Computers in Biology and Medicine 120 (2020) 103728

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: http://www.elsevier.com/locate/compbiomed

MV-RAN: Multiview recurrent aggregation network for echocardiographic sequences segmentation and full cardiac cycle analysis

Ming Li [a,b], Chengjia Wang [c], Heye Zhang [d], Guang Yang [e,f,*]

[a] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
[b] Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen, China
[c] BHF Centre for Cardiovascular Science, The University of Edinburgh, Edinburgh, EH16 4TJ, UK
[d] School of Biomedical Engineering, Sun Yat-Sen University, Shenzhen, China
[e] National Heart and Lung Institute, Imperial College London, London, SW7 2AZ, UK
[f] Car-diovascular Research Centre, Royal Brompton Hospital, London, SW3 6NP, UK

ARTICLE INFO

Keywords:
Echocardiography
Multiview learning
Cardiac segmentation
Ultrasound
Machine learning

ABSTRACT

Multiview based learning has generally returned dividends in performance because additional information can be extracted for the representation of the diversity of different views. The advantage of multiview based learning fits the purpose of segmenting cardiac anatomy from multiview echocardiography, which is a non-invasive, low-cost and low-risk imaging modality. Nevertheless, it is still challenging because of limited training data, a poor signal-to-noise ratio of the echocardiographic data, and large variances across views for a joint learning. In addition, for a better interpretation of pathophysiological processes, clinical decision-making and prognosis, such cardiac anatomy segmentation and quantitative analysis of various clinical indices should ideally be performed for the data covering the full cardiac cycle. To tackle these challenges, a multiview recurrent aggregation network (MV-RAN) has been developed for the echocardiographic sequences segmentation with the full cardiac cycle analysis. Experiments have been carried out on multicentre and multi-scanner clinical studies consisting of spatio-temporal (2D + t) datasets. Compared to other state-of-the-art deep learning based methods, the MV-RAN method has achieved significantly superior results (0.92 ± 0.04 Dice scores) for the segmentation of the left ventricle on the independent testing datasets. For the estimation of clinical indices, our MV-RAN method has also demonstrated great promise and will undoubtedly propel forward the understanding of pathophysiological processes, computer-aided diagnosis and personalised prognosis using echocardiography.

1. Introduction

Echocardiography plays an integral role in clinical cardiology, with important applications in diagnosis, patient management, and decision-making for a wide range of cardiovascular diseases (CVD) [29]. Compared to other imaging modalities, e.g., cardiac MR and CT, echocardiography has a number of advantages such as its non-invasive low-risk nature without using ionizing radiation, convenient and portable equipment, simplicity and low-cost imaging. In addition, recent development of transesophageal echocardiography and the technological advance of transthoracic transducers have resulted in significant improvement of the image quality. Therefore, nowadays, echocardiography is the most widely used method for the interpretation of pathophysiological processes, clinical decision-making and longitudinal therapy efficacy assessment for the patients with CVD [34].

Accurate delineation of the left ventricle (LV) is a prerequisite for reliable measurement of the cardiac morphology and function, including the extraction of the ejection fraction (EF). Recently as deep learning (DL) based methods have generally returned dividends in performance for various medical image analysis problems including segmentation of echocardiographic images. Therefore, we have broadly classified existing methods of LV delineation into *non-DL methods and DL based methods* and provide a concise literature review, and detailed surveys on this topic can be found elsewhere, e.g., for both 2D [5,26] and 3D [2,19] echocardiographic images.

(1) *Non-DL Methods*: Early attempts on LV segmentation from 2D or 3D echocardiographic images included using geometrical

* Corresponding author. National Heart and Lung Institute, Imperial College London, London, SW7 2AZ, UK.
E-mail address: g.yang@imperial.ac.uk (G. Yang).

https://doi.org/10.1016/j.compbiomed.2020.103728
Received 11 February 2020; Received in revised form 21 March 2020; Accepted 21 March 2020
Available online 24 March 2020

deformable models, shape-free clustering and level sets based methods [19], and statistical shape modelling, e.g., the active contour [6] and active appearance models [4]. However, the lack of standardized and publicly-accessible dataset has restrained a thorough assessment and fair comparison of these proposed methods. In 2016, Bernard et al. [2] carried out a benchmark study, in which several non-DL methods were compared fairly using the same 3D echocardiography dataset published by the MICCAI Challenge on Echocardiographic Three-Dimensional Ultrasound Segmentation (CETUS). Wang and Smedby [43]. proposed a semi-automatic segmentation method combining a multi-scale quadrature filter method with the model based level set to segment the LV endocardium. A real-time automated LV tracking method was proposed by Smistad and Lindseth [38], in which Kalman filtering and edge detection were used to track the mesh of the LV in every single frame of the 3D echocardiographic sequence. Barbosa et al. [1] presented an optical flow based LV tracking method on 4D echocardiographic sequences with local block matching. A graph cut based interactive segmentation method was demonstrated by Bernier et al. [3]. In another work, Milletari et al. [23] reported a learning based segmentation method using Hough forest with a shape-prior. Active appearance models [41] and multi-atlas [28] based statistical shape analysis methods were also applied to segment the LV from 3D echocardiographic data. In addition, random forest based machine learning algorithms were developed and tested on solving this challenging segmentation task [8,15]. Among these nine benchmarked methods, four of them were semi-automated [3,8,28,43] and these semi-automated methods tended to perform better with highest Dice score of 0.89 at end-diastolic (ED) and 0.87 at end-systolic (ES) phases, respectively. However, the processing time was from a few seconds up to 32 min for these non-DL methods. Beyond the studies tested by the CETUS open challenge, Huang et al. [12,13] proposed a sparse representation and dictionary learning based dynamical appearance model, which was constrained by inherent spatiotemporal coherence of individual data, to track both endocardial and epicardial contours of the LV from echocardiographic sequences. Recently, Leclerc et al. [17] investigated a machine learning solution based on structured random forest algorithm to accomplish the segmentation task for 2D echocardiographic images. Pedrosa et al. [32] extended the B-spline explicit active surfaces approach originally proposed by Barbosa et al. [1] and incorporated a shape prior derived from principal component analysis.

(2) *DL Based Methods*: By utilizing deep belief networks, Carneiro et al. [5] designed a two-step LV segmentation framework that consisted of an automatic region of interest (ROI) detection and selection step and an LV boundary delineation step. Their framework was evaluated on a dataset containing 400 annotated diseased cases and 80 annotated normal cases where all the cases presented long axis views of the LV. By varying the amount of the used training cases, their proposed method could achieve the best average Hausdorff distance of ~17 mm and the best average mean absolute distance of ~8 mm when they used 400 training cases. Yu et al. [44] proposed a dynamic convolutional neural networks (CNN) based method using multi-scale information and fine-tuning pretrained model for fetal LV segmentation from echocardiographic sequences, in which the dynamic CNN was fine-tuned with deep tuning from the first frame and shallow tuning with the following frames to adapt to the individual fetus. Their proposed algorithms achieved a Dice score of 0.95 compared to the ground truth. In Ref. [7], the authors used a multi-domain regularized fully convolutional networks (FCN) and transfer learning to implement an end-to-end learning framework for LV detection and segmentation. They obtained consistently good results (Dice: 0.86–0.89) for the echocardiographic data acquired from different views. Smistad et al. [39] demonstrated a U-Net [35] based CNN method for segmenting LV in 2D echocardiographic images. To reduce the scale of annotated training data, a pre-training strategy with automatic Kalman filter based segmentation was applied. In doing so, the Dice score was improved slightly (average Dice of 0.87 using CNN vs, 0.86 using the Kalman filter based segmentation), but the Hausdorff distance was significantly better with an average of 5.9 mm for the CNN vs. 7.5 mm derived from the Kalman filter method. Recently, Oktay et al. [27] reported an anatomically constrained neural network (ACNN), which combined 3D U-Net with prior anatomical knowledge, to solve the LV segmentation problems for both cardiac MRI and 3D echocardiographic images (CETUS datasets). They obtained average Dice scores of 0.91 and 0.87 for the end-diastole (ED) and end-systole (ES) phases, respectively. Veni et al. [42] coupled the U-Net with level set based segmentation to realise a shape-guided deformable model driven by an end-to-end trained fully convolutional network acting as the prior. By evaluating on a private 2D echocardiographic dataset, 0.86 ± 0.06 Dice score was obtained for the LV segmentation. Generative adversarial networks (GAN) [9] based models were modified to segment paediatric echocardiographic data [10] and image quality transfer followed by U-Net based LV segmentation [14]. In addition, recently, a bilateral segmentation network incorporating attention mechanisms was proposed to solve the fully automatic segmentation of paediatric echocardiography images in the four chamber view [11]. Promising results were yielded from all these studies. More recently, Leclerc et al. [18] evaluated several DL methods with encoder-decoder-based architectures, e.g., U-Net [35], ACNN [27], Stacked Hourglasses [25], U-Net++ [45], for the segmentation of cardiac anatomy from 2D echocardiographic images. This study found that (1) DL based methods outperformed state-of-the-art non-DL methods and (2) all DL based methods achieved similar results; however, more sophisticated encoder-decoder architectures, e.g., U-Net++, did not improve the results compared to the segmentation obtained by the U-Net.

Among various echocardiography techniques, 2D echocardiography is the most widely accessible and relatively cheaper option for patients with CVD. Quantitative analysis of 2D echocardiography such as myocardial motion analysis [30,31,33] is of the essence in clinical routine to estimate the cardiac morphology and function. However, such quantification is better to be performed using multiview (MV) 2D spatio-temporal (2D + t) echocardiographic sequences covering the full cardiac cycle, from which complementary information from different views and motion features can be extracted. The apical views are the most widely used views in the echocardiographic exam that provide crucial cardiac functional indices (e.g., EF, wall thickening, and strain) for the assessment of ventricular diastolic and systolic function and atrioventricular valve function and structure estimation [40]. The apical views transect the true cardiac apex that align parallel to the cardiac long-axis, which have typically acquired as MV datasets for the patients including (1) the apical two-chamber (A2C) view, (2) the apical three-chamber (A3C) or apical long axis (ALAX) view, (3) the apical four-chamber (A4C) view, and (4) the apical five-chamber (A5C) view [29,40]. These MV echocardiographic sequences are crucial in clinical decision-making [21].

Segmentation of the LV from MV 2D + t echocardiographic sequences is an essential step for reproducible and reliable quantitative cardiac functional analysis [16]. However, robust and accurate automatic segmentation of the LV is still very challenging for MV 2D + t echocardiographic sequences because of gross intensity inhomogeneities, various image quality affected by artefacts (e.g., attenuation, deformations, shadows, and signal dropout) [22], and poor contrast between regions of interest [12]. Moreover, discrepancies of

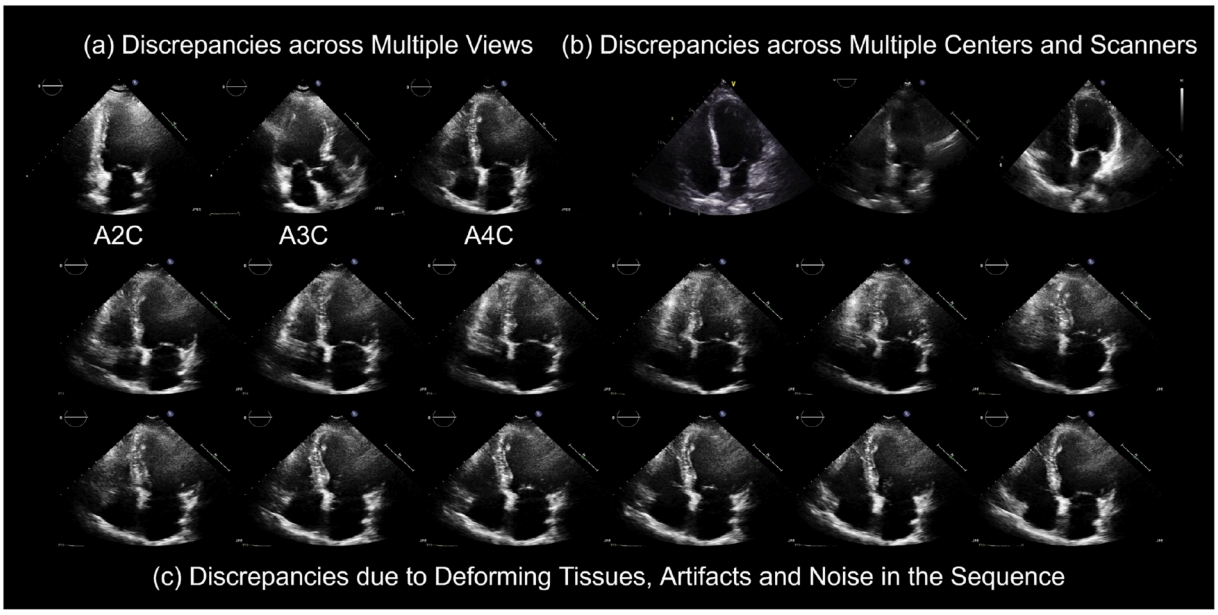

Fig. 1. Challenges of the discrepancies of image appearance in (a) multi-view datasets, e.g., A2C, A3C, and A4C, (b) multiple centres and scanners datasets and (c) echocardiographic sequences with various deforming tissues, artefacts and noise.

