Going Deeper into Cardiac Motion Analysis to Model Fine Spatio-Temporal Features

Ping Lu[1(✉)], Huaqi Qiu[2], Chen Qin[2], Wenjia Bai[3,4], Daniel Rueckert[2], and J. Alison Noble[1]

[1] Department of Engineering Science, University of Oxford, Oxford, UK
ping.lu@eng.ox.ac.uk
[2] Department of Computing, Imperial College London, London, UK
[3] Data Science Institute, Imperial College London, London, UK
[4] Department of Medicine, Imperial College London, London, UK

Abstract. This paper shows that deep modelling of subtle changes of cardiac motion can help in automated diagnosis of early onset of cardiac disease. In this paper, we model left ventricular (LV) cardiac motion in MRI sequences, based on a hybrid spatio-temporal network. Temporal data over long time periods is used as inputs to the model and delivers a dense displacement field (DDF) for regional analysis of LV function. A segmentation mask of the end-diastole (ED) frame is deformed by the predicted DDF from which regional analysis of LV function endocardial radius, thickness, circumferential strain (Ecc) and radial strain (Err) are estimated. Cardiac motion is estimated over MR cine loops. We compare the proposed technique to two other deep learning-based approaches and show that the proposed approach achieves promising predicted DDFs. Predicted DDFs are estimated on imaging data from healthy volunteers and patients with primary pulmonary hypertension from the UK Biobank. Experiments demonstrate that the proposed methods perform well in obtaining estimates of endocardial radii as cardiac motion-characteristic features for regional LV analysis.

Keywords: Cardiac MRI sequences · Cardiac motion · U-Net · Convolutional LSTM · Dense displacement field · Left ventricular function

1 Introduction

Magnetic resonance imaging (MRI) is widely used to assess cardiac function for cardiovascular disease diagnosis. Cardiac motion estimation highlights regional deformation of the myocardium, which is related to the severity of cardiovascular disease. Cardiac motion can be determined from the displacement field in MRI. Moreover, cardiac motion estimation can be regarded as an image registration

This work is supported by SmartHeart. EPSRC grant EP/P001009/1.

B. W. Papież et al. (Eds.): MIUA 2020, CCIS 1248, pp. 294–306, 2020.
https://doi.org/10.1007/978-3-030-52791-4_23

problem. Shen et al. [8] proposed a spatio-temporal 4D deformable registration method for cardiac motion estimation in MR image sequences. De Craene et al. [3] estimated motion and strain in $3D$ echocardiography by finding the $4D$ velocity field with spatio-temporal B-Spline kernels.

In recent years, deep learning-based methods have achieved promising results for deformable registration-based motion characterization. Zheng et al. [10] estimated cardiac motion using a variant of U-Net [7] with a semi-supervised learning strategy. Qin et al. [6] suggested a Siamese style recurrent spatial transformer network for cardiac motion estimation, to guide cardiac segmentation. Both of these works required expert manual segmentation of the left ventricle.

A major challenge is to estimate the effect of cardiac functional changes via automated cardiac motion analysis. The early onset of symptoms already causes an increased strain on the heart, but the strain-related changes are not always easy to see by eye until more significant cardiac structural changes occur. Motion-characteristic features, such as time series of the endocardial radius, thickness, circumferential strain (Ecc) and radial strain (Err) are related to cardiac disease and they are easy to explain as characteristics of pathological cardiac motion. Motion analysis is therefore also useful for early stage characterization of disease.

In this paper, we propose a deep learning-based architecture with a self-supervised strategy to characterise the spatio-temporal patterns of left ventricular (LV) cardiac motion in cardiac MR cine loops for improving the characterization of heart conditions. We compare the proposed method with two other state-of-the-art methods. Specifically, we extract motion-characteristic features and time series of the endocardial radius, thickness, Ecc and Err, based on the output dense displacement field (DDF) of the proposed method, and compare these features between a healthy group and a primary pulmonary hypertension (PPH) pathological group.

Contributions. The contributions of this work are as follows. (1) To our knowledge, this is the first attempt to exploit $2D + t$ spatio-temporal patterns with convolutional Long Short-Term Memory (ConvLSTM) in LV cardiac motion with a self-supervised strategy. (2) The predicted DDF of this method can be used to determine motion-characteristic features, namely a time series of the endocardial radius, thickness, Ecc and Err. These features are able to characterize different cardiac motion in health and pathologies. (3) We demonstrate that spatio-temporal patterns achieve better performance than the spatial-only pattern for cardiac motion estimation and regional analysis of LV function.

