Bios 740- Chapter 10. Image Registration

Acknowledgement: Many thanks to Mr. Mingchen Hu for preparing some of these slides. I also drew on material from Dr. Gang Li's presentation and the lecture presentations of StanfordCS231n as well as content generated by ChatGPT.

Content

1. Introduction to Image Registration
2. ConvNets based Registration
3. Network Architectures for Registration
4. Applications of Image Registration

Content

1. Introduction to Image Registration

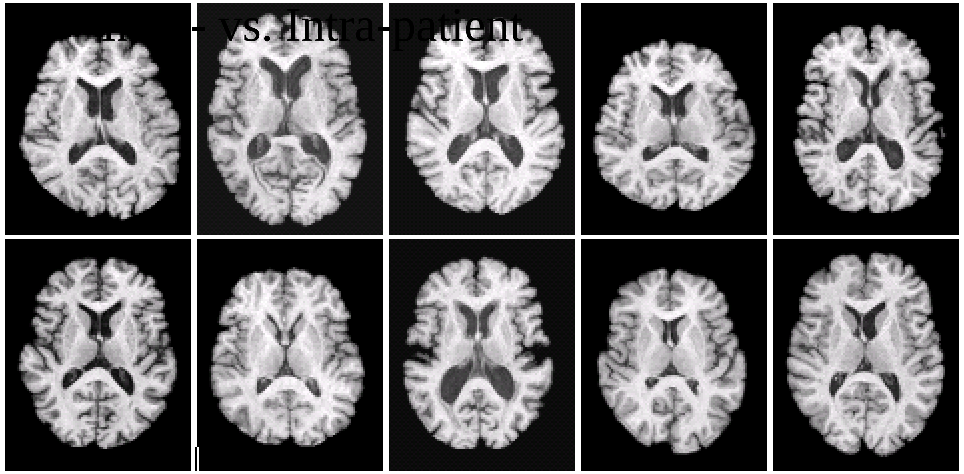
2. ConvNets based Registration

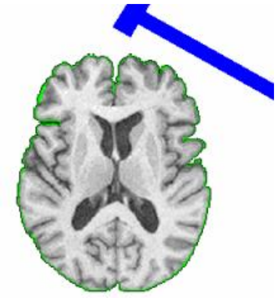
3. Network Architectures for Registration

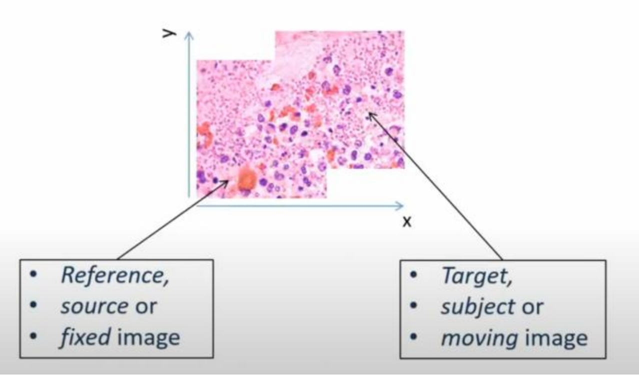
4. Applications of Image Registration

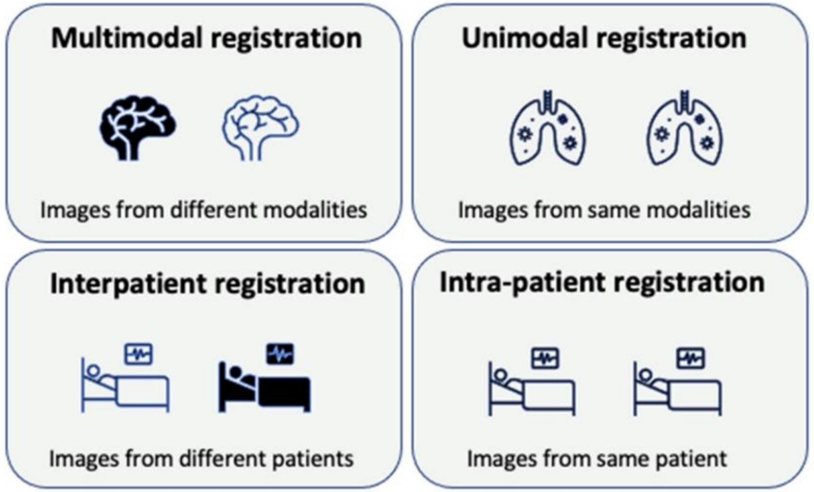
Image Registration

- **Definition:** Image registration is the process of aligning two or more images into a common coordinate system so that transformed images are similar to each other.
- **Applications:**
  - Medical imaging (e.g., MRI to CT alignment, longitudinal studies, tumc
  - Remote sensing (e.g., satellite image change detection)
  - Object tracking and video stabilization
  - Augmented reality and autonomous navigation
- **Key Types:**
  - Rigid vs. Non-rigid
  - Intensity-based vs. Feature-based
  - Intra-modal vs. Inter-modal
  - In[illegible] vs. Intra-patient









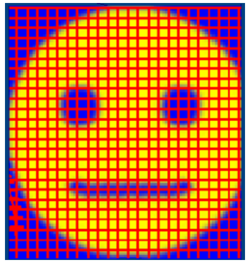
Darzi, F. and Bocklitz, T., 2024. A review of medical image registration for different modalities. *Bioengineering*, *11*(8), p.786.

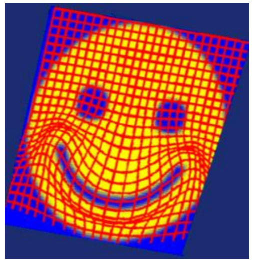
Registration vs. Other Image Transformation

- **Image Registration**:
  - Aligns images spatially using geometric transformations (e.g., translation, rotation, deformation).
  - Requires modeling spatial correspondences and often uses optimization.
  - Aims to overlay structures between images.
- **Other Image Transformations**:
  - Include operations like contrast enhancement, histogram equalization, filtering.
  - Do not alter spatial coordinates of pixels.
  - Aim to improve image quality or extract visual features.
- **Key Difference**: Registration manipulates image geometry to match another image; other transformations adjust pixel intensities or features without spatial alignment.

$$\tilde{f}(\tilde{x}) = f(T(\tilde{x})), \quad \text{for all } x \in \Omega$$

when $\tilde{x}' = T(\tilde{x})$ is a one-to-one transformation of $\tilde{x}$.

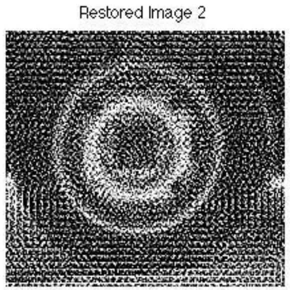




$$\tilde{f}(i,j) = T[f(i,j)]$$

$$\tilde{f}(\tilde{x}) = T[f(\tilde{x})], \quad \text{for all } \tilde{x} \in \Omega$$

when $T[\tilde{y}]$ is a monotonic function of $\tilde{y}$.

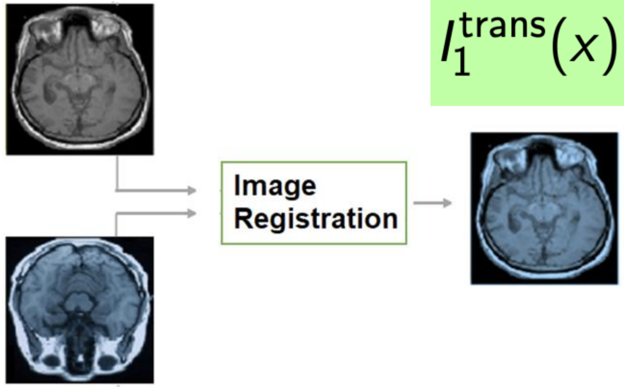


Key Components of Registration

- Let $I_1 : \Omega \to \mathbb{R}^d$ be the moving image.
- Let $I_2 : \Omega \to \mathbb{R}^d$ be the fixed/reference image.
- d indicates the number of channels (e.g., $d = 3$ for RGB).
- $\Omega \subseteq \mathbb{R}^n$ is the image domain.

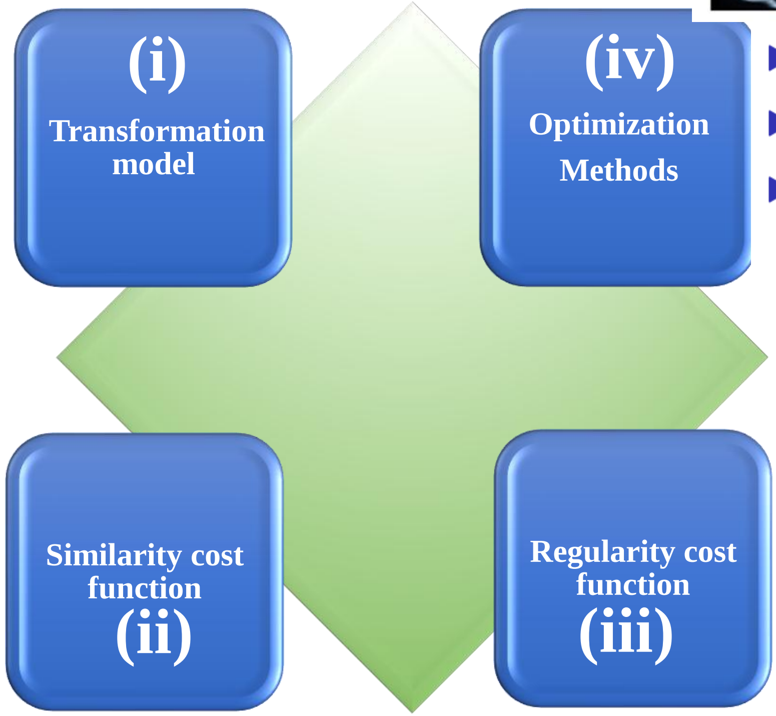
$I_2 : \Omega \to \mathbb{R}^d$

$I_1 : \Omega \to \mathbb{R}^d$



$I_1^{\text{trans}}(x) = I_1(\varphi(x))$

- $y \in \Omega$: coordinates in the moving image $I_1 = \{I_1(y), y \in \Omega\}$.
- $x \in \Omega$: coordinates in the fixed image (reference system) $I_2 = \{I_2(x), x \in \Omega\}$.
- $\varphi : \Omega \to \Omega$: push-forward (Lagrangian) transformation aligning I_1 to I_2 such that $y = \varphi(x)$.
- $\varphi^{-1} \equiv h$: pull-back transformation such that $x = \varphi^{-1}(y)$.



- Goal: Find φ^* that minimizes $\mathcal{L}(\varphi)$.
- Methods: gradient descent, Gauss-Newton, L-BFGS, etc.
- Often uses multiresolution strategies.

$$\varphi^* = \arg\min_{\varphi} \mathcal{L}(\varphi)$$

$$\mathcal{E}(\varphi) = \mathcal{S}(I_1(\varphi(x)), I_2(x)) + \lambda \mathcal{R}(\varphi)$$

Similarity Cost Function:

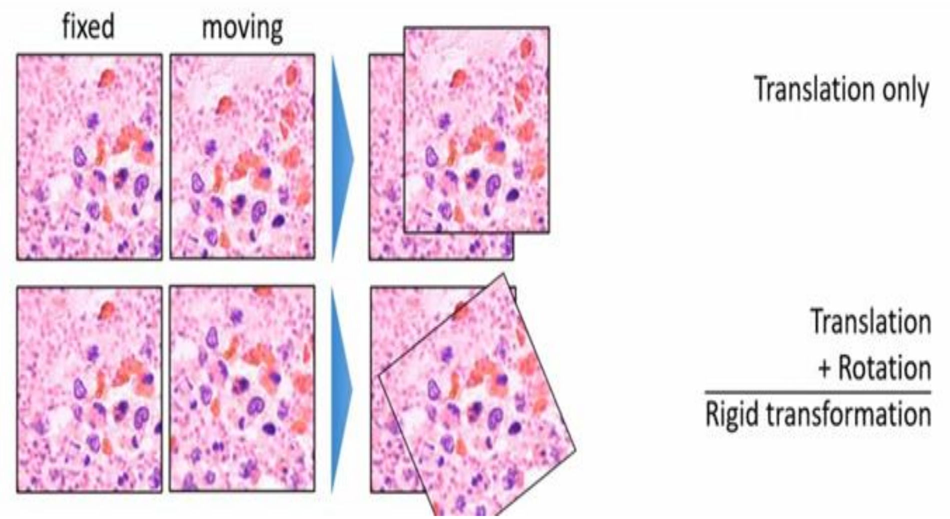
- Measures alignment: e.g., SSD, Mutual Information
- $\mathcal{S}(I_1(\varphi(x)), I_2(x))$

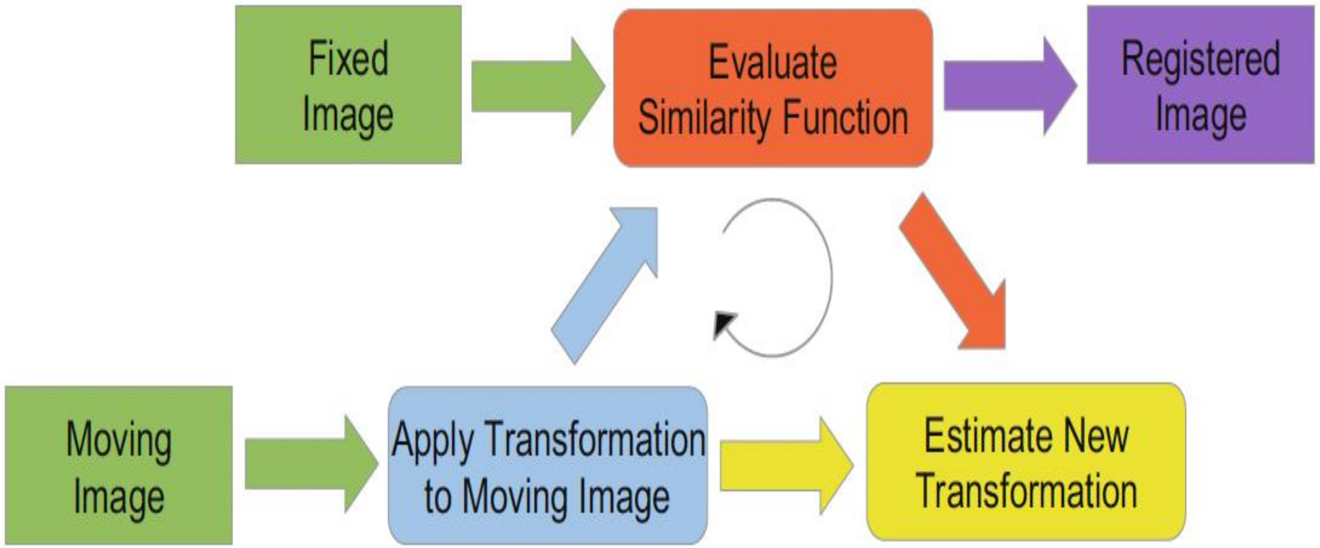
Regularity Cost Function:

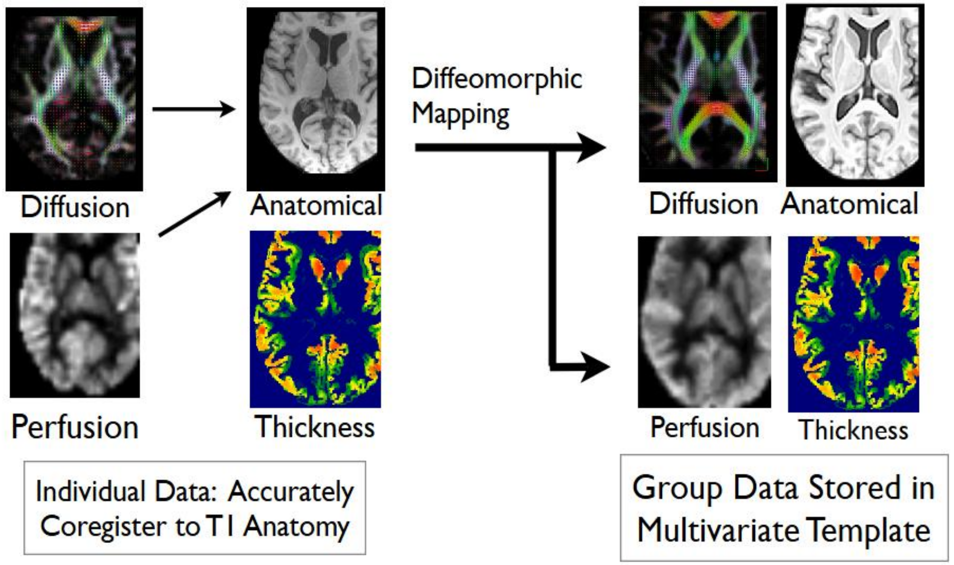
- Imposes smoothness or topology preservation
- $\mathcal{R}(\varphi)$

Four Key Questions of Registration

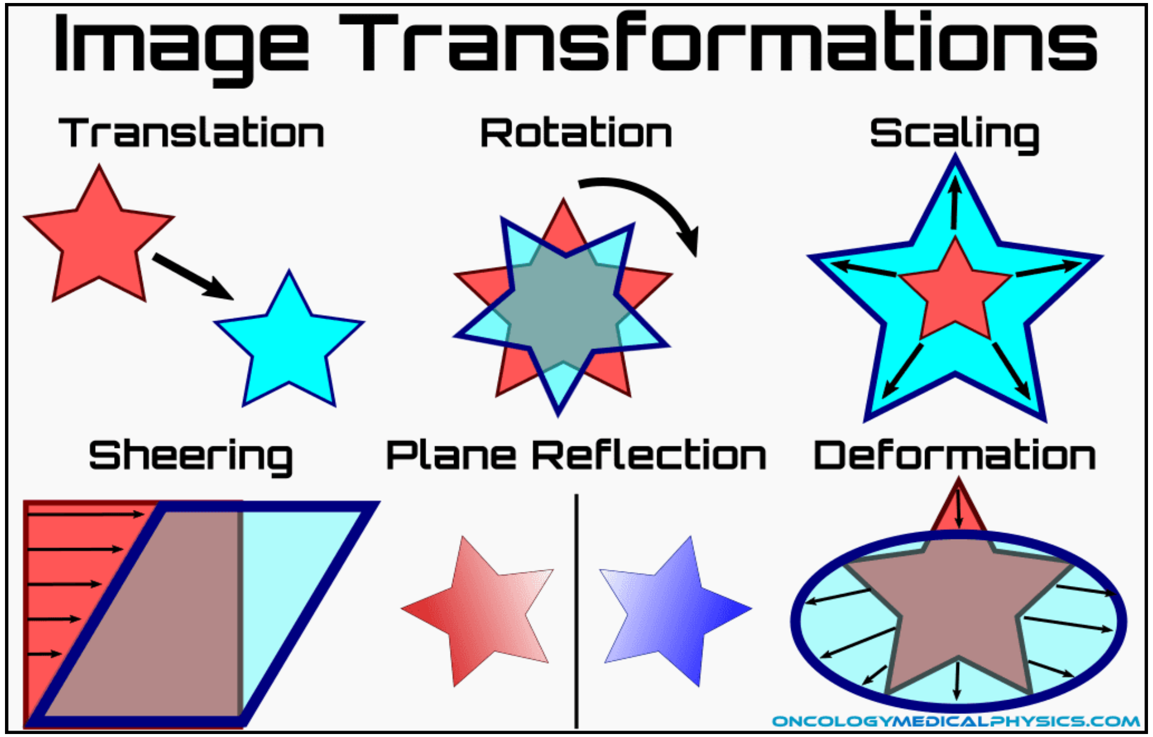
- **What is a transformation model** $\varphi : \Omega \rightarrow \Omega$ **?**
- **What is the similar cost function** $\mathcal{S}(I_1(\varphi(x)), I_2(x))$
- **What is the reasonability/regularity function** $\mathcal{R}(\varphi)$
- **How to optimize** $\varphi^* = \arg\min_{\varphi} \mathcal{L}(\varphi)$ **?**







Parametrized Transformations



https://oncologymedicalphysics.com/image-registration/

$$\text{scaling} \quad \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

$$\text{translation} \quad \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

$$\text{shear} \quad \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & u_x & 0 \\ u_y & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

$$\text{rotation} \quad \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

$$\text{general affine} \quad \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}.$$

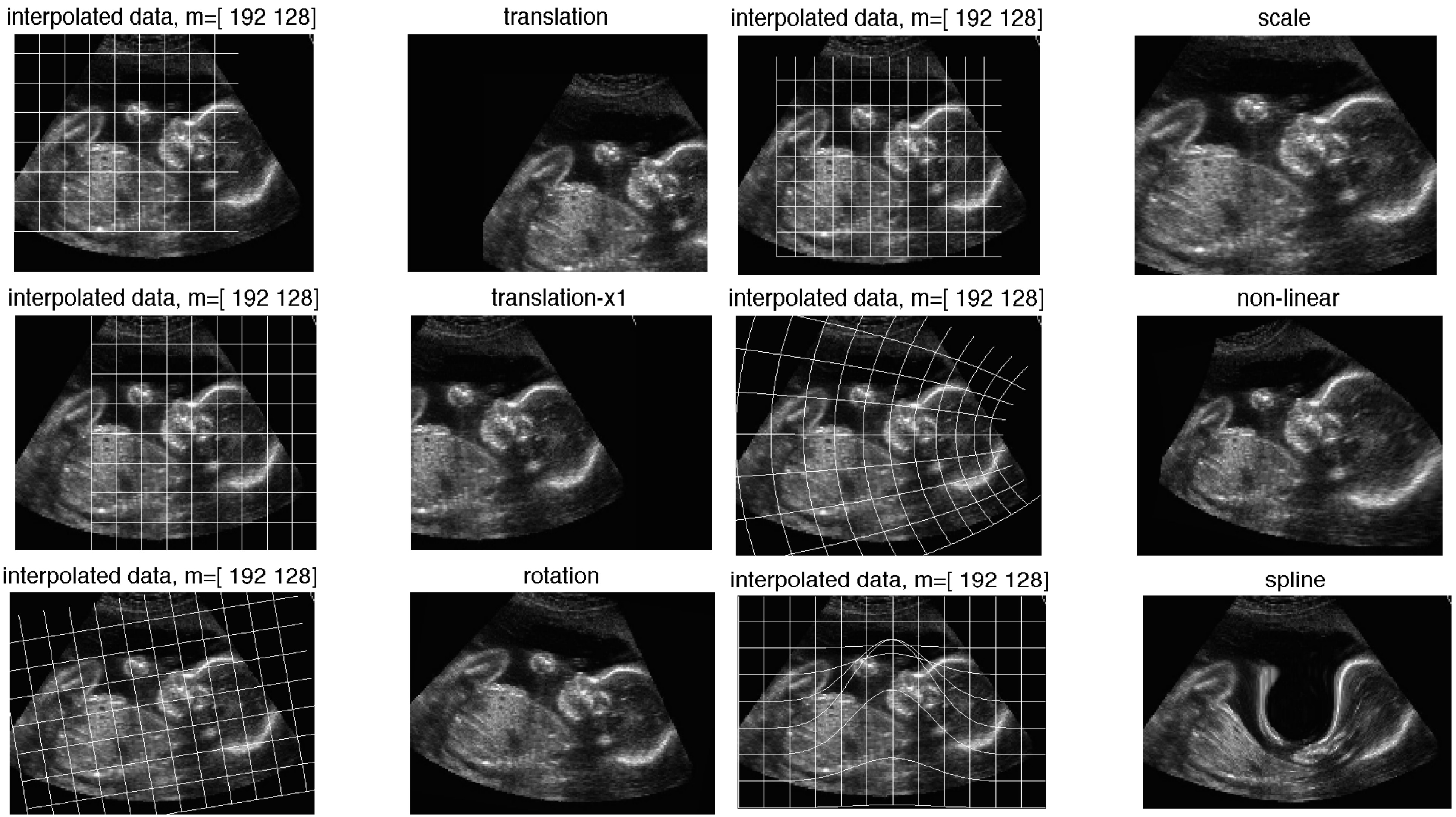
$$\textit{Rotation around x axis} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}$$

$$\textit{Rotation around y axis} \quad \begin{bmatrix} a_{11} & 0 & a_{13} \\ 0 & 1 & 0 \\ a_{31} & 0 & a_{33} \end{bmatrix} = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}$$

$$\textit{Rotation around z axis} \quad \begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

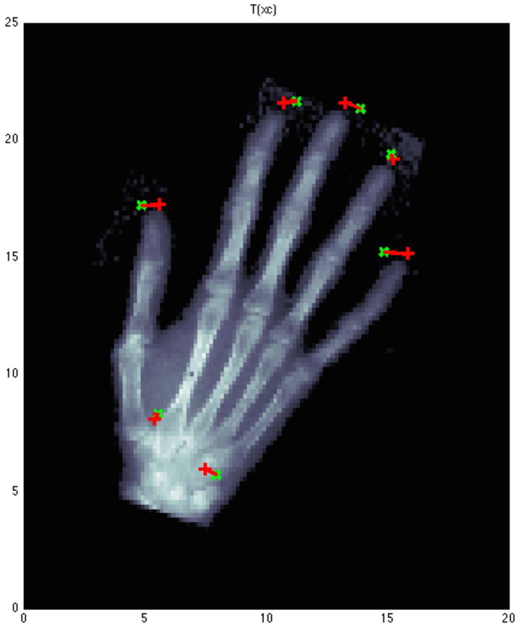
$$\text{general affine} \quad \begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & t_x \\ a_{21} & a_{22} & a_{23} & t_y \\ a_{31} & a_{32} & a_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}.$$

Parametrized Transformations: Example

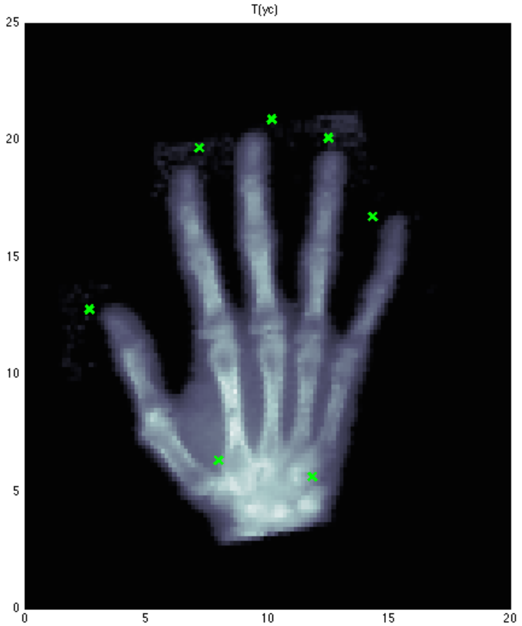


Landmark-based Registration

fixed



moving



- Given landmark pairs $\{(x_i, y_i)\}_{i=1}^{N}$
- Find φ such that: $\varphi(x_i) = y_i$ for all i
- Aligns images based on corresponding landmark points.
- Landmarks are user-defined or automatically detected key points.
- Useful in medical imaging, anthropometry, and morphometry.

The basic idea of landmark-based registration is to determine a transformation φ such that, for a finite number of distinctive features (landmarks), any feature of the moving image is mapped onto the corresponding feature of the reference image.

$I_2 : \Omega \to \mathbb{R}^d$

$I_1 : \Omega \to \mathbb{R}^d$

$I_2 = \{I_2(x), x \in \Omega\}$

$I_1 = \{I_1(y), y \in \Omega\}$

$$\varphi(x_i) \approx y_i = \varphi(x_i) + \epsilon_i, \quad \forall i = 1, \ldots, N$$

$$\mathcal{E}(\varphi) = \mathcal{S}(I_1(\varphi(x)), I_2(x)) + \lambda \mathcal{R}(\varphi)$$

$$\min_{\varphi} \sum_{i=1}^{N} \|\varphi(x_i) - y_i\|^2 + \lambda \mathcal{R}(\varphi)$$

Nonparametric Regression

Landmark-based Registration

- **Rigid:** $\varphi(x) = Rx + t, \quad R = \exp(\theta[\omega]_\times), \; t \in \mathbb{R}^3$

where $[\omega]_\times$ is the skew-symmetric matrix of rotation vector ω.

