



Instantiation-Net: 3D Mesh Reconstruction from Single 2D Image for Right Ventricle

Zhao-Yang Wang¹(✉), Xiao-Yun Zhou¹, Peichao Li¹, Celia Theodoresi-Riga^{3,4}, and Guang-Zhong Yang^{1,2}

¹ The Hamlyn Centre for Robotic Surgery, Imperial College London, London, UK
zhaoyang.wang18@imperial.ac.uk

² Institute of Medical Robotics, Shanghai Jiao Tong University, Shanghai, China

³ Regional Vascular Unit, St Marys Hospital, London, UK

⁴ The Academic Division of Surgery, Imperial College London, London, UK

Abstract. 3D shape instantiation for reconstructing the 3D shape of a target from limited 2D images or projections is an emerging technique for surgical navigation. It bridges the gap between the current 2D intra-operative image acquisition and 3D intra-operative navigation requirement in Minimally Invasive Surgery (MIS). Previously, a general and registration-free framework was proposed for 3D shape instantiation based on Kernel Partial Least Square Regression (KPLSR), requiring manually segmented anatomical structures as the pre-requisite. Two hyper-parameters including the Gaussian width and component number also need to be carefully adjusted. Deep Convolutional Neural Network (DCNN) based framework has also been proposed to reconstruct a 3D point cloud from single 2D image, with end-to-end and fully automatic learning. In this paper, an Instantiation-Net is proposed to reconstruct the 3D mesh of a target from its single 2D image, by using DCNN to extract features from the 2D image and Graph Convolutional Network (GCN) to reconstruct the 3D mesh, and using Fully Connected (FC) layers as the connection. Detailed validation on the Right Ventricle (RV), with a mean 3D distance error of $2.21mm$ on 27 patients, demonstrates the practical strength of the method and its potential clinical use.

Keywords: 3D shape instantiation · Intra-operative 3D navigation · Right Ventricle · Graph Convolutional Neural Network

1 Introduction

Recent advances in Minimally Invasive Surgery (MIS) bring many advantages to patients including reduced access trauma, less bleeding and shorter hospitalization. However, they also impose challenges to intra-operative navigation,

Z.-Y. Wang and X.-Y. Zhou contribute equally to this paper. This work was supported by EPSRC project grant EP/L020688/1.

© Springer Nature Switzerland AG 2020

A. L. Martel et al. (Eds.): MICCAI 2020, LNCS 12264, pp. 680–691, 2020.

https://doi.org/10.1007/978-3-030-59719-1_66

where acquisition of 3D data in real-time is challenging and in most clinical practices, 2D projections or images from fluoroscopy, cross sectional Magnetic Resonance Imaging (MRI) and ultrasound are used. It is difficult to use these 2D images to resolve complex 3D geometries and therefore there is a pressing need to develop real-time techniques to reconstruct 3D structures from limited or even a single 2D projection or image in real-time intra-operatively.

For example, pre-operative 3D context from MRI or Computed Tomography (CT) was registered to intra-operative 2D ultrasound images with both spatial and temporal alignment, which facilitates intra-operative navigation for cardiac MIS [9]. Pre-operative 3D meshes from CT were adapted to intra-operative 2D X-ray images with as-rigid-as-possible method, which acts as an arterial road map for Endovascular Aneurysm Repair [14]. The 3D shape of a stent-graft at three different status: fully-compressed [17], partially-deployed [16] and fully-deployed [18] was instantiated from single 2D fluoroscopic projection with stent-graft modelling, graft gap interpolation, the Robust Perspective-n-Point method, Graph Convolutional Network (GCN), and mesh manipulation, which improves the navigation in Fenestrated Endovascular Aneurysm Repair. A review of bony structures reconstruction from multi-view X-ray images could be found in [7].