2 Spatio-Temporal Network

In this paper, cardiac motion estimation is considered as an image registration problem. The goal then becomes to estimate the spatial transformation of each point in the cardiac structure over the whole cardiac cycle. Let $\{I_t\}_{t=0,1,2,\ldots,N}$ indicate the cardiac MR cine loop frames, where N is the total number of frames. Each pixel-wise point x_0 from the end-diastole (ED) frame I_0 corresponds to a

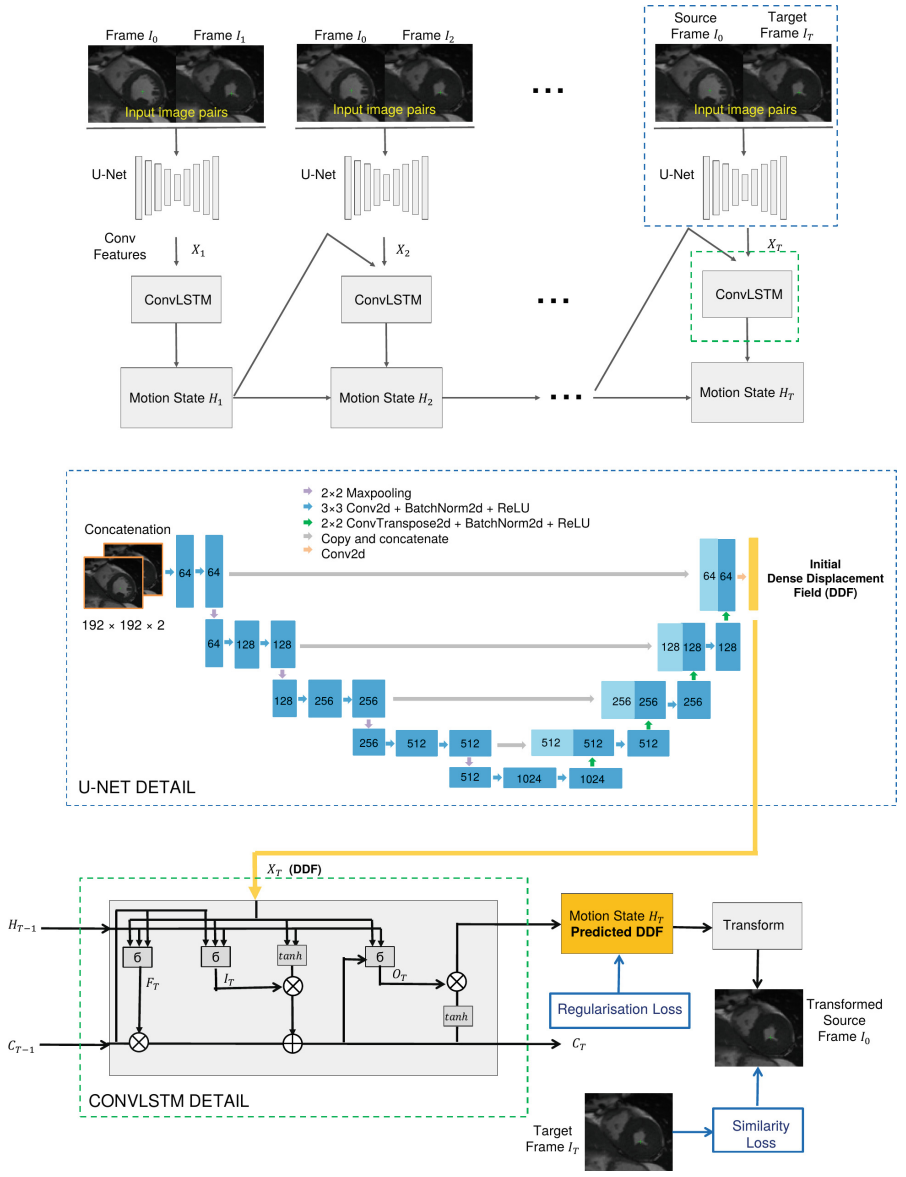

Fig. 1. Network Overview. A sequence of image pairs $\{(I_0, I_t)\}_{t=1,2,3,\ldots,n}$ is given as input to the U-Net convolutional network. The output of the U-Net, an initial dense displacement field (DDF), is fed to the convolutional LSTMs (ConvLSTM) to update the hidden states. The final output (predicted DDF) is used in subsequent analysis.

certain point x_t at the time frame t. In image registration, $I_t(x_t)$ and $I_0(T(x_0))$ denote the pixel value at same physical location. The spatial transformation T is represented by a DDF, described as u_t where $u_t(x_0) = x_t - x_0$.

