## RESEARCH



# Hybrid method for automatic initialization and segmentation of ventricular on largescale cardiovascular magnetic resonance images

Check for updates

Ning Pan<sup>1,2,3†</sup>, Zhi Li<sup>1,2,3†</sup>, Cailu Xu<sup>1,2,3</sup>, Junfeng Gao<sup>1,2,3</sup> and Huaifei Hu<sup>1,2,3\*</sup>

### Abstract

**Background** Cardiovascular diseases are the number one cause of death globally, making cardiac magnetic resonance image segmentation a popular research topic. Existing schemas relying on manual user interaction or semi-automatic segmentation are infeasible when dealing thousands of cardiac MRI studies. Thus, we proposed a full automatic and robust algorithm for large-scale cardiac MRI segmentation by combining the advantages of deep learning localization and 3D-ASM restriction.

**Material and methods** The proposed method comprises several key techniques: 1) a hybrid network integrating CNNs and Transformer as a encoder with the EFG (Edge feature guidance) module (named as CTr-HNs) to localize the target regions of the cardiac on MRI images, 2) initial shape acquisition by alignment of coarse segmentation contours to the initial surface model of 3D-ASM, 3) refinement of the initial shape to cover all slices of MRI in the short axis by complex transformation. The datasets used are from the UK BioBank and the CAP (Cardiac Atlas Project). In cardiac coarse segmentation experiments on MR images, Dice coefficients (Dice), mean contour distances (MCD), and mean Hausdorff distances (HD95) are used to evaluate segmentation performance. In SPASM experiments, Point-to-surface (P2S) distances, Dice score are compared between automatic results and ground truth.

**Results** The CTr-HNs from our proposed method achieves Dice coefficients (Dice), mean contour distances (MCD), and mean Hausdorff distances (HD95) of 0.95, 0.10 and 1.54 for the LV segmentation respectively, 0.88, 0.13 and 1.94 for the LV myocardium segmentation, and 0.91, 0.24 and 3.25 for the RV segmentation. The overall P2S errors from our proposed schema is 1.45 mm. For endocardium and epicardium, the Dice scores are 0.87 and 0.91 respectively.

**Conclusions** Our experimental results show that the proposed schema can automatically analyze large-scale quantification from population cardiac images with robustness and accuracy.

Keywords Cardiac image analysis, Deep learning, Statistical shape models, Complex transformation

<sup>†</sup>Ning Pan and Zhi Li contributed equally to this work.

\*Correspondence: Huaifei Hu huaifeihu@mail.scuec.edu.cn <sup>1</sup>College of Biomedical Engineering, South-Central Minzu University, Wuhan 430074, China
<sup>2</sup>Hubei Key Laboratory of Medical Information Analysis and Tumor Diagnosis & Treatment, Wuhan 430074, China
<sup>3</sup>Key Laboratory of Cognitive Science, State Ethnic Affairs Commission, Wuhan 430074, China



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

#### Introduction

Exponential growth of cardiac data is due to continuous progress in biomedical devices and technologies, which creates an opportunity for exploring the underlying mechanisms of disease, as well as a challenge for current capabilities to extract objective and quantitative cardiac phenotypes. Being the leading cause of death worldwide [1], cardiovascular diseases are an important societal health concern and burden.

Quantitative analysis of cardiac function requires establishing global or regional parameters of cardiac performance such as: left ventricular End-diastolic Volume (LVEDV) and left ventricular End-systolic Volume (LVEDV) for the blood pool, left ventricular mass (LVM) of the myocardium, left ventricular ejection fraction (LVEF), left ventricular stroke volume (LVSV), and wall thickening or wall thinning. To compute any of these parameters, the left ventricle must be segmented. However, it is tedious and time consuming task for cardiologists, radiologists or technicians to manually, or semi-manually (aided by software) identify and delineate the relevant cardiac structures for further analysis. Inter and intra-observer variability also undermines the validity of the derived parameters. Therefore, methods are desperately needed to accelerate and facilitate the process of image segmentation to support diagnosis, treatment evaluation and patient follow-up. A number of algorithms have been proposed for automatic and semi-automatic cardiac MRI (CMR) segmentation: image-based [2–4], pixel-/voxel-level classification [5], deformable models [6-9], atlas construction [10], and machine learning [11-13]. For a detailed account of previous work we refer the reader to recent topical reviews [14–16].

However, most of the above algorithms cannot meet the needs faced when dealing with large-scale heterogeneous populations. To address these scenarios, a robust technique known as ASM (active shape model) [17] can be employed for visualizing and quantifying both geometric and functional patterns of the heart. This method leverages prior knowledge by encoding the distinct shape and appearance variations present in the images. When the shape models are adopted for segmentation, the 3D-ASM surface needs to be initialized within the capture range of the intended boundaries for a robust and accurate fit. To get the initial shape for a single patient, Catalina et al. adopted a simple mechanism to roughly scale and position the mean shape of the model [18]. Three points are manually selected, two epicardial points at the basal level, and a third one at the apex. Corresponding anatomical landmarks of the mean shape were previously defined by an experienced operator. Using a similarity transformation, the initial shape can be derived after the mean shape is aligned to the landmarks. However, manual initialization becomes infeasible when dealing with thousands of CMR volumes. Xènia et al. proposed a fully automatic method for initializing cardiac MRI segmentation, by using image features and random forests regression to predict an initial position of the heart and key anatomical landmarks in an MRI volume [19]. However, this method relies on the intersection of the two LA (long axis) and the SA (short axis) images. This may result in failure when intersections cannot be obtained from the needed images. In addition, initial shapes relying on landmark detection sometimes cannot cover all slice images, especially at the basal and apex levels.

