

# Modelling Cardiac Motion via Spatio-temporal Graph Convolutional Networks to Boost the Diagnosis of Heart Conditions

Ping Lu<sup>1</sup>, Wenjia Bai<sup>2,3</sup>, Daniel Rueckert<sup>4</sup>, and J.Alison Noble<sup>1</sup>

<sup>1</sup> Institute of Biomedical Engineering, University of Oxford, UK  
`ping.lu@eng.ox.ac.uk`

<sup>2</sup> Data Science Institute, Imperial College London, London, UK

<sup>3</sup> Department of Brain Sciences, Imperial College London, London, UK

<sup>4</sup> Department of Computing, Imperial College London, London, UK

**Abstract.** We present a novel spatio-temporal graph convolutional networks (ST-GCN) approach to learn spatio-temporal patterns of left ventricular (LV) motion in cardiac MR cine images for improving the characterization of heart conditions. Specifically, a novel GCN architecture is used, where the sample nodes of endocardial and epicardial contours are connected as a graph to represent the myocardial geometry. We show that the ST-GCN can automatically quantify the spatio-temporal patterns in cine MR that characterise cardiac motion. Experiments are performed on healthy volunteers from the UK Biobank dataset. We compare different strategies for constructing cardiac structure graphs. Experiments show that the proposed methods perform well in estimating endocardial radii and characterising cardiac motion features for regional LV analysis.

**Keywords:** Spatio-Temporal Graph Convolutional Networks · Cardiac MR · Endocardium · Epicardium · Motion-characteristic features

## 1 Introduction

Cardiac motion estimation plays an important role in the diagnosis of heart condition [6][7][8][10]. Motion-characteristic features, such as time series of the endocardial radius, thickness, and radial strain (Err), can be evaluated by the sampling nodes of the endocardium and the epicardium from magnetic resonance imaging (MRI). These features are related to cardiac disease and they are easy to explain as characteristics of pathological cardiac motion [6]. The motion trajectory estimation of these sampling nodes can be regarded as motion analysis in computer vision.

In recent years, geometric deep learning-based methods have achieved promising results in motion recognition. Yan et al. [9] proposed a spatial-temporal graph neural network (ST-GCN) for human action recognition via motion classification. Li et al. [5] suggested actional-structural GCN for human action prediction using skeleton and joint trajectories of human bodies. Huang et al. [3] predicted human trajectories with a spatio-temporal graph attention network (STGAT).

These applications of GCNs [4] extend convolutional neural networks (CNNs) to graphs of arbitrary structures and model dynamic graphs over time sequences, which can be applied to cardiac motion analysis.

For the diagnosis of heart conditions, especially for early stage characterisation, a major challenge is estimating the effect of cardiac functional changes via automated cardiac motion analysis. Before more significant cardiac structural changes occur, functional changes such as motion and strain may occur, which are subtle but often indicates the early onset of cardiac disease symptoms. But the early onset of symptoms already causes an increased strain on the heart [6].

In this paper, we propose a geometric deep learning-based architecture with a self-supervised strategy. This method characterises the spatio-temporal patterns of left ventricular (LV) cardiac motion in cine MR image sequences to boost the diagnosis of heart conditions. We investigate two strategies of graph construction and cardiac motion estimation. The proposed method is to predict node locations on the endocardium and the epicardium separately. This requires cardiac structure graphs constructed from the endocardium and the epicardium respectively. The other method is node locations prediction on the endocardium and the epicardium simultaneously, which needs a graph constructed from both endocardium and epicardium. Moreover, we extract motion-characteristic features and time series of the endocardial radius, thickness, and radial strain (Err), based on the output cardiac motion trajectory, and compare these features between the prediction and the ground truth in a group of healthy volunteers.