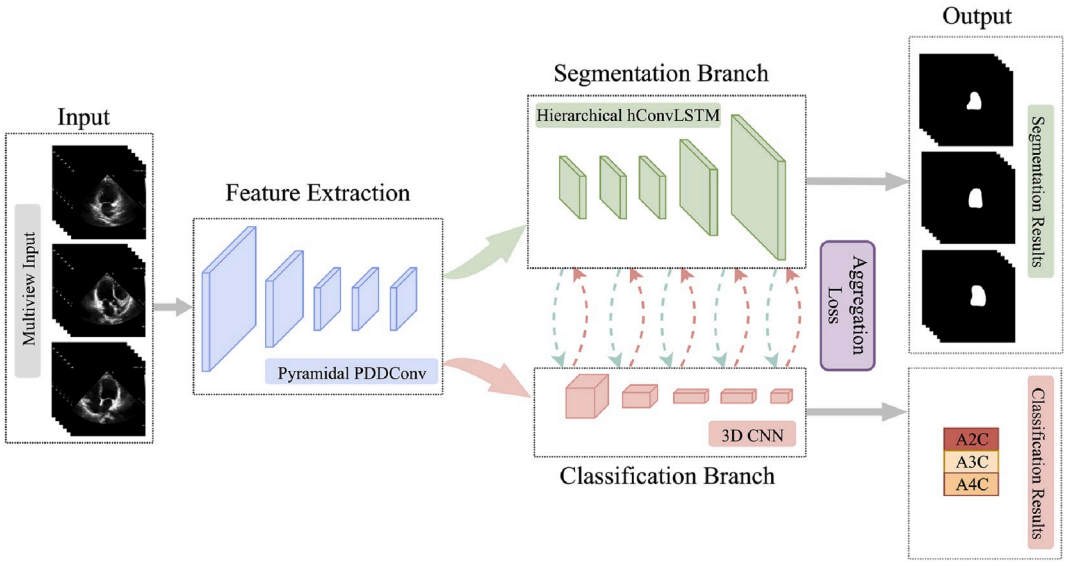

Fig. 2. Schema of our MV-RAN method.

image appearance due to differences across MV acquisitions and among multicentre and multi-scanner studies could further complicate the segmentation process (Fig. 1).

Application scenarios of existing methods are always limited, as mentioned in *non-DL methods and DL based methods*, they mostly focus on a single view, on specific frames (ED and ES), or on a single centre and scanner. Besides, when it comes to sequence segmentation, existing methods try to capture spatial-temporal information by leveraging a deformable model combined with optical flow or fine-tuning pretrained CNN dynamically from the first frame to the last one. The major disadvantages are that these methods are computational cumbersome and not a unified framework with an end-to-end manner.

To tackle the aforementioned challenges, in this study, a multiview recurrent aggregation network (MV-RAN) has been developed for the echocardiographic sequences segmentation with the full cardiac cycle analysis. The workflow of MV-RAN is depicted in Fig. 2. The feature extraction part captures multi-level and multi-scale spatial-temporal information by the pyramid dilated dense convolution (PDDConv), endowing MV-RAN with distinguished feature extraction capacity and the LV detection ability in multi-level and multi-scale space. Next, the segmentation branch fuses multi-level and multi-scale spatial-temporal information by the hierarchical convolutional layers with Long short-term memory (LSTM) recurrent units (hConvLSTM), helping to constrain the LV boundaries in the sequence and achieve better semantic representation. They jointly harness the knowledge embedded in the multicentre and multi-scanner dataset. Further, a novel double-branch aggregation mechanism is introduced to perform simultaneous MV segmentation and classification, which enables MV-RAN to cope with the differences across MV data. Different from existing methods, MV-RAN fully exploits the long term spatial-temporal information in an end-to-end manner and does not depend on any deformable model or optical flow or pretrained segmentation models. MV-RAN can handle

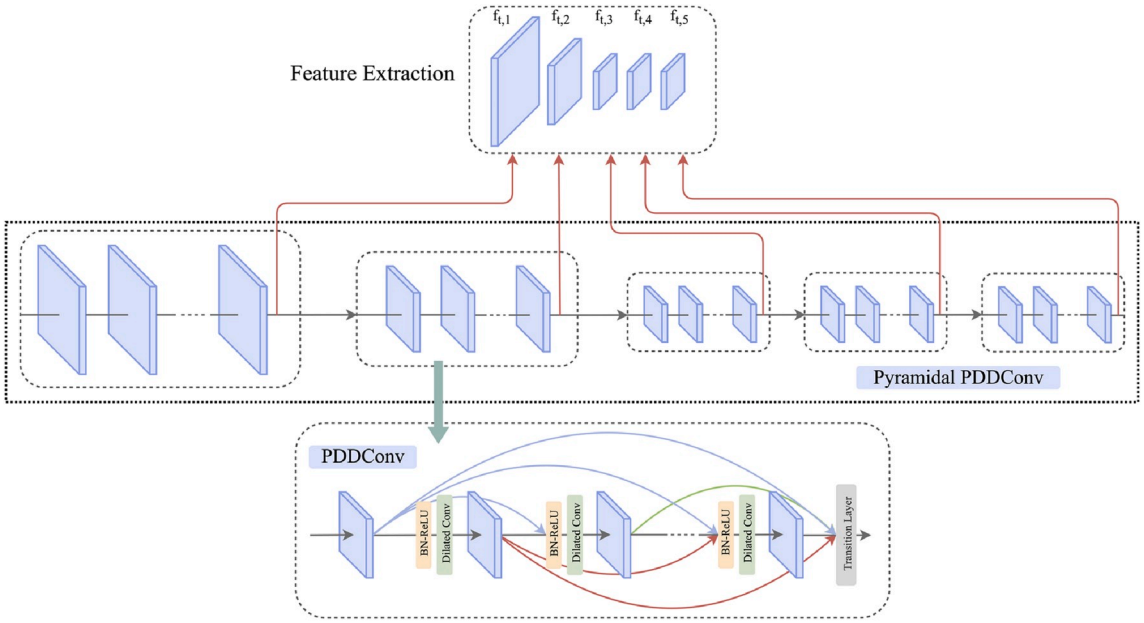

Fig. 3. Schema of the pyramidal PDDConv layers.

heterogeneous data, not only generate accuracy segmentation and classification but also provide full cardiac cycle analysis.

This work is an extension of the work that has been partially presented in our conference paper [20]. In this paper, we elaborate our work with further details of theories, optimization, implementation, comparison studies and limitations which were not covered in our conference paper. The salient extensions and main contributions of this work are summarised as follows:

- The proposed MV-RAN provides a unified framework for both MV echocardiographic sequences segmentation and classification and full cardiac cycle analysis.
- PDDConv joint hConvLSTM endows MV-RAN powerful ability to extract and fuse multi-level and multi-scale spatial-temporal information, harnessing the knowledge embedded in the multicentre and multi-scanner dataset.
- A double-branch aggregation mechanism is designed to perform simultaneous MV segmentation and classification. Two branches exert mutual promotion to cope with the differences across MV data.
- MV-RAN fully exploits the long term spatial-temporal information in an end-to-end manner without using any deformable model or optical flow or pretrained segmentation models.
- We have extended our evaluation to a larger clinical dataset compared to our conference publication. In addition, we also incorporate the newly proposed self-regularized non-monotonic neural (MISH) activation function [24] to further improve the performance of our MV-RAN method.

2. Method

MV-RAN was designed as an end-to-end framework and was comprised of three major components, including the feature extraction module, the segmentation branch, and the classification branch (Fig. 2). First, we designed a pyramid dilated dense convolution (PDDConv) for the feature extraction module. Second, our segmentation branch was developed using novel hierarchical convolutional with LSTM recurrent units (hConvLSTM) inspired by Shi et al. [37]. Finally, we proposed a series of aggregated downsample and fully connected layers for the classification branch. The details of these three major components of our framework are provided as follows.

2.1. Multi-level and multi-scale features extraction

In order to extract multi-level and multi-scale features, we designed a pyramid architecture in our feature extraction module, which consisted of five PDDConv layers. The rationale is that the multi-level feature can provide the global LV geometric characteristic, and multi-scale feature can help to strengthen thin and small areas, further refine the boundaries for an accurate delineation of the LV. In doing so, the discrepancies across different views, scanners and centres can be lessened, and our architecture can be assumed to be more robust to adapt various image qualities and anatomical structure variations.

Each PDDConv was comprised of L densely connected dilated convolution layers, as depicted in Fig. 3, which can enlarge the receptive field and maintain the resolution of feature maps at the same time, while the transition layer changes channels and resolution of feature maps by convolution and pooling operations.

For traditional CNNs, the output feature map of the l^{th} layer $x_l = H_l(x_{l-1})$, H is a non-linear transformation of the previous layer and is usually defined as a convolution followed by batch normalisation (BN) and rectified linear unit (ReLU). However, with the increasing depth of traditional CNNs, the gradient vanishing problem may happen. To alleviate the gradient vanishing, ResNet utilised shortcut connections and the output $x_l = H_l(x_{l-1}) + x_{l-1}$. In addition, DenseNet introduced a different connectivity pattern where any layer is directly connected to all subsequent layers, the output of the l^{th} layer $x_l = H_l([x_1, x_2, \ldots, x_{l-1}])$. This dense connection pattern allows feature reuse and can strengthen gradient information propagation.

To take full advantage of multi-level and multi-scale features as well as expand the receptive field without losing the resolution of feature maps, we introduce the dilated convolution to replace the plain convolution in H. Thus, in our model, the new non-linear transformation $\widetilde{\mathcal{H}}(\cdot)$ is a composite function of three joint operations including BN, ReLU and dilated convolution. The feedforward information propagation from previous l layers to $(l+1)^{th}$ layer can be formulated as

$$x_l = \widetilde{\mathcal{H}}(\mathfrak{C}(x_1, x_2, \ldots, x_{l-1})) \quad , \tag{1}$$

where x_l are the output of the l^{th} layer, $\mathfrak{C}(\cdot)$ refers to the concatenation of the outputs of previous layers. Five PDDConv layers generate multi-level and multi-scale features $f_t = \{f_{t,1}, f_{t,2}, f_{t,3}, f_{t,4}, f_{t,5}\}$ for the frame t in the sequence as shown in Fig. 3.

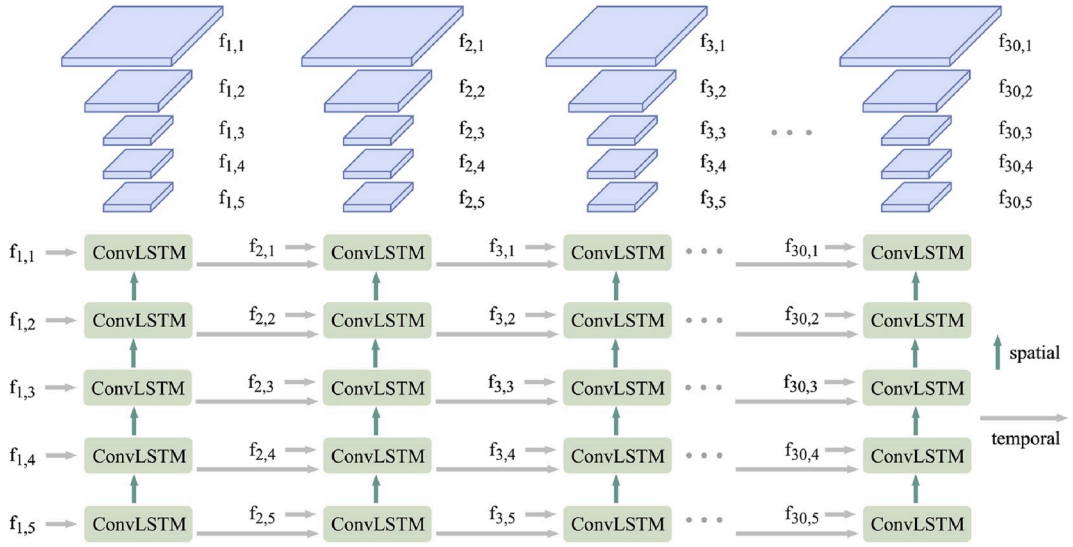

Fig. 4. Schema of the proposed hConvLSTM architecture.

