Implicit Neural Representations for Joint Decomposition and Registration of Gene Expression Images in the Marmoset Brain

Michal Byra[1,2(✉)], Charissa Poon[1], Tomomi Shimogori[3], and Henrik Skibbe[1]

[1] Brain Image Analysis Unit, RIKEN Center for Brain Science, Wako, Japan
`michal.byra@riken.jp`
[2] Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland
[3] Laboratory for Molecular Mechanisms of Brain Development, RIKEN Center for Brain Science, Wako, Japan

Abstract. We propose a novel image registration method based on implicit neural representations that addresses the challenging problem of registering a pair of brain images with similar anatomical structures, but where one image contains additional features or artifacts that are not present in the other image. To demonstrate its effectiveness, we use 2D microscopy *in situ* hybridization gene expression images of the marmoset brain. Accurately quantifying gene expression requires image registration to a brain template, which is difficult due to the diversity of patterns causing variations in visible anatomical brain structures. Our approach uses implicit networks in combination with an image exclusion loss to jointly perform the registration and decompose the image into a support and residual image. The support image aligns well with the template, while the residual image captures individual image characteristics that diverge from the template. In experiments, our method provided excellent results and outperformed other registration techniques.

Keywords: brain · deep learning · gene expression · implicit neural representations · registration

1 Introduction

Image registration is a crucial prerequisite for image comparison, data integration, and group studies in contemporary medical and neuroscience research. In research and clinical settings, pairs of images often show similar anatomical structures but may contain additional features or artifacts, such as specific staining, electrodes, or lesions, that are not present in the other image. This difficulty of finding corresponding structures for automatically aligning images complicates image registration. In this work, we address the challenging problem

of the gene expression image registration in the marmoset brain. Brain atlases of gene expression, created using images of brain tissue processed through *in situ* hybridization (ISH), offer single-cell resolution of spatial gene expression patterns across the entire brain [2,7]. However, accurately quantifying gene expression requires brain image registration to spatially align ISH images to a common atlas space. The diversity of gene expression patterns in ISH images causes variations in visible anatomical brain structures with respect to the template image. ISH microscopy images are also susceptible to tissue processing artifacts, resulting in non-specific staining and tissue deformations.

Traditional pair-wise image registration methods use optimization algorithms to find the deformation field that maximizes the similarity between a pair of images. While several deep learning methods based on convolutional neural networks (CNNs) have been proposed for calculating the deformation field between two images [3], such models typically require large training sets and may suffer from generalization issues when applied to images presenting texture patterns that diverge from the training data. Therefore, classic algorithms, such as Advanced Normalization Tools (ANTs) [1], are still preferred as off-the-shelf tools for image registration in neuroscience due to scarce experimental data and the diversity of data acquisition protocols and registration tasks. Recently, implicit neural representations (INRs) have been utilized for image registration in MRI and CT [14,16], offering a hybrid approach that connects modern deep learning techniques with per-case optimization as used in classical approaches. INRs are defined on continuous coordinate spaces, making them suitable for registration of images that differ in geometry.

In this work, we propose a novel INR-based framework well-suited to address the challenging problem of gene expression brain image registration. We associate the registration problem with an image decomposition task. We utilize implicit neural networks to decompose the ISH image into two separate images: a support image and a residual image. The support image corresponds to the part of the ISH image that is well-aligned with the registration template image in respect to the texture. On the contrary, the residual image presents features of the ISH image, such as artifacts or texture patterns (e.g. gene expression), which presumably undermine the registration procedure. The support image is used to improve the deformation field calculations. We also introduce an exclusion loss to encourage clearer separation of the support and residual images. The usefulness of the proposed method is demonstrated using 2D ISH gene expression images of the marmoset brain.

