Three-Dimensional Embedded Attentive RNN (3D-EAR) Segmentor for Left Ventricle Delineation from Myocardial Velocity Mapping

Mengmeng Kuang[1] ✉, Yinzhe Wu[1,3], Lufei Gao[4], Diego Alonso-Álvarez[5], David Firmin[1,2], Jennifer Keegan[1,2], Peter Gatehouse[1,2], and Guang Yang[1,2][0000−0001−7344−7733] ✉

[1] National Heart and Lung Institute, Imperial College London
[2] Cardiovascular Research Centre, Royal Brompton Hospital
[3] Department of Bioengineering, Imperial College London
[4] Department of Electronic and Information Engineering, The Hong Kong Polytechnic University
[5] Research Computing Service, Information and Communication Technologies, Imperial College London
mmkuang@live.com and g.yang@imperial.ac.uk

Abstract. Myocardial Velocity Mapping Cardiac MR (MVM-CMR) can be used to measure global and regional myocardial velocities with proved reproducibility. Accurate left ventricle delineation is a prerequisite for robust and reproducible myocardial velocity estimation. Conventional manual segmentation on this dataset can be time-consuming and subjective, and an effective fully automated delineation method is highly in demand. By leveraging recently proposed deep learning-based semantic segmentation approaches, in this study, we propose a novel fully automated framework incorporating a 3D-UNet backbone architecture with **E**mbedded multichannel **A**ttention mechanism and LSTM based **R**ecurrent neural networks (RNN) for the MVM-CMR datasets (dubbed 3D-EAR segmentor). The proposed method also utilises the amalgamation of magnitude and phase images as input to realise an information fusion of this multichannel dataset and exploring the correlations of temporal frames via the embedded RNN. By comparing the baseline model of 3D-UNet and ablation studies with and without embedded attentive LSTM modules and various loss functions, we can demonstrate that the proposed model has outperformed the state-of-the-art baseline models with significant improvement.

Keywords: Cardiac MRI · Myocardial Velocity Mapping · Segmentation · Attention Mechanism · LSTM

1 Introduction

Healthy functioning of the left ventricle requires complex motion and all parts of the myocardium must perform synergistically in order to ensure the heart

to pump efficiently. In certain pathologies, early regional myocardial instability may be compensated for by altered movement in other areas in order to maintain the ventricular function. Global myocardial velocities are widely used in clinical practice; however, the global measurement might only be detectable when the condition has advanced to a point where compensation is no longer possible. By additional measurement of local myocardial dynamics, myocardial stability can be more specifically quantified and potential cardiovascular disease can therefore be detected earlier [1].

Among different cardiac MR (CMR) techniques, there are a few methods that can be used to calculate global and regional myocardial dynamics [2]. Myocardial Velocity Mapping CMR (MVM-CMR) [3] can potentially provide both high spatial and temporal resolution, and has clear advantages compared to the blood velocity scans that are commonly used in clinical.

Accurate segmentation of the left ventricle (LV) is the first step and a prerequisite for robust and reproducible global and regional myocardial velocity estimation. Conventional manual segmentation on the MVM-CMR dataset can be extremely time-consuming considering both the high spatial and temporal resolution of the dataset. Such manual segmentation is also limited by clinician's experience and potential human operator fatigue may also affect the delineation accuracy. Therefore, an efficient and robust automated LV segmentation method for the multichannel (i.e., magnitude and three velocity channels) and multi-frame (i.e., temporal frames of the cardiac "movie" acquired) MVM-CMR data is necessary for the clinical deployment of the global and regional velocities estimation.

Development in deep learning represents a major leap for digital healthcare. Several research studies have demonstrated promising results for the anatomical and pathological segmentation of the heart from CMR images, e.g., whole heart segmentation [4], LV segmentation [5], left atrial segmentation [6] and atrial scar delineation [7]. A more detailed review of the segmentation of cardiac images can be found elsewhere [8].

