

RESEARCH

Open Access



Automatic segmentation of the great arteries for computational hemodynamic assessment

Javier Montalt-Tordera^{1†}, Endrit Pajaziti^{1†}, Rod Jones², Emilie Sauvage¹, Rajesh Puranik^{3,4}, Aakansha Ajay Vir Singh^{3,4}, Claudio Capelli¹, Jennifer Steeden¹, Silvia Schievano¹ and Vivek Muthurangu^{1*} 

Abstract

Background: Computational fluid dynamics (CFD) is increasingly used for the assessment of blood flow conditions in patients with congenital heart disease (CHD). This requires patient-specific anatomy, typically obtained from segmented 3D cardiovascular magnetic resonance (CMR) images. However, segmentation is time-consuming and requires expert input. This study aims to develop and validate a machine learning (ML) method for segmentation of the aorta and pulmonary arteries for CFD studies.

Methods: 90 CHD patients were retrospectively selected for this study. 3D CMR images were manually segmented to obtain ground-truth (GT) background, aorta and pulmonary artery labels. These were used to train and optimize a U-Net model, using a 70-10-10 train-validation-test split. Segmentation performance was primarily evaluated using Dice score. CFD simulations were set up from GT and ML segmentations using a semi-automatic meshing and simulation pipeline. Mean pressure and velocity fields across 99 planes along the vessel centrelines were extracted, and a mean average percentage error (MAPE) was calculated for each vessel pair (ML vs GT). A second observer (SO) segmented the test dataset for assessment of inter-observer variability. Friedman tests were used to compare ML vs GT, SO vs GT and ML vs SO metrics, and pressure/velocity field errors.

Results: The network's Dice score (ML vs GT) was 0.945 (interquartile range: 0.929–0.955) for the aorta and 0.885 (0.851–0.899) for the pulmonary arteries. Differences with the inter-observer Dice score (SO vs GT) and ML vs SO Dice scores were not statistically significant for either aorta or pulmonary arteries ($p = 0.741$, $p = 0.061$). The ML vs GT MAPEs for pressure and velocity in the aorta were 10.1% (8.5–15.7%) and 4.1% (3.1–6.9%), respectively, and for the pulmonary arteries 14.6% (11.5–23.2%) and 6.3% (4.3–7.9%), respectively. Inter-observer (SO vs GT) and ML vs SO pressure and velocity MAPEs were of a similar magnitude to ML vs GT ($p > 0.2$).

Conclusions: ML can successfully segment the great vessels for CFD, with errors similar to inter-observer variability. This fast, automatic method reduces the time and effort needed for CFD analysis, making it more attractive for routine clinical use.

Keywords: Segmentation, Machine learning, Neural network, Computational fluid dynamics, Congenital heart disease, Magnetic resonance imaging

[†]Montalt-Tordera Javier and Pajaziti Endrit contributed equally to this work

*Correspondence: v.muthurangu@ucl.ac.uk

¹ UCL Institute of Cardiovascular Science, UCL, London, UK
Full list of author information is available at the end of the article

Background

The past two decades have seen increasing interest in the use of computational fluid dynamics (CFD) for the assessment of cardiovascular disease, including congenital



heart disease (CHD) [1]. Computational fluid dynamics models enable realistic calculation of patient-specific blood flow conditions and provide valuable insights into pathological hemodynamics. These models can also be used to predict hemodynamic response to interventions, thereby aiding therapeutic planning.

The patient specific anatomies needed for CFD are often derived from three dimensional (3D) cardiovascular magnetic resonance (CMR), particularly cardiac and respiratory gated whole-heart sequences. This is because whole-heart images have the sharp borders and high contrast necessary for semi-automated segmentation of cardiovascular structures.

Automatic or semi-automatic segmentation methods based on shape models or level-set algorithms have existed for years [2–4]. However, they often require some type of user input or manual post-correction, rely on priors which are not readily available, or may struggle to adapt to abnormal anatomies (e.g., CHD). Thus, segmentation remains one of the most user-intensive and time-consuming parts of the CFD workflow, and one of the barriers to greater clinical use.

Recently, it has been shown that machine learning (ML) can accurately segment ventricles and great vessels from CMR images [5–7]. Quantitative metrics derived from ML segmentations (e.g., ventricular volumes) compare well with manual segmentations [8, 9] and these techniques are now entering clinical practice. However, the effectiveness of ML segmentation for CFD has not previously been investigated.

The aims of this study were to: (i) Develop a ML method for simultaneous segmentation of the aorta and pulmonary arteries from whole heart CMR images in patients with pediatric or adult CHD, (ii) Compare conventional and ML segmentations using traditional image-based scores, (iii) Compare CFD metrics derived from both conventional and ML segmentations, and (iv) Investigate the association between image-based scores and CFD errors.

Methods

Subjects

Ninety cardiac triggered, respiratory navigated, 3D whole heart, balanced, steady state, free precession (WH-bSSFP) data were collected from previously scanned children and adults with paediatric or CHD (excluding patients with single ventricles). All patients were scanned on a 1.5 T CMR scanner (Avanto, Siemens Healthineers AG, Erlangen, Germany) using a conventional WH-bSSFP sequence [10]. The imaging protocol was as follows: orientation: sagittal, matrix size: $256 \times 144 \times 96$ (head-foot, anterior–posterior,

left–right), acquired voxel size: 1.6 mm (isotropic), flip angle: 90° . Image acquisition was accelerated using GRAPPA (factor of 2 along phase encoding dimension) and partial Fourier (factor of 6/8 along both phase and slice encoding dimensions). The use of retrospectively collected training and test data was approved by the local research ethics committee, and written consent was obtained from all subjects/guardians (Ref: 06/Q0508/124).

Additionally, 10 external examples were retrospectively collected from a different centre. These were scanned on a 1.5 T CMR scanner (Ingenia, Philips Healthcare, Best, the Netherlands) with the following imaging protocol: orientation: axial, matrix size: $240 \times 240 \times 110$ (left–right, anterior–posterior, head-foot), acquired voxel size: 1.44 mm (isotropic), flip angle: 90° . Image acquisition was accelerated using SENSE (reduction factor of 2) and partial Fourier (6/8). Collection of this data and sharing with our site was approved by the local research ethics committee, and written consent was obtained from all subjects/guardians (Protocol No. X20-0237 & 2020/ETH01333).

Ground truth segmentation

Reference standard conventional segmentation of the aorta and pulmonary arteries was performed using a semi-automatic technique with manual correction (Plug-ins built in Horos v4.0, Horosproject.org sponsored by Nimble Co LLC d/b/a Purview, Maryland, USA). Initial segmentation was done using the fast level-set method [2]. This requires the user to: (i) set a threshold, (ii) place seeds in the vessel of interest and (iii) add blocking regions to prevent segmentation of unwanted structures. The quality of this initial segmentation is dependent on both the underlying anatomy and the image quality, but manual correction is always required to remove unwanted structures and clip vessels. The proximal limit of both the aortic and pulmonary artery segmentations was the semi-lunar valve. The distal limit of the segmentations were the diaphragmatic level of the aorta, and hilar branches of the pulmonary arteries. Head and neck arteries were manually removed at their origin.

All 90 datasets were segmented by a primary observer (RJ—10 years' experience in CMR post-processing). We refer to the primary observer's segmentations as the ground truth (GT). In addition, a secondary observer (VM—19 years' experience in CMR post-processing) segmented 10 of these images (test set, see below) to investigate inter-observer variability. We refer to these as the second observer (SO) data. The 10 external

examples were segmented by the secondary observer only following the same procedure.

Data preparation

Prior to ML training, the pixel intensities of the WH-bSSFP data were normalized (range [0, 1]). The aortic and pulmonary binary segmentation masks were concatenated in the channel dimension and combined with a third channel containing a binary “background” mask (one-hot encoding). The images and 3-channel segmentation masks were either centrally cropped or symmetrically zero-padded to a fixed matrix size of $160 \times 96 \times 64$ (superior-inferior, anterior-posterior, left-right). Finally, the image-label pairs were randomly split into a training set (70 examples, 78%), a validation set (10 examples, 11%) and a test set (10 examples, 11%). This split was used to maximize the size of the training set, while providing sufficient data for validation and statistical analysis.

