

New Multimodal Similarity Measure for Image Registration via Modeling Local Functional Dependence with Linear Combination of Learned Basis Functions

Joel Honkamaa¹[0000–0003–1532–9848] and Pekka Marttinen¹[0000–0001–7078–7927]

Aalto University, Finland

Abstract. The deformable registration of images of different modalities, essential in many medical imaging applications, remains challenging. The main challenge is developing a robust measure for image overlap despite the compared images capturing different aspects of the underlying tissue. Here, we explore similarity metrics based on functional dependence between intensity values of registered images. Although functional dependence is too restrictive on the global scale, earlier work has shown competitive performance in deformable registration when such measures are applied over small enough contexts. We confirm this finding and further develop the idea by modeling local functional dependence via the linear basis function model with the basis functions learned jointly with the deformation. The measure can be implemented via convolutions, making it efficient to compute on GPUs. We release the method as an easy-to-use tool and show good performance on three datasets compared to well-established baseline and earlier functional dependence-based methods.

Keywords: Deformable image registration · Multimodal similarity measure · Functional dependence

1 Introduction

The multimodal medical image registration aims to find a mapping between the anatomical coordinates of images of different modalities. This is an important prerequisite for the efficient utilization of complementary information provided by different imaging modalities. In deformable registration, the mappings are not limited to linear (affine). The problem is difficult and, while studied intensively for a few decades, it remains an active topic of research.

The registration of distorted images is traditionally formulated as an optimization problem that involves a similarity term and a regularization term [22]. In multimodal image registration, measuring similarity is particularly difficult due to the complex dependency between the registered images. Successful methods include measuring similarity with mutual information (MI) [30,24], a measure of statistical dependence. Another method of multimodal (affine) registration is the correlation ratio (CR) [27], which measures the more restrictive

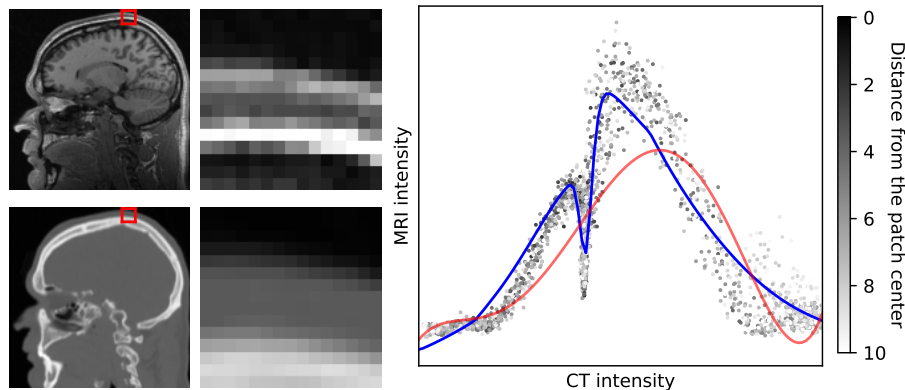


Fig. 1. We measure multimodal similarity via residuals of locally fitted functions (over sliding window). Each point on the right describes intensity value pair for voxels at identical locations in the patches (only one slice of the 3D volumes is shown). The points are weighted by the distance from the patch center. **Blue curve:** Learning basis (globally) allows reasonably good fit with very few basis functions (number of terms $J = 4$). **Red curve:** Polynomial functions (number of terms $J = 6$) struggle to fit the high frequencies. ©Copyright CERMEP – Imagerie du vivant, www.cermep.fr and Hospices Civils de Lyon. All rights reserved.

functional dependence between intensity values, addressing the fact that mutual information largely ignores the proximity of intensity values. In addition, CR is efficient to compute. However, functional dependence is, in general, too restrictive, especially for deformable registration.

Both MI and CR are global measures, making them sensitive to spatial changes in statistical relationships between the image intensities, e.g. due to non-uniform bias fields [24,12,25]. These measures have been defined in local contexts [24,17,25] for which CR is particularly attractive because it is cheaper to localize and functional dependency is often sufficient for small contexts. In this work, we build on this and develop a well-working image registration based on estimating local functional dependence with the following novel contributions.