Currently, deep learning techniques have been widely used by scholars in pattern recognition, computer vision and medical image computing [20-24]. Avendi et al. adopted deep convolutional neural networks (CNNs) to locate the LV, then inferred the LV shape using stacked auto encoders [25]. Excellent agreement with the ground truth was achieved for the endocardial contours using datasets from MICCAI 2009 LV segmentation challenge [3]. Inspired by combining their successful methods with the advantages of a model-based approach, here we investigate the analysis of a large population of images using both CNNs and SSMs. We design a new schema to build the initial shape for 3D-ASM, then apply the 3D-ASM method to introduce high-level knowledge on cardiac anatomy and deal with sparsely distributed multi-view slices in CMR. Our 3D-ASM implementation is aided by knowledge on the rough localization of the myocardial boundaries, produced by CNNs, to produce a stable and robust delineation of both endo and epicardial surfaces. Therefore, we propose a cardiac segmentation framework that, unlike some approaches requiring preprocessing operations on image data such as denoising and enhancement [26], is an end-to-end segmentation method based on CNNs, which streamlines the process by removing traditional preprocessing steps. Then, by combining CNNs with SSMs, this method features better localization and morphological accuracy in LV segmentation.

Our paper is organized as follows. In the next section, we detail our pipeline for cardiac image segmentation. Then data source used in this study is described. In the following results section, extensive comparisons are made to show the properties of our methods. Finally, we discuss the significance of our work.

#### Method

#### Overview

In this section, our work-flow exploits automatic initialization and segmentation of the left ventricle using 3D-ASM. Here the statistical shape model used is SPASM (sparse active shape model) [27]. Our algorithm includes three steps, i.e. Data pre-processing, Initial shape optimization and SPASM modeling & cardiac quantification,



Fig. 1 Pipeline for the proposed cardiac magnetic resonance segmentation



Fig. 2 The architecture of the fully neural network

as depicted in Fig. 1. In the beginning, cardiac MR datasets with ground truth are organized according to the time frames per subject, CTr-HNs (integrated CNNs and Transformer for heart segmentation networks) is applied to train these organized cases. Secondly, the test cases are sent to CTr-HNs to get segmentation. As a result, the masks for endo- and epi-cardial can be derived separately. Consider that CTr-HNs may cause some bad segmentation, the masks from CTr-HNs are refined subsequently. Then the mean Point Distribution Model (PDM) is fit to the endo- and epi-cardial points from CTr-HNs using point sets registration [28] to get an initial shape and the initial shape are refined using complex transformation subsequently. Distance maps are computed from the endo and epicardial walls obtained by CTr-HNs, which are subsequently used to drive the SPASM model towards image boundaries. Thirdly, SPASM is applied to refine the fit of the static shape model to the image data while penalizing large deviations from the ground truth, and the obtained results are employed for cardiac function analysis.

#### Initialization of SPASM

#### Cardiac localization and segmentation

In our task of cardiac segmentation, we adopt a hybrid network integrating CNNs and Transformer [29] as a hybrid encoder for the segmentation of cardiac on MRI images. This architecture is named as CTr-HNs. An overview of the network architecture can be seen in Fig. 2.

Given an image  $x \in \mathbb{R}^{H \times W \times C}$  with spatial resolution of  $H \times W$  and C-channels, the objective is to generate a prediction of the corresponding pixel-level labeled map with the size  $H \times W$ . Initially, the CNNs process MRI image to capture the local features. These features include details of edge, texture, and spatial information, which are progressively generated through convolutional and pooling operations to form multi-scale feature maps. Subsequently, the feature map is partitioned into  $\{f_p^i \in \mathbb{R}^{P^2.\mathcal{C}} | i = 1, \dots, N\}$  by a patch serialization operation, where each patch has a size of  $P \times P$  and the number of image patches is  $N = \frac{HW}{P^2}$ . Each patch is subsequently projected into a D-dimensional embedding space using a trainable linear transformation. Additionally, the spatial position information of each patches is encoded to obtain an embedding sequence of  $f_p^i = [f_p^1, f_p^2, \dots, f_p^N] i = 1, 2, \dots, N]$ , where the sequence dimension is  $f_p \in \mathbb{R}^{\frac{HW}{P^2} \times D}$ . Then this sequence is fed into the 12 Transformer Layers. One-layer Transformer structure consists of Multi-head Self-Attention Mechanism (MSA) and Multi-Layer Perceptron (MLP) blocks (See Fig. 3(a)). The Transformer effectively compensates for the limited receptive field of CNNs, generating features with global dependencies, thereby providing rich contextual information for the subsequent decoder.

After the hybrid encoder, we can obtain the sequence  $Z_L^i = [Z_L^1, Z_L^2, \ldots, Z_L^N | i = 1, 2, \ldots, N]$  with the size of  $Z_L \in \mathbb{R}^{\frac{HW}{P^2} \times D}$ . The sequence hidden features  $Z_L^i$  are fed into the bottleneck layer. To restore the spatial order of the sequence, the encoded features are reshaped from  $\frac{HW}{P^2} \times D$  to  $\frac{H}{P} \times \frac{W}{P} \times D$  to match the

input requirements of the subsequent decoder. In the decoder component, a cascaded structure of up-sampling and convolution operations is employed to progressively recover the resolution. Each level consists of 2 upsampling operations, one  $3 \times 3$  convolutional layer, and one ReLU layer, progressively restoring the feature map from size  $\frac{H}{D} \times \frac{W}{D}$  to the original resolution  $H \times W$ .