The examples from the external test set were reoriented, interpolated and cropped to match the orientation, matrix size and voxel size of the in-house data.

Network architecture

A U-Net [11] convolutional neural network was used to simultaneously segment the aorta and pulmonary arteries from WH-bSSFP data. The network architecture is shown in Fig. 1. Each convolutional layer was followed by a batch normalization layer and a rectified linear unit

(ReLU) activation. Downscaling was performed using max-pooling layers and upscaling was performed using transpose convolution layers. The number of convolutional filters after the first layer was set to double after each downscaling layer and halve after each upscaling step. The final convolutional layer has three filters (equaling the number of possible classes—aorta, pulmonary artery and background), followed by a softmax activation. Final predicted labels were obtained by assigning each pixel to the class with the highest probability.

Training, hyperparameter optimization and evaluation

The network implementation and training scheme were parametrized to allow investigation of multiple hyperparameter values, with the full search space shown in Table 1. We used the Hyperband algorithm [12] to perform efficient hyperparameter optimization. This method samples the search space randomly and adaptively allocates more computational resources to the most promising hyperparameters combinations. Dice score was used by the Hyperband to assess performance and choose the final model.

The neural network and related functionality were implemented and trained using TensorFlow [13]. In particular, the model implementation, losses and metrics are available in TensorFlow MRI [14], an open-source framework developed in-house. Weights were initialized

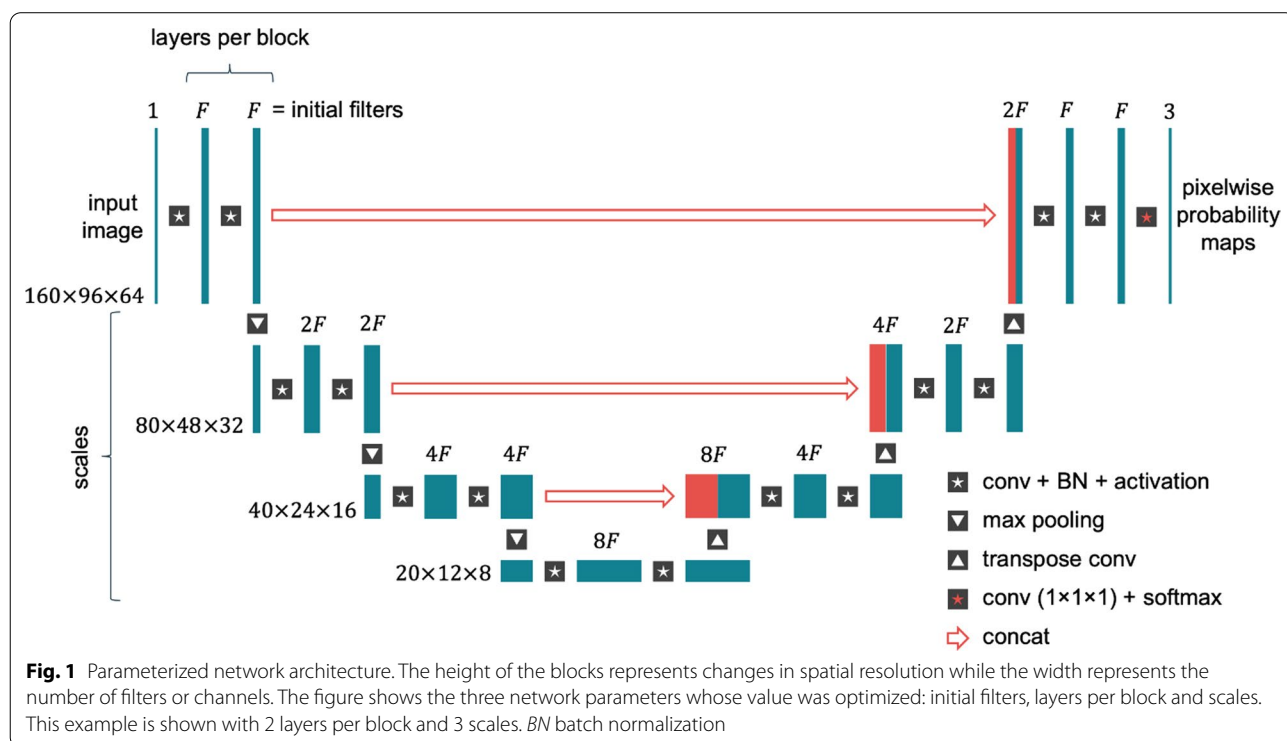


Table 1 Hyperparameter search space

| Parameter | Type | Domain |
|------------------|--------------|--|
| Scales | Architecture | {2, 3, 4} |
| Layers per block | Architecture | {2, 3, 4} |
| Initial filters | Architecture | {32, 64} |
| Learning rate | Training | [0.0001, 0.01] |
| Batch size | Training | {2, 4} |
| Loss function | Training | {CCE, Dice, IoU, Tversky, focal Tversky} |

{ · } represents a discrete set of values; [·] represents a continuous interval of real values. The learning rate was sampled using a log-uniform distribution

CCE categorical cross-entropy

using He's method [15] and optimized using the Adam algorithm [16]. Training, including hyperparameter optimization, took ~ 24 h on an Nvidia Titan RTX GPU with 24 GB of onboard RAM (Nvidia Corporation, Santa Clara, California, USA).

The optimized ML model was evaluated on the test dataset against the GT segmentations (ML vs GT). The accuracy of segmentation was quantified using several image-based segmentation metrics: Dice score, Intersection over Union (IoU), Hausdorff distance (HD) and average surface distance (ASD). Each metric was computed independently for each vessel. Additionally, the same metrics were calculated for the secondary observer's segmentation against the GT (SO vs GT), and between the ML model and the SO (ML vs SO).

To assess generalization ability to data from other sources, the ML model was also evaluated on the external test set. We computed the same set of metrics (Dice, IoU, HD and ASD) between the ML predictions (Ext-ML)

and the GT segmentations (we refer to these as Ext-SO, because the manual segmentation was performed by the same person as the SO data).

Prior to evaluation, ML masks were filtered to remove all but the largest connected component, as identified using 3D connected component labelling with 26-connectivity [17]. This postprocessing step was used to eliminate small background regions which had been misclassified as vessels.

Surface and volume meshing pipeline

The resultant segmentation masks were converted into finite element volume meshes, using the processes shown in Fig. 2. The masks (GT, ML and SO) were first transformed into surface meshes, by applying the marching cubes algorithm (implemented in VMTK), and then remeshed and smoothed with consistent parameters. The surface meshes were clipped manually at inlets and outlets to create planar surfaces; 30 mm extensions were added to ensure the fluid flow in the region of interest was fully developed and capped at the ends to generate close surfaces. These were meshed with tetrahedral elements to build the final unstructured grid for CFD analysis. The grid resolution was determined through a sensitivity analysis (see Additional file 1).

To assess the effect of the manual clipping of the anatomies on the CFD, the simulations from the ML segmentations were also re-run using the same inlet and outlet locations as the GT data, defined by overlaying the GT clipped geometry to the corresponding ML geometry.

