Automated atlas-based multi-label fetal cardiac vessel segmentation in Congenital Heart Disease

Paula Ramirez Gilliland[1] PAULA.RAMIREZ_GILLILAND@KCL.AC.UK
Alena Uus[1] ALENA.UUS@KCL.AC.UK
Milou P.M. van Poppel[1] MILOU.VAN_POPPEL@KCL.AC.UK
Irina Grigorescu[1] IRINA.GRIGORESCU@KCL.AC.UK
Johannes K. Steinweg[1] JOHANNES.STEINWEG@KCL.AC.UK
David F.A. Lloyd[1] DAVID.LLOYD@KCL.AC.UK
Kuberan Pushparajah[1] KUBERAN.PUSHPARAJAH@KCL.AC.UK
Andrew P. King[1] ANDREW.KING@KCL.AC.UK
Maria Deprez[1] MARIA.DEPREZ@KCL.AC.UK
[1] *School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK*

Editors: Under Review for MIDL 2022

Abstract

Congenital heart disease (CHD) is the most commonly diagnosed birth defect. T2w black blood MRI provides optimal vessel visualisation, aiding prenatal CHD diagnosis. Common clinical practice involves manual segmentation of fetal heart and vessels for visualisation and reporting purposes. We propose an automated multi-label fetal cardiac vessels deep learning segmentation approach for T2w black blood MRI. Our network is trained using single-label manual segmentations obtained through current clinical practice, combined with a multi-label anatomical atlas with desired multi-label segmentation protocol. Our framework combines deep learning label propagation with 3D residual U-Net segmentation to produce high-quality multi-label output well adapted to the individual subject anatomy. We train and evaluate the network using forty fetal subjects with suspected coarctation of the aorta, achieving a dice score of 0.79 ± 0.02 for the fetal cardiac vessels region. The proposed network outperforms the label propagation, achieving a comparable performance to a 3D residual U-Net trained exclusively on manual single-label data.This multi-label framework therefore represents an advancement over the single-label approach, providing label-specific anatomical information, particularly useful for assessing specific anomaly areas in CHD.

Keywords: Automated Segmentation, Congenital Heart Disease, Fetal MRI, Fetal Cardiac Imaging, Label Propagation

1. Introduction

Congenital heart disease (CHD) encompasses a set of cardiac malformations present from birth (Liu et al., 2019). CHD is the most frequently diagnosed congenital condition, and the dominant cause of congenital malformation-related deaths (Mendis et al., 2011).

Fetal cardiac MR (CMR) has proven its efficacy as an adjunct to echocardiography for detecting CHD prenatally (Dong and Zhu, 2018; Salehi et al., 2021), with T2w black blood MRI being particularly advantageous for vascular assessments (Lloyd et al., 2016, 2019).

Routine clinical practice assessments involve manual segmentation of ROIs, which is laborious and time consuming. Hence, there is an important clinical need for automated

© 2022 P.R. Gilliland *et al.*

fetal CMR segmentation methods, with deep learning being the method of choice for state-of-the-art automated segmentation approaches (Hesamian et al., 2019) and adult CMR segmentation (Chen et al., 2020).

Training a CNN for segmentation requires a large number of manually labelled examples. In this paper we propose a novel deep learning framework that combines label propagation from a multi-label atlas with manual single-label vascular ROI segmentations (see Fig. 1) for training of a 3D residual U-Net to predict high quality multi-label segmentations, well adapted to the individual subject anatomy.

We show that this approach is able to propagate the multi-label protocol from the fetal cardiac atlas while achieving high accuracy in individual images driven by the manual single cardiac vessel labels. These highly accurate fully automated multi-label segmentations will aid visualisation of the fetal cardiac vessels for prenatal diagnostic reporting purposes, and provide the basis for automated vessel biometry and detection.

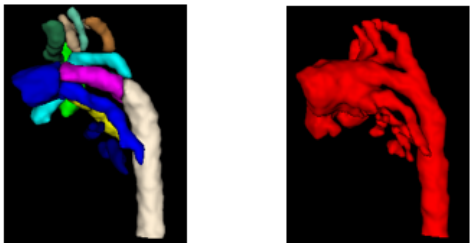


Figure 1: 3D models of the two label types used: a multi-label cardiac atlas (left), and single-label manual segmentations (right).

2. Related Work

2.1. Deep learning segmentation

U-Net (Ronneberger et al., 2015) is one of the most widely used deep neural network architectures, due to its performance, generalisability and efficiency (Isensee et al., 2018). U-Net based architectures have been successfully employed for adult cardiac CMR localisation (Payer et al., 2017), and segmentation in CHD (Xu et al., 2019).

Residual U-Net (Zhang et al., 2018) combines the advantages of Residual connections (He et al., 2016) with U-Net, to produce an enhanced network that eases training and information propagation without degradation. This type of architecture has demonstrated high segmentation performance on adult cardiac segmentation tasks (Kerfoot et al., 2018; Vesal et al., 2020). However Residual U-Net is restricted to the protocol of the available manual labels, which in our case is a single-label vessels ROI segmentation (Fig. 2 left).

2.2. Label propagation

In atlas-based label propagation, the label information from a given atlas is transferred to an individual subject via image registration (Heckemann et al., 2006). With deep learning becoming an increasingly popular choice for image registration due to high computational

efficiency and optimal performance (Tustison et al., 2019), we propose the use of VoxelMorph (Balakrishnan et al., 2019) for label propagation from a multi-label atlas (Fig. 2 right) to individual subjects. While label propagation can transfer any desired segmentation protocol from the atlas to the individual images, its accuracy is limited by the registration quality, and thus residual U-Net may produce more accurate results.

2.3. Automated fetal segmentation

U-Net based architectures have been employed for fetal brain, eye and thorax MRI segmentation (Payette et al., 2020; Salehi et al., 2018; Uus et al., 2021b,a). The only automated fetal cardiac segmentation method in literature consists of a random forest for global cardiac, liver and lungs detection (Keraudren et al., 2015). To our knowledge, the present paper proposes the first multi-label segmentation of fetal cardiac structures in CHD.

2.4. Contribution

Our deep learning framework combines advantages of highly accurate residual U-Net segmentation with deep learning label propagation of any desired protocol, fine-tuned by the available single-label manual segmentations (see Fig. 3). We show that the resulting multi-label residual U-Net segmentation outperforms the label propagation, while achieving comparable accuracy to the single-label residual U-Net trained on manual labels.

3. Methods

3.1. Data specifications

Our dataset consists of 40 fetal subjects with suspected coarctation of the aorta (CoA), 31.73±1.77 weeks mean GA. The datasets were acquired at Evelina London Children's Hospital using a 1.5 T Tesla Ingenia MRI system, T2-weighted SSFSE sequence (see Appendix Section A). All research participants provided written informed consent. The raw datasets comprised 6-12 multi-slice 2D stacks, covering the fetal thorax in three orthogonal planes, and were reconstructed using Slice-to-Volume Registration (SVR) (Kuklisova-Murgasova et al., 2012; Kainz et al., 2015) to 0.75 mm isotropic resolution. Each subject image was manually segmented by trained clinicians using ITK-SNAP (Yushkevich et al., 2006), encompassing the main cardiac vessels region (single label).

We employ 13 manually segmented vascular regions from a 3D CoA cardiac atlas[1], including innominate and azygos veins which are not present in the manual segmentations. We crop all data to the cardiac vessels region (see Fig. 2). We split the subjects into a training set (n=32), validation set (n=3) and testing set (n=5), and normalise and rescale the intensity between 0 and 1 for network training.

3.2. Overall Architecture

The overall framework consists of two main steps: label propagation and automated segmentation (see Fig. 3). First, the label propagation network (VoxelMorph) is fully trained.

