Deep learning for medical image registration

Ruisheng Su

Maureen van Eijnatten

Outline for today:

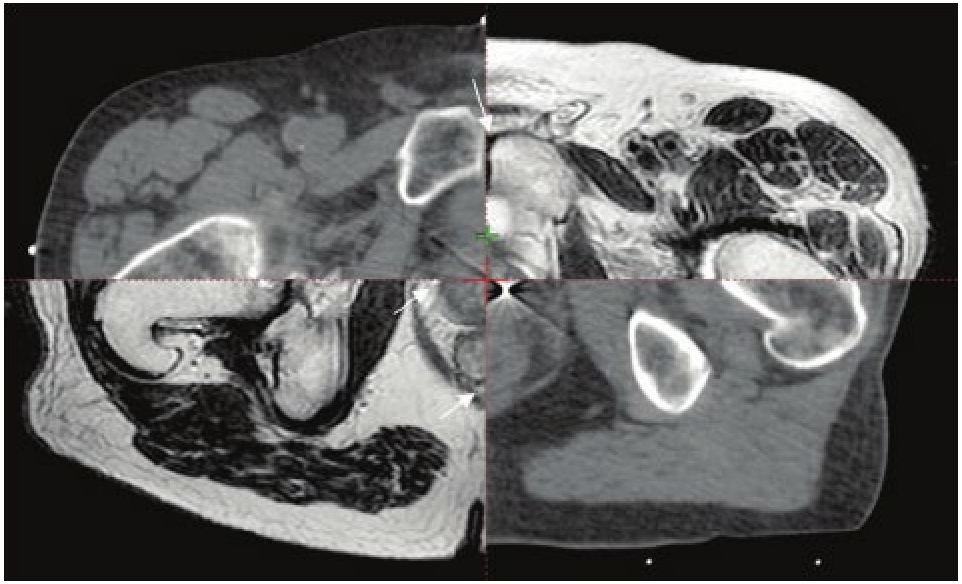
- Quick recap of medical image registration
- Introduction non-parametric / deformable image registration
- Deep learning for medical image registration:
  - Deep iterative registration
  - Supervised methods
  - Unsupervised (optimization-based) methods

Recap: medical image registration

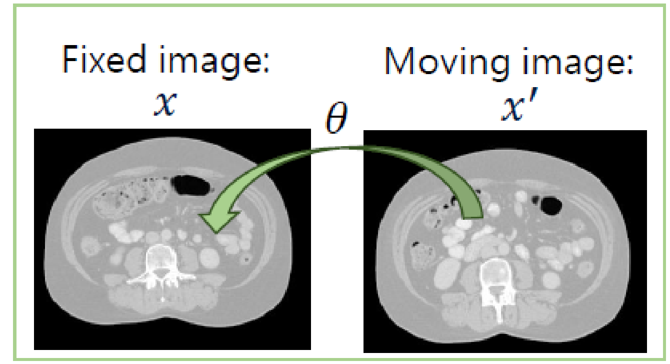
Why important?
- Multi-modal registration (e.g. CT on MRI)
- Inter-subject (e.g. atlas registration)
- Longitudinal (e.g. treatment evaluation)

Remaining challenges:
- Large 3D volumes
- Accuracy vs. efficiency

- Intensity inhomogeneities and discontinuities
- Outlier rejection



Medical image registration: general “recipe”



Transformation model:
(e.g., rigid, affine, deformable)

$$T(x', \theta)$$

Image registration

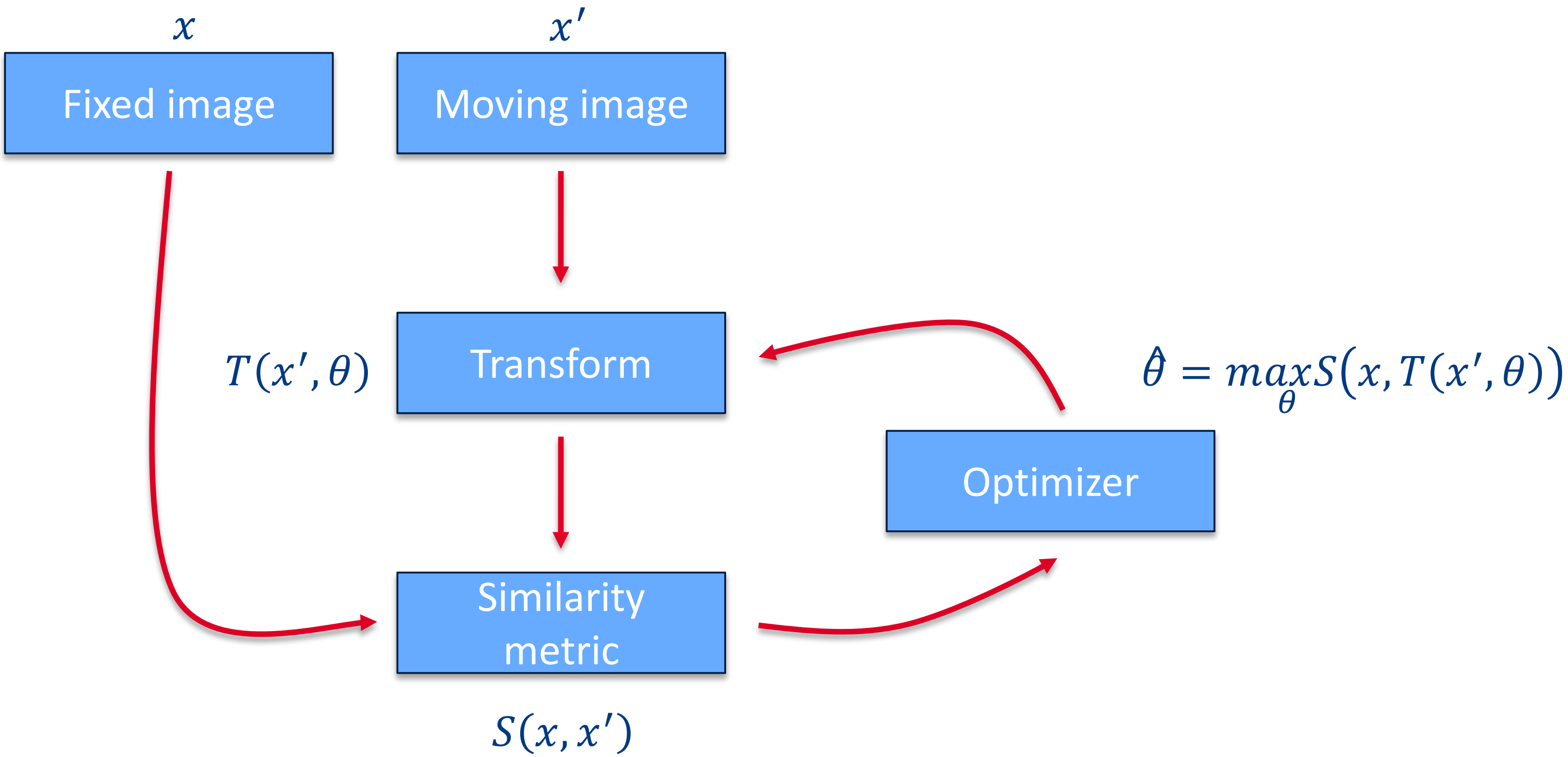
Similarity measure:
(e.g. SSD, CC, MI, MSE)

$$S(x, x')$$

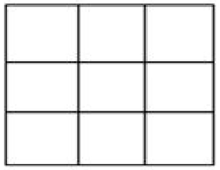
Optimization:

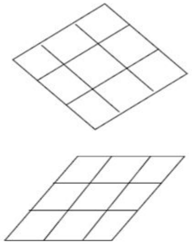
$$\hat{\theta} = \max_{\theta} S\big(x, T(x', \theta)\big)$$

Finding the optimal transformation

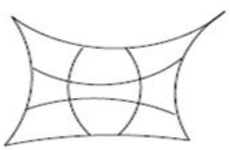


Transformation models

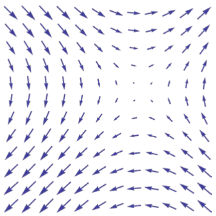




Rigid & Affine
- Point or feature-based (e.g. landmarks, fiducials)
- Intensity-based
- Gradient/edge-based