Computational fluid dynamics and boundary conditions

CFD simulations were carried out using the solver Fluent (v19.0, Ansys, Canonsburg, Pennsylvania, USA). Blood

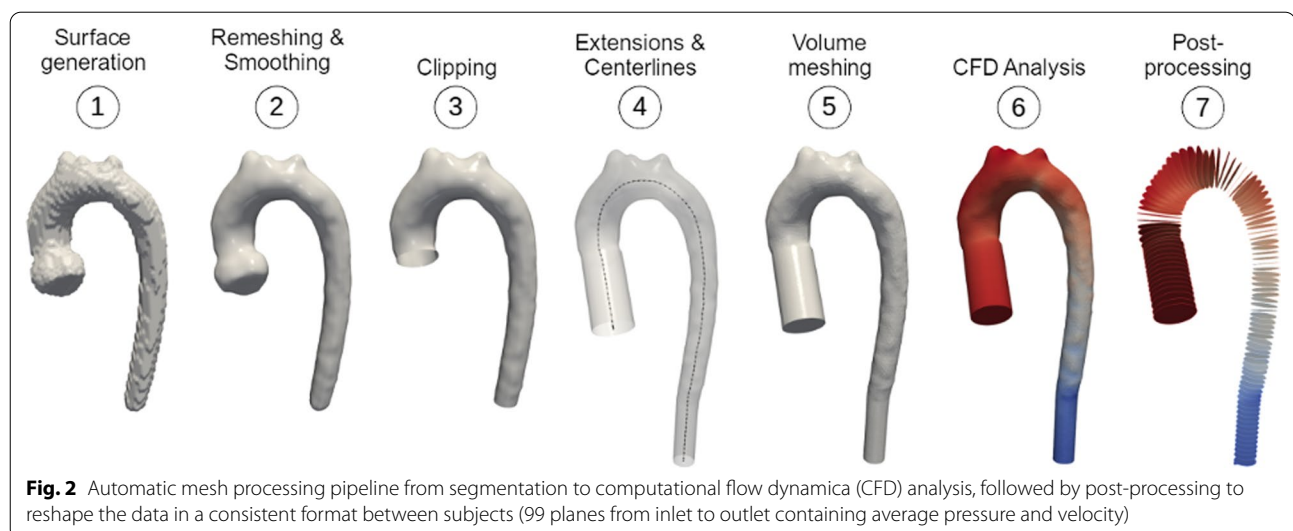


Fig. 2 Automatic mesh processing pipeline from segmentation to computational flow dynamics (CFD) analysis, followed by post-processing to reshape the data in a consistent format between subjects (99 planes from inlet to outlet containing average pressure and velocity)

was modelled as an incompressible Newtonian fluid with density 1060 kg/m^3 and $0.004 \text{ Pa}\cdot\text{s}$ dynamic viscosity [18]. Vessel walls were considered rigid, and no-slip conditions were imposed. A laminar, steady-state model was selected to simulate blood flow at peak systole [19, 20]. A generalisable inlet condition for the aorta and pulmonary artery was applied to all subjects with a uniform (plug) inlet velocity profile of 0.66 m/s for the aorta and 0.57 m/s for the pulmonary artery [21]. The outlets for all cases were assumed to be at zero pressure and the convergence criteria was set at 10^{-4} for the residual errors. Simulations were run on a Dell workstation, with a Xeon CPU E5-2630 (24 processors at 2.3 GHz), 32 Gb RAM and an Nvidia GeForce GTX 1080 Ti.

CFD post-processing and analysis

To compare the flow field between the pairs of different unstructured meshes (ML vs GT, SO vs GT and ML vs SO), correspondence was created by subdividing each vessel with 99 planes orthogonal to the centrelines, calculated in VMTK, and equally distanced. Static pressure and velocity magnitude were averaged in each plane (see Fig. 2) and a percentage error was calculated for each plane pair, using ML as reference (GT in the comparison between SO and GT). The mean absolute percentage errors (MAPE) for pressure and velocity were computed for each vessel pair.

Statistical analysis

Shapiro–Wilk tests were used to test the normality of the different segmentation metrics and CFD errors, grouped by vessel (aorta and pulmonary artery), and segmentation pair (ML vs GT, ML vs SO and SO vs GT). Wilcoxon signed rank tests were used to compare the pressure and velocity errors for the ML vs GT group. Mann–Whitney U-tests were used to compare segmentation metrics and flow field errors between the aorta and the pulmonary artery, for the ML vs GT group. Friedman tests for repeated measurements were performed to compare segmentation metrics and flow field errors between the ML vs GT, ML vs SO and SO vs GT groups, for both aorta and pulmonary artery segmentations. Significant Friedman test results were followed up by pairwise Wilcoxon post-hoc tests. Additionally, Wilcoxon signed rank tests were used to compare ML vs GT and SO vs GT metrics for both aorta and pulmonary artery segmentations. Mann–Whitney U-tests were used to compare Ext-ML vs Ext-SO segmentation metrics against ML vs SO metrics. Wilcoxon signed rank tests were used to compare the pressure and velocity errors for the manually clipped ML vs GT data and the equally clipped ML vs GT data. Pearson's correlation coefficient was used to measure the linear relationship between each pair of a segmentation

metric (i.e., Dice, IoU, HD or ASD) and a flow field error (pressure or velocity MAPEs), for both aorta and pulmonary artery segmentations. The p -value was calculated for each comparison to test non-correlation. Throughout this work, a p -value < 0.05 was considered statistically significant.

Results

Hyperparameter optimization

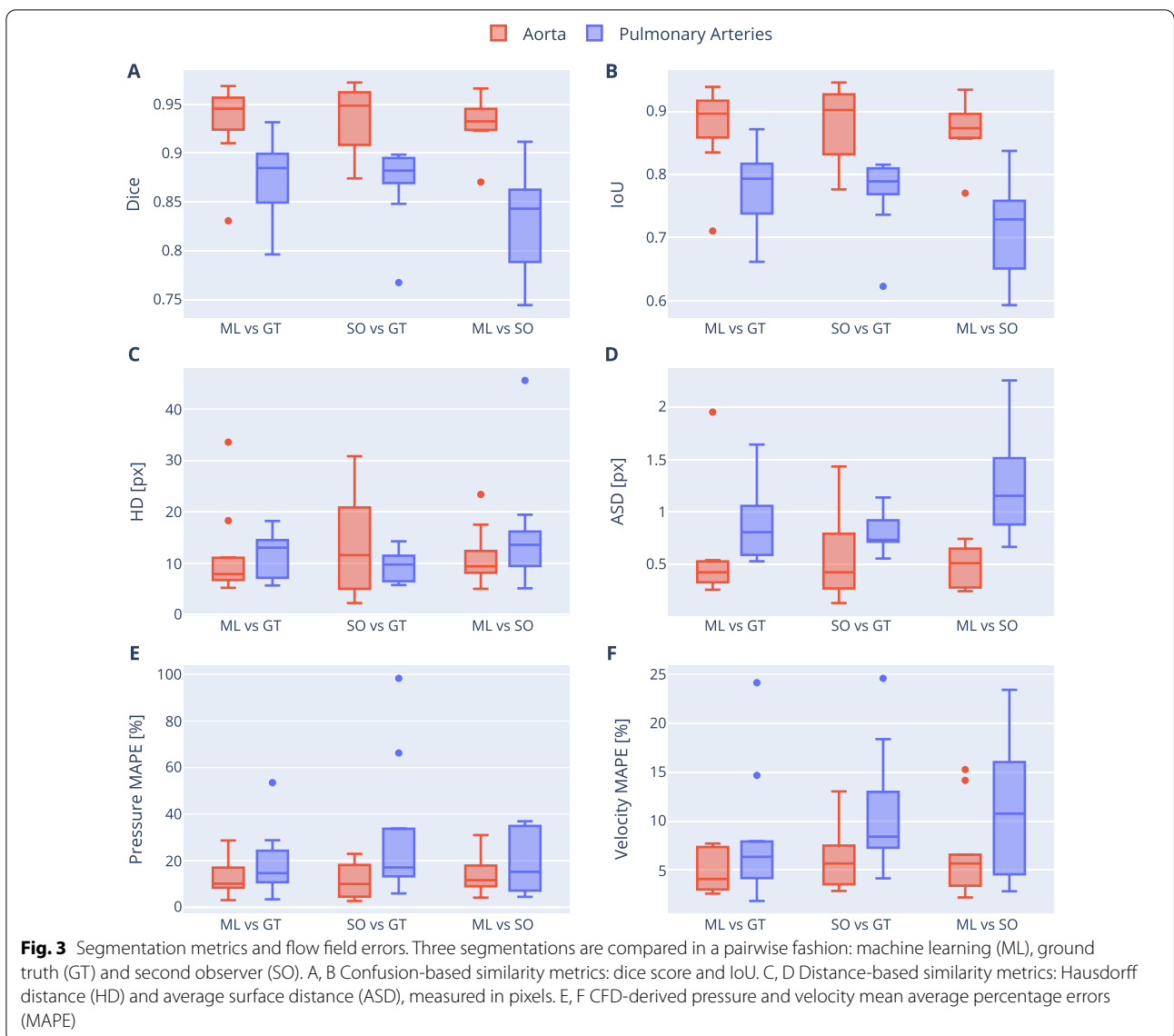
A total of 124 hyperparameter configurations were sampled during the neural network optimization procedure (see Additional file 2). The best performing configuration was as follows: scales = 3, layers per block = 2, initial filters = 64, learning rate = 3.46×10^{-4} , batch size = 2, and loss function = focal Tversky. This model was selected and used in all further experiments.

