invasive aggressive treatment. Mitral annulus (MA) motion and velocities represent the cardiac contraction and relaxation, and hold potential to improve the detection of subtle cardiac dysfunction. However, conventional interpretation of CMR images requires expert manipulation and is often operator-dependent. In this paper, we propose an end-to-end MA Point Tracking Network (MANet) to automatically detect and track MA motion during cardiac cycle. The MANet model consists of MA point detection module and motion tracking module. In MA point detection, we design the convolutional-based feature extraction and elastic regression to detect MA points frame by frame of each CMR video. Then, in MA tracking, we adopt the Deep SORT model to capture spatio-temporal continuity between frames and fine-tune the coordinate position of MA points. 171 CMR videos with 4275 frames are used in comparison experiments, and the results demonstrate that our MANet model achieves promising performance in reference to clinical ground truth (r=0.71, P<0.001). This work provides an important preamble for cardiac motion tracking and cardiac function evaluation.

Index Terms— Mitral annulus point tracking, cardiac motion analysis, medical image analysis.

1. INTRODUCTION

Cardiovascular disease (CVD) has become the main killer besides cancer [1, 2]. Cardiac magnetic resonance (CMR) imaging is the gold standard for noninvasive evaluation of cardiac structure and function, which has important clinical significance for CVD [2]. However, the routine interpretation of CMR requires expert manipulation, usually depends on the operator, and only the overall characteristics of stroke volume and ejection fraction are reported. The mitral annulus (MA, Fig. 1) displacement and velocities in multiple CMR views (i.e., 2-chamber, 3-chamber, and 4-chamber) during the cardiac cycle provide a means to differentiate the phasic diastolic and systolic functions [3, 4].



(a) 2-chamber (b) 3-chamber (c) 4-chamber

Fig. 1. Examples of MA points: anterior and inferior in 2-chamber, anteroseptal and inferolateral in 3-chamber, and inferoseptal and anterolateral in 4-chamber.

At present, various machine learning (ML) and deep learning (DL) algorithms have been applied to the field of cardiac function analysis, including cardiac tissue segmentation, cardiac strain tracking, and myocardial motion evaluation [5, 6, 7]. Most current approaches to track MA points are based on conventional semi-automated feature detection and tracking. In addition, since MA points exist in different chamber views, and there are different numbers of MA observed in each view, it is difficult to automatically and accurately detect and track MA points. Moreover, the limited amount of available training data further hinders the application of AI algorithms. To the best of our knowledge, there is no AI solution that enables MA point detection and tracking simultaneously in all CMR image views.

In this paper, we propose an MA point tracking network (MANet) to adaptively detect target MA points from CMR videos of different chambers, and robustly track the trajectory of these points during the cardiac cycle. The contributions of this paper are summarized as follows.

- This is the first work to provide an end-to-end AI solution to track the motion trajectories of MA points in multiple CMR image views. The MANet model consists of an MA point detection module and an MA mo-

*Contributed equally, †Corresponding author. Email: gmszl@nus.edu.sg.

