Research Article

Diffeomorphic unsupervised deep learning model for mono- and multi-modality registration

Anis Theljani[1] 🅐 and Ke Chen[2]

Abstract

Different from image segmentation, developing a deep learning network for image registration is less straightforward because training data cannot be prepared or supervised by humans unless they are trivial (e.g. pre-designed affine transforms). One approach for an unsupervised deep leaning model is to self-train the deformation fields by a network based on a loss function with an image similarity metric and a regularisation term, just with traditional variational methods. Such a function consists in a smoothing constraint on the derivatives and a constraint on the determinant of the transformation in order to obtain a spatially smooth and plausible solution. Although any variational model may be used to work with a deep learning algorithm, the challenge lies in achieving robustness. The proposed algorithm is first trained based on a new and robust variational model and tested on synthetic and real mono-modal images. The results show how it deals with large deformation registration problems and leads to a real time solution with no folding. It is then generalised to multi-modal images. Experiments and comparisons with learning and non-learning models demonstrate that this approach can deliver good performances and simultaneously generate an accurate diffeomorphic transformation.

Introduction

Image registration consists in constructing a reasonable geometrical correspondence between given two or more images of the same object taken at different times or using the same or different devices in order to locate different or complementary information. Applications of image registration include diverse fields such as astronomy, optics, biology, chemistry, remote sensing and particularly in medical imaging. For an overview of image registration methodology, approaches and applications, we refer to Fischer and Modersitzki[1]; Gigengack et al.[2]; Modersitzki[3]; Oliveira et al.[4]; Sotiras.[5] Though the topic is actively studied and useful models exist, there remain many challenges to be tackled mathematically, particularly in registration of images from different modalities. There exist various deformable variational models for image registration where the unknown displacement field **u** is sought in a properly chosen functional space.[6–11] Generally speaking, the variational problem consists in solving the optimisation problem

$$\min_{\mathbf{u}} \left\{ \mathcal{L}(\mathbf{u}) = S(\mathbf{u}) + \frac{\lambda}{2} D(T(\varphi), R) \right\} \qquad (1)$$

where $\varphi(\boldsymbol{x}) = \boldsymbol{x} + \mathbf{u}(\boldsymbol{x})$. In (1), $S(\mathbf{u})$ is a regularisation term which controls the smoothness of **u** and reflects our expectations by penalising unlikely

[1]EPSRC Liverpool Centre for Mathematics in Healthcare, University of Liverpool, Liverpool, UK
[2]Department of Mathematical Sciences and EPSRC Liverpool Centre for Mathematics in Healthcare, University of Liverpool, Liverpool, UK

Corresponding author:
Ke Chen, Department of Mathematical Sciences and EPSRC Liverpool Centre for Mathematics in Healthcare, University of Liverpool, Liverpool, UK.
Email: k.chen@liverpool.ac.uk

transformations. The second part $D(T(\varphi), R)$ is a similarity term which measures the goodness of the registration. These models are called non-learning based models as the optimisation problem (1) should be solved for each pair for images T and R. Although various non-learning-based models have been proposed in the recent years and many numerical and computational algorithms have been developed to accelerate the numerical resolution of these models, it remains a very challenging question of achieving both an accurate solution and fast speed for real time applications.

In recent years, deep learning approaches were proposed where the aim is to optimise and learn spatial transformations between pairs of images to be registered,[12–16] often, they require ground-truth deformation fields for the training task. They are called supervised models and their main drawback is the inability to predict transformations that may not be in the same range or class of the training transformations. As example, a deep learning model, which is learned and trained on a dataset where the ground-truth contains only small displacement fields, fails to predict and to give accurate results for large displacement.

In order to remedy these drawbacks, another class of deep leaning models was proposed. These unsupervised models do not require ground-truth deformation fields for training. The deformation fields are self-trained and driven by image similarity metrics computed on the input data. In Jaderberg,[12] a spatial transformer network is developed to learn transformations for 2D images; however only affine and thin plate spline transformations were used. More general nonparametric transformations were considered in Haskins et al.[13]; Li and Fan[14]; Theljani and Chen [6,25] for monomodal images. In these approaches, the transformations are controlled by penalising the derivatives of the deformation **u** in the loss function, which promotes smoothness of predicted transformations. However, this does not guarantee some physicality desired geometric properties of the deformation such as the topology preservation and invertibility. Thus, folding problem can occur in the results which is in practise inappropriate for real life problems where deformation is large and folding can occur. In terms of application, these approaches were only trained and tested on mono-modal images.

To overcome the folding problem, some deep learning approaches have addressed the question of getting diffeomorphic transformations, i.e. topology-preserving and invertible. These conditions were guaranteed by either adding an extra layer to the networks to enforce the output deformation to be diffeomorphic as in Krebs et al.,[17] or by controlling the back-and-forth registration as in Kuang.[18] In this case, the

network is learning a deformation that maps T to R and its possible inverse deformation that maps R to T. Moreover, these approaches were only applied for mono-modal images

In this paper, we propose an alternative approach to overcome the folding problem. The model is unsupervised, i.e. do not require any ground truth data, and is applied for registering mono- and multi-modal images. It can deliver diffeomorphic transformations without adding any extra layer to the neural network or enforcing it compute a deformation and its inverse. In fact, the diffeomorphisms are guaranteed by using a suitable loss function for training that controls the folding in the deformations.