- We model the intensity relationship with the linear basis function model. The coefficients are fitted locally in closed form, and the basis functions are learned jointly with the deformation. Earlier metrics measuring local functional dependence (RaPTOR [25] and SRWCR [8]) estimate the intensity relationship using Parzen windowing approximations of conditional expectation (Eq. 1). RaPTOR uses a smooth variant of Eq. 2 whereas SRWCR estimates full joint distribution for all subregions, which is very costly.
- We note that the measure can be implemented via a convolution operation, allowing for an easy but efficient GPU implementation.
- We release the method as an easy-to-use tool, *Locor* (available at <https://github.com/honkamj/locor>), and show good performance on three datasets.

2 Notation

We view the images to be registered as mappings from physical coordinates to intensity values $I_A : \mathbb{R}^n \rightarrow \mathbb{R}$ and $I_B : \mathbb{R}^n \rightarrow \mathbb{R}$ where n is the dimensionality of the image. Trilinear interpolation is used to make the images continuous in practice. We limit the mathematical analysis to single channel images for clarity. The task is then to find the deformation $d : \mathbb{R}^n \rightarrow \mathbb{R}^n$, describing the mapping from the anatomical locations of the image I_A to the corresponding locations in the image I_B . Only part of the domain \mathbb{R}^n contains a valid image, but for clarity, we do not explicitly mention this in the formulas. In general, any comparison measures are applied over the overlapping valid region between the images.

When taking the expected value or variance of images, we interpret the images as random variables on the intensity values (generated by uniformly sampling a coordinate), as is common in the literature [30,26].

3 Related work

The correlation ratio (CR) [27] is based on the assumption that there exists a functional relationship between the intensity values of the registered images $f(I_A) \approx I_B$ for some $f : \mathbb{R} \rightarrow \mathbb{R}$. The variance of $f(I_A) - I_B$ could then be used as a similarity measure. Since f is unknown a priori, they propose to compute the measure for optimal f (with which the variance is minimized), yielding

$$1 - CR(I_A, I_B) = \frac{\min_f \text{Var}[f(I_A) - I_B]}{\text{Var}[I_B]} = \frac{\text{Var}[\mathbb{E}[I_B|I_A] - I_B]}{\text{Var}[I_B]}. \quad (1)$$

Assuming that the intensity values are discrete, one can compute the measure exactly as variance of I_B over the isosets (voxels of equal value) of the image I_A . Denoting isosets with intensity a as Ω_a , one obtains (here $N := \sum_a |\Omega_a|$):

$$1 - CR(I_A, I_B) = \frac{\sum_a \sum_{x \in \Omega_a} \left(I_B(x) - \frac{1}{|\Omega_a|} \sum_{\tilde{x} \in \Omega_a} I_B(\tilde{x}) \right)^2}{N \text{Var}[I_B]} \quad (2)$$

To obtain a differentiable measure, one can use soft windowing to define the isosets (often called Parzen windowing). Alternatively, one could assume a specific parametric form for the function f , and in the appendix of [26] and later in [28] Roche et al. explored the function assuming a polynomial form. It should also be noted that the very popular cross-correlation (for intramodality registration) can be derived by assuming an affine form for f [26,3].

The correlation ratio, applicable to rigid or affine multimodal registration, was extended to deformable registration in RaPTOR [25] by computing the measure over small randomly sampled patches. The idea was further explored in SRWCR [8] by Gong et al. where patches are instead sampled over a regularly spaced grid, and voxels within each patch are given less weight further they are from the patch center. The method was implemented for a GPU to make it computationally tractable. Both RaPTOR and SRWCR compute the local CR measure by estimating conditional expectation with Parzen windowing.

4 Methods

Maximum likelihood has been used to derive global CR measures in [26]. Here, we first extend the framework to local CR measures by modeling local neighborhoods separately and subsequently aggregate the local losses into a global loss, which we maximize to learn the unknown deformation and other parameters.