1. https://gin.g-node.org/SVRTK/

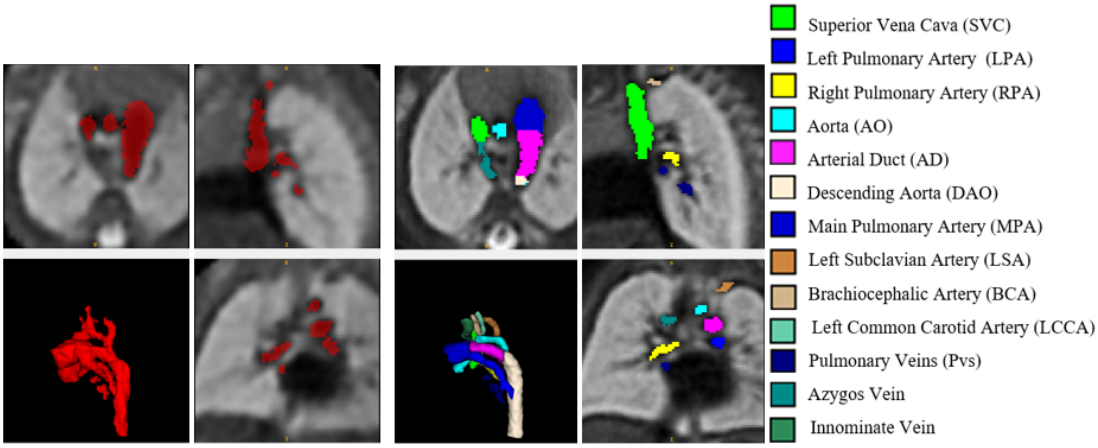


Figure 2: Left: 3D SVR reconstructed image with single manual label. Right: 3D T2w MRI CoA atlas with multi-label vessel segmentations.

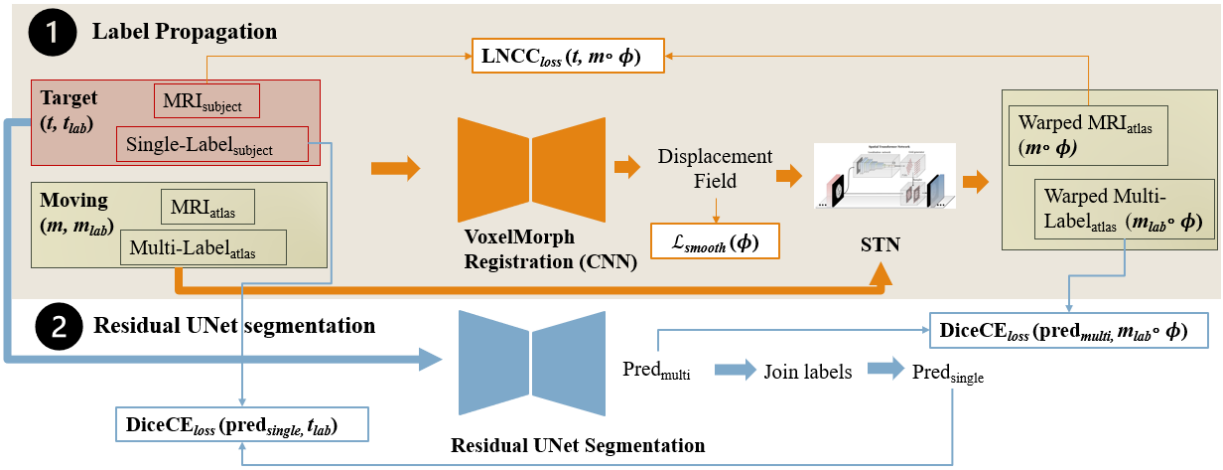


Figure 3: Overall architecture presented. The top half and orange arrows depict the label propagation framework, while the lower half and blue arrows illustrate the segmentation pipeline. Arrows of increased thickness depict the various stages t and m go through, while thinner arrows indicate the volumes used in the losses.

The resulting propagated multi-class labels are then used to train a residual U-Net on the task of automated multi-label segmentation.

3.3. Label Propagation

3.3.1. Registration network

We use VoxelMorph (Balakrishnan et al., 2019), which is a CNN-based registration method, for label propagation. VoxelMorph optimises the aligning deformation field ϕ by considering

a global function composed of shared parameters of a dataset of volume pairs. This global function is parametrised via the kernels of the convolutional layers of a CNN.

A Spatial Transformer Network (STN) (Jaderberg et al., 2015) then uses the resultant deformation field to warp the moving image. We use the subject images as target images (t), and the atlas as moving image (m).

3.3.2. Label propagation loss functions

We use Local Normalised Cross Correlation loss (LNCC$_{loss}$) (Balakrishnan et al., 2019) as a similarity loss function:

$$\mathrm{LNCC}_{loss}(t, m \cdot \phi) = -\sum_{\mathbf{p}\in\Omega} \frac{\left(\sum_{\mathbf{p}_i}(t(\mathbf{p}_i) - \hat{t}(\mathbf{p}))([m \cdot \phi](\mathbf{p}_i) - [\hat{m} \cdot \phi](\mathbf{p}))\right)^2}{\left(\sum_{\mathbf{p}_i} t(\mathbf{p}_i) - \hat{t}(\mathbf{p}))^2\right)\left(\sum_{\mathbf{p}_i}([m \cdot \phi](\mathbf{p}_i) - [\hat{m} \cdot \phi](\mathbf{p}))^2\right)} \quad (1)$$

Here Ω is the target and moving image spatial domain; $\hat{m} \cdot \phi(\mathbf{p})$ and $\hat{t}(\mathbf{p})$ are the local mean intensities of an n^3 volume of moving and target image, respectively; and $\mathbf{p}_i$ are the voxels contained within this n^3 volume, which are iterated through.

We also include a **smoothness loss** ($\mathcal{L}_{smooth}(\phi)$) for the displacement field, to ensure the registration is physically realistic, as described in (Balakrishnan et al., 2019).

We explored the use of an auxiliary segmentation loss between joined atlas labels and manual single-labels, however this resulted in insignificant registration overlap improvement.

Thus, in summary, **the total registration loss** $\mathcal{L}_{reg}$ may be expressed as

$$\mathcal{L}_{reg} = \mathcal{L}_{sim}(t, m \cdot \phi) + \lambda_1 \mathcal{L}_{smooth}(\phi), \quad (2)$$

where λ_1 is the smoothing loss weight.

3.3.3. Registration network implementation details

We employ a U-Net based encoder-decoder architecture with skip connections for registration. Output channels are 16, 32, 32, 64 for the encoder (blocks of 3D strided convolutions with leaky ReLU activations); and 32, 32, 32 for the decoder (blocks of strided transpose 3D convolutions and leaky ReLU activations). This is followed by two convolutional blocks. See Appendix Section B for further details.

We train VoxelMorph registration network for 64,000 iterations, using a linearly decaying learning rate initialised at 5×10^{-4}, and an Adam optimiser (default parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$), with a weight decay of 1×10^{-5}. We implement the LNCC$_{loss}$ using $n = 9$, and $\lambda_1 = 0.2$ for $\mathcal{L}_{smooth}(\phi)$, following hyperparameter tuning. We set the standard deviation of the smoothing kernel (applied to the velocity field) to 2.

We affinely register all subject images to the atlas prior to training. We use TorchIO (Pérez-García et al., 2021) data augmentation during training (random motion, spike, bias, blurring, ghosting, gamma and noise).

3.4. Multi-label segmentation

3.4.1. Segmentation network

We use a 3D Residual U-Net (Zhang et al., 2018) for automated segmentation, with five encoder-decoder blocks (output channels 32, 64, 128, 256 and 512). We use Project MONAI

implementation[2], with convolution and upsampling kernel size of 3, two residual unit convolutions, PReLU activation and instance normalisation, and a dropout ratio of 0.5.

We employ an AdamW optimiser (Loshchilov and Hutter, 2017) with learning rate of 1×10^{-4}, default β parameters and weight decay of 1×10^{-5}. We train the network for 50,000 iterations, using intensity and spatial augmentations from Project MONAI (noise, smoothing, bias field, affine deformations, intensity shifts).