A learning model

In this section, we introduce a learning based diffeomorphic model for both mono- and multi-modal image registration. The idea is that the deformation fields are self-trained by minimizing a specific loss function that guarantees the transformation to be physical. In Figure 1, we present an overview of the method:

(i) The pair of images T and R are concatenated and fed to the registration network $N(T, R)$ as a single multichannel image. The registration network, with weight parameters θ, processes the two images through a series of convolutional and pooling layers, and outputs a 2-channel map representing the 2D deformation field, denoted by $\mathbf{u}_\theta$. Any
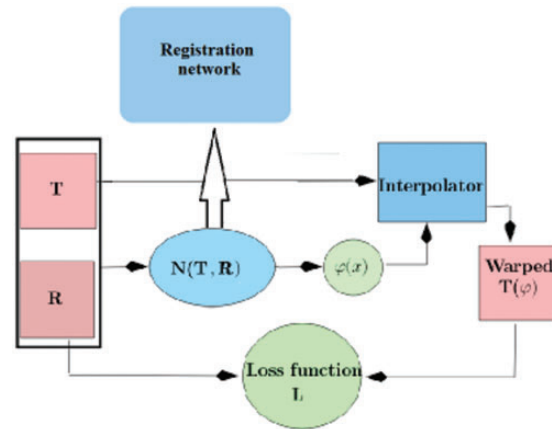

Figure 1. The work-flow of the registration model. The fixed and moving images *T* and *R* and first concatenated and fed to the registration network *N(T, R)*. The latter composed of many convolution layers, followed by activation functions, and ended by a final layer that produces the deformation $\mathbf{u}_\theta$. The moving image *T* and the obtained deformation **u** are passed to an interpolation layer that warps *T*, i.e. compute $T(\varphi_\theta)$. After getting the warped image, it is used to compute the loss function $\mathcal{L}(\cdot)$ and then backpropagate in order to minimise it.

network that can capture the image features may work well in our model. Here, we used a light version of the U-Net network with a ReLU activation at the end of each block in the network. The last block is responsible for generating the deformation $\mathbf{u}_\theta$.

(ii) Once the estimated transformation $\mathbf{u}_\theta$ is computed based on the features extracted by the the registration network $N(T, R)$. It is passed to a second component which is a differentiable warping module called Interpolator, that uses $\mathbf{u}_\theta$ to warp the image T, producing a warped image $T(\varphi_\theta)$, where $\varphi_\theta(\mathbf{x}) = \mathbf{x} + \mathbf{u}_\theta(\mathbf{x})$.
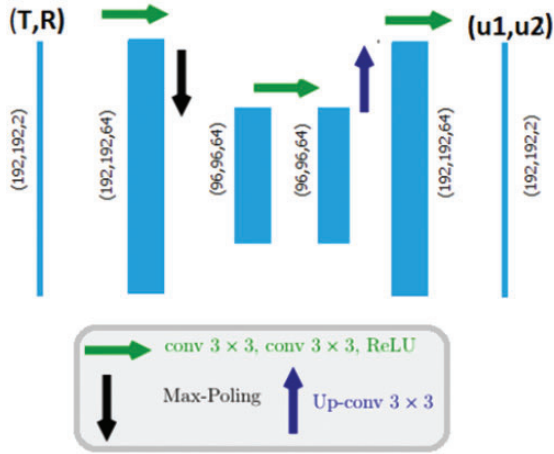

Figure 2. The architecture of the SN network for registration. The image *T* and *R* are concatenated to form a 2-channel image that will be fed to the network. The later output the deformation $\mathbf{u} = (u_1, u_2)$.

(iii) During training, the parameters θ of the registration network $N(T, R)$ are adjusted by minimising a loss function $\mathcal{L}(\cdot)$ which has form of the energy (1) i.e.

$$\min_\theta \left\{ \mathcal{L}(\mathbf{u}_\theta) = S(\mathbf{u}_\theta) + \frac{\lambda}{2} D(T(\varphi_\theta), R) \right\} \quad (2)$$

Generally, only the similarity measure $D(\cdot)$ depend on the image modality, whereas same regularisation $S(\cdot)$ could be used for both mono- and multi-modal images. In the sequel, we discuss the choice of the regulariser and the similarity measures.

Regularisation $S(\mathbf{u})$

The regularisation that we use consists in a smoothing constraint on the first- and second order derivatives and a constraint on the determinant of the transformation in order to obtain spatially smooth and plausible solutions Droske and Rumpf[19]; Burger et al.[7] The second-order derivatives allows getting smooth transformations and penalise affine linear transformations which are not included in the kernel of the first-order derivatives based regularises.[20] The term depending on $\det(\nabla \varphi_\theta(x))$ ensures the map to be locally invertible and then help avoiding the mesh folding problem. More precisely, we consider

$$S(\mathbf{u}_\theta) = ||\nabla \mathbf{u}_\theta||_2^2 + ||\nabla^2 \mathbf{u}_\theta||_2^2 + ||\phi(\det(\nabla \varphi_\theta))||_2^2 \quad (3)$$

