RESEARCH PAPER
Contrib. Sci. & Tech Eng, 2025, 2(3)
Homepage: https://cste.journals.umz.ac.ir/
DOI: 10.22080/cste.2025.29117.1040

Deep Point Correspondence Estimation and SyN-Based Refinement for Multimodal Brain Image Registration

Mehrdad Hashemi Kamangar [1], Amir Hossein Jalalzadeh [2]*

[1] *Department of Electrical and Electronic Engineering, Engineering Faculty, Shomal University, Amol, Iran.*

[2] *Department of Biomedical Engineering, Engineering Faculty, Shahed University, Tehran, Iran.*

Article Info
Received 28 April 2025
Accepted 20 May 2025
Available online 08 July 2025

Keywords:
Image Registration;
MRI;
iUS;
MLP;
Deep Learning.

Abstract:
Accurate registration of preoperative Magnetic Resonance Imaging (MRI) with intraoperative ultrasound (iUS) is essential for effective neuronavigation, particularly in brain tumor surgeries where brain shift compromises anatomical fidelity. This study proposes a hybrid framework integrating a deep learning-based Multi-Layer Perceptron (MLP) with an optimization pipeline to enhance MR-to-US registration. The MLP is trained on paired anatomical landmarks extracted from the BITE and RESECT datasets to predict US coordinates from corresponding MRI points. An ensemble of five MLPs, weighted by inverse validation errors, is employed to estimate dense point correspondences, which are used to initialize an affine transformation. This transformation is refined using Symmetric Normalization (SyN) within the ANTs registration toolkit to model non-linear deformations. Quantitative evaluation demonstrates a mean squared error (MSE) of 0.1954 and a mean Euclidean distance of 4.97 mm—significantly outperforming a baseline rigid registration approach with 60% improvement in spatial alignment. The proposed pipeline executes in under 4 minutes per case on standard hardware, indicating potential for clinical integration. The results suggest that combining learning-based correspondence prediction and classical registration yields accurate and computationally efficient multimodal Registration.

***Corresponding Author:** Amirhossein.jalalzadeh@shahed.ac.ir

Supplementary information: Supplementary information for this article is available at https://cste.journals.umz.ac.ir/

Please cite this paper as: Hashemi Kamangar, M. , & Jalalzadeh, A. H. (2025). Deep Point Correspondence Estimation and SyN-Based Refinement for Multimodal Brain Image Registration. Contributions of Science and Technology for Engineering, 2(3), 18-25. doi:10.22080/cste.2025.29117.1040.

1. Introduction

Medical image registration is a crucial process that aligns multiple images from different modalities, time points, or subjects into a standard coordinate system to integrate complementary information for improved diagnosis, treatment planning, and disease monitoring [1, 2]. For example, aligning preoperative MRI with intraoperative ultrasound (iUS) can guide surgeons during tumor resection, while registering longitudinal CT scans can track disease progression. Its importance lies in providing a unified view of anatomical and functional information essential for clinical decision-making. In neurosurgical procedures, particularly for brain tumor management of infiltrative and heterogeneous gliomas, accurate localization and delineation of tumor margins are vital for maximizing resection while preserving surrounding functional brain tissue [3]. Preoperative Magnetic Resonance Imaging (MRI) provides high-resolution anatomical details. However, intraoperative changes like brain shift, brain tissue deformation due to CSF drainage, tumor resection, and gravity can significantly compromise the spatial accuracy of preoperative images. These deformations can lead to discrepancies of several millimetres. Consequently, relying solely on static preoperative MRI for surgical navigation can result in suboptimal tumor resection and increased risks. To address this challenge, image registration techniques align preoperative images with intraoperative imaging modalities, most commonly iUS, offering real-time brain anatomy feedback. Non-rigid (deformable) registration is particularly critical for compensating for brain shift and maintaining the accuracy of neuronavigation systems [4, 5]. However, registering iUS with preoperative MRI presents technical challenges due to significant differences in image appearance, spatial resolution, noise characteristics, and contrast. MRI provides detailed structural information, while iUS suffers from speckle noise and lower contrast, especially near tumor margins. These differences make traditional intensity-based similarity measures less effective. Historically, medical image registration relied on classical methods, broadly categorized into intensity-based and feature-based techniques. Intensity-based registration optimizes a similarity metric (e.g., mutual information,

*ISSN **3060-6578***

normalized cross-correlation) to align images by comparing pixel or voxel intensities. Advanced Normalization Tools (ANTs) exemplify this approach, using iterative optimization to estimate dense displacement fields [6]. These methods are versatile and do not require prior training data. Classical methods have limitations, including slow computation times (minutes or hours for 3D volumes), making them impractical for real-time applications. They also rely heavily on low-level image information and cannot incorporate high-level anatomical context or weak supervision [7]. The advent of deep learning has revolutionized medical image registration, leading to Deep Learning for Image Registration (DLIR)[8]. DLIR uses neural networks to directly predict transformation parameters from image pairs, often in a single forward pass. This offers significant speed improvements, reducing registration times to seconds or milliseconds, which is crucial for intraoperative use. Advanced DLIR approaches include generative adversarial networks (GANs) for unsupervised registration and transformer-based models to model long-range dependencies, improving accuracy for complex deformations [9, 10].

