Learning-Based Spatiotemporal Regularization and Integration of Tracking Methods for Regional 4D Cardiac Deformation Analysis

Allen Lu[1(✉)], Maria Zontak[6], Nripesh Parajuli[2], John C. Stendahl[3],
Nabil Boutagy[3], Melissa Eberle[3], Imran Alkhalil[3], Matthew O'Donnell[5],
Albert J. Sinusas[3,4], and James S. Duncan[1,2,4]

[1] Department of Biomedical Engineering, Yale University, New Haven, CT, USA
allen.lu@yale.edu
[2] Department of Electrical Engineering, Yale University, New Haven, CT, USA
[3] Department of Internal Medicine, Yale University, New Haven, CT, USA
[4] Department of Radiology and Biomedical Imaging, Yale University,
New Haven, CT, USA
[5] Department of Bioengineering, University of Washington, Seattle, WA, USA
[6] College of Computer and Information Science, Northeastern University,
Seattle, WA, USA

Abstract. Dense cardiac motion tracking and deformation analysis from echocardiography is important for detection and localization of myocardial dysfunction. However, tracking methods are often unreliable due to inherent ultrasound imaging properties. In this work, we propose a new data-driven spatiotemporal regularization strategy. We generate 4D Lagrangian displacement patches from different input sources as training data and learn the regularization procedure via a multi-layered perceptron (MLP) network. The learned regularization procedure is applied to initial noisy tracking results. We further propose a framework for integrating tracking methods to produce better overall estimations. We demonstrate the utility of this approach on block-matching, surface tracking, and free-form deformation-based methods. Finally, we quantitatively and qualitatively evaluate our performance on both tracking and strain accuracy using both synthetic and in vivo data.

1 Introduction

Cardiac motion estimation and regional deformation analysis are important for detection of myocardial dysfunction. Tracking methods typically follow speckles (texture patterns) or image-derived features (e.g. surfaces) over the image cycle to produce a Lagrangian *dense* motion field, where the displacement vectors at each image frame references the material point of the initial image frame. However, inherent properties of ultrasound (US) can create image artifacts that cause speckle de-correlation and poor motion tracking results. Therefore, effective regularization of the raw tracking results is essential. Various methods of

speckle tracking have been proposed in the past, including block-matching [1], optical flow [2], and registration techniques [3,4]. These methods generally apply spatial and temporal regularization separately.

Several recently proposed approaches utilize joint spatiotemporal regularization. A free-form deformation (FFD) method with 2D spatiotemporal B-splines, proposed in [3], was extended to regularize velocities (instead of displacements) using diffeomorphi FFD with 3D B-splines [4]. However, B-splines require an priori carefully defined explicit grid of points that may yield bias in the tracking data. Recently, we proposed [5] learning joint spatiotemporal cardiac motion patterns via sparse dictionaries and reconstructing noisy tracking results with the learned dictionary. However, due to the inherent limitations of K-SVD dictionaries shown in [5], the dictionary representation was applied only to the high-error trajectories, yielding limited regularization performance.

In this work, we propose a neural-network based method that *eliminates the above limitation* and applies spatiotemporal regularization to the *entire* myocardium. The regularization procedure is learned by feeding 4D Lagrangian displacement patches to a multi-layered perceptron (MLP) network [6]. We demonstrate the effectiveness of our procedure for regularization of different tracking techniques, including block matching on radio-frequency (RF) images [1], non-rigid registration using FFD [7] and a graph-based tracking method with learned weights [8]. We further propose combining complementary tracking methods using a multi-view learning framework [9]. Our experiments show that combination of complementary tracking methods lead to the best overall estimation. Finally, we apply the combined architecture on a different set of 4D echocardiography images and show the plausibility of domain adaptation. This implies that the learned regularization procedures can be adapted and applied to other echocardiography datasets for improvement of tracking and strain estimations.

