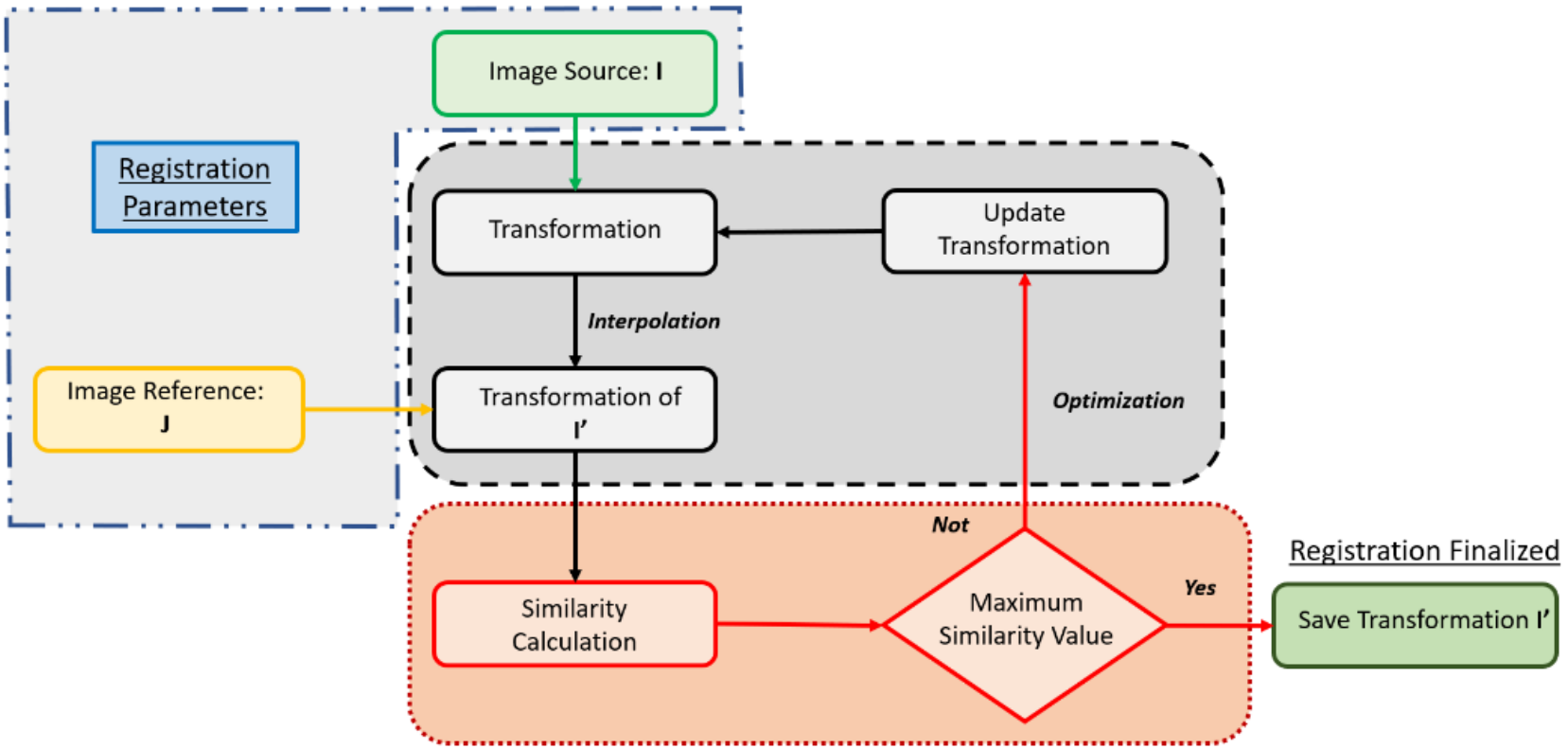
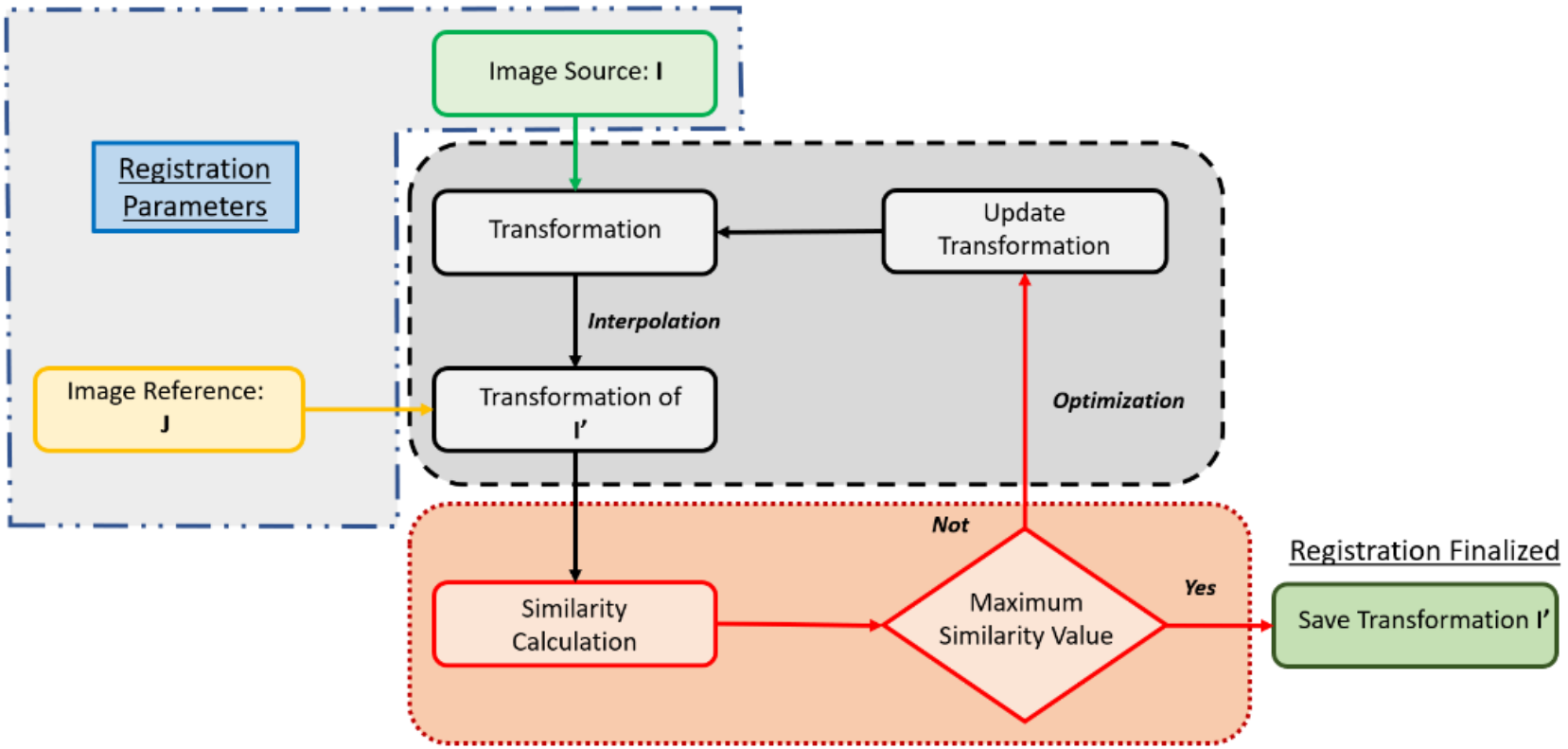
Deep Learning Approches

- Two categories:
  - Similarity Metrics
  - Transformation Parameters



Deep Learning Approches – Part 1

- Two categories:
  - **Similarity Metrics**
  - Transformation Parameters



	Algorithm	Neural Network	Type of Learning	Training Set	Optimization	Application	Dimension
Estimation of Similarity or Dissimilarity Metric	Cheng et al. [49]	AE and DNN	Supervised	4000 patches pairs	Gradient-Based	CT, MRI	2D
	Simonovsky et al. [11]	CNN	Supervised	Data augmentation and IXI	SGD	MRI	3D
	Sedgi et al. [52]	CNN	Semi-Supervised	1 million patches	Adam	Brain MRI	3D
	Grand Haskins et al. [51]	CNN	Supervised	670 images pairs	Adam	MR-TRUS	3D
	Wu et al. [26]	CAE	Unsupervised	7000 patches from 40 images	Gradient-Based	MRI	3D
Prediction of Transformation Parameters	Miao et al. [54]	CNN Regressors	Supervised	25000 pairs of synthetic images	SGD	CT	2D/3D
	Yang et al. [55]	CNN Regressors	Unsupervised	140000 patches from 373 images	SGD	Brain MRI	3D
	Bob D. de Vos et al. [56]	CNN using STN	Unsupervised	69540 pairs of images	Adam	Cardiac MRI	2D
	Mahapatra et al. [3]	GAN	Supervised	39000 pairs of images	Adam and Batch Normalization	Cardiac MRI, Retinal Image (FA)	2D
	Eppenhof et al. [57]	VGG based	Supervised	Synthetic, based on 7 pairs of images	SGD	Thoracic CT	3D
	Sheikhjafari et al. [58]	Fully Connected Generative NN	Unsupervised	30000 images from 100 cine MR sequences	Backpropagation using SGD	Cardiac MR	2D

