Temporally coherent cardiac motion tracking from cine MRI: Traditional registration method and modern CNN method

Mengyun Qiao, Yuanyuan Wang[a], and Yi Guo
Department of Electrical Engineering, Fudan University, Shanghai, China,

Lu Huang and Liming Xia
Department of Radiology, Tongji Hospital Tongji Medical College Huazhong University of Science and Technology, Wuhan, China

Qian Tao[a]
Department of Radiology, Leiden University Medical Center, Leiden, the Netherlands

(Received 10 February 2020; revised 7 June 2020; accepted for publication 10 June 2020;
published 6 July 2020)

Purpose: Cardiac motion tracking enables quantitative evaluation of myocardial strain, which is clinically interesting in cardiovascular disease research. However, motion tracking is difficult to perform manually. In this paper, we aim to develop and compare two fully automated motion tracking methods for the steady state free precession (SSFP) cine magnetic resonance imaging (MRI), and explore their use in real clinical scenario with different patient groups.

Methods: We proposed two automated cardiac motion tracking method: (a) a traditional registration-based method, named full cardiac cycle registration, which simultaneously tracks all cine frames within a full cardiac cycle by joint registration of all frames; and (b) a modern convolutional neural network (CNN)-based method, named Groupwise MotionNet, which enhances the temporal coherence by fusing motion along a continuous time scale. Both methods were evaluated on the healthy volunteer data from the MICCAI 2011 STACOM Challenge, as well as on patient data including hypertrophic cardiomyopathy (HCM) and myocardial infarction (MI).

Results: The full cardiac cycle registration method achieved an average end-point error (EPE) 2.89 ± 1.57 mm for cardiac motion tracking, with computation time of around 9 min per short-axis cine MRI (size 128×128, 30 cardiac phases). In comparison, the Groupwise MotionNet achieved an average EPE of 0.94 ± 1.59 mm, taking < 1 s for a full cardiac phases. Further experiments showed that registration method had stable performance, independent of patient cohort and MRI machine, while the CNN-based method relied on the training data to deliver consistently accurate results.

Conclusion: Both registration-based and CNN-based method can track the cardiac motion from SSFP cine MRI in a fully automated manner, while taking temporal coherence into account. The registration method is generic, robust, but relatively slow; the CNN-based method trained with heterogeneous data was able to achieve high tracking accuracy with real-time performance. © *2020 The Authors. Medical Physics published by Wiley Periodicals LLC on behalf of American Association of Physicists in Medicine.* [https://doi.org/10.1002/mp.14341]

Key words: cardiac MRI, CNN, motion tracking, registration

1. INTRODUCTION

Cardiac magnetic resonance imaging (MRI) is the current gold standard technique to visualize the heart, and provides important diagnosis and prognosis information for cardiovascular disease patients.[1] Cardiac motion can be observed from the steady state free precession (SSFP) cine MRI, the movie of heart over a cardiac cycle. Cine MRI reveals the contracting and relaxing pattern of the myocardium in fine spatial and temporal resolution. In clinical practice, myocardium strain and motion abnormalities are visually assessed by radiologists. However, the visual assessment is tedious, subjective, and lacking in quantitative nature. With the complexity and volume of the cine scan (usually more than 300 frames per scan), quantification of myocardium motion can hardly be manually performed. Advanced computer methods are demanded to automatically track the cardiac motion and generate clinically relevant quantitative parameters.

Established cardiac motion tracking methods include (a) local feature tracking (FT),[2] which tracks the myocardium tissue by following the texture pattern around local myocardium, and (b) registration-based method,[3] which tracks the myocardium tissue by aligning the myocardium between cardiac phases and inversely calculating the motion. The FT method is fast, however, it only focuses on local texture and can be sensitive to noise and image quality.[4] Moreover, the method was originally developed for the speckle-tracking echocardiography,[5] in which the myocardium exhibits unique speckle patterns that facilitates tracking. In cine MRI, there are no such patterns within the myocardium and the method does not have a solid ground. In contrast, the registration-based method[6,7] uses the information not only from the myocardium but also from the context, and is more robust to local noise or global quality of images. Registration between images is, however, computationally intensive, and a potential problem is that the resulting motion is lacking in consistency, as no temporal constraints are applied to regulate the continuous motion in a cardiac cycle.[8,9]

The latest development of deep learning convolutional neural networks (CNN) brings new opportunities to tackle the motion tracking problem. While it is not a common image classification or segmentation task,[10–12] cardiac motion tracking can be formulated as an optimal flow computation problem. In literature, Dosovitskiy et al.[13] proposed two prominent CNN models for optical flow: FlowNetS and FlowNetC, which shows the feasibility of directly estimating optical flow from raw images. Ilg et al.[14] presented FlowNet2.0 combined several FlowNetC and FlowNetS networks into one comprehensive model. More recently, Sun et al.[15] presented a more compact and effective CNN model for optical flow, called PWC-Net, and Ren et al.[16] proposed a method for multiframe optical flow estimation. All these methods focused on natural image or video streams with huge training datasets. In addition, some recent research studied automatic tracking of cardiac motion. Rohé et al.[17] applied the stationary velocity fields (SVF) parametrization for cardiac motion analysis and built affine subspaces on a manifold, where each point refers to a three-dimensional (3D) image and the geodesic distance between two points describes the deformation. Krebs et al.[18] proposed an unsupervised multiscale deformable registration approach that learns a low-dimensional probabilistic deformation model. In this work, we propose to develop a dedicated CNN, built upon the previous optical flow work, to address the real-time motion tracking problem for cardiac cine MRI.

We observed that in clinical practice, the radiologists typically view cine MRI in the movie mode, evaluating many temporal frames at the same time. They rarely evaluate the motion only by comparing two frames of cine. Inspired by the expert way, in this work, we take into consideration the temporal dimension in cine in both registration-based and CNN-based algorithms. We evaluated both methods using heterogeneous datasets: healthy volunteers with normal cardiac motion, hypertrophic cardiomyopathy (HCM) patients, and myocardial infarction (MI) patients with abnormal cardiac motion.

The contribution of the work is threefold: (a) A comparison between the traditional registration-based method and the modern CNN-based method for cardiac motion tracking, in terms of accuracy, efficiency, and generalization capability; (b) the integration of temporal information in both methods to imitate the radiologists' way of viewing a cine MRI; (c) the experimental setup to include both healthy volunteer and patient data to evaluate both motion tracking methods in a clinical scenario.

