An Unsupervised Learning Approach to Discontinuity-Preserving Image Registration

Eric Ng and Mehran Ebrahimi[✉]

Imaging Lab, Faculty of Science, Ontario Tech University,
2000 Simcoe Street North, Oshawa, ON L1H 7K4, Canada
{eric.ng,mehran.ebrahimi}@ontariotechu.ca
http://www.ImagingLab.ca

Abstract. Most traditional image registration algorithms aimed at aligning a pair of images impose well-established regularizers to guarantee smoothness of unknown deformation fields. Since these methods assume global smoothness within the image domain, they pose issues for scenarios where local discontinuities are expected, such as the sliding motion between the lungs and the chest wall during the respiratory cycle. Furthermore, an objective function must be optimized for each given pair of images, thus registering multiple sets of images become very time-consuming and scale poorly to higher resolution image volumes.

Using recent advances in deep learning, we propose an unsupervised learning-based image registration model. The model is trained over a loss function with a custom regularizer that preserves local discontinuities, while simultaneously respecting the smoothness assumption in homogeneous regions of image volumes. Qualitative and quantitative validations on 3D pairs of lung CT datasets will be presented.

1 Introduction

Image registration is an invaluable tool for medical image analysis and has received vast attention in imaging research for the past several decades. Image registration is used as a tool to find meaningful temporal transformations to align images taken at different time frames. Traditionally, registration algorithms assume smooth transformations. This assumption quickly falls apart for many cases, since different organs move, to a certain degree, independently from one another. Image misalignment becomes inevitable if smoothness is assumed at regions where discontinuities are expected, such as organ boundaries [4]. In this paper, we introduce an unsupervised learning model that learns the relationship between image pairs and a corresponding displacement field. We propose a regularizer that accounts for local image discontinuities while simultaneously respecting local homogeneity. This approach drastically decreases registration time, as the registration task is no longer an optimization task, but becomes a simple function evaluation.

2   Related Work

In traditional image registration, the most common approach is to solve an optimization problem, where the objective function is comprised of two terms, an image dissimilarity term and a regularization term to restrict the solution space. Common methods include elastic and diffusion models [16], free-form deformations using b-splines [3], and more recently, kernel methods [11–13]. Because all of these methods optimize an energy function for every image pair, large-scale or successive registration tasks becomes very time consuming. Specialized algorithms such as Thirion's Demons [5,17,22] allow significant reduction in computational time by estimating force vectors that acts to drive the deformation followed by Gaussian smoothing during the optimization process. Unfortunately, this algorithm restricts models to be diffusion-based models only.

With the rise of deep learning over the past decade, learning-based approaches have become extremely popular. Several models are trained in a supervised manner which required ground truth transformations to be available [6,15,18]. Although these methods showed promising results, the task of obtaining ground truth transformation fields is cumbersome and highly prone to error. Thus, recent methods have shifted to an unsupervised approach instead, where models are trained based on how transformation fields act on images, rather than strictly on the transformations [1,9,25]. For a survey of learning-based image registration methods, refer to the article by Haskins *et al.* [10].

3   Method

Our model follows a framework popularized by Voxelmorph [2]. Let I_F and I_M denote fixed and moving images. We find a function $g_\theta(I_F, I_M)$ that produces the displacement field $\mathbf{u}$, i.e. $\mathbf{u} = g_\theta(I_F, I_M)$. The deformation ϕ can then expressed as the mapping $\phi = Id + \mathbf{u}$ where Id is the identity mapping. The deformation field is applied to I_M to produce the warped image $I_M \circ \phi$ where $I_F(x)$ is similar to $[I_M \circ \phi](x)$ for all voxel locations $x \in \Omega$. Since ϕ may map the original coordinate system to non-integer valued voxel locations, interpolation is required to warp I_M under ϕ. For our experiments, we use trilinear interpolation due to its simplicity. An overview of the model is shown in Fig. 1.

