Registration by Regression (RbR): a framework for interpretable and flexible atlas registration

Karthik Gopinath[*1] [*], Xiaoling Hu[1*], Malte Hoffmann[1], Oula Puonti[1,2**], and Juan Eugenio Iglesias[1,3,4,**]

[1] Massachusetts General Hospital and Harvard Medical School
[2] Danish Research Centre for Magnetic Resonance, Copenhagen University Hospital
[3] Centre for Medical Image Computing, University College London
[4] Computer Science and AI Laboratory, Massachusetts Institute of Technology
[*]Contributed equally as first authors and [**]contributed equally as senior authors.

Abstract. In human neuroimaging studies, atlas registration enables mapping MRI scans to a common coordinate frame, which is necessary to aggregate data from multiple subjects. Machine learning registration methods have achieved excellent speed and accuracy but lack interpretability. More recently, keypoint-based methods have been proposed to tackle this issue, but their accuracy is still subpar, particularly when fitting nonlinear transforms. Here we propose Registration by Regression (RbR), a novel atlas registration framework that is highly robust and flexible, conceptually simple, and can be trained with cheaply obtained data. RbR predicts the (x, y, z) atlas coordinates for every voxel of the input scan (i.e., every voxel is a keypoint), and then uses closed-form expressions to quickly fit transforms using a wide array of possible deformation models, including affine and nonlinear (e.g., Bspline, Demons, invertible diffeomorphic models, etc.). Robustness is provided by the large number of voxels informing the registration and can be further increased by robust estimators like RANSAC. Experiments on independent public datasets show that RbR yields more accurate registration than competing keypoint approaches, while providing full control of the deformation model.

Keywords: Registration · atlas · brain · ML interpretability

1 Introduction

Image registration seeks to find a spatial mapping between two images, and is a crucial component of human neuroimaging pipelines. Image registration enables: measuring change between different timepoints of the same subject [17]; building population atlases or subject-specific templates via co-registration of multiple scans [25]; measuring structural differences via tensor-based morphometry [20]; mapping pre-, intra-, and post-operative images [1]; automated segmentation [22]; and many others.

---

[*] Corresponding author: K. Gopinath. **Email:** `kgopinath@mgh.harvard.edu`

One application of these methods is registering brain MRI scans to digital atlases. Atlases are prototypical brain images that are representative of a population and are often built by co-registering and averaging scans from a cohort of subjects (e.g., the ubiquitous MNI atlas). Atlas registration enables aggregation of spatial information of multiple subjects into a common coordinate frame (CCF). Such CCFs enable studies like tensor- or voxel-based morphometry [5]. Moreover, it also enables the mapping of population-wide information to the native space of a specific subject, e.g., label priors for image segmentation [12]. Atlas registration typically relies on the same methods as pairwise methods.

Most classical registration methods rely on numerical optimization [31] to minimize a cost function, often combining an image similarity term with a regularizer; the former seeks to align the images as close as possible, while the latter prevents excessively convoluted (unrealistic) deformation. Decades of research have provided medical image registration with sophisticated features, e.g., diffeomorphisms [6], symmetry [9], or inter-modality mapping [32].

In 2017, faster registration methods based on deep learning were initially published. The first attempts were supervised and sought to predict ground truth fields obtained, e.g., with accurate classical methods [39]. These algorithms then evolved into unsupervised methods, trained with the same losses used by classical algorithms – possibly augmented with segmentation losses [7,10]. By now, deep learning registration methods have incorporated many of the features of their classical counterparts (e.g., diffeomorphisms [26], symmetry [21], progressive warping [28], or inter-modality support [14]) and have also started to rely on more modern architectures, e.g., transformers [8] rather than U-nets [7].