3.4.2. Segmentation loss function

We train the segmentation network by combination of two losses. The first loss minimises the difference between individual manual single-label segmentations and predicted multi-label segmentations, where the multi-class labels are joined into a single label. The second loss minimises the difference between predicted and propagated multi-label segmentations. The overall **segmentation loss** function $\mathcal{L}_{seg}$ is expressed as

$$\mathcal{L}_{seg} = \text{DiceCE}_{loss}(\text{pred}_{multi}, [m \cdot \phi]_{\text{lab}}) + \lambda_2 \text{DiceCE}_{loss}(\text{pred}_{single}, t_{lab}) \tag{3}$$

where pred_{multi} are the multi-label residual U-Net predictions, pred_{single} are single-label predictions (multi-label output labels joined together), λ_2 is the single-class loss weight, $[m \cdot \phi]_{\text{lab}}$ are the propagated labels from the atlas, and t_{lab} are the manual labels (single). We use the **soft dice and cross entropy loss** (Hatamizadeh et al., 2021) expressed as

$$\text{DiceCE}_{loss}(P, G) = 1 - \frac{2}{J} \sum_{j=1}^{J} \frac{\sum_{i=1}^{N} G_{i,j} P_{i,j}}{\sum_{i=1}^{N} G_{i,j}^2 + \sum_{i=1}^{N} P_{i,j}^2} - \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} G_{i,j} \log P_{i,j}, \tag{4}$$

where P is the probabilistic prediction, G is the ground truth, N are the number of voxels, and J the number of classes. Refer to Fig. 3 for the summarised architecture with all losses.

3.5. Training experiments

Our experiments aim to quantitatively compare our proposed combined segmentation technique (*U-Net LP+manual*) with label propagation (*LP*), U-Net trained on propagated labels (*U-Net LP*) and U-Net trained on single-label manual segmentations (*U-Net manual*). These are summarised as

- *LP*: Label Propagation using VoxelMorph.
- *U-Net LP*: Residual U-Net trained with propagated labels ($\lambda_2 = 0$).
- *U-Net manual*: Residual U-Net trained with manual single-labels ($\lambda_2 = \infty$).
- *U-Net LP+manual*: Residual U-Net trained with both manual single-labels and propagated labels. We first train *U-Net LP+manual* using $\lambda_2 = 0.9$ for 50,000 iterations, to ensure accurate multi-label classification, followed by 1,200 iterations with $\lambda_2 = 5$. This is to refine the segmentation of the vessels ROI, reducing any *LP*-induced biases.

2. https://github.com/Project-MONAI/MONAI/.

4. Results

4.1. Quantitative evaluation

Table 1 includes results for mean Dice scores, 95% of Hausdorff Distance (HD95), and average surface distance on test set (n=5); comparing network predictions to manual single-labels (multi-class predictions are joined to form a single label).

Method	Dice	HD95	Average surface distance
LP	0.68 (0.02)	4.03 (0.82)	1.1 (0.1)
U-Net manual	**0.80 (0.02)**	**2.70 (0.40)**	**0.80 (0.06)**
U-Net LP	0.72 (0.01)	2.65 (0.33)	0.96 (0.04)
U-Net LP+manual	**0.79 (0.02)**	**2.39 (0.34)**	**0.81 (0.05)**

Table 1: Single-class mean similarity metrics (standard deviation) on test set.

Vessel	***U-Net manual***	***LP***	***U-Net LP***	***U-Net manual + LP***
SVC	0.78 (0.02)	0.75 (0.02)	0.76 (0.03)	**0.84 (0.04)**
LPA	0.60 (0.12)	0.54 (0.11)	0.59 (0.06)	**0.62 (0.06)**
RPA	**0.72 (0.03)**	0.65 (0.04)	0.65 (0.05)	0.69 (0.05)
Aorta	0.56 (0.05)	0.56 (0.04)	0.67 (0.04)	**0.77 (0.01)**
AD	0.72 (0.06)	0.71 (0.06)	0.72 (0.06)	**0.73 (0.07)**
DAO	0.88 (0.01)	0.86 (0.02)	0.86 (0.02)	**0.89 (0.02)**
MPA	0.67 (0.03)	0.67 (0.03)	0.78 (0.03)	**0.82 (0.04)**
LSA	0.48 (0.18)	0.44 (0.17)	0.60 (0.07)	**0.65 (0.06)**
BCA	0.36 (0.21)	0.32 (0.18)	0.47 (0.05)	**0.52 (0.11)**
LCCA	0.33 (0.21)	0.30 (0.19)	0.49 (0.18)	**0.53 (0.22)**
Pulmonary veins	0.66 (0.06)	**0.69 (0.03)**	0.53 (0.03)	0.59 (0.04)
Azygos vein	-	0.50 (0.15)	0.47 (0.15)	**0.50 (0.14)**
Innominate vein	-	0.62 (0.07)	0.59 (0.05)	**0.67 (0.06)**

Table 2: Mean dice scores (standard deviation) on test set. Highest scores are highlighted in bold. Azygos and innominate veins were not available in the manual labels.

The results in Table 1 show that *U-Net LP+manual* achieves a similar performance as *U-Net manual* for segmenting the whole cardiac vessels ROI, with the added value that using the propagated labels generates multi-class predictions.

Scores for each vessel compared to a multi-class GT are included in Table 2, with extended discussion and vessel score boxplots in Appendix Section D. We intersect our binary predictions from *U-Net manual* with the propagated atlas labels (*LP*) in order to obtain rough multi-class results for *U-Net manual* (label propagation *after* U-Net). Our multi-class scores (Table 2) demonstrate that *U-Net LP+manual* outperforms *U-Net manual*, particularly for small vessels (LSA, BCA, LCCA). Targeting these vessels individually leads to improved segmentation performance, as it imposes constraints on the shape of each vessel.

4.2. Visual inspection

Fig. 4 depicts an example prediction of each model. An important advantage of our multi-label approach over *U-Net manual* is the ability to target small vessels individually. For instance, LSA (white arrow in Fig. 4) remains unsegmented in the *U-Net manual* prediction. Adding multi-class propagated labels enables to detect this vessel.

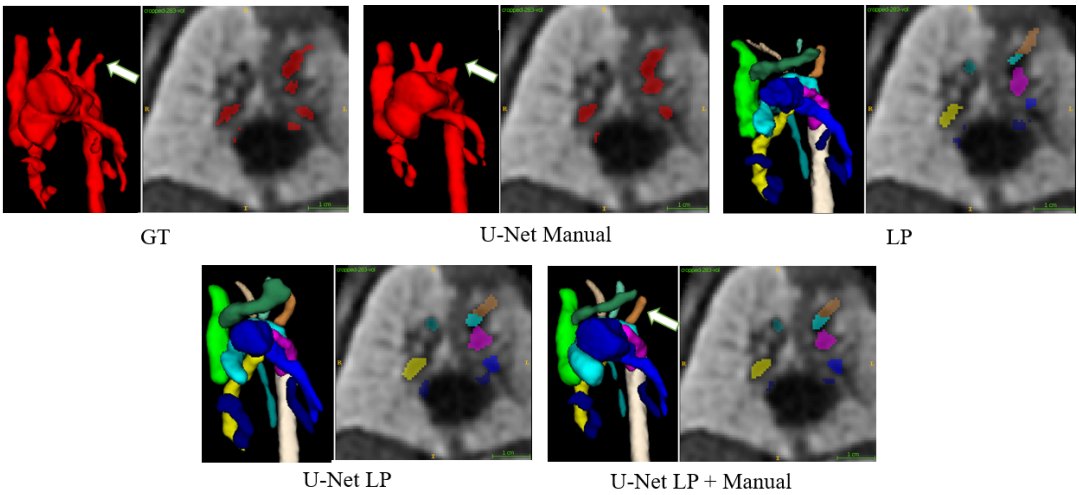


Figure 4: Single and multi-label prediction results for the different experiments.

5. Discussion

Our results show that the full framework which combines propagated labels with single-labels leads to synergistic training, with a vessel detection rate of 100%. Targeting labels individually leads to better segmentation of smaller vessel structures (see Appendix Sections C and D), while providing multi-class information. Including a vessels ROI manual label improves overall segmentation performance by refining areas of inaccurate label propagation.