Deep Learning as Similarity Metrics

(Cheng et al. 2015)

Deep Similarity Learning for Multimodal Medical Images

Xi Cheng, Li Zhang, and Yefeng Zheng

Siemens Corporation, Corporate Technology, Princeton, NJ, USA

Abstract. An effective similarity measure for multi-modal images is crucial for medical image fusion in many clinical applications. The underlining correlation across modalities is usually too complex to be modelled by intensity-based statistical metrics. Therefore, approaches of learning a similarity metric are proposed in recent years. In this work, we propose a novel deep similarity learning method that trains a binary classifier to learn the correspondence of two image patches. The classification output is transformed to a continuous probability value, then used as the similarity score. Moreover, we propose to utilize multi-modal stacked denoising autoencoder to effectively pre-train the deep neural network. We train and test the proposed metric using sampled corresponding/non-corresponding computed tomography (CT) and magnetic resonance (MR) head image patches from a same subject. Comparison is made with two commonly used metrics: normalized mutual information (NMI) and local cross correlation (LCC). The contributions of the multi-modal stacked denoising autoencoder and the deep structure of the neural network are also evaluated. Both the quantitative and qualitative results from the similarity ranking experiments show the advantage of the proposed metric for a highly accurate and robust similarity measure.

Proposal

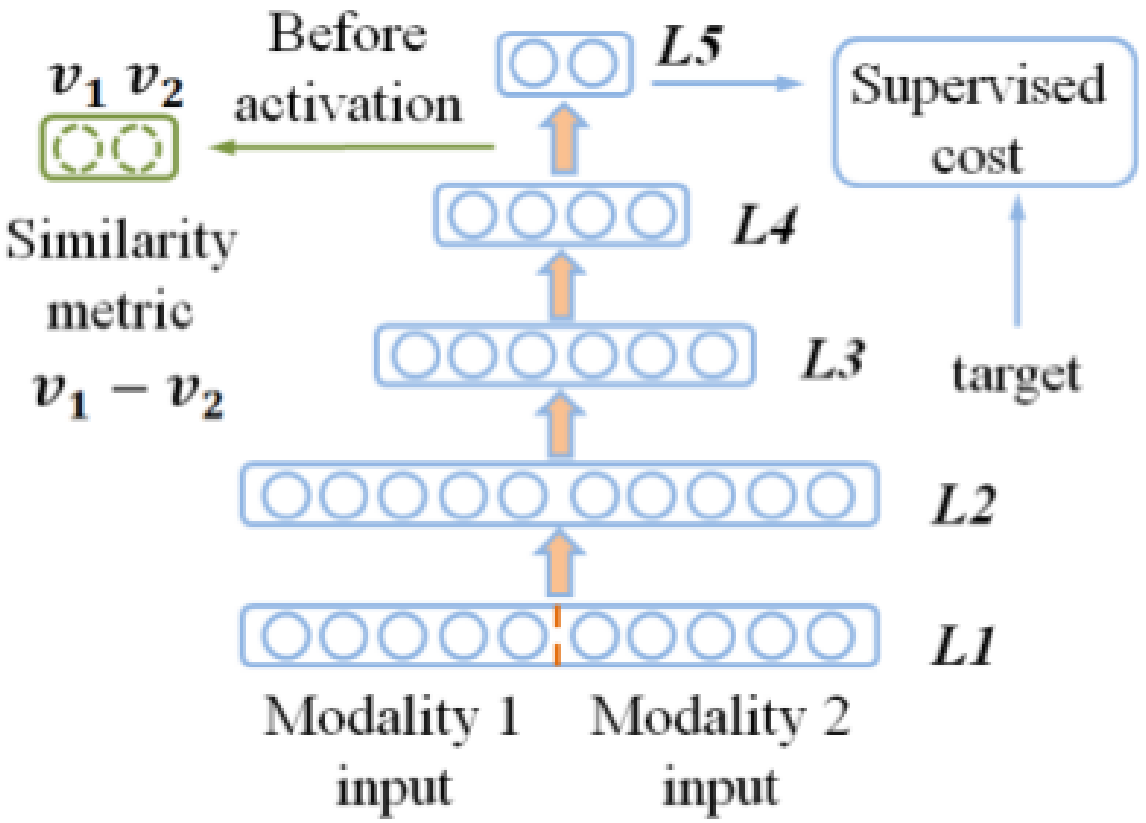


Fig. 1. The structure of a 5-layer DNN. While the 2-unit output is used for supervised training, their values before the activation, i.e., v_1, v_2 are used to form the proposed similarity metric.

Nodes

2

100

300

578

578 (17 x 17 x 2)

Modalities:
computed tomography (CT) and
magnetic resonance (MR)

Proposal

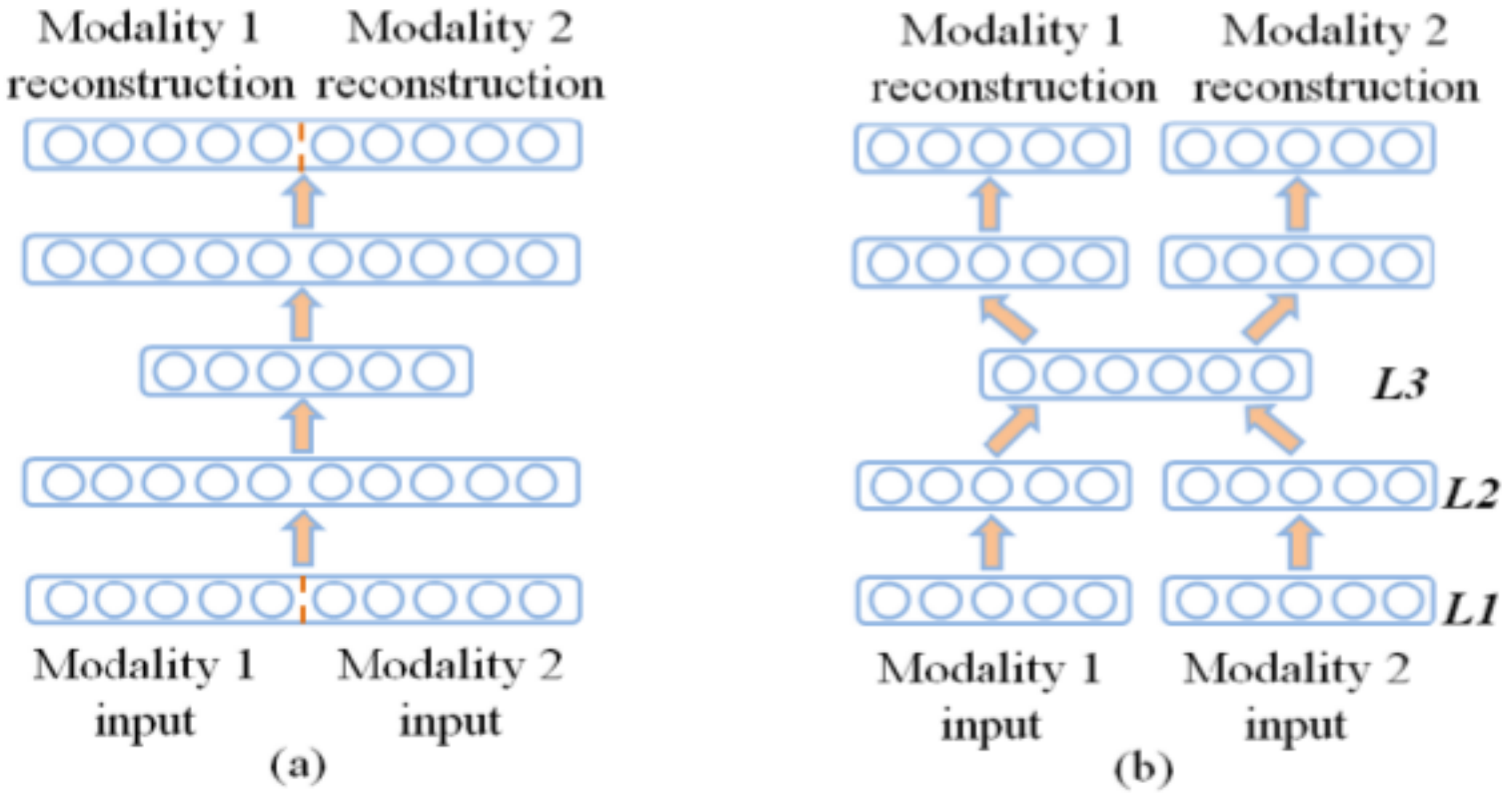


Fig. 2. (a) Uni-modal SDAE with two concatenated images as input; (b) Multi-modal SDAE. Both of them can be used for pre-training layers $L_1 \sim L_3$ of the DNN.

