Deep Information Theoretic Registration

Alireza Sedghi[1], Jie Luo[2,4], Alireza Mehrtash[2,5], Steve Pieper[2], Clare M. Tempany[2], Tina Kapur[2], Parvin Mousavi[1], and William M. Wells III[2,3]

[1] Medical Informatics Laboratory, Queen's University, Kingston, Canada
[2] Radiology Department, Brigham and Womens Hospital, Harvard Medical School, Boston, USA
[3] Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, USA
[4] Graduate School of Frontier Sciences, The University of Tokyo, Japan
[5] Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, Canada

Considering a pair of intensity values a,b of a pair of images, as random variables A and B. It is possible to estimate marginal p_A, p_B and joint distributions $p_{A,B}$ by normalization of the marginal and joint histograms $h(a,b)$ of the images:

$$\tilde{p}_A(a) = \sum_b \tilde{p}_{A,B}(a,b), \tilde{p}_B(b) = \sum_a \tilde{p}_{A,B}(a,b) \quad (5)$$

with

$$\tilde{p}_{A,B}(a,b) = h(a,b) \cdot \left\{ \sum_{c,d} p_{A,B}(c,d) \right\}^{-1}$$

Figure 2. MRI image, and joint histograms before and after registration.

Contents

Motivation

1. Mutual information (MI) and its variants have resulted in notable successes
2. Despite their strength, MI and its variants do not perform well for inter-modality image registration where, e.g., one modality has “tissue contrast" while the other has “boundary contrast” (e.g., CT to ultrasound registration).

$$\hat{\beta} = \operatorname*{argmax}_{\beta} \ln p(U, V; \beta, \hat{\theta}) = \operatorname*{argmax}_{\beta} \sum_{i} \ln p(u_i, v_i; \beta, \hat{\theta}) \ .$$

Pipeline

1. The network has a 5-layer architecture consisting of strided 3D convolutions of size $3\times3\times3$ and ReLU activation functions followed by an average pooling layer and a sigmoid.

Maximum Likelihood Registration

1. Out put of the network

$$p(z = 1|u, v; \theta) \doteq \sigma(f(u, v, \theta)) \ ,$$

Maximum Likelihood Registration

3. Cross entropy loss: maximum likelihood

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \sum_{i} \ln p(z_i | u_i, v_i; \theta) \ .$$

Maximum Likelihood Registration

4. Construct a joint distribution on registered patches that is based on the classifier. From Bayes' rule, and noting that zi does not depend on the parameters

$$p(u_i, v_i | z_i; \theta) p(z_i) = p(z_i | u_i, v_i; \theta) p(u_i, v_i; \theta) \ .$$

Maximum Likelihood Registration

5. Taking logs and subtracting over the two cases on the value of zi

$$p(u_i, v_i | z_i; \theta) p(z_i) = p(z_i | u_i, v_i; \theta) p(u_i, v_i; \theta) \ .$$

$$\ln p(u_i, v_i | z_i = 1; \theta) = \\ \ln p(u_i, v_i | z_i = 0; \theta) + \ln \left(\frac{p(z_i = 1 | u_i, v_i; \theta)}{p(z_i = 0 | u_i, v_i; \theta)} \right) - \ln \left(\frac{p(z_i = 1)}{p(z_i = 0)} \right) \ .$$

Maximum Likelihood Registration

6. Sigmoid function is a logistic function, so we apply a logit transformation

$$\begin{cases} \ln p(u_i, v_i | z_i = 1; \theta) = \\ \qquad \ln p(u_i, v_i | z_i = 0; \theta) + \ln\left(\frac{p(z_i = 1 | u_i, v_i; \theta)}{p(z_i = 0 | u_i, v_i; \theta)}\right) - \ln\left(\frac{p(z_i = 1)}{p(z_i = 0)}\right) . \\ p(z = 1 | u, v; \theta) \doteq \sigma(f(u, v, \theta)) , \end{cases}$$

$$\ln p(u_i, v_i | z_i = 1; \theta) = \ln(p(u_i, v_i | z_i = 0; \theta)) + f(u_i, v_i; \theta) + C .$$

