

PULPo: Probabilistic Unsupervised Laplacian Pyramid Registration

Leonard Siegert¹, Paul Fischer^{1,2}, Mattias P. Heinrich³, and Christian F. Baumgartner^{1,2}

¹ Cluster of Excellence – ML for Science, University of Tübingen, Germany

² Faculty of Health Sciences and Medicine, University of Lucerne, Switzerland

³ Medical Informatics, University of Lübeck, Germany

Abstract. Deformable image registration is fundamental to many medical imaging applications. Registration is an inherently ambiguous task often admitting many viable solutions. While neural network-based registration techniques enable fast and accurate registration, the majority of existing approaches are not able to estimate uncertainty. Here, we present PULPo, a method for probabilistic deformable registration capable of uncertainty quantification. PULPo probabilistically models the distribution of deformation fields on different hierarchical levels combining them using Laplacian pyramids. This allows our method to model global as well as local aspects of the deformation field. We evaluate our method on two widely used neuroimaging datasets and find that it achieves high registration performance as well as substantially better calibrated uncertainty quantification compared to the current state-of-the-art¹.

1 Introduction

Deformable registration aims to establish a non-linear mapping between pairs of unaligned images. It is a key component in numerous clinical tasks, including the alignment of pre- and postoperative scans, as well as the monitoring of disease progression over time. Traditionally, registration methods addressed this problem through optimization, whereby an energy function is minimised for each image pair [26]. In order to improve registration performance, the displacement field was often optimized on several hierarchically organised resolution levels allowing to capture global as well as local effects [17]. Substantial research was dedicated to diffeomorphic registration which ensures invertability and smoothness of the deformation fields [1,6].

While traditional methods are still widely used, they are very time consuming, often requiring many minutes or hours to align a 3D image pair. Recently, neural networks have emerged as a powerful alternative to traditional approaches. They allow for exceptionally fast deformation prediction with high accuracy [4,21]. Expanding upon the concept of hierarchies in traditional registration methods, a number of works introduced neural network-based diffeomorphic registration approaches which progressively estimate and combine the

¹ The code is available at <https://github.com/leonardsiegert/PULPo>.

deformation fields at different resolution levels [22,20]. These approaches have been shown to lead to performance improvements over non-hierarchical methods.

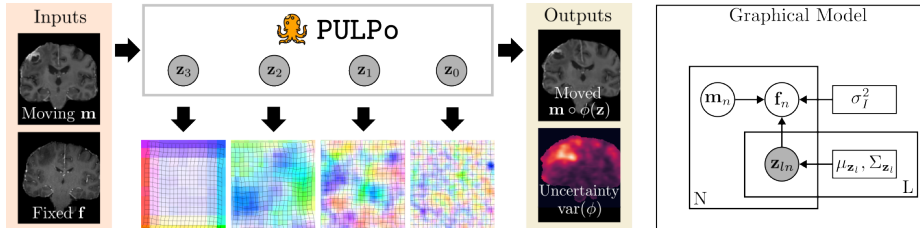


Fig. 1. Overview: (Left) Our probabilistic unsupervised Laplacian pyramid registration approach (PULPo) hierarchically models the deformation field at different resolutions allowing for accurate registration and well-calibrated uncertainty quantification particularly in the presence of lesions. (Right) Graphical model for our method.

An important but neglected area in registration is uncertainty quantification. The majority of existing approaches only output the most likely deformation. However, there may exist many viable alternative deformations that also lead to a good alignment [18]. The resulting aleatoric uncertainty in the deformation field is of high clinical significance for many applications such as disease monitoring or neurosurgery [25]. This problem is exacerbated when images contain unmatchable areas such as a preoperative scan containing a tumor and a follow-up scan where the tumor has been resected.

Early work on uncertainty quantification in traditional registration used probabilistic techniques to model the posterior of the deformation field given the input images, and used Markov-Chain Monte Carlo sampling to draw samples from this distribution [25,24,23,12].

