2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

UAG: Integrating R2UNet and Attention-Guided GNN for Robust Left Ventricle Motion Estimation

Junhao Wu[1][0000-0001-8284-7421], Huanbin Yao[1][0009-0003-4195-9883], Kai Li[1][0009-0004-8939-9851] and Muhammad Sadiq[2][0000-0003-2199-3702]

[1] The Department of Computer Science and Technology, College of Mathematics and Computer Science, Shantou University, Shantou, Guangdong 515063, China
[2] Corresponding author. The Shenzhen Institute of Information Technology, Shenzhen, Guangdong 518172, China. Email: `sadig@sziit.edu.cn` (Muhammad Sadiq).

Abstract. Cardiac diseases significantly affect the structure and function of the left ventricle (LV) during the cardiac cycle.Purpose: Develop a robust framework (UAG) for precise detection and correspondence estimation of aberrant LV myocardial motion, enhancing diagnostic accuracy in cardiac disease management. his paper proposes UAG, an innovative framework for LV motion estimation. The UAG framework integrates a U-shaped network architecture (R2UNet) for precise LV endocardial contour segmentation and a graph neural network (GNN) enhanced with attention mechanisms for robust feature matching. Initially, R2UNet is trained on cardiac magnetic resonance (CMR) images to extract discriminative features representing key points along the LV myocardial boundary. Subsequently, the GNN, combined with the Sinkhorn algorithm, establishes accurate correspondence between landmarks across diverse cardiac phases by leveraging both spatial and semantic feature relationships. Performance evaluation on two publicly available cardiac datasets demonstrates UAG's superiority over state-of-the-art methods. Using matching accuracy (ACC) and average perpendicular distance (APD) as evaluation metrics, UAG achieves the highest ACC and lowest APD values, outperforming existing techniques in both normal and pathological LV contour scenarios. xperimental results validate UAG's exceptional capability in LV motion estimation, particularly for images with abnormal contours. The integration of R2UNet's multi-scale feature extraction and the attention-guided GNN ensures robustness against morphological variations, highlighting its potential for clinical applications in cardiac diagnostics.

Keywords: Left Ventricle, Myocardial Motion, U-shaped Network, Graph Neural Network, Image Segmentation, Endocardial Contour

1 Introduction

Cardiovascular diseases (CVDs) are the leading cause of mortality globally, with a staggering toll of 19.8 million lives lost in 2022 alone [1]. Within the spectrum of CVDs, encompassing conditions like coronary artery disease, hypertension, and heart valve ailments, lies a significant impact on the left ventricle's (LV) structure and function throughout the cardiac cycle. This impact manifests through various phenomena, including morphological changes, anomalies in wall motion, and impaired LV functionality.

Recent research has illuminated the presence of dilatation and hypertrophic alterations in the LV among patients afflicted with coronary artery disease. The cascade initiated by myocardial ischemia inflicts damage upon myocardial cells, prompting subsequent remodeling and fibrosis processes within the myocardium. These transformations culminate in irreversible structural modifications in the LV [3]. In parallel, hypertension fosters persistent myocardial hypertrophy and chamber dilatation, disrupting typical ventricular wall motion and impairing diastolic function [4]. Consequently, individuals with hypertension often exhibit aberrant LV changes, adversely affecting both systolic and diastolic LV function. Similarly, individuals suffering from heart valve disease frequently present with LV dilatation and hypertrophy, potentially accompanied by proliferative changes [5]. These structural adaptations profoundly influence LV systolic and diastolic functions, thereby impacting cardiovascular health.

Moreover, the progression of heart failure manifests as left ventricular failure, hampering efficient blood ejection from circulation. Marked by diminished myocardial contractility and ventricular diastolic dysfunction, left ventricular failure represents a critical facet of heart failure's impact on cardiac function. Analyzing aberrant exercise patterns induced by cardiovascular diseases (CVDs) holds promise in identifying individuals at heightened risk for heart disease. This analysis offers an early warning mechanism, facilitating timely intervention and treatment to mitigate the risk of CVD-related mortality. At present, the acquisition of LV images predominantly involves techniques such as echocardiography, magnetic resonance imaging (MRI), and computed tomography (CT). Of these, MRI stands out as the preferred modality for analyzing LV myocardial motion owing to its inherent advantages. These include artifact-free imaging, elimination of the need for contrast agent injections, non-ionizing radiation, minimal impact on muscular tissues, and superior visualization of soft tissues [5,6]. Therefore, we utilize MRI images for LV motion estimation in this study.

Existing methods for LV motion estimation can be divided into three categories: deformation-model-based methods, image-registration-based methods, and feature tracking. Deformation-model-based methods construct a geometric or topological model of the heart to describe LV motion, and then calculate strain information in different myocardial segments for structure analysis. Image-registration-based methods utilize cardiac images extracted from different times and build a spatial transformation function or deformation map to estimate the LV motion. Feature tracking methods use feature point information to track the displacement of marker points on the LV in the image sequence, analyze the correspondence between marker points, and realize the estimation of LV motion. Compared to the other techniques, methods based on feature tracking eliminate the complex model building and parameterization process and can be flexibly adjusted according to different image types and qualities, etc. Due to the simplicity and flexibility of the feature tracking method, it is often applied in clinical scenarios.