where $\phi(v) = \frac{(v-1)^4}{v^2}$. This term is originally used in the non-learning hyper-elastic model and is known to be
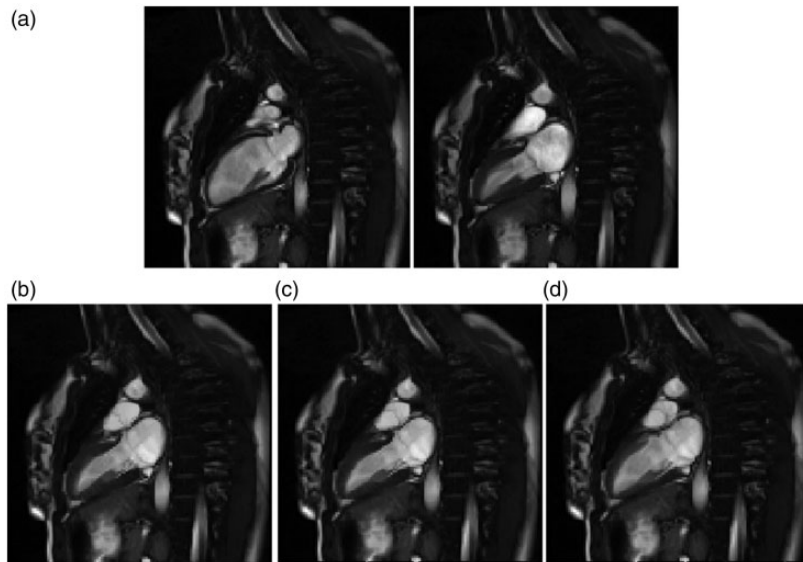

Figure 3. (a) The reference image *R*. (b) The moving image *T*. (c) Registered image $T(\varphi_\theta)$ for $\lambda = 200$, **MI** = 1.55. (d) Registered image for $\lambda = 60$, **MI** = 1.49. (e) Registered image for $\lambda = 10$, **MI** = 1.02.

very efficient in getting diffeomorphic maps, see Burger et al.[7]

The similarity measure $D(\cdot)$ depends on the image modality. In the sequel, we discuss the choice of $D(\cdot)$ for each image modality.

Measuring similarity for mono-modal images

Various similarity terms can be used to measure the goodness of the registration. In this work, we use an alternative measure to the correlation coefficient and which is well suitable for measuring linear dependence between images, hence mono-modal images. More precisely, we set

$$D(T(\varphi_\theta), R) = \mathbf{CLM}(T(\varphi_\theta), R)$$

where **CLM** is the following combined correlation-like measure:

$$\mathbf{CLM}(X, Y) = \int_\Omega \left(\frac{(X - m_X)}{\sqrt{\mathbf{Var}(X)}} - \frac{(Y - m_Y)}{\sqrt{\mathbf{Var}(Y)}} \right)^2 \mathrm{d}x \\ + \left(\sqrt{\mathbf{Var}(X)} + \sqrt{\mathbf{Var}(Y)} - \sqrt{\mathbf{Var}(X + Y)} \right)^2 \quad (4)$$

where m_X and $\mathbf{Var}(X)$ are the mean and the variance of X, respectively.

Measuring similarity for multi-modal images

Multi-modal images are often non-linearly correlated and the similarity measure **CLM** can't support multi-modal images as it only measures the linear correlation. In the sequel, we assume that the given multi-modal image pair is not random and has certain connections between them e.g. 'similar' shapes or edges.

Parallel level set measure. Various similarity terms can be used in the registration of multi-modal images such as mutual information Pluim et al.,[21] normalised gradient fields Ruhaak et al.[22] normalised gradients fitting Theljani and Chen[6]; Zhang et al.[23] In this work, we use parallel level sets similarity measure which is well suitable for measuring alignment between the gradients of two images. More precisely, consider

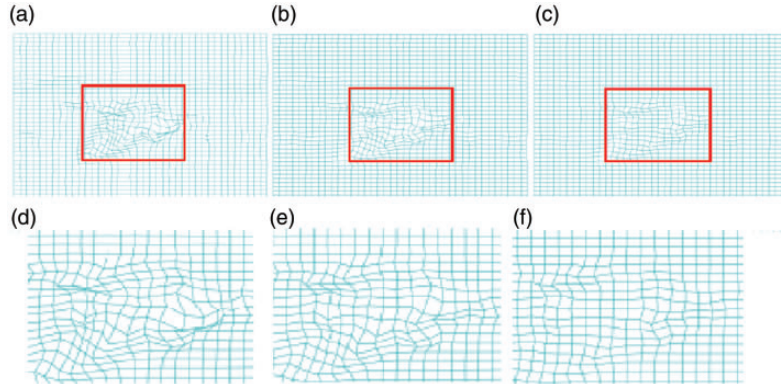$$D(T(\varphi_\theta), R) = \mathbf{PLS}(T(\varphi_\theta), R)$$

Figure 4. The transformed meshes for different values of λ for the of images in Figure 3. (a) $\lambda = 200$, $Q_{min} = -0.02$. (b) $\lambda = 60$, $Q_{min} = 0.13$. (c) $\lambda = 10$, $Q_{min} = 0.24$. (e) Zoom from the mesh in (a). (f) Zoom from the mesh in (b). Zoom from the mesh in (c). Clearly, a mesh folding occurs for $\lambda = 200$.
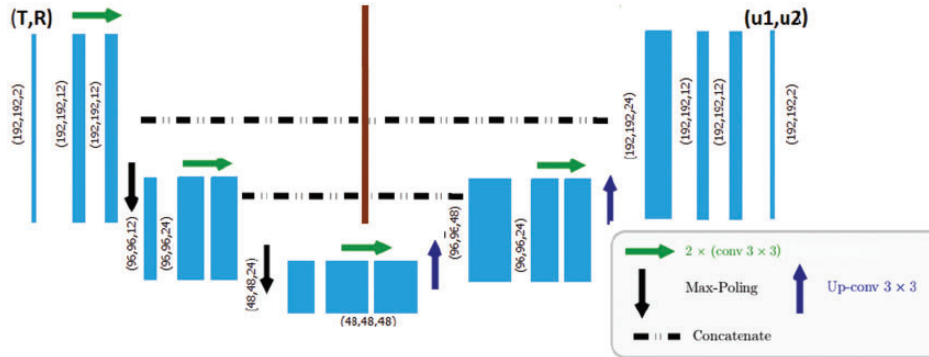