Additionally, the feature maps  $X_{in}$  obtained through upsampling are concatenated with the feature maps from the CNNs through the EFG (Edge feature guidance) module [30] along the channel dimension to achieve feature fusion. The skip connections combine high-resolution local features with global contextual information, and the EFG module further enhances edge features  $X_{out}$ , ultimately predicting the segmentation labels. The structure of the EFG module consists of a difference







Fig. 3 Transformer Layer and EFG Module. (a) Transformer Layer; (b) EFG Module

convolution operator and a spatial attention mechanism (See Fig. 3(b)). The difference operation extracts edge information from the image, while the spatial attention mechanism enhances the feature representation of edge regions, guiding the network to better localize and segment the target area, effectively avoiding the issue of blurred boundaries in traditional networks.

During the training process, the loss function of CTr-HNs is the sum of The Cross-Entropy (CE) loss and the Dice loss, as shown as follows:

$$Loss_{total} = Loss_{ce} + Loss_{dice} \tag{1}$$

To balance the CE loss and Dice loss, the final loss function is the weighted sum of CE and Dice loss, as shown in Eq. (2), the weights  $w_1$  and  $w_1$  are learnable parameters and subject to  $w_1 + w_2 = 1$ .

$$Loss_{total} = w_1 Loss_{ce} + w_2 Loss_{dice} \tag{2}$$

All coarse segmentation experiments are run on NVIDIA RTX A5000 GPU with 24GB RAM. CTr-HNs are trained for 300 epochs with a batch size of 6, and the Adam optimizer, with an initial learning rate of  $1e^{-4}$  and the weight decay constant of 3e-5, is used to iteratively update all parameters in the network. During training, the cosine annealing schedule to select the optimal learning rate. Additionally, to improve the robustness of CTr-HNs, in pre-processing, we also performed data augmentation operations on the training dataset, including rotation, translation, horizontal flipping, and vertical flipping.

To optimize initialization for SPASM, a slice-by-slice evaluation of the CTr-HNs segmentation starts from mid-slice and extends to the top-end slice and the bottom-end slice separately. For a slice image, if the CTr-HNs fails to process it, then the CTr-HNs results from neighbor slice are assigned to those of the current slice. Prior information about spatial relationships between slice segmentation is considered in this process, which makes the initialization accurate and robust.

Figure 4 shows the matching process for the initial shape of the SPASM. In Fig. 4(c), the initial shape is derived using a point-set registration algorithm [31]. However, the matching result is not optimal since the initial shape cannot cover all slices, which can be seen in Fig. 4(d). It is necessary to develop a technique to optimize the initial shape for SPASM. This refinement will be detailed in next step.

#### Initial shape refinement

Let's assume a points set P with n points each described by three-dimensional coordinates  $p_i(x_i, y_i, z_i)$  with  $i = 1 \dots n$ . Assume  $\overline{P}(\overline{x} \ \overline{y} \ \overline{z})$  is the center of points set P.

$$\begin{cases} \overline{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_{i} \\ \overline{\mathbf{y}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{y}_{i} \\ \overline{\mathbf{z}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{z}_{i} \end{cases}$$
(3)

Hence, the matrix X is

$$X = \begin{bmatrix} x_1 - \overline{x} & y_1 - \overline{y} & z_1 - \overline{z} \\ \dots & \dots & \dots \\ x_i - \overline{x} & y_i - \overline{y} & z_i - \overline{z} \\ \dots & \dots & \dots \\ x_n - \overline{x} & y_n - \overline{y} & z_n - \overline{z} \end{bmatrix}$$
(4)

Singular value decomposition is applied to X producing a diagonal matrix S, of the same dimension as X and with nonnegative diagonal elements in decreasing order, and unitary matrices U and V so that

$$X = U * S * V' \tag{5}$$

where  $V = (v_1, v_2, v_3)$ , and  $v_3$  is corresponding to the smallest singular value. A fitting plane *P*l passing through the center point  $\overline{P}(\overline{x} \ \overline{y} \ \overline{z})$  can be obtained with unit normal vector  $\overrightarrow{n}$  (See Fig. 5(a)).

$$\begin{cases} \overrightarrow{n} = (\cos\alpha \, \cos\beta \, \cos\gamma) \\ \overrightarrow{n_z} = (0 \ 0 \ 1) \end{cases}$$
(6)

Where  $\cos \alpha$ ,  $\cos \beta$  and  $\cos \gamma$  are directional cosines with x-, y- and z-axes respectively, is Z-axis unit normal vector.

Then the fitting plane Pl is rotated around the center point  $\overline{P}(\overline{x} \ \overline{y} \ \overline{z})$  helped by a complex transformation matrix T to ensure Pl perpendicular to Z-axis (See Fig. 5(b)).

$$T = T_1^{-1} * T_2^{-1}$$
(7)

Where  $\mathrm{T}_1$  and  $\mathrm{T}_2$  are two rotation transformation matrix defined as follows

$$\mathbf{T}_{1} = \begin{bmatrix} 1 & 0 & 0\\ 0 & \sqrt{\cos^{2}\alpha + \cos^{2}\gamma} & -\cos\beta\\ 0 & \cos\beta & \sqrt{\cos^{2}\alpha + \cos^{2}\gamma} \end{bmatrix}$$
(8)

$$T_2 = \begin{bmatrix}
 \frac{\cos \gamma}{\sqrt{\cos^2 \alpha + \cos^2 \gamma}} & 0 & \frac{-\cos \alpha}{\sqrt{\cos^2 \alpha + \cos^2 \gamma}} \\
 0 & 1 & 0 \\
 \frac{\cos \alpha}{\sqrt{\cos^2 \alpha + \cos^2 \gamma}} & 0 & \frac{\cos \gamma}{\sqrt{\cos^2 \alpha + \cos^2 \gamma}}
 \end{bmatrix}$$
(9)