In feature tracking methodologies, pivotal challenges revolve around feature extraction and subsequent matching. Feature extraction necessitates image segmentation to delineate the LV contours from intricate backgrounds, enhancing edge clarity to facilitate the extraction of discernible LV features. Traditional approaches to feature extraction encompass methods such as Scale-Invariant Feature Transform (SIFT) [7], Speeded-Up Robust Features (SURF) [8], and Histogram of Oriented Gradients (HOG) [9]. Atehortúa et al. [10] introduced a spatio-temporal saliency descriptor tailored for representing dynamic motion patterns observed in cardiac cine MRI sequences. This descriptor amalgamates spatial and temporal domain information, quantifying motion characteristics by assessing the saliency of motion. Nonetheless, the homogeneity of myocardial tissue and the sparse internal features within cine MRI [11] present challenges for these methods in establishing correspondence between myocardial contours.

The efficacy of feature matching profoundly influences the ultimate matching outcome. Graph matching algorithms serve as the cornerstone of LV feature matching methods. Wu et al. 12] proposed a left ventricular motion estimation approach leveraging Full Convolutional Network (FCN) feature descriptors for LV myocardial contour segmentation and position coordinate extraction. This approach is complemented by a graph matching algorithm for feature alignment. However, graph matching algorithms necessitate assumptions regarding the LV contour's circular nature. They derive graph edges by introducing auxiliary points, typically situated at the LV center, and connecting each point with the auxiliary point and its adjacent counterparts. Yet, deviations from the circular LV contour, such as those observed in abnormal cases, lead to misalignment between auxiliary points and the true LV center. Consequently, discrepancies arise between the derived graph edges and actual LV features, resulting in suboptimal feature matching outcomes.

The LV motion estimation framework primarily encompasses segmentation and feature matching processes. U-Net [13] has demonstrated efficacy in medical image segmentation. It incorporates skip connections during the up-sampling phase, facilitating the fusion of information from lower and higher levels. This amalgamation enables the acquisition of both local and global information simultaneously, enhancing the network's ability to discern relationships between image regions. Consequently, U-Net processes medical image data more efficiently, leading to improved segmentation accuracy and operational speed. Recursive Residual U-Net (R2UNet) [14] extends the U-Net architecture by introducing recursive residual convolution units. Its effectiveness in segmentation has been verified across various medical imaging domains, including eyeball, skin, and kidney segmentation. However, R2UNet's application in cardiac MRI remains limited.

Graph Neural Networks (GNN) iteratively enrich and update node representations within graph structures, thereby enhancing feature matching accuracy by incorporating global graph structure information. Sarlin et al. [15] introduced Superglue, a feature matching approach leveraging attentional GNNs. By integrating the attention mechanism with GNNs, the model can precisely focus on relevant features for the task at hand, thereby enhancing feature matching accuracy. Nonetheless, the applicability of attentional GNNs in cardiac MRI feature matching remains unexplored.

In this study, we propose a motion matching framework (UAG) for LV feature points. Our framework leverages feature segmentation by R2UNet [14] and feature matching using Attention Graph Neural Networks [15]. Experimental results demonstrate promising matching outcomes.

2 Methods

2.1 Dataset

In this paper, the database of 33 subjects [16] and the MICCAI 2009 challenge database [17] are used for training and evaluating the effectiveness of the proposed method. Details information of the datasets are shown in Table 1. The database of 33 subjects, for example, consisted of short-axis MRI cases from 33 patients. For each case, 8 to 15 image slices were taken from the atrioventricular ring to the apex. Each slice is a 256x256 image. Each case contains an image sequence consisting of exactly 20 frames. For all images, LV endocardial contours delineated by experienced cardiologists were used as the ground truth. Images slices in 8th and 20th phases for all cases are used as the training data for our motion estimation framework, and slices in 1st and 10th phases were used to validate the performance of the proposed method.

Table 1. Details of the databases.

	33 subjects	**MICCAI 2009**
Number of cases	33 cases	15 training cases, 15 test cases, and 15 online cases
Number of slices per case	8–15	6-12
Slice size	256 × 256	256 × 256
Number of phases	20 phases	20 phases
Ground truth	LV contours in all phases of each case	LV contours in ED and ES phases of each case
Training data	Slices in 8th and 20th phases for all cases	The 15 training cases
Validation data	Slices in 1st and 10th phases for all cases	The 15 test cases and the 15 online cases

2.2 Experimental Setting

The proposed framework, UAG, integrates R2UNet for image segmentation and attention GNN for feature matching. Initially, R2UNet is employed to segment the endocardial contour of the left ventricle from cardiac MRI images, with feature extraction conducted during down-sampling. Subsequently, the extracted features are subjected to analysis utilizing attention GNN and the Sinkhorn algorithm to estimate the correspondence of points across different MRI scans.

