Multimodal MRI Deformable Registration via Unsupervised Learning and Mutual Information

[1]*Kuan-Hua Chao* (趙冠華), [2,3]*Feng-Shiang Cheng* (鄭鳳翔), [1,*]*Chiou-Shann Fuh* (傅楸善)

[1]Department of Computer Science and Information Engineering,
National Taiwan University, Taipei, Taiwan,
[2]Department of Education and Research, Taipei City Hospital, Taipei, Taiwan,
[3]University of Taipei, Taipei, Taiwan
E-mail: chaonesskh@gmail.com a3989@tpech.gov.tw fuh@csie.ntu.edu.tw

ABSTRACT

Medical image registration is pivotal for enhancing diagnostic and therapeutic efficacy in healthcare. This study explores non-rigid registration methods for brain MRI, focusing on integrating Mutual Information (MI) with the learning-based VoxelMorph framework to handle multimodal data. The research evaluates traditional methods, VoxelMorph with Mean Squared Error (MSE) and Mutual Information Loss (MI Loss), and SynthMorph, employing the Dice similarity coefficient (DSC) for performance assessment. Results indicate that VoxelMorph with MSE excels in unimodal scenarios but exhibits potential overfitting in multimodal contexts. In contrast, VoxelMorph with MI Loss shows improved adaptability to multimodal data. SynthMorph demonstrates robust cross-modality registration but lacks the adaptability seen in VoxelMorph with MSE in atlas-specific tasks. The study also reveals the significance of lambda tuning in optimizing registration outcomes, with implications for future research in adaptive regularization within deep learning frameworks. The findings suggest the need for refined hyperparameter tuning strategies to develop advanced registration models capable of complex multimodal image analysis.

Keywords: *Medical Image Processing, Multimodal MRI, Deformable Registration, Unsupervised Learning, Mutual Information*

1. INTRODUCTION

Medical image registration is a critical procedure in healthcare, unifying diverse images into a coherent coordinate system to enhance diagnostic accuracy and therapeutic planning. This fundamentally geometric process adjusts a 'moving image' to conform to a 'fixed' reference [1]. Traditional methods involve manual feature identification, which, while effective, is laborious and prone to error [2]. Mutual Information (MI)-based techniques have risen as a superior, automated solution, particularly advantageous for multimodal contexts like Computed Tomography (CT) and MRI scans, by quantifying shared information to refine alignment processes.

This domain divides into rigid registration, maintaining morphology through translations and rotations, and non-rigid registration, which adapts to morphological variations. The advent of learning-based non-rigid registration, particularly integrated with MI for multimodal imaging, marks a pivotal shift. This study showcases the implementation of traditional MI within non-rigid registration [3], employing the VoxelMorph framework [6]—an unsupervised learning model, augmented for multimodal compatibility through MI-based image loss functions, and references SynthMorph's zero-shot learning [7] for cross-modality task comparison but is not the central theme of this exploration.

VoxelMorph has significantly advanced Deformable Image Registration (DIR), a task with profound implications in clinical settings such as aligning MRI or CT image pairs. Where discrete optimizations for individual image pairs characterized traditional DIR, VoxelMorph introduces a parametric function that yields a deformation field through a single streamlined computation, significantly expediting the registration process and facilitating global optimization over entire datasets.

This introduction sets the stage for an in-depth review of registration workflows, underscoring the role of entropy and mutual information. It also introduces the MRI brain registration application, which is central to this study. The subsequent sections detail the implementation of MI in non-rigid frameworks, VoxelMorph's augmentation with MI, and SynthMorph's benchmarking. The experimental analysis evaluates four non-rigid registration methods, with a focus on the Dice similarity coefficient as the assessment metric:

1. **Traditional techniques**, providing a baseline for comparison.
2. **Original VoxelMorph**, trained for adaptability via scan-to-scan and scan-to-atlas, albeit limited to single-modality.
3. **Enhanced VoxelMorph**, integrating MI to extend its application to multimodal data while maintaining foundational training strategies.
4. **SynthMorph**, evaluated for cross-modality registration, serving as a comparative benchmark.

The methods are rigorously tested on single-modality MRI T1 datasets from the Open Access Series of Imaging Studies (OASIS) [8, 9] and augmented by synthesized cross-modality data for comprehensive multimodal experimentation. The OASIS dataset provides a comprehensive cross-sectional collection of MRI brain data, representing a diverse demographic array of young, middle-aged, non-demented, and demented older adults. This comparative study is two-fold: to underscore the distinct strengths inherent in each registration approach and to investigate the influence of varying training strategies on the efficacy of sophisticated machine learning models within the domain of medical image registration. The inclusion of simulated multimodality data enhances the robustness and applicability of the findings, ensuring that the models' performance is not limited to the nuances of single-modality imaging but extends to the challenges presented by multimodal medical data analysis.

2. BACKGROUND

2.1. Registration Workflow

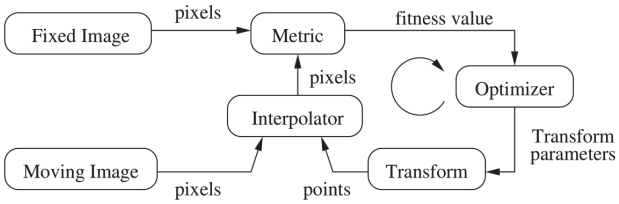


Fig. 1. Registration framework [4].

