Europe PMC Funders Group
Author Manuscript
Simpl Med Ultrasound (2022). Author manuscript; available in PMC 2024 September 19.

Published in final edited form as:
Simpl Med Ultrasound (2022). 2022 ; 13565: 86–95. doi:10.1007/978-3-031-16902-1_9.

Efficient Pix2Vox++ for 3D Cardiac Reconstruction from 2D echo views

David Stojanovski[1,*], **Uxio Hermida**[1], **Marica Muffoletto**[1], **Pablo Lamata**[1], **Arian Beqiri**[1,2], **Alberto Gomez**[1,2]

David Stojanovski: david.stojanovski@kcl.ac.uk; Uxio Hermida: uxio.hermida@kcl.ac.uk; Marica Muffoletto: marica.muffoletto@kcl.ac.uk; Pablo Lamata: pablo.lamata@kcl.ac.uk; Arian Beqiri: arian.beqiri@ultromics.com; Alberto Gomez: alberto.gomez@ultromics.com

[1]King's College London, School of Biomedical Engineering & Imaging Sciences, London, SE1 7EU, UK.

[2]Ultromics Ltd., Oxford, OX4 2SU, UK

Abstract

Accurate geometric quantification of the human heart is a key step in the diagnosis of numerous cardiac diseases, and in the management of cardiac patients. Ultrasound imaging is the primary modality for cardiac imaging, however acquisition requires high operator skill, and its interpretation and analysis is difficult due to artifacts. Reconstructing cardiac anatomy in 3D can enable discovery of new biomarkers and make imaging less dependent on operator expertise, however most ultrasound systems only have 2D imaging capabilities. We propose both a simple alteration to the Pix2Vox++ networks for a sizeable reduction in memory usage and computational complexity, and a pipeline to perform reconstruction of 3D anatomy from 2D standard cardiac views, effectively enabling 3D anatomical reconstruction from limited 2D data. We evaluate our pipeline using synthetically generated data achieving accurate 3D whole-heart reconstructions (peak intersection over union score > 0.88) from just two standard anatomical 2D views of the heart. We also show preliminary results using real echo images.

Keywords

Convolutional Neural Networks; 2D to 3D reconstruction; Deep learning; Ultrasound

1 Introduction

1.1 Motivation and background

The most common imaging modality for cardiac assessment worldwide is ultra-sound (US) [2] because it is more affordable and safer, and it has higher temporal resolution when

This work is licensed under a BY 4.0 International license.

*This work was supported by the Wellcome/EPSRC Centre for Medical Engineering [WT203148/Z/16/Z], by the British Heart Foundation [TG/17/3/33406] and the National Institute for Health Research (NIHR) Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and King's College London. Pablo Lamata holds a Wellcome Trust Senior Research Fellowship [209450/Z/17/Z]. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

compared to other modalities. The reliance on operator skill to acquire high quality standard views is the main limitation of US, particularly for cardiac applications (echocardiography).

Most US systems only have 2D capabilities, and most echo protocols are limited to 2D modes [13]. 3D data is desirable both for anatomical and functional understanding of the heart. The heart is a complex 3D structure, and subsequently most features cannot be fully captured within 2D planes. Ventricular and atrial walls are curved surfaces, valve hinges do not sit on a 2D circle but rather a 3D saddle shape, and as a result current protocols require many standard planes to be able to assess most features, while a high quality 3D reconstruction would capture them all. Cardiac function is quantified primarily via tissue motion and blood flow. Blood quantity is volumetric, but usually 2D approximations are used, e.g. 2D left ventricular delineation on a 4 chamber view to compute ejection fraction. This has been commonly used both in classical clinical disease detection and deep learning methods [10, 15]. Tissue motion is analysed using 2D components of the motion e.g. longitudinal, radial, etc., however cardiac motion is complex and combines torsion, vertical motion and displacement, which are fundamentally 3D phenomena. Capturing only 2D views can impede further analysis if the quality of the views is suboptimal from factors such as foreshortening and artifacts.

3D volumes would allow for more accurate quantification of the cardiac geometries, and in turn allow clinicians to reduce the rate of disease misdiagnosis as a result of the incorrect quantification arising from using 2D data. 3D ultrasound probes are available for clinical use. They are however very prone to artifacting, particularly in moving anatomies such as the heart, and as a result are very rarely used in practice [9].