Maximum Likelihood Registration

For the purpose of ML registration, we construct the joint distribution on patches conditioned on a transformation with parameters β as follows:

$$
\begin{cases}
p(u_i, v_i; \beta, \hat{\theta}) \propto p(u_i, {}^{\beta}v_i | z_i = 1; \hat{\theta}) \; . \\
\ln p(u_i, v_i | z_i = 1; \theta) = \ln(p(u_i, v_i | z_i = 0; \theta)) + f(u_i, v_i; \theta) + C \; .
\end{cases}
$$

$$\ln p(u_i, v_i; \beta, \hat{\theta}) = \ln p(u, {}^{\beta}v_i | z_i = 0) + f(u_i, {}^{\beta}v_i) + C \; .$$

Maximum Likelihood Registration

For the purpose of ML registration, we construct the joint distribution on patches conditioned on a transformation with parameters β as follows:

$$\ln p(u_i, v_i; \beta, \hat{\theta}) = \ln p(u, {}^{\beta}v_i | z_i = 0) + f(u_i, {}^{\beta}v_i) + C \ .$$

$$\boxed{\hat{\beta} = \underset{\beta}{\operatorname{argmax}} \sum_i \ln p(u_i, v_i; \beta, \hat{\theta}) \approx \underset{\beta}{\operatorname{argmax}} \sum_i f(u_i, {}^{\beta}v_i, \hat{\theta}) \ .}$$

Iterative Maximum Likelihood

$$\hat{\beta}^{n+1} = \underset{\beta}{\operatorname{argmax}} \sum_i \ln p(u_i, v_i; \beta, \theta^n) \approx \underset{\beta}{\operatorname{argmax}} \sum_i f(u_i, {}^{\beta}v_i, \hat{\theta^n})$$

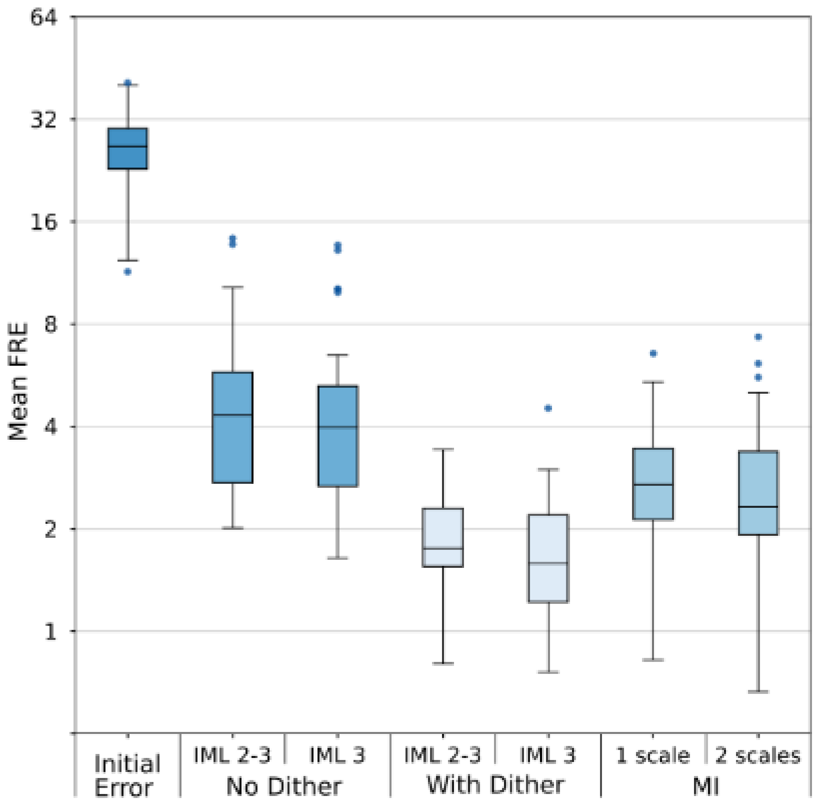
$$\hat{\theta}^{n+1} = \underset{\theta}{\operatorname{argmax}} \sum_i \ln p(u_i, v_i, \hat{\beta^n}, \theta) \approx \underset{\theta}{\operatorname{argmax}} \sum_i \ln p(z_i | u, {}^{\hat{\beta^n}}v_i; \theta) \ .$$