3.1   Network Architecture

The function g_θ is modeled using a convolutional neural network where θ denotes the network parameters. The neural network follows a modified version of U-Net [19], which contains an encoder and a decoder structure that mirror each other and are connected by skip connections at each layer (Fig. 2). The encoder/decoder architecture is motivated by image pyramid techniques in many computer vision algorithms, where each encoding and decoding layer operate from coarse to fine representations of the input.
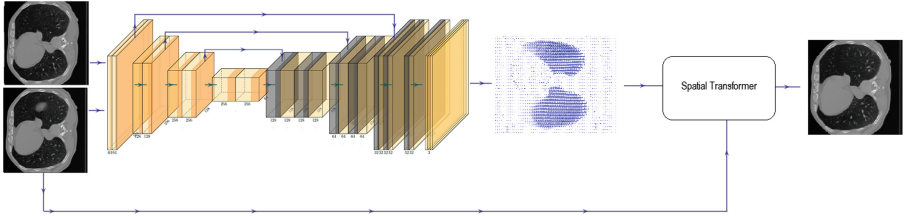
Fig. 1. Overview of the model. Fixed and moving images I_F, I_M are passed into a convolutional neural network which produces the displacement field $\mathbf{u}(x)$. The spatial transformer morphs I_M based on the displacement field. The loss is measured over the dissimilarity between the fixed and morphed moving images, as well as additional penalty functions defined over $\mathbf{u}$.

The encoder consists of three convolution layers by applying $3 \times 3 \times 3$ convolutions with stride 2 for downsampling, followed by LeakyReLU with slope of 0.2 at each layer. Each convolution layer has 32 output channels except the first layer which contains 16 output channels.

The decoder follows a similar structure as the encoder but in reverse order. In the first decoding layer, we simply use the output of the final encoding layer as the input. In subsequent decoding layers, we first upsample the output of the previous decoding layer. Skip connections are constructed by concatenating layer outputs with that of the mirroring encoding layer. This effectively uses representations of the encoding layers to enforce more precise outputs in the decoding layers. Similar to the encoder, each decoding layer applies $3 \times 3 \times 3$ convolutions followed by LeakyReLU of slope 0.2, but with stride 1 to preserve resolution at each layer. The output of the final decoding layer is passed into an additional convolution layer with 3 output channels, where each output channel contains the coordinate components of the displacement field $\mathbf{u}$.

3.2   Loss Function

We train our model using a loss function in the form

$$\mathcal{L}(I_F, I_M, \mathbf{u}) = \lambda_{sim}\mathcal{L}_{sim}(I_F, I_M, \mathbf{u}) + \lambda_{disc}\mathcal{L}_{disc}(\mathbf{u}) + \lambda_{mag}\mathcal{L}_{mag}(\mathbf{u}), \quad (1)$$

where $\mathcal{L}_{sim}$ measures image dissimilarity, $\mathcal{L}_{disc}$ is a discontinuity preserving regularizer, and $\mathcal{L}_{mag}$ is a second loss term that manages the (ir)regularities in the magnitude of the displacement fields. $\lambda_{sim}, \lambda_{disc}$, and λ_{mag} are corresponding regularization constants.

Similarity Loss. To measure image similarity/dissimilarity, we use a local normalized cross correlation which is defined as
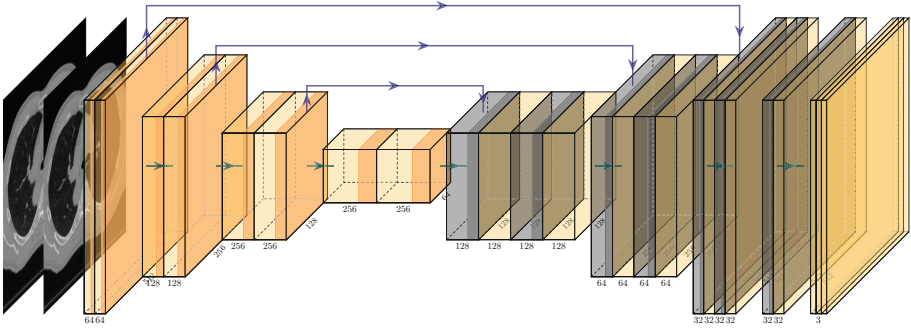
Fig. 2. Network architecture of g_θ based on a modified version of U-Net. The network receives I_M and I_F to produce the displacement field $\mathbf{u}$. The input and output of the network are of dimensions $D \times H \times W \times 2$ and $D \times H \times W \times 3$ respectively. The architecture consists of a contractive path (encoder) and a mirroring expansive path (decoder) connected by skip connections at each layer.