Very recently, keypoint methods [37,15] have been proposed, which predict blob-like feature maps such that their centers of gravity can be used to analytically compute a transform between the two images. This approach addresses two limitations of machine learning registration methods. First, the lack of interpretability associated with the "black box" nature of neural networks, by providing insight into the factors driving the alignment (i.e., the keypoints). And second, the lack of robustness of the affine component. Earlier learning approaches assumed that the inputs were already affinely aligned; some tried to predict the 12 parameter registration matrix [10,30], but these predictions were generally not robust (e.g., sensitive to initialization) and often failed to generalize well to new datasets. Instead, the affine matrix that best aligns two point clouds can be analytically estimated from keypoints in a robust fashion. Moreover, keypoints can also be used to fit nonlinear transforms, e.g., thin-plate splines (TPS) in KeyMorph [37]).

Here we present Registration by Regression (RbR), a novel framework that enables flexible, robust registration of brain MRI scans to digital atlases. In short, RbR is an evolution of keypoint methods where *every voxel* is a keypoint: it uses a convolutional neural network (CNN) that predicts, for each voxel of the input scan, its (x, y, z) atlas coordinates. Ground truth can be obtained cheaply via supervision with existing algorithms. This paradigm has a number of advantages: *(i)* Since the density of keypoints is much higher (over a million

vs ~500), nonlinear transforms can be fitted more accurately. *(ii)* It does not need pretraining to avoid clustering of keypoints in the center of the image. *(ii)* After a forward pass of the CNN, one can fit with minimal computation (closed-form expressions) a wide array of popular transforms and regularizers, including affine and nonlinear models like Bsplines [34], Demons [35], or the diffeomorphic log-polyaffine framework [2], among others. *(iv)* Every hyperparameter of the deformation model can be readily modified. This is in contrast with hypernetwork approaches [18], where only specific parameters can be tuned as the space of deformations is constant (e.g., control point spacing cannot be specified). *(v)* It can achieve enhanced robustness using robust estimators like RANSAC [13]. *(vi)* It is easy to implement, and ground truth for a new atlas or dataset can be easily obtained with classical registration methods.

2 Methods

Our proposed method is summarized in Figure 1. It comprises three different stages: data preparation, training, and test-time fitting.

2.1 Data preparation

The crux of RbR is a regression CNN that estimates, for every voxel of the input scan (with discrete coordinates x, y, z), its corresponding atlas coordinates x', y', z'. This CNN is trained in a supervised fashion using accurate nonlinear deformations obtained with a classical registration approach $\mathcal{R}$. If $I(x, y, z)$ is the input image and A is a reference atlas, the ground truth registration is:

$$\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3, \text{where} \quad \mathbf{F}(x, y, z) = [x', y', z'] = \mathcal{R}(x, y, z; I, A).$$

Therefore, a large labeled dataset can be obtained "for free" by simply registering a set of unlabeled images to the atlas at hand using a well-established registration algorithm. While there are no formal requisites on the deformation model of the registration method, diffeomorphic models that yield smooth and invertible fields are preferred [6]. Unlabeled images exist in abundance in public datasets – particularly 1mm isotropic MPRAGEs typically used in neuroimaging research.

2.2 Training

Training RbR is simple. Given a 3D voxel-wise regression CNN $\boldsymbol{N}(\cdot)$ with parameters $\boldsymbol{\theta}$, an atlas A, and a set of J images $\{I_j\}_{j=1,\ldots,J}$, the goal is to minimize the following loss function with stochastic gradient descent:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{j=1}^{J} \frac{1}{|\Omega_j|} \sum_{(x,y,z)\in\Omega_j} \ell_1[\mathcal{R}(x, y, z; I_k, A) - \boldsymbol{N}(x, y, z; I_k, \boldsymbol{\theta})], \qquad (1)$$