Uncertainty quantification in neural network-based approaches is currently understudied. Xu et al. [30] apply the popular Monte Carlo (MC) dropout approximation to get an estimate of epistemic uncertainty resulting from uncertainty in the model parameters. Recently, Chen et al. [8] applied the same technique to a transformer-based registration architecture. Dalca et al. [9,10] proposed the probabilistic VoxelMorph which is an extension of their previously proposed approach [4] by a variational inference of the deformation velocity field using a procedure inspired by conditional VAEs (cVAEs) [28]. Smolders et al. [27] extended this approach for use in adaptive proton therapy. Since VoxelMorph uses a latent space matches the resolution of the input images, it may predispose the model to capture the distribution of local deformations at the expense of capturing the distribution of global deformations. Conversely, Krebs et al. [15] proposed a related method based on cVAEs, in which the latent space is in a low-resolution bottleneck of the architecture. While this approach allows to capture the distribution of local deformations, it may be restricted in capturing the

distribution of local deformations. We note that Krebs et al. did not explore the uncertainty estimation aspect in their work.

Here, we propose a novel probabilistic unsupervised registration technique that can capture local as well as global deformations and their variability. The method uses a *hierarchical* variational inference approach to probabilistically model the deformation field at increasingly fine resolution levels. For each level we estimate stationary velocity fields (SVFs) [1], which we then integrate to obtain diffeomorphic deformations. The deformation fields from the different resolution levels are then combined by leveraging the principle of Laplacian pyramids.

Of the the neural network-based registration approaches capable of uncertainty estimation the probabilistic VoxelMorph [9,10] is arguably the most widely used due to its strong registration performance and the availability of easy-to-use, well-documented code. However, Grzech et al. [12] observed that the probabilistic VoxelMorph produces uncertainty estimates with very small magnitudes, putting their utility into question. We confirm these findings in our own experiments in Sec. 3, and demonstrate that our probabilistic unsupervised Laplacian pyramid registration (PULPo) approach substantially outperforms the probabilistic VoxelMorph approach in terms of uncertainty quantification on the OASIS-1 and BraTS-reg datasets, while still producing well-aligned registrations.

2 Method

We model pairwise registration as a probabilistic generative process in which the fixed image $\mathbf{f}$ is generated by transforming the moving image $\mathbf{m}$ with the deformation field $\phi(\mathbf{z})$ that depends directly on a set of latent variables $\mathbf{z} = \{\mathbf{z}_l | l \in \{0, \dots, L-1\}\}$ as shown in Fig. 1. The latent variables $\mathbf{z}_l$ are arranged hierarchically, with the resolution of each level l being half that of the level above $(l-1)$. This allows them to capture local as well as global effects in the resulting deformation field.

Following prior work [10,15], we define the likelihood of $\mathbf{f}$ as

$$p(\mathbf{f}|\mathbf{z}, \mathbf{m}) = \mathcal{N}(\mathbf{f}; \mathbf{m} \circ \phi(\mathbf{z}), \sigma_I^2 I) , \quad (1)$$

where $\circ$ is the transformation of the moving image $\mathbf{m}$ with the deformation field ϕ , and σ_I^2 denotes additive image noise. Our goal is to approximate the posterior distribution $p(\mathbf{z}|\mathbf{f}, \mathbf{m})$. As Maximum A Posteriori (MAP) estimation of the true posterior $p(\mathbf{z}|\mathbf{m}, \mathbf{f})$ is intractable, we use a variational approximation q of the true posterior distribution, optimizing a lower bound to the true posterior [14].