Despite successful applications of deep learning based techniques for CMR data segmentation, a plain deployment of existing methods for the MVM-CMR data can be challenging, but more informative. This is due to (1) multi-frame temporal MVM-CMR data may require a more complicated network design to ensure both accurate slice-wise delineation and continuity in the temporal dimension and (2) MVM-CMR data have multiple channels, e.g., magnitude and phase channels, that can provide richer information of the LV anatomy, but how to explore such informative multichannel data is still an open question.

Inspired by recent progress on semantic segmentation, e.g., UNet [9], we propose a novel **E**mbedded multichannel **A**ttentive **R**ecurrent neural networks (RNN), abbreviated 3D-EAR segmentor, for LV delineation from MVM-CMR datasets. The proposed 3D-EAR segmentor consists of three major components: (1) a 3D-UNet based backbone network, (2) embedded attention modules to enhance the network skip connections for more accurate localisation of the LV anatomy, and (3) long short-term memory (LSTM) based RNN modules to learn

the temporal information of the multi-frame context at the bottom of the U-shaped network. In doing so, the proposed method can leverage the amalgamation of magnitude and phase images as more informative input to realise an effective information fusion of this multichannel dataset. Besides, the temporal dimension continuity can be ensured by exploring the correlations of temporal frames via the embedded LSTM. In addition, by varying different loss functions (i.e., Cross-Entropy loss, Dice loss [10], and Dice-IoU [1] loss [11]), we perform ablation studies to find the optimal architecture of the proposed network.

By validation on MVM-CMR data collected from healthy controls, our proposed 3D-EAR segmentor achieves superior LV segmentation performance compared to state-of-the-art baseline models at the patient-level.

2 Method

2.1 Data Acquisition, Preprocessing and Augmentation

The training and testing datasets contain 26 MVM-CMR datasets with the data size of $50 \times 512 \times 512 \times 4$ which were acquired from 18 healthy subjects (8 of them were acquired twice, giving 26 datasets) at Royal Brompton Hospital. Each of the datasets consists of 3-5 cine slices, giving 121 cine slices in total. There are 50 temporal frames per cardiac cycle and 4 channels reconstructed by a non-Cartesian SENSE reconstruction channel (one magnitude encoding channel and three velocity encoding channels of orthogonal directions), constituting the multi-frame multichannel MVM-CMR data. The MVM-CMR slices have spatial resolutions of 0.85mm $\times$ 0.85mm that were reconstructed from the acquired 1.7mm $\times$ 1.7mm. The MVM-CMR were acquired in short-axis slices from base to apex of the LV. An experienced cardiac MRI physicist performed manual delineation of the LV myocardium to create the ground truth for this study. In addition, we augmented the data by random rotation (angle =[90°, 180°, 270°]) before model training. An example of our multichannel MVM-CMR dataset with the manual segmentation can be found in Figure 1.

2.2 Network Architectures

Attention Enhanced Skip Connections We propose embedded attention modules to realise the enhanced skip connections and synthesise features from the transferred image F of the convolutional layers in the original 3D-UNet structure. The objective is that the relevant features in a single slice can be enhanced during the attention process, while the less relevant ones can be less focused on. To achieve this, an attention map is computed in every frame of the input MVM-CMR images to represent the confidence of the transferred features for each position as expressed in Equation (1) as follows

$$\text{Layer}_i^{\text{Attention}} = \text{Softmax}(F_i \times \text{Conv}(F_i))) \cdot \text{Conv}(F_i)', \tag{1}$$

[1] IoU stands for Intersection-Over-Union, which is also known as the Jaccard index.