where Ω_j is the image domain of image I_j and is typically constrained to brain regions (i.e., excludes extracerebral tissue). Equation 1 is just the ℓ_1 norm between the predicted and ground truth atlas coordinates; we use ℓ_1 rather than
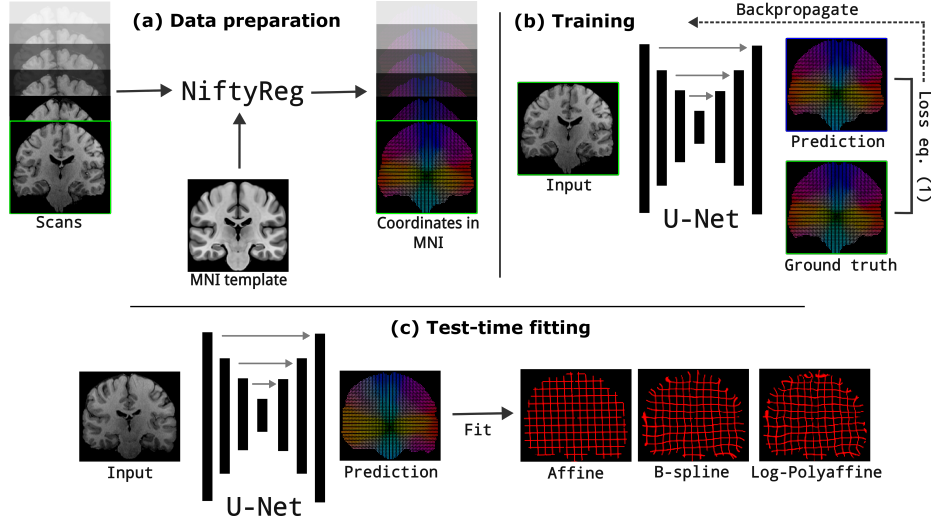
Fig. 1. (a) Training data are prepared by nonlinearly registering T1w scans from HCP and ADNI to the MNI template using NiftyReg. (b) A U-Net is trained to predict the MNI coordinates for every voxel of an input scan. (c) At test time, almost any transformation can be fitted between the input scan and predicted MNI coordinates.

ℓ_2 due to its higher robustness. We note that no regularization is needed: the CNN just does its best in a voxel-by-voxel basis, smoothness is provided by the deformation model and its parameters at test time. However, as with any other CNN, augmentation is crucial to endow the trained models with the best possible generalization ability. We use aggressive augmentation during training, including random blurring, bias field, noise, and deformation (affine and elastic).

2.3 Test-time fitting

The flexibility and robustness of RbR stem from its test-time fitting strategy. Given an input image I, one first processes it with the trained neural network to obtain $[x', y', z'] = \boldsymbol{N}(x, y, z; I, \hat{\boldsymbol{\theta}})$; we note that the atlas is not required at this point. Then, one can fit a deformation model to this predicted field to obtain the final atlas registration. We consider three different types of models in this article: a family of models based on basis functions, a model based on the Demons algorithm [35], and a diffeomorphic model based on the log-polyaffine framework [2]; we note that other models are possible.

Basis functions: a large class of transforms that can be written as a sum of spatial basis functions can be fitted to $[x', y', z']$ using regularized least squares, one spatial coordinate at the time. Let $\boldsymbol{d}$ be the flattened column vector of x', y', or z' coordinates. Let $\boldsymbol{\phi}$ be a set of B spatial functions evaluated on the discrete image domain Ω, such that each row corresponds to a spatial location (flattened), and each column corresponds to a basis function. Finally, let $\boldsymbol{c}$ be

a column vector with B elements, corresponding to the coefficients of the basis functions. Then, the fitted transform is $\boldsymbol{\phi c}$ and its squared coordinate error is:

$$E_{coord} = [\boldsymbol{d} - \boldsymbol{\phi c}]^t [\boldsymbol{d} - \boldsymbol{\phi c}] = \boldsymbol{d}^t \boldsymbol{d} + \boldsymbol{c}^t \boldsymbol{\phi}^t \boldsymbol{\phi c} - 2 \boldsymbol{d}^t \boldsymbol{\phi c}.$$