1. **Input Images**: The Fixed Image (reference image) is a static reference. The Moving Image transforms to align with the fixed one.
2. **Metric**: Pixels from both images are compared using Mutual Information (MI). MI measures shared information, which is ideal for multi-modal registration like CT and MRI.
3. **Interpolator**: As the moving image transforms, values at non-integer locations arise. The interpolator calculates these values using methods like nearest-neighbor or linear interpolation.
4. **Transform**: Defines how the moving image adjusts. Transformations can be rigid (translation, rotation), affine (shearing, scaling), or non-rigid (deformable). The initial parameters often derive from heuristics.
5. **Optimizer**: It tweaks transformation parameters to optimize MI between images. The process iterates to achieve the best MI "fitness value." It defines the best transformation once optimum or after a set iteration count.

Following this overview of the registration workflow, the review narrows its focus to delve into the Mutual Information metric and its pivotal role in image registration.

2.2. Multimodal Image Registration Based on Mutual Information

2.2.1. Joint Histogram

Given a transformation T and two images, I (moving image) and J (fixed image), the joint histogram represents the occurrences of intensity pairs from both images. Specifically, an entry at (a, b) in the joint histogram indicates the number of coordinates (u, v) where I_T (transformed moving image) has an intensity of a and J has an intensity of b. By dividing the joint histogram by the total number of pixels N, we obtain the joint probability density function $p_{I_T,J}$. The marginals, p_{I_T} and p_J, are obtained by summing over the rows or columns respectively.

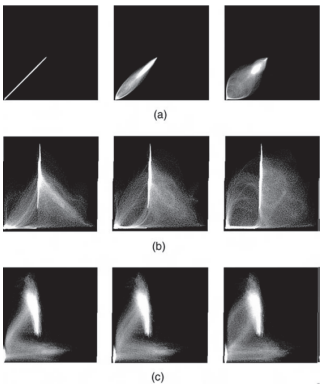


Fig. 2. Joint histograms demonstrating alignment across different modalities and levels of translation. From top to bottom: (a) Same modality (MRI-MRI), (b) Different modality (MRI-CT), (c) Different modality (MRI-PET). From left to right: increasing levels of translation. As the modality difference increases or the translation grows, the cluster in the joint histogram becomes more dispersed, providing a visual cue for alignment quality. [5].

2.2.2. Joint and Marginal Entropies

The joint entropy, $H(I_T, J)$, quantifies the uncertainty associated with the joint distribution and is given by:
$H(I_T, J) = -\sum_a \sum_b p_{I_T,J}(a, b) \log\left(p_{I_T,J}(a, b)\right)$.
The entropies for the individual images are:

$H(I_T) = -\sum_a p_{I_T}(a) \log\left(p_{I_T}(a)\right)$
and
$H(J) = -\sum_b p_J(b) \log\left(p_J(b)\right)$.
When the distributions p_{I_T} and p_J are similar, the joint entropy is typically less than the sum of their individual entropies. The objective is to identify the transformation that effectively minimizes this joint entropy, thereby aligning images with greater precision. A notable limitation of joint entropy lies in its relatively narrow capture range, particularly apparent during instances of significant misalignment. In conditions of pronounced misalignment, the joint histogram may retain a sharp profile, paradoxically resulting in lower joint entropy measures that can impede optimal registration effectiveness.

2.2.3. Mutual Information

Dependence$I(I_T, J)$, measures the statistical dependence between two images. It is expressed as: $I(I_T, J) = H(I_T) + H(J) - H(I_T, J)$ This definition is synonymous with the Kullback-Leibler divergence, reflecting the cost of assuming I_T and J to be independent. Maximizing mutual information ensures the best alignment between two images. This optimization task involves searching for a transformation T where I_T can be best predicted by J. Essentially if one knows the intensity $I_T(u, v)$, they can accurately predict $J(u, v)$.

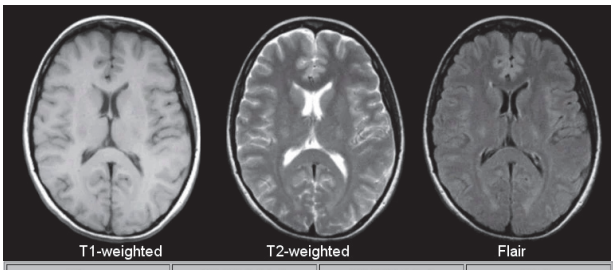
The joint entropy and mutual information provide a measure of similarity between two images by quantifying the statistical dependency between their pixel intensities. The process involves identifying a transformation that minimizes joint entropy or, conversely, maximizes mutual information, hence achieving optimal image alignment.

The cost function, denoted as $\theta^* = \underset{\theta}{\operatorname{argmin}}\, d(I_T, J)$, is integral to this alignment process, capturing the dissimilarity between a 'moving' image I_T and a 'fixed' reference image J. For mutual information, the cost function is the negative mutual information, $d(I_T, J) = -I(I_T, J)$,reflecting the objective to maximize mutual information and thus align the images more precisely. Conversely, if using joint entropy as a metric, the cost function is set as $d(I_T, J) = H(I_T, J)$, aligning with the goal of minimizing joint entropy to improve registration accuracy.

The pixel-based similarity measure $d(I_T, J)$ further refines the alignment through iterative minimization of the cost function, employing optimization methods such as Gradient Descent, Gauss-Newton, Newton-Raphson, and Levenberg-Marquardt [5]. This process meticulously adjusts the transformation parameters θ using the gradients and Hessians derived from the similarity measure, facilitating a progressively more accurate registration with each iteration.

2.3. Applications of MRI Brain Registration Across Varied Imaging Modalities

Magnetic Resonance Imaging (MRI) offers a suite of modalities, each elucidating unique aspects of brain anatomy and pathology. T1-weighted imaging (T1WI) provides excellent contrast between white and grey matter, making it ideal for delineating brain structure. T2-weighted imaging (T2WI), on the other hand, accentuates fluid-containing structures, thus highlighting areas of pathology such as edema or demyelination. FLuid-Attenuated Inversion Recovery (FLAIR) sequences suppress free fluid signals, relieving lesions such as those found in multiple sclerosis or other inflammatory processes.