2. MATERIALS

2.A. Dataset

The cardiac images used in this work are from three independent datasets: healthy volunteers, HCM, and MI. The healthy volunteer images are from the Motion Tracking Challenge datasets from MICCAI 2011 STACOM.[19] The MRI datasets were acquired using a 3.0-T Philips Achieva System (Philips Healthcare, Best, the Netherlands). SSFP datasets were scanned in multiple views (TR/TE = 2.9/1.5 ms, flip angle = 40). This dataset consisting of a whole-heart steady state free precession (SSFP) sequence gated at end-diastole and end-expiration from 15 healthy volunteers, where each acquisition consists of 14 short-axis levels with 30 cardiac phases. Image resolution is $1.2 \times 1.2 \times 8\ \mathrm{mm}^3$.

Additionally, cine MRI images of two patient cohorts were collected at Tongji Hospital, China, including 15 HCM patients and 15 MI patients. The datasets were both acquired by a 1.5-T Avando System (Siemens Medical Solutions, Erlangen, Germany). Each acquisition consists of 6-8 short-axis levels with 30-45 cardiac phases. The in-plane image resolution ranges from 1.679 to 1.979 mm, and through-plane image resolution is 10 mm. For all images, we only focused on the region-of-interest (ROI) of the heart, and cropped image from the center with 128 × 128 pixels, removing the background with little motion.

2.B. Ground truth

Development of motion tracking method usually suffers from the lack of ground-truth data. Gold standard of motion tracking is difficult to acquire due to the intensive labor needed; and for cine MRI it is especially difficult as the myocardium has a homogeneous texture (Fig. 1). We used an established algorithm to generate the motion ground truth on a dense grid.[20] The method matches pixel-wise scale-invariant feature transform (SIFT) features,[21] which establishes dense, semantically meaningful correspondence between two images. The core of this algorithm is based on Refs. [22] and [23]. The SIFT flow method follows the computational framework of the classical optical flow, but instead of matching pixels, it matches the transform-invariant SIFT features, and demonstrates improved performance over the original optical flow.[20] Because of its solid physical ground and state-of-the-art accuracy, the method has been widely chosen as reference in many work related to flow and motion, including optical flow least square estimation,[24] red blood cell tracking,[25] and cardiac motion estimation.[26] Figure 1 shows the ground truth obtained by the reference method, superimposed on the cine MRI, with the grid deforming at different cardiac phases. It can be observed that salient points, such as the right-ventricular insertion points, are well followed by the method.

3. METHODS

3.A. Cardiac motion tracking by registration

3.A.1. Full cardiac cycle registration

In a conventional way, by registering a pair of neighboring cine MRI frames (one as reference), displacement can be

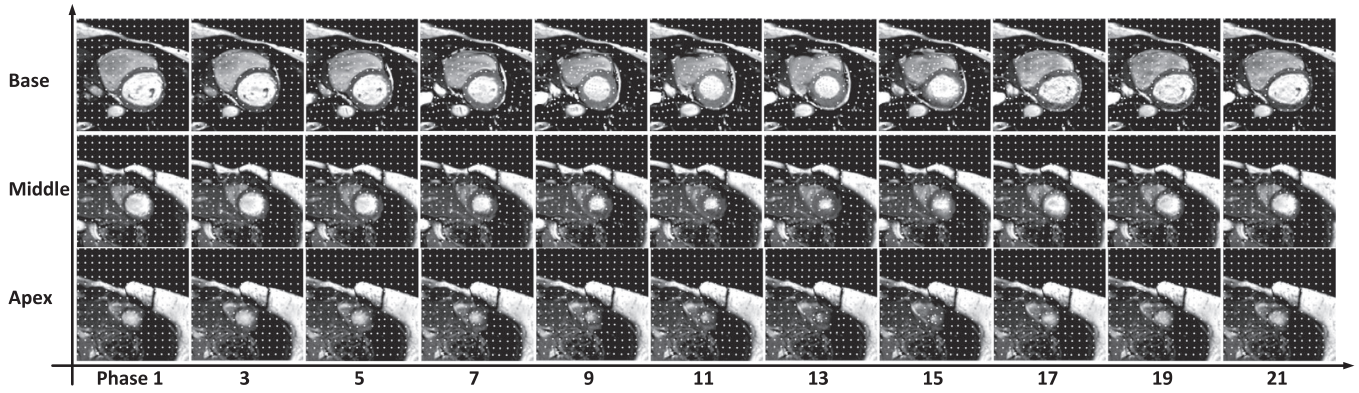

FIG. 1. An example of short-axis cine magnetic resonance imaging acquisition, with the x-axis being the cardiac phases, the y-axis being the location on the left ventricle: base, middle, and apex. The yellow grids show the motion tracking ground truth by the reference optical flow method.[20] The ground-truth motion fields were showed in sparse grid (8 × 8 pixels) for visualization, while the actual grid spacing in our experiments and calculation was 1 × 1 pixel.

computed at every voxel in the reference image and thereby motion can be estimated. If we have N frames in a cine MRI sequence, we can perform $N-1$ times pairwise registration between the neighboring frames and obtain the motion through a cardiac cycle.

This process, however, only takes into account two neighboring frames each time and cannot guarantee temporal coherence of the motion. Instead of $N-1$ separate independent registration processes, we propose to register all N frames in one registration step:

$$\rho = \arg\min_{\rho} C(T_\rho; I_1, I_2, \ldots, I_N) \tag{1}$$

where the cost function C can be defined as a groupwise registration metric.[27,28] In our work, we chose to minimize the variance in the group of images by:

$$C(\rho) = \sum_{i=1}^{N} \left(T_\rho(I_i) - \bar{I}_\rho\right)^2 \tag{2}$$

with $\bar{I}_\rho$ named as the “mean-shape” image:

$$\bar{I}_\rho = \frac{1}{N}\sum_{i=1}^{N} T_\rho(I_i) \tag{3}$$

where ρ is the transformation parameter applied to N images I_i. The transformation parameters are computed by minimizing (1), using the stochastic gradient descend method. The mean-shape image is conceptually an average “shape” of the heart across different phases, which is computed as the mean intensity image of groupwise registered images. It is different from the mean intensity image, which directly averages the original images and therefore suffers from motion of heart 6.