$$\text{LNCC}(I_M, I_F) = \sum_{x \in \Omega} \frac{\left[\sum_{y \in \mathcal{N}(x)} \left(I_M(y) - \mu_M(x)\right)\left(I_F(y) - \mu_F(x)\right)\right]^2}{\left[\sum_{y \in \mathcal{N}(x)} \left(I_M(y) - \mu_M(x)\right)^2\right]\left[\sum_{y \in \mathcal{N}(x)} \left(I_F(y) - \mu_F(x)\right)^2\right]} \tag{2}$$

where x is any voxel in the image domain Ω, and $y \in \mathcal{N}(x)$ are the neighborhood points around voxel x, and $\mu_M(x)$ and $\mu_F(x)$ are the average local intensities around x in the moving and fixed images, respectively. LNCC is maximized when $I_F = I_M$ which measures similarity, thus we define the dissimilarity measure as $\mathcal{L}_{\text{sim}} = 1 - \text{LNCC}$.

Discontinuous Loss. In designing the discontinuous loss, we first assume that there are no topological changes, i.e. no new tissue is introduced nor destroyed. We then consider the requirements based on these physical scenarios: 1. Homogeneous movement, 2. Movement along rigid structures, and 3. Sliding organs.

These scenarios help us define the requirements for our regularizer. Firstly, the regularizer must preserve smooth local deformations that occur locally within organ interiors. Secondly, the regularizer must not penalize large local changes in deformation magnitude as long as the movement is in a similar direction. This is to mimic the movement of soft tissues or organs against rigid structures such as the rib cage or the spinal column. Finally, the regularizer must be able to account for movements in the opposite directions along organ boundary. This final requirement is perhaps the most significant as there are many scenarios where sliding organs exist. Common examples include the sliding of the lungs against the chest wall during the respiratory cycle, and the movement of organs against one another in the abdominal region. Figure 3 visually summarizes possible desired behaviors of a discontinuous displacement field.
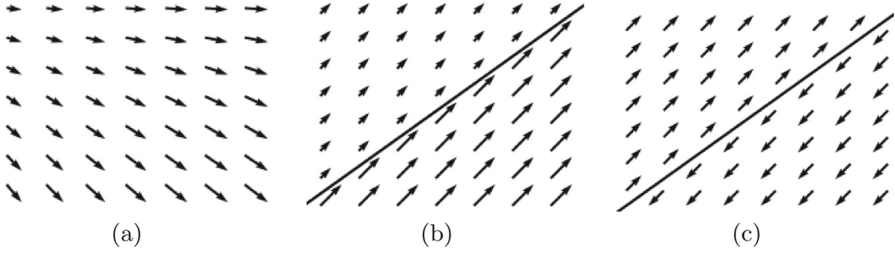
(a)   (b)   (c)

Fig. 3. Desired behaviors of the discontinuous displacement field. Figure 4(a) demonstrates local homogeneity which is expected within organs. Figure 4(b) allows displacement vectors of different magnitudes as long as they are in a similar direction, which represents soft tissue moving against rigid structures. (c) depicts sliding boundary conditions as displacement vectors on opposite sides of the boundary travel in opposite directions.