Recently, a general and registration-free framework for 3D shape instantiation was proposed [20] with three steps: 1) 3D volumes were pre-operatively scanned for the target at different time frames during the deformation cycle. 3D meshes were segmented and expressed into 3D Statistical Shape Models (SSMs). Sparse Principal Component Analysis (SPCA) was used to analyze the 3D SSMs to determine the most informative and important scan plane; 2) 2D images were scanned synchronously at the determined optimal scan plane. 2D contours were segmented and expressed into 2D SSMs. Kernel Partial Least Square Regression (KPLSR) was applied to learn the relationship between the 2D and 3D SSMs; 3) the KPLSR-learned relationship was applied to the intra-operative 2D SSMs to reconstruct the instantaneous 3D SSMs for navigation. Two deficiencies exist: 1) manual segmentation is essential; 2) two hyper-parameters including the Gaussian width and component number require to be carefully and manually adjusted. To avoid these drawbacks, a one-stage and fully automatic Deep Convolutional Neural Network (DCNN) was proposed to reconstruct the 3D point cloud of a target from its single 2D projection with PointOutNet and Chamfer loss [19]. However, 3D mesh with more details of the surface is more helpful and vital than point cloud.

In this paper, we propose an Instantiation-Net to reconstruct the 3D mesh of a target from its single 2D projection. DenseNet-121 is used to extract abundant features from the 2D image input. Graph Convolutional Network (GCN) is used to reconstruct the 3D mesh. Fully Connected (FC) layers are used as the connection. Figure 1 illustrates the framework for 3D shape instantiation is evolving from two-stage KPLSR-based framework [20], the PointOutNet [19], to the Instantiation-Net proposed in this paper. 27 Right Ventricles (RVs), indicating 609 experiments, were used for validation. An average 3D distance error around $2mm$ was achieved, which is comparable to the performance in [20] but with end-to-end and fully automatic training.