Non-linear
- Linear or higher order polynomials
- Spline-based



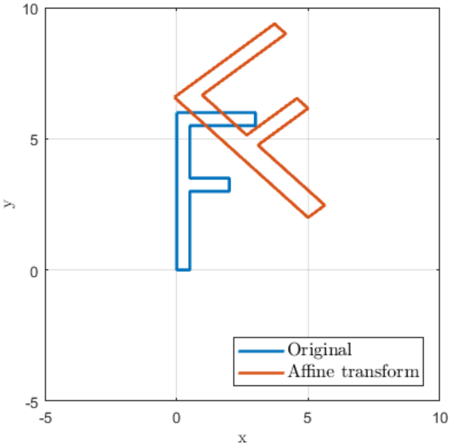
Non-parametric / deformable
- Allowing each image element to be displaced arbitrarily

Rigid transformations $\mathbf{x}' = \mathbf{R}\mathbf{x} + \mathbf{t}$

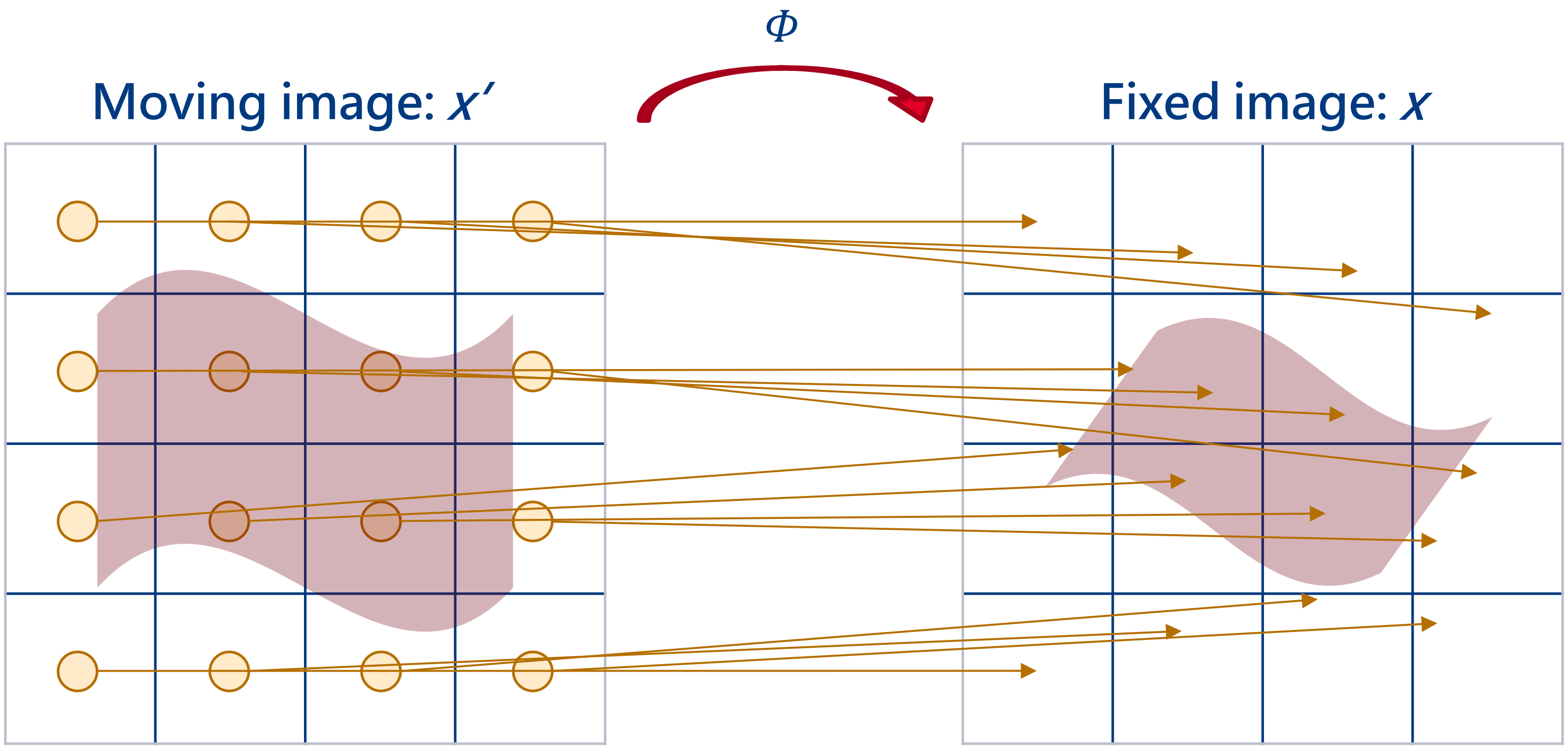
- Translation $\mathbf{t} = \begin{bmatrix} t_x \\ t_y \end{bmatrix}$
- Rotation $\mathbf{R} = \begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix}$

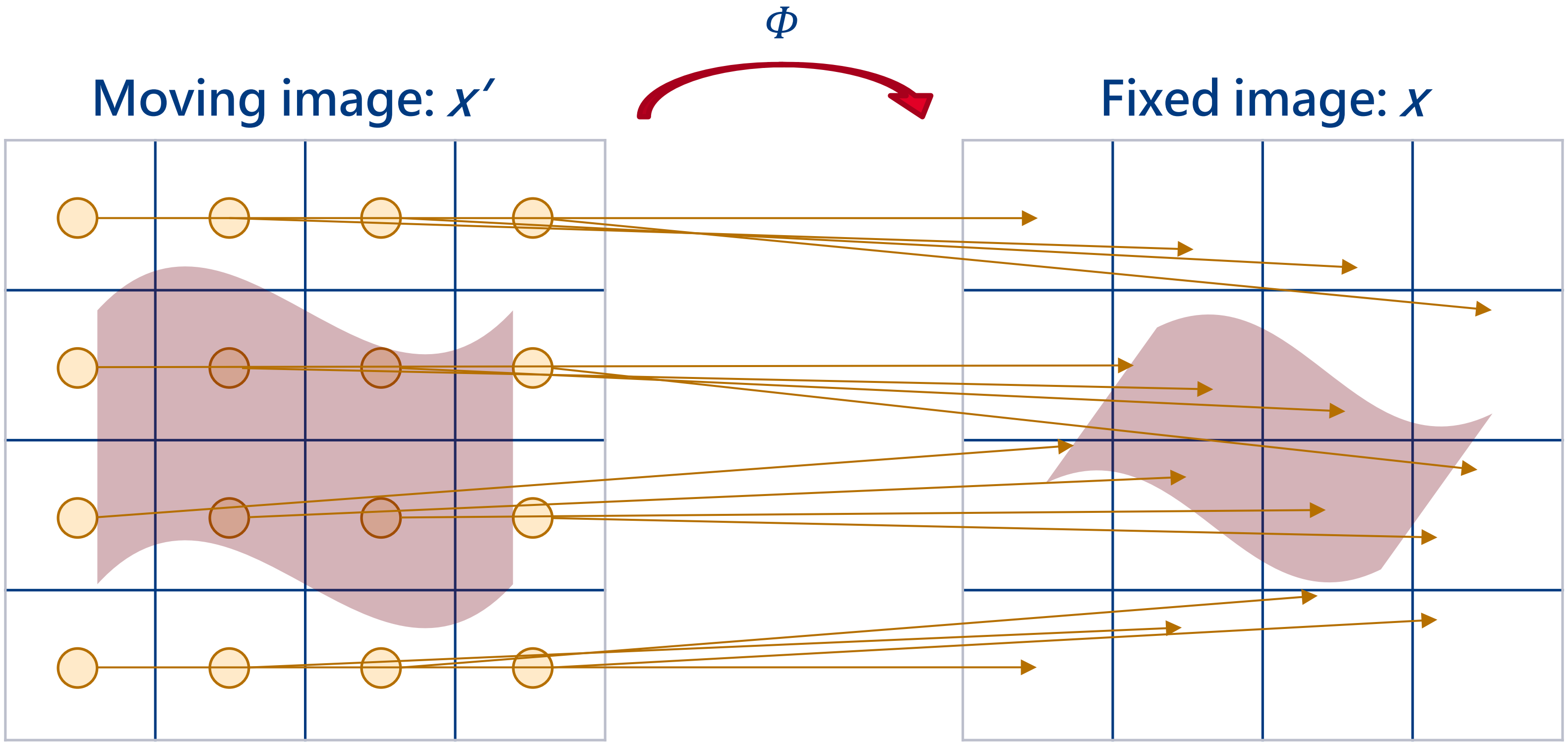
Affine transformations: translation, rotations and $\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{t}$

