IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 19 December 2022, accepted 5 January 2023, date of publication 19 January 2023, date of current version 26 January 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3238058

RESEARCH ARTICLE

GMCNet: A Generative Multi-Resolution Framework for Cardiac Registration

AMENEH SHEIKHJAFARI[1,2], MICHELLE NOGA[2], AHMED AHMED[3], NILANJAN RAY[1], AND KUMARADEVAN PUNITHAKUMAR[1,2], (Senior Member, IEEE)

[1]Department of Computing Science, University of Alberta, Edmonton, AB T6G 2G8, Canada
[2]Department of Radiology and Diagnostic Imaging, University of Alberta, Edmonton, AB T6G 2G8, Canada
[3]Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2G8, Canada

Corresponding author: Ameneh Sheikhjafari (sheikhja@ualberta.ca)

ABSTRACT Deformable image registration plays a crucial role in estimating cardiac deformation from a sequence of images. However, existing registration methods primarily process images as pairs instead of processing all images in a sequence together. This study proposes a novel end-to-end learning-free generative multi-resolution convolutional neural network (GMCNet) with the primary focus of registering images in a sequence. Even though learning-based methods have yielded high performance for image registration, their performance depends on their ability to learn information from a large number of samples which are difficult to obtain and might bias the framework to the specific domain of data. The proposed learning-free method eliminates the need for a dedicated training set while exploiting the capabilities of neural networks to achieve accurate deformation fields. Due to its capability of parameter sharing through the architecture, the GMCNet can be used as a groupwise registration as well as pairwise registration. The proposed method was evaluated on three different clinical cardiac magnetic resonance imaging datasets and compared quantitatively against nine other state-of-the-art learning and optimization-based algorithms. The proposed method outperformed other methods in all comparisons and yielded average Dice metric values ranging from 0.85 to 0.88 for the datasets. Different aspects of the GMCNet are also explored by assessing 1) the robustness; 2) performance on pairwise registration; 3) the influence of spatial transformation in a controlled environment; and 4) the impact of different multi-resolution structures. The results demonstrate that using temporal information to estimate the deformation fields leads to more accurate registration results and improved robustness under different noise levels. Moreover, the proposed method does not need images for training, and therefore, its prediction is not domain-specific and can be applied to any sequence of images.

INDEX TERMS Convolutional neural networks, cardiac cine MRI, deformable registration, generative network, learning-free framework, multi resolution.

I. INTRODUCTION

Deformable image registration is one of the crucial tasks in medical image analysis which aims to find the point-wise mapping between a pair of images. These techniques serve as the fundamental basis for procedures such as surgery, tumour growth monitoring, minimally invasive treatments, and many other challenging problems [1], [2], [3], [4], [5]. Studies have shown that deformable registration could also be used for the cardiac functional assessment and delineation of cardiac structures with magnetic resonance imaging (MRI) sequences. However, the lack of reliable, automatic and accurate tools to register images in a temporal sequence severely limits the use of registration for many medical image analysis applications including cardiac functional assessment. The majority of existing image registration methods primarily process input images pair-wise and do not take advantage of temporal information available in image sequences.

Typically, deformable image registration is modelled as an optimization problem consisting of an image similarity metric and regularization terms and applied to pairs of images. Without regularization, this may result in multiple physically

The associate editor coordinating the review of this manuscript and approving it for publication was Larbi Boubchir.

non-plausible solutions, for instance, tissue folding or/and tearing in human anatomy studies. Existing deformable registration algorithms could be broadly categorized into non-learning-based (classical) methods and learning-based methods.

II. BACKGROUND AND RELATED WORK

A. CLASSICAL IMAGE REGISTRATION METHODS

Traditionally, deformable registration is resolved by optimizing a similarity metric that measures the closeness between the fixed and the warped moving image. Several studies using optimization within the space of displacement vector field such as discrete methods [6], [7] and Demons [8], [9] have been proposed for the registration problem.

In [10] a combination of segmentation and registration is suggested based on nonlinear elasticity which uses a polyconvex for regularization. Another study [11] proposed a method based on the log-domain for spatial transformation where a physical constraint is applied to estimate the myocardial strain from cine MRI in the registration process. However, the lack of diffeomorphism properties in transformations could lead to image folding and twisting. In contrast, some methods rely on diffeomorphic transformation [1], computed using an artificial velocity over time governed by the Lagrange transport equation to model deformations. Due to the invertibility and folding-free transformation properties of the diffeomorphism, several other studies utilized diffeomorphic transformations in image registration applications [12], [13], [14].

In [15] and [16], a registration framework was proposed based on moving mesh (grid generation) to compute point-to-point correspondences, where the L^2 norm was used as a dissimilarity measure. The authors modelled the deformation using radial (divergence) and rotational (curl) components that is better suited for analyzing the heart as it closely matches actual cardiac motion.

Adapting the aforementioned methods to the specific application is challenging since most of these methods require users to find parameters on hand. The process of finding the parameters that match the characteristics of certain data depends entirely on the users' intuition and several tedious attempts to avoid defining the too flexible or too restrictive model. Besides, occasionally regularization terms need to be manually adjusted for each application.

B. DEEP LEARNING-BASED IMAGE REGISTRATION METHODS

In the past few years, learning-based predictive approaches, notably deep convolutional neural networks (CNNs) have been successfully applied to image registration problems. In these approaches, the regularization prior is implicitly learned by training a CNN on a large database of domain-specific images. Depending on how the networks are trained, they can be categorized into supervised and unsupervised learning.

In supervised-learning methods, a CNN is trained using examples of medical images along with their ground truth transformations to predict the transformations directly on test images. Inspired by U-Net network [17] and using mesh segmentation, Rohé et al. [18] predicted the deformation field for 3D cardiac magnetic resonance imaging (MRI). Cao et al. [19] proposed a CNN to estimate the displacement vector for 3D brain MRI where equalized active points guided sampling and similarity between image patches were used to guide the learning process. Even though the accuracy of these approaches is considerable, their performance is highly dependent on the quality of the ground truth [20]. One of the most significant challenges in applying supervised methods to medical imaging applications is that the actual ground truth of a desired neural network output is not often available. The limitation of supervised methods has motivated the investigation of unsupervised methods [21] and dual/weakly supervised transformation estimation [22], which still requires manually produced labels or segmentation. Unsupervised learning-based image registration (DLIR) has received a lot of attention because it bypasses the need for expert annotated data of any kind. In other words, the quality of the output is not dependent on the quality of labels. Inspired by the spatial transformer network (STN), unsupervised approaches were proposed for deformable registration [23], [24], [25] which relied on normalized cross-correlation (NCC) and bending-energy regularization term to train a fully convolutional neural network (FCN). Balakrishnan et al. [26] used similarity to train a general framework for unsupervised image registration. Hoopes et al. in [27] learnt the effects of registration hyperparameters on the deformation field, which leverage a secondary network to generate the conditioned weights for the entire network layers. This method adds an enormous number of parameters to the original image registration method. Alternatively, a more parameter-efficient and scalable approach based on conditional instance normalization is proposed in [28].

Deformable image registration could also be performed in multiple stages and multiple resolutions to speed up and reduce sensitivity to local optima and image folding [29], [30]. Multi-stage approaches commonly start with affine or other linear registrations followed by coarse-to-fine deformable image registration [24], [31], [32].

Recently generative learning-based models such as a generative adversarial network (GAN), stochastic variational autoencoder and adversarial autoencoder have shown promising results in medical imaging applications in learning data distribution from large image training set [1]. Following the success of these models, some research groups [33] used unsupervised adversarial learning for image registration. Fan et al. [34] proposed a GAN to perform deformable image registration of 3D MRI. Mahapatra et al. [35] proposed simultaneous segmentation and registration of 4D computed tomography chest X-rays using a GAN framework. Krebs et al. [1] proposed a generative and probabilistic model

for diffeomorphic image registration which first trains an encoder-decoder neural network to estimate the deformation field by providing a large dataset of training images. A diffeomorphic transformation can be achieved via time-dependent velocity field [36]. Also, the ease of implementation and relatively low computational cost cause almost all existing end-to-end learning-based registration models to adopt stationary velocity fields [28], [36]. However, in this study, we demonstrated how time-dependent velocity fields can be efficiently incorporated into an end-to-end deep neural network framework, which results in improved registration performance.

Despite the relative success of learning-based methods, 1) Parameterizing registration as an optimization of a similarity metric alone is an ill-posed problem that requires explicit regularization [1]. 2) They generally need to be trained on a set of images which could introduce bias in a specific data domain. 3) The need for a large and varied set of training data is crucial for these algorithms to perform well on new subjects.

C. MOTIVATION AND CONTRIBUTION

In 2D cardiac MRI image sequences, the temporal consistency between subsequent frames can be exploited to obtain more accurate results than registering them pairwise. This study proposes a novel end-to-end iterative framework that processes the image sequences batch-wise to generate temporally consistent deformation fields. A learning-free generative multiresolution convolutional neural network (GMCNet) architecture is proposed in this study which does not require a dedicated set for the neural network training. To control the smoothness of the deformation field, in each resolution the deformation field is followed by a Gaussian filter which eliminates the need for an external regularization term. The proposed framework can be applied as pair-wise registration. The results show that utilizing temporal information leads to more accurate registration as well as a more robust performance under different noise levels.

The rest of the paper is organized as follows. In section III, we introduce our novel GMCNet framework, and the architecture of the proposed multiresolution framework and also discuss the Groupwise registration, loss function, and implementation detail. Datasets and evaluations are discussed in section IV. We discuss and conclude in sections V and VI.