We model a function $g_\theta(I_0, I_t) = u_t$ using a deep learning architecture, where θ are the optimal parameters of the architecture that can be trained by optimising a function that considers the similarity of the source-target image pair (I_0, I_t) and a spatio-temporal smoothness constraint. We estimate the motion from the ED frame I_0 to all other time frames I_t, and generate a new image sequence $\{I_t^{'}\}_{t=0,1,2,\ldots,N}$. The complete pipeline of the proposed architecture is presented in Fig. 1, and is described in Sect. 2.1.

2.1 Network Architecture

Our deep learning architecture is a combination of a fully convolutional network (FCN) and a recurrent neural network (RNN). We describe the function of the FCN and RNN as follows.

U-Net. The FCN component explores the spatial information in each 2D slice (intra-slice information). U-Net [7] is employed due to its well-known ability to represent image features for biomedical image segmentation. It consists of encoder and decoder parts with skip connections. The U-Net detail is shown in the middle part of Fig. 1. A sequence of source-target image pairs $\{(I_0, I_t)\}_{t=1,2,3,\ldots,N}$ is input to the U-Net convolutional network. The image pair is concatenated into a 2-channel $2D$ image. The encoder uses blocks of the 2D convolutional layers (3×3 kernel size), 2D batch normalization, rectified linear unit (ReLu) and 2D max pooling layer (2×2 window size). The decoder uses blocks of the transposed 2D convolutional layers (2×2 kernel size), 2D batch normalization and ReLu. The output of the U-Net is an initial dense displacement field (DDF), which is fed to initialise the LSTM to update the hidden states.

Convolutional LSTMs. The RNN component learns temporal relationships along the timeline (inter-slice information). We stack multiple convolutional LSTMs (ConvLSTM) [9], in order to increase the likelihood of detecting long-term dependencies of the cardiac motion over the cardiac cycle. We ran our architecture with different numbers of layers and kernel sizes in the ConvLSTM. Based on the validation performance, we stack 2 ConvLSTM layers with a 3-pixel kernel size in each layer. The number of input channels and the number of hidden channels of the ConvLSTM are each 2, where information in one channel represents the displacement in the x direction and in the other represents the displacement in the y direction.

The ConvLSTM can learn which information to keep in the long-term state, which information to drop, and which information to read. We present the details

of the LSTM in Fig. 1. Let the current input be X_T, and the previous hidden state is H_{T-1}. Then,

$$
\begin{aligned}
&I_T = \sigma(W_{XI} * X_T + W_{HI} * H_{T-1} + W_{CI} \circ C_{T-1} + B_I),\\
&F_T = \sigma(W_{XF} * X_T + W_{HF} * H_{T-1} + W_{CF} \circ C_{T-1} + B_F),\\
&C_T = F_T \circ C_{T-1} + I_T \circ tanh(W_{XC} * X_T + W_{HC} * H_{T-1} + B_C),\\
&O_T = \sigma(W_{XO} * X_T + W_{HO} * H_{T-1} + W_{CO} \circ C_T + B_O),\\
&H_T = O_T \circ tanh(C_T).
\end{aligned}
$$

Here $*$ is the convolution operator and $\circ$ is the Hadamard product (also called element-wise product). $W_{XI}, W_{XF}, W_{XO}, W_{XC}, W_{HI}, W_{HF}, W_{HO}$ and W_{HC} represent the convolutional filters. B_I, B_F, B_O and B_C are the biases for each layers. The input gate I_T controls which part of the new input information will be kept in the long-term state. The forget gate F_T decides which part of the long-term state is removed. The output gate O_T decides which part of the long-term state is read. C_T is the long-term state. The short-term state H_T is the motion state in cardiac MR cine loop frames and indicates the output - predicted DDF.