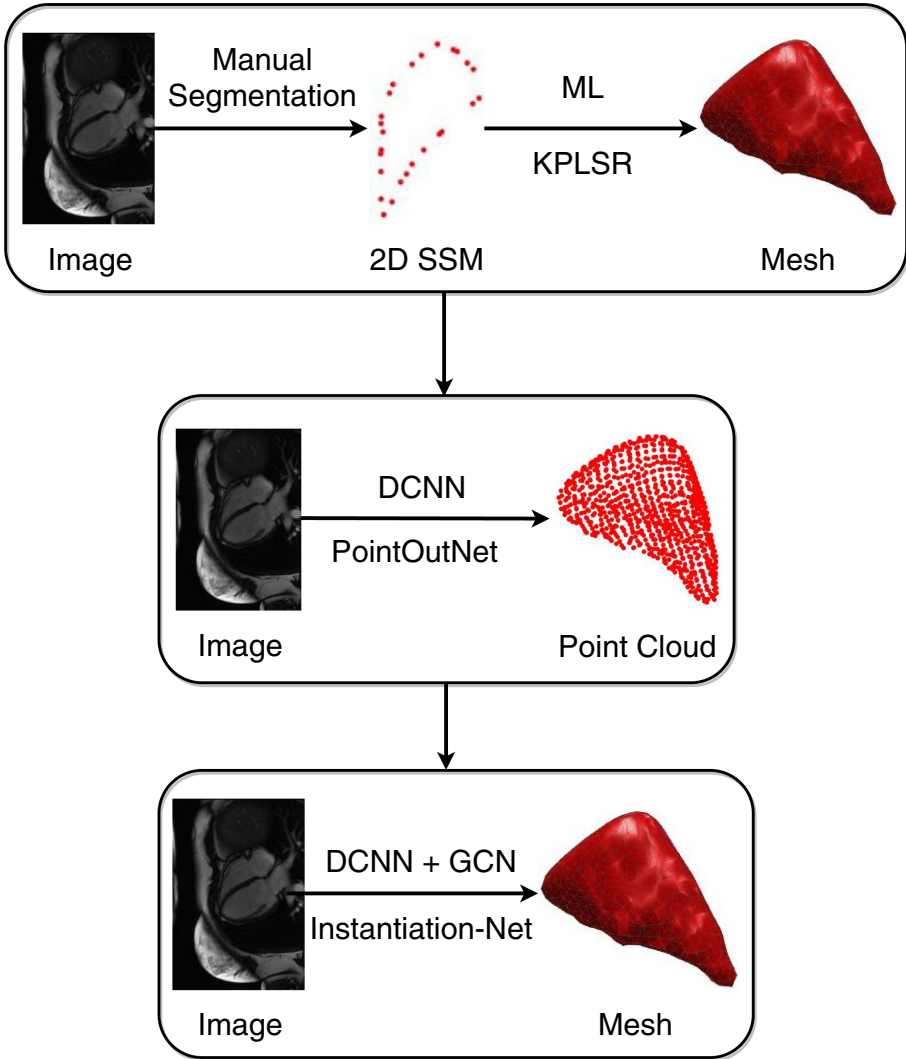


Fig. 1. An illustration of the evolution of 3D shape instantiation from two-stage approach based on KPLSR [20], PointOutNet based on deep learning [19] to the Instantiation-Net that could reconstruct 3D mesh from 2D image in end-to-end fashion.

2 Methodology

The input of Instantiation-Net is a single 2D image I with a size of 192×256 while the output is a 3D mesh \mathcal{F} with vertex V and the connectivity A . Three parts including DCNN, FC and GCN consist of the proposed Instantiation-Net.

2.1 DCNN

For an image input $I_{N \times H \times W \times C}$, where N is the batch size and is fixed at 1 in this paper, H is the image height, W is the image width, C is the image channel and is 1 for medical images, multiple convolutional layers, batch normalization layers, average-pooling layers and ReLU layers¹ consist of the first part of Instantiation-Net - DenseNet-121 [8] for extracting abundant features from the single 2D image input. Detailed layer configurations are shown in Fig. 2.

DCNN		FC + GCN	
Layers	Configuration	Layers	Configuration
Convolution	7×7 conv, stride 2	Fully Connected	$[1,6,8,1024] \rightarrow 8 \rightarrow (M/4^4) \times 64$
Pooling	3×3 max pool, stride 2	Up-Sampling	$(M/4^4) \times 64 \rightarrow (M/4^3) \times 64$
Dense Block (1)	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$, stride 1	Convolution	$(M/4^3) \times 64 \rightarrow (M/4^3) \times 64$
Transition Layer (1)	1×1 conv 2×2 average pool, stride 2	Up-Sampling	$(M/4^3) \times 64 \rightarrow (M/4^2) \times 64$
Dense Block (2)	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$, stride 1	Convolution	$(M/4^2) \times 64 \rightarrow (M/4^2) \times 64$
Transition Layer (2)	1×1 conv 2×2 average pool, stride 2	Up-Sampling	$(M/4^2) \times 64 \rightarrow (M/4) \times 64$
Dense Block (3)	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$, stride 1	Convolution	$(M/4) \times 64 \rightarrow (M/4) \times 64$
Transition Layer (3)	1×1 conv 2×2 average pool, stride 2	Up-Sampling	$(M/4) \times 64 \rightarrow M \times 64$
Dense Block (4)	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$, stride 1	Convolution	$M \times 64 \rightarrow M \times 3$

Fig. 2. Detailed layer configurations of the proposed Instantiation-Net.

2.2 GCN

For a 3D mesh \mathcal{F} with vertex of $\mathbf{V}_{M \times 3}$ and connectivity of $\mathbf{A}_{M \times M}$, where M is the number of vertex in the mesh, $\mathbf{A}_{M \times M}$ is the adjacency matrix with, $\mathbf{A}_{ij} = 1$ if the i th and j th vertex are connected by an edge, otherwise $\mathbf{A}_{ij} = 0$. The non-normalized graph Laplacian matrix is calculated $\mathbf{L} = \mathbf{D} - \mathbf{A}$, where $\mathbf{D}_{ii} = \sum_{j=1}^M \mathbf{A}_{ij}$, $\mathbf{D}_{ij} = 0$, if $i \neq j$, is the vertex degree matrix. For achieving Fourier transform on the mesh vertex, L is decomposed into Fourier basis as $\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, where \mathbf{U} is the matrix of eigen-vectors and $\mathbf{\Lambda}$ is the matrix of eigen-values. The Fourier transform on the vertex v is then formulated as $v_w = \mathbf{U}^T v$, while the inverse Fourier transform is formulated as $v = \mathbf{U}^T v_w$. The convolution in spatial domain of the vertex v and the kernel s can be inversely transformed from the spectral domain as $v * s = \mathbf{U}((\mathbf{U}^T v) \odot (\mathbf{U}^T s))$, where s is the convolutional filter.