III. METHODOLOGY

The main goal of deformable image registration is to find the optimal spatial deformation ϕ_θ parameterized by θ that warps a moving image I_M to align it with a fixed image I_F. The optimal values of θ could be obtained by minimizing an objective function of the form:

$$\theta^* = \underset{\theta}{\operatorname{argmin}}\, L(I_F, I_M \circ \phi_\theta)) + \lambda R(\phi_\theta) \tag{1}$$

where L is a metric that measures the dissimilarity between I_F and the warped moving image $I_M \circ \phi_\theta$. R is a regularization term that imposes smoothness on spatial deformation. λ is a weight coefficient corresponding to the regularization term. In unsupervised deep learning-based methods, a CNN is trained on a set of data to minimize a dissimilarity metric and a spatial transformer layer is used to warp the moving image I_M.

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \sum_{(I_F, I_M)} L(I_F, I_M \circ f_\theta(I_F, I_M)) + \lambda R(g_\theta(I_F, I_M)) \tag{2}$$

where f is the network and θ is the parameters of the network.

A. GENERATIVE MULTIRESOLUTION CONVOLUTIONAL NETWORK (GMCNet)

Inspired by [37], for each pair of moving and fixed images a latent variable z is introduced which is sampled from a random normal distribution and has the same dimensions as the input images $z \in \mathbb{R}^{H' \times W'}$. Where H and W are the height and width of images to be registered, respectively. The GMCNet maps the input latent variables to desired deformation fields and uses the corresponding batch of fixed and moving images for the loss function. Thus, during the optimization process, the network's parameters θ and z_i are optimized through back-propagation simultaneously.

So the new registration objective function can be formulated as follows:

$$\theta^*, z^* = \underset{\theta, z}{\operatorname{argmin}}\, L(I_F, I_M \circ f_\theta(z)), \tag{3}$$

where the θ^* and z^* are obtained using a neural network optimizer such as stochastic gradient descent. As can be seen in the new formulation there is no need for an external regularization term and a corresponding weight. Fig. 1 shows the optimization of an input latent variable z and the parameters of the network θ during the iterative processing for a sample pair of images. For the sake of simplicity, the optimization process is shown for one resolution framework.

The proposed architecture implements the general principle of the multiresolution framework, where computations follow coarse-to-fine resolutions [38]. It has three resolutions/scales denoted by a quarter $g^{1/4}_{\theta_1}$, a half $g^{1/2}_{\theta_1,\theta_2}$ and the original resolution, $g^{1}_{\theta_1,\theta_2,\theta_3}$. The $\theta_1, \theta_2, \theta_3$ are used to denote parameters of each resolution of GMCNet.

A spatial Gaussian smoothing kernel is integrated with the GMCNet to yield sufficiently smooth deformation fields. Such a strategy has been adopted in the Demons algorithm, where unconstrained optimization is followed by Gaussian filtering to impose a smoothness constraint. So that, at each resolution a Gaussian kernel is applied to the deformation field and then the smooth-scaled deformation field is upsampled and added to the next scale. The components of the proposed GMCNet architecture are shown in Fig. 2.

B. GROUPWISE REGISTRATION

Because of parameters sharing property, the GMCNet can be used as a groupwise registration framework when we have

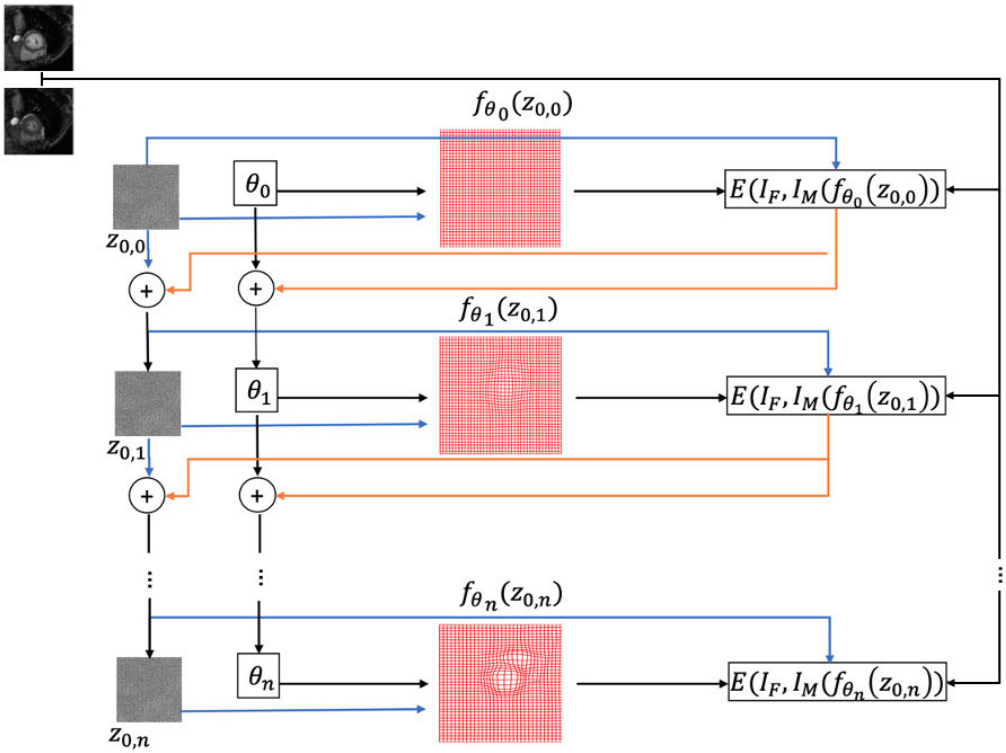


FIGURE 1. Iterations in the proposed GMCNet are shown for a pair of images for one resolution. The parameter set θ of the network f and latent variable z_0 for an image pair are iteratively updated through the back-propagation to minimize the registration cost (3) between the fixed image I_F and warped moving image I_M.

a sequence of images to be registered $\{I_i\}_{i=1}^N$. In groupwise image registration methods, several images are registered to a common coordinate system and therefore, a group of transformations has to be computed instead of a single geometric transformation [39]. A common practice for groupwise registration is to select the first (or any other) image as the reference image, and then register the rest of the images to the reference using pairwise image registration. The choice of the reference image is important in such an approach. In the proposed method each image in a cardiac sequence is registered to the next image in the sequence. To register a set of images $\{I_i\}_{i=1}^N$, the i^{th} image in a sequence is set as a moving image I_M^i and $i+1^{th}$ image is set as a fixed image I_F^i, for $i = 1, \ldots, N-1$ as shown in Fig. 3. Such a formulation eliminates any potential bias introduced by choosing a particular reference image. In our formulation, the weights θ are shared through the entire sequence. Such consistency may not be obtained using traditional pairwise approaches where the registration is performed by splitting the sequence into a set of image pairs. For a set of N image pairs $\{I_F^i, I_M^i\}_{i=1}^N$, N latent variables are defined $\{z_i\}_{i=1}^N$. Where N is the number of images in a sequence.

In the case of groupwise registration (3) is changed as follows:

$$\theta^*, \{z_i^*\}_{i=1}^N = \underset{\theta, \{z_i\}_{i=1}^N}{\operatorname{argmin}} \sum_{i=1}^{N} L(I_F^i, I_M^i(g_\theta(z_i))) \tag{4}$$

C. LOSS FUNCTION AND OPTIMIZATION

In each resolution a separate loss is defined to measure the dissimilarities between the fixed and warped moving images at that resolution as follows:

The loss function for the coarsest resolution:

$$Loss_{1/4} = \sum_{i=1}^{N} L(I_F^{1/4,i}, I_M^{1/4,i} \circ \phi_{\theta_1}^{1/4}(z_i)) \tag{5}$$

where $I_F^{1/4,i}$ and $I_M^{1/4,i}$ are respectively down-sampled fixed and moving images at the quarter resolution.

The loss function for the second (half) resolution:

$$Loss_{1/2} = \sum_{i=1}^{N} L(I_F^{1/2,i}, I_M^{1/2,i} \circ \phi_{\theta_1,\theta_2}^{1/2}(z_i)) \tag{6}$$

Similarly, the loss function for the full resolution is defined as:

$$Loss_1 = \sum_{i=1}^{N} L(I_F^{1,i}, I_M^{1,i} \circ \phi_{\theta_1,\theta_2,\theta_3}^{1}(z_i))) \tag{7}$$

The final loss function is a linear combination of loss functions in three resolutions:

$$Loss = \alpha Loss_{1/4} + \beta Loss_{1/2} + \gamma Loss_1, \tag{8}$$

We optimize the whole network as one stage to reduce computational time. All the parameters of the network $\theta_{1,2,3}$