- Scaling $\mathbf{S} = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix}$
- Shearing $\mathbf{H} = \begin{bmatrix} 1 & h_x \\ h_y & 1 \end{bmatrix}$



Displacement (vector) field (DVF) = Dense set of vectors representing the displacement in a given spatial domain





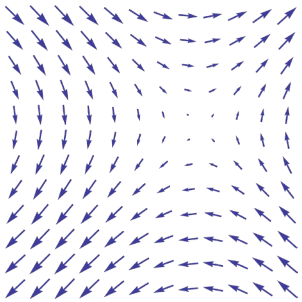
Deformation: $\varphi = Id + u, \varphi : \Omega \to \mathbb{R}^d$

or point-wise: $\varphi(x) = x + u(x)$

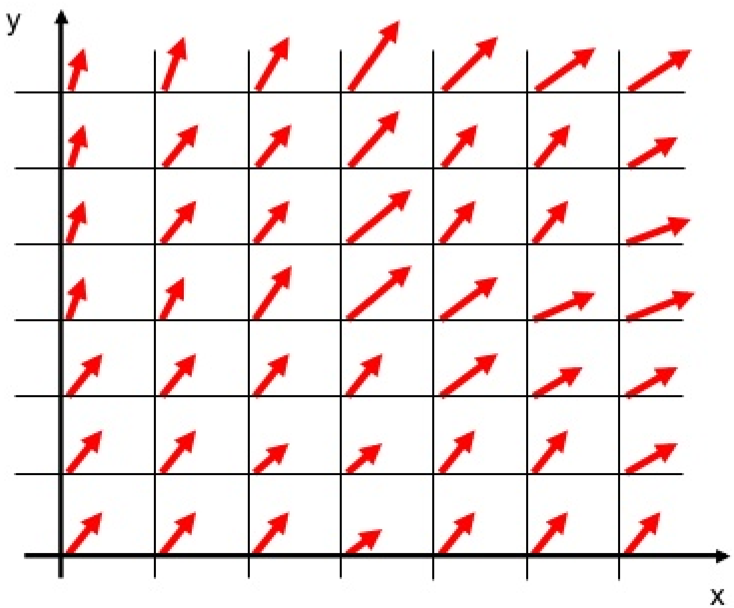
Displacement:

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

e.g. $u = [u_x, u_y, u_z]$



Deformable image registration: how many free parameters (DOF) in 3D?



Typical spatial resolution of a 3D medical image

(MRI: e.g. 1.5×1.5×4 mm^3 voxels)

DOF = 3*Nx*Ny*Nz (!)

Deformable image registration is still a very active area of research, and many different deformable registration models exist:

- Free-form deformation model
- Optical flow
- Demons
- Fluid flow
- Diffeomorphisms
- …

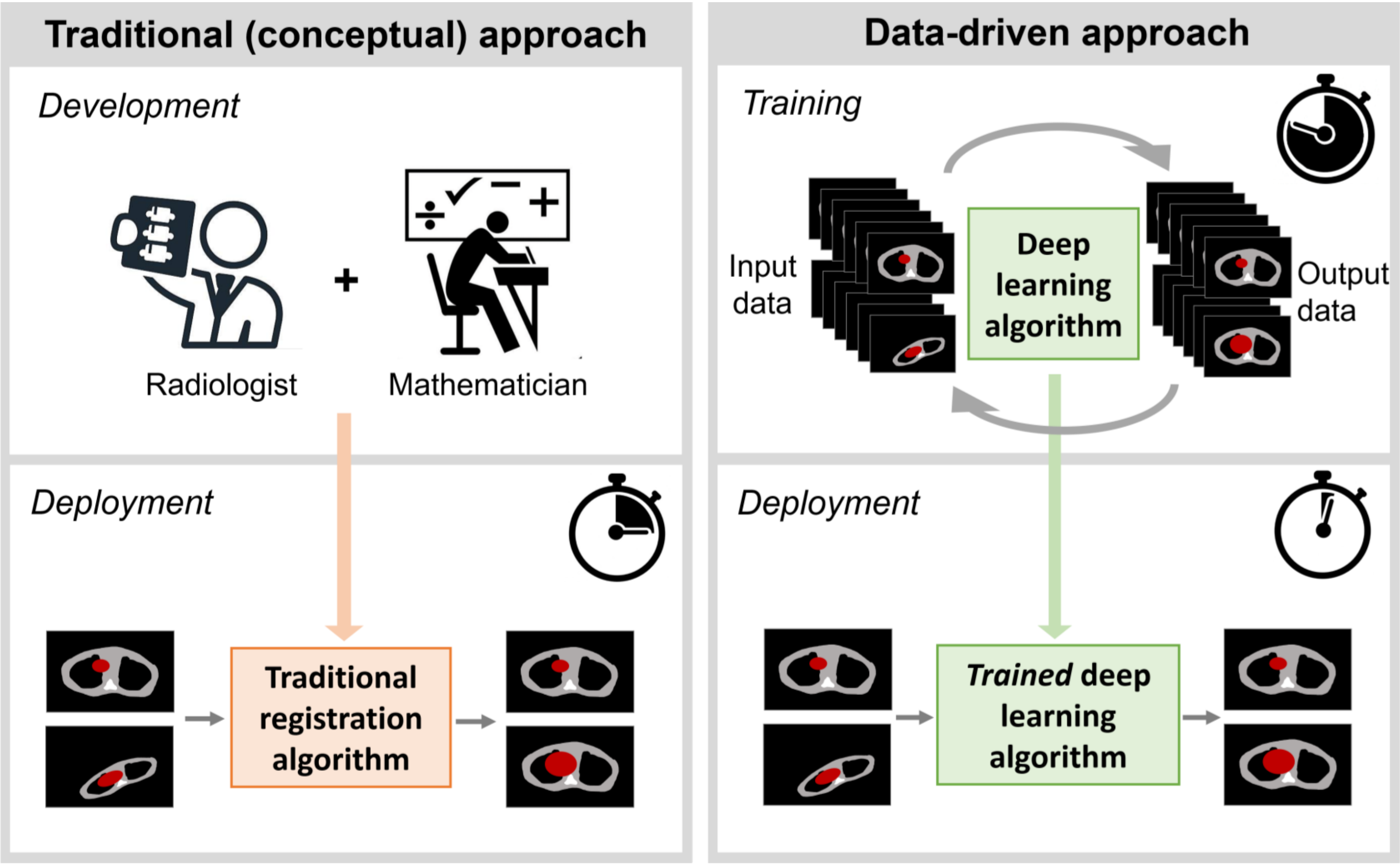
Note that the details of these models and their implementations are beyond the scope of this course.

Learning goals of this lecture

The students should be able to:

- Explain the difference between parametric and non-parametric image registration models
- Formulate deformable image registration as an optimization problem
- Understand displacement vector fields
- Explain the difference between supervised and unsupervised learning
- Understand why deep learning is an interesting technique to solve medical image registration tasks
- Understand the different ways in which deep learning can be used to perform (deformable) image registration

Why focus on deep learning for medical image registration?



Machine learning – different training strategies

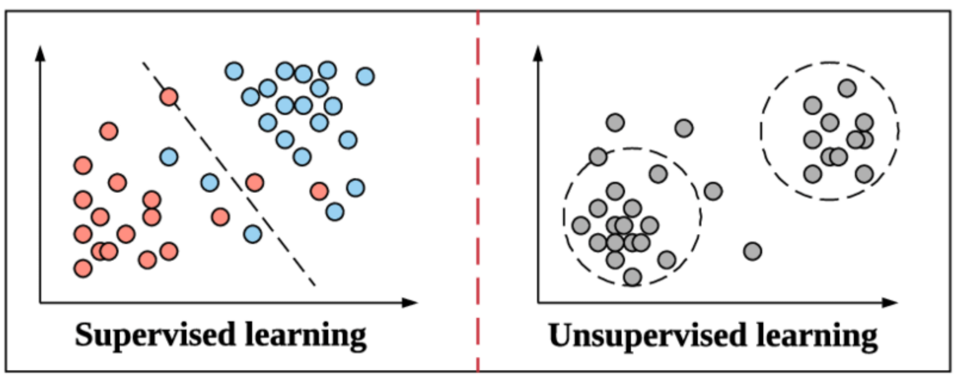
Supervised learning = develop predictive model based on input and output data (i.e., ground truth, "labels")

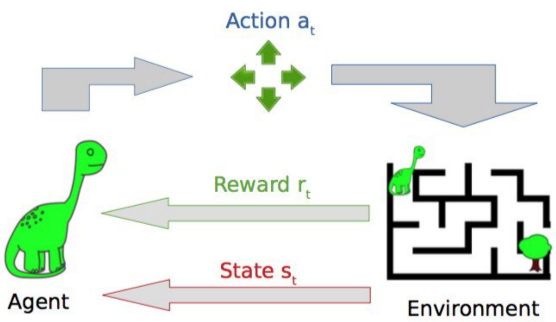
Examples: classification (e.g., skin lesion classification), regression (segmentation of vessel structures)

Unsupervised learning = group or interpret data based on input data alone

Example: clustering (e.g., k-means)

NB: also semi-supervised learning and reinforcement learning, not part of this course.