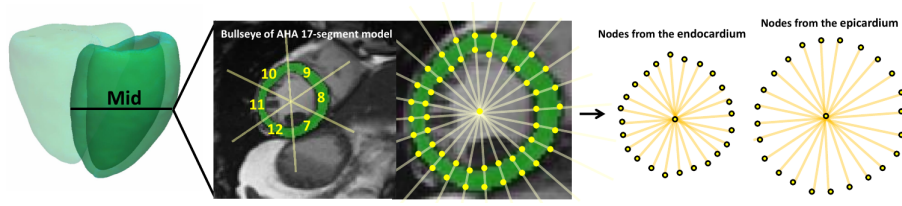
**Contributions.** The contributions of this work are as follows. (1) To our knowledge, this is the first time that cardiac motion has been modelled with a geometric deep learning-based architecture. (2) The predicted cardiac structure trajectory of this method can be used to evaluate motion characteristics, namely a time series of the endocardial radius, thickness, and Err. (3) We demonstrate that spatio-temporal patterns achieve good performance for cardiac motion estimation and regional analysis of LV function.

## 2 Method

### 2.1 Cardiac Structure Graph Construction

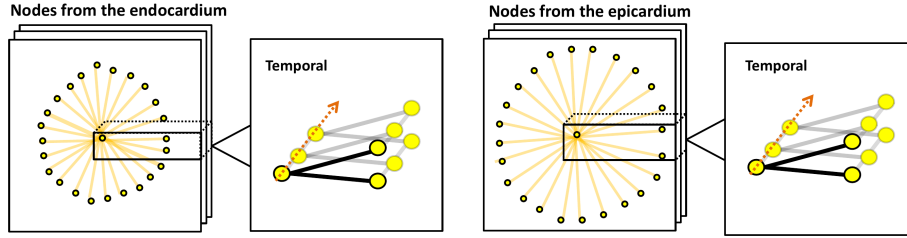
The high-level steps in cardiac structure graph construction are summarised in Fig.1. The cardiac structure sequence is represented by 2D coordinates of nodes on both the endocardium and epicardium in each cardiac MR frame. These nodes are chosen by the left and right ventricle geometry, according to the mid-slice 6-segments model of the 17-Segment AHA model [2]. Firstly, we select the barycenter of the left ventricle (LV) and the right ventricle (RV) in the middle slice of the short axis view image. Secondly, we define the straight line between these two nodes as the initial line [6]. Thirdly, we rotate this initial line around the barycenter node of the LV by every 15 degrees. The intersection nodes of the rotated line and the endocardial and epicardial borders are chosen. These

intersection node locations, 2D pixel coordinates  $(x, y)$ , are obtained in every single frame. Morphological transformations and finding the barycenter location of the LV and the RV are implemented using OpenCV. Later, these nodes will be classified into the mid-slice 6-segments model for the regional analysis of left ventricular function. Moreover, these selected node locations are the ground truth in our work.



**Fig. 1.** Overview of the proposed framework for the cardiac structure graph construction in the mid-ventricular of short-axis view cardiac MR image sequences. The barycenter of the left ventricle (LV) and the 24 node locations from both the endocardium and epicardium are considered as cardiac structure graphs.

We construct two undirected spatio-temporal graphs  $G = (V, E)$  on the cardiac structure with  $N$  nodes and  $T$  frames. As shown in Fig.2, one graph features both intra-endocardium and inter-frame connection, the other graph includes both intra-epicardium and inter-frame connection. These two graphs have the same structure. In the graph, the node set  $V = \{v_{it} | i = 1, \dots, N, t = 1, \dots, T\}$  includes all the 25 nodes in the cardiac structure sequence. The feature vector on a node  $F(v_{it})$  is the 2D pixel coordinate  $(x, y)$  of the  $i$ -th node on frame  $t$ .



**Fig. 2.** Overview of the cardiac structure as a spatio-temporal graph. One graph (left side) represents nodes from the endocardium, the other graph (right side) represents nodes from epicardium. In both graphs, the inter-frame edges connect the same nodes between consecutive sampling frames. Node locations, 2D pixel coordinates  $(x, y)$ , are defined as inputs to the spatio-temporal graph convolutional networks.