Pyramidal PDDConv layers endowed MV-RAN with the superior feature extraction ability and the LV region detection capacity in multi-level and multi-scale space, further contribute to capturing the global geometric characteristic of the LV and then establishing uniform semantic features. Therefore, the proposed MV-RAN can detect and extract the LV accurately and robustly from not only ED and ES frames but also from other frames in the sequence even when the boundary may not be clear due to various image qualities, low signal to noise ratio (SNR) and confounding tissues from other organs.

2.2. Recurrent features fusion for spatial-temporal modelling

For the segmentation of the 2D + t echocardiographic sequences, capturing the LV characteristic over time is essential for temporal stability of the automatic delineation. Recent studies on LSTM have shown great potential to model the temporal dependencies among data. The correlated temporal sequence features are mainly learned by the convolutional long-short term memory (ConvLSTM), which is a special recursive neural network architecture that can be defined mathematically as

$$f_t = \sigma\left(W_{xf} * x_t + W_{hf} * h_{t-1} + W_{cf} \circ c_{t-1} + b_f\right) , \tag{2}$$

$$i_t = \sigma\left(W_{xi} * x_t + W_{hi} * h_{t-1} + W_{ci} \circ c_{t-1} + b_i\right) , \tag{3}$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \mathrm{MISH}\left(W_{xc} * x_t + W_{hc} * h_{t-1} + b_c\right) , \tag{4}$$

$$o_t = \sigma\left(W_{xo} * x_t + W_{ho} * h_{t-1} + W_{cfo} \circ c_t + b_o\right) , \tag{5}$$

$$h_t = o_t \circ \mathrm{MISH}(c_t) , \tag{6}$$

where * represents convolutional operator and ∘ denotes the Hadamard product, W terms denote weight matrices, b terms denote bias vectors, σ represents a sigmoid function and MISH [24] is used in our study instead of tanh. The ConvLSTM uses three gates including the input gate i_t, the forget gate f_t and the output gate o_t. The memory cell c_t represents an accumulator of the state information, and h_t denotes the hidden states.

Inspired by Shi et al. [37], we constructed a novel hierarchical convolutional with LSTM recurrent units (hConvLSTM) architecture to exploit long term spatial-temporal modelling as depicted in Fig. 4. We added recurrence in the temporal domain to generate prediction S_t for frame t in the sequence, which carried forward the LV information from previous frames to the following frames and realised the matching between consecutive frames intuitively. In addition, we also added recurrence in the spatial domain for multi-level and multi-scale features fusion, which could help to integrate multi-level and multi-scale features more efficiently.

The output $y_{t,k}$ of the k^{th} ConvLSTM at frame t depends on the following variables:

(1) k^{th} level and scale feature $f_{t,k}$ from the feature extraction module;
(2) the output $y_{t,k-1}$ of preceding $(k-1)^{th}$ ConvLSTM at the same frame t;
(3) the output $y_{t-1,k}$ from the k^{th} ConvLSTM of the previous frame $(t-1)$;
(4) the hidden state representation $h_{t,k-1}$ from preceding $(k-1)^{th}$ ConvLSTM at the same frame t, which is the spatial hidden state;
(5) the hidden state representation $h_{t-1,k}$ from the k^{th} ConvLSTM of the previous frame $(t-1)$, which is the temporal hidden state.

The information flow can be formulated as

$$x_{\mathrm{input}} = \left[f_{t,k} \middle| \mathfrak{B}\left(y_{t,k-1}\right) \middle| y_{t-1,k}\right] , \tag{7}$$

$$h_{\mathrm{state}} = \left[h_{t,k-1} \middle| h_{t-1,k}\right] , \tag{8}$$

$$y_{t,k} = \mathrm{ConvLSTM}_k\left(x_{\mathrm{input}}, h_{\mathrm{state}}\right) , \tag{9}$$

where $\mathfrak{B}(\cdot)$ denotes the bilinear upsampling operator. At each time step, the ConvLSTM accepted hidden states h and encoded spatial-temporal features from previous ConvLSTM and frame, the corresponding extracted feature from the feature extraction module, it then output encoded spatial-temporal features to the next ConvLSTM and frame. Finally, predictions S_t are generated by the last ConvLSTM at every frame.

2.3. Double-branch aggregation learning

To further mitigate the discrepancies across MV and refine MV segmentation results, we developed a double-branch aggregation architecture for simultaneous segmentation and classification of MV echocardiographic sequences as depicted in Fig. 2. Features extracted from the last PDDConv were sent to the classification branch. Then, it went through successive convolution and pooling operators to deeply aggregate with multi-level and multi-scale spatial-temporal features from the segmentation branch. Finally, the classification results were produced using fully connected layers. Two branches exerted mutual promotion to cope with the differences across MV data.

In our framework, the double-branch aggregation mechanism can provide additional regularizations. The 3DCNN builds the classification branch while the hierarchical ConvLSTMs build the segmentation branch, they take full advantage of the features extracted from the pyramidal PDDConv. By focusing on different tasks, two branches develop different regularizations and share with each other. The segmentation branch generated multi-view segmentation results while the classification branch discriminated the specific view. They were mutually promoted by deep aggregation of multi-level and multi-scale spatial-temporal features, which brings multi-view discriminative regularization and supervision to refine the segmentation results. The segmentation branch provided multi-level and multi-scale spatial-temporal information to guide the classification while the classification branch, in turn, offered MV discriminative regularization to refine the segmentation results and further mitigated the discrepancies across views. This double-branch aggregation mechanism endowed our MV-RAN with outstanding performance and capability to adapt complex variations of anatomical structures of LV in a multicentre and multi-scanner study.

In addition, we proposed an aggregation loss to dynamically facilitate the communication between the segmentation and classification branches, as illustrated in Fig. 2. The aggregation loss comprised of a segmentation loss and a classification loss. The segmentation loss is a combination of a binary cross-entropy loss and a Dice loss, and the classification loss is a categorical cross-entropy loss. The binary cross-entropy loss regards the pixel-level similarity while the Dice loss regards the overlapping degree. Combination of two losses provides a more comprehensive way to evaluate the segmentation performance. The total aggregated loss function can be formulated as

$$\mathscr{L}_{\text{segmentation}} = -\left[\mathfrak{G}\cdot\log(\mathfrak{P}) + (1-\mathfrak{G})\cdot\log(1-\mathfrak{P})\right] + \frac{2\cdot\mathfrak{G}\cdot\mathfrak{P}}{\mathfrak{G}+\mathfrak{P}}\ , \tag{10}$$

$$\mathscr{L}_{\text{classification}} = -\sum_{i=1}^{3}\mathfrak{g}_i\log(\mathfrak{p}_i)\ , \tag{11}$$

$$\mathscr{L}_{\text{aggregation}} = \lambda_s\cdot\mathscr{L}_{\text{segmentation}} + \lambda_c\cdot\mathscr{L}_{\text{classification}}\ , \tag{12}$$

where $\mathfrak{G}$ and $\mathfrak{P}$ denote the ground truth and the prediction of the segmentation, respectively, and $\mathfrak{g}$ and $\mathfrak{p}$ refer to the ground truth and the prediction of the classification separately, where i represents the type of views. In addition, $\lambda_s$and λ_c denote the corresponding balance coefficients of the segmentation and classification tasks, and both were chosen carefully after a series of cross validation experiments during the training process.

3. Experiments and results

We conducted a comprehensive set of experiments with three goals: (1) evaluate the performance of the proposed MV-RAN method with various ablation studies to validate the robustness of the proposed method; (2) compare with other state-of-the-art deep learning based methods to prove the effectiveness of the proposed method; (3) quantify the derived clinical indexes of the segmented LV to test the reliability of the measured cardiac function using our proposed fully automated segmentation method.

3.1. Datasets and implementation details

In order to validate the effectiveness of our proposed MV-RAN method, we evaluated our methods on two datasets: (1) We obtained a large MV echocardiographic sequences dataset consists of data acquired from three hospitals in China (the Second People's Hospital of Shenzhen, the Third People's Hospital of Shenzhen and Peking University First Hospital) with three different scanners (i.e., Philips EPIQ 7C, GE VIVID E9 and Philips IE33). This retrospective study was approved by our institutional review board in accordance with local ethics procedures.

Table 1
Details of our multicentre MV 2D + t dataset.

Machines	Patients	Sequences	Frames (Images)
Philips EPIQ 7C	60	180 (150 A2C + 15 A3C + 15 A4C)	5400
GE VIVID E9	45	135 (135 A3C)	4050
Philips IE33	45	135 (135 A4C)	4050
Total	150	450	13500

Table 2
Details of the CAMUS dataset.

Machine	Patients	Sequences	Frames (Images)
GE VIVID E95	450	900 (450 A2C + 450 A4C)	1800 (ED and ES only)

Our multicentre dataset consists of 450 sequences (150 sequences for each of the A2C, A3C and A4C view, respectively) acquired from 150 patients. Each sequence contains 30 frames and in total there are 13500 frames entered for our model training and testing. All the MV 2D + t echocardiographic data were manually segmented by two echocardiography specialists. (2) We further tested our MV-RAN method on a recently developed large-scale dataset, i.e., CAMUS dataset, which is publicly available [18]. The CAMUS dataset consists of 2D echocardiographic sequences with two and four-chamber views of 500 patients; 450 patient's data with manually delineated ground truth were used in our study. Moreover, the CAMUS dataset contains only manually labelled ED and ES frames, which were acquired using the same machine (GE VIVID E95) from a single centre study in France (University Hospital of St Etienne). Details of the two datasets have been summarised in Table 1 and Table 2, and sketched in Fig. 5.

For the sake of computational efficiency, all frames were resized to 256 × 256 and pre-processed with zero-centre and normalisation: $I' = \frac{I-\text{mean}(I)}{\text{std}(I)}$, where I represents the pixel values of the image. We employ Adam as the optimizer and the dilated rate of five PDDConvs are 1, 1, 2, 4, 8 respectively. Additionally, we use a dynamical decay mechanism to reduce the learning rate. All experiments were performed on a Linux workstation equipped with two Intel Xeon 2.10 GHz CPU and four 12 GB Nvidia Titan XP GPU using the TensorFlow framework.

3.2. Training, validation and testing settings

Both our MV 2D + t echocardiographic data and CAMUS data have manually delineated ground truth. The labelled datasets were partitioned into disjoint training and cross-validation, and independent testing sets for evaluation as follows.

For the MV 2D + t echocardiographic data, we split the data into training and cross-validation dataset (randomly selected 80% of the data) and independent testing dataset (the rest of 20% of the data). This mainly because we have in total 13500 frames (i.e., 13500 2D images), and 80% of the data can provide us with enough amount of data for training. Compared to our MV 2D + t echocardiographic data, the CAMUS data has only 1800 frames, and therefore we split the data into 400 patients' data for training and cross-validation (~89%), and 50 patients' data for an independent testing (~11%) to ensure that we can have enough data for training (Fig. 5). It is of note that the exact partition of the data into training, validation, and test set should not be crucial for the developed model and should most likely yield very similar results. What is more important is that these two datasets are mutually disjoint, that they are chosen randomly, and what are the approximate sizes of each set [36].