Loss Function. The loss function (L) is defined as the sum of an image intensity-based similarity loss L_m and a regularisation loss L_s on the predicted DDF displacements. Namely, $\boldsymbol{L} = \boldsymbol{L_m} + \boldsymbol{L_s}$. $\boldsymbol{L_m}$ measures the mean squared error between each pixel in the registered source image $I_0^{'}$ and the target image I_t. $\boldsymbol{L_m} = \frac{1}{N}\sum_{t=1}^{N}(I_t - I_t^{'})^2$. According to the spatial transformation network [4], I_0 is transformed to $I_t^{'}$ using bilinear sampling. The second term, L_s, is the spatial and temporal smoothness penalty, which controls the variation of displacements over space and time via an approximated Huber loss [6]. Mathematically, $\boldsymbol{L_s} = \lambda_1 \boldsymbol{L_{spatial}} + \lambda_2 \boldsymbol{L_{temporal}}$, where $\boldsymbol{L_{spatial}}$ calculates first-order spatial derivatives and $\boldsymbol{L_{temporal}}$ calculates first-order temporal derivatives. λ_1 and λ_2 are regularization parameters which are chosen empirically.

2.2 The Regional Analysis of Left Ventricular Function

The high-level steps in regional analysis of LV function are summarised in Fig. 2. The segmentation mask of the ED frame is deformed to another frame based on the predicted DDF. Automatic post-processing is applied to identify the LV endocardial and epicardial borders. To smooth the borders of deformed masks on the mid-slice 6-segments model of the 17-Segment AHA model [2], we performed a morphological closing operation (kernel size = 2) on them.

We divide the resulting predicted myocardium mask into segments based on the 17-Segment Model (AHA). Firstly, we find the barycenter of the 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 points as the initial line. Thirdly, we rotate this initial line around the barycenter point of the LV by 60, 120, 180, 240, 300° and divide the middle slice into 6 segments. Morphological transformations

and barycenter location are implemented using OpenCV. The time series of the endocardial radius, thickness, Ecc and Err are measured in these 6 segments. In each segment, mean and standard deviation are used to show the rich detail. To this aim, we sample all the points on the endocardial border for the endocardial radius, 5 points by every 12° for the thickness and Err. Considering the small perimeter on the end-systolic (ES) frame, we divide the endocardial border into 3 sets instead of 5 sets for Ecc.

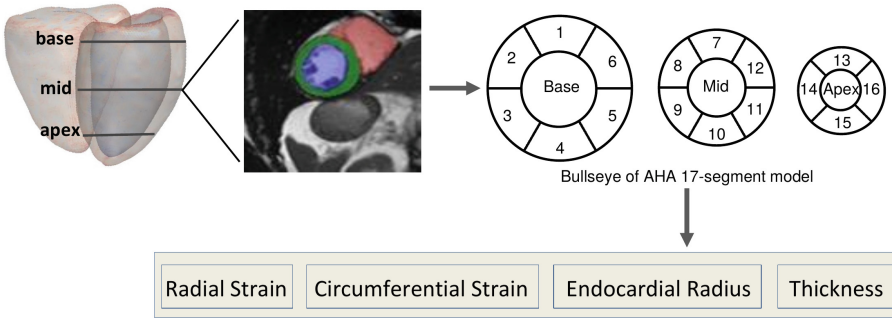

Fig. 2. Overview of the proposed framework for quantifying cardiac motion. The predicted DDF is applied to deform the segmentation mask of the ED frame from which the regional analysis of left ventricular endocardial radius, thickness, circumferential strain (Ecc) and radial strain (Err) can be estimated.

Strain Computation. Left ventricular strain indicates the deformation of the myocardium over the whole cardiac cycle and is shown in percentages. In each time frame T, circumferential strain (Ecc) and radial strain (Err) are computed as $E = \frac{d_T - d_{ED}}{d_{ED}} \times 100\%$. Here d_{ED} is the length on the ED frame, d_T is the length on the time frame T. In each sample, we choose the arc length of the endocardial border for the Ecc computation and LV wall thickness for the Err computation.

3 Experiments

3.1 Data Acquisition

Short-axis view cardiac MR image sequences from the UK BioBank[1] were used in this study. The CMR is obtained from a 1.5 T scanner (MAGNETOM Aera, Syngo Platform VD13A, Siemens Healthcare, Erlangen, Germany). A stack cine balanced steady-state free precession (bSSFP) of short-axis images, around 12 slices, covers the entire left and the right ventricles. In-plane resolution is $1.8 \times 1.8\,\text{mm}^2$, while the slice thickness is 8.0 mm and slice gap is 2.0 mm.