The spatio-temporal graphs on the cardiac structure sequences are constructed from the endocardium and the epicardium respectively. In the graph constructed from the endocardium, nodes from the endocardium in the same frame are connected to the barycenter of the LV with edges, which is presented in Fig. 1. After that, each node will be connected to the same node in the consecutive sampling frame, as described in Fig. 2. The graph constructed from the epicardium follows the same steps as the graph constructed from the endocardium. Based on the validation performance, every third frame is selected.

The edge set  $E$  consists of two subsets. The first subset describes the intra-structural connection at each cardiac MR frame, denoted as  $E_s = \{(v_{it}, v_{jt}) | (i, j) \in H\}$ , where  $H$  is the set of nodes from both the endocardium and epicardium separately. The second subset includes the inter-frame edges,  $E_f = \{(v_{it}, v_{i(t+1)})\}$ , which connect the same nodes in consecutive sampling frames. Hence, for every node  $n$ , the edge in  $E_f$  records its trajectory over the time period.

## 2.2 Spatio-Temporal Graph Convolutional Neural Network

Let  $A \in \{0, 1\}^{N \times N}$  represent the adjacency matrix of the cardiac structure graph. If the  $i$ -th and the  $j$ -th nodes are connected,  $A_{i,j} = 1$ . Otherwise,  $A_{i,j} = 0$ . Let  $D \in R^{N \times N}$  indicate the diagonal degree matrix where  $D_{i,i} = \sum_j A_{i,j}$ .

**Spatio-temporal GCN.** Spatio-temporal GCN (ST-GCN) contains a range of the ST-GCN blocks [9]. Each block includes a spatial graph convolution and then a temporal convolution, which extracts spatial and temporal features respectively. The crucial element of ST-GCN is the spatial graph convolution operation, which provides a weighted average of neighboring features for each node.

Let  $F_{in} \in R^{C \times T \times N}$  be the input features, where  $C$  is the number of channels,  $T$  is the temporal length and  $N$  is the number of nodes in one frame. Let  $F_{out} \in R^{C \times T \times N}$  be the output features obtained from the spatial graph convolution. The spatial graph convolution is described as

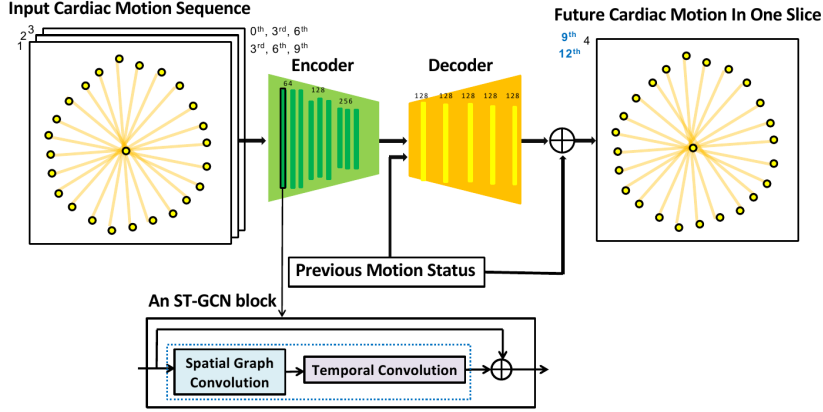
$$F_{out} = \sum M \circ \tilde{A} F_{in} W.$$

Here  $M$  is the edge weight matrix,  $W$  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.

In the temporal dimension, there are 2 corresponding nodes in the adjacent frames. Therefore, the number of neighbors for each node is 2. A 1D temporal convolution ( $K_t \times 1$  convolution) is applied to extract features,  $K_t$  is the kernel size of the temporal dimension.

**Encoder.** First of all, we normalize input cardiac structures with a batch normalization layer. The ST-GCN model contains 9 layers of spatial-temporal graph convolution blocks (ST-GCN blocks). The output channel for these layers are 64, 64, 64, 128, 128, 128, 256, 256, 256. Each layer has a temporal kernel size of 9. The ResNet mechanism is applied to each ST-GCN block. The strides of the 4-th and the 7-th temporal convolution layers are set to 2 as the pooling layer.

