PAPER • **OPEN ACCESS**

SOLID: a novel similarity metric for mono-modal and multi-modal deformable image registration

To cite this article: Paris Tzitzimpasis *et al* 2024 *Phys. Med. Biol.* **69** 015020

IOP Publishing Phys. Med. Biol. 69 (2024) 015020 https://doi.org/10.1088/1361-6560/ad120e

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

OPEN ACCESS

RECEIVED
17 August 2023

REVISED
27 October 2023

ACCEPTED FOR PUBLICATION
4 December 2023

PUBLISHED
26 December 2023

PAPER

SOLID: a novel similarity metric for mono-modal and multi-modal deformable image registration

Paris Tzitzimpasis[1,*], **Cornel Zachiu**[1], **Bas W Raaymakers**[1] **and Mario Ries**[2]

[1] Department of Radiotherapy, UMC Utrecht, Heidelberglaan 100, 3508 GA, Utrecht, The Netherlands
[2] Imaging Division, UMC Utrecht, Heidelberglaan 100, 3508 GA, Utrecht, The Netherlands
[*] Author to whom any correspondence should be addressed.

E-mail: p.tzitzimpasis@umcutrecht.nl

Keywords: image registration, similarity metric, curvature, second fundamental form, multi-modal

Supplementary material for this article is available online

Abstract

Medical image registration is an integral part of various clinical applications including image guidance, motion tracking, therapy assessment and diagnosis. We present a robust approach for mono-modal and multi-modal medical image registration. To this end, we propose the novel shape operator based local image distance (SOLID) which estimates the similarity of images by comparing their second-order curvature information. Our similarity metric is rigorously tailored to be suitable for comparing images from different medical imaging modalities or image contrasts. A critical element of our method is the extraction of local features using higher-order shape information, enabling the accurate identification and registration of smaller structures. In order to assess the efficacy of the proposed similarity metric, we have implemented a variational image registration algorithm that relies on the principle of matching the curvature information of the given images. The performance of the proposed algorithm has been evaluated against various alternative state-of-the-art variational registration algorithms. Our experiments involve mono-modal as well as multi-modal and cross-contrast co-registration tasks in a broad variety of anatomical regions. Compared to the evaluated alternative registration methods, the results indicate a very favorable accuracy, precision and robustness of the proposed SOLID method in various highly challenging registration tasks.

1. Introduction

The recent trend towards personalized precision medicine requires increasingly a broad spectrum of imaging modalities for diagnostics, therapy-guidance as well as response assessment. Precise and robust fusion of image information of ultrasound (US), computed tomography (CT), positron emission tomography (PET), x-ray or magnetic resonance imaging (MRI) becomes therefore of considerable importance for a growing number of clinical work-flows.

As a consequence, state-of-the-art medical image registration has to encompass not only mono-modal but also increasingly multi-modal image registration tasks: In mono-modal registration, we usually seek to align images of the same anatomy that have been obtained at different times (multi-temporal data) but with the same imaging modality and contrast. In this case, a straightforward comparison of the anatomical structures or even the gray-level intensities of the given images is possible. A prime example of such an algorithm is the optical flow algorithm suggested by Horn and Schunck (1981), which relies on the assumption of intensity conservation. A more exhaustive overview of these types of algorithms can be found in Maintz and Viergever (1998), Mani and Arivazhagan (2013) or Song *et al* (2017). In comparison, multi-modal image registration seeks to establish spatial correspondence between images obtained with different medical imaging modalities and/or imaging contrasts and represents therefore a significantly more challenging task.

One of the first proposed solution strategies for multi-modal image registration is mutual information (MI). Proposed by both Viola and Wells (1997) and Maes *et al* (1997) independently, MI-based multi-modal image registration was well received in the medical imaging community and triggered a considerable amount of follow-up research underlining the substantial potential, as summarized by Pluim *et al* (2003). An alternative approach is to convert all images into modality-independent scalar representations and to subsequently apply mono-modal registration algorithms to the pre-processed datasets. Several modality-independent representations such as the local phase by Mellor and Brady (2005), local entropy by Wachinger and Navab (2012) or the gradient orientation by Haber and Modersitzki (2006) have been suggested.

A more recent approach for multi-modal DIR, which has shown considerable potential is the modality independent neighborhood descriptor (MIND) by Heinrich *et al* (2012), which is based on a vectorial representation of the input images, encoding the relationship of each voxel to the surrounding voxel perimeter as suggested by Buades *et al* (2005). One of the principal requirements is that the method hypothesizes that each anatomical structure in one image has its exact counterpart in form of a similar descriptor in the other, which is frequently in multi-modal datasets only partially fulfilled. Another family of conceptually different DIR algorithms relies on edge alignment and has been originally proposed in the field of video image registration and subsequently translated to medical imaging as suggested by Pluim *et al* (2000), Sun *et al* (2004) and Sutour *et al* (2015). In this approach, the underlying assumption is that boundaries and details of anatomical structures display in both types of images contrast with respect to the surrounding tissue. One of the first data-fidelity metrics to successfully implement this principle for multi-modal registration is the normalized gradient fields (NGF) (Haber and Modersitzki 2006). A more recent variation of this type of algorithm has been EVolution, which implemented an edge-based detector in a variational algorithm by de Senneville *et al* (2016).

Finally, a recently emerging school of thought which deviates significantly from the aforementioned ideas rests on the employment of neural networks for the task of image registration (Fu *et al* 2020). With convolutional neural networks being the most frequently utilized architecture so far, deep learning approaches have attracted significant research interest due to their unexplored potential. In principle, such algorithms learn a function whose input is a pair of images and whose output is the sought after deformation vector field. Bypassing the necessity of repeating an optimization scheme for each image pair, learning-based approaches allow for lower computation times. There have been both supervised (Krebs *et al* 2017, Yang *et al* 2017) and unsupervised (de Vos *et al* 2019) approaches presented. Despite their apparent disparity, it has been pointed out that the traditional and deep learning approaches should not be regarded as incompatible since the insights obtained from traditional algorithms and their theoretical foundation have the potential of being applied in a deep learning setup (Dalca *et al* 2019). In particular, recent alternative approaches to convolutional neural networks incorporated these concepts directly as a similarity metric into variational algorithms, such as for example suggested by Simonovsky *et al* (2016) or Krebs *et al* (2019) for multi-modal image registration. In this study, we return to the more traditional approach and propose a deformable image registration algorithm, which combines the variational method with a novel similarity metric, treating the input images as hypersurfaces and comparing the geometric information encoded in the shape operator as described below.

As already mentioned, various image similarity metrics have relied on gradient information to codify structural information. Gradient direction has been utilized to align images of the same or different modalities. Normalized gradient fields and EVolution are examples of registration algorithms employing this idea. In a different spirit, the vector field similarity (VFS) technique attempts to create anatomically accurate structural representations (Jaouen *et al* 2021). Gradient based methods, despite their overall success, can be suboptimal in anatomies dense in gradient-based features since contiguous anatomical landmarks can be mismatched leading to registration errors, particularly in the presence of large initial displacements. This reduced discriminative ability of gradient fields suggests that a metric measuring higher order information of the shapes appearing in medical images could be beneficial. Our contributions in this work can be summarized as follows:

- We introduce SOLID, a novel image similarity metric that uses second order information of the given images.
- We incorporate SOLID in a variational image registration algorithm.
- We evaluate the performance of the proposed algorithm by comparing it to other state-of-the-art methods in a series of challenging mono-modal and multi-modal tasks.