¹ For more details of these layers, please refer to [10, 12].

However, this computation is very expensive as it involves matrix multiplication. Hence Chebyshev polynomial is used to reformulate the computation with a kernel g_θ :

$$g_\theta(\mathbf{L}) = \sum_{k=0}^{K-1} \theta_k T_k(\tilde{\mathbf{L}}) \quad (1)$$

where $\tilde{\mathbf{L}} = 2\mathbf{L}/\Lambda_{max} - \mathbf{I}_n$ is the scaled Laplacian, θ is the Chebyshev coefficient. T_k is the Chebyshev polynomial and is recursively calculated as [2]:

$$\mathbf{T}_k(\tilde{\mathbf{L}}) = 2\tilde{\mathbf{L}}\mathbf{T}_{k-1}(\tilde{\mathbf{L}}) - \mathbf{T}_{k-2}(\tilde{\mathbf{L}}) \quad (2)$$

where $\mathbf{T}_0 = 1$, $\mathbf{T}_1 = \tilde{\mathbf{L}}$. The spectral convolution is then defined as:

$$y_j = v * s = \sum_{i=1}^{F_{in}} g_{\theta_{i,j}}(\mathbf{L})v \quad (3)$$

where F_{in} is the feature channel number of the input V , $j \in (1, F_{out})$, F_{out} is the feature channel number of the output Y . Each convolutional layer has $F_{in} \times F_{out} \times K$ trainable parameters.

Except graph convolutional layers, up-sampling layers are also applied to learn the hierarchy mesh structures. First, the mesh \mathcal{F} is down-sampled or simplified to a simplified mesh with $M//S$ vertices, where S is the stride, and is set as 4 or 3 in this paper. Several mesh simplification algorithms can be used in this stage, such as Quadric Error Metrics [13], and weighted Quadric Error Metrics Simplification (QEMS) [5]. The connectivity of the simplified meshes is recorded and used to calculate \mathbf{L} for the graph convolution at different resolutions. The discarded vertexes during the mesh simplification are projected back to the nearest triangle, with the projected position computed with the barycentric coordinates. More details regarding the down-sampling, up-sampling and graph convolutional layers can be found in [13].

2.3 Instantiation-Net

For the DCNN part, DenseNet-121 from [8] is imported from Keras, with parameters pre-trained on ImageNet [3]. For the FC part, two FC layers with an output feature dimension of 8 are used. For the GCN part, four up-sampling and graph convolutional layers are adopted [13]. Detailed configurations of each layer are shown in Fig. 2. An intuitive illustration of the example Instantiation-Net with compacting multiple layers into blocks is shown in Fig. 3. The input is generated by tiling the 2D MRI image three times along the channel dimension. A 3D mesh can be reconstructed directly from the single 2D image input by the proposed Instantiation-Net in a fully automatic and end-to-end learning fashion.

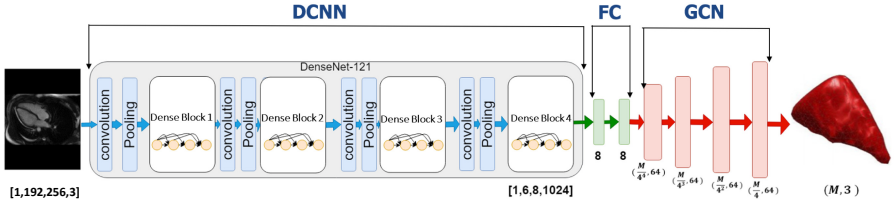


Fig. 3. An intuitive illustration of the proposed Instantiation-Net with compacting multiple layers into blocks.