The ability to quantify cardiac function using standard 2D US systems, without the need for external trackers to define a spatial geometry, can be of great clinical use, but is limited by lack of paired US and native 3D ground truth anatomy. The aim of this work is to develop a method for reconstructing a full 3D representation of the heart from 2D image views.

1.2 Related works

The advent of Deep Learning (DL) has led to development of 3D image reconstruction methods that have been shown to successfully learn shape and structure from partial observations, achieving start-of-the-art (SOTA) performance in natural and medical imaging.

2D to 3D reconstruction in natural images: Current SOTA methods using DL allow accurate 2D to 3D reconstruction of objects from one or more RGB images, without the aid of camera positioning calibration information. A popular standardized dataset used to perform comparisons of reconstruction techniques is ShapeNetCore [4]. ShapeNetCore covers 55 object categories with over 51,300 unique 3D models. In 2019, ShapeNetCore reconstruction SOTA accuracy was achieved by Xie et al. who were able to reach 0.706 Intersection over Union (IoU) using the Pix2Vox++ (PiVox) network [16]. Two network variations, Pix2Vox++/Fast (PiVox/Fast) and Pix2Vox++/Accurate, or (PiVox/Accurate), were proposed with the PiVox/Fast network being a lighter weight, albeit lower performing, version of the full PiVox/Accurate network. However, PiVox/Fast is still an expensive model

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

to train and perform inference, thus severely limiting the possible reconstruction resolution on currently available single Graphics Processing Units (GPUs).

Ultrasound specific 2D to 3D reconstruction: A number of US-specific 3D reconstruction algorithms have been developed showing effective performance but have not been applied to cardiac reconstruction. Cerrolaza et al. [1] performed fetal skull reconstruction using conditional hierarchical generative network Variational Auto-Encoders (VAEs) and achieved a Dice Coefficient (DC) of 0.91 using three orthogonal US views. In contrast to our goal of performing cardiac reconstructions, their reconstruction target had a generally regular shape which does not deform during imaging.

Prevost et al. [12] combine DL with an Inertial Measurement Unit (IMU), using two consecutive frames with optical flow as a channel to the network along with the Euler angles provided by the IMU. On a phantom they achieved a minimum, medium and maximum drift of 1.70, 18.30 and 36.90 mm respectively.

1.3 Contributions of this study

No previous work has addressed 3D cardiac shape reconstructions from untracked 2D echo views. As a first step towards this goal, we present an exploratory investigation using synthetic echo data: 1) A pipeline to synthesize 2D echo views with corresponding 3D ground truth; 2) A demonstration that PiVox can produce accurate 2D to 3D reconstructions of realistic synthetic hearts; 3) A simple modification to both PiVox networks which vastly reduces memory and computational expense, while still achieving high reconstruction accuracy.

2 Methods

To explore reconstruction of a 3D heart from 2D data, we used two types of synthetic data: 1) A segmentation dataset, containing binary tissue masks (segmentations) on 2D standard views. Masks are simulated by slicing 3D computational models (which allows using 3D ground truths for evaluation); 2) A synthetic US dataset, containing synthetic 2D standard US views, generated from the tissue masks. We used both for model training and testing. An overview of the proposed pipeline is shown in Figure 1.

2.1 Synthetic 2D segmentations from 3D heart models

We obtain synthetic 2D segmentations by slicing 3D mesh models of the heart at standard echocardiographic views. These will be used to asses both how the 2D to 3D reconstruction networks are affected by the large amounts of variation that will be present in real ultrasound images, and also to provide insight into the feasibility of performing training on synthetic data and testing on real ultrasound data. If it can be shown that training using just synthetic data and testing on real data is feasible then it would allow generation of large training datasets and greatly reduce the amount of paired 3D CT/2D ultrasound data required to validate methodologies.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Our slice extraction technique utilizes a combination of the Visualization Toolkit (VTK) [8] and the PyVista Python packages [3]. The mesh data used was a set of 1000 synthetic cardiac meshes created by Rodero et al. [14].

The first step in extracting the standard plane cardiac views requires defining either 3 landmarks for each view to which a plane can be fit, or 2 landmarks and a projection along a vector. We defined these automatically-calculated landmarks based on their relation to various cardiac structures present in the meshes/segmentations (cf. Table 1). A trained sonographer examined 10 sets of segmentations to confirm the chosen landmarks resulted in suitably realistic extracted slices from our synthetic meshes.