2. Theory

In this section, we describe our data-similarity term and the optimization framework used in the proposed registration algorithm. To avoid any potential confusion coming from the introduction of various mathematical objects, we also provide a notation table 1 at the end of the section. Our method employs the notion of *shape*

Table 1. Notation table.

Symbol	Meaning
I	Image intensity
I_x, I_y, I_z	Spatial derivatives of the image intensity
u^i	Deformation field components for x, y, z directions
g^{ij}	First fundamental form (metric tensor) of the reference image
$\mathcal{D}$	SOLID image similarity metric
$h^{\text{ref}}, h^{\text{mov}}$	Second fundamental forms of the reference/moving images
S^m_j	Shape operator associated to a given image
f^i_k	Numerical gradient of the data similarity term with respect to the ith component of the vector field and evaluated at the kth iteration
$\tilde{\gamma}$	Numerical stability factor that we set to 10^{-8}
γ	Numerical stability factor that is determined from $\tilde{\gamma}$ and image information
β	Smoothness regularization weight
τ	Time step used in our DCT optimizer (automatically computed)
$\langle A, B\rangle$	Frobenius product of the tensor fields A, B on a curved space with metric g^{ij}
$T(u^i)$	Spatial transformation associated to the vector field with components u^i
$\mathbf{A}$	Matrix of eigenvalues of the Laplace operator

operator to measure the similarity of given images. We think of a 2D image as a surface in $\mathbb{R}^3$ by using the image intensity I: $\mathbb{R}^2 \to \mathbb{R}$ as a height function so that $z = I(x, y)$. In this case, the curvature of the surface is dictated by the intensity patterns that appear the image and in particular by first and second order intensity information. In order to encode this information, we use the shape operator which is a local measure of the surface's curvature. Although the rest of our work focuses on 3D images, this intuitive 2D picture is useful. Given a 3D image with intensity function $I(x, y, z)$, the first quantity that needs to be computed is the *metric tensor* or *first fundamental form* which is defined as follows:

$$g^{mk} = \frac{1}{I_x^2 + I_y^2 + I_z^2 + 1} \begin{bmatrix} 1 + I_y^2 + I_z^2 & -I_x I_y & -I_x I_z \\ -I_x I_y & 1 + I_x^2 + I_z^2 & -I_y I_z \\ -I_x I_z & -I_y I_z & 1 + I_x^2 + I_y^2 \end{bmatrix} \tag{1}$$

The first fundamental form encodes the first order information of a given image. The second order information is captured by the *second fundamental form* h_{ij}:

$$h_{ij} = \frac{1}{\sqrt{I_x^2 + I_y^2 + I_z^2 + 1}} \begin{bmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{xy} & I_{yy} & I_{yz} \\ I_{xz} & I_{yz} & I_{zz} \end{bmatrix} \tag{2}$$

Our strategy encompasses the alignment of the second fundamental forms of the given 3D input images. In order to achieve this, we need to define an appropriate inner product for the (2,0)-tensors h_{ij}. Given a vector space of $m \times n$ matrices, a standard choice for an inner product is the Frobenius product defined by $\langle A, B\rangle = \text{tr}(A^T \cdot B)$ where A^T denotes the transpose of A and $\cdot$ is the standard matrix multiplication. The Frobenius inner product can be naturally generalized to (2,0)-tensors in the following way:

$$\langle h_{ij}, h'_{mn}\rangle = \sum_{i,j,m,n} g^{mi} g^{nj} h_{ij} h'_{mn} = \sum_{mj} S^m_j S^j{}_m \tag{3}$$

where $S^m_j = \sum_i g^{mi} h_{ij}$, $S^j{}_m = \sum_n g^{nj} h'_{mn}$ are called the *shape operators* associated to h_{ij} and h'_{mn} respectively. The summation indices run from one to the dimension of the images (in our case 3). With those definitions being in place, we can introduce our data-fidelity term. Given a reference and moving image with corresponding second fundamental forms h^{ref} and h^{mov} respectively (calculated as in equation (2)), we define the following scalar object called the *shape operator based local image distance* (SOLID)

$$\mathcal{D}\{h^{\text{ref}}, h^{\text{mov}}\}(x, y) = -\frac{\int_\omega \langle h^{\text{ref}}, h^{\text{mov}}\rangle^2 + \gamma}{\int_\omega \langle h^{\text{ref}}, h^{\text{ref}}\rangle \langle h^{\text{mov}}, h^{\text{mov}}\rangle + \gamma}, \tag{4}$$

where the inner products are computed as in 3 and the integrations are carried out in a neighborhood of size $2\omega + 1$ around the the pixel with coordinates (x,y). A value of $\omega = 1$ is usually the optimal choice for images with high SNR. For images heavily contaminated by noise, $\omega = 2$ provides in general more robust results. The parameter γ is introduced on computational grounds to avoid division by zero. For our purposes, we chose

$$\gamma = \tilde{\gamma} \cdot \langle \langle h^{\text{ref}}, h^{\text{ref}} \rangle \langle h^{\text{mov}}, h^{\text{mov}} \rangle \rangle_{\Omega} \tag{5}$$

where the outer brackets denote averaging over the entire domain Ω and $\tilde{\gamma}$ is a numerical parameter to be specified. The precise value of $\tilde{\gamma}$ does not affect the efficacy of the resulting registration algorithm. We elaborate on this in our experiment 3.5.

From equation (4) follows that the proposed similarity metric is *conformally invariant*, namely $\mathcal{D}\{\kappa h^{\text{ref}}, \lambda h^{\text{mov}}\}(x, y) = \mathcal{D}\{h^{\text{ref}}, h^{\text{mov}}\}(x, y)$ for arbitrary $\kappa, \lambda \in \mathbb{R}$. This is highly desirable since in multi-modal scenarios the magnitudes of the derivatives of the reference and moving images are in general not comparable. Expressed in a more intuitive way, our ultimate goal is to align the second fundamental form tensors h^{ref} and h^{mov} up to an overall factor, but not to match them directly. The aforementioned theoretical arguments are local so that they are valid within individual neighborhoods. This is crucial, since assuming that the curvatures of the two (aligned) images are a multiple of each other globally would be a very strict assumption that would not hold in many cases, in particular in complex multi-modal images.

A caveat of the definition in equation (4) is the ambiguity in the inverse metric g^{ij} used to compute the inner product due to the fact that the Frobenius product in equation (3) for tensors is introduced on the same surface but h^{ref} and h^{mov} are defined on different surfaces. However, since our goal is to align the two surfaces, we hypothesize that a common inverse metric for this computation is sufficiently precise, in particular considering that this becomes an increasingly valid approximation during the registration process as the surfaces are pulled closer to each other. In our implementation, we used the inverse metric associated to the reference image. This choice offers some computational efficacy since this object remains unchanged during the registration. The geometric interpretation of this assumption is that we view h^{mov} as a tensor field on the surface associated to the reference image.

