



Deformable MRI-Ultrasound Registration Using 3D Convolutional Neural Network

Li Sun^(✉) and Songtao Zhang

Southern University of Science and Technology, Shenzhen 518055, China
lisun97@foxmail.com

Abstract. Precise tracking of intra-operative tissue shift is important for accurate resection of brain tumor. Alignment of pre-interventional magnetic resonance imaging (MRI) to intra-operative ultrasound (iUS) is required to access tissue shift and enable guided surgery. However, accurate and robust image registration needed to relate pre-interventional MRI to iUS images is difficult due to the very different nature of image intensity between modalities. Here we present a framework that can perform non-rigid MRI-ultrasound registration using 3D convolutional neural network (CNN). The framework is composed of three components: feature extractor, deformation field generator and spatial sampler. Our automatic registration framework adopts unsupervised learning approach, allows accurate end-to-end deformable MRI-ultrasound registration. Our proposed method avoids the downfall of intensity-based methods by considering both image intensity and gradient. It achieves competitive registration accuracy on RESECT dataset. In addition, our method takes only about one second to register each image pair, enabling applications such as real time registration.

Keywords: MRI-ultrasound registration · 3D CNN · Deep learning

1 Introduction

Radiological imaging is commonly used for diagnosis, treatment and scientific research. Different modalities of techniques are often used concordantly in practice because they complement with each other. MRI measures the relaxation times of the ^1H nuclei, it can provide visualization for the overall structure and anatomy, while iUS measures the changes in acoustic impedance, it is relative inexpensive and allows for intra-operative detection.

Image registration refers to the spatial alignment of images into the same coordinate system. It can greatly facilitate a wide range of medical applications from diagnosis to therapy. As far as brain tumor resection is concerned, accurate registration can provide the boundary of brain tumor and corresponding tissue shift. Many algorithms and software toolkits have been developed for image registration [1, 5]. However, most current methods focus on registration within

modality and are based on intensity values. These intensity-based registration methods may fail in inter-modality registration tasks, such as MRI-iUS image registration. This is due to the different underlying principles of imaging techniques and striking difference in field of views. Inter-modality image registration poses special challenges and robust and accurate methods are still desired.

In recent years, deep convolutional neural networks (CNNs) have achieved great success in the field of computer vision. Inspired by the biological structure of visual cortex, CNNs are artificial neural networks with multiple hidden convolutional layers between the input and output layers. They have non-linear property and are capable of extracting higher level representative features. CNNs have been applied into a wide range of fields and achieved state-of-the-art performance on tasks such as image recognition, instance detection, and semantic segmentation. In this paper, we propose a novel learning-based framework for MRI-iUS image registration. It is composed of three parts: feature extractor, deformation field generator and spatial sampler. Our automatic registration framework allows accurate and fast MRI-ultrasound registration.

2 Related Work

2.1 Intensity-Based Approaches for Registration

To date, a lot of traditional intensity-based methods have been reported for medical image registration [1, 5]. These methods usually include the following steps. First, a transformation model is selected to deform the moving image and spatially align the intensity between fixed image and deformed moving image. The choice of transformation model depends on the complexity of deformations required. For example, simple transformation such as rigid, affine and B-spline transformation are enough to recover underlying rigid deformations. In more complicated cases, more flexible non-parametric transformation models are used to recover complex deformations.

Second, a similarity metric is defined to how well two images are matched after transformation. The selection of the similarity metric, also called the cost function, depends on the intrinsic properties of images to be registered and deformation complexity. Commonly used metrics include sum of squared distances, normalized cross-correlation (NCC), mutual information (MI) and others.

Finally, iterative optimization method is applied to update the transformation parameters to minimize the cost function. Traditional medical image registration methods have achieved acceptable result in many registration tasks. But there are two downfalls for these methods. First, most of methods focus on aligning image intensity, which may fail in inter-modality image registration. For example, MRI and iUS image have strikingly different fields of view, which is due to different nature in imaging principles. In addition, minimizing cost function by iterative optimization is slow, which may hinder application of image registration.

2.2 Learning-Based Approaches for Registration

Several studies have exploited learning-based approaches for image registration [6, 8]. Recently, CNNs have been applied to many computer vision tasks, including image registration. Deep CNNs contain many hidden layers so that they can non-linearly transform input data and extract higher level features, thus by training it can learn to determine the optimal decision boundary in the high-dimensional feature space. Wu et al. [8] utilize convolutional stacked auto-encoder to select deep feature representations in image patches, then estimate the deformation pathway. Miao et al. [6] use convolutional neural network to predict a transformation matrix, which is then used to perform rigid registration. In this paper, we follow these ideas and propose an end-to-end model for deformable image registration in an unsupervised learning way.