**Decoder.** There are 5 layers of ST-GCN blocks and the output channel of each layer is 128. The output of these blocks and the previous motion status are concatenated and put into an ST-GCN block. The output channel of this block is 2. The sum of the output of the decoder and the previous motion status predicts the future 2D node locations which represent the future cardiac motion trajectory. The loss function for future prediction in one frame is the mean squared error (MSE) of the node locations.



**Fig. 3.** Network overview. The cardiac motion sequence is given as input to the ST-GCN encoder-decoder framework. The output of the encoder and the previous motion status are fed into the decoder. 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 the left ventricular function evaluation.  $\oplus$  denotes the element-wise addition.

### 3 Experiments

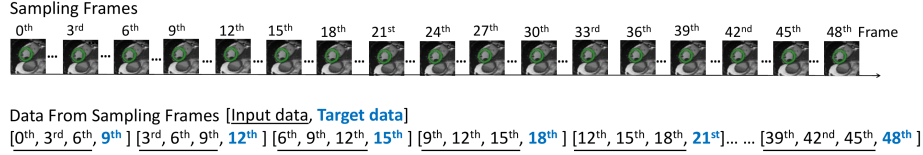
#### 3.1 Data Acquisition

In this study, we use short-axis view cardiac MR image sequences from the UK BioBank<sup>5</sup>. A stack of short-axis images, around 12 slices, cover the entire left and right ventricles. The 3 slices from the middle slice of short-axis images are selected for experiments. In-plane resolution is  $1.8 \times 1.8 \text{ mm}^2$ , while the slice gap is  $2.0 \text{ mm}$  and the slice thickness is  $8.0 \text{ mm}$ . Each sequence contains 50 consecutive time frames per cardiac cycle. We randomly selected image sequences of 1071 subjects for training, 270 subjects for validation and 270 subjects for testing.

<sup>5</sup> UK BioBank. <https://www.ukbiobank.ac.uk/>

### 3.2 Implementation Details

**Pre-processing.** The segmentation of the LV endocardial and epicardial borders and the RV was generated from using the FCN method proposed by Bai et al. [1] and used for nodes extraction. For training and testing, we obtained 1 barycenter node location of the LV and 24 node locations from both the endocardial and epicardial borders of the myocardium from the sampling frames shown in Fig.4. The detail of nodes extraction are described in section 2.1 and Fig.1. The input features are described with tensors (C, T, N). Here  $C$  denotes 2 channels for the 2D pixel coordinate  $(x, y)$ .  $T$  denotes 3 consecutive sampling time frames.  $N$  denotes 25 nodes in each time frame.

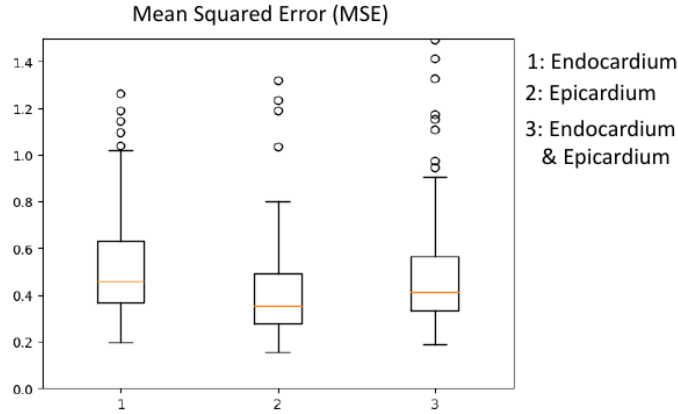


**Fig. 4.** Data preparation. Every third frame is selected from a cardiac MR image sequence. After sampling, 17 selected time frames denote a cardiac cycle for a sequence. The input cardiac structure data are extracted from 3 consecutive sampling frames and are used for cardiac motion estimation in the next sampling frame.