From the definition 4 it follows that when either the reference or moving second fundamental form is identically zero at some point, then SOLID attains its minimum value. Such points convey no information for the registration and it is solely the regularization that determines the deformation then. In this sense, homologous anatomical structures must be present in both images in order for the data fidelity to have a non-trivial contribution to the estimated deformation vector field. In this way, SOLID is designed to be robust with respect to absent correspondences in the co-registered images.

We now move on to describe our registration algorithm that incorporates the SOLID similarity metric. Due to the under-determined nature of the registration problem, the introduction of a regularization term is imperative. For that reason, we use a smoothness regularization and our full functional reads

$$\mathcal{L} = \int_{\Omega} \mathcal{D}\{h^{\text{ref}}, T(u^i) \cdot h^{\text{mov}}\} + \frac{\beta}{2} \int_{\Omega} \sum_{i=1}^{3} ||\nabla u^i||^2 \tag{6}$$

where the integrations are carried over the entire image domain Ω. Here, β is the regularization weight and $T(u^i)$ denotes the transformation generated by the vector field $(u^1(x, y, z), u^2(x, y, z), u^3(x, y, z))$ applied on h^{mov}. It is important to note that we transform the components of h_{ij}^{mov} as if they were scalars without taking their tensorial properties into account. In principle, given a diffeomorphism f: $S \to S$ the pullback of the (2,0)-tensor h_{ij}^{mov} is given by

$$(f^* h^{\text{mov}})_{ij}(x) = \sum_{\alpha\beta} \frac{\partial f^{\alpha}}{\partial x^i} \frac{\partial f^{\beta}}{\partial x^j} h_{\alpha\beta}^{\text{mov}}(f(x)), \tag{7}$$

where f^{α} are the coordinates of f. Since in the end we are computing a scalar (the data fidelity term in equation (4)) the transformation matrices from all the covariant and contravariant indices will cancel out. However, owing to the hybrid nature of the data fidelity term in equation (4), h^{mov} is the only term that transforms while the inverse metric and h^{ref} remain fixed, both being computed on the reference image. Therefore, if we apply the complete transformation in equation (7), the data fidelity term will not transform as a scalar which is highly undesirable. This justifies the transformation of the components h_{ij}^{mov} as independent scalars.

In summary, under the outlined assumptions, our registration problem reduces to the minimization of the functional in equation (6) with respect to the vector field u^i. Requiring that $\delta\mathcal{S} = 0$ results in the following Euler–Lagrange equations

$$\frac{\delta\mathcal{D}}{\delta u^i}(x, y, z) - \beta\Delta u^i(x, y, z) = 0 \quad i = 1, 2, 3, \tag{8}$$

where Δ is the Laplace operator. We complete the set of equations by imposing Neumann boundary conditions. To solve equations (8) we employ the following iterative scheme introduced in Fischer and Modersitzki (2004) and effectively used in Larrey-Ruiz *et al* (2007):

$$u_{k+1}^i = (\mathbf{I} - \tau\beta\mathbf{A})^{-1}(u_k^i - \tau f_k^i) \quad i = 1, 2, 3 \tag{9}$$

where k is the iteration index, τ is the time step, $\mathbf{I}$ is the identity operator, $\mathbf{A}$ is the matrix representation of the Laplace operator and f_k^i is the force term

$$f_k^i = \frac{\delta \mathcal{D}_k}{\delta u^i}(x, y, z) = \frac{\mathcal{D}(u_k^i + 1) - \mathcal{D}(u_k^i - 1)}{2} \qquad i = 1, 2, 3 \tag{10}$$

Following Larrey-Ruiz *et al* (2007), a discrete cosine transform (DCT) is implemented to diagonalize the Laplacian matrix **A**. In order to enable the estimation of larger scale motion and a faster convergence, we implemented a coarse-to-fine scheme. At each resolution level, the components of the inverse metric g^{ij} and the fundamental forms h^{ref} and h^{mov} are down-sampled by a factor of 2. The number of resolution levels is determined by the requirement that no dimension becomes smaller than 16 pixels. Finally, the time step τ is automatically determined at the first iteration of each resolution level as follows

$$\tau = 1 \Big/ \max_{\vec{x} \in \Omega} \{ \sqrt{f_1^1(\vec{x})^2 + f_1^2(\vec{x})^2 + f_1^3(\vec{x})^2} \} \tag{11}$$

This choice ensures that the term τf_k^i in (9) is bounded, therefore avoiding potential numerical instabilities.

3. Experiments

We have performed a number of experiments to evaluate the performance of the proposed algorithm. Our experiments are tailored towards highly challenging and clinically relevant applications. For comparison, other widely used algorithms have been subjected to the same tests. These are the publicly available Elastix (Klein *et al* 2009, Shamonin *et al* 2014) using B-splines and a mutual information based similarity metric, the gradient-based EVolution (de Senneville *et al* 2016) and MIND (Heinrich *et al* 2012) whose data-fidelity metric relies on the concept of self-similarity. In order to evaluate differences originating strictly from the data fidelity term, we opted to implement EVolution and MIND using the same numerical scheme as SOLID. In the following, these algorithms will be denoted as EVolution (DCT) and MIND (DCT) to stress the registration framework used. In order to enable comparison with other established methods, we have used publicly available datasets where possible. Finally, the procedure followed to calibrate the parameters of the tested algorithms can be found as supplementary material of this article.

3.1. Multi-modal (MR-CT) alignment of abdominal images

This experiment evaluates the registration of intra-patient abdominal MR and CT scans. This task is highly challenging due to the structural differences in MR and CT scans. On top of that, the human abdomen comprises a variety of organs whose position and shape are affected by respiration and other anatomical changes. However, MR to CT registration is highly relevant in clinical practice, for example in radiotherapy treatment planning. We used the 8 training image pairs from the Learn2Reg challenge (Clark *et al* 2013). The images have a spatial resolution of $192 \times 160 \times 192$ and isotropic voxel size of 2 mm. For more information we refer the reader to the website.[3] We measured the DSC overlap score on the liver, kidneys and spleen using the segmentations provided. We make the following remarks regarding the use of the dataset:

- In case 1, the left kidney was not taken into account, due to being present only in one of the scans.
- In case 2, both kidneys and the liver are only partially visible in the MR scan, making it less useful for validation purposes. Therefore, this case was used as a 'training' dataset that we utilized to tune the parameters of the tested algorithms (see figure S1 of supplementary material).

3.2. Cross-contrast abdominal MRI images with variable FOV

In this experiment we evaluate the SOLID algorithm on the task of T1w/T2w abdominal MR image registration. A notable difficulty of this dataset arises from the highly variable field of view between the two images. This poses a great challenge for image registration algorithms but can occur frequently in clinical imaging. Furthermore, the images were acquired at different time points, and are subject to additional misalignment due to respiratory and peristaltic motion. The liver and kidneys of all images have been segmented by medical experts and the DSC score before and after registration is used to evaluate the registration accuracy. Parameter selection for this experiment was done using one of the image pairs (see figure S2 of supplementary material).

3.3. Inter-subject alignment of mammillary bodies in neonatal brains

The mammillary bodies are small brainstem nuclei located at the posteroinferior aspect of the hypothalamus. Inter-subject registration and accurate contour propagation of these structures can be useful in population statistics analyses (see for example Vann *et al* (2022)). In order to evaluate the ability of SOLID to match similar anatomical features across images of different individuals, we used 5 T1w MR brain scans of neonatal subjects.