In our experiments, we use 8 synthetic cardiac sequences from [10], that simulate different physiological conditions, including one **normal**, 4 sequences with occlusions in the proximal (**ladprox**) and distal (**laddist**) parts of the left anterior descending coronary artery, left circumflex (**lcx**), right coronary artery (**rca**), 3 sequences with dilated geometry with 1 synchronous (**sync**) and 2 dyssynchronous (**lbbb**, **lbbbsmall**) sequences. These sequences contain realistic US features that simulate challenges for tracking. Each sequence contains 2250 ground truth trajectories. In [11], five speckle tracking algorithm are validated and compared using these **sparse** ground-truth trajectories located at grid-intersection points. However, *regional tracking and strain validation is more appropriate for precise localization of myocardial injury.* In this work, we spatially interpolate the sparse ground truth trajectories to produce dense ground-truth displacement fields for evaluation of both dense tracking and regional strain accuracy.

2   Method

2.1   Initial Tracking Methods

We demonstrate our method on three widely used, distinctive cardiac tracking methods: radio-frequency image-based block matching (RFBM), free-form deformation (FFD), and flow-network tracking (FNT).

RFBM is a block-matching method applied to 3D radio-frequency (RF) echocardiography images in a spherical coordinate system. Given two subsequent frames, the algorithm maximizes normalized cross-correlation (NCC) between a 3D block defined around every voxel in the first RF frame and a 3D block within a search region in the second frame [1]. **FNT** tracks discrete points on the myocardium surfaces while enforcing spatial and temporal consistency in the resulting trajectories. The tracking problem is defined with a graphical framework, where the nodes represent points on the endocardial and epicardial surfaces, and edges define spatial and temporal connections among points. The edge weights are learned using a Siamese network. The objective function finds optimal trajectories that adhere to edge weights while subjected to physiological constraints [8]. **FFD** finds a global transformation given a set of fixed grid points. The grid points parameterize a B-spline transformation that best minimizes the difference between a reference frame and an adjacent frame. Spatial regularization is imposed both implicitly, via the smooth B-splines, and explicitly, via minimizing the bending energy and uses a coarse to fine optimization scheme [7]. For each method, the resulting frame-to-frame displacement field is temporally interpolated and propagated to produce Lagrangian displacement fields and sampled into X_{train} and X_{test} as illustrated in Fig. 1.

We spatially interpolate the sparse set of ground-truth trajectories, provided in [10], with radial basis functions (RBFs) using the method described in [12]. The resulting frame-to-frame displacement fields are temporally interpolated and propagated to produce the Lagrangian displacement field, where ground-truth trajectory patches Y_{train} and Y_{test} are sampled as illustrated in Fig. 1.
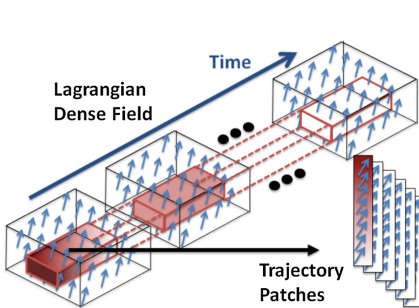
Fig. 1. Extraction of 4D spatiotemporal patches from dense displacement field
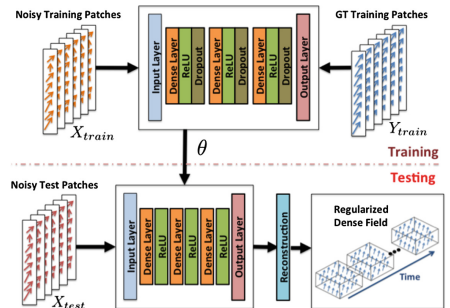
Fig. 2. Process diagram for training and testing of MLP architecture

2.2   Spatiotemporal Displacement Regularization Learning

In the training stage, given initial Lagrangian noisy tracking data, the optimal parameters θ^* are found by solving:

$$\theta^* = \arg\min_\theta \frac{1}{N} \sum_{i=0}^{N-1} \log\cosh[Y_{train}^{(i)} - f_\theta(X_{train}^{(i)})] \ , \tag{1}$$

where $Y_{train}^{(i)}$ is the ground-truth trajectory patch, and $f_\theta(X_{train}^{(i)})$ is the regularized trajectory patch for sample i over N samples. While L_2 norm (*i.e.* sum of squared distances between the patches' pixels) is widely used, we use the Mean Log-Cosh error, which is more robust to noise and outliers [13].