The left ventricular apex (LVA) was found by performing a ray casting from the centre of mass of the mitral valve to all mesh cell faces in the LV mesh. If there was a pair of intersection points along the ray line (i.e. the ray penetrates both the endo- and epicardial wall) the distance between these intersection points was calculated. The shortest distance was chosen as the LVA, to find the thinnest location of the cardiac wall and to minimize apical foreshortening.

2.2 Synthetic Ultrasound images

Synthetic Ultrasound images (i.e. with real appearance) are generated from the same 3D heart models to study feasibility in realistic echo data while having an exact ground truth. We adapted the technique proposed by Gilbert et al. [5] using the CAMUS dataset [7]. In brief, pseudo-images, i.e. tissue masks corresponding to 2 and 4 chamber images, with noise and Gaussian blurring are used as input to a CycleGAN network [17] with unpaired US images to create the final synthetic images, as exemplified in Fig. 2.

Producing high quality synthetic US images required small variations in 1) noise parameters 2) sequence of additive noise and blurring, and 3) size of the Gaussian blur kernel. These parameters were changed due to a different resolution input image, and our requirement to not perform any geometric transforms, e.g. changing of input anatomy dimensions.

2.3 Efficient Pix2Vox++

The general Pix2Vox++ (PiVox) network architecture is composed of a series of parallel encoder and decoder branches for each input view, before being passed into the fusion and refiner modules. The decoder, in particular, is memory and computationally expensive due to the 3D deconvolutional kernels. This greatly limits the resolution of the 3D volumes that can be used, thus necessitating a more efficient network with high reconstruction accuracy. It also contains a very large number of learnable parameters, in turn requiring a very large training set.

Xie et al. proposed a lighter weight, albeit lower performing, variant of the PiVox/Accurate network, referred to as PiVox/Fast, for reducing memory and computational complexity by 1) Using ResNet-18 instead of ResNet-50 2) Decreased de/convolution kernel sizes 3) Removal of the refiner module.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

We propose a simple adaptation of the PiVox networks, referred to as E-PiVox, to greatly decrease both memory usage and computational expense, with minimal impact to performance. We achieve this by adding a 3D convolution at the end of the encoder module along the input image dimension to reduce the decoder to a single branch for any number of input images. All information is propagated through the parallel decoder branches in the PiVox networks passes through a compressed latent space, in theory allowing for minimal decreases in performance from having just a single decoder branch rather than multiple branches.

2.4 2D to 3D reconstruction of synthetic hearts

Implementation details: The dataset of 1000 cardiac meshes was divided into a 70/15/15% train/validation/test split and remained unchanged for each of the different training runs. The models were trained with binary segmentation masks, the label maps converted to realistic echo-like images using the Cycle-GAN network and also on the ShapeNet dataset. All training was performed using Pytorch 1.9.1 [11] on an Nvidia RTX 3090 for 200 epochs. The Adam optimizer was used with a β_1 of 0.9 and a β_2 of 0.999 [6]. The code is available at https://github.com/david-stojanovski/E-Pix2Vox-reconstruction

3 Experiments and Results

Reconstruction accuracy was assessed using the thresholded Intersection over Union (IoU) [16].

The final reconstruction results of the ShapeNet and cardiac training runs are shown in Tables 2 and 3 respectively. Table 2 shows PiVox/Accurate was consistently the best performing network, however E-PiVox/Fast and E-PiVox/Accurate were consistently within 1.2% and 2% respectively of PiVox/Fast and PiVox/Accurate. This was achieved while being much more memory and computationally efficient.

It can be seen in Table 2 that our E-PiVox networks was able to come very close to the performance of the much more computationally expensive PiVox networks. The relationship between computational expense and memory usage in regard to number of input views is shown in Figure 3 and highlights the importance of efficient network architectures when using a larger number of views, such as in video recordings, or higher resolutions e.g. full resolution CT.

Table 3 shows the results when using both binary and realistic ultrasound CycleGAN inferenced images. As shown in Table 3 both E-PiVox networks generally outperformed the PiVox networks in the relative comparisons.The reconstruction accuracy dropped most in areas where the walls of the heart were quite thin, generally in the atrial region. Small errors appearing over the entire structure appear to arise due to errors in the exact location discretization of the voxels.

We present examples of 3D heart reconstructions in Figure 4, including preliminary results using real 2D echo images from the CAMUS dataset. Synthetically trained reconstructions show close correspondence with the ground truth anatomy (black wireframe). However,

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

reduced number of input views resulted in decreased accuracy and clear non-physiological holes. Such phenomena were exacerbated when using US-like images (see Figure 4C). Larger differences were observed in the LVA and valve planes for all models.