Filter Visualization

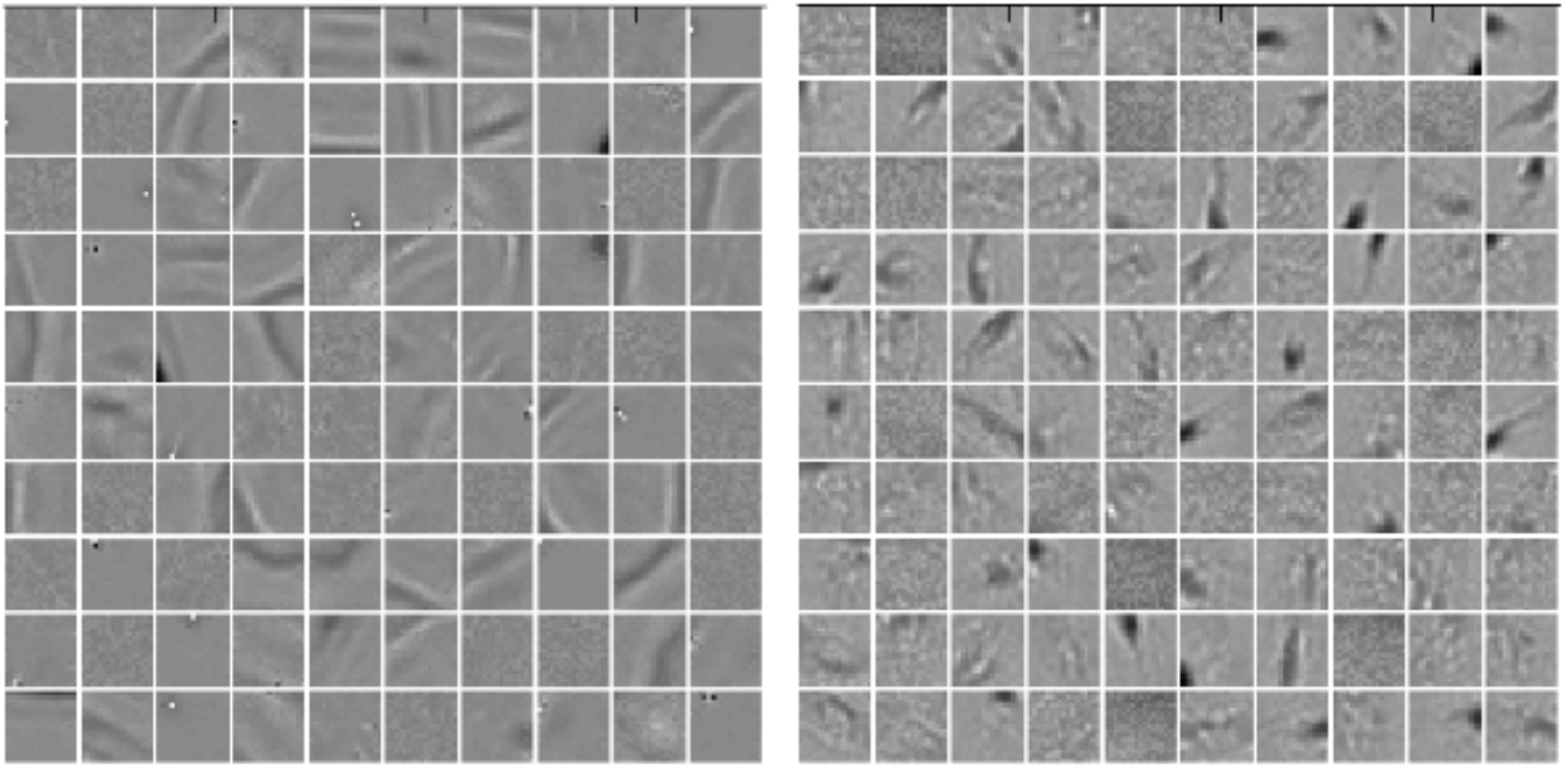


Fig. 3. Filter visualization. First 100 (out of 289) learned filters (17×17 in size) from multi-modal SDAE pre-training for CT image patches (left) and MR image patches (right) using the positive training data.

Results

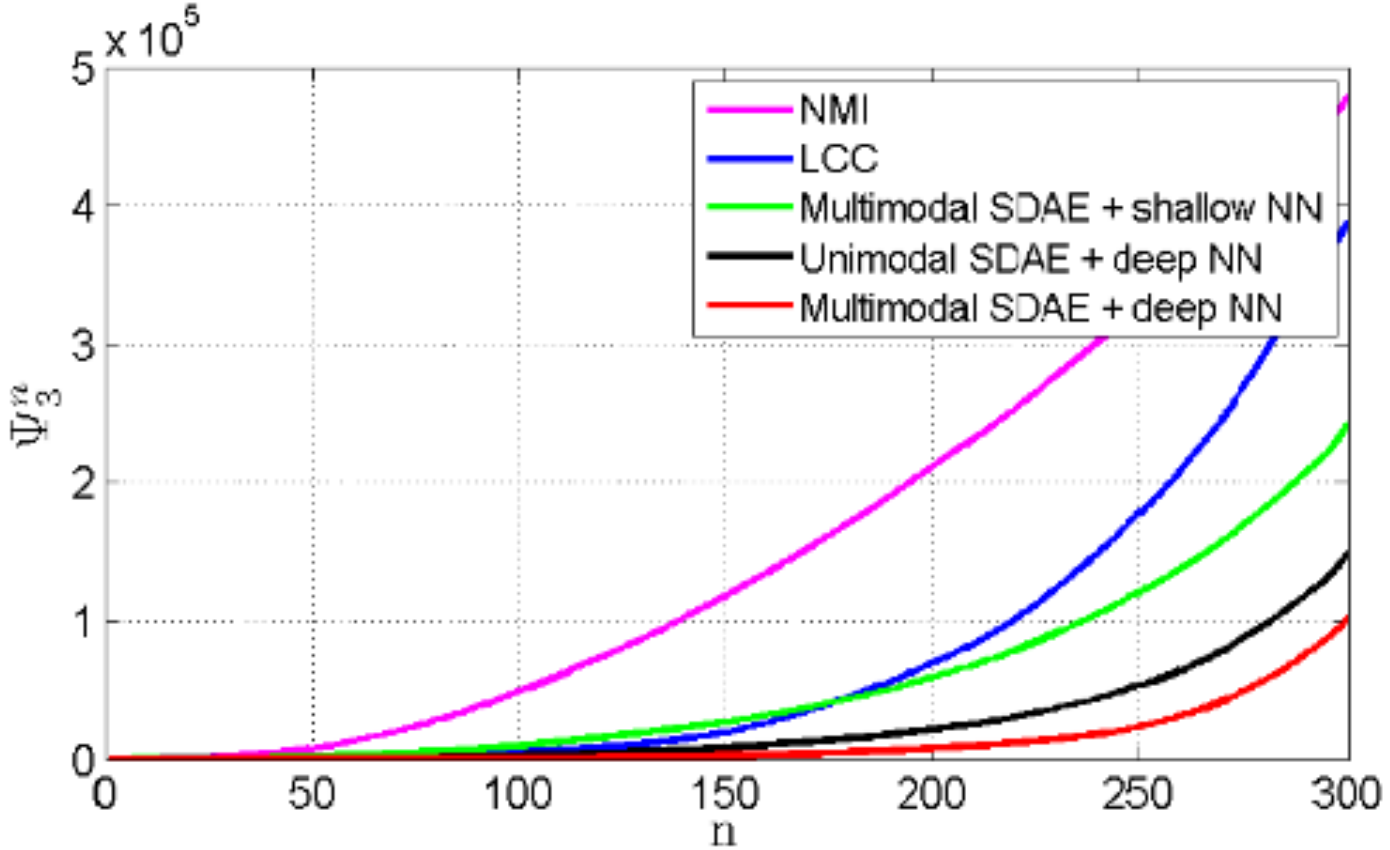
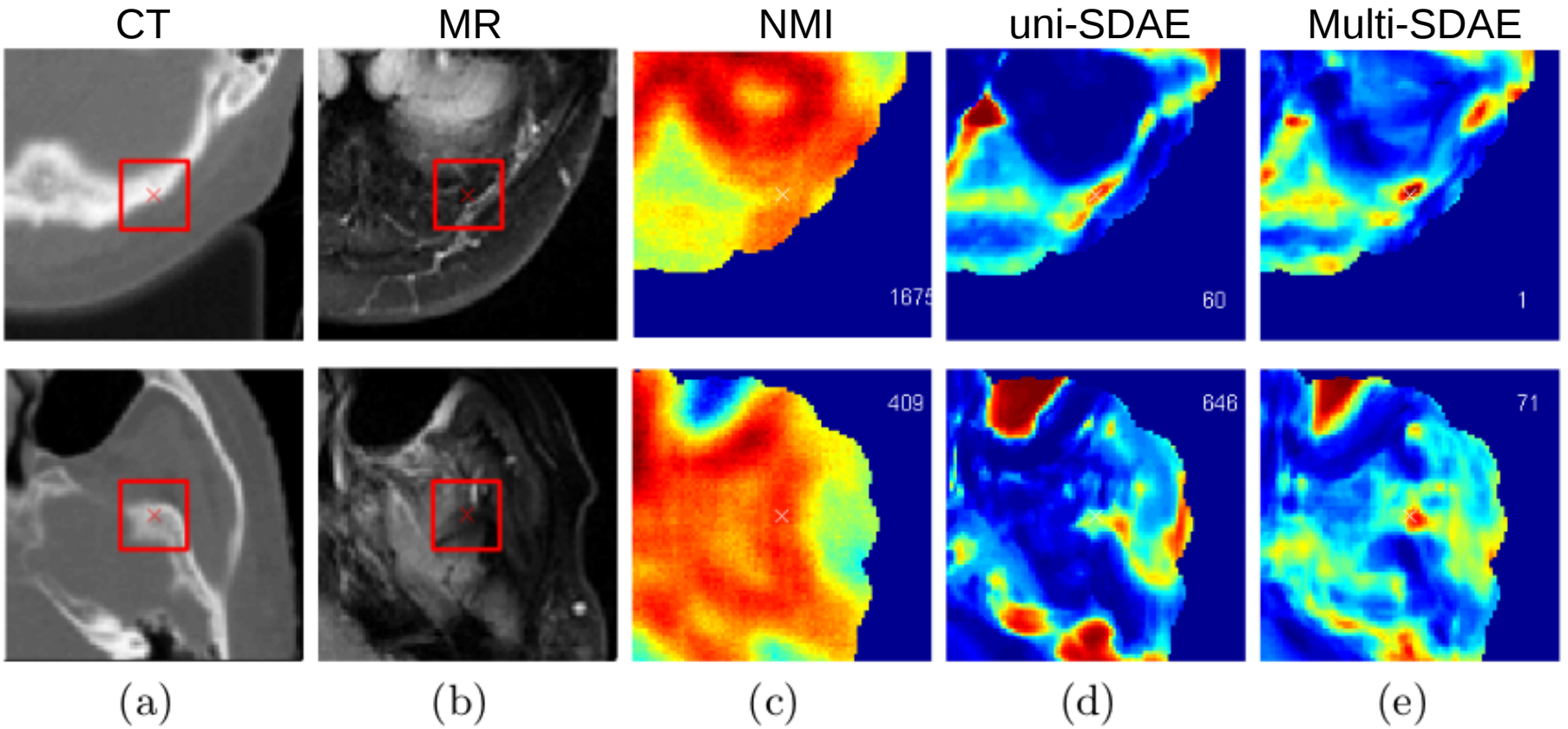