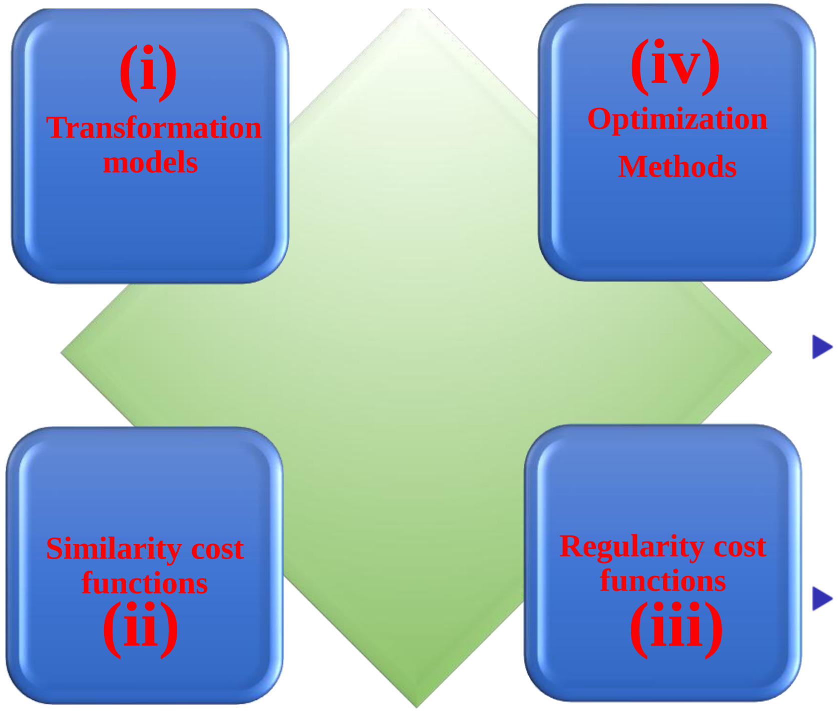
- **Affine:**

$$\varphi(x) = Ax + t, \quad A \in \mathbb{R}^{3\times 3}, \; t \in \mathbb{R}^3$$

- **Nonparametric (e.g., RKHS)**

$$\varphi(x) = Ax + t + \sum_{i=1}^{N} w_i \psi(\|x - x_i\|)$$

$$\psi(r) = \begin{cases} r^2 \log(r), & n = 2 \\ r, & n = 3 \end{cases}$$



$$\mathcal{S}(I_1(\varphi(x)), I_2(x)) \Longrightarrow \sum_{i=1}^{N} \|\varphi(x_i) - y_i\|^2 .$$

$$\mathcal{E}(\varphi)$$

$$\Downarrow$$

$$\min_{\varphi} \sum_{i=1}^{N} \|\varphi(x_i) - y_i\|^2 + \lambda \mathcal{R}(\varphi)$$

$$\frac{\partial \mathcal{E}(\varphi)}{\partial \varphi} = \mathbf{0}$$

- **Bending Energy (TPS):**

$$\mathcal{R}(\varphi) = \int_\Omega \left[\left(\frac{\partial^2 \varphi}{\partial x^2}\right)^2 + 2\left(\frac{\partial^2 \varphi}{\partial x \partial y}\right)^2 + \left(\frac{\partial^2 \varphi}{\partial y^2}\right)^2 \right] dxdy$$

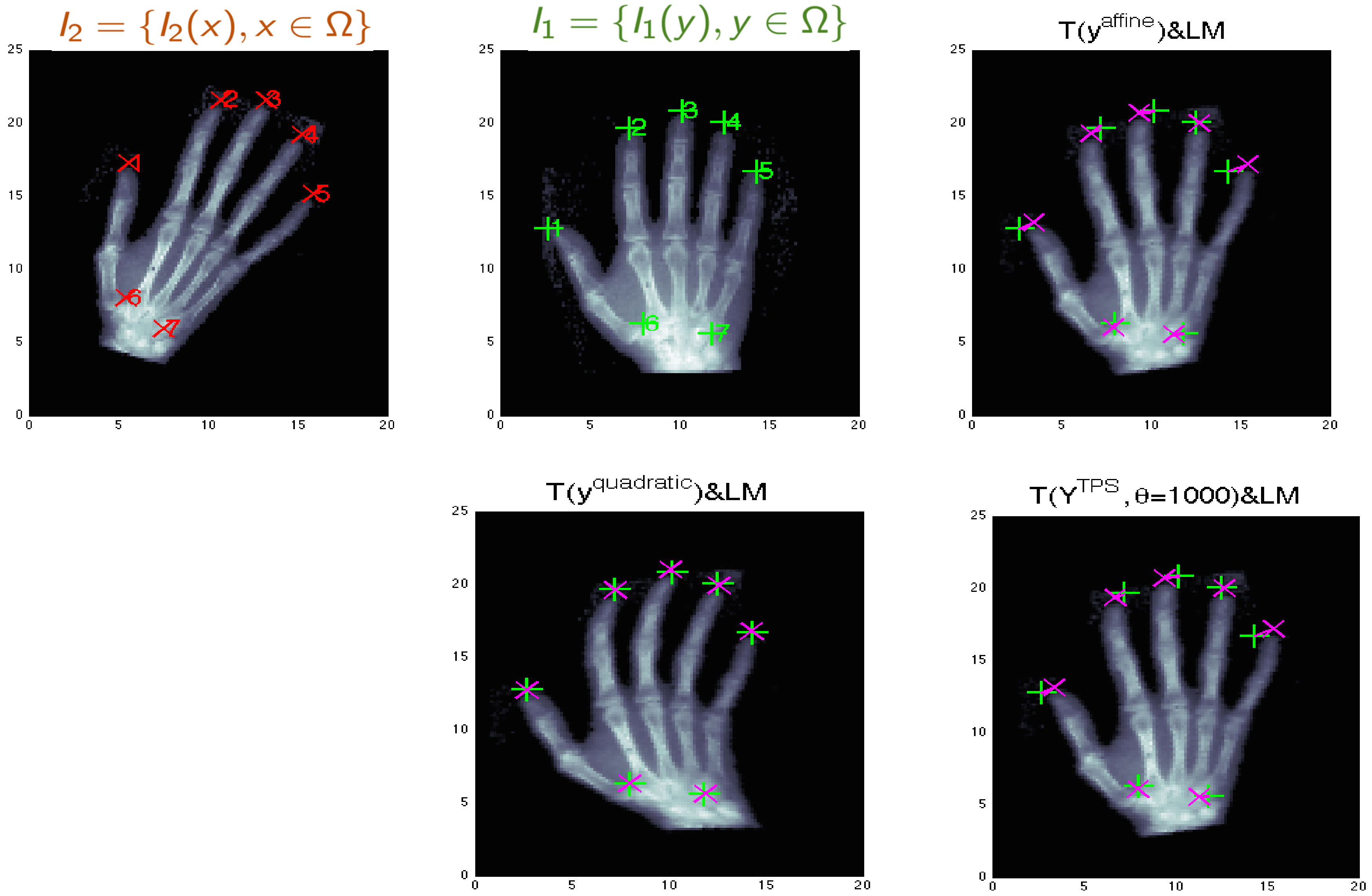
- **Elastic Regularization:**

$$\mathcal{R}(\varphi) = \int_\Omega \left(\mu \|\nabla \varphi\|^2 + \frac{\lambda + \mu}{2} (\nabla \cdot \varphi)^2 \right) dx$$

- **Diffusion Regularization:**

$$\mathcal{R}(\varphi) = \int_\Omega \|\Delta \varphi\|^2 dx$$

Landmark-based Registration: Example



(i) Small vs Large Transformation Models

▶ A **small** transformation model is characterized by small local rotations and small local strains.
▶ A **large** transformation model allows for large local rotations and large local strains.

Discussions:

▶ While large transformation models are more expressive and flexible, **small transformation models are often sufficient in practice.**

▶ In medical imaging, many anatomical structures differ only by small deformations, making **small transformation models very effective.**

▶ Small models are also simpler, involve **fewer degrees of freedom, and are computationally efficient to implement.**

Transformation Model	Small/Large Deformation	Degrees of Freedom 2D	Degrees of Freedom 3D
Rigid	Small	3	6
Affine	Small	6	12
d-th order polynomial	Small	$\binom{d+2}{2}$	$\binom{d+3}{3}$
Cubic B-splines	Small	$(\lfloor \frac{N}{n} \rfloor + 3)^2$	$(\lfloor \frac{N}{n} \rfloor + 3)^3$
Fourier Series	Small	$2(2h+1)^2$	$3(2h+1)^3$
Displacement Field	Small	$2N^2$	$3N^3$
Viscous Fluid (vector field)	Large	$2N^2$	$3N^3$
Stationary velocity (momenta)	Large	varies	varies
Stationary velocity (vector field)	Large	$2N^2$	$3N^3$
Time-dependent velocity field (momenta)	Large	varies	varies
Time-dependent velocity field (vector field)	Large	$2N^2T$	$3N^3T$

(Song, 2017)

Rigid, Affine, and Deformable Transforms

Rigid Transformation:

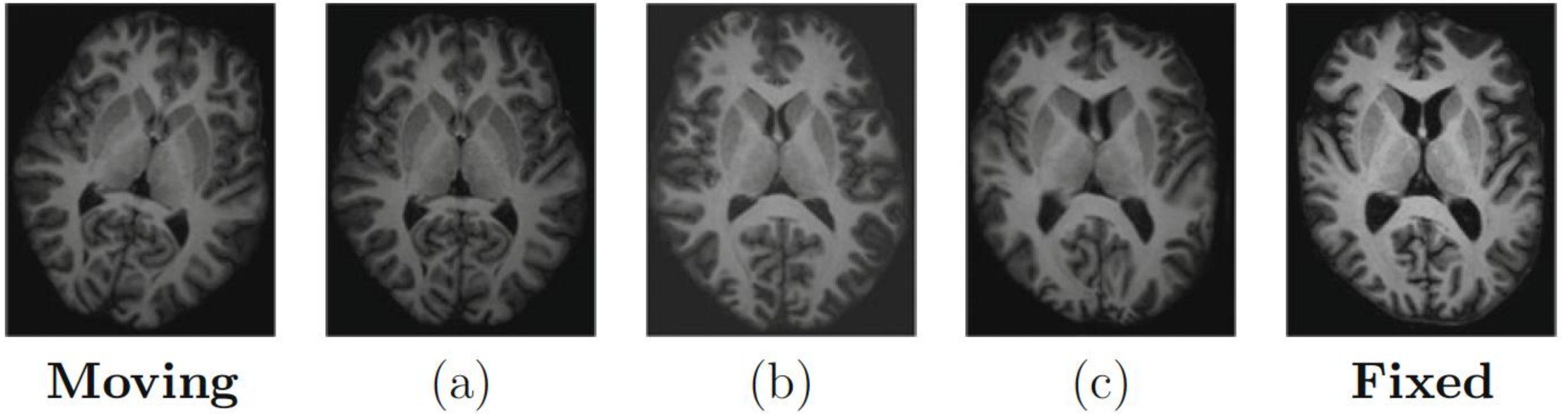
- Preserves distances and angles
- Involves translation and rotation (no scaling or shearing)
- Few parameters (e.g., 3 in 2D, 6 in 3D)
- Fast, often used for intra-subject alignment

Affine Transformation :

- Includes translation, rotation, scaling, and shearing
- More flexible than rigid
- 6 parameters in 2D, 12 in 3D
- Good for global alignment

Deformable Transformation:

- Allows local, nonlinear deformations
- High number of degrees of freedom
- Captures fine-grained anatomical variations
- Computationally more expensive



(a) rigid, (b) affine, and (c) deformable registration

Deformable Transformation Models

- Deformable transformations allow for spatially varying, nonlinear deformations of the image domain.
- Represented by a dense displacement field $\varphi : \Omega \to \mathbb{R}^n$ such that $x \mapsto \varphi(x)$.

Mathematical Formulation

- **Additive Form:** $\varphi(x) = x + u(x)$ where $u(x)$ is the displacement field.
- **Diffeomorphic Form:** $\varphi = \phi_1$ where $\{\phi_t\}_{t\in[0,1]}$ is a time-dependent flow satisfying:

$$\partial_t \phi_t(x) = v_t(\phi_t(x)), \quad \phi_0(x) = x$$

- A **diffeomorphism** is a smooth, invertible transformation with a smooth inverse: $\varphi : \Omega \to \Omega$.
- Ensures one-to-one mappings and topological consistency.
- Commonly used in medical image registration to preserve anatomical structure.

An example of B-spline transformation model is given by

$$\varphi(x) = x + \sum_{l=0}^{3}\sum_{m=0}^{3}\sum_{n=0}^{3} B_l(x_1)B_m(x_2)B_n(x_3)\, a_{i+l,j+m,k+n}$$

where x=(x_1, x_2, x_3), $i = \left\lfloor \frac{x_1}{\delta_1} \right\rfloor - 1, j = \left\lfloor \frac{y_2}{\delta_2} \right\rfloor - 1, k = \left\lfloor \frac{y_3}{\delta_3} \right\rfloor - 1, u = \frac{y_1}{\delta_1} - (i+1), v = \frac{y_2}{\delta_2} - (j+1), w = \frac{y_3}{\delta_3} - (k+1)$ and all $a \in R^3$ are the parameters, B-spline basis functions are defined as $B_0(t) = \frac{-t^3+3t^2-3t+1}{6}, B_1(t) = \frac{3t^3-6t^2+4}{6}$, $B_2(t) = \frac{-3t^3+3t^2+3t+1}{6}$, and $B_3(t) = \frac{t^3}{6}$ for $0 \le t \le 1$. These basis functions are derived using the *Cox-de Boor Recursive Formula*. The B-spline Transformation Model is often referred to as **Free-Form Deformation (FFD)**. FFD describes nonlinear deformations using a regular grid of control points and B-spline basis functions. Local control enables smooth, flexible modeling with a moderate number of parameters.

(ii) Similarity Cost Functions

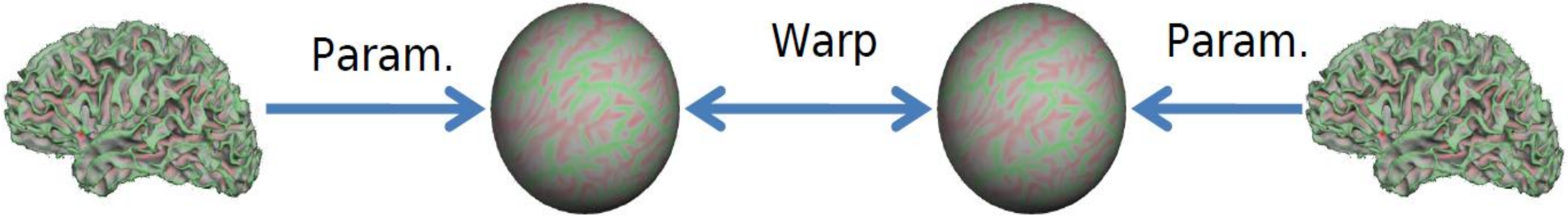
Similarity cost functions measure how well two images align after transformation. It is crucial for optimization-based registration algorithms. Choice depends on modality, noise level, prior segmentation, and specific registration goals.

Intensity-based:

- ▶ Compare voxel intensities directly across images.
- ▶ Assumes similar tissue types or structures have similar intensity patterns.
- ▶ Examples: Mean Squared Error (MSE), Normalized Cross-Correlation (NCC).
- ▶ Best suited for mono-modal registrations (same imaging modality).

Feature-based:

- ▶ Compare higher-level features such as edges, corners, contours, or landmarks.
- ▶ Extract salient image structures before similarity assessment.
- ▶ Examples: Mutual Information (MI) using gradient information, landmark-based distances.
- ▶ More robust to intensity distortions, multi-modal differences, and noise.

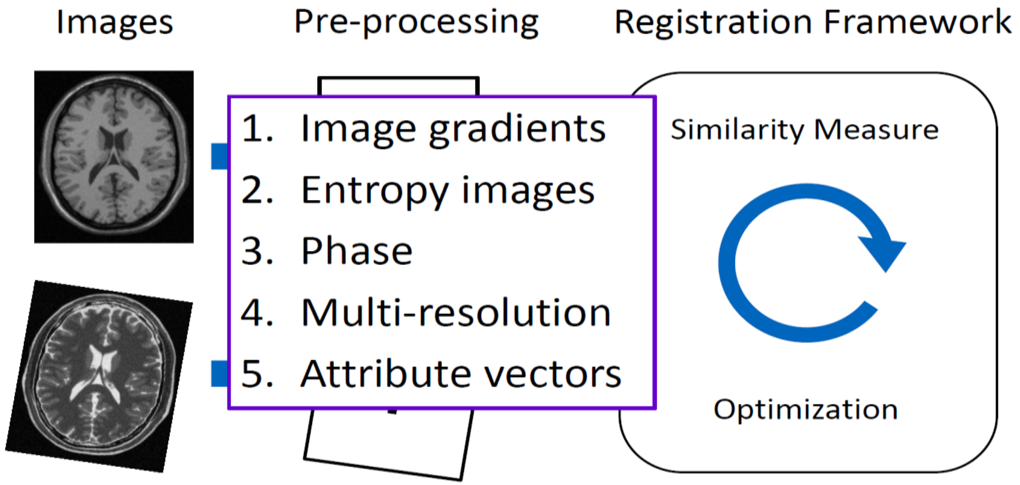


(ii) Similarity Cost Functions: Examples

Mean Squared Error (MSE)

$$\mathcal{D}_{\mathrm{MSE}}(I_1, I_2) = \frac{1}{|\Omega|}\int_{\Omega}(I_1(x) - I_2(x))^2\, dx$$

▶ Assumes corresponding points have similar intensities.
▶ Sensitive to global intensity differences (brightness/contrast changes).
▶ Simple to compute and differentiable, suitable for gradient-based optimization.
▶ **Applications:** Mono-modal rigid, affine, and deformable registration.



Normalized Cross-Correlation (NCC)

$$\mathcal{D}_{\mathrm{NCC}}(I_1, I_2) = -\frac{\left(\int_{\Omega}(I_1(x) - \bar{I}_1)(I_2(x) - \bar{I}_2)\, dx\right)^2}{\left(\int_{\Omega}(I_1(x) - \bar{I}_1)^2\, dx\right)\left(\int_{\Omega}(I_2(x) - \bar{I}_2)^2\, dx\right)}$$

▶ Measures the degree of linear correlation between intensity patterns.
▶ Invariant to linear brightness and contrast changes.
▶ **Applications:** Robust mono-modal registration under varying
lighting conditions.

Mutual Information

$$\mathcal{D}_{\mathrm{MI}}(I_1, I_2) = \sum_{i,j} p_{I_1,I_2}(i,j)\log\left(\frac{p_{I_1,I_2}(i,j)}{p_{I_1}(i)\,p_{I_2}(j)}\right)$$

▶ Captures the statistical dependence between intensities.
▶ High MI indicates strong dependency and good alignment.
▶ Suitable for multi-modal registration (e.g., CT-MRI).
▶ Sensitive to histogram estimation quality.
▶ **Applications:** Multi-modal rigid and deformable registration.

(iii) Regularity Cost Functions: Overview

▶ Regularity terms are added to prevent unrealistic deformations such as folding, tearing, or overly sharp transformations.
▶ They enforce smoothness, invertibility, topology preservation, and physical plausibility of deformation fields.

Common categories of regularity:

▶ **Diffusion Regularization:** Promotes first-order smoothness.
▶ **Elastic Regularization:** Models material-like deformation behavior.
▶ **Bending Energy Regularization:** Controls curvature and smooths second derivatives.

Regularization is typically weighted relative to similarity measures in variational formulations.

$$\mathcal{S}_{\text{diffusion}}(\varphi) = \int_\Omega \|\nabla \varphi(x)\|^2 dx$$

Diffusion Regularization

▶ Penalizes spatial gradients of the deformation.
▶ Encourages globally smooth, continuous transformations.
▶ Simple and computationally efficient, often used in non-rigidregistration frameworks.

$$\mathcal{S}_{\text{elastic}}(\varphi) = \int_\Omega \mu \|\text{sym}(\nabla \varphi)\|^2 + \lambda(\text{tr}(\nabla \varphi))^2 dx$$

Elastic Regularization

▶ Derived from linear elasticity theory.
▶ sym(∇φ): Symmetric part of the Jacobian matrix models local shear and stretch. μ controls shear resistance; λ controls resistance to volume change.

$$\boxed{\mathcal{S}_{\text{bending}}(\varphi) = \int_\Omega \|\nabla^2 \varphi(x)\|^2 dx}$$

Bending Energy Regularization

▶ Penalizes the Laplacian (second derivatives) of the deformation field.
▶ Leads to very smooth, nearly affine transformations locally.
▶ Frequently used in spline-based models such as B-spline.

Advanced Regularizations

- **Hyperelastic Regularization:** Extends elastic models to very large deformations, preserving topology.
- **Diffeomorphic Constraints:** Ensures transformations to be invertible and differentiable; critical for brain/organ mapping.
- **Sobolev Norm Regularization:** Combines multiple derivative orders for fine control over smoothness and stiffness.

$$\mathcal{S}_{\text{hyperelastic}}(\varphi) = \int_\Omega W(\nabla\varphi(x))\,dx$$

where W is a nonlinear strain energy density.

- Preserves topology (no folding or tearing).
- Suitable for highly deformable anatomical structures, e.g., abdominal organs.

$$\mathcal{S}_{\text{diffeo}}(v) = \int_0^1 \|v_t\|_V^2\,dt$$

where V is a reproducing kernel Hilbert space (RKHS) imposing smoothness.

$$\partial_t \varphi_t(x) = v_t(\varphi_t(x)), \quad \varphi_0(x) = x$$

- Critical for topology preservation, especially in brain mapping, longitudinal studies, and large deformation analysis.

$$\|\varphi\|_{H^k}^2 = \sum_{|\alpha|\leq k} \int_\Omega |D^\alpha \varphi(x)|^2\,dx$$

where α is a multi-index.

- Allows fine control over smoothness (first- and second-order together).
- Useful in large deformation models requiring flexible regularity constraints.

(iv) Optimization Techniques

$$u^* = \arg \min_{u(x)=\varphi(x)-x \in H} \left(\mathcal{E}(I_1(\varphi), I_2) + \lambda \mathcal{R}(\varphi) \right) = \arg \min_{\varphi(\cdot) \in H} E(\varphi)$$

Gradient Descent Methods:

- ▶ Compute gradients of the objective function w.r.t. deformation parameters. $\varphi_{k+1} = \varphi_k + h_k.$
- ▶ Iteratively update to minimize the total energy.

Newton and Quasi-Newton Methods:

- ▶ Use second-order derivatives (Hessian) or approximations.
- ▶ Faster convergence for well-behaved problems.

Multi-Resolution Schemes:

- ▶ Solve registration problem at coarse-to-fine scales.
- ▶ Improves convergence and avoids local minima.

Variational and PDE-based Methods:

- ▶ Formulate registration as solving Euler-Lagrange equations.
- ▶ Ensures strong theoretical grounding

$$\frac{d}{d\epsilon} E(u + \epsilon v) \Big|_{\epsilon=0} = 0 \quad \forall v$$

Demons fluid:

$$h = -G^\sigma * \tau \mathcal{F}(\nabla E_\mathrm{D})$$

Demons elastic:

$$h = -G^\sigma * \tau \left(P^{-1} \nabla E_\mathrm{D} + \nabla E_\mathrm{R} \right)$$

Sobolev H$^\infty$:

$$\mathcal{L}^*\mathcal{L} = \sum_{i=0}^{\infty} (-1)^i \, \sigma^{2i} / (i! 2^i) \; \Delta^i$$

$$\nabla_{\mathrm{H}^\infty} E = (\mathcal{L}^*\mathcal{L})^{-1} \; \nabla E$$
$$= G_\sigma * \nabla E$$

PDE-Inspired, semi-implicit:

$$h = -\tau \; (\mathrm{Id} + \tau \lambda \nabla E_R)^{-1} \; \nabla E$$

for diffusion:

$$h = -\tau \; (\mathrm{Id} - \tau \lambda \Delta)^{-1} \; \nabla E$$

Sobolev H^1:

$$\mathcal{L}^*\mathcal{L} = \mathrm{Id} - \lambda \Delta$$

$$\nabla_{\mathrm{H}^1} E = (\mathrm{Id} - \lambda \Delta)^{-1} \; \nabla E$$

Gauß-Newton:

$$h = -\tau \; (J_e^\top J_e)^{-1} \; \nabla E$$

for SSD+diffusion:

$$h = -\tau (\nabla I_\mathrm{S} \nabla I_\mathrm{S}^\top - \lambda \Delta)^{-1} \; \nabla E$$

Preconditioned Descent:

$$h = -\tau \; P^{-1} \; \nabla E$$

Image Registration Evaluation

▶ Evaluation measures the quality of the registration.

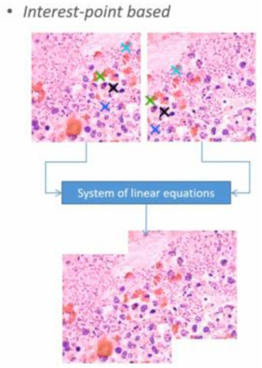
▶ **Key aspects to evaluate:**

▶ **Geometric accuracy:** how well anatomical features align.

▶ **Intensity consistency:** voxel-level similarity post-transformation.

▶ **Smoothness and physical plausibility**: absence of unrealistic folding or discontinuities.

Evaluation is critical for clinical applications and model validation.



Name	Form	Value for perfectly registered images
Landmark Error	$MLE = \sum_{i=1}^{N} \lvert\lvert \phi(p_i) - q_i \rvert\rvert$	0
ROI Overlap Evaluation	$Dice(S_i, T_i) = 2\frac{\lvert S_i \cap T_i \rvert}{\lvert S_i \rvert + \lvert T_i \rvert},\ IOU(S_i, T_i) = \frac{\lvert S_i \cap T_i \rvert}{\lvert S_i \cup T_i \rvert}$	1
Average Volume Difference	$AVD_{j,R} = \frac{1}{M}\sum_{i=1}^{M}\left(\frac{1}{\lvert R \rvert} T_i\left(h_{ij}(x)\right) - \frac{1}{\lvert R \rvert}\sum_{x \in R} T_j(x)\right)^2$	0
Average Sum of Squared Differences	$ASSD_{j,R} = \frac{1}{M}\sum_{i=1}^{M}\sum_{x \in R}\left(T_i\left(h_{ij}(x)\right) - T_j(x)\right)^2$	0
Intensity Variance	$IV_j(x) = \frac{1}{M-1}\sum_{i=1}^{M} T_i\left(h_{ij}(x) - Ave(x)\right)^2$ where $Ave(x) = \frac{1}{M}\sum_{i=1}^{M} T_i(h_{ij}(x))$	0
Average (Normalized) Correlation Coefficient	$ACC_{j,R} = \frac{1}{M}\sum_{i=1}^{M}\frac{\sum_{x \in R}\left(T_i(h_{ij}(x) - \bar{T}_i\right) \cdot \sum_{x \in R}\left(T_j(x) - \bar{T}_j\right)}{\sqrt{\sum_{x \in R}\left(T_i(h_{ij}(x) - \bar{T}_i\right)^2 \cdot \sum_{x \in R}\left(T_j(x) - \bar{T}_j\right)^2}}$	1
Average (Normalized) Mutual Information	$AMI_{j,R} = \frac{1}{M}\sum_{i=1}^{M}\sum_{x \in R} p_{ij}\left(T_i\left(h_{ij}(x)\right), T_j(x)\right) \log_2 \frac{p_{ij}(T_i(h_{ij}(x)), T_j(x))}{p_i\left(T_i\left(h_{ij}(x)\right)\right) \cdot p_j(T_j(x))}$	The higher the better

Major Limitations

Computational Burden:

- ❖ High computational cost due to per-pair optimization.
- ❖ Redundant calculations when registering multiple pairs.
- ❖ Real-time or large-scale applications become impractical.