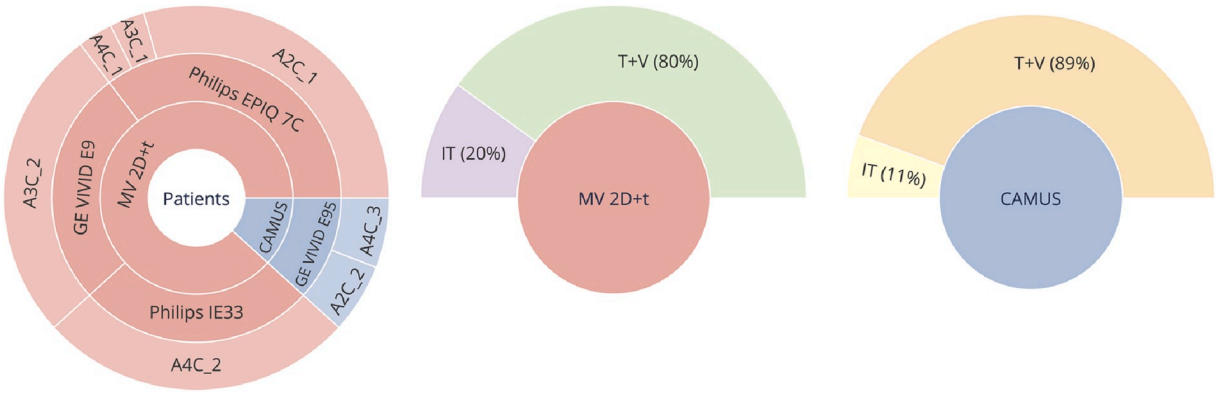

Fig. 5. Datasets and training, cross-validation and testing settings. T + V: Training and cross-validation; IT: Independent Testing.

3.3. Evaluation metrics

We used the following quantification metrics to evaluate our fully automated MV-RAN segmentation method: Hausdorff Distance (HD), Mean Absolute Distance (MAD) and Dice Similarity Coefficient (DSC).

Let $\mathscr{A}$ be a contour of automatic segmentation and $\mathscr{B}$ be the corresponding expert manual delineation. The HD for contours $\mathscr{A}$ and $\mathscr{B}$ is defined by

$$\mathrm{HD}(\mathscr{A},\mathscr{B}) = \max\{\max_{a\in\mathscr{A}}\mathbb{D}(a,\mathscr{B}),\ \max_{b\in\mathscr{B}}\mathbb{D}(b,\mathscr{A})\}\ , \tag{13}$$

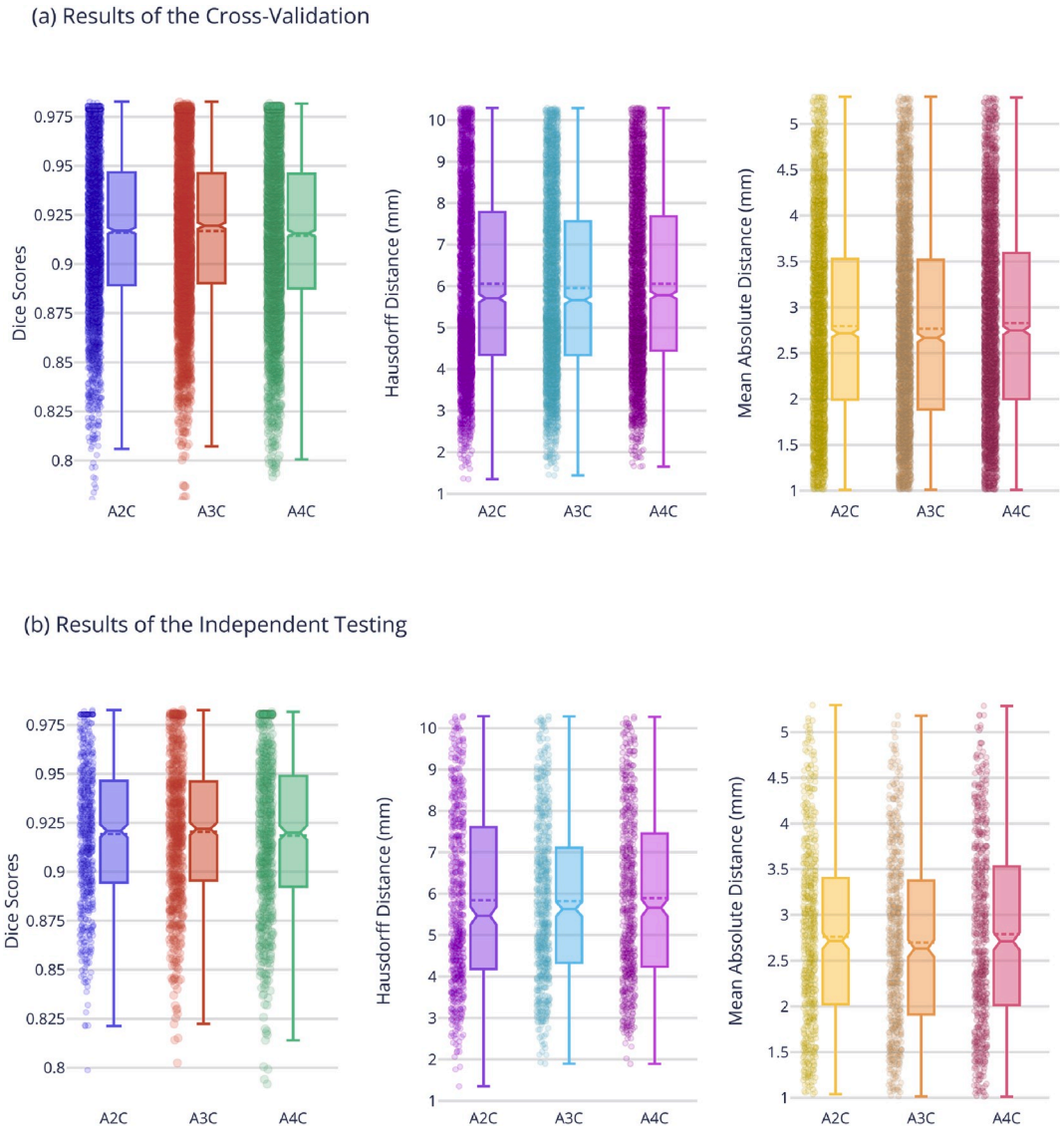

Fig. 6. Boxplots of the segmentation results. (a) results of the cross-validation and (b) results of the independent testing.

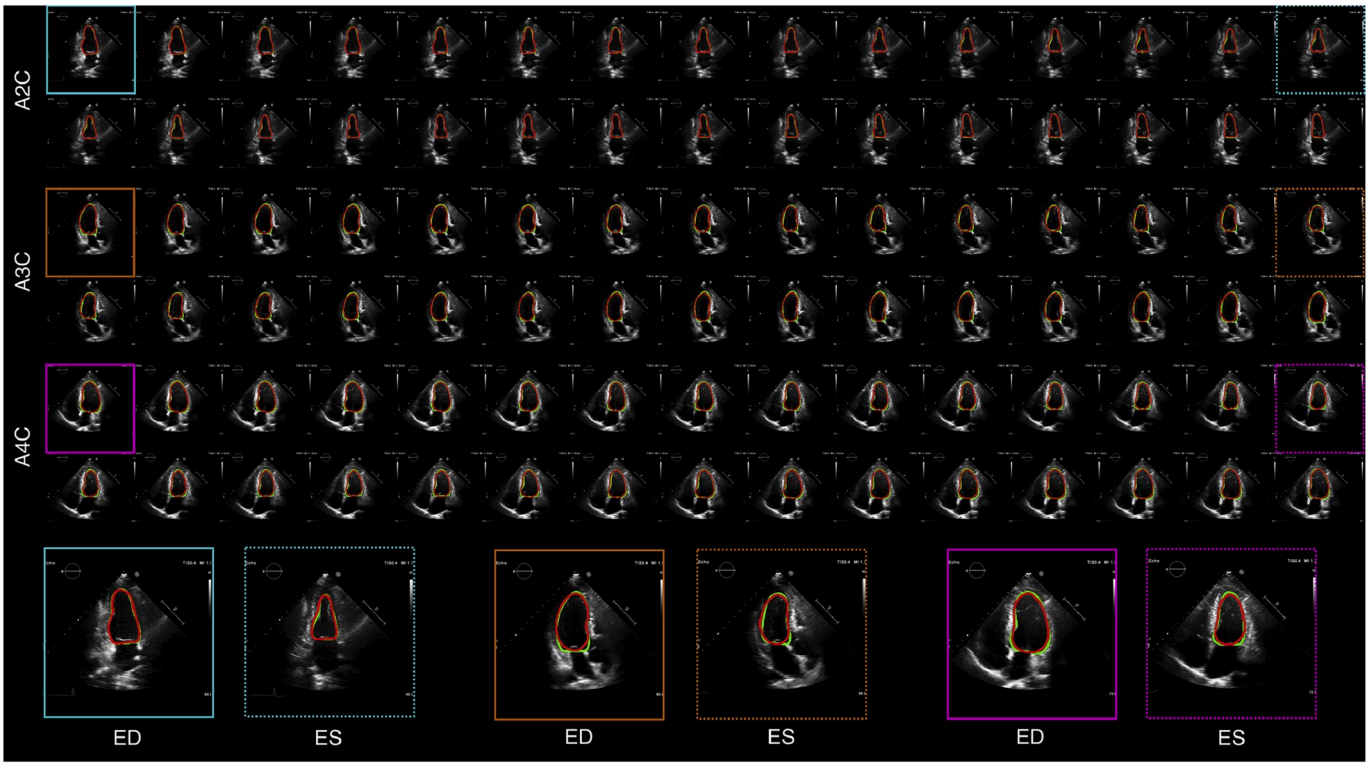

Fig. 7. Segmentation results of example independent testing cases using our MV-RAN method. Zoomed-in ED and ES frames are shown in the last column.

where $\mathbb{D}(a, \mathscr{B}) = \min_{b\in\mathscr{B}}\|b-a\|_2$, $(a\in\mathscr{A}, b\in\mathscr{B})$ are the discrete points on the contours $\mathscr{A}$ and $\mathscr{B}$ respectively, and it evaluates the maximum distance between the two contours. The MAD for contours A and B is given by

$$\mathrm{MAD}(\mathscr{A},\mathscr{B}) = \frac{1}{2}\left\{\frac{1}{\mathrm{N}_{\mathscr{A}}}\sum_{a\in\mathscr{A}}\mathbb{D}(a,\mathscr{B}) + \frac{1}{\mathrm{N}_b}\sum_{b\in\mathscr{B}}\mathbb{D}(b,\mathscr{A})\right\}, \tag{14}$$

in which $\mathrm{N}_{\mathscr{A}}$ and N_b are the numbers of points on the contours $\mathscr{A}$ and $\mathscr{B}$ respectively. The MAD measures the mean distance between the two contours. Let $r_{\mathscr{A}}$ and $r_{\mathscr{B}}$ denote the regions enclosed by the contours A and B. The Dice score can be expressed as

$$\mathrm{DCS}(\mathscr{A},\mathscr{B}) = \frac{2|r_{\mathscr{A}}\cap r_{\mathscr{B}}|}{|r_{\mathscr{A}}| + |r_{\mathscr{B}}|}. \tag{15}$$

DSC can be used to gauge the coincidence degree between our automated segmentation result and the manual delineated ground truth. Its value ranges from 0 to 1, and the closer the DSC to one, the better pixel level classification is obtained by the developed model.

The clinical indices, e.g., end-diastolic volume (EDV), end-systolic volume (ESV) and ejection fraction (EF), were derived from the segmentation results using the CAMUS datasets. Both linear regression and nonparametric Spearman correlation were calculated between the indices derived from the segmentation methods and from the ground truth. For the linear regression, the goodness of fit was estimated via R^2 and the standard deviation of the residuals Sy.x. The larger of the R^2 the closer of the clinical indices we measured from the segmentation results compared to the ground truth. The smaller the residual standard deviation Sy.x, the closer is the fit of the estimate (clinical indices derived from the segmentation methods) to the actual data (clinical indices derived from the ground truth). The Spearman correlation coefficient (r_s) can be used to measure linear or nonlinear monotonic relationships between two variables, and it can take on any value between 0 and 1, where higher value means higher agreement. Spearman correlation coefficients of <0.5, 0.5–0.9, and >0.9 were considered to indicate poor, fair-to-good, and excellent agreement, respectively. Either Wilcoxon matched-pairs signed rank test and unpaired Mann Whitney test was used to verify whether there was a significant difference ($P \geq 0.05$ not significant, $P = 0.01$–0.05 significant and $P < 0.01$ very significant) between the two test scores or estimations. All the statistical analyses were performed using GraphPad Prism version 8.2.1 (GraphPad Software, San Diego, CA).

3.4. Results

3.4.1. Cross-validation vs. independent testing using the MV 2D + t data

For our MV 2D + t Data, we have quantified the HD, MAD and DSC obtained using our MV-RAN method. The data has been split into training, cross-validation and independent testing datasets (Fig. 5), and we compared the results obtained from cross-validation and independent testing.