2.4 Experimental Setup

The data used are the same as [19,20]. Following [19,20], Instantiation-Net was trained patient-specifically with leave-one-frame-out cross-validation: one time frame in the patient was used in the test set, while all other time frames were used as the training set. Stochastic Gradient Descent (SGD) was used as the optimizer with a momentum of 0.9, while each experiment was trained up to 1200 epochs. The initial learning rate was $5e^{-3}$ and decayed with 0.97 every $5 \times M$ iterations, where M is the number of time frame for each patient. The kernel size of GCN was 3. For most experiments, the feature channel and stride size of GCN were 64 and 4 respectively, except that some experiments used 16 and 3 instead. The proposed framework was implemented on Tensorflow and Keras functions. L1 loss was used as the loss function, because L2 loss experienced convergence difficulty in our experiments. The average value of the 3D distance errors of all the vertices is used as the evaluation metric.

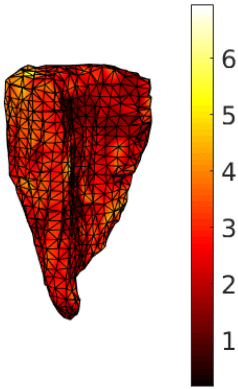
3 Results

To prove the stability and robustness of the proposed Instantiation-Net to each vertex inside a mesh and to each time frame inside a patient, the 3D distance error for each vertex of four meshes and 3D distance error for each time frame of 12 patients are shown in Sect. 3.1 and Sect. 3.2 respectively. To validate the performance of the proposed Instantiation-Net, the PLSR-based and KPLSR-based 3D shape instantiation in [20] are adopted as the baseline in Sect. 3.3.

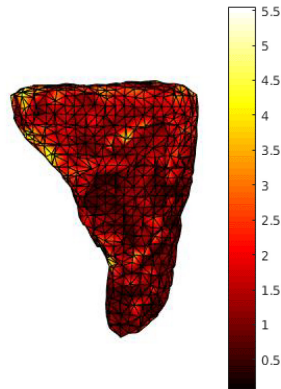
3.1 3D Distance Error for a Mesh

Four reconstructed meshes were selected randomly, showing the 3D distance error of each vertex in colors in Fig. 4. It can be observed that the error is distributed equally on each vertex and does not concentrate or cluster on one specific area. High errors appear at the top of the RV, which is normal, as the vertex number at the RV mesh top is less in the ground-truth than other areas.

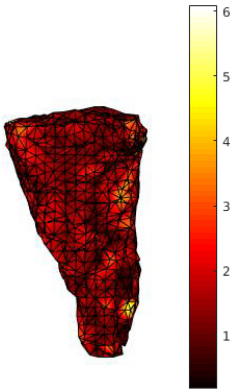
Subject 4 Time Frame 2
(3D distance error 3.44mm)



Subject 8 Time Frame 6
(3D distance error 1.19mm)



Subject 18 Time Frame 15
(3D distance error 2.35mm)



Subject 27 Time Frame 22
(3D distance error 2.31mm)

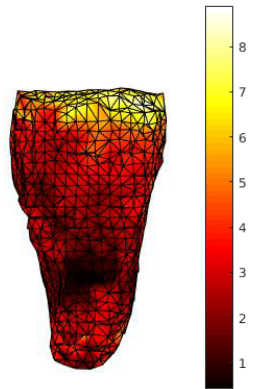


Fig. 4. The 3D distance error of each vertex of four randomly-selected meshes. The color bar is in a unit of *mm*.

3.2 3D Distance Error for a Patient

Figure 5 illustrates the 3D distance errors of each time frame of 12 subjects selected randomly. We can see that, for most time frames, the 3D distance errors are around $2mm$. High errors appear at some time frames, i.e. the time frame 1 and 25 of subject 7, the time frame 11 of subject 9, the time frame 9 of subject 5, the time frame 18 and 20 of subject 15, the time frame 13 of subject 26. This phenomenon was also observed in [19,20], which is the boundary effect. At systole or diastole of the cardiac cycle, the shape of cardiac reaches its smallest or largest size, resulting in extreme cases of 3D mesh compared with other time frames. In the cross validation, if these extreme time frames are not seen in the training data, but are tested, the accuracy of the prediction will be lower.

