Symmetric Multi-level Gradient-Inverse Consistency Network for Brain Image Registration with Large Deformation

Haoying Bai[1] HAOYINGBAI@MAIL.BNU.EDU.CN
Jichang Zhang[1] ZHANGJICHANG@MAIL.BNU.EDU.CN
Tongtong Che[*1] CHE@BNU.EDU.CN
Shuyu Li[*1] SHUYULI@BNU.EDU.CN
[1] *State Key Laboratory of Cognitive Neuroscience and Learning*
Beijing Normal University, Beijing, 100875, China.

Editors: Under Review for MIDL 2025

Abstract

Accurate and robust deformable image registration is crucial for brain image analysis. While deep learning has significantly advanced this field, existing methods often lack robustness for large deformations due to inter-subject variability, frequently requiring pre-registration and relying heavily on data-driven approaches. To address these limitations, we propose an end-to-end **Symmetric Multis-level Gradient-Inverse Consistency Network (SM-GICNet)** for accurate and robust brain image registration. SM-GICNet employs 1) a symmetric multi-level framework with an attention gate mechanism to capture complex deformations at multiple scales, 2) a symmetric registration strategy at each level to mitigate directional bias, and 3) a gradient inverse consistency strategy to reduce reliance on data-driven constraints and control deformation field complexity. Experimental results demonstrate that our method is able to eliminate the need for pre-registration and outperforms state-of-the-art methods on large deformation registration tasks, achieving a Dice similarity coefficient of 0.797. The implementation of our SM-GICNet is available online at https://github.com/LSYLAB/SM-GICNet.git.

Keywords: Symmetric registration, Consistency-Constrained, Inverse-Consistent, Multi-level

1. Introduction

Deformable image registration is a fundamental task in medical image analysis, aiming to establish a nonlinear spatial correspondence between a pair of images (source/moving and target/fixed images)(Sotiras et al., 2013). Traditional registration methods typically rely on iterative optimization strategies to maximize similarity metrics in the transformation space (Oliveira and Tavares, 2014; Shen and Davatzikos, 2002). While these methods achieve relatively stable performance, they suffer from low computational efficiency and limitations when handling complex deformation fields. Recently, deep learning-registration methods achieve significant performance improvements by leveraging large datasets and the powerful modeling capabilities of neural networks (Wang et al., 2023, 2022; Shi et al., 2022). These methods directly predict the deformation field via neural networks, demonstrating superior performance in nonlinear registration tasks.

[*] Corresponding authors

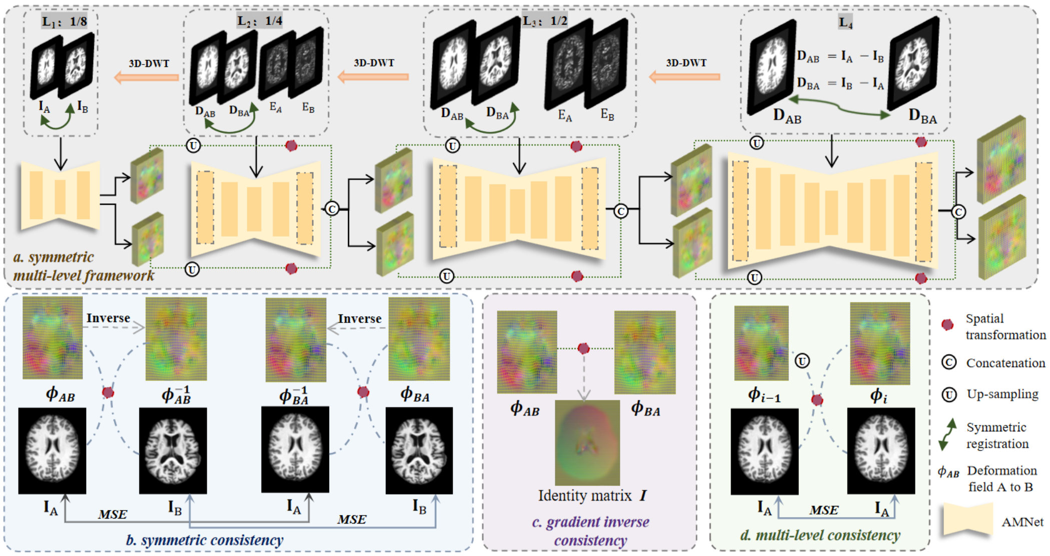

Figure 1: (a) A symmetric multi-level registration framework incorporating an attention gate mechanism; (b) symmetric deformation field consistency strategy at each level; (c) a deformation field constraint based on gradient inverse consistency; and (d) a multi-level consistency strategy.