4 Discussion and Conclusions

This work provides a proof of concept that a complex geometry like the human heart can be reconstructed with reasonable accuracy from a limited number of standard 2D anatomical views. And that this can be achieved with efficient training and inference.

Results in Table 3 show that the PiVox network can successfully reconstruct a full heart with a peak IoU of 0.903 for 9 segmented views, and a minimal decrease in performance to 0.881 for just 2 views. The 0.741 IoU achieved with just 2 synthetic US images represents an acceptable performance given the much more challenging input image data that emulates real-world acquisitions.

It is important to note that real US images present great variability in contrast and appearance. The preliminary results using real data (panel D in Figure 4) show reconstructions are possible, albeit needing further improvement. Given a suitably large enough set of real US images, accurate 3D reconstructions could be possible from a small number of standard clinical 2D views.

The keystone for this contribution is the synthesis of training data (i.e. 2D views) with idealized ground truth 3D reconstructions. This approach addresses the limitations in medical data availability, data privacy, and cost of expert annotation. Preliminary results show promise in the use on real data using models trained with synthetic data, but further research and evidence is needed.

The main limitation of our study is that it is based on the anatomical variability of a synthetic cohort of healthy subjects. As such, the performance could dramatically drop in the presence of disease.

In conclusion, this work demonstrates, in a synthetic workbench, the feasibility of 3D cardiac reconstruction from standard 2D views.

References

1. B, JJC, Li, Y, Biffi, C, Gomez, A, Sinclair, M, Matthew, J, Knight, C, Kainz, B, Rueckert, D. 3D Fetal Skull Reconstruction from 2DUS via Deep Conditional Generative Networks. Vol. 1. Springer International Publishing; 2018.
2. Braga JR, Leong-Poi H, Rac VE, Austin PC, Ross HJ, Lee DS. Trends in the Use of Cardiac Imaging for Patients with Heart Failure in Canada. JAMA Network Open. 2019; 2 (8) 1–13. DOI: 10.1001/jamanetworkopen.2019.8766 [PubMed: 31397858]
3. Castro, DdlI; Mader, K; Hanchen, Pandu-rao; Sullivan, B; Cole, J; Wall-baum, M; Mitchell, N; Movva, Sebastian M; McCartney, P; Jadhav, S; de Miguel, YB; , et al. Joskaaaa: daavoo/pyntcloud: v016. 2022; jan. doi: 10.5281/ZENODO.5841822
4. Chang AX, Funkhouser T, Guibas L, Hanrahan P, Huang Q, Li Z, Savarese S, Savva M, Song S, Su H, Xiao J, et al. ShapeNet: An Information-Rich 3D Model Repository. 2015. http://arxiv.org/abs/1512.03012

