2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

End to End Unsupervised Rigid Medical Image Registration by Using Convolutional Neural Networks

Huiying Liu[1], Yanling Chi[1], Jiawei Mao[2], Xiaoxiang Wu[3], Zhiqiang Liu[3], Yuyu Xu[3], Guibin Xu[3#], Weimin Huang[1#]

***Abstract*— In this paper, we focus on the issue of rigid medical image registration using deep learning. Under ultrasound, the moving of some organs, e.g., liver and kidney, can be modeled as rigid motion. Therefore, when the ultrasound probe keeps stationary, the registration between frames can be modeled as rigid registration. We propose an unsupervised method with Convolutional Neural Networks. The network estimates from the input image pair the transform parameters first then the moving image is wrapped using the parameters. The loss is calculated between the registered image and the fixed image. Experiments on ultrasound data of kidney and liver verified that the method is capable of achieve higher accuracy compared with traditional methods and is much faster.**

I. INTRODUCTION

Medical image registration has been investigated for tens of years due to its importance in medical diagnosis, screening, atlas construction, treatment planning, et. al. In the era of deep learning, this powerful tool has also been applied to medical image registration and there are fruitful results. In this paper, we focus on rigid ultrasound image registration. Compared with CT and MRI, ultrasound is fast and real time, the ultrasound facility is lightweight so it is widely used in ultrasound guided surgery [1]. The moving of the target organ makes some difficulties thus the registration is necessary for better application.

Image registration can be multi-modality and mono-modality registration [2]. It can be rigid and deformable registration [3]. It can be 2D and 3D registration [4]. For different application scenarios, corresponding schemes are needed. In this paper, we focus on the situation that the moving of the target organ is rigid motion, e.g., kidney and liver. It is mono-modality, 2D, and rigid registration. While there are already tens of papers targeting at medical image registration with deep learning, most of these methods are suitable for non-rigid registration which may be a more general situation. For a thorough survey, the readers are referred to [8].

In [6], the authors proposed a supervised deep learning method to do rigid registration. However, as it is well known that, the labeling data is not easy to be obtained. In the paper [6], the authors generated artificial data for training and test. In this paper, inspired by the work [9], we extend this work to unsupervised, by taking the difference between registered image and fixed image as loss function.

The novelty and contribution of this paper is that we propose an Convolutional Neural Network based rigid medical image registration method. This method has two major properties. First, it is end to end, meaning that the input is the image pairs, and output is the registered image instead of the transform parameters. Secondly, it is un-supervised. As we all know, the ground truth is not always available due to the labor intensive data labeling work. Thus a method with least labeled data is preferred in most cases.

In the rest of this paper, we will describe the proposed method in detail in Sec. 2. The experiments on ultrasound data of kidney and liver are presented in Sec. 3. Finally, we will conclude our work in Sec. 4.

II. THE METHOD

Rigid registration has three parameters, rotation (θ), horizontal translation (dx) and vertical translation (dy). Let $\mathbf{x} = (x, y)^T$ be a point in the image, then its position in the registered image, $\mathbf{x}' = (x', y')^T$, will be

$$\mathbf{x}' = \mathbf{R}\mathbf{x} + d\mathbf{x}.$$

Here $\mathbf{R} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$, $d\mathbf{x} = (dx, dy)^T$. Our target is to estimate the three parameters using deep learning. θ is in radian, ranges from $-\pi$ to π. dx and dy are normalized using the size of the image, range from -1 to 1.

The proposed method is extended from the work in [6], which proposed a supervised learning method for rigid medical image registration. In that work, the input to the network is the image pairs, each consists of a fixed image and a moving image. The network outputs the rigid parameters (θ, dx, dy), where θ is the rotation parameter, and dx and dy are the horizontal and vertical translation respectively. The mean square error (MSE) between the ground truth parameters and the estimated parameters is used as loss function. The method is illustrated in Fig. 1. The limitation of this work is that it needs the ground truth parameters which are not always available. In the paper, the author generated random motion of the images thus the parameters are known. But in real case, it is not an easy job to get the parameters. It usually requires labor intensive work.