Fig. 4. Comparison of 5 similarity metrics on Ψ_3^n. The cumulative sum of prediction errors is for $n \leq 300$ CT image patches. The worst ψ_3 for a 81×81 neighborhood is 6560, therefore 5×10^5 on the y axis corresponds to a prediction error of about 25% for the 300 patches.

NMI: Normalized Mutual Information
LCC: Local Cross Correlation

Results



- 81x81 CT and MR patches;
- 17x17 region of interest (ROI);
- The high values (red color) in figure (c), (d) and (e) indicate the higher probability to be the corresponding region (ROI);

Deep Learning as Similarity Metrics (Wu et al. 2016)

Scalable High Performance Image Registration Framework by Unsupervised Deep Feature Representations Learning

Guorong Wu, *Member, IEEE*, Minjeong Kim, Qian Wang, Brent C. Munsell, Dinggang Shen†, *Senior Member, IEEE, and for the Alzheimer's Disease Neuroimaging Initiative**

***Abstract*—Feature selection is a critical step in deformable image registration. In particular, selecting the most discriminative features that accurately and concisely describe complex morphological patterns in image patches improves correspondence detection, which in turn improves image registration accuracy. Furthermore, since more and more imaging modalities are being invented to better identify morphological changes in medical imaging data,, the development of deformable image registration method that scales well to new image modalities or new image applications with little to no human intervention would have a significant impact on the medical image analysis community. To address these concerns, a learning-based image registration framework is proposed that uses deep learning to discover compact and highly discriminative features upon observed imaging data. Specifically, the proposed feature selection method uses a convolutional stacked auto-encoder to identify intrinsic deep feature representations in image patches. Since deep learning is an unsupervised learning method, no ground truth label knowledge is required. This makes the proposed feature selection method more flexible to new imaging modalities since feature representations can be directly learned from the observed imaging data in a very short amount of time. Using the LONI and ADNI imaging datasets, image registration performance was compared to two existing state-of-the-art deformable image registration methods that use handcrafted features. To demonstrate the scalability of the proposed image registration framework, image registration experiments were conducted on 7.0-tesla brain MR images. In all experiments, the results showed the new image registration framework consistently demonstrated more accurate registration results when compared to state-of-the-art.**

Submitted to IEEE Transaction on Biomedical Engineering on March 14, 2015. † Corresponding author.

G. Wu, M. Kim, and D. Shen are with the Department of Radiology and BRIC, the University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA. D. Shen is also with Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, Republic of Korea. (e-mail: {grwu, mjkim, dgshen}@med.unc.edu).

Q. Wang is with the Med-X Research Institute of Shanghai Jiao Tong University, Shanghai, 200237, China. (e-mail: wangqian@sjtu.edu.cn).

B. Munsell is with the Department of Computer Science, the College of Charleston, Charleston, SC 29424, USA. (e-mail: munsellb@cofc.edu)

* Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

***Index Terms*—Deformable image registration, deep learning, hierarchical feature representation.**

I. Introduction

DEFORMABLE image registration is very important to neuroscience and clinical studies for normalizing individual subjects to the reference space [1-5]. In deformable image registration, it is critical to establish accurate anatomical correspondences between two medical images. Typically, a patch-based correspondence detection approach is often used, where a patch is a fixed-size symmetric neighborhood of pixel intensity values centered a point in the image. And if two different patches, from two different images, show similar morphological patterns, the two points (at each patch center) are considered to be well corresponded. Therefore, to improve correspondence detection, the problem becomes the one related to feature selection, i.e., how to *consistently* select a set of highly discriminative features that can *accurately*, and *concisely*, capture the morphological pattern presented in the image patch.

Intensity-based feature selection methods are widely used in medical image registration [6-11], however, two image patches that show similar, or even the same, distribution of intensity values do not guarantee the two points are corresponded from an anatomical point of view [4, 12-14]. Handcrafted features, such as geometric moment invariants [15] or Gabor filters [16], are also widely used by many state-of-the-art image registration methods [4, 11, 13, 14, 17]. In general, the major pitfall of using handcrafted features is that the developed model tends to be very ad-hoc. That is, the model is only intended to recognize image patches specific to an image modality or a certain imaging application [18].

Supervised learning-based methods have been proposed to select the best set of features from a large feature pool that may include plenty of redundant handcrafted features [18-24]. However, for this approach, the ground-truth data with known correspondences across the set of training images is required. Because human experts are typically needed to generate ground-truth data, it is well known that obtaining this type of data can be a very laborious and subjective process. In many cases, ground-truth data is simply not available, and even if it does exist, the size of the training population may be very