Using the above technique, the endocardial contour points set from CTr-HNs in base slice is fitted and get a plane (See Fig. 6(a) and (b)). Then the fitted plane is



Fig. 4 SPASM initialization. (a) Points from the mean PDM; (b) Points from CTr-HNs outputs; (c) Approximation of PDM to the points derived from CTr-HNs results; (d) Endocardial point sets derived from registered initial shape and CTr-HNs results respectively



Fig. 5 Plane fitting of points set and complex transformation. (a) Plane fitting from 3D points set; (b) Rotating the fitting plane perpendicular to Z-axis around the center point of the 3D points set



Fig. 6 Initial shape refinement procedure. (a) Endocardial contour points from PDM and CTr-HNs at their original position, (b) Plane fitting for points from CTr-HNs, (c) Rotated endocardial contour points from PDM and CTr-HNs, (d) Stretched & aligned initial shape with CTr-HNs points, (e) Refined initial shape and CTr-HNs results in their original position. Points with black circles are adopted for plane fitting and transformation

rotated to be perpendicular to Z-axis (See Fig. 6(c)). Assume  $Z_1$  and  $Z_4$  are the average Z-axis values for the marked points set from PDM in base and apex slices respectively,  $Z_2$  and  $Z_3$  are their counterparts from CTr-HNs. A scale is applied to stretch points from PDM defined as follows:

$$ratio = \frac{Z_2 - Z_4}{Z_1 - Z_3}$$
(10)

The points from PDM is stretched according to the ratio, and then aligned to the points from CTr-HNs (See Fig. 6(d)). A Procrustes analysis [32] is then employed to get a left ventricular model initialization in its original

position (See Fig. 6). Once the CTr-HNs is trained, we can segment the blood pool and myocardium on SA CMR images, and get the initial endo- and epicardial contours. Two distance maps are constructed from the initial endo- and epicardial contours for SPASM segmentation, which were used in our previously published work [6, 7, 33]. The distance maps are helpful to eliminate the long range deviations between the target LV and the trained active shape model.

#### Datasets

In this paper, UK BioBank dataset is adopted to train and test our CTr-HNs network. The UK Biobank encompasses short-axis and long-axis cine Cardiovascular Magnetic Resonance (CMR) images from 50,000 cardiac MRI cases, forming part of a large-scale, prospective, population-based study based in the United Kingdom. This initiative aims to investigate both genetic and nongenetic factors influencing a wide array of diseases. As part of this extensive research effort, CMR examinations are planned for an additional 100,000 participants, building upon the existing cohort of 500,000 middle-aged and older adults who have been recruited for comprehensive health studies.

The CTr-HNs network parameters are learned from short-axis (SA) view CMR images obtained from 700 subjects from the UK Biobank. In training phase, all MR images undergo a series of preprocessing steps, including slicing and the standardization of image dimensions. These images are then resized to a same size of  $256 \times 256$  pixels through a combination of cropping and padding. The primary objective of the CTr-HNs network is to accurately distinguish between four classes: background, LV cavity, RV cavity and myocardium. Each case is accompanied by expert-drawn endocardial and epicardial contours, providing high-quality ground truth annotations essential for supervised learning.

After the CTr-HNs network is trained, more than 1200 cardiac MRI cases from CAP (Cardiac Atlas Project) dataset [34] are used for the SPASM segmentation. CAP is a resource for cardiac image data sharing and atlasbased shape analysis for population studies which can be web-accessible (http://www.cardiacatlas.org). The cases used in our work include two cohorts: asymptomatic volunteers (AV) and patients with myocardial infarction (MI). Manual contours were also provided by the Cardiac Atlas Project. Readers can refer to literature [35] for the detail about the imaging protocols of CAP.

#### Results

## Evaluation of the method (segmentation accuracy measurement)

To validate the efficacy(performance) of the proposed model, we conducted the evaluation in two ways: 1) employing standard metrics for segmentation accuracy, such as the Dice coefficient, mean contour distance (MCD), and Hausdorff distance (HD95), and 2) utilizing clinically relevant measures derived from segmentations, including ventricular volume and mass. The Dice Coefficient serves as a metric to evaluate the overlap between the predicted segmentation and the ground truth. It ranges from 0 to 1, the closer the value is to 1, the higher overlap between the segmentation and ground truth. The mean contour distance quantifies the average distance between the contours derived from automatic segmentation and the ground truth, while the Hausdorff distance measures the maximum distance between the two segmentation contours. A lower the distance metric indicates a higher level of alignment between the two contours of the segmentation and the ground truth [36].

We have further conducted an evaluation of the accuracy of clinical metrics that are obtained from image segmentation. Specifically, we computed the left ventricular end-diastolic volume (LVEDV), end-systolic volume (LVESV), and myocardial mass (LVM) from the automated segmentations. These values were then compared to those derived from manual segmentations. The volumes were determined by summing the voxels corresponding to the relevant label class in the segmentation and multiplying by the volume per-voxel. As for the LV mass, it was calculated by multiplying the volume by a density of 1.05 g/mL [37].