Figure 5. The architecture of the light U-net network for registration.

where

$$\mathbf{PLS}(X, Y) = ||\psi(|\nabla X|_\epsilon |\nabla Y|_\epsilon) - \psi(|\nabla X \cdot \nabla Y|_\epsilon)||_2^2 \quad (5)$$

where $|\nabla X|_\epsilon = \sqrt{|\nabla X|^2 + \epsilon^2}$ is regularised version of $|\nabla X|$ and ϵ is small non-negative constant. Different
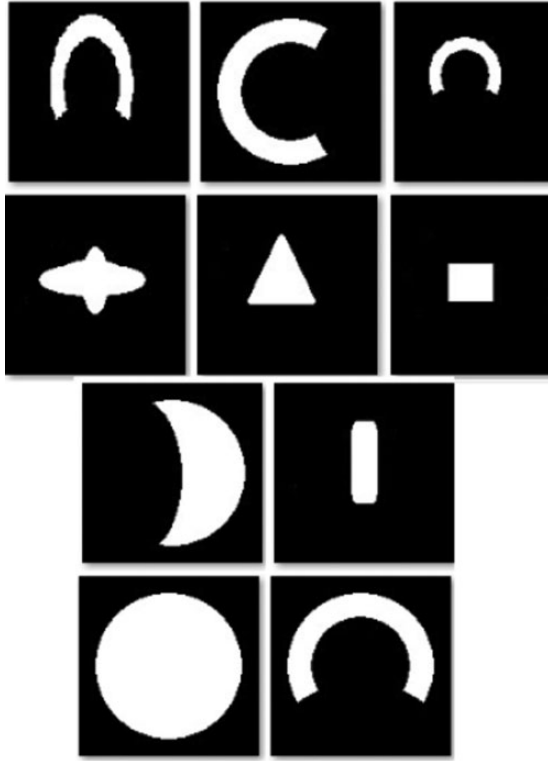

Figure 6. Dataset of 10 images used for the training, testing and comparing between different regularisers for large displacement.
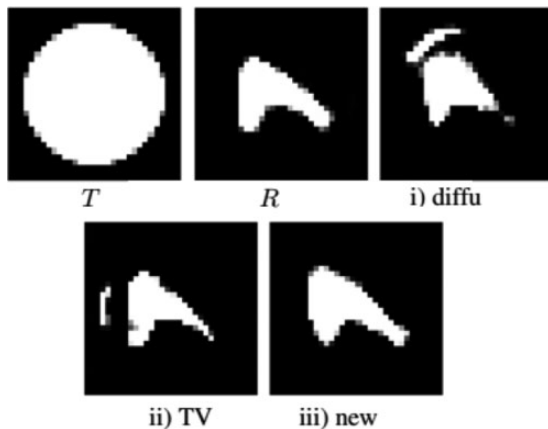

Figure 7. Comparison between different regularisers in registering an unseen pair of images. From left to right, Template *T*, Reference *R*, registered using different regularisation terms: i) the diffusion ($\mathbf{Q}_{min} = 0.38$), ii) Total-Variation ($\mathbf{Q}_{min} = -0.24$) and iii) $S(\mathbf{u})$ in (3) ($\mathbf{Q}_{min} = 0.71$), respectively. Clearly iii) gives the best result.

Table 1. Comparison between learning **DL Model** and non-learning **NL Model** for the images in Figure 10 in Speed (time) and Quality (MI). One sees that **DL Model** is about 100 times faster for a similar result.

	Examples			
	Exp 1	Exp 2	Exp 3	Exp 4
Time (s) for **DL model**	1.18±0.6			
Time (s) for **NL model**	112.54	120.31	107.11	112.5
MI for **DL model**	1.49	1.66	1.53	1.57
MI for **NL model**	1.52	1.64	1.54	1.59
$\mathbf{Q}_{min}$ for **DL model**	0.45	0.63	0.65	0.67
$\mathbf{Q}_{min}$ for **NL model**	0.61	0.47	0.69	0.54

Table 2. Comparison between the LU-net and SN networks in registering 4 mono-modal MRI heart images.

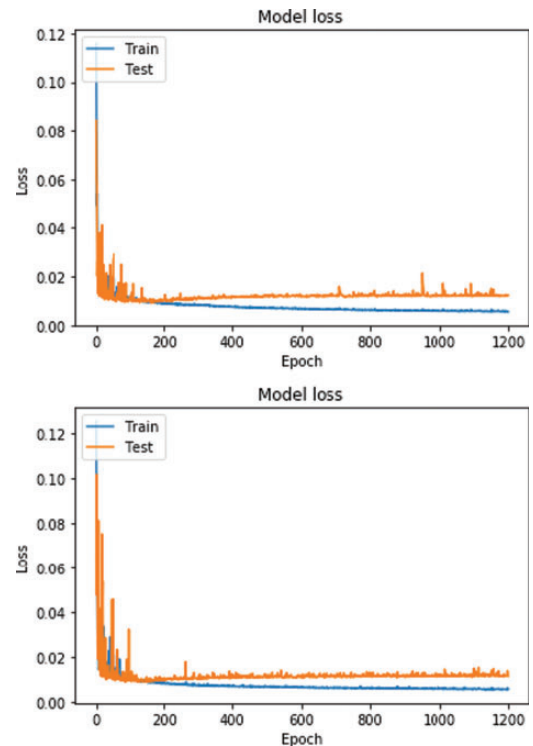|	Examples			
	Exp 1	Exp 2	Exp 3	Exp 4
MI for LU-net	1.53	1.68	1.52	1.62
MI for SN	1.49	1.66	1.53	1.57

Figure 8. The LU-net: Training and testing losses for the registration model for mono-modal (top) and multi-modal (bottom) images.

function $\psi(\cdot)$ can be used as in Ehrhardt et al.,[24] however, here we sued $\psi(x) = x$.