3.A.2. Hierarchical registration strategy

For stable registration performance, a hierarchical strategy is recommend. Firstly, to obtain a rough alignment of the heart size and orientation, we applied the affine registration, with the transformation parameters including one scaling and three rotational parameters. Subsequently, a B-spline nonrigid registration is performed to fit local deformation, where T_ρ is defined as:

$$T_\rho(x) = x + \sum_{x \in \mathrm{N}_x} p_k \beta^3\left(\frac{x - x_k}{\sigma}\right) \tag{4}$$

where x_k is the control points and σ is the x_k spacing. The definition of x_k is based on a regular grid k, which is defined by the physical space between the control points. p_k is the coefficient vector of the B-spline, β^3 is the cubic B-spline polynomial, and N_x is the set of all control points within the compact support of the B-spline at x. The transformation parameters set ρ is a combination of all these parameters. The local support of B-splines allow the transformation of a point to be computed from a limited number of surrounding control points. A multiresolution pyramid approach was applied to perform the nonrigid registration in a coarse-to-fine manner.[28]

3.A.3. Back propagation

After obtaining the transformation parameters for each image in the full cardiac cycle, the pixel positions of first image are propagated to the mean-shape image. The inverse transformation from the mean-shape image $\bar{I}$ to the other images can be obtained by solving $T'_\rho\left(T_\rho(I)\right) = I$.[26] Subsequently, the pixel position on the mean-shape image $\bar{I}$ can be back propagated onto the given image I. The motion during one cardiac circle is calculated based on the relative change of pixel position at neighboring phases.

3.B. Cardiac Motion Tracking by CNN

3.B.1. Optical flow

Optical flow is a well-established method to infer the vertical and horizontal motion of each pixel between two temporally neighboring frames,[29] expressed by the following equation:

$$I(x,y,t) = I(x + \Delta x, y + \Delta y, t + \Delta t) \tag{5}$$

where $I(x,y,t)$ is the signal intensity at the location (x,y) in the image plane at time t. Using the chain rule of differentiation

$$\frac{\partial I}{\partial x}V_x+\frac{\partial I}{\partial y}V_y+\frac{\partial I}{\partial t}=0 \qquad (6)$$

where V_x and V_y are the flow in two directions. Figure 2 shows two examples of motion field estimation from cine MRI by the optimal flow method, at the diastolic and systolic phases, respectively.

3.B.2. MotionNet

The MotionNet, dedicated to cardiac motion tracking, employs a context network that uses contextual information to refine the optical flow.[15] Given two input images, the network extracts features at each pyramid level by CNN with multiple channels. At each level, it utilizes the two up-sampled flow from the next level to warp the second feature image toward the first feature image. The cost volume layer is constructed using the features that store the matching costs for connecting a pixel with its corresponding pixel in the next frame. The optical flow estimator is a multilayer CNN enhanced with the DenseNet connections. The context network consists of seven convolutional layers with different dilation constants to refine the flow estimation. To solve the cardiac motion tracking problem, which is more constrained than natural or video scenes, the proposed network is lighter and faster than the traditional FlowNet.[14]

Figure 3 shows the detailed architecture of the MotionNet, each convolution followed by a leaky ReLU unit. The training loss is defined as:

$$L(\theta)=\sum_{l=l_0}^{L}\alpha_l\sum_{x}\left(\left|\,f_\theta^l(x)-f_{GT}^l(x)\right|+\varepsilon\right)^\rho+\delta\|\,\theta\,\|_2^2 \qquad (7)$$

where θ is the set of all learnable parameters, f^l is the flow field at the l th pyramid level. α_l is the weight at pyramid level l. ε is a small constant as weight decay. The parameter ρ controls the penalty for outliers to improve network robustness. δ weighs the regularization to reduce overfitting.

3.B.3. Groupwise MotionNet

With MotionNet as the building block, the Groupwise MotionNet integrates motion information over a time span to enhance accuracy and consistency of motion estimation. Figure 4 shows the architecture of the Groupwise MotionNet. Given N input cardiac image frames $I_{t-N-1},\ldots,I_t$, the Groupwise MotionNet estimates the optical flow $f_{t-1\to t}$ from frame I_{t-1} to frame I_t. The computation resource was the bottleneck for the max number frames that can be input into the MotionNet. Previous research 16 showed that using up till four frames improves the performance, while the relative improvement saturates when more than four frames are added. In this work, we therefore used four frames as an empirical choice to balance accuracy and speed of training.

The step is as follows: First, we use the pretrained MotionNet to estimate multiple motion fields: $f_{t-1\to t-3}, f_{t-3\to t-1}, f_{t-2\to t-1}, f_{t-1\to t-2}, f_{t-1\to t}$. Second, we warp the opposite flow pair to enhance the flow estimation[30]:

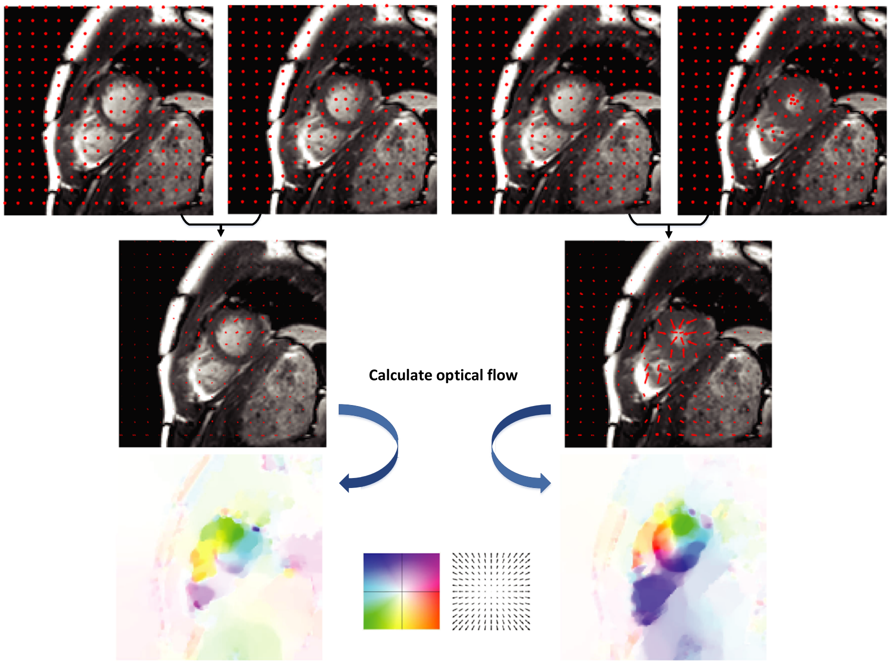

FIG. 2. Two examples of the optical flow computed between cine magnetic resonance imaging temporal frames. The flow field (third row) reflects the local displacement (second row) between two frames (first row), where red arrows represent example points motion direction. Flow image colors indicate directions of motion, and shades indicate the speed of motion. The first example illustrates the motion field during the diastolic phase, and the second example illustrates the more intense motion field during the systolic phase. [Color figure can be viewed at wileyonlinelibrary.com]