However, current deep-learning-based registration methods still face the challenges posed by large deformations. In common multi-level image registration approaches, this is typically addressed with a coarse-to-fine registration strategy (Hering et al., 2019; Mok and Chung, 2020; Eppenhof et al., 2019). Specifically registration proceeds from a coarse alignment that captures large deformations to a fine alignment that incorporates local details. Unfortunately, the fixed directionality in existing multi-level methods inevitably leads to asymmetric and biased mapping, which may negatively impact subsequent image analysis tasks such as computational anatomy. Some symmetric single-level registration methods have been proposed to simultaneously predict forward ($I_A \rightarrow I_B$) and backward ($I_B \rightarrow I_A$) transformations between images (Xu et al., 2023; Chen et al., 2023; Tian et al., 2024), enhancing the invertibility and accuracy of the deformation field. However, these methods mainly rely on similarity-driven optimizations, leading to slow convergence and high computational costs. Furthermore, without rigid or affine pre-registration, the registration performance for large deformation fields remains unsatisfactory, particularly in scenarios with high inter-subject heterogeneity or significant lesion variations (Mok and Chung, 2022; Balakrishnan et al., 2019).

To address these issues, we propose a **Symmetric Multis-level Gradient-Inverse Consistency Network (SM-GICNet)** capable of directly handling large deformation registration tasks with high variability without pre-registration. Specifically, we introduce a method that combines a **symmetric multi-level** registration framework with an **attention gate mechanism** to capture deformation features at different scales and focus attention on complex deformation regions during the high-resolution stage. At each level, we employ a **symmetric deformation field consistency strategy**, where images A and B are simultaneously registered as each other's moving image, predicting both forward and backward deformation fields to ensure the stability and consistency of the transformations. Furthermore, we incorporate a **gradient inverse consistency constraint** to directly regularize the gradient alignment of the forward and backward deformation fields, limiting deformation field complexity and mitigating reliance on purely data-driven optimizations.

Our main contributions are as follows:

- Our SM-GICNet is a symmetric multi-level registration framework with attention-gate mechanism for learning a progressively refined representations transformations, which eliminates the bias of generic directional image registration.
- Our SM-GICNet embraces the gradient inverse consistency constraint to replace conventional regularizers, which reduces reliance on purely data-driven optimization.
- Our SM-GICNet achieves accurate and fast registration without any pre-registration (rigid or affine registration), which demonstrates superior performance in large deformation registration tasks with high heterogeneity in brain MRI.

2. Related Work

2.1. Symmetric Diffeomorphic Registration

Symmetric registration is crucial for accurate medical image registration, particularly in estimating deformations between image pairs, improving geometric consistency and precision (Greer et al., 2021). Early methods independently estimated forward and backward transformations, lacking guaranteed inverse consistency (Zheng et al., 2021; Kim et al., 2021). Most deformable methods use displacement fields (Brauwers and Frasincar, 2021; Cao et al., 2018; Yang et al., 2017), neglecting differential properties like topology preservation and invertibility (Tian et al., 2025), hindering true symmetry. Diffeomorphic registration, using stationary velocity fields, offers a solution (Avants et al., 2008), ensuring smooth, invertible mappings. The diffeomorphic deformation field, ϕ_t (parameterized by $t \in [0, 1]$), is generated from the velocity field as:

$$\frac{d\phi_v}{dt} = v_t(\phi_t) = v_t \circ \phi_t \tag{1}$$

Diffeomorphic models are advantageous for constructing symmetric registration networks due to their inherent invertibility.

