Deep Learning for Medical Image Analysis

COMP5423

Hao CHEN

Dept. of CSE,CBE&LIFS, HKUST

jhc@cse.ust.hk

THE DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Medical Image Registration

- Introduction
- Deep Similarity Metric
- Supervised Image Registration
- Unsupervised Image Registration
- Challenges and Future Directions

Medical Image Registration

- Introduction
- Deep Similarity Metric
- Supervised Image Registration
- Unsupervised Image Registration
- Challenges and Future Directions

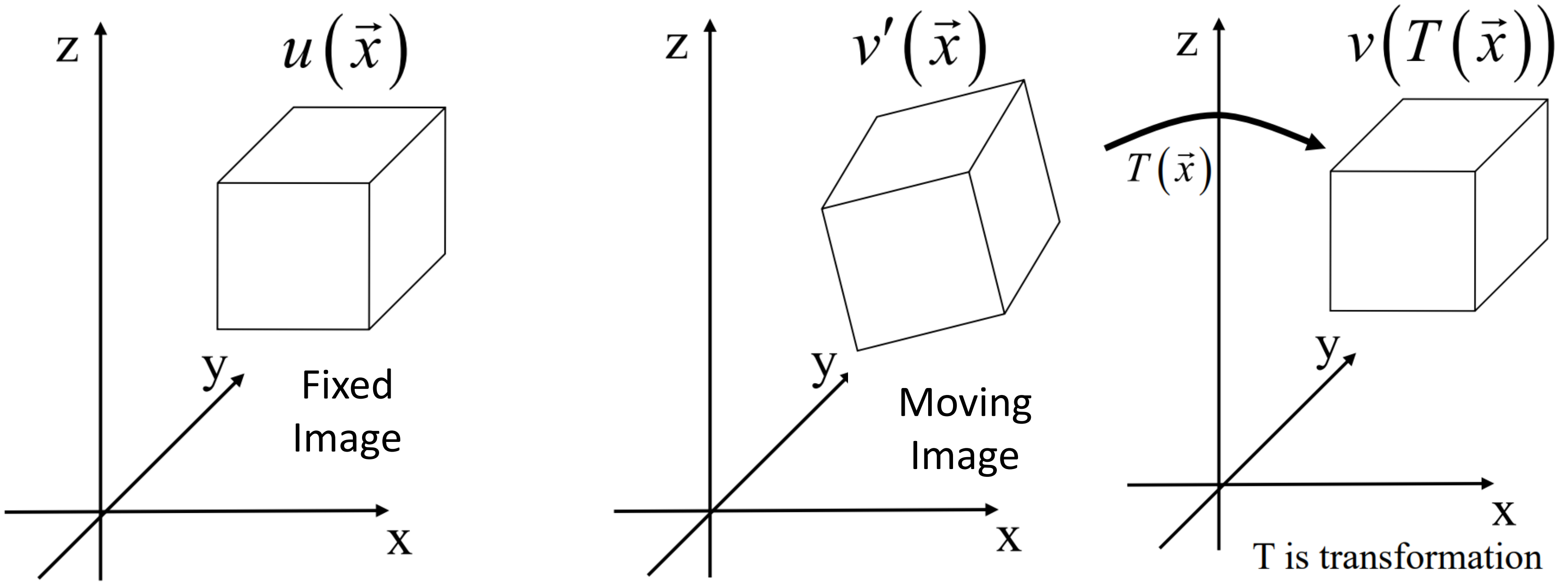
Introduction

- Image Registration

Image registration is **the process of identifying a spatial transformation** that maps two (pair-wise registration) or more (group-wise registration) images to a common coordinate space such that **corresponding anatomical structures are optimally aligned** (i.e., a pixel/voxel-wise 'correspondence' is established between the images).

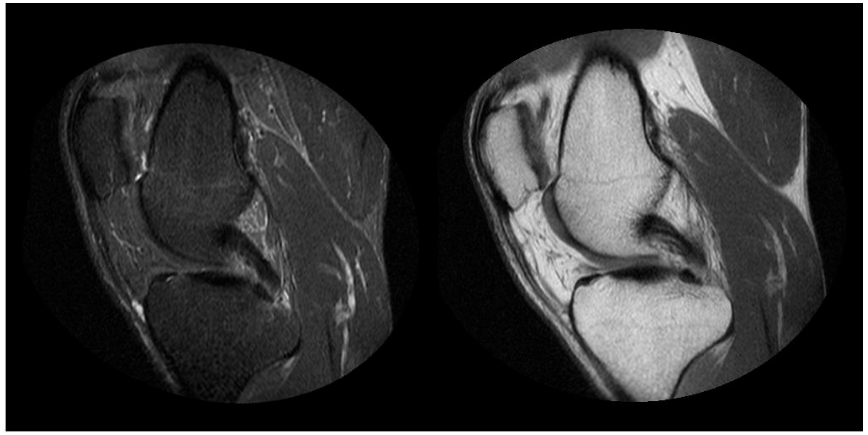
Introduction

- Image Registration

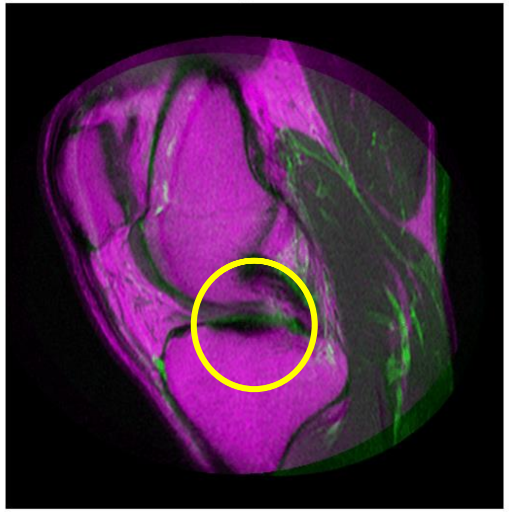


Introduction

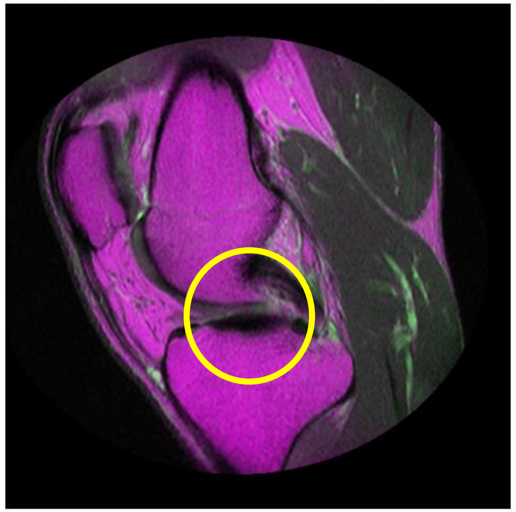
- Example: Multimodal medical image registration



Unregistered



E: Registration from Affine Model Based on Similarity Initial Condition



Introduction

- In pair-wise image registration, registering the two images can be regarded as an optimization problem

$$\hat{T} = argmin_T\, \mathcal{S}(F, T(M))$$

Here, *F*: fixed image, *M*: moving image.

$\hat{T}$: desired spatial transformation maps pixels/voxels of *M* to those of *F*.

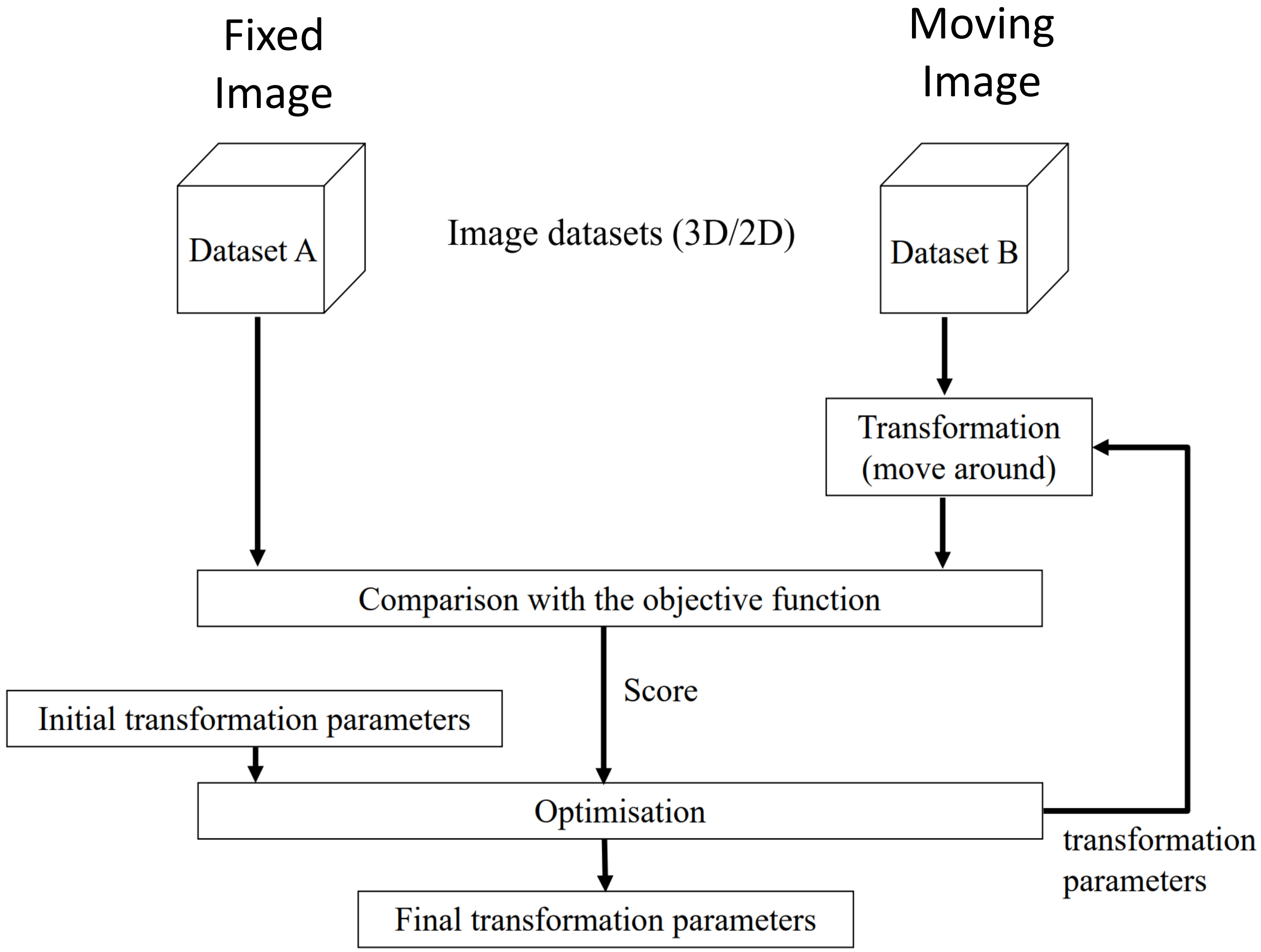
$\mathcal{S}$: a measure of (dis)similarity between the fixed and warped moving image.

