**An Ensemble of U-Net Architecture Variants for Left Atrial Segmentation**

C.Wang1, M. Rajchl­3, A. D. C. Chan1, E. Ukwatta1,2­­­­

1Department of System and Computer Engineering, Carleton University, Ottawa, Canada

2 School of Engineering, University of Guelph, Guelph, ON, Canada

3Department of Computing, Imperial College London, London, UK

**ABSTRACT**

Segmentation of the left atrium and proximal pulmonary veins is an important clinical step for diagnosis of atrial fibrillation. However, the automatic segmentation of the left atrium from late gadolinium-enhanced magnetic resonance (LGE-MRI) images remains a challenging task due to differences in acquisition and large variability between individuals. Deep learning has shown to outperform traditional methodologies for segmentation in numerous tasks. A popular deep learning architecture for segmentation is the U-Net, which has shown promising results biomedical segmentation problems. Many newer network architectures have been proposed that leverage the base U-Net architecture such as attention U-Net, dense U-Net and residual U-Net. These models incorporate updated encoder blocks into the U-Net architecture to incrementally improve performance over the base U-Net. Currently, there is no comprehensive evaluation of performance between these models. In this study we (1) explore approaches for the segmentation of the left atrium based on different-Net architectures. (2) We compare and evaluate these on the STACOM 2018 Atrial Segmentation Challenge dataset and (3) ensemble these models to improve overall segmentation by reducing the internal variance between models and architectures. (4) Lastly, we define and build upon a U-Net framework to simplify development of novel U-Net inspired architectures. Our ensemble achieves a mean Dice similarity coefficient (DSC) of 92.1 ± 2.0% on a test set of twenty 3D LGE-MRI images, outperforming other fully automatic segmentation methodologies.

**Keywords**: U-Net, Deep learning, Convolutional Neural Network, Left Atrial Segmentation, Atrial Fibrillation

1. **INTRODUCTION**

Atrial fibrillation (AF) is a common form of arrhythmia in the left atrium (LA) that is associated with substantial morbidity and increased mortality [1, 2]. Furthermore, due to its prevalence among older populations AF is expected to increase its already substantial burden on the healthcare system [3, 4]. Ablation therapies are a common treatment for AF, however they are not always effective [5]. Atrial segmentations from magnetic resonance imaging (MRI) is an important step for clinical diagnosis and prognosis of AF [6]. In particular, segmentations from late gadolinium-enhanced magnetic resonance imaging (LGE-MRI) have applications in atrial fibrosis analysis and quantification of the atrial wall [7]. Segmenting LGE-MRI images is a challenging task because the size of the left atrium is thin, there is low contrast between atrial tissue and background, and there is a large morphological variability between individuals [8]. Manual segmentations are commonly used for most applications; however, these segmentations are labor-intensive and prone to interobserver variability. Therefore, there is a need for an automatic algorithm that can accurately segment the left atrial cavity and this algorithm has important clinical and research applications.

The objective of our work was to develop an algorithm for the automatic segmentation of the LA for patients with AF. We based our work on top of Olaf Ronneberger et al. [11] on their development of U-Net, a convolutional neural network for biomedical image segmentation and K. Kamnitsas et al. [20] on their development of Ensemble of Multiple Models and Architectures (EMMA), an ensemble of fully convolutional neural networks for brain tumor segmentation. Our solution proposed an ensemble of multiple U-Net architecture variants with the goal of averaging away model bias. We first defined a U-Net Framework to simplify the design process of building U-Net architecture variants. We then investigated the effects of different U-Net architecture variants extended from the framework on the STACOM 2018 Atrial Segmentation Challenge dataset. Next, we combined the results of these models into a classical ensemble approach via majority voting. Finally, we evaluated the performance of the ensemble compared to its composite models. We investigated U-Net variants built using the feature encoding blocks of convolutional layers, dense connections [14], inception modules [15], residual connections [17] and squeeze-and-excitation units [18]. Furthermore, we examined different skip connection variants such as elementwise addition, concatenation and attention connections [19].

1. **BACKGROUND**

The U-Net architecture follows an encoder/decoder pattern in which an input is transformed to a compressed hidden representation (encoding) before being transformed into the target solution space (decoding) [11]. In base U-Net, semantic features are extracted by convolutional layers into feature maps. Feature maps are max pooled to reduced dimensionality between every two convolutional layers, thus allowing the extraction of higher-level representations. In addition to this spatial reduction, the number of filters increase at each scale, increasing the number of features extracted. In the decoder, the feature maps are progressively upsampled to the desired image size. The number of filters in each feature extractor block is proportional to the previous feature extractor blocks by a factor of 2. A final convolutional layer is used to reduce the dimensionality of the outputted image. The U-Net architecture includes skip connections, in which previous feature maps where concatenated during the image decoding thus preserving some of the previously extracted features. This allows to employ features in the encoder to aid the reconstruction in the decoder.