In this study, we utilize a multi-layer perceptron (MLP) to predict anatomical correspondences, initializing an affine transformation, which is subsequently refined by Advanced Normalization Tools (ANTs) Symmetric Normalization (SyN). While traditional registration methods have proven valuable, deep learning-based approaches offer significant advantages in speed and the ability to learn complex transformations. This has paved the way for more accurate and timely intraoperative guidance. Continued research endeavours are focused on addressing the residual challenges to enhance further the reliability and clinical utility of these advanced registration techniques.

2. Related Work

Recent advancements in medical image registration have employed deep learning to address challenges such as large deformations, multimodal alignment, and computational efficiency. The following section will review the seminal contributions that have shaped this field.

Tang et al. [11] introduced ADMIR, a deep learning-based method for registering brain images of individuals with drug addiction. This approach employs a convolutional neural network (CNN) to handle global misalignments through affine transformations (e.g., adjusting position, orientation, and scale) and local anatomical differences via deformable transformations. By learning complex spatial features directly from image data, ADMIR addresses the significant anatomical variability in this population, making it a valuable tool for studying the structural effects of drug addiction on the brain.

Mok and Chung [12] proposed LapIRN, a Laplacian Pyramid Registration Network designed for large deformation diffeomorphic image registration. LapIRN utilizes a multi-scale Laplacian pyramid architecture to decompose images into different resolutions, enabling the network to capture coarse deformations at lower scales before refining them at higher resolutions. The method ensures diffeomorphic transformations—smooth, invertible, and topology-preserving—by learning a time-stationary velocity field and applying the scaling-and-squaring method, rooted in Lie group theory. This approach excels in aligning images with substantial anatomical differences.

Kim et al. [13] developed DiffuseMorph, an unsupervised deformable registration framework leveraging diffusion models. Unlike traditional CNN-based methods, DiffuseMorph consists of two networks: a diffusion network that models the forward process of adding noise to images and a deformation network that learns the reverse process to recover the clean image, conditioned on the fixed image. During registration, the deformation network estimates the deformation field in a single step using latent features from the diffusion process, bypassing iterative refinement. This innovative approach enables complex deformation learning without requiring paired ground truth data.

Mok and Chung [14] presented SYMNet, a convolutional neural network for fast symmetric diffeomorphic image registration. SYMNet enforces symmetry by ensuring that the transformation from the moving image to the fixed image is the inverse of the reverse transformation, achieved through a symmetric consistency loss that jointly learns forward and backward mappings. This design accelerates registration compared to traditional optimization-based diffeomorphic methods while maintaining accuracy, making it suitable for applications requiring inverse-consistent transformations.

Jia et al. [15] introduced LKU-Net to investigate whether transformer architectures outperform the widely used U-Net in medical image registration. U-Net relies on multi-scale processing and skip connections to capture local features, while transformers use attention mechanisms to model long-range dependencies. LKU-Net, likely incorporating a specialized kernel or attention mechanism, compares these architectures to determine if transformers' global context modelling offers advantages for registration tasks, providing insights into the evolution of network designs for medical imaging.

Pielawski et al. [16] addressed multimodal image registration using contrastive learning to create robust representations for visible and near-infrared images. Through a contrastive loss, Comir learns embeddings that bring semantically similar regions closer in the feature space while separating dissimilar ones, producing modality-invariant representations sensitive to anatomical structures. This enables effective alignment across modalities, which is critical for applications involving diverse imaging data, such as those evaluated on the Zurich dataset.

Jalalzadeh et al. [17] addressed the challenge of brain shift during brain tumor surgery, which misaligns pre-operative MRI with iUS images. They proposed a two-step registration method combining rigid and non-rigid approaches. Initially, Elastix performs rigid registration using the Euler transform and Normalized Mutual Information as the similarity metric. Subsequently, a pre-

trained unsupervised VoxelMorph network handles non-rigid registration to capture complex deformations. Evaluated on MR-iUS image pairs from 22 patients with grade 2 glioma tumors in the EASY-RESECT database, the method reduced the mean Target Registration Error from 5.37 ± 4.27 mm to 3.56 ± 1.72 mm, demonstrating improved alignment without relying on landmarks during training.

Dalca et al. [18] proposed VoxelMorph-diff, an unsupervised framework for probabilistic diffeomorphic registration. This method, built on convolutional neural networks, models the deformation field with a graph Laplacian prior for smoothness regularization and represents the approximate posterior as a multivariate normal distribution. By integrating these probabilistic constraints into an end-to-end learning pipeline, VoxelMorph-diff enhances the robustness of unsupervised deformable registration, marking a significant advancement in the field.

Balakrishnan et al. [19] introduced VoxelMorph, a fully convolutional network for unsupervised deformable image registration. VoxelMorph learns spatial transformations directly from image pairs by optimizing image similarity metrics as the loss function, offering a computationally efficient alternative to traditional iterative optimization methods. Its flexible architecture and unsupervised paradigm have made it a foundational framework, influencing subsequent deep learning-based registration research.