The images are registered by iteratively improving estimates for the desired $\hat{T}$, such that the defined similarity $\mathcal{S}$ is maximized.

Chen, Xiang, et al. Deep learning in medical image registration. Progress in Biomedical Engineering, 2021.

Introduction

- Components:
  - A transformation model
  - Similarity metric
  - An optimization algorithm



Introduction

Algorithm:

1. Design of a suitable transformation model and initialization of its associated parameters.
2. Use of the transformation model to warp the moving image.
3. Evaluation of the similarity between the warped moving image and the fixed image.
4. Update of the parameters in the transformation model by optimizing the cost function formulated using the similarity metric, via a suitable optimization algorithm.
5. Iterates from step 2 to step 4 until a convergence criterion is satisfied.

Chen, Xiang, et al. Deep learning in medical image registration. Progress in Biomedical Engineering, 2021.

Introduction

Similarity Metrics

- Sum of Squared Difference (SSD)

$$SSD = \frac{1}{N} \sum_{\vec{x} \in D} \left[u(\vec{x}) - v\big(T(\vec{x})\big) \right]^2$$

where D represents the coordinate domain in space. SSD is normalized so that it is invariant to the number of voxels in D.

Introduction

Similarity Metrics

- Normalized Cross Correlation (NCC)

$$CC = \frac{\sum_{\vec{x}\in D}(u-\bar{u})(v-\bar{v})}{\sqrt{\sum_{\vec{x}\in D}(u-\bar{u})^2}\sqrt{\sum_{\vec{x}\in D}(v-\bar{v})^2}}$$

where *D* represents the coordinate domain in space, $\bar{u}$ and $\bar{v}$ represent means of u and v, respectively. CC assumes that there is a linear relationship between the intensity values in the images (-1 $\leq$ CC$\leq$1).

Introduction

Similarity Metrics

- Mutual Information (MI)

$$MI(u, v) = h(u) + h(v) - h(u, v)$$

where $h()$ represents the entropy (a measure of uncertainty) of a random variable and $h(x,y)$ represents the joint entropy of two random variables

$$h(x, y) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(x, y)$$

$$h(x) = -\sum_{x \in X} p(x) \log p(x)$$

Introduction

Similarity Metrics

- Mutual Information (MI)

$$MI(u, v) = h(u) + h(v) - h(u, v)$$

The third term is the joint entropy which encourages image matching.

The two images are matched when MI is maximized.

Introduction

Similarity Metrics

- Dice Similarity Coefficient (DSC)

$$DSC = \frac{2|M_A \cap M_B|}{|M_A| + |M_B|}$$

where M_A and M_B denote masks for specific anatomical structures or tissues in the warped moving image and fixed image, respectively.

Introduction

Image registration can be categorized as:

Rigid registration

- Translation, rotation (Euclidean)

Affine registration

- Translation, rotation, scaling and shear

Non-rigid/Deformable registration

- Establish the dense pixel/voxel-wise non-linear spatial correspondence between the fixed image and moving image.
- The warped moving image $M' = M(\varphi(v) + v)$ should be similar to fixed image $F(v)$. Here v is a voxel, φ denotes the deformation field.

Introduction

Toolboxes for medical image registration methods:

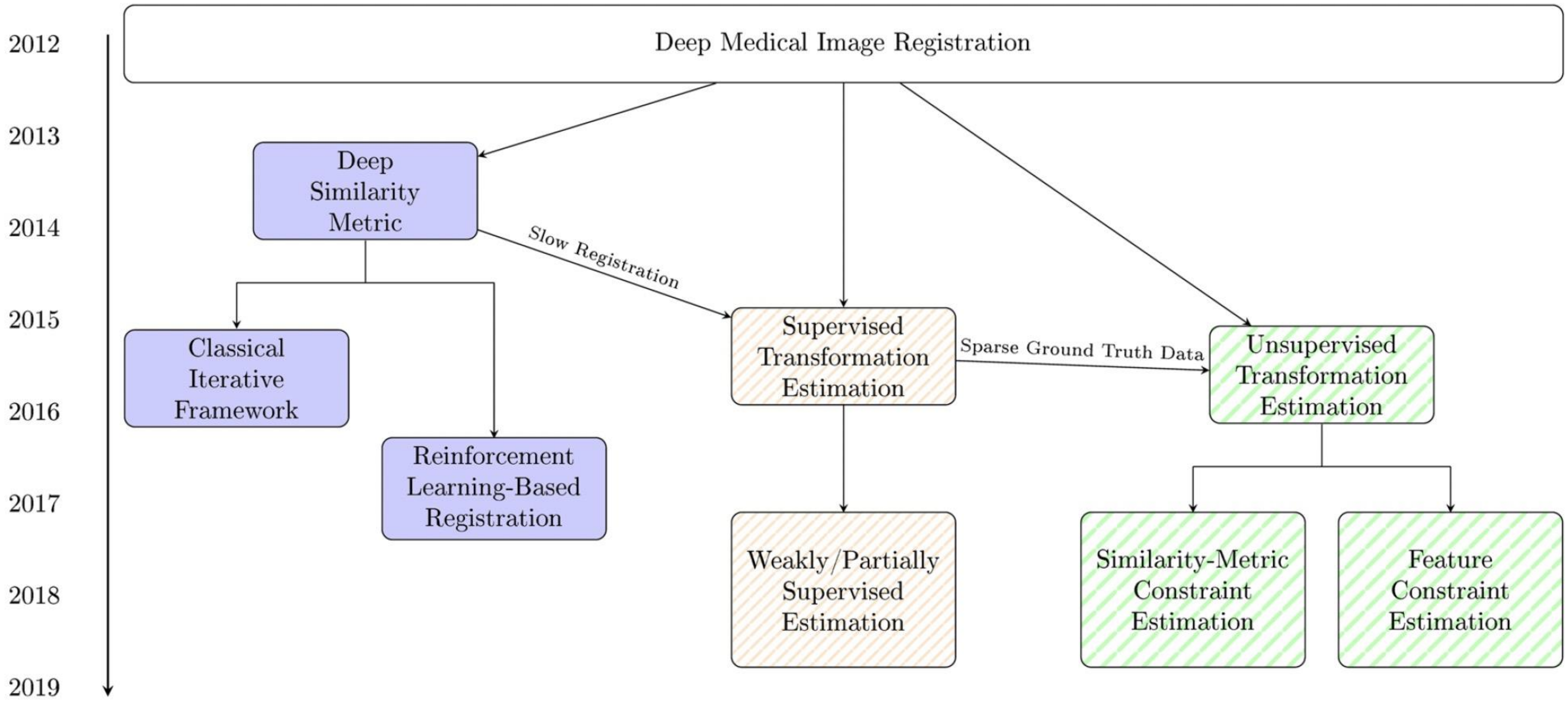
- **Elastix** [1]: A collection of algorithms that are commonly used to solve intensity-based medical image registration tasks.

- **Advanced Neuroimaging Tools** (ANTs) [2]: A suite of state-of-the-art image registration, segmentation and template building tools for quantitative morphometric analysis.

[1] Klein, Stefan, et al. Elastix: a toolbox for intensity-based medical image registration. TMI 2009.
[2] Avants, Brian B., et al. A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage 2011.

Introduction

- An overview of deep learning-based medical image registration:



Haskins, Grant, et al. Deep learning in medical image registration: a survey. Machine Vision and Applications, 2020.

Medical Image Registration

- Introduction
- Deep Similarity Metric
- Supervised Image Registration
- Unsupervised Image Registration
- Challenges and Future Directions

Deep Similarity Metric

- In traditional registration, the problem is formulated to minimize the objective:

$$E(\theta) = M\left(I_f, I_m\big(\mathcal{T}(\theta)\big)\right) + R(\mathcal{T}(\theta))$$

where *M* is a metric quantifying the cost (dissimilarity) of the alignment by transformation $\mathcal{T}$ parameterized by θ, *R* is a regularization term.

The minimization is solved iteratively to get optimal θ for transformation.

Simonovsky, Martin, et al. A deep metric for multimodal registration. MICCAI 2016.

Deep Similarity Metric

- In traditional registration, the problem is formulated to minimize the objective:

$$E(\theta) = \boxed{M\left(I_f, I_m(\mathcal{T}(\theta))\right)} + R(\mathcal{T}(\theta))$$

The (dis)similarity metric is crucial for optimizing registration accuracy.

Traditional similarity metrics (e.g., CC, and MI) can work well for the same modality with not complex variations.

Simonovsky, Martin, et al. A deep metric for multimodal registration. MICCAI 2016.

Deep Similarity Metric

- In traditional registration, the problem is formulated to minimize the objective:

$$E(\theta) = \boxed{M\left(I_f, I_m(\mathcal{T}(\theta))\right)} + R(\mathcal{T}(\theta))$$

The (dis)similarity metric is crucial for optimizing registration accuracy.

Multimodal registration is challenging due to the high variability of tissue appearance under different imaging modalities.