**Training.** The model is trained over 500 epochs using stochastic gradient descent with a learning rate of 0.01. The learning rate is decayed by 0.1 after every 10 epochs until the 50<sup>th</sup> epoch, after which the learning rate is constant. We set the input cardiac structure sequence length to 3 frames. As shown in Fig.4, 3 consecutive sampling time frames are selected for each sample. 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 compared three cardiac structure graphs and measured the predicted node locations with the ground truth using the mean square error (MSE). In the first cardiac structure graph, 24 nodes from the endocardium connect the centre node of the LV with edges. In the second graph, 24 nodes from the epicardium connect the centre node of the LV with edges. In the third graph, 48 nodes from both the endocardium and epicardium connect the centre node of the LV with edges. As shown in Fig.5, the second graph, graph nodes from the epicardium, have the least MSE and achieved better-predicted performance.



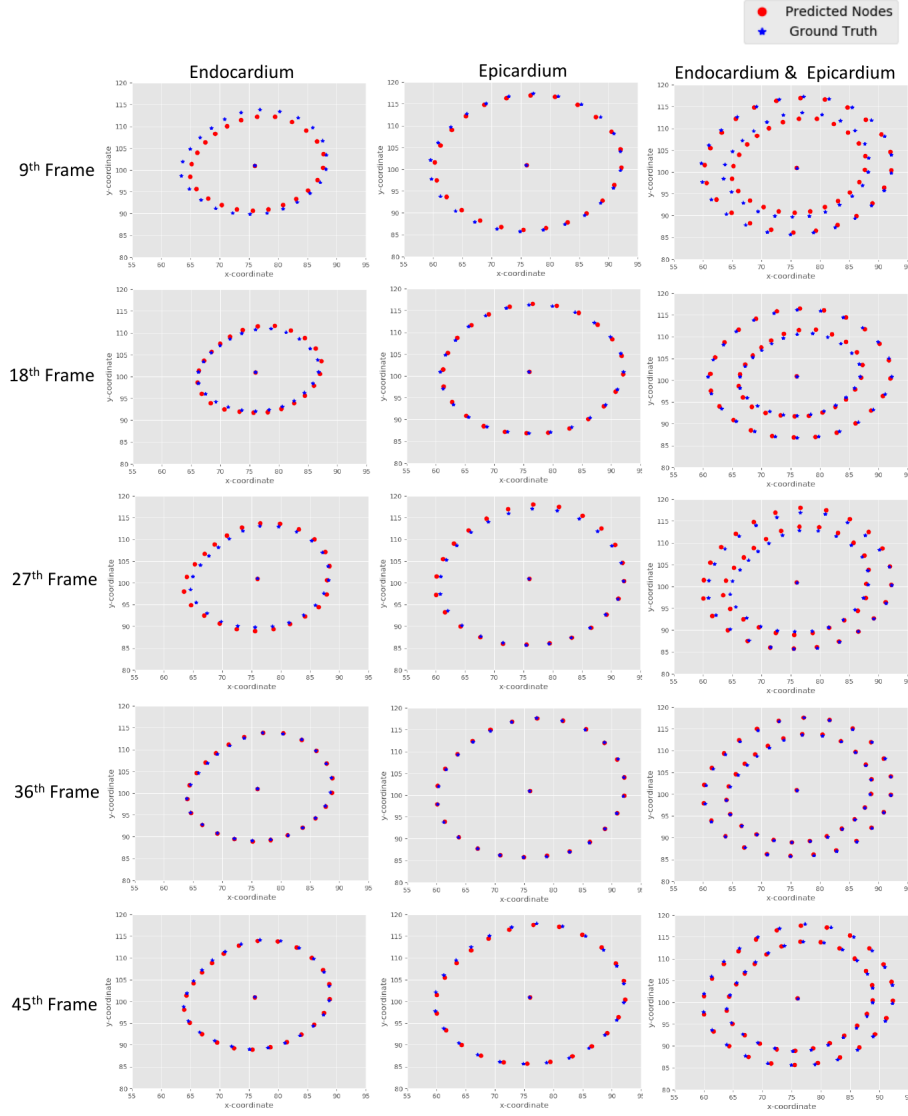
**Fig. 5.** Bar plot of the mean square error (MSE) for three different cardiac structure graphs using the proposed method. Graph nodes are from the endocardium, the epicardium, and both endocardium and epicardium. Each cardiac structure graph includes a node from the centre of the left ventricle.