Reinforcement learning

Image 1 source: Qian et al. (2019). Orchestrating the Development Lifecycle of Machine Learning-Based IoT Applications: A Taxonomy and Survey.
Image 2 source: https://twimlai.com/whats-hot-in-ai-deep-reinforcement-learning-edition/

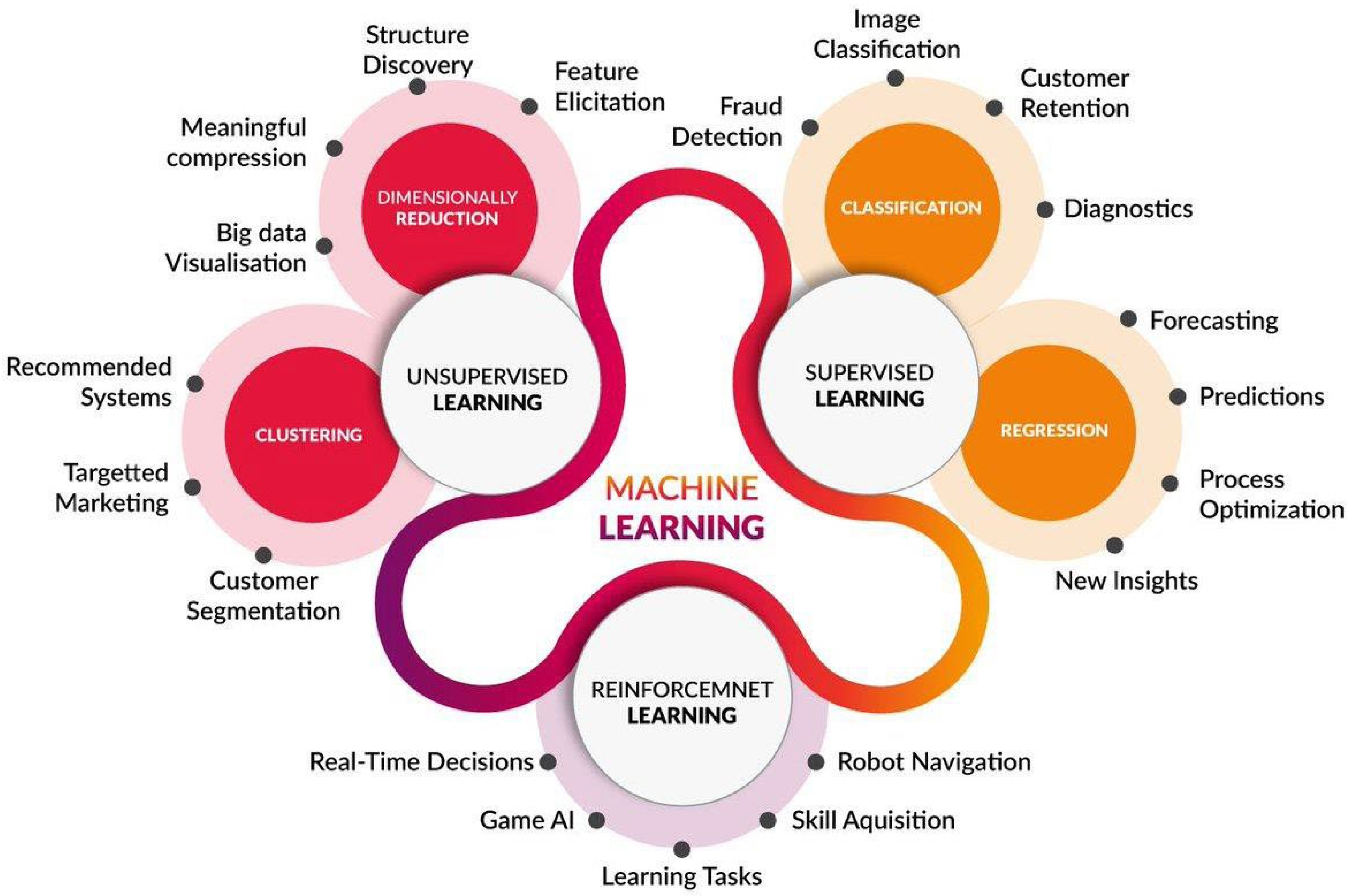


Image source: http://www.cognub.com/index.php/cognitive-platform/

Learning image registration: how does it work?

Problem: how can we obtain the ground truth displacement ?

A. Deep iterative registration

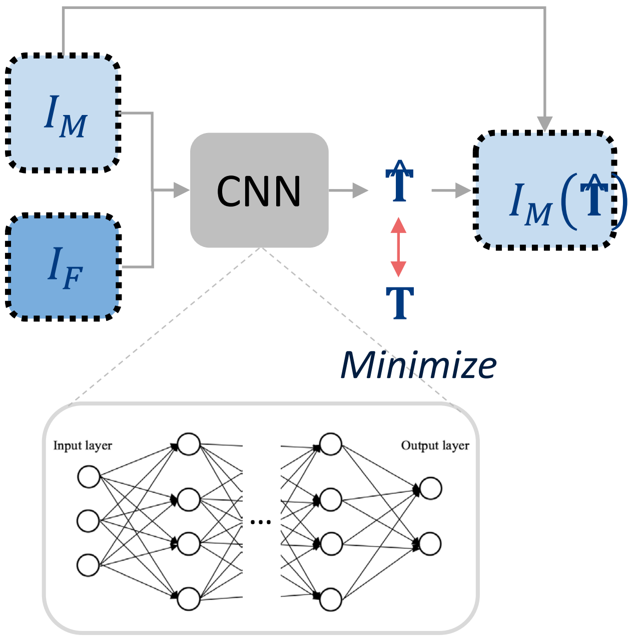
- Learn a component of a classical registration method

