2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

3D Attention M-net for Short-axis Left Ventricular Myocardium Segmentation in Mice MR cardiac Images

Luojie Huang[1], Andrew Jin[1], Jinchi Wei[1], Dnyanesh Tipre[1],
Chin-Fu Liu[1], Robert G. Weiss[2], and Siamak Ardekani[1,*]

***Abstract*— Small rodent cardiac magnetic resonance imaging (MRI) plays an important role in preclinical models of cardiac disease. Accurate myocardial boundaries delineation is crucial to most morphological and functional analysis in rodent cardiac MRIs. However, rodent cardiac MRIs, due to animal's small cardiac volume and high heart rate, are usually acquired with sub-optimal resolution and low signal-to-noise ratio (SNR). These rodent cardiac MRIs can also suffer from signal loss due to the intra-voxel dephasing. These factors make automatic myocardial segmentation challenging. Manual contouring could be applied to label myocardial boundaries but it is usually laborious, time consuming, and not systematically objective. In this study, we present a deep learning approach based on 3D attention M-net to perform automatic segmentation of left ventricular myocardium. In the deep learning architecture, we use dual spatial-channel attention gates between encoder and decoder along with multi-scale feature fusion path after decoder. Attention gates enable networks to focus on relevant spatial information and channel features to improve segmentation performance. A distance derived loss term, besides general dice loss and binary cross entropy loss, was also introduced to our hybrid loss functions to refine segmentation contours. The proposed model outperforms other generic models, like U-Net and FCN, in major segmentation metrics including the dice score (0.9072), Jaccard index (0.8307) and Hausdorff distance (3.1754 pixels), which are comparable to the results achieved by state-of-the-art models on human cardiac ACDC17 datasets.**

Clinical relevance **Small rodent cardiac MRI is routinely used to probe the effect of individual genes or groups of genes on the etiology of a large number of cardiovascular diseases. An automatic myocardium segmentation algorithm specifically designed for these data can enhance accuracy and reproducibility of cardiac structure and function analysis.**

I. Introduction

Cardiac magnetic resonance imaging (MRI) is the current gold standard for clinical quantitative cardiac analysis due to its accurate measurement of both anatomy and function [1], [2], which is crucial to any reliable cardiac analysis [3], [4].

Traditional segmentation tasks include image processing and machine learning methods, such as active shape models [5] and atlas-based methods [6]. However, these methods often require manual intervention, extensive feature engineering, and prior-knowledge incorporation.

This work was supported by several grants from the National Institutes of Health (HL130292, HL61912, HL63030)

[1]The Center for Imaging Science, Johns Hopkins University, Baltimore, MD 21218, USA.

[2]The Department of Medicine, Section of Cardiology, Johns Hopkins Medical Institutions, Baltimore, MD 21218, USA.

[*]Corresponding author: Siamak Ardekani sardekani@jhu.edu

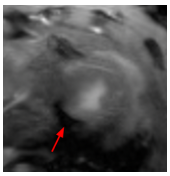


Fig. 1. **Signal loss within myocardium:** An example of left ventricular chamber with dark band (Red arrow) due to magnetic field inhomogeneity-induced signal void in myocardium.

In recent decades, deep learning algorithms, especially Convolutional Neural Networks (CNN), have succeeded in various automatic medical image segmentation tasks [7], [8]. Among those, Fully Convolutional Network (FCN) [9] and U-Net [10] have been utilized for cardiac segmentation. These works, however, mainly focused on 2D slices rather than 3D volumes due to the low out-of-plane resolution and motion artifacts in clinical cardiac MR scans. As a consequence, they do not account for inter-slice dependencies by performing slice-by-slice workflow. More recent works such as U-net [11], V-net [12], and M-net [13] have focused on network structure refinement to enhance feature learning. Compared to original FCN, these models reuse encoded features from inputs more effectively. Therefore, they are widely applied in biomedical image segmentation, especially in low signal-to-noise ratio (SNR) applications. Other state-of-the-art approaches to improve networks' performance include supervision enhancement, loss function modification, and attention techniques. For example, deep supervision in [14], [15], [16] is utilized to regularize network to capture more meaningful high-level features. Many loss functions, such as weighted cross-entropy, weighted Dice loss, focal loss [17], and distance loss [18], were proposed to overcome imbalanced data issues and to refine the boundaries of segmentation. More recently, attention gates [19] were highlighted in conditioning and regularizing deep learning networks to capture better local features in segmentation.