**Representative examples.** There are 14 predicted sampling frames (frames 9 - 48, in increments of 3 frames (i.e 9, 12, 15 etc...) ) of node locations in a cardiac cycle for each sequence. Fig.6 shows an example of cardiac motion estimation on frames 9, 18, 27, 36, 45 of the MRI sequence between the proposed method with cardiac structure graph constructed from the endocardium and the epicardium, and a graph constructed from both the endocardium and epicardium. Both methods better predict the location of the epicardium nodes than the endocardium nodes. The node locations are predicted very accurately on the epicardium by both methods, especially on the 36<sup>th</sup> and 45<sup>th</sup> frame.

**Left ventricular function evaluation.** Based on the 17-Segment AHA model, the prediction of 24 node locations from both the endocardium and epicardium in each frame is classified into 6 segments [2]. Fig. 7 shows an example of a time series of the endocardial radius, thickness, radial strain (Err) [6] in the six segments of myocardium from the prediction and the ground truth of nodes' locations of a healthy volunteer. Compared to the ground truth, the prediction of the endocardial radius achieves better performance. The thickness and the Err has a similar plot shape.

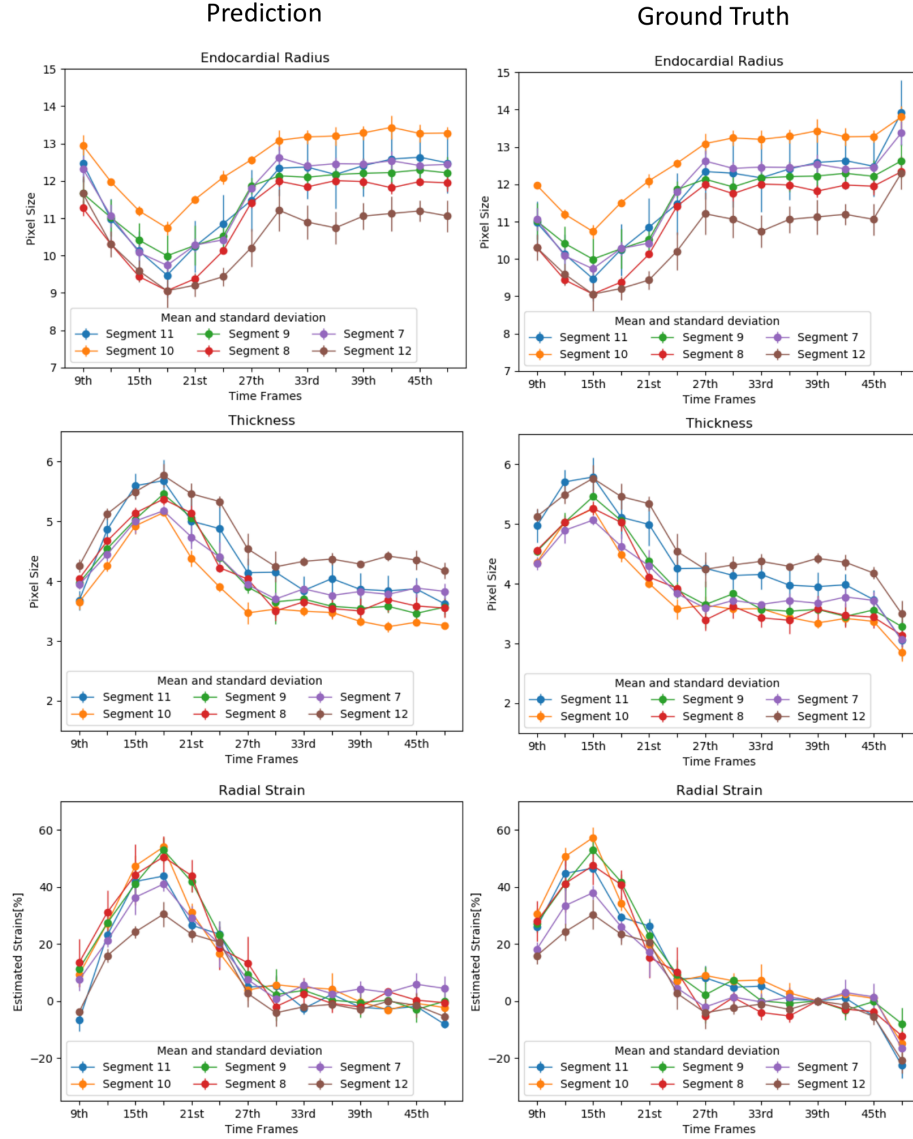
## 4 Discussion and Conclusion

In this work, we propose a spatio-temporal graph convolutional network (ST-GCN) to characterise cardiac motion. We evaluated and compared the effect of different cardiac structure graphs on cardiac motion estimation. In terms of accuracy, all the cardiac structure graphs yield good performance, especially



**Fig. 6.** Cardiac motion estimation comparison between the proposed method with different cardiac structure graphs and the ground truth. Both of these cases are modelled by three cardiac structure graphs constructed from the endocardium, the epicardium and both the endocardium and the epicardium. The predictions and the ground truth on the 9th, 18th(end-systolic), 27th, 36th and 45th frame of the MRI sequence are shown there.





**Fig. 7.** 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 9th, 12th, 15th, 18th(end-systolic), 21th, 24th, 27th, 30th, 33rd, 36th, 39th, 42th, 45th and 48th frame over a cardiac cycle. Results for prediction (top), and ground truth (bottom) in a healthy volunteer.

the prediction of the node locations on the epicardium. In the next step, we will compare the proposed method with other deep learning-based architectures, such as [6]. Future work will extend this method to analyse the basal and apical slices of the LV for both healthy heart and ischaemic heart disease.