ML segmentation

The ML segmentation was successful in all 10 test datasets. The specific diagnoses for these patients were: repaired tetralogy of Fallot ($n = 1$), repaired Tetralogy of Fallot with mild right pulmonary artery stenosis ($n = 1$), Marfan syndrome with dilated aorta ($n = 1$), Marfan syndrome with pectus excavatum ($n = 1$), dilated pulmonary artery ($n = 1$), bicuspid aortic valve with dilated aorta and unrepaired VSD ($n = 1$), repaired double outlet right ventricle with right sided arch ($n = 1$), unrepaired atrial septal defect ($n = 1$), aortic regurgitation with dilated aorta ($n = 1$), post Ross procedure with mechanical aortic valve ($n = 1$). Inference time for the ML model was approximately 160 ms for simultaneous segmentation of aorta and pulmonary arteries (compared to approximately 30 min for manual segmentation of aorta and pulmonary arteries). There was good agreement between the ML and GT segmentation with a median Dice score of 0.945 (interquartile range: 0.929–0.955) for the aorta and 0.885 (0.851–0.899) for the pulmonary arteries. The Dice score was significantly higher for the aorta than the pulmonary arteries ($p = 0.002$) with similar findings observed for IoU, HD and ASD (Fig. 3A–D).

The best, median and worst segmented images in terms of Dice score are shown in Fig. 4. The three main differences were: (i) the length of the vessel segmented, (ii) differences in pixel labelling that resulted in small deviations of the vessel border, and (iii) small protrusions at origin of the carotid and subclavian arteries in the ML segmentations of the aorta.

The aortic inter-observer Dice score (SO vs GT) was 0.949 (0.916–0.960) and was not significantly different from ML vs GT ($p = 0.575$). The pulmonary Dice score for the SO vs GT was 0.882 (0.870–0.894) and was also not significantly different from ML vs GT ($p = 0.721$).



The ML vs SO Dice score was 0.933 (0.924–0.944) for the aorta, which was not significantly different from ML vs GT and SO vs GT ($p=0.741$), and 0.843 (0.791–0.860) for the pulmonary arteries, which trended towards being lower than ML vs GT and SO vs GT ($p=0.061$).

The ML segmentation was also successful in the external dataset. The specific diagnoses for these patients were: cardiomyopathy ($n=4$), normal anatomy ($n=1$), repaired tetralogy of Fallot ($n=1$), left pulmonary artery stenosis ($n=1$), anomalous pulmonary venous drainage ($n=1$), repaired coarctation of the aorta with hypoplastic arch ($n=1$), bicuspid aortic valve with severe AR and dilated aortic root ($n=1$). The best, median and worst examples from the external test set are shown in Fig. 5. There was reasonable agreement between the Ext-ML

and Ext-SO segmentations, with a median Dice score of 0.913 (0.889–0.927) for the aorta and 0.751 (0.728–0.797) for the pulmonary arteries. Agreement was significantly lower than ML vs SO for the pulmonary arteries ($p=0.011$), but not for the aorta ($p=0.089$). Similar findings were observed for IoU, HD and ASD (Fig. 6).

CFD metrics

There was overall good agreement in CFD metrics calculated using ML and GT segmentations (Fig. 3E, F). The median MAPE for pressure and velocity in the aorta were 10.1% (interquartile range: 8.5–15.7%) and 4.1% (3.1–6.9%) respectively, and for the pulmonary arteries 14.6% (11.5–23.2%) and 6.3% (4.3–7.9%). Pulmonary artery MAPEs trended towards higher values

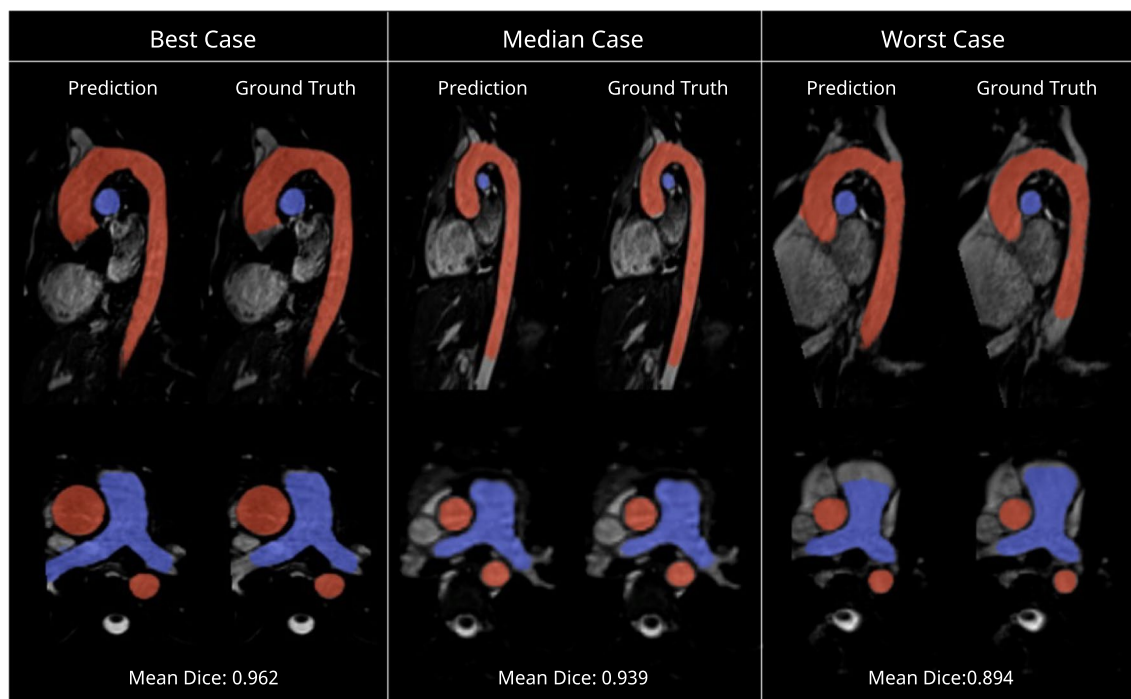


Fig. 4 Test set segmentation overlays. Predicted and ground truth masks are overlaid over the original images for the best, median and worst test cases. Aorta and pulmonary artery masks are shown in red and blue, respectively. Multiplanar reformats of the original 3D volume were manually selected on a case-by-case basis to be most informative. Best case had Ross procedure and mechanical aortic valve, the median case had an atrial septal defect and the worst case had a dilated pulmonary artery

compared to aortic MAPEs, but this did not reach statistical significance ($p=0.081$ for pressure and $p=0.093$ for velocity). However, pressure was more sensitive than velocity to different segmentations, with pressure MAPE being $\sim 2.5 \times$ greater than velocity MAPE ($p < 0.001$).

Figure 7 shows the surface meshes of test cases with the highest and lowest CFD MAPE, as well as pressure and velocities along the length of each vessel. Figure 8 shows pressure and velocity fields calculated using both ML and GT manual segmentations. The main difference in the surface meshes (particularly for the worst cases) were associated with the inlets and outlets (angle and size) and these differences propagated into pressure and velocity fields.

SO vs GT (inter-observer) and ML vs SO pressure and velocity MAPEs were of a similar magnitude to the errors from the ML segmentations (Fig. 3E, F, $p > 0.2$). When the clipping planes of the GT segmentations were used on the ML geometries, the median pressure and velocity MAPEs were reduced to 8.0/3.1% ($p < 0.01$) for the aorta, and to 1.04/3.7% for the pulmonary artery ($p < 0.01$) (see Additional file 3).