Proposal

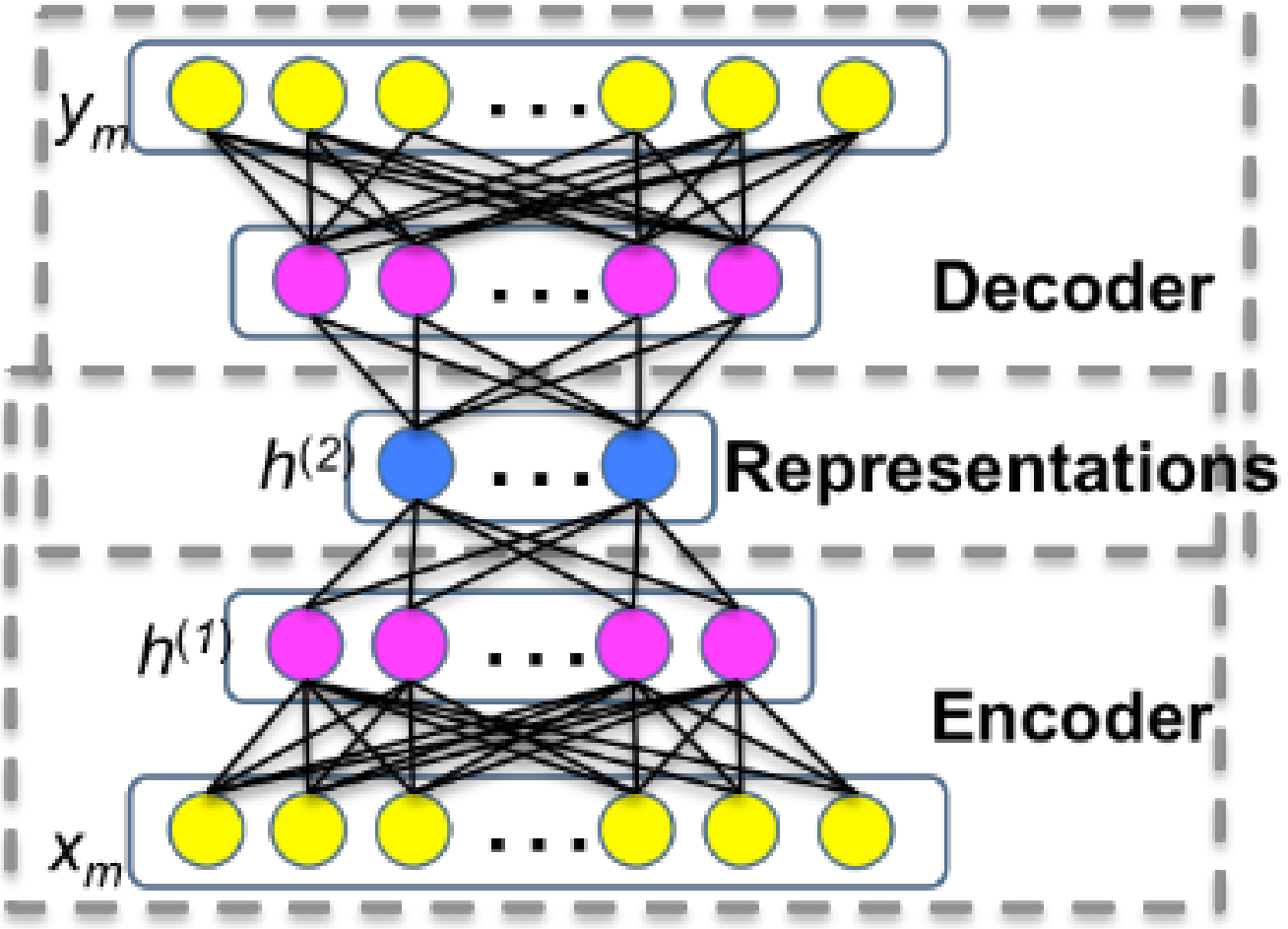


Fig. 2. The hierarchical architecture of Stacked Auto-Encoder (SAE).

Proposal

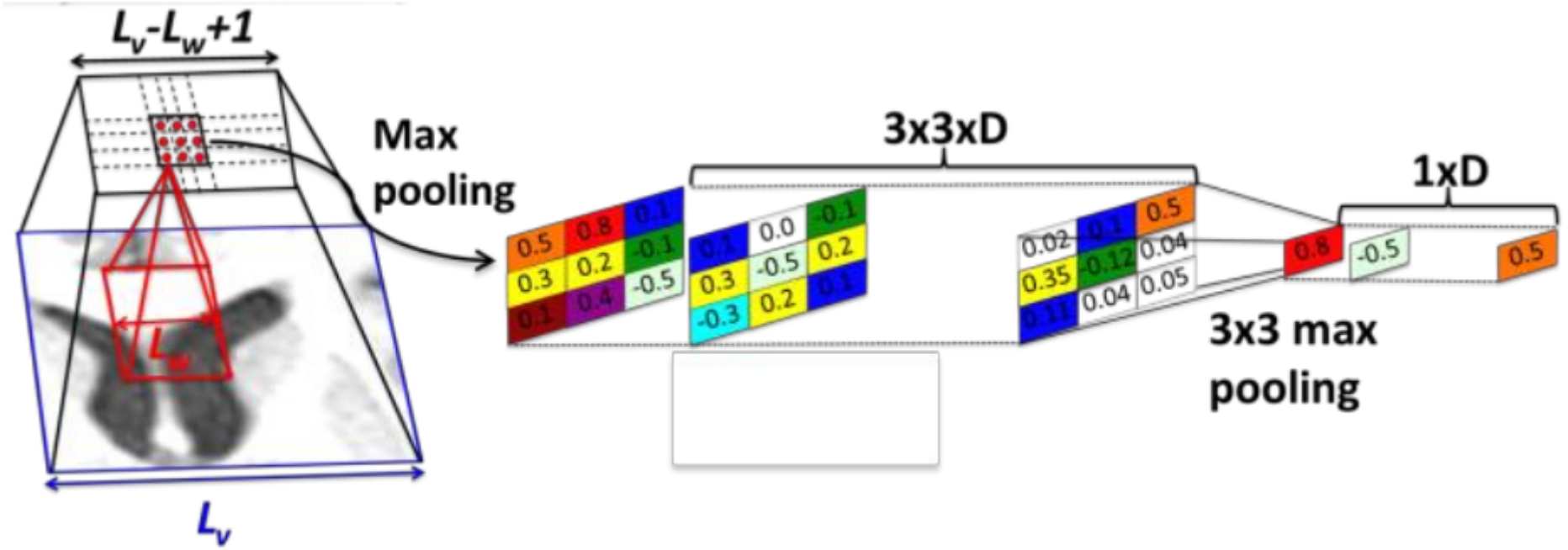
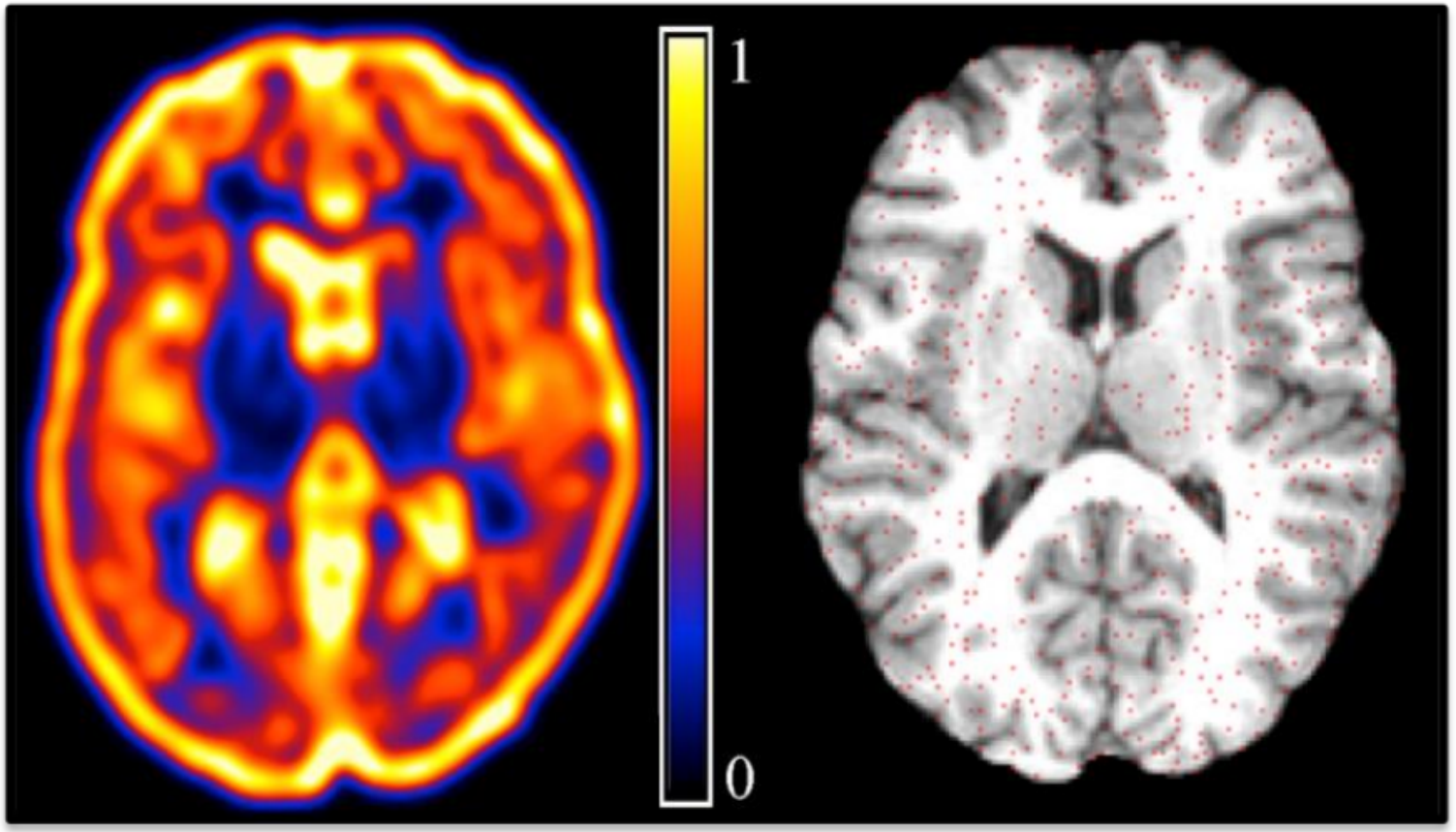


Fig. 3. The 3×3 max pooling procedure in convolutional network.

They used their SAE proposal to work as Multi-channel Demons (intensity-based) And Hammer (feature-based) methods.