[3] http://learn2reg.grand-challenge.org

Table 2. DSC scores on liver, kidneys and spleen for the abdominal MR/CT registrations. Each entry shows the mean (sd) of the DSC scores across the seven different cases.

	No registration	EVolution (DCT)	MIND (DCT)	Elastix	SOLID
Liver	0.54 (0.15)	0.74 (0.17)	0.75 (0.17)	0.82 (0.10)	0.82 (0.12)
Kidneys	0.33 (0.14)	0.67 (0.32)	0.72 (0.31)	0.59 (0.38)	0.82 (0.11)
Spleen	0.42 (0.20)	0.67 (0.27)	0.69 (0.29)	0.68 (0.20)	0.73 (0.18)

The scans were obtained from the 'Developing human connectome project' (dHCP) database[4] (Hughes *et al* 2017, Bastiani *et al* 2019, Fitzgibbon *et al* 2020). Due to the small size of the structures, we evaluated the registration accuracy using the Hausdorff 95 measure. Our dataset contains 5 subjects resulting in 20 registrations and 40 evaluation data points (2 mammillary bodies per registration). The original image resolution was $512 \times 512 \times 512$ with a voxel size of $0.28 \times 0.28 \times 0.2$ mm^3. We resampled the images to a resolution of $256 \times 256 \times 256$. The high resolution together with the small size of the mammillary bodies makes this dataset suitable for accuracy assessment. The parameters of the algorithms tested were optimized using a dataset of adult human brains of the same contrast (see figure S3 of supplementary material).

3.4. Pre-operative to post-operative brain registration

Registration of images with evolving pathologies and for therapy response assessment is a demanding task due to significant changes of the tissue. The existence of features that are not present in both images prohibits the use of intensity-matching algorithms unless appropriate steps are implemented, such as tumor masking where the similarity metric is not computed in the tumor region (Brett *et al* 2001). The brain tumor sequence registration (BraTS-Reg) challenge (Baheti *et al* 2021) comprises of pre-operative baseline and post-operative follow-up multi-parametric MRI scans of glioma patients. Ground truth annotations of 6–50 landmarks per patient are provided for a total of 140 patients which constitutes the challenge's training dataset. The images provided by the challenge organizers have been pre-processed with a rigid registration to a common template, resampling to a common isotropic resolution ($1 \times 1 \times 1$ mm^3) and skull-stripping. Out of the four MRI contrasts that are available, we chose to work with the contrast-enhanced T1-weighted (T1-CE) due to the richness in features and clinical relevance. One case (case 54) was discarded from our experiment because the images were corrupted. Algorithm configuration was done using a random single case (see figure S4 of supplementary material). In accordance with the BraTs-Reg challenge evaluation pipeline, we evaluated the tested algorithms using the median absolute error (MAE) and robustness criteria. We also included the MAE 30, namely the 30th percentile of largest MAE across all cases. Finally, for a more holistic evaluation, we report the cumulative distributions of the landmark L^1 distance.

3.5. Dependence of the results on the γ parameter

In order to assess the robustness of our proposed algorithm with respect to the choice of the $\tilde{\gamma}$ parameter, we co-registered an MR/synthetic CT pair of pelvic images for a wide range of $\tilde{\gamma}$ values. Being artificially generated, the CT is pre-aligned with the MR scan. The size of the images was $256 \times 256 \times 256$ with a voxel size of $1.7 \times 1.7 \times 0.35$ mm^3. Since we are interested in robustness and not anatomical plausibility, we applied a synthetically generated, B-spline based deformation to the moving image and performed a series of registrations for various values of $\tilde{\gamma}$. In this way, the ground truth deformation is known and we can evaluate the target registration error over the entire anatomy.

4. Results

4.1. Multi-modal (MR-CT) alignment of abdominal images

The results of the cross-modality experiment are shown in table 2. SOLID achieved a consistently optimal performance in all three organs of interest. Elastix achieved (on average) the same DSC score in the liver but performed considerably worse in the kidneys. We noticed that the most significant differences were observed in cases with small initial overlap where the large scale initial displacement makes it particularly challenging for the algorithms to match the corresponding anatomical structures. For further insight, figure 1 shows coronal slices from three cases before and after registration using the different algorithms. It can be observed that in those cases, SOLID delivers more accurate contour overlap, particularly on the lung-liver interface.

[4] http://www.developingconnectome.org/

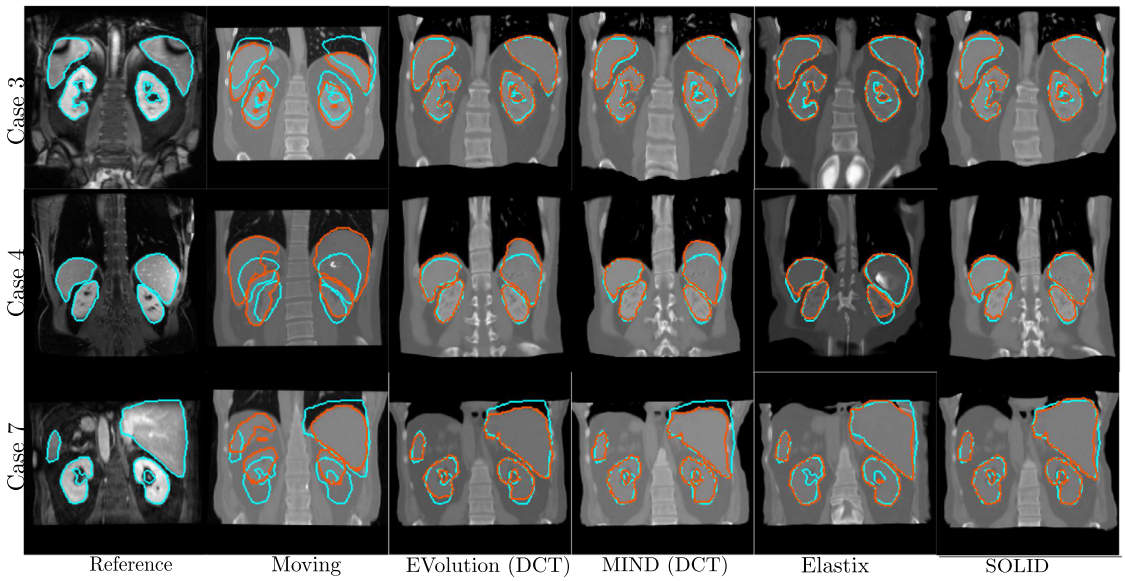

Figure 1. Three examples of MR-CT registration. From left to right: Reference (MR) image, Moving (CT) image and registered images using the different tested algorithms. The reference (blue) and moving/registered (orange) contours of the liver, spleen and kidneys are overlaid on the anatomy.

Table 3. DSC results on liver and kidneys for the abdominal T1/T2 registration test.

