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

An Unsupervised Non-rigid Registration Network for Fast Medical Shape Alignment

Jinyang Shi, Pen Wan, and Fang Chen*

Abstract— **Medical shapes alignment can provide doctors with abundant structure information of the organs. As for a pair of the given related medical shapes, the traditional registration methods often depend on geometric transformations required for iterative search to align two shapes. To achieve the accurate and fast alignment of 3D medical shapes, we propose an unsupervised and nonrigid registration network. Different from the existing iterative registration methods, our method estimates the point drift for shape alignment directly by learning the displacement field function, which can omit additional iterative optimization process. In addition, the nonrigid registration network can also adapt to the geometric shape transformations of different complexity. The experiments on two types of 3D medical shapes (liver and heart) at different-level deformations verify the impressive performance of our unsupervised and nonrigid registration network.**

Clinical Relevance—**This paper achieves the real-time medical shape alignment with high accuracy, which can help doctors to understand the pathological conditions of organs better.**

I. Introduction

Three-dimensional (3D) medical shapes can intuitively show the organ structures; thus, it is conducive for doctors to understand the pathological and physiological conditions of organs. Multiple medical shapes can be obtained for an organ through various medical imaging devices at different imaging time. Through the shape alignment, multiple medical shapes obtained at different imaging time can be fused together to acquire the comprehensive structure information of the organs [1]. For example, in cardiac valve replacement surgery navigation, aligning the patient's heart shapes acquired before and after the surgery can help doctors to analyze the structural changes brought by the surgical operation and adjust the treatment in time. Therefore, the real-time and precise shape alignment directly influences the success of the surgical operations. In this paper, we introduce an unsupervised and nonrigid shape alignment method based on deep-learning to achieve the real-time medical shape alignment with high accuracy.

II. Related Works

For medical shape registration task, a number of algorithms have been proposed, which can be divided into two categories: the iterative registration methods and the learning-based registration methods.

A. Iterative Registration

Iterative Closest Point (ICP) [2] is one of the most common registration methods which selects the corresponding points from the beginning to iteratively optimize the transformation until the distance between two-point sets is less than the given threshold. However, in early stage, ICP algorithm is applicable for the rigid shape alignment, and cannot handle the nonrigid cases. To improve the ICP algorithm, a new Go-ICP [3] method is proposed by using the branch-and-bound optimization, but this method is computationally expensive, which is chosen only for the practical scenarios where real-time performance is not critical. In addition, Myronenko *et al*. [4] proposed a nonparametric coherent point drift (CPD) algorithm, which leveraged Gaussian mixture likelihood and the penalization term on the velocity field to finish shape alignment. Thus, CPD algorithm is limited to the Gaussian kernel and relatively sensitive to the rotation of the target shape. In all, these iterative registration methods rely on the time-consuming iterations and cannot achieve real-time registration.

B. Learning-based Registration

Learning 3D shapes' features based on deep neural network has been widely used for 3D shape point set analysis. For example, PointNet [5] is the first deep neural network which processes 3D shape point clouds directly. PointNet cannot capture local structures induced by the metric shape space points. To solve this problem, Charles *et al*. [6] proposes the PointNet++, which exploits metric shape space distances to learn local features with increasing contextual scales. Based on these shape feature learning network, a method called PointNetLK [7], which minimizes the distance between the fixed-length global descriptors to achieve the 3D shape registration. However, the PointNetLK algorithm is not robust to noise, which influences the shape alignment accuracy. Different from the above methods, this paper introduces a new unsupervised and nonrigid registration network for the fast shape alignment. Especially, the proposed network can adapt to the different complexity of the target geometric transformation.

* Research supported by the National Nature Science Foundation of China grants (U20A20389, 61901214), China Postdoctoral Science Foundation (2021T140322, 2020M671484), Jiangsu Planned Projects for Postdoctoral Research Funds(2020Z024), High-level Innovation and Entrepreneurship Talents Introduction Program of Jiangsu Province of China.

