

Multiscale Graph Convolutional Networks for Cardiac Motion Analysis

Ping Lu¹[0000–0002–0199–3783], Wenjia Bai^{2,3}[0000–0003–2943–7698], Daniel Rueckert²[0000–0002–5683–5889], and J.Alison Noble¹[0000–0002–3060–3772]

¹ Institute of Biomedical Engineering, University of Oxford, UK
ping.lu@eng.ox.ac.uk

² Department of Computing, Imperial College London, London, UK

³ Department of Brain Sciences, Imperial College London, London, UK

Abstract. We propose a multiscale spatio-temporal graph convolutional network (MST-GCN) approach to learn the left ventricular (LV) motion patterns from cardiac MR image sequences. The MST-GCN follows an encoder-decoder framework. The encoder uses a sequence of multiscale graph computation units (MGCUs). The myocardial geometry is represented as a graph. The network models the internal relations of the graph nodes via feature extraction at different scales and fuses the feature across scales to form a global representation of the input cardiac motion. Based on this, the decoder employs a graph-based gated recurrent unit (G-GRU) to predict future cardiac motion. We show that the MST-GCN can automatically quantify the spatio-temporal patterns in cardiac MR that characterise cardiac motion. Experiments are performed on mid-ventricular short-axis view cardiac MR image sequence from the UK Biobank dataset. We compare the performance of cardiac motion prediction of the proposed method with ten different architectures and parameter settings. Experiments show that the proposed method inputting node positions and node velocities with multiscale graphs achieves the best performance with a mean squared error of 0.25 pixel between the ground truth node locations and our prediction. We also show that the proposed method can estimate a number of motion-related metrics, including endocardial radii, thickness and strain which are useful for regional LV function assessment.

Keywords: Spatio-Temporal Graph Convolutional Networks · Cardiac MR · Motion analysis

1 Introduction

Cardiac motion analysis plays an important role in the diagnosis of heart conditions [7] [11]. In diagnosis, displacement and strain are important biomarkers, which are sensitive to subtle changes in myocardial function and often indicate the early onset of cardiac disease [7]. Cardiac motion can be evaluated by tracking the contours of the endocardium and the epicardium from magnetic

resonance imaging (MRI). The aim of cardiac motion analysis is to perform an accurate estimation of the motion trajectories for the myocardial contours.

In recent years, geometric deep learning-based methods have achieved promising results in medical image analysis, such as artery and vein classification [8] and landmark detection [3]. Inspired by these applications of spatio-temporal graph neural network (ST-GCN) [4] [10], we propose to employ graph convolutional neural networks (GCN) to model cardiac motion in the geometry space of the GCN which can take advantage of a sparse representation of the cardiac myocardium using contours instead of images.

Our previous work [6] discussed different strategies for constructing cardiac structure graphs on a single scale. In this paper, we continue to explore relations between different parts of the endocardium and the epicardium, which convey essential information for cardiac motion. We represent myocardial geometry at different scales. At each scale, sample nodes on the left ventricular (LV) myocardial contour are connected as a graph. Then nodes across two scales are connected as a cross-scale graph, which is also named a bipartite graph. We propose a geometric deep learning-based architecture, named multiscale spatio-temporal graph convolutional networks (MST-GCN), with a self-supervised training strategy. This method predicts the future 2D LV cardiac motion given the previously observed motion trajectories. The cardiac motion is represented on a graph constructed from sample nodes on LV myocardial contours. The MST-GCN and a graph-based gated recurrent unit (G-GRU) are connected using an encoder-decoder framework. We investigate essential elements of the proposed architecture for modelling the spatio-temporal patterns of cardiac motion. We also analyse the regional function of the LV based on predicted cardiac motion trajectories.

Contributions. (1) We propose a multiscale geometric deep learning-based architecture for LV cardiac motion estimation. To our knowledge, this is the first method to explore the internal relations of the multiscale endo-epicardial geometry for feature extraction. (2) We evaluate the impact of different elements of the proposed architecture on the accuracy of the 2D LV cardiac motion prediction. (3) We demonstrate that multiscale spatio-temporal patterns achieve good performance for cardiac motion estimation and regional analysis of LV function.