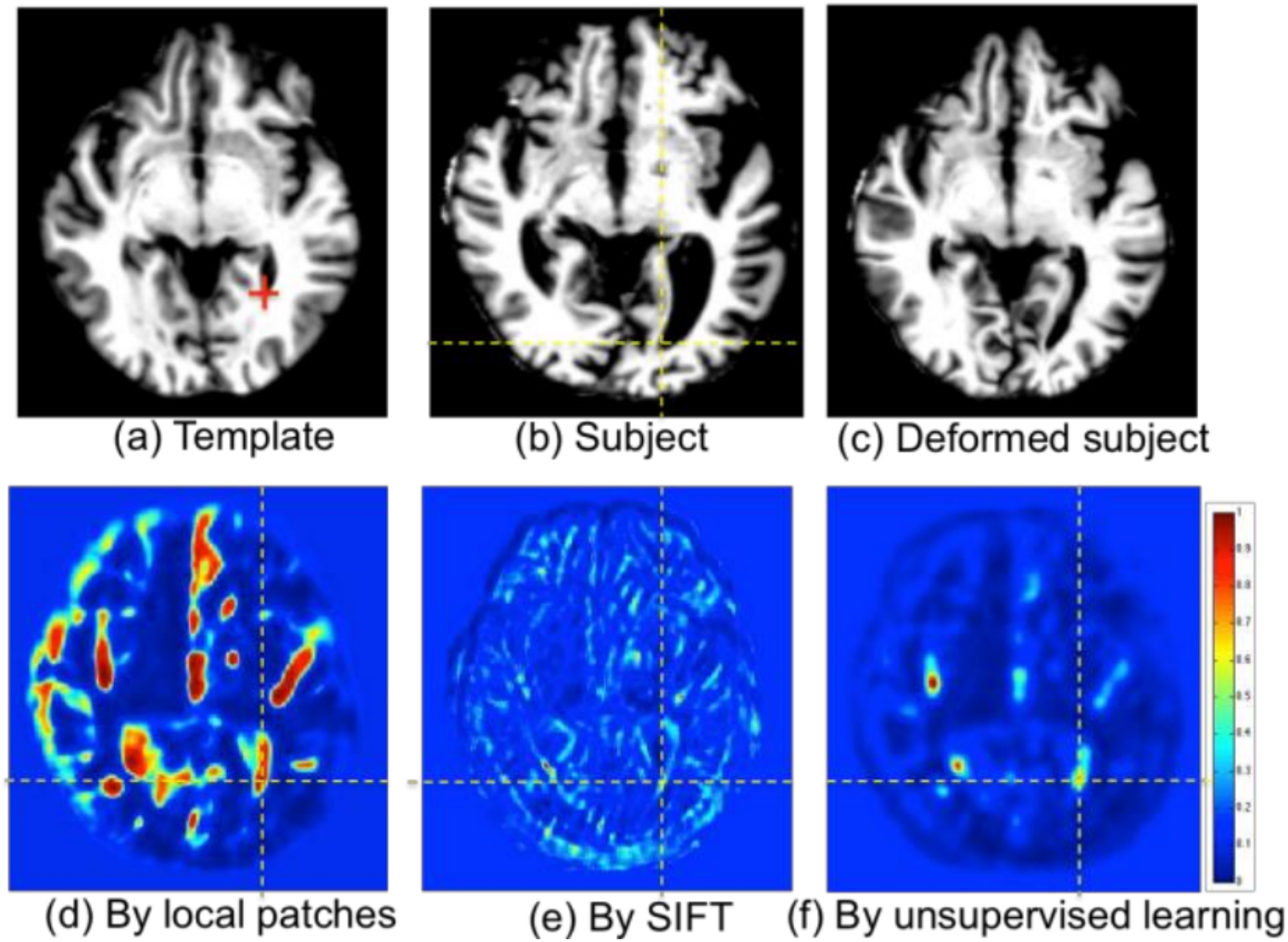
Results – Sample the image patches



(a) Importance map (b) Sampled points

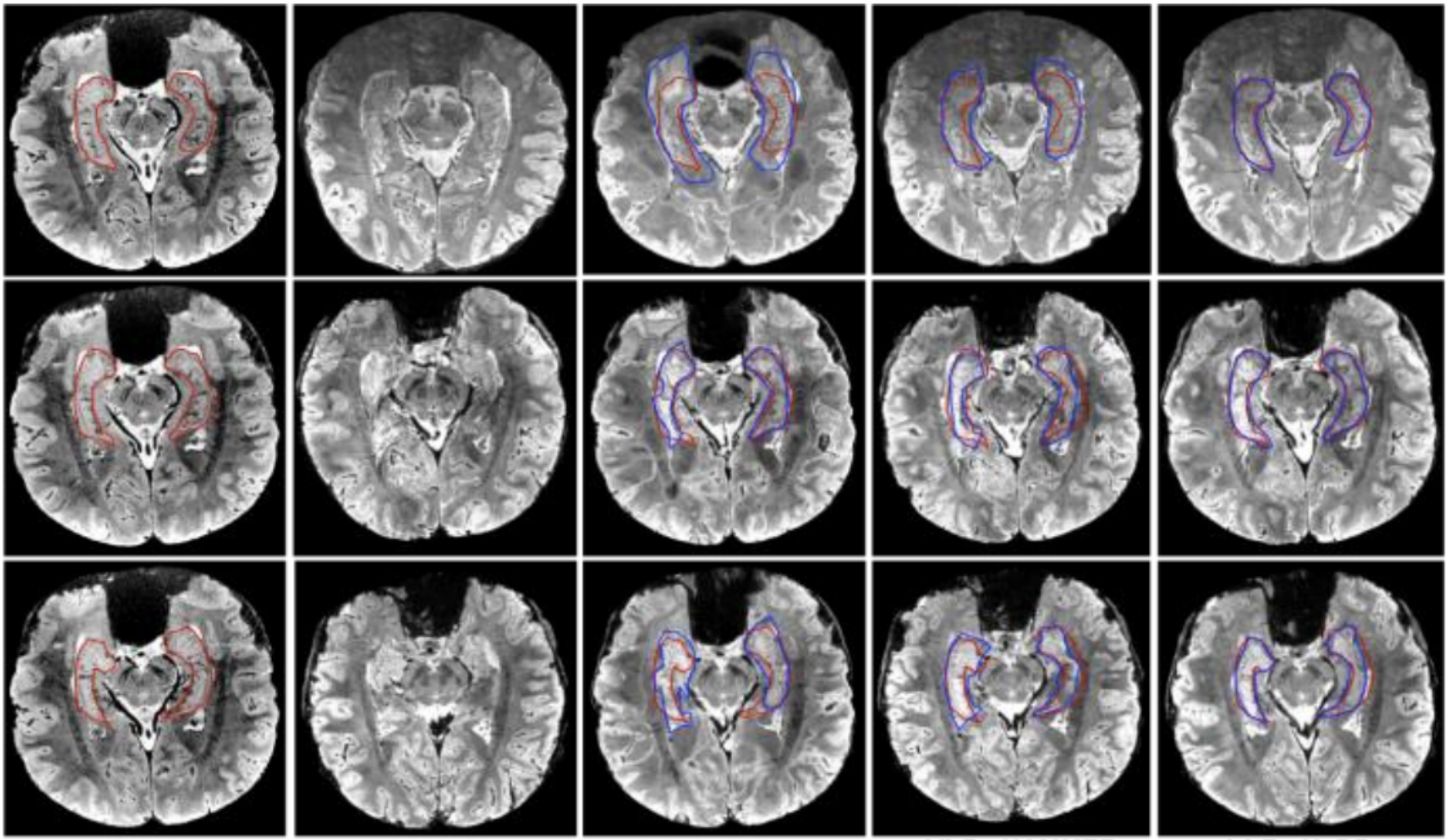
The sampled image patches (with the center point of each sampled patch denoted by the red dot in Fig. (b)) are more concentrated at the context-rich (or edge-rich) Regions, where the values of importance (or probability) are high.

Results



The similarity maps of identifying the correspondence for the red-crossed point in the templa
(a) w.r.t. the subject (b) by handcraft features (d-e) and
the learned features by unsupervised deep learning (f). The
registered subject image is shown in (c). It is clear that the in-accurate
registration results might undermine the supervised feature representation

Results



Typical registration results on 7.0-tesla MR brain images by Demons, HAMMER, and H+DP, respectively. Three rows represent three different slices in the template, subject, and registered subjects.

Results

TABLE I
THE DICE RATIOS OF WM, GM, AND VN ON ADNI DATASET (UNIT: %)

Method	WM	GM	VN	Overall
Demons	85.7	76.0	90.2	84.0
M+PCA	85.5	76.6	90.2	84.1
M+DP	85.8	76.5	90.9	84.4
HAMMER	85.4	75.5	91.5	84.1
H+PCA	86.5	76.9	91.7	85.0
H+DP	**88.1***	**78.6***	**93.0***	**86.6**

White Matter (WM)
Gray Matter (GM)
Cerebral-Spinal Fluid (CSF)
Ventricle (VN) from the CSF

Dice ratio is defined as:

$$D(R_A, R_B) = \frac{2|R_A \cap R_B|}{|R_A| + |R_B|},$$

where R A and R B denote two ROIs (Regions of Interest) and |·| stands for the volume of the region.

