Multi-channel Image Registration of Cardiac MR Using Supervised Feature Learning with Convolutional Encoder-Decoder Network

Xuesong Lu[1](✉) and Yuchuan Qiao[2]

[1] College of Biomedical Engineering, South-Central University for Nationalities,
Wuhan 430074, China
xslu-scuec@hotmail.com
[2] Laboratory of Neuro Imaging, Keck School of Medicine of USC,
Los Angeles, CA 90033, USA

Abstract. It is difficult to register the images involving large deformation and intensity inhomogeneity. In this paper, a new multi-channel registration algorithm using modified multi-feature mutual information (α-MI) based on minimal spanning tree (MST) is presented. First, instead of relying on handcrafted features, a convolutional encoder-decoder network is employed to learn the latent feature representation from cardiac MR images. Second, forward computation and backward propagation are performed in a supervised fashion to make the learned features more discriminative. Finally, local features containing appearance information is extracted and integrated into α-MI for achieving multi-channel registration. The proposed method has been evaluated on cardiac cine-MRI data from 100 patients. The experimental results show that features learned from deep network are more effective than handcrafted features in guiding intra-subject registration of cardiac MR images.

Keywords: Multi-channel image registration · Multi-feature mutual information · Supervised feature learning · Convolutional encoder-decoder network

1 Introduction

Image registration is an important technique in medical image analysis [1]. Many clinical applications, such as multi-modal image fusion, radiotherapy, and computer-assisted surgery, can benefit from this technique. However, large deformation and intensity inhomogeneity bring great challenges into this procedure. To deal with these problems, the standard metrics like sum of squared difference (SSD), correlation coefficient (CC), and mutual information (MI) are not sufficient for intensity-based registration.

Recently, some studies have focused on multi-channel image registration for these issues. Legg et al. [2] extracted several feature images from the original images, and subsequently incorporated these feature images into a dissimilarity measure based on regional mutual information for multi-modal image registration. Staring et al. [3] adopted *k*-nearest neighbors graph (KNNG) to implement multi-feature mutual

Ž. Špiclin et al. (Eds.): WBIR 2020, LNCS 12120, pp. 103–110, 2020.
https://doi.org/10.1007/978-3-030-50120-4_10

information (α-MI) in order to register cervical MRI data. Rivaz et al. [4] introduced a self-similarity weighted α-MI using local structural information to register multiple feature images. Li et al. [5] developed an objective function that relies on the auto-correlation of local structure (ALOST) into registration of intra-image with signal fluctuations. Guyader et al. [6] proposed to formulate multi-channel registration as a group-wise image registration problem, in which the modality independent neighborhood descriptor (MIND) was used as the feature images.

It is critical for these methods to select discriminative features that can establish accurate anatomical correspondences between two images. Most of multi-channel image registrations utilized handcrafted features, such as multi-scale derivatives or descriptor engineering, to achieve good performance. In general, handcrafted features need manually intensive efforts to design the model for specific task. Learning-based methods have been developed to select the best feature set from a large feature pool, which can be adapted to the data at hand [7]. Moreover, deep learning can automatically and hierarchically learn effective feature representation from the data. Shin et al. [8] applied the stacked auto-encoders to organ identification in MR images. Chmelik et al. [9] classified lytic and sclerotic metastatic lesions in spinal 3D CT images by deep convolutional neural network (CNN). Wu et al. [10] employed a convolutional stacked auto-encoder to identify intrinsic deep feature representations for multi-channel image registration.

In contrast, we propose an end-to-end feature learning method to improve the performance of α-MI based on minimal spanning tree (MST). The convolutional encoder-decoder architecture that combines semantic information from a deep, coarse layer with appearance information from a shallow, fine layer is trained in a supervised fashion. Various latent features can be learned by forward computation and backward propagation. The local feature representation of testing image extracted from the first layer of encoder part is integrated into α-MI metric. The proposed method is evaluated on intra-subject registration of cardiac MR images.