The aforementioned models aim at human cardiac MR segmentation. Hammouda et al. [20], recently proposed a novel FCN-based approach to perform the localization and segmentation of the LV cavity and myocardium. Compared to their works, we have focused on developing a deep learning technique that segments left ventricular myocardium in mice cardiac MRI at early and advanced disease stages (significant wall thinning at infacrion site), using higher magnetic fields. In comparison to human cardiac anatomy, a mouse

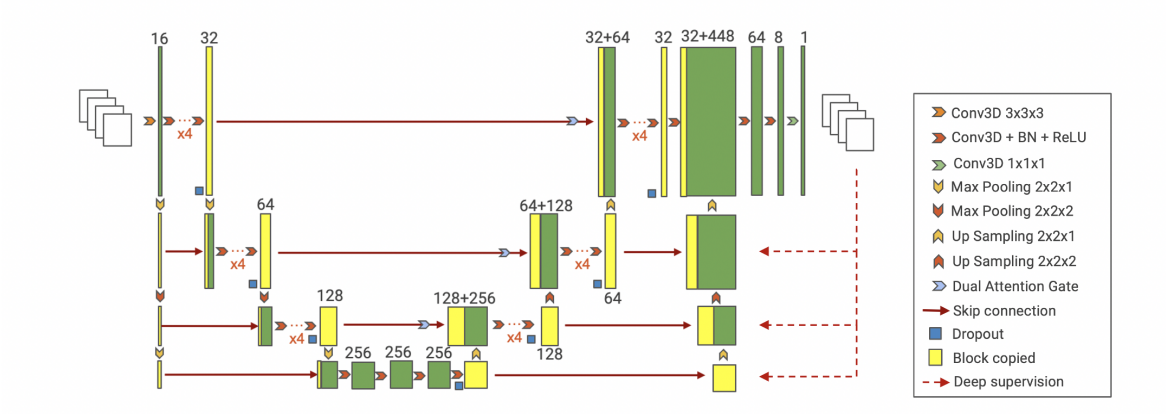


Fig. 2. **Attention M-net Architecture.** The model consists of encoder, decoder, and dual attention gates.

heart is much smaller, resulting in a lower SNR, relatively low spatial resolution, partial volume averaging and blurry boundaries. One way to improve SNR is to acquire images using higher magnetic field strength. However, magnetic field inhomogeneity that is typically observed at this higher field introduces image artifacts such as signal loss due to intra-voxel dephasing at the air-tissue boundaries in regions where the cardiac wall is in the vicinity of lung parenchyma (Figure 1). Similar to human cardiac MR studies, we also observe slice-to-slice misplacement due to sequential 2D acquisition scheme. To address these challenges, we have proposed a novel pipeline by applying modified M-net with dual attention gates to 3D volume segmentation, and introduced a distance derived loss function for optimal boundary refinement of segmentation.

The major contribution of our work is to use a unique in-house dataset of mice cine cardiac MRI data that contains both normal and diseased heart (myocardial infarction) to develop a new pipeline for 3D left ventricular myocardium segmentation based on 3D M-net architecture with attention to a new distance derived loss function. Our initial analysis shows that the proposed approach achieves better performance in Hausdorff distance than other current state-of-the-art generic models with a comparable dice score.

II. Methodology

The detailed model architecture is depicted in Figure 2. The proposed 3D Attention M-net consists of two major parts: an encoder and a decoder each with four levels. In addition, a dual attention gate is applied between the encoder and decoder at each level to refine the features from the encoder.