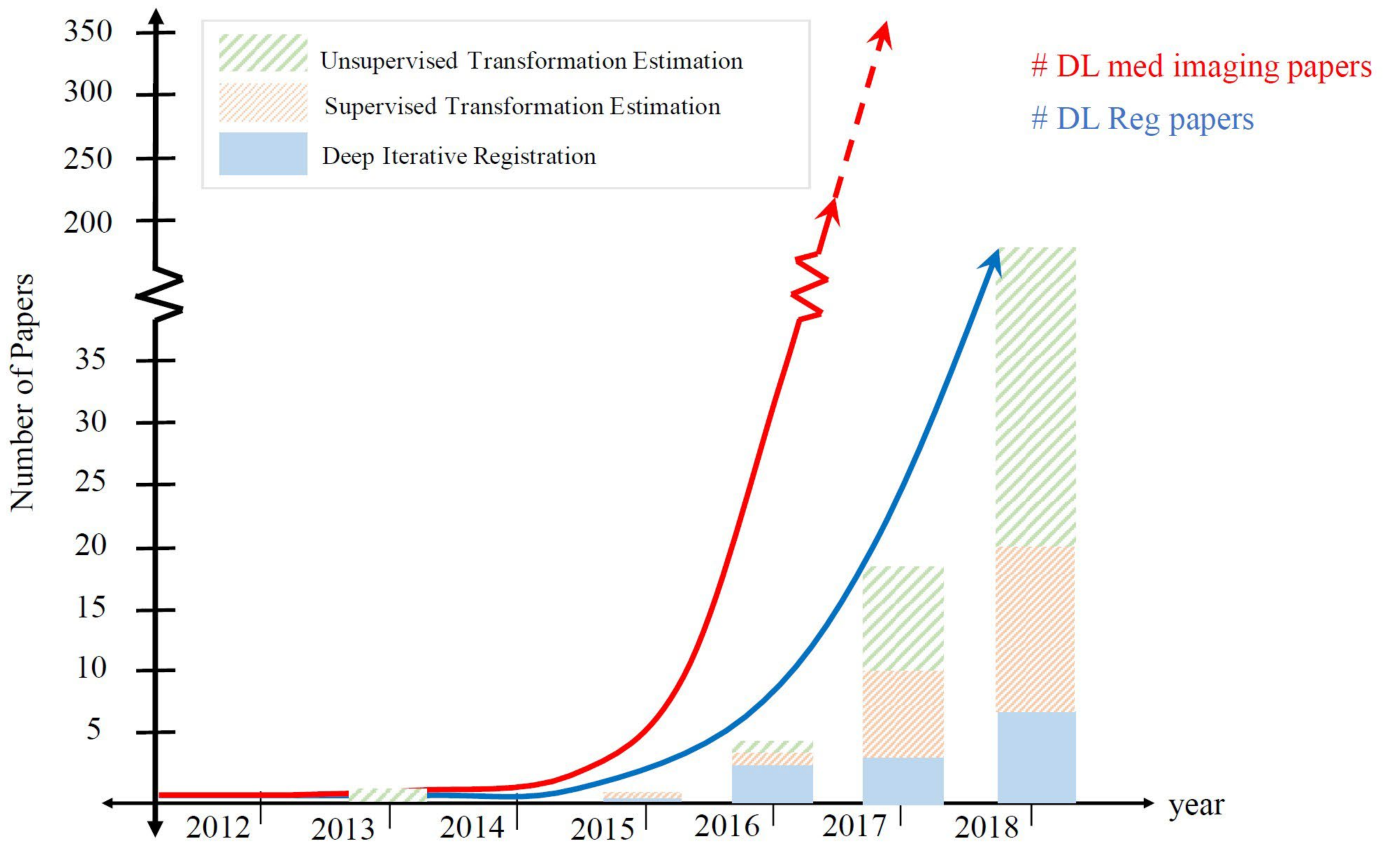
B. Supervised transformation estimation

- Obtain using classical registration method
- or make synthetic ground truth

C. Unsupervised transformation estimation

- Use similarity metric to judge





Haskins et al. (2019) Deep Learning in Medical Image Registration: A Survey. arXiv:1903.02026v1

A. Deep iterative registration

Use a traditional registration method and learn one component

Automatically extract (learn) similarity features

- *Unimodal*: Wu et al. (2013) Unsupervised deep feature learning for deformable registration of mr brain images (MICCAI); Wu et al. (2016) Scalable high-performance image registration framework by unsupervised deep feature representations learning (IEEE Transactions on Biomedical Engineering); Eppenhof et al. (2018) Error estimation of deformable image registration of pulmonary ct scans using convolutional neural networks (Journal of Medical Imaging)
- *Multimodal*: Mostly *rigid registration*; learn a similarity metric to evaluate or register multimodal images using gradient descent. E.g. Cheng et al. (2016&2018) Deep similarity learning for multimodal medical images (MICCAI)

Reinforcement learning

Mostly used for *rigid registration*

- Low-resolution transformation model for deformable registration

  Krebs et al. (2017) Robust non-rigid registration through agent-based action learning (MICCAI)

Where are we in the image registration "recipe"?

Transformation model:
(e.g., rigid, affine, deformable)

$$T(x', \theta)$$

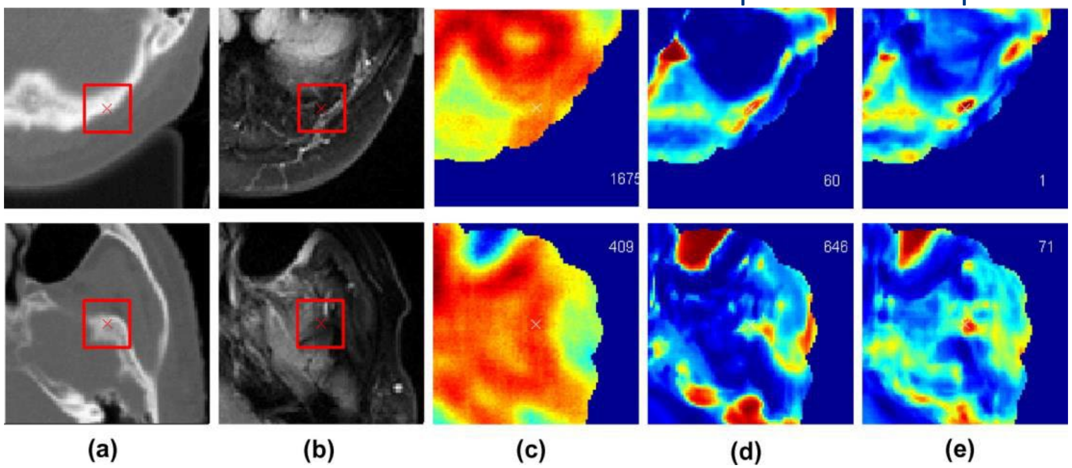
$\boxed{\begin{array}{l}\textbf{Similarity measure:}\\ \text{(e.g. SSD, CC, MI, MSE)}\\ \hfill S(x, x')\end{array}}$

Optimization:

$$\hat{\theta} = \max_{\theta} S\big(x, T(x', \theta)\big)$$

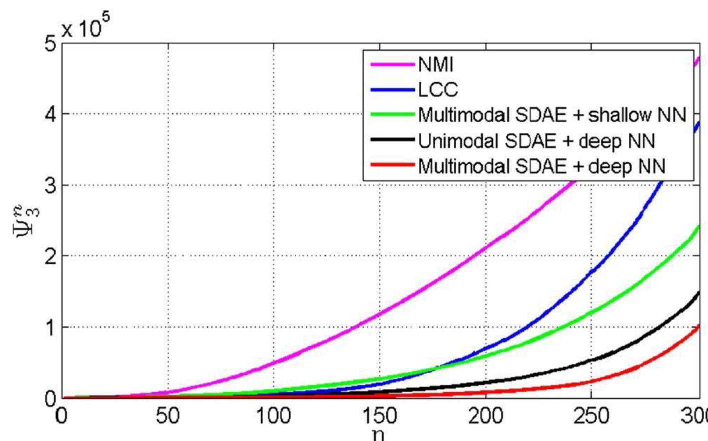
Example 1: *Learning multimodal feature extraction (Haskins et al., 2019)*

CT | MRI | MI | Uni-modal Deep NN | Multi-modal Deep NN



Similarity maps

Cumulative sum
of prediction errors



No. of patches

Cheng, Xi, Li Zhang, and Yefeng Zheng. "Deep similarity learning for multimodal medical images." *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* 6.3 (2018): 248-252

Where are we in the image registration “recipe”?

Transformation model:
(e.g., rigid, affine, deformable)

$T(x', \theta)$

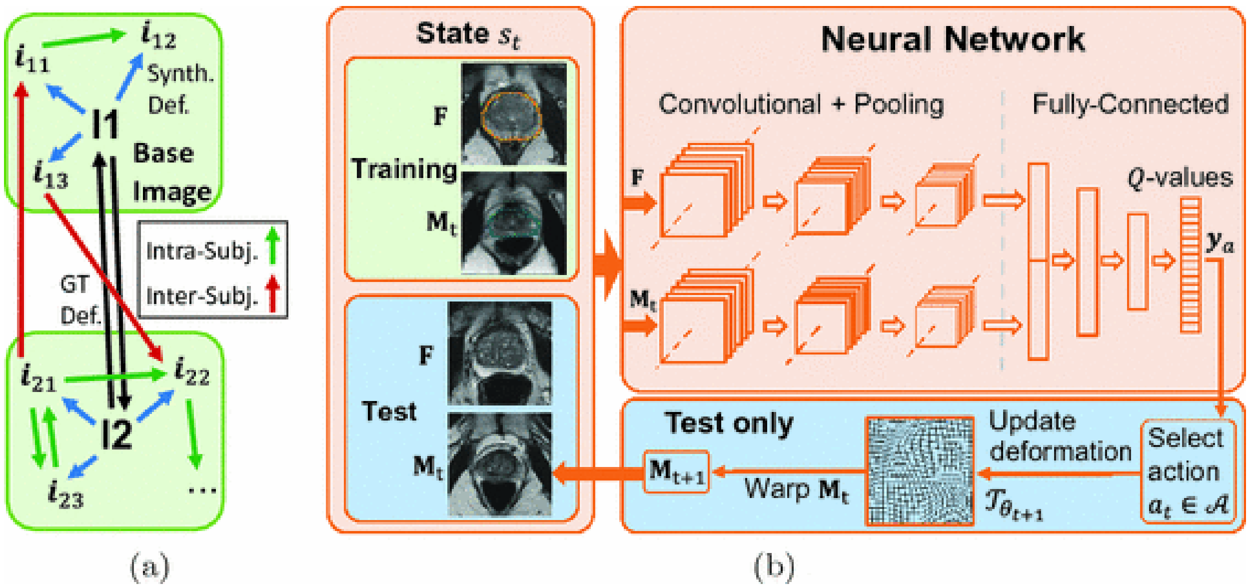
Similarity measure:
(e.g. SSD, CC, MI, MSE)