2 Method

2.1 α-MI Implementation Using MST

In the previous work [11], multi-channel registration of two images $I_f(\boldsymbol{x})$ and $I_m(\boldsymbol{x})$ can be formulated as $\hat{\mu} = \arg\min_{\mu} \alpha MI\left(T_\mu; I_f(\boldsymbol{x}), I_m(\boldsymbol{x})\right)$, where T_μ is the free-form deformation (FFD) model based on B-spline. Assume that $\boldsymbol{z}(x_i) = [z_1(x_i) \cdots z_d(x_i)]$ denotes a vector of dimension d containing all feature values at point x_i. Let $\boldsymbol{z}^f(x_i)$ be the feature vector of the fixed image at point x_i, and $\boldsymbol{z}^m\left(T_\mu(x_i)\right)$ be that of the moving image at the transformed point $T_\mu(x_i)$. Let $\boldsymbol{z}^{fm}\left(x_i, T_\mu(x_i)\right)$ be the concatenation of the two feature vectors: $\left[\boldsymbol{z}^f(x_i), \boldsymbol{z}^m\left(T_\mu(x_i)\right)\right]$. Three MST graphs with N samples can be constructed by:

$$L_f = min \sum\nolimits_{ij=1}^{N-1} \left\| z^f(x_i) - z^f\left(x_j\right) \right\|^{\gamma}, \tag{1}$$

$$L_m = min \sum\nolimits_{ij=1}^{N-1} \left\| z^m\left(T_\mu(x_i)\right) - z^m\left(T_\mu\left(x_j\right)\right) \right\|^{\gamma}, \tag{2}$$

$$L_{fm} = min \sum\nolimits_{ij=1}^{N-1} \left\| z^{fm}\left(x_i, T_\mu(x_i)\right) - z^{fm}\left(x_j, T_\mu\left(x_j\right)\right) \right\|^{2\gamma}, \tag{3}$$

where $\|\cdot\|$ is the Euclidean distance, and $\gamma \in (0, d)$. So α-MI based on MST can be expressed as:

$$\alpha\mathrm{MI} = \frac{1}{1-\alpha}\left(\log\frac{L_f}{N^\alpha} + \log\frac{L_m}{N^\alpha} - \log\frac{L_{fm}}{N^\alpha}\right), \tag{4}$$

where $\alpha = (d - \gamma)/d$.

2.2 Network Architecture

The network architecture like 2D U-Net [12] for deep feature learning consists of encoding and decoding branches connected with skip connections. The encoding stage contains padded 3 × 3 convolutions followed by rectified linear unit (ReLU) activation functions. A 2 × 2 maxpooling operation with stride 2 is applied after every two convolutional layers. After each downsampling, the number of feature channels is doubled. In the decoding stage, a 2 × 2 upsampling operation is applied after every two convolutional layers. The resulting feature map is concatenated to the corresponding feature map from the encoding part. After each upsampling, the number of feature channels is halved.

The input size of the encoder-decoder architecture should be divisible by 16, and equal to the output size. At the final layer, a 1 × 1 convolution is used to generate the same depth of feature map as the desired number of classes.

2.3 Feature Representation with Supervised Learning

To train the encoder-decoder network, the input images and their labels are used to optimize the weights of convolutional layers through the softmax classifier. For the class imbalance between the foreground and background, we adopt weighted cross entropy as the loss function:

$$\mathrm{L} = -\sum\nolimits_{x\in\Omega} \omega(x) y(x) log(\hat{y}(x)), \tag{5}$$

where $\mathrm{y}(x)$ is the true label, $\hat{y}(x)$ is the probability estimation by softmax, and $\omega(x)$ is the weight coefficient at the pixel x within domain Ω.

Due to supervised learning, global features containing semantic information are prone to be biased. Here local features containing appearance information are extracted from the first layer of our network for multi-channel registration. Figure 1 shows an example of 64 features from a 2D slice of cardiac MR image. Finally, we embed 65 features (original intensity image, 64 deep features) into α-MI based on MST metric.

Before performing registration, these features are normalized to have zero mean and unit variance. Note that feature extraction is executed in 2D manner, while registration is performed in 3D.