In our experimental setup, the R2UNet was initially trained to perform LV endocardial contour segmentation. To enhance the model's generalization capabilities, data augmentation techniques, including scaling, vertical flipping, and horizontal flipping, were employed to augment the training dataset. This augmentation strategy aimed to increase the number of training samples, thereby mitigating overfitting and enhancing the overall generalization performance of the R2UNet. Consequently, the training dataset comprised a total of 2088 images.

For assessing matching accuracy, the database comprising 33 subjects provided corresponding points located on the LV endocardial contour for all images. These manually delineated points served as the ground truth for evaluating the effectiveness of the GNN network. Additionally, as the MICCAI 2009 challenge database lacked ground truth correspondence between endocardial contours from end-diastole (ED) to end-systole (ES), a robust point matching algorithm [18] was employed to estimate the transformation function between the two endocardial contours. Subsequently, this transformation function was utilized to map manually outlined points from the source image to the target image, thereby establishing their corresponding points. Following the acquisition of the trained segmentation model, the GNN was subsequently trained on the same training dataset.

2.3 Image Segmentation

Inspired by the deep residual model [19], RCNN [20] and U-Net [13], R2UNet uses a recurrent residual convolutional unit in the network structure of U-Net. The network structure is shown in Fig 1. Each recurrent residual convolutional unit contains two cov.+ReLUs.

The recursive convolution operation employed in the cov.+ReLUs framework utilized in this study entails a single convolutional layer followed by two subsequent recursive convolutional layers. The adoption of Recurrent Convolutional Layers (RCLs), along with RCLs integrated with residual units, in lieu of conventional forward convolutional layers within both encoding and decoding units, facilitates a more efficient development of deep models. Additionally, RCLs integrated within the R2UNet architecture incorporate effective feature accumulation mechanisms. These mechanisms ensure the attainment of superior and more robust feature representations across different time steps, facilitated by feature accumulation within the model. Consequently, the model excels in extracting very low-level features crucial for segmentation tasks across diverse medical imaging modalities.

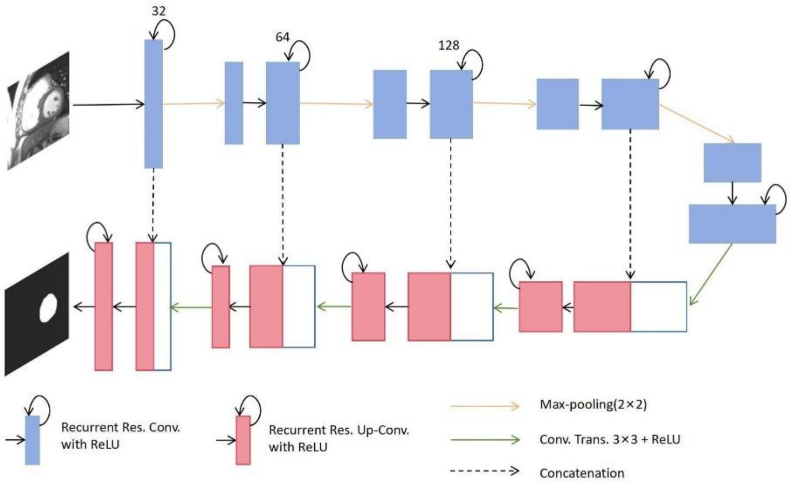
32
64
128
Recurrent Res. Conv. with ReLU
Recurrent Res. Up-Conv. with ReLU
Max-pooling(2×2)
Conv. Trans. 3×3 + ReLU
Concatenation

Fig. 1. The network architecture of R2UNet.

The initial layer of R2UNet comprises three components: convolution, activation, and max pooling. With an input image size of 128×128 pixels, the pooling operation compresses the image dimensions, while the augmentation in filter count per layer enhances feature depth. Consequently, the output entails a halved image size for sampling, with each pixel possessing a feature dimension of 32. This output serves as input for the subsequent layer, where a similar sequence of convolution, activation, and pooling steps further reduces image size while augmenting feature dimension to 128. This down-sampling process is iterated across subsequent layers, resulting in feature extraction at increasingly coarse scales. Notably, as down-sampling layers increase, extracted features become progressively coarser.

Feature extraction from the first three layers of the network is prioritized for identifying feature points. Within the trained R2UNet, features are extracted from points in each layer pre-max-pooling, with the number of features per layer determined by filters. To enhance results, multi-scale features from relevant points are extracted, combining features from the first layer with those from the second and third layers. Given the abundance of features at coarser scales, features from the finest scale of the first layer are also concatenated to yield the final feature set. Considering the feature dimensions extracted from the first, second, and third layers of R2UNet are 32, 64, and 128, respectively, the combined features from these layers yield a 256-dimensional representation of point characteristics within the image.

For image segmentation we use dice loss as our loss:

$$LOSS_1 = 1 - \frac{|X \cap Y|}{|X| + |Y|} \tag{1}$$

The set of pixels in the segmented image in the dataset serves as the ground truth X, and the network model predicts the set of pixels in the segmented image as the prediction value Y, where $|X \cap Y|$ denotes the size of the intersection of X and Y, while $|X|$ and $|Y|$ denote the sizes of the respective sets of pixels.

2.4 Feature Matching

Attentional GNN.