2 Method

We investigate various orders of motion difference fed into the proposed architecture. We define the node locations as the 0-order difference and the node velocities as the 1-order difference [5]. The node velocities are the differences of node locations between the current and the immediately previous cardiac MR frame, without the division of the time between each frame (since the time between each frame is constant).

2.1 Multiscale Cardiac Graph Construction

Fig.1 illustrates the construction of a three scale cardiac graph. These endo-epicardial nodes are chosen by the left and right ventricle geometry. The detail of how to sample nodes is described in our previous work [6]. These selected node locations are the ground truth in our work. In order to explore the further detail of the beating heart described by endo-epicardial nodes, we design a multiscale cardiac graph based on multiscale endo-epicardial node components. According to the mid-slice 6-segments model of the 17-Segment AHA model, we use three cardiac scales: endo-epicardial nodes scale $S1$, two high-level part scale, $S2$, $S3$.

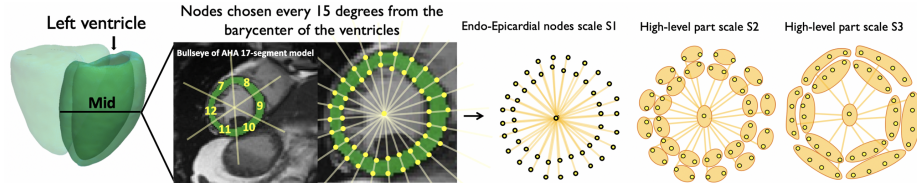


Fig.1: Overview of the proposed framework for the three scale cardiac structure graph construction in the mid-ventricular of short-axis view cardiac MR image sequences. At scale $S1$, the barycenter of the left ventricle (LV) and the 48 node locations from both the endocardium and epicardium are considered as cardiac structure graphs. At scale $S2$ and $S3$, 25 and 13 parts are considered respectively.

2.2 Multiscale Graph Computation Unit (MGCU)

The aim of the multiscale graph computation unit (MGCU) is to extract and fuse features at different scales based on a multiscale graph. Each MGCU includes spatio-temporal graph convolution blocks and cross-scale fusion blocks. The spatio-temporal graph convolution block extracts features at each scale with a single-scale graph. The cross-scale fusion block interchanges features from one scale to another.

Spatio-temporal graph convolution block (ST-GCB). The ST-GCB includes a spatial graph convolution and then a temporal convolution, which extracts spatial and temporal features respectively. Let $A \in \{0, 1\}^{N_S \times N_S}$ represent the adjacency matrix of the graph at the scale S , where N_S is the number of cardiac components. If the i -th and the j -th nodes are connected, $A_{i,j} = 1$. Otherwise, $A_{i,j} = 0$. Let the input feature at scale S be F_{in} , the spatial graph convolution is described as $F_{out} = \sum M \circ \tilde{A} F_{in} W$. Here M is the edge weight matrix using the trainable weights, W is the feature importance matrix, $\tilde{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ is the normalized adjacent matrix, \circ represents the Hadamard product. A $1D$ temporal convolution is applied to extract features along time.

Cross-scale fusion block (CS-FB). In order to obtain rich multiscale information, the SS-FB makes feature diffusion across different scales. It employs a bipartite graph [4] to deliver features from one scale graph to another. For instance, the features of a node at the segment 8 area in the high-level part scale $S2$ can offer the feature learning of a node at the segment 8 area in the endo-epicardial nodes scale $S1$. Here we describe CS-FB from $S1$ to $S2$ as an example.

Let $A_{S1S2} \in \{0, 1\}^{N_{S1} \times N_{S2}}$ denote the cross-scale relations. Let $(A_{S1S2})_{k,i}$ denote the edge weight between the i th node and k part. Let the feature of the i th node be $F_{S1,i}$ and the feature of the k th part be $F_{S2,k}$. $h_{S1,i} = \theta_{S1}(F_{S1,i})$, $h_{S2,k} = \theta_{S2}(F_{S2,k})$, $(A_{S1S2})_{k,i} = \text{softmax}(h_{S2,k}^T, h_{S1,i})$, where θ_{S1} and θ_{S2} are implemented by MLP [4] and convolution [4]. The softmax function softmax works on the result of the inner product matrix. After obtaining the adjacent matrix via inner product and softmax, we model the effects from the scale $S1$ to component in $S2$. The idea is to augment cardiac component features from the global related information. We obtain the edge weight from the inner product of two augmented features.