F. Chen, J. Yang and P. Wan are with the Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China. F. Chen is the corresponding authors of this paper (E-mail: chenfang@nuaa.edu.cn)

III. METHOD

Figure 1 shows the framework of the unsupervised and nonrigid registration network, that consists of two parts: Learning Global Descriptor and Point Displacement Prediction. Here, the point displacement vector is predicted directly by a deep neural network.

We define a training dataset D, which includes the source medical shape S_i and target medical shapes T_j, and $S_i, T_j \subset R^3$. Then we assume that there is a function $F_\theta(S_i, T_j) = \delta$ that can represent the displacement of the from S_i to T_j, and the θ means the weights of the neural network. We transform the input shapes based on the point displacement, and leverage the difference between the transformed source medical shapes and target medical shapes as the registration loss to update θ.For the weights θ, we use the SGD algorithm to optimize it, and the formula is as follows:

$$\theta^{optimal} = \underset{\theta}{\operatorname{argmin}}\left[L(S_i, T_j, F_\theta(S_i, T_j))\right] \quad (1)$$

, where L is a difference measure, i, j means the points in the source and target shapes, respectively. s

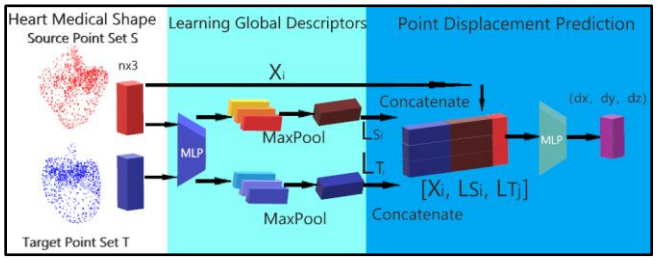

Figure 1. The structure of our unsupervised and nonrigid registration network.

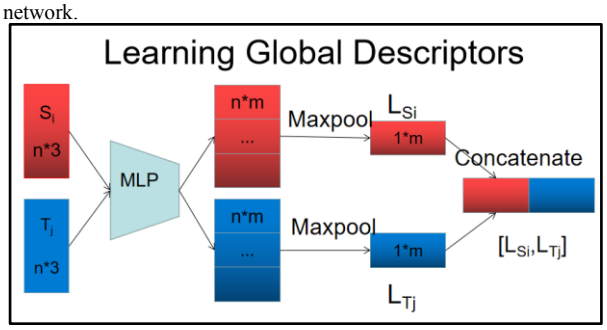

Figure 2. The schema of learning global descriptor.

A. Learning Global Descriptors

In this part, we firstly learn a global shape descriptor from the given medical shapes in order to capture shape representation. In order to solve the problem of irregular format of medical shapes, we use the multi-layer perceptron (MLP) and the maxpool functions. For MLP, its activation function is RELU. The aim of leveraging maxpool function is to extract the order-invariant features of the source and target medical shapes. We assume $\left(L_{S_i}, L_{T_j}\right)$ denotes shape descriptors of source and target medical shapes, and the network for Learning Global Descriptor is defined as follows:

$$L_{S_i} = Maxpool\{g_t g_{t-1} \cdots g(X_i)\}_{X_{i \in S_i}} \quad (2)$$

$$L_{T_j} = Maxpool\{g_t g_{t-1} \cdots g_1(X_i)\}_{X_i \in T_j} \quad (3)$$

, where the g_t is the RELU activation function, and can be described as $g_t: R^{W_t} \longrightarrow R^{W_{t+1}}$, the W_t and W_{t+1} are the dimensions of input layers and output layers. Then, L_{S_i} and L_{T_j} are concatenated to acquire a tensor with a dimension of $1 * 2m$ (here $m = 512$). The more detailed process of the Learning Global Descriptor is shown in figure 2.

B. Point Displacement Prediction

After capturing shape representative features, we further predict the point displacement function between the target and source shapes. In general, for each point x in medical shape X, the displacement function Ψ can be defined as:

$$\Psi(x) = x + \Delta(X) \quad (4)$$

, where Δ is the point displacement function for 3D shape data. Thus, the process of registration needs to calculate this displacement function. In this way, the source medical shape can be moved to the target medical shape continuously. And because we leverage deep neural network, the function is smooth and continuous.