2.2. Attention Gate Mechanism

The Attention Gate Mechanism (AGM) is an attention mechanism readily integrated into various Convolutional Neural Networks (CNNs) to enhance performance in tasks such as image segmentation, object detection, and image classification (Azad et al., 2024; Guo et al., 2022; Brauwers and Frasincar, 2021). By focusing on salient regions, AGMs guide networks to prioritize significant features (Li et al., 2023; Ranjbarzadeh et al., 2021). Numerous studies have shown that attention gate mechanisms significantly improve the accuracy of registration models in local regions, particularly in tasks requiring attention to fine anatomical structures (Chen et al., 2022; Tang et al., 2022). Attention gate networks based on U-Net architectures are especially prevalent in image registration. For example, Attention U-Net (Oktay et al., 2018) incorporates attention gate modules to focus on key regions of the input image, enhancing the model's response to specific areas while reducing interference from background noise and irrelevant features. This approach has demonstrates performance improvements on several public medical image datasets.

3. Methods

We present a symmetric multi-level gradient inverse consistency network (SM-GICNet) for large deformation image registration. As illustrated in Figure 1, SM-GICNet includes: (1) a symmetric multi-level registration framework incorporating the attention gate mechanism; (2) symmetric deformation field consistency strategy at each level; and (3) a deformation field constraint based on gradient inverse consistency.

3.1. Symmetric Multi-level Registration Framework

A novel symmetric multi-level registration framework with attention gate mechanism is proposed to effectively capture multi-scale deformation fields between image pairs. The multi-level architecture comprises four consecutive levels, efficiently capturing both global and complex local deformations within a single forward process. Symmetry is achieved by alternately using each image as the moving image in a single registration, yielding ϕ_{AB} and ϕ_{BA}. Simultaneously, both forward (ϕ_{AB}) and inverse (ϕ_{BA}^{-1}) deformation fields are directly obtained for each image, enabling multi-constraint network learning.

The multi-level network is constructed using a 3D discrete wavelet transform (3D-DWT) (Ghasemzadeh and Demirel, 2018) to leverage both low-frequency global and high-frequency local information, following the input method of AMNet (Che et al., 2023). An AGM is introduced at level 4 to automatically enhance deformation field learning in crucial regions while suppressing background influence, improving the model's ability to deal with large deformations and get finer structural details. Further details regarding the attention gate network architecture are provided in Appendix A.

3.2. Symmetric Deformation Field Consistency Strategy

Our network employs a symmetric deformation field consistency strategy at each level, promoting bidirectional symmetry during unidirectional registration to enhance deformation field stability and consistency. Following diffeomorphic principles, a stationary velocity field is used instead of a displacement field for parameterization. The deformation field is defined as in the equation1. The velocity field v is integrated over a unit time using a scaling and squaring operation with a time step $T = 7$ to obtain the final deformation field $\phi(1)$. The diffeomorphic model adapts to large or complex deformations, and since the output is a velocity field, the inverse velocity field, and subsequently the inverse deformation field, can be obtained by negating the velocity field. This forms the basis of the symmetric registration network.

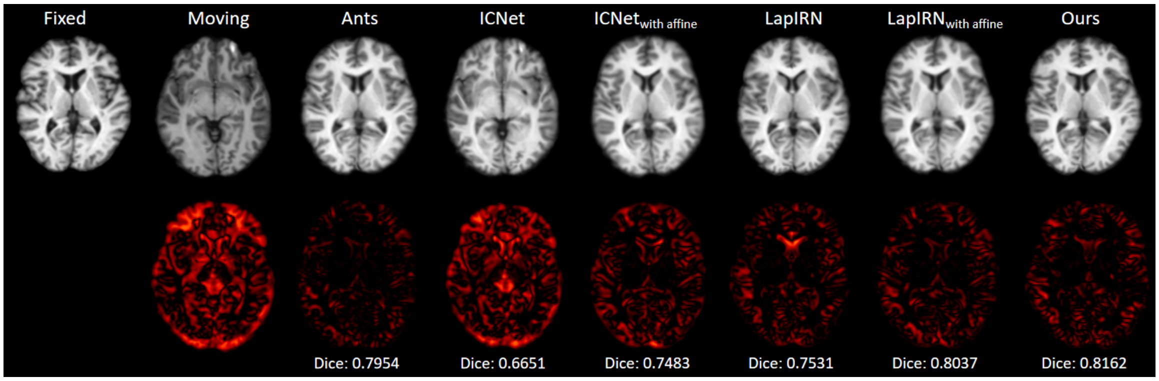

Figure 2: Qualitative registration results by three different methods and SM-GICNet. The second row shows corresponding heat maps of the absolute difference with respect to the fixed image.