Experiments

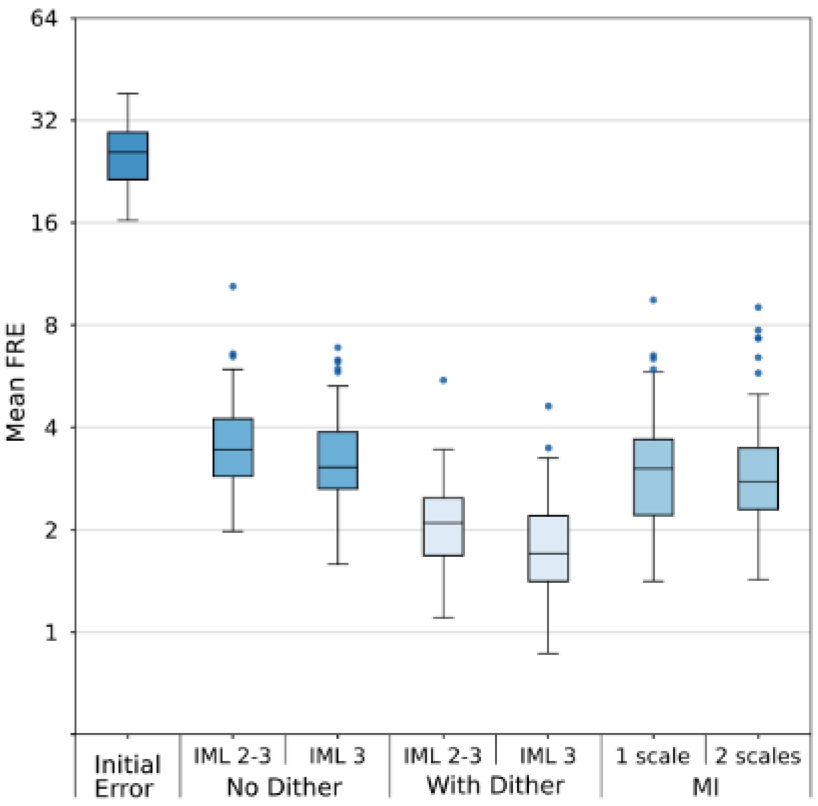
1. IXI Brain Development Dataset [17] which contains aligned T1-T2 image pairs from healthy subjects.
2. In the experiments, use 60 subjects for training and another 60 subjects for validation.
3. All images are resampled to 1×1×1 mm, and their intensity is normalized between the range of [0, 1]
4. 3D patches, ui, vi, of size 17×17×17 with labels zi = 1or 0
5. Learning rate of 5×10−5, batch size of 256 and L2-regularization (weight decay) of 0.005