Figure 9 illustrates the relationship between the segmentation metrics and the CFD errors on the ML vs GT comparison. No significant correlations were found

between any of the metrics, either for the aorta or the pulmonary arteries, for either manual or equally clipped data.

Discussion

In this study, a deep neural network was trained to simultaneously segment the aorta and pulmonary arteries from 3D CMR data. As its primary purpose was to provide patient specific anatomies for CFD models, we evaluated accuracy using conventional image-based segmentation metrics and resulting errors in CFD measures. The main findings were: (i) The proposed network achieved high performance in terms of image-based segmentation metrics, (ii) There was reasonable agreement between CFD models derived from the ML and GT manual segmentation, (iii) These errors were similar in magnitude to those observed between two different manual segmentations, and (iv) There was no relationship between the segmentation metrics and the resulting CFD errors.

ML segmentation

In data from the same distribution as the training data, the segmentation model achieved comparable or better performance than previously reported 3D ML

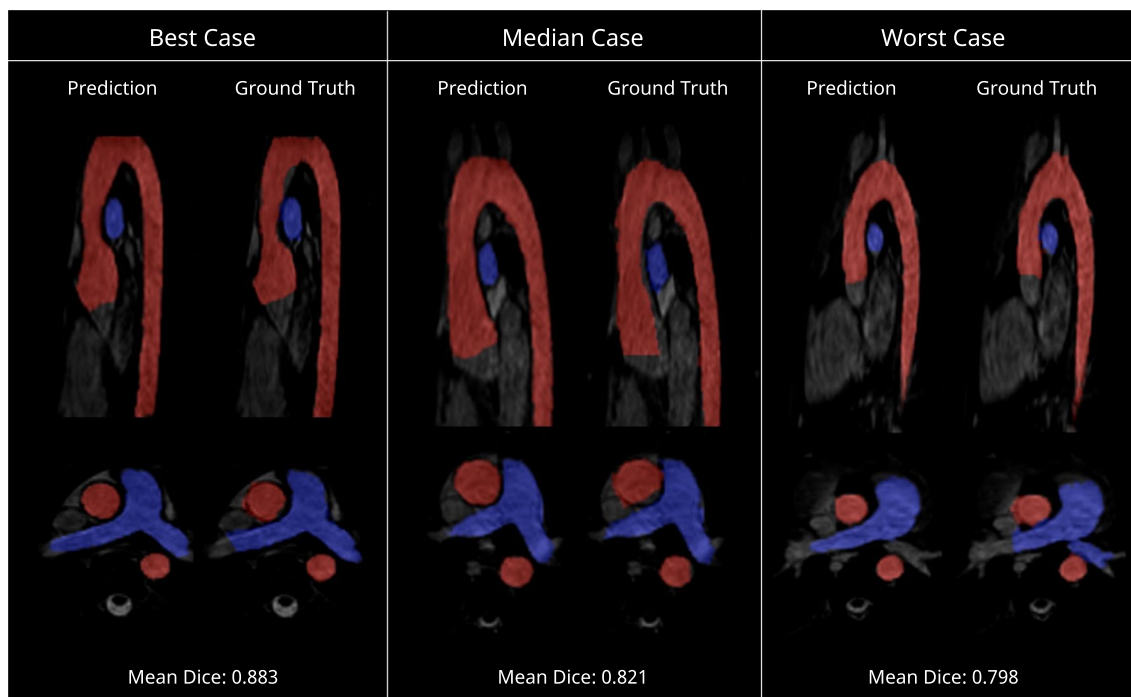


Fig. 5 External test set segmentation overlays. Predicted and ground truth masks are overlaid over the original images for the best, median and worst test cases. Aorta and pulmonary artery masks are shown in red and blue, respectively. Multiplanar reformats of the original 3D volume were manually selected on a case-by-case basis to be most informative. Best case had normal anatomy with gothic arch, the median case had bicuspid aortic valve and the worst case had mild left pulmonary artery stenosis