Specifically, assuming the network learns the deformation field $\phi_{A\to B}$ from moving image A to fixed image B, the inverse deformation field $\phi_{A\to B}$ is derived using diffeomorphic properties. The network then learns $\phi_{B\to A}$ from B to A, ensuring consistency between these two deformation fields. This reduces reliance on training data and improves the inverse representation capabilities during forward registration.

3.3. Gradient Inverse Consistency Constraint

In a single registration, images I_A and I_B are registered reciprocally, alternately serving as the moving image. This yields both forward (ϕ_{AB}) and backward (ϕ_{BA}) deformation fields. To avoid data-driven dependence, the gradient inverse consistency constraint is applied directly to the deformation fields, which replaces conventional regularizers. This ensures that the matrix resulting from the forward deformation field and its inverse is consistent with the identity grid $\mathbf{I}$, suppressing unreasonable complexities in the deformation field.

3.4. Losses and Objectives

Our method employs a multi-component loss function to optimize image registration. The total loss is a weighted sum of the following terms.

Consistency Loss: This term enforces consistency between forward and inverse transformations. It comprises three sub-components:
Symmetric Consistency:

$$\mathcal{L}_{Sy} = \mathcal{L}_{MSE}(I_A \circ \phi_{AB}, I_B \circ \phi_{BA}) \tag{2}$$

Inverse Consistency:

$$\mathcal{L}_{In} = \mathcal{L}_{MSE}(\phi_{AB} \circ \phi_{AB}^{-1}), \mathbf{I}) \tag{3}$$

Multi-level Consistency:

$$\mathcal{L}_{Mu} = \mathcal{L}_{MSE}(\mathbf{w}_i, (\mathbf{w}_i + \mathbf{w}_{i-1})/2) \tag{4}$$

And the Consistency Loss is the sum of them: $\mathcal{L}_{consistency} = \lambda_s \mathcal{L}_{Sy} + \lambda_i \mathcal{L}_{In} + \lambda_m \mathcal{L}_{Mu}$, where $\lambda_s = 0.001$, $\lambda_i = 0.01$ and $\lambda_m = 0.0005$.

Multi-level NCC Similarity Loss: This term maximizes the Normalized Cross-Correlation (NCC) similarity between warped and fixed images across multiple resolution levels. Let L_{NCC}^i denote the NCC loss at level i. Then: $L_{NCC} = \sum_i \alpha_i L_{NCC}^i$, where α_i are weights assigned to each level.

Smoothness Regularization Loss: This term penalizes overly complex deformations by minimizing the L2 norm of the deformation field gradients: $L_{smooth} = \sum_i \beta_i L_2^i$, where L_2^i is the loss at level i and β_i are weights.

The total loss function is given by: $L_{total} = L_{consistency} + L_{NCC} + L_{smooth}$.

4. Experiment

4.1. Data

We evaluated our method using the IXI[1] brain MRI dataset, which comprises T1-weighted 3D MRI scans. A subset of 314 subjects was selected and partitioned into training (n=269), validation (n=15), and testing (n=30) sets. All images undergo standard skull stripping and contrast correction preprocessing; however, no prior registration (e.g., affine or rigid transformations) is performed. Image segmentation was performed using FreeSurfer software, employing its default brain template (Fischl, 2012), resulting in 36 regions of interest (ROIs). All performance evaluations were based on the overlap of ROIs in the test images. For comparison, a separate training and testing dataset is created using affine registration with FSL's `flirt` command.

4.2. Training Details

Our models are trained using the Adam optimizer on a single NVIDIA A100 GPUs. We trained the network for 4, 4, 6 and 6 epochs in levels 1, 2, 3, and 4. We set the initial learning rate as 1×10^{-4} and then it is multiplied by 0.5 every 50k iterations after the first 60k iterations for each level.

4.3. Evaluation metrics

To evaluate registration accuracy, we employed the Dice Similarity Coefficient (DSC) and Jacobian Determinant(JD). We first evaluated our method using the Dice score of the subcortical segmentation maps. The closer the Dice value is to 1, the better the overlap of the two images, indicating better registration performance. The Jacobian determinant of the deformation field is used to evaluate the local properties of the deformation. For a deformation field ϕ, the Jacobian determinant J_ϕ at each spatial location measures the local volume change induced by the transformation. We also evaluated the computational efficiency of the registration method by measuring the processing time required for a single pair of images.