3.3 Comparison to Other Methods

Figure 6 shows the comparison of the reconstruction performance among the proposed Instantiation-Net, PLSR- and KPLSR-based 3D shape instantiation methods on the 27 subjects. These were evaluated by the mean of 3D distance errors across all time frames. We can see that the proposed Instantiation-Net out-performs PLSR-based 3D shape instantiation while under-performs KPLSR-based 3D shape instantiation slightly for most patients. The overall mean 3D distance error of the mesh generated by Instantiation-Net, PLSR-based and KPLSR-based 3D shape instantiation are $2.21mm$, $2.38mm$ and $2.01mm$ respectively. In addition, the performance of the proposed Instantiation-Net is robust across patients, no obvious outliers are observed.

All experiments were performed with a CPU of Intel Xeon[®] E5-1650 v4 and a GPU of Nvidia Titan Xp. The GPU memory consuming was around 11G which is larger than the 4G consumed by PointOutNet in [19], while PLSR-based and KPLSR-based method in [20] were trained on a CPU. The training time was around 1 h for one time frame which is longer than the 30mins of the PointOutNet in [19], while PLSR-based and KPLSR-based method in [20] took a few minutes. However, the inference of the end-to-end Instantiation-Net only took 0.5 seconds to generate a 3D mesh automatically, while KPLSR-based 3D shape instantiation needs manual segmentation.

4 Discussion

Due to limited coverage of the MRI images at the atrioventricular ring, less vertexes and sparse connectivity exist at the top of the 3D RV mesh, resulting in a higher error in this area, as shown in the right example in Fig. 4. In practical applications, the training data will cover all time frames pre-operatively, which can eliminate the boundary effect shown in Sect. 3.1.

DCNN has a powerful ability for feature extraction from images while GCN has a powerful ability for mesh deformation with both vertex deformation and connectivity maintenance. This paper integrates these two strong networks to achieve 3D mesh reconstruction from single 2D image, which crosses modalities. Based on the author's knowledge, this is one of the few pioneering works that achieve direct 3D mesh reconstruction from 2D images with end-to-end training. In medical computer vision, this is the first work that achieves 3D mesh reconstruction from single 2D image in an end-to-end and fully automatic training.

Apart from the baselines in this paper, there are also other works working on similar tasks, i.e. [1, 4, 6, 11, 15, 19]. However, 3D occupancy grid is reconstructed in [1], point cloud is reconstructed in [4, 11, 19] and 3D volume is reconstructed in [6, 15]. 3D occupancy grid, point cloud and 3D volume are different 3D data modalities compared to the 3D mesh reconstructed in this paper, hence it is difficult to conduct a fair comparison with them. In addition, two orthogonal X-rays are needed for a 3D volume reconstruction in [15] which can not work on a single image input in this paper.