Let $\mathbf{u}$ be represented by a collection of displacement vectors $\{u_i\}_{i=1,...,N}$, where N is the number of voxels in the image. Now consider two arbitrary vectors u_i and u_j, respectively corresponding to locations x_i and x_j in the image domain. The area of the parallelogram spanned by u_i and u_j is maximized when u_i and u_j is orthogonal to one another, and minimized when they are parallel. Thus the three conditions are encouraged for any regularizer in the form

$$\mathcal{L}_{disc} = \sum_{i,j=1}^{N} g(\mathcal{P}(u_i, u_j)) \tag{3}$$

where $\mathcal{P}$ the unsigned area of the parallelogram spanned by u_i and u_j, and $g : \mathbb{R} \to \mathbb{R}$ is a strictly increasing function satisfying $g(0) = 0$. $\mathcal{P}$ is computed as

$$\mathcal{P}(u_i, u_j) = \|u_i \times u_j\|_2 \tag{4}$$

where $\times$ denotes the cross product. We propose the regularizer

$$\mathcal{L}_{disc} = \sum_{i,j=1}^{N} \frac{1}{2} \log\left(1 + \mathcal{P}(u_i, u_j)^2\right) k(x_i, x_j) \tag{5}$$

where $k(x_i, x_j)$ is a decreasing weight function that depends on the proximity between the locations x_i and x_j. For our experiments, we choose the C^4 Wendland kernel [26] for $k(x_i, x_j)$.

Magnitude Loss. During preliminary stages of our experiments, we noticed that deformations in large dark image regions (background of CT image, for instance) behave erratically. We found that imposing an additional magnitude-based regularizer is needed to suppress this unpredictable behavior. Thus we add the following term to our loss function

$$\mathcal{L}_{mag}(u) = \max_i(\|u_i\|_2). \tag{6}$$

This effectively discourages large magnitudes of u. Evidently, this additional term may become problematic for coarse registration where large-scale movement may be expected. However, since this is aimed towards addressing local discontinuities, it is safe to assume that deformations remain relatively small.

4   Experiments

4.1   Setup

Our model is implemented using PyTorch 1.3.0 and trained using an NVIDIA GeForce GTX 1080Ti with 11 GB of graphics memory. CPU tests are performed on an Intel Xeon E5-1620 at 3.7 GHz. We trained our model using Adam optimizer [14] with $\lambda_{sim} = 100, \lambda_{disc} = \lambda_{mag} = 1$, and learning rate 10^{-4}.

The model is evaluated over 4DCT datasets provided by DIR-Lab [7,8] and the POPI-model [23]. The DIR-Lab Reference 4DCT datasets contain ten sets of image volumes of sizes 256×256 and 512×512 with various number of axial slices (average of 100 and 128 for the two respective resolutions). To account for these variations, we only keep the middle 96 axial slices of the 256×256 volume, and the middle 112 axial slices of the 512×512 volumes. Each set of image volumes are taken over 10 time steps over the period of a single respiratory cycle. Since the input is a pair of image volumes, I_F is chosen as the image volume with a randomly chosen case number and time step, and I_M is selected based on the same case number with a different time step. By choosing eight cases as training data, this allows $8 \times 10 \times 9 = 720$ training samples and $2 \times 10 \times 9 = 180$ test samples, despite only having ten available cases. The POPI-Model contains six image volumes of sizes 512×512 with 140 to 190 axial slices. For consistency, we only keep the middle 136 axial slices and use five of the six cases as training data. We follow the same approach as DIR-Lab in choosing I_F and I_M.

4.2   Results

We first compare our discontinuity-preserving model with one that assumes global smoothness. As a baseline, we trained a second model using the DIR-lab dataset with an identical configuration, with the exception where the discontinuous loss $\mathcal{L}_{disc}$ is replaced with a total variation loss $\mathcal{L}_{TV}$ defined as

