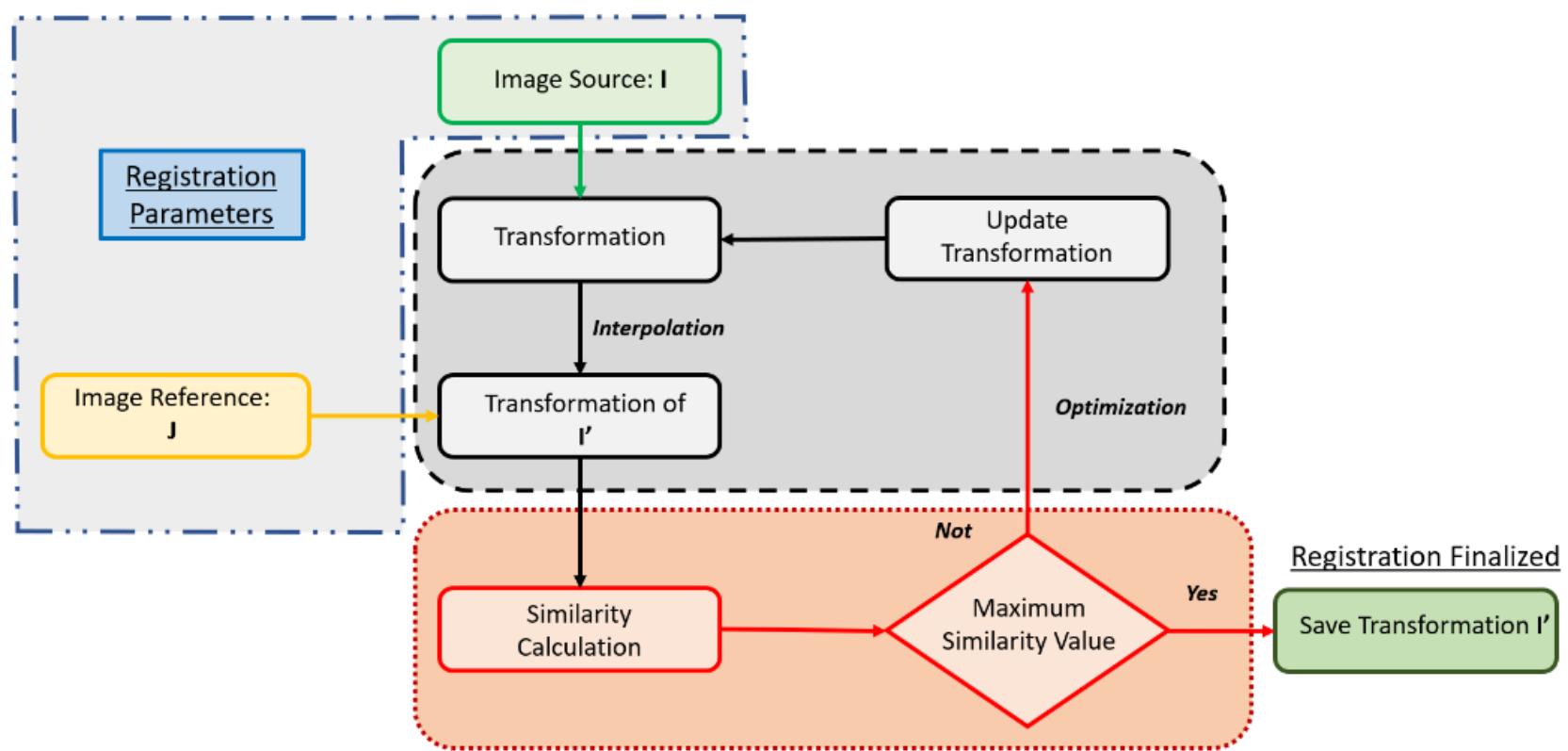


Deep Learning Approaches – Part 2

- Two categories:
 - Similarity Metrics
 - **Transformation Parameters**



Deep Learning - Transformation Parameters

(Miao et al. 2016)

A CNN Regression Approach for Real-time 2D/3D Registration

Shun Miao, *Member, IEEE*, Z. Jane Wang, *Senior Member, IEEE*, and Rui Liao, *Senior Member, IEEE*

Abstract—In this paper, we present a Convolutional Neural Network (CNN) regression approach to address the two major limitations of existing intensity-based 2-D/3-D registration technology: 1) slow computation and 2) small capture range. Different from optimization-based methods, which iteratively optimize the transformation parameters over a scalar-valued metric function representing the quality of the registration, the proposed method exploits the information embedded in the appearances of the digitally reconstructed radiograph and X-ray images, and employs CNN regressors to directly estimate the transformation parameters. An automatic feature extraction step is introduced to calculate 3-D pose-indexed features that are sensitive to the variables to be regressed while robust to other factors. The CNN regressors are then trained for local zones and applied in a hierarchical manner to break down the complex regression task into multiple simpler sub-tasks that can be learned separately. Weight sharing is furthermore employed in the CNN regression model to reduce the memory footprint. The proposed approach has been quantitatively evaluated on 3 potential clinical applications, demonstrating its significant advantage in providing highly accurate real-time 2-D/3-D registration with a significantly enlarged capture range when compared to intensity-based methods.

Index Terms—2-D/3-D Registration, Image Guided Intervention, Convolutional Neural Network, Deep Learning

I. INTRODUCTION

2-D/3-D registration represents one of the key enabling technologies in medical imaging and image-guided interventions [1]. It can bring the pre-operative 3-D data and intra-operative 2-D data into the same coordinate system, to facilitate accurate diagnosis and/or provide advanced image guidance. The pre-operative 3-D data generally includes Computed Tomography (CT), Cone-beam CT (CBCT), Magnetic Resonance Imaging (MRI) and Computer Aided Design (CAD) model of medical devices, while the intra-operative 2-D data is dominantly X-ray images. In this paper, we focus on registering a 3-D X-ray attenuation map provided by CT or CBCT with a 2-D X-ray image in real-time. Depending on the application, other 3-D modalities (e.g., MRI and CAD

model) can be converted to a 3-D X-ray attenuation map before performing 2-D/3-D registration.

In existing methods, accurate 2-D/3-D registration is typically achieved by intensity-based 2-D/3-D registration methods [2][3][4][5]. In these methods, a simulated X-ray image, referred to as Digitally Reconstructed Radiograph (DRR), is derived from the 3-D X-ray attenuation map by simulating the attenuation of virtual X-rays. An optimizer is employed to maximize an intensity-based similarity measure between the DRR and X-ray images. Intensity-based methods are known to be able to achieve high registration accuracy [6], but at the same time, they suffer from two major drawbacks: 1) long computation time and 2) small capture range. Specifically, because intensity-based methods involve a large number of evaluations of the similarity measure, each requiring heavy computation in rendering the DRR, they typically resulted in above 1 s running time, and therefore are not suitable for real-time applications. In addition, because the similarity measures to be optimized in intensity-based methods are often highly non-convex, the optimizer has a high chance of getting trapped into local maxima, which leads to a small capture range of these methods.

The small capture range of intensity-based methods is often addressed by employing initialization methods before registration [7][8]. However, initialization methods typically utilize dominant features of the target object for pose recovery and therefore are very application specific. For example, Varnavas *et al.* [7] applied Generalized Hough Transform (GHT) for initial pose estimation of spine vertebrae. This method is specific to applications where spine vertebrae edges are clearly visible in both the X-ray and CT images. Miao *et al.* [8] proposed to use shape encoding combined with template matching for initial pose estimation. This method can only be applied on metal implants, which are highly X-ray opaque objects that can be reliably segmented from X-ray images for shape encoding.

Some efforts have been made toward accelerating DRR generation for fast 2D/3D registration. One strategy for faster DRR generation is sparse sampling, where a subset of the pixels are statistically chosen for DRR rendering and similarity measure calculation [9][10]. However, only a few similarity

Proposal

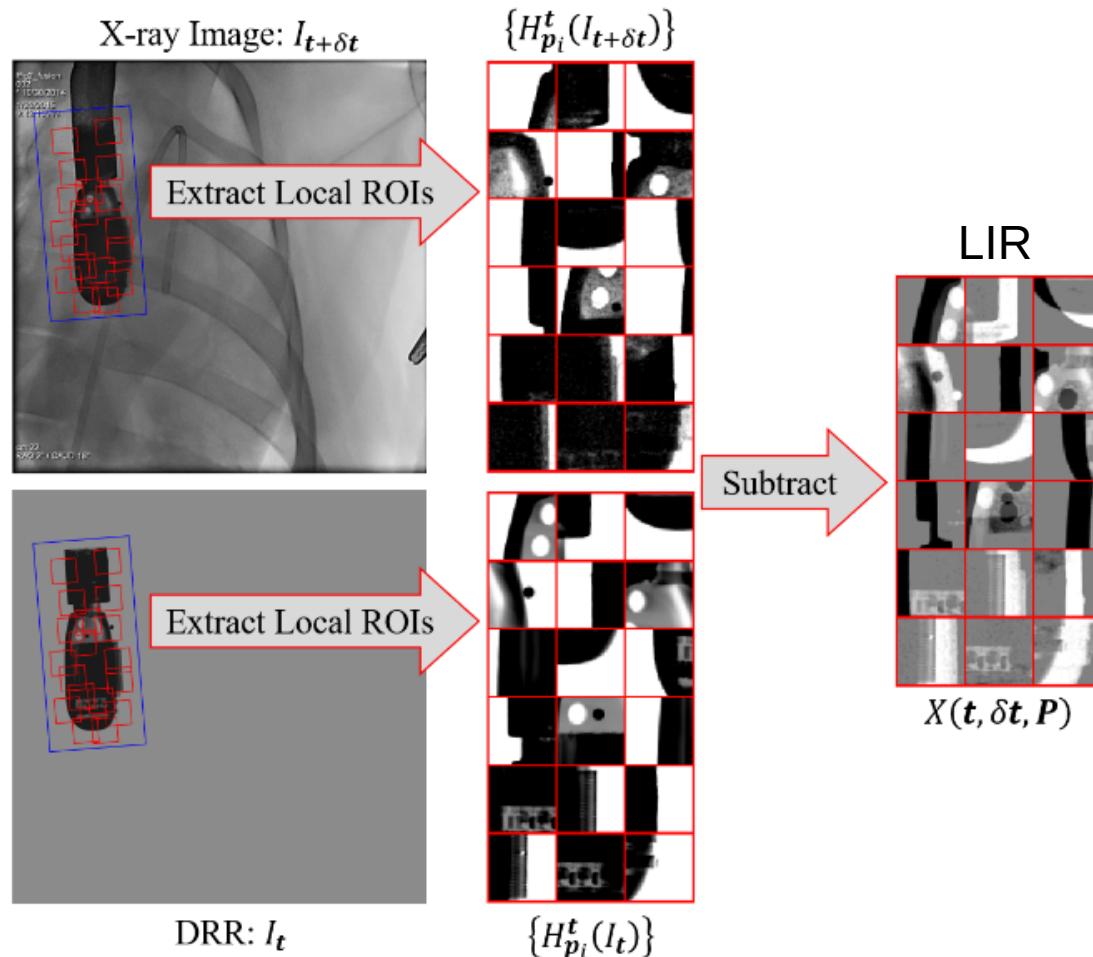
Pose Estimation via Hierarchical Learning (PEHL)

- Local image residual (LIR)
- Parameter space partitioning (PSP)
- Hierarchical parameter regression (HPR)

Local Image Residual (LIR)

Given N 3-D points, $P = \{p_1, \dots, p_N\}$, the LIR feature is then computed as:

$$X(t, I_{t+\delta t}, P) = \{H_{p_i}^t(I_t) - H_{p_i}^t(I_{t+\delta t})\}_{i=1,\dots,N}. \quad (9)$$



Digitally Reconstructed Radiograph (DRR), is derived from the 3-D X-ray attenuation map by Simulating the attenuation of virtual X-rays.