A. Encoder

Before the encoder, the cropped MRI volumes of size $96 \times 96 \times 36 \times 1$ are first inputted into a convolution layer to increase the feature channels to 16. Then, they are down-sampled by max-pooling layers as parallel inputs to corresponding encoder levels. Each encoder level consists of a cascade of 4 convolution blocks following a max-pooling layer for next-level encoders. Each convolution block includes a 3D convolution layer with a Batch Normalization layer and a 'ReLU' activation layer.

B. Decoder

The first half part of a decoder mirrors the encoder part. Then, a multi-scale fusion network concatenates and fuses the up-sampled features from decoders at each level into the final output. The fusion network includes two $3 \times 3 \times 3$ convolution layers and a $1 \times 1 \times 1$ convolution layer with a 'Sigmoid' activation layer to generate the final output.

C. Dual Attention Gate

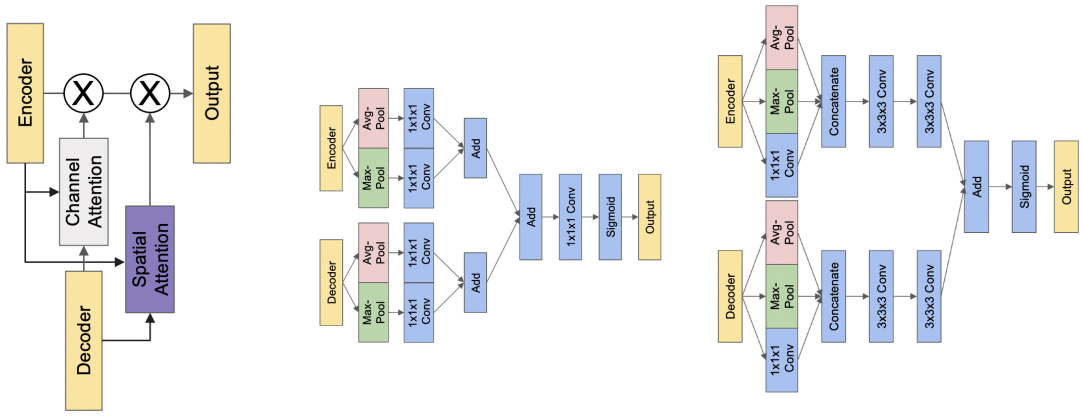
Between encoders and decoders at each level, we utilized the dual attention gate proposed by Khanh et al. [19] to regularize our network to focus on extracting meaningful contextual features from encoders and decoders along the spatial and channel dimensions. With the help of the attention gates, the decoders can more efficiently utilize encoded features to generate the final segmentation. As illustrated in Figure 3, the dual attention gate is made up of a spatial attention gate and a channel attention gate.

D. Loss functions

For our dataset, most errors occur around the segmentation boundaries due to the low SNR and artifacts from signal loss within myocardium. Therefore, in addition to a common loss function such as Generalized Dice Loss (GDL) and Balanced Cross-Entropy (BCE) to conquer the imbalanced voxel classes (relatively small myocardium volume compared to the background), we included an additional distance-derived loss (DDL) for contour refinement of the final segmentation.

1) ***Generalized Dice Loss (GDL):*** Sudre et al. [21] proposed an extension of dice loss with different weighting for each pixel class, which proved to be effective in unbalanced classes segmentation task.

2) ***Balanced Cross-Entropy (BCE):*** Xie et al. [22] proposed the BCE loss by adding a modulating factor in the original binary cross entropy loss to tackle class imbalance.



(a) Dual Attention Gate (b) Channel Attention (c) Spatial Attention

Fig. 3. **Architecture of Dual Attention Gate.** The dual attention gate (a), taking both encoder outputs and up-sampled decoder outputs as inputs, is made up by two sub-blocks: (b) channel attention block, and (c) spatial attention block.