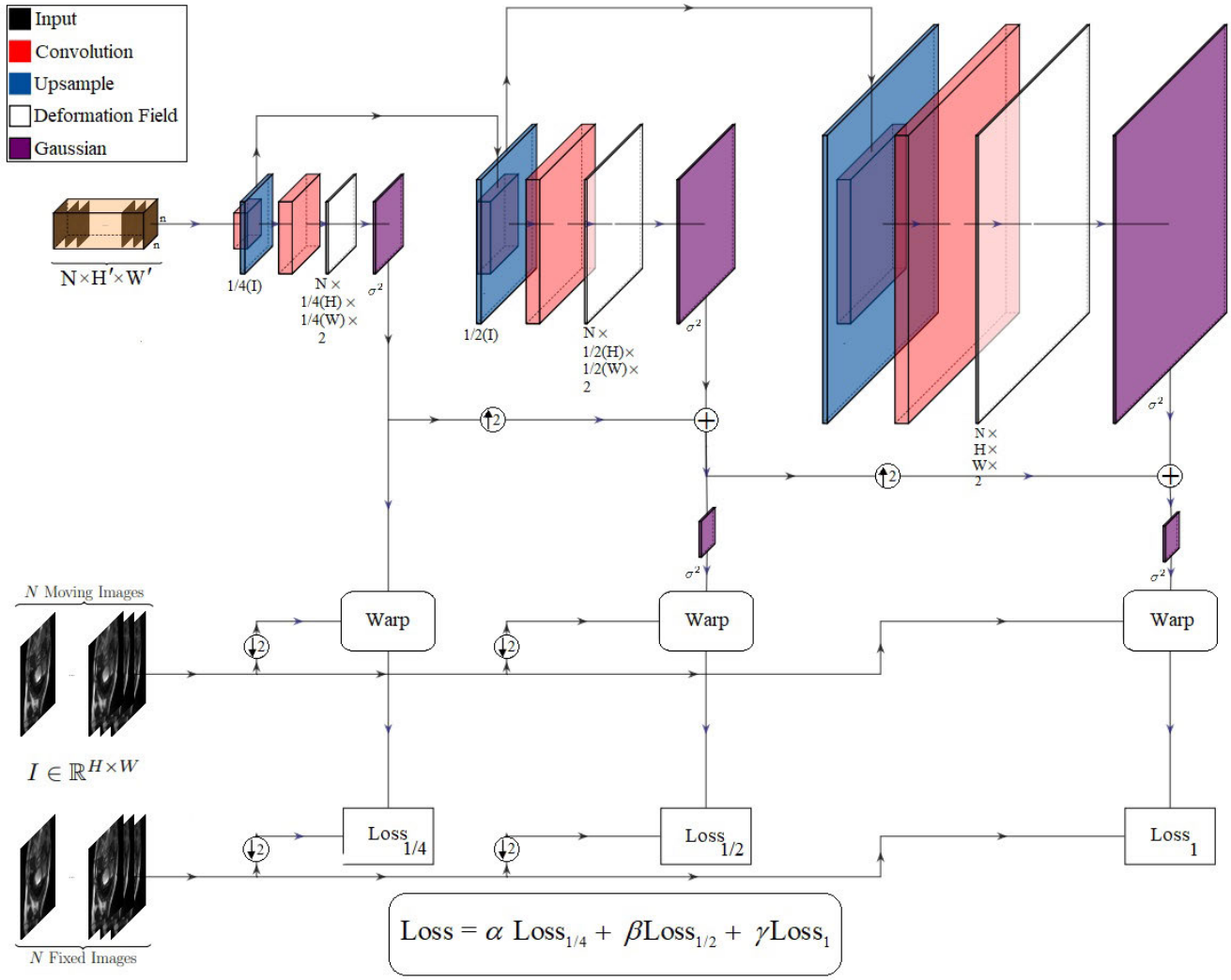


FIGURE 2. The proposed generative multiresolution convolutional network (GMCNet) architecture for deformable image registration. The network generates a deformation field based on low dimensional random inputs.

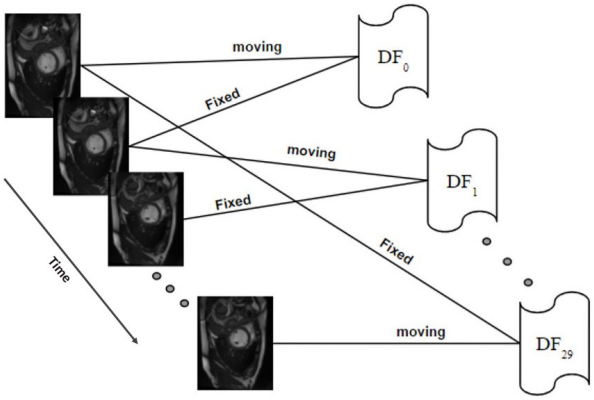


FIGURE 3. Illustration of choosing fixed and moving images in a sequence. The frame i is selected as a moving image and the next frame $i+1$ is selected as a fixed image.

are initialized only once and they are shared for all images in the sequence. It is possible to optimize the network in three phases so that the first stage optimizes it with respect to θ_1 and $\{z_i\}_{i=1}^N$, the second stage optimizes θ_1, θ_2 and $\{z_i\}_{i=1}^N$ and finally, the third stage optimizes loss (8) with regard to all parameters θ_1, θ_2, θ_3 and $\{z_i\}_{i=1}^N$.

D. LEARNING-FREE FRAMEWORK

The proposed GMCNet is learning-free and not trained on any given data sets. On this basis, to register any sequence/pair of images the z and the parameters of the network θ are optimized iteratively from scratch. The optimization is terminated based on two criteria, the step size and iteration number. In each iteration, the loss function is evaluated, and if it does not improve, the step size will be reduced. The process is terminated once the maximum number of iterations has been reached, or the step size is below the defined threshold.

E. IMPLEMENTATION DETAILS

The proposed GMCNet consists of six convolutional layers with strides (1, 1, 1, 1) and three upsample layers. The inputs

fit in a convolutional Gaussian layer with $\sigma = 2$ and kernel size 7. Each scale contains two convolutional layers, two upsample layers and a convolutional Gaussian layer with $\sigma = 3$ and kernel size 15 in front of the estimated deformation field and after adding a correction. Exponential linear unit (ELU) is selected as an activation function and NCC is used as a loss function. The values of α, β, and γ are chosen as 0.5, 0.25, and 0.25 respectively. We initialized a learning rate of 5×10^{-4} with the Adam optimizer and a batch size of 10. We used a grid search to find the maximum iteration number for all three data sets which is 800. The early stop method is also used, in case the network converges before 800 iterations. The minimum threshold of the step size is 0.5×10^{-7}. The framework has been implemented using Python programming language with the Tensorflow machine learning module and it is tested with an NVIDIA GTX 1080 Ti graphics processing unit with 11GB memory.

IV. EXPERIMENTS

The proposed framework is evaluated on clinical cardiac MRI images over three datasets.

A. DATASETS

The following three datasets are considered in this study:

1) AUTOMATED CARDIAC DIAGNOSIS CHALLENGE (ACDC) [40]

This dataset contains multiple temporal 2D short-axis cardiac cine MRI sequences acquired from 100 patients and is one of the publicly available datasets for cardiac MRI assessment. The spatial resolution varies from 1.37 to 1.68 mm^2/pixel with a slice thickness of 5 mm to 8 mm (in general 5mm) and sometimes an inter-slice gap of 5mm. Each sequence consists of 28 to 40 images that cover the cardiac cycle completely or partially.

2) THE SUNNYBROOK CARDIAC CHALLENGE DATA (SCD) [41]

This dataset contains multiple temporal 2D short-axis cardiac cine MRI scans acquired from 45 patients. Each cine sequence includes 20 frames to cover the cardiac cycle. The image resolution is 256×256, with a pixel spacing of 1.25 mm and slice thickness of 8 mm.

3) LEFT ATRIUM (LA)

This dataset includes 100 temporal 2D long-axis cine MRI steady-state sequences from the 2, 3 and 4-chamber views. It was acquired from the University of Alberta Hospital. Each cycle includes 25 or 30 frames with image resolutions 176×189 – 256×208 and image spacing 1.445 – 1.795 mm. The ground truth manual segmentation was initially performed by a medical student and edited by an experienced radiologist. The 2ch, 3ch and 4ch are used in the rest of the paper to denote 2, 3 and 4-chamber sequences, respectively.

B. EVALUATION

In this section, we evaluate and compare the performance of the proposed framework with both optimization-based state-of-the-art algorithms, SimpleElastix (Elastix) [42], Moving Mesh (MM) [16], Real-Time Image-based Tracker (RRT) [43], Fast Symmetric Forces Demons (Demons) [44], LCC-Demons [45], Symmetric Normalization [13] and also learning-based state-of-the-art algorithms VoxelMorph (VM) [46], learning probability model for diffeomorphic registration (LPM) [1] and DIRNet [23], a CNN-based, end-to-end unsupervised deformable image registration. We denote GMCNet as GMCNet_s and GMCNet_p to identify the sequential registration and pairwise registration.

1) QUANTITATIVE EVALUATION METRICS

The proposed method was evaluated quantitatively using four metrics, namely, Dice metric (DM), Hausdorff distance (HD in mm), determinant of Jacobian of the deformation field $\det(J)$, and reliability $R(d)$.

a: DICE METRIC

The DM [47] is a well-known segmentation based metric to measure the similarity (overlap) between two regions, warped moving and fixed image. The DM of two regions A and B is formulated as:

$$DM(A, B) = \frac{2|A \cap B|}{A + B} \tag{9}$$

b: HAUSDORFF DISTANCE

The HD [48] is another well-known metric which measures the maximum deviation between two regions' contours. The HD between two contours (C_A) and C_B is formulated as:

$$\begin{aligned} \mathrm{HD}(C_A, C_B) = \max(&\max_i(\min_j(d(p_A^i, p_B^j))), \\ &\max_j(\min_i(d(p_A^i, p_B^j)))) \end{aligned} \tag{10}$$

where p_A^i, p_B^j denote the set of all the points in C_A and C_B respectively. The term $d(\cdot)$ denotes the Euclidean distance. The DM and HD were computed by comparing the delineations obtained using the registration methods with expert manual contours. Table 1, 2 and 3 show the mean and standard deviations of DM and HD for all algorithms evaluated on LA, ACDC and SCD datasets, respectively. The reported values for methods indicated with $*$ were taken from previous publications [1], [23]. It should be noted that the presented method (GMCNet approach) has a good performance on all mentioned datasets (DM=$0.85 - 0.88$) and the reported values in Table 1, 2 and 3 demonstrate that our approach outperforms other registration methods in terms of both DM and HD.

c: RELIABILITY

We also evaluated the performance of the proposed algorithm using a reliability function computed based on DMs for each dataset. The complementary cumulative distribution function

TABLE 1. Quantitative evaluation of the results for cardiac MRI registration on the LA dataset. The evaluation is performed in terms of Dice (mean±standard deviation) and HD (mean). The 2ch, 3ch and 4ch stand for the 2, 3 and 4-chamber. Values in bold indicate the best performance. Undef stands for Undeformed.