Tissue	T1-Weighted	T2-Weighted	Flair
CSF	Dark	Bright	Dark
White Matter	Light	Dark Gray	Dark Gray
Cortex	Gray	Light Gray	Light Gray
Fat (within bone marrow)	Bright	Light	Light
Inflammation (infection, demyelination)	Dark	Bright	Bright

Fig. 3. Comparison of T1, T2, and FLAIR (Brain) [10].

Intra-modal Registration is crucial for applications like monitoring disease progression using successive T1WI scans. Precise alignment reveals minute anatomical variations over time, aiding in assessing treatment efficacy or disease evolution.

Inter-modal Registration converges the distinctive insights from different MRI sequences. By aligning T1WI with T2WI or FLAIR, it allows for a composite analysis that leverages the high-resolution anatomical detail of T1WI with the pathological sensitivity of T2WI or FLAIR. This multimodal approach is particularly beneficial in preoperative planning and targeted therapies, where a comprehensive view of the neuroanatomy is paramount.

The challenge lies in the intrinsic differences in image characteristics between modalities. Effective registration techniques must accommodate these variations to ensure the spatial correspondence accurately aligns anatomical landmarks. Recent advancements in registration algorithms, particularly those that integrate machine learning and Mutual Information, have significantly enhanced the precision and utility of multimodal MRI analyses.

These registration strategies are foundational in routine clinical practice and support evolving applications such as multi-parametric modeling, contributing to the burgeoning field of personalized medicine. By fusing various MRI modalities, clinicians

and researchers can construct a more detailed and dynamic picture of brain health and disease.

3. METHODS

3.1. Traditional Non-Rigid Multimodality MRI Registration Framework

This study implements a non-rigid registration algorithm for brain MRI across different patient scans, building on the pivotal work by Rueckert et al. [3]. The method is characterized by a dual-component transformation model comprising global affine $\mathbf{T}_{Global}$ and local Free-Form Deformations (FFD) $\mathbf{T}_{Local}$:

$$\mathbf{T}(x', y', z') = \mathbf{T}_{Global}(x', y', z') + \mathbf{T}_{Local}(x', y', z'), \quad (1)$$

specifically using **B-Spline** transformations for the latter. The approach omits the global affine transformation $\mathbf{T}_{Global}$ as preprocessing with OASIS dataset already addresses this. The registration process is orchestrated via a Python script reliant on the SimpleITK [4] library, aligned with the methodological framework of Rueckert et al., as follows:

1. **Data Input**: The script reads in the fixed (I_{fixed}) and moving (I_{moving}) MR images and their respective segmentations. The fixed image is the registration benchmark throughout the process.
2. **Similarity Assessment**: Normalized Mutual Information (NMI) is the metric for gauging the coherence between the fixed and moving images, adept for handling multimodal discrepancies. The similarity measure is calculated as
$$C_{similarity}(A, B) = \frac{H(A)+H(B)}{H(A,B)}, \quad (2)$$
where H denotes the Shannon entropy of the images A and B, and $H(A, B)$ represents the joint entropy.
3. **Transformation Application**: A **B-Spline**-based non-linear registration, articulated as
$$\mathbf{T}_{Local}(x', y', z') = \sum_{l'=0}^{3}\sum_{m'=0}^{3}\sum_{n'=0}^{3} B_{l'}(u_2)B_{m'}(v_2)B_{n'}(w_2)\phi_{i+l',j+m',k+n'}, \quad (3)$$
which allows for the elastic deformation of the moving image to match the fixed one where the indices are derived as:
$$i = \left\lfloor \frac{x'}{n_x} \right\rfloor - 1,\ j = \left\lfloor \frac{y'}{n_y} \right\rfloor - 1,\ i = \left\lfloor \frac{z'}{n_z} \right\rfloor - 1. \quad (4)$$
And the parameters are: $u_2 = \frac{x'}{n_x} - \left\lfloor \frac{x'}{n_x} \right\rfloor, v_2 = \frac{y'}{n_y} - \left\lfloor \frac{y'}{n_y} \right\rfloor, w_2 = \frac{z'}{n_z} - \left\lfloor \frac{z'}{n_z} \right\rfloor$.
The mesh ϕ stands for uniformly spaced control points arrayed in a $n_x \times n_y \times n_z$ matrix.
4. **Optimization and Resampling**: The algorithm refines transformation parameters by minimizing the cost function
$$C(\mathbf{T}) = -C_{similarity}\left(I_{fixed}, \mathbf{T}(I_{moving})\right) + \lambda C_{smooth}(\mathbf{T}), \quad (5)$$
where $C_{smooth} = \int\int\int_{V_m} \left(\frac{\delta \mathbf{T}}{\delta x'^2}\right)^2 + \left(\frac{\delta \mathbf{T}}{\delta y'^2}\right)^2 + \left(\frac{\delta \mathbf{T}}{\delta z'^2}\right)^2 + 2\left[\left(\frac{\delta^2 \mathbf{T}}{\delta x' y'}\right)^2 + \left(\frac{\delta^2 \mathbf{T}}{\delta y' z'}\right)^2 + \left(\frac{\delta^2 \mathbf{T}}{\delta x' z'}\right)^2\right]$ denotes the smoothness constraint that penalizes the second-order spatial derivatives of the transformation, promoting a smooth deformation field. Gradient descent and multi-resolution strategies are employed for resampling the moving segmentation, ensuring its precise alignment with the fixed segmentation. The balance between the similarity and smoothness components of the cost function, governed by the weight parameter λ, is key to accommodating the morphological variations between the fixed and moving images over time.
5. **Validation Metric**: Post-registration, the Dice similarity coefficient quantifies the overlap between the fixed and moved segmentations, directly indicating registration fidelity; higher values denote superior alignment.