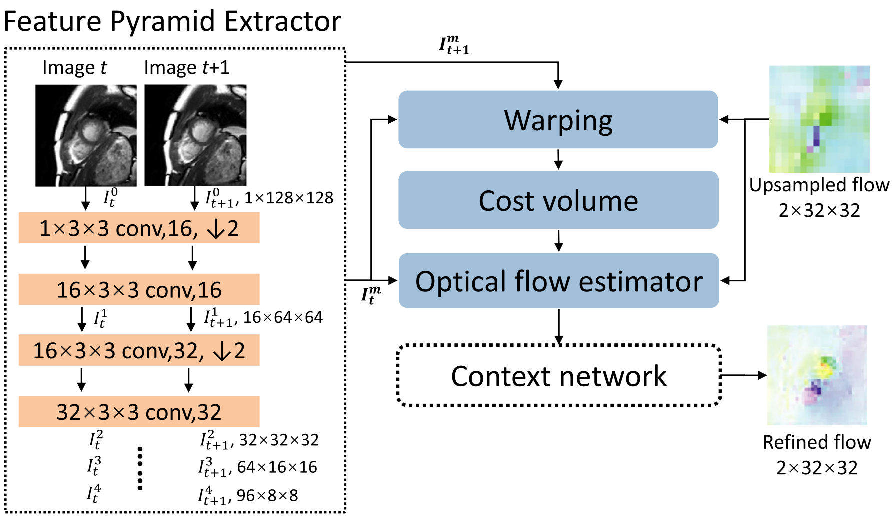

FIG. 3. The MotionNet architecture. The input of the network is two temporally adjacent cine frames and the output is the estimated optical flow. I_t^l denotes the extracted features of image I at level l. [Color figure can be viewed at wileyonlinelibrary.com]

specifically, we backward warped $f_{t-3\rightarrow t-1}$ to $f_{t-1\rightarrow t-3}$, and $f_{t-2\rightarrow t-1}$ to $f_{t-1\rightarrow t-2}$ to achieve a robust estimation of $f_{t-3\rightarrow t-1}$ and $f_{t-2\rightarrow t-1}$. We found the two-way estimation more accurate than the one-way estimation, as also suggested in literature.[16] The computation of $f_{t-1\rightarrow t}$, however, was not performed as $f_{t-1\rightarrow t}$ is our final goal of estimation and we intended to preserve the original computation in the direction of *t-1* -> *t* while using adjacent flows to correct for it. Third, we apply a light CNN to perform flow fusion integrating different estimates including: (a) the absolute image difference map $E_{t-3\rightarrow t-1}$, $E_{t-2\rightarrow t-1}$, and $E_{t-1\rightarrow t}$, where $E_{t-1\rightarrow t}$ is defined as $|I_{t-1} - W(I_t; f_{t-1\rightarrow t})|$, where $W(I;f)$ is the result of warping I_t to I_{t-1} using the flow field estimation. The map quantifies the difference between the first image and the second image warped with the estimated flow, indicating a measure of uncertainly for the estimated flow at each pixel, under the

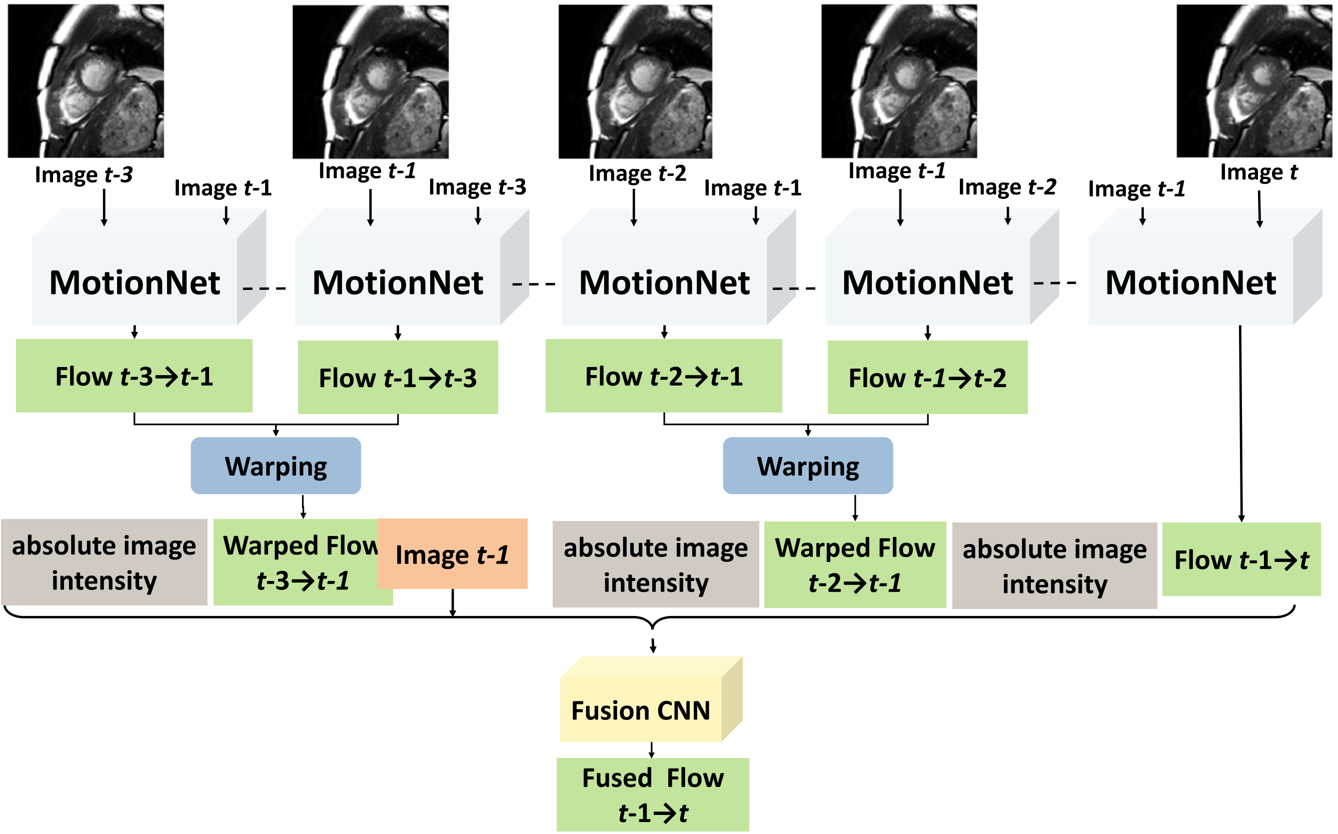

FIG. 4. The overall Groupwise MotionNet architecture. The input is four cine frames from I_{t-3} to I_t. Dashed lines imply shared weights of the same MotionNet. Multiple flows are estimated and fused by the fusion network to produce the final flow $f_{t-1\rightarrow t}$. [Color figure can be viewed at wileyonlinelibrary.com]

assumption that the difference between two cine frames was mainly due to cardiac motion, (b) the estimated flow fields $f_{t-3\to t-1}$, $f_{t-2\to t-1}$, and $f_{t-1\to t}$, and (3) the input image I_{t-1} The fusion step includes concatenating (1)–(3) as input to a CNN and pass them through a multilayer encoder–decoder CNN to produce the final flow estimation. Here the CNN acted as a way to fuse different channels of information by learning, as in Ref. [14].