2   Methods

2.1   Registration with Implicit Networks

The goal of the pairwise image registration is to determine a spatial transformation that maximizes the similarity between the moving image M and the target fixed template image F. INRs serve as a continuous, coordinate based approximation of the deformation field obtained through a fully connected neural network. In this study, as the backbone for our method, we utilized the standard

approach to registration with INRs, as described in [14,16]. We used a single implicit deformation network D to map 2D spatial coordinates $\bar{x} \in [-1, 1]^2$ of the moving image M to a displacement vector $\Delta\bar{x} \in \mathbb{R}^2$. Next, the transformation field was determined as $\Phi(\bar{x}) = \bar{x} + \Delta\bar{x}$ and the bilinear interpolation algorithm was applied to obtain the corresponding moved image $T_\Phi(M)$.

   To train the deformation network, the following loss function based on correlation coefficients was applied to assess the similarity between the moved image $T_\Phi(M)$ and the fixed template image F:

$$\mathcal{L}_{cc}(F, T_\Phi(M)) = \frac{1}{2N} \sum_{\bar{x}} \Big(\text{NCC}(F, T_\Phi(M)) + \text{LNCC}(F, T_\Phi(M)) \Big), \quad (1)$$

where NCC and LNCC stand for the normalized cross-correlation and local normalized cross-correlation based loss functions averaged over the entire image domain consisting of N elements. NCC was used to stabilize the training of the network, while LNCC ensured good local registration results. Additionally, following the standard approach to INR based registration, we regularized the deformation field based on the Jacobian matrix determinant $|J_{\Phi(\bar{x})}|$ using following equation [16]:

$$\mathcal{L}_{reg}(\Phi(\bar{x})) = \frac{1}{N} \sum_{\bar{x}} |1 - |J_{\Phi(\bar{x})}||. \quad (2)$$

2.2   Registration Guided Image Decomposition

Our aim is to improve the registration performance associated with the implicit deformation network D. The proposed framework is presented in Fig. 1. We assume that the moving image M can be decomposed with separate implicit networks, S and R, into two images: the support image M_S and the residual image M_R. Ideally, the support image should correspond to the part of the moving image that contributes to the registration performance. On the contrary, we expect the residual image to include image artifacts and texture patterns (e.g. ISH gene expression patterns) that diverge from the fixed template image and undermine the registration procedure. We impose the following condition based on the mean squared error loss function for the decomposition of the moving image:

$$\mathcal{L}_{rec}(M, M_S + M_R) = \frac{1}{N} \sum_{\bar{x}} (M - M_S - M_R)^2, \quad (3)$$

stating that the support M_S and residual M_R images should sum up to the moving image M. To ensure that the support image M_S contributes to the registration with respect to the fixed image F, we utilize the cross-correlation based loss function $\mathcal{L}_{cc}(F, T_\Phi(M_S))$ (Eq. 1), where $T_\Phi(M_S)$ stands for the transformed support image M_S. Therefore, the deformation network is trained to provide the transformation field $\Phi(x)$ both for the moving image and the support image
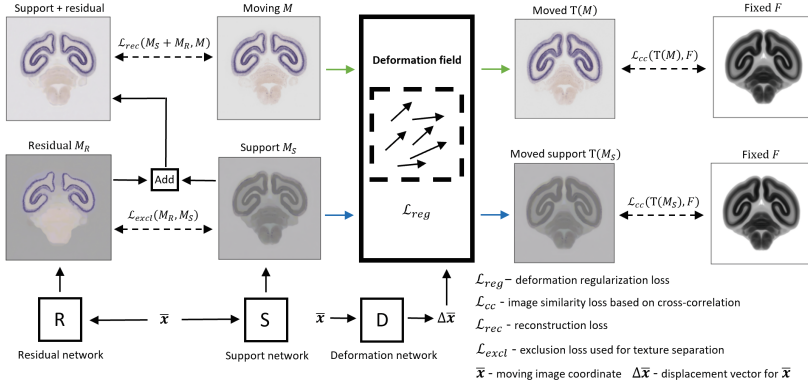
Fig. 1. We use implicit networks S and R to decompose the moving image into the support and residual images. The moving and support images are jointly registered to the fixed template image, which guides the image decomposition procedure to generate a support image that is well-aligned to the fixed image with respect to the texture. The residual image includes the remaining moving image contents that do not contribute to the registration, such as local gene expression patterns or image artifacts.

using two cross-correlation based loss functions. This way the training of the deformation network is guided to provide a more detailed transformation field for the contents of the moving image that actually correspond to the fixed template image. Moving image texture patterns that do not correspond to the fixed image have lower impact on the training of the deformation network.