$$\mathcal{L}_{TV} = \sum_i \|\nabla u_i\|_2 \tag{7}$$

where the summation is over all voxels indexed by i. Figure 4 shows a comparison between our model trained using $\mathcal{L}_{TV}$ and $\mathcal{L}_{disc}$. One can quickly identify sudden changes in the displacement field near the lung's boundaries especially near the lung/vertebrae interface. Additional registration results are shown in Fig. 5. We compare our results (Table 1) quantitatively to the following methods: Free-Form Deformations (FFD) [20], isotropic parametric Total Variation (pTV) [24], and Sparse Kernel Machines (SKM) [12]. For comparison, we fixed frame
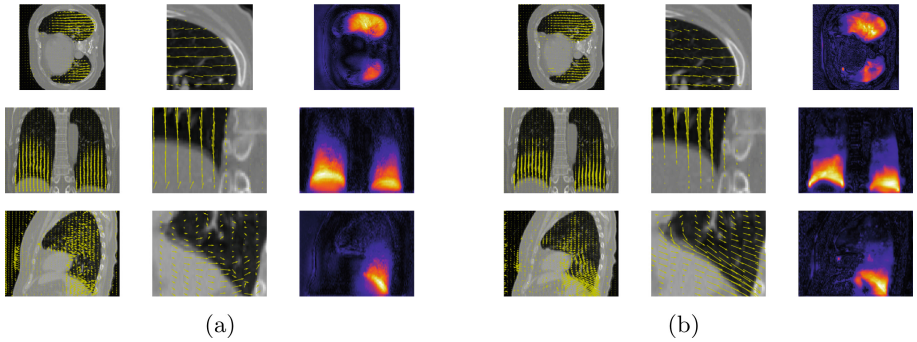
Fig. 4. Results obtained using $\mathcal{L}_{TV}$ (a) and $\mathcal{L}_{disc}$ (b). Columns 1 and 4 show an overlay of $\mathbf{u}$ over I_M. Columns 2 and 5 show a magnified local region where transformation discontinuities are expected. Columns 3 and 6 are heatmaps of the displacement field's local magnitudes.
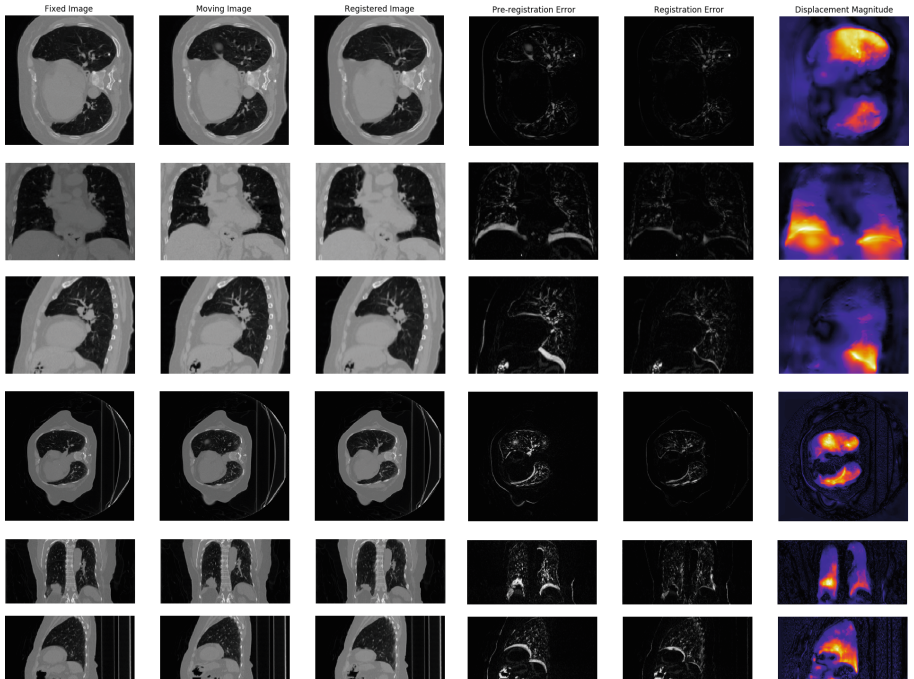

Fig. 5. Qualitative results of proposed model. From left to right: fixed image I_F, moving image I_M, registered image $I_M \circ \phi$, absolute error before registration $|I_F - I_M|$, absolute error after registration $|I_F - I_M \circ \phi|$, heatmap of displacement field $\mathbf{u}$.