Data and numerical tests

In the numerical validation, we assess the performance of the proposed model that we call **DL Model** in registering mono-and multi-modal images. Our proposed method was implemented with Tensorflow library. We applied Adam with momentum optimisation algorithm to train the models with a learning rate of 0.0001, and set the batch size 6. The model was trained using a NVIDIA GeForce GTX 1050 Ti GPU. The number of epochs was 1200 for mono-modal images and 400 epochs for the multi-modal case. For the training time, it takes approximately 2 hours to train the model for mono-modal images and 56 minutes for the multi-modal images.

Registration network

We have tested two networks for the registration task. The first one, in Figure 2, is a small network (SN) which contains four blocks containing different kernels with different sizes and ReLU activation is used at the end of each block. The second one is a light version of U-net network, denoted by LU-net [27]. The last block of each network is responsible for generating the deformation $\mathbf{u}_\theta$. We illustrate the architecture of this network in Figure 3.

Choice of λ in (2)

We have tested three different values of $\lambda = 10, 60, 200$. In Figure 4, we display a registration results for a pair of MRI images for these values. We also show the transformed grids and the value $Q_{\min} = \text{mindet}(\nabla\varphi_\theta)$ for all values. It is clear that the regularisation parameter λ affects the registration and the mesh qualities. We note that taking large value of λ makes it difficult to get a physical solution, i.e. no folding In fact, for $\lambda = 200$ we clearly get a good alignment with **MI** = 1.55, but this leads to a folded mesh as it can be seen in Figure 5(e), with $Q_{\min} = -0.02$. However, by decreasing the value of λ, the solution is getting more physically correct with no mesh folding (see the zooms in Figure 5), but the registration quality is not as good as the case where λ is large, i.e. 200. In the rest of the numerical examples, we consider $\lambda = 60$.

Part 1: Synthetic data

In the part of the numerical examples, we assess the importance of the term $||\phi(\det(\nabla\varphi_\theta))||_2^2$ in the used loss function, and its importance in getting large

diffeomorphic deformations by comparing between different common regularisers. We consider 10 synthetic images drawn in order to compare the regularisers for possible large deformations, where each image can serve as a template and reference in the same time (so that we have $2 \times 10^2 = 200$ pairs of images), see Figure 6. Then, we train and test three different models on this same data, where the losses in these models have the same similarity measures **CLM**, the same regularisation parameter $\lambda = 60$, but different regularisations terms $S(\cdot)$.

- New: It is our proposed model and where

$$S(\mathbf{u}_\theta) = ||\nabla\mathbf{u}_\theta||_2^2 + ||\nabla^2\mathbf{u}_\theta||_2^2 + ||\phi(\det(\nabla\varphi_\theta))||_2^2. \quad (6)$$

- Total variation: The regularisation in this model is given by $S(\mathbf{u}_\theta) = ||\nabla\mathbf{u}_\theta||_1$
- Diffusion (Diffu): We consider $S(\mathbf{u}_\theta) = ||\nabla\mathbf{u}_\theta||_2^2$ as regularisation term.

As shown in Figure 7, different $S(\mathbf{u}_\theta)$ lead to different results that we can distinguish between them visually and by also checking the value of $Q_{\min} = \text{mindet}(\nabla\varphi_\theta)$. Only the new model gives diffeomorphic
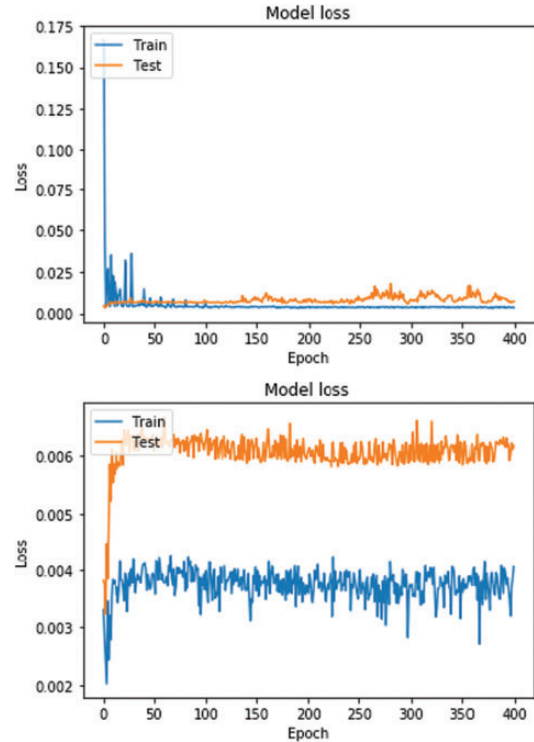

Figure 9. The SN network: Training and testing losses for the registration model for mono-modal (top) and multi-modal (bottom) images.

deformations since $\mathcal{Q}_{\min 2} > 0$ in this case, i.e. no mesh folding.

Part 2: Real data

We test the performance of the proposed **DL model** in registering real mono- and multi-modal images. Registration quality is evaluated using the mutual information (the larger the better) between the two images $T(\varphi_\theta)$ and R. We also assess if the map φ_θ is diffeomorphic by checking the minimum of the Jacobian determinant $\det(\nabla\varphi_\theta)$ i.e. yes if positive. We tested and compared between the two network architectures SN and LU-net (see Table 2).