Simonovsky, Martin, et al. A deep metric for multimodal registration. MICCAI 2016.

Deep Similarity Metric

- The (dis)similarity metrics *M* can be formulated as a well-trained CNN, to capture complex relationships between multimodal medical images:

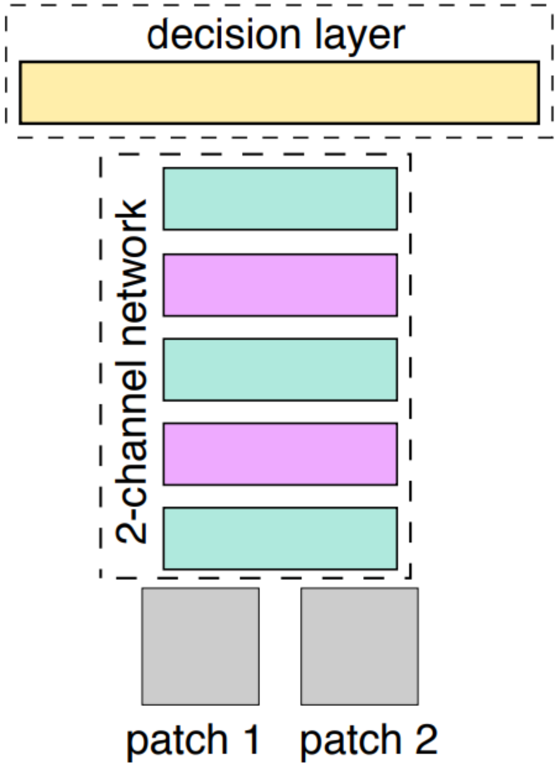
$$E(\theta) = \boxed{M\left(I_f, I_m(\mathcal{T}(\theta))\right)} + R(\mathcal{T}(\theta))$$

$$M(I_f, I'_m) = \sum_{P \in \mathcal{P}} N(I_f(P), I'_m(P))$$

where *N* is a CNN, $I'_m = I_m\big(\mathcal{T}(\theta)\big)$, $\mathcal{P}$ is the set of patch domains sampled on a dense uniform grid with significant overlaps.

Deep Similarity Metric

- *N* is trained on a set of *k* **aligned** pairs of training images $\{(A_j, B_j)\}_{j=1}^{k}$



Online augmentation to get practically unlimited amount of training data:

$$X_i = (A_j(\mathcal{T}_{i,A_j}(P)), B_j(\mathcal{T}_{i,B_j}(P)))$$

X_i is defined positive if $\mathcal{T}_{i,A_j} = \mathcal{T}_{i,B_j}$, negative otherwise.

Simonovsky, Martin, et al. A deep metric for multimodal registration. MICCAI 2016.

Deep Similarity Metric

- Comparison to mutual information methods. (The database consists of T1 and T2-weighted MRI scans of 20 neonatal brains.)

Table 1. Overlap scores (mean ± SD) after registration using the proposed metric (CNN) and mutual information with (MI+M) or without masking (MI)

	MI+M	MI	CNN $k = 557$	CNN $k = 11$	CNN $k = 6$	CNN $k = 3$
Dice	0.665 ± 0.096	0.497 ± 0.180	0.703 ± 0.037	0.704 ± 0.037	0.701 ± 0.040	0.675 ± 0.093
Jaccard	0.519 ± 0.091	0.369 ± 0.151	0.555 ± 0.041	0.556 ± 0.041	0.554 ± 0.044	0.527 ± 0.081

With online augmentation, even for small *k* (training data size) no obvious overfitting.

Simonovsky, Martin, et al. A deep metric for multimodal registration. MICCAI 2016.

Medical Image Registration

- Introduction
- Deep Similarity Metric
- Supervised Image Registration
- Unsupervised Image Registration
- Challenges and Future Directions

Supervised Image Registration

Traditional registration methods:

- Encode geometric transformations by optimizing a cost function for tens or hundreds of iterations.
- Computationally expensive, slow.

This motivates the development of networks that could estimate the transformation corresponding to optimal similarity in one step.

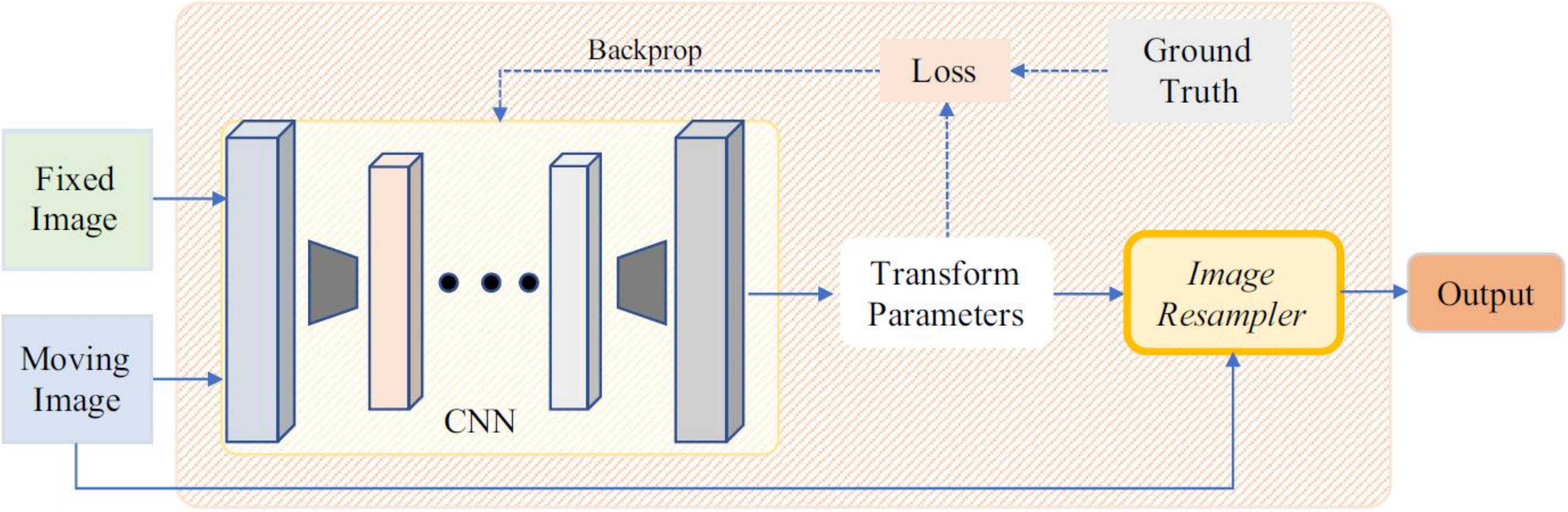
Supervised Image Registration

Supervised registration

- Use CNN to estimate the parameters associated with the transformation model adopted.
- Approaches for obtaining ground truth (i.e., target parameters):
    1. use traditional registration methods;
    2. use simulated images with known ground truth transformations;
    3. use markers (i.e., internal or external).
- After training, registration is achieved as *a single forward pass* through the network.

Supervised Image Registration

- A visualization of supervised single-step registration



Haskins, Grant, et al. Deep learning in medical image registration: a survey. Machine Vision and Applications, 2020.

Supervised Image Registration

- Formulation

The CNN is trained to learn the mapping $\mathcal{M}$ from image pair (e.g., a template $\mathcal{T}$ and a subject $\mathcal{S}$) to their final deformation field ϕ.

$$\mathcal{M} : \ (\mathcal{T}, \mathcal{S}) \Rightarrow \phi$$

Cao, Xiaohuan, et al. Deformable image registration based on similarity-steered CNN regression. MICCAI 2017.

Supervised Image Registration

- Training Set Preparation

A training sample $S_i = \{(p_{\mathcal{T}}(u), p_{\mathcal{S}}(u))|\phi(u)\}$, where $(p_{\mathcal{T}}(u), p_{\mathcal{S}}(u))$ is a patch pair extracted from the center location u, $\phi(u) = [d_x, d_y, d_z]$ is the displacement vector of u.

Sampling strategy is required due to the unbalanced distribution of displacement magnitude.

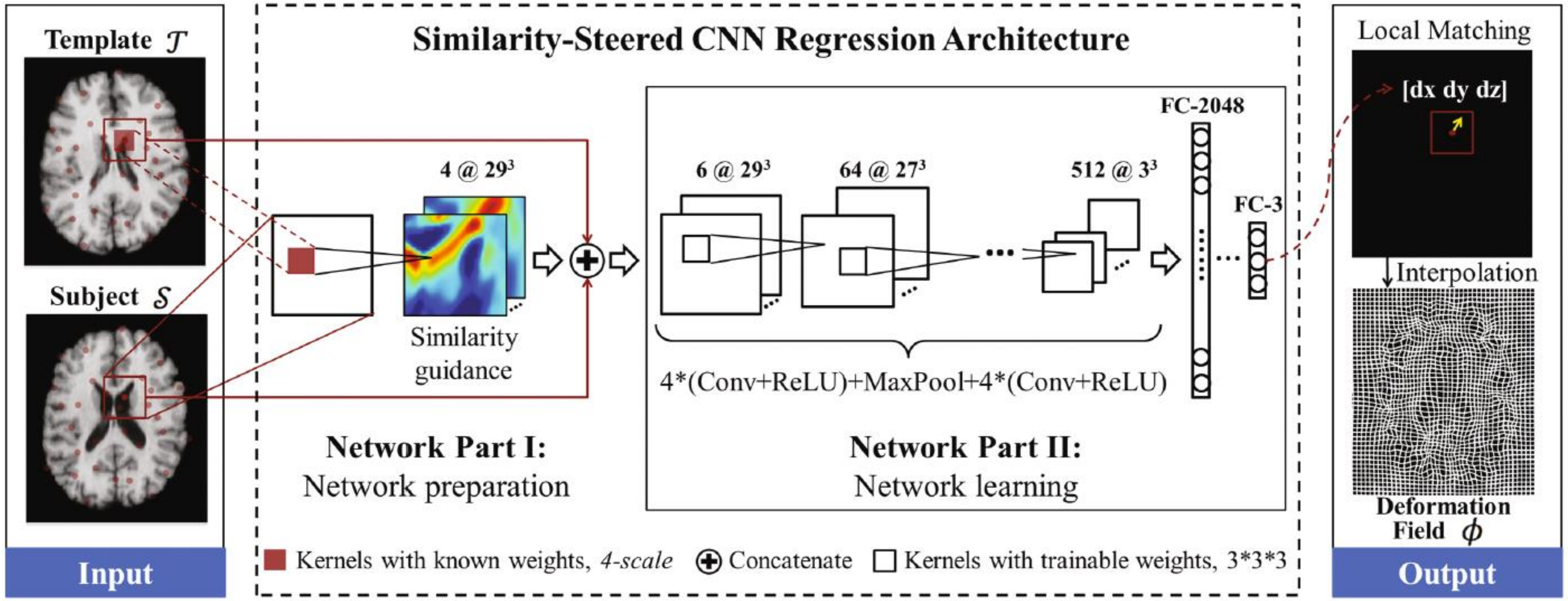
Cao, Xiaohuan, et al. Deformable image registration based on similarity-steered CNN regression. MICCAI 2017.