Data Augmentation

1. Dithering: randomly (normal distribution) moving images for mm

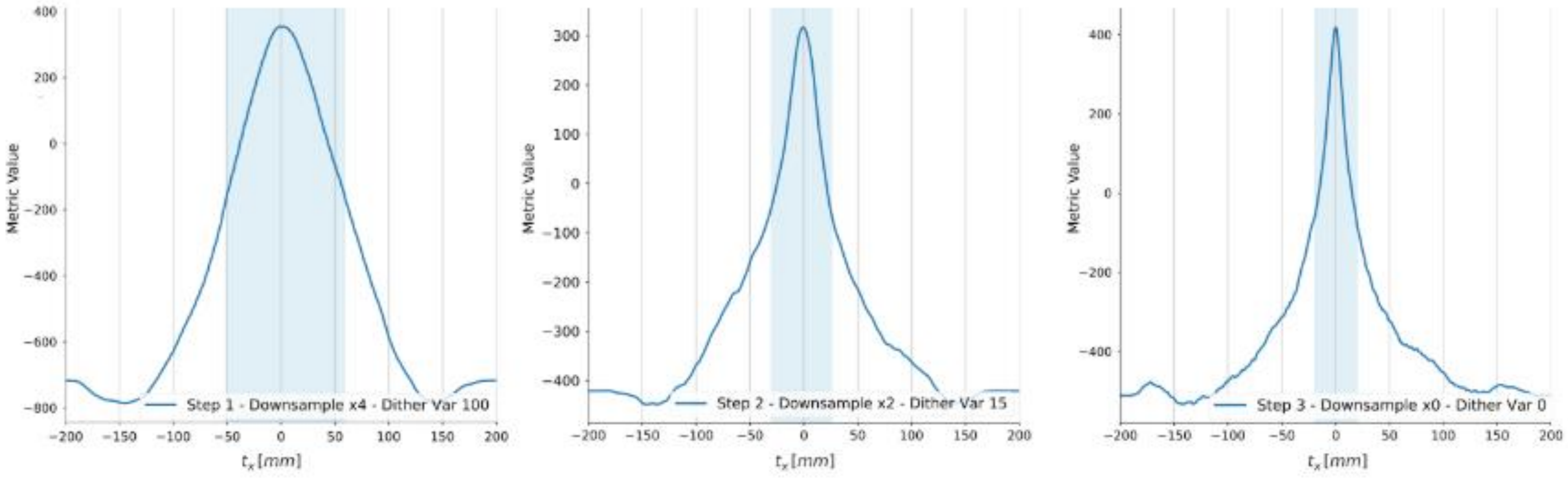


(a) Rigid Registration



(b) Affine Registration

Fig. 1. Box plots of mean FRE for rigid (a) and affine (b) registration between T1 and T2 images.



(a) (b) (c)

Fig. 2. Response functions for each iteration of IML for rigid registration plotted as a function of translation for a pair of registered fixed and moving images. Shading illustrates full-width half max.