Case	No registration	EVolution (DCT)	MIND (DCT)	Elastix	SOLID
1	0.44	0.92	0.68	0.61	0.92
2	0.35	0.69	0.66	0.38	0.77
3	0.52	0.64	0.54	0.76	0.85
4	0.50	0.91	0.86	0.65	0.94
5	0.58	0.93	0.85	0.78	0.94
Avg.	0.48	0.82	0.72	0.64	0.88

4.2. Cross-contrast abdominal MRI images with variable FOV

The DSC scores before and after registration are reported in table 3 and indicate the robustness of the proposed algorithm in this challenging task. As can be seen from the examples in figure 2, the initial misalignment is predominantly due to the variable FOV, a known and longstanding challenge in the field of image registration. The DSC scores together with the visual information from figure 2 illustrate the high anatomical accuracy of SOLID in spite of the large initial misalignment. The gradient-based EVolution scored the second best average DSC score but was prone to significant registration errors in certain cases such as case 3 illustrated in figure 2 (last case).

4.3. Inter-subject alignment of mammillary bodies in neonatal brains

The results of this experiment are summarized in the boxplots of the HD-95, shown in figure 3. We observe that the 75th percentile is well below 1 mm for the proposed algorithm indicating a highly consistent performance. SOLID also scored the lowest in terms of the average and median HD-95. In this experiment there were no notable outliers as all algorithms were similar in performance as can be inferred from the comparable ranges of the boxplots.

4.4. Pre-operative to post-operative brain registration

Cumulative distribution functions (CDF's) of the L1 landmark distance are shown in figure 4 and the evaluation metrics used in the original BraTS-Reg challenge in table 4. The proposed algorithm scored the best in terms of Mean MAE with MIND being second best by a small margin. The robustness metric also highlights a distinction between the MI-based Elastix and the other tested algorithms, that can possibly be attributed to the sensitivity of MI to absent correspondences. To further highlight the nuances in the performance of the different algorithms tested, figure 5 shows an example of a case from the BratS-Reg dataset.

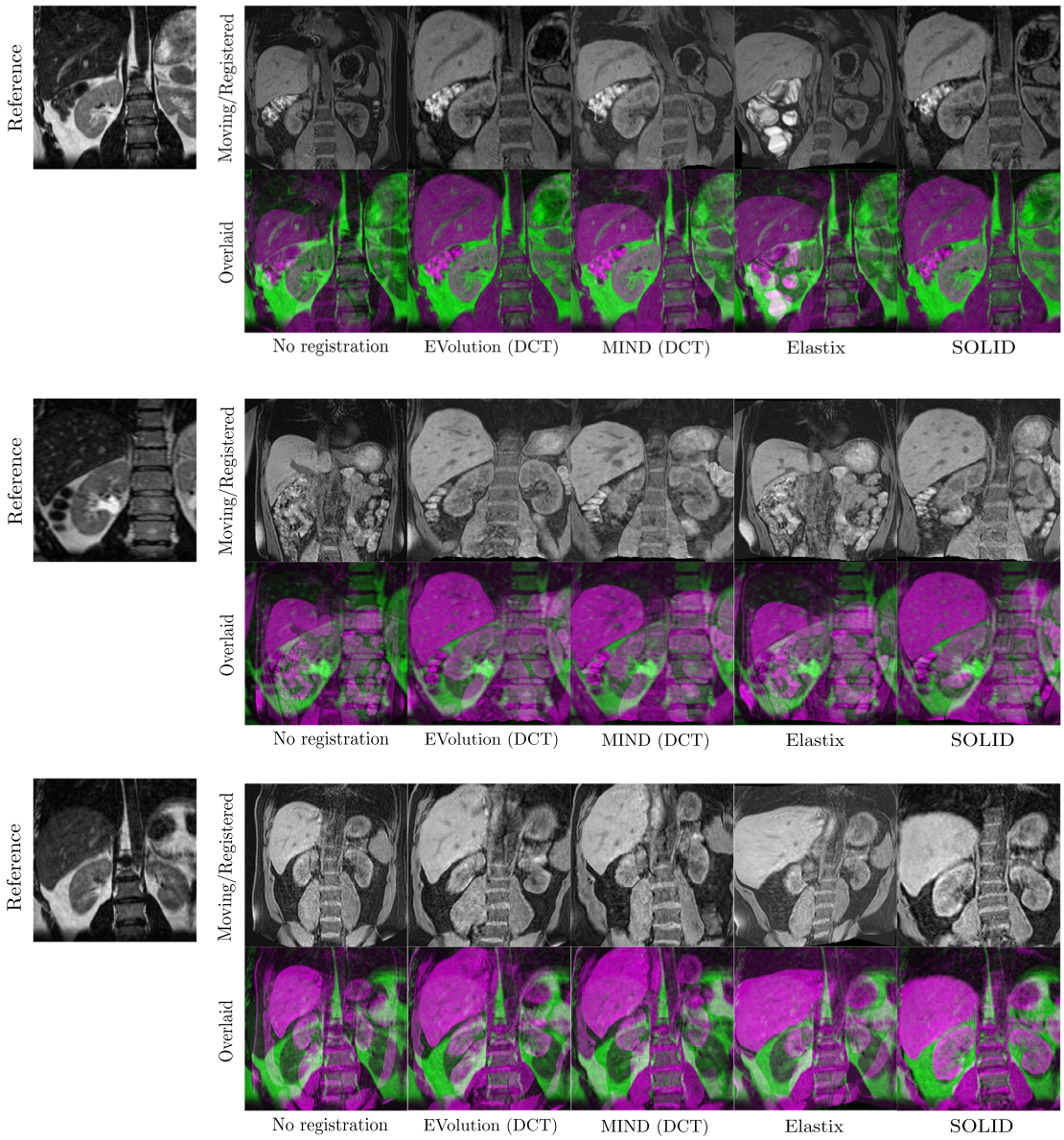

Figure 2. The first three cases (from top to bottom) from the cross-contrast abdominal MRI dataset. For each case, coronal slices of the reference, moving and registered volumes are shown together with their overlaid counterparts (with the reference image) to demonstrate the degree of alignment before and after registration.

4.5. Dependence of the results on the γ parameter

The results of this last experiment are shown in figure 6. The scale on the horizontal axis is logarithmic and the range of $\tilde{\gamma}$ values encompasses ten orders of magnitude. We observe that despite the large width of this range, the maximum TRE fluctuation is approximately 0.12 mm or 4.4%. This is well below the voxel size.

5. Discussion

In this study, we have investigated the use of second order information to locally quantify the similarity of medical images. We have attempted to include a significant amount of heterogeneous data but this was not always possible due to the limited availability of annotated data. The results obtained from our experiments, provide useful insights into the applicability of SOLID in a number of challenging registration tasks. In a brain anatomy for example, the presence of dense geometrical shapes and anatomical landmarks makes the registration problem particularly difficult. In order to expand on this point, figure 7 demonstrates the sensitivity of SOLID to changes in the shape of anatomical features inside the brain. As explained in this figure, the algorithm focuses on corners where the effect of curvature is prevalent. In the case of pre-operative to post-operative registration that we examined, the large local deformations require a similarity metric that is highly sensitive to those shape disparities and we believe that the shape operator based method proposed here can be