We approximate f_θ using a MLP network f with three fully-connected hidden layers and parameters θ. To accelerate learning, we use rectified linear units (ReLU) as our activation function. To avoid overfitting, we incorporate a dropout layer after each activation layer. Dropout randomly drops the output of each neuron during training in order to avoid co-adaptation among neurons [14].

During testing, we apply the neural network with the learned parameters θ^* onto the noisy trajectory patches X_{test} to produce corresponding regularized displacement trajectories. We then reconstruct the dense displacement field by averaging the overlapping regularized trajectories.

## 2.3   Soft-Threshold Outlier Regularization

Next, we outline our *soft-threshold regularization approach*. As described in Fig. 2, in order to train the network, we need to provide pairs of noisy - ground truth trajectory patches. However, similar to [5], we observed oversmoothing of initially well-tracked trajectories. Hence, better tracking performance was achieved when the learned regularization function was applied *only* on the outliers trajectory patches (that were detected via stacking an additional neural network). However, regularization of only selected trajectory patches created spatial displacement discontinuities that caused high derivatives and *noisy strain estimations*. Therefore, instead of applying hard threshold regularization (*i.e.* determining whether to regularize a certain trajectory), we implicitly learned *soft-threshold regularization* by simultaneously training our MLP architecture with *both ground truth-ground truth as well as noisy-ground truth pairs of data*. MLP learned to regularize both initially well-tracked trajectories via learned identity function and poorly-tracked trajectory patches via learned regularization function. Thus, the trade-off between good signal preservation and spatially smooth regularization is learned.

2.4   Combining Complementary Methods via Multiview Learning

Next we describe our multi-view MLP architecture. As illustrated in Fig. 4, RFBM performs better within the myocardium, while FNT performs better near
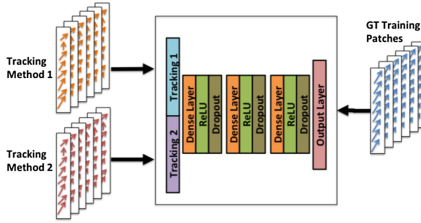
Fig. 3. Multiview Learning Architecture for integrating two tracking methods
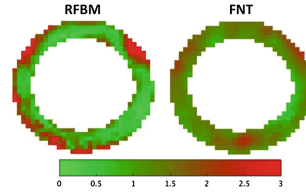
Fig. 4. RFBM vs. FNT tracking error at cross-sectional slice of myocardium for ladprox. RFBM error is higher near the boundaries but lower inside myocardium

the boundaries of the myocardium. Therefore, RFBM and FNT might complement each other in these regions. Inspired by the multi-view learning framework [9], we utilize trajectory patches from two different methods (i.e. RFBM and FNT) and combine them at the input layer of the regularization net as shown in Fig. 3.

3   Experiments and Results

We resampled each voxel to $0.5\,\mathrm{mm}^3$ with image size $75 \times 75 \times 61$ voxels. To test our method, we used a **leave-one-image-out** scheme, training on 7 images and testing on the 8th image. Training patches were sampled with a stride of 2 in each direction, and we used $5 \times 5 \times 5 \times 32 \times 3$ (3 for x–y–z directions) for normal geometry images and $5 \times 5 \times 5 \times 39 \times 3$ for dilated geometry images (around 100,000 patches). Test patches were sampled with a stride of 1 (around 22,000 patches). For each MLP, we utilized three hidden layers with 1000 neurons each along with dropout with probability of 0.2. Average test time is around $800\,\mathrm{s}$.

3.1   Quantitative Results

We quantitatively evaluated the performance of our algorithm on *dense* trajectories. Table 1 shows that applying the neural network-based spatiotemporal regularization (NNSTR) to RFBM, FNT, and FFD yielded significant improvements in tracking accuracy for all three methods over both initial tracking and dictionary learning-regularized trajectories (DL) [5]. In addition, combining RFBM and FNT in the multi-view learning framework further improved the tracking accuracy by leveraging the complementary nature of FNT and RFBM tracking.