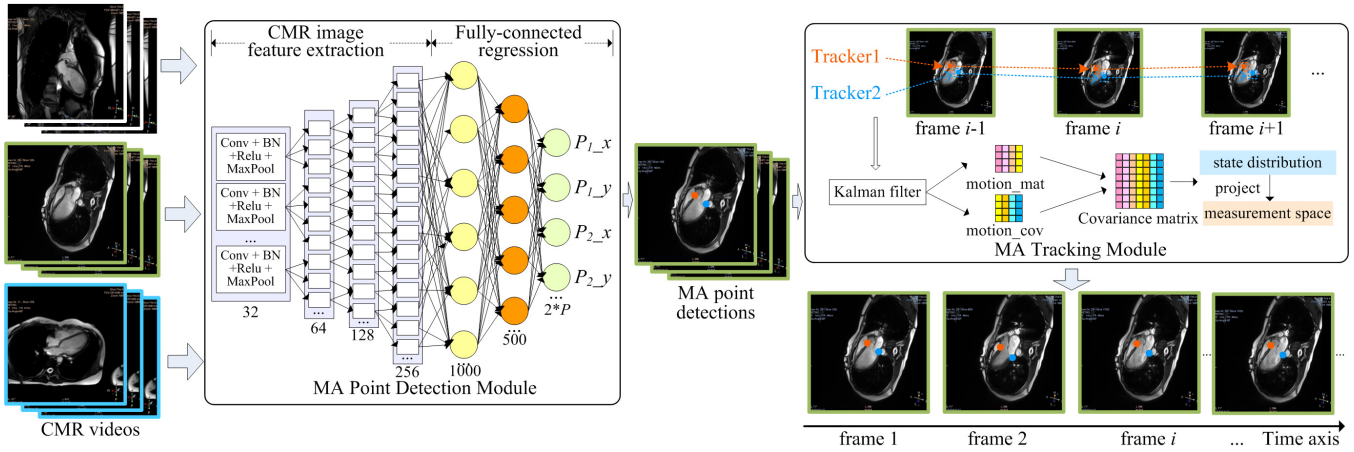


Fig. 2. Overview architecture of the proposed MANet model. CMR videos of multiple image views are input into the MA point detection module, and the preliminary positions of all candidate target MA points on each frame are detected by using convolutional-based feature extraction and elastic regression. The MA motion tracking module captures the spatio-temporal continuity among video frames and fine tunes the coordinate position of each MA point, so as to realize the accurate and robust trajectory tracking of each MA point in the cardiac cycle.

tion tracking module.

- We design the MA point detection module that uses convolutional-based feature extraction and elastic regression to effectively detect the preliminary positions of all candidate target MA points on each frame of the input CMR video.
- On this basis, we further design the MA motion tracking module to dynamically capture the spatio-temporal continuity between video frames, and fine-tune the coordinate position of each MA point to achieve accurate motion tracking.

2. RELATED WORK

In the field of cardiac function analysis, current DL and ML algorithms are mainly used in cardiac tissue segmentation, cardiac strain tracking, and myocardial motion evaluation [5, 6, 7, 8, 9, 10, 11]. For example, in [8], Mcleod *et al*. used sparse constraint tensor decomposition to observe the heart motion of healthy and diseased hearts. In [12], Parajuli *et al*. proposed a dynamic shape tracking method to obtain the Lagrangian trajectory derived from the boundary surface of the left ventricle. In [11], Lu *et al*. proposed a dynamic spatio-temporal map convolutional network for cardiac motion analysis to learn left ventricular (LV) motion patterns from CMR images. However, most of the existing studies focus on LV motion tracking and functional analysis, while ignoring other ventricles and atrium.

The recently published studies on cardiac valve annuli motion, tricuspid valve plane, and mitral valve plane [13, 14, 15] are most related to our work. In [13], the Dense block is stacked to construct a deep regression network for valve landmarks detection to estimate of the cardiac valve annuli motion. In [14] and [15], the Resnet block is stacked for the two points detection of mitral valve plane and tricuspid valve plane, respectively, through two-stage detection strategy. These methods are point detection based approaches, which perform the same function as the point detection module in our proposed MANet. While our proposed approach is employing multi-object tracking strategy based on the point detection results.

3. METHODOLOGY

3.1. Overview Architecture

We propose an end-to-end MA point tracking network (MANet) to automatically and accurately detect MA points from different CMR image views and track their motion trajectories during the cardiac cycle. The overview architecture of the proposed MANet model is shown in Fig. 2, which consists of an MA point detection module and an MA tracking module. The former uses convolutional-based feature extraction and elastic regression to effectively detect the preliminary positions of all candidate target MA points on each frame of CMR video. The latter module adopts the Deep SORT model [16, 17] to dynamically capture the spatio-temporal continuity between video frames, and fine-tune the coordinate position of each MA point.

3.2. MA Point Detection Module

In clinical CMR scanning, multiple CMR views (i.e., 2-chamber, 3-chamber, and 4-chamber) are collected for each

patient, and images in each image view are saved as a separate CMR video. To identify potential MA points from a given CMR video, we propose an MA point detection module that is trained on manually annotated images.

Firstly, each frame of a CMR video is input into the MA point detection module to identify MA points independently and used as a preliminary basis for subsequent point tracking. The module uses multiple feature extraction blocks with different scales to extract the global and local image features of each CMR frame. Each feature extraction block adopts a "Conv + BN + ReLU + Pool" structure, which is composed of a 2D convolutional layer, a batch normalization layer, a ReLU active function, and a 2D MaxPooling layer. The 2D MaxPooling layer uses the pool size of (2, 2) and the stride of (2, 2) to effectively achieve dimension reduction.