As shown in figure 1, we copy the concatenated descriptors $\left[L_{S_i}, L_{T_j}\right]$ learned from the former part firstly. Then, we concatenate them with the low-dimensional coordinate of every point as global descriptor features $\left[X_i, L_{S_i}, L_{T_j}\right]$. After that, we substitute the above features as input into the MLP architecture we designed to learn the displacement of the points in the source medical shape and the target medical shape. The architecture includes continuous MLP with RELU activation functions.

We define the predicted displacement of each point X_i from S_i as d_{X_i}:

$$d_{X_i} = f_s f_{s-1} \cdots f_1\left(\left[X_i, L_{S_i}, L_{T_j}\right]\right) \quad (5)$$

, where f_s represents a perceptron. The f_s can be described as $f_s: R^{V_s} \longrightarrow R^{V_{s+1}}$, the V_s and V_{s+1} are the dimensions of input layers and output layers. The notation $[*,*]$ indicates the concatenation of different data.

According to the above definitions, we can define the transformed source shape as S_i', as follows:

$$S_i' = \delta(S_i) = \{X_i + d_{X_i}\}_{X_i \in S_i} \quad (6)$$

, where S_i' is the point set converted from the source shape point set.

For two medical shapes, since there is no correspondence between each shape, pixel-level loss cannot be used for shape alignment. Therefore, based on Chamfer Distance proposed by Fen *et al*. [8], the Chamfer Distance Error is used and defined as:

$$L(S', T|\theta) = \sum_{X \in S'} \min_{Y \in T} \| X - Y \|_2^2$$

$$+\sum_{Y\in T} \min_{X\in S'} \| X - Y \|_2^2 \qquad (7)$$

, where S' and T are our transformed and target medical shapes, and the θ is the weights in our nonrigid registration network structure. X and Y are the points of S' and T.

C. Implementation Details

For the training of the entire network, our training set $\{(S^i, T^i) | (S^i, T^i) \in D\}_{i=1,2,\cdots,b}$ consists of batch data, where b represents the batch size and the b is set to 8 in the experiments. As we mentioned above, in order to learn the shape descriptor tensor, the input is two $n \times 3$ matrices, where n denotes the number of points in the medical shape.

In the Learning Global Description stage, the network consists of 5 MLP layers and a maxpool layer. The dimensions of the MLP layers are (16, 64, 128, 256, 512). And the role of the maxpool layer is to convert it into a 512-dimensional features. In the Point Displacement Prediction stage, we use 3 MLP layers with dimension (256, 128, 3). Except the output layer, for each layer, we implement batch normalization [9] and use RELU activation. In addition, the proposed unsupervised and nonrigid registration network was implemented using the popular deep learning framework Pytorch, and codes were run on the single GPU (i.e., NVIDIA RTX 2080, 8GB). Adaptive moment estimation (Adam) was employed for network optimization. The initial learning rate was set to 0.001. The network weights are trained for 100K iterations with a batch size of 10.

IV. Experiments and Results

A. Experimental Dataset and Evaluation Metrics

In this experiment, we used two types of medical shapes (i.e., 3D Heart and 3D Liver) to prepare the dataset. The used 3D Heart and Liver Shapes contain 1000 and 1013 points, respectively. To obtain the training and testing datasets, we used thin plate spline (TPS) transformation [10], a nonrigid geometric transformation, to deform the medical shapes, because TPS technique has special spline properties and has been widely used in nonrigid image registration [11]. For the training dataset, we performed TPS transformation on the normalized source medical shapes, and synthesized a set of 20k deformed shapes at various deformation levels (0.1, 0.4, 0.8, and 1.2), respectively. We change the deformation level by controlling the disturbance degree of the controlling points in the TPS deformation. In addition, to obtain the testing dataset, we used the TPS transformation to generate 1K testing shapes under different-level deformations, respectively.

Then, for each shape with different-level deformations, we analyzed the shape alignment results using quantitative evaluation and qualitative visual comparison. We used the Chamfer Distance Error between the source and target medical shapes before and after shape alignments to finish quantitative assessment.