After that, we fuse the endo-epicardial node features in the scale $S1$ to the high-level part scale $S2$ with A_{S1S2} . The features F_{S2} at the scale $S2$ can be described as $A_{S1S2}F_{S1}W + X_{S2}$, where W is the trainable weights. The fused F_{S2} is used to the ST-GCB of the next MGCU in $S2$.

2.3 Multiscale Spatio-Temporal Graph Convolutional Neural Network

Graph-based gated recurrent unit. The graph-based GRU (G-GRU) learns and updates hidden cardiac motion states with graph guidance. The graph is trained to regularise the states to generate a future cardiac structure. There are two inputs for the G-GRU: the initial state H_0 and the 2D cardiac structure-based feature $F_t \in R^{m \times d}$ at time t . We define $A_h \in R^{m \times m}$ as the adjacent matrix of the graph from time $t-1, \dots, t-n$. The G-GRU(F_t, H_t) is denoted as

$$\begin{aligned} z_t &= \sigma(z_{in}(F_t) + z_{hid}(A_h H_t W_h)), \\ r_t &= \sigma(r_{in}(F_t) + r_{hid}(A_h H_t W_h)), \\ g_t &= \tanh(g_{in}(F_t) + z_t \otimes g_{hid}(A_h H_t W_h)), \\ H_{t+1} &= r_t \otimes H_t + (1 - r_t) \otimes g_t. \end{aligned}$$

where the functions z_{in} , z_{hid} , r_{in} , r_{hid} , g_{in} , g_{hid} are linear transformations. The W_h is the trainable weights. A graph convolution is used on the hidden states H_t and generates the state for the next frame.

Encoder-decoder architecture. To initialize cardiac scales, we choose 2D average pooling to compute the average over endo-epicardial node clusters in $S1$ to corresponding components in coarser scales. For instance, we average two epicardial segment-8 nodes in $S1$ to the segment-8 part in $S2$. Then we use multiscale graph computation to extract spatio-temporal features. At the end,

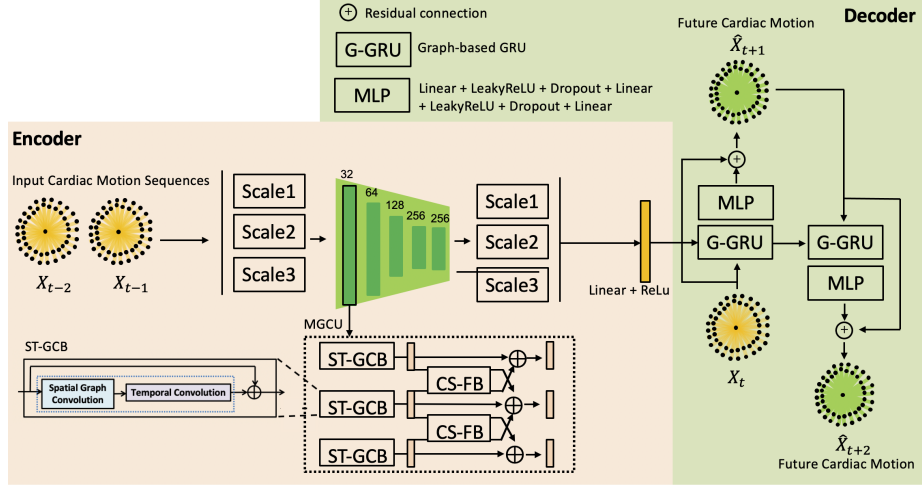


Fig. 2: *Network overview. The cardiac motion sequence is given as input to the MST-GCN encoder-decoder framework. In the encoder, each MGCU layer extracts spatio-temporal feature by multiscale graphs. The sum of the output of the decoder and the previous motion status represents the output future cardiac motion trajectory (predicted node locations), which is used in regional analysis of left ventricular function. \oplus denotes the element-wise addition.*

we combine three scales features. Let the output features of the scale $S1$, $S2$, $S3$ be H_{S1} , H_{S2} , H_{S3} respectively. The final feature H is $H = H_{S1} + H_{S2} + H_{S3}$.