Non-Convexity of Objective Function:

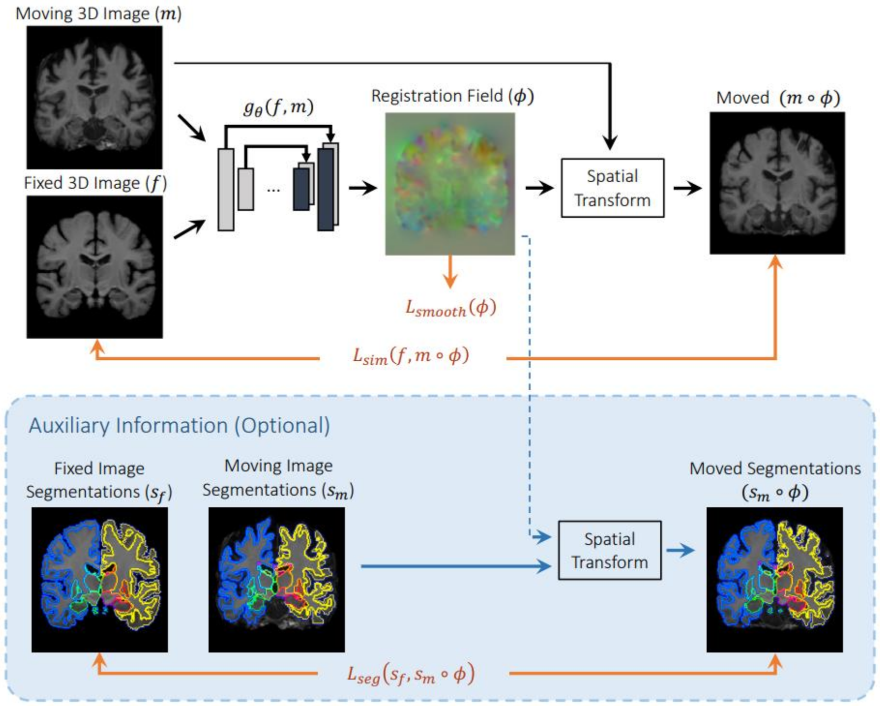
- ❖ The search space for transformations (e.g., displacement fields, diffeomorphisms) is highly non-linear.
- ❖ Objective functions have multiple local minima.
- ❖ Convergence depends heavily on initialization strategies.
- ❖ Regularization must be carefully balanced to avoid over-smoothing or instability.

Motivation for Newer Approaches:

▶ Development of deep learning models to directly predict deformations.

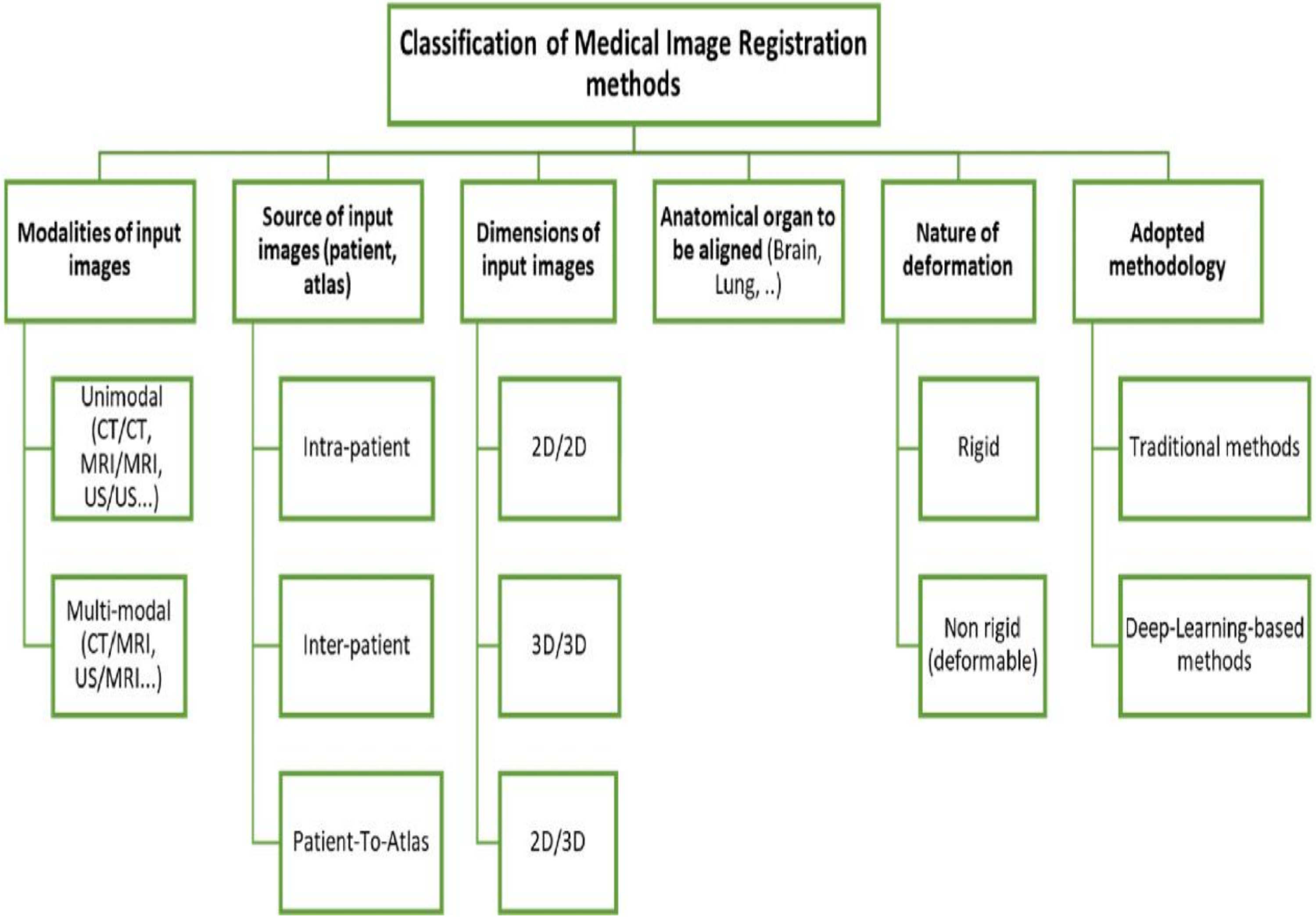
▶ Aim to bypass per-pair optimization with a single trained model.

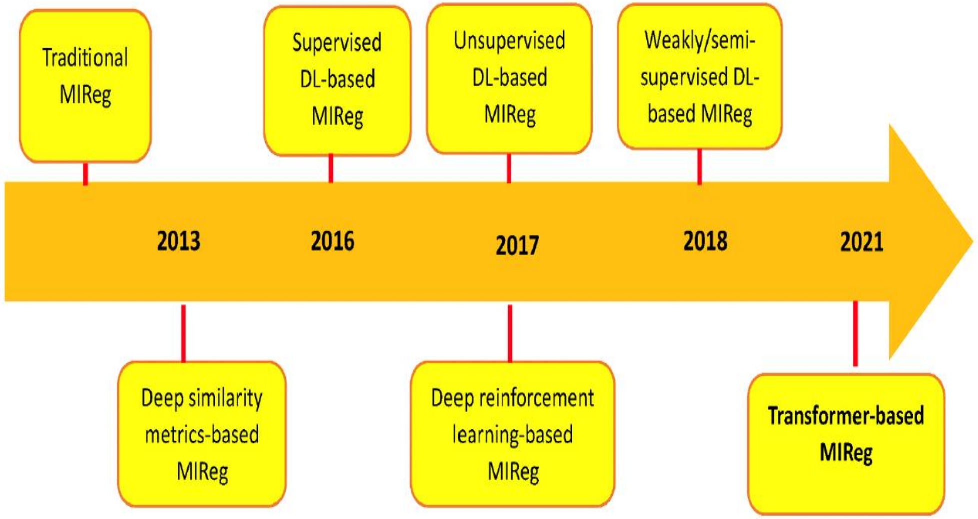
▶ Achieve faster inference and scalability for clinical or real-time use.

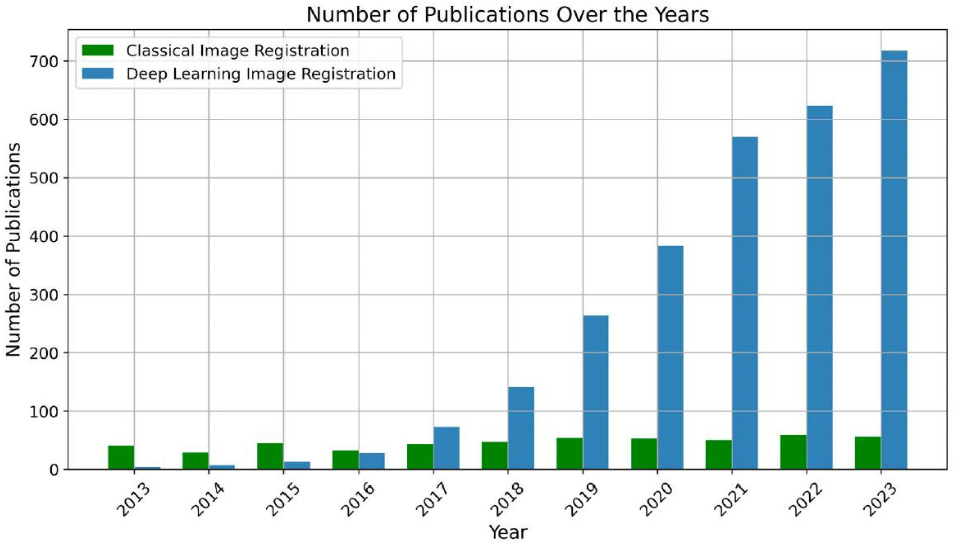


Content

Timeline of DL-based Registration







Learning-based Image Registration

Key Idea:

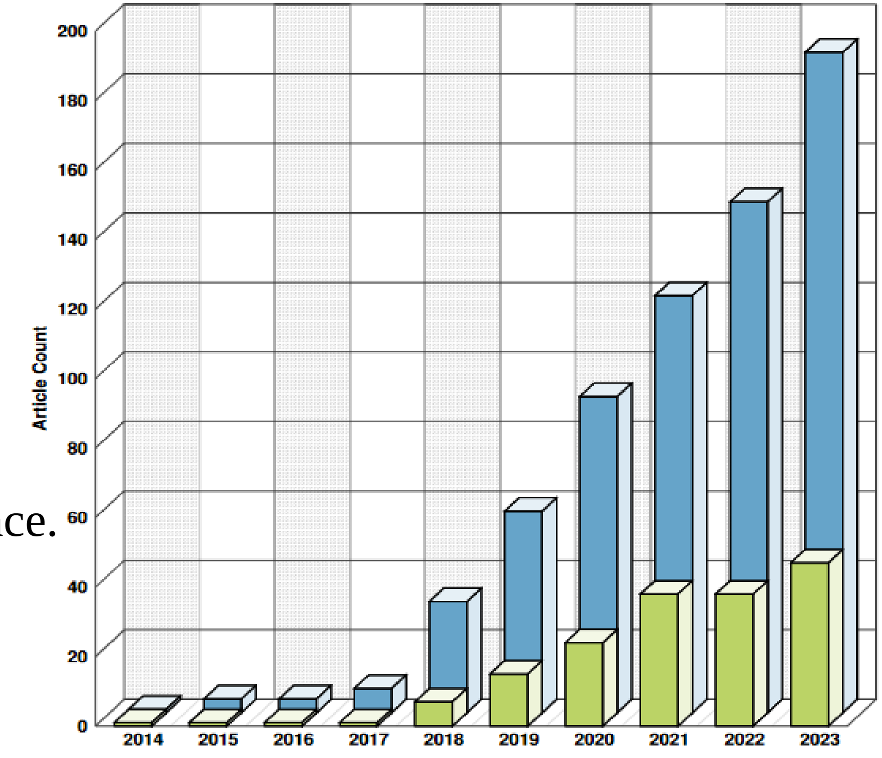
- ▶ Train a neural network on a dataset of image pairs by optimizing a global loss function.
- ▶ During inference, apply the fixed trained network weights directly to new image pairs without further optimization.

Advantages:

- ▶ **Implicit Regularization:**
- ▶ Diversity in training data smooths the loss landscape.
- ▶ Reduces overfitting to noise or local artifacts.

- ▶ **Better Optimization Landscape:**
- ▶ Pretrained weights help escape poor local minima.
- ▶ Transfer learning and advanced optimizers further improve convergence.

- ▶ **Fast Inference:**
- ▶ A single forward pass yields the transformation.
- ▶ Avoids time-consuming iterative optimization during testing.



(Chen et al., 2024)

Network Architectures for Deep Registration

Early Networks:
- ▶ Encoder-based architectures initially served mainly as feature extractors.
- ▶ Replaced hand-crafted features in traditional optimization frameworks.

Impact of U-Net:
- ▶ U-Net introduced encoder-decoder designs ideal for dense prediction tasks like deformable registration.
- ▶ Skip connections help preserve spatial information across scales.
- ▶ Allows for pixel-level accurate deformation field predictions.

Rigid/Affine Registration Networks:
- ▶ Encoder-only networks predict low-dimensional transformation parameters.
- ▶ Typically output 6 parameters (2D rigid) or 12 parameters (3D affine).
- ▶ Loss function minimizes alignment error between transformed and target images.

Supervision Targets:
- ▶ Dense displacement fields for training deformable registration models.
- ▶ Transformation matrices (rotation, translation, scaling) for rigid/affine registration.

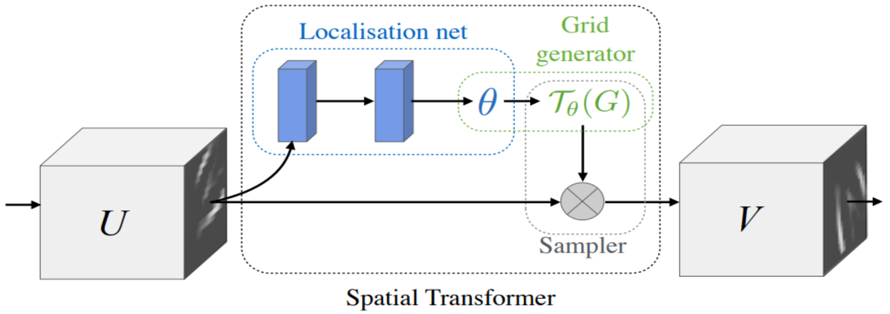
Spatial Transformer Network (STN)

Key Concept:

▶ STN is a differentiable neural network module that spatially transforms feature maps.

▶ Enables models to learn transformations (scaling, rotation, translation) during training.

▶ Allows end-to-end training without requiring manual preprocessing.

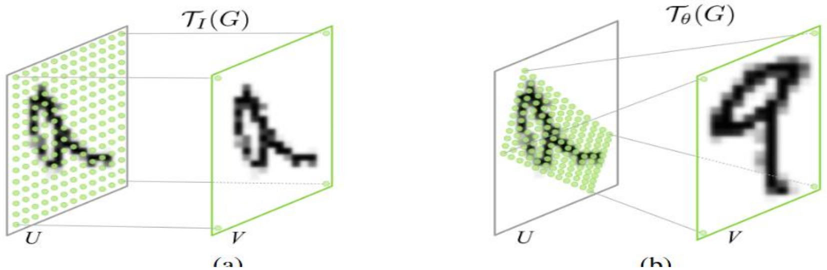


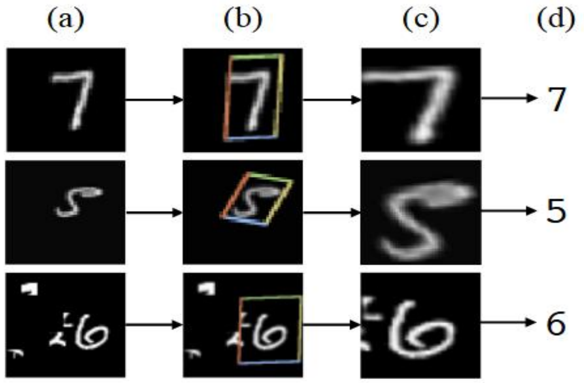
Components of STN:

▶ Localization Network: Predicts transformation parameters θ (e.g., 6 parameters for affine transformations).

▶ Grid Generator: Generates a sampling grid based on predicted θ.

▶ Sampler: Applies the grid to the input feature map to produce the transformed output.





Impact:

▶ Facilitates unsupervised registration by learning spatial transformations directly.

▶ Popular in tasks like image classification, object detection, and medical registration.

STN has led to a shift towards developing unsupervised methods that do not rely on ground-truth transformation.

Jaderberg, M., Simonyan, K., Zisserman, A., & Kavukcuoglu, K. (2015, June 5). *Spatial Transformer Networks*. arXiv.Org. https://arxiv.org/abs/1506.02025v3

Supervised vs. Unsupervised Learning

Two Broad Categories

- ▶ **Supervised Methods**
- ▶ Use ground-truth transformations (matrices or dense displacement fields).
- ▶ Approaches leveraging landmark correspondences or anatomical label maps are still supervised.
- ▶ **Unsupervised (Self-Supervised) Methods**
- ▶ Do not need ground-truth transformations.
- ▶ Train by minimising the discrepancy between the deformed moving image and the fixed image.

Rise of Unsupervised Methods via Spatial Transformer Networks (STN)

- ▶ Introduced a differentiable module to learn spatial transforms inside neural nets.
- ▶ Enabled true unsupervised/self-supervised registration: end-to-end training with image-similarity losses.

Benefits of Removing Ground-Truth Requirement

- ▶ Eliminates costly generation of target transformations.
- ▶ Allows networks to explore richer deformation spaces.
- ▶ Easier enforcement of smoothness, invertibility, and topology preservation.
- ▶ Provides flexibility to adapt across modalities and datasets

Paradigm for Learning-based Registration

(Chen et al., 2024)

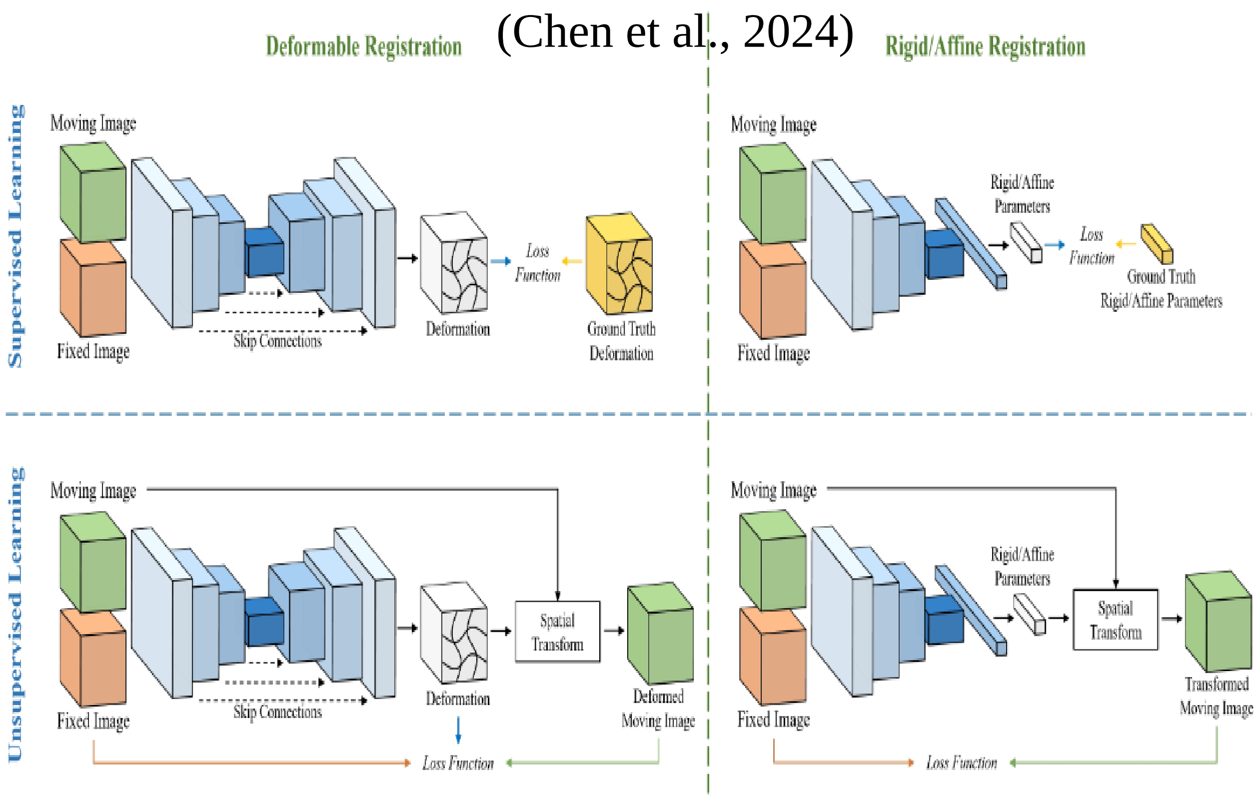


Figure above illustrates the conventional paradigm of learning-based rigid/affine and DIR with f the following components:

- **Moving and fixed images as input**
- **A deep neural network**
- **STN (for unsupervised methods)**
- **A loss function**

❖ For **affine/rigid** registration methods, neural network encoders are used for feature extraction and fully connected layers are used to output the parameters of the predicted transformation.

❖ For **deformable image registration (DIR)**, neural networks with both encoder and decoder are used. The result is a deformation field of equal sizes to the input images.

➢ In the **supervised** setting, the network output is compared to ground truth transformations generated from synthetic transformation or traditional image registration methods using a loss function.

➢ In the **unsupervised** setting, the predicted transformation is used by the STN to warp the moving image, and the transformed image is then evaluated against the fixed image using a loss function.

Local Similarity Measures in Deep Registration

Why move beyond MSE?

▶ Mean–squared error (MSE) ignores local intensity structure.
▶ Local similarity measures capture fine spatial correspondence.

Local Correlation Coefficient (LCC)

▶ Computes Pearson correlation in sliding windows W.
▶ Robust to bias–field and intensity non-uniformity in mono-modal MR.
▶ Implemented in deep nets via windowed convolutions ⇒ fully differentiable.

Local Mutual Information (LMI)

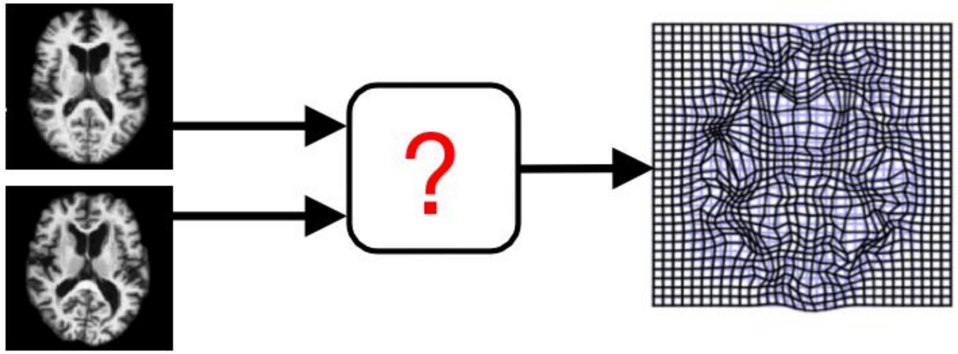
▶ Estimates mutual information within non-overlapping patches.
▶ Suited to multi-modal registration (e.g., CT–MRI).
▶ Patch-wise computation lowers memory vs. full 3-D histograms while remaining differentiable.

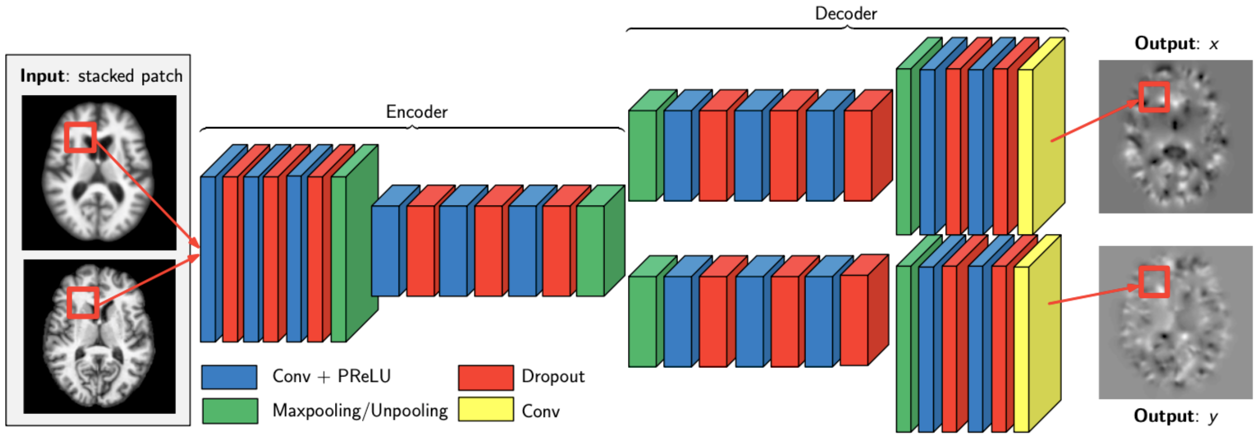
Trade-offs

▶ LCC & LMI improve alignment quality but increase computational cost compared with MSE.
▶ Choice depends on modality, GPU memory budget, and required accuracy.

Quicksilver: IR as a Regression Problem

- Idea: Optimization is slow, so let's do prediction instead





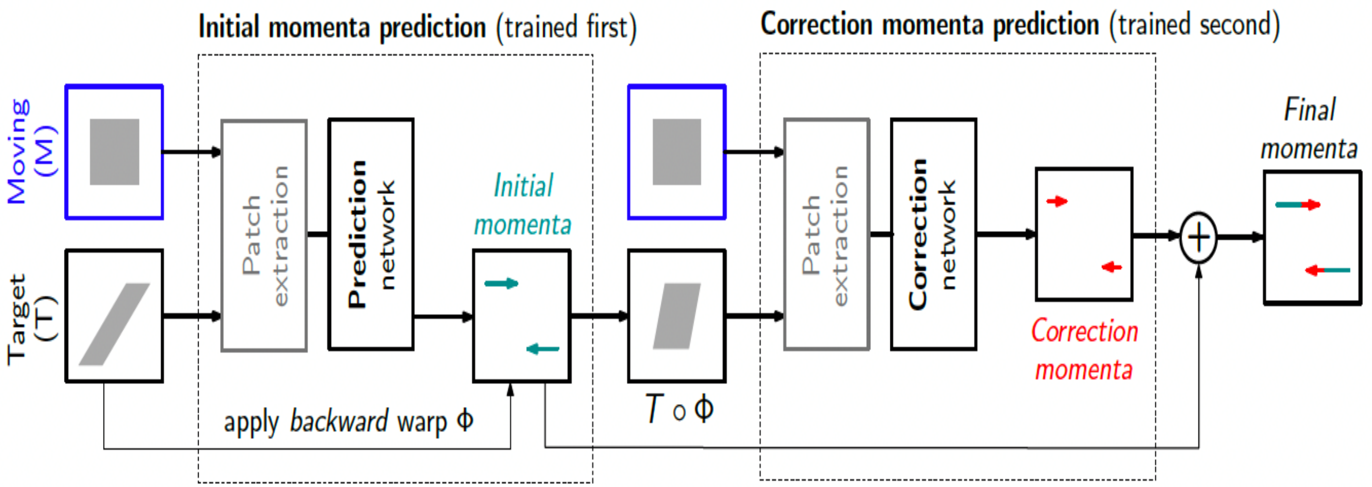
Possible choices for what to predict:

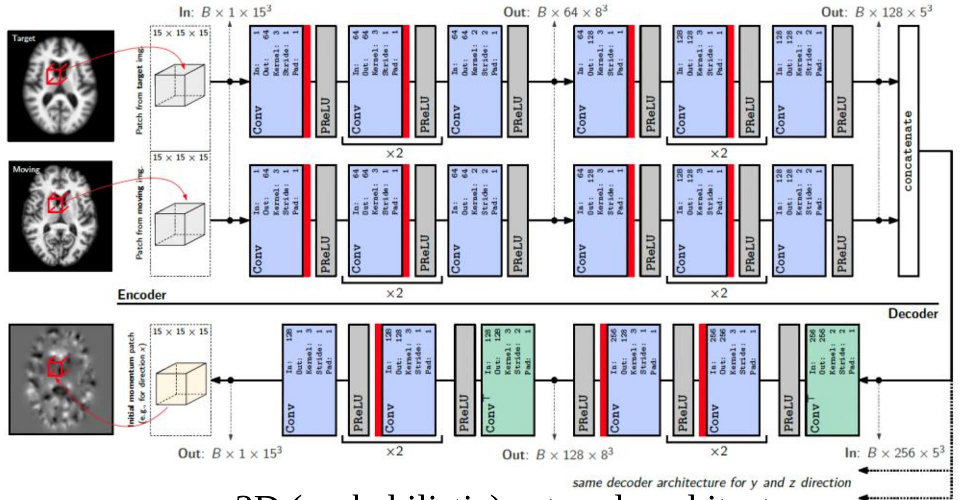
- Local displacement $\Phi(x) = x + u(x)$
- Stationary velocity field $\Phi_t = v \circ \Phi$
- Momentum fields $m = L^{\dagger}Lv$

- Introduces a deep learning-based approach for fast deformable image registration by predicting deformation models directly from image appearance.
- Predicts the momentum-parameterization of LDDMM, enabling patch-wise prediction while preserving theoretical guarantees like diffeomorphic mappings.
- Provides a probabilistic version of the prediction network to estimate uncertainties in predicted deformations during testing.