4. EXPERIMENTS AND RESULTS

4.A. Experiments

4.A.1. Training and Testing Datasets

The dataset from MICCAI 2011 STACOM and Tongji Hospital contain in total 45 subjects, including 15 healthy volunteers, 15 HCM patients, and 15 MI patients, described in Section 2.A. For each cohort, we randomly divided them into ten training and five testing datasets.

We performed three experiments to compare the Full Cardiac Cycle Registration and Groupwise MotionNet methods for motion tracking. As the former method is training free, the setting of training and testing is only related to the Groupwise MotionNet method. For fair comparison of the performance, we tested the methods on the same heterogeneous testing cohort: five volunteer data, five HCM data, and five MI data. The follow four experiments were performed:

i. The Full Cardiac Cycle Registration applied to the testing cohort.
ii. Groupwise MotionNet trained only with healthy volunteer data: using 10 volunteer data for training the CNN, and tested on the 15 testing data.
iii. Groupwise MotionNet trained with heterogeneous data: using 10 volunteer data, 10 HCM data, and 10 MI data for training the CNN, and testing the CNN on the 15 testing data.

4.A.2. Registration-based motion tracking

All registrations were performed using the Elastix toolbox[31] in the Matlab environment (R2018b, MathWorks, Natick, MA, USA). The spline grid was set to 10 mm, the number of pyramids 3, and the fixed number of iterations for each resolution 1000. We adopted the minimal variance (variance-over-last-dimension in the toolbox) metric and the stochastic gradient descent approach for optimization.

4.A.3. CNN-based motion tracking

We used four consecutive cine MRI frames (named hereafter as quadruplet) as input in all experiments. Parameters in the loss function n were set empirically as $\alpha = [0.32, 0.08, 0.02, 0.01], \varepsilon = 0.008, \delta = 0.0006$, and $\rho = 0.3$. For the Groupwise MotionNet trained with ten healthy volunteer data, we used 2900 quadruplets to training the CNN, in which 87 were used for validation of network parameters. For the Groupwise MotionNet trained with 30 heterogeneous data, we used 11 108 quadruplets to train the CNN, in which 2777 were used for validation. (In the latter scenario we kept a larger portion of data for validation purposes, as the training data are sufficient.) The heterogeneous testing cohort, including five subjects per category, have in total 5100 pairs of consecutive frames for evaluating the motion tracking performance.

4.A.4. Performance evaluation

The accuracy of the motion estimates was evaluated using the average end-point error (EPE) defined as the Euclidean distance between the predicted flow field f and the reference motion estimation by the reference method f^r over the heart ROI:

$$EPE = \frac{1}{M}\sum_{i=1}^{M}\sqrt{(f_x - f_x^r)^2 + (f_y - f_y^r)^2} \tag{8}$$

where M is the total number of pixels in the heart ROI.

4.B. Results: comparison to state-of-the-art CNNs

We evaluated the two proposed methods, namely, Full Cardiac Cycle Registration and Groupwise MotionNet, with reference to the basic pairwise registration method and the basic MotionNet without integrating temporal information. We also benchmarked the Groupwise MotionNet with three other state-of-the-art CNNs: FlowNetS 13, FlowNetC 13, and FlowNet2 14. The experiments were performed on the healthy volunteer data from the STACOM challenge, where all neural networks were trained on the same data as used to train the proposed CNNs. The pairwise registration resulted in higher EPE (3.24 ± 1.72 mm) than the Full Cardiac Cycle Registration method (2.89 ± 1.57 mm), $P<0.05$ by the paired t-test. The Groupwise MotionNet achieved a significantly lower average EPE (0.94 ± 1.59 mm) compared to the Full Cardiac Cycle Registration method (2.89 ± 1.57 mm), the MotionNet (1.17 ± 1.48 mm), the FlowNetS (2.62 ± 2.37 mm), FlowNetC (2.33 ± 2.48), and the FlowNet2 (2.21 ± 2.08 mm). The results of average EPE and running time are reported in Table I. Figure 5 shows a few examples of cardiac motion field by proposed Full Cardiac Cycle Registration and the Groupwise MotionNet. The boxplot showed the motion tracking performance in three groups: volunteer, HCM, and MI in Fig. 6. Three experiments were carried out for each group of subjects: Full Cardiac Cycle Registration, Groupwise MotionNet model trained by 10 volunteers, and by 30 subjects from each subject group: 10 volunteers, 10 HCM patients, and 10 MI patients.

4.C. Results: generalization to patient data

The generalizability of training-based method to new datasets is an important concern in clinical applications. The results of the two motion tracking method on different

TABLE I. Motion tracking performance.

		Registration-based methods		CNN-based methods				
EPE (mm)		Pairwise registration	Full cardiac cycle registration	FlowNetS	FlowNetC	FlowNet2	MotionNet	Groupwise MotionNet
EPE (mm)	Train			1.62 ± 1.83	1.59 ± 1.91	1.41 ± 1.82	0.86 ± 1.32	0.73 ± 1.27
	Test	3.24 ± 1.72	2.89 ± 1.57	2.62 ± 2.37	2.33 ± 2.48	2.21 ± 2.08	1.17 ± 1.48	0.94 ± 1.59
Time per frame(s)		7.63	19.13	0.05	0.06	0.13	0.02	0.03

Performance comparison among different motion tracking methods (registration based and CNN based), in terms of average end-point error (EPE) (mm) and running time.

cohorts are reported in Table II. The registration-based method achieved stable results with EPE of 2.73 ± 1.43, 2.87 ± 1.82, and 2.93 ± 1.74 mm for the heterogeneous testing data: five healthy volunteers, five HCM, and five MI subjects, respectively, without an issue of generalizability. In comparison, the Groupwise MotionNet trained with ten healthy volunteers showed excellent performance on the five healthy volunteer data (EPE 0.94 ± 1.59mm), but had significantly degraded performance on the HCM and MI cohorts, with an EPE of 3.14 ± 2.53mm and 3.61 ± 2.72mm, respectively. The results were also inferior to the Full Cardiac Cycle Registration method. However, when trained with 30 heterogeneous dataset, the Groupwise MotionNet showed improved generalization across the cohorts, with the EPE reduced to 0.87 ± 1.34, 0.90 ± 1.86, 0.89 ± 1.64 mm for healthy, HCM, and MI, respectively.