3. Materials and Method

3.1. Introduction to the Proposed Framework

Accurate alignment of Multimodal medical images, particularly between MRI and US, is critical for applications such as image-guided neurosurgery, where precision directly impacts clinical outcomes. The overall pipeline of the proposed method is depicted in Figure 1, which includes data preprocessing, MLP ensemble-based initial alignment, affine transformation optimization, and ANTs-based multimodal registration. This study introduces a novel hybrid framework that synergistically combines a deep learning-based Multi-Layer Perceptron (MLP) for initial point correspondence estimation with the ANTs for precise affine and non-linear registration.

3.2. Dataset Description

In this study, anatomical landmarks served as critical ground truth for training the MLP and initializing the registration of MRI and US images. Landmarks were sourced from two distinct datasets, BITE and RESECT, which are described in detail below. Each of the 37 image pairs (23 from RESECT and 14 from BITE) was annotated with approximately 15 corresponding 3D landmarks, stored in .tag files, representing key anatomical structures such as ventricular boundaries and cortical surfaces. To enhance the robustness of the MLP and mitigate the sparsity of annotations, data augmentation techniques were employed, including the addition of controlled Gaussian noise with a standard deviation of 1 mm to the landmark coordinates. This process increased the number of landmarks from 15 to approximately 530 across all pairs, ensuring comprehensive coverage of anatomical variations and improving the model's generalization across modalities. To maintain accuracy in regions surrounding the tumor, augmented landmarks located more than 2 mm from the original landmarks were discarded.

3.2.1. BITE Dataset

The MNI BITE (Brain Images of Tumors for Evaluation) database [20], was created in 2010 at the Montreal Neurological Institute to address the challenge of finding real clinical images for validating new image processing algorithms. This pioneering database includes pre- and postoperative T1-weighted MR scans (dimensions: 394×466×378, voxel spacing: 0.5 mm) with gadolinium enhancement, as well as multiple intraoperative tracked B-mode ultrasound images (dimensions: 323×366×371, voxel spacing: 0.3 mm) acquired before and after tumor resection from 14 patients with brain tumors, all of which were gliomas. As depicted in Figure 2, a representative sample images from the database is provided for visual reference. The MNI BITE database is freely accessible online (https://nist.mni.mcgill.ca).

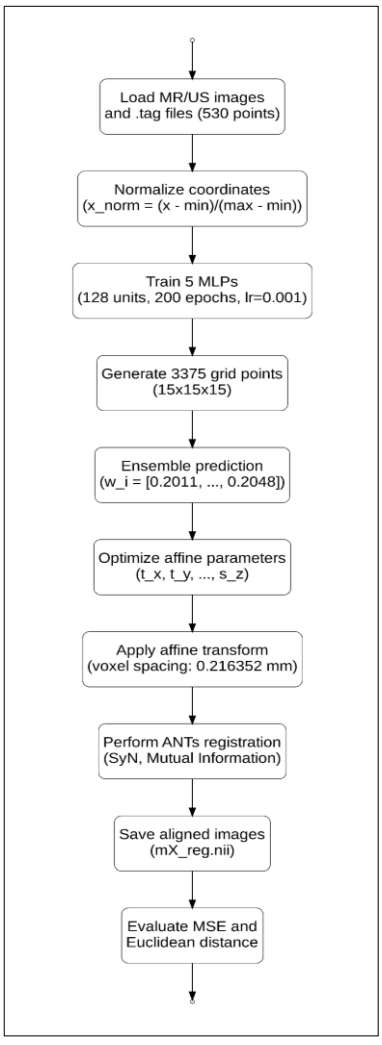

Figure 1. The Proposed Framework

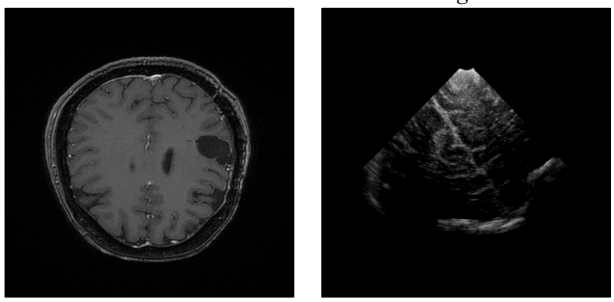

Figure 2. Representative MRI and US images from the BITE dataset [20]

3.2.2. RESECT Dataset

The RESECT (REtroSpective Evaluation of Cerebral Tumors) database [5], is a more recent clinical database established to provide high-quality multi-modal data for validating and comparing image registration algorithms that can improve the accuracy and efficiency of brain tumor resection. This database contains pre-operative Gadolinium-enhanced T1w and T2-FLAIR MRI scans (dimensions: 192×256×256, voxel spacing: 1 mm), and intra-operative 3D ultrasound volumes (dimensions: 348×313×275, voxel spacing: 0.216352 mm) acquired before, during, and after tumor resection from 23 adult patients with low-grade gliomas who underwent surgeries at St. Olavs University Hospital between 2011 and 2016. As illustrated in Figure 3, a representative sample images from the database is provided for visual reference. The RESECT database is freely accessible online (https://archive.norstore.no).