We define the prior $p(\mathbf{z})$ and variational posterior $q(\mathbf{z}|\mathbf{m}, \mathbf{f})$ of our generative model as products of Gaussians with diagonal covariance matrices

$$\begin{aligned} p(\mathbf{z}) &= \prod_{l=0}^{L-1} p(\mathbf{z}_l) = \prod_{l=0}^{L-1} \mathcal{N}(\mathbf{z}_l; \mathbf{0}, I) \\ q(\mathbf{z}|\mathbf{m}, \mathbf{f}) &= \prod_{l=0}^{L-2} q(\mathbf{z}_l|\mathbf{z}_{l+1}, \mathbf{m}, \mathbf{f}) = \prod_{l=0}^{L-2} \mathcal{N}(\mathbf{z}_l; \mu_{\mathbf{z}_l|\mathbf{z}_{l+1}, \mathbf{m}, \mathbf{f}}, \Sigma_{\mathbf{z}_l|\mathbf{z}_{l+1}, \mathbf{m}, \mathbf{f}}) , \end{aligned} \quad (2)$$

with $q(\mathbf{z}_{L-1}|\mathbf{m}, \mathbf{f}) = \mathcal{N}(\mathbf{z}_{L-1}; \mu_{\mathbf{z}_{L-1}|\mathbf{m}, \mathbf{f}}, \Sigma_{\mathbf{z}_{L-1}|\mathbf{m}, \mathbf{f}})$.

To optimize the variational lower bound, we minimize the KL-divergence between the approximate posterior $q(\mathbf{z}|\mathbf{m}, \mathbf{f})$ and the true posterior $p(\mathbf{z}|\mathbf{m}, \mathbf{f})$ leading to the following evidence lower bound (ELBO):

$$\begin{aligned} \min \text{KL}[p(\mathbf{z}|\mathbf{m}, \mathbf{f})||p(\mathbf{z}|\mathbf{m}, \mathbf{f})] &= \min \mathbf{E}_q[\log q(\mathbf{z}|\mathbf{m}, \mathbf{f}) - \log p(\mathbf{z}|\mathbf{m}, \mathbf{f})] \\ &= \min \sum_{l=0}^{L-1} \text{KL}[q(\mathbf{z}_l|\mathbf{z}_{l+1}, \mathbf{m}, \mathbf{f})|p(\mathbf{z}_l)] - \mathbf{E}_q[\log p(\mathbf{f}|\mathbf{z}, \mathbf{m})] . \end{aligned} \quad (3)$$

A detailed derivation can be found in the supplemental materials.

We represent the deformation field using the SVF model [1] where we first estimate a velocity field $\mathbf{v}_l$ for each level, and then integrate it with scaling and squaring in 7 steps to yield an approximately diffeomorphic deformation field ϕ_l .

2.1 Neural network implementation

We use neural networks to approximate $q(\mathbf{z}|\mathbf{m}, \mathbf{f})$ and the non-linear transformation $\phi(\mathbf{z})$. Following related work on hierarchical conditional VAEs [5,11] the model is organized in K total levels and $L \leq K$ latent levels. A schematic of the architecture is shown in Fig. 2. We used $K = 5$ and $L = 4$ for all experiments. The encoder predicts the parameters $\mu_{\mathbf{z}_l}$ and $\sigma_{\mathbf{z}_l}$ of each $\mathbf{z}_l$ based on the moving ($\mathbf{m}$) and fixed ($\mathbf{f}$) input images. A sampling layer samples $\mathbf{z}_l$ according to the predicted distribution. Those samples are then transformed into velocity fields $\mathbf{v}_l$ for each level. The velocity fields are processed by a vector integration layer to obtain deformation fields ϕ_l for each level, which are then transformed to the moved image $\mathbf{f}_l$ using a spatial transformer layer. The outputs ϕ_0 and $\mathbf{f}_0$ are the final deformation field and moved image and have the same resolution as the input images. The outputs $\mathbf{f}_l$ of the lower levels are used for deep supervision. Each level also informs the level above with a feedback connection containing a concatenation of $\mathbf{z}_l$, $\mathbf{v}_l$, ϕ_l and $\hat{\mathbf{f}}_l$.