Europe PMC Funders Author Manuscripts

5. Gilbert A, Marciniak M, Rodero C, Lamata P, Samset E, McLeod K. Generating Synthetic Labeled Data from Existing Anatomical Models: An Example with Echocardiography Segmentation. IEEE Transactions on Medical Imaging. 2021; 40 (10) 2783–2794. DOI: 10.1109/TMI.2021.3051806 [PubMed: 33444134]
6. Kingma, DP; Ba, JL. Adam: A method for stochastic optimization; 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings; 2015. 1–15.
7. Leclerc S, Smistad E, Pedrosa J, Ostvik A, Cervenansky F, Espinosa F, Espeland T, Berg EAR, Jodoin PM, Grenier T, Lartizien C, et al. Deep Learning for Segmentation Using an Open Large-Scale Dataset in 2D Echocardiography. IEEE transactions on medical imaging. 2019; 38 (9) 2198–2210. [PubMed: 30802851]
8. Lowekamp BC, Chen DT, Ibáñez L, Blezek D. The design of simpleITK. Frontiers in Neuroinformatics. 2013; DEC. 7: 1–14. DOI: 10.3389/fninf.2013.00045 [PubMed: 23386828]
9. Nelson TR, Pretorius DH, Hull A, Riccabona M, Sklansky MS, James G. Sources and impact of artifacts on clinical three-dimensional ultrasound imaging. Ultrasound in Obstetrics and Gynecology. 2000; 16 (4) 374–383. [PubMed: 11169316]
10. Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Langlotz CP, Heidenreich PA, Harrington RA, Liang DH, Ashley EA, Zou JY. Video-based AI for beat-to-beat assessment of cardiac function. Nature. 2020; 580 (7802) 252–256. DOI: 10.1038/s41586-020-2145-8 [PubMed: 32269341]
11. Paszke, A, Gross, S, Massa, F, Lerer, A, Bradbury, J, Chanan, G, Killeen, T, Lin, Z, Gimelshein, N, Antiga, L, Desmaison, A. , et al. Advances in Neural Information Processing Systems. Vol. 32. NeurIPS; 2019.
12. Prevost R, Salehi M, Jagoda S, Kumar N, Sprung J, Ladikos A, Bauer R, Zettinig O, Wein W. 3D freehand ultrasound without external tracking using deep learning. Medical Image Analysis. 2018; 48: 187–202. [PubMed: 29936399]
13. Robinson S, Rana B, Oxborough D, Steeds R, Monaghan M, Stout M, Pearce K, Harkness A, Ring L, Paton M, Akhtar W, et al. A practical guideline for performing a comprehensive transthoracic echocardiogram in adults: The British Society of Echocardiography minimum dataset. Echo Research and Practice. 2020; 7 (4) G59–G93. DOI: 10.1530/ERP-20-0026 [PubMed: 33112828]
14. Rodero C, Strocchi M, Marciniak M, Longobardi S, Whitaker J, O'Neill MD, Gillette K, Augustin C, Plank G, Vigmond EJ, Lamata P, et al. Linking statistical shape models and simulated function in the healthy adult human heart. PLoS Computational Biology. 2021; 17 (4) 1–28. DOI: 10.1371/journal.pcbi.1008851 [PubMed: 33857152]
15. Upton R, Mumith A, Beqiri A, Parker A, Hawkes W, Gao S, Porumb M, Sarwar R, Marques P, Markham D, Kenworthy J, et al. Automated Echocardiographic Detection of Severe Coronary Artery Disease Using Artificial Intelligence. JACC: Cardiovascular Imaging. 2022; 1–13. [PubMed: 34922865]
16. Xie, H; Yao, H; Sun, X; Zhou, S; Zhang, S. Pix2Vox: Context-aware 3D reconstruction from single and multi-view images; Proceedings of the IEEE International Conference on Computer Vision; 2019 Octob; 2690–2698.
17. Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. 2017; doi: 10.48550/ARXIV.1703.10593

Europe PMC Funders Author Manuscripts

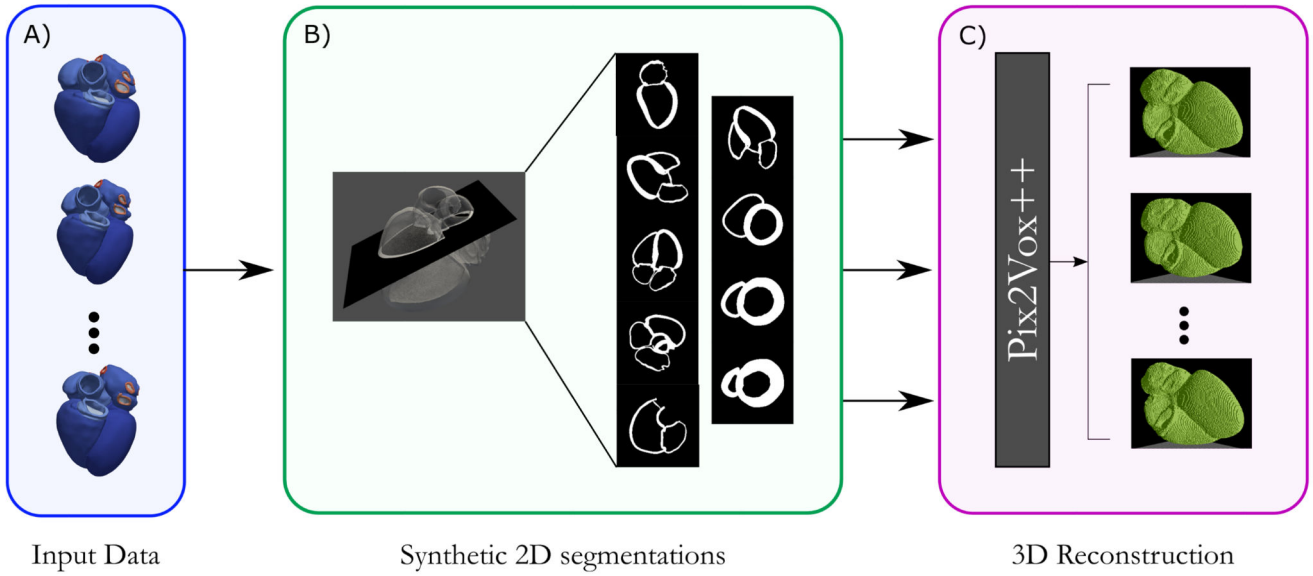


Fig. 1. Proposed pipeline for extracting 2D standard US views from a mesh and performing a reconstruction.