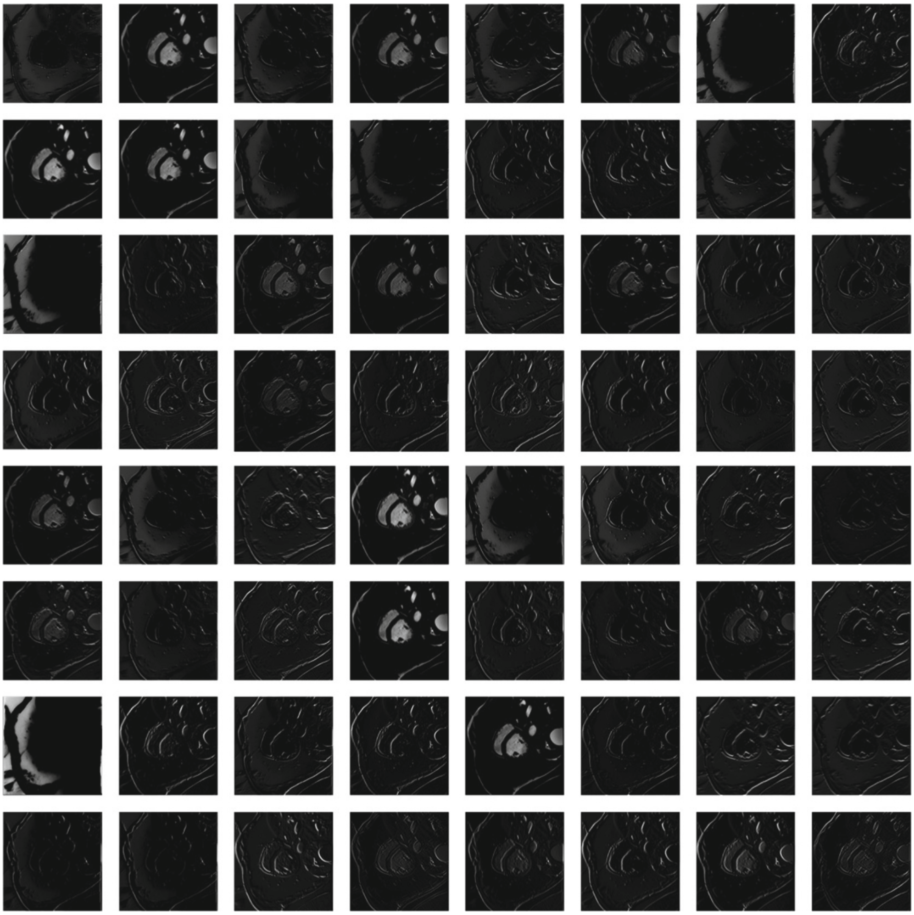

Fig. 1. An example of 64 local feature representations with supervised learning from a 2D slice of cardiac MR image.

3 Experiment and Result

The multi-feature mutual information using MST was implemented in the registration package *elastix* [13] with multi-threaded mode, which is mainly based on the Insight Toolkit. The registration experiments were run on a Windows platform with an Intel Dual Core 3.40 GHz CPU and 32.0 GB memory. A Tensorflow implementation of convolutional encoder-decoder network was trained on a Nvidia GeForce GTX 1070 GPU.

3.1 Dataset and Evaluation Method

To evaluate the performance of the proposed method, our experiments were on cardiac cine-MRI training data of the ACDC challenge [14], which consists of 100 patient scans. The image spacing varies from 0.70 × 0.70 × 5 mm to 1.92 × 1.92 × 10 mm. We resampled the data to an in-plane spacing of 1.37 × 1.37 mm, and then cropped all resampled images to an in-plane size of 224 × 224 pixels. The manual delineation of the left ventricle (LV), the left ventricle myocardium (LVM), and the right ventricle (RV) at the end-diastolic (ED) and end-systolic (ES) phases of each patient is provided as the ground truth for quantitative evaluation.

The data were divided into the training and validation set. The training set comprising 80 subjects was used to train the deep network in a slice-by-slice manner for feature extraction. The validation set with the remaining 20 subjects was performed registration between images at ED and ES. In total 40 different registration results were available for evaluation. The propagated segmentations can be generated by transforming the manual segmentation of the moving image to the fixed image domain, with obtained deformation field.

The Dice Similarity Coefficient (DSC) as a measure of overlap was calculated between propagated segmentation and ground truth of the fixed image. To compare two methods, a value of $p < 0.05$ in two-sided Wilcoxon tests is regarded as a statistically significant difference. The Hausdorff distance (HD) between the surface of propagated segmentation and the surface of ground truth was also used to measure the quality of registration.

3.2 Parameter Settings

The proposed α-MI based on MST using the deep feature representation (in total 65 features, called aMI+SDF) was compared to localized MI (called LMI) [15] and α-MI based on MST with the Cartesian feature set [3] (in total 15 features, called aMI+HCF). Since cardiac MR images only show local deformations between the time phases, initial rigid registration was not necessary.

For weighted cross entropy, we set a weight of 0.3 for the foreground class, and 0.1 for the background class. To train the encoder-decoder network, we used the Adam optimizer, where learning rate 1.0×10^{-3} and 60 epochs with batch size of 4 were set.

For all experiments on intra-subject registration, a multiresolution scheme using Gaussian smoothing was applied. Scales $\sigma = 4.0$, 2.0, and 1.0 voxels in the x and y directions were used. For the z direction, $\sigma = 2.0$, 1.0, and 0.5 voxel was used. As for transformation model, the parameterized B-splines with grid spacing of 20, 10, and 5 mm was employed for three resolution levels respectively.