We also analyzed our performance via regional strain analysis. We computed strain as $E_f = \frac{1}{2}[\nabla U_f + (\nabla U_f)^T + (\nabla U_f) \cdot (\nabla U_f)^T]$, where U_f is the Lagrangian dense displacement at frame f. We projected the strain tensor in clinically relevant radial (Rad.), circumferential (Cir.), and longitudinal (Long.) directions. We summarize strain performance improvements in Table 2.

Table 1. Median tracking error (mm) per frame compiled for all 8 studies for **all** trajectories within myocardium

| Methods	Error (mm)
RFBM	1.64 ± 1.78
RFBM-DL	1.48 ± 1.55
RFBM-NNSTR	0.90 ± 0.73
FNT	1.31 ± 0.95
FNT-DL	1.28 ± 0.86
FNT-NNSTR	1.05 ± 0.86
FFD	1.62 ± 1.14
FFD-DL	1.61 ± 1.12
FFD-NNSTR	1.16 ± 0.80
RFBM+FNT	$\mathbf{0.82 \pm 0.61}$

Table 2. Median strain error (%) per frame between estimated strain and ground-truth strain compiled for all 8 studies for **all** trajectories within myocardium

| Methods	Rad. (%)	Cir. (%)	Long. (%)
RFBM	21.3 ± 72.6	7.0 ± 44.0	5.9 ± 45.1
RFBM-DL	20.2 ± 33.9	4.9 ± 19.7	5.7 ± 17.5
RFBM-NNSTR	5.9 ± 10.7	2.3 ± 2.6	2.4 ± 3.4
FNT	8.1 ± 22.0	4.6 ± 12.4	6.1 ± 8.7
FNT-DL	8.2 ± 19.2	4.9 ± 10.2	6.0 ± 8.4
FNT-NNSTR	4.7 ± 11.4	2.6 ± 3.4	2.6 ± 3.7
FFD	12.3 ± 24.3	4.9 ± 6.0	7.0 ± 16.9
FFD-DL	12.1 ± 21.7	4.9 ± 5.8	6.9 ± 14.9
FFD-NNSTR	6.0 ± 10.4	3.0 ± 3.9	3.1 ± 4.1
RFBM+FNT	$\mathbf{4.0 \pm 9.8}$	$\mathbf{1.9 \pm 2.2}$	$\mathbf{2.2 \pm 2.9}$

3.2 Qualitative Results

Figure 5 shows the median strain curves within each segment of mid-cavity according to the American Heart Association (AHA) 17-segment standard. RFBM estimates radial strain poorly due to relatively high deformation (see also Table 1). FNT estimates radial strain well due to its restriction of the tracking space to myocardial surfaces and capturing high deformations. However, FNT tends to underestimate circumferential strain due to the lack of surface features that capture torsion, while RFBM captures rotational motion well. Applying NNSTR to RFBM and FNT individually indeed yielded improvement. Further combining RFBM and FNT using the proposed multi-view architecture, thus exploiting the complementary nature of these two methods, produced better overall results for both radial(Fig. 5a) and circumferential (Fig. 5b) strains. Figure 6 shows that NNSTR and combined method significantly reduced the spatial noise, producing more clinically plausible results. In the case of lcx, the combined method leveraged FNT to produce better estimation than regularizing RFBM only.

Finally, we trained the multi-view learning architecture combining RFBM and FNT with all 8 synthetic images, and we applied our learned network on a completely different set of in vivo open-chest canine data (N = 5 studies) acquired using our Philips iE33 scanner and X7-2 probe (conducted in compliance with Institutional Animal Care and Use Committee policies). For each canine study, we applied NNSTR to a baseline image and a corresponding image with occlusion in the Left Anterior Descending (LAD) artery for simulation of high stenosis. Figure 7 shows example displacements from RFBM and regularized displacements, which are smoother and physically plausible compared to the original RFBM results. Figure 8 shows example radial strain for both baseline and high stenosis case. We noticed again that our multi-view architecture learns the complementary nature of FNT and RFBM and produced radial strain that resembles