Methods	Dice			HD		
	2ch	3ch	4ch	2ch	3ch	4ch
Undef	0.79 ± 0.07	0.78 ± 0.08	0.78 ± 0.09	7.37	7.70	8.66
FCG [37]	0.81 ± 0.11	0.84 ± 0.13	0.80 ± 0.10	7.5	7.67	8.2
RRT [43]	0.81 ± 0.13	0.85 ± 0.08	0.80 ± 0.11	7.3	6.99	8.01
Elastix [42]	0.82 ± 0.11	0.86 ± 0.10	0.82 ± 0.10	7.28	6.82	7.56
MM [16]	0.84 ± 0.06	0.83 ± 0.06	0.83 ± 0.08	6.58	6.48	6.77
Demons [44]	0.84 ± 0.08	0.85 ± 0.06	0.82 ± 0.10	7.41	7.33	7.84
SyN [13]	0.87 ± 0.06	0.86 ± 0.13	0.84 ± 0.11	6.92	7.52	7.51
GMCNet_p	0.81 ± 0.12	0.84 ± 0.09	0.81 ± 0.11	7.1	5.1	6.3
GMCNet_s	$\mathbf{0.88 \pm 0.05}$	$\mathbf{0.87 \pm 0.05}$	$\mathbf{0.85 \pm 0.07}$	**6.2**	**4.9**	**6.2**

TABLE 2. Quantitative evaluation of the results for cardiac MRI registration on the ACDC dataset. The evaluation is performed over the left ventricle in terms of Dice (mean± standard deviation) and HD (mean). Values in bold indicate the best performance.

Method	Dice	HD
Undeformed	0.71 ± 0.145	10.1
LCC-Demon* [45]	0.79 ± 0.096	9.21
VM* [46]	0.79 ± 0.096	8.46
LPM S1* [1]	0.79 ± 0.091	7.58
FCG [37]	0.80 ± 0.150	8.80
SyN* [13]	0.80 ± 0.091	8.24
LPM S3* [1]	0.81 ± 0.085	6.88
RRT [43]	0.83 ± 0.161	5.75
MM [16]	0.83 ± 0.153	5.64
Elastix [42]	0.84 ± 0.162	5.51
GMCNet_p	0.84 ± 0.170	7.43
GMCNet_s	$\mathbf{0.86 \pm 0.141}$	**5.35**

* Results are reported in [1] in which standard deviation could not be estimated. N/A refers to the value that are not reported in the original paper.

TABLE 3. Quantitative evaluation of the results for cardiac MRI registration on the SCD dataset. The evaluation is performed in terms of Dice (mean± standard deviation) and HD (mean). Values in bold indicate the best performance.

Method	Dice	HD
Undeformed	0.62 ± 0.15	16.02
RRT [43]	0.71 ± 0.19	13.10
FCG [37]	0.71 ± 0.2	12.82
Demons [44]	0.71 ± 0.18	12.46
MM [16]	0.72 ± 0.12	12.53
Elastix [42]	0.79 ± 0.08	11.12
DIRNet* [23]	0.80 ± 0.08	N/A
SyN [13]	0.81 ± 0.16	8.9
GMCNet_p	0.71 ± 0.16	7.95
GMCNet_s	$\mathbf{0.87 \pm 0.09}$	**4.58**

* Results are reported in [23]. N/A refers to the value that are not reported in the original paper.

is defined for each $d \in [0, 1]$ as the probability of obtaining DM higher than d over all volumes.

$$R(d) = P_r(Dice > d) = \frac{\#\text{ Images segmented with DM higher than d}}{\text{total number of images}}. \quad (11)$$

$R(d)$ measures how reliable the algorithm is in yielding accuracy d. The corresponding reliability $R(d)$ is plotted as a function of Dice in Fig. 4. The reliability values at $d = 0.80, 0.85, 0.90$ and 0.95 are reported in Table 4, 5, and 6 for ACDC, SCD and LA datasets, respectively. Our algorithm led to a higher reliability curve on all ACDC, SCD, and LA datasets. We have 2%, 17% and 2% improvement in $R(0.80)$, respectively on ACDC, SCD and LA. For instance, on ACDC, we obtained $R(0.85) = 0.65$, i.e, an excellent agreement ($DM > 0.85$) in 65% of the cases, whereas the Elastix [42] achieved 63% of the cases with similar accuracy. Also, we obtained $R(0.85) = 0.60$ on SCD, whereas the SyN [13] method, which has higher accuracy among other methods, achieved 43%.

d: det(J)

To analyze deformation regularity in different algorithms, we show the determinant of the Jacobian $\det(J)$ [49]. If the value of $\det(J)$ equals 1, the area remains constant after the transformation, whereas the value smaller or larger than 1 indicates the local area shrinkage or expansion, respectively. The negative values of $\det(J)$ imply that local folding and twisting have occurred, which is physically not realizable and mathematically not invertible [46]. The minimum of $\det(J)$ of

TABLE 4. Reliability function ($R(d) = Pr(DM > d)$) for ACDC. The higher the R, the better the performance.

Method	R(0.80)	R(0.85)	R(0.90)	R(0.95)
RRT [43]	0.73	0.60	0.40	0.13
MM [16]	0.73	0.61	0.42	0.16
Elastix [42]	0.74	0.63	0.45	0.20
GMCNet_s	**0.74**	**0.65**	**0.48**	**0.23**

TABLE 5. Reliability function ($R(d) = Pr(DM > d)$) for SCD. The higher the R, the better the performance.

Method	R(0.80)	R(0.85)	R(0.90)	R(0.95)
RRT [43]	0.45	0.31	0.16	0.04
Demons [44]	0.22	0.16	0.08	0.02
MM [16]	0.46	0.30	0.20	0.09
Elastix [42]	0.46	0.33	0.19	0.04
SyN [13]	0.55	0.48	0.20	0.09
GMCNet_s	**0.75**	**0.61**	**0.37**	**0.13**

TABLE 6. Reliability function ($R(d) = Pr(DM > d)$) for LA. The higher the R, the better the performance.

Method	R(0.80)	R(0.85)	R(0.90)	R(0.95)
LA-2ch				
RRT [43]	0.64	0.44	0.26	0.06
Demons [44]	0.75	0.51	0.24	0.07
MM [16]	0.78	0.50	0.17	0.04
Elastix [42]	0.60	0.44	0.22	0.08
SyN [13]	0.88	**0.74**	0.44	0.14
GMCNet_s	**0.90**	0.72	**0.45**	**0.18**
LA-3ch				
RRT [43]	0.81	0.62	0.37	0.12
Demons [44]	0.82	0.54	0.26	0.04
MM [16]	0.68	0.44	0.14	0.03
Elastix [42]	0.75	0.59	0.36	0.08
SyN [13]	0.87	0.66	**0.41**	0.11
GMCNet_s	**0.87**	**0.68**	0.40	**0.13**
LA-4ch				
RRT [43]	0.56	0.33	0.15	0.03
Demons [44]	0.59	0.39	0.26	0.09
MM [16]	0.45	0.31	0.16	0.04
Elastix [42]	0.50	0.28	0.13	0.02
SyN [13]	0.79	0.58	**0.35**	0.09
GMCNet_s	**0.79**	**0.59**	0.34	**0.10**

deformation for each pair of images is reported. We use the det(J) computed by SimpleITK [50], [51], [52] to quantify deformation regularity. Fig. 5 reports the minimum det(J) values observed in each method. Since the negative value of the determinant of Jacobian implies folding or twisting in the deformation field, only the minimum value of the determinant of Jacobian is reported. Each dot shows the minimum value of the determinant of Jacobian for each image sample. No negative minimum values of det(J) were observed for test cases which indicates that the proposed method does not lead to local mesh folding or twisting. This implies that the estimated deformations are physically realizable. A Sample of warped moving images and corresponding Jacobian determinant with grid overlay is shown in Fig. 8. The determinant of jacobian at each pixel is shown as a color map. Pixels with yellowish color are showing the constant area, blue pixels indicate the shrinkage area, and orange and red pixels indicate expansion.

2) IMPACT OF MULTI-RESOLUTION

We assess the impact of using different multi-resolution structures with GMCNet on the performance and reported the corresponding evaluations in terms of DM and HD in Table 7. The results indicate that high performance was obtained with the use of three resolutions.

3) ASSESSMENT OF ROBUSTNESS AGAINST DIFFERENT NOISE LEVELS

To assess the robustness of the proposed algorithm against different noise levels, we selected a random sequence from SCD and applied the same level of noise (speckle noise with $\mu = 1$ and $\sigma_s = [0, 0.5, \ldots, 4]$) to every frame in the sequence except for two randomly selected frames which were corrupted with twice as much noise level than other frames (standard deviation $= 2\sigma$). The DM of each algorithm was measured and plotted in Fig. 6. The results show that our method is almost independent of the noise level and at all noise level, we have the highest DM. This happens because the proposed method uses shared weights for all frames in a sequence, which capture the correlation between frames. This, per se, leads to a more robust performance under different levels of noise than non-learning-based methods.