$S(x, x')$

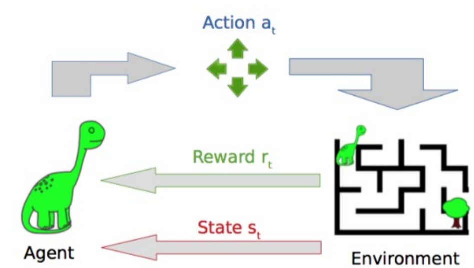
Optimization:

$\hat{\theta} = \max_{\theta} S\big(x, T(x', \theta)\big)$

Example 2: *Reinforcement learning for image registration*



(a) Training Data Generation: **Synthetic deformations** (blue arrows) and inter-subject GT deformations (black) are used for intra- (green) and inter-subject (red) image pairs for training. (b) Dual-stream network used for Q-value prediction including complete single-stage Markov Decision Process for testing (blue background).



Krebs, Julian, et al. "Robust non-rigid registration through agent-based action learning." *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham, 2017.

B. Supervised methods

Require many known transformations for training, use ground truth labels to calculate the loss

Fully supervised

2D (FlowNet) vs. 3D approaches (e.g. 3D U-net); mostly patch-based (e.g. 128^3)

- *Generation of ground truth transformations* ➔ many, e.g. Eppenhof et al. (2018)
  Pulmonary ct registration through supervised learning with convolutional neural networks (IEEE trans. on med. imaging)
- *Large Deformation Diffeomorphic Metric Mapping (LDDMM)* ➔Yang et al. (2017)
  Quicksilver: Fast Predictive Image Registration - a Deep Learning Approach (NeuroImage)

Weakly supervised

Use overlap between segmentations or a similarity metric between *M* and *F* combined with ground truth

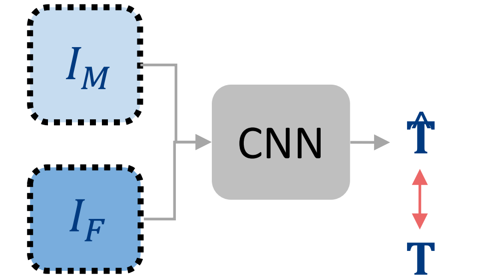
- MR-US registration
  - *CNN* ➔ Hu et al. (2018) Weakly-supervised convolutional neural networks for multimodal image registration (Medical Image Analysis)
  - *Generative Adversarial Networks (GANs)* ➔ Yan et al. (2018) Adversarial image registration with application for mr and trus image fusion (arXiv preprint: 1804.11024)

Supervised learning of the transformation requires a ground truth!

Questions:

- What is the ground truth of an image registration task?
  - Parametric?
  - Non-parametric?

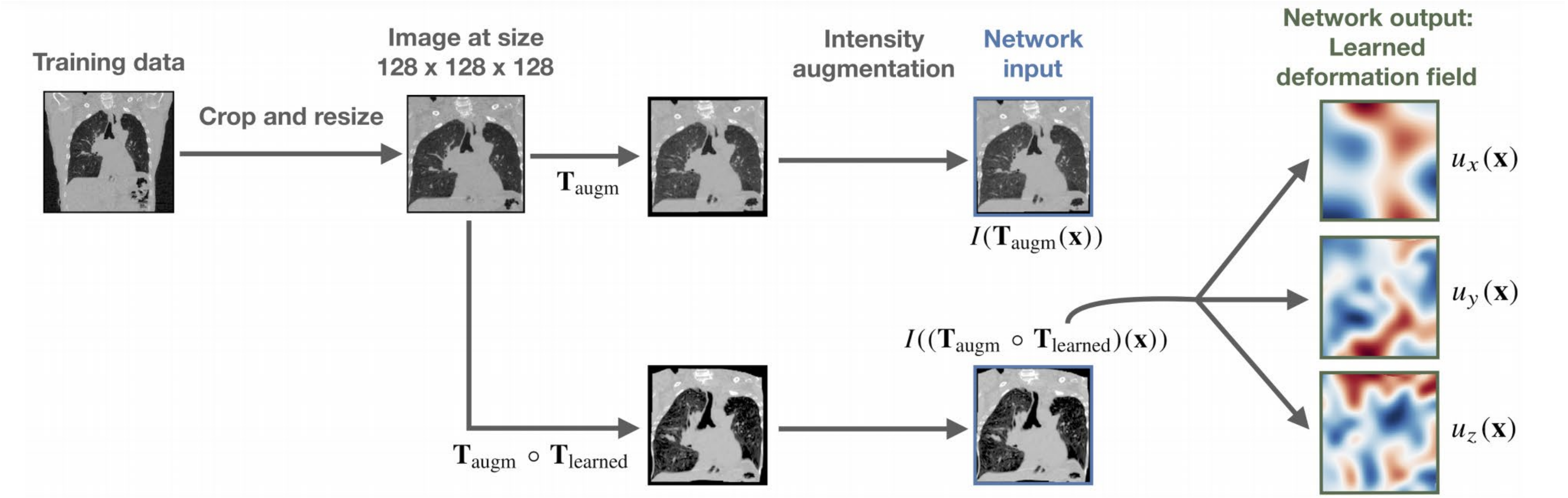
- How can we acquire this?



B. Supervised transformation estimation

Requires many known transformations for training, use ground truth labels to calculate the loss

Example 1: *"On-the-fly" simulation of displacement fields: T_{augm} (Eppenhof & Pluim, 2018)*



Eppenhof, Koen AJ, and Josien PW Pluim. "Pulmonary CT registration through supervised learning with convolutional neural networks." *IEEE transactions on medical imaging* 38.5 (2018): 1097-1105.

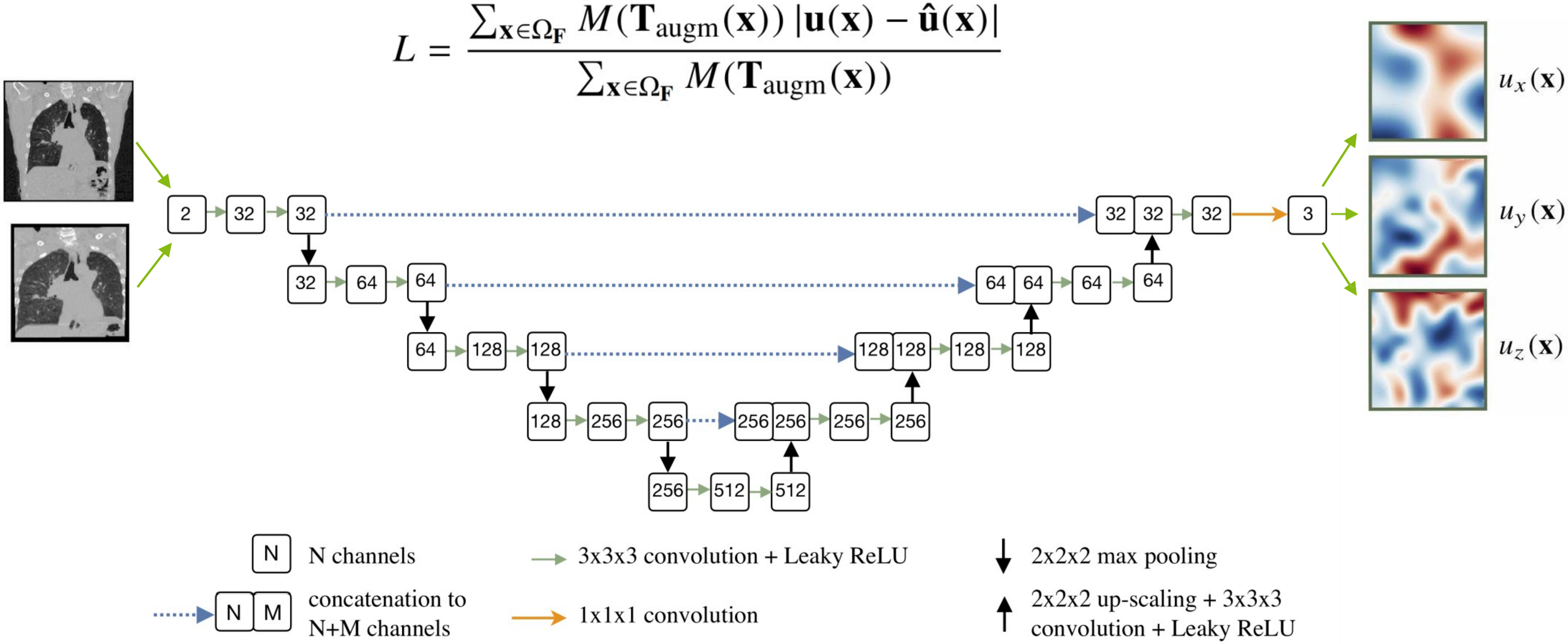
Learning the displacement vector field (Eppenhof & Pluim, 2018)