Based on the feature extraction blocks, a fully-connected regression block is embedded into the module to obtain the coordinate positions of the target MA points. Specifically, three dense connected layers are used, where the number of neurons is 1000, 500, and 4, respectively. To narrow the overfitting problem caused by limited training data, we add the dropout unit to each block. We use the Mean-Square Error (MSE, L_2 loss) as the loss function of the module.

To accelerate the convergence of the module and avoid falling into local optimality, we use the Adam optimizer [18] and set the learning rate to 0.001. In this way, the module can accurately locate the coordinates of the MA points between various heart tissues in different CMR views, thus providing the basis for subsequent MA object tracking.

3.3. MA Motion Tracking Module

Based on the detected position of the target MA points in each CMR video frame, we further design a MA motion tracking module to dynamically capture the Spatio-temporal continuity between video frames and fine-tune the coordinate position of each MA point.

The multi-MA object tracking module adopts the Deep SORT [16, 17] model as the network backbone, which is a simple and efficient multi-object tracking model. Since MA points are not occluded in CMR videos, we use a simple and efficient Kalman filtering method in the module for object tracking. It performs the Hungarian method and performs Kalman filter in frame-by-frame image space to calculate the bounding-box overlap among different frames. Our multi-MA object tracking scenario is defined on the three-dimensional state space (x, y, i, t) for each tracking target, where (x, y) is the center position of the target bounding box of the MA region, i is i-th tracker, and t is the length of the frames in the given CMR video.

For each track result OT_i of an MA point detection $D_i = (x_i, y_i)$, an appearance descriptor r_i is defined with $||r_i|| = 1$, and R^{OT} is the space of the appearance descriptors of all associated tracks. Then, we calculate the Mahalanobis distance between each track OT_i and target detection D_j as:

$$d_M(OT_i, D_j) = \min\{1 - r_i \times r_j^T | r_i \in R^{OT}\}.$$

By building the measurement-to-track associations in visual appearance space, the module can well track the position changes of each point between adjacent frames. In this way, the tracking module can automatically and accurately track the movement trajectory of each MA point in a cardiac cycle.

4. EXPERIMENTS

4.1. Dataset and Experimental Setting

(1) Dataset
The training set includes 45 patients with 135 CRM videos, each video consists of 25 frames, there are 3375 images in the training set, in which 20% of the images are used for evaluation during training process, since we perform extensive on-line augmentations (random rotation by 0 to 30 degree, and 10 percentage shifting in both x, y direction), which greatly increase the sample diversity and avoid potential overfitting. While the testing set includes 12 patients with 36 CRM videos, which extracts 900 images used for the prediction. The ground truth measurements were performed by one reader with ¿7 years of experience in CMR imaging, and results were reviewed by a second observer (a cardiologist with ¿20 years of experience in CMR). In the preprocessing stage, we perform normalization on the CMR images, and the size of all images is adjusted to 512 × 512 (pixel).
(2) Metrics
Point shift (PS) of the detected / tracked points from their ground truth is used to evaluate the accuracy of the detection or tracking results. Let $p_i = (x_i, y_i)$ be the ground truth of a point i and $p'_i = (x'_i, y'_i)$ be the detected or tracked positions, the PS between p_i and p'_i is defined as:

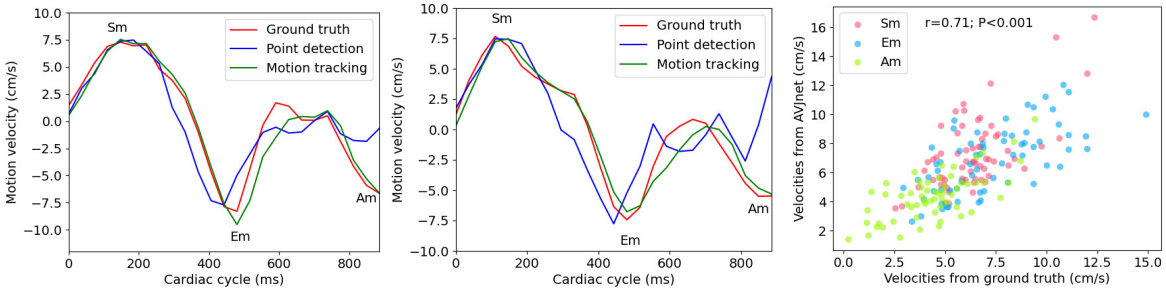
$$PS(p_i, p_{i'}) = d_{i,i'} \times Dp_i,$$

where $d_{i,i'}$ is the Euclidean distance between p_i and p'_i and Dp_i is the pixel spacing of the current image. A small shift value indicates better matching between the automatic detected / tracked MA points and the ground truth.