In practice, it might be beneficial, following INR based methods for obstruction and rain removal, to additionally constrain the image decomposition procedure to obtain more clearly separated support M_S and residual M_R images [10]. For this, we utilize the following exclusion loss to encourage the gradient structure of the implicit networks S and R to be decorrelated [4]:

$$\mathcal{L}_{excl}(M_S, M_R) = \frac{1}{N} \sum_{\bar{x}} \sum_{i,j} |\Gamma(J_S(\bar{x}), J_R(\bar{x}))| \tag{4}$$

$\Gamma(J_S(\bar{x}), J_R(\bar{x})) = \tanh(J_S(\bar{x})) \otimes \tanh(J_R(\bar{x}))$, $\otimes$ indicates element-wise multiplication and indices i, j go over all elements of the matrix Γ.

In our framework, we jointly optimize all three implicit networks (D, S and R) using the following composite loss function:

$$\begin{aligned} Loss = {} & \alpha_1 \mathcal{L}_{cc}(F, T_\Phi(M)) + \alpha_2 \mathcal{L}_{cc}(F, T_\Phi(M_S)) + \alpha_3 \mathcal{L}_{reg}(\Phi(\bar{x})) \\ & + \alpha_4 \mathcal{L}_{rec}(M, M_S + M_R) + \alpha_5 \mathcal{L}_{excl}(M_S, M_R). \end{aligned} \tag{5}$$

The first row of Eq. 5 can be perceived as a standard registration loss, while the second row stands for a regularized image reconstruction loss.

2.3 Evaluation

We designed the proposed method with the aim to address the problem of ISH gene expression image registration. For the evaluation, we used neonate marmoset brain ISH images collected at the Laboratory for Molecular Mechanisms of Brain Development, RIKEN Center for Brain Science, Wako, Japan (gene-atlas.brainminds.jp) [6,12]. We prepared manual annotations for 2D images from 50 gene expression datasets. Atlas template images were created using ANTs [1], based on semi-automatically aligned sets of 2D ISH images from 1942 gene expression datasets. ISH images used to generate the template were converted to gray-scale to meet ANTs requirements and better highlight brain tissue interfaces.

Performance of the proposed approach was compared to the SynthMorph network and the ANTs SyN registration algorithm based on mutual information metric, as these two methods do not require pre-training and can serve as off-the-shelf registration tools for neuroscience [1,5]. We conducted an ablation study to assess the effectiveness of the proposed representation decomposition approach with and without the exclusion loss. Registration methods were evaluated quantitatively based on Dice scores using manual 2D segmentations prepared for the following five brain structures ranging in size and shape complexity: aqueduct (AQ, 95 masks), hippocampus area (HA, 570 masks), dorsal lateral geniculate (DLG, 370 masks), inferior colliculus (IC, 70 masks) and visual cortex area (VCA, 68 masks). Segmentations were outlined both for the template and ISH 2D images, resulting in 1114 image pairs corresponding to the same brain regions. We also calculated the percentage of the non-positive Jacobian determinant values to assess the deformation field folding. Moreover, we determined the structural similarity index (SSIM) between the moved images and the template fixed images.

2.4 Implementation

We utilized sinusoidal representation networks to determine the implicit representations [13]. Each network contained five fully connected hidden layers with 256 neurons. We used the Fourier mapping with six frequencies to encode the input coordinates [15]. The coordinates and the encoded coordinates were additionally concatenated within the middle layer of the network. Weights of the networks were initialized following the original paper except for the last linear layer of the deformation network D, for which we uniformly sampled the weights from $[-0.0001, 0.0001]$ interval to ensure small deformations at initial epochs. Additional details about the network architecture can be found in the supplementary materials. Networks were trained for 1000 epochs using AdamW optimizer with learning rate of 0.0001 on a server equipped with several NVIDIA A100 GPUs [8]. ISH images of size 360×420 were downsampled to 256×256. Each epoch corresponded to a batch of all image pixel coordinates [13]. After some initial experiments, we set the composite loss function weights (Eq. 5) to