Binary mask of the lungs

True vector field

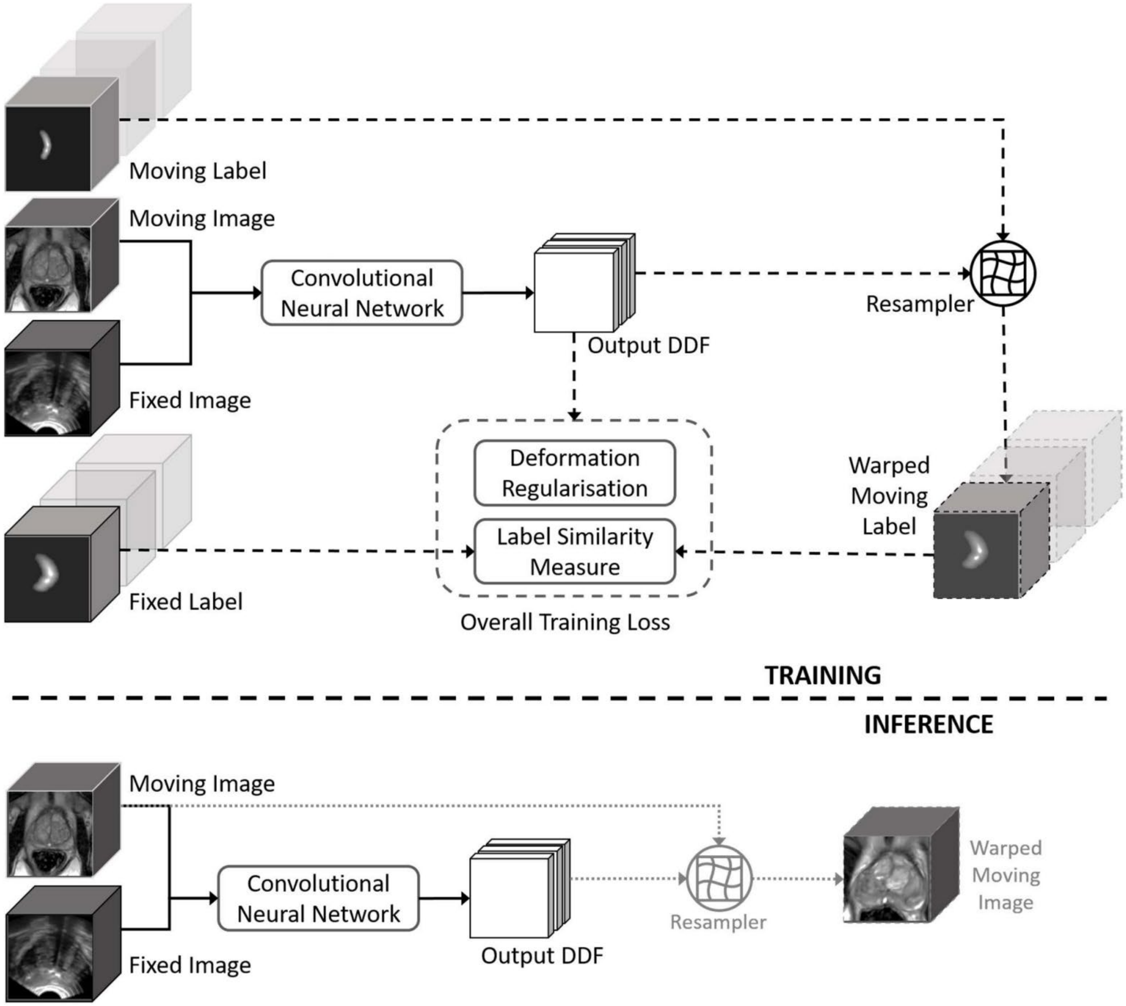
Estimated vector field

$$L = \frac{\sum_{\mathbf{x}\in\Omega_\mathbf{F}} M(\mathbf{T}_{\text{augm}}(\mathbf{x}))\,|\mathbf{u}(\mathbf{x}) - \hat{\mathbf{u}}(\mathbf{x})|}{\sum_{\mathbf{x}\in\Omega_\mathbf{F}} M(\mathbf{T}_{\text{augm}}(\mathbf{x}))}$$



Eppenhof, Koen AJ, and Josien PW Pluim. "Pulmonary CT registration through supervised learning with convolutional neural networks." *IEEE transactions on medical imaging* 38.5 (2018): 1097-1105.

Example 2: *Weakly-supervised* CNN for MR-US registration



Hu, Yipeng, et al. "Weakly-supervised convolutional neural networks for multimodal image registration." *Medical image analysis* 49 (2018): 1-13.

C. Unsupervised methods

No ground truth needed. Often use a spatial transformer layer.

Variational autoencoders ➔ Krebs et al. (2018) Unsupervised Probabilistic Deformation Modeling for Robust Diffeomorphic Registration. (DLMIA)

Generative Adversarial Networks (GANs) ➔ Tanner et al. (2018) Generative Adversarial Networks for MR-CT Deformable Image Registration (CVPR) & Hu et al. (2018) Adversarial Deformation Regularization for Training Image Registration Neural Networks (MICCAI)

Multi-scale methods:

- **RegNet** ➔ Sokooti et al. (2017) Nonrigid Image Registation Using Multi-scale 3D Convolutional Neural Networks (MICCAI). NB: dual path
- **ConvNet** ➔ De Vos et al. (2018) A deep learning framework for unsupervised affine and deformable image registration (Medical Image Analysis). NB: chain
- **pgCNN** ➔ Eppenhof et al. (2019) Progressive Growing Convolutional Networks for End-to-End Deformable Image Registration (SPIE medical imaging)

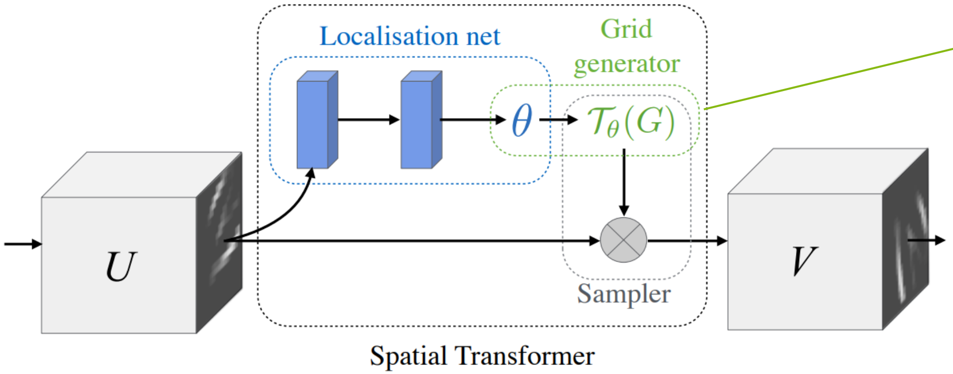
VoxelMorph (U-net) ➔ Balakrishnan et al. (2019) VoxelMorph: A Learning Framework for Deformable Medical Image Registration (IEEE tran.med.imaging)

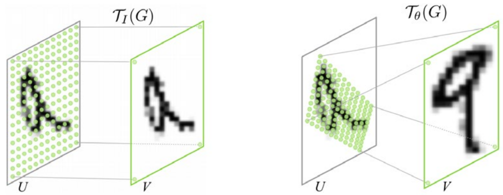
- **Cycle-consistent VoxelMorph** ➔ Kim et al. (2019) Unsupervised Deformable Image Registration Using Cycle-Consistent CNN (MICCAI)

Spatial transformer networks (Jaderberg et al., NIPS 2015)