[1] UK BioBank. https://www.ukbiobank.ac.uk/.

Each sequence contains 50 consecutive time frames per cardiac cycle. We randomly selected image sequences of 450 subjects for training, 47 subjects for validation and 100 subjects for testing.

3.2 Implementation Details

Pre-processing. For training and testing the deep learning architecture, all images were cropped to a size of 192×192 pixels because of GPU limitations, and the intensity normalisation applied to the cropped images. The segmentation mask of the LV endocardial and epicardial borders and the right ventricular (RV) endocardial borders at the ED frame was generated from using the FCN method proposed by Bai et al. [1] and used to quantify cardiac motion.

Training. The model is trained over 150 epochs using Adaptive Moment Estimation (Adam) optimisation [5] with learning rate 0.0001 and a batch size of 1. For the smoothness penalty of the loss, we set λ_1 to 0.002 and λ_2 to 0.0002 based on algorithm performance on the validation dataset. Further, we randomly select one frame in the selected slice to be frame I_0. We set the input image sequence length to 20 frames due to GPU memory limitations. The proposed network was implemented using Python 3.7 with Pytorch. All the experiments are run with computational hardware GeForce GTX 1080 Ti GPU 10 GB.

3.3 Evaluation Metrics.

To quantify the similarity between the predicted image and the target image, we use three image metrics: the normalised root mean-squared error (NRMSE), the mean structural similarity index (MSSIM) and the peak signal to noise ratio (PSNR). A two-sided Wilcoxon signed rank test is used to find where there is a statistically significant difference in these three metrics among three methods.

4 Results

4.1 Quantitative Results

Table 1 summarizes the comparative results on the MRI sequences and the ES frame between the proposed and other methods. It is observed that the proposed method is superior to Qin et al.'s method [6] and U-Net [7]. The proposed method achieves an accuracy with a NRMSE of 0.053 ± 0.017, MSSIM of 0.851 ± 0.049, and PSNR of 35.391 ± 2.976 on the MRI sequences, and a NRMSE of 0.065 ± 0.012, MSSIM of 0.836 ± 0.036, and PSNR of 33.399 ± 1.120 on the ES frame. U-Net yielded the lowest MSSIM and PSNR value and the highest NRMSE value on both the MRI sequences and the ES frame among the evaluated approaches. Using a two-sided Wilcoxon signed rank test, statistically significant greater results than Qin et al.'s and U-Net were obtained ($p < 0.05$) for all the measurements.