*3) **Distance Derived Loss (DDL):*** Caliva et al. [18] created a custom penalty based loss function, using distance maps derived from ground truth to weight the cross entropy loss. Inspired by their work, we propose a new term of loss function, focusing on segmentation contour refinement. We first generate a weight map W based on the ground truth mask. For each pixel in the weight map, the weight is equal to the reciprocal of the closest Euclidean distance to the ground truth mask. Let $P_G = \{g_1, g_2, \ldots, g_n\}$ be the set of valid positions from ground truth mask, where $g_i = (g_{ix}, g_{iy})$. The difference between prediction and ground truth is then multiplied by the weight map $W(x,y) = 1/(\min_i \sqrt{(x-g_{ix})^2 + (y-g_{iy})^2} + 1)$. The voxels closer to the ground truth are assigned larger weights so that this loss term enables the model to focus more on contours. This is defined as:

$$DDL = \frac{\sum_x \sum_y (W(x,y) \cdot \mid P(x,y) - G(x,y) \mid)}{\sum_x \sum_y G(x,y)} \quad (1)$$

where $P(x,y)$ and $G(x,y)$ represent the labels of predicted and ground truth pixels, respectively.

In our model, we have used a combination of three aforementioned losses to address class imbalance (GDL, BCE), and to perform contour refinement (DDL). The combined loss is defined as: $L_{comb} = \lambda_1 \cdot GDL + \lambda_2 \cdot BCE + \lambda_3 \cdot DDL$ Where $\cdot$ represents scalar product. We have also introduced deep supervision on the decoders at each level in our model. The final loss function is a weighted sum of the losses from the decoders at all 4 levels: $L_{deep} = \sum_{n=1}^{4} w_n \cdot L_{comb_n}$

III. Experiment

A. Dataset

As a part of an ongoing project, in-vivo heart MR images of adult male wild type mice and Galectin-3 knockout mice were acquired at different stages of disease following sham surgery or induction of myocardial infarction (pre-op, and days 1, 7, and 56) using an MRI spectrometer equipped with a 11.7T magnet. Details of image acquisition has been provided elsewhere [24]. The animal protocol was approved by the Institutional Animal Care and Use Committee of the Johns Hopkins University (protocol numbers: MO16A398 and MO19E374).

The myocardium was manually segmented in each short-axis image by two people (inter-rater agreement of 89% based on dice scores) using a free semi-automatic software package in MATLAB called 'Segment' [25]. In principal, manual segmentation of the left ventricle was performed following recommendations by Schulz-Menger et. al [26]. Endocardial and epicardial contours were traced on short-axis cine images at several time points during the cardiac cycle, with simultaneous viewing of short and long-axis images of the same region, if applicable. Papillary muscles and trabecular tissue were excluded. The most apical region was identified as a section where left ventricular epicardium was visible, while the most basilar section was selected at the level of outflow tract. Delineation excluded the aortic valve cusps resulting in a myocardial segmentation that resembled crescent shape. The contours were then interpolated across all time points of the cardiac cycle, and manually examined to correct for any interpolation errors. Segmentations were then reviewed by an expert with more than 14 years of experience in cardiac MR research. We used 1114 fully annotated volumes that were collected from the repeated acquisitions (each contained multiple time frames per cardiac cycle) of several animals who underwent surgery and followed over the course of disease. The data was split into three subsets: 700 for training, 200 for validation, and 214 for testing. To prevent information leakage, scans from same rodents are kept in the same partition while ensuring similar myocardial volume distribution across all groups. Since the gap between slices (0.8 mm) is significantly larger than in-plane pixel size (0.1307 mm), linear interpolation was performed to upsample the inter-slice gap to 0.16 mm for each image volume. Thus, the data size was augmented to $96 \times 96 \times 36 \times 1$.

To improve robustness and generalization, we also performed data augmentation by applying random shift (up to

TABLE I
SEGMENTATION EVALUATIONS OF COMPARATIVE MODELS

Model	Parameters	Dice Score	Jaccard Index	Hausdorff Distance	Sensitivity	Specificity	PPV	NPV
2D U-net	24,902,996	0.8677	0.7675	4.83 (0.6315)	0.8762	**0.9957**	0.8604	0.9963
3D U-net	24,502,812	0.8732	0.7728	4.62 (0.6037)	0.8975	0.9951	0.8475	**0.9969**
DeepMedic	24,505,161	0.8883	0.7995	4.59 (0.6003)	0.9109	0.9812	0.8693	0.9877
3D FCN	24,128,560	0.8894	0.8015	4.32 (0.5641)	0.9212	0.9867	0.8636	0.9927
3D Att-Mnet (w/o distance loss)	24,657,034	0.8953	0.8111	4.00 (0.5231)	**0.9286**	0.9869	0.8664	0.9934
3D Att-Mnet (w/ distance loss)	24,657,034	**0.9072**	**0.8307**	**3.18 (0.4150)**	0.9249	0.9891	**0.8913**	0.9927