[1] Huiying Liu, Weimin Huang, and Yanling Chi are with the Institute for Infocomm Research, A*STAR, Singapore. {liuhy, wmhuang, chiyl}@i2r.a-star.edu.sg

[2] Jiawei Mao is with Creative Medtech Solutions Pte Ltd. mao.jiawei@ultrastmedtech.com

[3] Xiaoxiang Wu, Zhiqiang Liu, Yuyu Xu, Guibin Xu are with the Department of Urology, Fifth Affiliated Hospital of Guangzhou Medical University, Guangzhou 510700, China. 348215424@qq.com, zhiqiang_liu2012@163.com, gyxyy@foxmail.com, gyxgb@163.com

[#] Correspondence authors

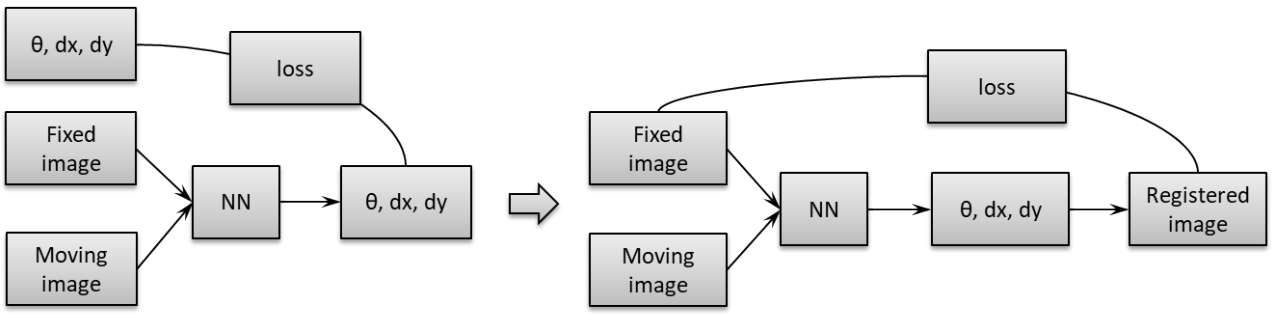

Fig. 1. Framework of the proposed work, compared with the one of [6].

In our case, what we have are ultrasound image sequences of the organ. The ultrasound probe is kept un-moved, the organ is moving rigidly. What we need to do is the registration between the frames. Due to the limited motion intensity, we model this issue as rigid registration. The ground truth parameters are not available in our case. Thus we extend the above mentioned work to unsupervised and end to end.

A. Image preprocessing

An example of ultrasound image of kidney is illustrated in Fig. 2. It consists of the ultrasound region and the surrounding region. In most sequences, the position of the ultrasound region is fixed that a mask is enough to crop the region. In some special cases, we need to crop the region manually.

The images are further cropped to exclude the region below the kidney, as shown in Fig. 2 (a). This process is done by a fixed rectangle which is suitable for most cases. After cropping, the image is enhanced by gray scale normalization.

In each image, the upper part is the skin fat layer, which keeps still due to the pressure of the ultrasound probe. The bottom part contains the organs, i.e., kidney and the surrounding structures, which are moving. The two parts are performing different types of motion thus we need to deal with them with different methods. Since what we are interested in is the lower part, i.e., the kidney, we detect the boundary between the skin fat layer and the organ layer then focus on the organ layer. Fig. 2 (b) shows an example of the boundary. In Fig. 2 (b), it is the average image of the ultrasound sequence. Because the skin fat layer keeps still, this part keeps sharp in the average image. The bottom part is blurred due to the motion of the organ.

B. Network

The network structure is like the fully convolutional network in [6]. Following the work in [6], we do not use average pooling, because it may reduce the network's sensitivity to subtle moving. Dropout layer with 0.5 is used instead. Batch normalization is performed at each layer.

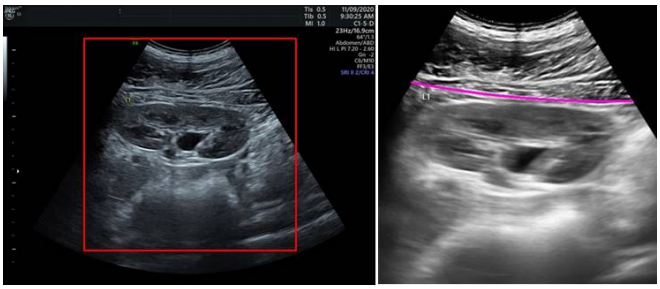

Fig. 2. An example of the ultrasound data and the boudnary between the skinfat layer and the organ layer.