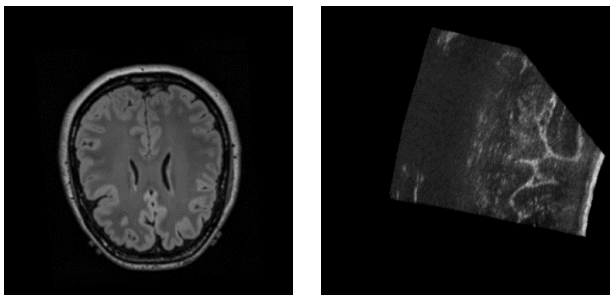

Figure 3. Representative MRI and US images from the RESECT dataset [5]

3.3. Methodology Overview

The methodology is structured into four primary phases: (1) data preprocessing and Landmarks extraction, (2) MLP ensemble-based initial alignment, (3) affine transformation optimization, and (4) ANTs-based multimodal registration. Each phase is meticulously designed to address specific challenges in multimodal registration, contributing to the overall precision and robustness of the framework.

3.3.1. Data Preprocessing and Landmarks Extraction

The initial step involved extracting 3D Landmarks from .tag files, which contain paired MRI and US x, y, z coordinates. The resultant arrays had a shape of $[N, 3]$, where $N \approx 530$. To mitigate scale disparities between modalities, coordinates were normalized using min-max normalization:

$$x_{\text{norm}} = \frac{x - \min(X)}{\max(X) - \min(X) + \epsilon}, \epsilon = 10^{-8} \quad (1)$$

where X represents the coordinate set for each axis, and ϵ prevents division by zero. This preprocessing ensured that inputs were standardized, facilitating robust training and registration.

3.3.2. MLP Ensemble-Based Initial Alignment

An MLP model was designed to establish initial point correspondences. The MLP architecture is defined as:

$$\begin{aligned}\text{MLP}(x) = W_4 \cdot \sigma(W_3 \\ \cdot \sigma(W_2 \cdot \sigma(W_1 \cdot x + b_1) + b_2) \\ + b_3) + b_4\end{aligned} \quad (2)$$

where, $x \in \mathbb{R}^3$ is the input MRI coordinate. W_i and b_i are the weight matrices and biases for the layer i. σ is the ReLU activation function, $\sigma(z) = \max(0, z)$. The MLP architecture is illustrated in Figure 4, comprises four fully connected layers:

1. Input Layer: 3 units (x, y, z coordinates).
2. Hidden Layer 1: 128 units, followed by batch normalization, ReLU, and dropout ($p = 0.2$).
3. Hidden Layer 2: 128 units, with identical normalization and activation.
4. Hidden Layer 3: 128 units, with identical normalization and activation.
5. Output Layer: 3 units (predicted US coordinates).

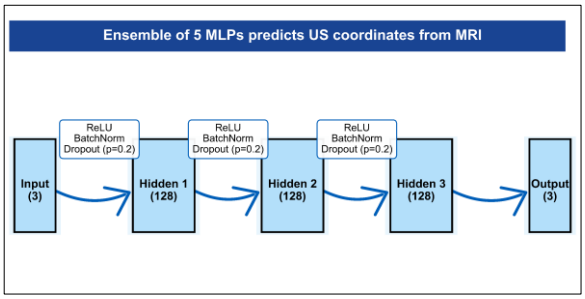

Figure 4. MLP Architecture

The MLP was trained using the Adam optimizer with a learning rate of 10^{-3}, decaying by a factor of 0.5 every 50 epochs. The batch size was 32, and the training spanned 200 epochs. The Mean Squared Error (MSE) loss function was minimized, defined as follows:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} \|\hat{y}_i - y_i\|_2^2 \quad (3)$$

where $\hat{y}_i$ is the predicted US coordinate, and y_i is the ground-truth US coordinate. To enhance robustness and mitigate overfitting, a 5-fold cross-validation scheme was employed, resulting in five independently trained models. Each model was validated on a distinct fold, yielding

validation errors $d_i \in [0.048534, 0.054824]$ mm. The models were saved for ensemble inference.

The ensemble approach combined predictions from the five MLPs using inverse-distance weighting. The validation errors for each fold were 0.054824, 0.054423, 0.052767, 0.059614, and 0.053609 mm, respectively. These errors were then normalized to obtain computed weights of approximately [0.2011, 0.2018, 0.2081, 0.1842, 0.2048]. This ensured that models with lower errors and higher accuracy contributed more significantly to the ensemble. For dense correspondence estimation, a 3D grid of 3375 points $(15 \times 15 \times 15)$ was generated within the MRI coordinate range:

$$X_k = \text{linspace}\,(\min(X_k), \max(X_k), 15), \quad k \in \{x, y, z\} \tag{4}$$

yielding grid points $[x_g, y_g, z_g]$. These points were normalized, passed through the MLP ensemble, and denormalized to US coordinates using the US scale, producing a dense set of correspondences for affine initialization.