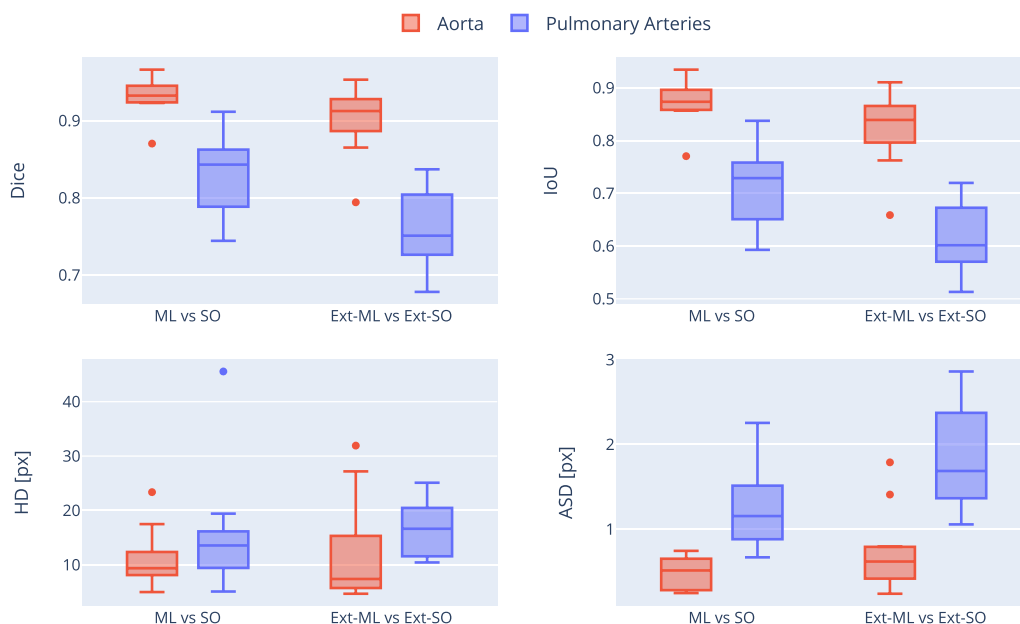
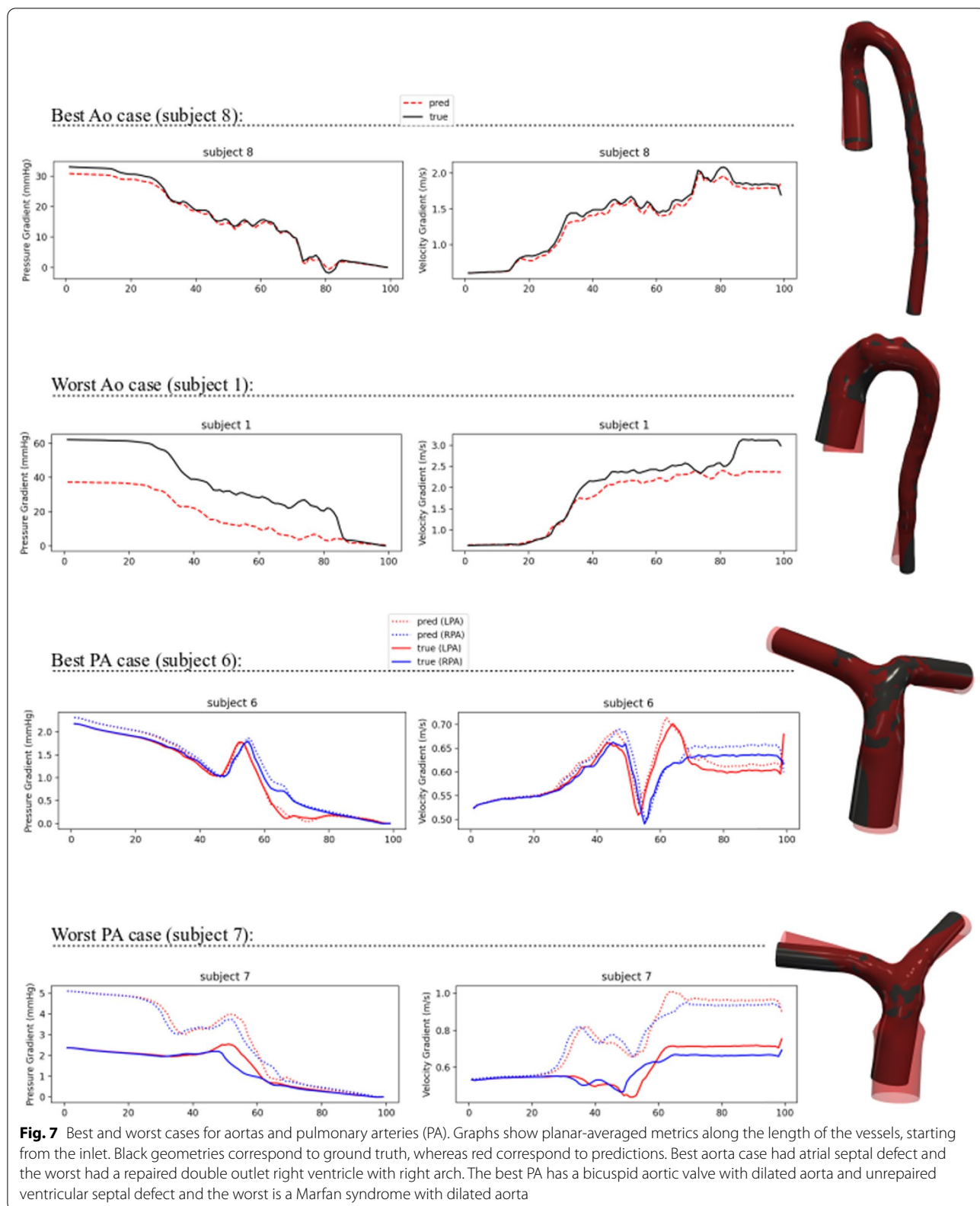
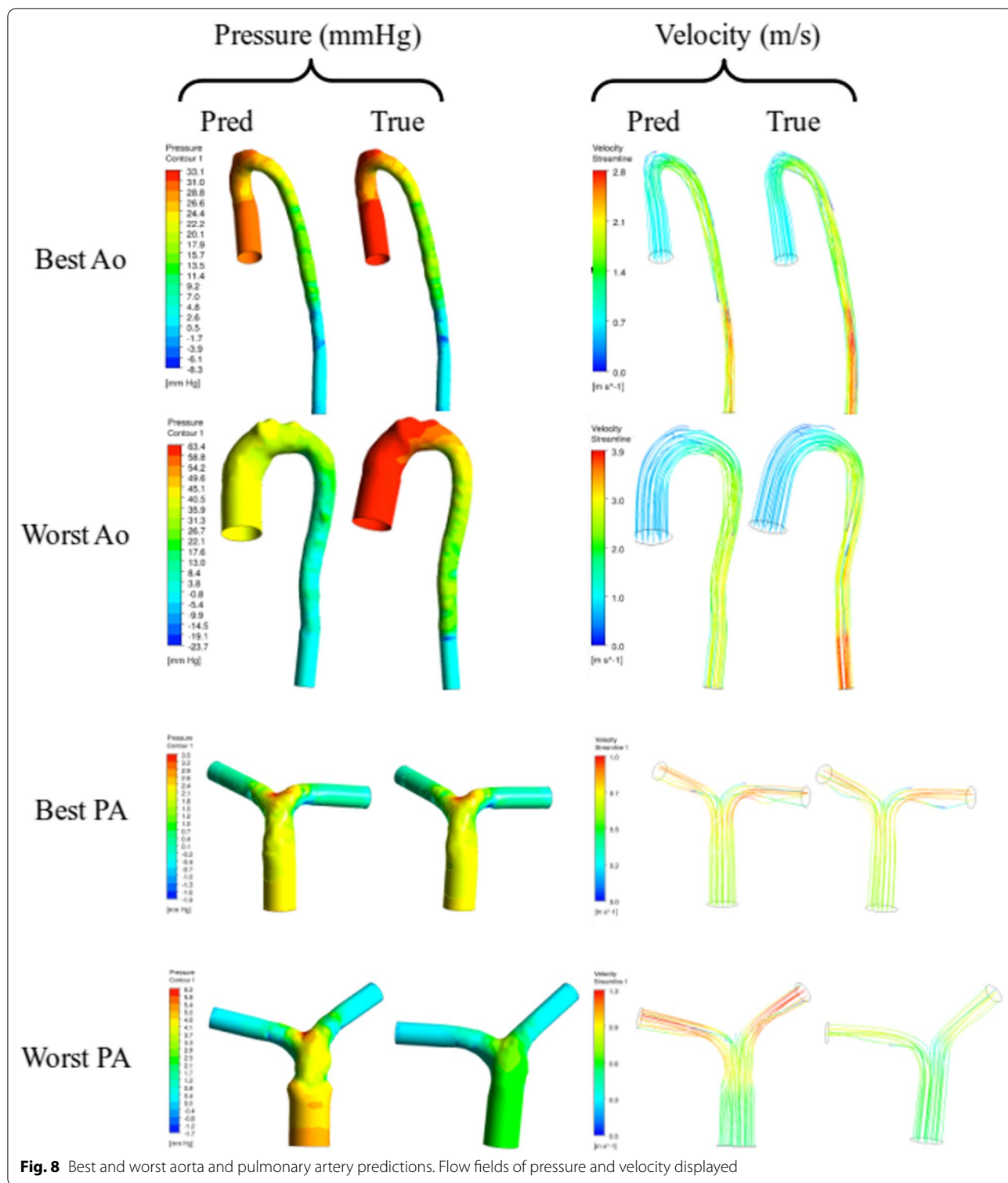


Fig. 6 Segmentation metrics. The model’s segmentation (ML) is compared with the an observer’s segmentation (SO) for two different datasets: our original data (ML vs SO) and an external test set from a different site and vendor (Ext-ML vs Ext-SO). A, B: Confusion-based similarity metrics: dice score and intersection-over-union (IoU). C, D Distance-based similarity metrics: Hausdorff distance (HD) and average surface distance (ASD), measured in pixels





segmentation of the great vessels, including in patients with CHD [6, 22]. This suggests that the chosen network architecture and subsequent hyperparameter optimization were sufficient for accurate segmentation.

Nevertheless, there were some differences between the GT and ML segmentations and visual inspection reveals three main types of error. The first error was a tendency for ML to start and stop segmenting at slightly different

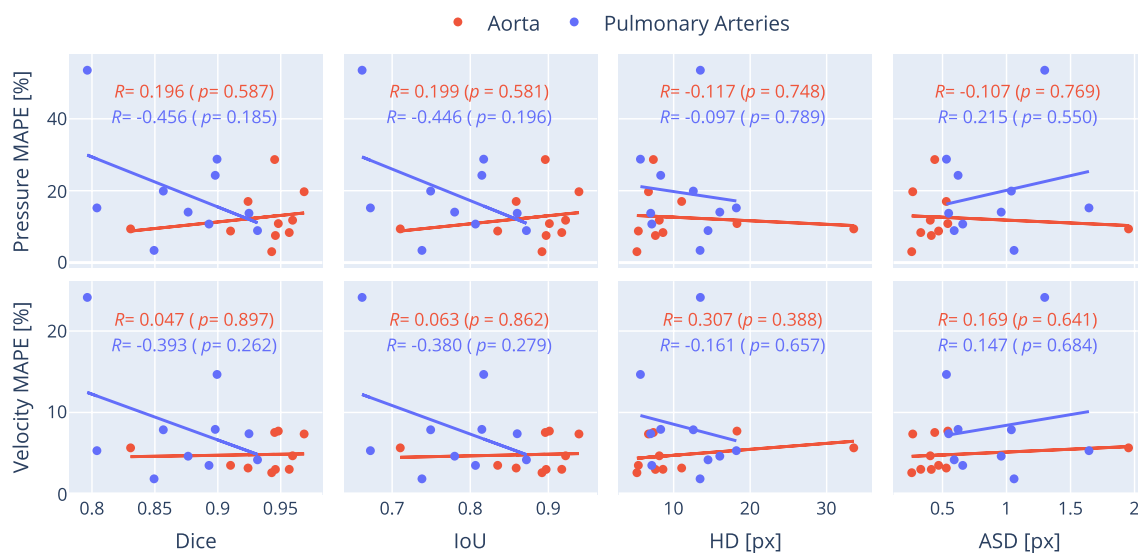


Fig. 9 Flow errors against similarity metrics. The figure shows a scatter-plot matrix where each point corresponds to a subject. In the abscissas, two confusion-based metrics, Dice and IoU, and two distance-based metrics, the Hausdorff distance and the average surface distance, measured in pixels. In the ordinates, the pressure and velocity mean average percentage errors (MAPE). All values are for the ML vs GT comparison. Red and blue colours identify aorta and pulmonary artery data, respectively. Trend lines are least-squares polynomial fits of degree 1. For Dice and IoU, higher is better (more similar). For Hausdorff distance, average surface distance and pressure and velocity MAPEs, lower is better