2.3 Spatial Transformer Network (STN)

Jaderberg et al. [4] proposed the spatial transformer network, which enables the learning of spatial transformation. STN is a fully differentiable module so that it can be inserted into existing convolutional neural networks, giving CNNs the ability to spatially transform feature maps. STN takes transformation parameters as input, then it generates a sampling grid according to the parameters. The sampling grid is used to spatially transform image by bilinear interpolation. By training with supervision, STN is capable to learn a dynamic mechanism to actively spatially transform an image by producing an appropriate transformation for each input voxel, including scaling, cropping, rotations, as well as non-rigid deformations. de Vos et al. [7] applied STN to handwritten digit registration, but it requires large amount of data for training.

3 Methodology

3.1 Problem Statement

In image registration, the moving image I_M , is deformed to match the corresponding image I_F called the fixed image. Thus, the deformed image \tilde{I} can be expressed as

$$\tilde{I} = I_M(x + u(x)) \quad (1)$$

where x denotes a three-dimensional coordinate and u represents the deformation field. In this work, we attempt to predict the optimal deformation field $u(x)$ to register MRI to corresponding iUS image.

3.2 Registration Framework

Our registration framework is composed of three components: feature extractor, deformation field generator and spatial sampler. The overall workflow is illustrated in Fig. 1:

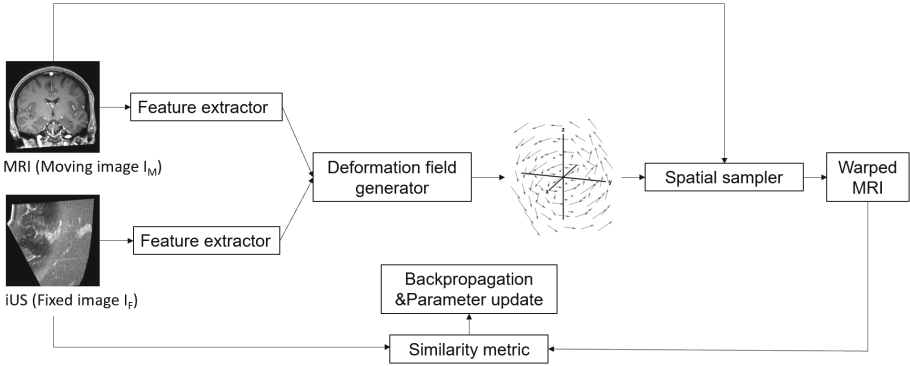


Fig. 1. Framework overview

For feature extractor, two fully convolutional neural networks are used to extract higher level representative features from MRI and iUS images respectively. Each network contains three convolutional layers with 16 kernels sized $3 \times 3 \times 3$, coupled with batch normalization and exponential linear units for activation. The extracted features are concatenated and fed into the deformation field generator.

The deformation field generator takes features extracted from both MRI and iUS images as input, and it produces a deformation field as output. The structure of deformation field generator is inspired by FlowNet [2], which is original used to estimate optic flow. It is composed of a contracting part and an expanding part. The contracting part includes three convolutional layers and a downsampling layer, which is used to capture context and deep level features. The expanding part is consisted of a upsampling layer and three convolutional layers, which is used to restore details and produce a deformation field the same size as the input image. Skip connections are also incorporated to integrate both high-level and low-level features. All layers contain 16 filters sized $3 \times 3 \times 3$, and are coupled with batch normalization and exponential linear units for activation, except for the last layer which use linear activation. The resulting deformation field is fed into the spatial sampler (Fig. 2).

Finally, a spatial sampler is used to apply the deformation field to regular spatial grid, resulting in the sampling grid. The MRI image is resampled by bilinear interpolation. And the deformed MRI image is aligned to the iUS image to calculate the similarity. The loss is backpropagated into the network and update the parameters. The training process is unsupervised as it does not need expert-labeled landmark data.

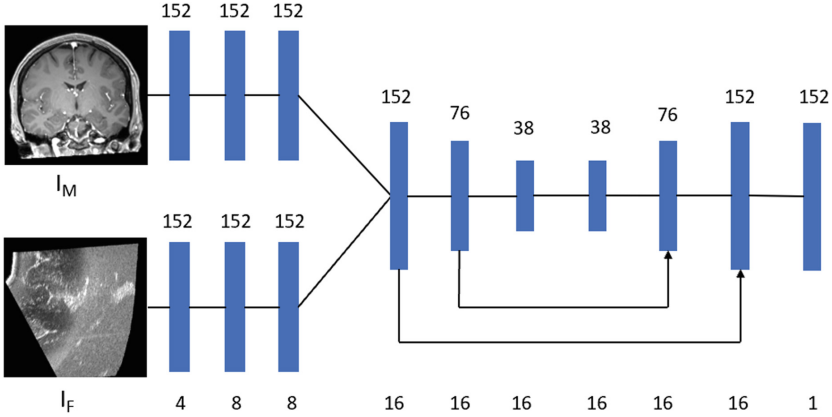


Fig. 2. Detailed structures of feature extractor and deformation field generator. Note that the size and number of channels of each feature map are shown on the top and bottom of figure respectively.