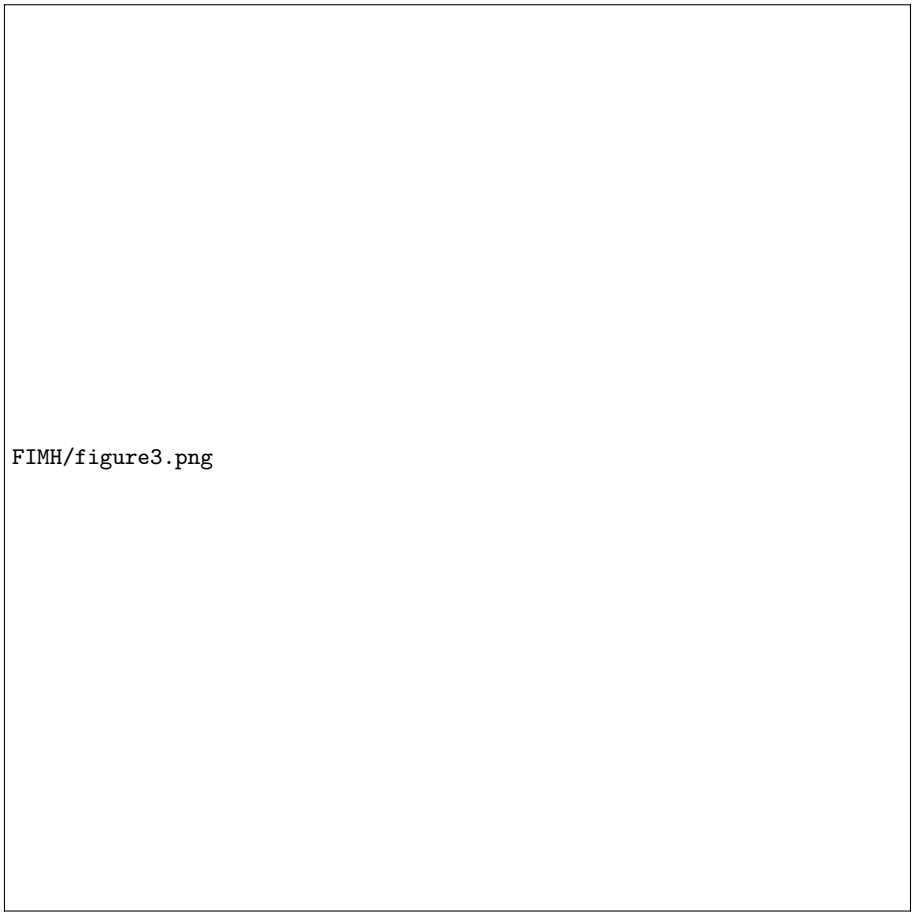

Fig. 1. A sample frame of our multi-channel MVM-CMR dataset with the manual segmentation. From left to right: magnitude channel, three phase channels and the manual LV segmentation.

where i donates the i-th frame in the 3D feature map. Conv and Softmax represent the convolutional layer and the Softmax activation operation, respectively (Figure 2 (a)).

LSTM Based Temporal Feature Extractor We also develop LSTM layers to capture the cross-frame features at the bottom of the U-shaped structure in the 3D-UNet network. We assume that the LSTM can learn the temporal correlations from the multi-frame MVM-CMR data. For our 3D (2D+t) MVM-CMR data, we need to convert them into sequences and then transfer back into 3D (i.e., 2D+t) images before and after the LSTM layer. The whole operation can be denoted as Equation (2) that is

$$\text{Layer}_{\text{out}} = \text{Reshape}(\text{LSTM}(\text{Reshape}(\text{Layer}_{\text{in}}))). \tag{2}$$

Fig. 2. The structure of the embedded in-slice attention block (a) and cross-frame LSTM module (b).

Figure 2 (b) shows the LSTM workflow and Figure 3 represents the overall structure of the proposed model (i.e., 3D-EAR segmentor). In this figure, we use rounded rectangles to denote the flow of feature maps, triangles of different colours to represent different neural network blocks and arrows to indicate the skip connections and up-sampling.