1. https://brain-development.org/ixi-dataset/

5. Results

5.1. Comparisons with the state-of-the-art methods

Table 1: Quantitative evaluation of different registration methods. Higher DSC values indicate better performance, while lower proportions of $|JD| \leq 0$ and registration times are preferred.

	SyN	**ICNet**	**ICNet w.affine**	**LapIRN**	**LapIRN w.affine**	**Ours**
DSC	0.753 (±0.103)	0.289 (±0.211)	0.714 (±0.081)	0.478 (±0.272)	**0.803 (±0.069)**	0.797 (±0.137)
$\|JD\| \leq 0$	0.000 (±0.000)	0.499 (±0.002)	0.489 (±0.014)	0.503 (±0.041)	0.488 (±0.028)	**0.483 (±0.027)**
Time	1 h	**0.252s (±0.021)**	7.893 s (±0.021)	1.100 s (±0.006)	8.743 s (±0.006)	0.272 s (±0.042)

We compared our method with three widely-used registration approaches:

SyN (Avants et al., 2009) – A widely used registration method from ANTs, using cross-correlation and a multi-resolution optimization strategy with an initial affine transformation.

ICNet (Zhang, 2018) – An inverse-consistent deep network for unsupervised deformable registration, trained using the authors' optimal hyperparameters.

LapIRN (Mok and Chung, 2020) – A multi-level diffeomorphic registration algorithm using a Laplacian pyramid architecture and three CNNs, trained using the authors' optimal hyperparameters.

Table 1 summarizes the quantitative results of our method and three comparison methods across all ROIs. For a fair comparison, LapIRN and ICNet are trained and tested on both unregistered and pre-registered datasets. Our method achieves the highest Dice score and the lowest JAD score on the unregistered dataset. While LapIRN achieved slightly better performance on the pre-registered dataset, it incurs a significantly higher computational cost with only marginal improvement in accuracy.

Figure 2 illustrates a prediction example for different methods. Our method demonstrates results closest to the fixed image, achieving high accuracy even without pre-registration. The energy map of high-frequency information provides useful guidance for further refining the registration of structural boundaries.

Figure 3(a) presents the results of reciprocal registration for a pair of images. Our network effectively registers these images, even in subjects with high heterogeneity, and the resulting deformation fields exhibit symmetry.

5.2. Ablation Study:

Figure 3(b) demonstrates the importance of each component of our model. By individually removing the gradient inverse consistency constraint, the multi-level structure, the

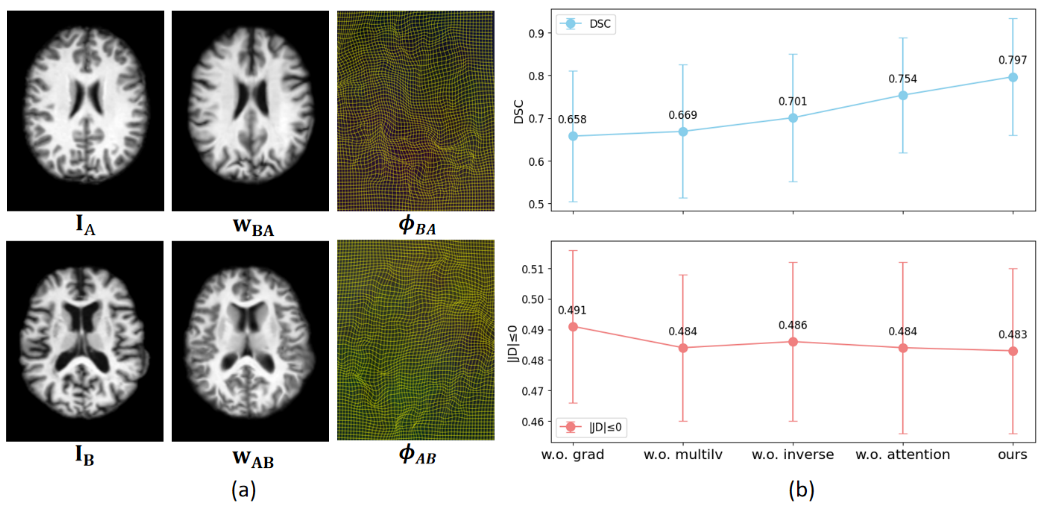

Figure 3: (a) The deformed images resulting from reciprocal registration of the image pair, along with the resulting symmetric deformation fields. (b) Ablation study results show DSC and JD values obtained after removing each component of the network individually.