Yang, Xiao et al. "Quicksilver: Fast predictive image registration - A deep learning approach." *NeuroImage* vol. 158 (2017): 378-396. doi:10.1016/j.neuroimage.2017.07.008

Two-step Training Pipeline of Quicksilver





3D (probabilistic) network architecture.

Quicksilver

Step 1: Train Prediction Network Train on original moving–target pairs using ground-truth initial momenta from full LDDMM optimization.

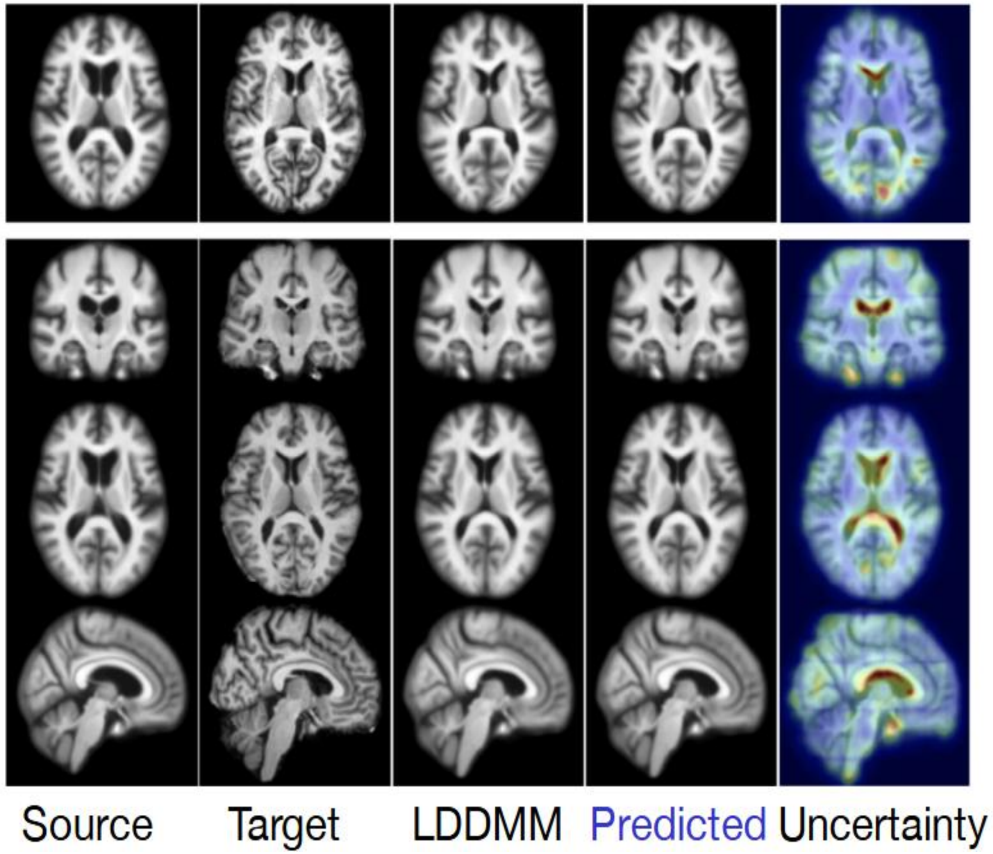
Step 2: Back-Warp Targets Shoot predicted momenta $\widehat{m}_0^{\text{pred}}$ to get deformation Φ and warp each target back: T ′ = T ∘Φ.

Step 3: Train Correction Network Feed (moving, T ′) patches; supervise with residual $m_0^\star - \widehat{m}_0^{\text{pred}}$ to learn the momentum error.

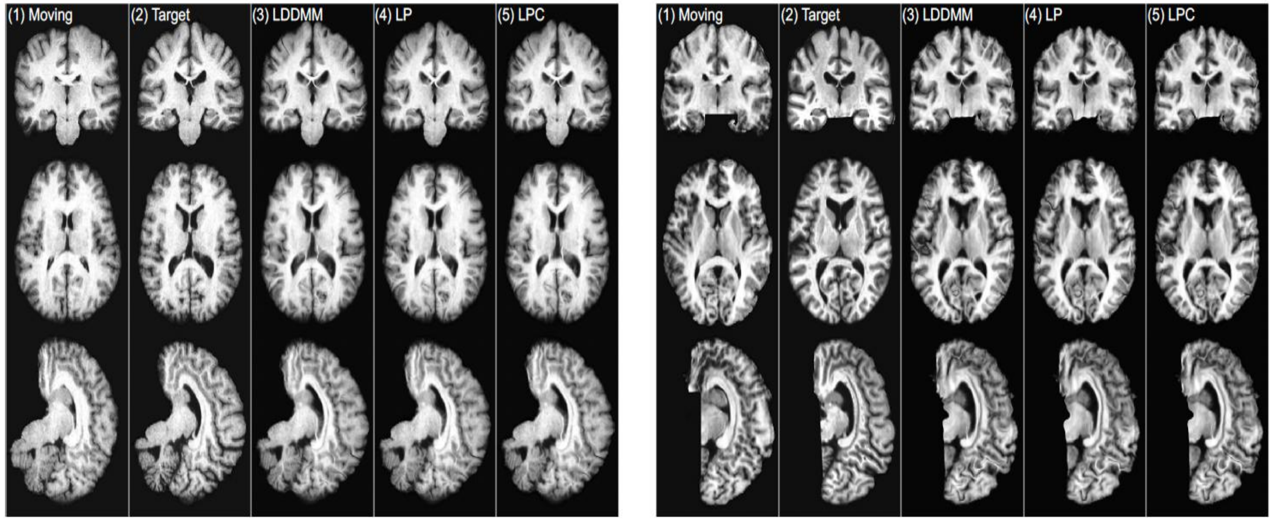
At inference: run Prediction → Correction, add the two momenta, then shoot once for the final diffeomorphic map.

- **Input patches** $P_M, P_T \in \mathbb{R}^{15\times15\times15}$ (moving / target)
- **Twin 3-D Encoders (no weight sharing)**
  - 2 blocks each: $[3 \times (3^3 \text{ Conv} + \text{PReLU}) \rightarrow 2^3 \text{ Conv}_{\text{stride}=2}]$
  - Channels: $1 \rightarrow 64 \rightarrow 128$
- **Feature fusion** – concatenate encoders → 256-ch latent tensor.
- **Three Symmetric Decoders** (m_x, m_y, m_z)
  - Mirror of encoder with transposed-conv unpooling Channels: $256 \rightarrow 128 \rightarrow 1$
  - Final conv *linear* (no activation)
- **Regularisation** Dropout $d = 0.2$ after every conv (Bayesian MC-Dropout)
- **Loss** – voxel-wise $\ell_1(\widehat{m}, m^\star)$
- **Capacity** – 97 360 kernels 21.8 M learnable params, trained with $>10^6$ patches.

Quicksilver



Atlas-to-image registration example. The coloring indicates the level of uncertainty, with red = high uncertainty and blue = low uncertainty.



(a) LPBA40 (b) IBSR18

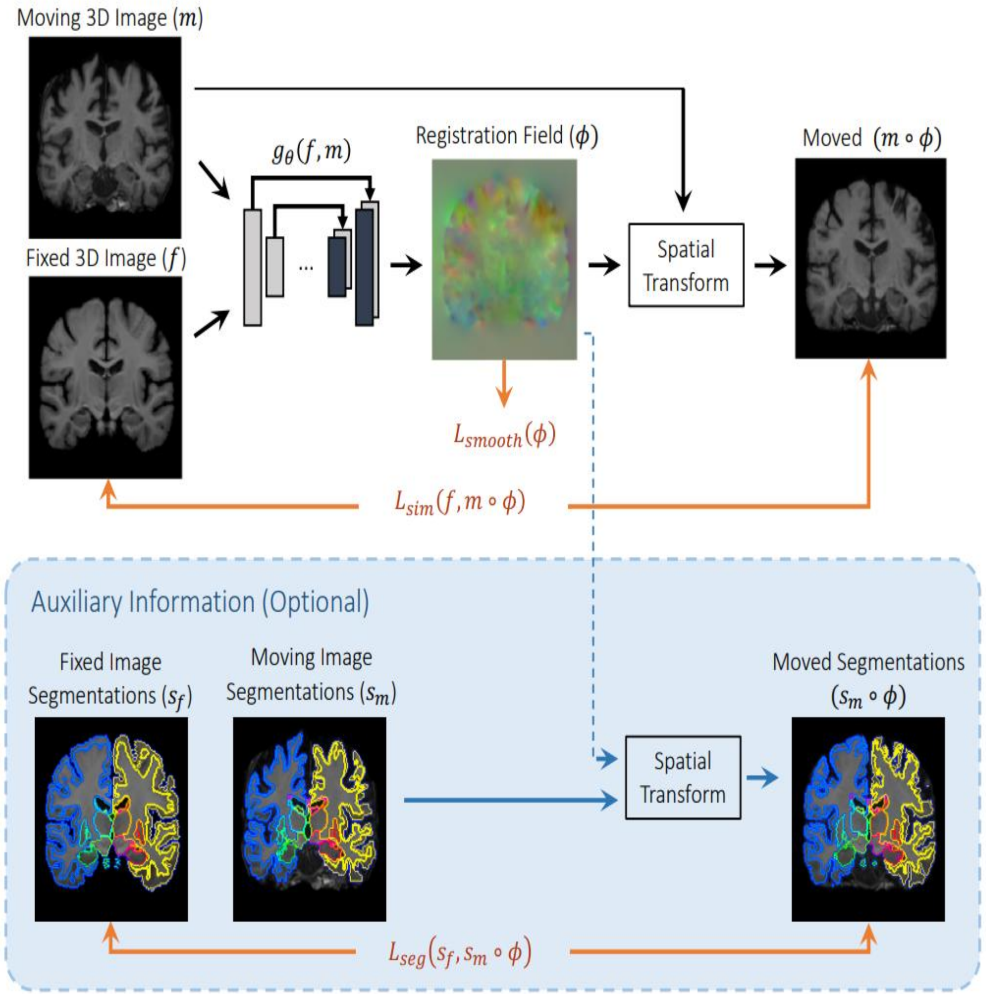
Example test cases for the image-to-image registration.

	Deformation Error w.r.t LDDMM optimization on T1w-T1w data [mm]						
Data percentile for all voxels	0.3%	5%	25%	50%	75%	95%	99.7%
Affine (Baseline)	0.1664	0.46	0.9376	1.4329	2.0952	3.5037	6.2576
T1w-T1w LP	0.0348	0.0933	0.1824	0.2726	0.3968	0.6779	1.3614
T1w-T1w LPC	0.0289	0.0777	0.1536	0.2318	0.3398	0.5803	1.1584
T1w-T2w LP	0.0544	0.1457	0.2847	0.4226	0.6057	1.0111	2.0402
T1w-T2w LPC	0.0520	0.1396	0.2735	0.4074	0.5855	0.9701	1.9322
T1w-T2w LP, 10 images	0.0660	0.1780	0.3511	0.5259	0.7598	1.2522	2.3496
T1w-T2w LPC, 10 images	0.0634	0.1707	0.3356	0.5021	0.7257	1.1999	2.2697

VoxelMorph

VoxelMorph is an unsupervised CNN-based DIR method for MRI brain atlas-based registration. The architecture uses a U-Net-like architecture.

1. **Inputs**: m: moving volume f: fixed volume
2. **CNN** $g_\theta(f, m)$: UNet-style encoder–decoder outputs dense displacement field u.
3. **Deformation map**: $\phi = \mathrm{Id} + u$ (voxel-wise offsets).
4. **Spatial Transformer**: Warps m to $m \circ \phi$ with trilinear interpolation (fully differentiable).
5. **Training losses**
   - Image similarity: MSE or local CC.
   - Smoothness: $\|\nabla u\|^2$.
   - Optional Dice term if segmentations available.
6. **Optimization**: Single SGD training on $\{(f_i, m_i)\}$ amortised registration.
7. **Inference**: One forward pass: <1 s GPU / <1 min CPU.



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. https://doi.org/10.1109/TMI.2019.2897538

VoxelMorph

- The unsupervised loss function consists of two components for a regularization parameter λ:

$$L_{us}(f,m,\phi) = L_{sim}(f,m,\phi) + \lambda L_{smooth}(\phi)$$

- L_{sim} can take either of two forms:
  - Mean squared error: $MSE(f, m\circ\phi) = \frac{1}{|\Omega|}\sum_{p\in\Omega}|f(p) - [m\circ\phi](p)]^2$
  - Local cross correlation $CC(f, m\circ\phi) = \sum_{p\in\Omega}\frac{\left[\sum_{p_i}\left(f(p_i)-\hat{f}(p)\right)\left([m\circ\phi](p_i)-[\widehat{m}\circ\phi](p)\right)\right]^2}{\left[\sum_{p_i}\left(f(p_i)-\hat{f}(p)\right)^2\right]\left[\sum_{p_i}\left([m\circ\phi](p_i)-[\widehat{m}\circ\phi](p)\right)^2\right]}$ where p_i is the intensity of the i-th voxel and the local region is an $n\times n\times n$ cube, $\hat{f}(p) = \frac{1}{n^3}\sum_{p_i} f(p_i)$ denote the local mean intensity image. This choice is more robust to intensity variations across scans and datasets.

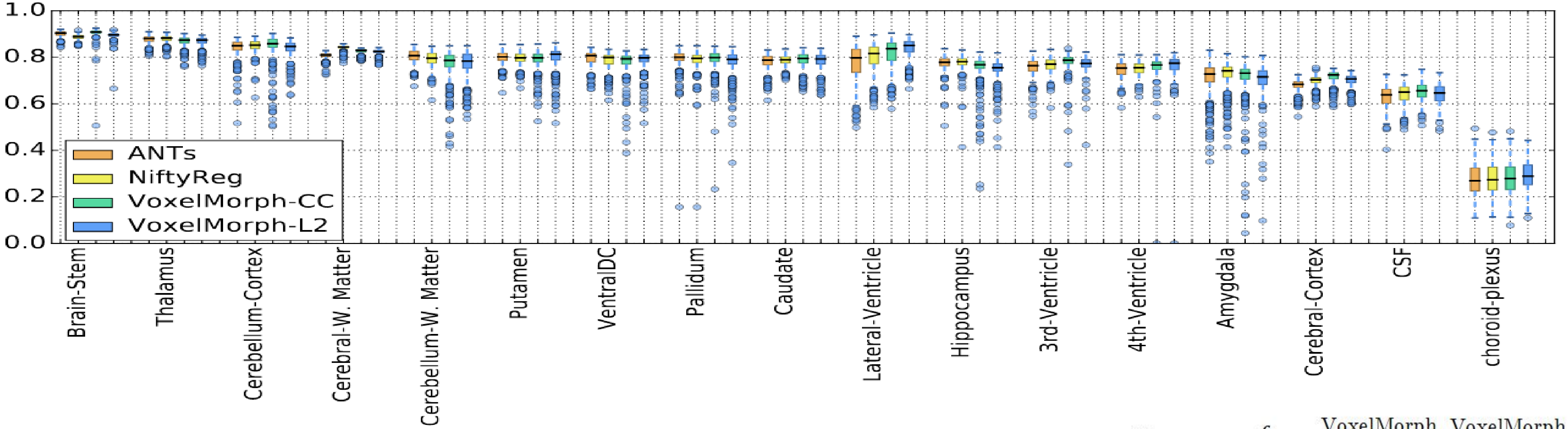
- We encourage a smooth displacement field ϕ using a **diffusion regularizer** on the spatial gradients: $L_{smooth}(\phi) = \sum_{p\in\Omega}\left|\left|\nabla \mathrm{u(p)}\right|\right|^2$

- Optionally, auxiliary information such as anatomical segmentations s_f, s_m can be leveraged during training. The loss function can be defined as follows, where γ is a regularization parameter:

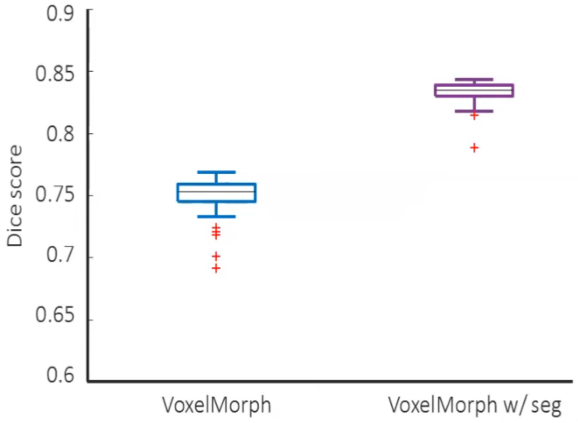
$$L_a(fm, m, s_f, s_m, \phi) = L_{us}(f,m,\phi) + \gamma L_{seg}(s_f, s_m\circ\phi)$$

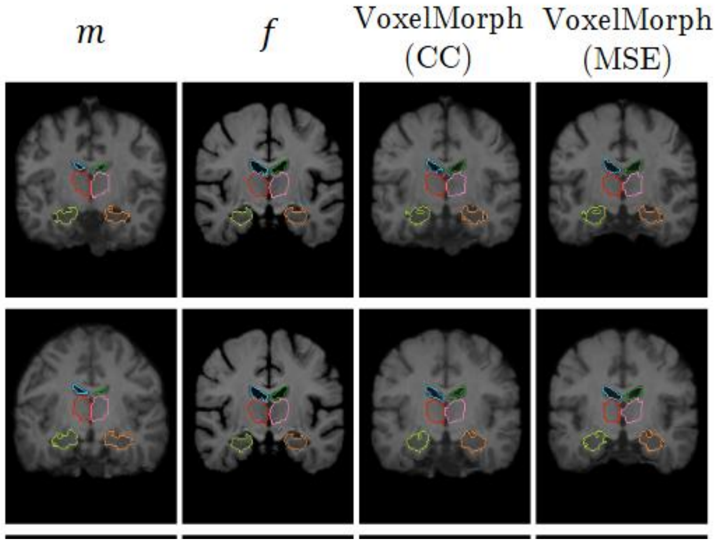
- The segmentation loss L_{seg} over all structures $k\in[1,K]$ is defined as $L_{seg}(s_f, s_m\circ\phi) = -\frac{1}{K}Dice(s_f^k, s_m^k\circ\phi)$

VoxelMorph: Performance



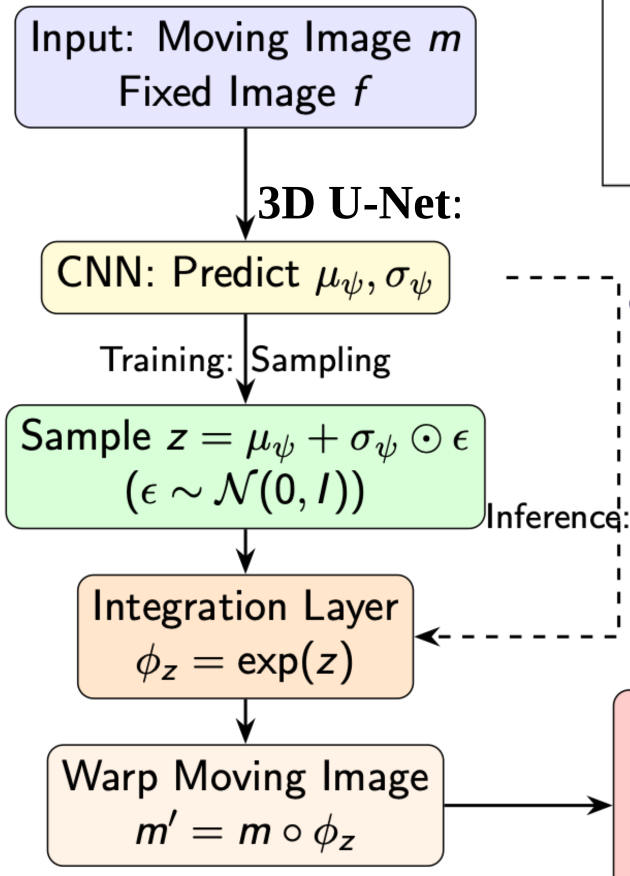
Method	Dice	GPU sec	CPU sec	$\lvert J_\phi \rvert \leq 0$	% of $\lvert J_\phi \rvert \leq 0$
Affine only	0.584 (0.157)	0	0	0	0
ANTs SyN (CC)	0.749 (0.136)	-	9059 (2023)	9662 (6258)	0.140 (0.091)
NiftyReg (CC)	0.755 (0.143)	-	2347 (202)	41251 (14336)	0.600 (0.208)
VoxelMorph (CC)	0.753 (0.145)	0.45 (0.01)	57 (1)	19077 (5928)	0.366 (0.114)
VoxelMorph (MSE)	0.752 (0.140)	0.45 (0.01)	57 (1)	9606 (4516)	0.184 (0.087)

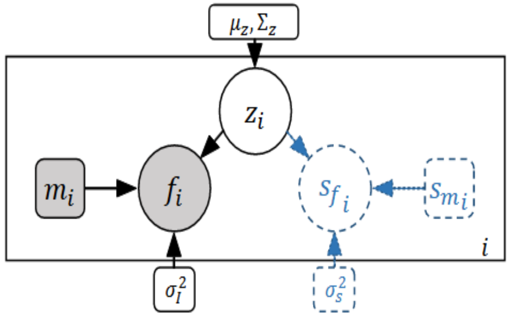




Probabilistic Diffeomorphic Registration

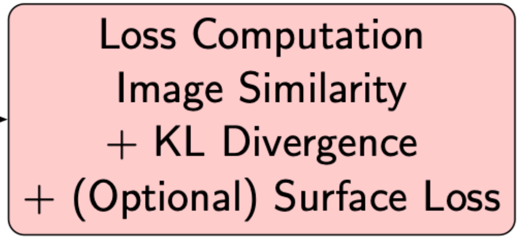
- **Likelihood:** $p(f \mid z; m) = \mathcal{N}(f; m \circ \phi_z, \sigma_I^2 I)$
- **Prior:** $p(z) = \mathcal{N}(0, \Sigma_z), \quad \Sigma_z^{-1} = \lambda L.$

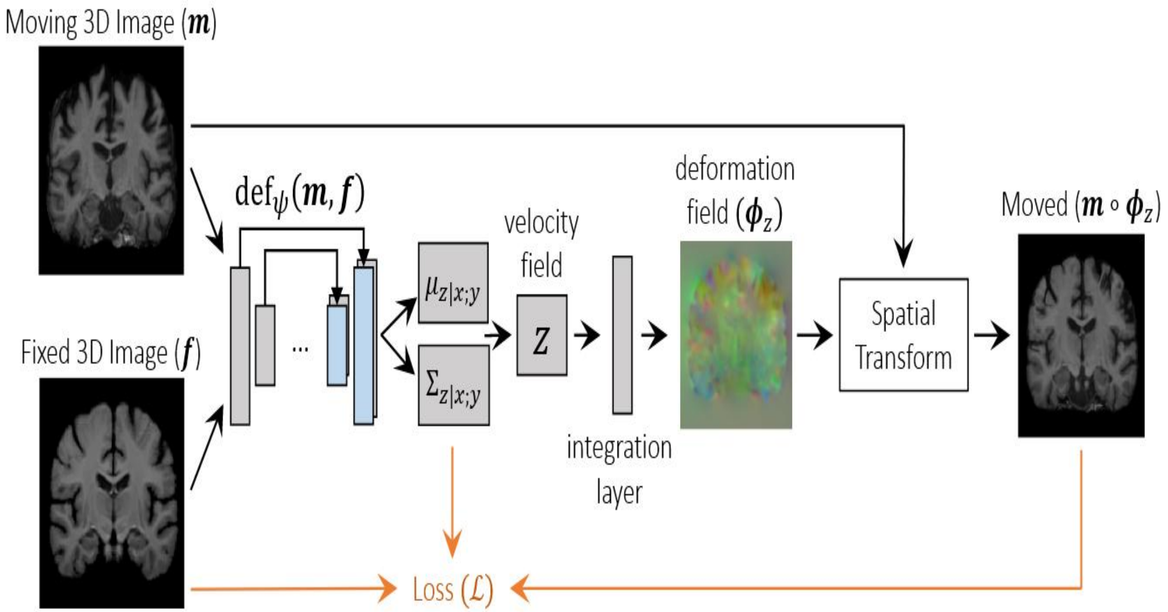




- **Approximate posterior**

$$q_\psi(z \mid f, m) = \mathcal{N}(z; \mu_\psi, \mathrm{diag}(\sigma_\psi^2))$$





- **Variational loss:**

$$\mathcal{L}(\psi) = \frac{1}{2\sigma_I^2}\|f - m \circ \phi_z\|^2 + \frac{1}{2}\left(\mathrm{tr}(\lambda L \Sigma_\psi) - \log|\Sigma_\psi| + \mu_\psi^\top \lambda L \mu_\psi\right)$$

- **Optional surface loss:**

$$\mathcal{L}_{\mathrm{surf}} = \frac{1}{2\sigma_s^2}\left(\sum_n d(s_f[n] \circ \phi_{-z}, s_m) + \sum_n d(s_m[n] \circ \phi_z, s_f)\right)$$

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. https://doi.org/10.1016/j.media.2019.07.006

Integration Layer and Performance

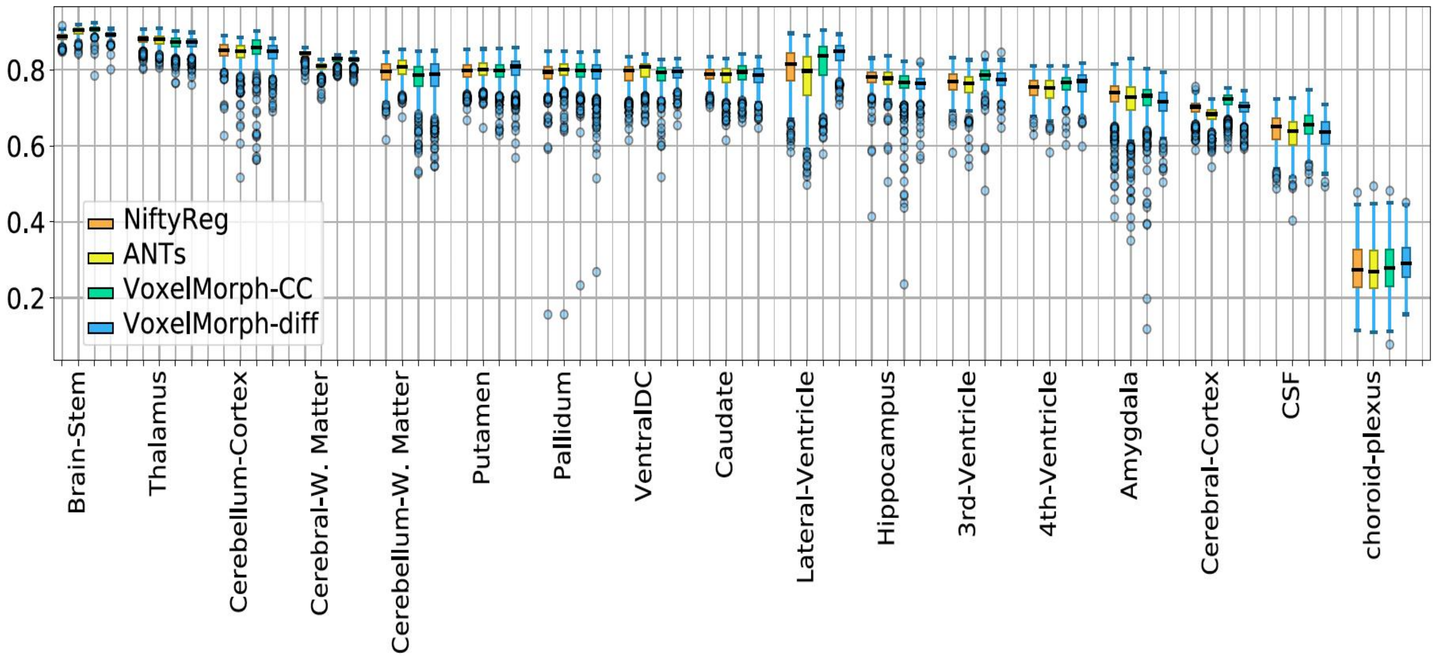
Scaling and Squaring Integration

Compute the **exponential map** $\phi_z = \exp(z)$ using **scaling-and-squaring**:

1. Scale: $z \to z/2^T$
2. Initialize: $\phi = \mathrm{Id} + z/2^T$
3. Repeat T times: $\phi \leftarrow \phi \circ \phi$

Ensures that ϕ_z is a **diffeomorphism** (smooth, invertible, topology-preserving).