We also compare with traditional deformable registration model Burger et al.,[7] which is a non-learning model and that we call **NL model**. It consists in solving the following optimisation problem

$$\min_{\mathbf{u}} \left\{ ||\nabla\mathbf{u}||_2^2 + ||\nabla^2\mathbf{u}||_2^2 + ||\phi(\det(\nabla\varphi))||_2^2 + \frac{\lambda}{2} D(T(\varphi), R) \right\} \tag{7}$$

where $D(T(\varphi_\theta), R) = \mathbf{CLM}(T(\varphi_\theta), R)$ for mono-modal images and $D(T(\varphi_\theta), R) = \mathbf{PLS}(T(\varphi_\theta), R)$ for multi-modal images. In this model, we solve an optimisation problem w.r.t $\mathbf{u}$ for each pair of images T, R. We used
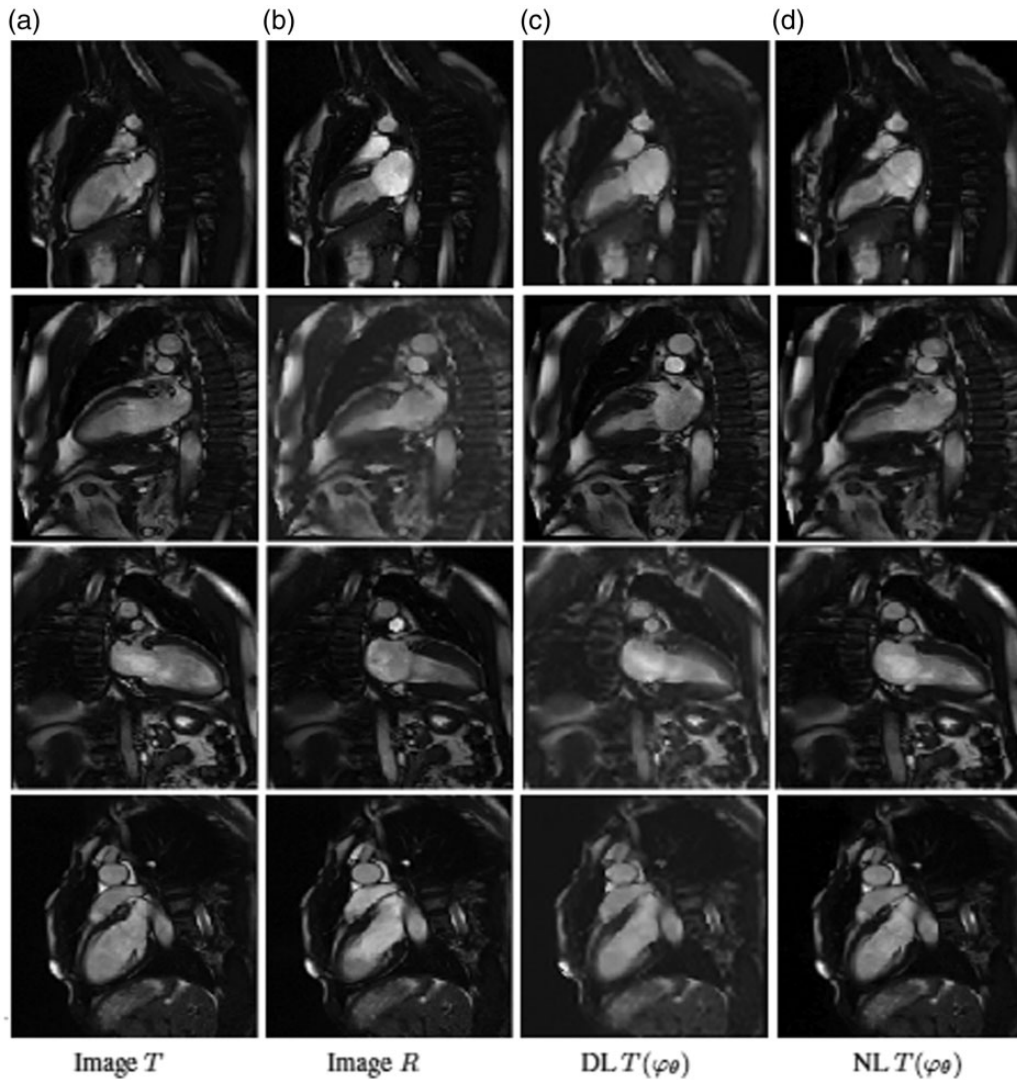

Figure 10. Pairwise registration results for 4 pairs of MRI images. Each row represents the result for a pair test. (a) Moving images *T*. (b) Fixed images *R*. (c) and (d) are the registered images using **DL Model** and **NL Model**, respectively. The mutual information errors and the values of $\mathbf{Q}_{min}$ related to these tests are given in Table 1.

the Matlab implementation of FAIR's Modersitzki[3] for the regularization and we have just implemented the similarity measures **CLM** and **PLS** under the framework of FAIR's package.

Mono-modal images. We trained our model on 160 mono-modal MRI heart images and compare with the classical variational model **NL model** where we solve the optimisation problem (7). We test our model on 20 images and we display 4 comparison tests with the **NL model** in term of runtime and accuracy, see Figure 10. **DL Model** is by far faster and predict the transformation in 1 second for a pair of images with a resolution of 192×192. The **NL model** achieves the same result in term of accuracy but it takes more than 1 minute. The comparison between both models is summarised in Table 1.