Rather than independently estimating the velocity field $\mathbf{v}_l$ for each resolution level, the velocity fields are incrementally added to the previously estimated velocity fields of the lower levels. This emulates the mechanics of a Laplacian pyramid [7], and allows a gradual refinement of the final deformation akin to traditional hierarchical registration approaches. The vector integration on each level ensures that the resulting deformation fields are approximately diffeomorphic.

2.2 Training

Since the loss terms in Eq. 3 contain a sum over all voxels, higher levels are weighted higher in the resulting loss function since they contain more voxels. To account for this difference, we scale the losses on each level with $w_l = 2^{3l}$.

We substitute the MSE loss resulting from the log-likelihood term in Eq. 1 with the normalized cross-correlation (NCC) [2]) with window size $k_l = 1 + 2(L - l)$ and a balancing factor $\gamma = 0.05$. In preliminary experiments we observed a

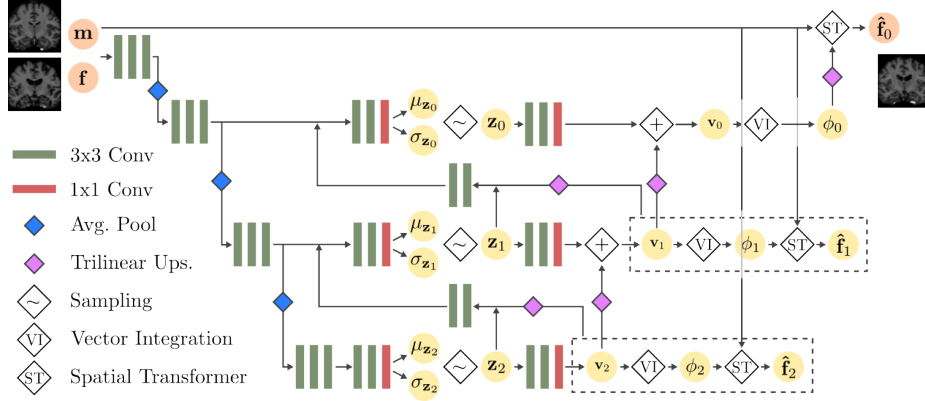


Fig. 2. Architecture: Schematic network architecture with 3 latent levels. In practice, we used 4 latent levels.

substantially better performance with the NCC-loss and used it for all experiments. We also use an auxiliary deep similarity loss to incentivize the method to use the lower levels. For this, we downscale the moving image $\mathbf{m}$ to the same resolution as the samples $\mathbf{z}_l$ and deformation fields $\phi(\mathbf{z}_l)$, and computed the NCC-loss with the level predictions $\hat{\mathbf{f}}_l = \mathbf{m}_l \circ \phi(\mathbf{z}_l)$. Lastly, to ensure smooth deformation fields, we use a diffusion regularizer on the spatial gradients of the deformation field ϕ_l on each level. The total training objective for our model can be written as

$$\mathcal{L} = \beta \sum_{l=0}^{L-1} w_l \text{KL}[q(\mathbf{z}_l | \mathbf{z}_{l+1}, \mathbf{m}, \mathbf{f}) | p(\mathbf{z}_l)] - \frac{\gamma}{\sigma_I^2} \sum_{l=0}^{L-1} w_l \text{NCC}(\mathbf{m}_l, \mathbf{f}_l, k_l) + \lambda \sum_{l=0}^{L-1} w_l \sum_v^{V_l} \|\nabla \phi_v\|^2. \quad (4)$$

We used $\beta = 0.1$ and $\lambda = 0.025$ for our experiments. We set $\sigma_I^2 = 0.25$ on the highest level and $\sigma_I^2 = 1$ on all others.

2.3 Inference

After training, stochastic samples of the deformation field ϕ can be obtained with repeated forward passes through our model. This allows us to calculate the variation of the deformation field $\text{var}(\phi(\mathbf{z}))$. Transforming the moving image $\mathbf{m}$ with the samples further allows us to obtain an uncertainty estimate on a voxel level, i.e. $\text{var}(\mathbf{m} \circ \phi(\mathbf{z}))$.