3.2. VoxelMorph for Unsupervised Unimodality MRI Registration

VoxelMorph [6] redefines the DIR task, traditionally split into two stages: initial affine transformations addressing gross movements and rotations, followed by fine non-linear adjustments. The focus herein is on the latter stage, employing pre-processed images that have already undergone initial affine adjustments. This allows for a detailed capture of complex, non-linear deformations pertinent to medical imaging, where organ and tissue movements are not rigid but free-form, influenced by various physiological factors.

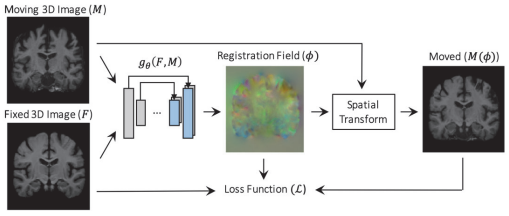


Fig. 4. VoxelMorph architecture [6].

VoxelMorph formulates the Deformable Image Registration (DIR) problem as a parametric function $g_\theta(F, M) = \boldsymbol{u}$ that maps an input pair of images to a displacement field $\boldsymbol{u}$, indicating the transformation from a moving image M to a fixed image F. This approach shifts the paradigm from optimizing each image pair separately to a global, dataset-level optimization, thus learning parameters θ across the entire dataset. It utilizes a deep learning model that requires minimal data for training and incorporates U-Net architecture to capture image features precisely. This flexibility extends to both

2D and 3D images, potentially including supervised losses if auxiliary data for images are available.

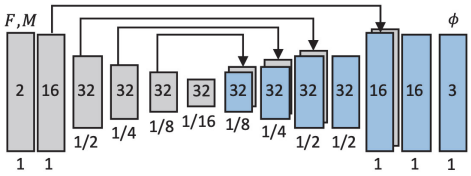


Fig. 5. UNet architecture [6].

VoxelMorph utilizes a UNet architecture that directly outputs the registration field Φ as $\Phi = Id + \boldsymbol{u}$, ensuring that the identity transformation Id is preserved where no displacement occurs. The architecture is designed such that f and m are of the same dimensions, and the model outputs a displacement vector $\boldsymbol{u}$ which, when added to Id, results in the field Φ. This setup effectively captures both the channel-wise features and the scale of inputs, with the application of a 3×3 kernel and a stride of 2 as the default configuration for the convolutional layers. A LeakyReLU activation function with an alpha value of 0.2 follows each convolution layer. The UNet structure ensures that localized, low-level image features captured by the Encoder are seamlessly integrated with the upsamples and refined features of the Decoder via skip connections. This approach allows for precise localization of features and guarantees the quality of the feature map.

Spatial Transformer Networks (STN), an essential aspect of the VoxelMorph framework, facilitate the transformation of the moving image M by applying the computed displacement field Φ. This process is crucial for verifying the image registration accuracy against the fixed image F. STN achieves this by interpolating pixel values from the closest neighbors, thus preserving the image features' inherent continuity and natural flow. The transformation for each voxel $\boldsymbol{p}$ is determined by the equation:

$$M \circ \phi(\boldsymbol{p}) = \sum_{q \in \mathcal{Z}(\boldsymbol{p}')} M(\boldsymbol{q}) \prod_{d \in (x,y,z)} (1 - |\boldsymbol{p}'_d - \boldsymbol{q}_d|), \quad \boldsymbol{p}' = \boldsymbol{p} + \boldsymbol{u}(\boldsymbol{p}) \tag{6}$$

Here, the interpolation considers the nearest voxel values $\boldsymbol{q}$, adjusting for the displacement $\boldsymbol{u}(\boldsymbol{p})$ to ensure that the modified image $M \circ \phi(\boldsymbol{p})$ is a weighted blend of neighboring pixels, maintaining image integrity.

The **loss function** in VoxelMorph supports unsupervised learning and can extend to supervised learning if anatomical segmentation data is available.

The **unsupervised loss**, $\mathcal{L}_{us}(F, M, \phi)$, consists of a similarity term $\mathcal{L}_{sim}(F, M \circ \phi)$ and a regularization term $\mathcal{L}_{smooth}(\phi)$, with the former ensuring visual consistency between the transformed and fixed images and the latter controlling the smoothness of the displacement field. The Mean Squared Error (MSE) metric is typically employed for the similarity loss. The unsupervised loss is given by:

$$\mathcal{L}_{us}(F, M, \phi) = \mathcal{L}_{sim}(F, M \circ \phi) + \lambda \mathcal{L}_{smooth}(\phi) \tag{7}$$

$$MSE(F, M \circ \phi) = \frac{1}{|\Omega|} \sum_{\boldsymbol{p} \in \Omega} [F(\boldsymbol{p}) - [M \circ \phi](\boldsymbol{p})]^2 \tag{8}$$

$$\mathcal{L}_{smooth}(\phi) = \sum_{\boldsymbol{p} \in \Omega} \|\nabla \boldsymbol{u}(\boldsymbol{p})\|^2, \quad \nabla \boldsymbol{u}(\boldsymbol{p}) = \left(\frac{\partial \boldsymbol{u}(\boldsymbol{p})}{\partial x}, \frac{\partial \boldsymbol{u}(\boldsymbol{p})}{\partial y}, \frac{\partial \boldsymbol{u}(\boldsymbol{p})}{\partial z} \right) \tag{9}$$

This regularization is crucial for ensuring the displacement vectors reflect realistic physical properties of image transition, promoting spatially coherent mappings by controlling the rate of change in the displacement vector $\boldsymbol{u}$. The term $\nabla \boldsymbol{u}(\boldsymbol{p})$ represents the gradient of the displacement at voxel $\nabla \boldsymbol{u}(\boldsymbol{p})$, imposing smoothness by penalizing high gradients, thus reinforcing the natural behavior of the transformation across the image domain.