symmetry structure, and the attention gate mechanism, we observe that each component contributes to the overall performance. Employing only gradient inverse consistency or symmetric consistency constraints individually resulted in decreased registration accuracy, hindering the model's ability to learn large deformations. The synergistic combination of both constraints was crucial for improved performance. Similarly, omitting the multi-level network architecture also led to reduced accuracy, as a single-level network struggled to capture the full range of deformations. Finally, incorporating attention at the highest resolution level enabled the network to better discriminate between large and small deformations, preserving more feature information and leading to improved results.

6. Conclusion

We introduce a novel Symmetric Multi-level Gradient-Inverse Consistency Network (SM-GICNet) for robust large deformation image registration, specifically addressing the challenges posed by high inter-subject variability in medical images. Unlike many existing deep learning-based methods, SM-GIC Net directly handles large deformations without requiring pre-registration steps. This is achieved through a synergistic combination of three key innovations: 1) a symmetric multi-level architecture incorporating an attention gate mechanism for efficient multi-scale deformation capture; 2) a symmetric deformation field consistency strategy to ensure bidirectional symmetry and stability; and 3) a gradient inverse consistency constraint to reduce reliance on purely data- driven optimization and complexity.

Acknowledgments

This work is supported by the National Natural Science Foundation of China [No. 32271146], the Startup Funds for Top-notch Talents at Beijing Normal University, China, and the fellowship of China National Postdoctoral Program for Innovative Talents. [No. BX20240039].

References

Brian B Avants, Charles L Epstein, Murray Grossman, and James C Gee. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, 12(1):26–41, 2008.

Brian B Avants, Nick Tustison, Gang Song, et al. Advanced normalization tools (ants). *Insight j*, 2(365):1–35, 2009.

Reza Azad, Ehsan Khodapanah Aghdam, Amelie Rauland, Yiwei Jia, Atlas Haddadi Avval, Afshin Bozorgpour, Sanaz Karimijafarbigloo, Joseph Paul Cohen, Ehsan Adeli, and Dorit Merhof. Medical image segmentation review: The success of u-net. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, 38(8):1788–1800, 2019.

Gianni Brauwers and Flavius Frasincar. A general survey on attention mechanisms in deep learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(4):3279–3298, 2021.

Xiaohuan Cao, Jianhua Yang, Jun Zhang, Qian Wang, Pew-Thian Yap, and Dinggang Shen. Deformable image registration using a cue-aware deep regression network. *IEEE Transactions on Biomedical Engineering*, 65(9):1900–1911, 2018.

Tongtong Che, Xiuying Wang, Kun Zhao, Yan Zhao, Debin Zeng, Qiongling Li, Yuanjie Zheng, Ning Yang, Jian Wang, and Shuyu Li. Amnet: Adaptive multi-level network for deformable registration of 3d brain mr images. *Medical Image Analysis*, 85:102740, 2023.

Junyu Chen, Eric C Frey, Yufan He, William P Segars, Ye Li, and Yong Du. Transmorph: Transformer for unsupervised medical image registration. *Medical image analysis*, 82: 102615, 2022.

Zeyuan Chen, Yuanjie Zheng, and James C Gee. Transmatch: A transformer-based multilevel dual-stream feature matching network for unsupervised deformable image registration. *IEEE transactions on medical imaging*, 43(1):15–27, 2023.

Koen AJ Eppenhof, Maxime W Lafarge, Mitko Veta, and Josien PW Pluim. Progressively trained convolutional neural networks for deformable image registration. *IEEE transactions on medical imaging*, 39(5):1594–1604, 2019.

Bruce Fischl. Freesurfer. *Neuroimage*, 62(2):774–781, 2012.

Aman Ghasemzadeh and Hasan Demirel. 3d discrete wavelet transform-based feature extraction for hyperspectral face recognition. *Iet Biometrics*, 7(1):49–55, 2018.

Hastings Greer, Roland Kwitt, François-Xavier Vialard, and Marc Niethammer. Icon: Learning regular maps through inverse consistency. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3396–3405, 2021.