Recently, multiple U-Net architecture variants have been proposed for different segmentation problems, each offering incremental developments over base U-Net. All these architectures follow the same basic design pattern of the original U-Net. We generalize the design philosophy of base U-Net by defining a U-Net Framework. The U-Net framework is composed of feature extractor blocks, followed by maxpooling or upsampling layers, with skip connections between feature maps of the same size. Feature extractor blocks serve to extract semantic features. They can be composed of one or more convolutional layers, depending on the block. Skip connections serve to relay previously extracted features to later blocks. Skip connections may also introduce new features, such as with attention connections [19]. Finally, the contraction of the feature space from the maxpooling layers serve to allow feature extractor blocks to extract higher level features and reduce the dimensionality of the image. Furthermore, maxpooling allows for a greater number of convolutional filters per feature extractor block by reducing the dimensionality and thus the memory size of the image. The pattern of contraction from maxpooling layers followed by expansion from upsampling layers give U-Net its name. Figure 1 is a visualization of the U-Net Framework we used to build our U-Net variant architectures and models.

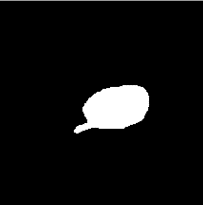
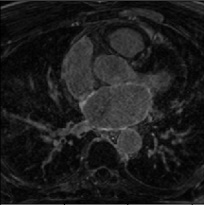
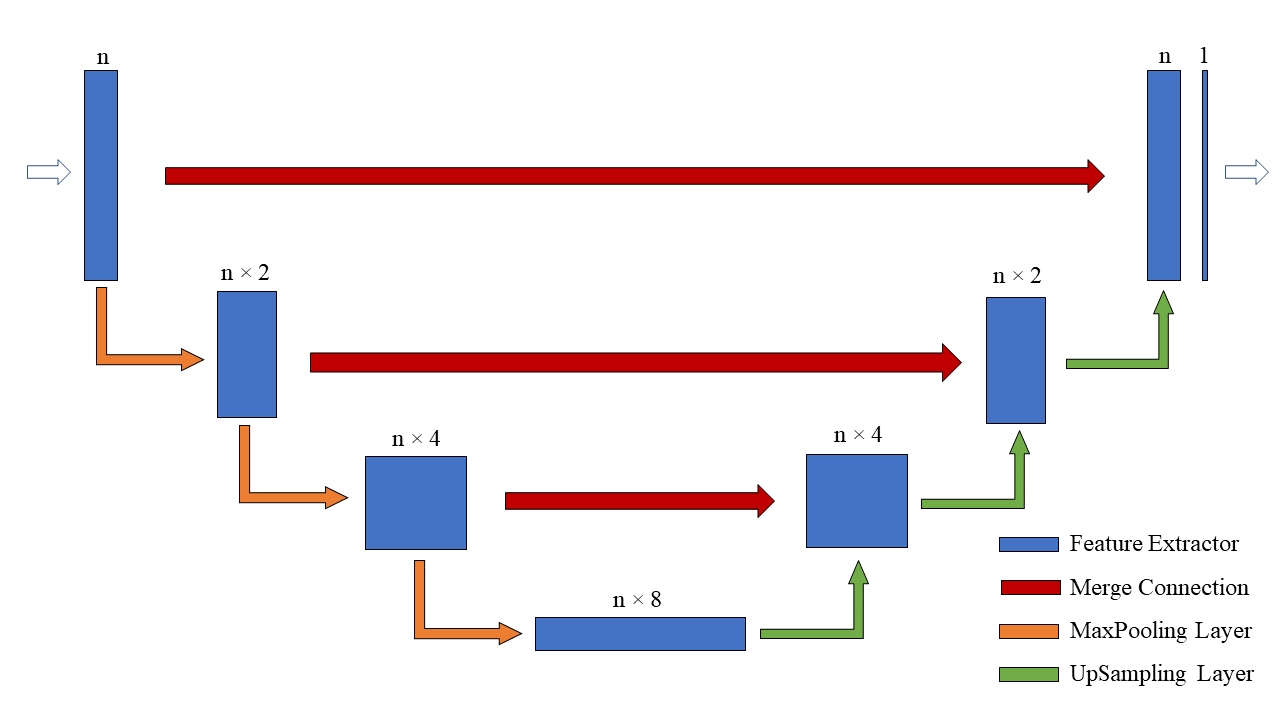
****

Figure 1. The U-Net Framework. Each feature extractor block generates n × 2m feature maps, which are then max pooled into the next feature extractor block. The U-Net Framework can be extended by using different feature extractor blocks and skip connections in conjunction.