Supervised Image Registration

- Network

Contextual cue by **similarity map** (local cross-correlation from the center location in template patch to the whole subject patch locations).

Patches and the cue are concatenated as the input to CNN.



Cao, Xiaohuan, et al. Deformable image registration based on similarity-steered CNN regression. MICCAI 2017.

Supervised Image Registration

- Results

Table 1. Mean DSC and ASSD of GM, WM and CSF based on 10 testing subject from LONI dataset, after deformable registration by Demons, SyN, and our proposed method. "*" indicates statistically significant improvement by our method, compared to other two methods ($p < 0.05$).

	DSC (%)			ASSD (mm)		
	GM	WM	CSF	GM	WM	CSF
Demons	72.8 ± 1.0	80.8 ± 7.3	**62.9 ± 5.1**	0.55 ± 0.34	**0.49 ± 0.15**	**0.43 ± 0.10**
SyN	72.7 ± 1.6	78.1 ± 0.7	61.5 ± 2.6	0.46 ± 0.03	0.58 ± 0.05	0.54 ± 0.05
Proposed	**75.3 ± 1.4***	**81.3 ± 0.6***	61.2 ± 2.4	**0.43 ± 0.04***	0.56 ± 0.05	0.48 ± 0.03

Averaged symmetric surface distance (ASSD)

Cao, Xiaohuan, et al. Deformable image registration based on similarity-steered CNN regression. MICCAI 2017.
Avants, et al.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. MIA 2008.

Supervised Image Registration

- Results

Table 2. Mean DSC and ASSD of GM, WM and CSF for ADNI dataset, after deformable registration by Demons, SyN, and our proposed method. "*" indicates statistically significant improvement, compared to other two methods ($p < 0.05$).

	DSC (%)			ASSD (mm)		
	GM	WM	CSF	GM	WM	CSF
Demons	64.8 ± 2.3	75.7 ± 1.0	54.3 ± 2.5	0.70 ± 0.04	**0.73 ± 0.05**	0.38 ± 0.04
SyN	64.6 ± 2.5	76.3 ± 1.8	55.2 ± 3.2	0.69 ± 0.04	0.81 ± 0.06	0.37 ± 0.05
Proposed	**65.1 ± 2.2***	**78.3 ± 0.6***	**56.0 ± 2.1***	**0.64 ± 0.04***	0.75 ± 0.04	**0.32 ± 0.04***

Averaged symmetric surface distance (ASSD)

Cao, Xiaohuan, et al. Deformable image registration based on similarity-steered CNN regression. MICCAI 2017.

Medical Image Registration

- Introduction
- Deep Similarity Metric
- Supervised Image Registration
- Unsupervised Image Registration
- Challenges and Future Directions

Unsupervised Image Registration

Limitation of supervised methods:

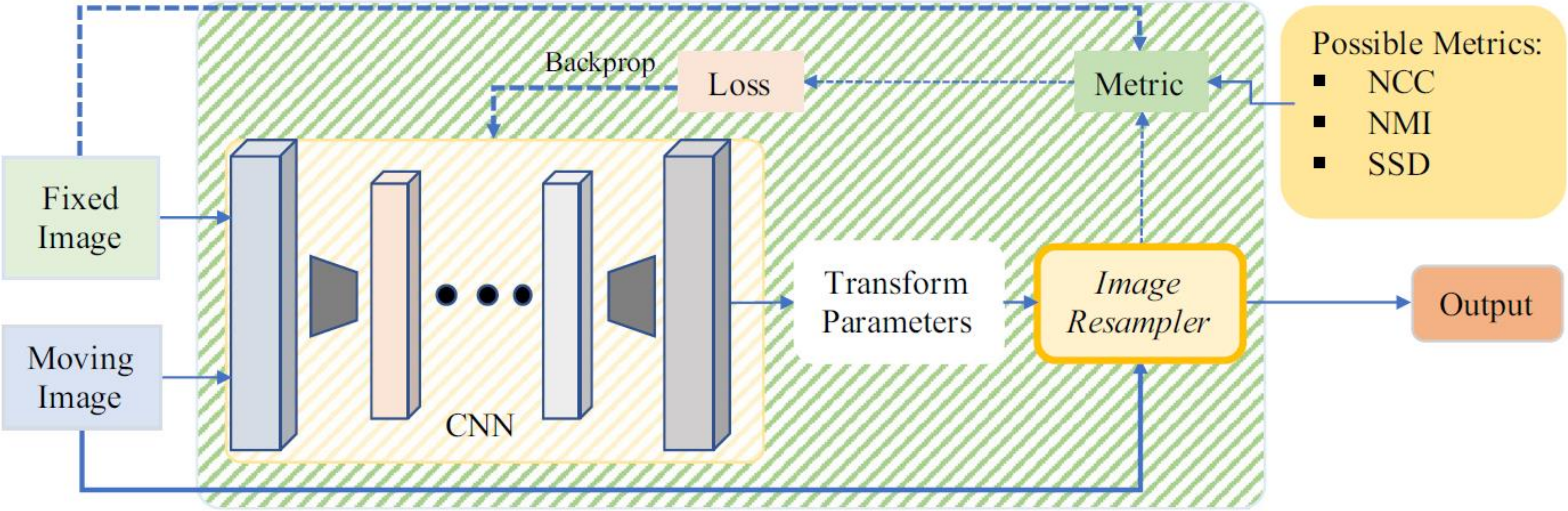
- Difficult to obtain plausible ground-truth transformations.

Therefore, researchers have explored *unsupervised methods* (hot topic):

- Require only fixed and moving images, with no ground truth for training.
- Predict **deformation fields** and warped moving images in just one forward pass.

Unsupervised Image Registration

A visualization of unsupervised deep single-step registration where the network is trained using a metric that quantifies image similarity.

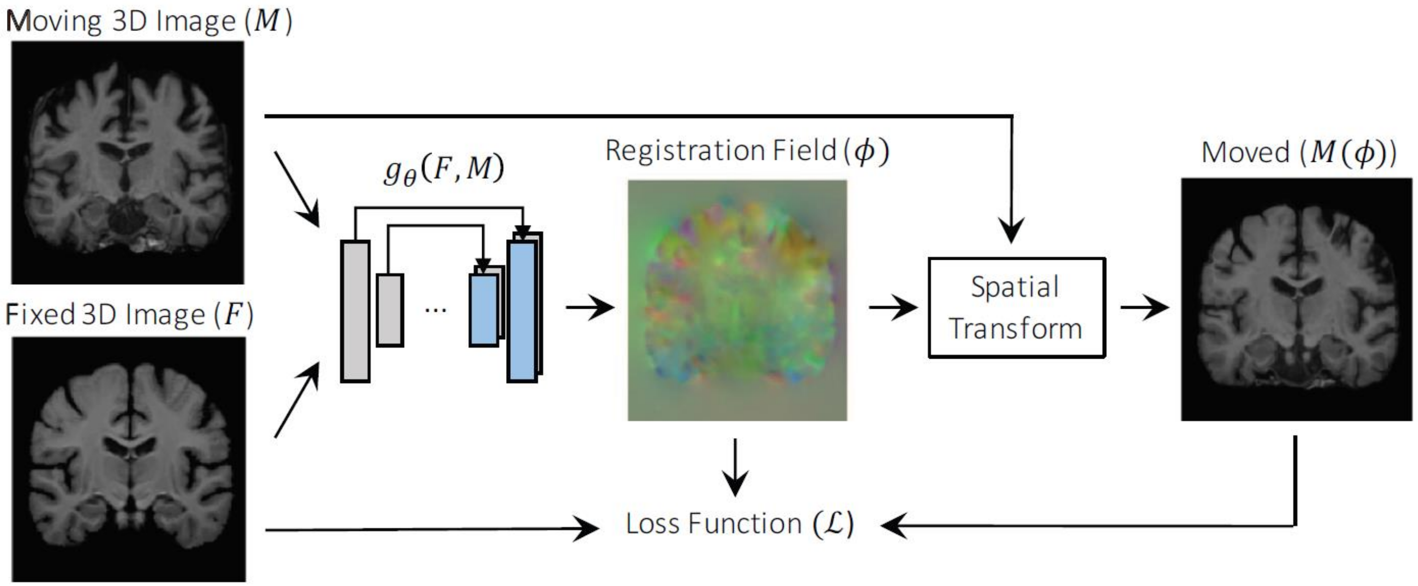


Haskins, Grant, et al. Deep learning in medical image registration: a survey. Machine Vision and Applications, 2020.

Unsupervised Image Registration

VoxelMorph

- The similarity (cross correlation, CC) between the warped moving image and fixed image is used to calculate the loss function for back-propagation.