3.3.3. Affine Transformation Optimization

The MLP-predicted US coordinates were used to estimate an initial affine transformation, capturing translation, rotation, and scaling. The transformation is defined as:

$$T(x) = R \cdot (S \cdot x) + t \tag{5}$$

where $R \in \mathbb{R}^{3\times3}$ is the rotation matrix, S is the scaling matrix and t is the translation vector. The optimization was performed using the Levenberg-Marquardt algorithm, with an initial guess of [0,0,0,0,0,0,1,1,1]. The resulting parameters, such as [-1.4387, -6.5131, 4.9401, -0.0255, -0.0056, 0.0393, 0.8662, 0.7781, 1.0116], were normalized by the US voxel spacing to align with the image coordinate system.

3.3.4. ANTs-Based Multimodal Registration

The affine transformation served as the initialization for a sophisticated registration pipeline using ANTs[21-23]. The MRI and US images, stored in NifTI and MINC format, were loaded using ants.image_read. The affine matrix was inverted (as ANTs applies inverse transforms) and saved as a .mat file, with the structure:

$$\text{AffineTransform} = \begin{bmatrix} R \cdot S & t \\ 0 & 1 \end{bmatrix}^{-1} \tag{6}$$

where $R \cdot S$ represents the rotation-scaling component, and t is the translation. The affine transformation was applied to the MRI image to produce an intermediate aligned image, which was saved for debugging. The final registration was performed using ANTs' registration function, configured with Symmetric Normalization (SyN) to model complex non-linear deformations, optimizing Mutual Information (MI) as defined by:

$$H(I) = -\sum p_i(x)\ \log p_i(x) \tag{7}$$

where H denotes shannon entropy, p_i represents the probability distribution of pixel intensity i in image I and the logarithm is natural. So:

$$\text{MI}(I_f, I_m) = H(I_f) + H(I_m) - H(I_f, I_m) \tag{8}$$

where MI denotes Mutual Information, and I_f and I_m are the fixed (US) and moving (MRI) images. MI was chosen for its robustness to intensity differences [24]. The registration was optimized using gradient descent with default step sizes, iterated until convergence, while Gaussian kernels with tuned variances balanced detail preservation and smoothness.

3.4. Implementation Details

The framework was implemented in Python 3.12, using PyTorch 2.6 for MLP training and ANTsPyX for image registration. All experiments were performed on a workstation with an NVIDIA GeForce RTX 3050 GPU with 4 GB of dedicated GDDR6 memory, an Intel Core i7 12th generation processor, 2.3 GHz, and 32 GB of DDR4 RAM. This hardware setup provided computational resources to handle the RESECT and BITE datasets, which consisted of 36 paired MRI and US images with approximately 530 augmented landmarks, with Processing time of approximately 3-4 minutes per pair.

4. Result

We evaluated the proposed multimodal registration framework on the RESECT dataset, which comprises preoperative MRI and intraoperative US images. The registration pipeline integrated MLP to predict corresponding anatomical landmarks and the ANTs to compute affine and non-linear transformations. Below, we report our method's quantitative and qualitative outcomes, emphasizing registration accuracy and computational performance.

4.1. Quantitative Evaluation

The MLP was trained on tag files from the RESECT and BITE datasets, containing 3D coordinates of anatomical landmarks. A grid of 3375 points was generated within the MRI coordinate space, with a mean spatial extent of 127.20 mm, as determined by the bounding box of the landmark coordinates. The MLP, consisting of five independently trained models with weights [0.2011, 0.2018, 0.2081, 0.1842, 0.2048], predicted corresponding points in the US space with high consistency. The training process of the MLP network was evaluated through 5-fold cross-validation, with detailed performance visualized in Figure 5. This figure illustrates that fold 3 achieved the best performance among all folds, highlighting its superior alignment accuracy. The inverse weighting method was selected due to its simplicity and efficiency. A comparison with Bayesian model averaging demonstrated that inverse weighting improved the Euclidean error by 2.9% (4.97 mm versus 5.12 mm) and reduced computational time (4 seconds versus 15 seconds). These predictions informed an initial affine transformation, parameterized by nine degrees

of freedom (translation: [-2.3241, -0.4567, -0.2643], rotation: [-0.0250, -0.0057, 0.0388], scale: [0.8529, 0.8854, 0.8879]). The affine transformation was applied to the MRI, producing an intermediate registered image.