Several previous methods have been reported on LA segmentation from LGE-MRI images. Zhu et al. described a fully automatic variational region growing method along with a global shape prior represented by Zernike moments [9]. Their method reported a Dice similarity coefficient (DSC) of 79 ± 5% for a dataset of 64 LA LGE-MRI images. Yang et al. [10] presented a semi-automated multi-atlas propagation based whole heart segmentation (WHS) to delineate the LA and PVs from LGE-MRI images. Multi-atlas propagation-based segmentation was first applied to the steady state free precision MRI image and the resulting segmentation is used as initial guess for multi-scale patch-based label fusion. Their method reported a mean DSC of 90.7 ± 7.5%.

1. **METHODOLOGY**

**3.1 Data Preprocessing**

In this study, we used 3D LGE MRI images of 100 patients with AF. These images were obtained from the STACOM 2018 Atrial Segmentation Challenge. Each MRI includes a binary label map of the left atrium and proximal pulmonary veins labelled by experts. All 3D MR images were collected using a whole-body MRI scanner with a resolution of 0.625 × 0.625 × 1.25 mm³. The challenge set was split into three groups, training (n = 70), validation (n = 10) and testing (n = 20). Prior to training, both the original and labelled MR images where center-cropped to 88 × 320 × 320 voxels then downsampled by a factor of two along the axial plane to reduce image memory size. Finally, individual 88 ×160 × 160 LGE MRI images where standardized by subtracting the image mean then dividing by the standard deviation to improve learning speed and the accompanying label maps were resampled via nearest neighbor interpolation.

**3.2 Training**

Every model in the ensemble was extended from the U-Net Framework and constructed with a single type of feature extractor block and skip connection. Every permutation of the five feature extractor blocks and the three skip connections was examined, resulting in a total of fifteen trained models. For every feature extractor we used 3D convolutions and 3D maxpooling to preserve spatial features across LGE MRI slices. Following every convolutional layer, we applied a rectified linear unit (ReLU) activation function then batch normalization. These models were trained for 50 epochs using an early stopping method based on validation data performance, the Adam optimizer [12] and a Jaccard’s distance loss function [13]. The Jaccard’s distance loss function was empirically found to converge faster than a Dice similarity coefficient loss function [13]. Models where trained on the downsampled images.

**3.3 Evaluation**

The predictions of the models were converted to binary images by thresholding the floating-point prediction values. The binary predictions of the models were compared to the manual segmentations of the LA using the Dice similarity coefficient. The predictions of the individual models were then ensembled though majority voting, according to Kamnitsas et al. [20]. Finally, all predictions where upsampled to the original size and evaluated using the original mask. Training and testing were computed on a NVIDIA P100 Pascal GPU using the Compute Canada cedar cluster. A complete visualization of the experimental design is shown in Figure 2.

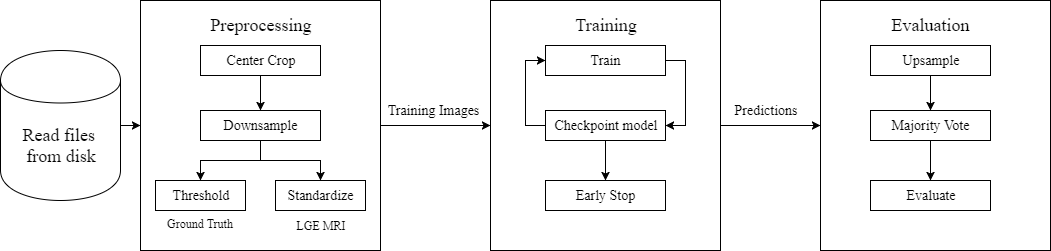


Figure 2. Complete experimental pipeline with preprocessing, training and evaluation.

1. **RESULTS**

The ensemble achieved a mean DSC of 92.1 ± 2.0% across the test set of 20 labeled LGE MRI images. While it is difficult to make comparisons between the methods due to differences in datasets, the results of the fully automatic method are comparable to those of the semi-automated method of Yang et al. [10]. The ensemble outperformed previous known fully automatic algorithms for left atrium segmentation on LGE MRI images. Figure 3 depicts an example segmentation cross-section from three of the 20 test LGE MRI images. Furthermore, the ensemble outperformed all the composite U-Net variant models, having the highest average DSC and lowest standard deviation.