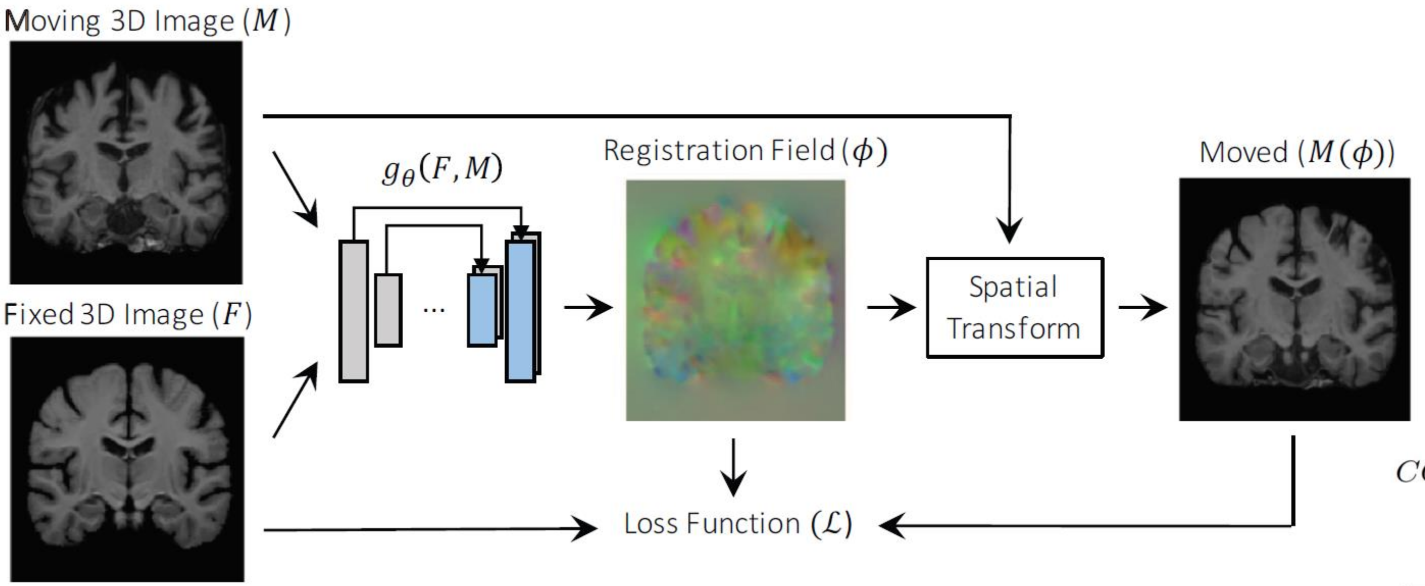
$$\mathcal{L}(F, M, \phi) = \mathcal{L}_{sim}(F, M(\phi)) + \lambda \mathcal{L}_{smooth}(\phi)$$

Balakrishnan, Guha, et al. An unsupervised learning model for deformable medical image registration. CVPR 2018.

Unsupervised Image Registration

VoxelMorph

- The similarity (cross correlation, CC) between the warped moving image and fixed image is used to calculate the loss function for back-propagation.



$$CC(F, M(\phi)) = \sum_{p\in\Omega} \frac{\left(\sum_{p_i}(F(p_i) - \hat{F}(p))(M(\phi(p_i)) - \hat{M}(\phi(p)))\right)^2}{\left(\sum_{p_i}(F(p_i) - \hat{F}(p))\right)\left(\sum_{p_i}(M(\phi(p_i)) - \hat{M}(\phi(p)))\right)}$$

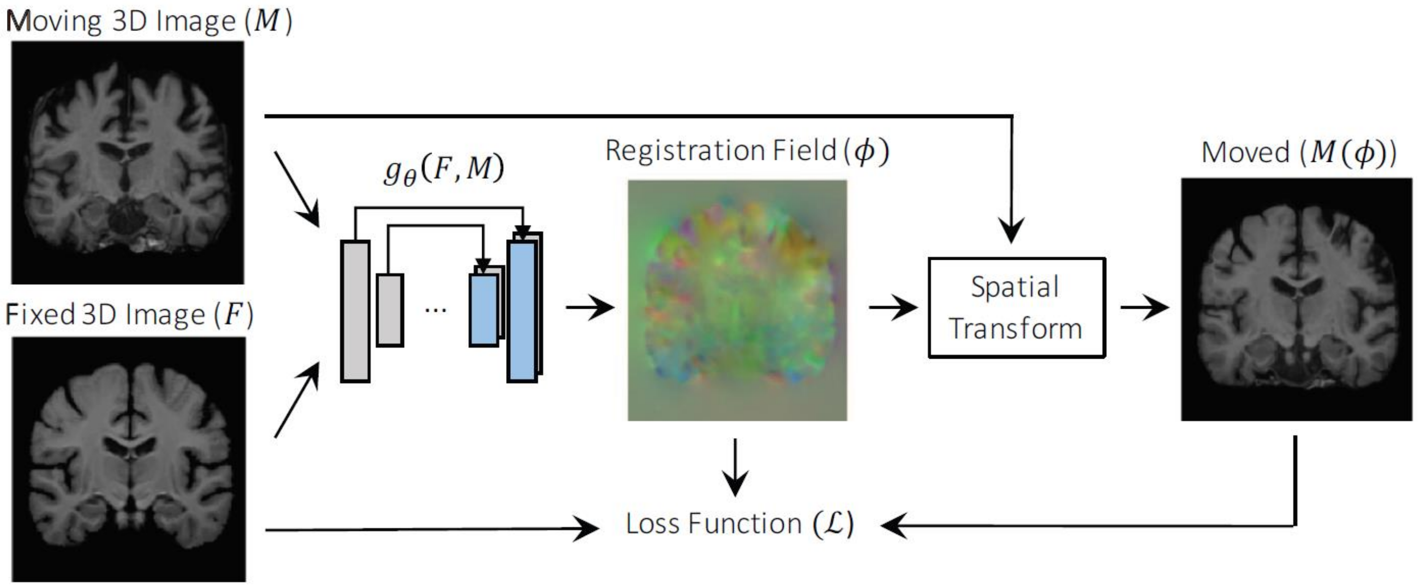
$$\mathcal{L}(F, M, \phi) = \boxed{\mathcal{L}_{sim}(F, M(\phi))} + \lambda \mathcal{L}_{smooth}(\phi)$$

Balakrishnan, Guha, et al. An unsupervised learning model for deformable medical image registration. CVPR 2018.

Unsupervised Image Registration

VoxelMorph

- The similarity (cross correlation, CC) between the warped moving image and fixed image is used to calculate the loss function for back-propagation.

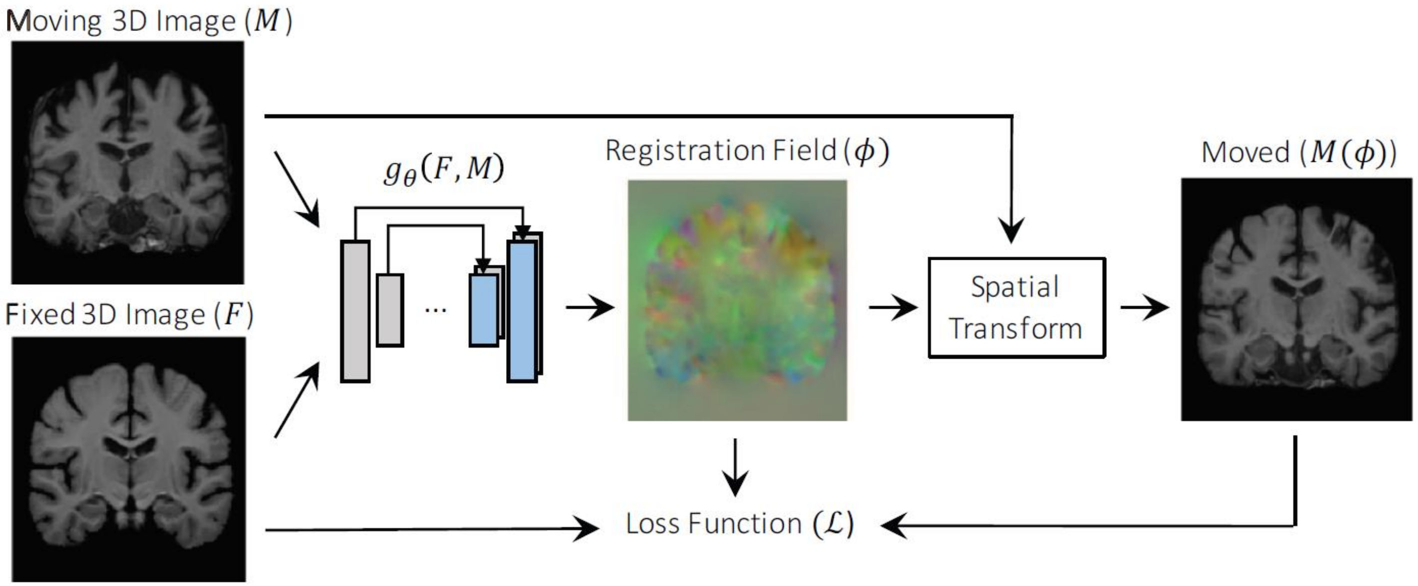


$$\mathcal{L}(F, M, \phi) = \mathcal{L}_{sim}(F, M(\phi)) + \boxed{\lambda \mathcal{L}_{smooth}(\phi)} \qquad \mathcal{L}_{smooth}(\phi) = \sum_{p \in \Omega} \|\nabla \phi(p)\|^2$$

Unsupervised Image Registration

VoxelMorph

- The similarity (cross correlation, CC) between the warped moving image and fixed image is used to calculate the loss function for back-propagation.