Method	Avg. Dice	GPU sec	CPU sec	mean $\lvert J_\Phi \rvert$	$\lvert J_\Phi \rvert \le 0$
Affine only	0.584 (0.157)	0	0	1	0
ANTs (SyN)	0.749 (0.136)	–	9059 (2023)	1.001 (0.036)	7,523 (4790)
NiftyReg (CC)	0.755 (0.143)	–	2347 (202)	1.072 (0.131)	33,838 (8307)
VoxelMorph (CC)	0.753 (0.145)	0.45 (0.01)	57 (1.0)	1.032 (0.074)	19,715 (3540)
Supervised-diff	0.730 (0.144)	0.35 (0.03)	82.6 (3.8)	1.088 (0.121)	0.05 (0.5)
VoxelMorph-diff	0.754 (0.139)	0.47 (0.01)	84.2 (0.1)	1.075 (0.124)	0.2 (1.0)



Content

Registration Neural Networks

Recent registration NN architectures for registration leverage powerful deep learning tools:

- **Adversarial learning** for better realism
- **Contrastive learning** for robust features
- **Transformers** for global interactions
- **Diffusion models** for uncertainty modeling
- **Hyperparameter conditioning** for adaptability

Future: Combine multiple paradigms into unified, efficient registration frameworks

Method	Anatomy	Modality	Network Infrast
AC-DMIR	Brain/Uterus	MRI	Transformer
ADMIR	Brain	MRI	CNN
Attention-Reg	Prostate	US/MRI	CNN(Self Attent)
CycleMorph	Faces/Brain/Liver	Photogra/MRI/CT	CNN
DiffuseMorph	Faces/Brain/Cardc	Photogra/MRI	DDPM
DLIR	Cardiac/Chest	MRI/CT	CNN
FAIM	Brain	MRI	CNN
Fourier-Net	Brain	MRI	CNN
HyperMorph	Brain	MRI	CNN
TransMorph	Brain/Abdomen	MRI/CT	Transformer
VoxelMorph	Brain	MRI	CNN
ViT-VoxelMorph	Brain	MRI	Transformer
XMorpher	Brain/Cardc	MRI / CT	Transformer

Table: Summary of selected registration methods: anatomy, modality, and network infrastructure.

Adversarial Learning

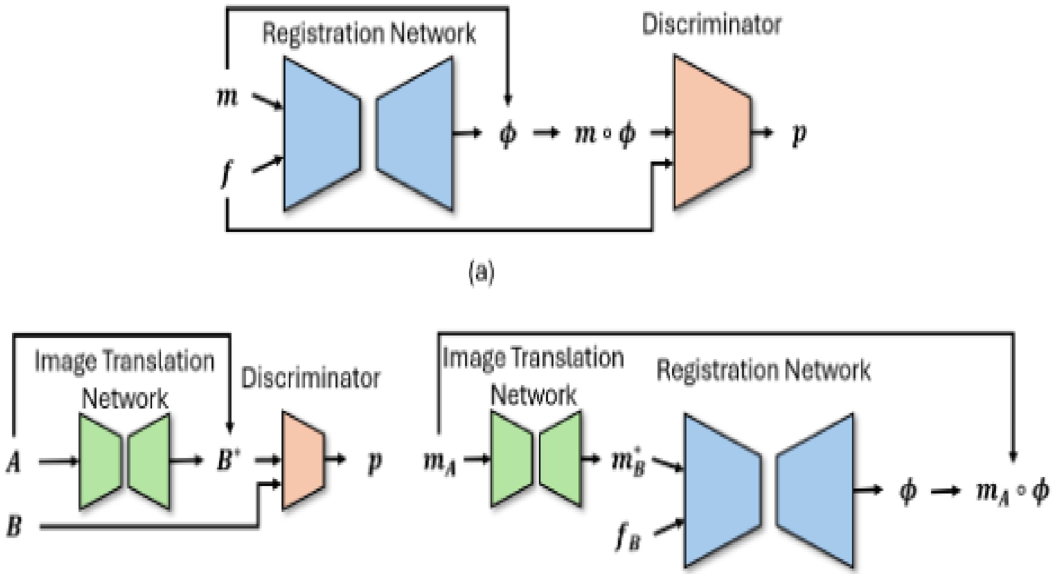
1. Deformation or Transformation Prediction:
- Generators predict deformation fields or affine transformation parameters.
- Discriminators judge alignment quality between warped moving images and fixed images, learning implicit similarity.

2. Inverse-Consistent Deformation Enforcement:
- Adversarial learning combined with cycle consistency constraints ensures that forward and backward deformations are consistent.

3. Incorporating Anatomical Label Maps:
- Label maps are warped alongside images, and discriminators evaluate anatomical alignment, improving structure preservation.

4. Flexible Positive Pair Definitions:
- Positive registration examples include blended images or pre-aligned multimodal pairs, relaxing strict identity assumptions.

5. Modality Synthesis and Registration:
- Images are first translated across modalities using GANs, then registered in a unified modality space.
- Symmetric pipelines and uncertainty-weighted fusion further improve registration robustness.



Two Roles of Adversarial Learning: (a) Metric Learning for Similarity: Discriminator D learns to differentiate well-aligned vs poorly-aligned pairs. p = D(f , m ∘ ϕ) used as similarity measure. **(b) Modality Synthesis for Multi-Modal Registration** Adversarial learning synthesizes images into a common modality space (e.g., A → B). Registration then proceeds in the synthesized space.

6. Knowledge Distillation via Adversarial Learning:
- A lightweight student network learns from a larger teacher network.
- Discriminator distinguishes deformation fields generated by student and teacher.
- After training, only the compact student network is retained, achieving comparable anatomical accuracy with significantly fewer parameters.

Contrastive Learning

Principle: DNNs learn by comparing positive pairs (similar) and negative pairs (dissimilar), without relying on task-specific similarity metrics.

Benefits for Registration:

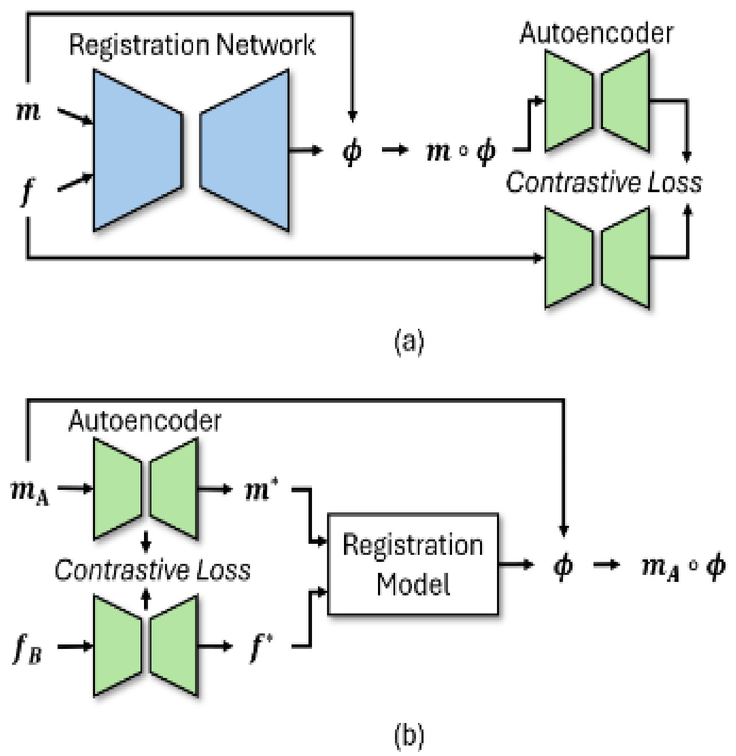
- Avoids manual selection of similarity measures for different modalities (e.g., MRI vs CT, mono- vs multi-modal).
- Learns registration-aware representations directly from data.

Contrastive Learning Strategies:

- **Keypoint Patch-Based:** Detect keypoints, extract patches, use Siamese networks and contrastive loss to optimize affine alignment.
- **Representation Space Alignment:** Map multi-modal images into contrastive representations using separate networks, maximize mutual information (InfoNCE loss), followed by conventional registration.
- **Intermediate Feature Contrastive Supervision:** Apply contrastive loss to intermediate or final layers of encoder networks to improve feature quality.
- **Synthesis-by-Registration:** Train a registration network first, then train an image synthesis network using patch-based contrastive loss (PatchNCE) to enhance geometric consistency.

Recent Extensions:

- **Mono-modal Registration:** Apply contrastive loss between unregistered moving and fixed images, leveraging consistency in anatomical structures.
- Positive pairs may include structurally similar but unaligned images to encourage learning of correspondence under small deformations.



In (a), the contrastive learning acts as a similarity metric. In (b), contrastive learning can be used to transform images from different modalities into a unified feature representation, upon which registration model operates. For the contrastive loss, we may minimize the distance between corresponding key points and maximizing the distance between non-corresponding key points.

Contrastive Learning: CNNFR

Objective: Improve the robustness and accuracy of **rigid registration** for **multi-modal images** (e.g., CT & MRI) using **deep learned descriptors** instead of hand-crafted features like SIFT or MIND.

Key Idea: Use a **Siamese CNN** trained with **contrastive loss** to learn discriminative keypoint descriptors:

•Minimize feature distance between matching keypoints

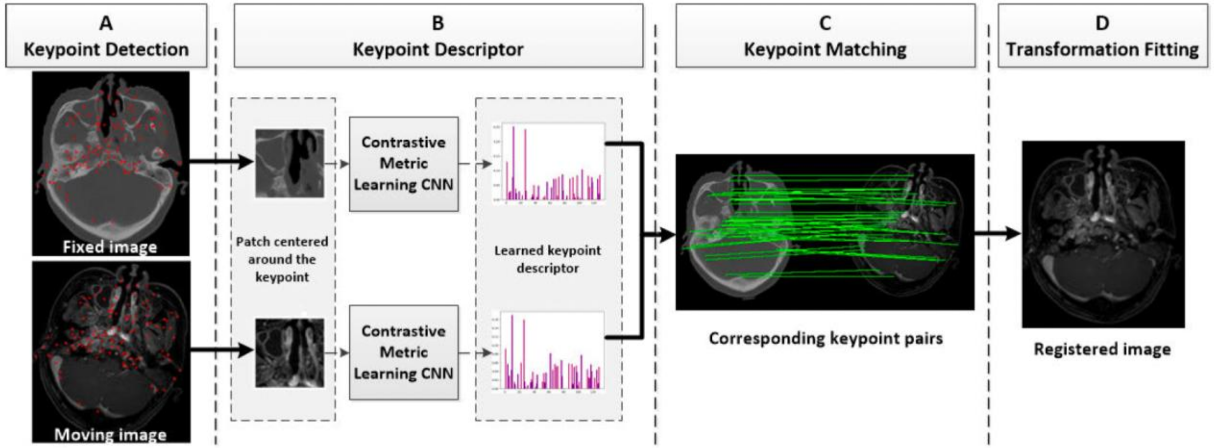
•Maximize distance between mismatches

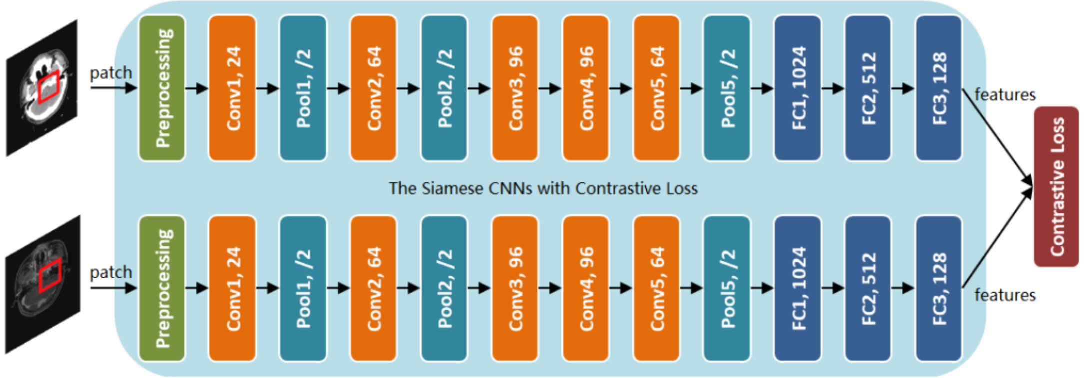
Contrastive Loss:

$$L = \frac{1}{2N} \sum_{i=1}^{N} y_i d_i^2 + (1 - y_i) max(margin - d_i, 0)^2$$

$$d_i = \|x_{i1} - x_{i2}\|_2$$

yi=1 if matched, 0 otherwise





Pipeline (CNNFR):

1. **Keypoint detection** via DoG
2. **Patch extraction** around keypoints
3. **Descriptor learning** using contrastive Siamese CNN
4. **Keypoint matching** based on descriptor distance
5. **Affine transformation fitting** using RANSAC

Hu, J., Sun, S., Yang, X., Zhou, S., Wang, X., Fu, Y., Zhou, J., Yin, Y., Cao, K., Song, Q., & Wu, X. (2019). Towards Accurate and Robust Multi-Modal Medical Image Registration Using Contrastive Metric Learning. *IEEE Access*, *7*, 132816–132827. https://doi.org/10.1109/ACCESS.2019.2938858

Contrastive Learning: CNNFR

Transfer Learning Variant (TrCNNFR):

•Pretrained on natural image patches (UBC dataset)

•Fine-tuned on CT-MR pairs → better generalization

Evaluation Metrics:

•**Target Registration Error (TRE)**

•**Precision-Recall** for keypoint matching

Key Results:

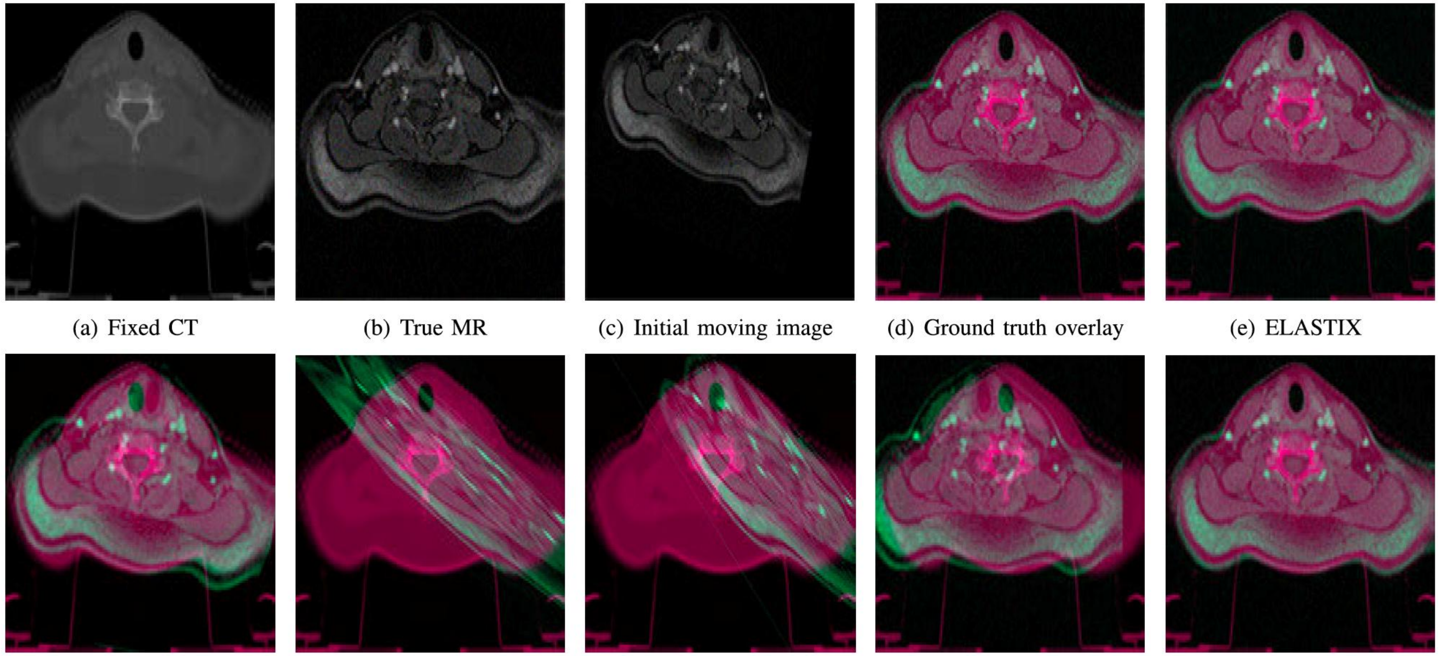
•TrCNNFR outperforms:

- **SIFT, MIND, AIRNet, ELASTIX**

•Robust to:

- **Image noise, scaling, rotation**
- **Missing data, low overlap regions**

•~29× faster than ELASTIX



(a) Fixed CT (b) True MR (c) Initial moving image (d) Ground truth overlay (e) ELASTIX

(f) AIRNet (g) SIFT (h) MIND (i) CNNFR (j) TrCNNFR

Generalization:

•Tested on **unseen body parts** (chin–shoulder) and **modalities** (T1–T2)

•Maintains competitive performance without retraining

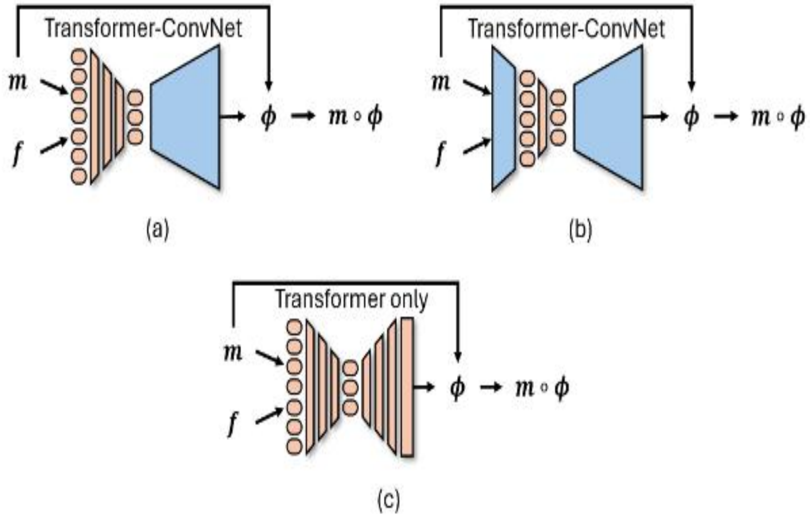
Transformers

1 Self-attention-Based:

❖ Transformers (e.g., ViT, Swin) replace or augment ConvNet encoders.
❖ Capture intra-image relations for registration tasks.
❖ Examples: Hybrid Transformer-ConvNet architectures; full Transformer encoders/decoders.

2 Cross-attention-Based:

❖ Cross-attention mechanisms correlate features between moving and fixed images.
❖ Enhance matching accuracy across modalities or anatomy differences.
❖ Dual-stream encoders, deformable cross-attention modules improve spatial correspondence.



3. Advanced Transformer Architectures:

❖ Coarse-to-Fine Strategies: Multi-resolution ViTs progressively refine deformations.
❖ Deformable Cross-Attention: Sample beyond fixed windows for better matching, reducing computational cost.
❖ Coordinate-Based Cross-Attention: Explicitly guide spatial correspondences (e.g., im2grid).
❖ Motion Decomposition: Predict multiple candidate deformation fields (e.g., ModeT), followed by competitive weighting.

4. ConvNet Evolution Inspired by Transformers:

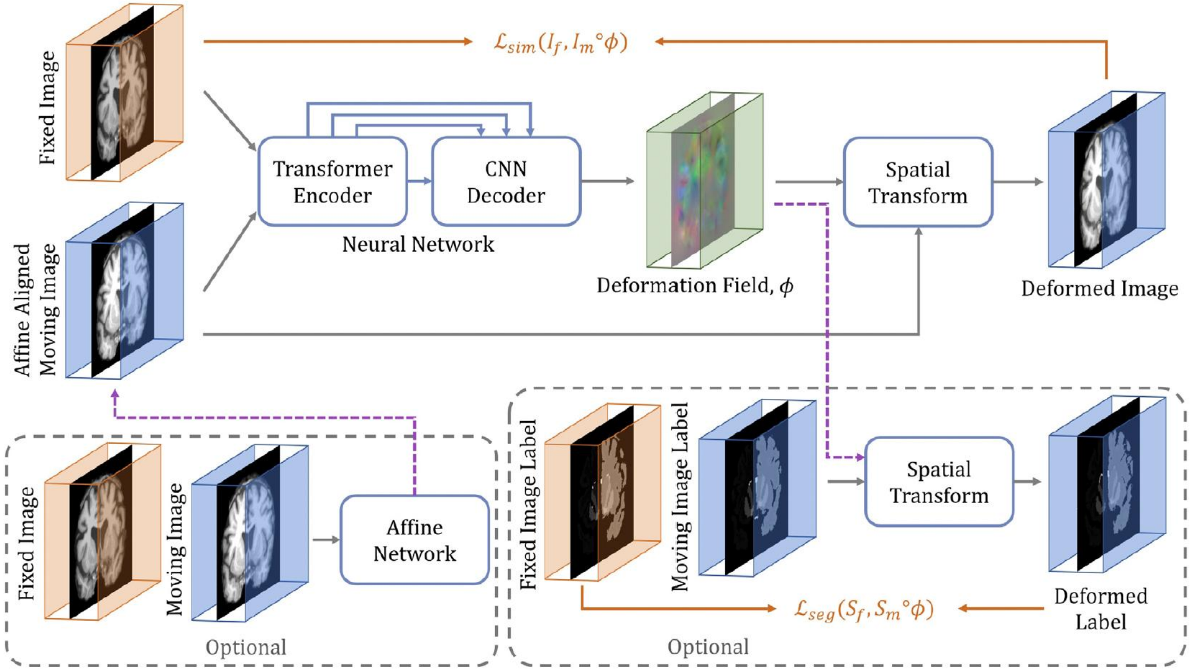
❖ New ConvNet models (e.g., ConvNeXt, RepLKNet) integrate Transformer concepts (e.g., large kernels).
❖ Enhanced U-Nets with large convolution kernels expand receptive fields and challenge Transformer dominance.
❖ ConvNets maintain advantages: invariance to input size, inductive bias, computational efficiency.

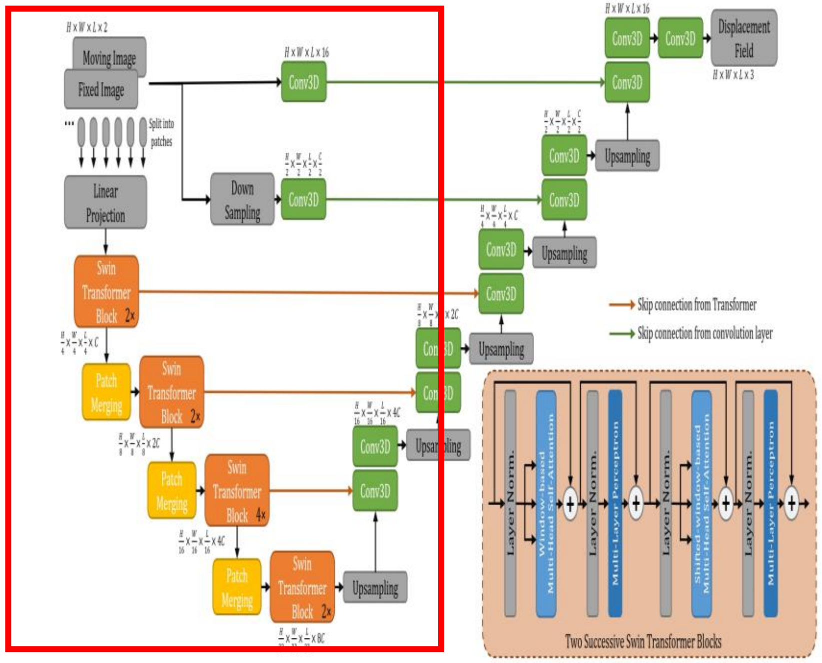
Future Direction:

Hybrid designs and improved ConvNets leveraging Transformer insights are promising for registration tasks.

Transformers: TransMorph

- **Goal:** Develop a Transformer-based deep learning framework for **unsupervised medical image registration**.
- **Model Architecture:** **TransMorph** is a **hybrid Transformer-ConvNet** framework:
    - **Encoder:** Swin Transformer extracts hierarchical features.
    - **Decoder:** ConvNet reconstructs dense deformation field φ.
    - **Skip Connections:** Preserve spatial details across encoder-decoder stages.