- The unit of Hausdorff distance is $pixels$(mm). Smaller Hausdorff distances mean better alignment between contours of ground truth and prediction.
- PPV: positive predictive value; NPV: negative predictive value.

15 pixels), uniformly varying rotation (within 15 degrees), and horizontal/vertical flips, with a probability of 0.5.

B. Implementation Details

The model is built by tensorflow and Keras in python on an Intel Xeon E5-2620 v4 CPU model with 2 Titan V GPUs. After hyperparameter tuning, the deep supervision loss L_{deep}, as described in the previous section, was set with weights w_i of the first level at 1 and others at 0.5. For the combined loss L_{comb} at each level, we set $\lambda_1, \lambda_2, \lambda_3$ to 1.0, 1.0 and 3.0. The model was optimized by Adam method with a learning rate of 10^{-4} and batch size of 8. For all experiments, we trained the networks from scratch for 150 epochs. The learning rate was automatically adjusted by monitoring callback on plateau, allowing the optimizer to more efficiently reach the local minimum. The monitored callback is the generalized dice loss of the final level output with setting parameters $patience$ to 8, $factor$ to 0.5 and min_delta to 0.0001.

C. Results and Comparison with other models

We compare the proposed model with 4 state-of-the-art segmentation models: 2D U-net [10], 3D U-net [11], DeepMedic [23] and 3D FCN [9]. All models use similar sizes of trainable variables for fair comparison and are trained with the loss functions proposed in their original publications. The segmentation performance is quantitatively evaluated by metrics including mean Dice Score, Jaccard Index, Hausdorff Distance, Sensitivity and Specificity, Positive predictive value (PPV) and Negative predictive value (NPV). Results are summarized in Table 1.

According to Table 1, our proposed models outperform all the state-of-the-art models in dice score, Jaccard index, and Hausdorff distance, with comparable performance in all other metrics. The mean dice score of our 3D attention M-net is 0.9072, which is better than other methods: 3D U-net (0.8732), DeepMedic (0.8883) and 3D FCN (0.8894) with similar Sensitivity and Specificity. This indicates that our model has better general myocardium segmentation performance than other models in the mice MR cardiac images. Furthermore, compared to the dice score agreement between our two annotators, which is 0.89 over 150 samples, our model has achieved comparable performance (0.9072). The proposed model has also demonstrated a potentially comparable performance on the mice dataset as compared to the state-of-the-art myocardium segmentation models on the human cardiac dataset ACDC17 [27], whose dice scores are around 0.9 on ACDC17 dataset.

In Table 1, there is a noticeable improvement from 2D U-net to 3D U-net in Hausdorff distances (-0.2142, $p = 0.0068$ using two-sided Wilcoxon rank sum test). This could be the benefit of the 3D model's ability to incorporate inter-slice contextual information to achieve better 3D segmentation when compared to the 2D model. The contextual information from adjacent slices could enable better segmentation in each slice, especially for the apical slice (Slice #8 in Fig. 4), which typically contains smaller myocardium and suffers from larger signal loss due to its adjacency to the lung pareynchyma. In the selected case shown in Figure 4, the segmentation of the 2D Unet, compared to all other 3D whole-volume networks, has noticeably worse performance and boundary refinement. DeepMedic also has the same problem at the last slice. This could be because DeepMedic is a local batch-wise 3D network, which does not fully utilize the whole volume contextual information.