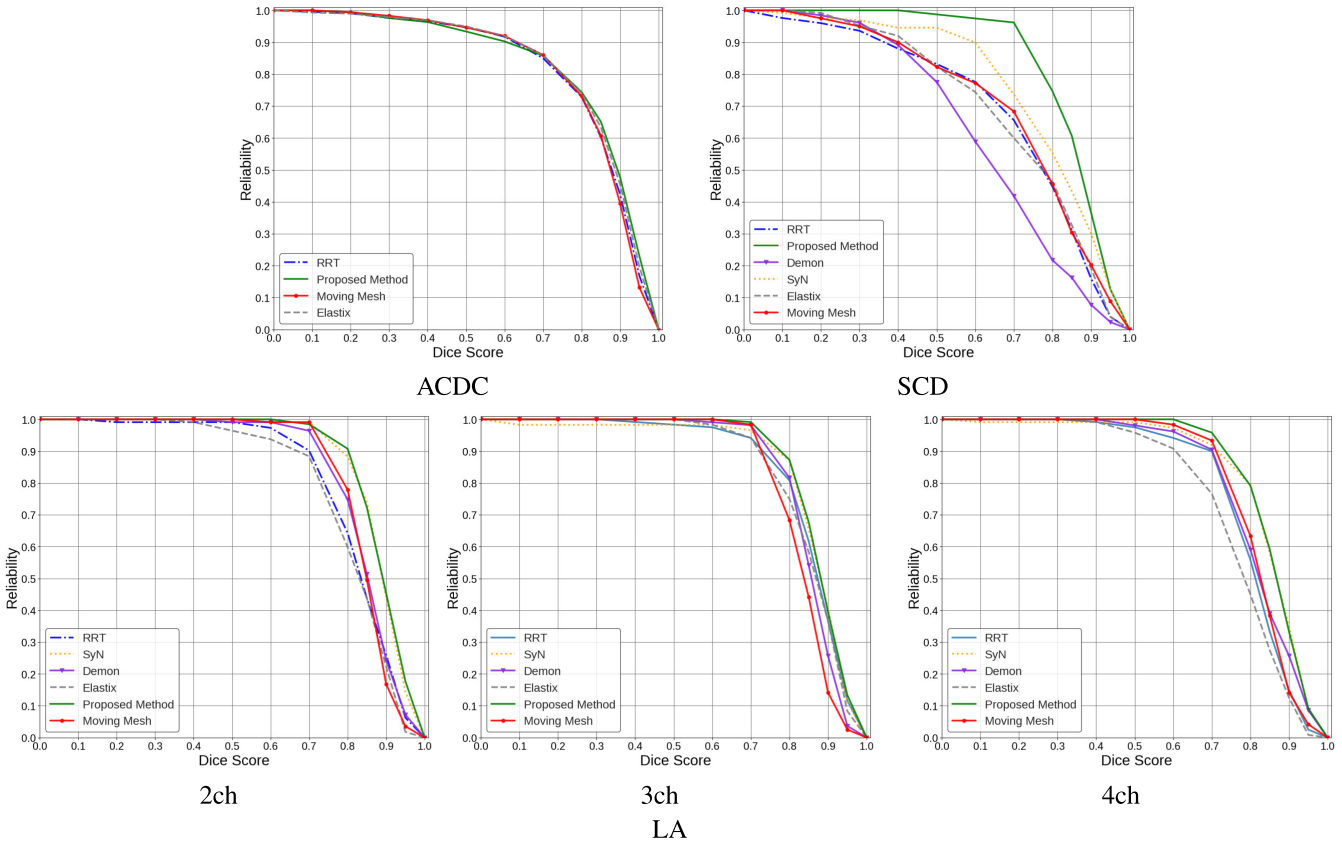


FIGURE 4. Reliability versus Dice metric of different algorithms: MM [16], RRT [43], Demons [46], SyN [13], SimpleElastix (Elastix) [42] and the proposed algorithm. The figures omit results for methods with no public implementation available. The corresponding reliability $R(d)$ is plotted as a function of Dice. As illustrated, our algorithm led to a higher reliability curve on all ACDC, SCD, and LA datasets.

TABLE 7. Quantitative cardiac MRI registration results on the ACDC, SCD and LA based on different number of resolution of GMCNet. The evaluation is performed in terms of Dice (mean±standard deviation) and HD (mean). The 2ch, 3ch and 4ch stand for the 2, 3 and 4-chamber. Values in bold indicate the best performance.

	Dataset	Dice	HD
One Resolution	ACDC	0.79 ± 0.15	8.01
	SCD	0.73 ± 0.09	9.43
	LA-2ch	0.80 ± 0.11	7.42
	LA-3ch	0.82 ± 0.08	7.27
	LA-4ch	0.79 ± 0.10	7.53
Two Resolutions	ACDC	0.81 ± 0.13	6.81
	SCD	0.77 ± 0.11	7.31
	LA-2ch	0.81 ± 0.09	7.11
	LA-3ch	0.84 ± 0.10	6.98
	LA-4ch	0.80 ± 0.09	7.01
Three Resolutions	ACDC	$\mathbf{0.86 \pm 0.14}$	**5.35**
	SCD	$\mathbf{0.87 \pm 0.09}$	**4.58**
	LA-2ch	$\mathbf{0.88 \pm 0.06}$	**6.2**
	LA-3ch	$\mathbf{0.87 \pm 0.05}$	**4.9**
	LA-4ch	$\mathbf{0.84 \pm 0.07}$	**6.2**
Four Resolutions	ACDC	0.75 ± 0.18	8.45
	SCD	0.85 ± 0.08	5.83
	LA-2ch	0.82 ± 0.09	7.3
	LA-3ch	0.85 ± 0.08	7.03
	LA-4ch	0.82 ± 0.10	6.83

Also, the drop rate for the other methods is higher than the proposed GMCNet method.

4) EVALUATION AGAINST GROUND TRUTH DEFORMATION BY THIN-PLATE SPLINE

To analyze the spatial transformation in a controlled environment, we used the thin-plate spline algorithm to generate a true deformation field [53]. The algorithm requires an initial contour at end-diastole which was obtained by randomly selecting patients from ACDC. First, the ground truth contour was obtained for the end-diastolic frame. Then, by using a diffeomorphic registration method a set of control points were generated. The contours include 20 equally spaced points which were subsampled from the generated sequence of control points. The thin-plate spline method was then used to produce a smooth interpolation between these sets of points. We defined the least bent surface that fits through the control points as follows:

$$f(x, y) = a_1 + a_2 x + a_3 y + \sum_{i=1}^{n} (w_i U(|P_i - (x, y)|), \quad (12)$$

where the first three terms $a_1 + a_2 x + a_3 y$ define the best-fitted plane through the control points and the $\sum_{i=1}^{n} (w_i U(|P_i - (x, y)|)$ term is correspondence to the bending forces provided by n control points.

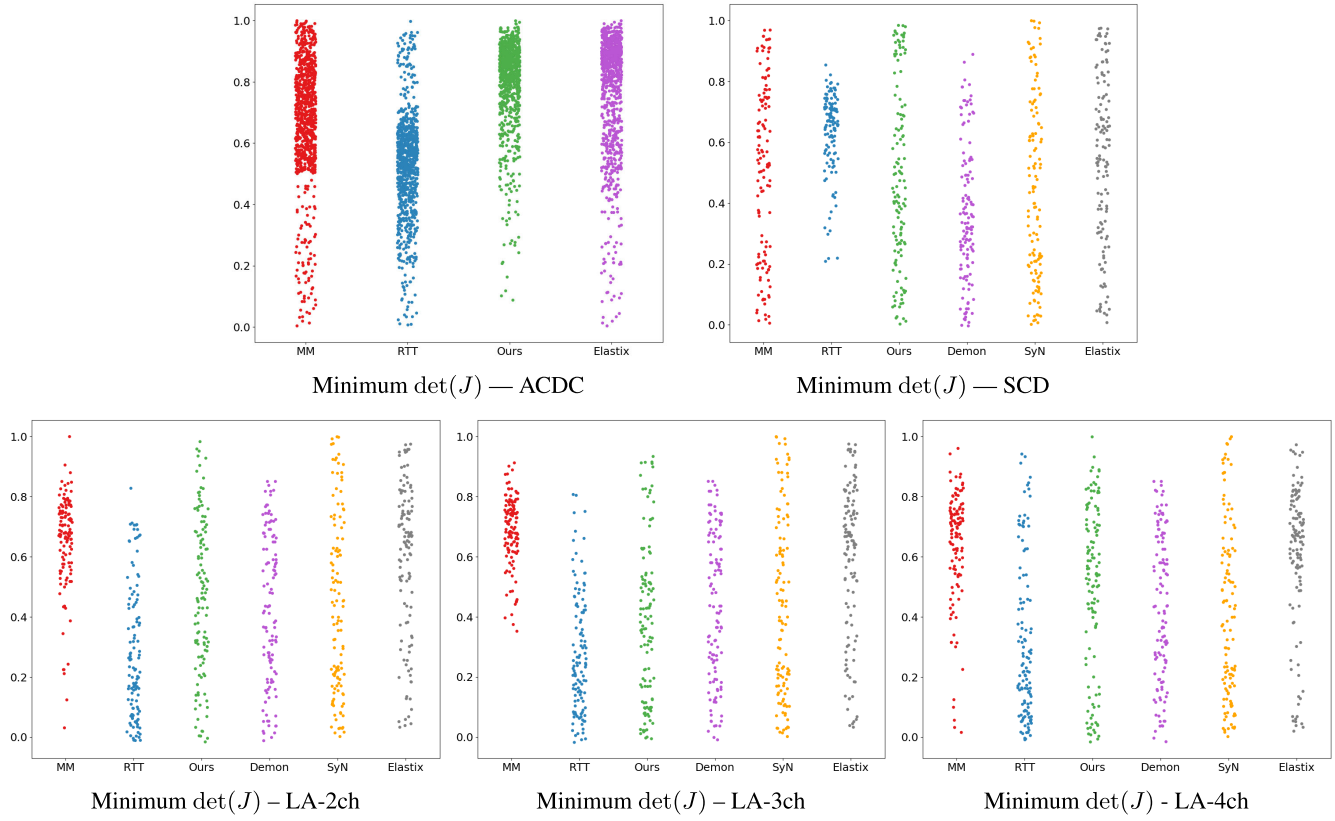


FIGURE 5. Minimum of determinant of Jacobian results for different algorithms: MM [16], RRT [43], Demons [46], SyN [13], SimpleElastix (Elastix) [42] and the proposed algorithm. The figures omit results for methods with no public implementation available. Since the negative value of the determinant of Jacobian implies folding or twisting in the deformation field, only the minimum value of the determinant of Jacobian is reported. Each dot shows the minimum value of the determinant of Jacobian for each image sample.