B. Results of the Liver Shape

In figure 3, we illustrated the qualitative results of shape alignment for liver shapes at different-level deformations. The first row showed the different levels of deformations. The second and the third rows illustrated the corresponding positions of the target and source medical shapes before and after shape alignments, respectively. We could see that our network was robust to different-level deformations. Even if we increased the deformation level to 1.2, with the difficulty of medical shape alignment increasing, our network could still transform the source medical shape to align the main part of the target medical shape reliably.

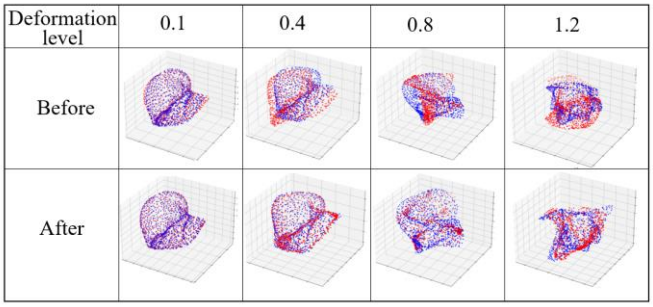

Figure 3. The qualitative results of shape alignments for liver shapes at different-level deformations. The target medical shape was shown in blue, and the source medical shape was shown in red.

In figure 4 (a), we also plotted the mean deviation of the Chamfer Distance Error between the source and target liver shapes before and after the medical shape alignments. It was obvious that the blue curve was below the red curve consistently, and the blue curve remained nearly stable. This indicated that, with the deformation level increasing, even under high deformation level, our network could still align medical shapes accurately.

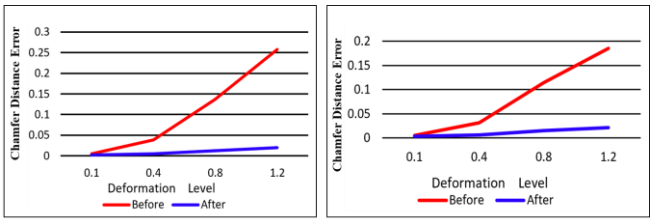

Figure 4. The Chamfer Distance Error between the source and target shapes, before (red line) and after (blue line) the medical shape alignments. (a) the result of liver shape alignment;(b) the result of heart shape alignment.

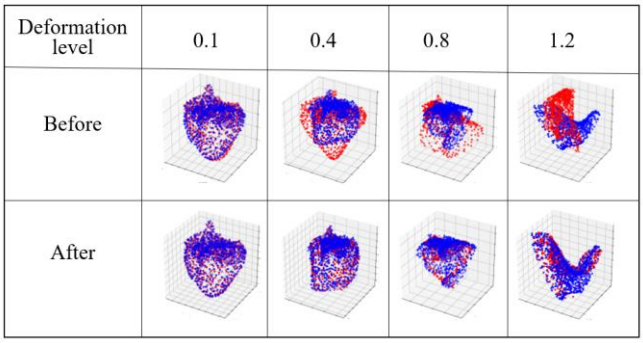

Figure 5. The qualitative results of the shape alignments for heart shapes at different-level deformations. The target medical shape was shown in blue, and the source medical shape was shown in red.

C. Results of Heart Shape

TABLE I. COMPARISON RESULTS WITH THE RELATED WORKS

Method	Liver		Heart	
	Chamfer Distance Error	***Time (Second)***	***Chamfer Distance Error***	***Time (Second)***
ICP [2]	0.027 ± 0.012	0.108	0.025 ± 0.010	0.109
GMMTree [12]	0.035 ± 0.012	48.756	0.027 ± 0.009	48.8785
GMMReg [13]	0.021 ± 0.006	424.221	0.026 ± 0.011	469.653
R-PointHop [14]	0.074 ± 0.036	0.397	0.035 ± 0.016	0.394
Ours	**0.005 ±0.001**	**0.014**	**0.008 ±0.001**	**0.014**

The best results are represented in blue bold.