In addition, we manually annotated the myocardium in 15 subjects, including five cases from each category of healthy

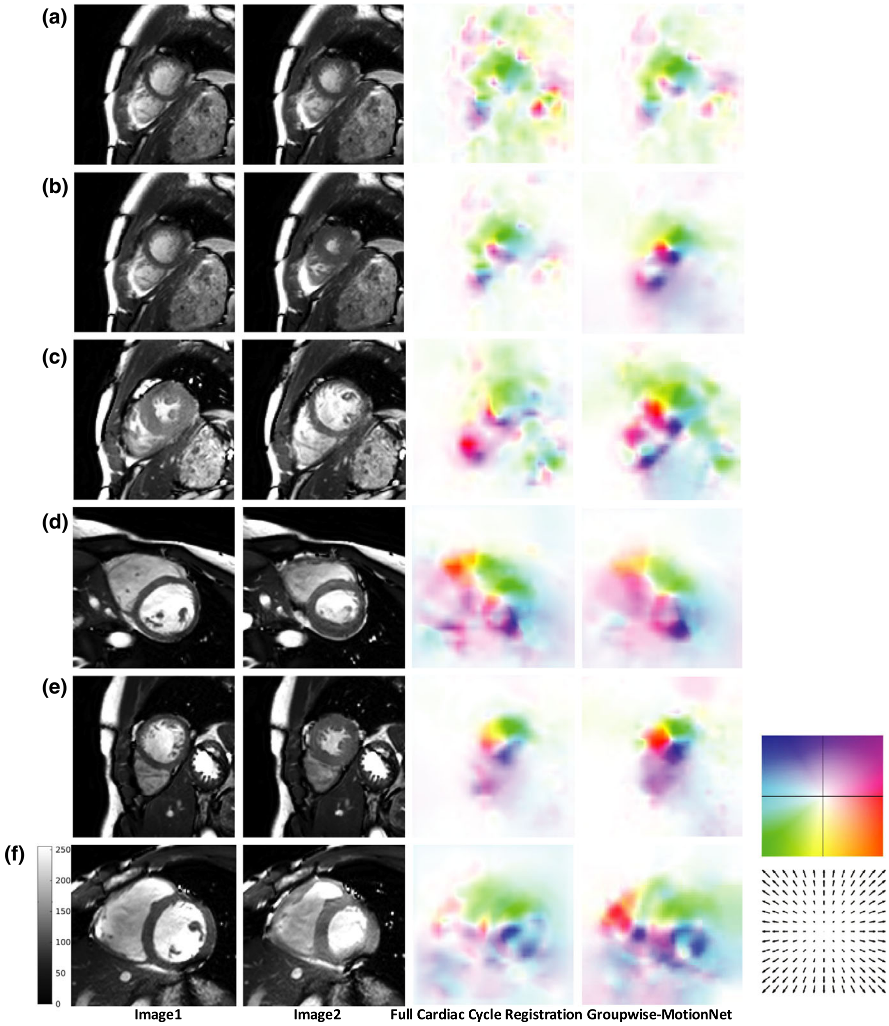

FIG. 5. The flow field estimation from input Image 1 to Image 2 by the two proposed methods: Full cardiac cycle registration and the Groupwise MotionNet. The colormap of flow field is shown at the rightmost panel, encoding both the amplitude and direction of motion. [Color figure can be viewed at wileyonlinelibrary.com]

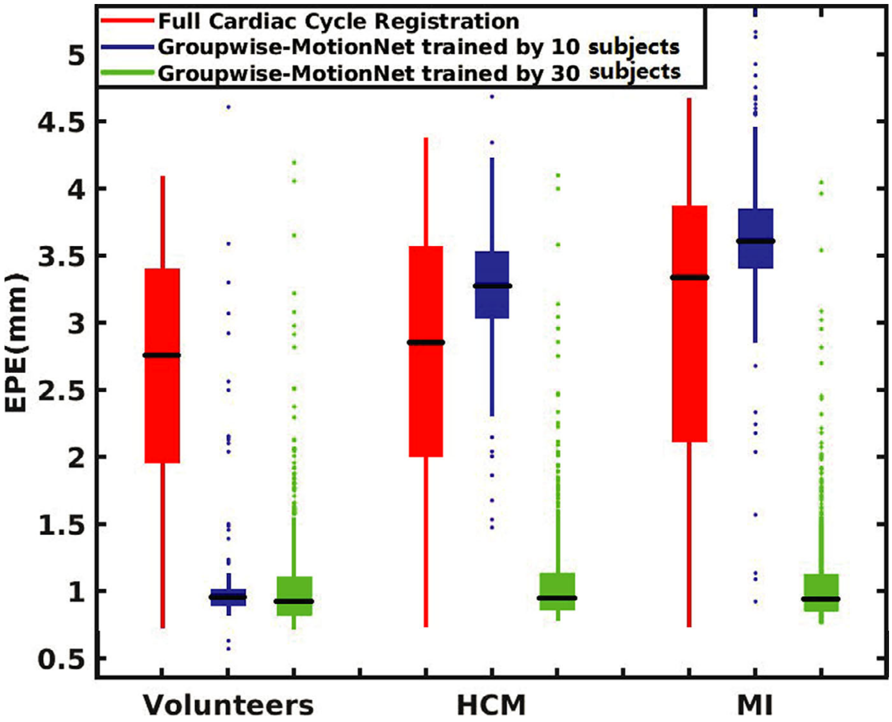

FIG. 6. Bar plots of the testing results of motion tracking in three groups: volunteer, HCM, and MI. Each group has three bars: Full cardiac cycle registration, Groupwise MotionNet model trained by 10 volunteers, and trained by 30 subjects: 10 volunteers, 10 HCM patients, and 10 MI patients. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE II. Generalizability of motion tracking methods.