For LMI, a local region of 50 × 50 × 25 mm was randomly selected. About the parameter optimization, $A = 200$, $\tau = 0.6$, $a = 2000$, and 2000 iterations were set. The number of random samples was set to $N = 2000$. For aMI+HCF and aMI+SDF, $A = 50$, $\tau = 0.602$, $a = 2000$, and 600 iterations were set. The number of random samples was set to $N = 5000$.

In multi-feature mutual information, the *k*D trees, a standard splitting rule, a bucket size of 50, and an errorbound value of 10.0 were selected. The $k = 20$ nearest neighbors were set. In addition, α value was set to 0.99.

3.3 Registration Accuracy

The boxplot of overlap scores using the three methods is shown in Fig. 2. It is clear that registration quality of LMI is the worst. Compared to aMI+HCF, the median overlap of aMI+SDF increases significantly from 0.898 to 0.921 ($p = 2.70 \times 10^{-3}$) for the LV, from 0.781 to 0.822 ($p = 4.57 \times 10^{-6}$) for the LVM, and from 0.775 to 0.813 ($p = 1.92 \times 10^{-5}$). The overall mean and standard deviation of the measures are summarized in Table 1. The same trend can be found in the HD measure. The median HD of aMI+SDF for the LV is as low as 9.171 mm. Figure 3 displays a typical example of registration results. It can be observed that aMI+SDF performs much better than aMI+HCF for these anatomical structures.

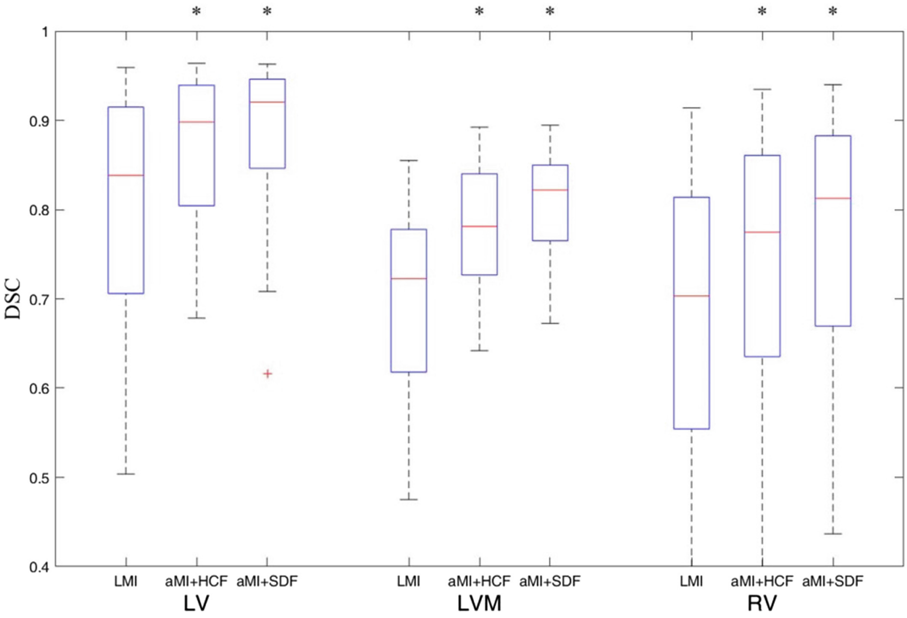

Fig. 2. The boxplot of overlap scores using different methods at different anatomical structures. A star indicates a statistical significant difference of the median overlap compared to the previous column.

Table 1. The mean and standard deviation of quantitative measures using the three methods for different anatomical structures.

Structures	Methods	DSC	HD (mm)
LV	LMI	0.797 ± 0.135	12.567 ± 4.111
	aMI+HCF	0.868 ± 0.085	10.072 ± 3.412
	aMI+SDF	**0.888 ± 0.080**	**9.614 ± 3.348**
LVM	LMI	0.696 ± 0.104	12.243 ± 3.804
	aMI+HCF	0.776 ± 0.069	10.481 ± 3.260
	aMI+SDF	**0.808 ± 0.055**	**10.009 ± 3.130**
RV	LMI	0.680 ± 0.168	19.065 ± 7.503
	aMI+HCF	0.732 ± 0.162	17.745 ± 8.095
	aMI+SDF	**0.765 ± 0.155**	**17.378 ± 7.513**