2.3 Loss Functions

For our 3D-EAR segmentor, we implement various loss functions, e.g., (1) Cross-Entropy loss, (2) Dice loss (with Laplace smoothing) and (3) Dice-IoU loss (with Laplace smoothing) to seek an optimal solution. The standard Cross-Entropy loss can be represented by Equation (3), that is

$$\text{Loss}_{\text{Cross-Entropy}} = -\frac{1}{n}\sum_{i}\sum_{c=1}^{n} y_{ic}\log\left(p_{ic}\right), \tag{3}$$

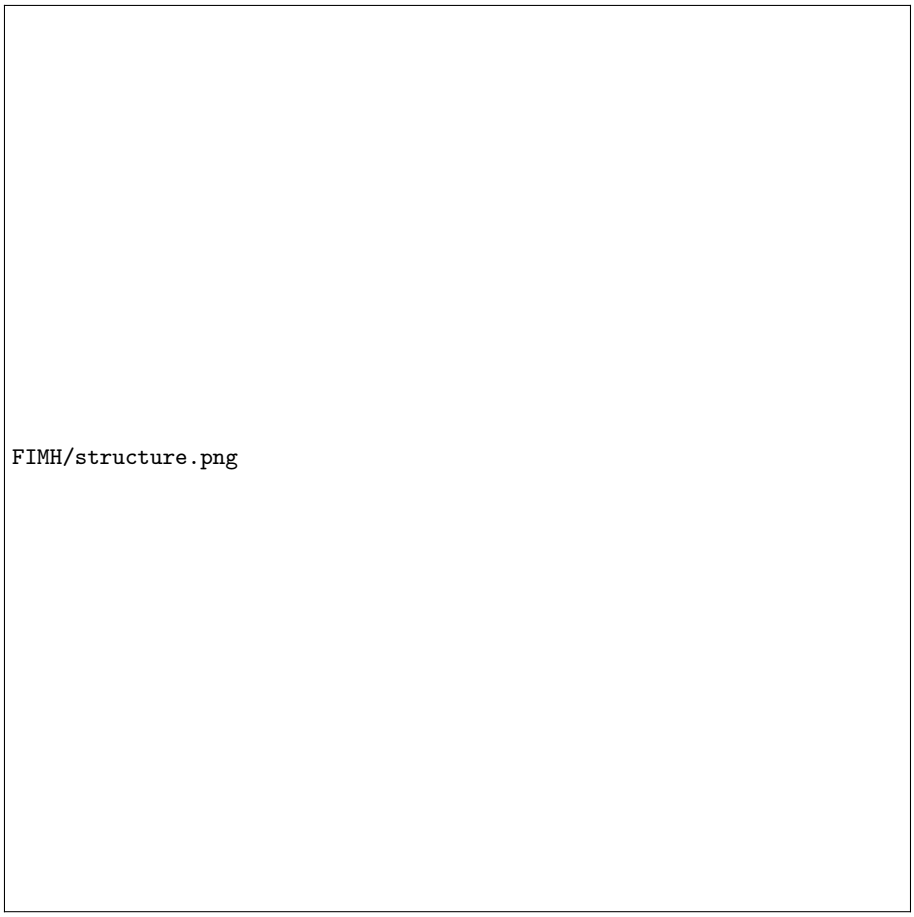

Fig. 3. The overall architecture of our proposed 3D-EAR segmentor.

where y and p represent the true and predicted labels in the n^{th} class, respectively.

The Dice loss (with Laplace smoothing factor f_{smooth}) can be denoted as Equation (4), that is

$$\text{Loss}_{\text{Dice}} = 1 - \frac{2 \times |\text{GT} \cap \text{Pred}| + f_{\text{smooth}}}{|\text{GT}| + |\text{Pred}| + f_{\text{smooth}}}, \tag{4}$$

where GT stands for the ground truth of the segmentation, Pred donates the prediction of the model and $|\bullet|$ represents the area of $\bullet$.

The Dice-IoU loss, which is a combination of the Dice loss with the IoU calculation, can be represented as Equation (5)

$$\text{Loss}_{\text{Dice-IoU}} = \frac{1}{n} \times \sum_{i=1}^{n} \text{Loss}_{\text{Dice}} \times \text{IoU}. \tag{5}$$

Equation (6) illustrates the IoU calculation, which is also smoothed by the Laplace factor f_{smooth}, that is

$$\text{IoU} = 1 - \frac{|\text{GT} \cap \text{Pred}| + f_{\text{smooth}}}{|\text{GT} \cup \text{Pred}| + f_{\text{smooth}}}. \tag{6}$$