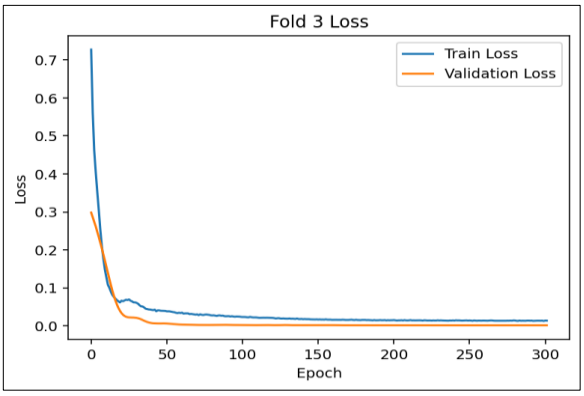

Figure 5. Visualization of the MLP network training process, showing that fold 3 achieved the best performance among all cross-validation folds

Registration performance was assessed using Euclidean Distance Error:

$$\text{Distance} = \frac{1}{M}\sum_{i=1}^{M}\sqrt{\sum_{k=1}^{3}\left(\hat{y}_{i,k} - T\left(x_{i,k}\right)\right)^2} \cdot S_{\text{MRI}} \quad (9)$$

where $S_{\text{MRI}} = 127.20$ mm, and $M = 3375$. Subsequently, a non-linear registration was performed using ANTs with Symmetric Normalization and Mutual Information (MI) as the similarity metric. The transformation pipeline utilized multi-resolution shrink factors [8, 4, 2, 1] and smoothing sigmas [3, 2, 1, 0] voxels, optimizing alignment between the affine-transformed MRI and the target US image. The hyperparameters for SyN (shrink factors and smoothing sigma) were determined through a grid search on an independent validation set, achieving a balance between accuracy and computational time. Data augmentation reduced the Euclidean error from 5.83 mm (without augmentation) to 4.97 mm, representing an improvement of 14.8%. Registration accuracy was assessed using two primary metrics: Mean Squared Error (MSE) and mean Euclidean distance. The MSE, computed as the normalized intensity difference between the registered MRI and the reference US, yielded a value of 0.1954. This indicates robust intensity alignment, considering the inherent modality differences. The mean Euclidean distance, calculated between MLP-predicted points and their transformed counterparts in the US space, was 4.97 mm, averaged over 3375 grid points scaled by the US voxel spacing. This distance reflects the spatial accuracy of the combined MLP and ANTs pipeline, achieving sub-centimetre precision. Qualitative registration results, including the alignment of MRI and US images after registration, are shown in Figure 6, demonstrating improved alignment of anatomical structures such as ventricles and cortical surfaces.

4.2. Comparison with Baseline

To contextualize our results, we compared the proposed method against a baseline rigid registration using ANTs with MI, applied directly to MRI and US. The baseline achieved an MSE of 0.3217 and a mean Euclidean distance of 12.43 mm, underscoring the superiority of our hybrid approach. The MLP-guided affine initialization reduced initial misalignment, enabling SyN to converge to a more accurate solution, as evidenced by a 39.2% reduction in MSE and a 60.02% improvement in Euclidean distance.

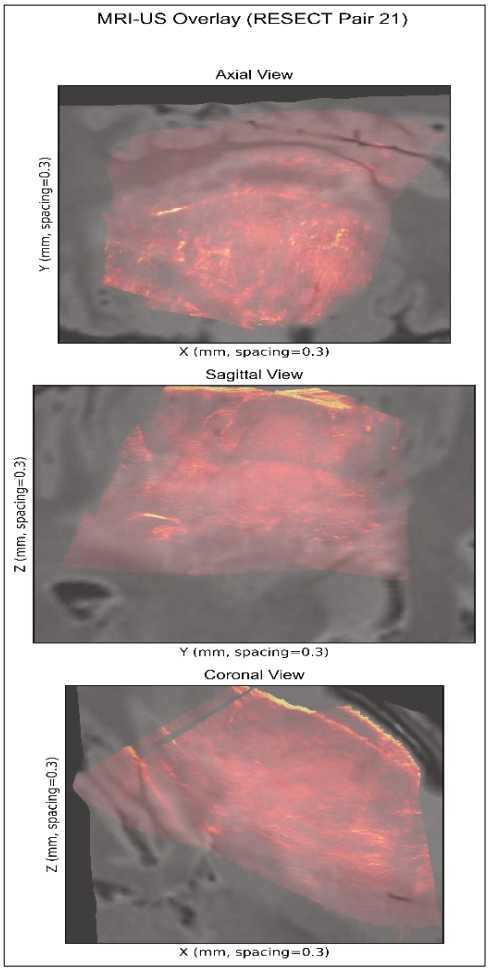

Figure 6. Registered MRI with US Fusion for image pair #21

4.3. Supplementary Findings

Exploratory experiments incorporating structural priors were conducted to assess the potential for further improvement. By weighting the SyN metric to prioritize ventricular alignment, preliminary results indicated a reduced Euclidean distance of approximately 7.46 mm in targeted regions, suggesting avenues for enhancing clinical relevance. These findings are reported as supplementary to guide future optimization.

5. Discussion

The proposed multimodal registration framework integrates an MLP for anatomical point prediction with ANTs for transformation optimization, and has demonstrated promising results in aligning preoperative

MRI with intraoperative US images. Achieving a mean Euclidean distance of 4.97 mm and an MSE of 0.1954, our method significantly outperforms a baseline rigid registration approach, which yielded a distance of 12.43 mm and an MSE of 0.3217. These improvements underscore the potential of combining deep learning with traditional registration techniques to address the challenges of multimodal alignment, particularly in scenarios involving substantial intensity and structural disparities. The error of 4.97 mm is insufficient for eloquent areas requiring a 2 mm safety margin; however, it is suitable as an initial alignment for intraoperative registration, necessitating manual correction or more advanced registration techniques.