A) Initial heart model with tissue labels, B) Mesh slicing with a plane and extracted segmentation-like images and C) Input images being trained/inferenced in the PiVox networks to produce 3D reconstructions.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

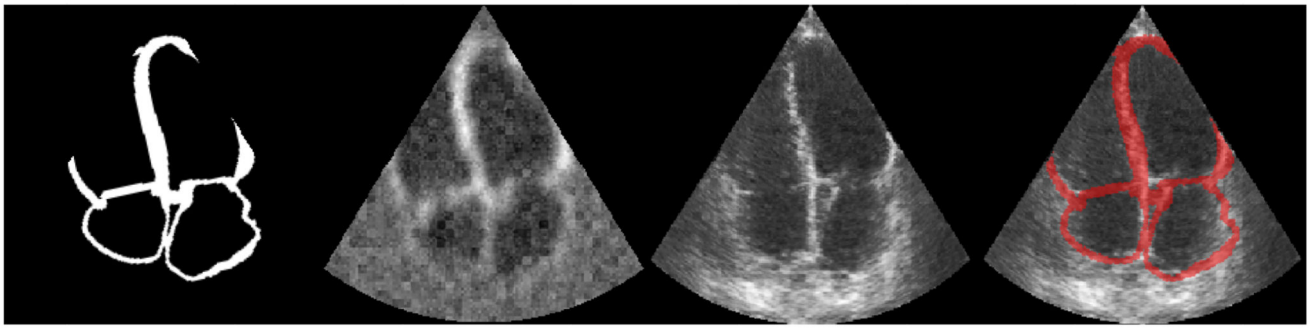


Fig. 2. Example case of data generated at each stage in the process of creating a synthetic ultrasound image from a segmentation mask.

From left to right, a) Initial segmentation image b) generated pseudo image c) generated synthetic US d) overlay of input mask on generated image.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