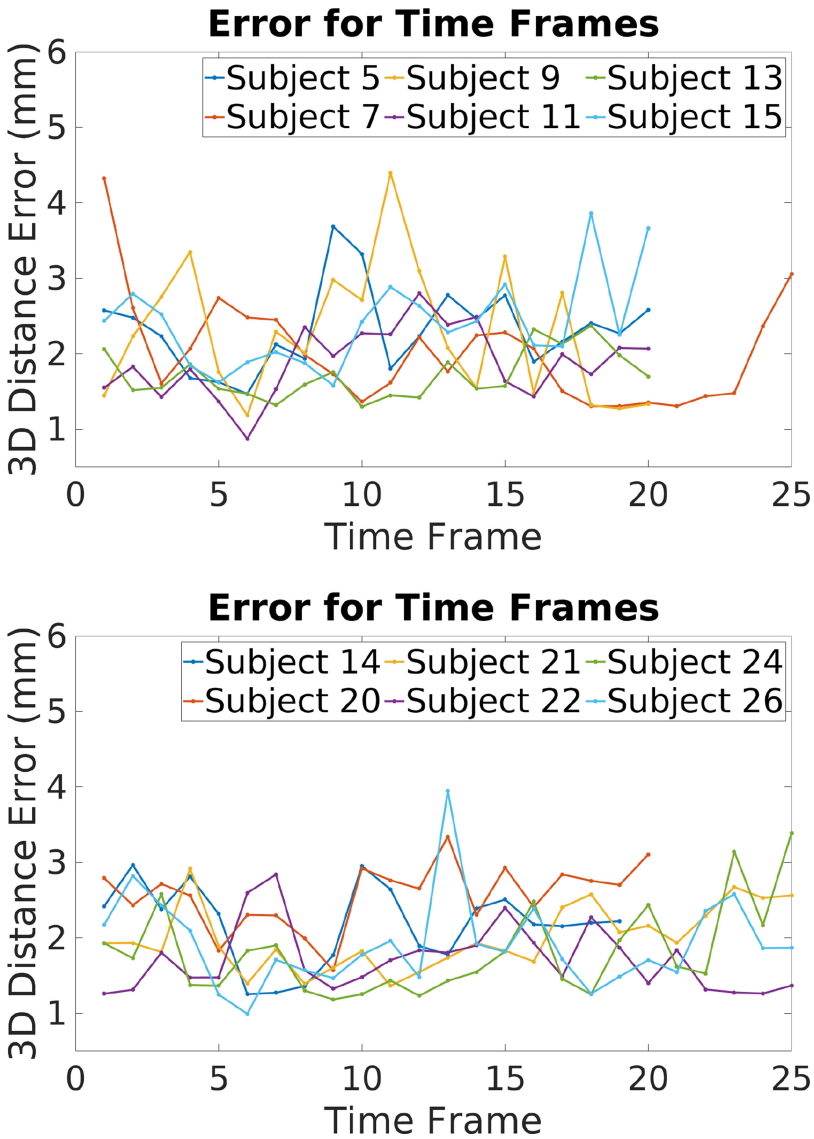


Fig. 5. The mean 3D distance errors of each time frame of 12 randomly selected patients.

One potential drawback of the proposed Instantiation-Net is that it requires both the larger consumption in GPU memory and the longer training time than that of the PointOutNet in [20] and the PLSR-based and KPLSR-based 3D shape instantiation in [19], but, the inference is quick and fully automatic.

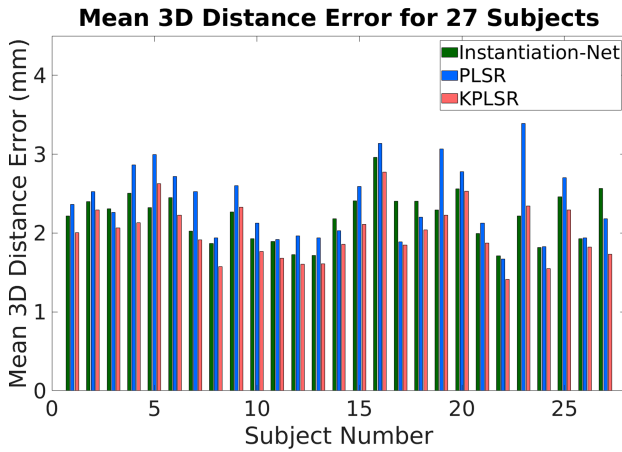


Fig. 6. The mean 3D distance errors of the mesh of 27 subjects generated by the proposed Instantiation-Net, PLSR- and KPLSR-based 3D shape instantiation.

5 Conclusion

In this paper, an end-to-end framework, called Instantiation-Net, was proposed to instantiate the 3D mesh of RV from its single 2D MRI image. DCNN is used to extract the feature map from 2D image, which is connected with 3D mesh reconstruction part based on GCN via FC layers. The results on 609 experiments showed that the proposed network could achieve higher/slightly lower accuracy in 3D mesh than PLSR-based/KPLSR-based 3D shape instantiation in [20]. According to the result, one-stage shape instantiation directly from 2D image to 3D mesh can be achieved by the proposed Instantiation-Net, obtaining comparable performance with the two baseline methods.