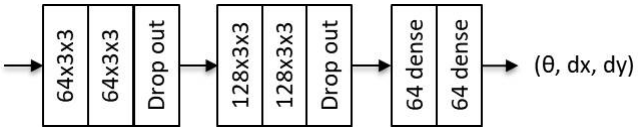

Fig. 3. The structure of the network.

C. Loss Function

Image registration is to minimize the difference between the registered image and the fixed image. Mean Square Error (MSE) between the fixed image and registered image is used as the loss function. Let I_y be the fixed image, I_x as the moving image, I_{xy} be the registered image, L_{xy} be the MSE between I_y and I_{xy}, then L_{xy} is calculated as

$$L_{xy} = \frac{1}{N} \sum_{(i,j)\in\Theta} \left(I_y(i,j) - I_{xy}(i,j)\right)^2$$

Here Θ is the set of valid pixels, meaning the organ region in the ultrasound image. N is the number of valid pixels. A regularization item is also added to the loss function to constrain the motion intensity. It it calculated as

$$L_{REGU} = \frac{1}{3}\left(\theta^2 + dx^2 + dy^2\right)$$

We also incorporate Inverse Consistency Error (ICE) into the loss function. Aiming at this purpose, we also perform registration from the fixed image (I_y) to the moving image (I_x). Let the registration result is I_{yx}, we get the loss L_{yx}.

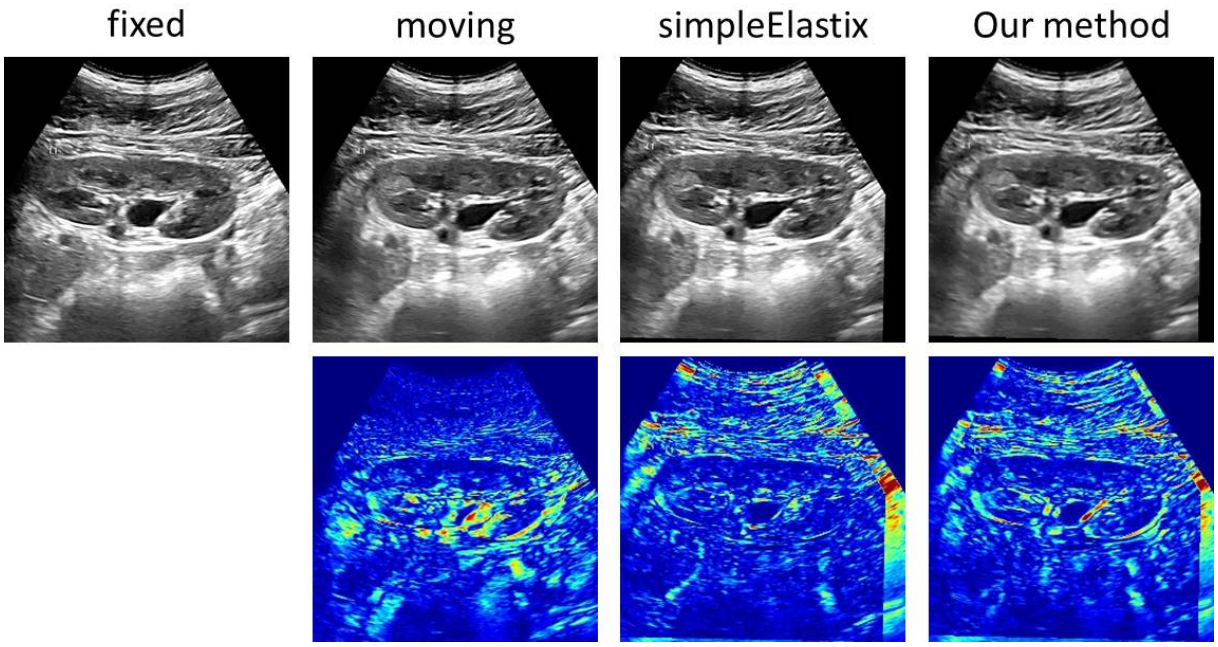

Fig. 4. Exampls of comparison on kidney dataset. In the bottom line are the residual images. This the 22^{th} frame of the same sequence as in Fig. 6.

The final loss function is

$$Loss = L_{xy} + L_{yx} + \lambda L_{REGU}$$

λ is set as 0.01 in the experiments.

D. Activation function

In most networks, activation function is not used at the last layer. In our case, although θ may range from π to π, it obeys $\theta \sim N(0, 0.01)$. dx and dy are normalized by the width and height of the image thus range between -1 and 1. The activation function is chosen to constrain the output to locate between -1 and 1. Tanh is chosen for the last layer for this purpose. In the previous layers, elu is chosen as activation function.