Deep Learning as Similarity Metrics

(Simonovsky et al. 2016)

A Deep Metric for Multimodal Registration

Martin Simonovsky[1], Benjamín Gutiérrez-Becker[2] Diana Mateus[2], Nassir Navab[2], and Nikos Komodakis[1]

[1] Imagine, Université Paris Est / École des Ponts ParisTech, France
{martin.simonovsky,nikos.komodakis}@enpc.fr
[2] Computer Aided Medical Procedures, Technische Universität München, Germany
gutierrez.becker@tum.de,{mateus,navab}@in.tum.de

Abstract. Multimodal registration is a challenging problem in medical imaging due the high variability of tissue appearance under different imaging modalities. The crucial component here is the choice of the right similarity measure. We make a step towards a general learning-based solution that can be adapted to specific situations and present a metric based on a convolutional neural network. Our network can be trained from scratch even from a few aligned image pairs. The metric is validated on intersubject deformable registration on a dataset different from the one used for training, demonstrating good generalization. In this task, we outperform mutual information by a significant margin.

Proposal

The problem is often solved by minimizing the energy

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

$\Omega_f \mapsto \mathbb{R}^d$ between a *fixed image* $I_f : \Omega_f \subset \mathbb{R}^d \mapsto \mathbb{R}$

$I_m : \Omega_m \subset \mathbb{R}^d \mapsto \mathbb{R}.$ $I'_m = I_m(\mathcal{T}(\theta)) : \Omega_f \subset \mathbb{R}^d \mapsto \mathbb{R}.$

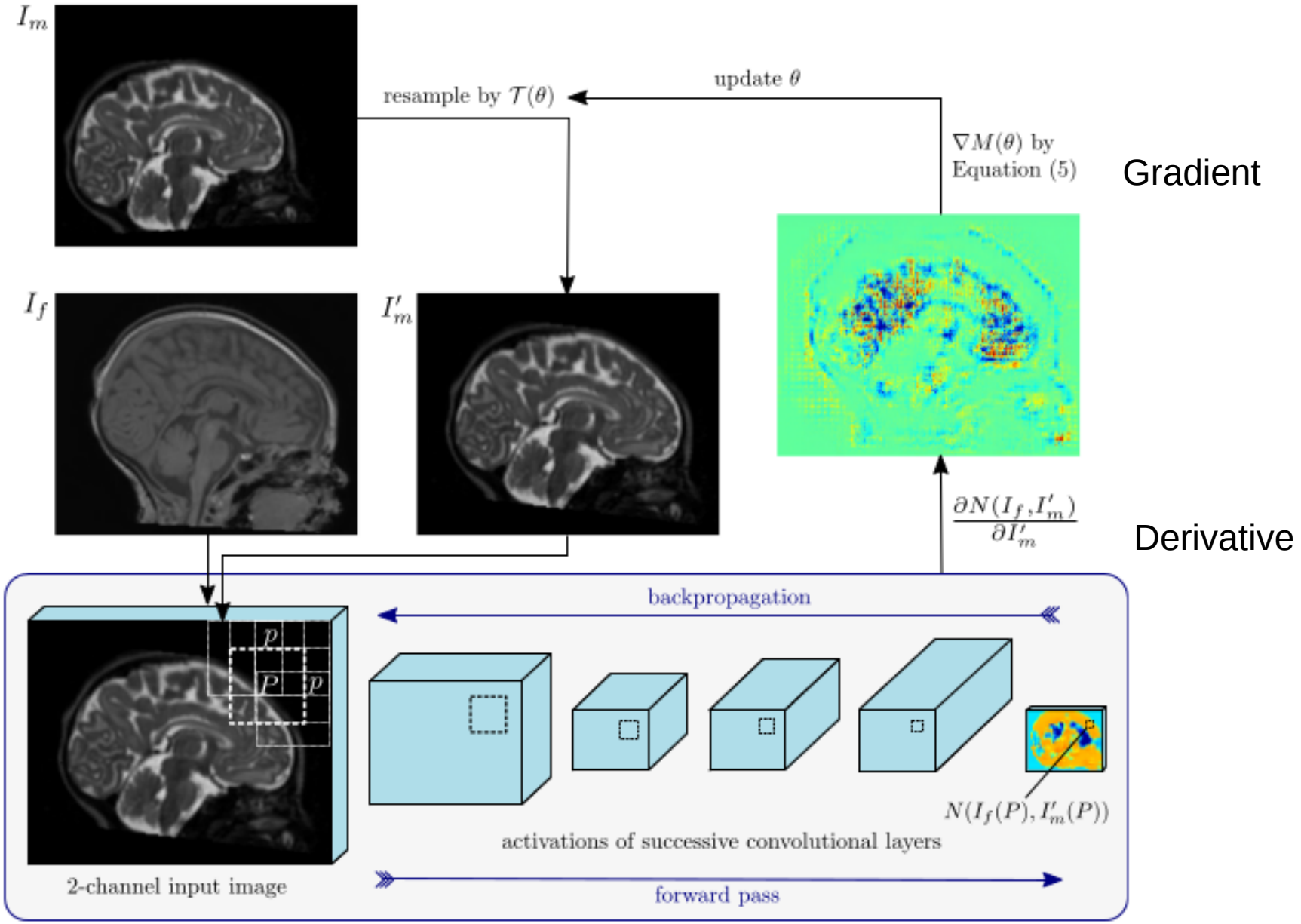
Proposal

Continuous optimization methods iteratively update parameters θ based on the gradient of the objective function $E(\theta)$. We restrict ourselves to first-order methods and use gradient descent in particular. Our metric is defined to aggregate local patch comparisons as

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

where $\mathcal{P}$ is the set of patch domains $P \subset \Omega_f$ sampled on a dense uniform grid with significant overlaps. The method is illustrated in Figure 1

Proposal


I_m
resample by $\mathcal{T}(\theta)$
update θ
$\nabla M(\theta)$ by Equation (5)
Gradient
I_f
I'_m
$\frac{\partial N(I_f, I'_m)}{\partial I'_m}$
Derivative
backpropagation
p
P
p
2-channel input image
$N(I_f(P), I'_m(P))$
activations of successive convolutional layers
forward pass

Results

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

Results

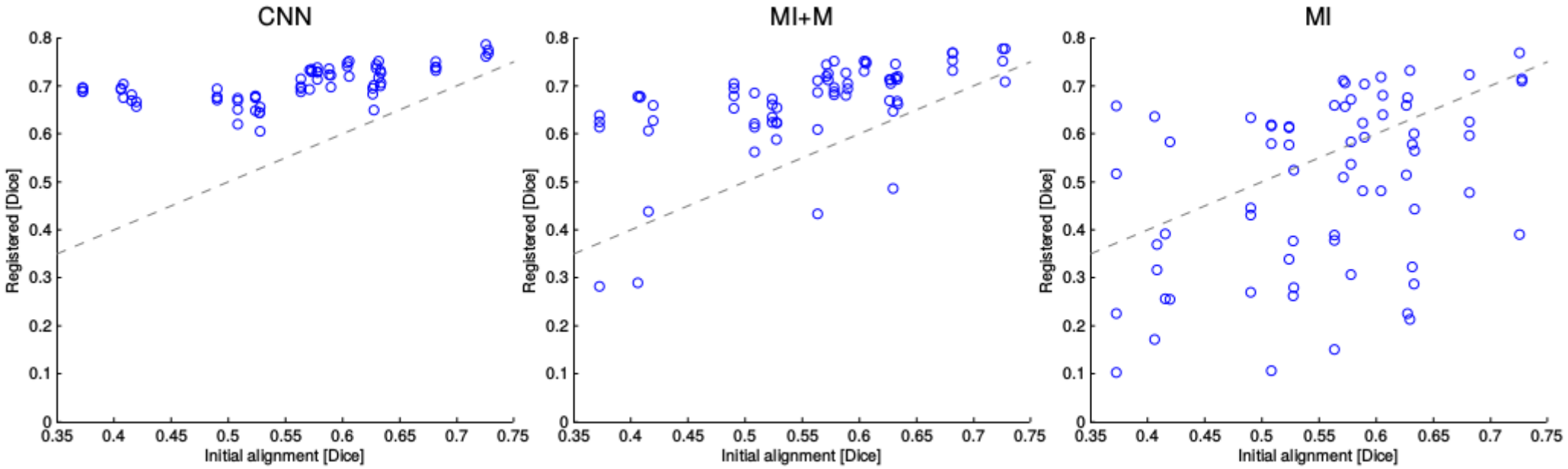


Fig. 2. Improvement in average Dice score due to registration using the proposed metric (CNN) and mutual information with (MI+M) or without masking (MI). Each data point represents a registration run. Dashed line denotes identity transformation.

Deep Learning as Similarity Metrics

(Grand Haskins et al. 2018)

IJCARS manuscript No.
(will be inserted by the editor)

Learning Deep Similarity Metric for 3D MR-TRUS Registration

Grant Haskins · Jochen Kruecker · Uwe Kruger · Sheng Xu · Peter A. Pinto · Brad J. Wood · Pingkun Yan

Received: date / Accepted: date

Abstract

Purpose The fusion of transrectal ultrasound (TRUS) and magnetic resonance (MR) images for guiding targeted prostate biopsy has significantly improved the biopsy yield of aggressive cancers. A key component of MR-TRUS fusion is image registration. However, it is very challenging to obtain a robust automatic MR-TRUS registration due to the large appearance difference between the two imaging modalities. The work presented in this paper aims to tackle this problem by addressing two challenges: (i) the definition of a suitable similarity metric and (ii) the determination of a suitable optimization strategy.