We also compared the two used networks in term of registration quality. The comparison details are reported in Table 2. The training curves of the registration model using the two networks are illustrated in Figure 8 and Figure 9. These curves present the learning rate values at each epoch, and we see that the model converges for the two networks after 800 epochs.

Multi-modal images. For multi-modal images, we trained the network on 120 pairs of CT and MRI images. In Figure 11, we display the prediction results for registering 4 pairs of MRI and CT images. To see the quality visually, we show the fused CT and MRI images for the four examples before and after registration. Clearly after the registration the images are well aligned. We give the mutual information errors and the run-time during the prediction in Table 3.
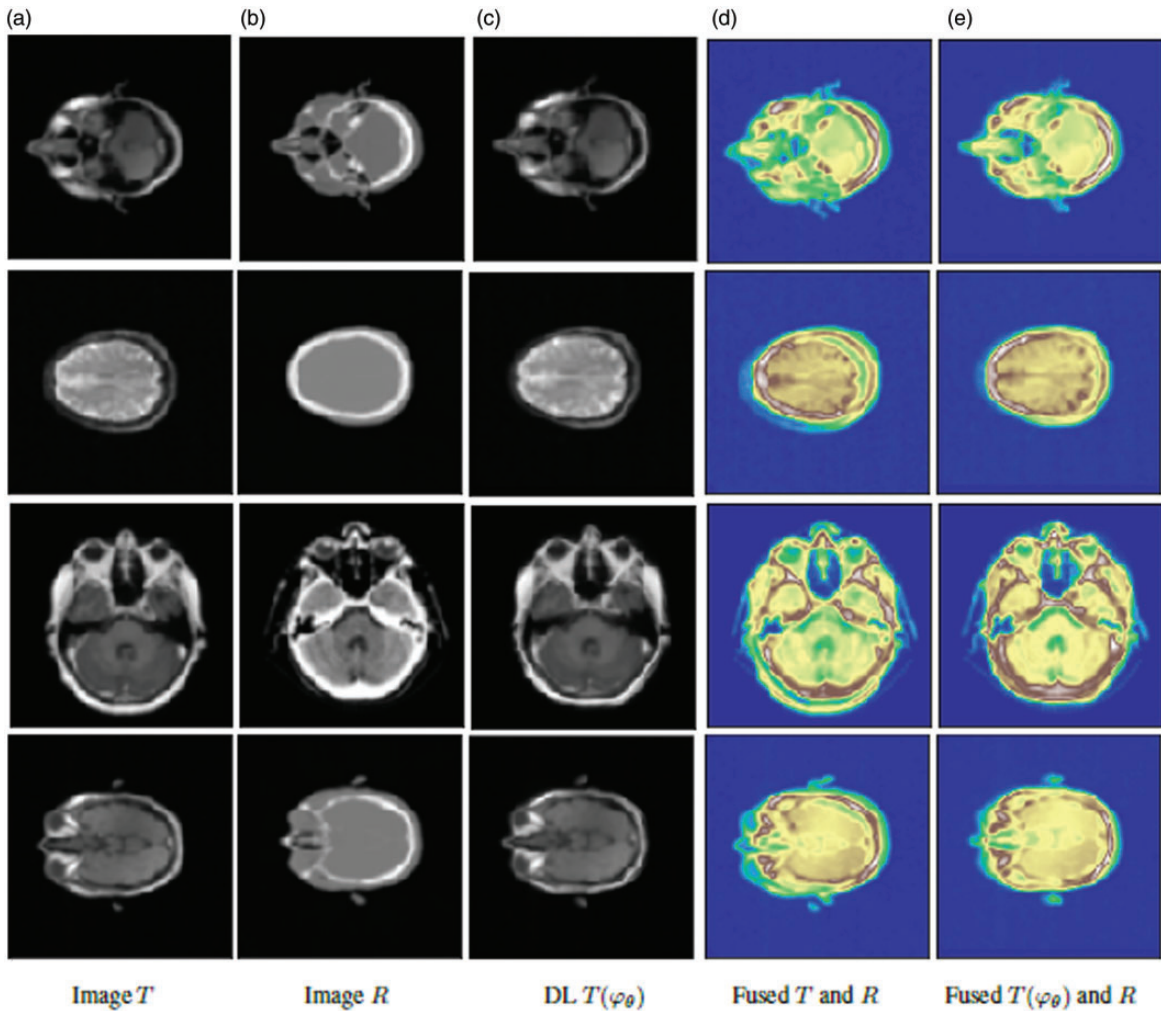

Figure 11. Pairwise registration results for 4 pairs of MRI-CT images. Each row represents the result for a pair test. (a) Moving images *T*. (b) Fixed images *R*. (c) registered images using **DL Model**. (d) and (e) Fused images before and after the registration. The mutual information errors and the values of $\mathbf{Q}_{min}$ related to these tests are given in Table 3.

Table 3. DL model: The Speed (time), Quality (MI) and Q_{min} vlaues for MR-CT images in Figure 11.

	MR-CT images			
	Pair 1	Pair 2	Pair 3	Pair 4
Time (s) for **DL model**	0.045 ± 0.002			
MI for **DL model**	1.95	1.93	1.94	1.94
Q_{min} for **DL model**	0.97	0.98	0.97	0.97

Conclusions

We have developed and presented an unsupervised deep learning approach for mono-and multi-modal images registration. We tested and compared different choices of regularisation constraints on the deformation fields. The results have shown that control on the Jacobian determinant of the deformation is necessary in the loss function in order to get a diffeomorphic map, mainly for large displacements. The learning model was first designed and tested for mono-modal images. The same learning approach works effectively for multi-modal images, by only changing the similarity measure to fit with the multi-modal setting. For for mono- and multi-modal images, we tested 2 choices of networks. We found that LU-net is recommended.