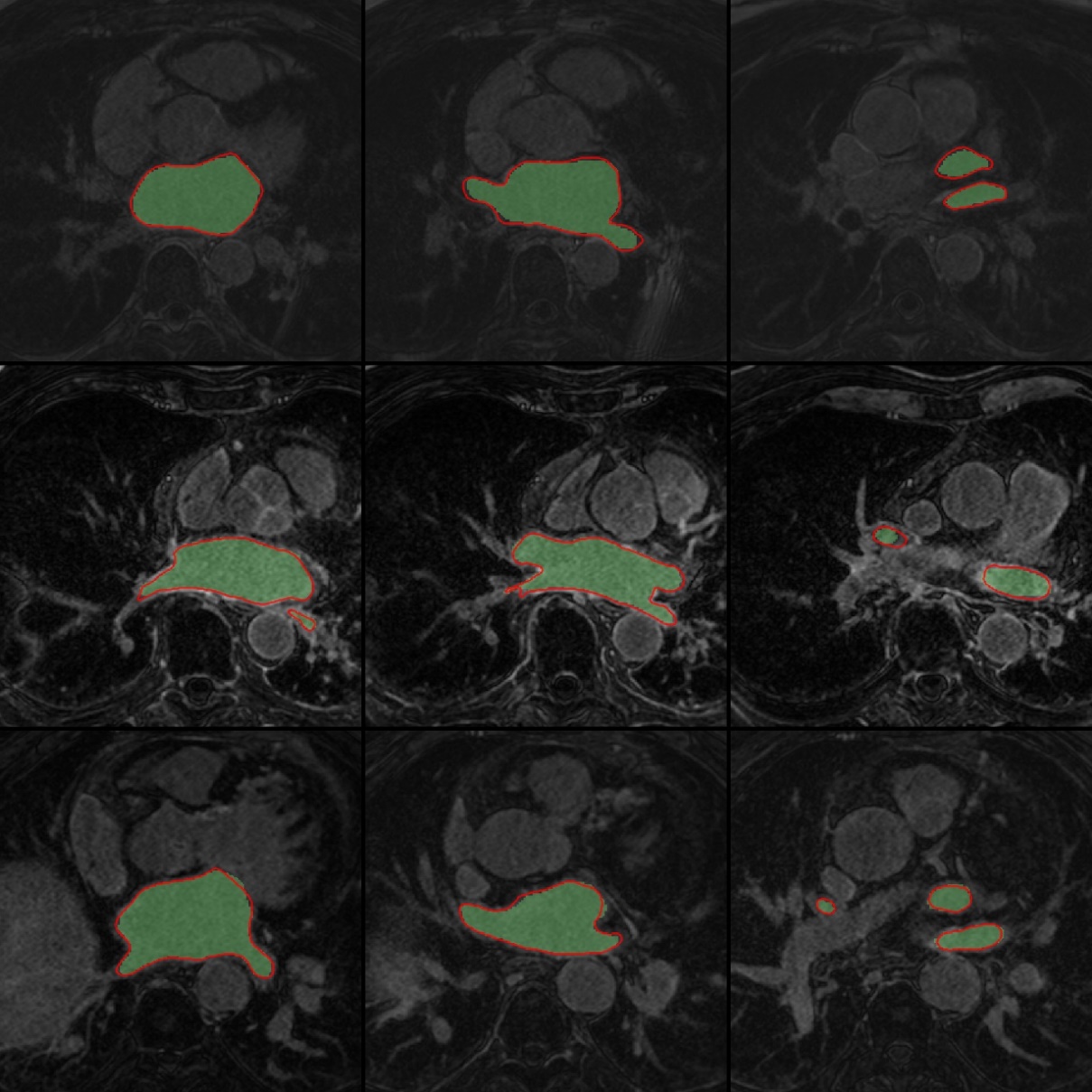


Figure 3. Segmentation slices from the 3D LGE-MRI images of three subjects. Each subject corresponds to a row. Ensemble segmentations are displayed in green and manual segmentation contours are displayed in red. Segmentations were chosen to show the variety of left atrial and proximal pulmonary vein morphologies.

Figure 5 provides a visualization of the performance differences across composite models. Apart from the ensemble, the Convolutional U-Net with addition connections, Dense U-Net with attention connections and Residual U-Net with Concatenation connections performed the best across the composite models, achieving a mean dice of 91.5 ± 2.5%, 90.6 ± 2.4%, and 91.4 ± 2.1% respectively. The Inception U-Net with concatenation connections performed the worst, achieving a mean dice of 87.8 ± 6.9%. Figure 4 depicts surface renderings of segmentations produced by the basic U-Net (Convolution/Concatenation), the EMMA ensemble and the expert manual segmentation.