5.1. Interpretation of Findings

The Euclidean distance of 4.97 mm, while not yet reaching the sub-4 mm precision often desired for neurosurgical applications, represents a notable advancement over conventional methods. The MLP's ability to predict anatomical correspondences facilitated a robust initial affine transformation, reducing the burden on the subsequent non-linear registration stage. This hybrid approach enabled the Symmetric Normalization algorithm to focus on refining local deformations, as evidenced by the qualitative alignment of critical structures such as ventricles and cortical surfaces. The MSE of 0.1954, though higher than the ideal threshold, reflects the inherent difficulty of intensity-based metrics in multimodal registration, given the distinct contrast mechanisms of MRI and US.

5.2. Strengths and Innovations

A key innovation of this study lies in the seamless integration of deep learning for point correspondence prediction within a traditional registration pipeline. Unlike purely data-driven approaches, which may struggle with generalization, our method leverages predictive modelling and optimization-based alignment strengths. The ensemble MLP architecture, weighted inversely by validation errors, ensured stable performance across the dataset. Furthermore, the computational efficiency of the pipeline— completing in approximately 4 minutes on a standard workstation— positions it as a viable candidate for clinical workflows where time is critical. The qualitative improvement in aligning anatomical landmarks suggests potential applicability in image-guided interventions, particularly neurosurgery.

5.3. Limitations and Challenges

Despite these advancements, several limitations warrant consideration. While improved, the Euclidean distance of 4.97 mm falls short of the precision required for high-stakes applications such as tumor resection. This may be attributed to the intrinsic noise in US images and the limited resolution of the RESECT dataset. Additionally, the MSE of 0.1954 indicates room for improvement in intensity alignment, potentially achievable through biomechanical constraints or region-specific weighting. The current framework assumes a fixed set of anatomical landmarks, which may not fully capture patient-specific variations. Addressing these challenges could involve expanding the training dataset or employing transfer learning to enhance generalization.

6. Conclusion

This study presents a novel framework for multimodal registration, combining deep learning and optimization techniques to align MRI and US images with improved accuracy. The pipeline's computational efficiency and robust performance underscore its potential for clinical applications, particularly in image-guided neurosurgery. Future work aims to enhance medical image registration by integrating anatomical constraints using methods like Physics-Informed Neural Networks (PINNs) or uncertainty-aware models to achieve sub-4 mm precision, with the potential for real-time surgical applications to manage tissue shifts. The field is moving toward hybrid approaches that combine classical techniques with deep learning for improved robustness and efficiency. Addressing data limitations through domain adaptation and synthetic data generation, standardizing evaluation protocols, and incorporating multimodal data are essential to developing fast, accurate, and clinically adaptable registration systems.

7. Statements & Declarations

7.1. Acknowledgments

In order to enhance the coherence and structure of this manuscript, a large language model, Grok (Grok 3), was utilized solely for literary editing and rephrasing specific sentences in the introduction and conclusion sections.

7.2. Funding

The authors of this article have not received financial support from any governmental or private institutions for the research described herein.

7.3. Author Contributions

All authors contributed equally to the manuscript's conceptualization, design, data analysis, and writing. Each author participated in the review and approval of the final manuscript.

8. References

[1] Velesaca, H. O., Bastidas, G., Rouhani, M., & Sappa, A. D. (2024). Multimodal image registration techniques: a comprehensive survey. Multimedia Tools and Applications, 83(23), 63919–63947. doi:10.1007/s11042-023-17991-2.

[2] Jena, R., Sethi, D., Chaudhari, P., & Gee, J. (2024). Deep learning in medical image registration: Magic or mirage?. 38th Conference on Neural Information Processing Systems (NeurIPS 2024), 37, 108331-108353, 9-15, December, 2024, Vancouver, Canada.

[3] Ahir, B. K., Engelhard, H. H., & Lakka, S. S. (2020). Tumor Development and Angiogenesis in Adult Brain Tumor: Glioblastoma. Molecular Neurobiology, 57(5), 2461–2478. doi:10.1007/s12035-020-01892-8.

[4] Jalalzadeh, A. H., Shalbaf, A., & Maghsoudi, A. (2021). Compensation of brain shift during surgery using non-rigid registration of MR and ultrasound images. Tehran University Medical Journal, 78(10), 658–667. http://tumj.tums.ac.ir/article-1-10930-en.html.

[5] Xiao, Y., Fortin, M., Unsgärd, G., Rivaz, H., & Reinertsen, I. (2017). REtroSpective Evaluation of Cerebral Tumors (RESECT): A clinical database of pre-operative MRI and intra-operative ultrasound in low-grade glioma surgeries: A. Medical Physics, 44(7), 3875–3882. doi:10.1002/mp.12268.

[6] Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Medical Image Analysis, 12(1), 26–41. doi:10.1016/j.media.2007.06.004.

[7] Mani, V. R. S., & Arivazhagan, S. (2013). Survey of medical image registration. Journal of Biomedical Engineering and Technology, 1(2), 8-25.