$\alpha_1 = \alpha_2 = \alpha_3 = \alpha_5 = 1$ and $\alpha_4 = 100$, partially following the previous studies on INRs [10,14,16]. The window size for the LNCC loss was set to [32, 32]. Our PyTorch implementation of the proposed INR based registration method is available at https://github.com/BrainImageAnalysis/ImpRegDec.
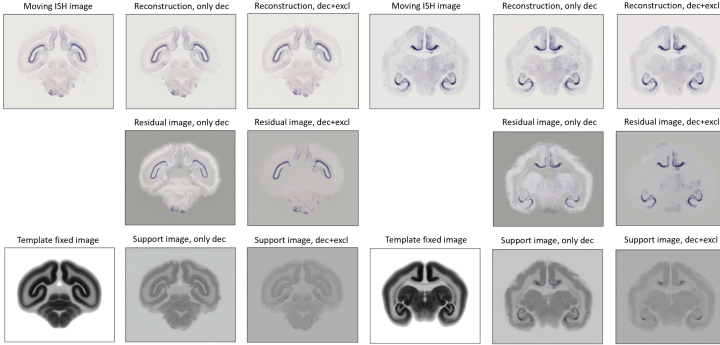

Fig. 2. Illustration of the moving image decomposition obtained with the proposed method (dec). Incorporation of the exclusion loss (excl) resulted in clearer separation of the gene expression texture patterns in the residual images.

3   Results

3.1   Qualitative Results

Support and residual images generated with the proposed method are shown in Fig. 2. The support images retain the main style and content of the fixed template image, while the residual images include the remaining image contents, along with gene expression patterns not present in the template image. Utilization of the exclusion loss resulted in a clearer and more visually plausible separation between the support and residual images, particularly for gene expression patterns. Figure 3 further highlights the usefulness of the proposed registration guided image decomposition technique. First, our method can be applied to extract microscopy image artifacts, and therefore mitigate their impact on the registration. Second, the proposed method is general and can also be applied to register an ISH gene expression image to a Nissl image. In this case, the color distribution of the support image corresponds to that of a Nissl image, while the residual image presents the local contents of the gene expression image. We also used the proposed method to register an ISH brain image to another ISH image with a different gene expression. For this example, the residual image highlighted the gene expression patterns of the moving image, while the support image showed the gene expression patterns of the fixed image.

Figure 4 visually compares the registration performance of the proposed technique, equipped with the exclusion loss, to ANTs. We found that the proposed

method provided good results both in respect to the image registration and the transformation of the manual segmentations.

3.2   Quantitative Results

Table 1 shows Dice scores obtained for the selected marmoset brain regions. Registration techniques based on INRs outperformed the other methods on four out of five brain regions. ANTs achieved better registration results for only one structure, the VCA, which was the largest among the annotated brain regions
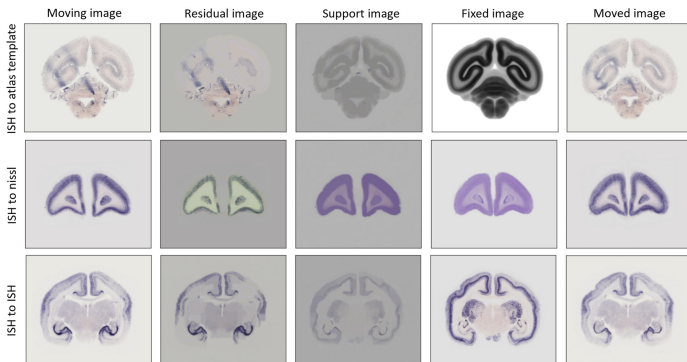