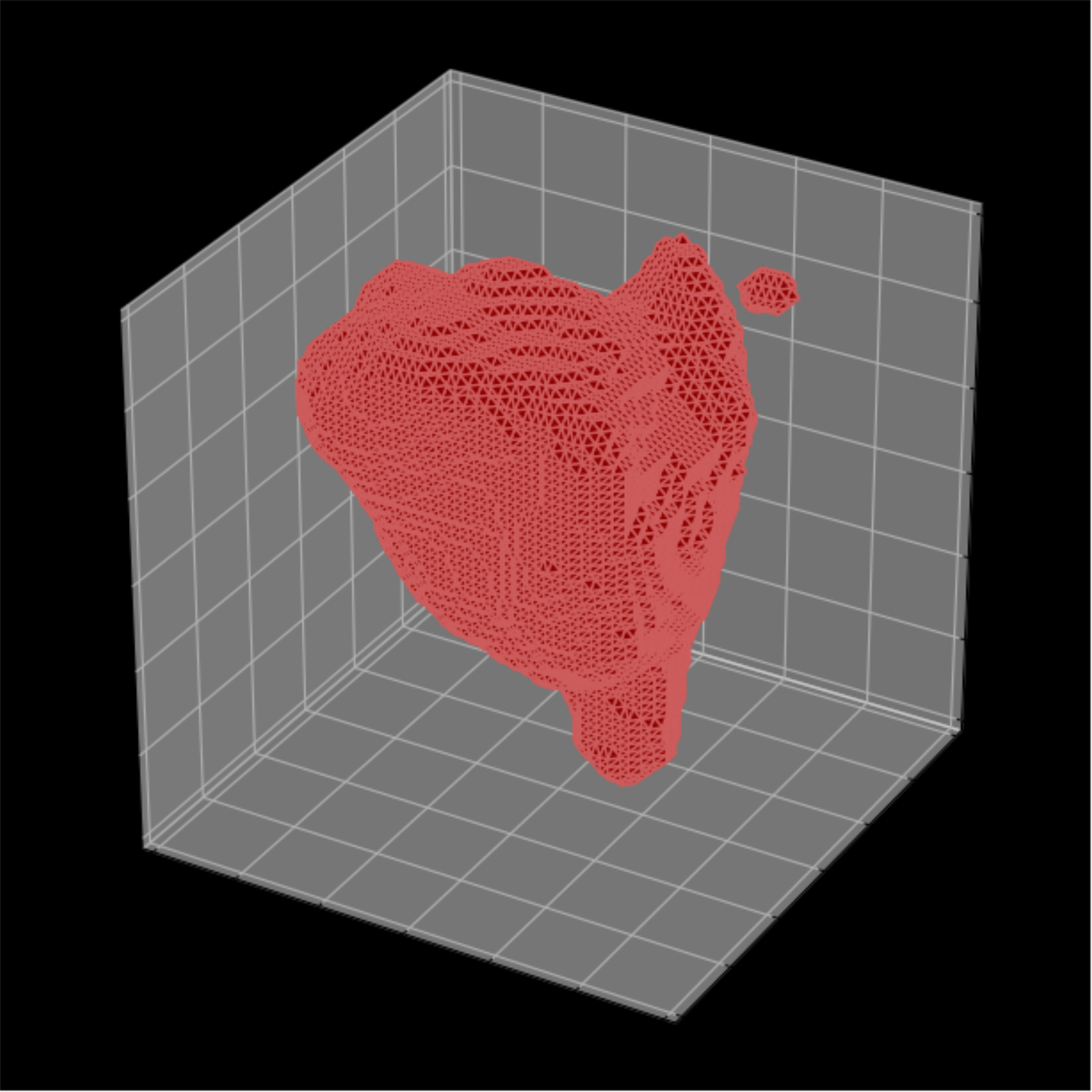
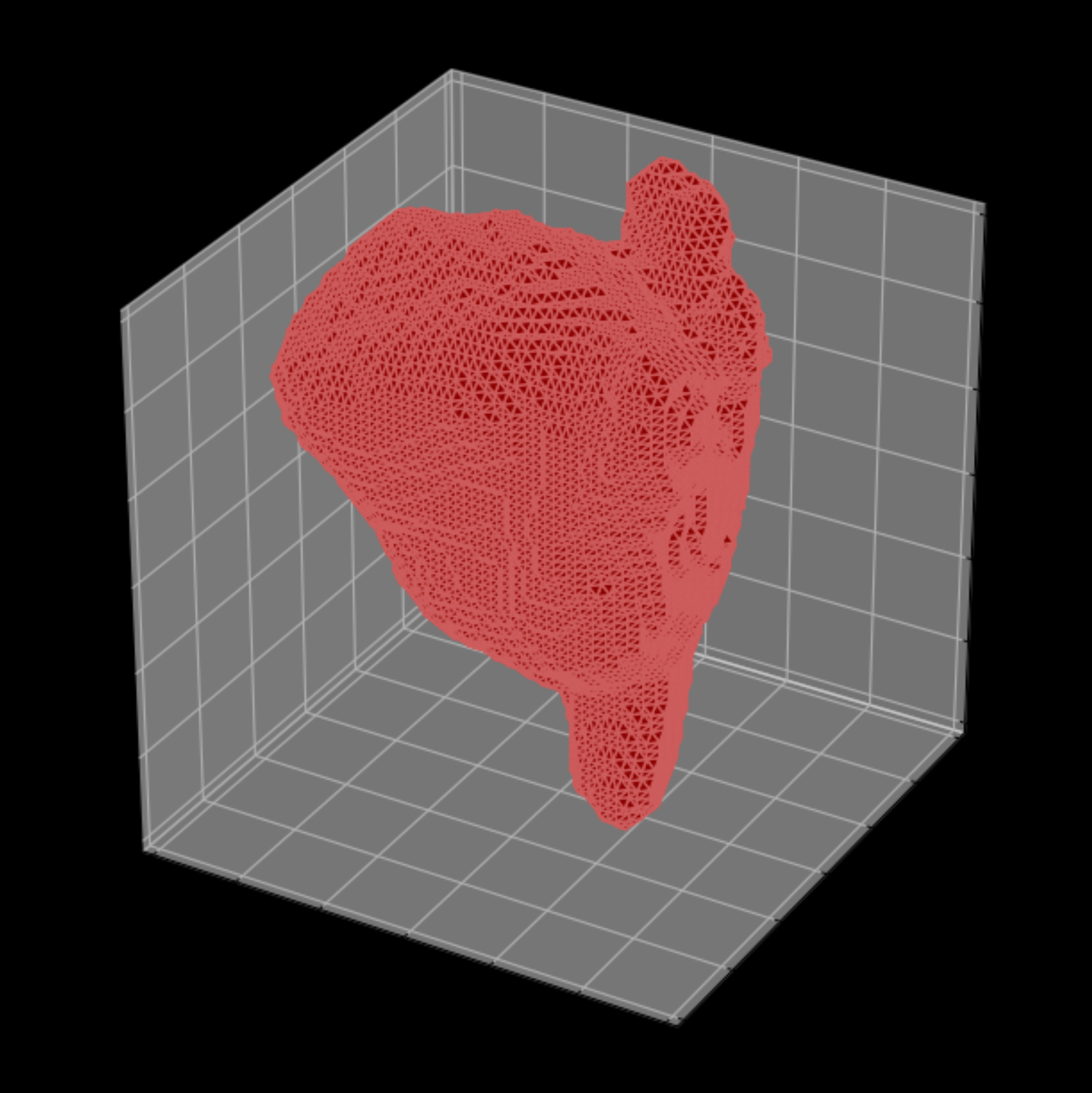