Predictions were inspected by a trained clinician, reporting overall optimal results, with partial overestimation of pulmonary arteries and upper pulmonary veins, due to scan quality and close proximity to other respiratory structures. Future work will address this using alternative reconstruction techniques (Uus et al., 2020).

6. Conclusion

This study demonstrates the feasiblity of deep learning for multi-label vessel segmentation in black blood T2w reconstructed images, combining label propagation using VoxelMorph with Residual 3D U-Net segmentation. Training experiments explored the combination of single labels of the whole vessels ROI (manual labels) with propagated multi-labels from the atlas, with the inclusion of both labelling information yielding improved predictions. With promising results for subjects with coarctation of the aorta, future work will extend this framework to target other cardiac anomalies, including right aortic arch and double aortic arch.

Acknowledgments

We would like to acknowledge funding from the EPSRC Centre for Doctoral Training in Smart Medical Imaging (EP/S022104/1).

We thank everyone who was involved in the acquisition and examination of the datasets and all participating mothers. This work was supported by the Rosetrees Trust [A2725], the Wellcome/EPSRC Centre for Medical Engineering at King's College London [WT 203148/Z/16/Z], the Wellcome Trust and EPSRC IEH award [102431] for the iFIND project, the NIHR Clinical Research Facility (CRF) at Guy's and St Thomas' and by the National Institute for Health Research Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

References

Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, 38(8):1788–1800, 2019.

Chen Chen, Chen Qin, Huaqi Qiu, Giacomo Tarroni, Jinming Duan, Wenjia Bai, and Daniel Rueckert. Deep learning for cardiac image segmentation: a review. *Frontiers in Cardiovascular Medicine*, 7:25, 2020.

Su-Zhen Dong and Ming Zhu. Utility of fetal cardiac magnetic resonance imaging to assess fetuses with right aortic arch and right ductus arteriosus. *The Journal of Maternal-Fetal & Neonatal Medicine*, 31(12):1627–1631, 2018.

Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger Roth, and Daguang Xu. Unetr: Transformers for 3d medical image segmentation. *arXiv preprint arXiv:2103.10504*, 2021.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *European conference on computer vision*, pages 630–645. Springer, 2016.

Rolf A Heckemann, Joseph V Hajnal, Paul Aljabar, Daniel Rueckert, and Alexander Hammers. Automatic anatomical brain mri segmentation combining label propagation and decision fusion. *NeuroImage*, 33(1):115–126, 2006.

Mohammad Hesam Hesamian, Wenjing Jia, Xiangjian He, and Paul Kennedy. Deep learning techniques for medical image segmentation: achievements and challenges. *Journal of digital imaging*, 32(4):582–596, 2019.

Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, Paul F Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, Sebastian Wirkert, et al. nnu-net: Self-adapting framework for u-net-based medical image segmentation. *arXiv preprint arXiv:1809.10486*, 2018.

Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. *Advances in neural information processing systems*, 28:2017–2025, 2015.

Bernhard Kainz, Markus Steinberger, Wolfgang Wein, Maria Kuklisova-Murgasova, Christina Malamateniou, Kevin Keraudren, Thomas Torsney-Weir, Mary Rutherford, Paul Aljabar, Joseph V Hajnal, et al. Fast volume reconstruction from motion corrupted stacks of 2d slices. *IEEE transactions on medical imaging*, 34(9):1901–1913, 2015.

Kevin Keraudren, Bernhard Kainz, Ozan Oktay, Vanessa Kyriakopoulou, Mary Rutherford, Joseph V Hajnal, and Daniel Rueckert. Automated localization of fetal organs in mri using random forests with steerable features. In *International conference on medical image computing and computer-assisted Intervention*, pages 620–627. Springer, 2015.

Eric Kerfoot, James Clough, Ilkay Oksuz, Jack Lee, Andrew P King, and Julia A Schnabel. Left-ventricle quantification using residual u-net. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 371–380. Springer, 2018.

Maria Kuklisova-Murgasova, Gerardine Quaghebeur, Mary A Rutherford, Joseph V Hajnal, and Julia A Schnabel. Reconstruction of fetal brain mri with intensity matching and complete outlier removal. *Medical image analysis*, 16(8):1550–1564, 2012.

Yingjuan Liu, Sen Chen, Liesl Zühlke, Graeme C Black, Mun-kit Choy, Ningxiu Li, and Bernard D Keavney. Global birth prevalence of congenital heart defects 1970–2017: updated systematic review and meta-analysis of 260 studies. *International journal of epidemiology*, 48(2):455–463, 2019.

David FA Lloyd, Joshua FP Van Amerom, Kuberan Pushparajah, John M Simpson, Vita Zidere, Owen Miller, Gurleen Sharland, Joanna Allsop, Matthew Fox, Maelene Lohezic, et al. An exploration of the potential utility of fetal cardiovascular mri as an adjunct to fetal echocardiography. *Prenatal Diagnosis*, 36(10):916–925, 2016.

David FA Lloyd, Kuberan Pushparajah, John M Simpson, Joshua FP Van Amerom, Milou PM Van Poppel, Alexander Schulz, Bernard Kainz, Maria Deprez, Maelene Lohezic, Joanna Allsop, et al. Three-dimensional visualisation of the fetal heart using prenatal mri with motion-corrected slice-volume registration: a prospective, single-centre cohort study. *The Lancet*, 393(10181):1619–1627, 2019.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

Shanthi Mendis, Pekka Puska, Bo Norrving, World Health Organization, et al. *Global atlas on cardiovascular disease prevention and control*. World Health Organization, 2011.

Christian Payer, Darko Štern, Horst Bischof, and Martin Urschler. Multi-label whole heart segmentation using cnns and anatomical label configurations. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 190–198. Springer, 2017.

Kelly Payette, Raimund Kottke, and Andras Jakab. Efficient multi-class fetal brain segmentation in high resolution mri reconstructions with noisy labels. In *Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis*, pages 295–304. Springer, 2020.

Fernando Pérez-García, Rachel Sparks, and Sébastien Ourselin. Torchio: a python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Computer Methods and Programs in Biomedicine*, page 106236, 2021. ISSN 0169-2607. doi: https://doi.org/10.1016/j.cmpb.2021.106236. URL https://www.sciencedirect.com/science/article/pii/S0169260721003102.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

Daniel Salehi, Katrin Fricke, Misha Bhat, Håkan Arheden, Petru Liuba, and Erik Hedström. Utility of fetal cardiovascular magnetic resonance for prenatal diagnosis of complex congenital heart defects. *JAMA network open*, 4(3):e213538–e213538, 2021.

Seyed Sadegh Mohseni Salehi, Seyed Raein Hashemi, Clemente Velasco-Annis, Abdelhakim Ouaalam, Judy A Estroff, Deniz Erdogmus, Simon K Warfield, and Ali Gholipour. Real-time automatic fetal brain extraction in fetal mri by deep learning. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 720–724. IEEE, 2018.

Nicholas J Tustison, Brian B Avants, and James C Gee. Learning image-based spatial transformations via convolutional neural networks: A review. *Magnetic resonance imaging*, 64:142–153, 2019.

Alena Uus, Tong Zhang, Laurence H Jackson, Thomas A Roberts, Mary A Rutherford, Joseph V Hajnal, and Maria Deprez. Deformable slice-to-volume registration for motion correction of fetal body and placenta mri. *IEEE transactions on medical imaging*, 39(9): 2750–2759, 2020.

Alena Uus, Irina Grigorescu, Milou van Poppel, Emer Hughes, Johannes Steinweg, Thomas Roberts, David Lloyd, Kuberan Pushparajah, and Maria Deprez. 3d unet with gan discriminator for robust localisation of the fetal brain and trunk in mri with partial coverage of the fetal body. *bioRxiv*, 2021a.

Alena Uus, Jacqueline Matthew, Irina Grigorescu, Samuel Jupp, Lucilio Cordero Grande, Anthony Price, Emer Hughes, Prachi Patkee, Vanessa Kyriakopoulou, Robert Wright, et al. Spatio-temporal atlas of normal fetal craniofacial feature development and cnn-based ocular biometry for motion-corrected fetal mri. In *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Perinatal Imaging, Placental and Preterm Image Analysis*, pages 168–178. Springer, 2021b.