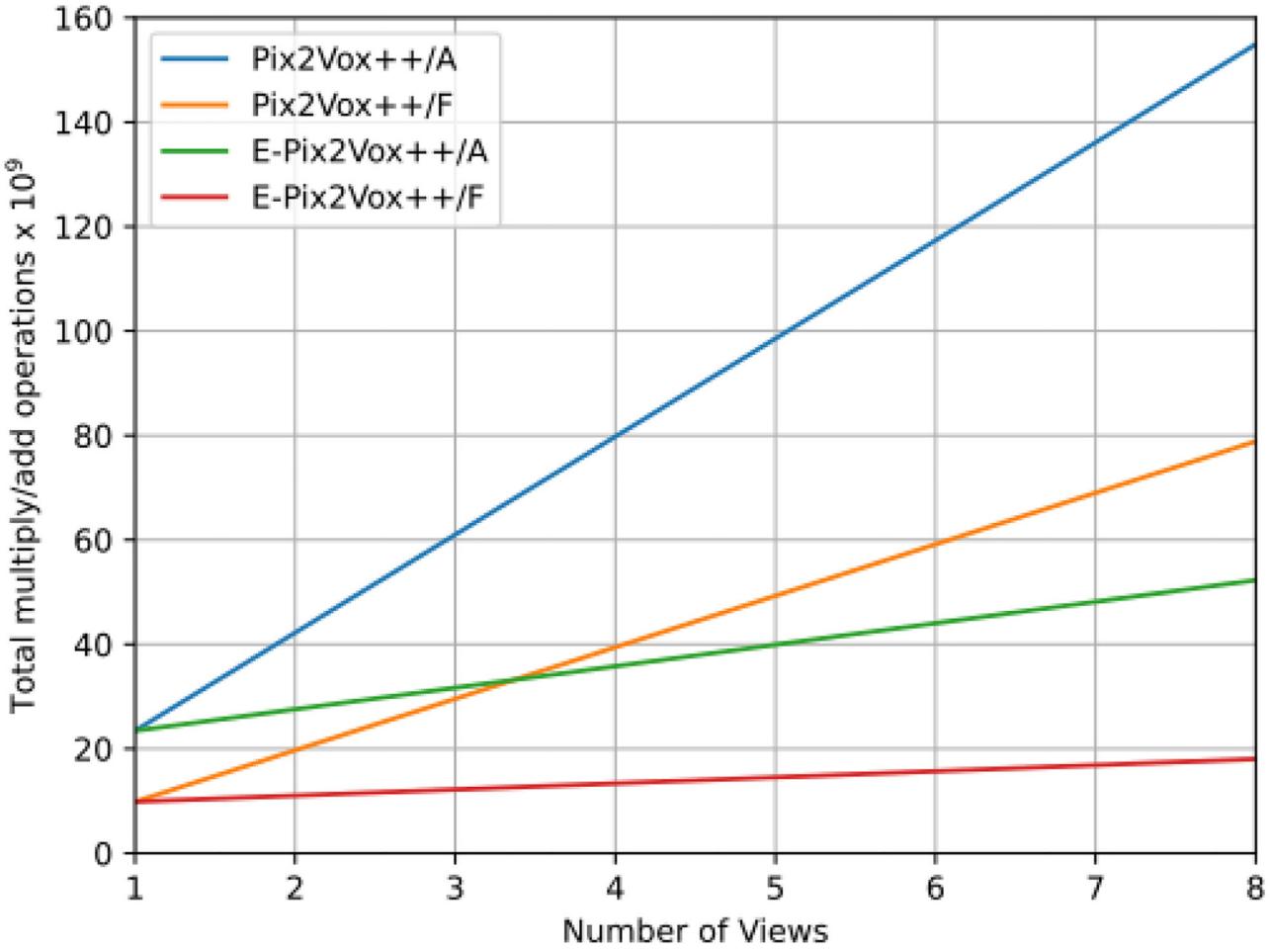


Fig. 3. Comparison of the 4 networks and their theoretical total Multiply/Add operations for each number of input views at a 64^3 resolution.