Consider a local neighborhood around a point, say $r \in \mathbb{R}^n$, and assume that intensity values of image I_A are modeled with a Gaussian distribution with a mean function f_r that takes as input the intensity values of an (aligned) image I_B , with noise variance increasing with the increasing distance from r :

$$I_A(x) \sim \mathcal{N}(f_r(I_B(x)), \phi(x - r)\sigma_r). \quad (3)$$

In Eq. 3, $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ is a function that increases with distance from the origin ($\phi(0) = 1.0$), and σ_r is the variance in r . As in [26], we assume that the errors in the predicted intensity values of I_A at different locations x are independent given I_B , and define a likelihood which we denote by $p_r(I_A|I_B, \sigma_r, f_r)$. This likelihood can be maximized to estimate the unknown f_r and σ_r . Furthermore, the same likelihood will be used to find the (local) deformation from I_A to I_B by finding the deformed I_B with the highest likelihood (combined with a regularization loss), but the deformation is suppressed in the notation for clarity.

The standard formula for the maximized log-likelihood for a Gaussian model (maximization w.r.t. σ_r), yields a local loss (similar to a weighted least squares):

$$\mathcal{L}_r := \max_{\sigma_r} \log p_r(I_A|I_B, \sigma_r, f_r) = -\frac{1}{2}|X| \log \sum_{x \in X} \frac{(f_r(I_B(x)) - I_A(x))^2}{\phi(x - r)} + C,$$

where X is the set of all considered spatial locations in the image.

To derive a global loss \mathcal{L} , we normalize and average local \mathcal{L}_r over all $r \in X$. Although this no longer corresponds to a proper log-likelihood, this nevertheless yields a meaningful global loss that is efficient to optimize. Different modeling assumptions for f_r lead to different loss functions [26]. Not assuming a particular form for f leads to the estimation of conditional expectation (Eq. 1), used by RaPTOR and SRWCR. Local cross-correlation (for intra-modality registration) can be derived from assuming affine form [26,3]. In the context of rigid registration and global CR, a polynomial f has been proposed [28].

To find the optimal compromise in modeling assumptions (strict vs. loose), we propose to model f as a linear combination of learned non-linear basis functions

$$f_{\theta_r, \omega}(x) := \sum_{j=1}^J \theta_r^{(j)} \psi_\omega^{(j)}(x) \quad (4)$$

where $\theta_r^{(j)} \in \mathbb{R}$ are the local coefficients, and ψ_ω the global basis functions parametrized by ω . We define ψ_ω as a small fully connected neural network, and the parameters can be learned jointly with the deformation (via gradients) while θ_r can be solved in closed form. In our experiments, relatively small number of

terms J (e.g. 4) is enough for good results (saving compute), and having too high J can be detrimental by allowing for poorly matching images with high likelihood.

Maximizing for coefficients θ_r is a standard weighted linear least squares with input-output pairs $(I_B(x), I_A(x))_{x \in X}$ and input transformation functions $\psi_\omega^{(j)}$, yielding the input matrix $\Psi_\omega \in \mathbb{R}^{|X| \times J}$, $(\Psi_\omega)_{i,j} := \psi_\omega^{(j)}(I_B(x_i))$, the output vector, $y \in \mathbb{R}^{|X|}$, $y_i := I_A(x_i)$, and the diagonal weight matrix $W_r \in \mathbb{R}^{|X| \times |X|}$, $(W_r)_{i,i} := \frac{1}{\phi(x_i - r)}$. The well-known solution is:

$$\hat{\theta}_r := \operatorname{argmax}_{\theta_r} \mathcal{L}_r = (\Psi_\omega^T W_r \Psi_\omega)^{-1} \Psi_\omega^T W_r y.$$

As the final loss we average the formula $\max_{\theta_r} \mathcal{L}_r$ over all $r \in X$ while removing the additive and multiplicative constants:

$$\mathcal{L} := \frac{1}{|X|} \sum_{r \in X} \log \sum_{x \in X} \frac{(f_{\hat{\theta}_r, \omega}(I_B(x)) - I_A(x))^2}{\phi(x - r)}. \quad (5)$$