In the decoder, the aim is to predict future cardiac structure sequences, which are represented by node locations on the myocardial geometry in time sequences. We choose the proposed graph-based GRU (G-GRU) and a multilayer perceptron (MLP) to model cardiac displacement between two consecutive frames. Then we add a residual connection between the input and the output of each G-GRU cell to get the estimated cardiac motion (see Fig.2). At time t , the decoder is described as $H_{t+1} = G-GRU(\hat{X}_t, H_t)$, $\hat{X}_{t+1} = \hat{X}_t + \tau(H_{t+1})$. Here the function τ is implemented by MLP (see Fig.2). The initial state $H_0 = H$, which is the final output feature of encoder.

3 Experiments

3.1 Data Acquisition

In this study, we use 1611 short-axis view cardiac MR image sequences from the UK BioBank⁴, acquired using a 1.5 Tesla scanner (Siemens Healthcare). A stack of short-axis images, around 12 slices, cover the entire left and right ventricles. In-plane resolution is $1.8 \times 1.8mm^2$, while the slice gap is $2.0mm$ and the slice

⁴ UK BioBank. <https://www.ukbiobank.ac.uk/>

thickness is $8.0mm$. Each sequence contains 50 consecutive time frames per cardiac cycle. We randomly split image sequences into training/validation/testing with 1071 / 270 / 270 subjects. We perform motion analysis on the 3 mid-ventricular slices.

3.2 Implementation Details

Pre-processing. The segmentation of the LV endo-epicardial borders and the RV was generated from using a publicly available FCN model [1] and used for node extraction. **Training.** The model is trained over 100 epochs via Adam with a learning rate 0.001 and a batch size of 1. In each training sample, we set the input difference operators length to 3 frames, and we predict future cardiac structure in 2 frames. The mean squared error (MSE) using the node locations is chosen as the evaluation metric. The proposed network was implemented using Python 3.7 with Pytorch. Experiments are run with computational hardware GeForce GTX 1080 Ti GPU 10 GB.

3.3 Results

Quantitative results. We compare the proposed method with ten different architectures and parameter settings with the baseline ST-GCN approach [6]. We also investigate key elements of the proposed architecture: effects of multiple scales, high-order motion difference, hyper-parameter in the fusion block balances the effect between the node-scale and other abstract scales. Based on the experiments on the validation data, we find that the best model is $P\&V, S_{123}$ (inputting the 0-order and 1-order difference features using 3 scale graphs without the fusion block), which achieves a mean squared error of 0.251 pixels on MRI sequences. From Figure 3 (a), the multiscale architecture works better than the single architecture. The feature velocity improves performance. The fusion block between scale 1 and 2 works slightly better with the input 0th order difference feature node locations. Based on the performance with hyper-parameter 0.3 and 0.5 in the fusion block, the accuracy is not improved significantly. Figure 3 (b) compares the MSE of the best two performing architectures and parameter settings using the proposed method in each MRI frame for an example subject. The MSE increases after the 19th frame. Later, we use the same example subject for left ventricular function evaluation.

Left ventricular function evaluation. Based on the 17-Segment AHA model, the predicted nodes are classified into 6 segments [2]. Fig. 4 shows an example of a time series of the endocardial radius, thickness, radial strain (Err) [7] in the six segments of myocardium from the ground truth of node locations, prediction and two other methods from Lu et al. [7] and Qin et al. [9] in a healthy volunteer. Compared to the ground truth, the prediction of the endocardial radius, the thickness and the Err has a similar plot shape.