We believe that the combination of DCNN and GCN will be very useful in the medical area, as it bridges the gap between the image and mesh modality. In the future, we will work on extending the proposed Instantiation-Net to broader applications, i.e., reconstructing 3D meshes directly from 3D volumes.

References

1. Choy, C.B., Xu, D., Gwak, J.Y., Chen, K., Savarese, S.: 3D-R2N2: A unified approach for single and multi-view 3D object reconstruction. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9912, pp. 628–644. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46484-8_38
2. Defferrard, M., Bresson, X., Vandergheynst, P.: Convolutional neural networks on graphs with fast localized spectral filtering. In: Advances in Neural Information Processing Systems, pp. 3844–3852 (2016)
3. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. IEEE (2009)

4. Fan, H., Su, H., Guibas, L.J.: A point set generation network for 3d object reconstruction from a single image. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 605–613 (2017)
5. Garland, M., Heckbert, P.S.: Surface simplification using quadric error metrics. In: Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, pp. 209–216 (1997)
6. Henzler, P., Rasche, V., Ropinski, T., Ritschel, T.: Single-image tomography: 3d volumes from 2d cranial x-rays. In: Computer Graphics Forum, vol. 37, pp. 377–388. Wiley Online Library (2018)
7. Hosseinian, S., Arefi, H.: 3d reconstruction from multi-view medical x-ray images—review and evaluation of existing methods. *Int. Archives Photogrammetry Remote Sensing Spatial Inf. Sci.* 40 (2015)
8. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708 (2017)
9. Huang, X., Moore, J., Guiraudon, G., Jones, D.L., Bainbridge, D., Ren, J., Peters, T.M.: Dynamic 2d ultrasound and 3d ct image registration of the beating heart. *IEEE Trans. Med. Imaging* 28(8), 1179–1189 (2009)
10. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: ICML, pp. 448–456 (2015). <http://proceedings.mlr.press/v37/ioffe15.html>
11. Jiang, L., Shi, S., Qi, X., Jia, J.: Gal: Geometric adversarial loss for single-view 3d-object reconstruction. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 802–816 (2018)
12. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
13. Ranjan, A., Bolkart, T., Sanyal, S., Black, M.J.: Generating 3d faces using convolutional mesh autoencoders. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 704–720 (2018)
14. Toth, D., Pfister, M., Maier, A., Kowarschik, M., Hornegger, J.: Adaption of 3D models to 2D X-ray images during endovascular abdominal aneurysm repair. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9349, pp. 339–346. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24553-9_42
15. Ying, X., Guo, H., Ma, K., Wu, J., Weng, Z., Zheng, Y.: X2CT-GAN: reconstructing CT from biplanar x-rays with generative adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 10619–10628 (2019)
16. Zheng, J.Q., Zhou, X.Y., Riga, C., Yang, G.Z.: Real-time 3D shape instantiation for partially deployed stent segments from a single 2-d fluoroscopic image in fenestrated endovascular aortic repair. *IEEE Robot. Autom. Lett.* 4(4), 3703–3710 (2019)
17. Zhou, X., Yang, G., Riga, C., Lee, S.: Stent graft shape instantiation for fenestrated endovascular aortic repair. In: Proceedings of the The Hamlyn Symposium on Medical Robotics. The Hamlyn Symposium on Medical Robotics (2016)
18. Zhou, X.Y., Lin, J., Riga, C., Yang, G.Z., Lee, S.L.: Real-time 3D shape instantiation from single fluoroscopy projection for fenestrated stent graft deployment. *IEEE Robot. Autom. Lett.* 3(2), 1314–1321 (2018)

19. Zhou, X.-Y., Wang, Z.-Y., Li, P., Zheng, J.-Q., Yang, G.-Z.: One-stage shape instantiation from a single 2D Image to 3D point cloud. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11767, pp. 30–38. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32251-9_4
20. Zhou, X.Y., Yang, G.Z., Lee, S.L.: A real-time and registration-free framework for dynamic shape instantiation. *Med. Image Anal.* **44**, 86–97 (2018)