points in the vessel compared to the GT. The second type of error was the presence of “bumps”, due to the segmentation masks bleeding out at the locations of arterial branches, particularly in the aorta. Both these errors can be considered failures to properly demarcate vessel limits, rather than failures to correctly label blood pool pixels. The third type of error was inaccurate labelling of blood pixels at the vessel border, resulting in subtle differences in surface geometry. It should be noted that none of these patients had very abnormal pulmonary vascular or aortic anatomy, which was necessary to ensure that CFD models could be created from the segmentations. However, further testing on complex CHD is necessary if segmentation models are to be used more widely. Extension to complex CHD may require further enhancements, and several strategies could potentially help improve the ML segmentation accuracy and generalizability. These include increasing the amount and heterogeneity of training data, or performing data augmentation, both of which improve generalizability and performance of ML models [23, 24]. Another interesting option might be the inclusion of statistical shape models [25, 26], which could help ensure that the segmented shapes conform to common patterns.

The model was also tested on 3D data acquired on a different vendor scanner. Although the type of sequence (3D WH-bSSFP) and imaging protocol were similar to the original data, there were visually apparent differences in image quality and characteristics. Nevertheless, we

observed reasonable segmentation quality. For the aorta, agreement with a human observer was only slightly lower than agreement with the same observer in our original data. However, there was a larger reduction in agreement for the pulmonary arteries. This suggests there is scope for improving the generalizability of the model. One of the best solutions for this is to include multi-site, multi-vendor data in the training set, but this would incur obvious labelling costs and potential data sharing difficulties. Other approaches to improve robustness to out-of-distribution data might be the use of data augmentation (e.g., domain translation methods to generate multi-vendor datasets [27, 28]) and the use of strategies that incorporate additional domain knowledge [29].

Of course, segmentation is a challenging task and we demonstrated that the agreement between two humans was similar to the agreement between ML and the GT human segmentation. This suggests that ML “errors” are approximately at the level of the inter-observer variability and similar observations have previously been made for aortic segmentation [6]. Thus, we believe ML can provide segmentation with ‘real world’ accuracy. Furthermore, there are significant advantages of ML over manual segmentation including very fast segmentation without user interaction and perfect reproducibility, due to its deterministic nature. This makes ML particularly useful for removing clinical bottlenecks and accelerating population-based research.

Relationship between CFD and segmentation errors

We demonstrated reasonable agreement in velocity and pressure fields calculated from ML and manual segmentations. Importantly, the differences in CFD metrics using ML vs manual segmentations were of a similar magnitude to those between two independent manual segmentations. This suggests that ML can be successfully used to provide a starting point for CFD simulations, with accuracy similar to inter-observer variability. However, there were some differences in CFD metrics between ML vs GT segmentations, particularly for pressure calculations. We think pressure errors are higher because local deviations in surface geometry tend to cause only local velocity field derangement, but have a global effect on upstream pressures. This can be seen in the worst-case aorta, where a kink in the GT descending aorta results in localized flow acceleration, and significantly altered upstream pressures.

Interestingly, we found no significant correlations between image-based segmentation metrics and errors in the pressure and velocity fields. This suggests that neither overlap-based (Dice, IoU) nor boundary distance-based (HD, ASD) metrics can accurately capture the features that ensure CFD accuracy. This may be because CFD models are highly sensitive to local geometric errors, while segmentation metrics are global and therefore may not fully capture these localized deviations. Another reason may be that differences in clipping (which were not accounted for by segmentation metrics) are responsible for some of the CFD errors, as shown by our analysis of equally clipped data. However, significant CFD errors remained after removing this confounding factor, and these errors were still not correlated with image-based segmentation metrics. Irrespective of the cause, the poor correlation between segmentation and CFD errors has some important implications. Specifically, in our application it might be better to combine conventional global image-based losses with more CFD specific objective measures during training.

Computational fluid dynamics can benefit in several ways from machine learning. Firstly, ML segmentation is completely automated and very fast, enabling significant reduction in pre-processing time, one of the major impediments to clinical uptake. Secondly, ML segmentations are completely reproducible, and this is important as we have shown significant human inter-observer variability. Finally, there has been recent work demonstrating the use of ML to accelerate the CFD simulations. Combined with ML segmentation this would substantially reduce the time taken to perform CFD and make CFD much more attractive for routine clinical use.

Limitations

Our study has several limitations. One of the main limitations of this study was that a simplified CFD model was applied across all subjects (laminar, steady state with no patient-specific parameters). This was done to better isolate the effect of segmentation differences on the resulting CFD model. However, it does limit the patient specific aspect of these comparisons and in the future, boundary conditions for each subject (such as velocity profiles taken from phase contrast CMR) could be incorporated into the model. Furthermore, now that we have demonstrated good agreement using simple CFD models, the utility of ML segmentation for more complex CFD models should be investigated.

Another limitation is that the methods used for comparison don't necessarily account for the full flow field. We used plane-averaged pressures and velocities along the length of the centreline to quantitatively compare different CFD models. However, this averaging does lead to a loss of localized details in the flow fields. Additionally, the slice locations were determined independently for ML, GT and SO models, so there may not be an exact one-to-one correspondence. In future studies, particularly if using more complex CFD models, new metrics of CFD errors that capture subtle deviations will need to be developed.

Conclusions

A convolutional neural network was developed, optimized and trained for segmentation of the aorta and the pulmonary arteries in 3D CMR. The segmentation network was validated for its primary purpose: the creation of CFD models and calculation of flow fields. Segmentation errors in terms of Dice, IoU, HD and ASD as well as derived pressure and velocity field errors were in the range of human inter-observer variability. The proposed method could help to automate clinical hemodynamic assessment workflows and improve their robustness.

Abbreviations

3D: Three-dimensional; ASD: Average surface distance; bSSFP: Balanced steady state free precession; CFD: Computational fluid dynamics; CHD: Congenital heart disease; CMR: Cardiovascular magnetic resonance; GPU: Graphics processing unit; GT: Ground truth; HD: Hausdorff distance; IoU: Intersection-over-union; MAPE: Mean absolute percentage error; ML: Machine learning; RAM: Random-access memory; ReLU: Rectified linear unit; SO: Second observer; WH: Whole-heart.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12968-022-00891-z>.

Additional file 1. Supplementary Materials A.

Additional file 2. Supplementary Materials B.

Additional file 3. Supplementary Materials C.

Acknowledgements

Not applicable.