Sulaiman Vesal, Andreas Maier, and Nishant Ravikumar. Fully automated 3d cardiac mri localisation and segmentation using deep neural networks. *Journal of Imaging*, 6(7):65, 2020.

Xiaowei Xu, Tianchen Wang, Yiyu Shi, Haiyun Yuan, Qianjun Jia, Meiping Huang, and Jian Zhuang. Whole heart and great vessel segmentation in congenital heart disease using deep neural networks and graph matching. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 477–485. Springer, 2019.

Paul A. Yushkevich, Joseph Piven, Heather Cody Hazlett, Rachel Gimpel Smith, Sean Ho, James C. Gee, and Guido Gerig. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage*, 31 (3):1116–1128, 2006.

Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, 2018.

Appendix A. Data Acquisition Details

The datasets were acquired at Evelina London Children's Hospital using a 1.5 T Tesla Ingenia MRI system, T2-weighted SSFSE sequence (repetition time = 20,000 ms, echo time = 50ms, flip angle = 90°, voxel size = 1.25 × 1.25 mm, slice thickness = 2.5 mm and slice overlap = 1.25 mm). Mothers were scanned in a left lateral tilt position using an anterior torso coil array in combination with a posterior spine coil array to measure signal in 28 receiver channels.

We recruited via the iFIND project, and another long-running fetal MRI study (REC: 14/LO/1806 and REC: 07/H0707/105).

Appendix B. Registration CNN Architecture

Fig. 5 illustrates the CNN architecture employed for label propagation. The input images (atlas and subject) are concatenated along the channel dimension and inputted into the network, with Gaussian smoothing being applied to the resultant velocity field, to ensure diffeomorphism. This is followed by scaling and squaring layers, which yield the final deformation field, used to warp the moving image via a Spatial Transformer Network (STN) (Jaderberg et al., 2015).

Appendix C. Extended Qualitative Evaluation

Here we provide further examples to showcase the advantages of our full multi-label framework (*U-Net LP+manual*) over the alternative models investigated. Fig. 6 showcases an example where *U-Net LP+manual* provides a much more realistic prediction of the shape of the circled vessel (BCA) over *U-Net manual*. The fact that each vessel is targeted individually in our multi-class approach (*U-Net LP+manual*) therefore leads to favourable predictions regarding the accuracy of the shape and structure of small vessels.

This is further validated by Fig. 7, where *U-Net manual* provides a biologically unrealistic prediction: BCA is joined to SVC (see arrow showcasing joined vessels), in contrast with the prediction from *U-Net LP+manual*. Thus our full framework also proves advantageous for ensuring biologically realistic predictions, again due to vessels being targeted separately.

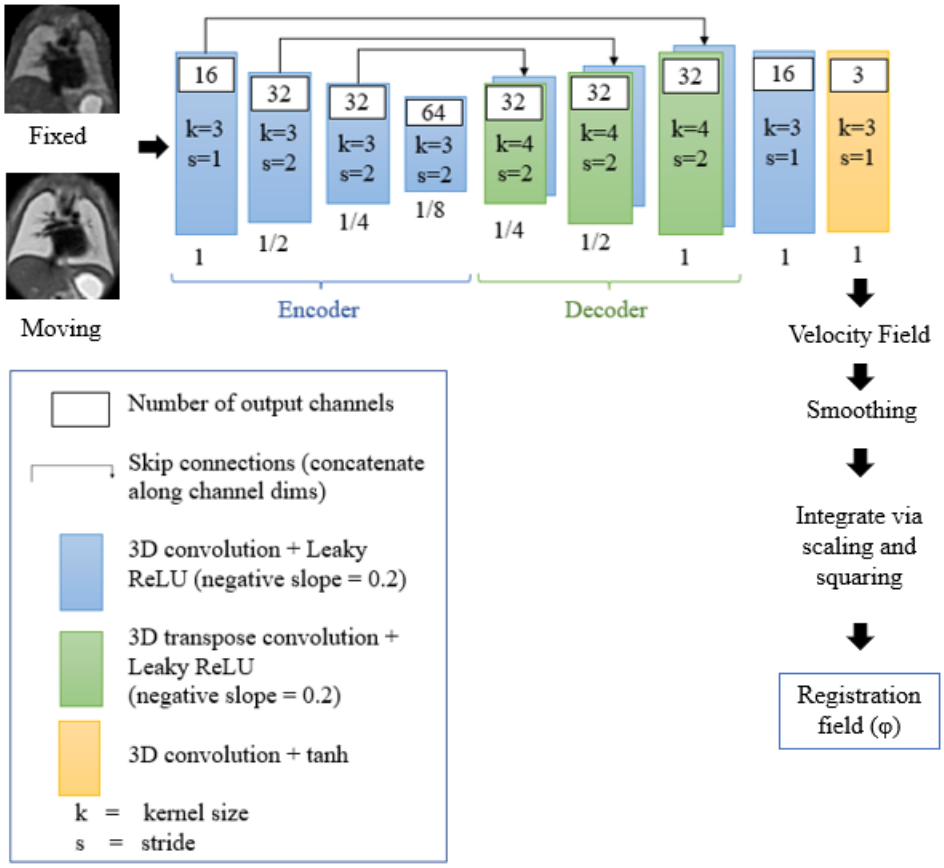


Figure 5: CNN architecture used for registration. The numbers under each convolution representation indicate the volume spatial resolution relative to the input volume. k=kernel size, s=stride.

Appendix D. Multi-class Results

Figures 10 to 22 include boxplots displaying dice scores on test set for all the vessels segmented, against a multi-class ground truth. The mean is indicated by a green triangle, and the median by an orange intersecting line. Our different implementations are colour coded, to aid visualisation.

Note that Figures 21 and 22 do not contain results for U-Net trained on manual single class labels (*U-Net manual*), as these two vessels (innominate vein and azygos vein) were not segmented in the single-class manual labels.

Figure 8 includes a visual example of multi-class prediction from our full framework (*U-Net manual + LP*) against multi-class GT.

Our full framework (*U-Net manual + LP*) generally achieves higher scores than *U-Net manual* for most vessels with most notable differences in SVC, Aorta, AD, DAO, MPA, LSA, BCA, LCCA. Importantly, for the condition investigated, which is coarctation of the aorta, the vessels of interest (i.e. surrounding the anomaly area), which are the aorta, head

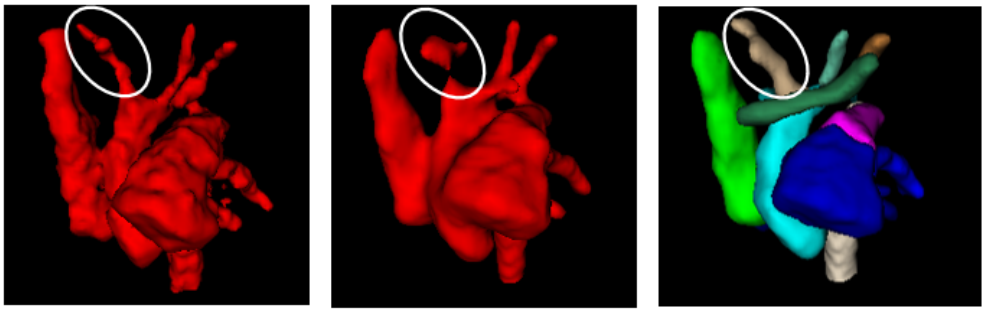


Figure 6: 3D Model predictions with BCA circled, showcasing shape prediction improvements with our full framework.