Proposal

Pose Estimation via Hierarchical Learning (PEHL)

- Local image residual (LIR)
- Parameter space partitioning (PSP)
- Hierarchical parameter regression (HPR)

Parameter space partitioning (PSP)

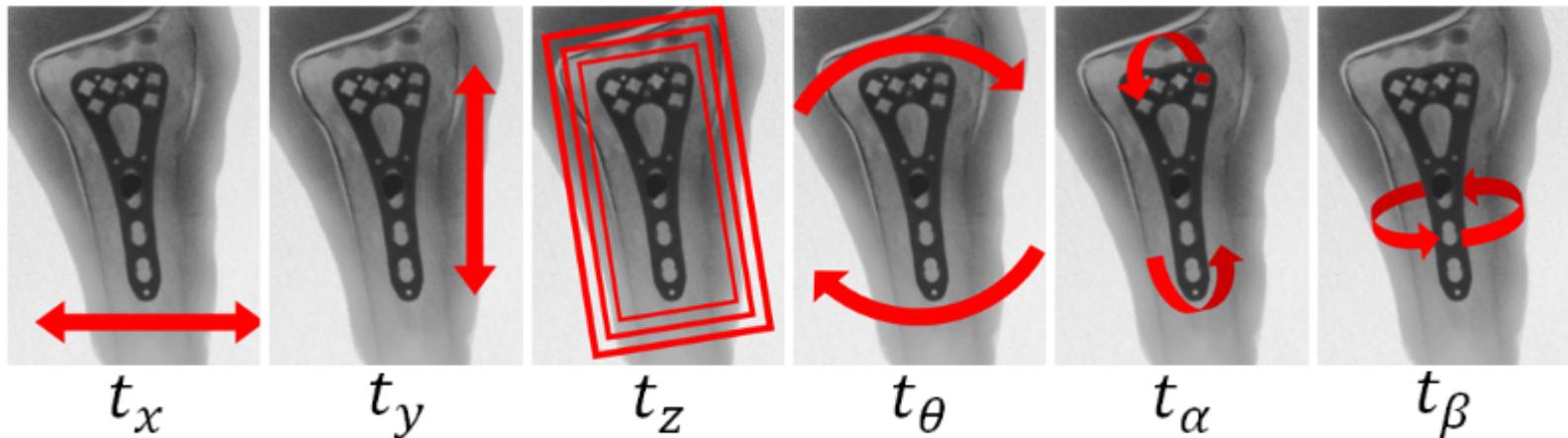


Fig. 1: Effects of the 6 transformation parameters.

$$X(\mathbf{t}_1, I_{\mathbf{t}_1+\delta\mathbf{t}}) \approx X(\mathbf{t}_2, I_{\mathbf{t}_2+\delta\mathbf{t}}) \quad \forall (\mathbf{t}_1, \mathbf{t}_2).$$

To use ROIs to make $X(t, I|t+\delta t)$ invariant to the in-plane and scaling parameters, (x, y, z, θ). However, they are unable to make $X(t, I|t+\delta t)$ insensitive to α and β , because they cause complex appearance changes in the projection image.
Other partitions the parameter space spanned by α and β has been created (k -th zone):

$$X_k(\mathbf{t}_1, I_{\mathbf{t}_1+\delta\mathbf{t}}) \approx X_k(\mathbf{t}_2, I_{\mathbf{t}_2+\delta\mathbf{t}}) \quad \forall (\mathbf{t}_1, \mathbf{t}_2) \in \Omega_k,$$

Proposal

Pose Estimation via Hierarchical Learning (PEHL)

- Local image residual (LIR)
- Parameter space partitioning (PSP)
- Hierarchical parameter regression (HPR)

Hierarchical parameter regression (HPR)

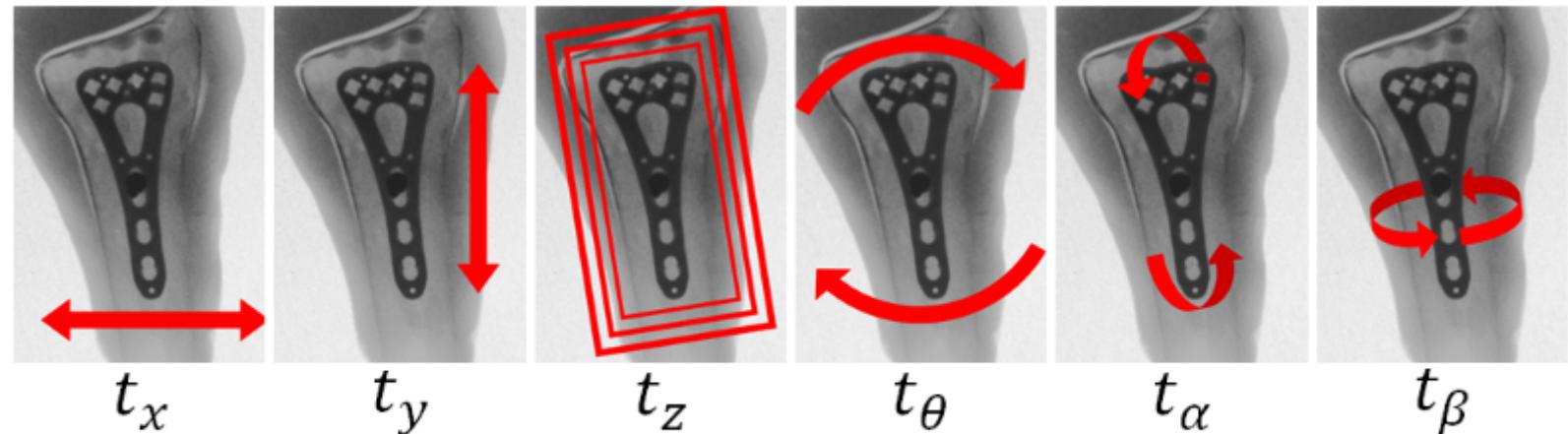


Fig. 1: Effects of the 6 transformation parameters.

- *Group 1:* In-plane parameters: $\delta t_x, \delta t_y, \delta t_\theta$
- *Group 2:* Out-of-plane rotation parameters: $\delta t_\alpha, \delta t_\beta$
- *Group 3:* Out-of-plane translation parameter: δt_z

Pose Estimation via Hierarchical Learning

Algorithm 1 PEHL: Application Stage

```
1: procedure REGISTER( $t, I, k$ )
2:   repeat
3:     Retrieve  $\mathbf{P}$  for the zone covering  $(t_\alpha, t_\beta)$ 
4:     Retrieve  $\mathbf{f}(\cdot)$  for the zone covering  $(t_\alpha, t_\beta)$ 
5:     Calculate  $X(t, I_{t+\delta t}, \mathbf{P})$             $\triangleright$  Eqn. (9)
6:      $\mathbf{t}_{\{x,y,\theta\}} \leftarrow \mathbf{t}_{\{x,y,\theta\}} + \mathbf{f}_{\{x,y,\theta\}}(X)$     ←
7:     Calculate  $X(t, I_{t+\delta t}, \mathbf{P})$             $\triangleright$  Eqn. (9)
8:      $\mathbf{t}_{\{\alpha,\beta\}} \leftarrow \mathbf{t}_{\{\alpha,\beta\}} + \mathbf{f}_{\{\alpha,\beta\}}(X)$     ←
9:     Calculate  $X(t, I_{t+\delta t}, \mathbf{P})$             $\triangleright$  Eqn. (9)
10:     $t_z \leftarrow t_z + f_z(X)$                   ←
11:   until reaching  $k$  iterations
12:   return  $t$ 
```

\mathbf{P} : 3-D points;
 $\mathbf{F}(\cdot)$: set of regressors are
trained to reveal the mapping
from a feature $X(\cdot)$;

Compute

Group 1

Group 2

Group 3

Hierarchical Parameter Regression

- *Group 1*: In-plane parameters: $\delta t_x, \delta t_y, \delta t_\theta$
- *Group 2*: Out-of-plane rotation parameters: $\delta t_\alpha, \delta t_\beta$
- *Group 3*: Out-of-plane translation parameter: δt_z

Proposal

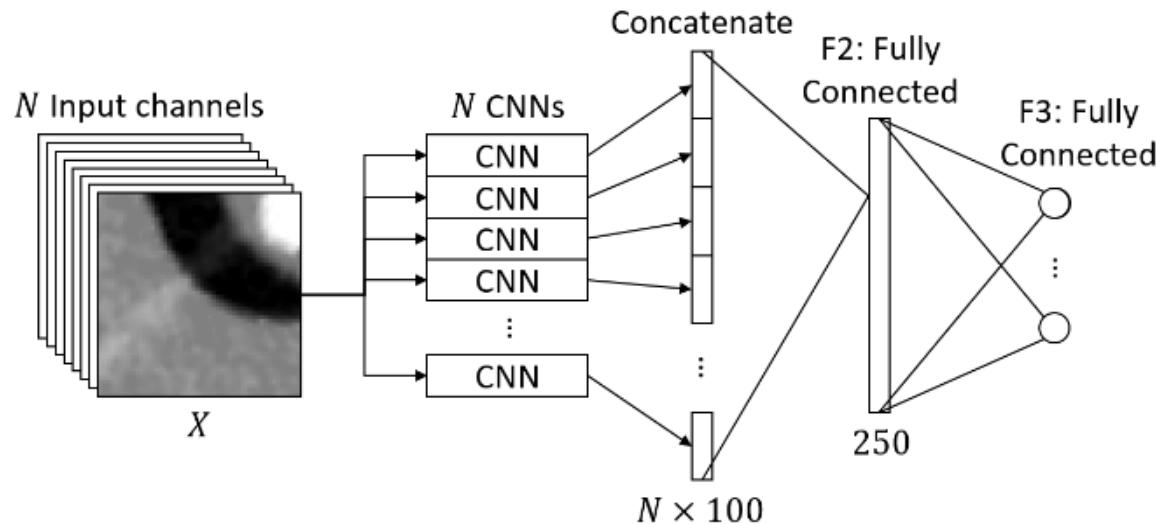


Fig. 3: Structure of the CNN regression model.

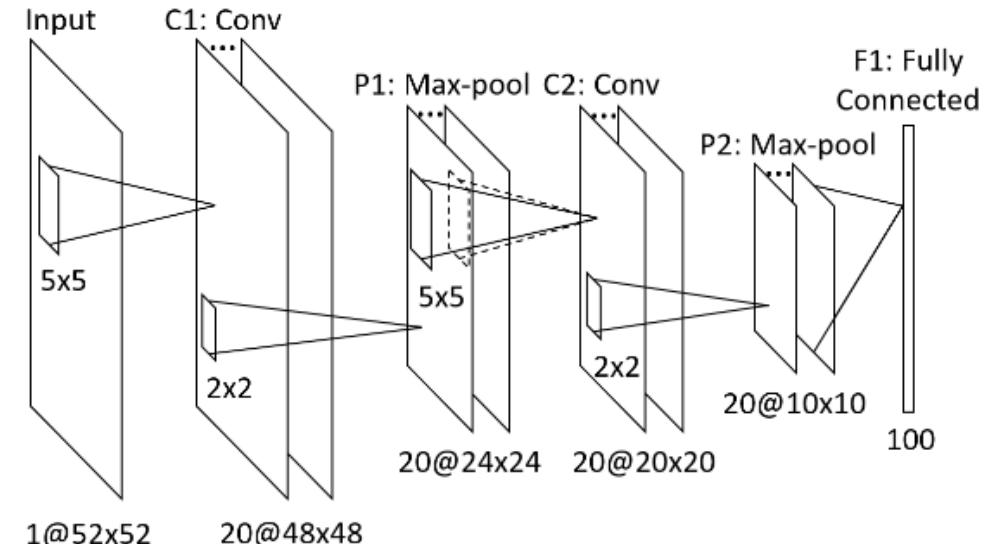
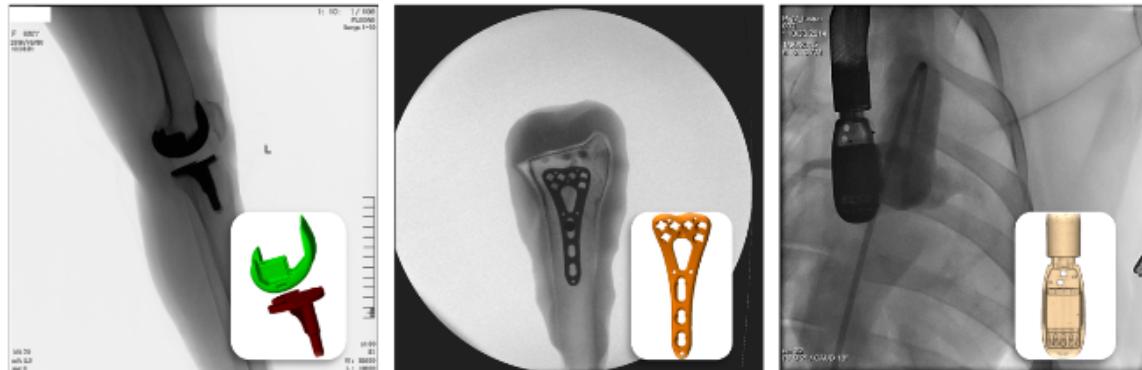


Fig. 4: Structure of the CNN applied for each input channel.