We extracted features of feature points from two different cardiac MRIs, one for end-diastolic(ED) D and one for end-systolic(ES) S, m and n feature points on the left ventricular silhouette according to the method used in this paper, denoted as $D' \coloneqq \{p_1, \ldots, p_m\}$ and $S' \coloneqq \{p_1, \ldots, p_n\}$, respectively. The information of each feature point can be composed of two parts, the position information of feature point i, $p_i \coloneqq (x_i, y_i, z_i)$, and the feature information d_i, where (x_i, y_i) is the coordinates of the feature point in the original image, z_i is the feature point confidence, and d_i is a 256-dimensional feature vector. After obtaining p_i and d_i, the similarity of feature description and location should be taken into account at the same time when performing feature matching via the graph neural network, combining the location information and feature vectors for location coding. In this paper, we use the Multilayer Perceptron (MLP) to fuse p_i into d_i, and obtain the joint feature c_i that fuses the location information and feature information:

$$c_i = d_i + MLP(p_i) \tag{2}$$

After obtaining the features c_i of the feature points, they are fed into a GNN that integrates an attention mechanism. This GNN architecture comprises several attention aggregation structures, each comprising a self-attention layer and a cross-attention layer. The self-attention layer enhances the specificity of input features for matching purposes, while the cross-attention layer determines matching target points by assessing the similarity between feature points across two images. Through iterative processes, the feature similarity between feature points and their corresponding target points is progressively refined. Given the relatively sparse features on the endocardium of the LV in cine MRI, the learning process is comparatively straightforward [9], this paper uses three attention aggregation structures. There are two types of undirected edges in the graph [21,22], E_{self} denotes an internal edge in one image, indicating that feature point i connects to other feature points in the same image. E_{cross} denotes an external edge between two different images, indicating that feature point i connects to other feature points in the other image. All the undirected edges $E\epsilon\{E_{self}, E_{cross}\}$.Information is propagated along two types of edges using the message passing formulation [23,24]. The feature information $m_{E\rightarrow i}$ is computed by attention aggregation in the GNN.$m_{E\rightarrow i}$ is the result of aggregation of all feature points $\{ j: (i,j)\epsilon E\}$:

$$m_{E\rightarrow i} = \sum_{j:(i,j)\epsilon E} W_{ij}\, V_{ij} \tag{3}$$

$$W_{ij} = Softmax_i\left(q_i^T k_j\right) \tag{4}$$

The process of obtaining $m_{E\to i}$ is analogous to a database search by querying q_i for the values v_j of certain elements based on their keys k_j.We obtain the attention weight w_{ij} from the Softmax function of the q_i and k_j. For $(Q,O)\epsilon\{D,S\}^2$, we can denote the obtained q_i, k_j and v_j as:

$$q_i = W_1 c_i^{l^Q} + b_1, k_i = W_2 c_i^{l^Q} + b_2, v_i = W_3 c_i^{l^Q} + b_3 \tag{5}$$

l denotes the information in layer l of the network where l is computed for E_{self} for odd numbers and E_{cross} for even numbers. The representation $c_i^{(l+1)^D}$ of feature point i in image D iteratively updated at layer (l+1) is calculated by the following equation:

$$c_i^{(l+1)^D} = c_i^{l^D} + MLP\left(\left[c_i^{l^D}, m_{E\to i}\right]\right) \tag{6}$$

Where $[\,,]$ denotes tandem operation. All feature points in image S are similarly updated. A number of iterative updates were performed to obtain the feature descriptors $y_i^D, y_j^S(i\epsilon D', j\epsilon S')$ for the two sets of linear projections, take as an example:

$$y_i^D = W c_i^{L^D} + b \tag{7}$$

The structure of the whole network is shown in Fig 2.

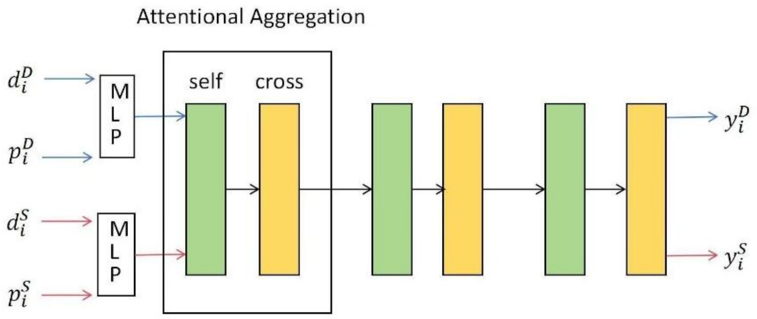

Fig. 2. The network architecture of Attentional GNN.

Feature matching using Sinkhorn algorithm.