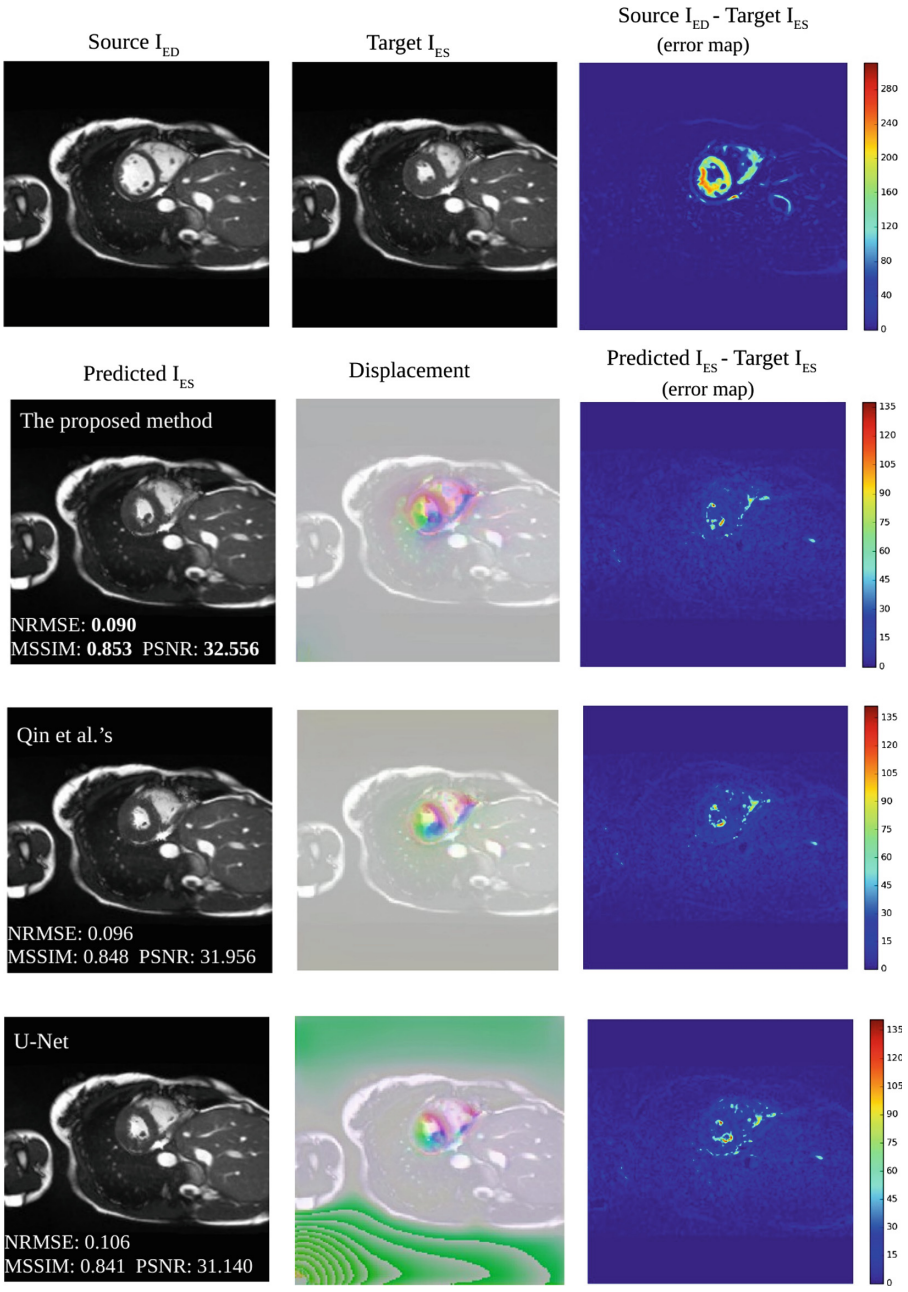

Fig. 3. Cardiac motion estimation comparison on the ES frame of the MRI sequences between (top row to bottom row) the proposed method, Qin et al.'s method [6] and U-Net [7].

Table 1. Quantitative comparison on the MRI sequences and the ES frame between our method and two other methods, Qin et al.'s [6] and U-Net [7]. The results are presented as mean ± standard deviation. The best performance is indicated in bold. The ⋆ indicates that our method results are statistically significant greater ($p < 0.05$) than other methods using a two-sided Wilcoxon signed ranks test.

Method	Proposed method	Qin et al.'s	U-Net
The MRI sequences			
NRMSE	**0.053 ± 0.017** ⋆	0.059 ± 0.020	0.075 ± 0.030
MSSIM	**0.851 ± 0.049** ⋆	0.848 ± 0.050	0.825 ± 0.060
PSNR	**35.391 ± 2.976** ⋆	34.633 ± 3.301	32.651 ± 3.725
ES Frame			
NRMSE	**0.065 ± 0.012** ⋆	0.075 ± 0.014	0.091 ± 0.024
MSSIM	**0.836 ± 0.036** ⋆	0.829 ± 0.036	0.806 ± 0.048
PSNR	**33.399 ± 1.120** ⋆	32.168 ± 1.325	30.595 ± 1.876

4.2 Representative Examples

Cardiac Motion Estimation. Figure 3 shows an example cardiac motion estimation comparison on the $19th$ frame (ES) of the MRI sequence between the proposed method, Qin et al.'s method [6] and U-Net [7], using spatial-only patterns. It is observed that the proposed method provides a higher MSSIM 0.853 and PSNR 32.556 and a lower NRMSE 0.090 than the other methods on the predicted ES frame. The displacement image visualizes the DDF. Different colours describe the different motion directions, and the colour intensity expresses the magnitude of the displacement. The proposed method estimates higher displacements (visualised as a stronger colour in Fig. 3 middle column) compared to other methods, especially at the centre area of the LV blood pool. The U-Net seems to be less accurate, because it has strong background noise (shown in green) compared to the proposed method and the Qin et al.'s method. The displacement error maps show that the U-Net has the largest difference at the LV and the surrounding area, followed by the method of Qin et al.