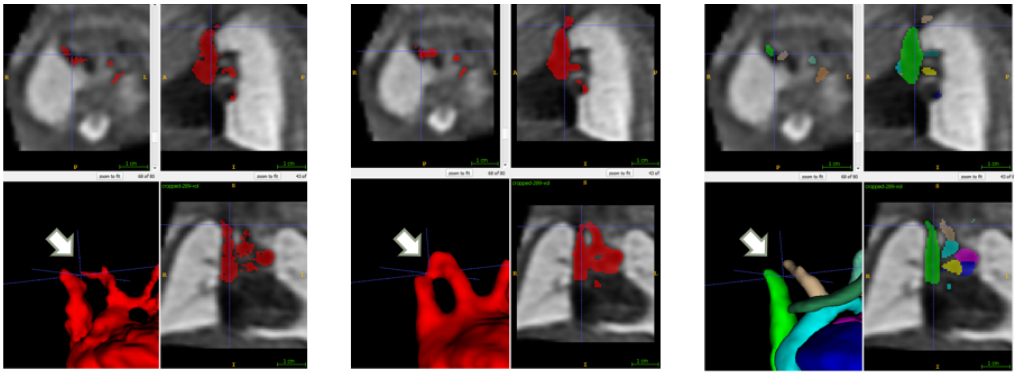
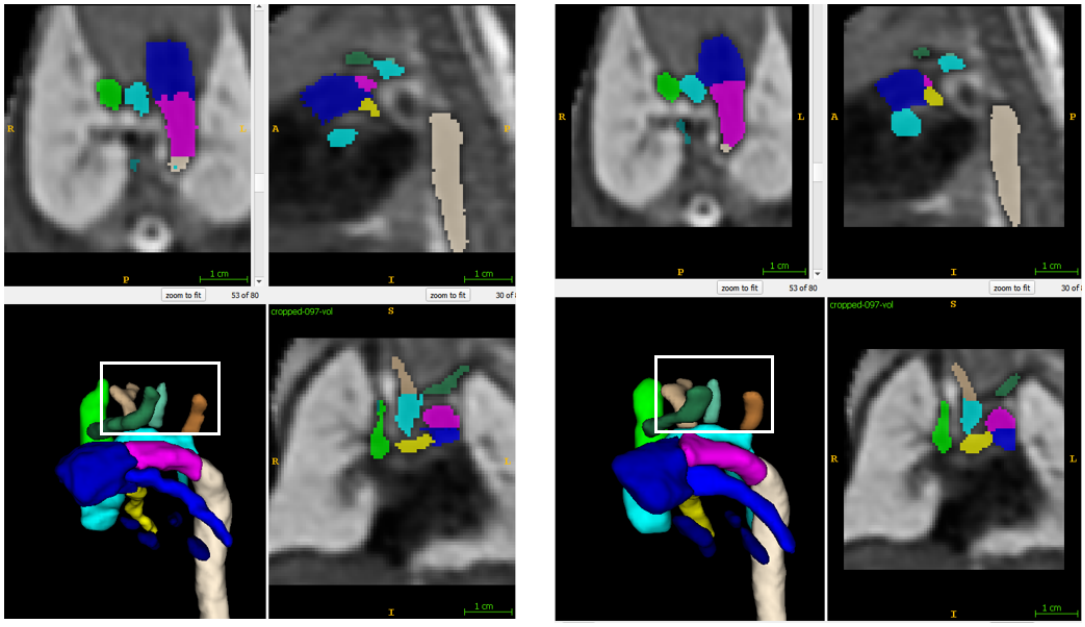


Figure 7: Predictions showcasing inaccurately joined SVC and BCA in the *U-Net manual* prediction (white arrow).

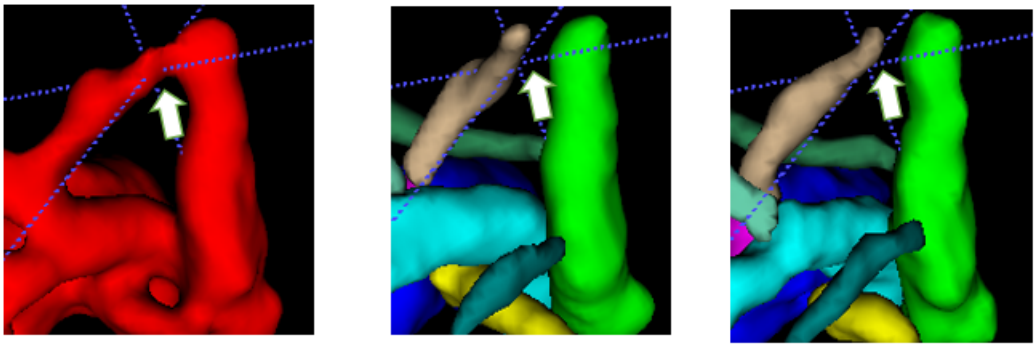
and neck vessels (LSA, BCA, LCCA), DAO and AD, present much higher scores in our full implementation (*U-Net manual + LP*) than in *U-Net manual*.

Having multi-class predictions allows to focus on each vessel at a time, which imposes constraints for generating realistic predictions, as opposed to just having a single-class prediction such as *U-Net manual*, where certain vessels may be missed (see Discussion in Section 4.2). This is quantitatively seen by the higher dice scores from our full framework (*U-Net manual + LP*) in the smallest vessels (LSA, BCA, LCCA, vessels in rectangle in Figure 8), compared to *U-Net manual*, as well as in Figure 9, where BCA is joined to SVC in *U-Net Manual* predictions, the same situation as in Figure 7. Targeting each vessel individually minimises these issues, as can be seen by our correct predictions from *U-Net manual + LP* in Figure 9.



GT U-Net manual + LP

Figure 8: Example test set subject predictions from our full framework (*U-Net manual + LP*) and multi-class GT. BCA, LCCA and LSA are highlighted in a rectangular box, to display their small size relative to other vessels, which explains their lower scores.



U-Net manual U-Net manual + LP GT

Figure 9: Arrow depicts BCA joined to SVC in *U-Net Manual* predictions, as opposed to *U-Net Manual + LP*, which aside from providing multi-class information, imposes constraints, as each vessel is learnt individually.