In this paper, we use the Sinkhorn algorithm [25] to compute the final feature point matching results. According to Eq.7 to obtain $y_i^D, y_j^S(i\epsilon D', j\epsilon S')$,it is necessary to compute to obtain an allocation matrix $A\epsilon[0,1]_{m\times n}$.There are two characteristics of feature point matching on the LV endocardium: Feature points on the ED picture to the corresponding feature points on the ES picture have and only one feature point corresponds to it. Matching is performed based on the labeled expert points in the existing dataset, the number of feature points on the two images is the same, and all feature points can be matched, $m = n$. So we get an allocation matrix A that needs to satisfy $A1_n = 1_m$ and $A^T 1_m = 1_n$. In order to obtain the allocation matrix A, we also need to compute the score matrix $B \epsilon R^{m \times n}$ and maximize the score$\sum_{i,j} A_{i,j} B_{i,j}$. Constructing separate representations for all the m × n potentially possible matches is very difficult, so we need to utilize the feature vector similarity of the different feature points between the results y_i^D and y_j^S obtained from the attentional GNN as the score:

$$B_{i,j} = y_i^D \cdot y_j^S, \forall (i,j) \epsilon D' \times S' \tag{8}$$

where · denotes the inner product. To compute the distribution matrix A is to compute the optimal transmission problem between discrete distributions 1m and 1n of the score matrix B [26]. By iteratively updating the rows and columns with Sinkhorn algorithm, the entropy regularization formulation naturally yields the desired allocation, and it can be viewed as a microscopic Hungarian algorithm [27]. The final allocation matrix $A(i,j)$ is obtained, and $A(i,j)$ denotes the matching probability between feature point i in image D and feature point j in image S.

Loss.
In the feature matching process, it is trained by supervised approach through the truth value $T = \{(i,j)\} \subset D' \times S'$. The ground truth T is obtained based on the correspondence of expert points between images in the dataset. With this data we minimize the negative log likelihood loss:

$$LOSS_2 = -\sum_{(i,j) \epsilon T} \log A_{i,j} \tag{9}$$

3 Results

To assess the performance of the proposed feature matching method, 16 points sampled uniformly along the LV endocardial contour of each image serve as reference points, with matching accuracy serving as the evaluation metric. This experiment compares the performance of UAG with several state-of-the-art graph matching (GM) methods, including Spectral Matching with Affine Constraints (SMAC) [28], Integer Projected Fixed Point (IPFP) [29], Re-weighted Random Walk Matching (RRWM) [30], and Convex Cost Function Graph Matching (CGM) [12], utilizing FCN descriptors [12]. Matching accuracy for each case is determined by aggregating accuracy scores across all slices.

The matching accuracy on two databases, the database of 33 subjects and the MICCAI 2009 challenge database, is depicted in Fig 3 and Fig 4, respectively. The results demonstrate that UAG consistently achieves significantly higher matching accuracy compared to other GM methods. This observation validates UAG's superiority over state-of-the-art GM algorithms in addressing the LV correspondence estimation problem.

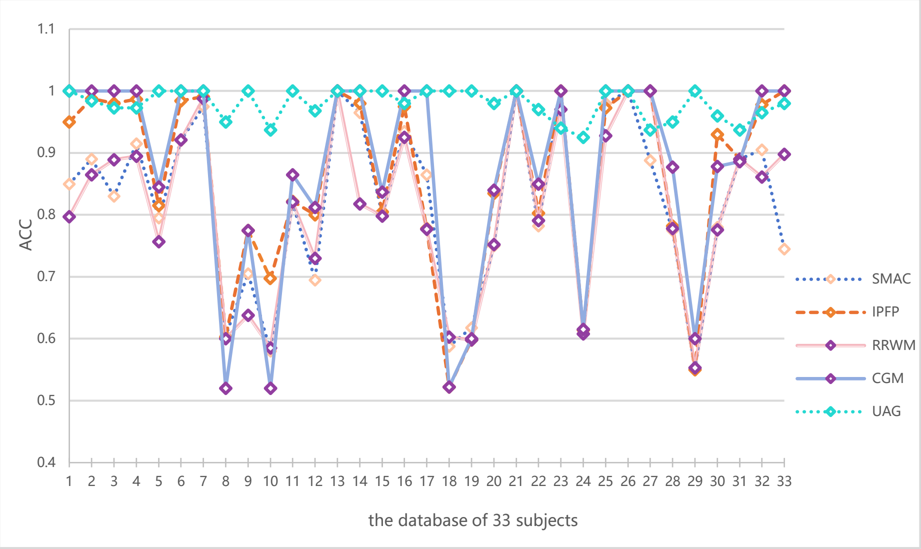

Fig. 3. Comparison of matching accuracy obtained by UAG and other methods on the 33-subject database.