Among all 3D models, our proposed models stands out with a noticeably smaller Hausdorff distance. The Att-Mnet without the proposed distance loss has already successfully refined the final segmentation boundary to achieve 4.0026px in Hausdorff distance, which is at least 0.3px less than all other competitive models in Table 1. This shows that our proposed M-net with dual attention gates (Att-Mnet) has better segmentation accuracy than others with similar computational complexity (the number of the trainable parameters). Moreover, our Att-Mnet with the proposed distance loss further reduces the Hausdorff distance to 3.1754px. This is a major improvement -0.8272px from our Att-Mnet without the proposed distance loss function. This demonstrates that the proposed distance loss has enabled our model to effectively refine segmentation contours. This will improve the accuracy of global left ventricular volumetric measurements such as end-diastole and ends-stole volumes, ejection fraction, and myocardial mass, as well as the accuracy of computational analysis of ventricular shape and motion.

IV. CONCLUSION

In this study, a 3D M-net model with dual attention gates and a distance derived loss term was proposed for myocardium segmentation, tackling the challenges of low image quality and signal loss in mice MR cardiac images. The experiment results indicate that our proposed model not

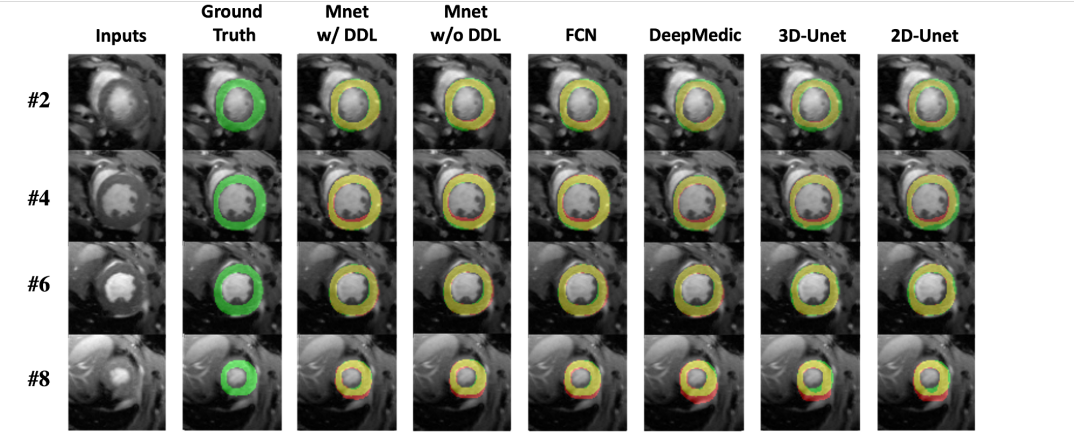


Fig. 4. **Comparison of partial segmentation results on a selected volume.** Column 1: MRI; Column 2: manual ground truth; Column 3-8: yellow, green and red masks represent true positive, false negative and false positive, respectively. Mnet: 3D attention M-net; DDL: distance derived loss.

only outperforms other benchmark models, including U-Net and FCN, in dice score and contour agreement (Hausdorff distance), but also has very comparable performance as experienced human annotators. The proposed model enables a fast, objective, and accurate myocardial boundaries delineation of rodent cardiac MRIs, and has strong potential application in morphological and functional analysis for preclinical cardiac models.

REFERENCES

[1] C. B. Marcu, A. M. Beek, and A. C. van Rossum, "Clinical applications of cardiovascular magnetic resonance imaging," Canadian Medical Association Journal, vol. 175, no. 8, pp. 911–917, Oct. 2006.

[2] D. Pennell, "Clinical indications for cardiovascular magnetic resonance (CMR): Consensus Panel report?," European Heart Journal, vol. 25, no. 21, pp. 1940–1965, Nov. 2004.

[3] J. Schwitter et al., "MR-IMPACT II: Magnetic Resonance Imaging for Myocardial Perfusion Assessment in Coronary artery disease Trial: perfusion-cardiac magnetic resonance vs. single-photon emission computed tomography for the detection of coronary artery disease: a comparative multicentre, multivendor trial," European Heart Journal, vol. 34, no. 10, pp. 775–781, Mar. 2013.