Datasets

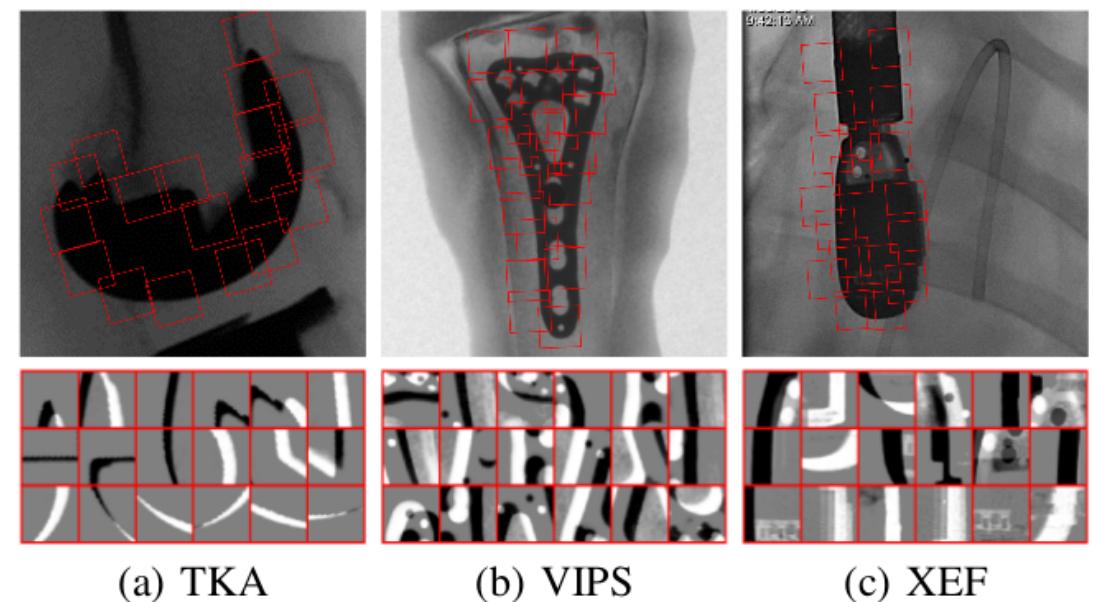


(a) TKA

(b) VIPS

(c) XEF

Fig. 5: Example data, including a 3-D model and a 2-D X-ray image of the object.



(a) TKA

(b) VIPS

(c) XEF

Fig. 6: Examples of local ROIs and LIRs.

Results

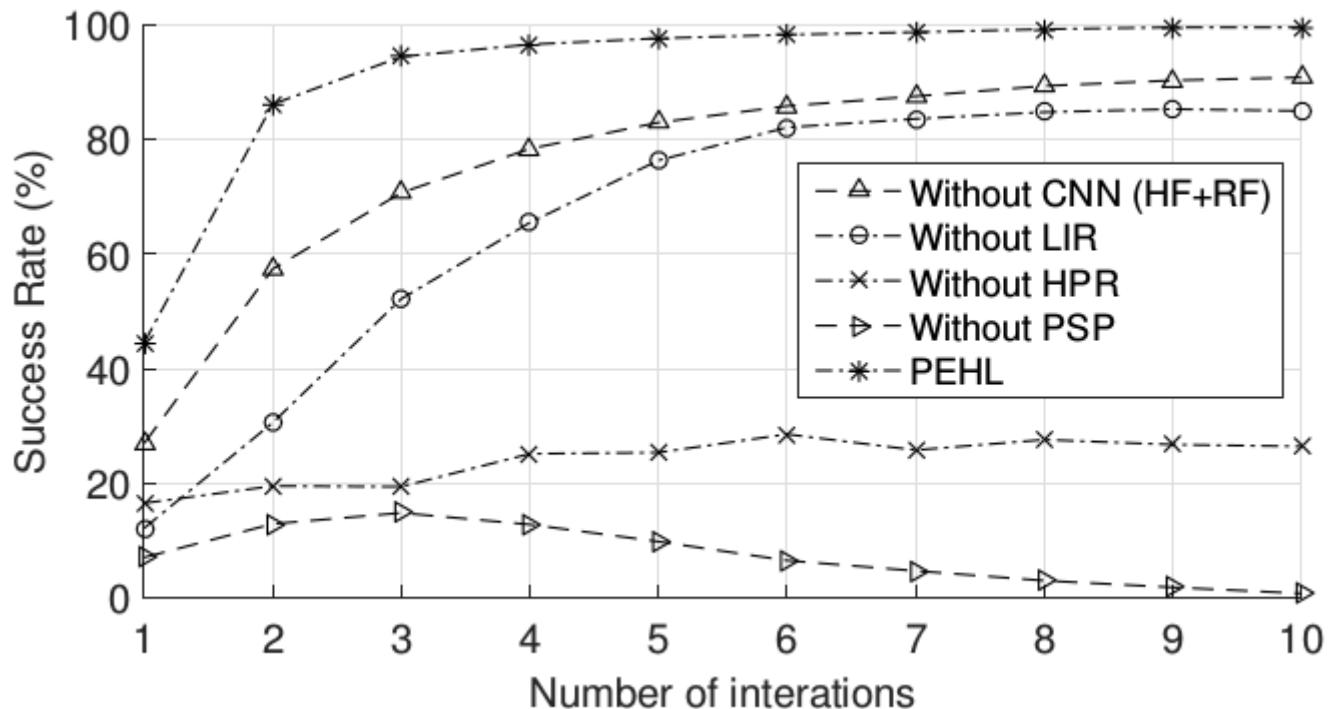


Fig. 9: Success rates of PEHL with 1 to 10 iterations. Four individual core components of PEHL, i.e., CNN, LIR, HPR and PSP, were disabled one at a time to demonstrate their detrimental effects on performance. Harr feature (HR) + Regression Forest (RF) was implemented to show the effect on performance without CNN. These results were generated on the XEF dataset.

TABLE II: RMSE of the 6 transformation parameters yielded by PEHL and CLARET on the training data for XEF. The units for (t_x, t_y, t_z) and $(t_\theta, t_\alpha, t_\beta)$ are mm and degree, respectively.

	t_x	t_y	t_z	t_θ	t_α	t_β
Start	0.86	0.86	8.65	1.71	8.66	8.66
PEHL	0.04	0.04	0.32	0.06	0.18	0.18
CLARET	0.51	0.88	34.85	2.00	19.41	17.52

RMSE = Root Mean Squared Error

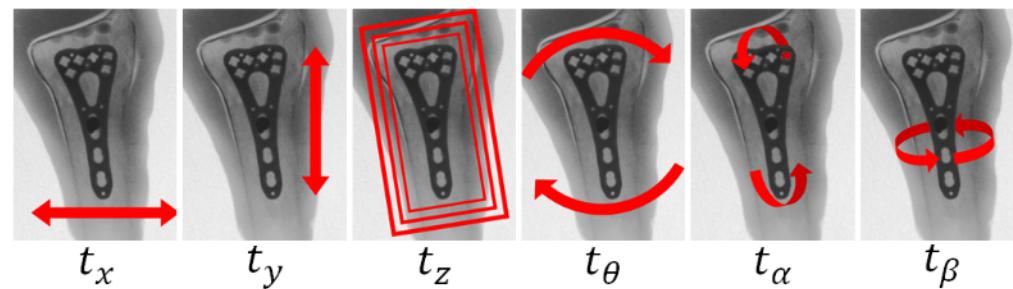


Fig. 1: Effects of the 6 transformation parameters.

Results

TABLE III: Quantitative experiment results of PEHL and baseline methods. Success rate is the percentage of successful registrations in each experiment. Capture range is the initial mTREproj for which 95% of the registrations were successful. The 10th, 25th, 50th, 75th and 90th percentiles of mTREproj are reported. Running time records the average and standard deviation of the computation time for each registration computed in each experiment. Capture range is only reported for experiments where there are more than 20 samples within the capture range.

Application	Method	Success Rate	Capture Range (mm)	mTREproj Percentile (mm)					Running Time (s)
				10th	25th	50th	75th	90th	
TKA	Start	N/A	N/A	3.285	4.627	6.979	10.050	12.667	N/A
	MI_Powell	36.2%	N/A	0.437	0.746	1.693	6.238	8.421	1.37±0.44
	CC_Powell	43.8%	1.88	0.348	0.637	1.359	6.321	8.398	0.92±0.27
	GC_Powell	45.2%	2.14	0.330	0.588	1.313	7.615	9.765	2.52±1.22
	MI_GC_Powell	51.8%	2.83	0.299	0.521	1.048	6.408	8.614	3.11±0.94
	PEHL	79.6%	7.23	0.333	0.444	0.593	0.903	6.733	0.11±0.00
VIPS	Start	N/A	N/A	1.180	1.521	2.003	2.594	3.101	N/A
	MI_Powell	75.1%	N/A	0.156	0.234	0.375	0.604	0.917	1.66±0.60
	CC_Powell	57.7%	0.89	0.187	0.303	0.535	0.851	1.293	0.91±0.31
	GC_Powell	78.7%	1.12	0.121	0.207	0.325	0.543	2.283	3.91±1.55
	MI_GC_Powell	92.7%	2.77	0.106	0.170	0.259	0.367	0.535	4.71±1.59
	PEHL	99.7%	5.51	0.151	0.181	0.244	0.389	0.451	0.10±0.00
XEF	Start	N/A	N/A	1.048	1.369	1.826	2.307	2.790	N/A
	MI_Powell	69.7%	N/A	0.165	0.207	0.280	0.403	0.598	0.79±0.29
	CC_Powell	54.8%	N/A	0.117	0.168	0.321	0.893	1.173	0.40±0.10
	GC_Powell	56.9%	N/A	0.071	0.135	0.279	1.055	3.150	2.06±1.05
	MI_GC_Powell	89.1%	0.84	0.047	0.098	0.174	0.273	0.380	2.03±0.69
	PEHL	94.5%	3.33	0.082	0.113	0.148	0.195	0.243	0.10±0.00

Deep Learning - Transformation Parameters

(Bob D. de Vos et al. 2107)

End-to-End Unsupervised Deformable Image Registration with a Convolutional Neural Network

Bob D. de Vos¹, Floris F. Berendsen², Max A. Viergever¹, Marius Staring², and Ivana Eguim¹

¹Image Sciences Institute, University Medical Center Utrecht, the Netherlands

²Division of Image Processing, Leiden University Medical Center, the Netherlands

Abstract. In this work we propose a deep learning network for deformable image registration (DIRNet). The DIRNet consists of a convolutional neural network (ConvNet) regressor, a spatial transformer, and a resampler. The ConvNet analyzes a pair of fixed and moving images and outputs parameters for the spatial transformer, which generates the displacement vector field that enables the resampler to warp the moving image to the fixed image. The DIRNet is trained end-to-end by unsupervised optimization of a similarity metric between input image pairs. A trained DIRNet can be applied to perform registration on unseen image pairs in one pass, thus non-iteratively. Evaluation was performed with registration of images of handwritten digits (MNIST) and cardiac cine MR scans (Sunnybrook Cardiac Data). The results demonstrate that registration with DIRNet is as accurate as a conventional deformable image registration method with substantially shorter execution times.