(a) **Radial** Strain for **lbbb** (dyssynchronous)

(b) **Circumferential** Strain for **ladprox**(occlusion in LAD)
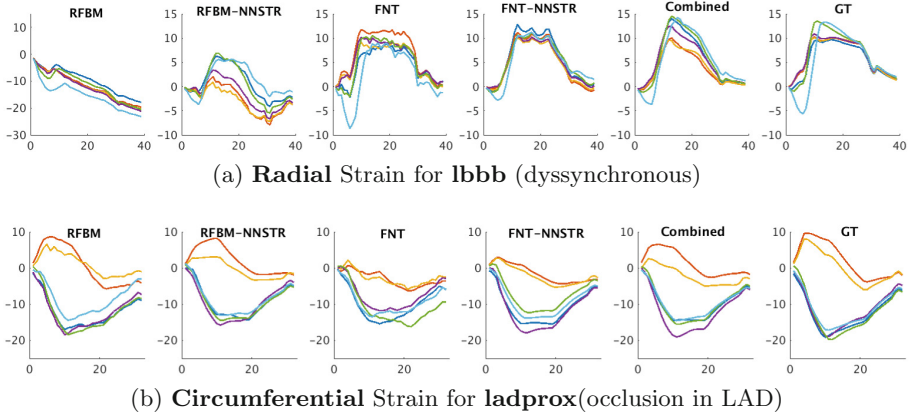
Fig. 5. Strain (%) vs. time in the mid-cavity according to the AHA 17-segment model

(a) Radial Strain for **normal**   (b) Radial Strain for **lcx**
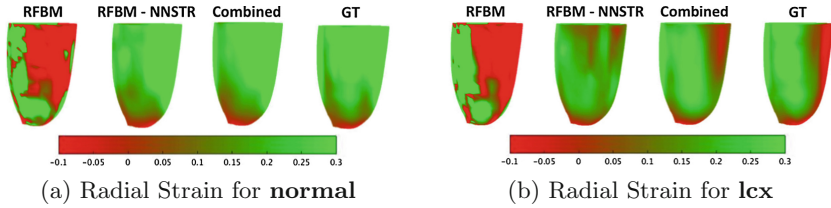
Fig. 6. Radial strain during end-systole produced with RFBM, regularized RFBM, combined method, and GT interpolated at epicardium

FNT. Finally, we expected to see motion abnormalities at the Left Ventricle(LV)-Right Ventricle(RV) junction due to occlusion in the LAD. This is captured in the radial strain map of the combined method.
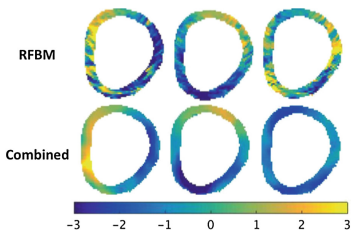
Fig. 7. Displacements at end-systole from **canine images** in the horizontal, vertical, and longitudinal directions for RFBM and regularized RFBM with combined architecture
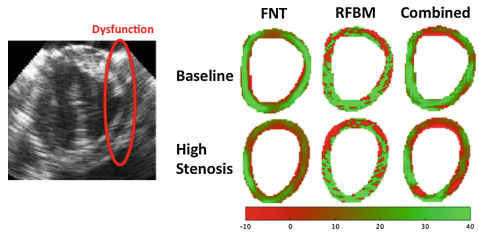
Fig. 8. Radial strain (%) at end-systole from FNT, RFBM, and combined architecture estimated on **canine images**. Strain from combined method shows expected dysfunction from LV-RV junction

4   Conclusions

In this work, we proposed a learning-based method for spatiotemporal regularization of myocardial tracking. The regularization procedure was learned by feeding 4D Lagrangian displacement trajectories to a multi-layered perceptron (MLP) network. We showed effectiveness of our method on three distinct tracking methods: RF-block matching (RFBM), non-rigid registration (FFD), and a graph-based myocardial surface tracking method (FNT). We further proposed a multi-view learning framework that learned to leverage the complementary nature of FNT and RFBM to produce better estimations than individual regularization. Finally, we showed how our learned regularization model can potentially be applied to other echocardiography datasets via domain adaptation.