Meng-Hao Guo, Tian-Xing Xu, Jiang-Jiang Liu, Zheng-Ning Liu, Peng-Tao Jiang, Tai-Jiang Mu, Song-Hai Zhang, Ralph R Martin, Ming-Ming Cheng, and Shi-Min Hu. Attention mechanisms in computer vision: A survey. *Computational visual media*, 8(3):331–368, 2022.

Alessa Hering, Bram Van Ginneken, and Stefan Heldmann. mlvirnet: Multilevel variational image registration network. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22*, pages 257–265. Springer, 2019.

Boah Kim, Dong Hwan Kim, Seong Ho Park, Jieun Kim, June-Goo Lee, and Jong Chul Ye. Cyclemorph: cycle consistent unsupervised deformable image registration. *Medical image analysis*, 71:102036, 2021.

Xiang Li, Minglei Li, Pengfei Yan, Guanyi Li, Yuchen Jiang, Hao Luo, and Shen Yin. Deep learning attention mechanism in medical image analysis: Basics and beyonds. *International Journal of Network Dynamics and Intelligence*, pages 93–116, 2023.

Tony CW Mok and Albert CS Chung. 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*, pages 211–221. Springer, 2020.

Tony CW Mok and Albert CS Chung. Unsupervised deformable image registration with absent correspondences in pre-operative and post-recurrence brain tumor mri scans. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 25–35. Springer, 2022.

Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

Francisco PM Oliveira and Joao Manuel RS Tavares. Medical image registration: a review. *Computer methods in biomechanics and biomedical engineering*, 17(2):73–93, 2014.

Ramin Ranjbarzadeh, Abbas Bagherian Kasgari, Saeid Jafarzadeh Ghoushchi, Shokofeh Anari, Maryam Naseri, and Malika Bendechache. Brain tumor segmentation based on deep learning and an attention mechanism using mri multi-modalities brain images. *Scientific Reports*, 11(1):1–17, 2021.

Dinggang Shen and Christos Davatzikos. Hammer: hierarchical attribute matching mechanism for elastic registration. *IEEE transactions on medical imaging*, 21(11):1421–1439, 2002.

Jiacheng Shi, Yuting He, Youyong Kong, Jean-Louis Coatrieux, Huazhong Shu, Guanyu Yang, and Shuo Li. Xmorpher: Full transformer for deformable medical image registration via cross attention. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 217–226. Springer, 2022.

Aristeidis Sotiras, Christos Davatzikos, and Nikos Paragios. Deformable medical image registration: A survey. *IEEE transactions on medical imaging*, 32(7):1153–1190, 2013.

Linfeng Tang, Yuxin Deng, Yong Ma, Jun Huang, and Jiayi Ma. Superfusion: A versatile image registration and fusion network with semantic awareness. *IEEE/CAA Journal of Automatica Sinica*, 9(12):2121–2137, 2022.

Lin Tian, Hastings Greer, Roland Kwitt, François-Xavier Vialard, Raúl San José Estépar, Sylvain Bouix, Richard Rushmore, and Marc Niethammer. unigradicon: A foundation model for medical image registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 749–760. Springer, 2024.

Lin Tian, Hastings Greer, Raúl San José Estépar, Roni Sengupta, and Marc Niethammer. Nephi: Neural deformation fields for approximately diffeomorphic medical image registration. In *European Conference on Computer Vision*, pages 213–237. Springer, 2025.

Alan Q Wang, M Yu Evan, Adrian V Dalca, and Mert R Sabuncu. A robust and interpretable deep learning framework for multi-modal registration via keypoints. *Medical Image Analysis*, 90:102962, 2023.

Xuqing Wang, Yanan Su, Ruoyu Liu, Qianhui Qu, Hao Liu, and Yi Gu. Medical image registration method based on simulated ct. In *International Conference on Intelligent Computing*, pages 719–728. Springer, 2022.

Tao Xu, Ting Jiang, and Xiaoning Li. Ic-cvit: Inverse-consistent convolutional vision transformer for diffeomorphic image registration. In *2023 International Joint Conference on Neural Networks (IJCNN)*, pages 1–10. IEEE, 2023.