EPE (mm)	Volunteer	HCM	MI
Full cardiac cycle registration	2.73 ± 1.43	2.87 ± 1.82	2.93 ± 1.74
Groupwise MotionNet			
Trained by data from healthy volunteers	0.94 ± 1.59	3.14 ± 2.53	3.61 ± 2.72
Trained by data from heterogeneous cohort	0.87 ± 1.83	0.90 ± 1.86	0.89 ± 1.64

Generalizability of the registration-based and CNN-based motion tracking methods on different cohorts: healthy volunteer, hypertrophic cardiomyopathy (HCM), and myocardial infarction (MI). Performance was quantified by the average endpoint error (EPE) (mm).

volunteers, hypertrophic cardiomyopathy (HCM), and myocardial infarction (MI). The results of the EPE measure in the myocardium region are reported in Table III.

4.D. Execution performance

For the Full Cardiac Cycle Registration, the typical computation time was around 9 min for a cardiac cycle, implying around 18 s per frame on a computer with a CPU of Intel Xeon E5-2687W processor (3.00GHz) and 64GB RAM. In comparison, the CNN-based approaches dramatically reduced the computation time to 0.03 s per cine frame. Computation time of motion in one cine MRI acquisition was therefore reduced from 9 min to <1 s.

5. DISCUSSION

In this work, we developed and validated two fully automated motion tracking methods, using different methodologies: registration based and CNN based. In both methods, we integrated the temporal information to make the motion tracking more coherent and accurate. As CNN motion tracking method is based on training, we further tested its generalizability on different patient cohorts: healthy, HCM, and MI. Experiment results show that the proposed Groupwise MotionNet, if trained with heterogeneous data, could achieve fast and accurate motion tracking performance. The Full Cardiac Cycle Registration method also showed stable performance on all cohorts of data, without need of training.

A characteristic of the proposed Full Cardiac Cycle Registration method is that it takes into consideration of all cine frames in one full cardiac cycle. With pairwise registration, the motion tracking can be sensitive to the artifact in one occasional frame, that is, the motion related to this particular frame may fail and shows up as abrupt, unrealistic movements. To estimate the motion over the full cardiac cycle as one optimization problem, the situation is largely alleviated as the continuity of motion is an intrinsic constraint during the computation. The registration-based motion tracking method does not require training, therefore can be used in situations where training data are not available.

Convolutional neural network-based approaches are increasingly popular in recent years, symbolizing the paradigm shift in medical image analysis. Compared to the

TABLE III. Motion tracking performance within myocardium.

EPE (mm)	Volunteer (N = 5)	HCM (N = 5)	MI (N = 5)
Full cardiac cycle registration	3.91 ± 1.74	4.21 ± 2.15	4.08 ± 2.28
Groupwise MotionNet trained by data from heterogeneous cohort	2.03 ± 1.91	2.15 ± 2.01	2.32 ± 1.87

The average end-point error (EPE) within the annotated myocardium region in five subjects from each of the three categories: healthy volunteer, hypertrophic cardiomyopathy (HCM), and myocardial infarction (MI).

registration-based method, the CNN-based method showed higher motion tracking accuracy, when the training and testing data are from the same cohort. Another big advantage of CNN-based methods over registration-based methods is that they run real time. This makes it feasible for evaluating motion in clinical practice, where the radiologists do not need to wait for a long time to report quantitative numbers. In the proposed Groupwise MotionNet, we also integrated the temporal information from neighboring frames to boost its performance by the groupwise strategy. Our experiments showed that the Groupwise MotionNet indeed achieved better motion tracking performance than the original MotionNet alone.

As CNN is training based, it is very important that the generalization capability is rigorously evaluated, not only on independent testing data from the same cohort but also on completely unseen data from patient cohorts. In this work, we showed that the first Groupwise MotionNet trained only with healthy volunteer data worked well on other volunteer data, but degraded significantly on unseen patient data. However, if we enlarge the training dataset to include patient data, the generalization significantly improved on all cohorts. This is in line with a previous multicenter multivendor study, showing that including data from heterogeneous origin is a simple and effective way to improve generalization.[32] We argue that the enlarged training dataset is helpful in two senses: the first reason is that the training datasets now include different motion patterns covering a wider population, and the second reason is that the training dataset now also include data from different MRI machines (i.e., Siemens) that have different image characteristics, for example, contrast and sharpness. Without the presence of the physical tissue patterns as in MRI tagging, the tracking error of the CNN method was around 1.5 pixels, slightly higher than that reported previously from MRI tagging (around 1 pixel).[33]

The improvement in accuracy of CNN-based methods over registration-based methods can be attributed to its flexibility: in our registration-based method, for regularization we used B-Spline, which may limit the estimated displacement at the systolic phase where the motion is too large to be covered by smooth Spline; in comparison, CNN-based methods do not apply such internal parametrization. In Fig. 5 it can be observed that for neighboring frames with large changes (e.g., systole), the motion amplitude estimated by the Groupwise MotionNet was higher than that by the Full Cardiac Cycle Registration method. In the registration-based motion tracking method, we adopted the Elastix implementation; however, we note that the registration could also be performed by optical flow-based algorithms, which are likely to yield motion tracking performance closer to the optical flow-based ground truth.

There are a few limitations in the presented study. Firstly, due to the lack of gold standard, we used a widely used flow estimation method from the computer vision society, based on physical principles, to generate the ground truth. Secondly, strictly speaking, the motion of heart is in 3D,[34–37] but our computed flow was in 2D because most clinical cine MRI is 2D + t, acquired per breath-hold. Nevertheless, if there are 3D + t cine MRI data available for clinical use, our method can be adapted in two ways: firstly, we can create multiorientation 2D + t data (e.g., short-axis, two-chamber, four-chamber) and directly apply the methods, secondly, the methodology of both methods can be extended to one dimension higher, with the same rationale but increased computation and memory use. Another limitation is that to improve the CNN generalization, we need to include sufficient data from different cohorts (although no annotation is needed) for training the CNN. Further studies on transfer learning or domain adaptation are warranted, which may lead to a more generic solution.

6. CONCLUSION

In this paper, we developed and compared two fully automatic cardiac motion tracking method for SSFP cine MRI, namely, Full Cardiac Cycle Registration and Groupwise MotionNet. In designing both methods, we incorporated temporal information for more accurate and coherent motion tracking. We evaluated both methods on the heterogeneous datasets including healthy volunteers, HCM patients, and MI patients. Experiments showed that the registration method had stable performance independent of patient cohort and MRI machine, while the CNN-based method had low generalizability when the training data were limited. However, the CNN-based method trained with heterogeneous data achieved high accuracy in different patient groups, with real-time performance.

ACKNOWLEDGMENTS

This work was supported by National Key Research and Development Program of China under Grant 2018YFC0116303.

[a] Authors to whom correspondence should be addressed. Electronic mails: yywang@fudan.edu.cn and q.tao@lumc.nl.