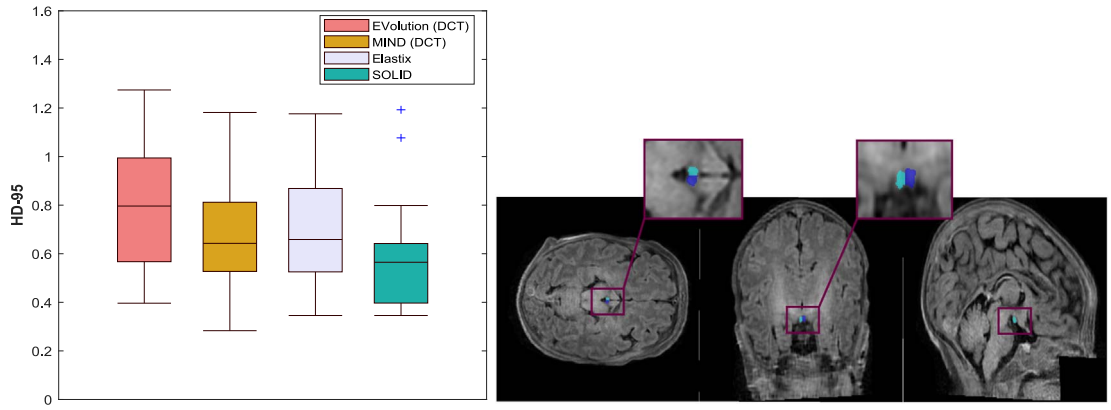

Figure 3. Left: Results from the inter-subject alignment of mammillary bodies. Each data point of the boxplots represents the Hausdorff 95 (HD-95) distance of a single mammillary body. The median (range) before registration was 5.64 (2.03–8.22). Those datapoints are omitted because their large values would skew the *y*-axis limits. We observe that SOLID achieves the lowest median of all the tested algorithms. Right: Example of a neonatal brain from the dataset. The two mammillary bodies are indicated inside the box.

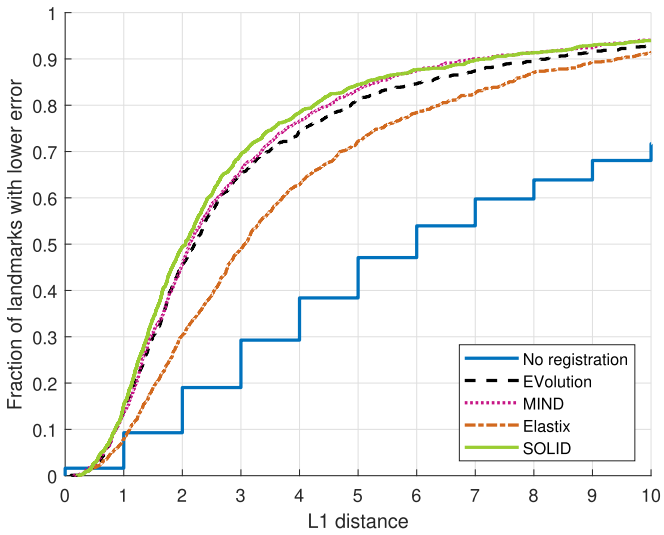

Figure 4. Registration of pre-operative to post-operative T1-CE scans from the BraTS-Reg challenge. Cumulative distributions of all the landmarks L1 distance (1240 data points in total).

Table 4. Summary of pre-operative to post-operative registration task. The table shows the median absolute error (in mm), robustness metric *R* and MAE 30 (in mm) for the cases of the BraTS-Reg challenge training dataset. SOLID achieved the lowest mean MAE and the lowest standard deviation.

Method	Mean MAE (sd)	Mean *R* (sd)	MAE 30
No registration	8.14 (7.62)	—	9
EVolution (DCT)	2.64 (1.61)	0.78 (0.22)	2.86
MIND (DCT)	2.50 (1.42)	0.78 (0.23)	2.57
Elastix	3.37 (1.50)	0.68 (0.25)	3.84
SOLID	2.37 (1.24)	0.79 (0.20)	2.58

advantageous over other methods that utilize first order information (gradients) or statistical means. This is also highlighted in figure 5. For the same experiment, the MAE of SOLID is also lower than the one reported in Mok and Chung (2022). Notably, the latter was the top performing algorithm of the BratS-Reg challenge (based on the validation cases). We believe that those are strong indications towards the favorable applicability of SOLID in

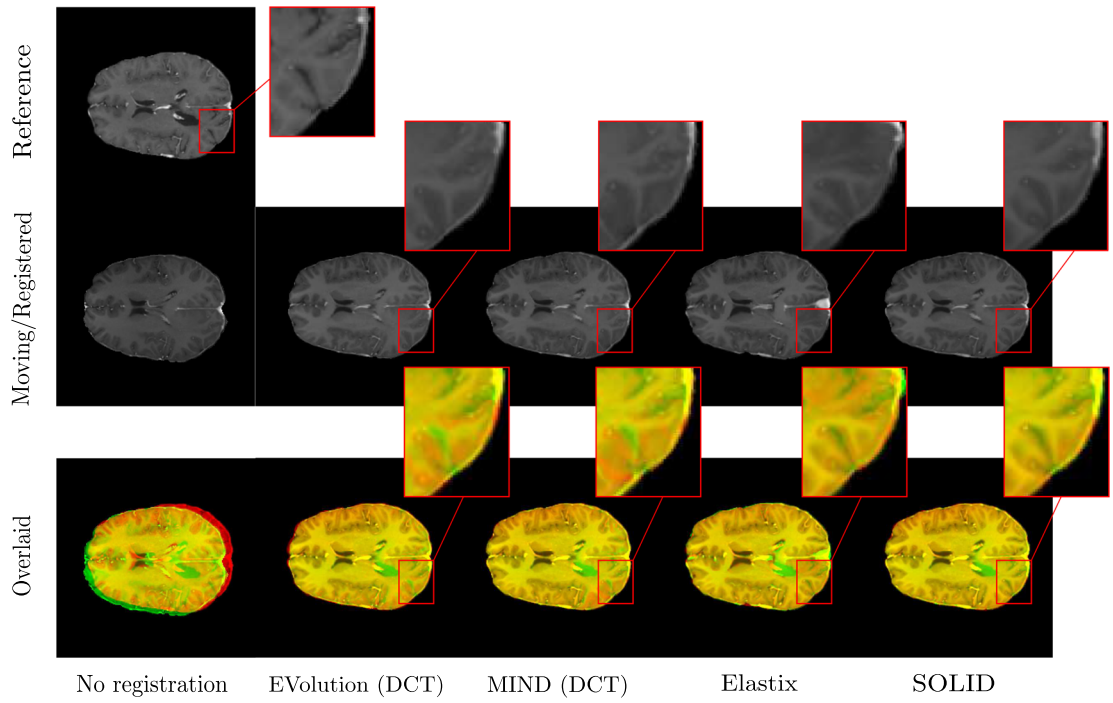

Figure 5. Visual comparison of registration results for a single case from the BratS-Reg dataset. The case shown (case 101) was the 'training' dataset that we used to calibrate the hyperparameters of the registration algorithms. Middle axial slices of the reference and moving volumes are shown along with the corresponding slices of the registered volumes using the four algorithms tested. The last row shows overlaid images of the reference (green) and moving/registered (red) images. An anatomical region has been been highlighted to illustrate the different local accuracy of the various algorithms.