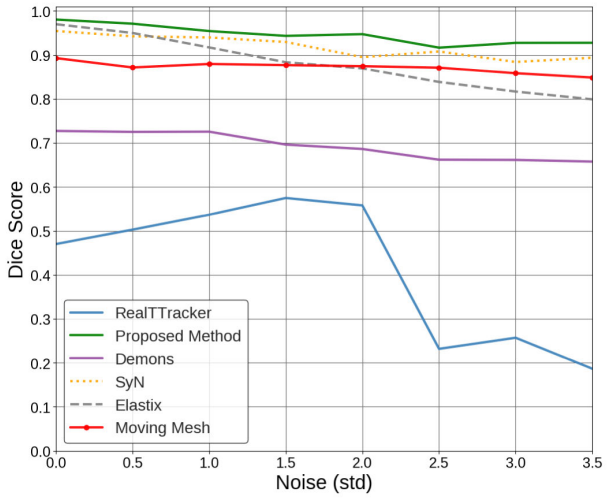


FIGURE 6. Dice scores (y-axis) for a randomly selected sequence from SCD corrupted with speckle noise with increasing standard deviations (x-axis). See details in the text.

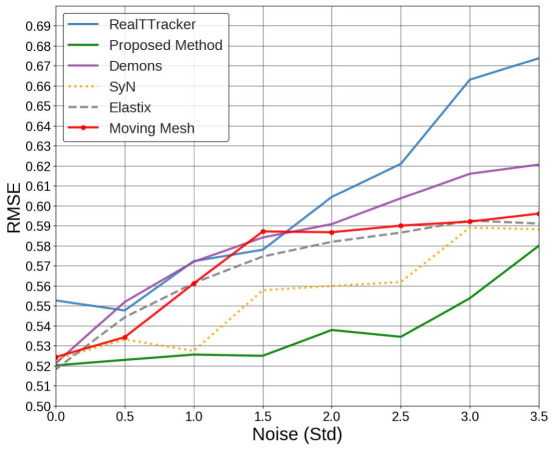


FIGURE 7. Root mean square error (RMSE) for quantitative evaluation of the accuracy between the ground truth deformation field and the six registration methods. The lower the values of RMSE, the more accurate the registration.

The deformation fields generated by the thin plate spline approach were used to verify and compare the performance of different parameterization approaches: RRT [43], Demons [44] and MM [16], Elastix [42], SyN [13], and the proposed GMCNet method.

The root mean squared error (RMSE) was calculated between the deformation fields for both directions (x, y) and summed. The RMSE is defined by:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2} \qquad (13)$$

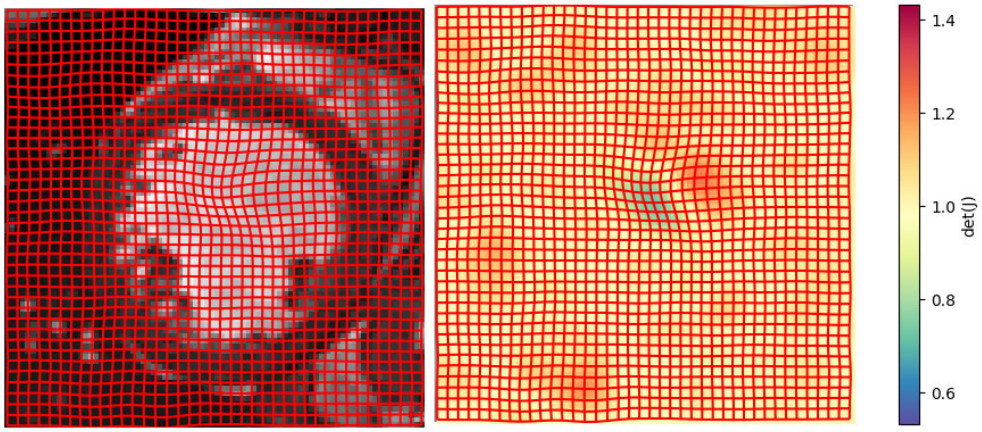


FIGURE 8. Registration results by the proposed GMCNet method showing a warped moving image with grid overlay and determinant of Jacobian. The determinant of jacobian at each pixel is shown as a color map. Pixels with yellowish color are showing the constant area, blue pixels indicate the shrinkage area, and orange and red pixels indicate expansion.

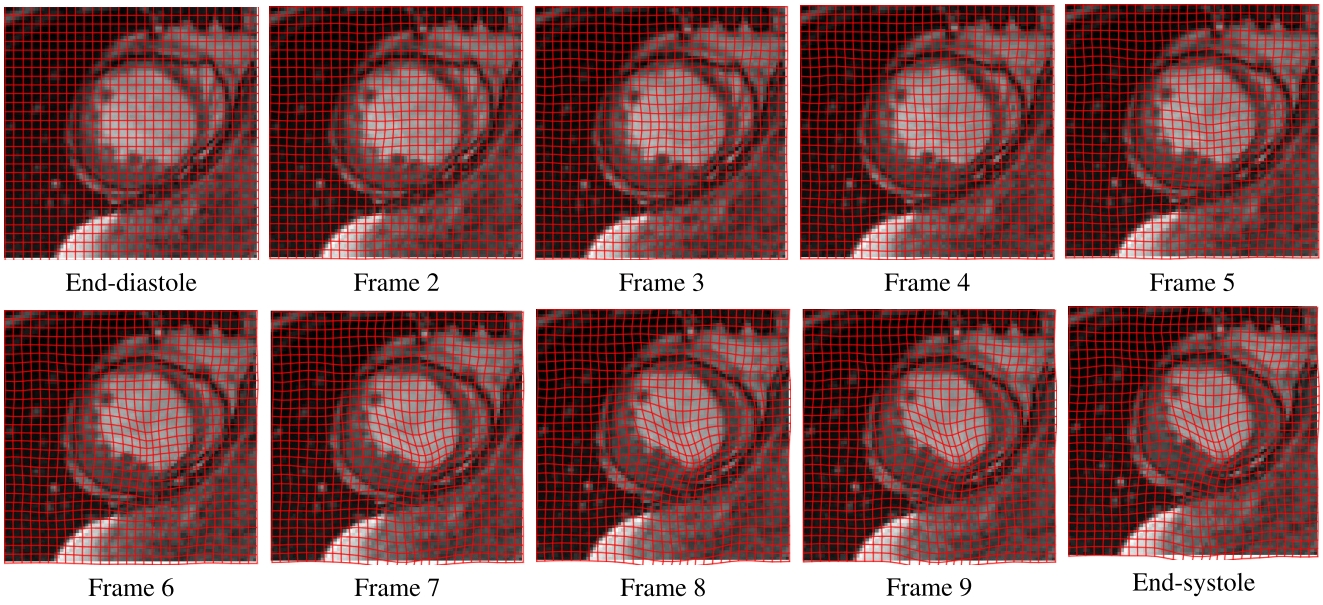


FIGURE 9. An example showing the predicted deformation results over the systolic phase of the cardiac cycle. It start from End-diastole (Frame 1) to End-systole which is 10th in this sequence. The grid deformation over the sequence shows the impact of the application of the smoothness. Each frame is registered to the next frame with a smooth transformation field computed using the proposed method.

where n is the total number of pixels in the deformation field and (x_i, y_i) and $(\hat{x}_i, \hat{y}_i)$ denote the true and estimated deformed points, respectively.

Fig. 7 displays the performance of the various registration algorithms compared to the ground truth thin plate spline method. These algorithms were evaluated by the difference in the deformation grids, RMSE, where images were corrupted with different degrees of noise. It can be seen that when there is no added noise, most of the methods yield similar RMSE in the range of $0.52 - 0.53$. However, the algorithm performance is highly dependent on the noise level and even a small increase in the value of noise would highly increase the RMSE. On the contrary, the presented method demonstrates less sensitivity to noise levels with the standard deviation values of $0.0 - 1.5$. Even for the greater noise levels with standard deviation values of $1.5 - 3.5$, our method still yields the least RMSE.

5) PAIRWISE REGISTRATION

To assess the effects of groupwise registration, we also applied our method as pairwise registration on images acquired at end-systolic and end-diastolic phases which have large deformation compared to neighbour frames in a sequence that has a small difference. Tables 1, 2 and 3 show the mean and standard deviations of DM and HD for pairwise

GMCNet_p, which is evaluated on ACDC, SCD and LA datasets, respectively. It can be seen from these tables that the proposed sequential GMCNet_s outperforms GMCNet_p in each tested case in terms of DM and HD.