In the experimental setup described, despite the OASIS dataset containing segmentation data, the focus is on unsupervised learning. The inclusion of the supervised loss in the model formulation is to illustrate the potential for further accuracy enhancements. However, in this study, the comparisons are made solely based on the outcomes of unsupervised learning approaches.

3.3. Enhanced Multimodal MRI Registration via Vectorized MI

In advancing VoxelMorph's established efficacy in single-modality image registration, this study extends its application to multimodal contexts. This necessitates transcending the constraints of linearly assumed cost functions, such as Mean Squared Error (MSE) based similarity loss. To this end, Mutual Information (MI) is employed as a non-linear alternative that eschews such linear constraints. The implementation of MI is adeptly suited for the intricacies of multimodal data, enabling the capture of complex intensity relationships between diverse imaging modalities. This adaptation enhances VoxelMorph's versatility, aligning it with the nuanced demands of multimodal image registration.

The approach acknowledges that voxel contributions to histogram bins are not discrete but follow a continuous distribution, a concept brought to fruition through **Parzen windowing** [11]. This technique employs a Gaussian weighting function, parameterized by σ, to smoothly assign voxel intensities to histogram bins. The Gaussian function is defined as:

$$G(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2} \right) \tag{10}$$

where x is the voxel intensity, μ is the bin center, and σ is the standard deviation of the Gaussian function. It's an essential component in the computation of individual image intensity distributions and their joint histogram, facilitating the accurate estimation of mutual information.

The integration of MI into VoxelMorph leverages vectorization to enhance computational efficiency. The

approach is optimized for capturing the complex relationships inherent in multimodal data through global implementation.

Global Mutual Information (Global MI) vectorizes the entire image into intensity bins, applying Gaussian functions to create a continuous histogram. This involves:

1. **Reshape and Tile**: Images I_A and I_B are first flattened and tiled, establishing a voxel-to-bin relationship for the entire image.
2. **Gaussian Weighted Histograms**: Gaussian functions determine voxel contributions to intensity bins, yielding a smooth and differentiable histogram for individual and joint intensities. The Gaussian-weighted histograms are computed as:
$$H_A = G(I_A; \mu, \sigma)$$
$$H_B = G(I_B; \mu, \sigma) \quad (11)$$
where H_A and $H_B$represent the Gaussian-weighted histograms for images I_A and I_B, respectively.
3. **Normalization and Mean Calculation**: Gaussian-derived matrices H_A and H_B are normalized to form probability distributions P_A and P_B, from which marginal distributions $p_A(x)$ and $p_B(y)$ are extracted.
4. **Joint Distribution and MI Computation**: The joint distribution $p(x, y)$ is derived by the matrix product of the normalized matrices, and mutual information is computed by aggregating over the joint distribution, applying a logarithmic ratio against the product of the marginal distributions, with epsilon ϵ added to avert division by zero.
$$MI(I_A, I_B) = \sum_{x,y} p(x,y) \log\left(\frac{p(x,y)}{p_A(x) \cdot p_B(y) + \epsilon}\right) \quad (12)$$

The vectorized calculations of Global MI maintain the integrity of mutual information as a measure while enabling its use in a high-throughput, learning-based registration framework like VoxelMorph, thus making the registration process more computationally feasible and efficient.

4. EXPERIMENTS

4.1. Dataset & Evaluation Metrics

For a robust and realistic evaluation, we utilize brain T1 MRIs from the Open Access Series of Imaging Studies (OASIS) [8, 9], which includes a cross-sectional collection from young, middle-aged, nondemented, and demented older adults. This open-access dataset consists of 414 subjects and has undergone extensive preprocessing, such as intensity normalization, affine alignment, and skull stripping. These steps are crucial for emphasizing deformable registration and ensuring the precision of alignment with brain tissue.

Dataset Configuration: The OASIS dataset is divided into 353 training images, 30 validation images, and 30 testing images, all anchored to a single atlas image serving as a stable reference for the fixed image. Each subject in the dataset comes with a set of automated label segmentations, providing a 4-label tissue-type segmentation that includes the cortex, subcortical gray matter, white matter, and Cerebro-Spinal Fluid (CSF). Working with a multimodal dataset is essential to evaluate multi-modal image registration methods. Since an alternative modality atlas is unavailable, a synthetic modality is generated by inverting and scaling the intensity values of the original atlas within the positive range by a factor of 1.1.

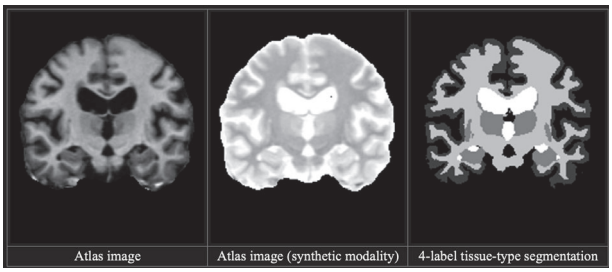


Fig. 6. The figure displays (from left to right): the original atlas image, the synthetic modality image with intensity values adjusted, and the segmentation image showing the 4-label tissue types.