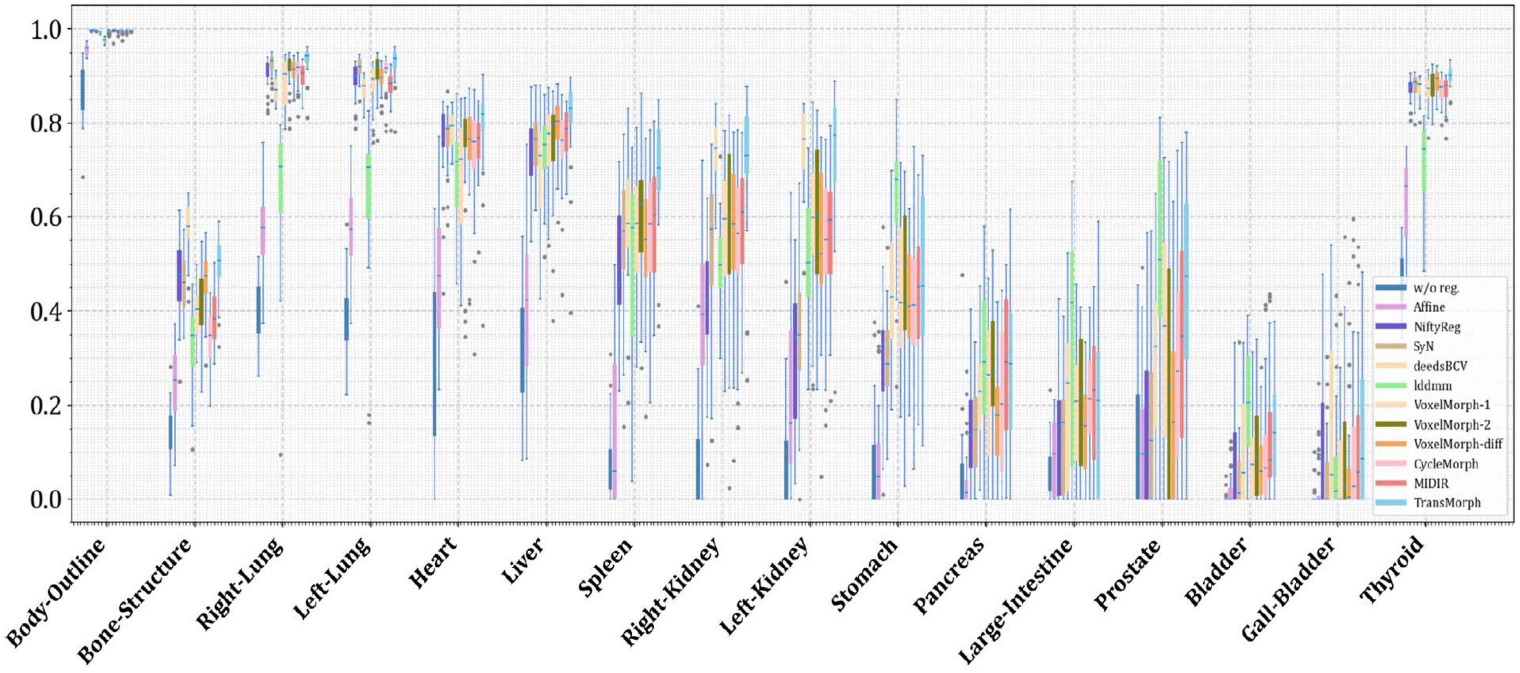


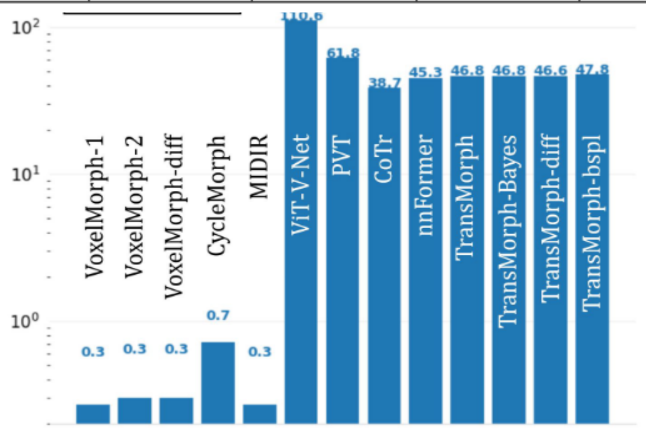
Chen, J., Frey, E. C., He, Y., Segars, W. P., Li, Y., & Du, Y. (2022). TransMorph: Transformer for unsupervised medical image registration. *Medical Image Analysis*, *82*, 102615. https://doi.org/10.1016/j.media.2022.102615

Transformers: TransMorph

- In inter-subject and atlas-to-subject brain MRI registration, it achieved significantly improved registration performance when compared to top-performing traditional and ConvNet-based registration models.
- Even though certain networks (ViT-V-Net) had almost twice the number of trainable parameters, TransMorph still outperformed all the Transformer-based models in Dice score, demonstrating Swin-Transformer's superiority over other Transformer architectures.

	Inter-patient MRI		Atlas-to-patient MRI	
Model	DSC	% of $\lvert J_\phi \rvert \leq 0$	DSC	% of $\lvert J_\phi \rvert \leq 0$
Affine	0.572±0.166	-	0.386±0.195	-
SyN	0.729±0.127	<0.0001	0.645±0.152	<0.0001
NiftyReg	0.723±0.131	0.061±0.093	0.645±0.167	0.020±0.046
LDDMM	0.716±0.131	<0.0001	0.680±0.135	<0.0001
deedsBCV	0.719±0.130	0.253±0.110	0.733±0.126	0.147±0.050
VoxelMorph-1	0.718±0.134	0.426±0.231	0.729±0.129	1.590±0.339
VoxelMorph-2	0.723±0.132	0.389±0.222	0.732±0.123	1.522±0.336
VoxelMorph-diff	0.715±0.137	<0.0001	0.580±0.165	<0.0001
CycleMorph	0.719±0.134	0.231±0.168	0.737±0.123	1.719±0.382
MIDIR	0.710±0.132	<0.0001	0.742±0.128	<0.0001
ViT-V-Net	0.729±0.128	0.402±0.249	0.734±0.124	1.609±0.319
PVT	0.729±0.130	0.427±0.254	0.727±0.128	1.858±0.314
CoTr	0.725±0.131	0.415±0.258	0.735±0.135	1.292±0.342
nnFormer	0.729±0.128	0.399±0.234	0.747±0.135	1.595±0.358
TransMorph-Bayes	*0.744±0.125*	0.389±0.241	0.753±0.123	1.560±0.333
TransMorph-diff	0.730±0.129	<0.0001	0.594±0.163	<0.0001
TransMorph-bspl	0.740±0.123	<0.0001	**0.761±0.122**	<0.0001
TransMorph	**0.745±0.125**	0.396±0.240	*0.754±0.124*	1.579±0.328





Diffusion Models

Background:

- Diffusion models have gained popularity in computer vision for tasks, such as image synthesis and super-resolution.
- They learn to reverse a forward process where noise gradually diffuses image information—analogous to thermodynamic diffusion.
- **Advantage:** no restrictions on training data variability or modality.

Application to Image Registration:

- Combine a diffusion network (to learn semantic priors via score function) with a registration network.
- The score function captures features of the fixed image and guides deformation of the moving image.
- This approach enables robust, continuous deformation estimation.

Examples:

DiffuseMorph (Kim et al., 2022):

❖ Diffusion network learns a conditional score function ∇x log p(x|I_f).

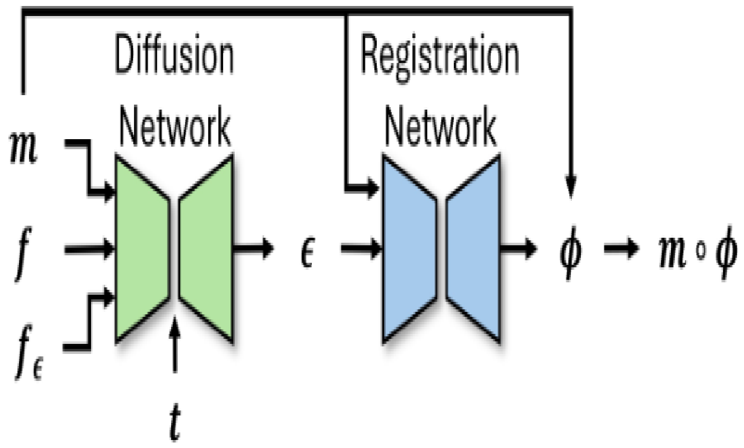
❖ Score used by deformation network.

❖ Enhances semantic representation in registration.

Qin and Li (2023):

❖ Use the score as a spatial weighting function for similarity terms in the loss.

❖ Depart from conventional Gaussian noise modeling.



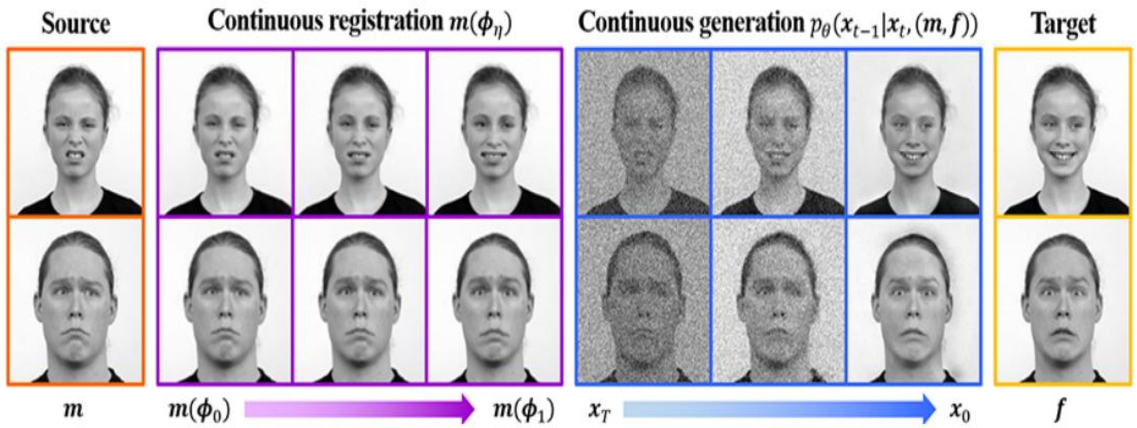
Challenges:

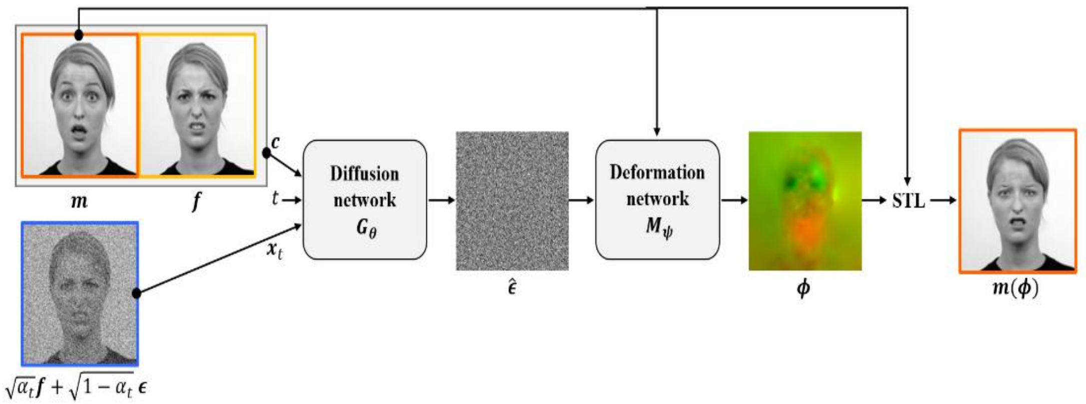
❖ High computational cost due to thousands of sampling steps.

❖ Few existing works in registration; adaptation requires non-trivial reformulations.

Diffusion Models: DiffuseMorph

- DiffuseMorph involves a diffusion network and a deformation network.
  - The diffusion network learns a conditional score function (added noise)
  - The deformation network uses the latent feature in the reverse diffusion process to estimate the deformation field.
- The registration process is a one-step procedure, as the fixed mage is the target image at the end of the reverse diffusion process ($t = 0$), and it is already given. As a result, there is no need for time-consuming reverse diffusion steps to synthesize a target image from the moving image.
- Furthermore, DiffuseMorph offers the added capability of producing continuous deformations through the interpolation of the learned space.



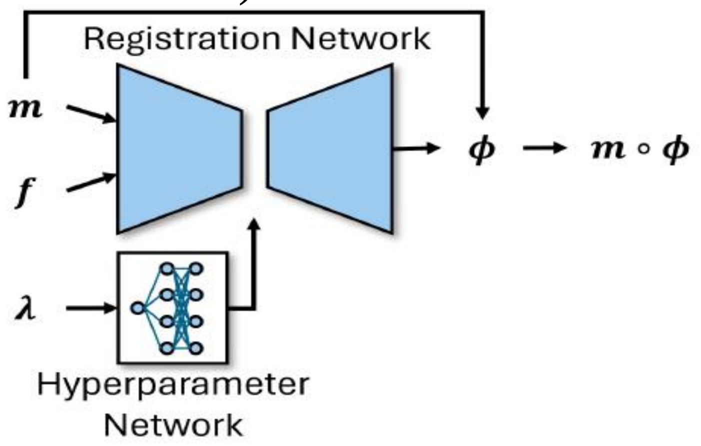


Kim, B., Han, I., & Ye, J. C. (2022). *DiffuseMorph: Unsupervised Deformable Image Registration Using Diffusion Model* (No. arXiv:2112.05149). arXiv. https://doi.org/10.48550/arXiv.2112.05149

Hyperparameter Conditioning

Motivation:
- Traditional registration models require re-training for each hyperparameter setting (e.g., regularization weight).
- Inspired by HyperNetworks (Ha et al., 2017) and Hyperparameter Optimization (Franceschi et al., 2018).

Key Idea:
- Condition the registration network on hyperparameter values (e.g., deformation smoothness).
- Sample hyperparameters during training and generate deformation field.
- Compute loss with same sampled hyperparameter value to update network.



Benefits:
- Efficient hyperparameter tuning without training multiple models.
- Enables dynamic control of deformation regularization.

HyperMorph (Hoopes et al., 2022a):
- **Two-network system:**
  - ❖ Hypernetwork: Takes in regularization hyperparameter, outputs weights for the U-Net.
  - ❖ U-Net (VoxelMorph): Generates deformation field for image warping.
- Hyperparameter sampled from uniform distribution during training.
- Best hyperparameter value selected via gradient descent on validation Dice score.

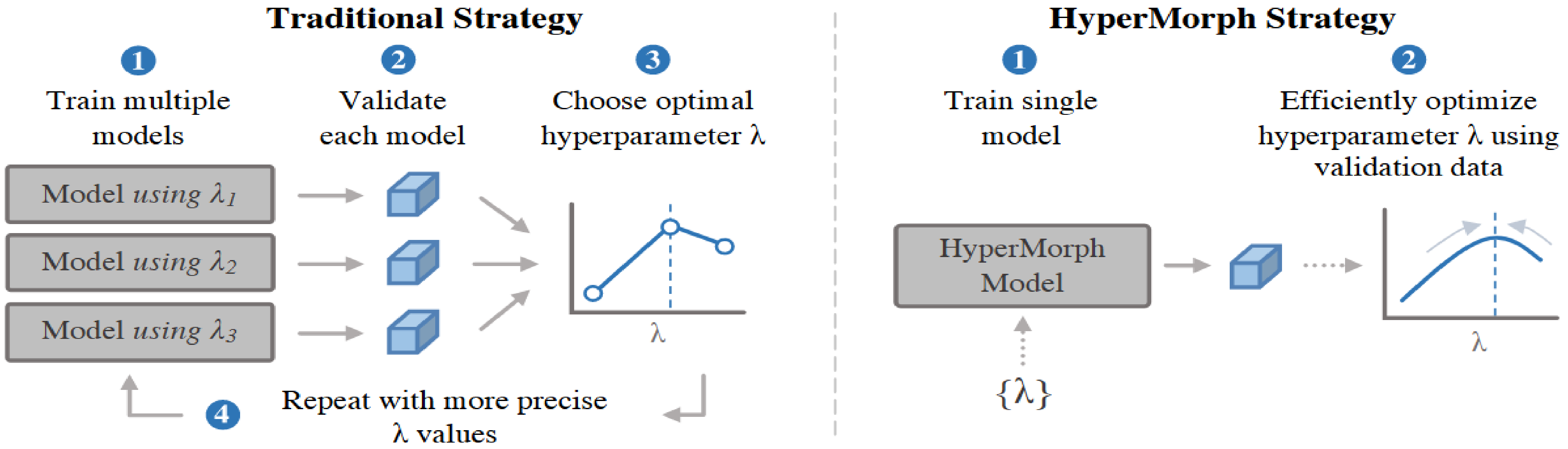
Other Approaches:

Mok and Chung (2021b): Affine transformation of regularization maps based on sampled hyperparameter. Lightweight mapping network used for conditioning.

Chen et al. (2023b): Extended conditioning to Transformer-based models via conditional layer normalization. Both use grid search to select optimal hyperparameter.

HyperMorph

- The HyperMorph learns a hypernetwork that takes in an input hyperparameter and modulates a registration network to produce the optimal deformation field for that hyperparameter value.
- HyperMorph comprises two ConvNets: a hypernetwork and a UNet-like registration network such as VoxelMorph.
    - The hypernetwork estimates the weights of the U-Net based on the provided hyperparameter value for the diffusion regularizer
    - The U-Net generates a deformation field to warp the moving image.
- In each training step, the hyperparameter value is randomly sampled from a uniform distribution, and the loss is computed using the same sampled value
- After training, the best-performing hyperparameter value is acquired using gradient descent. In this process, the network weights are fixed, and an optimizer iteratively updates the hyperparameter based on a target objective function such as the Dice score.



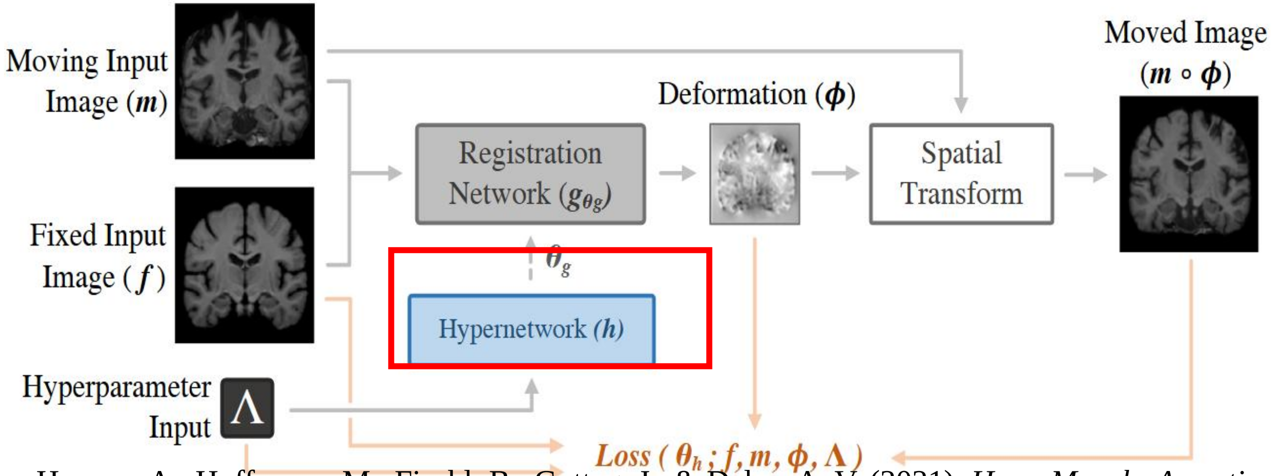
Hoopes, A., Hoffmann, M., Fischl, B., Guttag, J., & Dalca, A. V. (2021). *HyperMorph: Amortized Hyperparameter Learning for Image Registration* (No. arXiv:2101.01035). arXiv. https://doi.org/10.48550/arXiv.2101.01035

HyperMorph

- **Goal:** Model how loss hyperparameters Λ influence the registration.
- Define a hypernetworks function $h_{\theta_h}(\Lambda) = \theta_g$ with parameters that takes as input sample values for Λ and outputs the parameters of the registration network θ_g.
- To learn the optimal parameter θ_h, we optimize the loss

$$L_h(\theta_h; D) = E_{\Lambda \sim p(\Lambda)}[L(\theta_h; D, \Lambda)]$$

where D is the dataset of images, $p(\Lambda)$ is a prior probability over the hyperparameters (uniform distribution here), and L is a registration loss involving hyperparameters Λ.



Datasets:

ABIDE, GSP, PPMI, ADNI, UK Biobank — 3D T1-weighted brain MRIs.

Main Results:

Accuracy: Comparable Dice scores to grid search: ABIDE: HyperMorph Dice = 0.833; Grid Search Dice = 0.831; GSP: HyperMorph Dice = 0.845; Grid Search Dice = 0.846

Efficiency:

1 HP tuning: 5.2× fewer GPU-hours

2 HPs (e.g., λ, learning rate): 10.5× fewer GPU-hours

Robustness: Lower standard deviation in Dice across random initializations.

Adaptivity:

Optimal λ varies across populations and brain structures.

Enables personalized tuning: different λ values for hippocampus vs. cerebellum.

Hoopes, A., Hoffmann, M., Fischl, B., Guttag, J., & Dalca, A. V. (2021). *HyperMorph: Amortized Hyperparameter Learning for Image Registration* (No. arXiv:2101.01035). arXiv. https://doi.org/10.48550/arXiv.2101.01035

Symmetric and Cycle Consistency

Objective: Impose structural constraints to ensure invertibility and improve regularity in deformation-based registration models.

Symmetric Consistency: Focuses on the deformation field ϕ, not just the transformation T: $\phi_{A\to B} \circ \phi_{B\to A} = Id$

- ❖ Encourages the forward and backward deformation fields to be mutual inverses.
- ❖ Typically implemented using a single shared network to predict both directions.

Cycle Consistency: - A special case of transitivity, often with C = A, $T_{B\to A} \circ T_{A\to B}(A) = A$

- ❖ Ensures that registering an image to another and back yields the original image.
- ❖ Used in unsupervised learning and multi-domain settings (e.g., GAN-based registration)

Intuition: Enforcing these consistencies implicitly regularizes learned deformations and helps preserve anatomical plausibility.

Implementation Approaches

- ➢ **Symmetric Consistency Loss:**

$$\mathcal{L}_{\text{sym}} = \|\phi_{A\to B} \circ \phi_{B\to A} - \text{Id}\|_F^2$$

- ➢ **Cycle Consistency Loss:**

$$\mathcal{L}_{\text{cyc}} = \|I_A - I_A \circ T_{B\to A} \circ T_{A\to B}\|^2$$

Neural Network Setup:

A single network outputs both ϕ:A→B and ϕ:B→A.

The total loss may include:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{sim}} + \lambda_1 \mathcal{L}_{\text{sym}} + \lambda_2 \mathcal{L}_{\text{cyc}}$$

Key Benefits:

- ❖ Encourages invertibility of deformation fields.
- ❖ Enhances registration accuracy and stability.
- ❖ Complements smoothness.

Symmetric Consistency: GradICON

Goal: Learn diffeomorphic image registration mappings without explicit spatial regularization.

Key Idea: Use gradient-based inverse consistency

$$L_{GradICON} = \left\| \nabla[\Phi_\theta^{AB} \circ \Phi_\theta^{BA}] - I \right\|_F^2 \quad v.s. \quad L_{ICON} = \left\| \Phi_\theta^{AB} \circ \Phi_\theta^{BA} - Id \right\|_2^2$$

Motivation:

Avoid instability of pixel-wise inverse consistency.

Operate on Jacobians to ensure smooth transformations.

Implicit Regularization:

$$\mathbb{E}[\mathcal{L}_{\text{GradICON}}] \approx \epsilon^2 \left\| [\nabla\Phi^{AB}]^{-1} \sqrt{\det \nabla\Phi^{AB}} \right\|_F^2 + \epsilon^2 \left\| [\nabla\Phi^{BA}]^{-1} \right\|_F^2$$

Benefits:

Enforces smoothness and topology preservation.

Avoids hand-tuning of regularization weights.

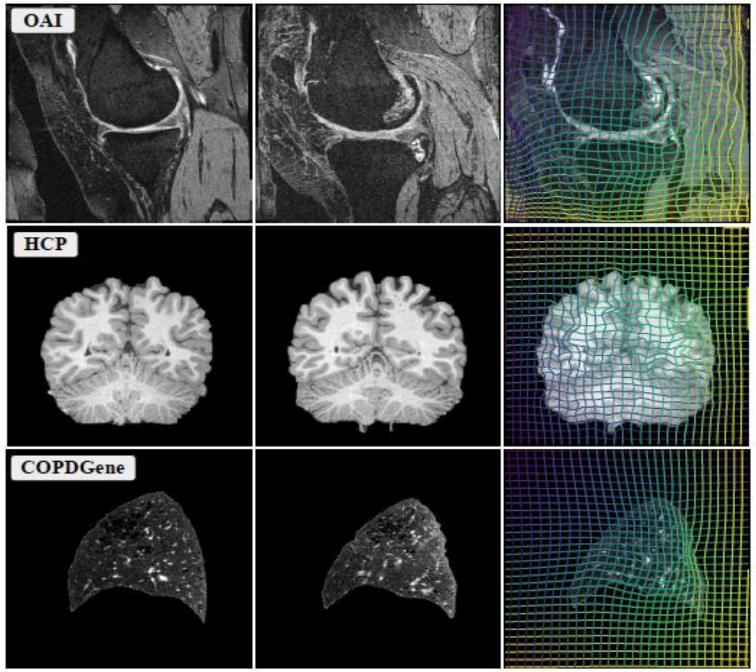


Figure 1. Example source (*left*), target (*middle*) and warped source (*right*) images obtained with our method, trained with a **single protocol**, using the proposed GradICON regularizer.

Network:

Multi-resolution U-Net-style architecture.

Predicts forward and backward deformation fields.

Loss Function:

$$\mathcal{L} = -\text{LNCC}(I_A, I_B \circ \Phi^{AB}) + \lambda \mathcal{L}_{\text{GradICON}}, \quad \lambda = 1$$

Benchmark Results:

OAI (Knee MRI): Dice = 71.2% (vs. 68.4% baseline)

HCP (Brain MRI): Dice = 80.5% (vs. 79.8%)

COPDGene (Lung CT): TRE = 2.68mm (vs. 3.01mm)

DirLab (CT): TRE = 1.31mm, Negative Jacobian = 0.0002%

Tian, L., Greer, H., Vialard, F.-X., Kwitt, R., Estépar, R. S. J., Rushmore, R. J., Makris, N., Bouix, S., & Niethammer, M. (2023). *$\texttt{GradICON}$: Approximate Diffeomorphisms via Gradient Inverse Consistency* (No. arXiv:2206.05897). arXiv. http://arxiv.org/abs/2206.05897

CycleMorph: Cycle Consistency

Key Architecture:

- Two networks $G_X : (X, Y) \to \phi_{XY}$ and $G_Y : (Y, X) \to \phi_{YX}$ generate forward and reverse deformation fields.
- Deformed images: $\hat{Y} = T(X, \phi_{XY})$, $\hat{X} = T(Y, \phi_{YX})$
- Cycle: $\tilde{X} = T(\hat{Y}, \hat{\phi}_{YX})$, $\tilde{Y} = T(\hat{X}, \hat{\phi}_{XY})$

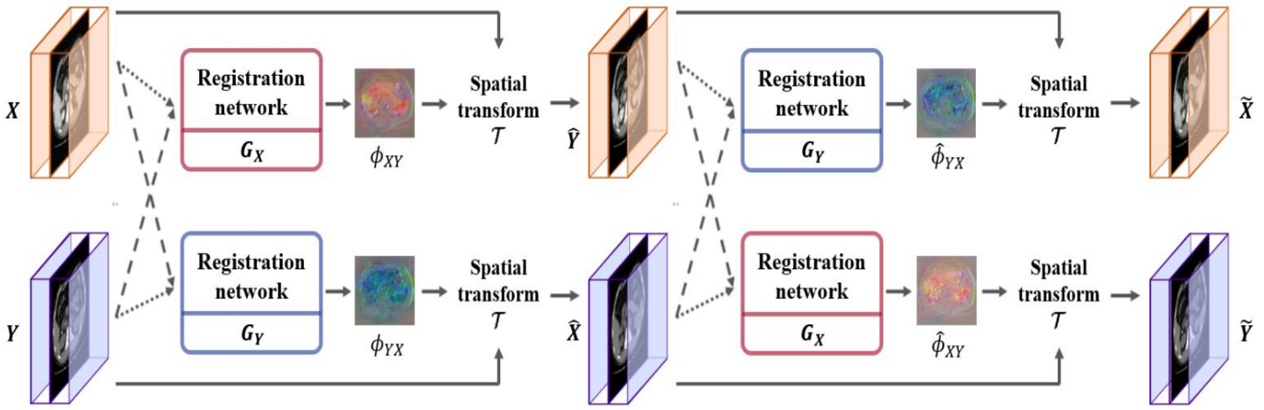
Cycle Construction:

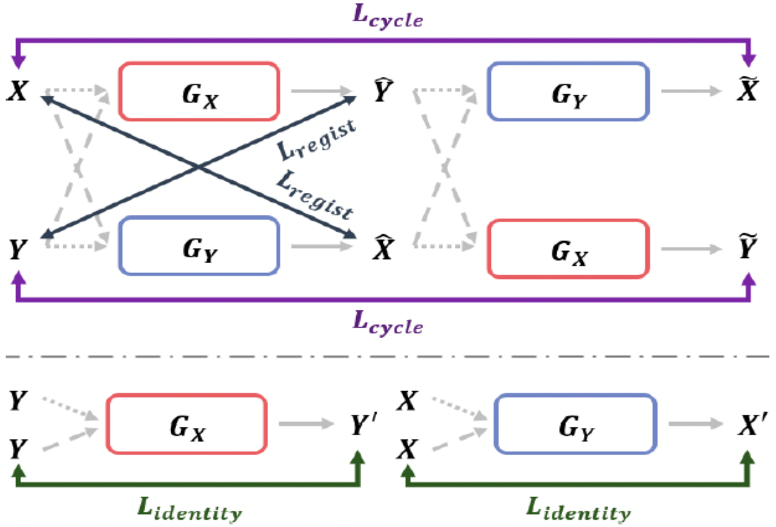
$$\hat{Y} = T(X, \phi_{XY}), \quad \hat{X} = T(Y, \phi_{YX})$$
$$\tilde{X} = T(\hat{Y}, \phi_{YX}), \quad \tilde{Y} = T(\hat{X}, \phi_{XY})$$

Total Loss:

$$L(X,Y,G_X,G_Y) = L_{regist}(X,Y,G_X) + L_{regist}(Y,X,G_Y) + \alpha L_{cycle}(X,Y,G_X,G_Y) + \beta L_{\mathrm{identity}}(X,Y,G_X,G_Y)$$

where $\mathrm{L}_{regist}, L_{cycle}$ and $\mathrm{L}_{identity}$ are the registration loss, cycle loss and identity loss, respectively, and α, β are hyperparameters.