Author contributions

JMT contributed to the conception and design of this work, pre-processed the images and labels, designed, implemented, trained and optimized the neural network and related functionality, extracted the reported metrics, interpreted the results, performed the statistical analysis and was a major contributor in drafting the manuscript. EP contributed to the conception and design of this work, designed and implemented the semi-automatic mesh processing pipeline, performed the computational fluid dynamics simulations, extracted the reported pressure and velocity field data, interpreted the results and was a major contributor in drafting the manuscript. RJ performed the reference ground truth segmentation. ES contributed to the conception and design of this work. RP contributed to the preparation of the external test set. AS contributed to the preparation of the external test set. CC contributed to the conception and design of this work. JS contributed to the conception and design of this work and substantially revised the manuscript. SS contributed to the conception and design of this work and substantially revised the manuscript. VM contributed to the conception and design of this work, pre-processed the images and labels, performed the secondary observer segmentation and contributed to draft and substantially revise the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by UK Research and Innovation (MR/S032290/1), Heart Research UK (RG2661/17/20), the British Heart Foundation (NH/18/1/33511, PG/17/6/32797), the Engineering and Physical Sciences Research Council (EP/N02124X/1) and the European Research Council (ERC-2017-StG-757923). This report incorporates independent research from the National Institute for Health Research Biomedical Research Centre Funding Scheme. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health. The funders had no role in study design, collection, analysis and interpretation of the data, decision to publish, or preparation of the manuscript.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The use of retrospectively collected training and test data was approved by the local research ethics committee, and written consent was obtained from all subjects/guardians (Ref.: 06/Q0508/124).

Consent for publication

Written consent to use the data for publication was obtained from all subjects/guardians (Ref.: 06/Q0508/124).

Competing interests

The authors declare that they have no competing interests.

Author details

¹UCL Institute of Cardiovascular Science, UCL, London, UK. ²Great Ormond Street Hospital, London, UK. ³Children's Hospital at Westmead, Sydney, Australia. ⁴Faculty of Medicine and Health, University of Sydney, Sydney, Australia.

Received: 29 November 2021 Accepted: 3 October 2022

Published online: 07 November 2022

References

- Biglino G, Capelli C, Bruse J, Bosi GM, Taylor AM, Schievano S. Computational modelling for congenital heart disease: how far are we from clinical translation? *Heart*. 2017;103:98–103. <https://doi.org/10.1136/heartjnl-2016-310423>.
- Wang C, Frimmel H, Smedby Ö. Fast level-set based image segmentation using coherent propagation. *Med Phys*. 2014;41. <https://doi.org/10.1118/1.4881315>.
- Weese J, Kaus M, Lorenz C, Lobregt S, Truyen R, Pekar V. Shape Constrained Deformable Models for 3D Medical Image Segmentation. *Lect Notes Comput Sci (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2001;2082:380–7. https://doi.org/10.1007/3-540-45729-1_38.
- Peters J, Ecabert O, Meyer C, Schramm H, Kneser R, Groth A, et al. Automatic Whole Heart Segmentation in Static Magnetic Resonance Image Volumes. *Lect Notes Comput Sci (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 2007;4792 LNCS:402–10. https://doi.org/10.1007/978-3-540-75759-7_49.
- Chen C, Qin C, Qiu H, Tarroni G, Duan J, Bai W, et al. Deep learning for cardiac image segmentation: a review. *Front Cardiovasc Med*. 2020;25. <https://doi.org/10.3389/FCVM.2020.00025>.
- Berhane H, Scott M, Elbaz M, Jarvis K, McCarthy P, Carr J, et al. Fully automated 3D aortic segmentation of 4D flow MRI for hemodynamic analysis using deep learning. *Magn Reson Med*. 2020;84:2204–18. <https://doi.org/10.1002/MRM.28257>.
- Zhuang X, Li L, Payer C, Štern D, Urschler M, Heinrich MP, et al. Evaluation of algorithms for multi-modality whole heart segmentation: an open-access grand challenge. *Med Image Anal*. 2019;58: 101537. <https://doi.org/10.1016/J.MEDIA.2019.101537>.
- Bai W, Sinclair M, Tarroni G, Oktay O, Rajchl M, Vaillant G, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J Cardiovasc Magn Reson*. 2018;20:1–12. <https://doi.org/10.1186/S12968-018-0471-X>.
- Böttcher B, Beller E, Busse A, Cantré D, Yücel S, Öner A, et al. Fully automated quantification of left ventricular volumes and function in cardiac MRI: clinical evaluation of a deep learning-based algorithm. *Int J Cardiovasc Imaging*. 2020;36:2239–47. <https://doi.org/10.1007/S10554-020-01935-0>.
- Steeden JA, Quail M, Gotschy A, Mortensen KH, Hauptmann A, Arridge S, et al. Rapid whole-heart CMR with single volume super-resolution. *J Cardiovasc Magn Reson*. 2020;22:56. <https://doi.org/10.1186/s12968-020-00651-x>.
- Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. vol. 9351, Springer Verlag; 2015, p. 234–41. https://doi.org/10.1007/978-3-319-24574-4_28.
- Li L, Jamieson K, Rostamizadeh A, Talwalkar A. Hyperband: a novel bandit-based approach to hyperparameter optimization. *J Mach Learn Res*. 2018;18:1–52.
- Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems. 2015.
- Montalt-Tordera J. TensorFlow MRI 2021. <https://doi.org/10.5281/zenodo.5151590>.
- He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In: Proceedings of the IEEE International Conference on Computer Vision, Institute of Electrical and Electronics Engineers Inc.; 2015, p. 1026–34. <https://doi.org/10.1109/ICCV.2015.123>.
- Kingma DP, Ba JL. Adam: A method for stochastic optimization. In: 3rd International Conference on Learning Representations, ICLR 2015—Conference Track Proceedings, International Conference on Learning Representations, ICLR; 2015.
- Silversmith W. cc3d: connected components on multilabel 3D images 2021.
- Gijssen FJH, Van De Vosse FN, Janssen JD. The influence of the non-Newtonian properties of blood on the flow in large arteries: steady flow in a carotid bifurcation model. *J Biomech*. 1999;32:601–8. [https://doi.org/10.1016/S0021-9290\(99\)00015-9](https://doi.org/10.1016/S0021-9290(99)00015-9).

19. Caballero AD, Laín S. A review on computational fluid dynamics modeling in human thoracic aorta. *Cardiovasc Eng Technol*. 2013;4:103–30. <https://doi.org/10.1007/S13239-013-0146-6>.
20. Bonfanti M, Balabani S, Greenwood JP, Puppala S, Homer-Vanniasinkam S, Díaz-Zuccarini V. Computational tools for clinical support: a multi-scale compliant model for haemodynamic simulations in an aortic dissection based on multi-modal imaging data. *J R Soc Interface* 2017;14. <https://doi.org/10.1098/RSIF.2017.0632>.
21. Gabe IT, Gault JH, Ross J, Mason DT, Mills CJ, Schillingford JP, et al. Measurement of instantaneous blood flow velocity and pressure in conscious man with a catheter-tip velocity probe. *Circulation*. 1969;40:603–14. <https://doi.org/10.1161/01.CIR.40.5.603>.
22. Payer C, Štern D, Bischof H, Urschler M. Multi-label whole heart segmentation using CNNs and anatomical label configurations. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 2017;10663 LNCS:190–8. https://doi.org/10.1007/978-3-319-75541-0_20.
23. Sun C, Shrivastava A, Singh S, Gupta A. Revisiting unreasonable effectiveness of data in deep learning era. In: *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2017–October, Institute of Electrical and Electronics Engineers Inc.; 2017, p. 843–52. <https://doi.org/10.1109/ICCV.2017.97>.
24. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Adv Neural Inform Process Syst*. 2012;25.
25. Bhalodia R, Elhabian SY, Kavan L, Whitaker RT. DeepSSM: a deep learning framework for statistical shape modeling from raw images. *Lecture Notes Comput Sci. (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 2018;11167 LNCS:244–57.
26. Raju A, Miao S, Cheng C-T, Lu L, Han M, Xiao J, et al. Deep implicit statistical shape models for 3D medical image delineation 2021. *AAAI*. 2022;36:2135–43.
27. Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the IEEE International Conference on Computer Vision 2017*;2017–October:2242–51. <https://doi.org/10.48550/arxiv.1703.10593>.
28. Yan W, Wang Y, Gu S, Huang L, Yan F, Xia L, et al. The domain shift problem of medical image segmentation and vendor-adaptation by Unet-GAN 2019. <https://doi.org/10.48550/arxiv.1910.13681>.
29. Liu L, Wolterink JM, Brune C, Veldhuis RNJ. Anatomy-aided deep learning for medical image segmentation: a review. *Phys Med Biol*. 2021;66:11TR01. <https://doi.org/10.1088/1361-6560/ABFBF4>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