To use the loss to learn the deformation d one has to simply substitute I_B with $I_B \circ d$. We use average over the log-likelihood instead of likelihood followed by local normalization (as with CR) since the log-form has the same scale invariance property (with respect to the derivatives) and avoids computing the local variance (see the ablation study in Section 5.3). Computing Eq. 5 naively is expensive. However, assuming that the considered locations X lie on a regular grid, substituting $f_{\hat{\theta}_r, \omega}$ in Eq. 4 into Eq. 5, and rearranging terms yields a formulation of the loss \mathcal{L} that can be calculated efficiently with convolutions:

$$\begin{aligned} \mathcal{L} = \frac{1}{|X|} \sum_{r \in X} \log & \left[\sum_{j=1}^J \sum_{k=1}^J \hat{\theta}_r^{(j)} \hat{\theta}_r^{(k)} \left((\psi_\omega^{(j)} \circ I_B)(\psi_\omega^{(k)} \circ I_B) * \frac{1}{\phi} \right) (r) \right. \\ & \left. - 2 \sum_{j=1}^J \hat{\theta}_r^{(j)} \left((\psi_\omega^{(j)} \circ I_B) I_A * \frac{1}{\phi} \right) (r) + \left(I_A^2 * \frac{1}{\phi} \right) (r) \right]. \end{aligned} \quad (6)$$

Similarly, the terms required to calculate the least squares solution $\hat{\theta}$ can be expressed with convolutions as

$$\begin{cases} (\Psi_\omega^T W_r \Psi_\omega)_{j,k} &= \left((\psi_\omega^{(j)} \circ I_B)(\psi_\omega^{(k)} \circ I_B) * \frac{1}{\phi} \right) (r) \\ (\Psi_\omega^T W_r y)_j &= \left((\psi_\omega^{(j)} \circ I_B) I_A * \frac{1}{\phi} \right) (r). \end{cases} \quad (7)$$

NB. convolution, given any functions $h_1, h_2 : \mathbb{R}^n \rightarrow \mathbb{R}$, is defined as the sum $(h_1 * h_2)(x) := \sum_{x \in X} h_1(x) h_2(x - r)$.

4.1 Further details

Incorporating derivative information: The method allows efficient registration of multichannel data. We take advantage of this and augment input images

with additional spatial derivative magnitude channel. This slightly improved the results (see the ablation study in Section 5.3). A similar approach was used with the polynomial f for the rigid global CR registration in [28].

Practical implementation: We incorporate the metric into a multi-resolution (see, e.g., [4,21]) registration pipeline with bidirectional formulation via scaling and squaring [2]. We apply the similarity loss in both directions and take the average. We use convolutions with stride 3 to save computational cost, but we shift the sampling grid randomly over iterations, leading to equal weighting for all spatial locations. We define the kernel $\frac{1}{\phi}$ as the Gaussian kernel (radially symmetric and separable) and truncate it at 3 standard deviations. The method is implemented in PyTorch and the optimization is performed with gradients of loss w.r.t. the deformation d and the neural network parameters ϕ via automated differentiation and Adam [16] optimizer. For regularization of the deformation we use the bending energy penalty [29]. In the experiments, one registration took under 2 minutes on a modern GPU (V100), and under 1 minute with 2 GPUs. The evaluation of similarity and its derivative took (combined time) for the highest resolution level approximately from 50ms to 500ms (volume size differed a lot between the datasets). The method is available at <https://github.com/honkamj/locor>.

5 Experiments

We conducted experiments on three datasets: two real and one semi-synthetic. The code for all experiments is available at <https://github.com/honkamj/locor-experiments>.

We evaluated **abdomen MRI-CT** registration on Learn2Reg 2021 [6,1,7,18,14] data (CC BY 3.0 license) originally from The Cancer Imaging Archive (TCIA) project, which contains 8 sets of MRI-CT image pairs with evaluation based on anatomical segmentation masks of abdominal organs (dice score). We omitted one subject (TCIA 0006) from the test set due to a very different appearance compared to the other images, rendering our hyperparameter optimization setup meaningless.

We used CERMEP-IDB-MRXFDG database [19] (CCO license) with 37 subjects for evaluating **head MRI-CT** registration. Since the database contains no labels for evaluation, we generated pseudo-CT images using a deep learning image-to-image translation method designed for geometrically accurate cross-modality synthesis [15]. The mean absolute error (MAE) between the pseudo-CT and registered CT images was used for evaluation. Based on our qualitative visual analysis, the metric corresponds well to the actual registration performance.

We generated a semi-synthetic dataset for **head MRI T2-PD** registration from IXI¹ dataset (CC BY-SA 3.0 license) by deforming PD images with random synthetic deformations, consisting of rigid and elastic components; the latter

¹ <http://brain-development.org/ixi-dataset/>

Table 1. Results of the main experiment showing mean and standard deviation over the test cases of each dataset. N_b/N refers to the ratio of test subjects for which the method beats our method. corrField was not tested on Head MRI-CT since the implementation could not handle images with different coordinate systems.