Kim, B., Kim, D. H., Park, S. H., Kim, J., Lee, J.-G., & Ye, J. C. (2020). *CycleMorph: Cycle Consistent Unsupervised Deformable Image Registration* (No. arXiv:2008.05772). arXiv. https://doi.org/10.48550/arXiv.2008.05772

CycleMorph: Cycle Consistency

- **Registration loss:** $L_{regist}(X,Y,G_X) = -(T(X,\phi_{XY}) \otimes Y) + \lambda \sum \left\|\nabla \phi_{XY}\right\|^2$

where λ is a hyperparameter, $\otimes$ denotes the local cross correlation.

- **Cycle loss:** $L_{cycle}(X,Y,G_X,G_Y) = \left\|T(\hat{Y},\widehat{\phi_{YX}}) - X\right\|_1 + \left\|T(\hat{X},\widehat{\phi_{XY}}) - Y\right\|_1$
- **Identity loss:** $L_{identity}(X,Y,G_X,G_Y) = -[T\big(Y,G_X(Y,Y)\big) \otimes Y$
- $+T\big(X,G_Y(X,X)\big) \otimes X]$

Multiscale Refinement:

- **Global Network:** Coarse registration at low resolution
- **Local Patch Network:** Refines deformation in 643 local 3D volumes
- **Final Deformation:** ϕ = ϕ_global + ϕ_local

Datasets Evaluated:

Brain MRI (IBSR and LPBA40): Inter-subject registration across anatomical regions

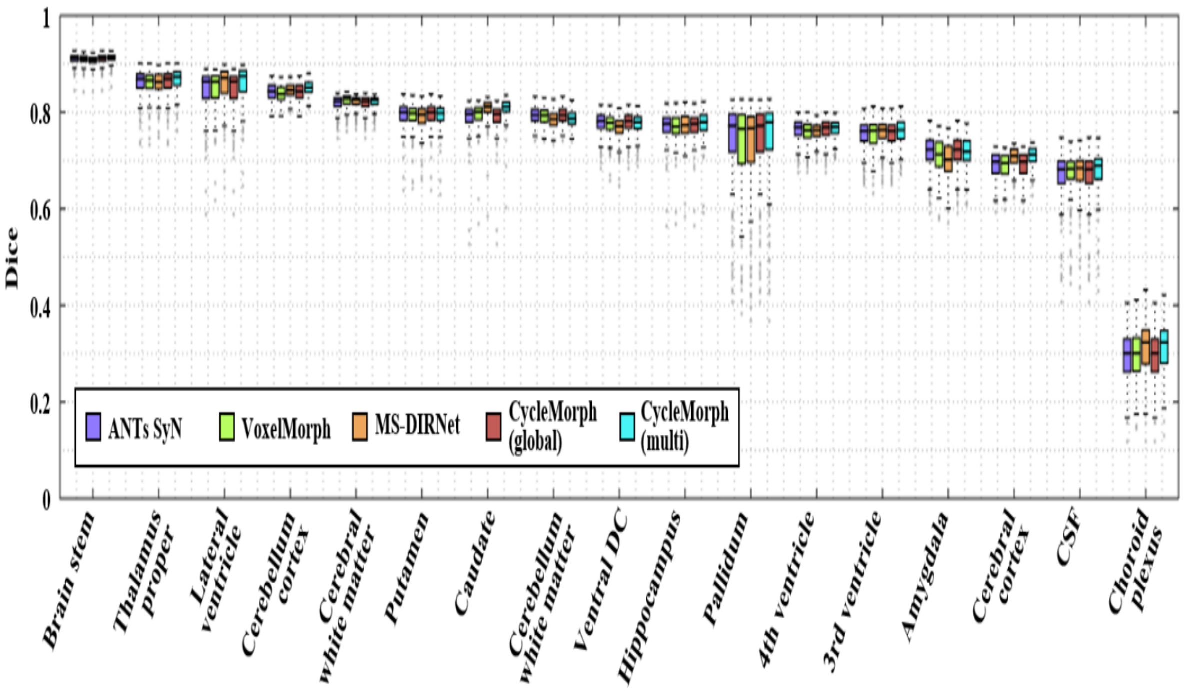
Liver CT (LiTS): Multiphase intra-subject organ alignment

Facial Expression: Landmark alignment for facial emotion transfer

Performance Highlights:

Brain MRI: Dice = 0.756 (CycleMorph) vs. 0.749 (VoxelMorph) vs. 0.752 (ANTs)

Liver CT: Target Registration Error (TRE) = 3.9 mm vs. 4.7 mm (Elastix), 30x faster



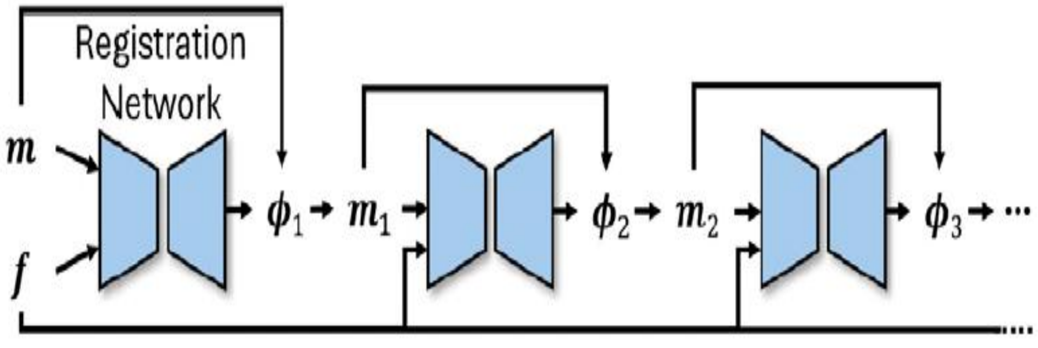
Progressive and Multi-Scale Image Registration

Two Major Strategies:

- ❖ Progressive Registration: Sequentially refine deformation fields by cascading registration networks.
- ❖ Multi-Scale Registration: Employ image pyramids to learn coarse-to-fine deformations across resolutions.

Progressive Framework (e.g., VTN, VR-Net):

- Decomposition of large displacement into smaller steps.
- Each subnetwork G_i predicts ϕ_i and updates the moving image: $I_i = T(I_{i-1}, \phi_i)$
- Final deformation field: $\Phi = \phi_n \circ \cdots \circ \phi_1$



(a)

Panel (a) outlines the framework for progressive image registration

Cycle-Based Optimization (VR-Net):

- ❖ Linearizes nonlinear registration objective with first-order Taylor expansion.
- ❖ Solves two convex problems: (1) similarity update and (2) regularization.
- ❖ Each network block refines deformation iteratively:

$$\phi^{(k+1)} = \phi^{(k)} + \Delta\phi^{(k)}$$

Progressive and Multi-Scale Image Registration

Multi-Scale Pyramid Frameworks:

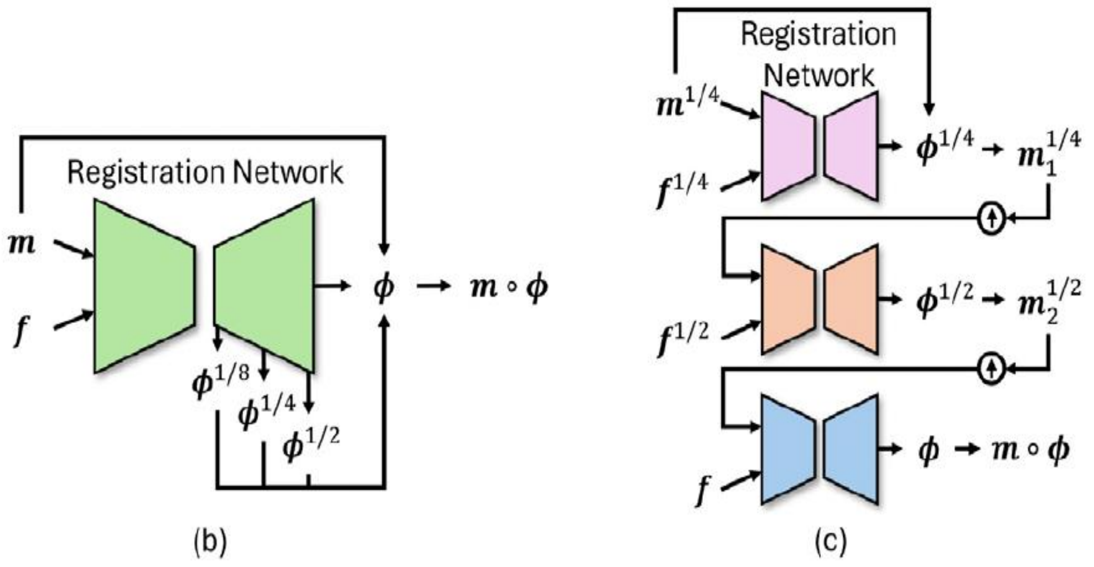
- **LapIRN:** 3 networks at increasing resolution with skip connections and progressive refinement.
- **Self-Recursive Contextual Net (Hu et al.):** Shared weights; recursively refines ϕ using same network at different scales.

Progressive Training Techniques:

- **De Vos et al.:** Train ConvNets at different resolutions stage-wise; no regularizer due to B-spline.
- **Eppenhof et al.:** Gradually increase input resolution and network depth during training.

Transformer-Based Approaches:

- **NICE-Trans:** Dual-path ConvNet encoder + Transformer decoder predicts both affine + deformable fields.
- **Ma et al. (2023):** Swin Transformer blocks at bottleneck refine ϕ progressively; final ϕ formed via upsampling and convolution.



Panels (b) and (c) illustrate two representative strategies for multi-scale image registration in learning-based methods: (b) a single-network approach that aggregates deformation fields across scales (e.g., im2grid), and (c) a multi-network approach where each resolution scale is handled by a separate network (e.g., DLIR and LapIRN).

Vision Transformer for Affine Registration

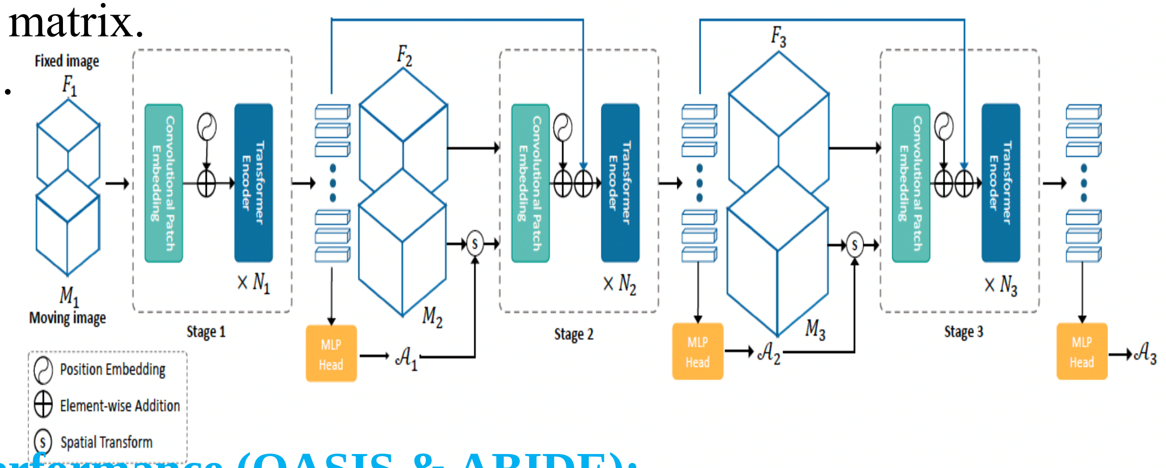
Motivation:

- ❖ Traditional affine methods are accurate but computationally intensive.
- ❖ CNNs lack global context, struggle with large misalignments.

Goal: Design a fast and robust model for 3D affine registration using Vision Transformers.

Architecture:

Three-stage coarse-to-fine pyramid.
Each stage: Patch embedding → Transformer → MLP → Affine matrix.
The moving image is warped progressively before the next stage.



Progressive Multi-Scale Training:

- Use 3 scales: $64\times64\times64$, $128\times128\times128$, $192\times192\times192$.
- Deformation refinement across levels:

$$A_i = \mathrm{MLP}_i(\mathrm{Transformer}_i(\mathrm{Embed}(F_i, M_i))), \quad M_{i+1} \leftarrow \phi(A_i)(M_{i+1})$$

- Residual skip connections for feature propagation.

Loss Function:

$$\mathcal{L}_{\mathrm{sim}} = \sum_{i=1}^{3} \frac{-1}{2^{3-i}} \cdot \mathrm{NCC}_w(F_i, M_i(\phi)), \quad \mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{sim}} + \lambda \cdot \mathcal{L}_{\mathrm{seg}}$$

Performance (OASIS & ABIDE):

Dice score: 0.757 (OASIS), 0.724 (ABIDE) — best among 6 baseline methods.
HD95 (mm): 3.12 (OASIS), 3.59 (ABIDE)
Runtime: 0.09s (GPU, C2FViT) vs. 6.6–38s (ANTs/Elastix)

Mok, T.C., Chung, A., 2022a. Affine medical image registration with coarse-to-fine vision transformer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20835–20844.

Content

1. Introduction to Image Registration
2. ConvNets based Registration
3. Network Architectures for Registration
4. Applications of Image Registration

Overview: : Applications of Image Registration

Core Goals:
- ❖ Match anatomical or structural features across time, modality, or subjects.
- ❖ Enable direct voxel- or pixel-wise comparison between aligned images.

Application Domains:
- ❖ **Medical Imaging:** Diagnosis, image-guided surgery, treatment planning.
- ❖ **Remote Sensing:** Satellite image alignment for temporal analysis.
- ❖ **Computer Vision:** Image stitching, motion tracking, 3D modeling.
- ❖ **Augmented/Virtual Reality:** Overlay alignment between virtual and real scenes

Types of Registration:
- ❖ **Modality:** Intra-modal (e.g., MRI-MRI), Inter-modal (e.g., CT-MRI)
- ❖ **Transformation:** Rigid, affine, deformable (non-rigid)
- ❖ **Dimensionality:** 2D-2D, 3D-3D, or 2D-3D registration

1. Remote Sensing and Environmental Monitoring:
- ➢ Align multi-temporal satellite images for land-use change, disaster assessment, deforestation tracking.
- ➢ **Tools:** Sentinel-2, Landsat series, Google Earth Engine.

2. Augmented and Virtual Reality (AR/VR):
- ➢ Align real-world scenes with virtual objects using visual SLAM and marker tracking.
- ➢ **Example:** Microsoft HoloLens, Meta Quest Pro.

3. Robotics and Autonomous Navigation:
- ❖ Use LiDAR and camera data fusion via registration to build and update 3D maps.
- ❖ Core to SLAM (Simultaneous Localization and Mapping) frameworks.

4. Industrial Inspection and Manufacturing:
- ❖ Register 3D CAD models to sensor data for defect detection or quality control.

Applications in Biomedical Sciences

1. Longitudinal Studies:
- ❖ Track progression of neurodegenerative diseases (e.g., Alzheimer’s) by aligning baseline and follow-up MRIs.

2. Multi-Modal Fusion:
- ❖ Fuse PET (functional) with MRI (structural) for tumor detection and monitoring.
- ❖ Example: PET-MRI registration enhances precision in oncology.

3. Intra-Operative Guidance:
- ❖ Register pre-operative MRI with real-time ultrasound during brain surgery.

4. Radiotherapy Planning:
- ❖ Align planning CT with daily Cone-Beam CT (CBCT) for precise dose delivery in cancer treatment.

5. Atlas-Based Analysis:
- ➢ Build anatomical atlases (e.g., MNI atlas) by deformably registering subjects to a common template.
- ➢ Enables population-wide analysis of brain shape and volume.

6. Genotype-Phenotype Association:
- ➢ Align imaging-derived phenotypes with genotypic data (GWAS, eQTL). E.g., detect genetic variants associated with hippocampal volume.

7. Disease Subtyping and Progression Modeling:
- ➢ Register multi-subject, multi-timepoint scans to identify disease trajectories.

8. Inter-Group Comparison:
- ➢ Align scans to compare aging, disease, or treatment effects across cohorts. Applications in aging research, psychiatry, and developmental neuroscience.

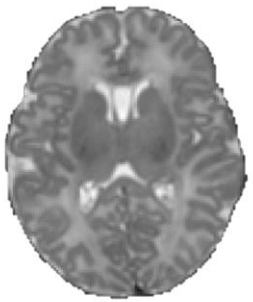
Generation of Anatomy-Realistic 4D Infant Brain Atlases with Tissue Maps Using Generative Adversarial Networks

Dr. Gang Li

Introduction: Background

- **Brain development during infancy**
  - Complex and dynamic
  - Significant **structural** and **volumetric** changes
- **Infant brain atlas construction**
  - Crucial to generate **spatiotemporal (4D) volumetric atlases** with **continuously sampled** time points
  - Essential for **downstream tasks**, e.g., atlas-guided segmentation and spatial normalization
- **Infant brain MR images (T1w/T2w)**
  - **Low** tissue contrast and **dynamic** change in appearance
- **Challenging** to generate **accurate and anatomically meaningful** 4D infant atlases, particularly, for younger ages

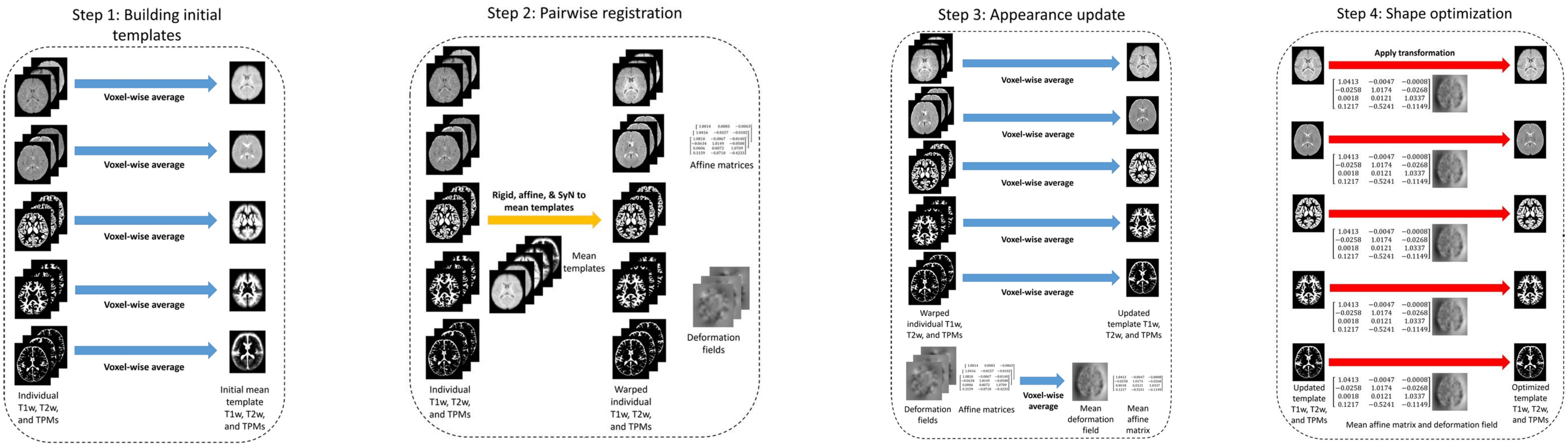
00 Months



00 Months

Introduction: Existing Methods and Limitations

- **Traditional methods**
  - Iterative atlas construction using symmetric group-wise normalization (SyGN) (*Chen, L., et al., NeuroImage 2022*)

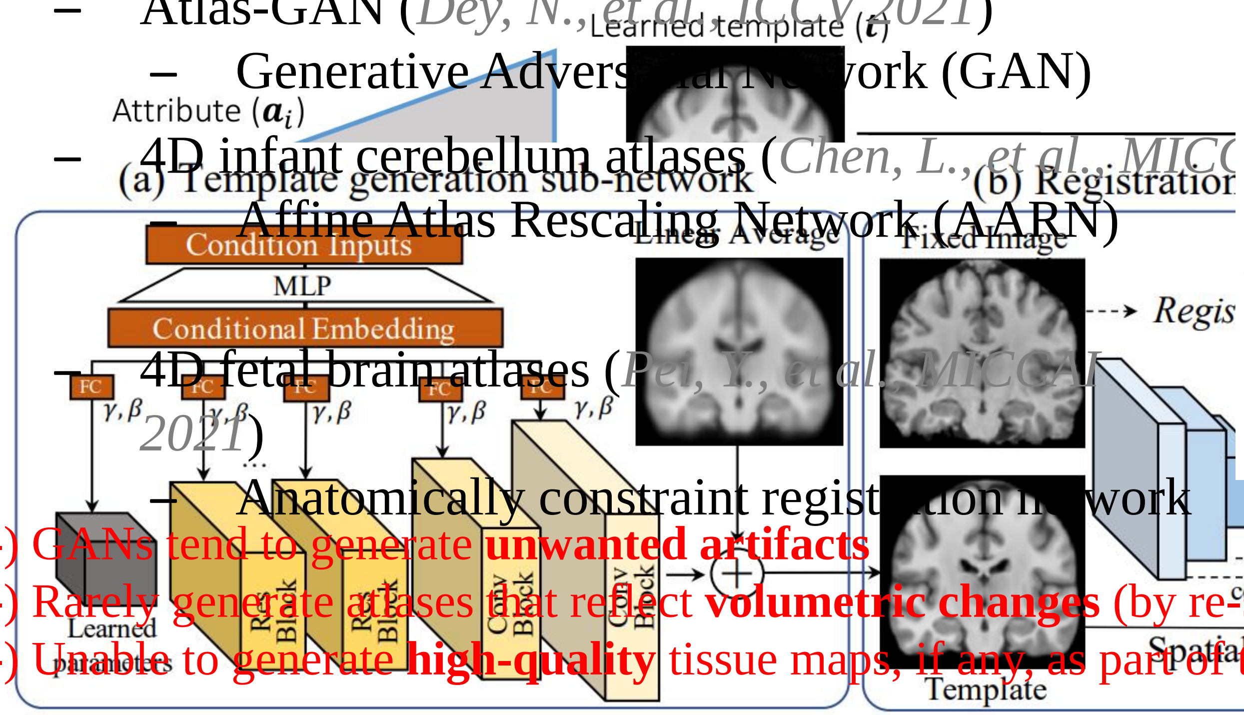


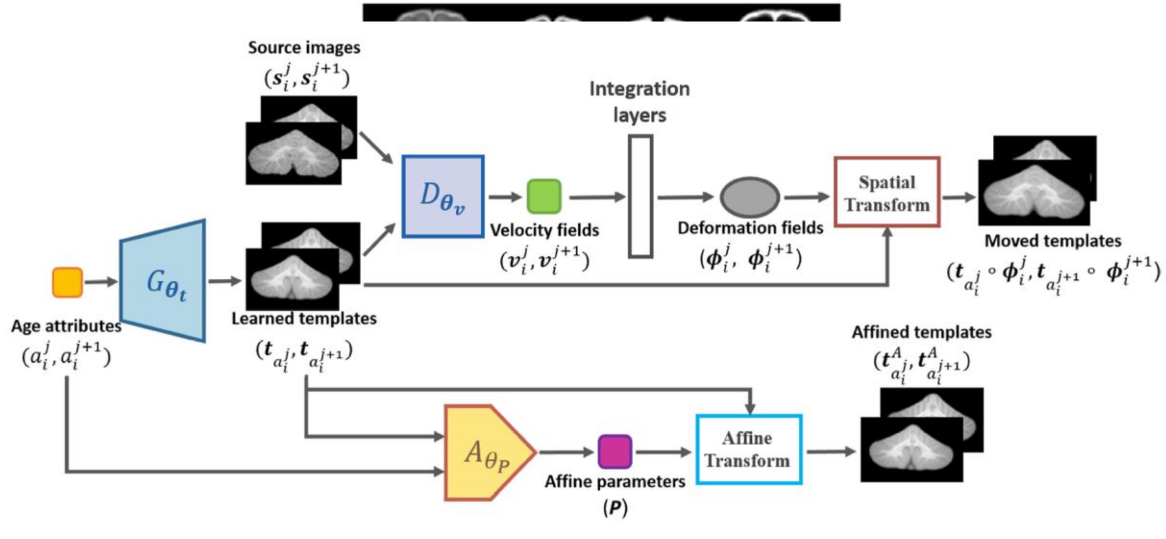
(-) Separately built at **discrete** time points
(-) Require **iterative** and **computationally expensive** non-linear registration

Chen, L., et al., A 4D Infant Brain Volumetric Atlas Based on the UNC/UMN Baby Connectome Project (BCP) Cohort. NeuroImage (2022).
Also see: https://www.nitrc.org/projects/uncbcp_4d_atlas/

Introduction: Existing Methods and Limitations

- **Deep learning-based methods**
  - Conditional atlas building using VoxelMorph (*Dalca, A., et al., NIPS 2019*)
  - Atlas-GAN (*Dey, N., et al., ICCV 2021*)
    - Generative Adversarial Network (GAN)
  - 4D infant cerebellum atlases (*Chen, L., et al., MICCAI 2021*)
    - Affine Atlas Rescaling Network (AARN)
  - 4D fetal brain atlases (*Pei, Y., et al., MICCAI 2021*)
    - Anatomically constraint registration network





(-) GANs tend to generate **unwanted artifacts**

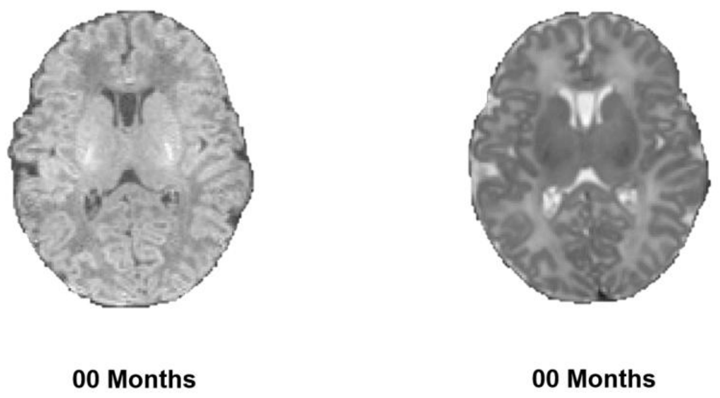
(-) Rarely generate atlases that reflect **volumetric changes** (by re-scaling the generated atlases back to age-specific spaces)

(-) Unable to generate **high-quality** tissue maps, if any, as part of the atlas

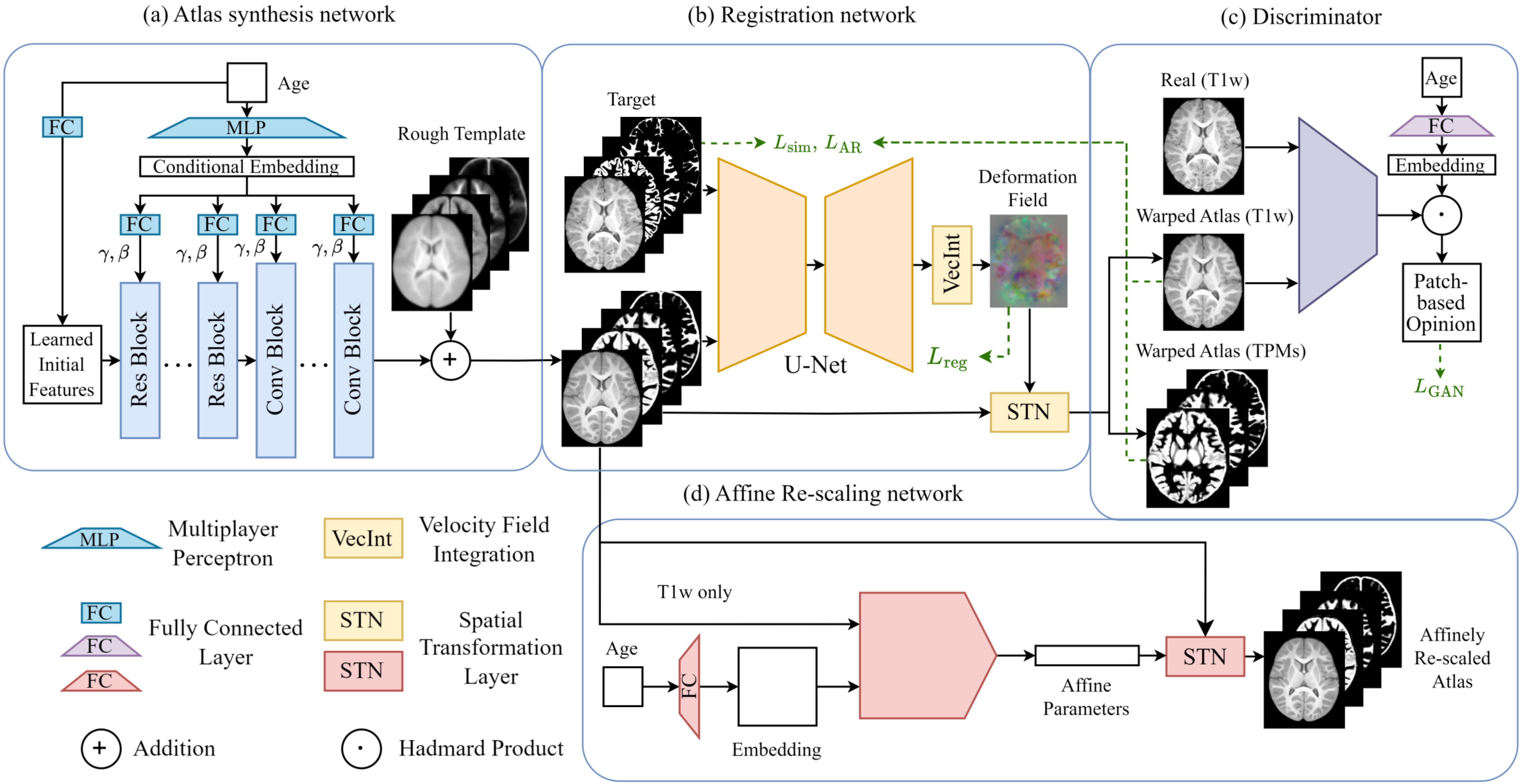
Dalca, A., et al., Learning Conditional Deformable Templates with Convolutional Networks. NIPS (2019).
Dey, N., et al., Generative Adversarial Registration for Improved Conditional Deformable Templates. ICCV (2021).
Chen, L., et al., Construction of Longitudinally Consistent 4D Infant Cerebellum Atlases Based on Deep Learning. MICCAI (2021).
Pei, Y., et al., Learning Spatiotemporal Probabilistic Atlas of Fetal Brains with Anatomically Constrained Registration Network. MICCAI (2021).