Left Ventricular Function Evaluation. In our dataset, we do not have manual image segmentation. In order to do regional analysis of LV function, we ran Bai et al.'s algorithm [1] to get the segmented ED frame. Then we warped the segmented ED frame to other frames in the sequence. Table 2 and Fig. 4 shows an example of a healthy volunteer and a primary pulmonary hypertension (PPH) patient with the proposed method. Figure 4 shows an example of a time series of the endocardial radius, thickness, Err and Ecc in the six segments of myocardium estimated for a healthy volunteer and a PPH patient. Compared to a healthy volunteer, the LV of the PPH patient has poor contraction over the whole cardiac cycle, and as a result, the endocardial radius of a hypertension patient is larger than that of a healthy volunteer. For instance, the endocardial radius (orange)

of segment 1 contracts less. Table 2 shows that on the $19th$ frame (ES), the mean radius of segment 1 is 10.69 pixel from the PPH patient, while the mean radius of segment 1 is 9.79 pixel from the healthy volunteer. In clinical practice, the endocardial radius should take on its smallest value over the cardiac cycle on the ES frame, because the volume of the LV blood pool reaches the minimum value then. Moreover, the LV wall thickness from all six segments is smaller for the PPH patient, compared to the healthy one. Due to the reduced thickness, we conclude that this left ventricle exhibits atrophy.

Table 2. Example results of peak mean value on the ES frame of the motion- characteristic features, time series of the endocardial radius (Endo radius), and thickness, circumferential (Ecc) and radial strain (Err) for cardiac segments (Seg) (0–5) over a cardiac cycle for a healthy volunteer and a primary pulmonary hypertension (PPH) patient in the proposed method.

Healthy						
Feature	Seg 0	Seg 1	Seg 2	Seg 3	Seg 4	Seg 5
Endo radius	9.344	9.789	9.836	9.145	9.333	9.537
Thickness	6.721	6.451	8.265	7.865	7.249	7.437
Err	57.154	60.842	66.751	60.259	27.565	22.271
Ecc	−11.728	−14.485	−8.657	−9.314	−11.071	−9.586
PPH						
Endo radius	9.643	10.693	9.139	9.675	10.465	9.174
Thickness	4.143	6.203	5.395	5.069	4.643	5.218
Err	39.992	97.947	58.334	52.159	55.657	31.872
Ecc	−8.657	−5.586	−9.414	−6.899	−8.657	−7.414

5 Discussion

In this work we have proposed a deep learning-based approach to cardiac MR motion analysis that uses a self-supervised paradigm to learn spatio-temporal features in cardiac MR cine loops. The results show the ability of the proposed approach to capture spatio-temporal patterns and predict a dense displacement field (DDF) over a full cardiac cycle. The proposed method has higher accuracy than the method of Qin et al. and U-Net which we attribute to the use of spatio-temporal features. According to our experiments, the best DDF results are obtained when we stack 2 ConvLSTM layers with a 3-pixel kernel size in each layer.

The predicted DDF is employed to deform an ED myocardium mask to other frames and perform regional LV endocardial radius, thickness, Ecc and Err time-series analysis. The results show the potential of the proposed approach to evaluate the clinical parameters for cardiovascular diseases. Currently, we do not