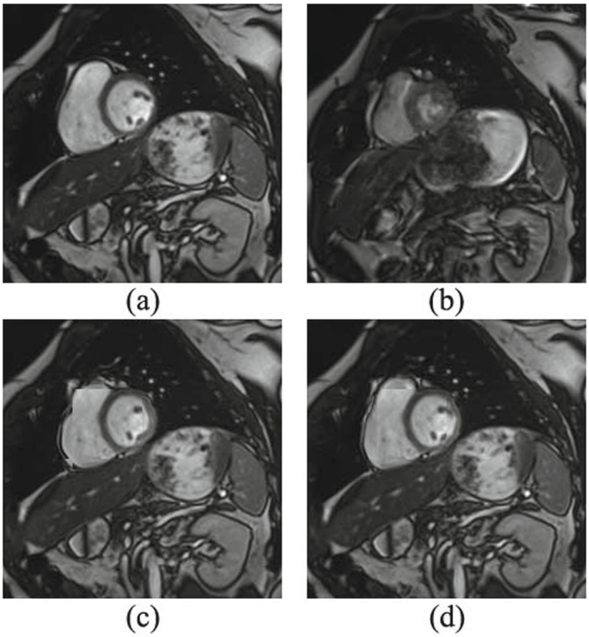

Fig. 3. (a) The fixed image. (b) The moving image. (c) The fusion result by aMI+HCF registration. (d) The fusion result by aMI+SDF registration. The fixed image is combined with the warped moving image, using a checkerboard pattern.

4 Conclusion

In this paper, we present a multi-channel registration algorithm for cardiac MR images. To make the feature representation more robust to large appearance variations of cardiac substructures, we propose to extract the features with convolutional encoder-decoder network. Afterwards, the learned features in a supervised fashion are incorporated into multi-feature mutual information framework. With experiments on cardiac cine-MRI data, the proposed method demonstrates the superior performance regarding to intra-subject registration accuracy.

References

1. Aristeidis, S., Christos, D., Nikos, P.: Deformable medical image registration: a survey. IEEE Trans. Med. Imag. **32**(7), 1153–1190 (2013)
2. Legg, P.A., Rosin, P.L., Marshall, D., Morgan, J.E.: A robust solution to multi-modal image registration by combining mutual information with multi-scale derivatives. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 616–623. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04268-3_76
3. Staring, M., Heide, U.A., Klein, S., Viergever, M.A., Pluim, J.P.W.: Registration of cervical MRI using multifeature mutual information. IEEE Trans. Med. Imag. **28**(9), 1412–1421 (2009)
4. Rivaz, H., Karimaghaloo, Z., Collins, D.L.: Self-similarity weighted mutual information: a new nonrigid image registration metric. Med. Image Anal. **18**, 343–358 (2014)
5. Li, Z., Mahapatra, D., Tielbeek, J.A.W., Stoker, J., Vliet, L.J., Vos, F.M.: Image registration based on autocorrelation of local structure. IEEE Trans. Med. Imag. **35**(1), 63–75 (2016)
6. Guyader, J.M., et al.: Groupwise multichannel image registration. IEEE J. Biomed. Health Inform. **23**(3), 1171–1180 (2019)
7. Bengio, Y., Courville, A., Vincent, P.: Representation learning: a review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell. **35**(8), 1798–1828 (2013)
8. Shin, H., Orton, M.R., Collins, D.J., Doran, S.J., Leach, M.O.: Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data. IEEE Trans. Pattern Anal. Mach. Intell. **35**(8), 1930–1943 (2013)
9. Chmelik, J., et al.: Deep convolutional neural network-based segmentation and classification of difficult to define metastatic spinal lesions in 3D CT data. Med. Image Anal. **49**, 76–88 (2018)
10. Wu, G.R., Kim, M.J., Wang, Q., Munsell, B.C., Shen, D.G.: Scalable high-performance image registration framework by unsupervised deep feature representations learning. IEEE Trans. Biomed. Eng. **63**(7), 1505–1516 (2016)
11. Lu, X.S., Zha, Y.F., Qiao, Y.C., Wang, D.F.: Feature-based deformable registration using minimal spanning tree for prostate MR segmentation. IEEE Access **7**, 138645–138656 (2019)
12. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
13. Shamonin, D.P., Bron, E.E., Lelieveldt, B.P.F., Smits, M., Klein, S., Staring, M.: Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer's disease. Front. Neuroinform. **7**(50), 1–15 (2014)
14. Bernard, O., et al.: Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE Trans. Med. Imag. **37**(11), 2514–2525 (2018)
15. Klein, S., Heide, U.A., Lips, I.M., Vulpen, M., Staring, M., Pluim, J.P.: Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. Med. Phys. **35**(4), 1407–1417 (2008)