While the prediction could be obtained by averaging multiple samples, we propose an accelerated strategy: We perform a single forward pass, but instead of sampling $\mathbf{z}_l$, we directly propagate the predicted mean $\mu_{\mathbf{z}_l}$. Since the variational

posterior q is Gaussian when fixing the conditions, this procedure can be shown to approximate the MAP of the true posterior.

3 Experiments & Results

We performed experiments on two 3D brain datasets and evaluated registration as well as uncertainty quantification performance.

3.1 Data

OASIS-1 We used the OASIS-1 3D brain MRI dataset [13,19], containing both subjects with Alzheimer’s disease and healthy controls, aged 18 to 96. We separated the subjects into a 354/30/30 train/validation/test split. The images were skull-stripped, bias-corrected and affinely registered. The resolution was isotropic 1mm resulting in volume sizes of $160 \times 224 \times 192$. The dataset also contains segmentation maps with 35 anatomical labels, which we used to calculate the DSC metric. We used the anatomical landmarks from Taha et al. [29]. We used only one scan per subject and employed a pairwise inter-subject training scheme.

BraTS-Reg We also used the BraTS-Reg 2022 dataset [3], consisting of pre-operative and follow-up brain MRI scans of the same patient diagnosed with glioma. We used only the contrast-enhanced T1-weighted (T1CE) MR images, as Meng et al. [20] found T1CE-sequences to yield better performance than T1, T2, FLAIR, and multi-channel inputs. The data was split into training/validation/test following a 120/20/20 ratio. Landmarks in baseline and follow-up scans were manually annotated by clinical experts. After preprocessing, the volumes had an isotropic voxel size of 1mm and dimensions $144 \times 192 \times 160$. We used a pairwise intra-subject training scheme, aligning the follow-up to the pre-operative scan.

3.2 Metrics

We evaluated alignment in terms of the image intensity-based Root Mean Squared Error (RMSE) and the Target Registration Error (TRE) of the anatomical landmarks. For OASIS, we also evaluated the soft Dice Similarity Coefficient (DSC) of the anatomical segmentation maps. We further evaluated the percentage of negative voxels of the Jacobian determinant (JD) of the deformation fields $\%|J_\phi| \leq 0$ as a measure of smoothness and invertibility.

To evaluate the uncertainty estimation, we computed the NCC between the variance of the output and mean squared error over $N = 20$ samples [11,5]. This provides a measure for the calibration, i.e. how well the method allocates uncertainty where it makes mistakes [16]. We report the NCC for the output image intensities (NCC_{VX}) and the anatomical landmarks (NCC_{LM}).

3.3 Experiments

We trained our proposed PULPo method as described in Sec. 2.2.

We used the probabilistic diffeomorphic version of VoxelMorph (DIF-VM) [10] as baseline method. Building on the authors’ publicly available code, we implemented the probabilistic extension as described by Dalca et al. [10]. We used the same hyperparameters that Dalca et al. used in their experiments. We trained DIF-VM using our NCC loss instead of the MSE as originally proposed, since we found in preliminary results that the NCC version was superior across all metrics.

In order to better understand the effect of the probabilistic hierarchical structure of our method, we performed a non-hierarchical ablation of our method, PULPo_{NH}. While the base architecture was the same as that of the full PULPo, probabilistic sampling was only enabled on the highest level ($l = 0$), and lower levels didn’t contribute to the final deformation field. The losses on all other levels were set to zero, and $\mu_{\mathbf{z}_l}$ was propagated deterministically directly instead of sampling a $\mathbf{z}_l$.

3.4 Evaluation of Registration Performance

Table 1. Results: Registration performance and uncertainty quantification metrics for all methods.