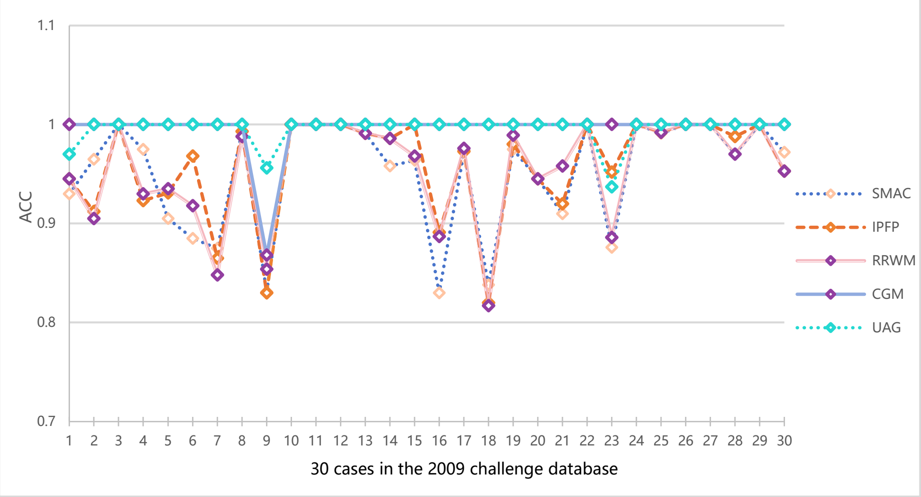

Fig. 4. Comparison of matching accuracy obtained by UAG and other methods on the MICCAI 2009 database.

Furthermore, we evaluate the performance of UAG for LV motion estimation by estimating the transformation function between cine MRI images at different image slices. The endocardial contour of a given slice, annotated by an expert, is mapped to a corresponding slice based on the estimated transformations. The resultant mapping errors are indicative of the performance of LV motion estimation, measured by comparing the mapped contour to the original endocardial contour.

Table 2. Comparison of APD between UAG and other GM algorithms using the database of 33 subjects, with optimal results shown in bold.

	SMAC	IPFP	RRWM	CGM	UAG
1	1.93	1.64	1.62	1.62	**1.49**
2	0.96	0.95	0.96	**0.94**	1.16
3	1.29	1.23	1.28	**1.22**	1.27
4	1.32	1.33	1.35	1.43	**1.16**
5	1.51	1.62	1.63	1.62	**1.50**
6	1.51	**1.49**	1.60	**1.49**	1.85
7	2.23	2.14	**2.08**	2.14	2.21
8	2.05	1.95	1.89	1.73	**1.49**
9	2.85	**1.39**	1.61	1.47	1.86
10	2.03	**1.73**	1.84	**1.73**	1.97
11	2.61	2.59	2.51	2.64	**2.20**
12	1.22	1.11	1.12	**1.07**	1.16
13	**0.86**	**0.86**	**0.86**	**0.86**	1.04
14	1.67	1.81	1.80	1.89	**0.92**
15	1.61	1.31	1.55	**1.26**	1.51
16	1.87	1.79	1.88	1.76	**1.36**
17	1.94	1.93	2.04	1.68	**1.18**
18	0.95	**0.94**	0.96	**0.94**	1.04
19	1.35	1.25	1.30	1.25	**1.06**
20	1.66	1.53	1.76	1.53	**1.52**
21	2.09	1.99	2.23	2.05	**1.34**
22	2.07	2.04	2.06	2.04	**1.64**
23	4.94	2.15	2.45	2.29	**2.10**
24	2.07	1.80	2.24	1.83	**0.98**
25	1.77	1.77	**1.69**	1.75	2.11
26	2.53	2.22	2.31	2.24	**1.38**
27	3.81	3.64	3.56	3.52	**3.21**
28	1.84	1.48	1.84	1.57	**1.20**
29	1.75	1.35	1.58	1.34	**1.04**
30	2.09	1.96	2.07	1.94	**1.06**
31	2.13	2.13	2.36	2.13	**1.29**
32	**1.52**	**1.52**	1.92	1.55	1.91
33	2.01	1.97	1.93	1.99	**1.76**
Average	1.97	1.71	1.81	1.71	**1.52**

Table 3. Comparison of APD between UAG and other GM algorithms using the MICCAI 2009 challenge database, with optimal results shown in bold.

	SMAC	IPFP	RRWM	CGM	UAG
1	2.85	1.71	1.92	1.62	**0.85**
2	1.80	1.78	1.83	1.71	**1.21**
3	2.36	2.00	2.11	1.98	**1.21**
4	1.72	1.59	1.84	1.56	**1.44**
5	2.41	1.93	2.22	1.89	**1.41**
6	1.40	1.42	1.53	**1.23**	1.33
7	2.82	2.41	2.79	**2.32**	**2.32**
8	2.83	2.75	2.60	2.76	**2.11**
9	2.21	2.23	2.19	2.17	**1.80**
10	2.98	2.86	2.91	2.87	**1.16**
11	3.30	2.93	3.29	2.92	**2.23**
12	4.33	4.30	4.64	**4.29**	5.82
13	2.29	2.21	2.22	**2.19**	2.23
14	**2.67**	2.72	2.71	2.71	3.84
15	2.10	1.99	2.01	2.02	**1.17**
16	2.47	1.76	1.81	**1.60**	1.89
17	1.54	1.28	1.41	**1.21**	1.73
18	2.26	1.86	2.22	1.69	**1.22**
19	2.05	1.75	1.97	2.03	**1.34**
20	2.10	**1.97**	2.22	1.99	2.15
21	3.37	2.28	2.83	2.17	1.65
22	**2.05**	2.07	2.08	2.07	3.13
23	2.06	2.05	2.45	1.91	**1.83**
24	1.88	**1.86**	1.87	1.87	2.28
25	2.91	2.80	2.75	2.77	**2.21**
26	2.34	2.22	2.31	2.22	**1.92**
27	2.15	2.04	2.04	2.08	**1.37**
28	1.76	1.81	1.79	1.80	**1.44**
29	2.46	**2.44**	2.46	**2.44**	2.46
30	2.15	1.88	2.01	**1.83**	2.02
Average	2.39	2.16	2.30	2.13	**1.96**

For each case, a trained segmentation model is employed to predict the endocardial contours of both the source and target images, subsequently estimating the correspondence matrix between these contours. The average perpendicular distance (APD) [17] is utilized as a metric to assess the performance of LV motion estimation, where lower APD values signify superior performance.