[4] P. S. Douglas et al., "Outcomes of Anatomical versus Functional Testing for Coronary Artery Disease," N Engl J Med, vol. 372, no. 14, pp. 1291–1300, Apr. 2015.

[5] C. Petitjean et al., "Right ventricle segmentation from cardiac MRI: A collation study," Medical Image Analysis, vol. 19, no. 1, pp. 187–202, Jan. 2015.

[6] V. Tavakoli and A. A. Amini, "A survey of shaped-based registration and segmentation techniques for cardiac images," Computer Vision and Image Understanding, vol. 117, no. 9, pp. 966–989, Sep. 2013.

[7] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges," J Digit Imaging, vol. 32, no. 4, pp. 582–596, Aug. 2019.

[8] D. Ciresan et al., "Deep neural networks segment neuronal membranes in electron microscopy images", Adv. Neural Inf. Process. Syst., pp. 2843-2851, 2012.

[9] P. V. Tran, "A fully convolutional neural network for cardiac segmentation in short-axis MRI," 2016, [online] Available: https://arxiv.org/abs/1604.00494.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in MICCAI 2015, vol. 9351, pp. 234–241.

[11] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," in MICCAI 2016, vol. 9901, pp. 424–432.

[12] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," in 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, Oct. 2016, pp. 565–571.

[13] R. Mehta and J. Sivaswamy, "M-net: A Convolutional Neural Network for deep brain structure segmentation," ISBI 2017, Melbourne, VIC, Australia, 2017, pp. 437-440.

[14] C.-Y. Lee, S. Xie, P. Gallagher et al., "Deeply-Supervised Nets," 2021, [Online] Available: http://arxiv.org/abs/1409.5185.

[15] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, vol. 11045, 2018, pp. 3–11.

[16] H. Huang et al., "UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation," 2021, http://arxiv.org/abs/2004.08790.

[17] S. Jadon, "A survey of loss functions for semantic segmentation," In: CIBCB 2020, Via del Mar, Chile, pp. 1-7, 2020.

[18] F. Caliva, C. Iriondo, A. M. Martinez, et al., "Distance map loss penalty term for semantic segmentation," 2019, https://arxiv.org/abs/1908.03679.

[19] T. L. B. Khanh et al., "Enhancing U-Net with Spatial-Channel Attention Gate for Abnormal Tissue Segmentation in Medical Imaging," Applied Sciences, vol. 10, no. 17, p. 5729, Aug. 2020.

[20] K. Hammouda et al. "A New Framework for Performing Cardiac Strain Analysis from Cine MRI Imaging in Mice," Sci Rep 10, 7725, 2020.

[21] C. H. Sudre et al., "Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations," In: DLMIA 2017, vol. 10553, pp. 240–248, 2017.

[22] S. Xie and Z. Tu, "Holistically-nested edge detection," In ICCV 2015, pp. 1395–1403.

[23] K. Kamnitsas, C. Ledig, et al., "Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation," Medical Image Analysis, 2016.

[24] B. C. Lee et al., "Diffeomorphic Upsampling of Serially Acquired Sparse 2D Cross-Sections in Cardiac MRI," Conf Proc IEEE Eng Med Biol Soc, vol. 2019, pp. 4491–4495, Jul. 2019, doi: 10.1109/EMBC.2019.8856317.

[25] E. Heiberg, J. Sjögren, M. Ugander, M. Carlsson, H. Engblom, and H. Arheden, "Design and validation of Segment - freely available software for cardiovascular image analysis," BMC Med Imaging, vol. 10, no. 1, p. 1, Dec. 2010.

[26] J. Schulz-Menger et al., "Standardized image interpretation and post-processing in cardiovascular magnetic resonance - 2020 update: Society for Cardiovascular Magnetic Resonance (SCMR): Board of Trustees Task Force on Standardized Post-Processing," J Cardiovasc Magn Reson, vol. 22, no. 1, p. 19, Mar. 2020.

[27] O. Bernard et al., "Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved?," in IEEE Transactions on Medical Imaging, vol. 37, no. 11, pp. 2514-2525, Nov. 2018.