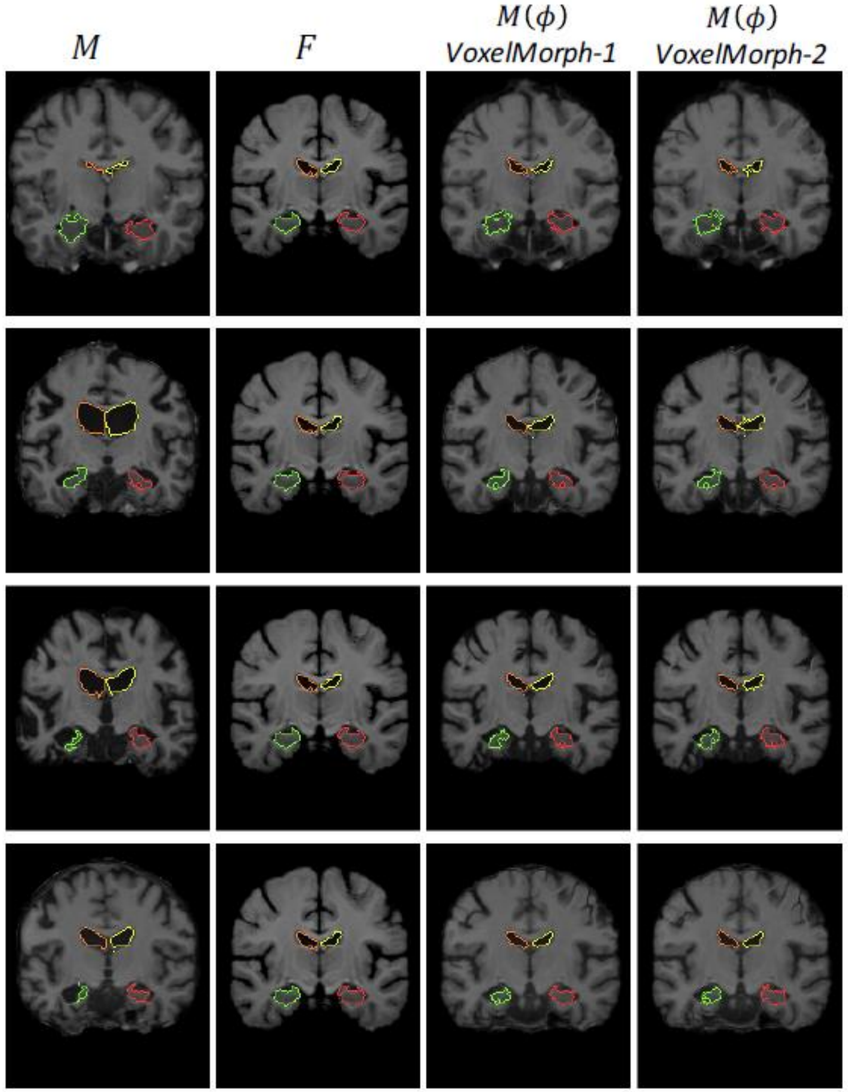


$$\mathcal{L}(F, M, \phi) = -CC(F, M(\phi)) + \lambda \sum_{p \in \Omega} \|\nabla \phi(p)\|^2$$

Balakrishnan, Guha, et al. An unsupervised learning model for deformable medical image registration. CVPR 2018.

Unsupervised Image Registration

Method	Avg. Dice	GPU sec	CPU sec
Affine only	0.567 (0.157)	0	0
ANTs	0.749 (0.135)	-	9059 (2023)
VoxelMorph-1	0.742 (0.139)	0.365 (0.012)	57(1)
VoxelMorph-2	0.750 (0.137)	0.554 (0.017)	144 (1)

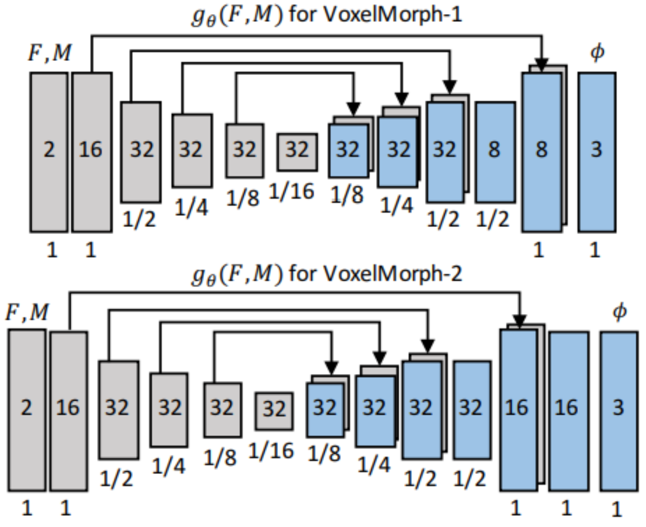


- Yield comparable results to ANTs in Dice score.
- Operate orders of magnitude faster during testing.

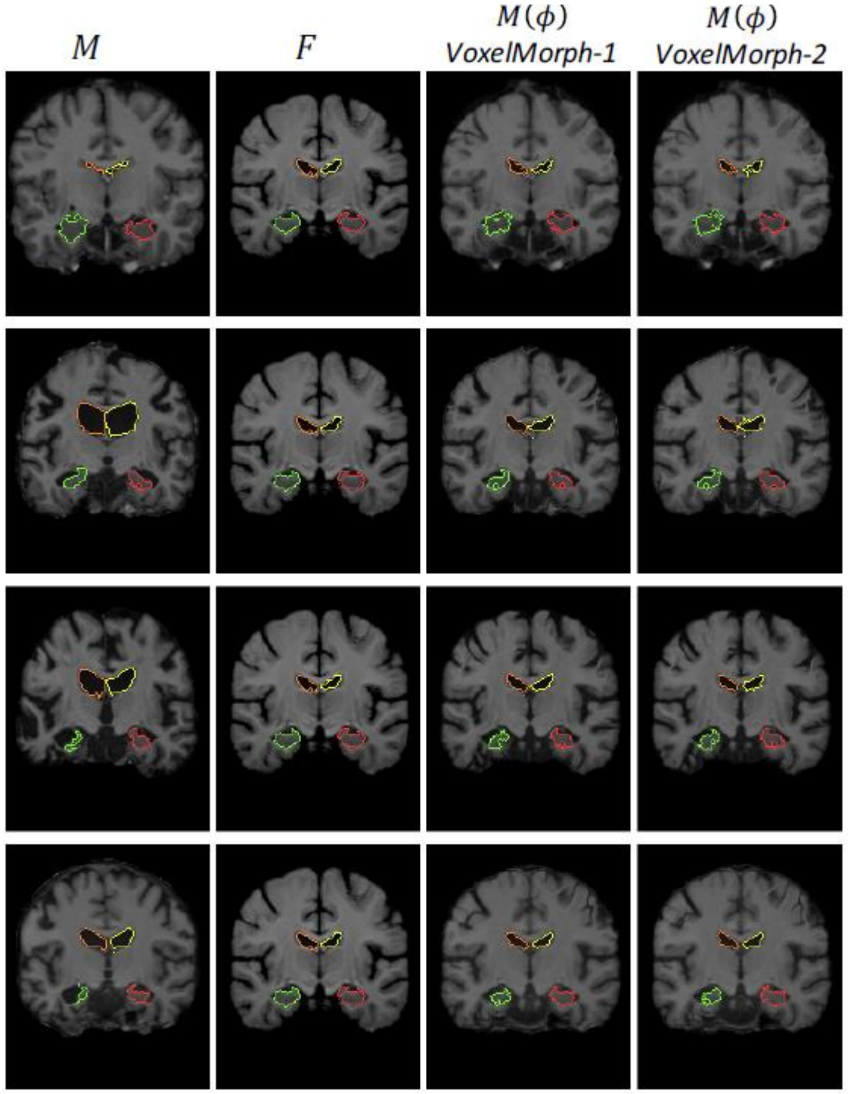
Balakrishnan, Guha, et al. An unsupervised learning model for deformable medical image registration. CVPR 2018.

Unsupervised Image Registration

Method	Avg. Dice	GPU sec	CPU sec
Affine only	0.567 (0.157)	0	0
ANTs	0.749 (0.135)	-	9059 (2023)
VoxelMorph-1	0.742 (0.139)	0.365 (0.012)	57(1)
VoxelMorph-2	0.750 (0.137)	0.554 (0.017)	144 (1)



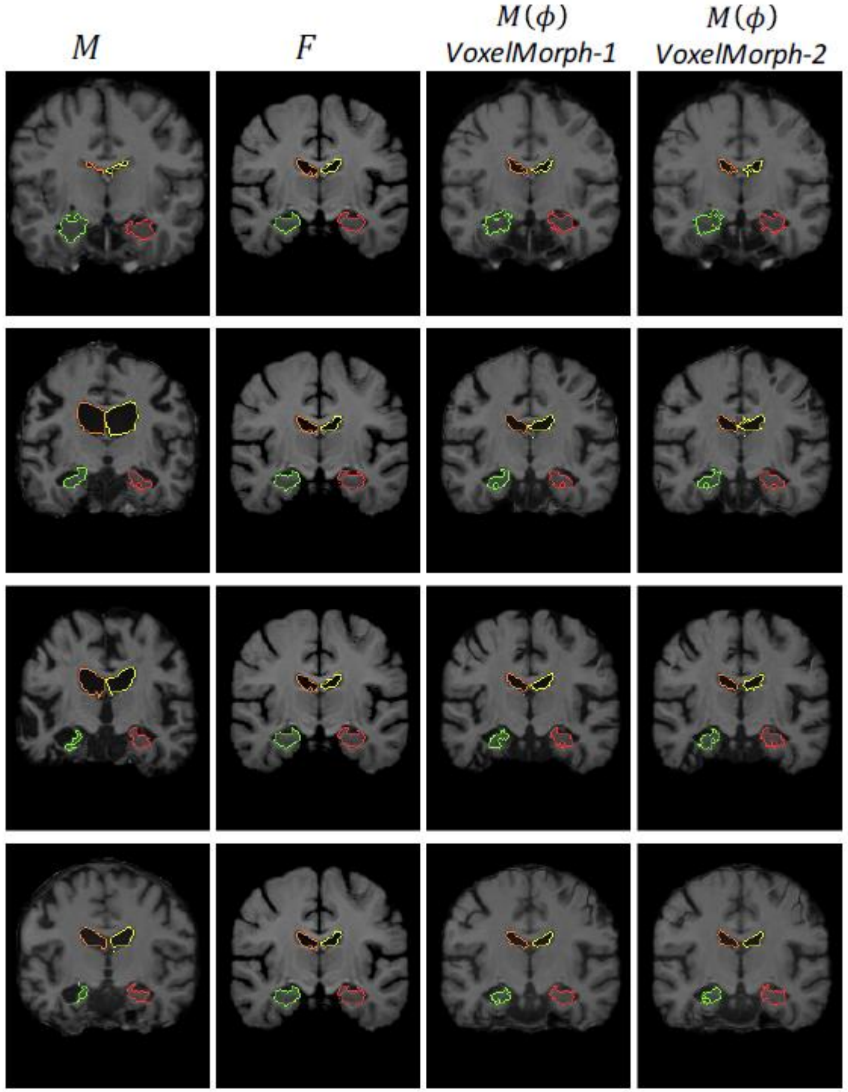
- VoxelMorph-2 uses a larger architecture, using one extra convolutional layer at the output resolution, and more channels for later layers.
- ANTs: SyN implementation in the publicly available ANTs software package.