Experimental results obtained using UAG are compared with SMAC, IPFP, RRWM, and CGM methods. The comparison results are summarized in Tables 2 and 3. It is evident that UAG consistently exhibits lower APD values compared to other methods in the majority of cases, indicating its superiority for LV motion estimation using MRI images.

4 Discussion

As depicted in Fig. 3 and Fig. 4, UAG exhibits superior matching accuracy results on two publicly available cardiac MRI image databases. Furthermore, UAG demonstrates better and more consistent performance across specific cases. Particularly noteworthy is its performance on the 33-subject database, where in cases such as 8 and 29, the matching accuracy of alternative methods falls below 0.7, while UAG maintains a stable accuracy level above 0.9.

The limitations of graph matching methods become apparent when applied to MRI scans depicting abnormal LV) contours. In these methods, the LV contour is typically assumed to conform to a circular shape. Consequently, auxiliary points are introduced at the LV center to establish the graph's edges, connecting each contour point with the auxiliary point and its adjacent counterparts, thereby representing the LV's topology structure. However, deviations from this circular assumption, often observed in LV contours with non-circular structures, can cause auxiliary points to stray from the LV contour center. This discrepancy may lead to significant disparities between edges, thereby compromising the performance of graph matching (GM) methods in the presence of LV anomalies.

In contrast, the proposed UAG framework does not rely on auxiliary points for constructing the LV graph. Consequently, UAG achieves remarkable matching accuracy even in cases with abnormal LV contours, underscoring its robustness and efficacy in accommodating variations in LV morphology (as illustrated in Fig. 5).

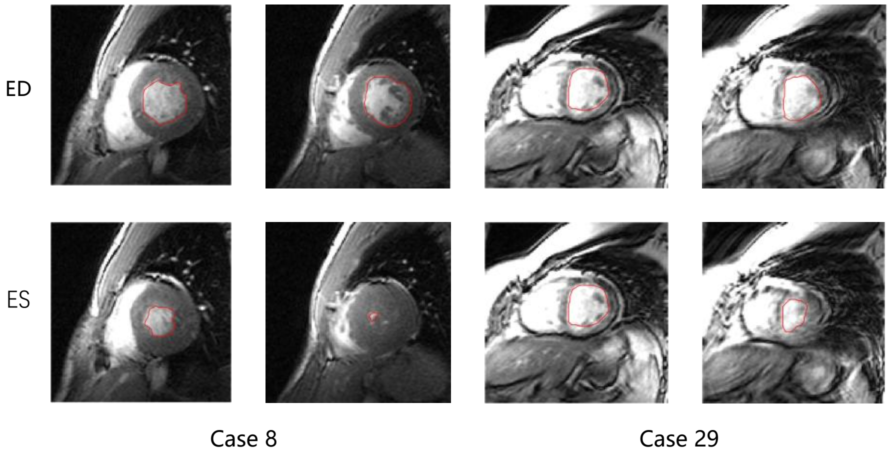

Fig. 5. Demonstrate abnormal LV contours. The first and the second lines show corresponding slices in ED and ES phases, respectively, for case 8 and case 29.

Furthermore, as can be seen from Tables 2 and 3, UAG possesses the lowest average APD compared to other graph matching methods, and average APD was 0.19 and 0.17

lower than CGM in the two publicly available cardiac MRI image databases. For each case, the UAG yielded the highest number of cases with the lowest APD.

5 Conclusion

This paper proposes a LV motion estimation framework, leveraging R2UNet for myocardial segmentation and feature extraction, while employing a graph matching network with an attention mechanism for feature matching. Experimental evaluations are conducted on two publicly available cardiac MRI image databases to assess the performance of the proposed framework, termed UAG. Results demonstrate that UAG surpasses other state-of-the-art methods in terms of feature matching accuracy and LV motion estimation accuracy using cardiac MRI images. Notably, UAG exhibits superior performance, particularly in matching accuracy and motion estimation accuracy, when applied to images with abnormal contours. This finding underscores the robustness of the proposed framework against LV contour abnormalities, highlighting its potential for clinical applications.

Acknowledgments. This work is supported by Shantou University Scientific Research Start-up Fund Project (NTF21005), and Science and Technology Ph.D. Research Startup Project under Grant No. SZIIT2023KJ016.