As shown in Fig. 6, using MV-RAN, we obtained mean DSC of ~0.92 ± 0.04 for all A2C, A3C and A4C views for both cross-validation (A2C vs. A3C: $P = 0.3058$, A2C vs. A4C: $P = 0.2188$ and A3C vs. A4C: $P = 0.0228$ by Wilcoxon matched-pairs signed rank test) and independent testing (A2C vs. A3C: $P = 0.4075$, A2C vs. A4C: $P = 0.9602$ and A3C vs. A4C: $P = 0.5105$ by Wilcoxon matched-pairs signed rank test). Comparing the results of using cross-validation and independent testing, we found that $P = 0.1037$, 0.0640 and 0.0214 (by unpaired Mann Whitney test) for A2C, A3C and A4C data, respectively.

We obtained mean HD of 6.06 ± 2.11 mm, 5.96 ± 2.07 mm and 6.06 ± 2.04 for A2C, A3C and A4C views for the cross-validation (A2C vs. A3C: $P = 0.1032$, A2C vs. A4C: $P = 0.9775$ and A3C vs. A4C: $P = 0.0194$ by Wilcoxon matched-pairs signed rank test) and 5.84 ± 2.06 mm, 5.82 ± 1.92 mm and 5.90 ± 2.02 for A2C, A3C and A4C views for the independent testing (A2C vs. A3C: $P = 0.7867$, A2C vs. A4C: $P = 0.6186$ and A3C vs. A4C: $P = 0.5141$ by Wilcoxon matched-pairs signed rank test). Comparing the results of using cross-validation and independent testing, we found that $P = 0.0260$, 0.2383 and 0.0740 (by unpaired Mann Whitney test) for A2C, A3C and A4C data, respectively.

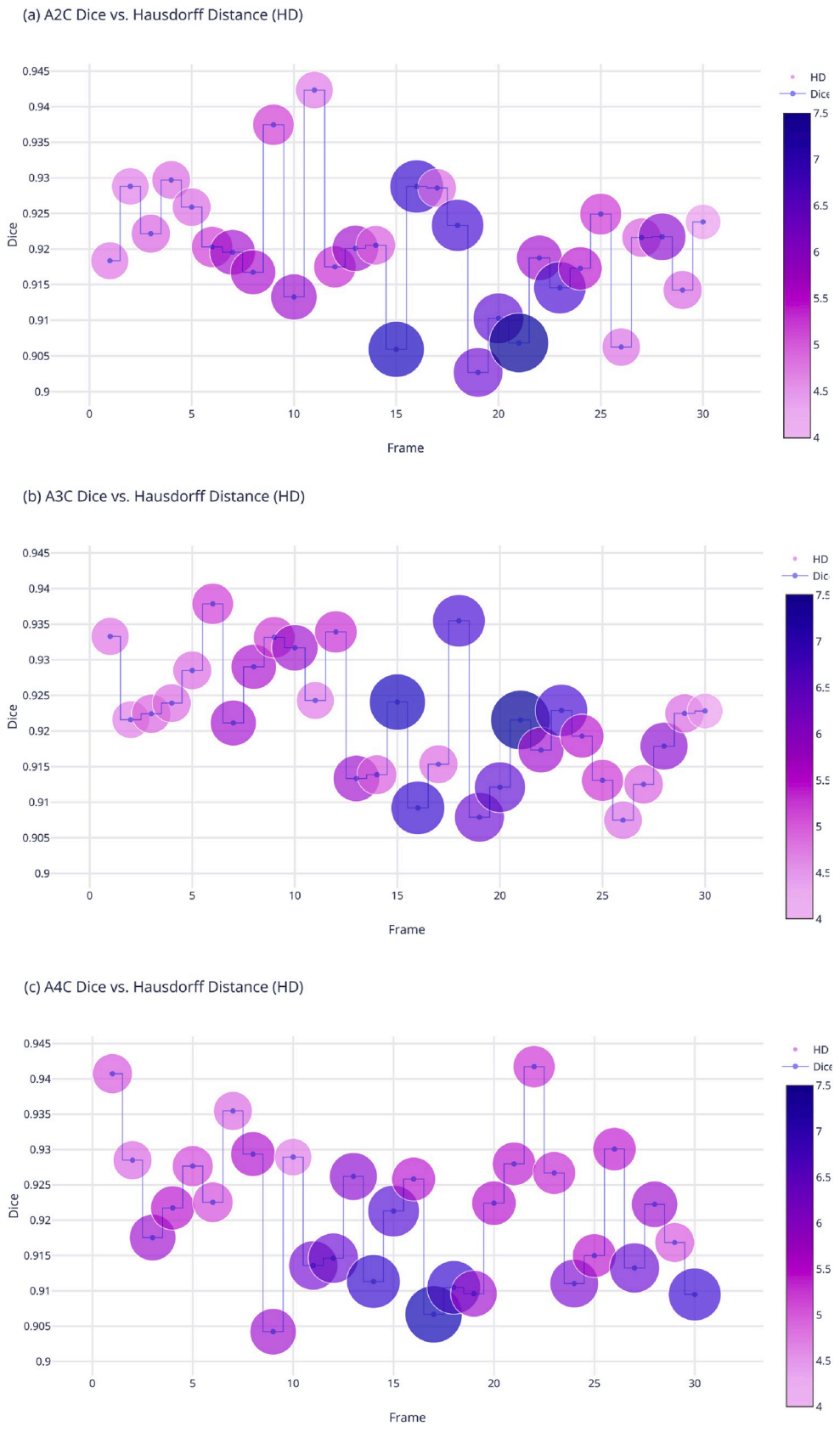

Fig. 8. Per-frame comparison results: Dice vs. Hausdorff Distance (HD) for our independent testing datasets. The line indicates the variate of Dice, the size of the circle means the deviation of HD and the colour of the circle shows the mean value of HD. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article).

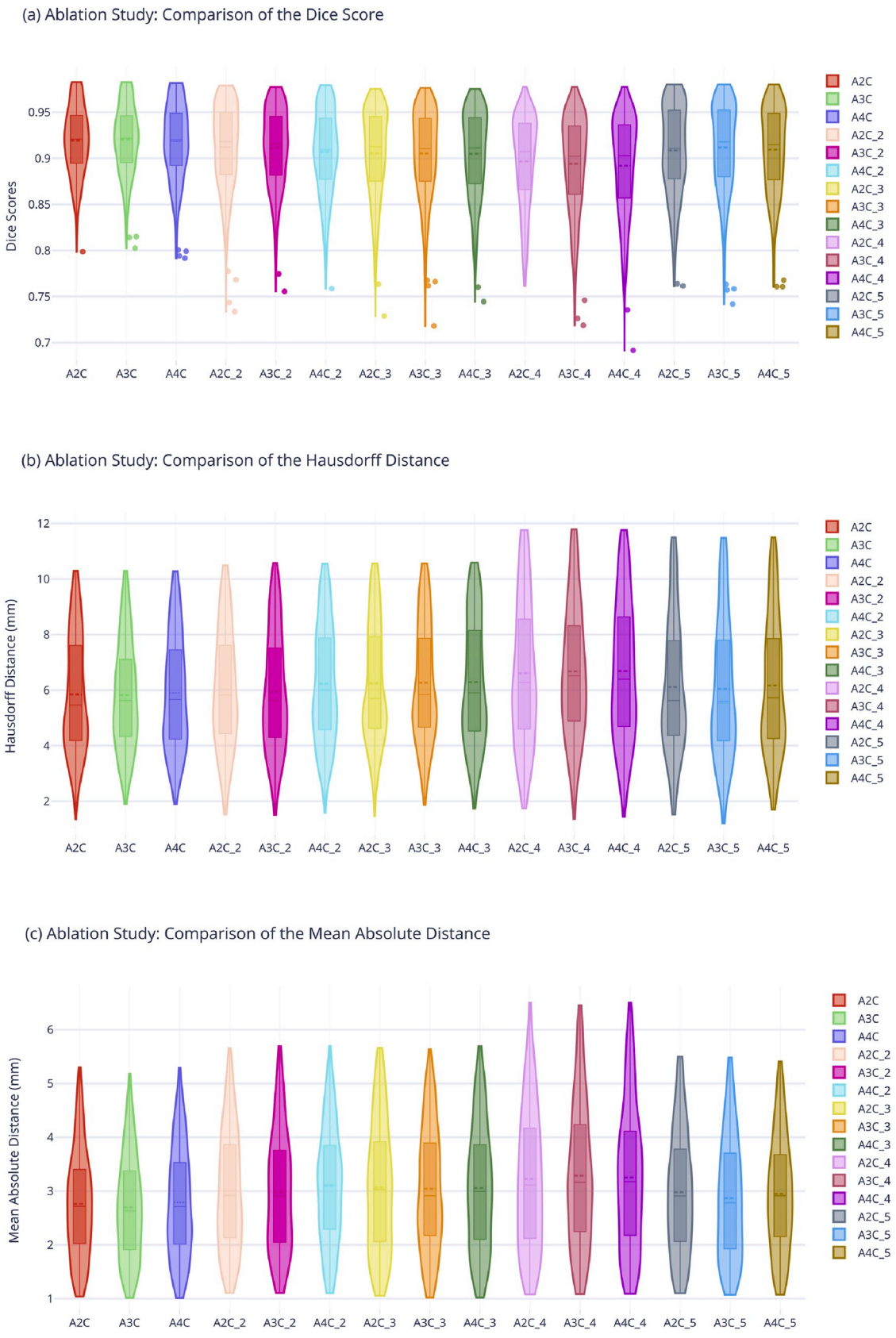

Fig. 9. Ablation studies results and comparison. Results of the full MV-RAN framework (A2C, A3C and A4C); results of the framework without the classification branch (A2C_2, A3C_2 and A4C_2); results of the framework without the PDDConv (A2C_3, A3C_3 and A4C_3); results of the framework without the spatial modelling (A2C_4, A3C_4 and A4C_4); results of the framework without the temporal modelling (A2C_5, A3C_5 and A4C_5).

Fig. 10. Comparison study results. Top: Comparison of the DSC; Middle: Comparison of the HD; Bottom: Comparison of the MAD.

In addition, we obtained mean MAD of 2.80 ± 1.02 mm, 2.77 ± 1.05 mm and 2.83 ± 1.04 for A2C, A3C and A4C views for the cross-validation (A2C vs. A3C: P = 0.2272, A2C vs. A4C: P = 0.2488 and A3C vs. A4C: P = 0.0109 by Wilcoxon matched-pairs signed rank test) and 2.76 ± 0.98 mm, 2.70 ± 0.95 mm and 2.79 ± 1.01 for A2C, A3C and A4C views for the independent testing (A2C vs. A3C: P = 0.6631, A2C vs. A4C: P = 0.2616 and A3C vs. A4C: P = 0.1022 by Wilcoxon matched-pairs signed rank test). Comparing the results of using cross-validation and independent testing, we found that P = 0.5780, 0.2916 and 0.4739 (by unpaired Mann Whitney test) for A2C, A3C and A4C data, respectively.

3.4.2. *Per-frame comparison using the MV 2D + t data*

Fig. 7 shows the per-frame qualitative segmentation results using our MV-RAN method. In general, the segmentation results are accurate throughout the whole cardiac cycle. In addition, the segmentation results are consistent for A2C, A3C and A4C views.

We further analysed the per-frame DSC vs. HD, as shown in Fig. 8. For all three views, we obtained over 0.9 mean DSC for ES frames; however, the mean HD we obtained is relatively large, e.g., >7 mm. For both A2C and A3C view (Fig. 8(a) and (b)), we achieved high mean DSC (>0.92) and low mean HD (~4 mm) at ED frames (frames 1 and 30). For A4C view, we achieved high mean DSC (>0.94) at frame 1 with low mean HD; however, for frame 30, the mean DSC was lower but still ~0.91 with relatively higher mean HD (Fig. 8 (c)).