Xiao Yang, Roland Kwitt, Martin Styner, and Marc Niethammer. Quicksilver: Fast predictive image registration–a deep learning approach. *NeuroImage*, 158:378–396, 2017.

Jun Zhang. Inverse-consistent deep networks for unsupervised deformable image registration. *arXiv preprint arXiv:1809.03443*, 2018.

Yuanjie Zheng, Xiaodan Sui, Yanyun Jiang, Tongtong Che, Shaoting Zhang, Jie Yang, and Hongsheng Li. Symreg-gan: symmetric image registration with generative adversarial networks. *IEEE transactions on pattern analysis and machine intelligence*, 44(9):5631–5646, 2021.

Appendix A. Attention gate network architecture

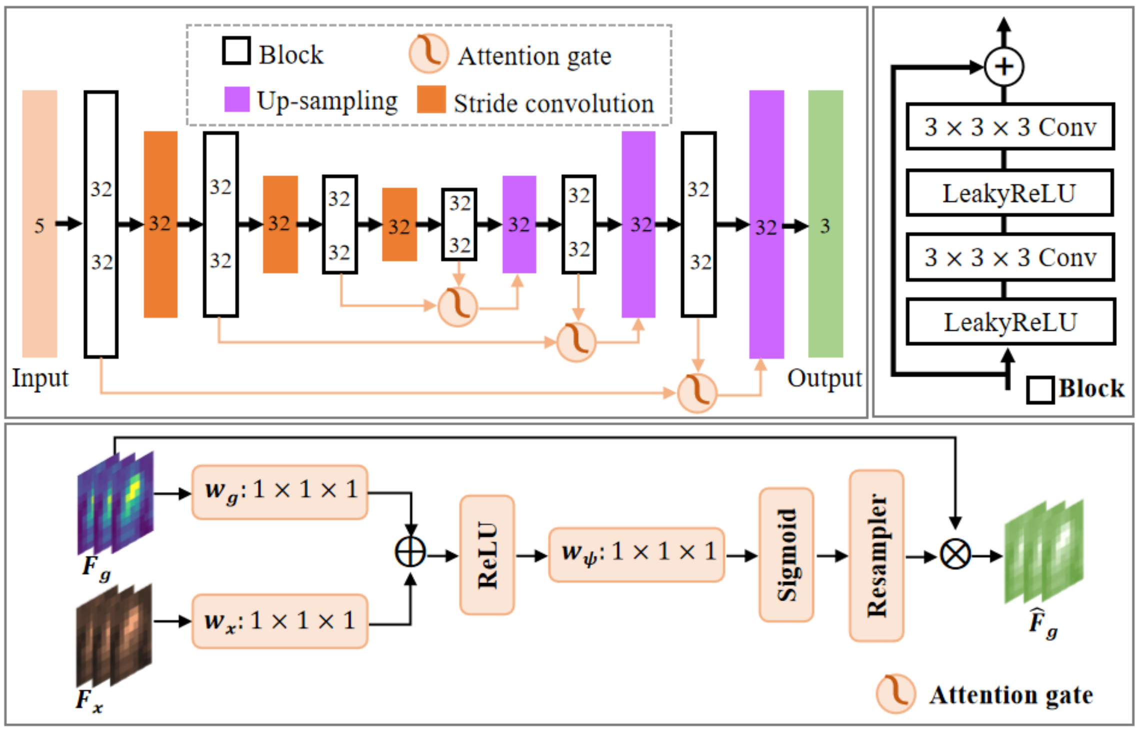

Figure 4: Attention gate network architecture.

The level 4 architecture retains an input layer, three residual blocks, and three convolutional layers with stride 2 in the encoder, along with three residual blocks, three upsampling layers, and an output layer in the decoder. Each residual block contains two consecutive convolutional layers. Three skip connections between the encoder and decoder incorporate attention gates. The attention gate first applies a 1x1x1 convolution to the input feature F_g from the encoder and, similarly, to the downsampled feature from the corresponding encoder branch. These two outputs are then summed, followed by ReLU activation and another 1x1x1 convolution to reduce the channel dimension to 1. A sigmoid activation is applied to the result, resampled to match the original feature size, creating a 1D weight matrix. Finally, this weight matrix is multiplied with the input feature, producing a new feature map. This process enhances focus on registration-relevant local features, resulting in more accurate deformation fields and improved network convergence speed.