This fit can be regularized with quadratic penalty terms while remaining closed form. Here we choose the membrane energy, which penalizes the squared norm of the gradient of the deformation field. Let $\boldsymbol{G}_x, \boldsymbol{G}_y, \boldsymbol{G}_z$ be the spatial gradients of the basis functions; as for $\boldsymbol{\phi}$, every row corresponds to a spatial location, and every column to a basis function. The regularizer is then given by:

$$E_{reg} = [\boldsymbol{G}_x \boldsymbol{c}]^t [\boldsymbol{G}_x \boldsymbol{c}] + [\boldsymbol{G}_y \boldsymbol{c}]^t [\boldsymbol{G}_y \boldsymbol{c}] + [\boldsymbol{G}_z \boldsymbol{c}]^t [\boldsymbol{G}_z \boldsymbol{c}] = \boldsymbol{c}^t [\boldsymbol{G}_x^t \boldsymbol{G}_x + \boldsymbol{G}_y^t \boldsymbol{G}_y + \boldsymbol{G}_z^t \boldsymbol{G}_z] \boldsymbol{c}.$$

E_{coord} and E_{reg} are combined into an objective function E using a relative weight λ. Setting the gradient to zero yields the regularized least square estimate:

$$\nabla E = \nabla (E_{coord} + \lambda E_{reg}) = 2 \boldsymbol{\phi}^t \boldsymbol{\phi c} - 2 \boldsymbol{\phi}^t \boldsymbol{d} + 2\lambda [\boldsymbol{G}_x^t \boldsymbol{G}_x + \boldsymbol{G}_y^t \boldsymbol{G}_y + \boldsymbol{G}_z^t \boldsymbol{G}_z] \boldsymbol{c} = 0,$$
$$\Rightarrow \boldsymbol{c} = [\boldsymbol{\phi}^t \boldsymbol{\phi} + \lambda (\boldsymbol{G}_x^t \boldsymbol{G}_x + \boldsymbol{G}_y^t \boldsymbol{G}_y + \boldsymbol{G}_z^t \boldsymbol{G}_z)]^{-1} \boldsymbol{\phi d}.$$

Within this family of transforms, we consider three in this article:

- **Affine**: a standard affine transform, fitted with all available voxels. We note that, in this case, the basis function matrix is simply $\boldsymbol{\phi} = [\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \mathbb{1}]$ (where $\mathbb{1}$ is the all-ones vector), and the gradients of the basis functions are ignored.
- **Affine-RANSAC**: an affine transform fitted robustly with RANSAC.
- **Bsplines** [34]: with control point spacing specified by the user.

Other basis previously used in registration that are straightforward to implement include lower-order polynomials [38] and the discrete cosine transform [4].

Demons-like model: The demons algorithm [35] computes a nonlinear registration by alternating between: *(i)* estimating force vectors via an optical-flow-like algorithm [19]; *(ii)* smoothing these force vectors with a Gaussian kernel; and *(iii)* applying the deformation before recomputing the force vectors. Here, we propose a similar algorithm with the difference that, since the output of the CNN is constant, we only take one step, i.e., we just filter the output of the CNN with a Gaussian kernel. The standard deviation is specified by the user.

Log-polyaffine model: Diffeomorphic models that can be analytically inverted are desirable in registration. Here we explore a log-polyaffine model [2] that is fitted to the output of the CNN as follows. First, we subdivide the image domain into cubic supervoxels of fixed, user-defined width W; this parameter controls the flexibility of the transform. Next, we compute an affine transform for every supervoxel s, assuming that it contains a minimum number of voxels. If we denote this affine transform by $\boldsymbol{T}_s$ in homogeneous coordinates, then [2]:

$$\log(\boldsymbol{T}_s) = \begin{pmatrix} \boldsymbol{L}_s & \boldsymbol{v}_s \\ 0 & 0 \end{pmatrix},$$

and it can be shown that the affine transform can be represented by a stationary velocity field (SVF) given by: $\boldsymbol{\Psi}_s(x, y, z) = \boldsymbol{v}_s + \boldsymbol{L}_s[x, y, z]^t$. The log-polyaffine framework computes a global SVF $\boldsymbol{\Psi}$ as a weighted sum of SVFs:

$$\boldsymbol{\Psi}(x, y, z) = \sum_s w_s(x, y, z)(\boldsymbol{v}_s + \boldsymbol{L}_s[x, y, z]^t),$$

where $w_s(x, y, z)$ are normalized non-negative weights obtained with a Gaussian function of the distance between (x, y, z) and the center of supervoxel s; we set the standard deviation of this Gaussian to $W/2$. The final SVF $\boldsymbol{\Psi}$ can be integrated with the scale-and-square algorithm to obtain the final deformation field [3]; the inverse field can be obtained by integrating the negated SVF $(-\boldsymbol{\Psi})$.
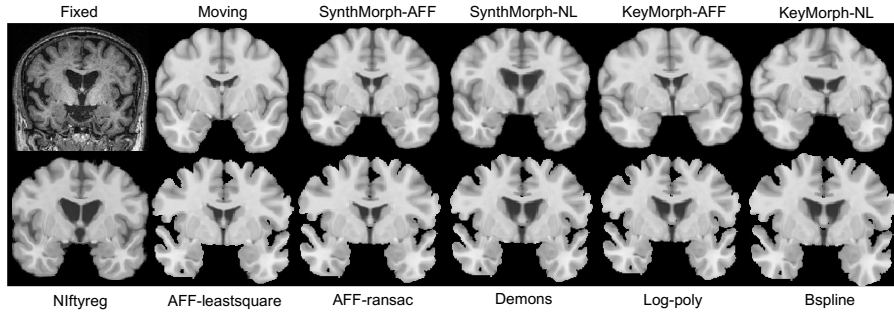
2.4 Implementation details

The CNN is a standard U-net [33], with a design inspired by nnU-net [23]. It has four resolution levels with two convolutional layers (comprising $3 \times 3 \times 3$ convolutions and a ReLu) followed by $2 \times 2 \times 2$ max pooling (in the encoder) or upconvolution (decoder). The final activation layer is linear, to regress the atlas coordinates in decimeters (which roughly normalizes them from -1 to 1). In addition to predicting atlas coordinates, the CNN also learns to predict a brain mask, which is used to define the domain Ω at test time. For this, we add a segmentation loss to Equation 1 with relative weight 0.5; the segmentation loss itself combines a Dice loss (weight 0.75) and cross-entropy loss (weight 0.25). The model is trained with stochastic gradient descent with learning rate 0.01, weight decay 3e−5, and momentum 0.99 for 100 epochs, setting aside 20% of the data for validation (selecting the best model). Input images are normalized with their median value inside the brain mask. At test time, nonlinear models are fitted on the residual of the Affine-RANSAC fit, clipped at 10mm to minimize the impact of the worst outliers. The RANSAC algorithm uses 50,000 randomly selected voxels, a maximum of 100 iterations, tests 500 voxels in every attempt, and requires 20% of voxels to be inliers to consider any given solution. The minimum number of voxels per supervoxel in the log-polyaffine model is 100.

3 Experiments and Results

The training data consists of high-resolution, isotropic, T1-weighted scans of 897 subjects from the HCP dataset [36] and 1148 subjects from the ADNI [24]. The test data consists of high-resolution, isotropic, T1 of the first 100 subjects from both the ABIDE [11] and OASIS3 [27] datasets, for a total of 200 test subjects. All scans were segmented into 36 regions using FreeSurfer [12]. Details on the acquisitions can be found in the corresponding publications.