Dataset	Method	RMSE	DSC	$\% J_\phi \leq 0$	TRE	NCC _{VX}	NCC _{LM}
OASIS-1	DIF-VM	0.039	0.804	0.002	2.943	0.210 ± 0.015	-0.171 ± 0.201
	PULPo _{NH}	0.036	0.770	0.084	3.202	0.510 ± 0.051	0.092 ± 0.247
	PULPo	0.036	0.777	0.068	4.097	0.533 ± 0.039	0.302 ± 0.110
BraTS-Reg	DIF-VM	0.044	-	$< 10^{-6}$	2.321	0.264 ± 0.043	-0.016 ± 0.366
	PULPo _{NH}	0.055	-	$< 10^{-6}$	3.030	0.474 ± 0.089	0.231 ± 0.352
	PULPo	0.044	-	0.001	2.434	0.497 ± 0.073	0.229 ± 0.324

We evaluated the registration performance in terms of the metrics described in Sec. 3.2. The results in Tab. 1 show that DIF-VM resulted in better alignment in terms of DSC and TRE, and smoother DFs in terms of $\%|J_\phi| \leq 0$, however PULPo achieved better RMSE. Visual inspection of qualitative results (see Fig. 3) suggested that both methods achieved good alignment with smooth deformations. Additional qualitative results can be found in the appendix.

The non-hierarchical ablation PULPo_{NH} resulted in better alignment than PULPo in terms of the TRE, but in worse alignment in terms of the DSC on OASIS-1. On BraTS-Reg, PULPo_{NH} resulted in significantly worse alignment than PULPo.

3.5 Evaluation of uncertainty quantification

The uncertainty calibration metrics NCC_{VX} and NCC_{LM} are shown in Tab. 1 on the right. PULPo showed a substantially better calibration than DIF-VM, both

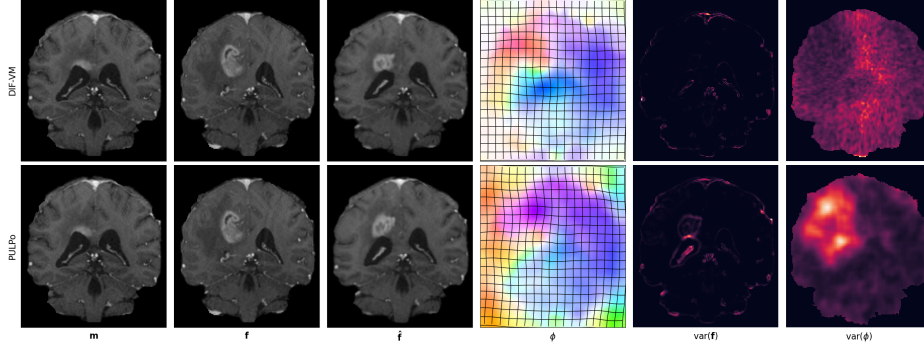


Fig. 3. Registration Results: The moving image $\mathbf{m}$, fixed image $\mathbf{f}$, prediction $\hat{\mathbf{f}}$, predicted DF ϕ , voxel variance $\text{var}(\mathbf{f})$ and DF variance $\text{var}(\phi)$. Value range of $\text{var}(\mathbf{f})$ for DIF-VM/PULPo: $[0, 3.98\text{e-}6]/[0, 0.042]$. Value range of $\text{var}(\phi)$ for DIF-VM/PULPo: $[0, 3.37\text{e-}5]/[0, 1.407]$. Shown: Coronal slice 76 of T1CE-scan 5 on BraTS-Reg.

in terms of image intensities (NCC_{VX}) and anatomical landmarks (NCC_{LM}) on both datasets. The negative NCC_{LM} even indicates a negative correlation between landmark MSE and landmark uncertainty for DIF-VM. PULPo also generally performed better than the non-hierarchical ablation PULPo_{NH} except on the BraTS-Reg landmarks, where the two methods performed similarly. This indicates that the probabilistic hierarchy allowed better calibrated uncertainty quantification.