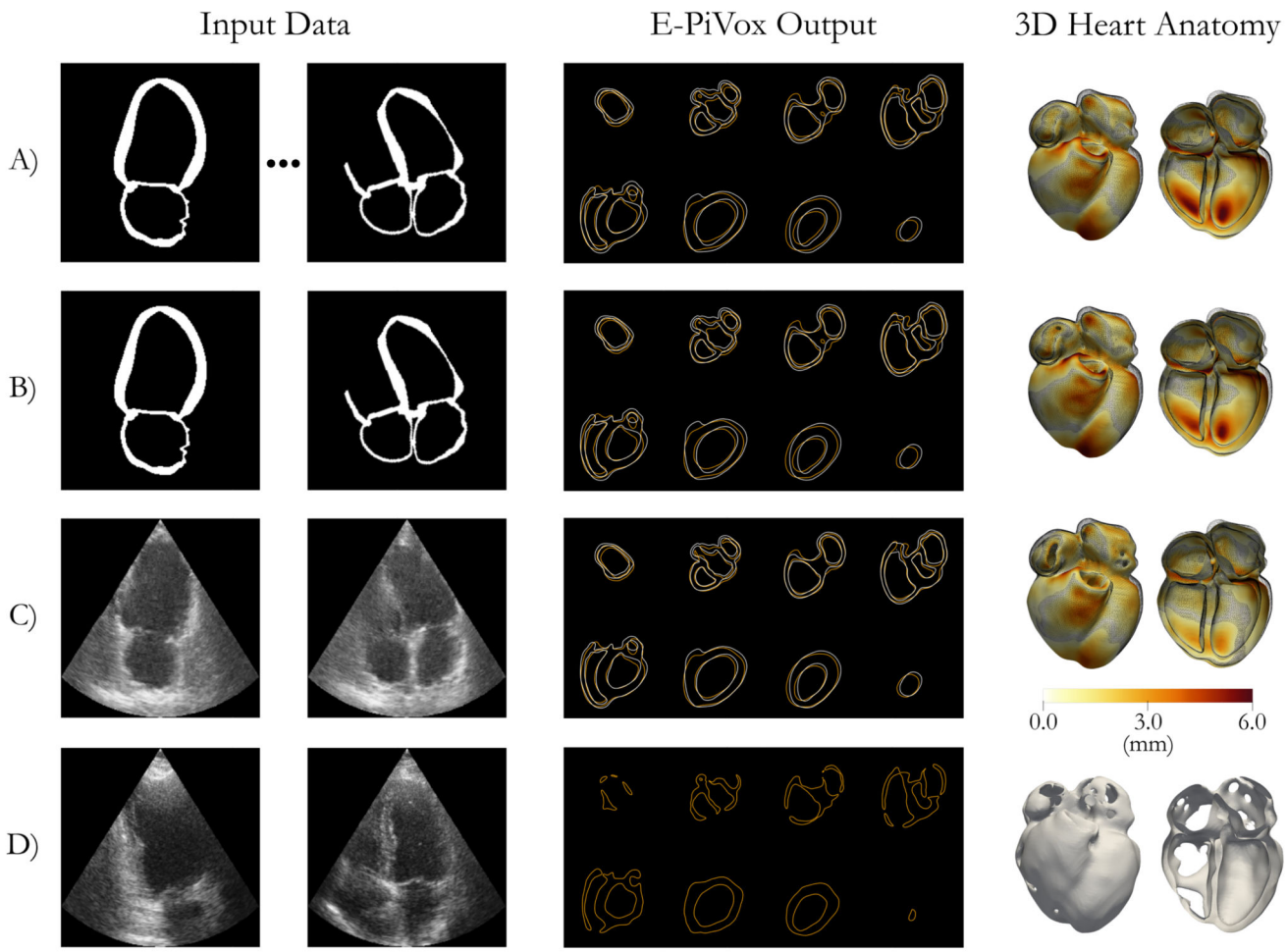


Fig. 4.

A) 9 binary views; B) Binary Apical 2 and 4 chamber views; C) CycleGAN inferenced Apical 2 and 4 chamber views; D) Example case from the CAMUS dataset. The same case is shown in A, B and C. The E-PiVox output column shows the ground truth (White) and the network predictions (Orange). The 3D Heart Anatomy column shows the absolute error in relation to the ground truth mesh (Black wireframe).

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 1 Ultrasound views and the corresponding landmarks to define these planes.

Ultrasound View	Landmarks
Right Ventricular Inflow	Right Ventricle, Right Atrium, Pulmonary Valve
Left Ventricular Parasternal Long-Axis	Left Ventricle Apex, Mitral Valve, Aortic Valve
Parasternal Short-Axis Aortic Valve Level	Left Atrium, Right Atrium, Aortic Valve
Parasternal Short-Axis Mitral Valve Level	Left Ventricle Apex, Whole Heart
Parasternal Short-Axis Papillary Muscle Level	Left Ventricle Apex, Whole Heart
Parasternal Short-Axis Apex Level	Left Ventricle Apex, Whole Heart
Apical 4 Chamber	Left Atrium, Right Atrium, Left Ventricle Apex
Apical 5 Chamber	Left Atrium, Aortic Valve, Left Ventricle Apex
Apical 2 Chamber	Left Ventricle Apex, Mitral Valve, Right Ventricle

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 2 IoU Comparison of multi-view 2D to 3D object reconstruction on the ShapeNet data set for a 32^3 resolution.

Method	1 View	2 Views	4 Views	8 Views
PiVox/Fast	0.645	0.669	0.682	0.690
E-PiVox/Fast	0.641	0.663	0.675	0.682
Percentage diff.(%)	-0.620	-0.897	-1.026	-1.16
PiVox/Accurate	0.670	0.695	0.708	0.715
E-PiVox/Accurate	0.6687	0.6836	0.6949	0.701
Percentage diff.(%)	-0.19	-1.64	-1.85	-1.96

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 3 IoU Comparison of multi-view 2D to 3D object reconstruction on the cardiac data set for a 64^3 resolution.

A2C and A4C refer to atrial 2 and 4 chamber image views respectively, and 9 views denotes all 9 of the previously described standard ultrasound views.

	Binary				CycleGAN inferenced		
method	**9 views**	**A2C**	**A4C**	**A2C & A4C**	**A2C**	**A4C**	**A2C & A4C**
PiVox/Fast	0.859	0.828	0.841	0.818	0.644	0.652	0.7
E-PiVox/Fast	0.86	0.837	0.85	0.822	0.651	0.66	0.712
PiVox/Accurate	0.895	0.85	0.867	0.873	0.673	0.732	0.741
E-PiVox/Accurate	0.903	0.868	0.878	0.881	0.674	0.734	0.74