Method	Abdomen (MRI-CT)		Head (MRI-CT)		Head (MRI T2-PD)	
	Dice \uparrow	N_b/N	MAE \downarrow	N_b/N	TRE \downarrow	N_b/N
NiftyReg (MIND)	0.72(0.31)	1/4	85.67(10.18)	0/34	0.51(0.08)	0/49
NiftyReg (NMI)	0.67(0.30)	0/4	72.07(18.59)	0/34	0.98(0.28)	0/49
ANTs (Mattes MI)	0.67(0.33)	0/4	80.07(13.12)	0/34	1.20(0.32)	0/49
corrField	0.79(0.11)	0/4	-	-	1.76(1.72)	0/49
SRWCR [†]	0.84(0.09)	1/4	64.51(13.51)	1/34	0.30(0.06)	0/49
MINDSSC [†]	0.88(0.05)	2/4	67.28(20.20)	1/34	0.59(0.55)	0/49
MI [†]	0.85(0.04)	0/4	73.48(13.20)	0/34	2.04(6.61)	0/49
Locor polynom. (ours)	0.86(0.07)	0/4	63.03(13.16)	2/34	0.20(0.04)	1/49
Locor (ours)	0.88(0.04)		55.01(10.14)*		0.18(0.02)*	

* Wilcoxon signed-rank test has p-value < 0.005 compared to all the baselines.

[†] Our reimplementations to a setup similar with our method.

were generated by scaling and squaring [2] from Gaussian smoothed white noise [2]. The target registration error (TRE) in the voxels with respect to the known deformation was used as an evaluation metric (with background masked out). We used a randomly chosen subset of 52 subjects.

5.1 Hyperparameter optimization

For all the datasets, we separated 3 subjects into validation set, and for each method (including ours), we ran 100 trials on each validation set to find the best hyperparameters. We used a Gaussian process (GP) based optimizer and to choose the final parameters, and fitted the same Gaussian process to the best 60% (poor performing parameters were omitted due to their high variance) of the parameters and picked the parameters with the best posterior mean. The procedure was done to avoid "lucky" parameter values with poor average performance. The setup is fully documented by the codebase. All the results shown in the paper are averages of 5 runs with the same optimized hyperparameters.

5.2 Comparison to earlier methods

As baselines we used well-established methods NiftyReg [29,23,21,20] (with both MI and MIND [12]) and Advance Normalization Tools (ANTs) [4], as well as corrfield[11,10] which did well on Abdomen MRI-CT in Learn2reg 2021 [14]. We implemented SRWCR [8] into identical multi-resolution setup as our method, and for fair comparison, we did that also for mutual information via Parzen windowing (implementation from [9]) and MINDSSC [13,5] (later variant of MIND).

Table 2. Results of the ablation study showing mean and standard deviation over the validation cases of each dataset. **Log vs. Normalized:** Whether to use the logarithmic form (Eq. 5) or to exponentiate and normalize the local log-likelihoods. **Learned vs. Polynom.:** Whether to learn g or to use the fixed polynomial basis. **Deriv.:** Augment input volumes with derivative magnitude channel.

	Abdomen (MRI-CT)	Head (MRI-CT)	Head (MRI T2-PD)
	Dice \uparrow	MAE \downarrow	TRE \downarrow
Log + Learned	0.902 (0.006)	56.0 (11.9)	0.187 (0.019)
Log + Learned + Deriv.	0.904 (0.006)	55.0 (11.9)	0.175 (0.012)
Log + Polynom.	0.895 (0.011)	61.0 (12.5)	0.193 (0.018)
Log + Polynom. + Deriv.	0.900 (0.007)	60.4 (12.2)	0.184 (0.015)
Normalized + Learned	0.900 (0.007)	56.2 (11.9)	0.214 (0.016)
Normalized + Learned + Deriv.	0.903 (0.006)	55.3 (11.7)	0.193 (0.011)

We also compared against our method replaced with the fixed polynomial basis (Locor polynom.).