Balakrishnan, Guha, et al. An unsupervised learning model for deformable medical image registration. CVPR 2018.

Unsupervised Image Registration

Method	Avg. Dice	GPU sec	CPU sec
Affine only	0.567 (0.157)	0	0
ANTs	0.749 (0.135)	-	9059 (2023)
VoxelMorph-1	0.742 (0.139)	0.365 (0.012)	57(1)
VoxelMorph-2	0.750 (0.137)	0.554 (0.017)	144 (1)

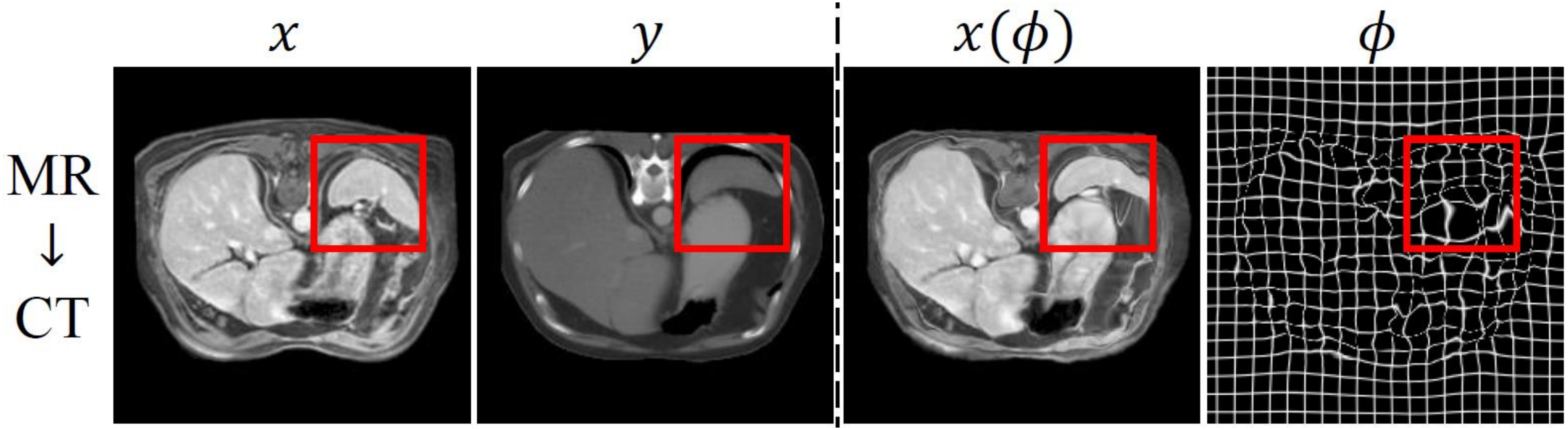


- Yield comparable results to ANTs in Dice score.
- Operate orders of magnitude faster during testing.

Balakrishnan, Guha, et al. An unsupervised learning model for deformable medical image registration. CVPR 2018.

Unsupervised Multi-modal Image Registration

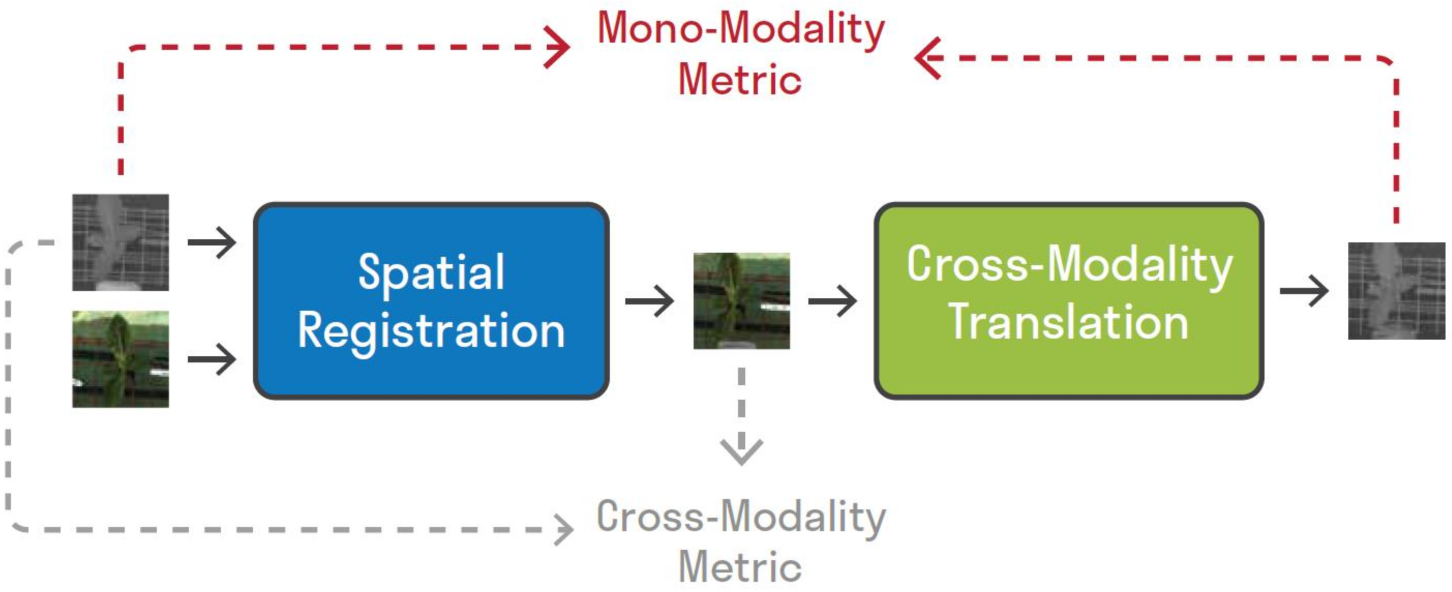
- Modality: a way of representing information in some medium
- Multi-modality: More than two modalities with different medium
- Most multi-modal registration methods struggle computing the spatial correspondence between the images using prevalent cross-modality similarity measures (e.g., NCC, NMI).



Chen, et al. Unsupervised multi-modal medical image registration via discriminator-free image-to-image translation. IJCAI 2022.

Unsupervised Multi-modal Image Registration

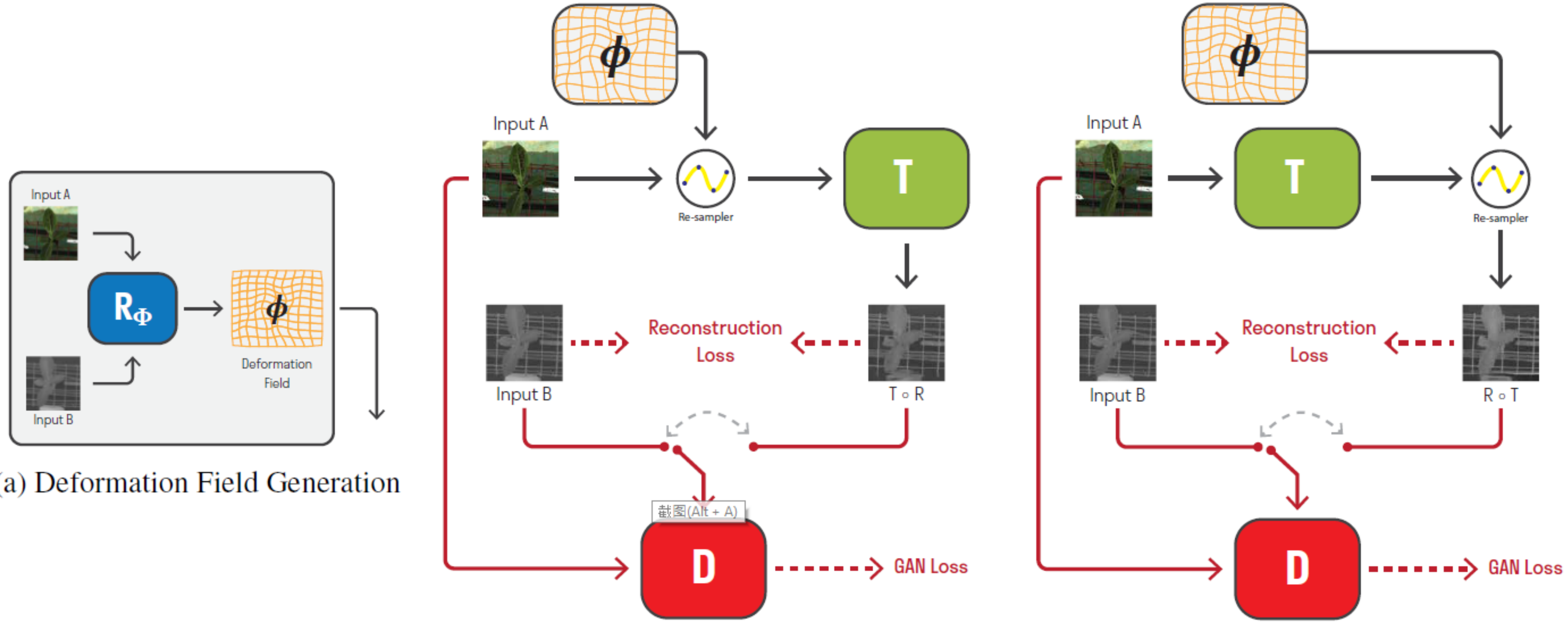
- Multi-modal registration using two networks: a **spatial transformation network** and a **translation network**.
- Bypass the difficulties of developing cross-modality similarity measures by training an **image-to-image translation** network on the two input modalities.