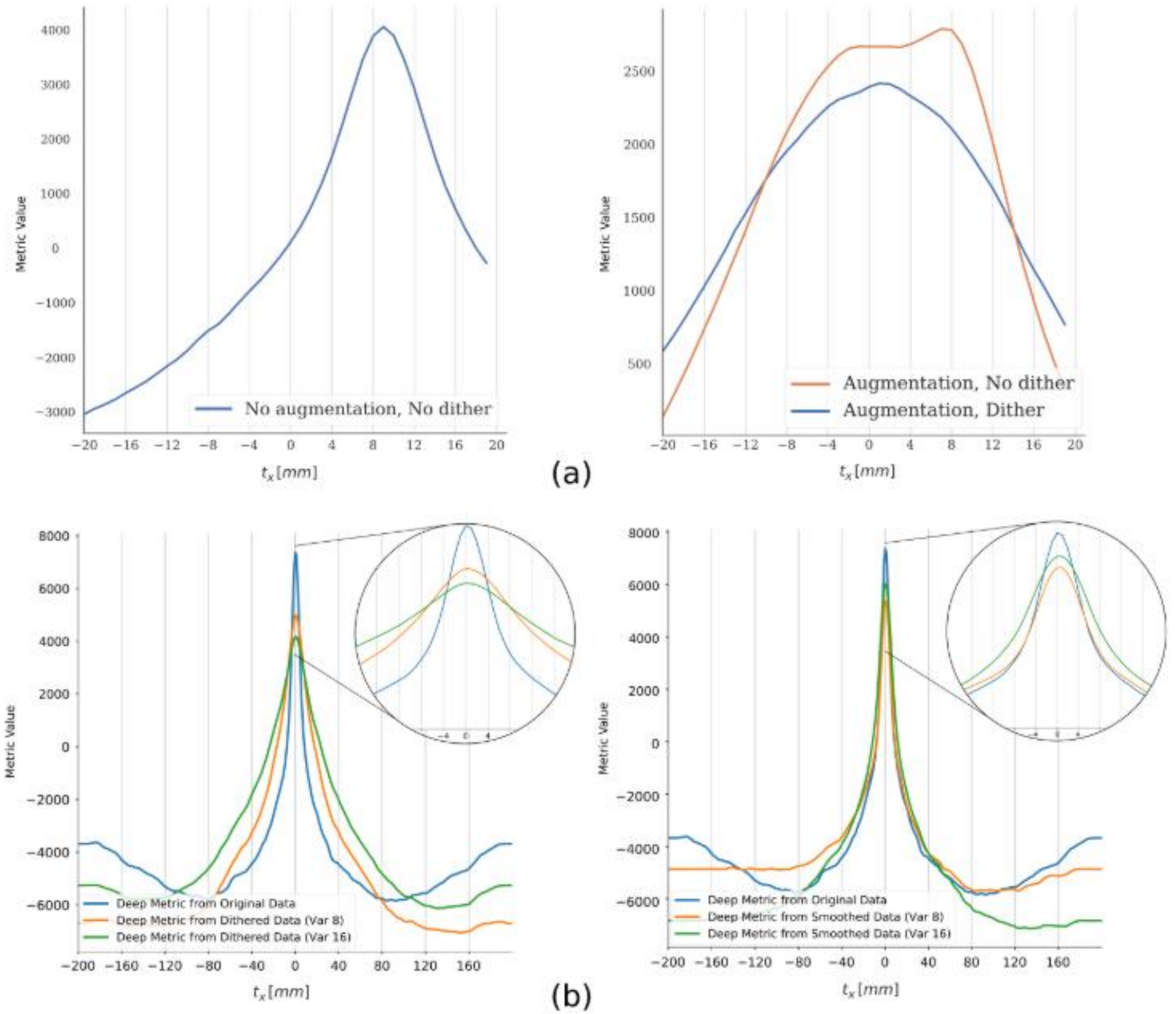
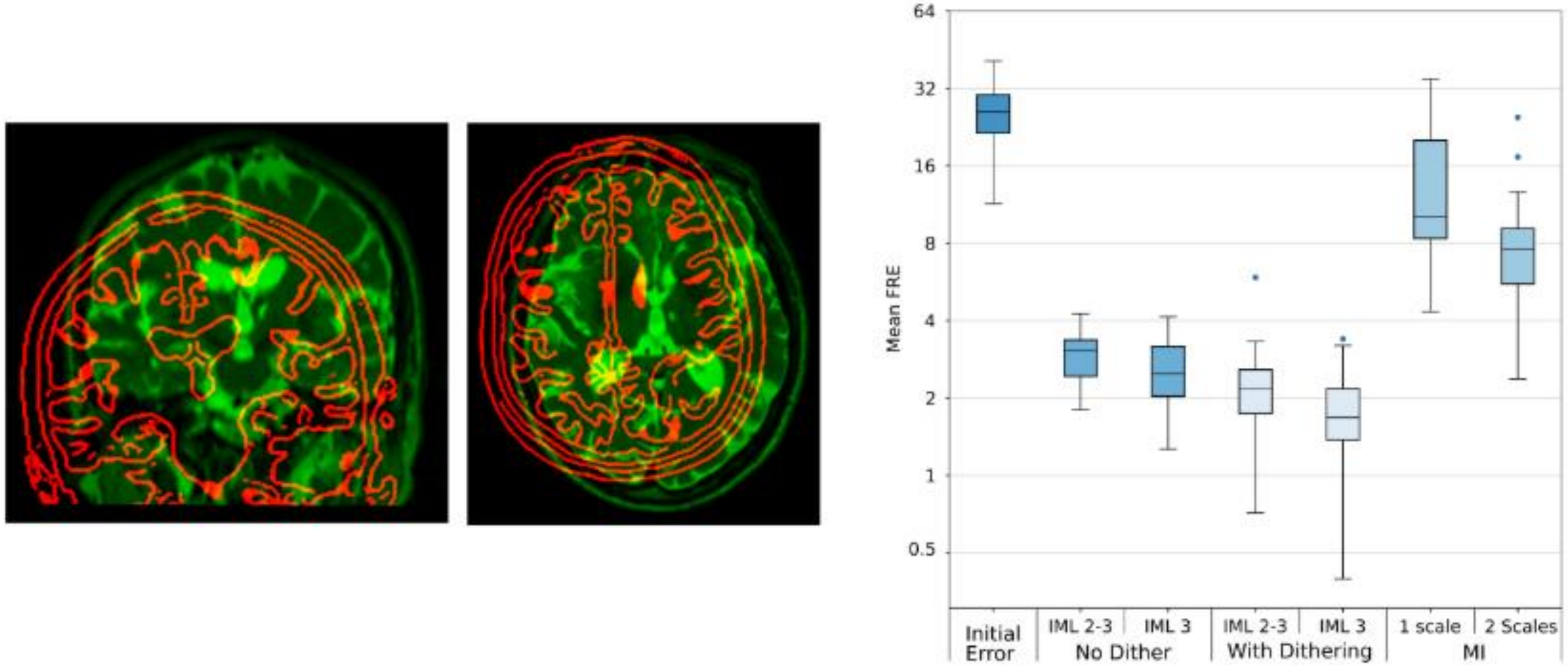


Fig. 3. (a) Characteristics of the deep metric learned from different training data. (b) Impact of dithering (left) and smoothing (right) on the deep metric response function.



(a) sample training patient (b) rigid registration

Fig. 4. (a) An example case from the edge-to-image registration experiment (b) mean FRE achieved by different methods.

Conclusions

1) An information theoretical (IT) foundation for iterated maximum likelihood (IML) registration with deep image metrics, DITR.
2) Focused on the analysis of registration objective functions rather than transformation modeling and optimization methods