Challenge and Aims

- **Challenge**
  - **Low and dynamic tissue contrast** of infant brain MR images
- **Aims**
  - Provide **explicit guidance** from tissue maps to help generate anatomically more realistic intensity atlases
  - Produce **tissue maps** alongside intensity atlases
  - Affinely scale the predicted atlas automatically to accurately **reflect volumetric change**

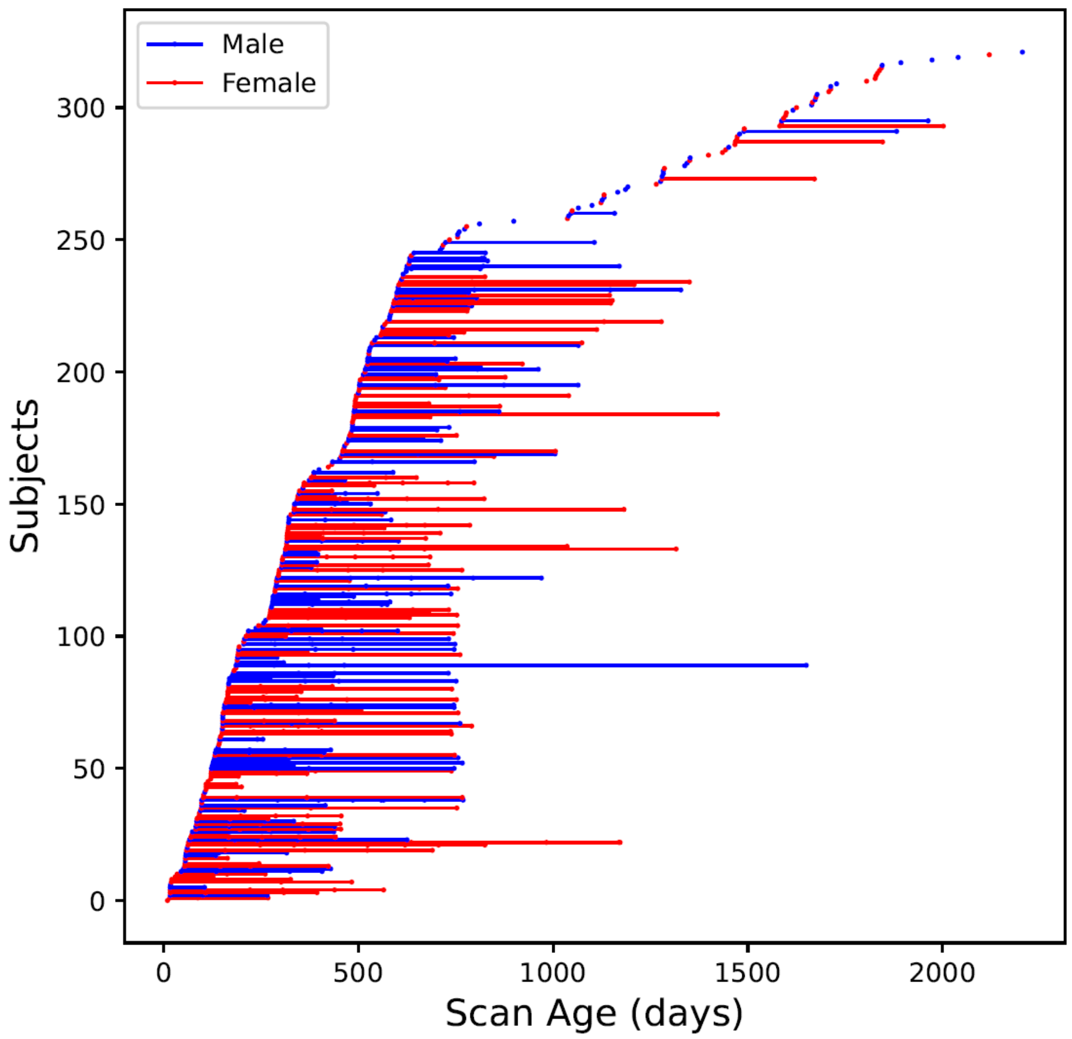


Method: Deformable Atlas Construction and Affine Re-scaling Network



Experiments

- **Dataset**
  - 699 MRI scans (T1w) from 322 subjects from the **UNC/UMN Baby Connectome Project (BCP)** (*Howell, B.R., et al., NeuroImage 2019*)
  - $0.8 \times 0.8 \times 0.8\ mm^3$
  - Bias-corrected, skull-stripped, and segmented into white matter (WM), cortical gray matter (GM), and cerebrospinal fluid (CSF) using iBEAT V2.0 at http://www.ibeat.cloud/ (*Wang, L., et al., Nat Protoc 2023*)
- **Comparison**
  - Atlas-GAN (*Dey, N., et al., ICCV 2021*)
- **Evaluation Metric**
  - Dice Similarity Coefficient (DSC)



Howell, B.R., et al., The UNC/UMN Baby Connectome Project (BCP): An Overview of the Study Design and Protocol Development. NeuroImage (2019).
Wang, L., et al., iBEAT V2.0: A Multisite-applicable, Deep Learning-based Pipeline for Infant Cerebral Cortical Surface Reconstruction. Nat Protoc (2023).
Dey, N., et al., Generative Adversarial Registration for Improved Conditional Deformable Templates. ICCV (2021).

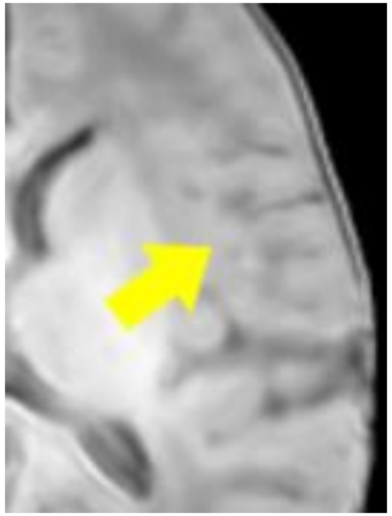
Results: Quantitative

- **Experiments**
  - 699 scans are split by subject into 629 and 70 scans for training and testing, respectively
- **Result**
  - Our method yields greatly improved performance in terms of Dice Similarity Coefficient (DSC)

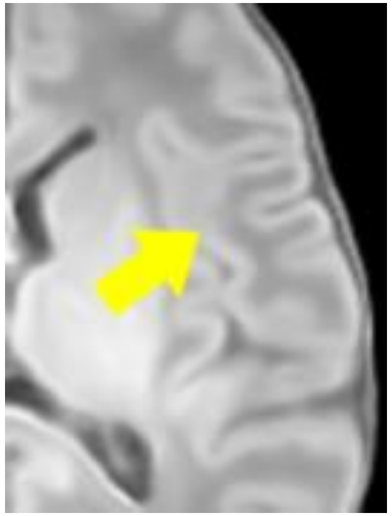
	DSC, %, $\bar{x}$ (s)		
	White Matter	Cortical Gray Matter	Cerebrospinal fluid
Atlas-GAN	56.96 (2.39)	51.28 (2.61)	34.17 (3.71)
Ours	**81.39 (1.86)**	**83.90 (2.32)**	**60.22 (4.68)**

Results: Qualitative

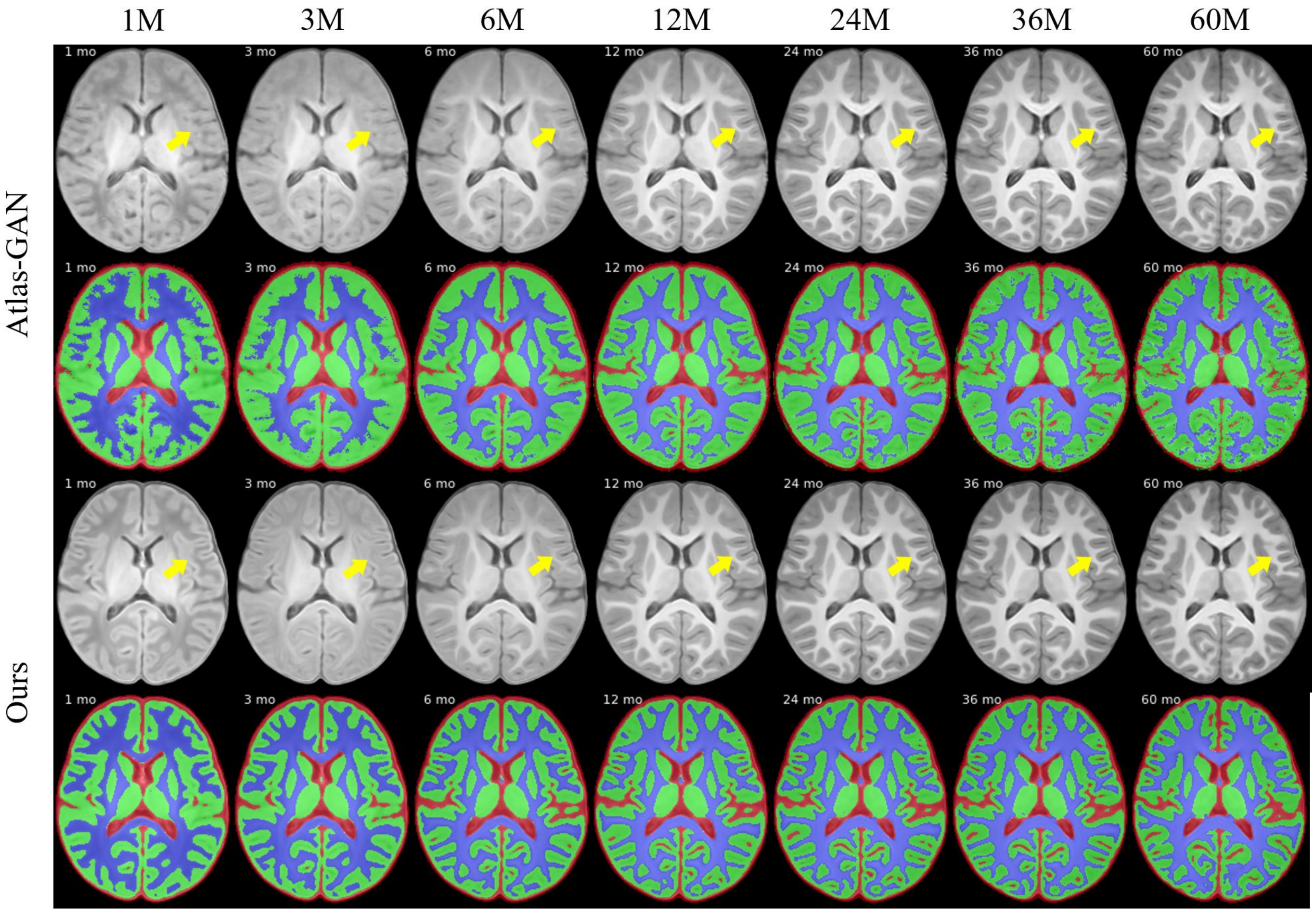
- **Result**
  - Improved tissue maps with **more accurate details**
  - **Sharper** and **anatomically more realistic** intensity atlases



Atlas-GAN, 1M

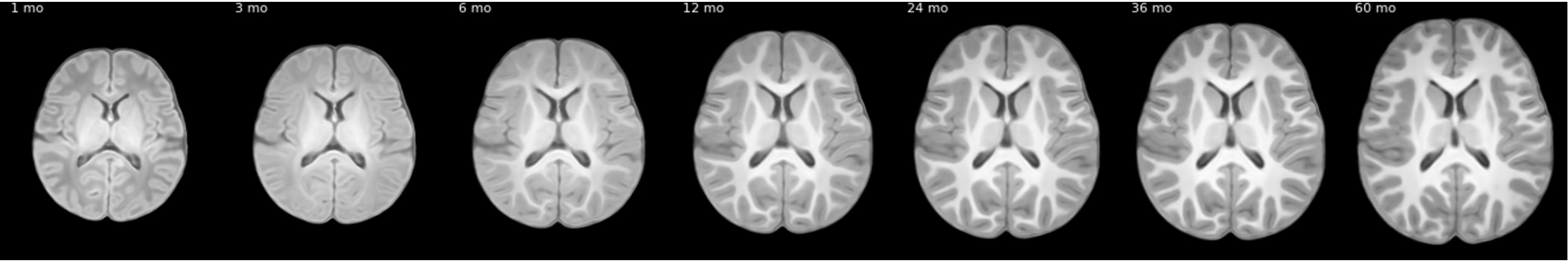


Ours, 1M



Results: Qualitative

- **Result**
  - Generated atlases at representative ages re-scaled from the population common space to the **age-specific spaces** using the affine re-scaling network



Conclusion

- We present a deep learning-based framework with **explicit anatomical guidance** for the construction of **4D** infant brain volumetric atlases, which can jointly
  - Produce **tissue maps** alongside **anatomically realistic** intensity atlases, and
  - Affinely scale the predicted atlas to **reflect volumetric change** during early development.

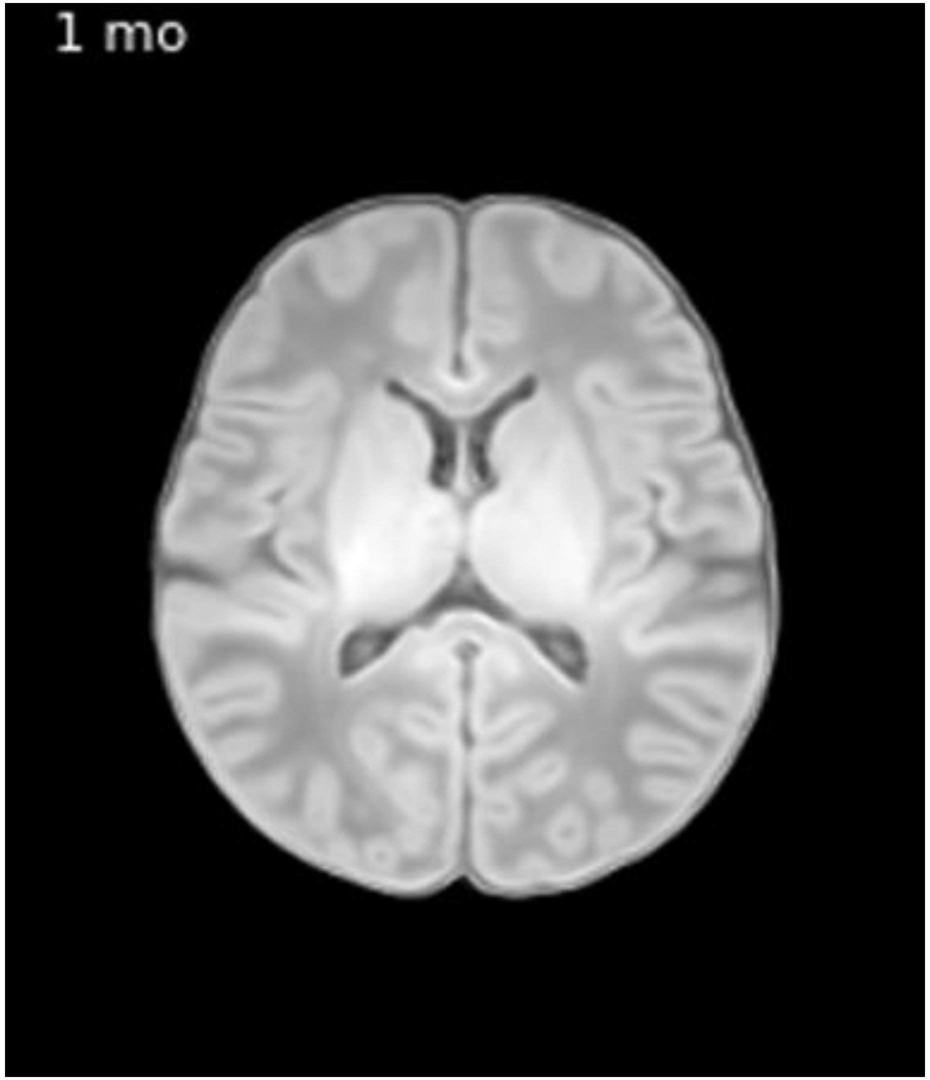


Table 3. A summary of the publicly available benchmark dataset for medical image registration.

Dataset	Anatomy	Cohort Type	Modality	Highlights
IXI[a]	Brain	Healthy Controls	T1w, T2w, PDw MRI	Nearly 600 MRI images with cortical and subcortical label maps from prior studies (Liu et al., 2024; Chen et al., 2022b; Hoopes et al., 2022c).
LUMIR (Dorent et al., 2024)	Brain	Healthy Controls	T1w MRI	Part of Learn2Reg 2024 (Dorent et al., 2024), using the OpenBHB dataset (Dufumier et al., 2022); 4,014 MRIs from ten public datasets with label maps and landmarks.
LPBA40 (Shattuck et al., 2008)	Brain	Healthy Controls	T1w MRI	40 MRI scans affine-transformed to a common atlas with 50 manually delineated brain structures.
Mindboggle (Klein and Tourville, 2012)	Brain	Healthy Controls	T1w MRI	101 MRIs affine-aligned to an atlas with 106 manually delineated brain structures.
OASIS (Marcus et al., 2007; Hoopes et al., 2022b)	Brain	Alzheimer's disease	T1w MRI	416 MRIs from OASIS-1 (Marcus et al., 2007) with label maps generated using FreeSurfer and SAMSEG, used in Learn2Reg 2021 (Hering et al., 2022).
BraTS-Reg (Baheti et al., 2021)	Brain	Glioma	T1w, T1ce, T2w, FLAIR MRI	140 training, 20 validation, and 50 testing cases with manual landmarks across baseline and follow-up scans.
CuRIOUS (Hering et al., 2022)	Brain	Glioma	T1w, T2-FLAIR MRI, 3D US	Part of Learn2Reg 2020, 22 subjects with pre-op MRI, and intra-op 3D US with annotated landmarks from EASY-RESECT (Xiao et al., 2017).
ReMIND2Reg (Juvekar et al., 2024)	Brain	Tumor resection	T1w, T2w MRI, 3D US	Part of Learn2Reg 2024 (Dorent et al., 2024), 104 intra-operative US, 98 T1ce, and 67 T2 MRIs from 104 patients, with manual landmarks.
Hippocampus-MR (Hering et al., 2022)	Brain	Non-affective psychosis	T1w MRI	Part of Learn2Reg 2020, 394 MR scans of the hippocampus region with manually tracings for evaluation.
DIR-Lab (Castillo et al., 2013, 2009a)	Lung	COPD, cancer	Breath-hold and 4DCT	20 CTs (COPDgene and 4DCT subsets) with 7,000+ manually paired landmarks for evaluating deformable registration.
NLST (Team, 2011)	Lung	Smokers	Spiral CT	100 paired inhale-exhale CTs with lung masks and keypoints; 10 test images with manual landmarks for Learn2Reg 2022 (Heinrich et al., 2022).
Lung-CT (Hering et al., 2022)	Lung	Healthy Controls	Inspiratory, expiratory CT	30 paired lung CTs with lung masks and keypoints; evaluation with manual landmarks from vessels and airways for Learn2Reg 2021 (Hering et al., 2022).
EMPIRE10 (Murphy et al., 2011)	Lung	Healthy Controls	Inspiratory, expiratory CT	30 lung CT pairs with 100 manual landmarks for each, covering different scan types to evaluate registration methods.
Thorax-CBCT (Hugo et al., 2016)	Lung	Cancer Patients	CT, CBCT	18 paired CTs from TCIA-4D-Lung with manual organ and target delineations for interventional registration in Learn2Reg 2023 (Heinrich et al., 2023).
Lung250M-4B (Falta et al., 2024)	Lung	Mixed	CT	248 paired CTs from seven datasets with 4 billion voxels and 250M keypoints, providing ground truth displacements and nnUNet segmentations.
ACDC (Bernard et al., 2018)	Heart	Cardiac diseases	4D cine-MRI	150 subjects with manual LV, RV, and Myo segmentations at ED and ES phases for intra-patient registration.
M&Ms (Campello et al., 2021)	Heart	Cardiac diseases	4D cine-MRI	375 subjects from multiple centers with LV, RV, and Myo segmentations at ED and ES phases for intra-patient registration.
MM-WHS (Zhuang et al., 2019)	Heart	Cardiac diseases	CT, MRI	120 cardiac scans (CT and MRI) from 60 subjects with 7 key heart structures manually annotated for mono- and multi-modal registration.
Abdomen-CT-CT (Hering et al., 2022)	Abdomen	Cancer Patients	CT	Part of Learn2Reg 2020 (Hering et al., 2022), featuring 50 CT images with 13 manually labeled structures from (Xu et al., 2016).
Abdomen-MR-CT (Hering et al., 2022)	Abdomen	Cancer Patients	CT, MR	Part of Learn2Reg 2021 (Hering et al., 2022), containing 16 CT/MR pairs with 4 labeled structures.
ACROBAT (Weitz et al., 2024)	Breast	Breast Cancer	Pathological images	4,212 whole-slide-images from 1,152 breast cancer patients.
ANHIR (Borovec et al., 2020)	Body-wide	Cancer tissue samples	Pathological images	355 images with 18 different stains, resulting in 481 valid image registration pairs.
COMULISglobe SHG-BF (Dorent et al., 2024)	Breast / Pancreas	Cancer tissue samples	Pathological images	Part of Learn2Reg 2024 (Dorent et al., 2024), featuring paired second-harmonic generation and bright field pathology images.
COMULISglobe 3D-CLEM (Dorent et al., 2024)	Cell	Mitochondria, nuclei	Microscopy	Part of Learn2Reg 2024 (Dorent et al., 2024), featuring 3 pre-processed microscopy datasets with manually annotated landmarks.

[a] https://brain-development.org/ixi-dataset/

References

Avants, B. B., Tustison, N., & Johnson, H. (2014). *Advanced Normalization Tools (ANTS)*.

Chen, J., Liu, Y., Wei, S., Bian, Z., Subramanian, S., Carass, A., Prince, J. L., & Du, Y. (2024). *A survey on deep learning in medical image registration: New technologies, uncertainty, evaluation metrics, and beyond* (No. arXiv:2307.15615). arXiv. http://arxiv.org/abs/2307.15615

Fu, Y., Lei, Y., Wang, T., Curran, W. J., Liu, T., & Yang, X. (2020). *Deep learning in medical image registration: A review*. Physics in Medicine & Biology, 65(20), 20TR01. https://doi.org/10.1088/1361-6560/ab843e

Gonzalez, R. C., Woods, R. E., and Eddins, S. L. (2006). *Digital Image Processing Using MATLAB(R)*. Prentice Hall.

Grenander, U. and Miller, M. (2007). *Pattern Theory: From Representation to Inference*. Oxford University Press.

haesleinhuepf (Director). (2020, July 5). *14a Image Registration* [Video recording]. https://www.youtube.com/watch?v=3CGC-5vwraM

Kellis, M. (Director). (2021, May 12). *Deep Learning Image Registration and Analysis—Lecture 21—MIT ML in Life Sciences (Spring 2021)* [Video recording]. https://www.youtube.com/watch?v=c4dvyTBvysQ

Lester, H. and Arridge, R. (1999). *A Survey of Hierarchical Non-linear Medical Image Registration*. Pattern Recognition, 32, 129–149.

Maintz, J. B. A., & Viergever, M. A. (1998). *A survey of medical image registration*. Medical Image Analysis, 2(1), 1–36.

Modersitzki, J. (2004). *Numerical Methods for Image Registration*. Oxford University Press, New York.

Modersitzki, J. (2009). *FAIR. Flexible Algorithms for Image Registration*. SIAM.

Niethammer, M. (2020). *Higher Order Models, Uncertainties, and Predictions*.

Pluim, J. P. W., Maintz, J. B. A., and Viergever, M. A. (1999). *Mutual-information based registration of medical images: a survey*. IEEE Transactions on Medical Imaging, 986–1004.

Song, J. H. (2017). *Methods for evaluating image registration* [Doctor of Philosophy, University of Iowa]. https://doi.org/10.17077/etd.v0vailob

Toga, A. and Thompson, P. (2001). *The role of image registration in brain mapping*. Image and Vision Computing, 19, 3–24.

Xiao, H., Teng, X., Liu, C., Li, T., Ren, G., Yang, R., Shen, D., & Cai, J. (2021). *A review of deep learning-based three-dimensional medical image registration methods*. Quantitative Imaging in Medicine and Surgery, 11(12), 4895–4916. https://doi.org/10.21037/qims-21-175

Zitova, B. and Flusser, J. (2003). *Image Registration Methods: A Survey*. Image and Vision Computing, 21, 977–1000.