Utilization of Anatomical Segmentations for Evaluation: Anatomical segmentations are employed as a rigorous metric for evaluation to address the limitations of mere visual inspection and ensure anatomically coherent results. The process involves using the trained UNet model to register fixed and test images, generating a deformation field. This field is then applied via a Spatial Transformer Network (STN) to produce registered test images and labels. Subsequently, the Dice Similarity Coefficient (DSC) between the registered test labels and fixed labels is calculated to assess registration accuracy quantitatively. This provides a measure of the spatial accuracy of the registration and ensures the anatomical plausibility of the deformation fields, thereby addressing the concerns of anatomical relevance neglected by visual assessments alone.

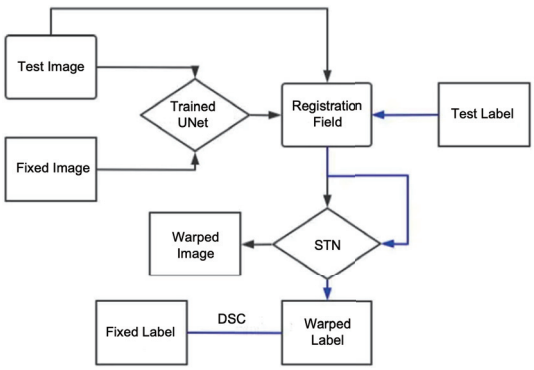


Fig. 7. Evaluation workflow.

4.2. Implementation and hyperparameter details

Traditional Non-Rigid Multimodality MRI Registration Framework:
For baseline comparisons, the registration framework is based on the Free-Form Deformation (FFD) or B-Spline transform utilizing the Normalized Mutual Information (NMI) metric. The SimpleITK [4] library's ImageRegistrationMethod() is configured with the following parameters:

1. **Similarity metric**: Mattes Mutual Information with 50 histogram bins.
2. **Sampling strategy**: Random sampling with a 10% voxel sample rate.
3. **Interpolator**: Linear.
4. **Optimizer**: Gradient Descent, fine-tuned with a learning rate of 1.0, 100 iterations, a convergence minimum value of 1e-6, and a convergence window size of 10.
5. **Optimizer Scales**: Derived from physical shifts to maintain relevance to the image space.
6. **Multi-resolution framework**: Configured with shrink factors of [4, 2, 1] and smoothing σ of [2, 1, 0], allowing for hierarchical refinement.
7. **Initial Transform**: Set using a B-Spline initializer with a mesh size based on the moving volume's dimensions.

VoxelMorph Training:
Implemented using TensorFlow Keras [12] and Neurite packages [13], VoxelMorph's training involves the ADAM optimizer with a learning rate of 10^{-4}, a batch size of 1, spanning 200 training epochs. Each model's loss function necessitates fine-tuning the regularization parameter, affecting the smoothness of the deformation field. Regularization weights λ are set at 0.01 for Mean Squared Error (MSE) and 2 for mutual information, with the latter deliberately chosen for later comparative analysis. While not optimized, this strategic selection of lambda weight for mutual information is intended for further discussion on hyperparameter tuning, which will be covered in subsequent sections, providing insight into the regularization weight's role in model performance.

Mutual information calculation:
With the calculation of Mutual Information, the study eschews the memory-intensive Local Mutual Information in favor of Global Mutual Information. The configuration is defined as follows:

1. **Bin Centers**: Not specified, allowing for flexible histogram shaping.
2. **Histogram Bins**: Limited to 16 due to memory constraints.
3. **Soft Bin Alpha**: Set at $\frac{1}{2*\sigma^2}$, with σ calculated relative to the bin count, integral to the soft quantization and MI calculation process.
4. **Data Clipping**: Normalization parameters are set to accommodate the full data range, with clipping thresholds set to negative and positive infinity.

SynthMorph Pretrained Weights:
SynthMorph's pretrained weights [7] are employed strictly for comparative evaluation, providing a benchmark to assess the performance enhancements of my implemented VoxelMorph models. This comparison aims to highlight the effectiveness of the model adjustments and hyperparameter optimizations in the context of multimodal MRI registration tasks.

4.3. Result

4.3.1. Evaluation of Image Registration Methods

The Dice Similarity Coefficient (DSC) is used to evaluate various non-rigid image registration methods to measure the overlap between the segmented cerebral structures post-registration. The DSC quantifies the similarity between the predicted and actual segmentations, with values ranging from 0 (no overlap) to 1 (perfect overlap). It is calculated as $\text{Dice}(P,T) = \frac{2|P_1 \wedge T_1|}{|P_1|+|T_1|}$ where P and T represent the predicted and true segmentation masks. Higher DSC values indicate better alignment.

This study assesses 30 test subjects by registering their images as moving images against either a standard atlas or a synthesized atlas used as fixed images. The results, detailed in Table 1, summarize the performance. Regularization weights (λ) are set at 0.01 for Mean Squared Error (MSE) and 2 for Mutual Information (MI). The MI weight is selected for initial testing without fine-tuning to provide a baseline for later comparisons of regularization effects.

Table 1: Comparison of Non-Rigid Image Registration Methods Based on DSC (↑ indicates higher is better; **without fine-tuning**: $\lambda = 2$ for MI loss).