Keywords: convolution neural network, deformable image registration, spatial transformer, cardiac MRI

Proposal

The **DIRNet** is trained by optimizing an image similarity metric (normalized cross correlation)

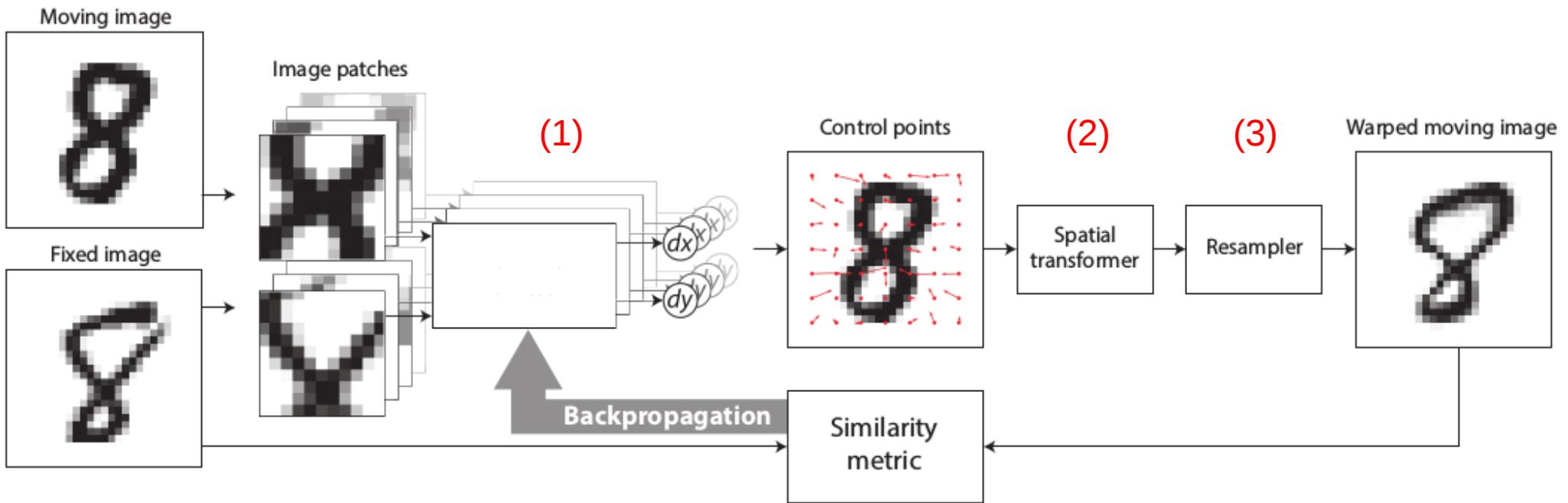
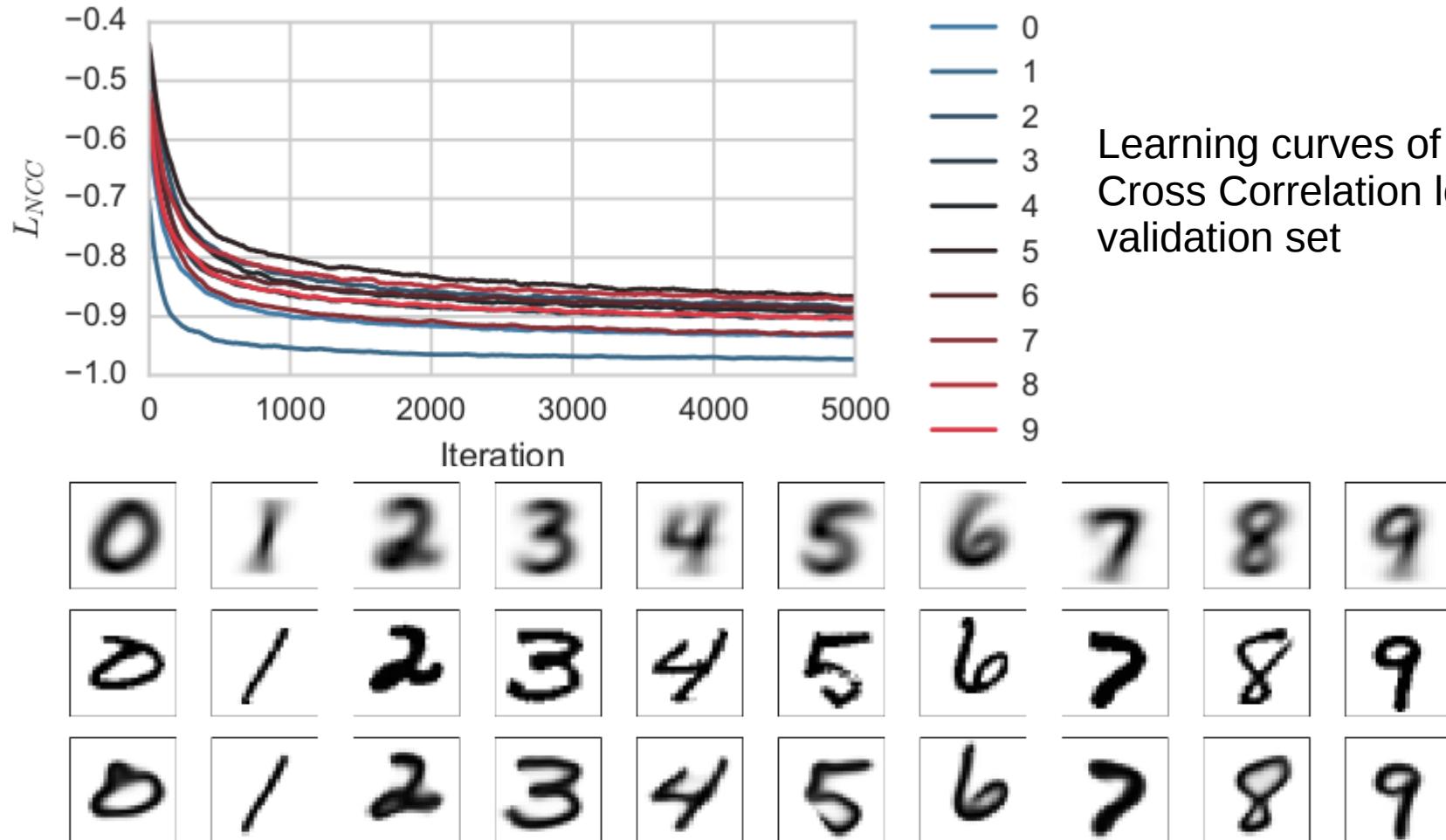


Fig. 1. Schematics of the DIRNet with two input images from the MNIST data. The DIRNet takes one or more pairs of moving and fixed images as its inputs. The fully convolutional ConvNet regressor analyzes spatially corresponding image patches from the moving and fixed images and generates a grid of control points for a B-spline transformer. The B-spline transformer generates a full displacement vector field to warp a moving image to a fixed image. Training of the DIRNet is unsupervised and end-to-end by backpropagating an image similarity metric as a loss.

Results



- 1- The top row shows an average of all moving images per class;
- 2- the middle row shows one randomly chosen fixed image per class;
- 3- the bottom row shows an average of the registration results of the moving images to the chosen fixed image.

Results

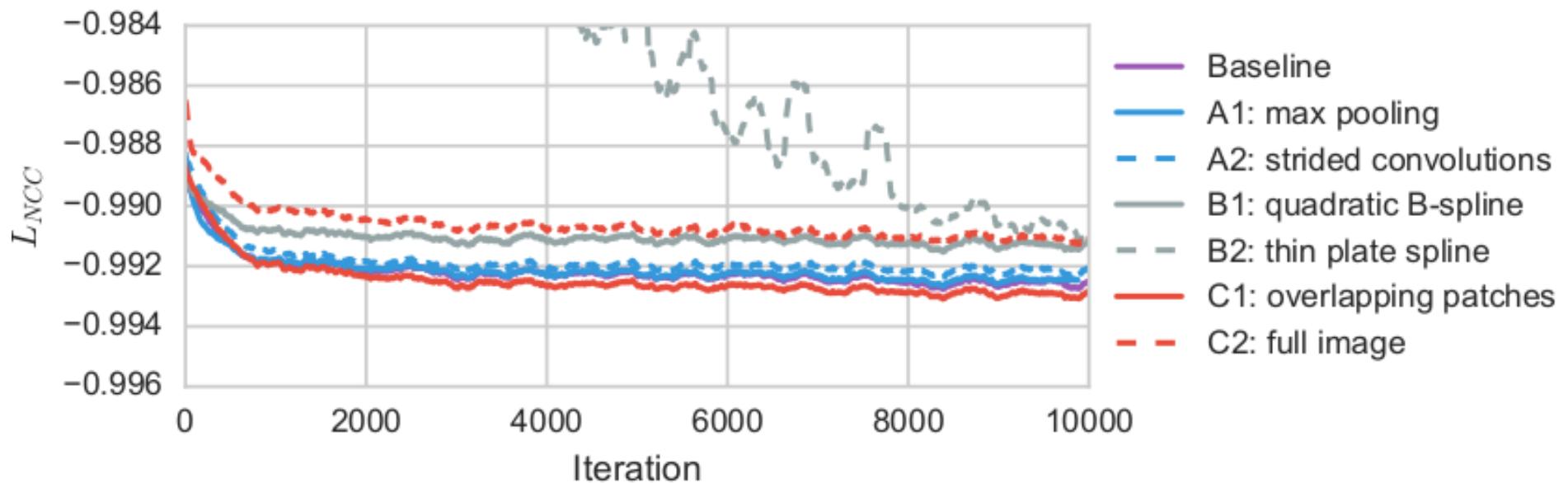


Fig. 3. Validation loss over 10,000 iteration for the baseline DIRNet, DIRNets with different downsampling techniques (A1, A2), DIRNets with different spatial transformers (B1, B2), and DIRNets with different receptive fields (C1, C2).

Results

Table 1. Quantitative cardiac MRI registration results by comparing reference annotations in fixed images and warped annotations of the moving to the fixed images. The table lists mean and standard deviation for the Dice score, 95th percentiles of the surface distance (95thSD), and mean absolute surface distance (MAD). The rows show results before registration, with conventional iterative image registration using SimpleElastix, and registration using the DIRNet. The best obtained results are shown in bold.

		Dice	95 th SD	MAD
Before		0.62 ± 0.15	7.79 ± 2.92	2.89 ± 1.07
SimpleElastix		0.79 ± 0.08	5.09 ± 2.36	1.91 ± 0.94
DIRNet	BL	0.79 ± 0.08	5.20 ± 2.30	1.92 ± 0.89
	A1	0.78 ± 0.08	5.26 ± 2.16	1.95 ± 0.85
	A2	0.78 ± 0.08	5.30 ± 2.28	1.97 ± 0.87
	B1	0.72 ± 0.11	6.41 ± 2.61	2.40 ± 0.96
	B2	0.78 ± 0.09	5.48 ± 2.36	2.01 ± 0.89
	C1	0.80 ± 0.08	5.03 ± 2.30	1.83 ± 0.89
	C2	0.76 ± 0.09	5.55 ± 2.24	2.10 ± 0.90

Results

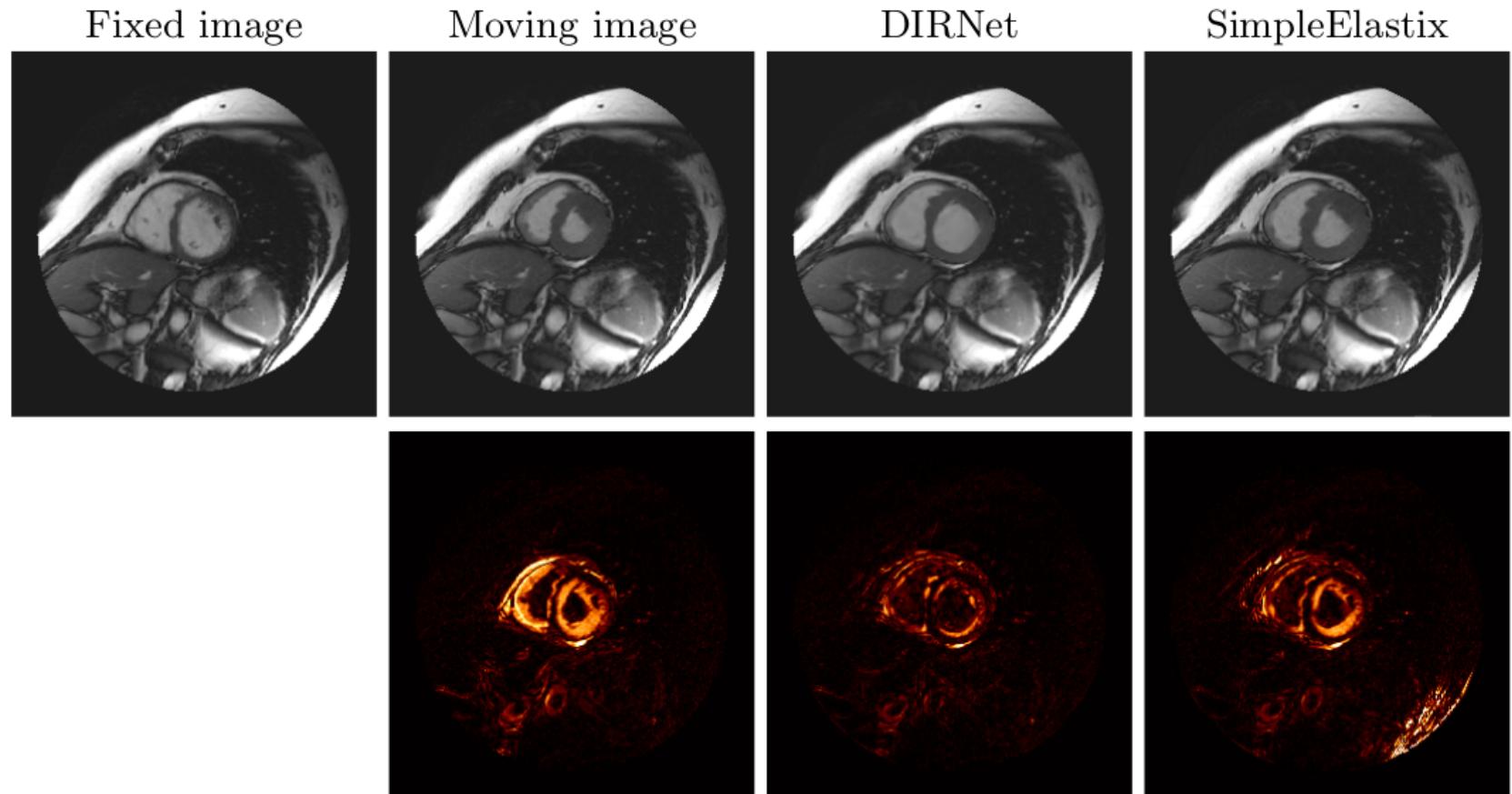


Fig. 4. Top, from left to right: The fixed (ED), the moving (ES), the DIRNet-C1 warped, and the SimpleElastix warped images. Bottom: Heatmaps showing absolute difference images between the fixed image and (from left to right) the original, the DIRNet warped, and the SimpleElastix warped moving images.