#### The CTr-HNs segmentation

To verify the performance of the proposed model, we compare the results between Bai and our CTr-HNs on a same test set of 600 subjects. Table 1 presents the experimental results of cardiac MRI image segmentation conducted on the UK Biobank dataset. The results demonstrate that our proposed approach achieves Dice coefficients (Dice), mean contour distances (MCD), and mean Hausdorff distances (HD95) of 0.95, 0.10, and 1.54

Table 1 Our proposed Segmentation Scheme on UK BioBank dataset

	Bai (4275 cases for training)			Proposed (700 cases for training)		
	Dice	MCD(mm)	HD95(mm)	Dice	MCD(mm)	HD95(mm)
LV cavity	0.94(0.04)	1.04(0.35)	3.16(0.98)	0.95(0.03)	0.10(0.09)	1.54(0.70)
LV myocardium	0.88(0.03)	1.14(0.40)	3.92(1.37)	0.88(0.03)	0.13(0.14)	1.94(0.86)
RV cavity	0.90(0.05)	1.78(0.70)	7.25(2.70)	0.91(0.03)	0.24(0.26)	3.25(1.84)
Avg	0.906(0.04)	1.32(0.48)	4.78(1.68)	0.913(0.03)	0.16(0.16)	2.24(1.13)

Table 2 Point to surface errors for the clinical cases (mm)

Method		Proposed	Alba's (2018)
Diastolic phase	LV Endo	1.12±0.39	1.64±0.46
	LV Epi	$1.32 \pm 0.39$	$1.77 \pm 0.47$
	LV myocardium	$1.11 \pm 0.27$	1.47±0.31
Systolic phase	LV Endo	$1.54 \pm 0.70$	2.21±0.79
	LV Epi	$1.97 \pm 0.76$	$2.48 \pm 0.33$
	LV myocardium	$1.62 \pm 0.56$	$3.06 \pm 1.02$
	Overall	$1.45 \pm 0.51$	$2.11 \pm 0.56$

for the LV segmentation, respectively; 0.88, 0.13, and 1.94 for the myocardium segmentation; and 0.91, 0.24, and 3.25 for the RV segmentation. For the Dice, our method achieves outperformance compared to the Bai's method in both LV and RV segmentation, while achieving comparable results in myocardium segmentation. Specifically, the Dice for LV cavity and RV cavity are improved by 0.01. For the MCD, all segmentation metrics showed significant improvement. Compared to Bai's method, our method demonstrates a reduction in MCD by 0.94 for LV, 1.54 for RV, and 1.01 for myocardium. For the HD95, our proposed method also showed significant advantages, with achieving reductions of 1.62, 1.98, and 4.00 in the LV, myocardium and RV, respectively.

#### The SPASM segmentation

To evaluate the precision of SPASM, we conducted a comparative analysis of point-to-surface (P2S) distances and the Dice score between the automated segmentation outcomes and the ground truth on CAP dataset.

To show the advantages of the proposed technique, P2S errors are calculated between ground truth and automatic shapes in Table 2. The overall P2S errors is  $1.45 \pm 0.51$  mm for the proposed schema, while they are  $2.11 \pm 0.56$  for SPASM adopted by Alba et al [19].

The cumulative P2S error distribution curves are drawn in Fig. 7 for endocardium, epicardium and myocardium, which represent the cumulative percentages corresponding to the percentage of test cases for which the error is less than a specific value. In our schema, 90% of the P2S error are detected with a 1.9 mm for endocardium and myocardium, 2.4 mm for epicardium; they are greater than 2.8 mm for previous work respectively.

Table 3 shows the results of the clinical cases for two methods. In our proposed method, average Dice scores from endo- and epi-cardial contours are 0.87 and 0.91, respectively.

Figure 8 displays segmentation of one case using different methods with/without refined initial shape. It can be seen that base and apex slices may fail to be segmented for the case without optimized initial shape. However, only adopting initial shape refinement techniques may



Fig. 7 Cumulative point-to-surface error distribution curves for LV myocardium, endo- and epi-cardial for the results from the proposed and Alba's (2018) [13]

		Endo			Ері		
		ED	ES	Average	ED	ES	Average
Dice	Proposed	$0.88 \pm 0.07$	0.85±0.09	$0.87 \pm 0.08$	0.92±0.06	$0.90 \pm 0.06$	$0.91 \pm 0.06$
	Alba's (2018)	$0.82 \pm 0.09$	$0.66 \pm 0.15$	$0.74 \pm 0.12$	$0.79 \pm 0.13$	$0.68 \pm 0.15$	$0.74 \pm 0.14$

 Table 3
 Dice score for the clinical cases, ED (End-diastole), ES (End-systole)

Fig. 8 Short axis slice segmentation of one patient using different strategies (Y: using initial shape refinement; N: without initial shape refinement)

	From experts	Proposed	MADif
LVEDV (ml)	171.67±49.39	166.33±47.81	7.68
LVESV (ml)	$97.92 \pm 45.11$	$95.93 \pm 43.67$	6.28
LVSV (ml)	73.76±28.13	$70.41 \pm 28.73$	7.34
LVM (g)	$148.74 \pm 44.97$	$149.22 \pm 44.23$	8.30
LVEF (%)	44.17±16.83	43.34±17.71	3.22

cause poor segmentation (see the third row images), and, hence, the coarse segment CTr-HNs result is employed to drive the contours to the correct location (see the second row images).

To evaluate cardiac function, clinic parameters are calculated for LVEDV, LVESV, LVSV, LVM and LVEF. In Table 4, it can been seen that the results from ours are close to those of experts.