Method	Train Method	Test on Atlas	Test on Synth Atlas
Traditional	-	0.6239 +/- 0.0495	0.6326 +/- 0.0352
VoxelMorph w/ MSE Loss	scan-to-scan	0.7947 +/- 0.0196	0.3763 +/- 0.0148
	scan-to-atlas	**0.8067 +/- 0.0178**	0.3256 +/- 0.0156
	train on synth	0.5955 +/- 0.0270	0.4229 +/- 0.0016
VoxelMorph w/ MI Loss (ours)	scan-to-scan	0.7615 +/- 0.0268	0.4069 +/- 0.0128
	scan-to-atlas	**0.7718 +/- 0.0237**	0.4625 +/- 0.0235
	train on synth	0.5864 +/- 0.0344	**0.7291 +/- 0.0286**
SynthMorph	-	0.7625 +/- 0.0278	**0.7634 +/- 0.0321**

1. **Scan-to-scan**: A random pair from the training data is used as the fixed and moving images.
2. **Scan-to-atlas**: An atlas image is designated as the fixed image and paired with each piece of training data as the moving image.
3. **Train on synth**: A synthetic atlas is used as the fixed image, pairing with each training datum as the moving image.

The outcomes indicate:

1. **VoxelMorph with MSE Loss**: showcases optimal performance when trained on atlas images and tested on atlas images, possibly revealing a bias towards the atlas used during training. Notably, its performance markedly declines when evaluated on synthetic atlas images, suggesting potential overfitting to the training modality, as even training on synthetic data does not bolster accuracy, indicating a lack of multimodality training capability.
2. **VoxelMorph with MI Loss**: reveals a minor decrease in DSC scores on atlas testing compared to MSE Loss but exhibits a pronounced improvement in scores on synthetic atlas images. This suggests an augmented capacity to accommodate multimodality data.
3. **Traditional Methods**: maintain steady performance across both test atlas and synthetic atlas images, indicating their competence in addressing multimodality challenges.
4. **SynthMorph**: attains the highest DSC on synthetic atlas images, reflecting its substantial multimodality proficiency. However, it falls short of reaching peak scores seen in VoxelMorph with MSE Loss during atlas-specific training, signaling potential areas for enhancement in model versatility.

4.3.2. Lambda Tuning Sensitivity

The sensitivity analysis of the regularization parameter λ in VoxelMorph with Mutual Information (MI) Loss highlights its significant impact on model performance. The key findings are summarized in Table 2.

Table 2. Sensitivity Analysis of Lambda Values Based on DSC (↑) for VoxelMorph with MI Loss.

Lambda	Train Atlas, Test Atlas	Train Atlas, Test Synth	Train Synth, Test Synth
2	0.7718 +/- 0.0237	**0.4625 +/- 0.0235**	0.7291 +/- 0.0286
1.5	0.7878 +/- 0.0207	0.4044 +/- 0.0190	0.7516 +/- 0.0244
1	0.8088 +/- 0.0188	0.3668 +/- 0.0142	0.7699 +/- 0.0206
0.5	0.8326 +/- 0.0175	0.4076 +/- 0.0141	0.7979 +/- 0.0167
0.1	**0.8512 +/- 0.0152**	0.4315 +/- 0.0141	**0.8118 +/- 0.0155**
0.01	0.8469 +/- 0.0131	0.4286 +/- 0.0165	0.7655 +/- 0.0155

1. **Atlas Training and Testing**: Setting λ to 0.1 improves the DSC by 0.08 compared to λ of 2, even surpassing the DSC of 0.8067 +/- 0.0178 achieved with MSE loss.
2. **Synth Atlas Training and Testing**: A λ of 0.1 enhances the DSC by 0.09, outperforming SynthMorph (0.7634 +/- 0.0321), signifying the importance of a well-tuned regularization parameter.
3. **General Trend**: Reducing lambda from 2 to 0.1 generally benefits the DSC for both Atlas and Synth testing scenarios. However, further reducing λ to 0.01 adversely impacts DSC, suggesting an ideal range for λ exists, not merely the smallest value.
4. **Complex Patterns**: The **Train Atlas, Test Synth** scenario exhibits the lowest DSC at λ of 1, with improvements seen with both an increase and a decrease in λ, underscoring a nuanced relationship between λ and performance that warrants detailed exploration.

The findings indicate that lower λ values can enhance the model's performance up to a point, after which further reduction is detrimental. This non-linear effect of λ underscores the need for careful experimentation to determine the optimal λ for each registration task. Future research should explore methods for the model to autonomously fine-tune λ, which could offer a significant advancement in medical image registration.

4.3.3. Visualization and Case Analysis

Figure. 8 in the Appendix showcases various outcomes of image registration experiments, comparing Mutual Information (MI) and Mean Squared Error (MSE) methods across different training and testing scenarios. For MI, successful cases are evident when trained and tested on atlas images, as well as in synthetic data scenarios, demonstrating robust alignment and effective capture of shared information between scan and atlas images. Conversely, MI struggles with direct scan-to-scan registration on synthetic atlases without alignment training, resulting in poorer outcomes. MSE, however, consistently faces challenges across synthetic data, highlighting its inadequacy in handling variations and difficulties in maintaining brain morphology. Overall, MI proves superior to MSE in preserving brain structure integrity and achieving accurate image alignment, particularly when transitioning between training on atlas and testing on synthetic datasets.

5. CONCLUSION

This study compares non-rigid brain MRI registration methods: traditional approaches, VoxelMorph (with MSE and MI losses), and SynthMorph. The Dice similarity coefficient (DSC) is used to evaluate accuracy on both standard and synthetic atlas images. VoxelMorph with MSE loss excels on atlas images but struggles with synthetic data, indicating overfitting. Using MI loss, VoxelMorph better handles multimodal data. Traditional methods consistently perform well, with Normalized Mutual Information proving effective for multimodal registration. SynthMorph achieves the highest DSC on synthetic data, excelling in cross-modality registration, although it falls short on atlas-specific tasks compared to VoxelMorph with MSE loss. Optimal tuning of λ is crucial for enhancing model performance, underscoring the importance of precise calibration for transformation field smoothness. Future work should focus on adaptive regularization techniques.