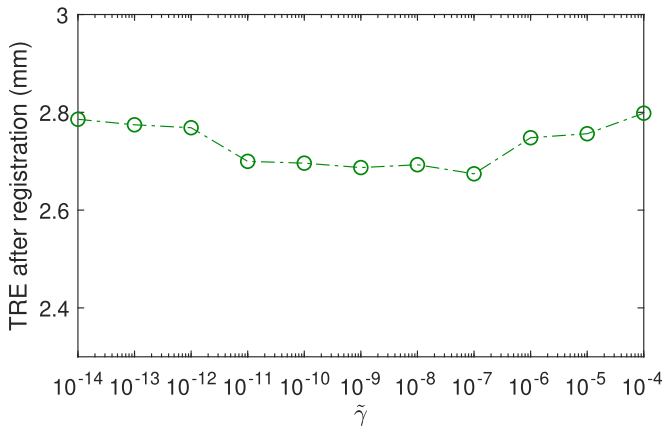

Figure 6. The target registration error (TRE) after an MR/synthetic CT registration as a function of the $\tilde{\gamma}$ parameter used. The scale of the x-axis is logarithmic and encompasses ten orders of magnitude.

this task. The downside of utilizing higher order information of the images is that the resulting method is more sensitive to noise and often computationally expensive.

The experiments conducted in the present study indicate that the SOLID metric has ample potential to be used as a similarity metric both for mono-modal and multi-modal tasks. While various modern methods in image processing are highly specialized for certain tasks and specific types of data, the proposed approach is suitable for a broad spectrum of applications. An additional merit of the proposed method is its easy adaptation to different tasks using a very small set of tuning parameters. Indeed, as we have previously noted, the only truly unspecified parameter is the regularization weight β in (6). The integration window ω in (4) can be set to $\omega = 1$ in most applications. The parameter $\tilde{\gamma}$ was shown in section 4.5 to only affect the registration to a negligible degree. This conclusion is corroborated by all examples that we examined. On this basis, no fine tuning was required throughout the experimental process. The value that we used was $\tilde{\gamma} = 10^{-8}$. Finally, concerning the computation time, the results reported in this work were obtained using our initial CPU implementation which

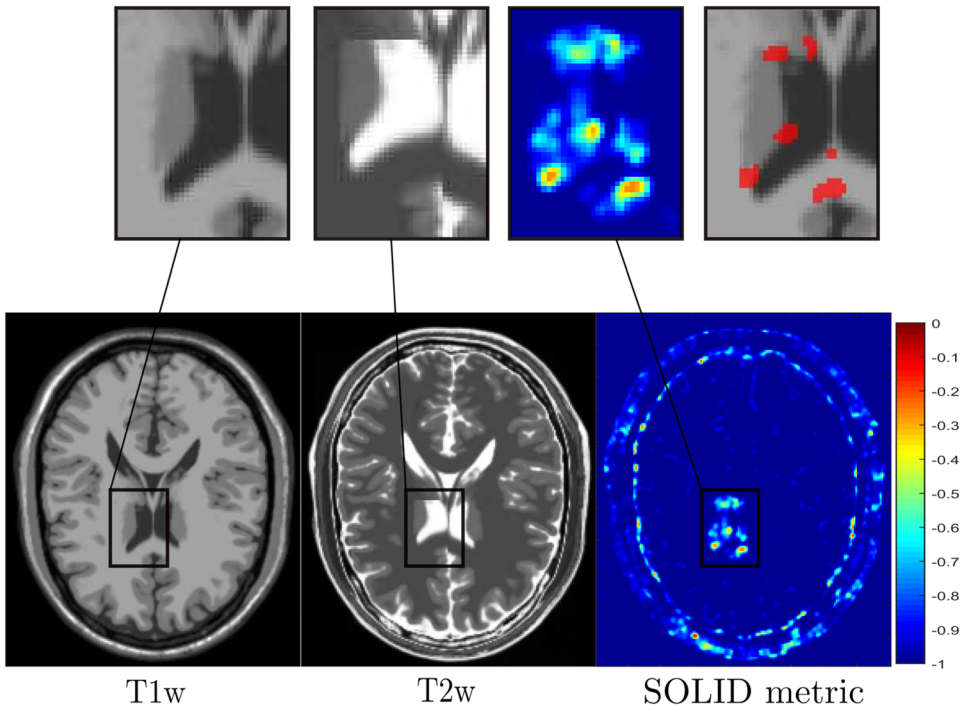

Figure 7. Axial slices of a T1w and a T2w brain scan. The original 3D volumes are obtained from the Brain Web database (Kwan *et al* 1996, Cocosco *et al* 1997, Collins *et al* 1998, Kwan *et al* 1999). The images are simulated and by construction pre-aligned. We applied a smooth local deformation in the vicinity of the lateral ventricle of the T2w image. The deformation is zero outside the region enclosed in the rectangle. In the third image, the SOLID metric is depicted with red corresponding to the maximum value of zero and blue corresponding to the minimum value of −1 as indicated in the colorbar. The magnified images demonstrate the sensitivity of the SOLID metric to the changes in the shape of anatomical features. The last image on the top row shows the area of interest in the T1w anatomy with the red points denoting the locations where the SOLID metric attains its highest values (top percentile). We observe that those points appear at the corners of geometric features, in agreement with the theoretical analysis of the shape operator as an object which encodes curvature information.

is computationally suboptimal, with typical registration times of a few minutes. However, the algorithm is highly parallelizable and a prototype GPU implementation that we have developed can achieve typical computation times of less than 20 seconds for registering $256 \times 256 \times 256$ images.

Finally, it should be pointed out that although the efficacy of the proposed SOLID similarity metric has been demonstrated in the context of the presented registration algorithm, its potential application domain is much broader. For instance, we already alluded to the interplay between traditional registration algorithms and learning-based approaches. In this sense, the proposed metric could be utilized as a similarity metric for unsupervised network training. More generally, the SOLID metric provides a method for locally quantifying the similarity of mono-modal and multi-modal images and can potentially be of interest in other image processing tasks.

6. Conclusion

We have proposed SOLID, a novel image registration similarity metric that identifies common shapes in images and performs very favorably for challenging cases such as variable FOV, different contrast, different acquisition modality and tissue appearance changes. Using second order information, the proposed registration algorithm is also highly sensitive to local shape differences, delivering improved accuracy over the evaluated alternative methods tested. The robustness and accuracy of the proposed method has been demonstrated through a series of diverse experiments encompassing a variety of registration tasks. Considering the increasing need for greater accuracy and the constantly improving imaging techniques, we are confident that the introduction of higher order data-similarity metrics can open additional possibilities in the development of clinically relevant registration methods.

Acknowledgments

This work has been supported in part by the Topconsortia for Knowledge and InnovationLifeSciences and Health, project no. LSHM19031 (POPCORN). Data were provided by the developing Human Connectome

Project, KCL-Imperial-Oxford Consortium funded by the European Research Council under the European Union Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement no. [319456]. We are grateful to the families who generously supported this trial.