Deep Learning - Transformation Parameters

(Mahapatra et al. 2018)

ELASTIC REGISTRATION OF MEDICAL IMAGES WITH GANS

Dwarikanath Mahapatra, Bhavna Antony, Suman Sedai, Rahil Gamavi

IBM Research - Australia, Melbourne

ABSTRACT

Conventional approaches to image registration consist of time consuming iterative methods. Most current deep learning (DL) based registration methods extract deep features to use in an iterative setting. We propose an end-to-end DL method for registering multimodal images. Our approach uses generative adversarial networks (GANs) that eliminates the need for time consuming iterative methods, and directly generates the registered image with the deformation field. Appropriate constraints in the GAN cost function produce accurately registered images in less than a second. Experiments demonstrate their accuracy for multimodal retinal and cardiac MR image registration.

Index Terms— GANs, deformable registration, displacement field

on simulated deformations to generate displacement vector fields for a pair of unimodal images. Vos et. al. [7] propose the deformable image registration network (DIR-Net) which takes pairs of fixed and moving images as input, and outputs a transformed image non-iteratively. Training is completely unsupervised and unlike previous methods it is not trained with known registration transformations.

While RegNet and DIRNet are among the first methods to achieve registration in a single pass, they have some limitations such as: 1) using spatially corresponding patches to predict transformations. Finding corresponding patches is challenging in low contrast medical images and can adversely affect the registration task; 2) Multimodal registration is challenging with their approach due to the inherent problems of finding spatially corresponding patches; 3) DIRNet uses B-splines for spatial transformations which limits the extent of recovering a deformation field; 4) Use of intensity based cost functions limits the benefits that can be derived from a DL

Proposal – CycleGAN

The adversarial loss function for G is given by:

$$L_{cycGAN}(G, D_Y, X, Y) = E_{y \in p_{data}(y)} [\log D_Y(y)] + E_{x \in p_{data}(x)} [\log (1 - D_Y(G(x)))] , \quad (3)$$

We retain notations X, Y for conciseness. There also exists $L_{cycGAN}(F, D_X, Y, X)$ the corresponding adversarial loss for F and D_X .

$$X \rightarrow G(x) \rightarrow F(G(x)) \sim x$$

Cycle Consistency Loss

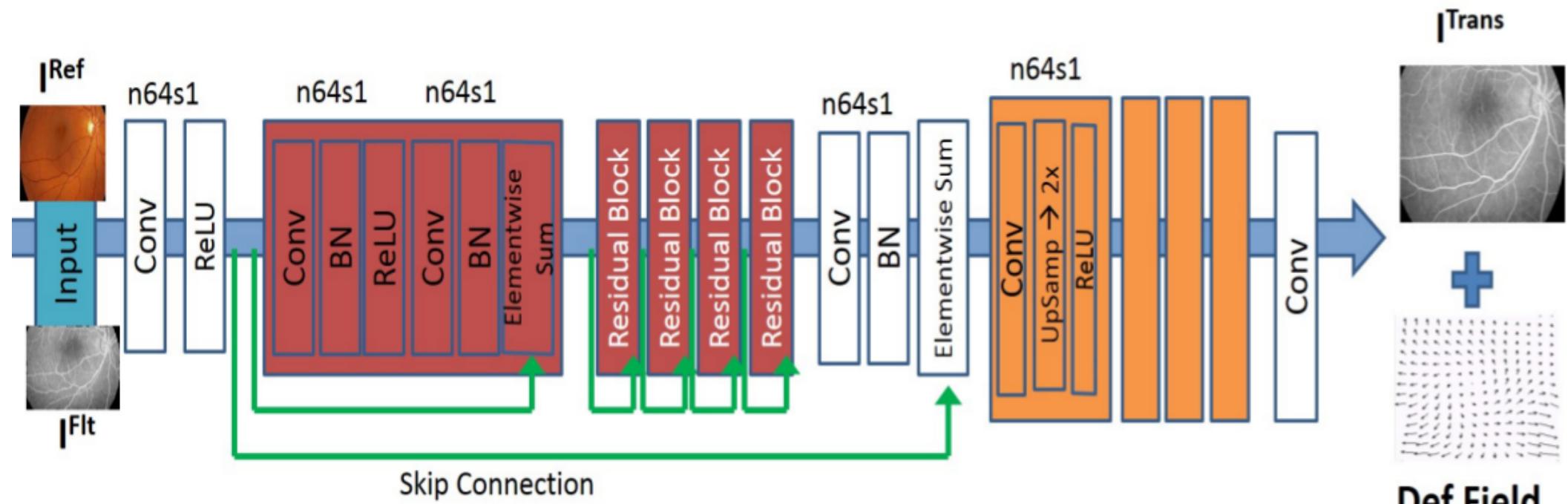
$$L_{cyc}(G, F) = E_x \|F(G(x)) - x\|_1 + E_y \|G(F(y)) - y\|_1 , \quad (4)$$

The full objective function is

$$L(G, F, D_{Flt}, D_{Ref}) = L_{cycGAN}(G, D_{Ref}, I^{Flt}, I^{Ref}) + L_{cycGAN}(F, D_{Flt}, I^{Ref}, I^{Flt}) + \lambda L_{cyc}(G, F) \quad (5)$$

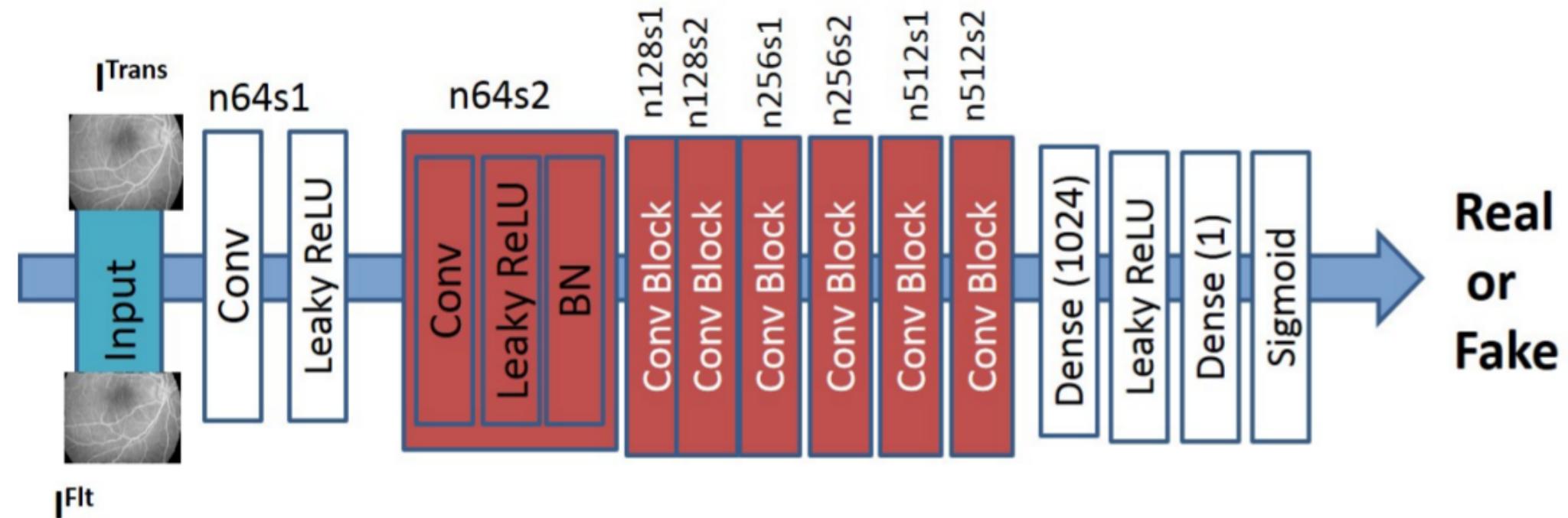
A generative network tries to fool other discriminative model (discriminator) to compose a minimax Algorithm. GANs face a really big problem that it is the creation of non-realistic samples.

Generator Network



n64s1 denotes 64 feature maps (n) and stride (s) 1 for each convolutional layer.

Discriminator network



n64s1 denotes 64 feature maps (n) and stride (s) 1 for each convolutional layer.

Results

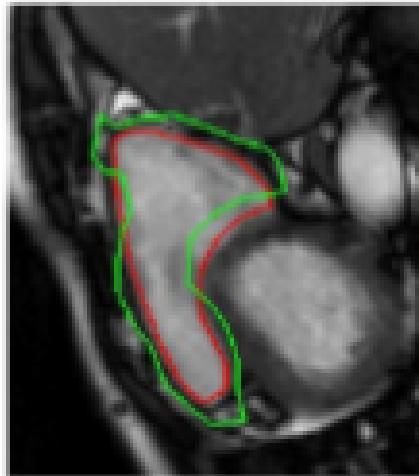
Table 1. Comparative average performance of different methods before and after retinal image registration. *Time* denotes time in seconds taken to register a test image pair.

	Bef. Reg.	After Registration			
		<i>GAN</i> <i>Reg</i>	DIRNet [7]	Elastix [18]	<i>GAN</i> <i>Reg_{nCyc}</i>
Dice	0.843	0.946	0.911	0.874	0.887
Err_{Def}	14.3	5.7	7.3	12.1	9.1
HD_{95}	11.4	4.2	5.9	9.7	8.0
MAD	9.1	3.1	5.0	8.7	7.2
MSE	0.84	0.09	0.23	0.54	0.37
<i>Time</i> (s)		0.7	0.9	15.1	0.7

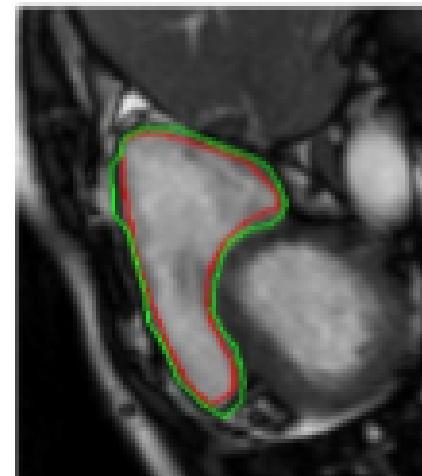
Table 2. Comparative average performance of different methods before and after cardiac image registration. *Time* denotes time in seconds taken to register a test image pair..