Table 1. Target Registration Error (TRE) in millimeters (mm) against FFD [20], pTV [24], and SKM [12] on the DIR-Lab and POPI 4DCT Model. Baseline model is the same configuration but trained with $\mathcal{L}_{TV}$ in place of $\mathcal{L}_{disc}$.

Frame	DIR-Lab 256					DIR-Lab 512					POPI				
	FFD	pTV	SKM	Base	Ours	FFD	pTV	SKM	Base	Ours	FFD	pTV	SKM	Base	Ours
0	1.01	0.92	1.06	1.10	1.04	0.79	0.62	0.59	0.77	0.65	0.79	0.72	0.66	0.77	0.76
2	0.99	0.89	0.85	0.94	0.91	0.81	0.63	0.57	0.73	0.64	0.81	0.71	0.65	0.73	0.74
3	1.29	1.34	1.32	1.26	1.24	1.14	0.99	1.01	0.97	1.00	1.14	1.12	1.17	1.08	1.11
4	1.26	1.27	1.25	1.23	1.26	1.11	0.92	0.93	0.96	0.95	1.11	1.01	1.07	1.04	1.07
5	1.27	1.29	1.35	1.29	1.31	1.11	1.00	1.01	0.99	1.02	1.11	1.11	1.13	1.10	1.16
6	1.31	1.17	1.18	1.27	1.25	1.20	0.90	0.89	1.02	0.92	1.20	1.03	1.00	1.11	1.06
7	1.36	1.19	1.22	1.25	1.30	1.20	0.95	0.93	1.00	1.01	1.20	1.06	1.05	1.09	1.13
8	1.10	1.05	0.94	1.04	1.07	0.88	0.73	0.67	0.78	0.79	0.88	0.84	0.75	0.88	0.90
9	1.09	0.97	0.99	1.07	1.09	0.92	0.70	0.75	0.78	0.80	0.92	0.81	0.83	0.86	0.89

Table 2. Comparison of registration time between learning-based model and inverse model. For the learning-based model, we used our proposed model for evaluation. For the inverse model, we perform pairwise registration with diffusion regularizer over 1,000 iterations. The inverse model is evaluated using the AIRLab framework [21]. The CPU time for the classical model over DIR-Lab 512 and POPI Model is not computed, as they were much higher than the corresponding GPU time. Time is measured in seconds.

	Learning-based		Inverse Model	
	GPU	CPU	GPU	CPU
DIR-lab 256	0.33	15.70	82.57	5724.36
DIR-lab 512	1.38	63.14	532.41	–
POPI Model	1.67	76.45	702.86	–

1 as the fixed image, and register all remaining frames to the reference. Finally, we compare the time required to register a pair of images using our approach versus a classical registration algorithm using minimization (Table 2). Classical registration is applied using the AIRLab framework [21] via diffusion regularizer.

5   Conclusion and Future Work

We presented an unsupervised learning-based model for discontinuity preserving image registration. Although the training set was relatively small, our model performed on par with existing methods while begin able to handle locations where discontinuities may occur. Furthermore, our model significantly reduced computation by several orders of magnitude, allowing successive registration to be performed within a relatively short time frame. A drawback of the model is its sensitivity to noise. In particular, since $\mathcal{L}_{disc}$ is computed by comparing local displacement vectors with neighboring displacement vectors individually, there are no mechanisms to discourage local chaotic behaviors in the displacement field. A possible remedy is to extend the current model to incorporate

additional information, such as segmentation masks and edge information. This allows image discontinuities to be defined rather than relying on only image intensities to predict boundary regions.