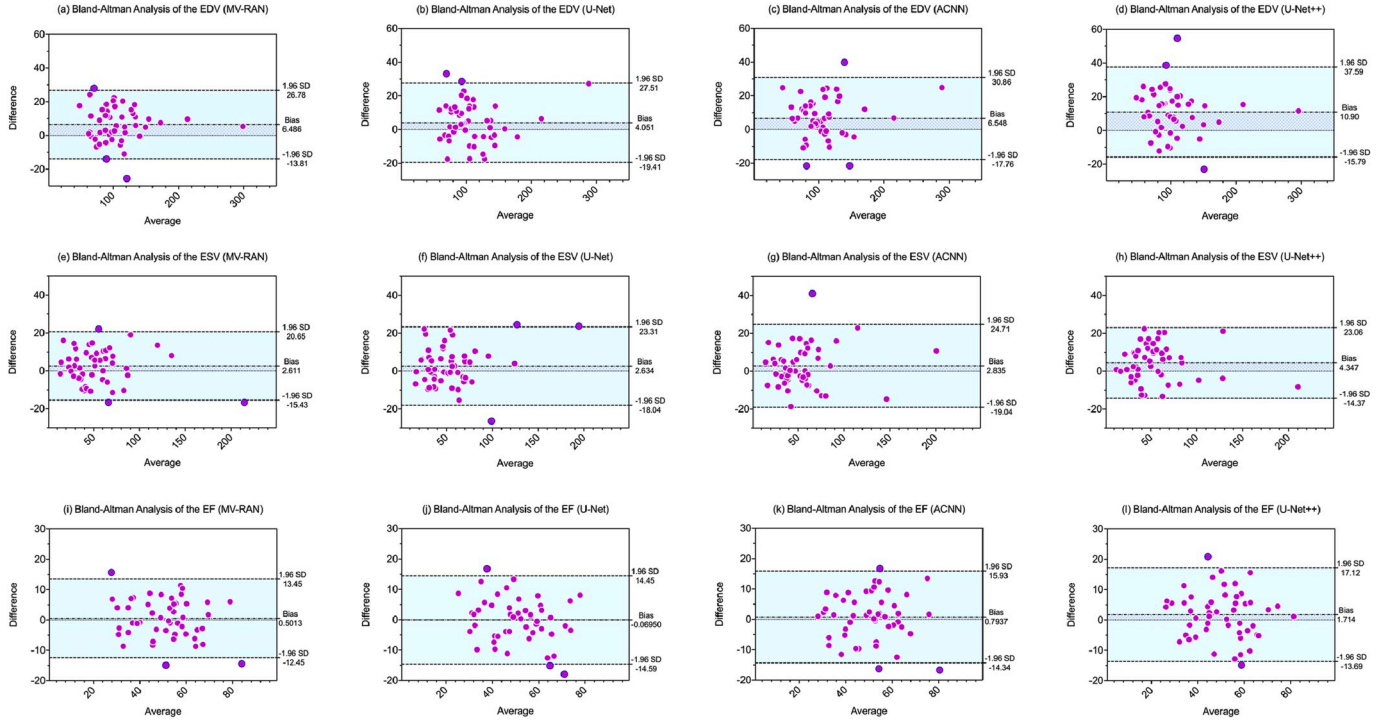

Fig. 11. Bland-Altman analysis results.

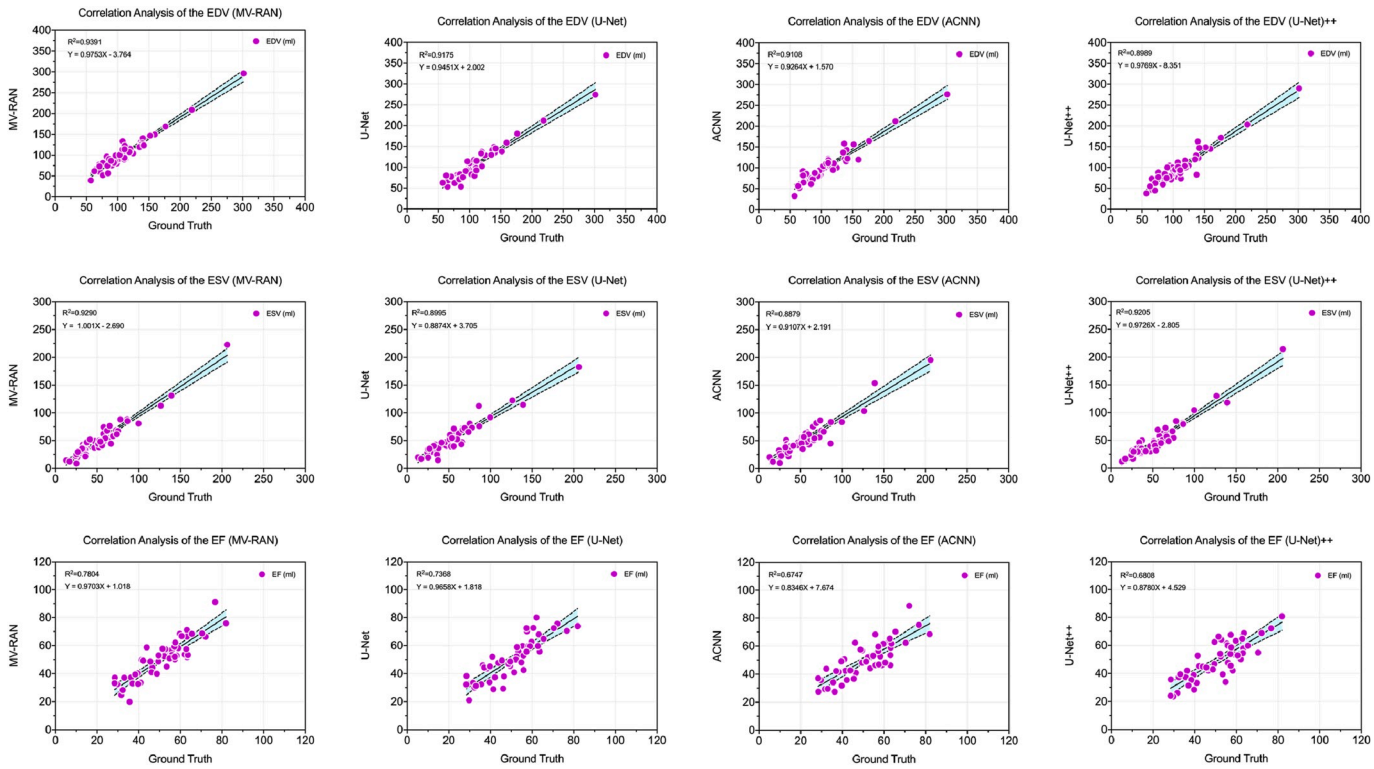

Fig. 12. Linear regression based correlation analysis results.

MV-RAN achieves relative stable mean values of DSC and HD in the cardiac cycle and there only exists moderate fluctuation in the middle of the sequence. The coherence of consecutive frames in the sequence is good. This observation indicates that MV-RAN can obtain efficient spatial-temporal modelling.

3.4.3. *Ablation study results using the MV 2D + t data*

In order to test the effectiveness of each component in the proposed framework, we performed comprehensive ablation studies. We evaluated our MV-RAN method using different configurations to corroborate the necessity of each component in the proposed MV-RAN. We tested our framework without the classification branch, PDDConv, spatial modelling or temporal modelling, respectively. Fig. 9 shows the ablation study results, and we found that our full MV-RAN framework has achieved higher mean DSC, lower mean HD and MAD with fewer variations across all metrics compared against other configurations. The full MV-RAN framework has also achieved the best classification accuracy (0.933). Every single component has brought in important improvement for the LV segmentation, especially when adding recurrence in the temporal domain.

3.4.4. *Comparison results*

We compared our MV-RAN with recently proposed and well-validated U-Net [35], ACNN [27] and U-Net++ [45] based methods on our MV 2D + t data. As shown in Fig. 10, our MV-RAN framework outperformed other methods for all the metrics, achieving the highest mean DSC (~0.92), the lowest mean HD (<5.9 mm) and MAD (<2.8 mm), and significantly lower standard deviations across all the metrics. These findings provided compelling evidence that our MV-RAN is able to accomplish the best region coverage, the highest contour accuracy, and the minimum distance error when processing MV echocardiographic sequences across multicentre and multi-scanner datasets.

We further validated the performance of various segmentation methods based on the clinical indices derived from the delineation results of the CAMUS datasets. First, we calculated the EDV (ml), ESV (ml) and EF (%) from the segmentation results. Second, we performed a Bland-Altman analysis. From the results in Fig. 11, we can observe that compared to ACNN and U-Net++, our MV-RAN framework has always achieved less bias and narrower (better) limits of agreement. Compared to our MV-RAN framework, U-Net has achieved lower bias for the EDV and EF estimation; however, our MV-RAN framework has still got narrower (better) limits of agreement. In addition, correlation analyses using linear regression (Fig. 12) and Spearman correlation (Table 3) have also provided clear evidence that our proposed MV-RAN framework achieved promising results compared to the clinical indices derived from the segmentation ground truth. Compared to other methods, our MV-RAN framework has obtained the highest R^2 and the lowest Sy.x. For the Spearman correlation, bias analysis and mean absolute error, our MV-RAN framework has achieved similar results compared to the best performed method (Table 3).

Table 3
Comparison of the clinical indices results using the CAMUS datasets. EDV: end-diastolic volume (ml); ESV: end-systolic volume (ml); EF: ejection fraction (%); r_s: Spearman correlation coefficient; Sy.x: standard deviation of the residuals; MAE: mean absolute error.

Methods	EDV				ESV				EF			
	r_s	Sy.x	Bias	MAE	r_s	Sy.x	Bias	MAE	r_s	Sy.x	Bias	MAE
MV-RAN	0.923	**10.41**	−7.5 ± 11.0	**8.8**	**0.921**	**9.30**	−3.8 ± 9.2	7.1	0.873	**6.66**	−0.9 ± 6.8	**5.0**
U-Net	0.919	11.87	−6.9 ± 11.8	9.8	0.902	9.97	**−3.7 ± 9.0**	**6.8**	**0.878**	7.74	−1.0 ± 7.1	5.3
ACNN	**0.924**	12.15	**−6.7 ± 12.9**	10.8	0.874	10.87	−4.0 ± 10.8	8.3	0.796	7.50	**−0.8 ± 7.5**	5.7
U-Net++	0.836	13.73	−11.4 ± 12.9	13.2	0.884	9.60	−5.7 ± 10.7	8.6	0.807	7.78	−1.8 ± 7.7	5.6

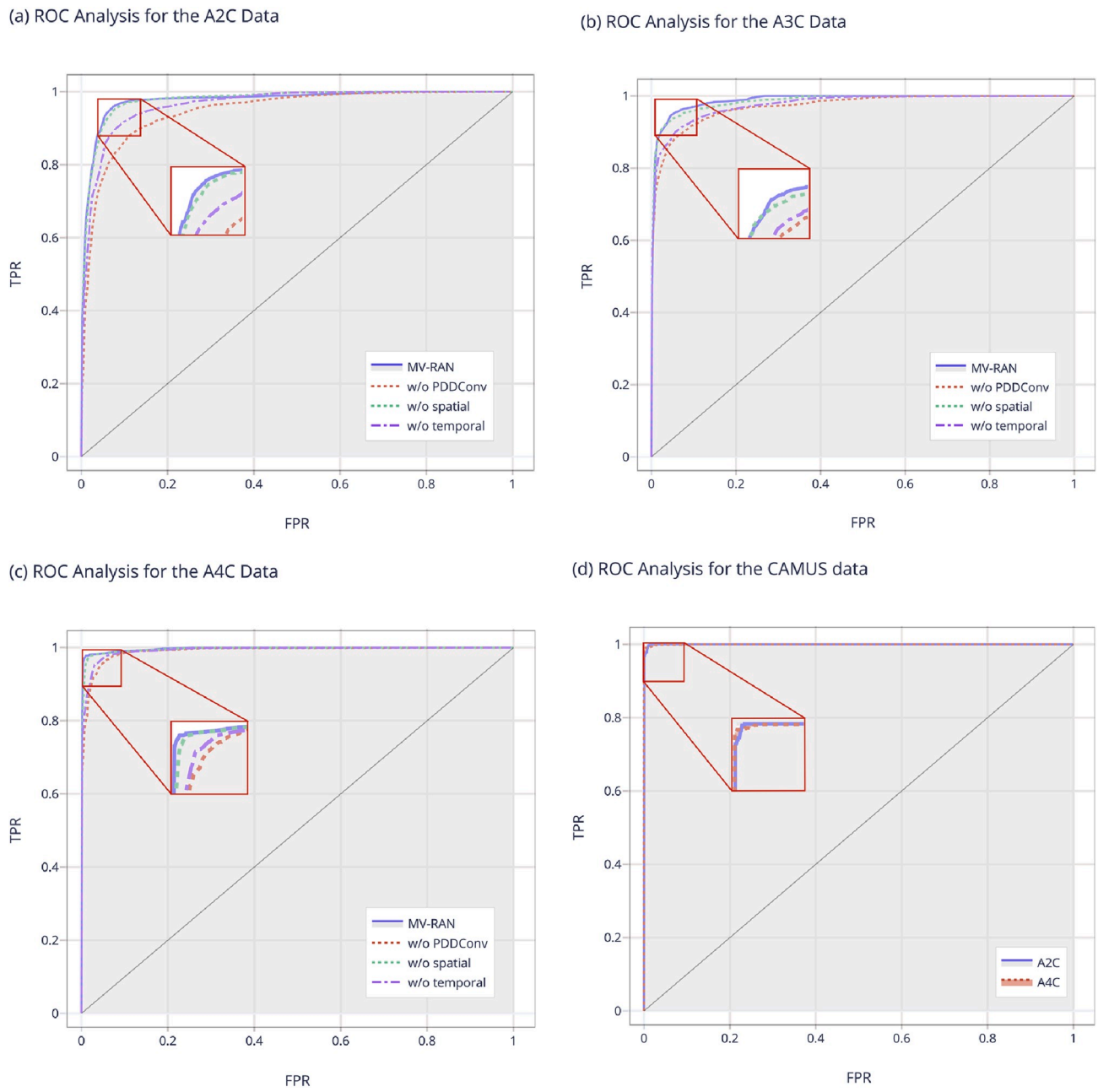

Fig. 13. Comparison of the receiver operating characteristic (ROC) analysis curves.