Fig. 3. Proposed technique can be useful for the extraction of microscope image artifacts (e.g. diagonal lines in the first row of images). It can also be applied to register ISH brain images to Nissl images or other ISH images. For such cases the support image presents image style of the fixed image, while the residual image includes local image patterns of the moving image.
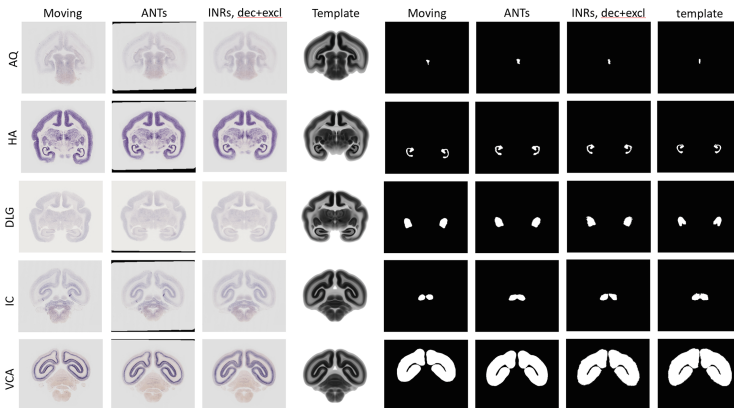

Fig. 4. Comparison of the proposed registration technique based on implicit networks and ANTs. AQ, HA, DLG, IC and VCA indicate the aqueduct, hippocampus area, dorsal lateral geniculate, inferior colliculus and visual cortex area, respectively.

Table 1. Dice scores (mean ± std) determined for the aqueduct (AQ), hippocampus area (HA), dorsal lateral geniculate (DLG), inferior colliculus (IC) and visual cortex are (VCA). Best results are shown in bold. dec and excl stand for the proposed image decomposition technique and the exclusion loss.

Method	AQ ↑	HA ↑	DLG ↑	IC ↑	VCA ↑
None	0.497 ± 0.194	0.311 ± 0.132	0.612 ± 0.141	0.742 ± 0.169	0.848 ± 0.051
ANTs SyN	0.673 ± 0.102	0.644 ± 0.141	0.757 ± 0.130	0.831 ± 0.133	$\mathbf{0.941} \pm \mathbf{0.015}$
SynthMorph	0.625 ± 0.129	0.503 ± 0.190	0.719 ± 0.146	0.798 ± 0.157	0.922 ± 0.034
INRs	0.734 ± 0.071	0.657 ± 0.127	0.756 ± 0.130	0.804 ± 0.191	0.922 ± 0.022
INRs, dec	0.748 ± 0.067	0.662 ± 0.138	$\mathbf{0.767} \pm \mathbf{0.125}$	0.839 ± 0.142	0.916 ± 0.033
INRs, dec+excl	$\mathbf{0.749} \pm \mathbf{0.063}$	$\mathbf{0.665} \pm \mathbf{0.134}$	0.766 ± 0.128	$\mathbf{0.845} \pm \mathbf{0.143}$	0.920 ± 0.017

Table 2. Structural similarity index (SSIM) and the percentage of the non-positive Jacobian determinant values (mean ± std) calculated for the investigated registration methods. Best results are shown in bold. dec and excl indicate the proposed image decomposition technique and the exclusion loss, respectively.

Method	SSIM ↑	$|J_\Phi| \leq 0$ [%] ↓
None	0.619 ± 0.046	–
ANTs	0.656 ± 0.059	$<\mathbf{0.001}$
SynthMorph	0.683 ± 0.039	$<\mathbf{0.001}$
INRs	0.713 ± 0.052	0.353 ± 0.459
INRs, dec	0.725 ± 0.054	0.359 ± 0.415
INRs, dec+excl	$\mathbf{0.727} \pm \mathbf{0.054}$	0.429 ± 0.460

and already similar in unregistered images with an initial Dice score of 0.848. Additionally, the Dice score for the VCA was high and comparable across all investigated registration methods. Our approach achieved significantly better Dice scores compared to the standard INRs for AQ, HA, DLG and IC (t-test's p-values < 0.05). Furthermore, incorporating the exclusion loss slightly improved the Dice scores for three structures.