REFERENCES

1. de Roos A, Higgins CB. Cardiac radiology: centenary review. *Radiology*. 2014;273:S142–S159.

2. Hor KN, Gottliebson WM, Carson C, et al Comparison of magnetic resonance feature tracking for strain calculation with harmonic phase imaging analysis. *JACC Cardiovasc Imaging*. 2010;3:144–151.
3. Elen A, Choi HF, Loeckx D, et al Three-dimensional cardiac strain estimation using spatio–temporal elastic registration of ultrasound images: a feasibility study. *IEEE Trans Med Imaging*. 2008;27:1580–1591.
4. Vavilin A, Ha LM, Jo KH, et al Camera motion estimation and moving object detection based on local feature tracking. Lecture Notes in Computer Sciences; 2012;544–552.
5. Obokata M, Nagata Y, Wu VCC, et al Direct comparison of cardiac magnetic resonance feature tracking and 2D/3D echocardiography speckle tracking for evaluation of global left ventricular strain. *Eur Heart J Cardiovasc Imaging*. 2016;17:525–532.
6. Qiao M, Wang Y, Berendsen FF, et al Fully automated segmentation of the left atrium, pulmonary veins, and left atrial appendage from magnetic resonance angiography by joint-atlas-optimization. *Med Phys*. 2019;46:2074–2084.
7. Royuela VJ, Cordero GL, Simmross WF, et al Nonrigid groupwise registration for motion estimation and compensation in compressed sensing reconstruction of breath-hold cardiac cine MRI. *Magn Reson Med*. 2016;75:1525–1536.
8. Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging*. 2004;23:903–921.
9. Langerak TR, van der Heide UA, Kotte ANTJ, et al Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE). *IEEE Trans Med Imaging*. 2010;29:2000–2008.
10. Voulodimos A, Doulamis N, Doulamis A, et al Deep learning for computer vision: a brief review. *Comput Intell Neurosci*. 2018;2018:1–13.
11. Martin JF, Volfson LB, Kirzon-Zolin VV, et al Application of pattern recognition and image classification techniques to determine continuous cardiac output from the arterial pressure waveform. *IEEE Trans Biomed Eng*. 1994;41:913–920.
12. Yan Z, Yang X, Cheng KT. Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation. *IEEE Trans Biomed Eng*. 2018;65:1912–1923.
13. Dosovitskiy A, Fischer P, Ilg E, et al Flownet: Learning optical flow with convolutional networks. *Proceedings of the IEEE International Conference on Computer Vision*; 2015: 2758–2766
14. Ilg E, Mayer N, Saikia T, et al Flownet 2.0: Evolution of optical flow estimation with deep networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2017: 2462–2470.
15. Sun D, Yang X, Liu MY et al Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2018: 8934–8943.
16. Ren Z, Gallo O, Sun D, et al A fusion approach for multi-frame optical flow estimation. *IEEE Winter Conference on Applications of Computer Vision (WACV)*; 2019:2077–2086.
17. Rohé MM, Sermesant M, Pennec X. Low-dimensional representation of cardiac motion using barycentric subspaces: a new group-wise paradigm for estimation, analysis, and reconstruction. *Med Image Anal*. 2018; 45:1–12.
18. Krebs J, Delingette H, Mailhé B, et al Learning a probabilistic model for diffeomorphic registration. *IEEE Trans Med Imaging*. 2019;38: 2165–2176.
19. Tobon-Gomez C, De Craene M, Mcleod K, et al Benchmarking framework for myocardial tracking and deformation algorithms: an open access database. *Med Image Anal*. 2013;17:632–648.
20. Liu C. Beyond pixels: exploring new representations and applications for motion analysis. Doctoral Thesis, Massachusetts Institute of Technology, May; 2009.
21. Lowe DG. Object recognition from local scale-invariant features. *Proceedings of the seventh IEEE International Conference on Computer Vision*; 1999: 1150–1157.
22. Brox T, Bruhn A, Papenberg N, et al High accuracy optical flow estimation based on a theory for warping. *European Conference on Computer Vision*; 2004:25–36.
23. Bruhn A, Weickert J, Schnörr C. Lucas/Kanade meets Horn/Schunck: combining local and global optic flow methods. *Int J Comput Vision*. 2005;61:211–231.
24. Guo D, Van de Ven AL, Zhou X. Red blood cell tracking using optical flow methods. *IEEE J Biomed Health Inform*. 2013;18:991–998.
25. Yan W, Wang Y, Li Z, et al Left ventricle segmentation via optical-flow-net from short-axis cine MRI: preserving the temporal coherence of cardiac motion. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham; 2018: 613–621.
26. Metz CT, Klein S, Schaap M, et al Nonrigid registration of dynamic medical imaging data using nD+ t B-splines and a groupwise optimization approach. *Med Image Anal*. 2011;15:238–249.
27. Shahzad R, Tao Q, Dzyubachyk O, et al Fully-automatic left ventricular segmentation from long-axis cardiac cine MR scans. *Med Image Anal*. 2017;39:44–55.
28. Thevenaz P, Ruttimann UE, Unser M. A pyramid approach to subpixel registration based on intensity. *IEEE Trans Image Process*. 1998;7:27–41.
29. Horn BKP, Schunck BG. Determining optical flow. Techniques and applications of image understanding. *Intl Soc Opt Phot*. 1981;281:319–331.
30. Wang TC, Zhu JY, Kalantari NK, et al Light field video capture using a learning-based hybrid imaging system. *ACM Trans Graph (TOG)*. 2017;36:1–13.
31. Klein S, Staring M, Murphy K, et al Elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging*. 2009;29:196–205.
32. Tao Q, Yan W, Wang Y, et al Deep learning–based method for fully automatic quantification of left ventricle function from cine MR images: a multivendor, multicenter study. *Radiology*. 2019;290:81–88.
33. Xu C, Pilla J, Isaac G, et al Deformation analysis of 3D tagged cardiac images using an optical flow method. *J Cardiovasc Magn Reson*. 2010;12:19.
34. Mcveigh ER. MRI of myocardial function: motion tracking techniques. *Magn Reson Imaging*. 1996;14:137–150.
35. Osman NF, Kerwin WS, Mcveigh ER, et al Cardiac motion tracking using CINE harmonic phase (HARP) magnetic resonance imaging. *Magn Reson Med*. 1999;42:1048–1060.
36. Pan L, Prince JL, Lima JAC, et al Fast tracking of cardiac motion using 3D-HARP. *IEEE Trans Biomed Eng*. 2005;52:1425–1435.
37. Chen T, Wang X, Metaxas D, et al 3D cardiac motion tracking using robust point matching and meshless deformable models. IEEE ISBI; 2008.