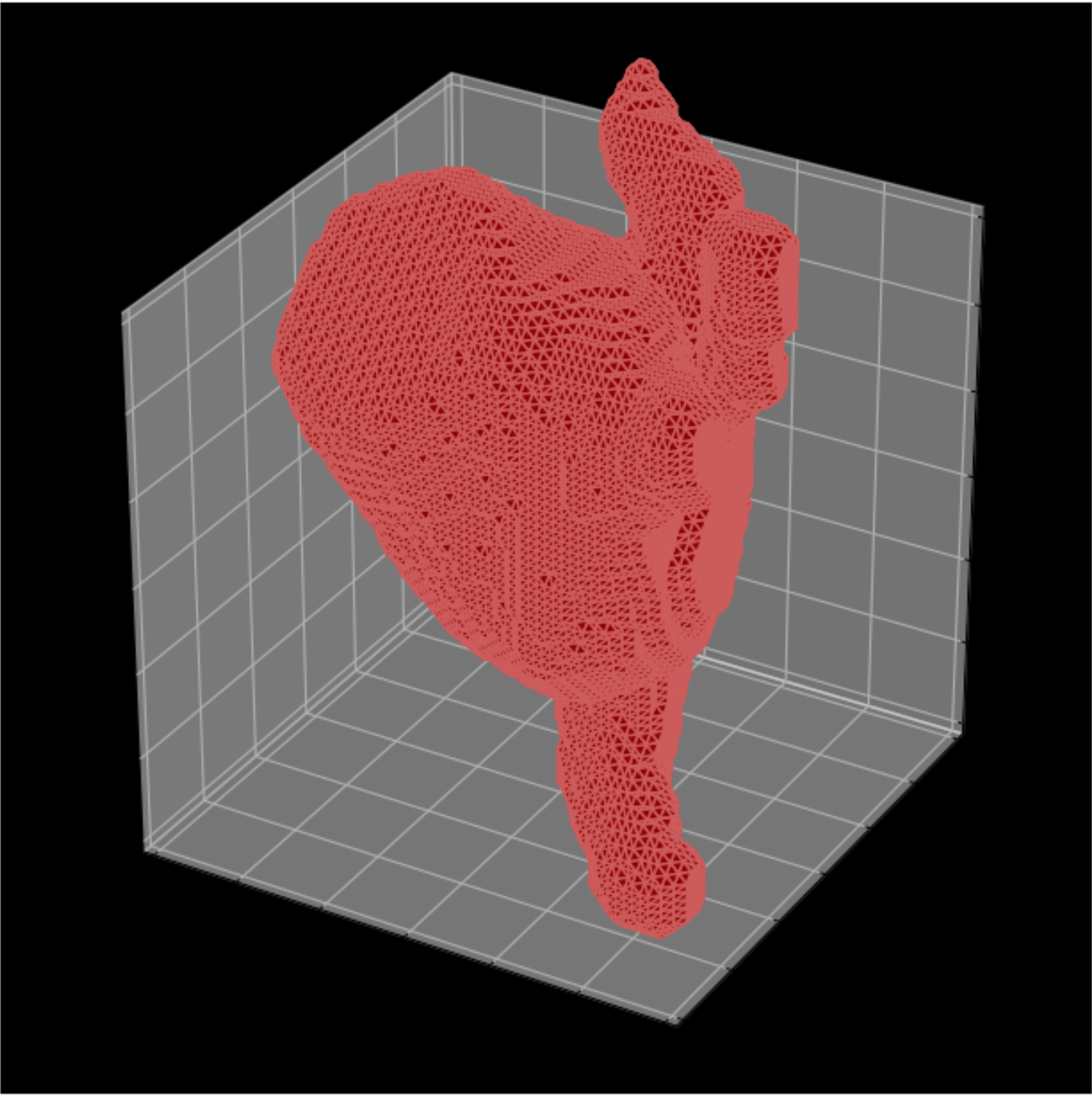
  