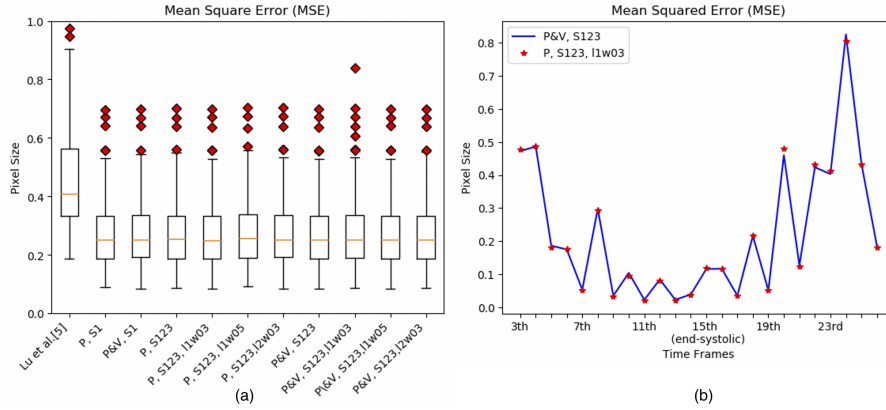


Fig. 3: (a) Using the proposed method - with ten different architectures and parameter settings - and the baseline, a comparison was made between the MSE on the predicted node locations on the endo-epicardial borders in the MRI sequences with the ground truth. The key elements of the architecture are inputting the 0-order difference features (P), 1-order difference features (V) in single scale 1 (S1), scale 1 to 3 (S123), fusion block in 1st MGCU layer (l1), fusion block in 1st and 2nd MGCU layers (l2), weight parameter 0.3 (w3) and 0.5 (w5). (b) Comparison of the mean squared error (MSE) for the best two performing architectures and parameter settings using the proposed method in each MRI frame. *P&V, S123* means inputting the 0-order and 1-order difference features using 3 scale graphs without the fusion block and *P, S123, l1w03* means inputting the 0-order features using 3 scale graphs with the fusion block (hyper-parameter 0.3) in 1st MGCU layer.