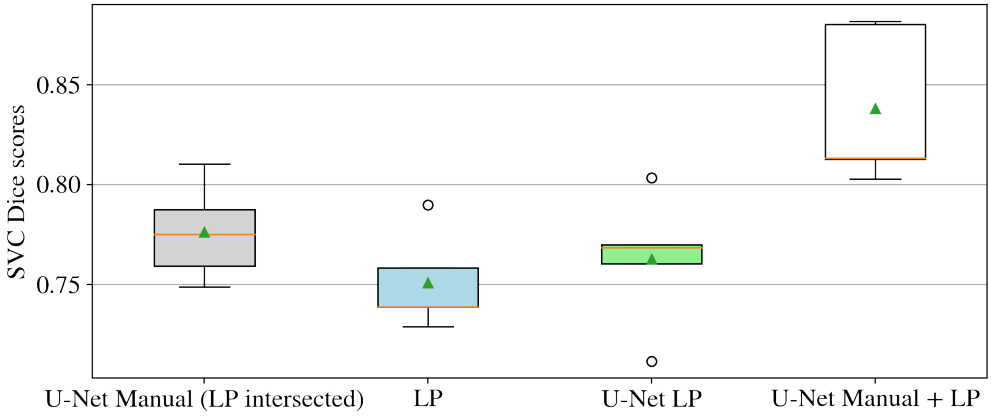


Figure 10: Superior Vena Cava (SVC) Dice scores

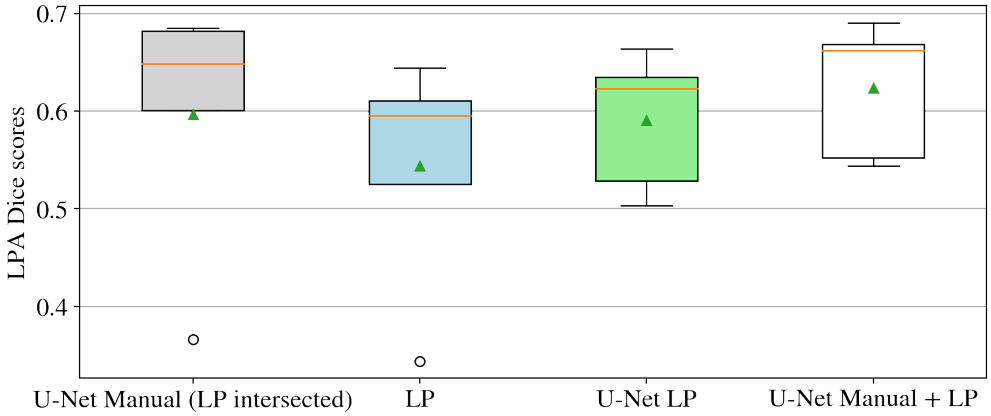


Figure 11: Left Pulmonary Artery (LPA) Dice scores

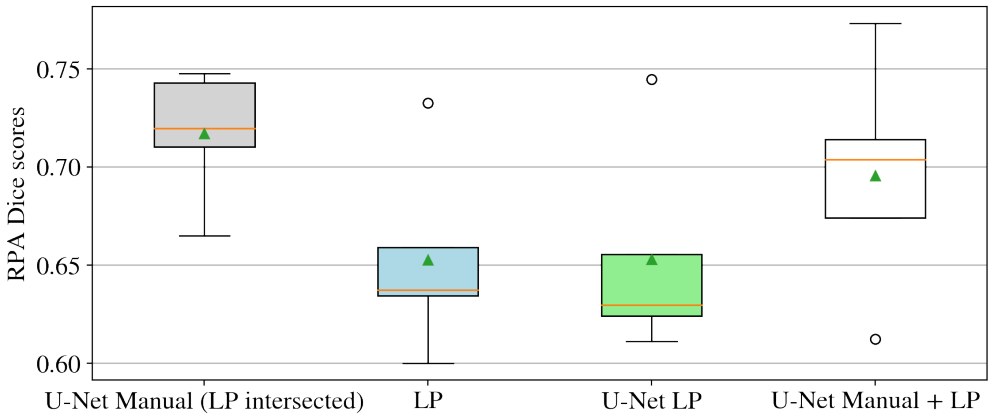


Figure 12: Right Pulmonary Artery (RPA) Dice scores

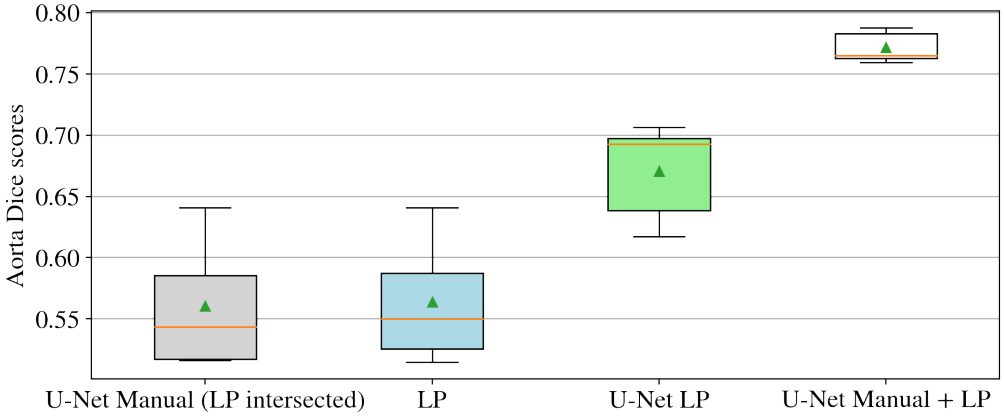


Figure 13: Aorta Dice scores

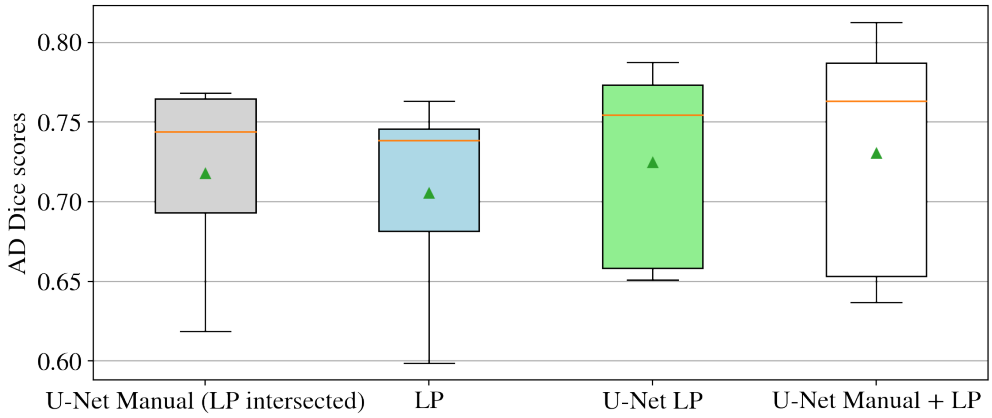


Figure 14: Arterial Duct (AD) Dice scores

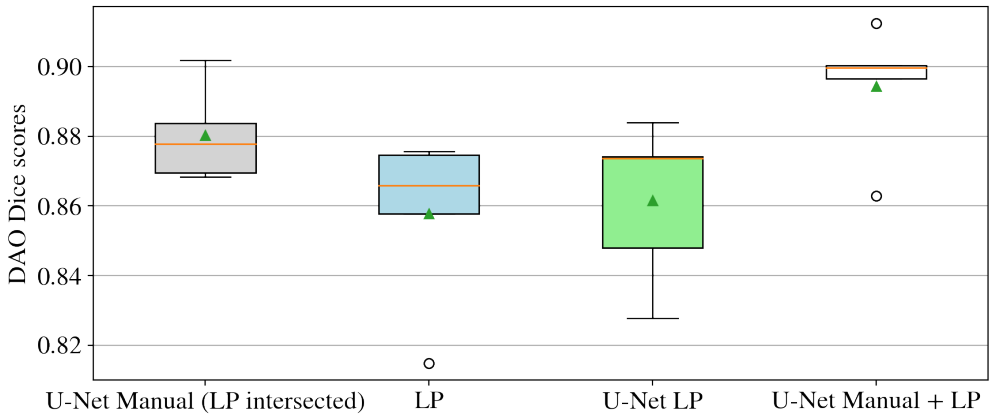


Figure 15: Descending Aorta (DAO) Dice scores

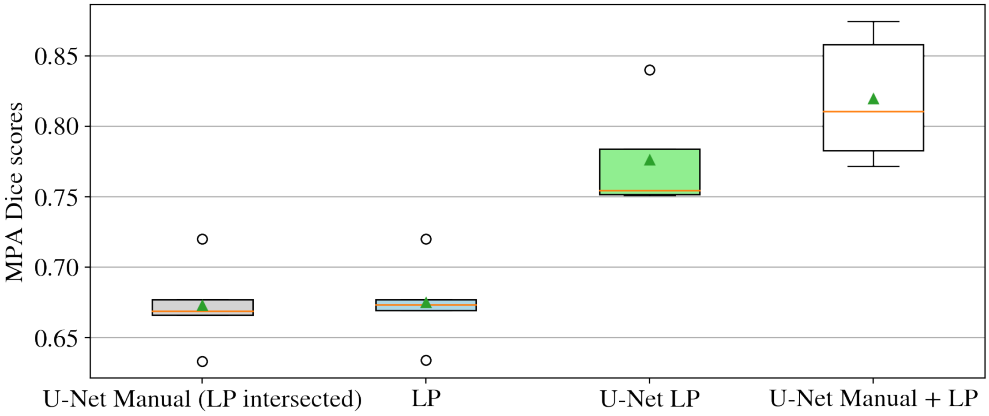


