Two-stage Unsupervised Learning Method for Affine and Deformable Registration

Dongdong Gu, Guocai Liu, Juanxiu Tian, Qi Zhan

Department of Electrical and Information Engineering, Hunan University, China

Introduction

Image registration aims to establish spatial correspondences between a pair of images. The process is typically formulated as an optimization problem which seeks spatial transformation φ by maximizing similarity of the fixed image F and moving image M.

$$\underset{\varphi}{\operatorname{argmin}}\,\Re(F,M,\varphi) = \underset{\varphi}{\operatorname{argmin}}\,(\underbrace{Sim(F,M(\varphi))}_{\text{Image similarity}} + \gamma \underbrace{Reg(\varphi)}_{\text{Regularization}})$$

Our goal: learn the global and local spatial transformation parameters in an unsupervised CNN model.

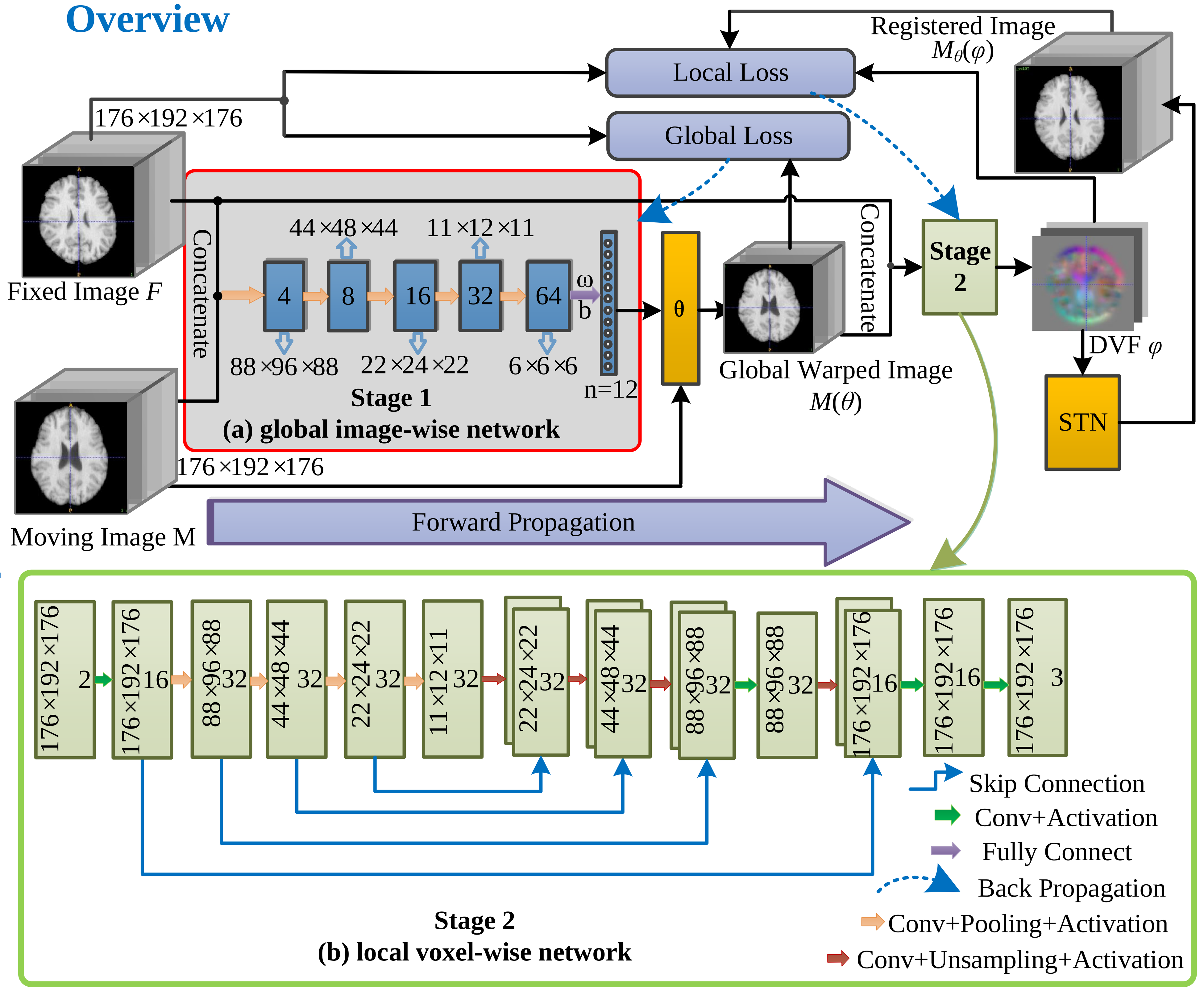
Problems and Solutions

- ➢ Iterative optimization methods are computationally expensive and time consuming. Registration performance is sensitive and requires user interaction to tune parameters.
- ✓ We model registration pipelines as an optimizable function parameterized by deep network weights which are learned in the training period.
- ➢ Preparing the supervised information for medical image registration is time consuming and requires specialized knowledge.
- ✓ We proposed a two-stage unsupervised learning method to find the global displacement map and local deformation field. The training does not need anatomical labels or ground-truth deformations fields. The fixed image is not restricted to a template image. The preprocess step does not need aligning images

Approach

- In the first stage, we learn a global image-wise affine map θ by a deep network. In the second stage, we learn a local voxel-wise deformation vector field φ by an encoder-decoder architecture.