3.3 Similarity Metric

We evaluate the registration quality by considering both the image intensity and gradient. Many conventional intensity-based methods are not appropriate for this inter-modality registration task, because MRI and iUS images have very different nature in intensity values. To tackle this, we assume that the US intensity value $u_i = I_M(x + u(x))$ for voxel i is either correlated with the corresponding MRI intensity value or with the MRI image gradient magnitude $g_i = |\nabla p_i|$. As suggested by Fuerst et al. [3] that, ultrasound intensity values may describe different properties of internal fluids and tissues as well as represent tissue interfaces or gradients. Thus, we define the loss function as:

$$\sum_{x \in \phi} (I_F(x) - (\alpha p_i + \beta g_i + \gamma))^2 \quad (2)$$

in which α , β and γ are learnt parameters during training. We assume that the network will automatically find the optimal parameter to make the deformed MRI image best fit with the iUS image.

4 Experiments

4.1 Dataset

We use the publicly available RESECT dataset [9] for training and validation. The dataset provides pre-operative T1w and T2-FLAIR MRI and iUS images from 23 patients. It also provides expert-labeled homologous anatomical landmarks, defined on all image modalities. All data were acquired for routine clinical care at St Olavs University Hospital, after patients gave their informed consent. The imaging data are available in both MINC and NIFTI formats.

4.2 Preprocessing

We use T1w MRI scans and before resection intra-operative US images for training and validation, which account for 22 image pairs. We split 18 cases for training phase and 4 cases for validation phase. We downsample all images to $150 \times 150 \times 150$ to reduce memory usage and suppress speckle noise. In order to augment the training data, we applied random flipping, rotation, cropping, as well as random gaussian noise to the images.

4.3 Result

In order to evaluate the performance of our method, we applied the trained model on validation dataset and calculated the mean target registration errors (mTREs) between the predicted landmark positions on the iUS images and ground truth. The evaluation results in training phase and validation phase are listed as follows (Table 1):

Table 1. Evaluation result

Phase	mTREs (mm)	Std. (mm)	Process time per image (sec)
Training	4.73	2.71	2.66
Validation	3.91	0.53	1.21

4.4 Implementation Details

To implement the algorithm, we use Tensorflow framework and a NVIDIA Tesla M40 GPU accelerator. We use stochastic gradient descent optimizer with momentum 0.9, and set initial learning rate to 0.001. We also set the number of epoch for training 20 and batch size to 3 for training.

5 Conclusion

In this paper, we present a framework that can perform non-rigid MRI-ultrasound registration using 3D convolutional neural network. This framework is composed of feature extractor, deformation field generator and spatial sampler. Our fully automatic registration framework adopts a learning-based approach and it avoids the downfall of intensity-based methods by considering both image intensity and gradient. In addition, our method only takes one second to register each image pair. Moreover, our method is unsupervised, without the requirement for expert-curated landmarks for training. The evaluation result on RESECT dataset demonstrated that our proposed method achieves competitive registration accuracy, and it can be applied to other cross-modality image registration tasks. In the future, we will explore more possibilities of optimizing network structure and penalizing shadow regions as suggested by Fuerst et al. [3].

References

1. Avants, B.B., Tustison, N., Song, G.: Advanced Normalization Tools (ANTs). *Insight J.* **2**, 1–35 (2009)
2. Dosovitskiy, A., et al.: FlowNet: learning optical flow with convolutional networks. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2758–2766 (2015)
3. Fuerst, B., Wein, W., Müller, M., Navab, N.: Automatic ultrasound-MRI registration for neurosurgery using the 2D and 3D LC² metric. *Med. Image Anal.* **18**(8), 1312–1319 (2014)
4. Jaderberg, M., Simonyan, K., Zisserman, A., et al.: Spatial transformer networks. In: *Advances in Neural Information Processing Systems*, pp. 2017–2025 (2015)
5. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.: Elastix: a toolbox for intensity-based medical image registration. *IEEE Trans. Med. Imaging* **29**(1), 196–205 (2010)
6. Miao, S., Wang, Z.J., Zheng, Y., Liao, R.: Real-time 2D/3D registration via CNN regression. In: *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pp. 1430–1434. IEEE (2016)
7. de Vos, B.D., Berendsen, F.F., Viergever, M.A., Staring, M., Išgum, I.: End-to-end unsupervised deformable image registration with a convolutional neural network. In: Cardoso, M.J. (ed.) *DLMIA/ML-CDS -2017*. LNCS, vol. 10553, pp. 204–212. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67558-9_24
8. Wu, G., Kim, M., Wang, Q., Munsell, B.C., Shen, D.: Scalable high-performance image registration framework by unsupervised deep feature representations learning. *IEEE Trans. Biomed. Eng.* **63**(7), 1505–1516 (2016)
9. Xiao, Y., Fortin, M., Unsgård, G., Rivaz, H., Reinertsen, I.: REtroSpective Evaluation of Cerebral Tumors (RESECT): a clinical database of pre-operative MRI and intra-operative ultrasound in low-grade glioma surgeries. *Med. Phys.* **44**(7), 3875–3882 (2017)