To demonstrate whether the cardiac functional indexes derived from the ground truth align with those generated by our novel algorithm, we present Bland–Altman plots (displayed in the first row of Fig. 9) and correlation plots (shown in the second row of Fig. 9). These visualizations reveal a strongly match between our results and those from manual delineation. Correlations of cardiac indexes range from 0.89 to 0.99, demonstrating a strong relationship between manual and automatic methods.

#### Discussion

This paper presents a fully automatic approach which can analyze cardiac MRI in large MRI studies. Our schema combines a deep learning neural network, an initial shape refinement algorithm, and a SPASM segmentation method. Different from other approaches, the initial shape derived from CTr-HNs results are rotated and scaled to cover all short slices using complex transformation techniques. Subsequently, the refined initial shape is adopted to obtain a three-dimensional LV segmentation based on a SPASM search.

In CTr-HNs segmentation experiment, we can observe that the standard deviation from ours is notably smaller in Dice metrics, which indicate a stable results and a certain level of performance improvement. Moreover, there is a significantly decrease on the MCD metric, demonstrating that CTr-HNs effectively optimize the segmentation boundaries of various tissues, thereby achieving more precise boundary localization. Additionally, the



Fig. 9 Plots of Bland–Altman and correlation of cardiac functional indexes between the manual and automatic results. In the top row, the mean difference (i.e., bias) and limits of agreement (LoA, i.e., ± 1.96 standard deviations from the mean) are denoted by black horizontal lines and the two red dashed lines respectively; while in the second row, the red lines represent correlation plots

experimental results also reveal a significant improvement in HD95 on RV, further verifying that CTr-HNs can accurately capture the structural boundaries. These outcomes exhibit that our proposed method is capable of leveraging both global contextual information and local boundary features through the hybrid CNNs-Transformer architecture. Furthermore, by incorporating the edge feature guidance (EFG) module, it achieves more precise boundary information localization.

In SPASM, an initial estimate, denoted initial shape, describes the LV position. Considering the similar shapes and edge information between the endo- and epi-cardial contours, if initial shapes are in incorrect LV positions, failures with cardiac image segmentation using SPASM are inevitable. To get the initial shape for SPASM, point-sets registration method is used to align the points of mean shape to the counterparts from CTr-HNs. However, base or apex slice may be missing in the cover of the initial shape, this can be seen in Fig. 7 that poor results are obtained for SPASM when the initial shapes are failed to cover all short slices.

To overcome these difficulties, points from CTr-HNs in the base slice is fitted into a plane, and the fitted plane is rotated to be perpendicular to Z-axis. In the meantime, points from CTr-HNs and initial shape is rotated with the same angle. Note that the rotation is purposely designed, because the initial shape is easily to be scaled and moved in Z-axis direction only.

At last, CTr-HNs segmentation results are used in building distance maps and combined with an image intensity model to drive the initial shapes to the LV position. As a result, a 3D shape which represents an accurate segmentation for the LV is generated.

To confirm it is the same distribution of cardiac functional indexes from manual and automatic methods, Kolmogorov-Smirnov test analysis is adopted for the corresponding clinical parameters. It can be seen in the distribution plots a common distribution, common location and scale, similar distributional shapes.

A limitation of our framework lies in the heavy reliance of our algorithms on model-fitting techniques that utilize 3D active shapes to align cardiac contours across 2D imaging plane stacks. Consequently, the deep learning algorithm employed in this study is geared towards a trainable 2D segmentation model that integrates CNNs and Transformers as an encoder. This approach was chosen because the implemented SPASM method proves effective for increasingly sparse image datasets, encompassing various orientations and originating from different MRI acquisition protocols [27]. The incorporation of an update propagation scheme and a fuzzy inference system enabled application of SPASM to multi-protocol cardiac sparse data sets with a segmentation performance that is better than or comparable to other 3D model-based segmentation methods operating on a full data set with parallel image planes.

#### Conclusion

This study introduces a hybrid schema that can automatically build initial shapes to cover all short slices for SPASM. Deep learning algorithms are employed not only for myocardial detection, but also to drive the shape model to the LV endo- and epi-cardial contours. Results indicate that our method can overcome technical difficulties and obtain robust segmentation for cardiac MRI studies with subvoxel accuracy. Our approach still can be improved in some aspects. For example, the detection of cardiac images with LVOT (left ventricular outflow) and how to use images with LVOT to optimize the initial shape and enhance segmentation using SPASM.

#### Abbreviations

CNNs	Convolutional neuron networks
CTr-HNs	Integrated CNNs and Transformer for heart segmentation
	networks
LVEDV	Left ventricular End-diastolic Volume
LVEDV	Left ventricular End-diastolic Volume
LVM	Left ventricular mass
LVEF	Left ventricular ejection fraction
LVSV	Left ventricular stroke volume
ASM	Active shape model
LA	Long axis
SA	Short axis
CAP	Cardiac Atlas Project
CMR	Cardiac MRI
SPASM	Sparse active shape model
PDM	Point Distribution Model
IIM	Image intensity model
AV	Asymptomatic volunteers
ED	End-diastole
ES	End-systole
MI	Myocardial infarction
P2S	Point-to-surface
MCD	Mean contour distance

#### Acknowledgements

The authors are grateful to all UK Biobank participants and staff. The authors also thank Dr M Pereanez, Dr R Attar, and M Hoz from centre for Computational Imaging and Simulation Technologies in Biomedicine (CISTIB) at the University of Sheffield for this work.

#### Author contributions

Ning Pan performed Methodology and Writing. Zhi Li designed and partially performed the Methodology and Software. Cailu Xu partially performed Writing and Software. Junfeng Gao provided Funding acquisition. Huaifei Hu provided funding acquisition, supervision, and manuscript review. All authors read and approved the final manuscript.