Data availability statement

The data cannot be made publicly available upon publication because they contain sensitive personal information. The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Paris Tzitzimpasis https://orcid.org/0000-0003-4787-0145
Cornel Zachiu https://orcid.org/0000-0001-9755-4584

References

Baheti B *et al* 2021 The brain tumor sequence registration challenge: establishing correspondence between pre-operative and follow-up mri scans of diffuse glioma patients (https://doi.org/10.48550/arXiv.2112.06979)
Bastiani M *et al* 2019 Automated processing pipeline for neonatal diffusion MRI in the developing human connectome project *Neuroimage* **185** 750–63
Brett M, Leff A P, Rorden C and Ashburner J 2001 Spatial normalization of brain images with focal lesions using cost function masking *Neuroimage* **14** 486–500
Buades A, Coll B and Morel J M 2005 A review of image denoising algorithms, with a new one *Multiscale Model. Simul.* **4** 490–530
Clark K *et al* 2013 The cancer imaging archive (tcia): maintaining and operating a public information repository *J. Digit. Imaging* **26** 1045–57
Cocosco C A, Kollokian V, Kwan R K-S, Pike G B and Evans A C 1997 Brainweb: online Interface to a 3d MRI simulated brain database *NeuroImage* **5** 4 https://brainweb.bic.mni.mcgill.ca/
Collins D L, Zijdenbos A P, Kollokian V, Sled J G, Kabani N J, Holmes C J and Evans A C 1998 Design and construction of a realistic digital brain phantom *IEEE Trans. Med. Imaging* **17** 463–8
Dalca A V, Balakrishnan G, Guttag J and Sabuncu M R 2019 Unsupervised Learning of Probabilistic Diffeomorphic Registration for Images and Surfaces *Med. Image Anal.* **57** 226–36
de Senneville B D, Zachiu C, Ries M and Moonen C 2016 Evolution: an edge-based variational method for non-rigid multi-modal image registration *Phys. Med. Biol.* **61** 7377
de Vos B D, Berendsen F F, Viergever M A, Sokooti H, Staring M and Išgum I 2019 A deep learning framework for unsupervised affine and deformable image registration *Med. Image Anal.* **52** 128–43
Fischer B and Modersitzki J 2004 A unified approach to fast image registration and a new curvature based registration technique *Linear Algebr. Appl.* **380** 107–24
Fitzgibbon S P *et al* 2020 The developing human connectome project (dhcp) automated resting-state functional processing framework for newborn infants *Neuroimage* **223** 117303
Fu Y, Lei Y, Wang T, Curran W J, Liu T and Yang X 2020 Deep learning in medical image registration: a review *Phys. Med. Biol.* **65** 20TR01
Haber E and Modersitzki J 2006 *Int. Conf. on Medical Image Computing and Computer-Assisted InterventionIntensity Gradient Based Registration and Fusion of Multi-Modal Images* (Springer) pp 726–33
Heinrich M, Jenkinson M, Bhushan M, Matin T, Gleeson F, Brady S and Schnabel J 2012 Mind: modality independent neighbourhood descriptor for multi-modal deformable registration *Med. Image Anal.* **16** 1423–35
Heinrich M P, Jenkinson M, Bhushan M, Matin T, Gleeson F V, Brady M and Schnabel J A 2012 Mind: modality independent neighbourhood descriptor for multi-modal deformable registration *Med. Image Anal.* **16** 1423–35
Horn B K and Schunck B G 1981 Determining optical flow *Artif. Intell.* **17** 185–203
Hughes E J *et al* 2017 A dedicated neonatal brain imaging system *Magn. Reson. Med.* **78** 794–804
Jaouen V, Conze P-H, Dardenne G, Bert J and Visvikis D 2021 *arXiv,* 10.48550/arXiv.2111.15509 30 Nov 2021Regularized directional representations for medical image registration
Klein S, Staring M, Murphy K, Viergever M A and Pluim J P 2009 Elastix: a toolbox for intensity-based medical image registration *IEEE Trans. Med. Imaging* **29** 196–205
Krebs J, Delingette H, Mailhe B, Ayache N and Mansi T 2019 Learning a probabilistic model for diffeomorphic registration *IEEE Trans Med Imaging* **38** 2165–76
Krebs J, Mansi T, Delingette H, Zhang L, Ghesu F C, Miao S, Maier A K, Ayache N, Liao R and Kamen A 2017 *Int. Conf. on Medical Image Computing and Computer-Assisted InterventionRobust non-rigid registration through agent-based action learning* (Springer) pp 344–52
Kwan R K-S, Evans A C and Pike G B 1996 An extensible mri simulator for post-processing evaluation *Int. Conf. on Visualization in Biomedical Computing* (Springer) pp 135–40
Kwan R-S, Evans A C and Pike G B 1999 MRI simulation-based evaluation of image-processing and classification methods *IEEE Trans. Med. Imaging* **18** 1085–97
Larrey-Ruiz J, Morales-Sánchez J and Verdú-Monedero R 2007 Generalized regularization term for non-parametric multimodal image registration *Signal Process.* **87** 2837–42
Maes F, Collignon A, Vandermeulen D, Marchal G and Suetens P 1997 Alignment by maximization of mutual information *IEEE Trans. Med. Imaging* **16** 187–98
Maintz J and Viergever M 1998 A review on medical image registration as an optimization problem *J. Biomed. Eng. Technol.* **2** 1–36
Mani V and Arivazhagan S 2013 Survey of medical image registration *J. Biomed. Eng. Technol.* **1** 8–25
Mellor M and Brady M 2005 Phase mutual information as a similarity measure for registration *Med. Image Anal.* **9** 330–43

Mok T C and Chung A C 2022 *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th Int. Conf., Proc., Part VIUnsupervised deformable image registration with absent correspondences in pre-operative and post-recurrence brain tumor mri scans (September 18–22, 2022) (Springer)* pp 25–35

Pluim J, Maintz J and Viergever M 2000 Image registration by maximization of combined mutual information and gradient information *IEEE Trans. Med. Imaging* **19** 809–14

Pluim J, Maintz J and Viergever M 2003 Mutual-information-based registration of medical images: a survey *IEEE Trans. Med. Imaging* **22** 986–1004

Shamonin D P, Bron E E, Lelieveldt B P, Smits M, Klein S and Staring M 2014 Fast parallel image registration on cpu and gpu for diagnostic classification of alzheimer's disease *Front. Neuroinformatics* **7** 50

Simonovsky M, Gutiérrez-Becker B, Mateus D, Navab N and Komodakis N 2016 A deep metric for multimodal registration *19th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2016), ser. Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016* ed S Ourselin *et al* (Springer) pp 10–8 Available:https://hal.archives-ouvertes.fr/hal-01576914

Song G, Han J, Zhao Y, Wang Z and Du H 2017 A review on medical image registration as an optimization problem *Curr. Med. Imaging* **13** 274–83

Sun Y, Jolly M and Moura J 2004 *IEEE ICIPIntegrated registration of dynamic renal perfusion mr images* pp 1923–6

Sutour C, Aujol J-F, Deledalle C A and Denis de Senneville B 2015 Integrated registration of dynamic renal perfusion mr images *J. Math. Imaging Vis.* **53** 131–50

Vann S D, Zachiu C, Meys K M, Ambrosino S, Durston S, de Vries L S, Groenendaal F and Lequin M H 2022 Normative mammillary body volumes: from the neonatal period to young adult *NeuroImage: Rep.* **2** 100122

Viola P and Wells W 1997 Alignment by maximization of mutual information *Int. J. Comput. Vision* **24** 137–54

Wachinger C and Navab N 2012 Entropy and laplacian images: structural representations for multimodal registration *Med. Image Anal.* **16** 1–17

Yang X, Kwitt R, Styner M and Niethammer M 2017 Quicksilver: fast predictive image registration–a deep learning approach *NeuroImage* **158** 378–96