4.2. Performance Evaluation

We conduct comparison experiments to evaluate the tracking performance of our MANet model by comparing the MA point detection module (the first module of MANet) and the whole MANet model. The PS value of of each MA point in 2-chamber, 3-chamber, and 4-chamber is recorded, and the experimental results are shown in Table 1.

As shown in Table 1, the point detection module can extract the CMR image features of MA points. In addition, we perform motion tracking based on the point detection results to further improve the accuracy of the model. For the point



(a) MA anteroseptal (b) MA inferolateal (c) Correlation plot

Fig. 3. (a) and (b) MA motion velocities of a 3-chamber case based on MA point detection and motion tracking; (c) Correlation plot between velocities from MANet and ground truth. Sm, peak systolic velocity; Em, peak early diastolic velocity; Am, peak late diastolic velocity during atrial contraction.

Table 1. Performance evaluation of MA point detection and motion tracking (point shift: mean±standard deviation, mm).

Modules	**2-chamber**		**3-chamber**		**4-chamber**	
	Anterior	Inferior	Anteroseptal	Inferolateral	Inferoseptal	Anterolateral
Point detection	1.86 ±1.84	1.75 ±1.60	1.49 ±1.78	1.88 ±1.57	2.68 ±1.45	2.06 ±1.72
Motion tracking	1.13 ±0.89	1.20 ±1.11	1.28 ±1.01	1.21 ±0.78	1.25 ±1.16	1.41 ±0.86
p-value	0.8021	0.8326	0.8219	0.8104	0.8436	0.8774

shift results, the range of the statistical significance (based on p-values) of the differences is [0.8021, 0.8774], that indicates no significant difference between ground truth and prediction. Examples of the visualization of the MA motion tracking results at end-diastolic (ED) and end-systolic (ES) frames in different CMR views are shown in Fig. 4. We can see that the MANet model with the motion tracking module can effectively identify the motion pattern of each MA point in the time dimension, so it can obtain accurate position in different frames of different CMR views.

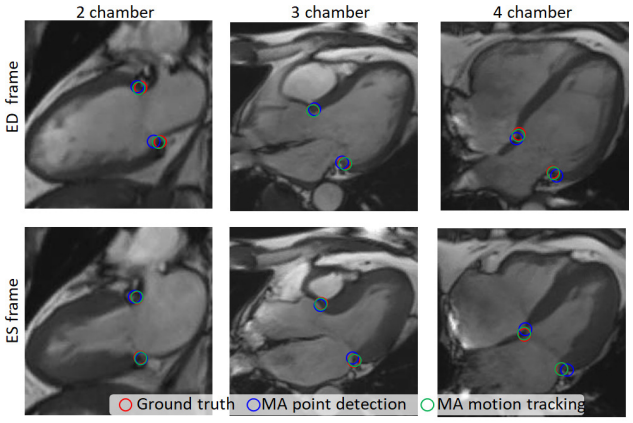


Fig. 4. Visualization of MA detection and tracking results at ED and ES frames in multiple CMR views.

4.3. MA Motion and Velocity Analysis

MA motion velocities are calculated after obtaining the motion trajectory of each MA point in a cardiac cycle. Taking two MA points of a 3-chamber case as an example, the results of their real motion velocity and the values calculated by our MANet model are shown in Fig. 3.

The red lines in Fig. 3(a) and (b) represent the MA motion velocities calculated based on the positions marked by the cardiologist. The blue lines come from the results of the MA point detection module, and the green lines come from the results of the whole MANet model. We can see that compared with the results of the MA point detection module, the velocities of the two MA points obtained by the whole MANet model are closer to the actual values. Fig. 3(c) shows the correlation plot between MA velocities derived from MANet and ground truth. The results demonstrate that our MANet model achieves promising performance in reference to clinical ground truth (r=0.71, P<0.001).

5. CONCLUSION

This paper presented an end-to-end MANet model, consisting of point detection and motion tracking modules, for mitral annulus point detection and tracking. We evaluated the performance of MANet model based on point shift and correlation coefficients against the clinical ground truth. The experimental results show that MANet obtains a promising performance and is worthy of further clinical applications. Though the application is specific, the proposed methodologies are generalizable to the detection and tracking of various anatomical structures of the heart or other human organs. Furthermore, our proposed video-based AI techniques can be easily extended to other industrial fields beyond medical imaging.