Spatial transformer = a learnable module that explicitly allows the spatial manipulation of data within the network

- *Differentiable*
- Can be inserted into existing convolutional neural networks
- Actively transforms feature maps (conditional on the feature map itself)



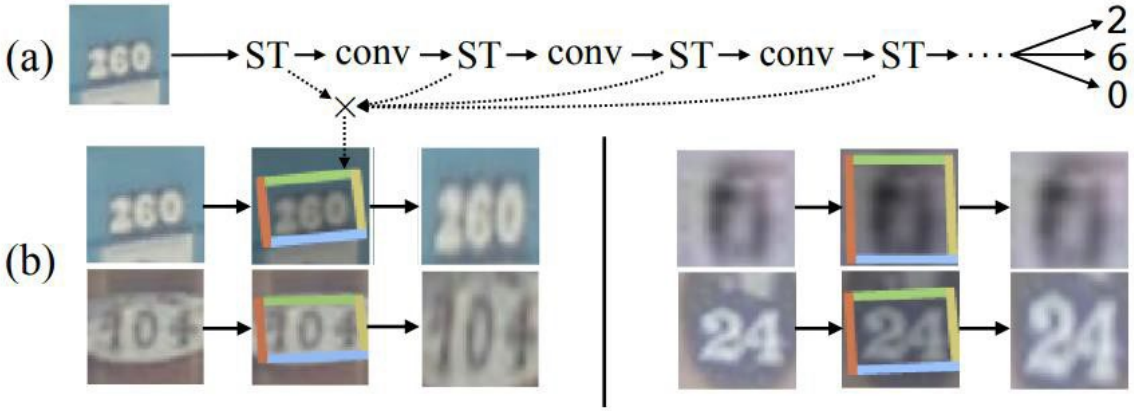


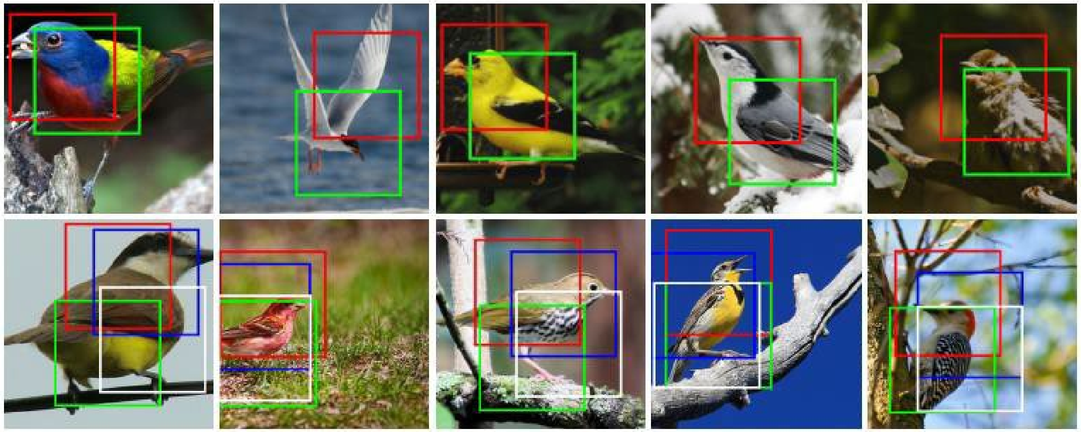
Example $T_\theta(G)$ for an affine transformation:

$$\begin{pmatrix} x_i^s \\ y_i^s \end{pmatrix} = \mathcal{T}_\theta(G_i) = \mathtt{A}_\theta \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix}$$

Jaderberg, Max, Karen Simonyan, and Andrew Zisserman. "Spatial transformer networks." *Advances in neural information processing systems*. 2015.

CNNs are in essence not invariant to translation, scale, rotation and more generic warping of the input data

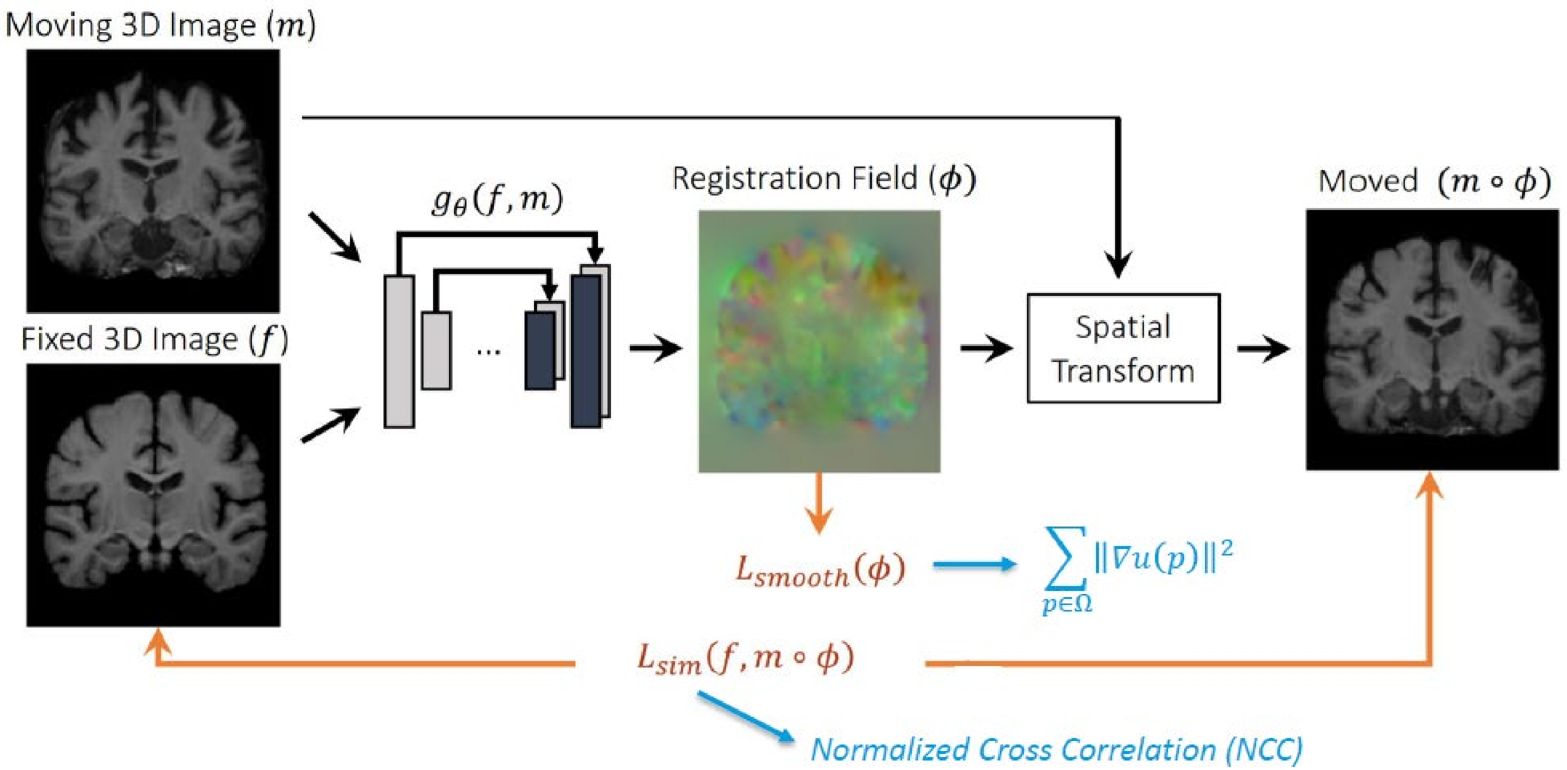




Jaderberg, Max, Karen Simonyan, and Andrew Zisserman. "Spatial transformer networks." *Advances in neural information processing systems*. 2015.

VoxelMorph

Example of a popular unsupervised framework for deformable image registration



$$\hat{\Phi} = \underset{\Phi}{\operatorname{argmin}}\, L_{sim}(f, m \circ \Phi) + \lambda L_{smooth}(\Phi)$$

Balakrishnan et al. (2018): VoxelMorph: A Learning Framework for Deformable Medical Image Registration. arXiv:1909.05231v3

Summary

- Parametric vs. non-parametric image registration
  - Displacement vector fields
- Formulation of deformable image registration
- Different ways to use deep learning for image registration
  - Deep iterative registration
  - Supervised learning
  - Unsupervised learning
- Disadvantages of deep learning for image registration (performance, # of training data, ground truth, …)