V. DISCUSSION

In this study, we proposed a CNN based registration approach to obtain accurate results by exploiting temporal information from 2D image sequences. Registering a sequence of images plays an important role in many applications including cardiac functional assessment for MRI sequences. For instance, left ventricle dysfunction is a significant condition for adults and often requires the assessment of the regional function [54]. A point-to-point registration approach could be used for detecting regional left ventricular function abnormality [55]. In addition to the functional assessment of the left ventricle from short-axis MRI sequences, several other applications could benefit from image registration applied to 2D sequences. This includes functional assessment of left ventricle from short-axis MRI sequences [55], the functional assessments of left and right atria from long-axis MRI sequences [56], [57].

An example of the prediction results of the GMCNet method can be visualized in Fig. 9 where the images were acquired from a patient's MRI sequence. This example shows that the proposed method leads to deformation fields with no local folding or twisting. We evaluated the approach on cine-MRI registration and compared registration performance in terms of DICE and Hausdorff distances to nine popular learning and non-learning based algorithms, [1], [13], [16], [23], [42], [43], [44], [45], [46], and our conference version method [37]. The performance of our approach showed significant improvements in terms of registration accuracy.

The proposed method is applied as a sequential registration as well as pairwise registration for the datasets tested in this study. In sequential registration, all images were processed together, and therefore, we computed the equivalent time for pairwise registration by dividing the total time by the number of image pairs. In this case, the mean computational times were around 13, 10, and 17 seconds per image pair for ACDC, SCD, and LA datasets, respectively. In the case of pairwise registration, the mean computational time was around 25, 30, and 47 seconds for ACDC, SCD and LA datasets, respectively. One of the trade-offs of not having a dedicated training set is the additional time required for the neural net to converge at each inference which leads to increased computational complexity to obtain final results. However, the proposed method is suitable for medical applications that are not time-sensitive and priority is given to robustness and accuracy.

VI. CONCLUSION

We proposed a learning-free fully automated approach using the structure of an untrained generative multi-resolution convolutional neural network for deformable medical image registration. Optimizing the latent variables during the registration eliminates the need for regularization and tuning. The proposed method has yielded promising results on cardiac MRI images in comparison to the learning and non-learning based methods. Moreover, the proposed method's prediction is not domain-specific and could be applied to any medical image sequence without resorting to annotated training data. The proposed method was evaluated using a left atrial dataset and yielded an average Dice score of 0.88, 0.87, and 0.85 for 2, 3 and 4 chamber sequences, respectively. It also yielded an average Dice score of 0.86 and 0.87 for the left ventricular segmentation for the ACDC and Sunnybrook datasets, respectively. Our algorithm captures the correlation between frames in a sequence which leads to a more robust performance under different levels of noise. Currently, the presented methodology is applied to the temporal 2D registration problem. In the future, we will extend the proposed method for temporal 3D sequences.

ACKNOWLEDGMENT

(Nilanjan Ray and Kumaradevan Punithakumar are co-first authors.)

REFERENCES

[1] J. Krebs, H. Delingette, B. Mailhe, N. Ayache, and T. Mansi, "Learning a probabilistic model for diffeomorphic registration," *IEEE Trans. Med. Imag.*, vol. 38, no. 9, pp. 2165–2176, Sep. 2019.
[2] G. Haskins, U. Kruger, and P. Yan, "Deep learning in medical image registration: A survey," 2019, *arXiv:1903.02026*.
[3] A. Sheikhjafari, H. A. Talebi, and M. Zareinejad, "3D visual stabilization for robotic-assisted beating heart surgery using a thin-plate spline deformable model," in *Proc. 3rd RSI Int. Conf. Robot. Mechatronics (ICROM)*, Oct. 2015, pp. 743–748.
[4] A. Sheikhjafari, H. A. Talebi, and M. Zareinejad, "Robust and efficient 3D motion tracking in robotic assisted beating heart surgery," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2015, pp. 1828–1833.
[5] A. Sheikhjafari, D. Krishnaswamy, M. Noga, N. Ray, and K. Punithakumar, "Deep learning based parametrization of diffeomorphic image registration for the application of cardiac image segmentation," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2022, pp. 1164–1169.
[6] A. V. Dalca, A. Bobu, N. S. Rost, and P. Golland, "Patch-based discrete registration of clinical brain images," in *Proc. Int. Workshop Patch-Based Techn. Med. Imag.* Cham, Switzerland: Springer, 2016, pp. 60–67.
[7] B. Glocker, N. Komodakis, G. Tziritas, N. Navab, and N. Paragios, "Dense image registration through MRFs and efficient linear programming," *Med. Image Anal.*, vol. 12, no. 6, pp. 731–741, 2008.
[8] J.-P. Thirion, "Image matching as a diffusion process: An analogy with Maxwell's demons," *Med. Image Anal.*, vol. 2, no. 3, pp. 243–260, 1998.
[9] P. Cachier and X. Pennec, "3D non-rigid registration by gradient descent on a Gaussian-windowed similarity measure using convolutions," in *Proc. IEEE Workshop Math. Methods Biomed. Image Anal.*, Jun. 2000, pp. 182–189.
[10] C. Le Guyader and L. A. Vese, "A combined segmentation and registration framework with a nonlinear elasticity smoother," *Comput. Vis. Image Understand.*, vol. 115, no. 12, pp. 1689–1709, Dec. 2011.
[11] T. Mansi, X. Pennec, M. Sermesant, H. Delingette, and N. Ayache, "iLogDemons: A demons-based registration algorithm for tracking incompressible elastic biological tissues," *Int. J. Comput. Vis.*, vol. 92, no. 1, pp. 92–111, 2011.
[12] M. Zhang and P. T. Fletcher, "Finite-dimensional lie algebras for fast diffeomorphic image registration," in *Proc. Int. Conf. Inf. Process. Med. Imag.* Cham, Switzerland: Springer, 2015, pp. 249–260.
[13] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Med. Image Anal.*, vol. 12, no. 1, pp. 26–41, Feb. 2008.

[14] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Symmetric log-domain diffeomorphic registration: A demons-based approach," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2008, pp. 754–761.

[15] K. Punithakumar, M. Noga, I. B. Ayed, and P. Boulanger, "Right ventricular segmentation in cardiac MRI with moving mesh correspondences," *Computerized Med. Imag. Graph.*, vol. 43, pp. 15–25, Jul. 2015.

[16] K. Punithakumar, P. Boulanger, and M. Noga, "A GPU-accelerated deformable image registration algorithm with applications to right ventricular segmentation," *IEEE Access*, vol. 5, pp. 20374–20382, 2017.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[18] M.-M. Rohé, M. Datar, T. Heimann, M. Sermesant, and X. Pennec, "SVF-Net: Learning deformable image registration using shape matching," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2017, pp. 266–274.

[19] X. Cao, J. Yang, J. Zhang, D. Nie, M. Kim, Q. Wang, and D. Shen, "Deformable image registration based on similarity-steered CNN regression," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2017, pp. 300–308.

[20] Y. Sang, X. Xing, Y. Wu, and D. Ruan, "Imposing implicit feasibility constraints on deformable image registration using a statistical generative model," *Proc. SPIE*, vol. 11313, Dec. 2020, Art. no. 113132V.

[21] C. Doersch, A. Gupta, and A. A. Efros, "Unsupervised visual representation learning by context prediction," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1422–1430.

[22] M. P. Heinrich, "Closing the gap between deep and conventional image registration using probabilistic dense displacement networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 50–58.

[23] B. D. de Vos, F. F. Berendsen, M. A. Viergever, M. Staring, and I. Išgum, "End-to-end unsupervised deformable image registration with a convolutional neural network," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2017, pp. 204–212.

[24] B. D. de Vos, F. F. Berendsen, M. A. Viergever, H. Sokooti, M. Staring, and I. Išgum, "A deep learning framework for unsupervised affine and deformable image registration," *Med. Image. Anal.*, vol. 52, pp. 128–143, Feb. 2019.

[25] A. Sheikhjafari, M. Noga, K. Punithakumar, and N. Ray, "A training-free recursive multiresolution framework for diffeomorphic deformable image registration," *Appl. Intell.*, vol. 2022, pp. 1–10, Feb. 2022.

[26] G. Balakrishnan, A. Zhao, M. R. Sabuncu, A. V. Dalca, and J. Guttag, "An unsupervised learning model for deformable medical image registration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9252–9260.

[27] A. Hoopes, M. Hoffmann, B. Fischl, J. Guttag, and A. V. Dalca, "HyperMorph: Amortized hyperparameter learning for image registration," in *Proc. Int. Conf. Inf. Process. Med. Imag.* Cham, Switzerland: Springer, 2021, pp. 3–17.

[28] T. C. Mok and A. Chung, "Conditional deformable image registration with convolutional neural network," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2021, pp. 35–45.

[29] J. A. Schnabel, D. Rueckert, M. Quist, J. M. Blackall, A. D. Castellano-Smith, T. Hartkens, and G. P. Penney, "A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Berlin, Germany: Springer, 2001, pp. 573–581.

[30] E. I. Zacharaki, D. Shen, S.-K. Lee, and C. Davatzikos, "ORBIT: A multiresolution framework for deformable registration of brain tumor images," *IEEE Trans. Med. Imag.*, vol. 27, no. 8, pp. 1003–1017, Aug. 2008.

[31] Q. Xing, P. Chitnis, S. Sikdar, J. Alshiek, S. A. Shobeiri, and Q. Wei, "M3 VR—A multi-stage, multi-resolution, and multi-volumes-of-interest volume registration method applied to 3D endovaginal ultrasound," *PLoS ONE*, vol. 14, no. 11, Nov. 2019, Art. no. e0224583.