6. REFERENCES

[1] J.P.W. Pluim, and D. Mattes, "Mutual Information Based Registration of Medical Images," Available: https://courses.cs.washington.edu/courses/cse577/11au/notes/Z5-Mutual-Information.pdf.

[2] D. Sengupta, P. Gupta, and A. Biswas, "A survey on mutual information based medical image registration algorithms," *Neurocomputing*, Vol. 486, pp. 174-188, 2022.

[3] D. Rueckert, L.I. Sonoda, C. Hayes, D.L. Hill, M.O. Leach, and D.J. Hawkes, "Nonrigid registration using free-form deformations: application to breast MR images," *IEEE Transactions on Medical Imaging*, Vol. 18, No. 8, pp. 712-721, 1999.

[4] H.J. Johnson, M.M. McCormick, L. Ibanez, and J. Cates, *Template: The ITK Software Guide Book 1: Introduction and Development Guidelines-Volume 1*, Kitware Inc., 2015.

[5] J.V. Hajnal, and D.L.G. Hill, *Medical Image Registration*, CRC Press, 2001.

[6] G. Balakrishnan, A. Zhao, M.R. Sabuncu, J. Guttag, and A.V. Dalca, "VoxelMorph: a learning framework for deformable medical image registration," *IEEE Transactions on Medical Imaging*, Vol. 38, No. 8, pp. 1788-1800, 2019.

[7] M. Hoffmann, B. Billot, D.N. Greve, J.E. Iglesias, B. Fischl, and A.V. Dalca, "SynthMorph: learning contrast-invariant registration without acquired images," *IEEE Transactions on Medical Imaging*, Vol. 41, No. 3, pp. 543-558, 2021.

[8] A. Hoopes, M. Hoffmann, D.N. Greve, B. Fischl, J. Guttag, and A.V. Dalca, "Learning the Effect of Registration Hyperparameters with HyperMorph," *Medical Image Learning and Biomedical Applications (MELBA)*, 2022.

[9] D.S. Marcus, T.H. Wang, J. Parker, J.G. Csernansky, J.C. Morris, and R.L. Buckner, "Open Access Series of Imaging Studies (OASIS): Cross-Sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults," *Journal of Cognitive Neuroscience*, Vol. 19, pp. 1498-1507.

[10] "Magnetic Resonance Imaging (MRI) of the Brain and Spine: Basics," Available: https://case.edu/med/neurology/NR/MRI%20Basics.htm.

[11] W.M. Wells III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information," *Medical Image Analysis*, Vol. 1, No. 1, pp. 35-51, 1996.

[12] F. Chollet, et al., "Keras," 2015. Available: https://keras.io.

[13] M. Hines, A.P. Davison, and E. Muller, "NEURON and Python," *Frontiers in Neuroinformatics*, Vol. 3, p. 391, 2009.

7. APPENDIX

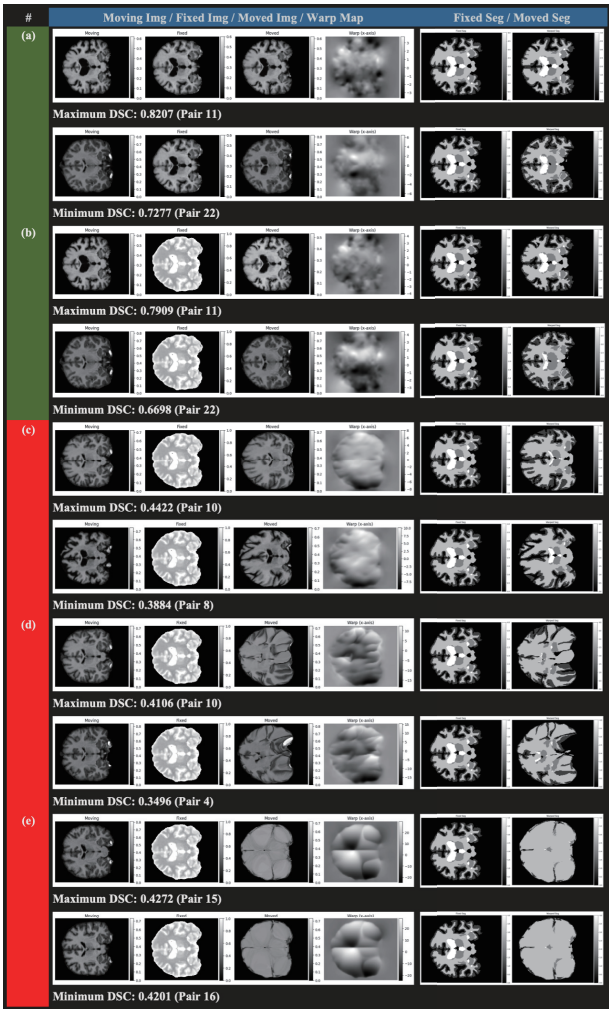


Fig. 8. The figure illustrates the outcomes of different image registration experiments, highlighting both successful and failed cases. It presents a series of moving images, fixed images, the resulting moved images, and warp maps. Additionally, it compares fixed and moved segmentations across various configurations.

Successful Cases:

(a) Mutual Information (MI) Scan-to-Atlas: Training on Atlas, Testing on Atlas;

(b) Mutual Information (MI) Scan-to-Atlas: Training on Synth, Testing on Synth

Failed Cases:

(c) Mutual Information (MI) Scan-to-Scan: Testing on Synth Atlas;

(d) Mean Squared Error (MSE) Scan-to-Scan: Testing on Synth Atlas;

(e) Mean Squared Error (MSE) Scan-to-Atlas: Training on Synth, Testing on Synth.