**Acknowledgements.** This work is supported by SmartHeart. EPSRC grant EP/P001009/1. This research has been conducted using the UK Biobank Resource under Application Number 40119.

## References

1. Bai, W., Sinclair, M., Tarroni, G., Oktay, O., Rajchl, M., Vaillant, G., Lee, A.M., Aung, N., Lukaschuk, E., Sanghvi, M.M., et al.: Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *Journal of Cardiovascular Magnetic Resonance* **20**(1), 65 (2018)
2. Cerqueira, M.D., Weissman, N.J., Dilsizian, V., Jacobs, A.K., Kaul, S., Laskey, W.K., Pennell, D.J., Rumberger, J.A., Ryan, T., 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. Huang, Y., Bi, H., Li, Z., Mao, T., Wang, Z.: Stgat: Modeling spatial-temporal interactions for human trajectory prediction. In: *Proceedings of the IEEE International Conference on Computer Vision*. pp. 6272–6281 (2019)
4. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016)
5. Li, M., Chen, S., Chen, X., Zhang, Y., Wang, Y., Tian, Q.: Actional-structural graph convolutional networks for skeleton-based action recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 3595–3603 (2019)
6. Lu, P., Qiu, H., Qin, C., Bai, W., Rueckert, D., Noble, J.A.: Going deeper into cardiac motion analysis to model fine spatio-temporal features. In: *Annual Conference on Medical Image Understanding and Analysis*. pp. 294–306. Springer (2020)
7. Qin, C., Bai, W., Schlemper, J., Petersen, S.E., Piechnik, S.K., Neubauer, S., Rueckert, D.: Joint learning of motion estimation and segmentation for cardiac mr image sequences. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 472–480. Springer (2018)
8. Qiu, H., Qin, C., Le Folgoc, L., Hou, B., Schlemper, J., Rueckert, D.: Deep learning for cardiac motion estimation: Supervised vs. unsupervised training. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*. pp. 186–194. Springer (2019)
9. Yan, S., Xiong, Y., Lin, D.: Spatial temporal graph convolutional networks for skeleton-based action recognition. In: *Thirty-second AAAI conference on artificial intelligence* (2018)
10. Zheng, Q., Delingette, H., Ayache, N.: Explainable cardiac pathology classification on cine MRI with motion characterization by semi-supervised learning of apparent flow. *Medical image analysis* (2019)