We observed that DIF-VM produced samples with very little variance. This likely explains the poor performance on the uncertainty metrics. Upon closer inspection, we found that the $\sigma_{\mathbf{z}}$ of DIF-VM’s random latent variable converged to a very small number making the model almost deterministic in behaviour. Visual inspection of qualitative results in Fig. 3 confirmed that for PULPo the deformation field and image variance allocated large uncertainty around the tumor area where large uncertainties are expected. DIF-VM, on the other, hand produced a uniform DF variance map on which difficult to register areas cannot be discerned from others.

4 Discussion and Conclusion

Uncertainty quantification in medical image registration is of high importance for clinical applications such as disease monitoring or neurosurgery. However, the problem is severely understudied with only one prior work addressing this problem in the neural network setting. We have introduced a novel method for fast diffeomorphic 3D registration that can model global and local uncertainties by probabilistically modelling the deformation fields at different resolution levels, and combining them using a Laplace pyramid approach. We showed that our method produces accurate registrations as well as calibrated uncertainty quantification.

Our experiments revealed a trade-off between accurate uncertainty estimation and high registration performance. While the probabilistic diffeomorphic voxelmorph (DIF-VM) achieved better registration performance on some metrics, its near deterministic behaviour severely limited its performance on uncertainty quantification. Future work will focus on further improving our architecture to improve registration performance, while maintaining well-calibrated uncertainty estimates.

Acknowledgments. This work was supported by the Excellence Cluster 2064 “Machine Learning — New Perspectives for Science”, project number 390727645). The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting Paul Fischer.

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Ashburner, J.: A fast diffeomorphic image registration algorithm. *Neuroimage* **38**(1), 95–113 (2007)
2. Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C.: A reproducible evaluation of ants similarity metric performance in brain image registration. *Neuroimage* **54**(3), 2033–2044 (2011)
3. Baheti, B., Waldmannstetter, D., Chakrabarty, S., Akbari, H., Bilello, M., Wiestler, B., Schwarting, J., Calabrese, E., Rudie, J.D., Abidi, S.A.R., Mousa, M.S., Villanueva-Meyer, J.E., Marcus, D.S., Davatzikos, C., Sotiras, A., Menze, B.H., Bakas, S.: The brain tumor sequence registration challenge: Establishing correspondence between pre-operative and follow-up mri scans of diffuse glioma patients. *ArXiv abs/2112.06979* (2021), <https://api.semanticscholar.org/CorpusID:245131368>
4. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging* **38**(8), 1788–1800 (2019)
5. Baumgartner, C.F., Tezcan, K.C., Chaitanya, K., Hötter, A.M., Muehlmatter, U.J., Schawkat, K., Becker, A.S., Donati, O., Konukoglu, E.: Phiseg: Capturing uncertainty in medical image segmentation. In: *Medical Image Computing and Computer Assisted Intervention—MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II* 22. pp. 119–127. Springer (2019)
6. Beg, M.F., Miller, M.I., Trounev, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International journal of computer vision* **61**, 139–157 (2005)
7. Burt, P.J., Adelson, E.H.: The laplacian pyramid as a compact image code. In: *Readings in computer vision*, pp. 671–679. Elsevier (1987)
8. Chen, J., Frey, E.C., He, Y., Segars, W.P., Li, Y., Du, Y.: Transmorph: Transformer for unsupervised medical image registration. *Medical image analysis* **82**, 102615 (2022)
9. Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R.: Unsupervised learning for fast probabilistic diffeomorphic registration. In: *Medical Image Computing*