Overview



Loss Functions

- The first stage measures NCC of the whole fixed image F and the moving image M to capture global information:

$$L_{global}(F,M,\theta) = -Ls_{im}(F,M(\theta))$$
$$= -\frac{\left(\sum_{x_i\in\Omega}(I_F(x_i)-\bar{I}_F)\,(I_{M_\theta}(x_i)-\bar{I}_{M_\theta})\right)^2}{\sum_{x_i\in\Omega}(I_F(x_i)-\bar{I}_F)^2\sum_{x_i\in\Omega}(I_{M_\theta}(x_i)-\bar{I}_{M_\theta})^2 + 1e^{-5}}$$

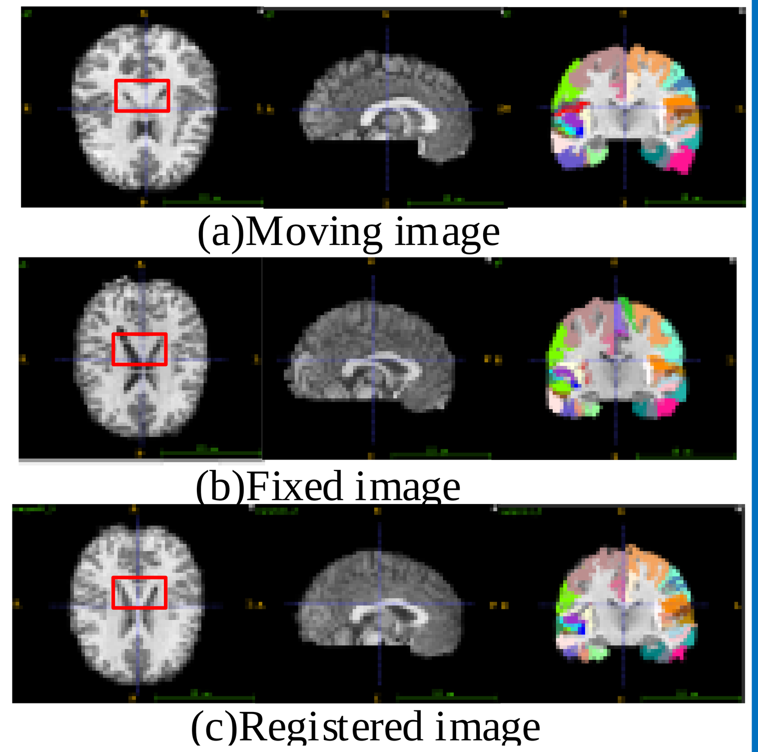
- In the second stage, loss function is defined as local NCC in addition with a regularization term (L_2 norm of φ on its gradients):

$$L_{local}(F,M_\theta,\varphi) = -Ls_{im}(F,M_\theta(\varphi)) + \gamma L_{r(\varphi)}$$
$$= -\sum_{k\in\Omega}\frac{\left(\sum_{x_i\in\Omega^k}(I_F(x_i)-\bar{I}_{Fk})\left(I_{M_{\theta(\varphi)}}(x_i)-\bar{I}_{M_{\theta(\varphi)}k}\right)\right)^2}{\sum_{x_i\in\Omega^k}(I_F(x_i)-\bar{I}_{Fk})\sum_{x_i}\left(I_{M_{\theta(\varphi)}}(x_i)-\bar{I}_{M_{\theta(\varphi)}k}\right)+1e^{-5}} + \gamma\sum_{k\in\Omega}\|\nabla\varphi(k)\|^2$$

Experimental Results

We used four public T1 weighted 3D brain MR images datasets: LPBA40, IBSR18, CUMC12 and MGH10, with 58-22 patient-level splits for training and testing.

Experiment results of MGH10 dataset



Mean DSCs (%) of 4 datasets by the two-stage network

Dataset	Original	Global	Registered
LPBA40	54.5	61.4	70.5
IBSR18	46.1	51.4	65.1
CUMC12	31.3	32.8	46.5
MGH10	40.6	41.4	43.9

Comparison with state-of-the-art methods(mean DSCs and standard deviations)

Dataset	Our method	Affine	BSpline	Fsf_ Demons	SyN	Voxel Morph
LPBA40	0.705 ±0.099	0.644 ±0.101	0.683 ±0.088	0.699 ±0.101	0.724 ±0.086	0.659 ±0.127
IBSR18	0.651 ±0.232	0.493 ±0.227	0.498 ±0.221	0.677 ±0.214	0.695 ±0.228	0.658 ±0.233
CUMC12	0.465 ±0.206	0.382 ±0.185	0.436 ±0.185	0.476 ±0.206	0.507 ±0.196	0.446 ±0.210
MGH10	0.439 ±0.185	0.402 ±0.176	0.399 ±0.169	0.406 ±0.180	0.236 ±0.178	0.436 ±0.184

The proposed approach outperforms several state-of-the-art methods in terms of accuracy and efficiency. It achieves promising results in volume overlap values and errors between fixed labels and warped moving labels.

Conclusions

The two-stage unsupervised CNN architecture can speed up 3D medical image registration pipelines. Anatomical labels or ground-truth deformation fields are not required. It can capture the global map and local deformation vector field. It can be tailored to different datasets without iteration and tuning parameters.