#### Funding

This study is funded by National Natural Science Foundation of China (No. 62076257), Key Research and Development Plan of Hubei Province (No.2022BCA025), Hubei Provincial Key Research Projects (No. 2022BA037), Applied Basic Research Programme of Wuhan (No. 20200601012239), Fund for Scientific Research Platforms of South-Central Minzu University (No. PTZ24008), and Fundamental Research Funds for the Central Universities, South-Central Minzu University (No. CZQ24015).

#### Data availability

This research has been conducted using the UK Biobank Resource under Applications 11350 and the Cardiac Atlas Project (CAP), which can be webaccessible (http://www.cardiacatlas.org).

#### Declarations

#### Ethics approval and consent to participate

This research has been conducted using the UK Biobank Resource under Applications 11350 and the Cardiac Atlas Project (CAP), which can be webaccessible (http://www.cardiacatlas.org).

#### **Consent for publication**

Not applicable

#### Clinical trial number

Not applicable.

#### **Competing interests**

The authors declare no competing interests.

Received: 3 September 2024 / Accepted: 21 April 2025 Published online: 07 May 2025

#### References

- Timmis A, Vardas P, Townsend N, Torbica A, Katus H, De Smedt D, Gale CP, Maggioni AP, Petersen SE, Huculeci R. European society of cardiology: cardiovascular disease statistics 2021. Eur Heart J. 2022;43(8):716–99.
- Liu H, Hu H, Xu X, Song E. Automatic left ventricle segmentation in cardiac MRI using topological stable-state thresholding and region restricted dynamic programming. Acad Radiol. 2012;19(6):723–31.
- 3. Lu Y, Radau P, Connelly K, Dick A, Wright G. Automatic image-driven segmentation of left ventricle in cardiac cine MRI. The MIDAS J. 2009;49:2.
- Hu H, Pan N, Wang J, Yin T, Ye R. Automatic segmentation of left ventricle from cardiac MRI via deep learning and region constrained dynamic programming. Neurocomputing. 2019;347:139–48.
- Hu H, Liu H, Gao Z, Huang L. Hybrid segmentation of left ventricle in cardiac MRI using gaussian-mixture model and region restricted dynamic programming. Magn Reson Imaging. 2013;31(4):575–84.
- Hu H, Pan N, Yin T, Liu H, Du B. Hybrid method for automatic construction of 3D-ASM image intensity models for left ventricle. Neurocomputing. 2020;396:65–75.
- Huaifei H, Pan N, Haihua L, Liman L, Tailang Y, Zhigang T, Frangi AF. Automatic segmentation of left and right ventricles in cardiac MRI using 3D-ASM and deep learning. Signal Process Image Commun. 2021;96:116303.
- Shi X, Li C. Anatomical knowledge based level set segmentation of cardiac ventricles from MRI. Magn Reson Imaging. 2022;86:135–48.
- Khamechian M-B, Saadatmand-Tarzjan M. FoCA: a new framework of coupled geometric active contours for segmentation of 3D cardiac magnetic resonance images. Magn Reson Imaging. 2018;51:51–60.
- Zhang Y, Wei H. Atlas construction of cardiac fiber architecture using a multimodal registration approach. Neurocomputing. 2017;259:219–25.
- Ankenbrand MJ, Shainberg L, Hock M, Lohr D, Schreiber LM. Sensitivity analysis for interpretation of machine learning based segmentation models in cardiac MRI. BMC Med. Imaging. 2021;21(1).
- Diller GP, Vahle J, Radke R, Vidal MLB, Fischer AJ, Bauer UMM, Sarikouch S, Berger F, Beerbaum P, Baumgartner H, et al. Utility of deep learning networks for the generation of artificial cardiac magnetic resonance images in congenital heart disease. BMC Med. Imaging. 2020;20(1).
- Wu J, Gan Z, Guo W, Yang X, Lin A. A fully convolutional network feature descriptor: application to left ventricle motion estimation based on graph matching in short-axis MRI. Neurocomputing. 2019.
- 14. Frangl AF, Rueckert D, Duncan JS. Three-dimensional cardiovascular image analysis. IEEE Trans Med Imaging. 2002;21(9):1005–10.
- Petitjean C, Dacher JN. A review of segmentation methods in short axis cardiac MR images. Med Image Anal. 2011;15(2):169–84.
- Peng P, Lekadir K, Gooya A, Shao L, Petersen SE, Frangi AF. A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging. Magn Reson Mater Phys Biol Med. 2016;29(2):155–95.
- El-Rewaidy H, Fahmy AS, Khalifa AM, Ibrahim E-SH. Multiple two-dimensional active shape model framework for right ventricular segmentation. Magn Reson Imaging. 2022;85:177–85.
- Catalina TG, Constantine B, Santiago A, Federico S, Gloria M, Frangi AF. Automatic construction of 3D-ASM intensity models by simulating image acquisition: application to myocardial gated SPECT studies. IEEE Trans Med Imaging. 2008;27(11):1655–67.
- Alba X, Lekadir K, Pereanez M, Medranogracia P, Young AA, Frangi AF. Automatic initialization and quality control of large-scale cardiac MRI segmentations. Med Image Anal. 2018;43:129–41.
- Ansari MY, Mangalote IAC, Meher PK, Aboumarzouk O, Al-Ansari A, Halabi O, Dakua SP. Advancements in deep learning for b-mode ultrasound segmentation: a comprehensive review. IEEE Trans Emerging Top Comput Intell. 2024;8(3):2126–49.
- Kar J, Cohen MV, McQuiston SP, Malozzi CM. A deep-learning semantic segmentation approach to fully automated MRI-based left-ventricular deformation analysis in cardiotoxicity. Magn Reson Imaging. 2021;78:127–39.
- Wang Z, Xie L, Qi J. Dynamic pixel-wise weighting-based fully convolutional neural networks for left ventricle segmentation in short-axis MRI. Magn Reson Imaging. 2020;66:131–40.
- Lin A, Wu J, Yang X. A data augmentation approach to train fully convolutional networks for left ventricle segmentation. Magn Reson Imaging. 2020;66:152–64.
- Huaifei H, Pan N, Wang J, Yin T, Ye R. Automatic segmentation of left ventricle from cardiac MRI via deep learning and region constrained dynamic programming. Neurocomputing. 2019;347(28):139–48.