Conflict of Interest Statement. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. G. A. Mensah, V. Fuster, C. J. Murray, G. A. Roth, G. B. of Cardiovascular Diseases, and R. Collaborators, Global burden of cardiovascular diseases and risks, 1990-2022, Journal of the American College of Cardiology **82**, 2350–2473 (2023).
2. B. D. Powell, M. M. Redfield, K. A. Bybee, W. K. Freeman, and C. S. Rihal, Association of obesity with left ventricular remodeling and diastolic dysfunction in patients without coronary artery disease, The American journal of cardiology **98**, 116–120 (2006).
3. C. Russo, Z. Jin, S. Homma, T. Rundek, M. S. Elkind, R. L. Sacco, and M. R. Di Tullio, Effect of obesity and overweight on left ventricular diastolic function: a community-based study in an elderly cohort, Journal of the American College of Cardiology **57**, 1368–1374 (2011).
4. T. Mitsui, Circulating DPP4 Activity Predicts Systolic Left-ventricular Dysfunction in Heart Failure Patients., Journal of Cardiac Failure **19**, S132–S133 (2013).
5. Z. Zhang, X. Yang, C. Tan, W. Guo, and G. Chen, Surface structure feature matching algorithm for cardiac motion estimation, BMC medical informatics and decision making **17**, 11–24 (2017)

6. F. M. Parages, T. S. Denney, H. Gupta, S. G. Lloyd, L. J. Dell'Italia, and J. G. Brankov, Estimation of left ventricular motion from cardiac gated tagged MRI using an image-matching deformable mesh model, IEEE Transactions on Radiation and Plasma Medical Sciences 1, 147–157 (2017).
7. D. G. Lowe, Distinctive image features from scale-invariant keypoints, International journal of computer vision **60**, 91–110 (2004).
8. H. Bay, T. Tuytelaars, and L. Van Gool, Surf: Speeded up robust features, in *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9*, pages 404–417, Springer, 2006.
9. D. Navneet, Histograms of oriented gradients for human detection, in *International Conference on Computer Vision & Pattern Recognition, 2005*, volume 2, pages 886–893, 2005.
10. A. Atehort´ua, E. Romero, and M. Garreau, Characterization of motion patterns by a spatio-temporal saliency descriptor in cardiac cine MRI, Computer Methods and Programs in Biomedicine **218**, 106714 (2022).
11. G. Pedrizzetti, P. Claus, P. J. Kilner, and E. Nagel, Principles of cardiovascular magnetic resonance feature tracking and echocardiographic speckle tracking for informed clinical use, Journal of cardiovascular magnetic resonance **18**, 51 (2016).
12. J. Wu, Z. Gan, W. Guo, X. Yang, and A. Lin, A fully convolutional network feature descriptor: Application to left ventricle motion estimation based on graph matching in short-axis MRI, Neurocomputing **392**, 196–208 (2020).
13. O. Ronneberger, P. Fischer, and T. Brox, U-net: *Convolutional networks for biomedical image segmentation, in Medical image computing and computer-assisted intervention-MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241, Springer, 2015.
14. M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, Recurrent residual U-Net for medical image segmentation, Journal of Medical Imaging **6**, 014006–014006 (2019).
15. P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, Superglue: Learning feature matching with graph neural networks, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4938–4947, 2020.
16. A. Andreopoulos and J. K. Tsotsos, Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI, Medical image analysis **12**, 335–357 (2008).
17. P. Radau, Y. Lu, K. Connelly, G. Paul, A. J. Dick, and G. A. Wright, Evaluation framework for algorithms segmenting short axis cardiac MRI., The MIDAS Journal (2009).
18. H. Chui and A. Rangarajan, A new point matching algorithm for non-rigid registration, Computer Vision and Image Understanding 89, 114–141 (2003).
19. K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
20. M. Liang and X. Hu, Recurrent convolutional neural network for object recognition, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3367–3375, 2015.
21. P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, Community structure in time-dependent, multiscale, and multiplex networks, science **328**, 876–878 (2010).
22. V. Nicosia, G. Bianconi, V. Latora, and M. Barthelemy, Growing multiplex networks, Physical review letters **111**, 058701 (2013).
23. J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, Neural message passing for quantum chemistry, in *International conference on machine learning*, pages 1263–1272, PMLR, 2017.

24. P. W. Battaglia et al., Relational inductive biases, deep learning, and graph networks, arXiv preprint arXiv:1806.01261 (2018).
25. C. M. S. Distances, Lightspeed Computation of Optimal Transport, Advances in neural information processing systems **26**, 2292–2300 (2013).
26. G. Peyr´e et al., Computational optimal transport: With applications to data science, Foundations and Trends® in Machine Learning **11**, 355–607 (2019).
27. J. Munkres, Algorithms for the assignment and transportation problems, Journal of the society for industrial and applied mathematics **5**, 32–38 (1957).
28. T. Cour, P. Srinivasan, and J. Shi, Balanced graph matching, Advances in neural information processing systems **19** (2006).
29. M. Leordeanu, M. Hebert, and R. Sukthankar, An integer projected fixed point method for graph matching and map inference, Advances in neural information processing systems **22** (2009).
30. K. Daniilidis, P. Maragos, and N. Paragios, *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part V*, volume 6315, Springer, 2010.