6. REFERENCES

[1] David Ouyang, Bryan He, Amirata Ghorbani, et al., "Video-based ai for beat-to-beat assessment of cardiac function," *Nature*, vol. 580, pp. 252–256, 2020.

[2] Ghalib A Bello, Timothy JW Dawes, Jinming Duan, et al., "Deep-learning cardiac motion analysis for human survival prediction," *Nat. Mach. Intell.*, vol. 1, no. 2, pp. 95–104, 2019.

[3] Shuang Leng, Xiaodan Zhao, Feiqiong Huang, et al., "Automated quantitative assessment of cardiovascular magnetic resonance-derived atrioventricular junction velocities," *Am. J. Physiol.-Heart Circul. Physiol.*, vol. 309, pp. H1923–35, 2015.

[4] Rongzhen Ouyang, Shuang Leng, Aimin Sun, et al., "Detection of persistent systolic and diastolic abnormalities in asymptomatic pediatric repaired tetralogy of fallot patients with preserved ejection fraction: a cmr feature tracking study," *Eur. Radiol.*, vol. 31, pp. 6156–6168, 2021.

[5] Kristin McLeod, Kristin Tondel, Lilian Calvet, et al., "Cardiac motion evolution model for analysis of functional changes using tensor decomposition and cross-sectional data," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 12, pp. 2769–2780, 2018.

[6] Sulaiman Vesal, Mingxuan Gu, Andreas Maier, et al., "Spatio-temporal multi-task learning for cardiac mri left ventricle quantification," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 7, pp. 2698–2709, 2021.

[7] Shuang Leng, Xiaodan Zhao, Angela S Koh, et al., "Age-related changes in four-dimensional cmr-derived atrioventricular junction velocities and displacements: Implications for the identification of altered annular dynamics for ventricular function assessment," *Int. J. Cardiol. Heart. Vasc.*, vol. 22, pp. 6–12, 2018.

[8] Kristin Mcleod, Maxime Sermesant, Philipp Beerbaum, et al., "Descriptive and intuitive population-based cardiac motion analysis via sparsity constrained tensor decomposition," in *MICCAI'15*, 2015, pp. 419–426.

[9] Xulei Yang, Gabriel Tjio, Feng Yang, et al., "A multi-channel deep learning approach for segmentation of the left ventricular endocardium from cardiac images," in *EMBC'19*, 2019, pp. 122–125.

[10] Xulei Yang, Yi Su, Si Yong Yeo, Liang Zhong, and Ru San Tan, "Segmentation of cardiac magnetic resonance (cmr) images using a memory persistence approach," *US Patent US10235750B2*, 2019.

[11] Ping Lu, Wenjia Bai, Daniel Rueckert, et al., "Dynamic spatio-temporal graph convolutional networks for cardiac motion analysis," in *ISBI'21*, 2021, pp. 122–125.

[12] Nripesh Parajuli, Allen Lu, John C Stendahl, et al., "Integrated dynamic shape tracking and rf speckle tracking for cardiac motion analysis," in *MICCAI'16*, 2016, pp. 431–438.

[13] Eric Kerfoot, Carlos Escudero King, Tefvik Ismail, et al., "Estimation of cardiac valve annuli motion with deep learning," in *STACOM'20*, 2020, pp. 146–155.

[14] Ricardo A. Gonzales, Felicia Seemann, Jerome Lamy, et al., "Mvnet: automated time-resolved tracking of the mitral valve plane in cmr long-axis cine images with residual neural networks: a multi-center, multi-vendor study," *J. Cardiov. Magn. Reson.*, vol. 23, no. 1, pp. 137, 2021.

[15] Ricardo A. Gonzales, Jerome Lamy, Felicia Seemann, et al., "Tvnet: Automated time-resolved tracking of the tricuspid valve plane in mri long-axis cine images with a dual-stage deep learning pipeline," in *MICCAI'21*, 2021, pp. 567–576.

[16] Alex Bewley, Zongyuan Ge, Lionel Ott, et al., "Simple online and realtime tracking," in *ICIP'16*, 2016, pp. 3464–3468.

[17] Nicolai Wojke, Alex Bewley, and Dietrich Paulus, "Simple online and realtime tracking with a deep association metric," in *ICIP'17*, 2017, pp. 3645–3649.

[18] Diederik P. Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," in *ICLR'15*, 2015, pp. 18–25.