	Bef. Reg.	After Registration			
		<i>GAN</i> <i>Reg</i>	DIRNet [7]	Elastix [18]	<i>GAN</i> <i>Reg_{nCyc}</i>
Dice	0.62	0.85	0.80	0.77	0.79
HD_{95}	7.79	3.9	5.03	5.21	5.12
MAD	2.89	1.3	1.83	2.12	1.98
<i>Time</i> (s)		0.8	0.8	11.1	0.8

Results

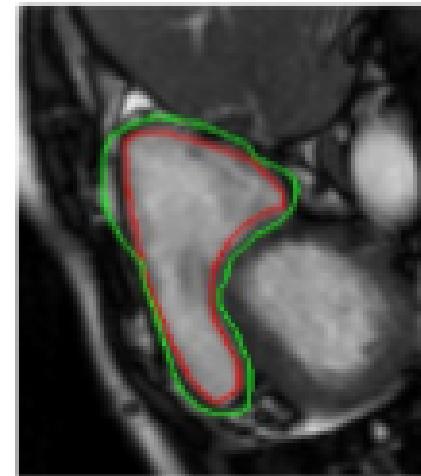


(a)



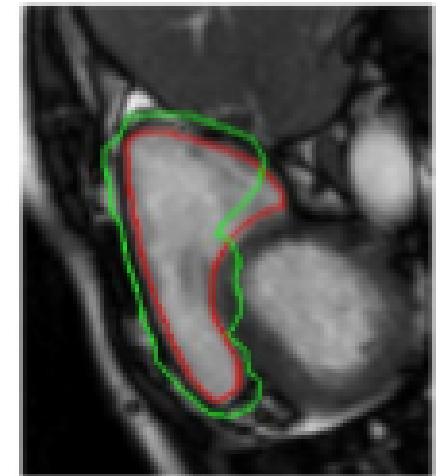
(b)

GAN



(c)

DIRNet



(d)

Elastix

Fig. 3. Example results for cardiac RV registration. Superimposed contours of the ground truth (red) and deformed segmentation mask of moving image (green): (a) before registration; after registration using (b) GAN_{Reg} ; (c) $DIRNet$; (d) Elastix.

right ventricular (RV)

Deep Learning - Transformation Parameters

(Eppenhof et al.2018)

Deformable image registration using convolutional neural networks

Koen A.J. Eppenhof, Maxime W. Lafarge, Pim Moeskops, Mitko Veta, and Josien P.W. Pluim

Medical Image Analysis Group (IMAGE/e), Department of Biomedical Engineering,
Eindhoven University of Technology, Eindhoven, The Netherlands

ABSTRACT

Deformable image registration can be time-consuming and often needs extensive parameterization to perform well on a specific application. We present a step towards a registration framework based on a three-dimensional convolutional neural network. The network directly learns transformations between pairs of three-dimensional images. The outputs of the network are three maps for the x, y, and z components of a thin plate spline transformation grid. The network is trained on synthetic random transformations, which are applied to a small set of representative images for the desired application. Training therefore does not require manually annotated ground truth deformation information. The methodology is demonstrated on public data sets of inspiration-expiration lung CT image pairs, which come with annotated corresponding landmarks for evaluation of the registration accuracy. Advantages of this methodology are its fast registration times and its minimal parameterization.

Keywords: deformable image registration, machine learning, convolutional networks, thoracic CT

Proposal

$$\mathbf{T}(\mathbf{x}) = \mathbf{x} + A\mathbf{x} + \mathbf{t} + \sum_k \mathbf{c}_k \phi(\|\mathbf{d}_k\|)$$

where A is an affine matrix, \mathbf{t} is a translation vector, and \mathbf{c}_k are spline coefficients. The parameters A , \mathbf{t} , and \mathbf{c}_k are computed from the displacements \mathbf{d}_k .

$$\mathbf{T}_{\text{offset}}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b} = (\mathbb{I} + \mathbf{B})\mathbf{x} + \mathbf{b},$$

Thin plate splines (TPS) are a spline-based technique for data interpolation and smoothing

Offset = data augmentation

Proposal

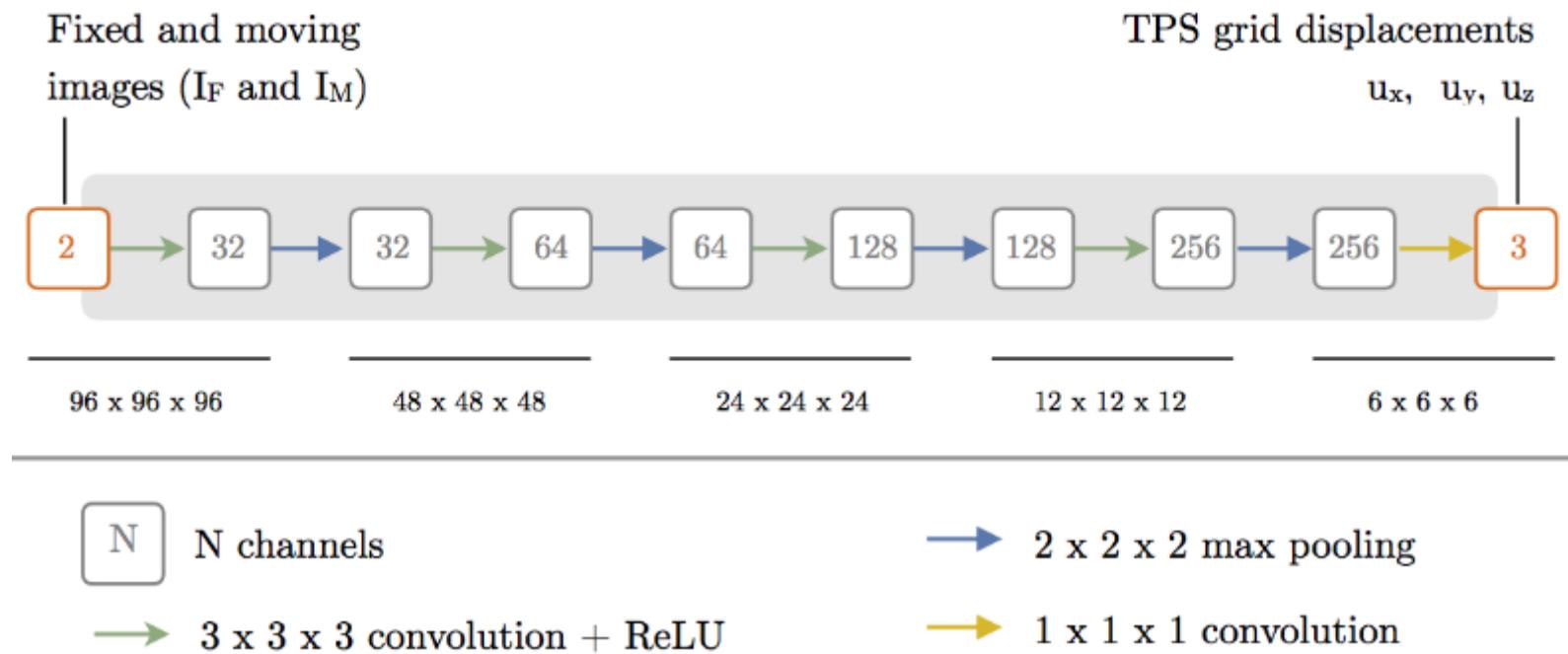


Figure 1. Network architecture. The network learns the displacements on a $6 \times 6 \times 6$ grid as three maps: one for every component of the thin plate spline grid. The input to the network is an image with two channels, where the channels correspond to the fixed and moving image of the registration.

Proposal

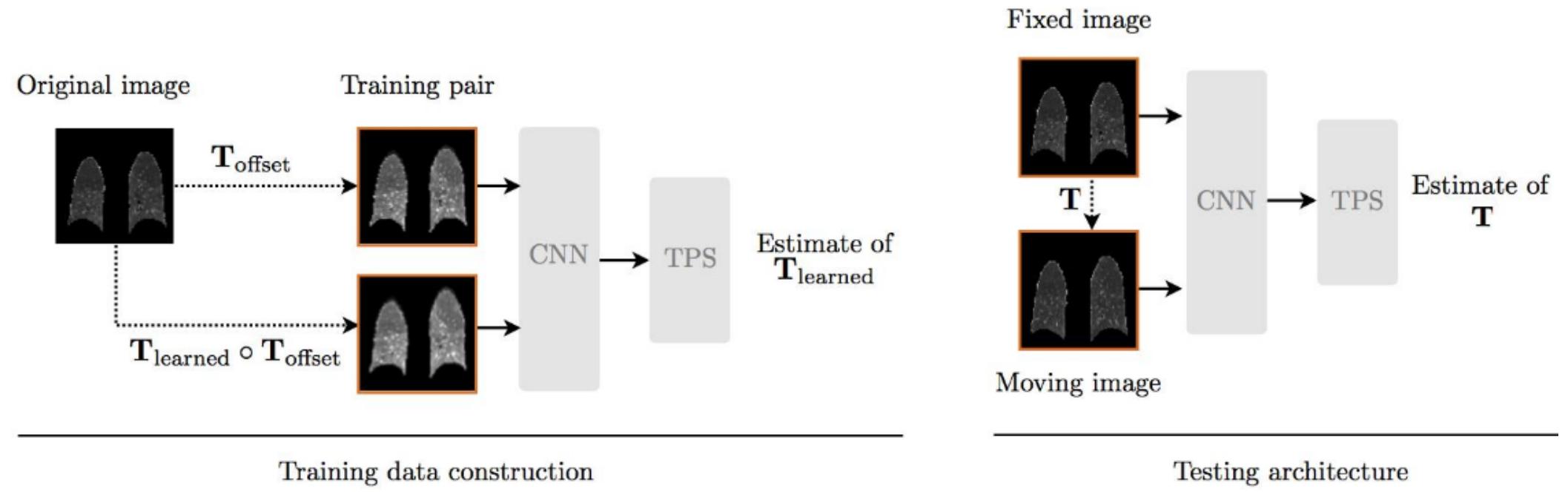


Figure 2. During training the inputs to the network are two transformed versions of the same image. During testing, the inputs are the fixed and moving images.

Results

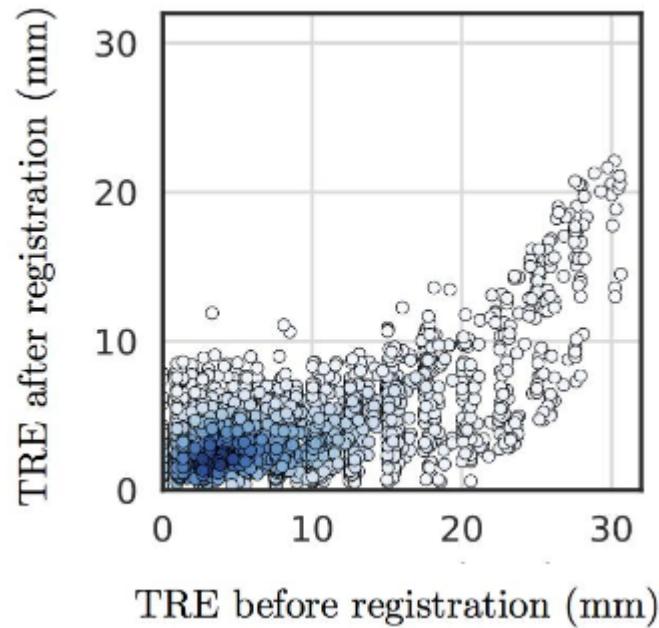


Figure 3. Post-registration TREs against pre-registration TREs. Every point corresponds to a landmark in the ten test images. Darker colors indicate a higher density of points.