Table 4
Comparison of the area under the ROC curves (AUC).

AUC	MV 2D + t Data			CAMUS Data	
	A2C	A3C	A4C	A2C	A4C
MV-RAN	**0.9802**	**0.9895**	**0.9974**	0.9997	0.9997
w/o PDDConv	0.9479	0.9721	0.9908		
w/o spatial	0.9780	0.9862	0.9963		
w/o temporal	0.9650	0.9776	0.9930		

3.4.5. *Classification results*

Receiver operating characteristic (ROC) analysis and area under curves (AUC) results are shown in Fig. 13 and Table 4. We can observe that MV-RAN has performed better compared to other model variations with AUC values > 0.98 using MV 2D + t data. The ROC analysis also shows that our MV-RAN can provide accurate classification results on CAMUS data. Combine with the results in 3.3.4 Ablation Study, we can see that the classification branch takes advantage of the multi-level and multi-scale spatial-temporal information from the segmentation branch.

3.4.6. *Learning curves analysis*

Fig. 14 shows the learning curves recorded from training and cross-validation using different methods. We can observe slight overfitting of all the tested methods; however, our MV-RAN method has achieved less loss compared to other segmentation methods.

4. Discussion

In this study, a multiview recurrent aggregation network has been proposed for achieving a simultaneous segmentation and classification for MV echocardiographic sequences data. The proposed method is also capable of full cardiac cycle analysis. To the best of our knowledge, the proposed MV-RAN is the first fully automatic method for the classify and

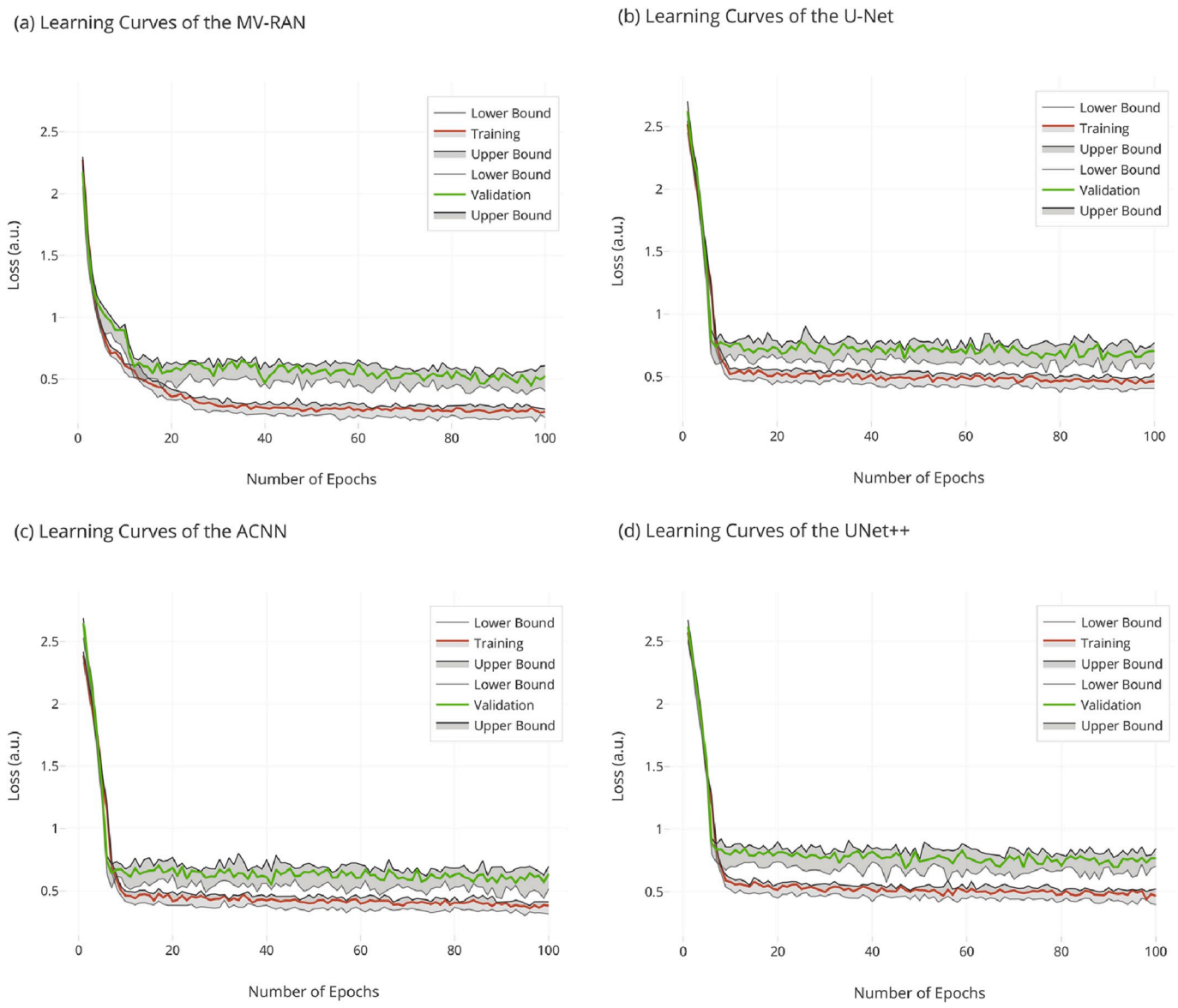

Fig. 14. Comparison of the learning curves during training and cross-validation.

segment MV echocardiographic sequences data in one single framework. The proposed framework is based on a hybrid approach consisting of pyramid dilated dense convolution for accurate spatial feature extraction, hierarchical convolutional with LSTM recurrent units for temporal information retrieval, and a series of aggregated downsample and fully connected layers for the benefits of a simultaneous classification of multiview data with the more accurate LV delineation. The experiments and results lead to the following observations.

By incorporating spatial-temporal information learned from our propose framework, we obtained much better segmentation accuracy (mean DSC >0.92) compared to previous methods that could only obtain mean DSC <0.9. Further validation using HD and MAD also have shown that our MV-RAN framework has produced superior results compared to recently proposed deep learning based methods (mean HD < 6 mm using MV-RAN vs. ~7.5 mm–~9.5 mm using other methods). More importantly, the proposed MV-RAN framework performed consistently over cross-validation and independent testing datasets (mean DSC of $\sim 0.92 \pm 0.04$ vs. $\sim 0.92 \pm 0.04$, $P > 0.01$ for A2C and $P > 0.05$ for A3C and A4C data using unpaired Mann Whitney test). For different views, our MV-RAN framework also produced consistent and accurate accuracy that the mean DSC scores are all about 0.92 ± 0.04 with $P > 0.05$ when comparing A2C vs. A3C, A2C vs. A4C and A3C vs. A4C using the independent testing dataset. Similar findings have been obtained using HD and MAD. While accurate segmentation results have been obtained by our MV-RAN, the classification of MV data is also precise (Fig. 13 and Table 4). The ROC analysis and AUC results have demonstrated that adding pyramid dilated dense convolution and incorporating both spatial and temporal information can improve the classification accuracy with superior automated delineation obtained simultaneously. In addition, qualitative and quantitative visualisation also further confirmed the validity and stabilization of our segmentation over different views and multiple frames (Figs. 7 and 8).

Further validation has been done by calculating and comparing the clinical indices from the segmentation results. For all the estimated EDV, ESV and EF, our MV-RAN framework has produced accurate quantification compared to the results obtained by using the ground truth (MAE = 8.8 ml and 7.1 ml for the EDV and ESV and 5% for the EF). Correlation analysis via linear regression shows that we obtained $R^2 > 0.93$ for both EDV and ESV and $R^2 = 0.78$ for the EF, which are accurate and is capable of a fully automated whole cardiac cycle analysis. Spearman correlation also confirmed that the clinical indices produced by our MV-RAN framework have high correlation compared to the ones derived from

the ground truth (r_s=0.92, 0.92 and 0.87 for EDV, ESV and EF, respectively). Comparison studies have also shown that our MV-RAN framework has outperformed other state-of-the-art methods by multiple quantification metrics.

The current study still has some limitations. One limitation is that we have only one set of manually segmented ground truth (delineated by two echocardiography specialists) for our MV echocardiographic sequences dataset. Having multiple specialists to delineate the dataset separately can provide a more comprehensive intra-observer validation. Another limitation is that the sequential process of our LSTM brings errors forward causing the accumulation of temporal errors in the sequence which can be observed in Fig. 8 that the segmentation results show at the beginning of the frame tends to perform better. Nevertheless, the fluctuating of DSC and HD is moderate, and the overall segmentation performance is satisfactory. This limitation could be alleviated via further development to incorporate the bi-direction LSTM. One more limitation is that MV-RAN is a supervised learning-based method, it still relies on a lot of annotated data. However, building a big annotated dataset requires laborious manual interpretation.

5. Conclusions

An MV-RAN, i.e., multiview recurrent aggregation network, has been proposed for simultaneous segmentation and classification of MV echocardiographic sequences that can also enable a full cardiac cycle analysis. Our MV-RAN has been trained and evaluated on a private dataset containing 13500 2D echocardiographic images collected from 150 patients that each patient has been scanned in 3 views, including 30 frames covering the whole cardiac cycle. The MV-RAN has achieved 0.92 ± 0.04 Dice scores with Hausdorff distance of 5.87 ± 3.46 mm and an MV classification accuracy of 0.93. In addition, our MV-RAN has obtained superior correlation scores for the derived clinical indices compared to other state-of-the-art DL based methods using the open access CAMUS dataset. All these results have shown compelling evidence that our MV-RAN method can be an efficient and accurate clinical tool to segment LV from MV echocardiographic sequences data collected from multiple centres using different scanners and provide timely interpretations for full cardiac cycle analysis.

In the future work, we will try to overcome the current limitations and extend the application to other acquisition settings in a large-scale multicentre and multi-observer study to validate the reproducibility and comparing both the inter- and intra-observer performance. We will also develop an integrated multimodal learning to exploit the knowledge distillation, which transfers what has been learnt by a large-scale model to a smaller-scale model using soft-label supervision, that may result in improvement of the overall segmentation performance and decrease of the reliance on large amounts of annotated data.

Acknowledgements

This work is supported in part by the Fundamental Research Funds for the Central Universities, Key R&D Program of Guangdong Province (2019B010110001), Key Program for International S&T Cooperation Projects of Guangdong Province (2018A050506031), and the National Natural Science Foundation of China (61771464, U1801265).

References

[1] D. Barbosa, D. Friboulet, J. D'hooge, O. Bernard, Fast tracking of the left ventricle using global anatomical affine optical flow and local recursive block matching, in: Proceedings MICCAI Challenge on Echocardiographic Three-Dimensional Ultrasound Segmentation (CETUS), MIDAS Journal, Boston, 2014, pp. 17–24.