Figure 16: Main Pulmonary Artery (MPA) Dice scores

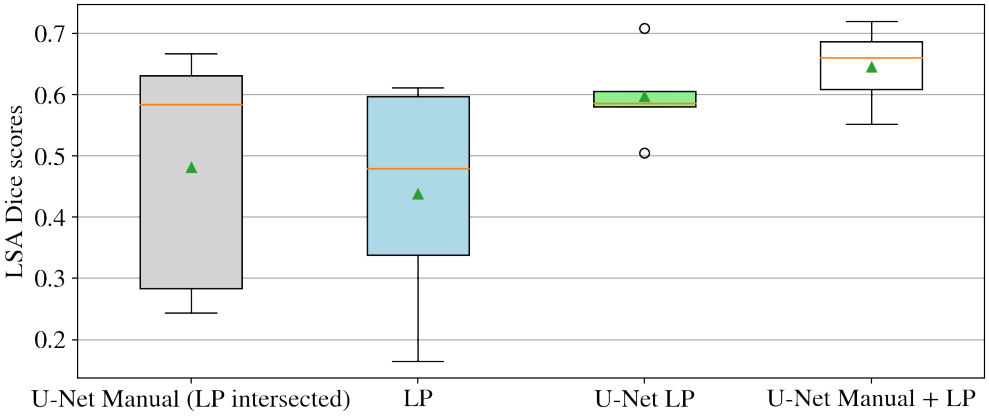


Figure 17: Left Subclavian Artery (LSA) Dice scores

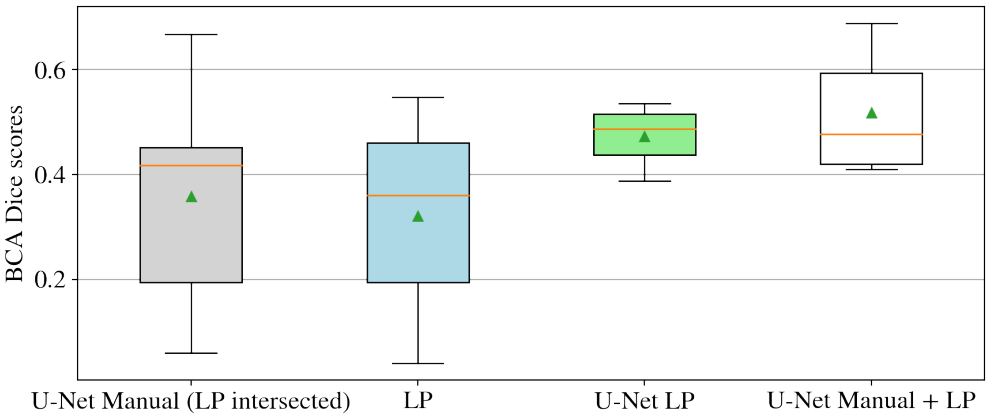


Figure 18: Brachiocephalic Artery (BCA) Dice scores

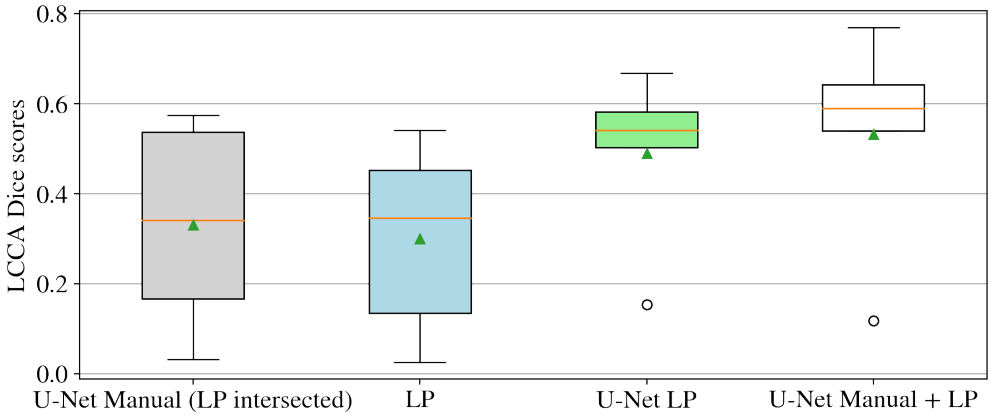


Figure 19: Left Common Carotid Artery (LCCA) Dice scores

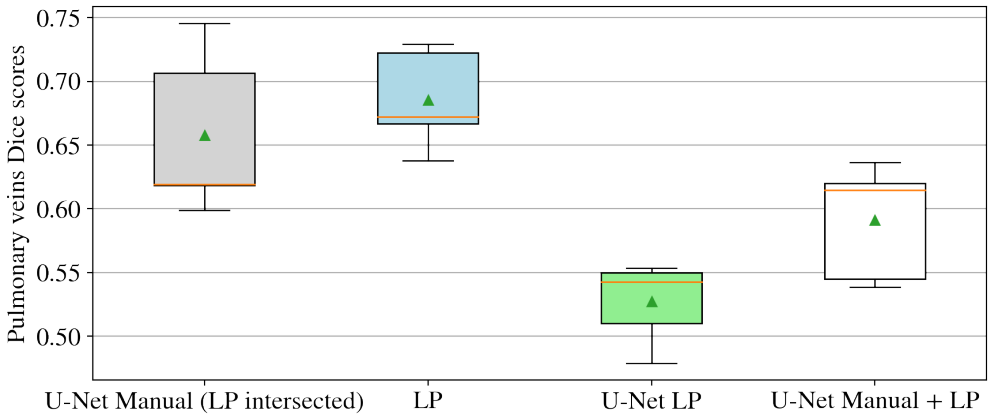


Figure 20: Pulmonary Veins Dice scores

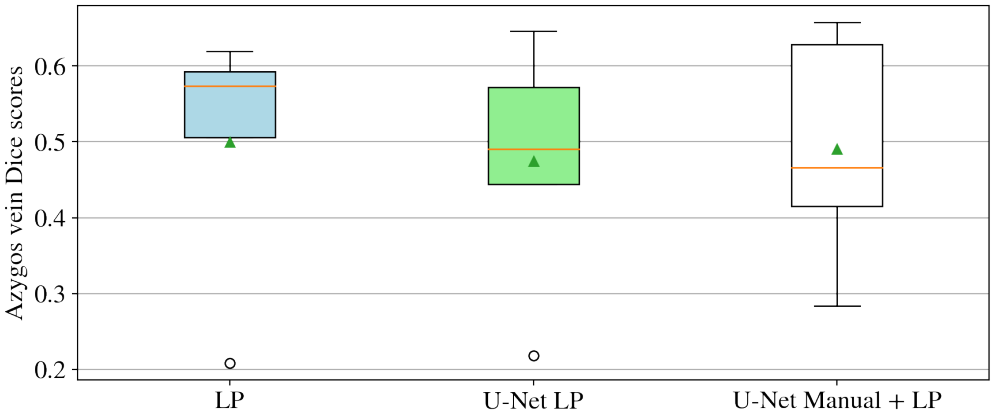


Figure 21: Azygos Vein Dice scores

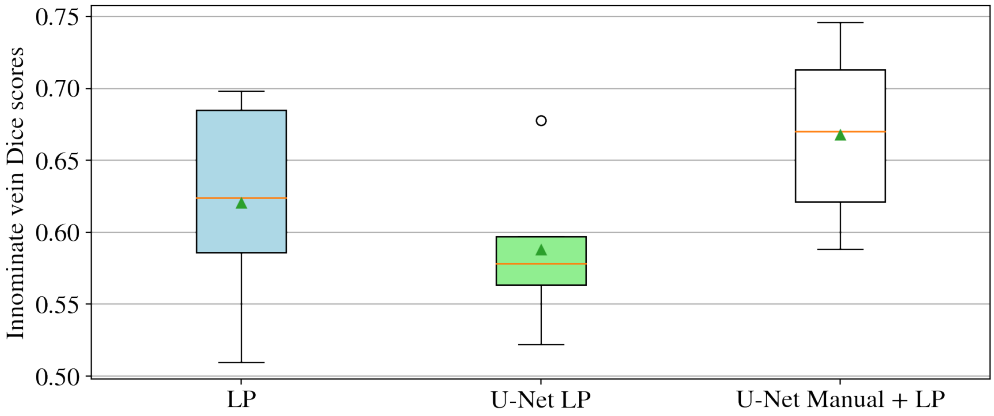


Figure 22: Innominate Vein Dice scores