Proposal

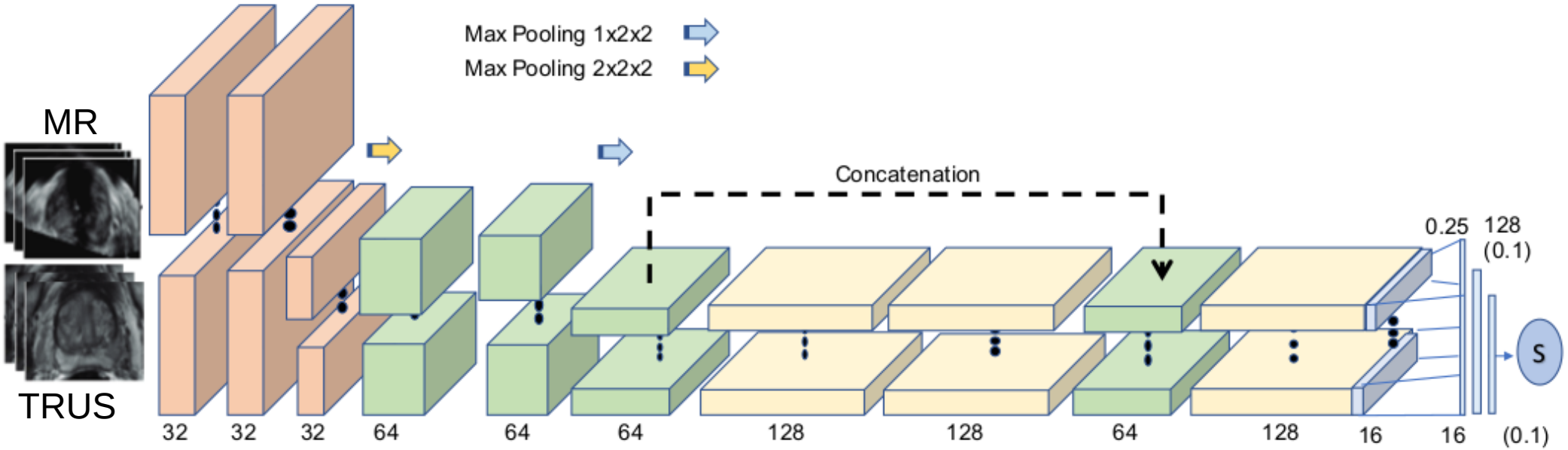


Fig. 1 The architecture of the designed CNN that is used to learn the similarity metric.

- transrectal ultrasound (TRUS)
- Multi-parametric magnetic resonance (MR)

Proposal

Throughout the optimization that is used to perform the registration, the moving image is slightly perturbed N times and the average of the associated TRE estimates is used as the objective function evaluation, defined as

$$E(I_{moving}, I_{fixed}) = \frac{1}{N} \sum_{n=1}^{N} CNN(g(I_{moving}, \theta_n), I_{fixed}),$$

where g(·) is a resampling function to resample the moving image by using the giving parameter θ n .

$$\theta_R = \{T_x, T_y, T_z, r_x, r_y, r_z\}$$

Three **Traslations** and three **rotations** parameters;

Target Registration Error (TRE) between 3D MR and TRUS images;

They proposed a differential evolution initialized Newton-based optimization (DINO) method;

Results

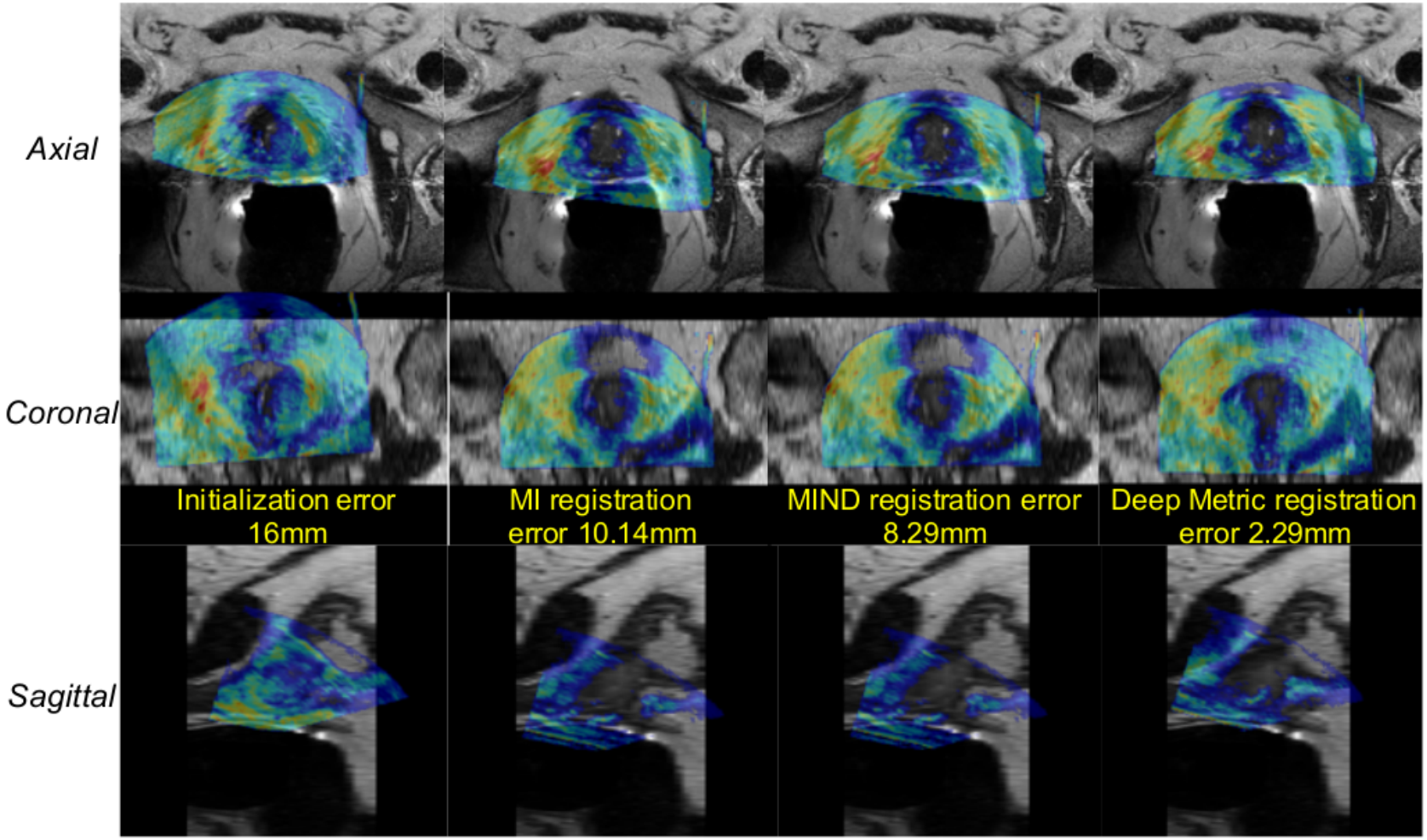


Fig. 4 Axial, coronal, and sagittal views of example registration results: (a) Initial alignment; (b) Registration performed by optimizing the mutual information (MI) using DINO; (c) Registration performed by optimizing the MIND similarity using DINO; (d) Registration performed by optimizing the learned metric using DINO.

Results

Table 1 Comparison of the TREs obtained using several different similarity metrics and optimization strategies. Note that (sp) refers to the single pass approach and (mp) refers to the multi-pass approach. All the numbers are in millimeter (mm).

Similarity Metric	Optimizer	Initial	Final mean±std	[min, max]
Mutual Information	DINO	8mm	8.96 ± 1.28	[5.50, 13.45]
SSD MIND	DINO		6.42 ± 2.86	[1.75, 10.64]
Deep Metric (sp)	DINO		3.97 ± 1.67	[0.77, 8.51]
Deep Metric (mp)	BFGS		7.31 ± 0.61	[6.63, 9.11]
Deep Metric (mp)	Powell		6.11 ± 4.62	[1.89, 12.98]
Deep Metric (mp)	DINO		**3.82 ± 1.63**	[0.65, 8.80]
Mutual Information	DINO	16mm	10.07 ± 1.40	[8.82, 14.09]
SSD MIND	DINO		6.62 ± 2.96	[1.58, 13.63]
Deep Metric (sp)	DINO		4.21 ± 1.64	[1.08, 8.68]
Deep Metric (mp)	BFGS		14.27 ± 0.61	[13.11, 15.25]
Deep Metric (mp)	Powell		11.05 ± 5.32	[2.11, 24.09]
Deep Metric (mp)	DINO		**3.94 ± 1.47**	[1.35, 9.37]

Deep Learning Approches – Part 2

- Two categories:
  - Similarity Metrics
  - **Transformation parameters**