Results

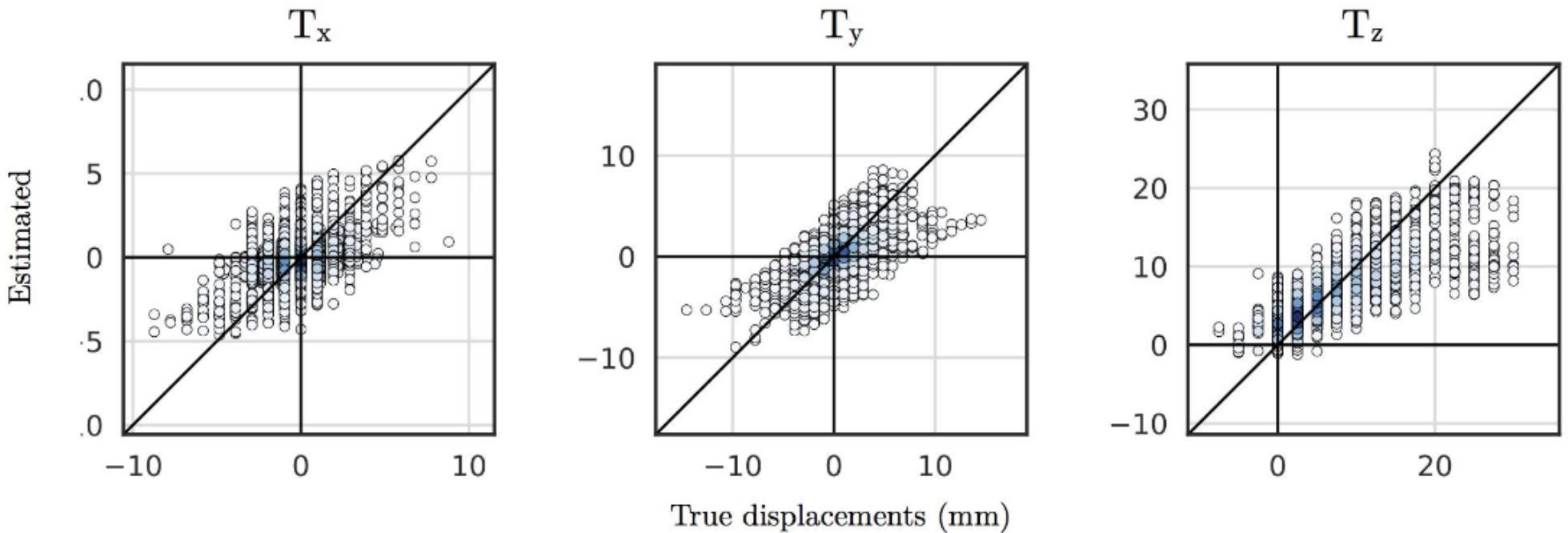


Figure 4. Correlation plots for each of the vector field components. Every point corresponds to a landmark in the ten test images. Darker colors indicate a higher density of points.

Results

Table 1. Target registration errors in millimeters measured on annotated corresponding landmarks, comparing our method with state-of-the-art methods for lung registration. Standard deviations over all 300 landmarks in parentheses.

Image	No registration	Schmidt-Richtberg et al. ¹⁷	Wu et al. ¹⁸	Delmon et al. ¹⁶	Berendsen et al. ¹⁹	Proposed
1	3.89 (2.78)	1.22 (0.64)	1.1 (0.5)	1.2 (0.6)	1.00 (0.52)	1.65 (0.89)
2	4.34 (3.90)	1.14 (0.65)	1.0 (0.5)	1.1 (0.6)	1.02 (0.57)	2.26 (1.16)
3	6.94 (4.05)	1.36 (0.81)	1.3 (0.7)	1.6 (0.9)	1.14 (0.89)	3.15 (1.63)
4	9.83 (4.86)	2.68 (2.79)	1.5 (1.0)	1.6 (1.1)	1.46 (0.96)	4.24 (2.69)
5	7.48 (5.51)	1.57 (1.23)	1.9 (1.5)	2.0 (1.6)	1.61 (1.48)	3.52 (2.23)
6	10.9 (6.97)	2.21 (1.66)	1.6 (0.9)	1.7 (1.0)	1.42 (0.89)	3.19 (1.50)
7	11.0 (7.43)	3.81 (3.06)	1.7 (1.1)	1.9 (1.2)	1.49 (1.06)	4.25 (2.08)
8	15.0 (9.01)	3.42 (4.25)	1.6 (1.4)	2.2 (2.3)	1.62 (1.71)	9.03 (5.08)
9	7.92 (3.98)	1.83 (1.19)	1.4 (0.8)	1.6 (0.9)	1.30 (0.76)	3.85 (1.86)
10	7.30 (6.35)	2.06 (1.92)	1.6 (1.2)	1.7 (1.2)	1.50 (1.31)	5.07 (2.31)
mean	8.46 (5.48)	2.13 (1.82)	1.47 (0.96)	1.66 (1.14)	1.36 (1.01)	4.02(3.08)

Deep Learning - Transformation Parameters

(Sheikhjafari et al. 2018)

Unsupervised Deformable Image Registration with Fully Connected Generative Neural Network

Ameneh Sheikhjafari

Department of Computing Science
University of Alberta
Servier Virtual Cardiac Centre
Mazankowski Alberta Heart Institute
sheikhja@ualberta.ca

Michelle Noga

Radiology and Diagnostic Imaging
University of Alberta
Servier Virtual Cardiac Centre
Mazankowski Alberta Heart Institute
mnoga@ualberta.ca

Kumaradevan Punithakumar

Radiology and Diagnostic Imaging
University of Alberta
Servier Virtual Cardiac Centre
Mazankowski Alberta Heart Institute
punithak@ualberta.ca

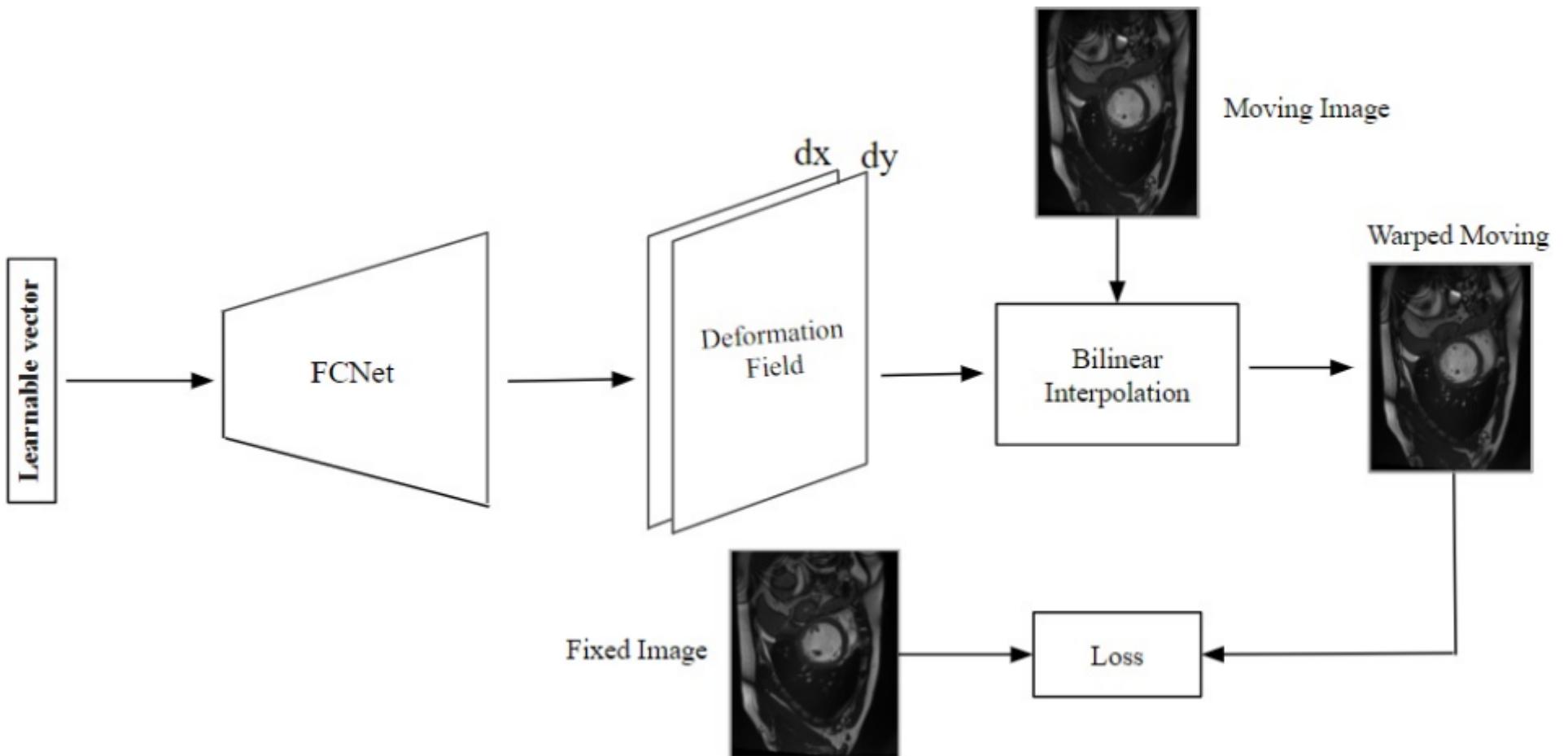
Nilanjan Ray

Department of Computing Science
University of Alberta
nray1@ualberta.ca

Abstract

In this paper, a new deformable image registration method based on a fully connected neural network is proposed. Even though a deformation field related to the point correspondence between fixed and moving images are high-dimensional in nature, we assume that these deformation fields form a low dimensional manifold in many real world applications. Thus, in our method, a neural network generates an embedding of the deformation field from a low dimensional vector. This low-dimensional manifold formulation avoids the intractability associated with the high dimensional search space that most other methods face during image registration. As a result, while most methods rely on explicit and handcrafted regularization of the deformation fields, our algorithm relies on implicitly regularizing the network parameters. The proposed method generates deformation fields from latent low dimensional space by minimizing a dissimilarity metric between a fixed image and a warped moving image. Our method removes the need for a large dataset to optimize the proposed network. The proposed method is quantitatively evaluated using images from the MICCAI ACDC challenge. The results demonstrate that the proposed method improves performance in comparison with a moving mesh registration algorithm, and also it correlates well with independent manual segmentations by an expert.

Proposal



$$E_{data}(\theta, \{t_i\}_{i=1}^n) = \sum_i |I_i - I_{mov(i)}(f_\theta(t_i))|$$

Auto-encoder (AE) brings the data from high dimensional input to low dimensional (latent vector) output.

Proposal

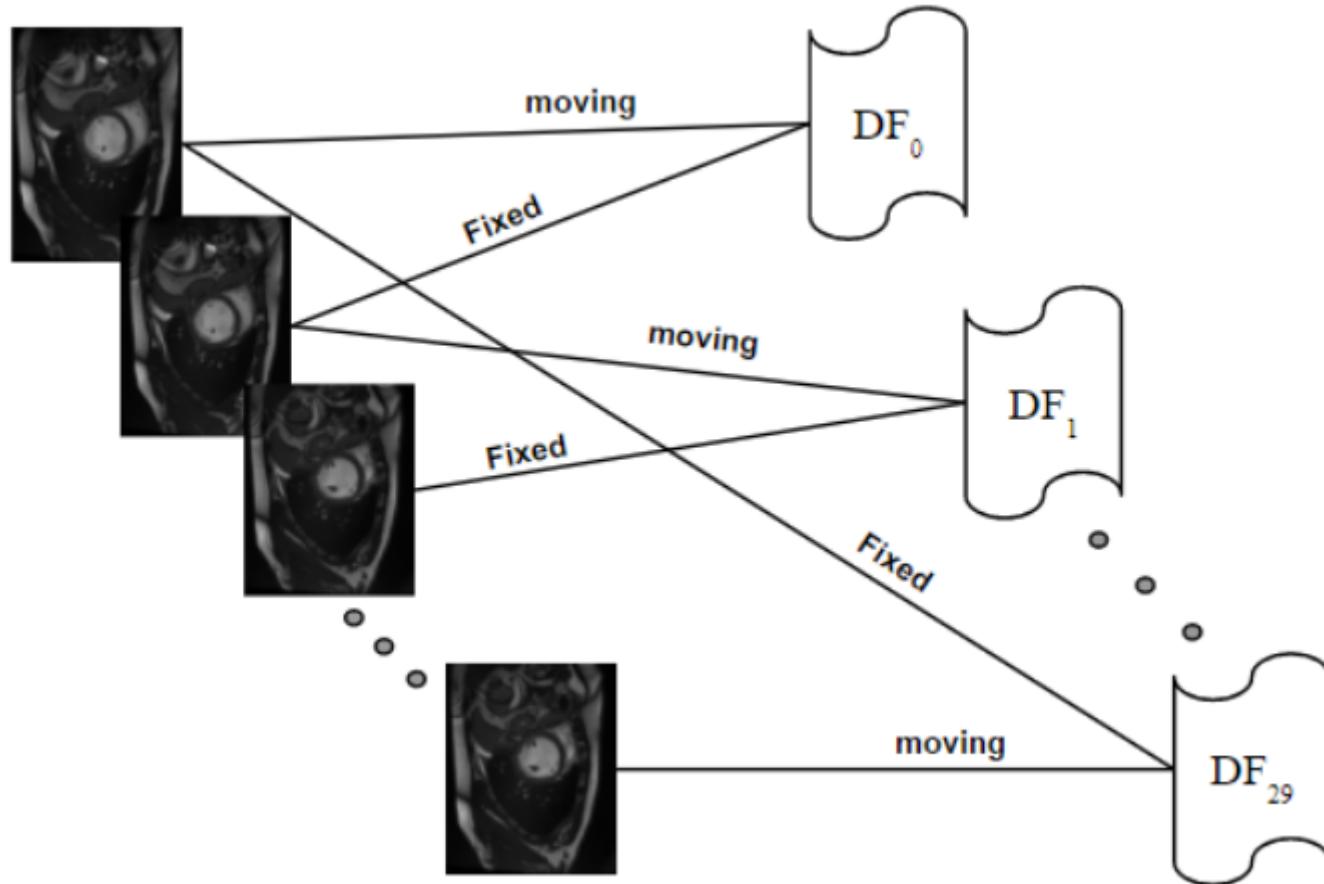


Figure 2: Selection of fixed and moving images in one sequence of the medical image. $DF_i (i = 0, 1, 2, \dots, K)$ are deformation fields generated by the network for each pair of fixed and moving images.