Arar, et al. Unsupervised multi-modal image registration via geometry preserving image-to-image translation. CVPR 2020.

Unsupervised Multi-modal Image Registration

- Geometric Preserving Translation Network: encourages T to be geometric preserving that T and R are commutative.



(a) Deformation Field Generation

(b) Register First

(c) Translate First

Arar, et al. Unsupervised multi-modal image registration via geometry preserving image-to-image translation. CVPR 2020.

Unsupervised Multi-modal Image Registration

- Geometric Preserving Translation Network: encourages T to be geometric preserving that T and R are commutative.
- Translation first:

$$O_{RT} = R_S\left(O_T, \phi\right) = R\left(T\left(I_a\right), R_\Phi\left(I_a, I_b\right)\right)$$

- Register first:

$$O_{TR} = T\left(R_S\left(I_a, \phi\right)\right) = T\left(R_S\left(I_a, R_\Phi\left(I_a, I_b\right)\right)\right)$$

- Loss function:

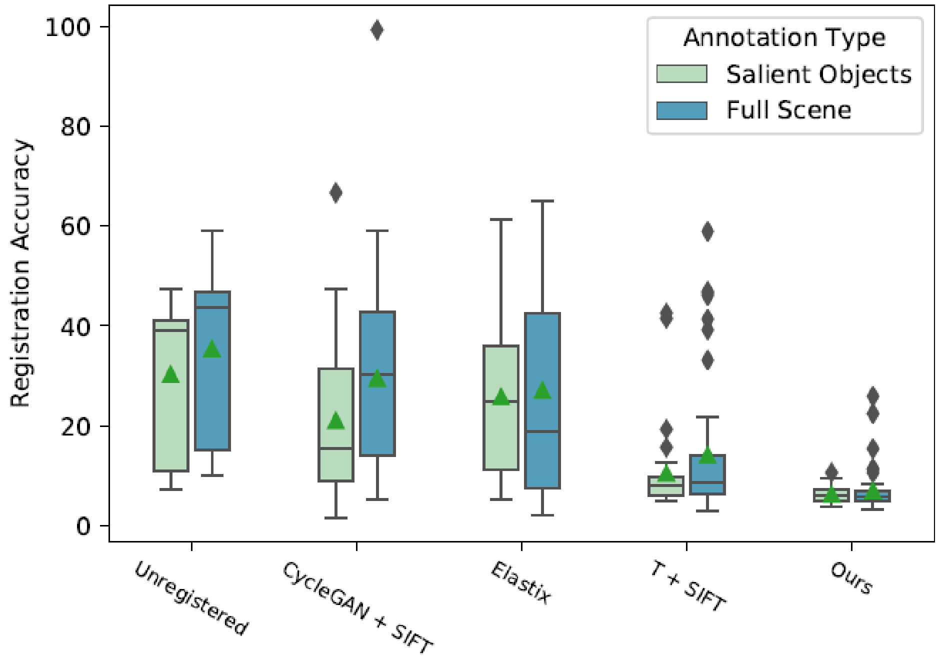
$$\mathcal{L}_{recon}(T, R) = \|O_{RT} - I_b\|_1 + \|O_{TR} - I_b\|_1$$

Arar, et al. Unsupervised multi-modal image registration via geometry preserving image-to-image translation. CVPR 2020.

Unsupervised Multi-modal Image Registration

- Evaluation

STN	SSIM on edges	NCC on edges	NCC	Ours
R_{Affine}	X / X	19.44 / 19.45	20.56 / 13.26	13.53 / 8.5
R_{TPS}	X / X	32.47 / 26.82	28.68 / 26.47	10.01 / 7.02
R	28.41 / 26.12	27.41 / 16.78	29.91 / 15.8	**6.93 / 6.27**

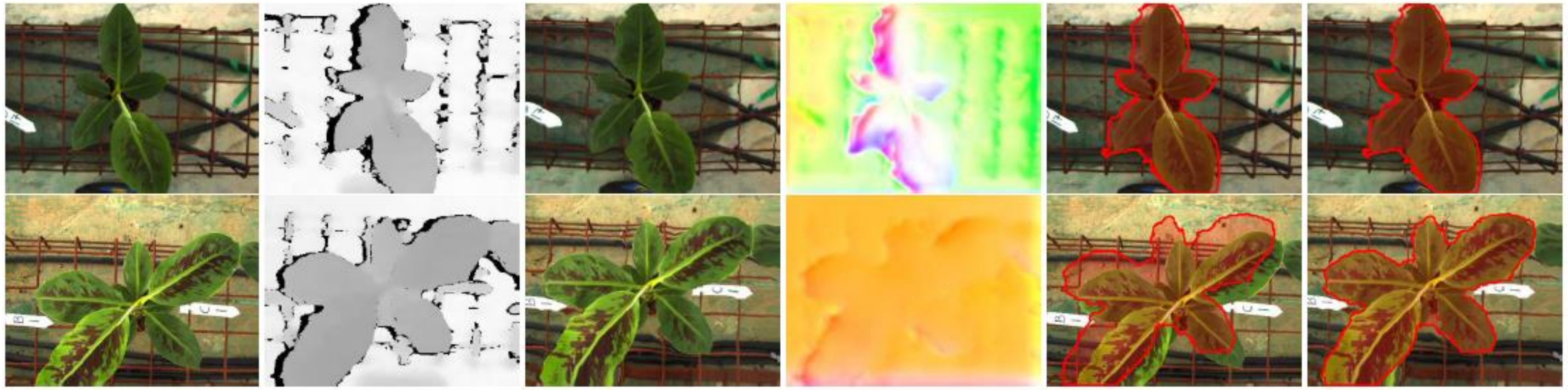


The accuracy of the registration network R is simply the **average Euclidean distance** between the target points and their matching deformed source points.
Full scene annotation (left)/the other based on salient objects only annotation (right).
X denotes cases where the network degenerates.

Arar, et al. Unsupervised multi-modal image registration via geometry preserving image-to-image translation. CVPR 2020.

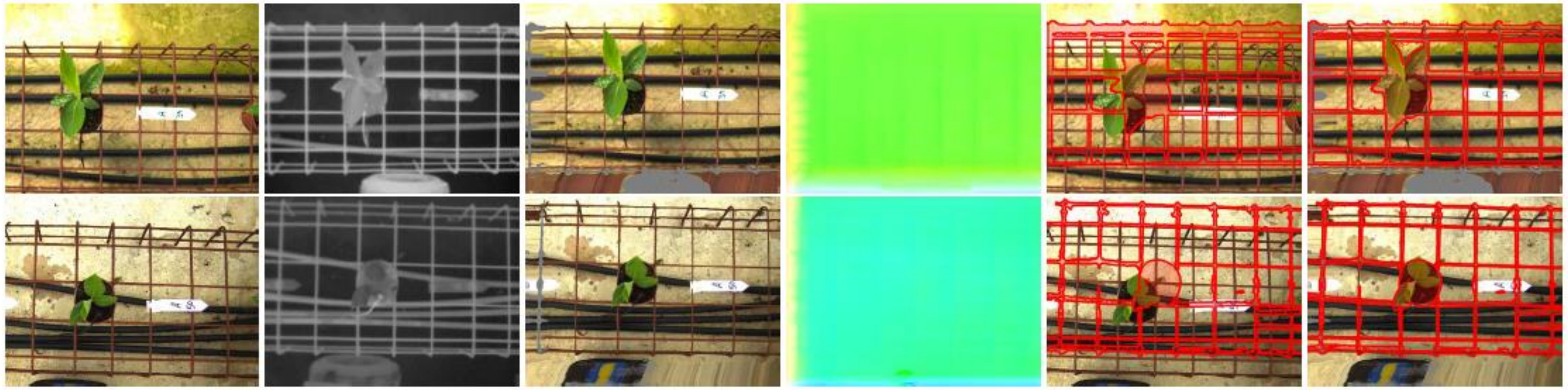
Unsupervised Multi-modal Image Registration

- Evaluation



Input A | Input B | Registered | Deform. Field | Before | After

(a) Image registration between RGB and Depth modalities.



Input A | Input B | Registered | Deform. Field | Before | After

(b) Image registration between RGB and IR modalities.

Arar, et al. Unsupervised multi-modal image registration via geometry preserving image-to-image translation. CVPR 2020.

Medical Image Registration

- Introduction
- Deep Similarity Metric
- Supervised Image Registration
- Unsupervised Image Registration
- Challenges and Future Directions

Challenges and Future Directions

- Preprocessing

Different preprocessing steps (e.g., skull-stripping, affine registration, spatial resampling, image enhancement, intensity normalization and cropping) with different adapted parameters for each step (e.g., voxel size, smoothing factor, etc.) may lead to different registration results, even using the same datasets.

- Clinical Deployment

Due to the lack of interpretation of DL-based methods, further research and systematic means for assessing registration uncertainty are necessary.

Chen, Xiang, et al. Deep learning in medical image registration. Progress in Biomedical Engineering, 2021.

Challenges and Future Directions

Goals for all registration methods: *Accuracy*, *robustness* and *speed*.

- Improve the **accuracy** and **generalization capability** of the networks.
- The superiority of classical methods (e.g., diffeomorphic attributes and robust registration) can not be overlooked.
- Boosting performance with priors, such as expected type of deformation, topology and morphology of anatomical structures.
- Feature-based unsupervised registration gains significant interests from the research community. Extension to the multimodal case is likely to be a prominent research focus in the next few years.