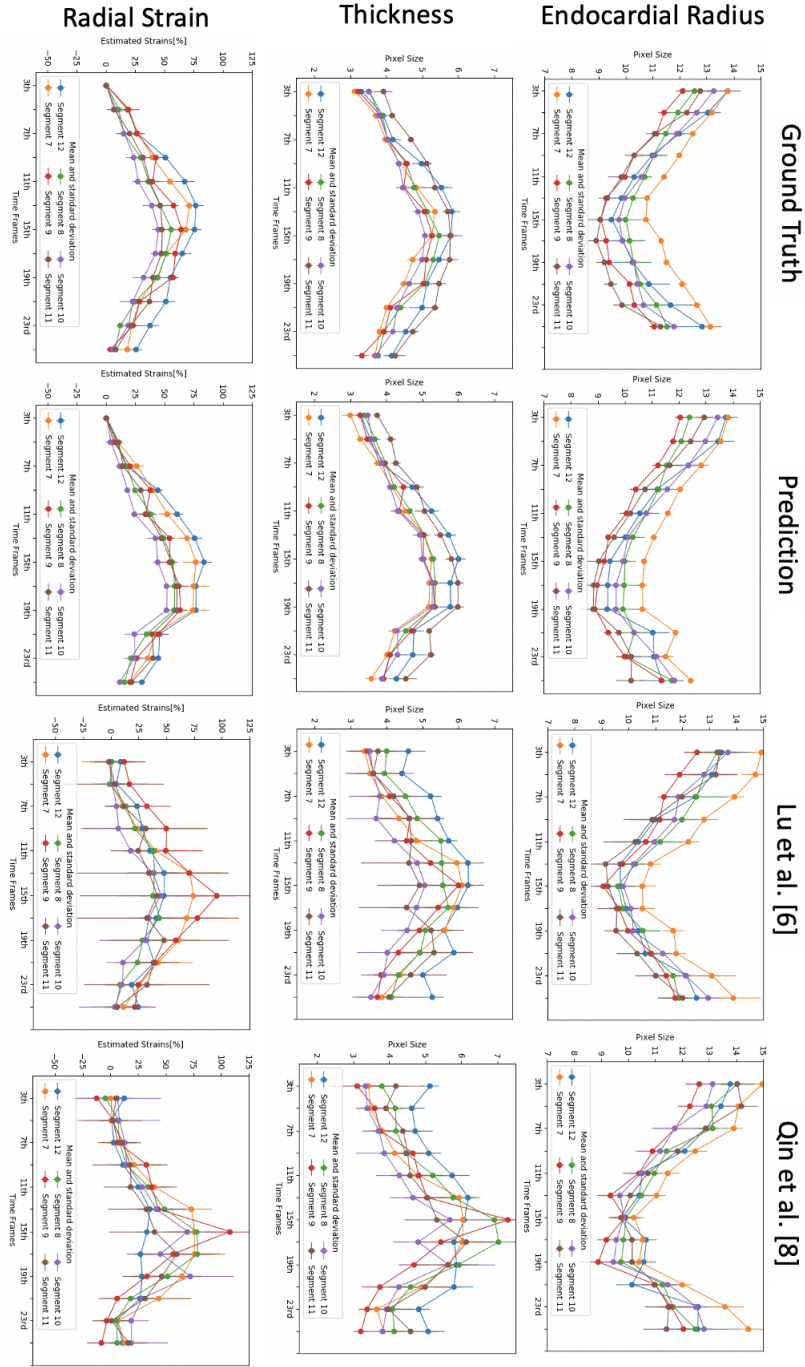


Fig. 4: Example results of the estimated endocardial radius (mean and standard deviation shown), thickness (mean and standard deviation shown) and radial strain (mean and standard deviation shown) for cardiac segments (7-12) plotted on frames 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 over a cardiac cycle. Results (left to right) for ground truth, prediction, methods from Lu et al. [7] and Qin et al. [9] in a healthy volunteer.

4 Discussion and Conclusion

The focus of this paper is to investigate the effectiveness of graph convolutional networks for cardiac motion prediction. In clinical applications, motion trajectories may directly come from a motion tracking algorithm. Future work would extend the graph convolutional networks to the analysis of cardiac motion trajectories for disease diagnosis and motion-based biomarker discovery.

In this work, we propose a multiscale spatio-temporal graph convolutional network (MST-GCN) to characterise cardiac motion. We investigated the factors which can improve the accuracy of cardiac motion estimation. We found that the proposed method achieved a mean squared error of 0.25 pixel on MRI sequences. The accuracy is not significantly improved by a cross-scale fusion block. The proposed methods can characterise cardiac motion features for regional LV analysis.

Acknowledgements. This work is supported by SmartHeart. EPSRC grant EP/P001009/1. This study used data from the UK Biobank Resource under Application Number 40119. The authors declare no conflicts of interest related to this research study.

References

1. Bai, W., et al: Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *JCMR* **20**(1), 65 (2018)
2. Cerqueira, M.D., et al: Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: a statement for healthcare professionals from the cardiac imaging committee of the council on clinical cardiology of the american heart association. *Circulation* **105**(4), 539–542 (2002)
3. Lang, Y., et al: Automatic localization of landmarks in craniomaxillofacial cbct images using a local attention-based graph convolution network. In: *MICCAI*. pp. 817–826. Springer (2020)
4. Li, M., et al: Dynamic multiscale graph neural networks for 3d skeleton based human motion prediction. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 214–223 (2020)
5. Lu, P., et al: Dynamic spatio-temporal graph convolutional networks for cardiac motion analysis. In: *ISBI 2021*
6. Lu, P., et al: Modelling cardiac motion via spatio-temporal graph convolutional networks to boost the diagnosis of heart conditions. In: *STACOM Workshop, MICCAI 2020*
7. Lu, P., et al: Going deeper into cardiac motion analysis to model fine spatio-temporal features. In: *MIUA*. pp. 294–306. Springer (2020)
8. Noh, K.J., et al: Combining fundus images and fluorescein angiography for artery/vein classification using the hierarchical vessel graph network. In: *MICCAI*. pp. 595–605. Springer (2020)
9. Qin, C., et al: Joint learning of motion estimation and segmentation for cardiac mr image sequences. In: *MICCAI*. pp. 472–480. Springer (2018)
10. Yan, S., et al: Spatial temporal graph convolutional networks for skeleton-based action recognition. In: *Thirty-second AAAI conference on artificial intelligence* (2018)

11. Zheng, Q., et al: Explainable cardiac pathology classification on cine MRI with motion characterization by semi-supervised learning of apparent flow. Medical image analysis (2019)