3.1 Data preparation for RbR training

The scans in the training dataset were masked and registered to the ICBM 2009b Nonlinear Symmetric MNI template using NiftyReg [29]. Specifically, we first

| Fixed | Moving | SynthMorph-AFF | SynthMorph-NL | KeyMorph-AFF | KeyMorph-NL |

| NIftyreg | AFF-leastsquare | AFF-ransac | Demons | Log-poly | Bspline |

Fig. 2. Coronal slice of sample fixed image and corresponding registered MNI slice, using the different approaches.

ran the block matching algorithm for affine registration (`reg_aladin`) with the `-noSym` option, and subsequently used the fast free-form deformation algorithm (`f3d`) to compute the nonlinear registration. `f3d` was run in diffeomorphic mode (`-vel`) and local normalized correlation coefficient ($\sigma = 5$) as similarity metric. The processing time for the whole dataset was less than 24h on a 64-core desktop.

3.2   Experimental setup

We quantify registration accuracy with average Dice scores between the segmentations of the scans and the segmentation of the atlas, deformed with the estimated registrations. We further evaluate the regularity of the deformations using the membrane energy. Using these metrics, we compared RbR with:

NiftyReg: We use the same parameters as for the training data, Since we used it as ground truth in training, NiftyReg provides a ceiling for the performance that RbR can achieve.

KeyMorph [37]: Our main competing method is KeyMorph. We test KeyMorph variants as trained by the original authors (2024-03-06 version). For affine registration, we use 128 keypoints with weights optimized for affine Dice overlap. To estimate non-linear transforms, we use the TPS model with 512 keypoints, Dice-specific weights, and $\lambda_{\text{TPS}} = 0$ (which yields the highest Dice in their article). We intensity normalize and skull-strip scans as in the original publication.

SynthMorph [14]: for completeness, we also consider a non-interpretable state-of-the-art algorithm trained to maximize segmentation overlap (Dice scores) with synthetic images. SynthMorph has an affine model based on keypoints [15] and a nonlinear model based on regression [16]. We finetuned SynthMorph with the MNI atlas permanently selected as moving image, and evaluated its affine and nonlinear modules separately. We used its default trade-off parameter $\lambda = 0.5$.
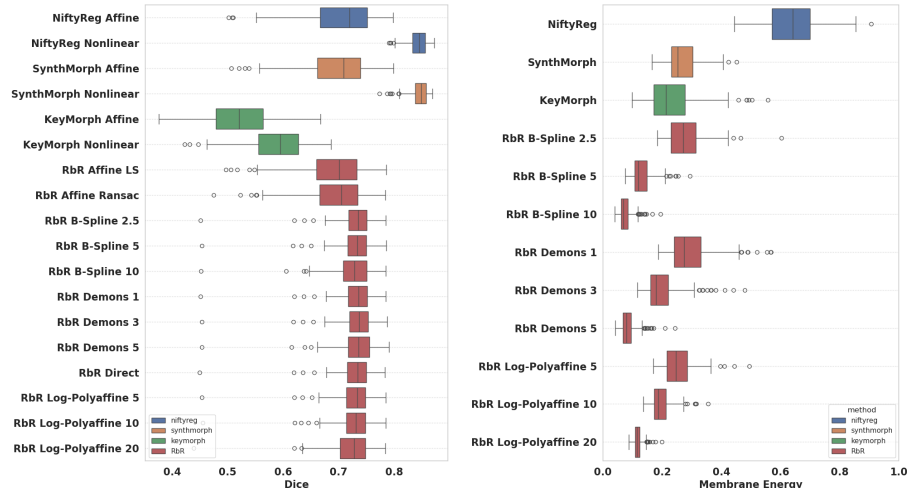
Fig. 3. Left: Average Dice between segmentations of scans and registered MNI. For RbR, the numbers indicate control point spacing (Bsplines), Gaussian σ (demons), or supervoxel width (log-polyaffine). Right: membrane energy for nonlinear models.

3.3 Results

Figure 2 shows qualitative results for the different methods, while Figure 3 shows box plots for the average Dice and membrane energy. KeyMorph with TPS yields an average Dice of 0.60, which is 7 points lower than the value reported in the original publication. This could be a domain gap issue – including both the test datasets and the MNI atlas, which is blurrier than a regular scan. Errors are apparent in Figure 2, where KeyMorph fails to correct the rotation about the A-P axis in linear mode and introduces wrong deformations in nonlinear mode.