References

1. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: An unsupervised learning model for deformable medical image registration. In: Proceedings of the IEEE Conference on CVPR, pp. 9252–9260 (2018)
2. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: VoxelMorph: a learning framework for deformable medical image registration. IEEE Trans. Med. Imaging **38**, 1788–1800 (2019)
3. Balci, S.K., Golland, P., Shenton, M.E., Wells, W.M.: Free-form B-spline deformation model for groupwise registration. In: MICCAI (2007)
4. Berendsen, F.F., Kotte, A.N., Viergever, M.A., Pluim, J.P.: Registration of organs with sliding interfaces and changing topologies. In: Medical Imaging 2014, vol. 9034, p. 90340E. International Society for Optics and Photonics (2014)
5. Cahill, N.D., Noble, J.A., Hawkes, D.J.: A demons algorithm for image registration with locally adaptive regularization. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 574–581. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04268-3_71
6. Cao, X., et al.: Deformable image registration based on similarity-steered CNN regression. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 300–308. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_35
7. Castillo, E., Castillo, R., Martinez, J., Shenoy, M., Guerrero, T.: Four-dimensional deformable image registration using trajectory modeling. Phys. Med. & Biol. **55**(1), 305 (2009)
8. Castillo, R., et al.: A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. Phys. Med. Biol. **54**(7), 1849 (2009)
9. Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R.: Unsupervised learning for fast probabilistic diffeomorphic registration. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11070, pp. 729–738. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_82
10. Haskins, G., Kruger, U., Yan, P.: Deep learning in medical image registration: a survey. arXiv preprint arXiv:1903.02026 (2019)
11. Jud, C., Möri, N., Bitterli, B., Cattin, P.C.: Bilateral regularization in reproducing kernel hilbert spaces for discontinuity preserving image registration. In: Wang, L., Adeli, E., Wang, Q., Shi, Y., Suk, H.-I. (eds.) MLMI 2016. LNCS, vol. 10019, pp. 10–17. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-47157-0_2
12. Jud, C., Mori, N., Cattin, P.C.: Sparse kernel machines for discontinuous registration and nonstationary regularization. In: Proceedings of the IEEE Conference on CVPR Workshops, pp. 9–16 (2016)

13. Jud, C., Sandkühler, R., Möri, N., Cattin, P.C.: Directional averages for motion segmentation in discontinuity preserving image registration. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 249–256. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_29

14. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: International Conference on Learning Representations (ICLR) (2015)

15. Krebs, J., et al.: Robust non-rigid registration through agent-based action learning. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 344–352. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_40

16. Modersitzki, J.: FAIR: flexible algorithms for image registration, vol. 6. SIAM (2009)

17. Pennec, X., Cachier, P., Ayache, N.: Understanding the "Demon's Algorithm": 3D non-rigid registration by gradient descent. In: Taylor, C., Colchester, A. (eds.) MICCAI 1999. LNCS, vol. 1679, pp. 597–605. Springer, Heidelberg (1999). https://doi.org/10.1007/10704282_64

18. Rohé, M.-M., Datar, M., Heimann, T., Sermesant, M., Pennec, X.: SVF-Net: learning deformable image registration using shape matching. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 266–274. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_31

19. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

20. 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)

21. Sandkühler, R., Jud, C., Andermatt, S., Cattin, P.C.: AirLab: autograd image registration laboratory. arXiv preprint arXiv:1806.09907 (2018)

22. Thirion, J.P.: Image matching as a diffusion process: an analogy with Maxwell's demons. Med. Image Anal. **2**(3), 243–260 (1998)

23. Vandemeulebroucke, J., Rit, S., Kybic, J., Clarysse, P., Sarrut, D.: Spatiotemporal motion estimation for respiratory-correlated imaging of the lungs. Med. Phys. **38**(1), 166–178 (2011)

24. Vishnevskiy, V., Gass, T., Szekely, G., Tanner, C., Goksel, O.: Isotropic total variation regularization of displacements in parametric image registration. IEEE Trans. Med. Imaging **36**(2), 385–395 (2016)

25. de Vos, B.D., Berendsen, F.F., Viergever, M.A., Staring, M., Išgum, I.: End-to-end unsupervised deformable image registration with a convolutional neural network. In: Cardoso, M.J., et al. (eds.) DLMIA/ML-CDS -2017. LNCS, vol. 10553, pp. 204–212. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67558-9_24

26. Wendland, H.: Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. Adv. Comput. Math. **4**(1), 389–396 (1995)