III. Experiments

The experiments are performed on two datasets, one kidney dataset and one liver dataset. The baseline method chosen for comparison is SimpleElastix [7]. The parameter map is set as "rigid" and the B-spline interpolation order is set as 0.

A. Implementation

The method is implemented using TensorFlow [10]. NVidia Quadro K5100M GPU and CUDA 9.0 are used for acceleration. Adam optimizer is chosen with learning rate of 0.0001 at the training stage. For both the two datasets, 500 iterations are trained. The image size is 64×64, the batch size is 64.

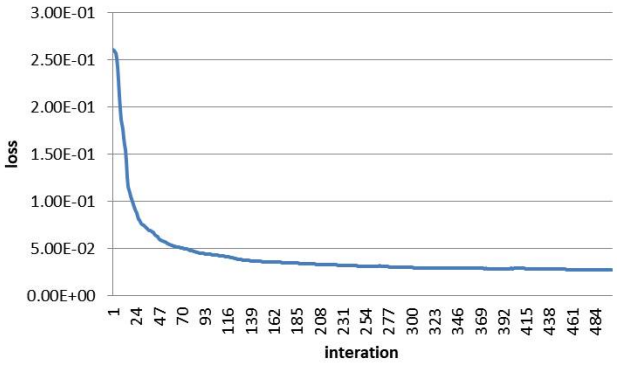

Fig. 5. The loss curve at training stage, on the kidney dataset.

B. Experiments on Kidney Ultrasound

The kidney dataset are collected at hospital. Data of 10 patients are used. Each patient has 4 sequences, each sequence covers 60-100 frames. From each sequence, 2 times image pairs are chosen for training, meaning that if the image sequence includes 60 frames, 120 image pairs are chosen for training. 3904 image pairs in total are used for training. At the test stage, all the frames are registered to the first one. Fig. 4, Fig. 5 and Fig. 6 show the comparison examples, loss curve and MSE comparison, respectively. The average MSE of SimpleElastix and the proposed method are 1.70×10^3 and 1.59×10^3, respectively.

From the result, we can see that the network is able to get comparable or better result in means of MSE. On the same PC, SimpleElastix needs 6 seconds for an image pair, CNN needs about 0.1 seconds for an image pair.

An example is shown in Fig. 4. From the result, we can

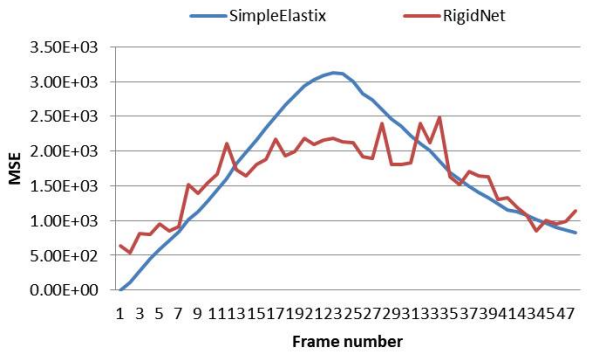

Fig. 6. MSE of SimpleElastix and our method on kidney dataset.

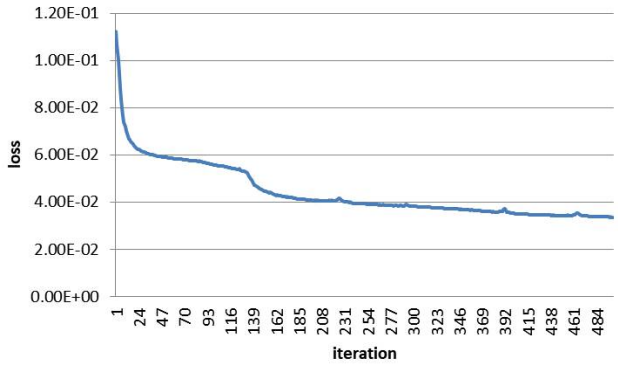

Fig. 7. The loss curve at training stage, on the liver dataset.

find a drawback of the proposed method. It is that, although the MSE shows better performance compared with SimpleElastix, its matching maybe not better than SimpleElastix. This is because MSE is an overall metric to measure the similarity between two images, thus it may be not suitable for local matching. We will try other metrics as loss function.