Compared with KeyMorph, RbR does not require skull-stripping and yields better Dice, with medians of $\sim$0.70 (affine) and $\sim$0.73 (nonlinear) – higher than Keymorph's $\sim$0.60 on our dataset or even the value reported in their publication ($\sim$0.67). While the improvement from linear to nonlinear models in RbR may not be large in terms of median, it is very noticeable in terms of the first quartile and outliers. In Figure 2, the difference between the nonlinear versions is subtle, but noticeable e.g., around the third ventricle (less aggressively "closed" by the Bspline model). The ability to more accurately follow contours is paid in terms of membrane energy, which is noticeably higher for the more flexible models.

Finally, we note that, while RbR-affine and Synthmorph-affine (both based on keypoints) achieve the same Dice as NiftyReg in the affine model, keypoint methods still trail non-interpretable approaches (classical and learning-based) in nonlinear registration: NiftyReg and SynthMorph achieve Dice scores 10 points higher than RbR and $\sim$20 points higher than KeyMorph in this setting – and SynthMorph does so with similar levels of membrane energy.

3.4 Discussion and conclusion

RbR offers a new perspective on interpretable keypoint-based registration, posing it as a coordinate regression problem. Allowing every voxel to inform the model fit enables RbR to outperform standard keypoint approaches and fit less parsimonious nonlinear models. RbR also has disadvantages, particularly the need to retrain the model for each new atlas. Future work will include fitting other models, adding topological losses in training (e.g., penalizing negative Jacobians), investigating RbR's value as a feature extractor (i.e., as pretraining for other tasks), and exploring other improvements that help close the gap with non-interpretable approaches. We believe that RbR will be particularly valuable in scenarios requiring robust and interpretable registration, such as fetal imaging.

Acknowledgments

References

1. Alam, F., Rahman, S.U., Ullah, S., Gulati, K.: Medical image registration in image guided surgery: Issues, challenges and research opportunities. Biocybern Biomed Eng **38**(1) (2018)
2. Arsigny, V., Commowick, O., Ayache, N., Pennec, X.: A fast and log-euclidean polyaffine framework for locally linear registration. J Math Imaging Vis **33** (2009)
3. Arsigny, V., Commowick, O., Pennec, X., Ayache, N.: A log-euclidean framework for statistics on diffeomorphisms. In: MICCAI 2006, Proceedings, Part I 9. Springer (2006)
4. Ashburner, J., Friston, K.J.: Nonlinear spatial normalization using basis functions. Hum Brain Mapp **7**(4) (1999)
5. Ashburner, J., Friston, K.J.: Voxel-based morphometry—the methods. Neuroimage **11**(6) (2000)
6. 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. Med Image Anal **12**(1) (2008)
7. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: VoxelMorph: a learning framework for deformable medical image registration. IEEE T Med Imaging **38**(8) (2019)
8. Chen, J., Frey, E.C., He, Y., Segars, W.P., Li, Y., Du, Y.: Transmorph: Transformer for unsupervised medical image registration. Med Image Anal **82** (2022)
9. Christensen, G.E., Johnson, H.J.: Consistent image registration. IEEE T Med Imaging **20**(7) (2001)
10. De Vos, B.D., Berendsen, F.F., Viergever, M.A., Sokooti, H., Staring, M., Išgum, I.: A deep learning framework for unsupervised affine and deformable image registration. Med Image Anal **52** (2019)

11. Di Martino, A., et al.: The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Mol Psychiatry **19**(6) (2014)

12. Fischl, B., Salat, D.H., et al.: Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron **33**(3) (2002)

13. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Comm of the ACM **24**(6) (1981)

14. Hoffmann, M., Billot, B., Greve, D.N., Iglesias, J.E., Fischl, B., Dalca, A.V.: SynthMorph: learning contrast-invariant registration without acquired images. IEEE T Med Imaging **41**(3) (2021)