[32] X. Zhou, J. Wang, and G. Gong, "Two stage registration-based automatic left ventricle myocardium segmentation of cardiac 4DCT images," in *Proc. 11th Int. Conf. Graph. Image Process. (ICGIP)*, Jan. 2020, Art. no. 113731.

[33] D. Mahapatra, B. Antony, S. Sedai, and R. Garnavi, "Deformable medical image registration using generative adversarial networks," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 1449–1453.

[34] J. Fan, X. Cao, Z. Xue, P.-T. Yap, and D. Shen, "Adversarial similarity network for evaluating image alignment in deep learning based registration," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 739–746.

[35] D. Mahapatra, Z. Ge, S. Sedai, and R. Chakravorty, "Joint registration and segmentation of Xray images using generative adversarial networks," in *Proc. Int. Workshop Mach. Learn. Med. Imag.* Cham, Switzerland: Springer, 2018, pp. 73–80.

[36] J. Chen, E. C. Frey, and Y. Du, "Unsupervised learning of diffeomorphic image registration via TransMorph," in *Proc. Int. Workshop Biomed. Image Registration*. Cham, Switzerland: Springer, 2022, pp. 96–102.

[37] A. Sheikhjafari, M. Noga, K. Punithakumar, and N. Ray, "Unsupervised deformable image registration with fully connected generative neural network," in *Medical Imaging with Deep Learning*, 2018.

[38] N. Ray, "Computation of fluid and particle motion from a time-sequenced image pair: A global outlier identification approach," *IEEE Trans. Image Process.*, vol. 20, no. 10, pp. 2925–2936, Oct. 2011.

[39] C. Twining, T. Cootes, S. Marsland, V. Petrovic, R. Schestowitz, and C. Taylor, "Information-theoretic unification of groupwise non-rigid registration and model building," in *Medical Image Understanding and Analysis*, vol. 2, 2006, pp. 226–230.

[40] O. Bernard et al., "Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved?" *IEEE Trans. Med. Imag.*, vol. 37, no. 11, pp. 2514–2525, Nov. 2018.

[41] P. Radau, Y. Lu, K. Connelly, G. Paul, A. J. Dick, and G. A. Wright, "Evaluation framework for algorithms segmenting short axis cardiac MRI," *MIDAS J.*, vol. 49, pp. 1–10, Jul. 2009.

[42] K. Marstal, F. Berendsen, M. Staring, and S. Klein, "SimpleElastix: A user-friendly, multi-lingual library for medical image registration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2016, pp. 134–142.

[43] C. Zachiu, N. Papadakis, M. Ries, C. Moonen, and B. D. de Senneville, "An improved optical flow tracking technique for real-time MR-guided beam therapies in moving organs," *Phys. Med. Biol.*, vol. 60, no. 23, p. 9003, 2015.

[44] M. McCormick, X. Liu, J. Jomier, C. Marion, and L. Ibanez, "ITK: Enabling reproducible research and open science," *Frontiers Neuroinform.*, vol. 8, p. 13, Jan. 2014.

[45] M. Lorenzi, N. Ayache, G. B. Frisoni, and X. Pennec, "LCC-Demons: A robust and accurate symmetric diffeomorphic registration algorithm," *NeuroImage*, vol. 81, pp. 470–483, Nov. 2013.

[46] A. V. Dalca, G. Balakrishnan, J. Guttag, and M. R. Sabuncu, "Unsupervised learning for fast probabilistic diffeomorphic registration," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 729–738.

[47] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[48] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 9, pp. 850–863, Sep. 1993.

[49] J. Ashburner, J. L. R. Andersson, and K. J. Friston, "High-dimensional image registration using symmetric priors," *NeuroImage*, vol. 9, no. 6, pp. 619–628, Jun. 1999.

[50] R. Beare, B. Lowekamp, and Z. Yaniv, "Image segmentation, registration and characterization in R with SimpleITK," *J. Stat. Softw.*, vol. 86, no. 8, pp. 1–35, 2018.

[51] Z. Yaniv, B. C. Lowekamp, H. J. Johnson, and R. Beare, "SimpleITK image-analysis notebooks: A collaborative environment for education and reproducible research," *J. Digit. Imag.*, vol. 31, no. 3, pp. 290–303, Jun. 2018.

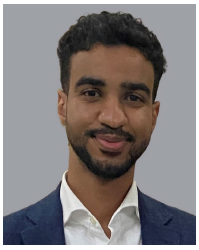
[52] B. C. Lowekamp, D. T. Chen, L. Ibáñez, and D. Blezek, "The design of SimpleITK," *Front. Neuroinform.*, vol. 7, p. 45, Dec. 2013.

[53] D. Krishnaswamy, M. Noga, and K. Punithakumar, "Validation of a diffeomorphic registration algorithm using true deformation computed from thin plate spline interpolation," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 1351–1354.

[54] N. J. Weissman, V. Dilsizian, A. K. Jacobs, S. Kaul, W. K. Laskey, D. J. Pennell, J. A. Rumberger, T. Ryan, and M. S. Verani, "Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: A statement for healthcare professionals from the cardiac imaging committee of the council on clinical cardiology of the American heart association," *Circulation*, vol. 105, no. 4, pp. 539–542, 2002.
[55] K. Punithakumar, I. B. Ayed, A. Islam, A. Goela, I. G. Ross, J. Chong, and S. Li, "Regional heart motion abnormality detection: An information theoretic approach," *Med. Image Anal.*, vol. 17, no. 3, pp. 311–324, Apr. 2013.
[56] X. Zhang, M. Noga, D. G. Martin, and K. Punithakumar, "Fully automated left atrium segmentation from anatomical cine long-axis MRI sequences using deep convolutional neural network with unscented Kalman filter," *Med. Image Anal.*, vol. 68, Feb. 2021, Art. no. 101916.
[57] M. Regehr, A. Volk, M. Noga, and K. Punithakumar, "Machine learning and graph based approach to automatic right atrial segmentation from magnetic resonance imaging," in *Proc. IEEE 17th Int. Symp. Biomed. Imag. (ISBI)*, Iowa City, IA, USA, Apr. 2020, pp. 826–829.

AMENEH SHEIKHJAFARI received the B.S. degree in computer engineering and the M.Sc. degree in robotics engineering from the Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, in 2011 and 2015, respectively. She is currently pursuing the Ph.D. degree with the Department of Computing Science and the Department of Radiology and Diagnostic Imaging, University of Alberta, Edmonton, Canada, under the supervision of Dr. Nilanjan Ray and Dr. Kumaradevan Punithakumar. She was a Co-Op/Computer Vision Scientist at Neurosoph Extend Intelligence, Edmonton, in 2019. From July 2021 to February 2022, she was a Co-Op/Deep Learning Specialist at Vivid Machines Inc., Toronto, Canada. Her research interests include medical image analysis, computer vision, and deep learning.

MICHELLE NOGA is currently a Radiologist trained with the University of Alberta and British Columbia's Children's and Women's Hospital. She is also a Professor with the Department of Radiology and Diagnostic Imaging, University of Alberta, and a Radiologist with medical imaging consultants. She established the first cardiac MRI program at Alberta, and has successfully established the first center for virtual 3D display of cardiac imaging in a clinical setting, a product of a $1,000,000 Grant from Servier Canada. She has published over 80 publications and conference proceedings and holds three patents. Her research interests include pediatric cardiac MRI and CT, post-processing of cross-sectional cardiac imaging, 3D visualization, rapid prototyping, pediatric airway, and finite element analysis.



AHMED AHMED is currently pursuing the bachelor's degree in computer engineering with the University of Alberta. In 2019, he joined the Servier Virtual Cardiac Centre, where he worked on medical image analysis. In 2020, he was a Teaching Assistant at the Faculty of Computing Science, University of Alberta, and since then he has been exploring a variety of different fields within industry. His research interests include medical image analysis, electronics, and telecommunications.

NILANJAN RAY received the Bachelor of Mechanical Engineering degree from Jadavpur University, India, in 1995, the M.Tech. degree in computer science from the Indian Statistical Institute, India, in 1997, and the Ph.D. degree in electrical engineering from the University of Virginia, USA, in 2003. He is currently a Professor of computing science with the University of Alberta, Canada. His research interests include computer vision and medical image analysis. He published more than 125 peer-reviewed articles in these areas. He served for the AI-GI-CRV 2017 as the Co-Chair and BMVC 2022 as a PC Member. He served as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING, from 2013 to 2017, and *IET Image Processing*, from 2016 to 2021.

KUMARADEVAN PUNITHAKUMAR (Senior Member, IEEE) received the B.Sc.Eng. degree (Hons.) in electronic and telecommunication engineering from the University of Moratuwa, and the M.A.Sc. and Ph.D. degrees in electrical and computer engineering from McMaster University.

From 2001 to 2002, he was an Instructor at the Department of Electronic and Telecommunication Engineering, University of Moratuwa. From 2002 to 2007, he was a Teaching Assistant. In 2008, he was a Postdoctoral Research Fellow at the Department of Electrical and Computer Engineering, McMaster University. From 2008 to 2012, he was an Imaging Research Scientist at GE Healthcare, Canada. He is currently an Associate Professor with the Department of Radiology and Diagnostic Imaging, University of Alberta, and an Operational and Computational Director of the Servier Virtual Cardiac Centre, Mazankowski Alberta Heart Institute. His research interests include medical image analysis and visualization, information fusion, object tracking, and nonlinear filtering.

Dr. Punithakumar was a recipient of the Industrial Research and Development Fellowship by the National Sciences and Engineering Research Council of Canada, in 2008.