C. Experiments on Liver Ultrasound

CLUST [5] is a dataset of liver ultrasound. It consists of 64 2D-t sequences and 22 4D sequences. Our experiment was performed on the 2D-t data of 9 patients which are publicly available. These sequences are of healthy volunteers acquired during free-breathing over a period of 5-10 minutes. The liver is moving periodically with breathing. The first 300 frames from each sequence are used in our experiments. The first 8 sequences are used for training and the rest one is used for test.

The loss curve and the comparison result on this dataset is shown in Fig. 7 and Fig. 8, respectively. The average MSE of SimpleElastix and the proposed method are 1.40×10^2 and 5.81×10^1, respectively.

IV. CONCLUSIONS

In this paper, we propose an unsupervised rigid registration method with convolutional neural networks. It does not require labeled data for training. Also it outputs the registration result directly instead of the transform parameters.

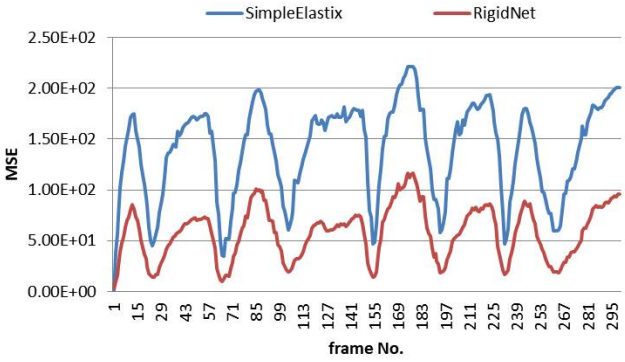

Fig. 8. MSE of SimpleElastix and our method on liver dataset.

It achieves comparable performance with traditional method, simpleElastix, and is faster than simpleElastix.

REFERENCES

[1] Arko, Darja, Nina as Sikoek, Nejc Kozar, Monika Soboan, and Iztok Taka. "The value of ultrasound-guided surgery for breast cancer." European Journal of Obstetrics & Gynecology and Reproductive Biology 216 (2017): 198-203.

[2] Cao, Xiaohuan, Jianhua Yang, Yaozong Gao, Qian Wang, and Dinggang Shen. "Region-adaptive deformable registration of CT/MRI pelvic images via learning-based image synthesis." IEEE Transactions on Image Processing 27, no. 7 (2018): 3500-3512.

[3] de Vos, Bob D., Floris F. Berendsen, Max A. Viergever, Marius Staring, and Ivana Igum. "End-to-end unsupervised deformable image registration with a convolutional neural network." In Deep learning in medical image analysis and multimodal learning for clinical decision support, pp. 204-212. Springer, Cham, 2017.

[4] Kuang, Dongyang, and Tanya Schmah. "Faima convnet method for unsupervised 3d medical image registration." In International Workshop on Machine Learning in Medical Imaging, pp. 646-654. Springer, Cham, 2019.

[5] L. Petrusca, P. Cattin, V. De Luca, F. Preiswerk, Z. Celicanin, V. Auboiroux, M. Viallon, P. Arnold, F. Santini, S. Terraz, K. Scheffler, C. D. Becker, R. Salomir, "Hybrid Ultrasound/Magnetic Resonance Simultaneous Acquisition and Image Fusion for Motion Monitoring in the Upper Abdomen", Investigative Radiology, Vol. 48, No. 5, pp.333-340.

[6] Sloan, J.M., Goatman, K.A. and Siebert, J.P., 2018. Learning rigid image registration-utilizing convolutional neural networks for medical image registration.

[7] Marstal, Kasper, Floris Berendsen, Marius Staring, and Stefan Klein. "SimpleElastix: A user-friendly, multi-lingual library for medical image registration." In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 134-142. 2016.

[8] Fu, Yabo, Yang Lei, Tonghe Wang, Walter J. Curran, Tian Liu, and Xiaofeng Yang. "Deep learning in medical image registration: a review." Physics in Medicine & Biology 65, no. 20 (2020): 20TR01.

[9] de Vos, Bob D., Floris F. Berendsen, Max A. Viergever, Hessam Sokooti, Marius Staring, and Ivana Igum. "A deep learning framework for unsupervised affine and deformable image registration." Medical image analysis 52 (2019): 128-143.

[10] Martn Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Rafal Jozefowicz, Yangqing Jia, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Man, Mike Schuster, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Vigas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.