SSIM values in Table 2 show that the registration based on implicit networks provided the most structurally similar results to the template images. With respect to the SSIM metric, our method significantly outperformed other approaches (t-test's p-values < 0.05). ANTs and SynthMorph provided smoother deformation fields compared to the implicit networks, with significantly lower percentage of folding (t-test's p-values < 0.05). However, the percentage of the folding obtained for the implicit networks was small and acceptable, as defined by folds in 0.5% of all pixels [11]. The main disadvantage of the proposed approach was the relatively long optimization time of about 90 s for a single pairwise registration, resulting from the requirement to jointly train three implicit networks.

4   Conclusion

Our approach based on implicit networks and registration-guided image decomposition has demonstrated excellent performance for the challenging task of registering ISH gene expression images of the marmoset brain. The results show that our approach outperformed pairwise registration methods based on ANTs and SynthMorph CNN, highlighting the potential of INRs as versatile off-the-shelf tools for image registration. Moreover, the proposed registration-guided image decomposition mechanism not only improved the registration performance, but also could be used to effectively separate the patterns that diverge from the target fixed image. In the future, we plan to investigate the possibility of using image decomposition for simultaneous registration and pattern segmentation, and methods to speed up the training [9]. We also plan to extend our technique to 3D and test it on medical images that include pathologies.

References

1. 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)
2. Corrales, M., et al.: A single-cell transcriptomic atlas of complete insect nervous systems across multiple life stages. Neural Dev. **17**(1), 8 (2022)
3. Fu, Y., Lei, Y., Wang, T., Curran, W.J., Liu, T., Yang, X.: Deep learning in medical image registration: a review. Phys. Med. Biol. **65**(20), 20TR01 (2020)
4. Gandelsman, Y., Shocher, A., Irani, M.: "Double-dip": unsupervised image decomposition via coupled deep-image-priors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11026–11035 (2019)
5. Hoffmann, M., Billot, B., Greve, D.N., Iglesias, J.E., Fischl, B., Dalca, A.V.: SynthMorph: learning contrast-invariant registration without acquired images. IEEE Trans. Med. Imaging **41**(3), 543–558 (2021)
6. Kita, Y., et al.: Cellular-resolution gene expression profiling in the neonatal marmoset brain reveals dynamic species-and region-specific differences. Proc. Natl. Acad. Sci. **118**(18), e2020125118 (2021)
7. Lein, E.S., et al.: Genome-wide atlas of gene expression in the adult mouse brain. Nature **445**(7124), 168–176 (2007)
8. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
9. Mehta, I., Gharbi, M., Barnes, C., Shechtman, E., Ramamoorthi, R., Chandraker, M.: Modulated periodic activations for generalizable local functional representations. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 14214–14223 (2021)

10. Nam, S., Brubaker, M.A., Brown, M.S.: Neural image representations for multi-image fusion and layer separation. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds.) ECCV 2022, Part VII. LNCS, vol. 13667, pp. 216–232. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-20071-7_13

11. Qiu, H., Qin, C., Schuh, A., Hammernik, K., Rueckert, D.: Learning diffeomorphic and modality-invariant registration using B-splines. In: Medical Imaging with Deep Learning (2021)

12. Shimogori, T., et al.: Digital gene atlas of neonate common marmoset brain. Neurosci. Res. **128**, 1–13 (2018)

13. Sitzmann, V., Martel, J., Bergman, A., Lindell, D., Wetzstein, G.: Implicit neural representations with periodic activation functions. Adv. Neural. Inf. Process. Syst. **33**, 7462–7473 (2020)

14. Sun, S., Han, K., Kong, D., You, C., Xie, X.: MIRNF: medical image registration via neural fields. arXiv preprint arXiv:2206.03111 (2022)

15. Tancik, M., et al.: Fourier features let networks learn high frequency functions in low dimensional domains. Adv. Neural. Inf. Process. Syst. **33**, 7537–7547 (2020)

16. Wolterink, J.M., Zwienenberg, J.C., Brune, C.: Implicit neural representations for deformable image registration. In: International Conference on Medical Imaging with Deep Learning, pp. 1349–1359. PMLR (2022)