[2] O. Bernard, J.G. Bosch, B. Heyde, M. Alessandrini, D. Barbosa, S. Camarasu-Pop, F. Cervenansky, S. Valette, O. Mirea, M. Bernier, P.-M. Jodoin, J.S. Domingos, R. V. Stebbing, K. Keraudren, O. Oktay, J. Caballero, W. Shi, D. Rueckert, F. Milletari, S.-A. Ahmadi, E. Smistad, F. Lindseth, M. van Stralen, C. Wang, O. Smedby, E. Donal, M. Monaghan, A. Papachristidis, M.L. Geleijnse, E. Galli, J. D'hooge, Standardized evaluation system for left ventricular segmentation algorithms in 3D echocardiography, IEEE Trans. Med. Imag. 35 (2016) 967–977, https://doi.org/10.1109/TMI.2015.2503890.

[3] M. Bernier, P. Jodoin, A. Lalande, Automatized evaluation of the left ventricular ejection fraction from echocardiographic images using graph cut, in: Proceedings MICCAI Challenge on Echocardiographic Three-Dimensional Ultrasound Segmentation (CETUS), MIDAS Journal, Boston, 2014, pp. 25–32.

[4] J.G. Bosch, S.C. Mitchell, B.P.F. Lelieveldt, F. Nijland, O. Kamp, M. Sonka, J.H. C. Reiber, Automatic segmentation of echocardiographic sequences by active appearance motion models, IEEE Trans. Med. Imag. 21 (2002) 1374–1383.

[5] G. Carneiro, J.C. Nascimento, A. Freitas, The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods, IEEE Trans. Image Process. 21 (2011) 968–982.

[6] V. Chalana, D.T. Linker, D.R. Haynor, Y. Kim, A multiple active contour model for cardiac boundary detection on echocardiographic sequences, IEEE Trans. Med. Imag. 15 (1996) 290–298.

[7] H. Chen, Y. Zheng, J.-H. Park, P.-A. Heng, S.K. Zhou, Iterative multi-domain regularized deep learning for anatomical structure detection and segmentation from ultrasound images, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2016, pp. 487–495.

[8] J. Domingos, R. Stebbing, J. Noble, Endocardial segmentation using structured random forests in 3D echocardiography, in: Proceedings MICCAI Challenge on Echocardiographic Three-Dimensional Ultrasound Segmentation (CETUS), MIDAS Journal, Boston, 2014, pp. 33–40.

[9] I. Goodfellow, J. Pouget-Abadie, M. Mirza, Generative Adversarial Networks, 2014 arXiv Prepr. arXiv … 1–9.

[10] L. Guo, Y. Hu, B. Lei, J. Du, M. Mao, Z. Jin, B. Xia, T. Wang, Dual network generative adversarial networks for pediatric echocardiography segmentation, in: International Workshop on Preterm, Perinatal and Paediatric Image Analysis, 2019, pp. 113–122, https://doi.org/10.1007/978-3-030-32875-7_13.

[11] Y. Hu, L. Guo, B. Lei, M. Mao, Z. Jin, A. Elazab, B. Xia, T. Wang, Fully automatic pediatric echocardiography segmentation using deep convolutional networks based on BiSeNet, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2019, pp. 6561–6564, https://doi.org/10.1109/EMBC.2019.8856457.

[12] X. Huang, D.P. Dione, C.B. Compas, X. Papademetris, B.A. Lin, A. Bregasi, A. J. Sinusas, L.H. Staib, J.S. Duncan, Contour tracking in echocardiographic sequences via sparse representation and dictionary learning, Med. Image Anal. 18 (2014) 253–271, https://doi.org/10.1016/j.media.2013.10.012.

[13] X. Huang, B.A. Lin, C.B. Compas, A.J. Sinusas, L.H. Staib, J.S. Duncan, Segmentation of left ventricles from echocardiographic sequences via sparse appearance representation, in: 2012 IEEE Workshop on Mathematical Methods in Biomedical Image Analysis, 2012, pp. 305–312.

[14] M.H. Jafari, Z. Liao, H. Girgis, M. Pesteie, R. Rohling, K. Gin, T. Tsang, P. Abolmaesumi, Echocardiography segmentation by quality translation using anatomically constrained CycleGAN, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2019, pp. 655–663, https://doi.org/10.1007/978-3-030-32254-0_73.

[15] K. Keraudren, O. Oktay, W. Shi, J.V. Hajnal, D. Rueckert, Endocardial 3d ultrasound segmentation using autocontext random forests, in: Proceedings of the MICCAI Challenge on Echocardiographic Three-Dimensional Ultrasound Segmentation (CETUS), 2014, pp. 41–48.

[16] R.M. Lang, L.P. Badano, V. Mor-Avi, J. Afilalo, A. Armstrong, L. Ernande, F. A. Flachskampf, E. Foster, S.A. Goldstein, T. Kuznetsova, P. Lancellotti, D. Muraru, M.H. Picard, E.R. Rietzschel, L. Rudski, K.T. Spencer, W. Tsang, J.-U. Voigt, Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging, J. Am. Soc. Echocardiogr. 28 (2015) 1–39, https://doi.org/10.1016/j.echo.2014.10.003, e14.

[17] S. Leclerc, T. Grenier, F. Espinosa, O. Bernard, A fully automatic and multi-structural segmentation of the left ventricle and the myocardium on highly heterogeneous 2D echocardiographic data, in: 2017 IEEE International Ultrasonics Symposium (IUS), 2017, pp. 1–4.

[18] S. Leclerc, E. Smistad, J. Pedrosa, A. Østvik, F. Cervenansky, F. Espinosa, T. Espeland, E.A.R. Berg, P.-M. Jodoin, T. Grenier, others, Deep learning for segmentation using an open large-scale dataset in 2d echocardiography, IEEE Trans. Med. Imag. 38 (9) (2019) 2198–2210, 9.

[19] K.Y.E. Leung, J.G. Bosch, Automated border detection in three-dimensional echocardiography: principles and promises, Eur. J. Echocardiogr. 11 (2010) 97–108, https://doi.org/10.1093/ejechocard/jeq005.

[20] M. Li, W. Zhang, G. Yang, C. Wang, H. Zhang, H. Liu, W. Zheng, S. Li, Recurrent aggregation learning for multi-view echocardiographic sequences segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2019, pp. 678–686, https://doi.org/10.1007/978-3-030-32245-8_75.

[21] A. Madani, Ramy Arnaout, M. Mofrad, Rima Arnaout, Fast and accurate view classification of echocardiograms using deep learning, npj Digit. Med. 1 (2018) 6, https://doi.org/10.1038/s41746-017-0013-1.

[22] S. Mazaheri, R. Wirza, P.S. Sulaiman, M.Z. Dimon, F. Khalid, R.M. Tayebi, Segmentation methods of echocardiography images for left ventricle boundary detection, J. Comput. Sci. 11 (2015) 957–970.

[23] F. Milletari, M. Yigitsoy, N. Navab, Left ventricle segmentation in cardiac ultrasound using hough-forests with implicit shape and appearance priors, in: Proceedings MICCAI Challenge on Echocardiographic Three-Dimensional Ultrasound Segmentation (CETUS), MIDAS Journal, Boston, 2014, pp. 49–56.

[24] D. Misra, Mish: A Self Regularized Non-monotonic Neural Activation Function, 2019 arXiv Prepr. arXiv1908.08681.

[25] A. Newell, K. Yang, J. Deng, Stacked hourglass networks for human pose estimation, in: European Conference on Computer Vision, 2016, pp. 483–499, https://doi.org/10.1007/978-3-319-46484-8_29.
[26] A. Noble, D. Boukerroui, Ultrasound image segmentation: a survey, IEEE Trans. Med. Imag. 25 (2006) 987–1010.
[27] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, S.A. Cook, A. De Marvao, T. Dawes, D.P. O'Regan, Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation, others, IEEE Trans. Med. Imag. 37 (2017) 384–395.
[28] O. Oktay, W. Shi, K. Keraudren, J. Caballero, D. Rueckert, J. Hajnal, Learning shape representations for multi-atlas endocardium segmentation in 3D echo images, in: Proceedings MICCAI Challenge on Echocardiographic Three-Dimensional Ultrasound Segmentation (CETUS), MIDAS Journal, Boston, 2014, pp. 57–64.
[29] C. Otto, Textbook of Clinical Echocardiography, sixth ed., Elsevier, 2018.
[30] N. Ouzir, A. Basarab, H. Liebgott, B. Harbaoui, J.-Y. Tourneret, Motion estimation in echocardiography using sparse representation and dictionary learning, IEEE Trans. Image Process. 27 (2018) 64–77, https://doi.org/10.1109/TIP.2017.2753406.
[31] N. Parajuli, A. Lu, K. Ta, J. Stendahl, N. Boutagy, I. Alkhalil, M. Eberle, G.-S. Jeng, M. Zontak, M. O'Donnell, A.J. Sinusas, J.S. Duncan, Flow network tracking for spatiotemporal and periodic point matching: applied to cardiac motion analysis, Med. Image Anal. 55 (2019) 116–135, https://doi.org/10.1016/j.media.2019.04.007.
[32] J. Pedrosa, S. Queirós, O. Bernard, J. Engvall, T. Edvardsen, E. Nagel, J. D'hooge, Fast and fully automatic left ventricular segmentation and tracking in echocardiography using shape-based B-spline explicit active surfaces, IEEE Trans. Med. Imag. 36 (2017) 2287–2296.
[33] K. Rajpoot, V. Grau, J. Alison Noble, H. Becher, C. Szmigielski, The evaluation of single-view and multi-view fusion 3D echocardiography using image-driven segmentation and tracking, Med. Image Anal. 15 (2011) 514–528, https://doi.org/10.1016/j.media.2011.02.007.
[34] R. Ribes, P. Kuschnir, A. Luna, J.C. Vilanova, J.M. Jimenez-Hoyuela (Eds.), Learning Cardiac Imaging, Learning Imaging, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, https://doi.org/10.1007/978-3-540-79084-6.
[35] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention, 2015, pp. 234–241, https://doi.org/10.1007/978-3-319-24574-4_28.
[36] Y. Shenkman, B. Qutteineh, L. Joskowicz, A. Szeskin, A. Yusef, A. Mayer, I. Eshed, Automatic detection and diagnosis of sacroiliitis in CT scans as incidental findings, Med. Image Anal. 57 (2019) 165–175, https://doi.org/10.1016/j.media.2019.07.007.
[37] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W. Wong, W. Woo, in: Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting, NIPS, 2015.
[38] E. Smistad, F. Lindseth, Real-time tracking of the left ventricle in 3D ultrasound using Kalman filter and mean value coordinates, in: Proceedings MICCAI Challenge on Echocardiographic Three-Dimensional Ultrasound Segmentation (CETUS), MIDAS Journal, Boston, 2014, pp. 65–72.
[39] E. Smistad, A. Østvik, others, 2D left ventricle segmentation using deep learning, in: 2017 IEEE International Ultrasonics Symposium (IUS), 2017, pp. 1–4.
[40] S.D. Solomon, J. Wu, L.D. Gillam, Essential Echocardiography: A Companion to Braunwald's Heart Disease, first ed., Elsevier, 2018 https://doi.org/10.1016/C2014-0-01316-0.
[41] M. Van Stralen, A. Haak, K.E. Leung, G. van Burken, J.G. Bosch, Segmentation of multi-center 3d left ventricular echocardiograms by active appearance models, in: Proceedings MICCAI Challenge on Echocardiographic Three-Dimensional Ultrasound Segmentation (CETUS), MIDAS Journal, Boston, 2014, pp. 73–80.
[42] G. Veni, M. Moradi, H. Bulu, G. Narayan, T. Syeda-Mahmood, Echocardiography segmentation based on a shape-guided deformable model driven by a fully convolutional network prior, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE, 2018, pp. 898–902, https://doi.org/10.1109/ISBI.2018.8363716.
[43] C. Wang, Ö. Smedby, Model-based left ventricle segmentation in 3D ultrasound using phase image, in: Proceedings MICCAI Challenge on Echocardiographic Three-Dimensional Ultrasound Segmentation (CETUS), MIDAS Journal, Boston, 2014, pp. 81–88.
[44] L. Yu, Y. Guo, Y. Wang, J. Yu, P. Chen, Segmentation of fetal left ventricle in echocardiographic sequences based on dynamic convolutional neural networks, IEEE Trans. Biomed. Eng. 64 (2016) 1886–1895.
[45] Z. Zhou, M.M. Rahman Siddiquee, N. Tajbakhsh, J. Liang, UNet++: a nested U-net architecture for medical image segmentation, in: International Workshop on Deep Learning in Medical Image Analysis, 2018, pp. 3–11, https://doi.org/10.1007/978-3-030-00889-5_1.