References

1. Chen, X., Xie, H., Erkamp, R., Kim, K., Jia, C., Rubin, J., O'Donnell, M.: 3-D correlation-based speckle tracking. Ultrason. Imaging **27**(1), 21–36 (2005)
2. Alessandrini, M., Liebgott, H., Barbosa, D., Bernard, O.: Monogenic phase based optical flow computation for myocardial motion analysis in 3D echocardiography. In: Camara, O., Mansi, T., Pop, M., Rhode, K., Sermesant, M., Young, A. (eds.) STACOM 2012. LNCS, vol. 7746, pp. 159–168. Springer, Heidelberg (2013). doi:10. 1007/978-3-642-36961-2_19
3. Ledesma-Carbayo, M.J., Kybic, J., Desco, M., Santos, A., Suhling, M., Hunziker, P., Unser, M.: Spatio-temporal nonrigid registration for ultrasound cardiac motion estimation. IEEE Trans. Med. Imaging **24**(9), 1113–1126 (2005)
4. De Craene, M., Piella, G., Camara, O., Duchateau, N., Silva, E., Doltra, A., Dhooge, J., Brugada, J., Sitges, M., Frangi, A.F.: Temporal diffeomorphic free-form deformation: application to motion and strain estimation from 3D echocardiography. Med. Image Anal. **16**(2), 427–450 (2012)
5. Lu, A., Zontak, M., Parajuli, N., Stendahl, J.C., Boutagy, N., Eberle, M., ODonnell, M., Sinusas, A.J., Duncan, J.S.: Dictionary learning-based spatiotemporal regularization for 3D dense speckle tracking. In: SPIE Medical Imaging, International Society for Optics and Photonics, p. 1013904 (2017)
6. Hornik, K., Stinchcombe, M., White, H.: Multilayer feedforward networks are universal approximators. Neural Netw. **2**(5), 359–366 (1989)
7. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J.: Non-rigid registration using free-form deformations: application to breast MR images. IEEE Trans. Med. Imaging **18**(8), 712–721 (1999)
8. Parajuli, N., Lu, A., Stendahl, J.C., Zontak, M., Boutagy, N., Eberle, M., Alkhalil, I., ODonnell, M., Sinusas, A.J., Duncan, J.S.: Flow network based cardiac motion tracking leveraging learned feature matching. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (2017)
9. Sun, S.: A survey of multi-view machine learning. Neural Comput. Appl. **23**(7–8), 2031–2038 (2013)

10. Alessandrini, M., De Craene, M., Bernard, O., Giffard-Roisin, S., Allain, P., Waechter-Stehle, I., Weese, J., Saloux, E., Delingette, H., et al.: A pipeline for the generation of realistic 3D synthetic echocardiographic sequences: methodology and open-access database. IEEE Trans. Med. Imaging. **34**(7), 1436–1451 (2015)
11. Alessandrini, M., Heyde, B., Queirós, S., Cygan, S., Zontak, M., Somphone, O., Bernard, O., Sermesant, M., Delingette, H., Barbosa, D., et al.: Detailed evaluation of five 3D speckle tracking algorithms using synthetic echocardiographic recordings. IEEE Trans. Med. Imaging **35**(8), 1915–1926 (2016)
12. Parajuli, N., Compas, C.B., Lin, B.A., Sampath, S., O'Donnell, M., Sinusas, A.J., Duncan, J.S.: Sparsity and biomechanics inspired integration of shape and speckle tracking for cardiac deformation analysis. In: van Assen, H., Bovendeerd, P., Delhaas, T. (eds.) FIMH 2015. LNCS, vol. 9126, pp. 57–64. Springer, Cham (2015). doi:10.1007/978-3-319-20309-6_7
13. He, R., Hu, B., Yuan, X., Wang, L.: Robust Recognition via Information Theoretic Learning. SCS. Springer, Cham (2014)
14. Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**(1), 1929–1958 (2014)