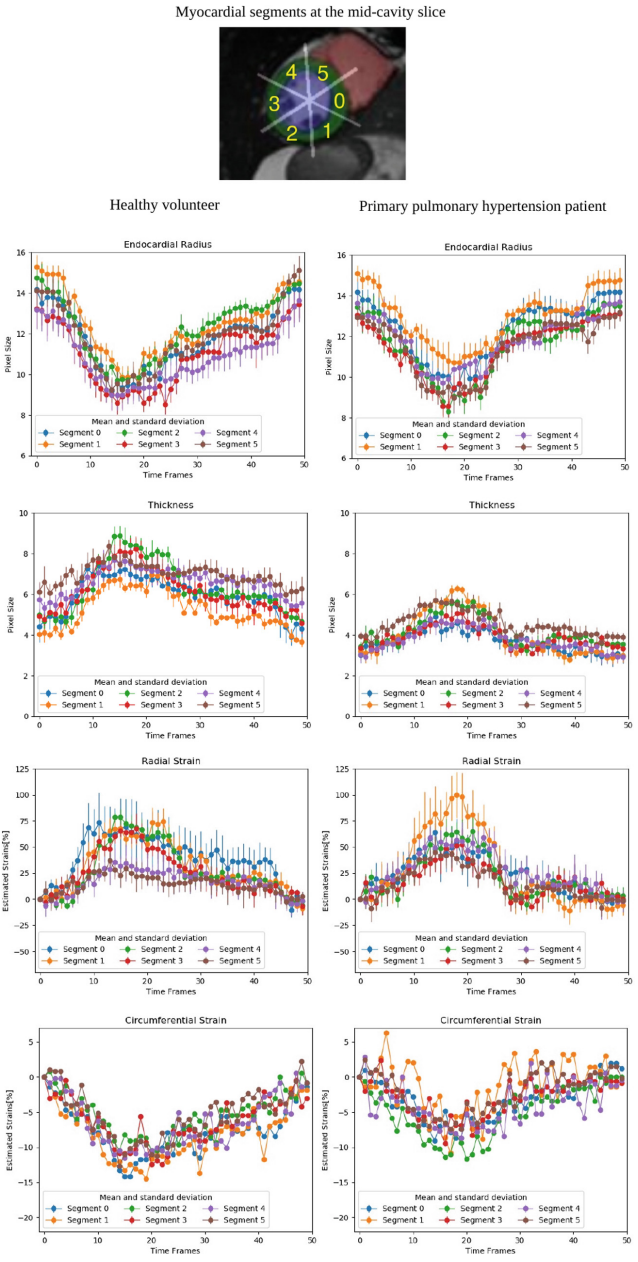

Fig. 4. Example results of estimated endocardial radius (mean and standard deviation shown), thickness(mean and standard deviation shown), radial strain (mean and standard deviation shown) and circumferential strain (mean and standard deviation shown) for cardiac segments (0–5) plotted over a cardiac cycle. Myocardial segment notation (top); and results for a healthy volunteer (left column), and a primary pulmonary hypertension patient (right column).

use interpolation to smooth feature time series. In our experiments, we find that it is not necessary to smooth the curve. We can use the unsmoothed curve of the endocardial radius to explain the abnormal motion phenomenon in the PPH pathological group.

There are some limitations of this work. The UK BioBank consists of mainly healthy volunteers, and has a sparse number of PPH patients. The model may not well represent the motion and strain patterns typically seen in PPH patients.

6 Conclusion

We present a novel spatio-temporal network to characterise cardiac motion, visualise the dense displacement field and explain motion-characteristic features in a healthy group and a pathological group. The model learns meaningful spatio-temporal patterns of the cardiac motion that can be used for LV regional function analysis. Future work will extend this method to analyse the basal, mid-cavity and apical slices of the LV. The motion and strain analysis method is not disease-specific and could be extended extend to other cardiac conditions such as ischaemic health disease, assuming suitable training examples are available.

References

1. Bai, W., et al.: Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. J. Cardiovasc. Magn. Reson. **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. De Craene, M., et al.: Temporal diffeomorphic free-form deformation: application to motion and strain estimation from 3D echocardiography. Med. Image Anal. **16**(2), 427–450 (2012)
4. Jaderberg, M., Simonyan, K., Zisserman, A., et al.: Spatial transformer networks. In: Advances in Neural Information Processing Systems, pp. 2017–2025 (2015)
5. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
6. Qin, C., et al.: Joint learning of motion estimation and segmentation for cardiac MR image sequences. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11071, pp. 472–480. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00934-2_53
7. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
8. Shen, D., Sundar, H., Xue, Z., Fan, Y., Litt, H.: Consistent estimation of cardiac motions by 4D image registration. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 902–910. Springer, Heidelberg (2005). https://doi.org/10.1007/11566489_111

9. Xingjian, S., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.C.: Convolutional LSTM network: a machine learning approach for precipitation nowcasting. In: Advances in Neural Information Processing Systems, pp. 802–810 (2015)
10. Zheng, Q., Delingette, H., Ayache, N.: Explainable cardiac pathology classification on cine MRI with motion characterization by semi-supervised learning of apparent flow. Med. Image Anal. **56**, 80–95 (2019)