Future work will consider generalisations to 3 dimensional images registration.

ORCID iD

Anis Theljani ⓘ https://orcid.org/0000-0002-9126-2158

References

1. Fischer B and Modersitzki J. Ill-posed medicine - – an introduction to image registration. *Inverse Problems* 2008; 24: 034008.
2. Gigengack F, Ruthotto L, Burger M, et al. Motion correction in dual gated cardiac pet using mass-preserving image registration. *IEEE Trans Med Imaging* 2012; 31: 698–712.
3. Modersitzki J. *FAIR: flexible algorithms for image registration*. Philadelphia: SIAM, 2009.
4. Oliveira FP, Manuel J and Tavares RS. Medical image registration: a review. *Comput Methods Biomech Biomed Eng* 2014; 17: 73–93.
5. Sotiras A, Davatzikos C and Paragios N. Deformable medical image registration: a survey. *IEEE Trans Med Imaging* 2013; 32: 1153–1190.
6. Theljani A and Chen K. An augmented Lagrangian method for solving a new variational model based on gradients similarity measures and high order regulariation for multimodality registration. *Inverse Problems Imaging* 2019; 13: 309–335.
7. Burger M, Modersitzki J and Ruthotto L. A hyperelastic regularization energy for image registration. *SIAM J Sci Comput* 2013; 35: B132–148.
8. Droske M and Ring W. A Mumford–Shah level-set approach for geometric image registration. *SIAM J Appl Math* 2006; 66: 2127–2148.
9. Henn S. A multigrid method for a fourth-order diffusion equation with application to image processing. *SIAM J Sci Comput* 2005; 27: 831–849.
10. Mang A and Biros G. Constrained h^1-regularization schemes for diffeomorphic image registration. *SIAM J Imaging Sci* 2016; 9: 1154–1194.
11. Theljani A and Chen K. A Nash game based variational model for joint image intensity correction and registration to deal with varying illumination. *Inverse Problems* 2020; 36: 034002.
12. Jaderberg M, Simonyan K, Zisserman A, et al. Spatial transformer networks. In: C Cortes, N Lawrence, D. Lee, M Sugiyama and R Garnett (eds) *Advances in neural information processing systems*, 2015, pp.2017–2025.
13. Haskins G, Kruger U and Yan P. Deep learning in medical image registration: a survey. *arXiv preprint arXiv:1903.02026*, 2019.
14. Li H and Fan Y. Non-rigid image registration using fully convolutional networks with deep self-supervision. *arXiv preprint arXiv:1709.00799*, 2017.
15. de Vos BD, Berendsen FF, Viergever MA, et al. A deep learning framework for unsupervised affine and deformable image registration. *Med Image Anal* 2019; 52: 128–143.
16. Cao X, Yang J, Wang L, et al. Deep learning based inter-modality image registration supervised by intra-modality similarity. In: *International workshop on machine learning in medical imaging*. Cham: Springer, 2018, pp.55–63.
17. Krebs J, Delingette H, Mailhé B, et al. Learning a probabilistic model for diffeomorphic registration. *IEEE Trans Med Imaging* 2019; 38: 2165–2176.
18. Kuang D. Cycle-consistent training for reducing negative Jacobian determinant in deep registration networks. In: Burgos N, Gooya A and Svoboda D (eds) *Simulation and synthesis in medical imaging. SASHIMI 2019. Lecture notes in computer science*. Vol. 11827. Cham: Springer, Cham, pp 120-129.
19. Droske M and Rumpf M. A variational approach to nonrigid morphological image registration. *SIAM J Appl Math* 2004; 64: 668–687.

20. Zhang J, Chen K and Yu B. A novel high-order functional based image registration model with inequality constraint. *Comput Math Appl* 2016; 72: 2887–2899.

21. Pluim JP, Maintz JA and Viergever MA. Mutual-information-based registration of medical images: a survey. *IEEE Trans Med Imaging* 2003; 22: 986–1004.

22. Ruhaak J, König L, Hallmann M, et al. A fully parallel algorithm for multimodal image registration using normalized gradient field. In: *Biomedical 754 Imaging (ISBI), 2013 IEEE 10th international symposium on Biomedical 754 Imaging (ISBI)*. Piscataway: IEEE, 2013, pp.572–575.

23. Zhang D, Theljani A and Chen K. On a New Diffeomorphic Multi-Modality Image Registration Model and Its Convergent Gauss-Newton Solver. Journal of Mathematical Research with Applications. 2019; 39: 633-656.

24. Ehrhardt MJ, Markiewicz P, Liljeroth M, et al. PET reconstruction with an anatomical MRI prior using parallel level sets. *IEEE Trans Med Imaging* 2016; 35: 2189–2199.

25. Theljani A and Chen K . An unsupervised deep learning method for diffeomorphic mono-and multi-modal image registraon. In: *23rd conference in medical imaging, understnding and analysis*. Cham: Springer, 2020, pp.317–326.

26. Hwan Jin K, McCann MT, Froustey E, et al. Deep convolutional neural network for inverse problems in imaging. *IEEE Trans Image Process* 2017; 26: 4509–4522.

27. Ronneberger O, Fischer P and Brox T. U-net: convolutional networks for biomedical image segmentation. In: *International conference on medical image computing and computer-assisted intervention*. Berlin: Springer, 2015, pp.234–241.