- Regaya Y, Amira A, Dakua SP. Development of a cerebral aneurysm segmentation method to prevent sentinel hemorrhage. Network Model Anal Health Inf Bioinf. 2023;12(1):18.
- Assen VHH, Danilouchkine MM, Frangi AA, Ordás SS, Westenberg JJ, Reiber JJ, Lelieveldt BB. SPASM: a 3D-ASM for segmentation of sparse and arbitrarily oriented cardiac MRI data. Med Image Anal. 2006;10(2):286–303.
- Radu H, Florence F, Manuel Y, Guillaume D, Jian Z. Rigid and articulated point registration with expectation conditional maximization. IEEE Trans Pattern Anal Mach Intell. 2011;33(3):587–602.
- Jieneng Chen YL, Qihang Y, Luo X, Adeli E, Wang Y, Lu L, Yuille AL, Zhou Y. TransUNet: transformers make strong encoders for medical image segmentation. arXiv 2021, 2102.04306.
- Zhou S, Wang K-N, Zhou G-Q. Edge-enhanced feature guided joint segmentation of left atrial and scars in LGE MRI images: left atrial and scar quantification and segmentation: 2023//. Cham: Springer Nature Switzerland; 2023. p. 93–105
- Myronenko A, Song X. Point set registration: coherent point drift. IEEE Trans Pattern Anal Mach Intell. 2010;32(12):2262–75.
- Chen Y, Chan AB, Lin Z, Suzuki K, Wang G. Efficient tree-structured SfM by RANSAC generalized Procrustes analysis. Comput Vision Image Understanding. 2017;157:179–89.
- Hu H, Pan N, Frangi AF. Fully automatic initialization and segmentation of left and right ventricles for large-scale cardiac MRI using a deeply supervised network and 3D-ASM. Comput Methods Programs Biomed. 2023;240:107679.
- Fonseca CG, Backhaus M, Bluemke DA, Britten RD, Chung JD, Cowan BR, Dinov ID, Finn JP, Hunter PJ, Kadish AH. The cardiac atlas project—an imaging database for computational modeling and statistical atlases of the heart. Bioinformatics. 2011;27(16):2288–95.
- Zhang X, Cowan BR, Bluemke DA, Finn JP, Fonseca CG, Kadish AH, Lee DC, Lima JA, Suinesiaputra A, Young AA. Atlas-based quantification of cardiac remodeling due to myocardial infarction. PloS One. 2014;9(10):e110243.
- Bai W, Sinclair M, Tarroni G, Oktay O, Rajchl M, Vaillant G, Lee AM, Aung N, Lukaschuk E, Sanghvi MM, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. J Cardiovasc Magn Reson 2018;20:65.
- Grothues F, Smith GC, Moon JCC, Bellenger NG, Collins P, Klein HU, Pennell DJ. Comparison of interstudy reproducibility of cardiovascular magnetic resonance with two-dimensional echocardiography in normal subjects and in patients with heart failure or left ventricular hypertrophy. Am J Cardiol. 2002;90(1):29–34.

#### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ning Pan** is currently in School of Biomedical Engineering at South-Central Minzu University. He worked as a Post-Doctoral Fellow in School of Computer Science and Technology at Huazhong University of Science and Technology (HUST) for two years. He received his MS in Computer Software and Theory from Chongqing University (CQU) and his Ph.D. in Computer Science from HUST, in 2007 and 2013, respectively. His research interests include medical image processing, data mining and deep learning etc.

**Zhi Li** received the Ph.D. in biomedical engineering from Huazhong University of Science & Technology in 2014. He is currently in College of Biomedical Engineering, South-Central Minzu University, Wuhan, Hubei, China. His current research interests include AI in healthcare, computer vision and image processing.

**Cailu Xu** is currently a master candidate in College of Biomedical Engineering at South-Central Minzu University, Wuhan, Hubei, China. Her current research interests include computer vision and medical image processing.

Junfeng Gao received the Ph.D. degree in Biomedical Engineering from the School of Life Science and Technology, Xi'an jiaotong University, Xi'an, China, in 2011. He is currently a Professor with the College of Biomedical Engineering, South-Central Minzu University, Wuhan. He has authored over 50 research articles in domestic and foreign academic journals. His current research interests include EEG signal processing, medical information analysis, pattern recognition and machine learning.

**Huaifei Hu** received the M.S. and Ph.D. in computer science and technology from Huazhong University of Science & Technology in 2007 and 2012, respectively. He is currently in College of Biomedical Engineering, South-Central Minzu University, Wuhan, Hubei, China. He was a visiting Scholar with the University Of Sheffield, UK, from 2015 to 2017. His current research interests include statistical modeling, medical image processing and deep learning.