In figure 5, the results validated that our network was robust to the different-level deformations, and when the deformation level reached to 1.2, the most part of the heart medical shapes could still be aligned. In addition, in the figure 4 (b), we plotted the average deviation of the Chamfer Distance Error between the source and target heart shapes, before and after the shape alignments. Compared with the Chamfer Distance Error before the shape alignment, the Chamfer Distance Error after shape alignment is much less. These experimental results validates that the network is applicable for the alignment of heart medical shapes. The experimental results also show that this network structure can be applied to different medical shapes.

D. Comparison with the Related Works

To demonstrate the advantage of our proposed nonrigid registration approach, we designed the comparison experiments with the related works, including the non-learning based iterative methods (i.e., ICP [2], GMMTree [12], GMMReg [13]), and the learning-based shape matching method (i.e., R-PointHop [14]). In this experiment, we use the Liver and Heart shapes at the deformation of 0.4 to prepare the dataset, and take 20 Liver and Heart shapes as the testing data. For comparison with R-PointHop method, we take another 1000 Liver and Heart shapes as the training data. Table I listed the quantitative comparison results. As shown in Table.1, it can be found that for both liver and heart shapes, the proposed nonrigid registration network outperforms all the compared algorithms with the smallest Chamfer distance error. In addition, our nonrigid registration method consumed significantly less time than the traditional iterative methods. Therefore, our proposed method is suitable for the fast shape alignment tasks.

V. CONCLUSION

We have proposed an unsupervised and nonrigid registration network that can learn the displacement field function for the automatic medical shape alignments. Our experiments verify the impressive performance of our unsupervised and nonrigid registration network on two categories of 3D medical models at different-level deformations. This new solution to the medical shape alignments is especially helpful for doctors to acquire the abundant structure information accurately and quickly.

REFERENCES

[1] E. C. S. Chen, A. J. McLeod, J. S. H. Baxter." et al. Registration of 3D shapes under anisotropic scaling" [J]. International journal of computer assisted radiology and surgery, 2015, 10(6): 867-878.

[2] P. J. Besl, N. D. McKay. "Method for registration of 3-D shapes." *Sensor fusion IV: control paradigms and data structures*. Vol. 1611. International Society for Optics and Photonics, 1992.

[3] J. Yang, H. Li, D. Campbell, et al. "Go-ICP: A globally optimal solution to 3D ICP point-set registration." *IEEE transactions on pattern analysis and machine intelligence* 38.11 (2015): 2241-2254.

[4] A. Myronenko, X. Song, M A. Carreira-Perpinán. "Nonrigid point set registration: Coherent point drift." *Advances in neural information processing systems* 19 (2007): 1009.

[5] C. R. Qi, H. Su, K. Mo, et al. Pointnet: Deep learning on point sets for 3d classification and segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 652-660.

[6] C. R. Qi, H. Su, K. Mo, et al. "Pointnet++: Deep hierarchical feature learning on point sets in a metric space." *arXiv preprint arXiv:1706.02413* (2017).

[7] Y. Aoki, H. Goforth, R. A. Srivatsan, et al.. "Pointnetlk: Robust & efficient point cloud registration using pointnet." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019.

[8] H. Fan, H. Su, L. J. Guibas. "A point set generation network for 3d object reconstruction from a single image." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

[9] S. Ioffe, C. Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." *International conference on machine learning*. PMLR, 2015.

[10] F. L. Bookstein. "Principal warps: Thin-plate splines and the decomposition of deformations." *IEEE Transactions on pattern analysis and machine intelligence* 11.6 (1989): 567-585.

[11] W. Sun, W. Zhou, M. Yang. "Medical image registration using thin-Plate spline for automatically detecting and matching of point sets." *2011 5th International Conference on Bioinformatics and Biomedical Engineering*. IEEE, 2011.

[12] B. Eckart, K. Kim, J. Kautz. "Fast and accurate point cloud registration using trees of gaussian mixtures." *arXiv preprint arXiv:1807.02587* (2018).

[13] B. Jian, B. C. Vemuri. "Robust point set registration using gaussian mixture models." *IEEE transactions on pattern analysis and machine intelligence* 33.8 (2010): 1633-1645.

[14] P. Kadam, M. Zhang, S. Liu, et al. "R-PointHop: A Green, Accurate and Unsupervised Point Cloud Registration Method." *arXiv preprint arXiv:2103.08129* (2021).