- and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16–20, 2018, Proceedings, Part I. pp. 729–738. Springer (2018)
10. Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R.: Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. *Medical image analysis* **57**, 226–236 (2019)
 11. Fischer, P., Küstner, T., Baumgartner, C.F.: Uncertainty estimation and propagation in accelerated mri reconstruction. *arXiv preprint arXiv:2308.02631* (2023)
 12. Grzech, D., Azampour, M.F., Qiu, H., Glocker, B., Kainz, B., Loïc, L.F.: Uncertainty quantification in non-rigid image registration via stochastic gradient markov chain monte carlo. In: *MELBA Special Issue: Uncertainty for Safe Utilization of Machine Learning in Medical Imaging (UNSURE) 2020*. vol. 12443, p. 3. MELBA (2021)
 13. Hoopes, A., Hoffmann, M., Fischl, B., Guttag, J., Dalca, A.V.: Hypermorph: Amortized hyperparameter learning for image registration. In: *Information Processing in Medical Imaging: 27th International Conference, IPMI 2021, Virtual Event, June 28–June 30, 2021, Proceedings 27*. pp. 3–17. Springer (2021)
 14. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013)
 15. Krebs, J., Delingette, H., Mailhé, B., Ayache, N., Mansi, T.: Learning a probabilistic model for diffeomorphic registration. *IEEE transactions on medical imaging* **38**(9), 2165–2176 (2019)
 16. Laves, M.H., Ihler, S., Fast, J.F., Kahrs, L.A., Ortmaier, T.: Recalibration of aleatoric and epistemic regression uncertainty in medical imaging. *arXiv preprint arXiv:2104.12376* (2021)
 17. Lester, H., Arridge, S.R.: A survey of hierarchical non-linear medical image registration. *Pattern recognition* **32**(1), 129–149 (1999)
 18. Luo, J., Sedghi, A., Popuri, K., Cobzas, D., Zhang, M., Preiswerk, F., Toews, M., Golby, A., Sugiyama, M., Wells, W.M., et al.: On the applicability of registration uncertainty. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22*. pp. 410–419. Springer (2019)
 19. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *Journal of cognitive neuroscience* **19**(9), 1498–1507 (2007)
 20. Meng, M., Bi, L., Feng, D., Kim, J.: Brain tumor sequence registration with non-iterative coarse-to-fine networks and dual deep supervision. *arXiv preprint arXiv:2211.07876* (2022)
 21. Meng, M., Bi, L., Feng, D., Kim, J.: Non-iterative coarse-to-fine registration based on single-pass deep cumulative learning. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 88–97. Springer (2022)
 22. Mok, T.C., Chung, A.C.: Large deformation diffeomorphic image registration with laplacian pyramid networks. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part III 23*. pp. 211–221. Springer (2020)
 23. Parisot, S., Wells III, W., Chemouny, S., Duffau, H., Paragios, N.: Concurrent tumor segmentation and registration with uncertainty-based sparse non-uniform graphs. *Medical image analysis* **18**(4), 647–659 (2014)

24. Risholm, P., Janoos, F., Norton, I., Golby, A.J., Wells III, W.M.: Bayesian characterization of uncertainty in intra-subject non-rigid registration. *Medical image analysis* **17**(5), 538–555 (2013)
25. Risholm, P., Pieper, S., Samset, E., Wells III, W.M.: Summarizing and visualizing uncertainty in non-rigid registration. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 554–561. Springer (2010)
26. 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)
27. Smolders, A., Lomax, T., Weber, D.C., Albertini, F.: Deformable image registration uncertainty quantification using deep learning for dose accumulation in adaptive proton therapy. In: *International Workshop on Biomedical Image Registration*. pp. 57–66. Springer (2022)
28. Sohn, K., Lee, H., Yan, X.: Learning structured output representation using deep conditional generative models. *Advances in neural information processing systems* **28** (2015)
29. Taha, A., Gilmore, G., Abbass, M., Kai, J., Kuehn, T., Demarco, J., Gupta, G., Zajner, C., Cao, D., Chevalier, R., et al.: Magnetic resonance imaging datasets with anatomical fiducials for quality control and registration. *Scientific Data* **10**(1), 449 (2023)
30. Xu, Z., Luo, J., Lu, D., Yan, J., Frisken, S., Jagadeesan, J., Wells III, W.M., Li, X., Zheng, Y., Tong, R.K.y.: Double-uncertainty guided spatial and temporal consistency regularization weighting for learning-based abdominal registration. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 14–24. Springer (2022)