Figure 4. 3D visualizations of LA and PV segmentation predictions on a test image. Left is the base U-Net (Convolution/Concatenation) prediction with a DSC of 80.0%. Center is the ensemble prediction with a DSC of 90.0%. Right is the ground truth.

Qualitatively, we found that each model had produced different outliers. Furthermore, feature extractor blocks performed differently when paired with different skip connection methods. The goal of the ensemble was to average away the bias of individual architectures, thus improving the average performance.

1. Convolution/Addition

2. Convolution/Attention

3. Convolution/Concatenation

4. Dense/Addition

5. Dense/Attention

6. Dense/Concatenation

7. Inception/Addition

8. Inception/Attention

9. Inception/Concatenation

10. Residual/Addition

11. Residual/Attention

12. Residual/Concatenation

13. Squeeze Excitation/ Addition

14. Squeeze Excitation/ Attention

15. Squeeze Excitation/Concatenation

16. Multi-model Ensemble

Figure 5. Performance differences across composite models and the ensemble. Error bars show the highest and lowest segmentation DSC. The upper and lower edges of the box show the third and first quartile segmentation DSC respectively. The line shows the median segmentation DSC of the model.

(i)

(g)

(h)

1. **CONCLUSION**

The rapid growth of deep learning in the field of computer vision has resulted in new developments regarding convolutional neural networks, each improving over previous findings. In this study we compare and evaluate the performance of multiple U-Net variants and their EMMA ensemble on the STACOM 2018 Atrial Segmentation Challenge dataset. We show that feature blocks and skip connections differing impacts on the performance of atrial segmentations. Furthermore, we introduce the idea of a framework for the development of new U-Net architecture variants. Finally, we present an algorithm for the automatic segmentation of the left atrium; an ensemble of 15 U-Net variants. Our ensemble has a mean DSC of 92.1 ± 2.0% on a test set of twenty 3D MRI images, being competitive with other state of the art te­­­chniques. Left atrial segmentations have applications in ablation therapies for patients with atrial fibrillation. An automatic algorithm for this process is a step towards more successful treatments for one of the most common forms of heart arrhythmia.

1. **ACKNOWLEDGEMENTS**

Eranga Ukwatta would like to acknowledge the funding support through Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery grant program. Christopher Wang acknowledges the NSERC Undergraduate Student Research Award (USRA) and the Internship-Carleton University Research Experience for Undergraduate Students (I-CUREUS). Martin Rajchl holds an Imperial College Research Fellowship.