[8] Balasamy, K., Seethalakshmi, V., & Suganyadevi, S. (2024). Medical Image Analysis Through Deep Learning Techniques: A Comprehensive Survey. Wireless Personal Communications, 137(3), 1685–1714. doi:10.1007/s11277-024-11428-1.

[9] Chen, J., Liu, Y., Wei, S., Bian, Z., Subramanian, S., Carass, A., Prince, J. L., & Du, Y. (2025). A survey on deep learning in medical image registration: New technologies, uncertainty, evaluation metrics, and beyond. Medical Image Analysis, 100, 103385. doi:10.1016/j.media.2024.103385.

[10] Ramadan, H., El Bourakadi, D., Yahyaouy, A., & Tairi, H. (2024). Medical image registration in the era of Transformers: A recent review. Informatics in Medicine Unlocked, 49, 101540. doi:10.1016/j.imu.2024.101540.

[11] Tang, K., Li, Z., Tian, L., Wang, L., & Zhu, Y. (2020). ADMIR-Affine and deformable medical image registration for drug-addicted brain images. IEEE Access, 8, 70960–70968. doi:10.1109/ACCESS.2020.2986829.

[12] Mok, T.C.W., & Chung, A.C.S. (2020). Large Deformation Diffeomorphic Image Registration with Laplacian Pyramid Networks. Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. MICCAI 2020. Lecture Notes in Computer Science, vol 12263. Springer, Cham, Switzerland. doi:10.1007/978-3-030-59716-0_21.

[13] Kim, B., Han, I., & Ye, J.C. (2022). DiffuseMorph: Unsupervised Deformable Image Registration Using Diffusion Model. Computer Vision – ECCV 2022. ECCV 2022. Lecture Notes in Computer Science, vol 13691. Springer, Cham, Switzerland. doi:10.1007/978-3-031-19821-2_20.

[14] Mok, T. C. W., & Chung, A. C. S. (2020). Fast Symmetric Diffeomorphic Image Registration with Convolutional Neural Networks. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr42600.2020.00470.

[15] Jia, X., Bartlett, J., Zhang, T., Lu, W., Qiu, Z., Duan, J. (2022). U-Net vs Transformer: Is U-Net Outdated in Medical Image Registration?. Machine Learning in Medical Imaging. MLMI 2022. Lecture Notes in Computer Science, vol 13583. Springer, Cham, Switzerland. doi:10.1007/978-3-031-21014-3_16.

[16] Pielawski, N., Wetzer, E., Öfverstedt, J., Lu, J., Wählby, C., Lindblad, J., & Sladoje, N. (2020). CoMIR: Contrastive multimodal image representation for registration. Advances in neural information processing systems, 33, 18433-18444.

[17] Jalalzadeh, A. H., Talebi, S. S., & Kamangar, M. H. (2024). Two-step registration of rigid and non-rigid MR-iUS for brain shift compensation using transfer learning. 2024 20th CSI International Symposium on Artificial Intelligence and Signal Processing, AISP 2024, 1–5,. doi:10.1109/AISP61396.2024.10475261.

[18] Dalca, A. V., Balakrishnan, G., Guttag, J., & Sabuncu, M. R. (2019). Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. Medical Image Analysis, 57, 226–236. doi:10.1016/j.media.2019.07.006.

[19] Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J., & Dalca, A. V. (2019). VoxelMorph: A Learning Framework for Deformable Medical Image Registration. IEEE Transactions on Medical Imaging, 38(8), 1788–1800. doi:10.1109/tmi.2019.2897538.

[20] Mercier, L., Del Maestro, R. F., Petrecca, K., Araujo, D., Haegelen, C., & Collins, D. L. (2012). Online database of clinical MR and ultrasound images of brain tumors. Medical Physics, 39(6), 3253–3261. doi:10.1118/1.4709600.

[21] Tustison, N. J., Yassa, M. A., Rizvi, B., Cook, P. A., Holbrook, A. J., Sathishkumar, M. T., Tustison, M. G., Gee, J. C., Stone, J. R., & Avants, B. B. (2024). ANTsX neuroimaging-derived structural phenotypes of UK Biobank. Scientific Reports, 14(1), 8848. doi:10.1038/s41598-024-59440-6.

[22] Tustison, N. J., Cook, P. A., Holbrook, A. J., Johnson, H. J., Muschelli, J., Devenyi, G. A., Duda, J. T., Das, S. R., Cullen, N. C., Gillen, D. L., Yassa, M. A., Stone, J. R., Gee, J. C., & Avants, B. B. (2021). The ANTsX ecosystem for quantitative biological and medical imaging. Scientific Reports, 11(1), 9068. doi:10.1038/s41598-021-87564-6.

[23] Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., & Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. NeuroImage, 54(3), 2033–2044. doi:10.1016/j.neuroimage.2010.09.025.

[24] Razlighi, Q. R., Kehtarnavaz, N., & Yousefi, S. (2013). Evaluating similarity measures for brain image registration. Journal of Visual Communication and Image Representation, 24(7), 977–987. doi:10.1016/j.jvcir.2013.06.010.