15. Hoffmann, M., Hoopes, A., Fischl, B., Dalca, A.V.: Anatomy-specific acquisition-agnostic affine registration learned from fictitious images. In: Medical Imaging 2023: Image Processing. vol. 12464. SPIE (2023)

16. Hoffmann, M., Hoopes, A., Greve, D.N., Fischl, B., Dalca, A.V.: Anatomy-aware and acquisition-agnostic joint registration with synthmorph. arXiv preprint arXiv:2301.11329 (2023)

17. Holland, D., Dale, A.M., Initiative, A.D.N., et al.: Nonlinear registration of longitudinal images and measurement of change in regions of interest. Med Image Anal **15**(4) (2011)

18. Hoopes, A., Hoffmann, M., et al.: Learning the effect of registration hyperparameters with hypermorph. Machine Learning for Biomedical Imaging **1** (2022)

19. Horn, B.K., Schunck, B.G.: Determining optical flow. Artif Intell **17**(1-3) (1981)

20. Hua, X., Lee, S., Yanovsky, I., et al.: Optimizing power to track brain degeneration in alzheimer's disease and mild cognitive impairment with tensor-based morphometry: an adni study of 515 subjects. Neuroimage **48**(4) (2009)

21. Iglesias, J.E.: A ready-to-use machine learning tool for symmetric multi-modality registration of brain MRI. Scientific Reports **13**(1) (2023)

22. Iglesias, J.E., Sabuncu, M.R.: Multi-atlas segmentation of biomedical images: a survey. Med Image Anal **24**(1) (2015)

23. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods **18**(2) (2021)

24. Jack Jr, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L. Whitwell, J., Ward, C., et al.: The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J. Magn Reson Imaging **27**(4) (2008)

25. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage **23** (2004)

26. Krebs, J., Delingette, H., Mailhé, B., Ayache, N., Mansi, T.: Learning a probabilistic model for diffeomorphic registration. IEEE T Med Imaging **38**(9) (2019)

27. LaMontagne, P.J., et al.: Oasis-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease. MedRxiv (2019)

28. Lv, J., Wang, Z., et al.: Joint progressive and coarse-to-fine registration of brain MRI via deformation field integration and non-rigid feature fusion. IEEE T Med Imaging **41**(10) (2022)

29. 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. Comput Methods Programs Biomed **98**(3) (2010)

30. Mok, T.C., Chung, A.: Affine medical image registration with coarse-to-fine vision transformer. In: Proceedings of the IEEE/CVF on CVPR (2022)

31. Nocedal, J., Wright, S.J.: Numerical optimization. Springer (1999)
32. Pluim, J.P., Maintz, J.A., Viergever, M.A.: Mutual-information-based registration of medical images: a survey. IEEE T Med Imaging **22**(8) (2003)
33. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI 2015, Proceedings, Part III 18. Springer (2015)
34. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J.: Nonrigid registration using free-form deformations: application to breast MR images. IEEE T Med Imaging **18**(8) (1999)
35. Thirion, J.P.: Image matching as a diffusion process: an analogy with Maxwell's demons. Med Image Anal **2**(3) (1998)
36. Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., Consortium, W.M.H., et al.: The WU-minn human connectome project: an overview. Neuroimage **80** (2013)
37. Wang, A.Q., Evan, M.Y., Dalca, A.V., Sabuncu, M.R.: A robust and interpretable deep learning framework for multi-modal registration via keypoints. Med Image Anal **90** (2023)
38. Woods, R.P., Grafton, S.T., Watson, J.D., Sicotte, N.L., Mazziotta, J.C.: Automated image registration: II. intersubject validation of linear and nonlinear models. J Comput Assist Tomogr **22**(1) (1998)
39. Yang, X., Kwitt, R., Styner, M., Niethammer, M.: Quicksilver: Fast predictive image registration – a deep learning approach. NeuroImage **158** (2017)