# **REFERENCES**

|  |  |
| --- | --- |
| [1] | E. Benjamin, P. Wolf, R. D'Agostino, H. Silbershatz, W. Kannel and D. Levy, "Impact of atrial fibrillation on the risk of death: the Framingham Heart Study," *Circulation,* vol. 10, no. 98, pp. 946-52, 1998. |
| [2] | R. Ana, S. Johansson, M.-A. Wallander and L. A. G. Rodríguez, "Risk of mortality in a cohort of patients newly diagnosed with chronic atrial fibrillation," *BMC Cardiovascular Disorders,* vol. 2, p. 5, 2002. |
| [3] | D. J. O'Reilly, R. B. Hopkins, J. S. Healey, P. Dorian, L. Sauriol, J. E. Tarride, N. Burke and R. A. Goeree, "The burden of atrial fibrillation on the hospital sector in Canada," *Canadian Journal of Cardiology,* vol. 29, no. 2, pp. 229-235, 2013. |
| [4] | S. Dai, A. Bienek, P. Walsh and A. Wielgosz, "The Burden of Hospitalized Atrial Fibrillation and Flutter in Canada," *Canadian Journal of Cardiology,* vol. 29, no. 10, p. 276, 2013. |
| [5] | H. Calkins, K. H. Kuck, R. Cappato et al. "“2012 HRS/EHRA/ECAS expert consensus statement on catheter and surgical ablation of atrial fibrillation: recommendations for patient selection, procedural techniques, patient management and follow-up, definitions, endpoints, and research trial design," *Europace,* vol. 14, no. 4, p. 528–606, 2012. |
| [6] | C. J. McGann, E. G. Kholmovski, R. S. Oakes et al. "New Magnetic Resonance Imaging-Based Method for Defining the Extent of Left Atrial Wall Injury After the Ablation of Atrial Fibrillation," *Journal of the American College of Cardiology,* vol. 52, no. 15, p. 1263–1271, 2018. |
| [7] | C. McGann, N. Akoum, A. Patel et al. "Atrial Fibrillation Ablation Outcome Is Predicted by Left Atrial," *Circulation: Arrhythmia and Electrophysiology,* vol. 7, no. 1, pp. 23-30, 2014. |
| [8] | L. Zhu, Y. Gao, A. Yezzi, R. MacLeod, J. Cates and A. Tannenbaum, "2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society," *Conf Proc IEEE Eng Med Biol,* 2012. |
| [9] | L. Zhu, Y. Gao, A. Yezzi and A. Tannenbaum, "Automatic Segmentation of the Left Atrium from MR Images via Variational Region Growing With a Moments-Based Shape Prior," *IEEE Transactions on Image Processing,* vol. 22, no. 12, pp. 5111-5122, 2013. |
| [10] | G. Yang, X. Zhuang, H. Khan et al "Multi-atlas propagation based left atrium segmentation coupled with super-voxel based pulmonary veins delineation in late gadoliniumenhanced cardiac MRI," in *SPIE MEDICAL IMAGING*, Orlando, 2017. |
| [11] | O. Ronneberger, P. Fischer and Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *edical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Munich, 2015. |
| [12] | D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization,* arXiv:1412.6980 [cs.LG], 2014. |
| [13] | A. H. Curiale, F. D. Colavecchia, P. Kaluza, R. A. Isoardi and G. Mato, *Automatic Myocardial Segmentation by Using A Deep Learning Network in Cardiac MRI,* arXiv:1708.07452 [cs.CV], 2017. |
| [14] | G. Huang, Z. Liu, L. v. d. Maaten and K. Q. Weinberger, *Densely Connected Convolutional Networks,* arXiv:1608.06993 [cs.CV], 2016. |
| [15] | C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, *Rethinking the Inception Architecture for Computer Vision,* arXiv:1512.00567 [cs.CV], 2015. |
| [17] | K. He, X. Zhang, S. Ren and J. Sun, *Deep Residual Learning for Image Recognition,* arXiv:1512.03385 [cs.CV], 2015. |
| [18] | J. Hu, L. Shen and G. Sun, *Squeeze-and-Excitation Networks,* arXiv:1709.01507 [cs.CV], 2017. |
| [19] | O. Oktay, J. Schlemper, L. Le Folgoc et al. *Attention U-Net: Learning Where to Look for the Pancreas,* arXiv:1804.03999 [cs.CV], 18. |
| [20] | K. Kamnitsas, W. Bai, E. Ferrante et al. *Ensembles of Multiple Models and Architectures for Robust Brain Tumour Segmentation,* arXiv:1711.01468 [cs.CV], 2017. |