The results of the main experiment are given in Table 5. Our method clearly outperformed the baselines in head MRI-CT and head T2-PD registration, while in the abdominal MRI-CT dataset our method performed similarly to MINDSSC, although the dataset is too small to draw statistically significant conclusions.

5.3 Ablation study

Before the main experiment, we performed an ablation study on the validation set on different variants of the algorithm. The results are shown and the variants are explained in Table 5.3. For polynomial basis we searched degrees ≤ 5 and ≤ 3 for variants without and with derivative magnitude channel, respectively (maximum of 6 and 10 terms, respectively).

In conclusion, using the learned g improved the results compared to using the fixed polynomial basis. This was further confirmed by including the polynomial basis variant (with derivative channel) in the main experiment on the test set ("Locor polynom." in Table 5). Including the derivative magnitude channel also improved the results (we make no claims of statistical significance here). The logarithmic variant performed slightly better than the normalized one, making it a trivial choice due to cheaper computation.

6 Discussion

We provided further evidence that local functional dependence is a good similarity measure for generic multimodal registration. Modeling the function parametrically as a linear combination of learned basis functions improved the performance further, and the method outperformed all the baselines clearly on two

out of three datasets. The main downside of the method is the additional complexity due to the learning of the small neural network in conjunction, making optimization dynamics more difficult to understand or predict.

Acknowledgments. This work was supported by the Research Council of Finland (Flagship programme: Finnish Center for Artificial Intelligence FCAI, and grants 352986, 358246) and EU (H2020 grant 101016775 and NextGenerationEU). We also acknowledge the computational resources provided by the Aalto Science-IT Project.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Akin, O., Elnajjar, P., Heller, M., Jarosz, R., Erickson, B.J., Kirk, S., Lee, Y., Linehan, M.W., Gautam, R., Vikram, R., Garcia, K.M., Roche, C., Bonaccio, E., Filippini, J.: The cancer genome atlas kidney renal clear cell carcinoma collection (TCGA-KIRC) (2016)
2. Arsigny, V., Commowick, O., Pennec, X., Ayache, N.: A log-euclidean framework for statistics on diffeomorphisms. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006: 9th International Conference, Copenhagen, Denmark, October 1-6, 2006. Proceedings, Part I 9. pp. 924–931. Springer (2006)
3. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis* **12**(1), 26–41 (2008)
4. Avants, B.B., Tustison, N., Song, G., et al.: Advanced normalization tools (ants). *Insight j* **2**(365), 1–35 (2009)
5. Chen, J., Frey, E.C., He, Y., Segars, W.P., Li, Y., Du, Y.: Transmorph: Transformer for unsupervised medical image registration. *Medical Image Analysis* p. 102615 (2022). <https://doi.org/https://doi.org/10.1016/j.media.2022.102615>, <https://www.sciencedirect.com/science/article/pii/S1361841522002432>
6. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., et al.: The cancer imaging archive (tcia): maintaining and operating a public information repository. *Journal of digital imaging* **26**, 1045–1057 (2013)
7. Erickson, B.J., Kirk, S., Lee, Y., Bathe, O., Kearns, M., Gerdes, C., Rieger-Christ, K., Lemmerman, J.: The cancer genome atlas liver hepatocellular carcinoma collection (TCGA-LIHC) (2016)
8. Gong, L., Zhang, C., Duan, L., Du, X., Liu, H., Chen, X., Zheng, J.: Nonrigid image registration using spatially region-weighted correlation ratio and gpu-acceleration. *IEEE journal of biomedical and health informatics* **23**(2), 766–778 (2018)
9. Guo, C.K.: Multi-modal image registration with unsupervised deep learning. Ph.D. thesis, Massachusetts Institute of Technology (2019)
10. Hansen, L., Heinrich, M.P.: Graphregnet: Deep graph regularisation networks on sparse keypoints for dense registration of 3d lung cts. *IEEE Transactions on Medical Imaging* **40**(9), 2246–2257 (2021)
11. Heinrich, M.P., Handels, H., Simpson, I.J.: Estimating large lung motion in copd patients by symmetric regularised correspondence fields. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International

- Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part II 18. pp. 338–345. Springer (2015)
12. Heinrich, M.P., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F.V., Brady, M., Schnabel, J.A.: Mind: Modality independent neighbourhood descriptor for multi-modal deformable registration. *Medical image analysis* **16**(7), 1423–1435 (2012)
 13. Heinrich, M.P., Jenkinson, M., Papież, B.W., Brady, S.M., Schnabel, J.A.: Towards realtime multimodal fusion for image-guided interventions using self-similarities. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013: 16th International Conference, Nagoya, Japan, September 22-26, 2013, Proceedings, Part I* 16. pp. 187–194. Springer (2013)
 14. Hering, A., Hansen, L., Mok, T.C., Chung, A.C., Siebert, H., Häger, S., Lange, A., Kuckertz, S., Heldmann, S., Shao, W., et al.: Learn2reg: comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning. *IEEE Transactions on Medical Imaging* **42**(3), 697–712 (2022)
 15. Honkamaa, J., Khan, U., Koivukoski, S., Valkonen, M., Latonen, L., Ruusuuvuori, P., Marttinen, P.: Deformation equivariant cross-modality image synthesis with paired non-aligned training data. *Medical Image Analysis* **90**, 102940 (2023)
 16. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)
 17. Klein, S., Van Der Heide, U.A., Lips, I.M., Van Vulpen, M., Staring, M., Pluim, J.P.: Automatic segmentation of the prostate in 3d mr images by atlas matching using localized mutual information. *Medical physics* **35**(4), 1407–1417 (2008)
 18. Linehan, M., Gautam, R., Kirk, S., Lee, Y., Roche, C., Bonaccio, E., Filippini, J., Rieger-Christ, K., Lemmerman, J., Jarosz, R.: The cancer genome atlas cervical kidney renal papillary cell carcinoma collection (TCGA-KIRP) (2016)
 19. Mérida, I., Jung, J., Bouvard, S., Le Bars, D., Lancelot, S., Lavenne, F., Bouillot, C., Redouté, J., Hammers, A., Costes, N.: Cermep-idb-mrxfdg: a database of 37 normal adult human brain [18 f] fdg pet, t1 and flair mri, and ct images available for research. *EJNMMI research* **11**, 1–10 (2021)
 20. Modat, M., Cash, D.M., Daga, P., Winston, G.P., Duncan, J.S., Ourselin, S.: Global image registration using a symmetric block-matching approach. *Journal of medical imaging* **1**(2), 024003–024003 (2014)
 21. Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., Fox, N.C., Ourselin, S.: Fast free-form deformation using graphics processing units. *Computer methods and programs in biomedicine* **98**(3), 278–284 (2010)
 22. Oliveira, F.P., Tavares, J.M.R.: Medical image registration: a review. *Computer methods in biomechanics and biomedical engineering* **17**(2), 73–93 (2014)
 23. Ourselin, S., Roche, A., Subsol, G., Pennec, X., Ayache, N.: Reconstructing a 3d structure from serial histological sections. *Image and vision computing* **19**(1-2), 25–31 (2001)
 24. Pluim, J.P., Maintz, J.A., Viergever, M.A.: Mutual-information-based registration of medical images: a survey. *IEEE transactions on medical imaging* **22**(8), 986–1004 (2003)
 25. Rivaz, H., Chen, S.J.S., Collins, D.L.: Automatic deformable mr-ultrasound registration for image-guided neurosurgery. *IEEE transactions on medical imaging* **34**(2), 366–380 (2014)
 26. Roche, A., Malandain, G., Ayache, N.: Unifying maximum likelihood approaches in medical image registration. *International Journal of Imaging Systems and Technology* **11**(1), 71–80 (2000)

27. Roche, A., Malandain, G., Pennec, X., Ayache, N.: The correlation ratio as a new similarity measure for multimodal image registration. In: Medical Image Computing and Computer-Assisted Intervention—MICCAI'98: First International Conference Cambridge, MA, USA, October 11–13, 1998 Proceedings 1. pp. 1115–1124. Springer (1998)
28. Roche, A., Pennec, X., Malandain, G., Ayache, N.: Rigid registration of 3-d ultrasound with mr images: a new approach combining intensity and gradient information. *IEEE transactions on medical imaging* **20**(10), 1038–1049 (2001)
29. 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 transactions on medical imaging* **18**(8), 712–721 (1999)
30. Viola, P., Wells III, W.M.: Alignment by maximization of mutual information. *International journal of computer vision* **24**(2), 137–154 (1997)