Proposal

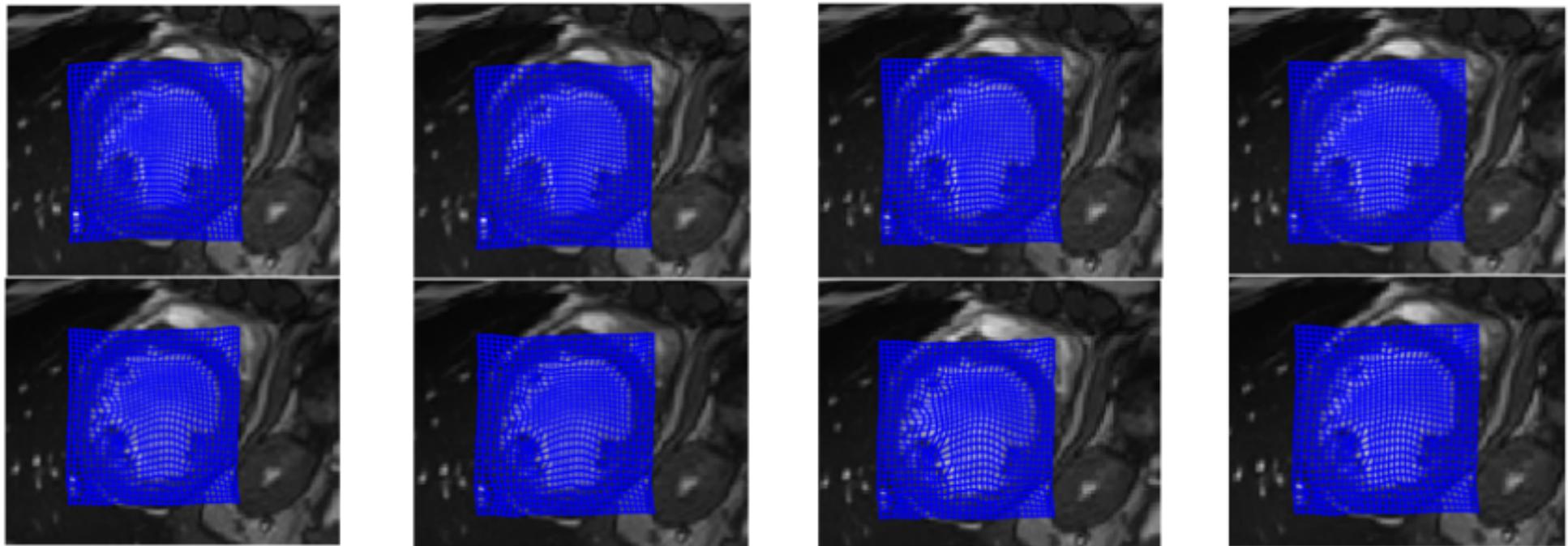


Figure 3: Representative examples of the displacement field obtained by the proposed method

Proposal

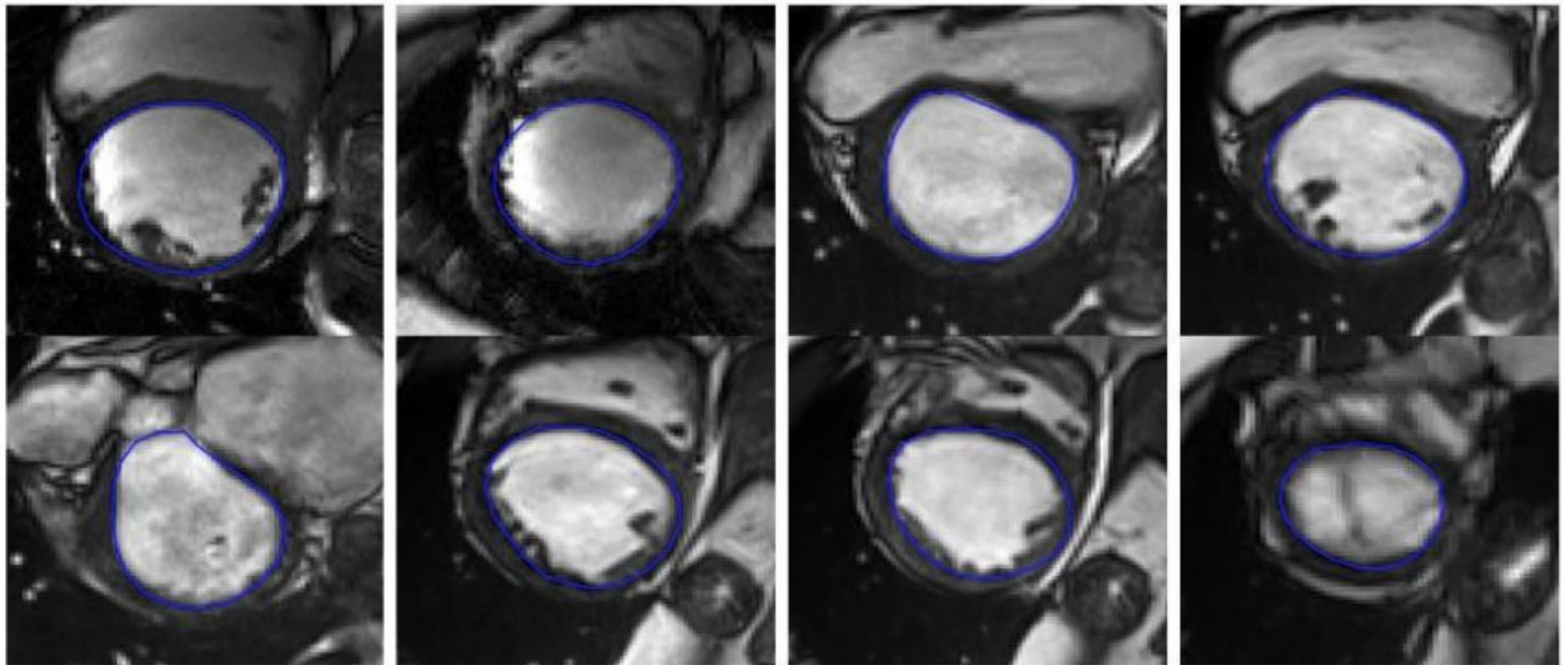


Figure 4: Sample of the boundary results with the proposed FCNet.

Results

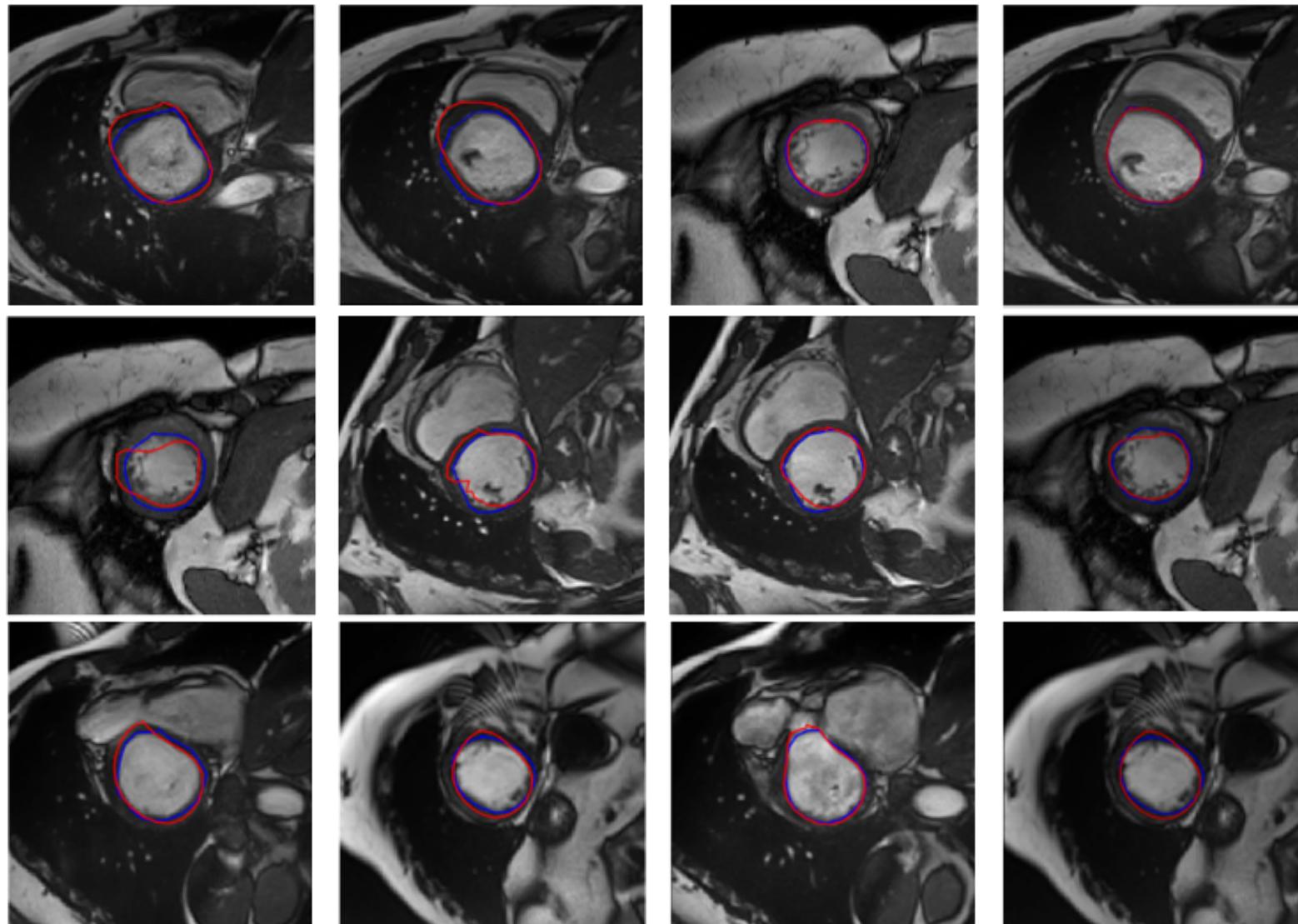


Figure 5: Representative examples of obtained borders of the LV with FCNet (blue) and moving mesh correspondence [Punithakumar et al. [2017]] (red) methods where FCNet provided significantly more accurate results than [Punithakumar et al. [2017]].

Trends

- 1) Fusion of similarity metrics or machine learning techniques;
- 2) Creation of free large datasets with correctly segmented and evaluated images in order to support future solutions;
- 3) Creation of new protocols for image registration validation and quantitative evaluation;
- 4) Optimizing deformation parameters instead of using a similarity metric;
- 5) New approaches based on Deep Learning to improve the quality of the registration itself. So far, current advances are more related to the increase of the speed and robustness of the methods;
- 6) New proposals based on meta-registration procedure, in which different registration techniques work in a collaborative and complementary way to improve results;
- 7) Generation of image-label or multi-scale-patch network registration methods to estimate larger deformations;
- 8) Implementation of CNN ensembles to learn regression parameters for each released methodology.

Postgraduate Program in Computer Science

October 15th to November 30th – Period to application;



**Group for Innovation Based on
Images and Signals.**



<http://www.unifesp.br/campus/sjc/selecao-ppgcc/mestrado-e-doutorado/editais-abertos-ppgcc.html>

Acknowledgment

- Institute of Science and Technology
- Unifesp
- GIBIS Team
- Capes
- CNPq
- FAPESP
- NVIDIA by Titan XP devices

Thank you very much!



Group for Innovation Based on
Images and Signals.

References

- Refer to our SIBGRAPI2018 tutorial paper:

“A Practical Review on Medical Image Registration: from Rigid to Deep Learning based Approches”.