2.4 Implementation Details

The input of our model (and compared models) was a 4-channel 50-frame 512 $\times$ 512 MVM-CMR dataset, and the output prediction was with the same size as the input but had 2 different labels (i.e., LV and non-LV). We divided the MVM-CMR cine slices into two sets for experiments, one consisting of 80% of the subjects for model training and the other one consisting of the remaining 20% as independent testing. During the training process, we also performed the 5-fold cross-validation. The training was carried out on two standard NVIDIA GEFORCE RTX 2080 Ti GPUs. Our implementation was based on Keras and TensorFlow backend. The implementation and pre-trained models will be open source (on Github) for a reproducible study.

3 Results

3.1 Experiments and Evaluation Metrics

We performed the following comparison and ablation studies, including UNet3D (the baseline model 3D-UNet), UNet3D-Attention (3D-UNet with attention) and our proposed 3D-EAR segmentator with various loss functions. We evaluated model performance using (1) Dice scores, (2) Sensitivities and (3) Positive Predictive Values (PPV).

3.2 Quantitative Results

Quantitative results of our comparison study can be found in Table 1.

Table 1. Dice scores of our comparison studies and ablation studies using various loss functions.

Structures \ Losses	Cross-Entropy	Dice	Dice-IoU
UNet3D	0.84±0.02	0.85±0.03	0.88±0.02
UNet3D-Attention	0.87±0.03	0.87±0.02	0.89±0.02
3D-EAR	0.88±0.03	0.89±0.02	0.91±0.03

Table 1 and Table 2 show outstanding segmentation performance of using our proposed 3D-EAR model. Compared to the baseline model, our proposed

Table 2. Sensitivities and PPV of our comparison studies and ablation studies using various loss functions.

Losses / Structures	Cross-Entropy		Dice		Dice-IoU	
	Sensitivity	PPV	Sensitivity	PPV	Sensitivity	PPV
UNet3D	0.75±0.03	0.95±0.02	0.81±0.04	0.90±0.05	0.86±0.03	0.91±0.01
UNet3D-Attention	0.84±0.02	0.90±0.02	0.84±0.01	0.91±0.02	0.85±0.02	0.93±0.02
3D-EAR	0.80±0.01	0.98±0.01	0.86±0.01	0.93±0.02	0.87±0.01	0.96±0.02

3D-EAR has achieved significantly higher Dice scores, sensitivities and PPV. We can also find that with the LSTM, our 3D-EAR has further improvement on the model with only the attention module.

An automated segmentation example result obtained by our 3D-EAR model is shown in Figure 4. Followed by a morphological post-processing stage, we were able to generate their LV myocardium global velocity curves from the predicted results and compare it with the ones derived from the ground truth. We are able to observe close alignments of curves and little differences in the peak velocities generated from these curves.

4 Discussion and Conclusion

In this study, we have developed and validated a novel 3D-EAR segmentor for the delineation of LV from MVM-CMR data. The proposed model incorporated embedded attention enhanced skip connections to filter our irrelevant features from images and LSTM based RNN for accounting correlations among temporal frames of the MVM-CMR data. The experimental results have shown promising quantification and visualisation that can facilitate accurate and reliable estimation of global and local myocardial velocities. More detailed method descriptions, comparison results with and without multichannel input data, ablation studies of network parameters, and velocity comparisons in all the slices will be presented.

5 Acknowledgement

This study was supported in part by BHF (TG/18/5/34111, PG/16/78/32402), in part by Heart Research UK RG2584, in part by the ERC IMI [101005122], and in part by ERC H2020 [952172].

References

1. Robin Simpson, Jennifer Keegan, Peter Gatehouse, Michael Hansen, and David Firmin. Spiral tissue phase velocity mapping in a breath-hold with non-cartesian SENSE. *Magnetic Resonance in Medicine*, 72(3):659–668, 2014.
2. Robin M Simpson, Jennifer Keegan, and David N Firmin. MR assessment of regional myocardial mechanics. *J. Magn. Reson. Imaging*, 37(3):576–599, 2013.

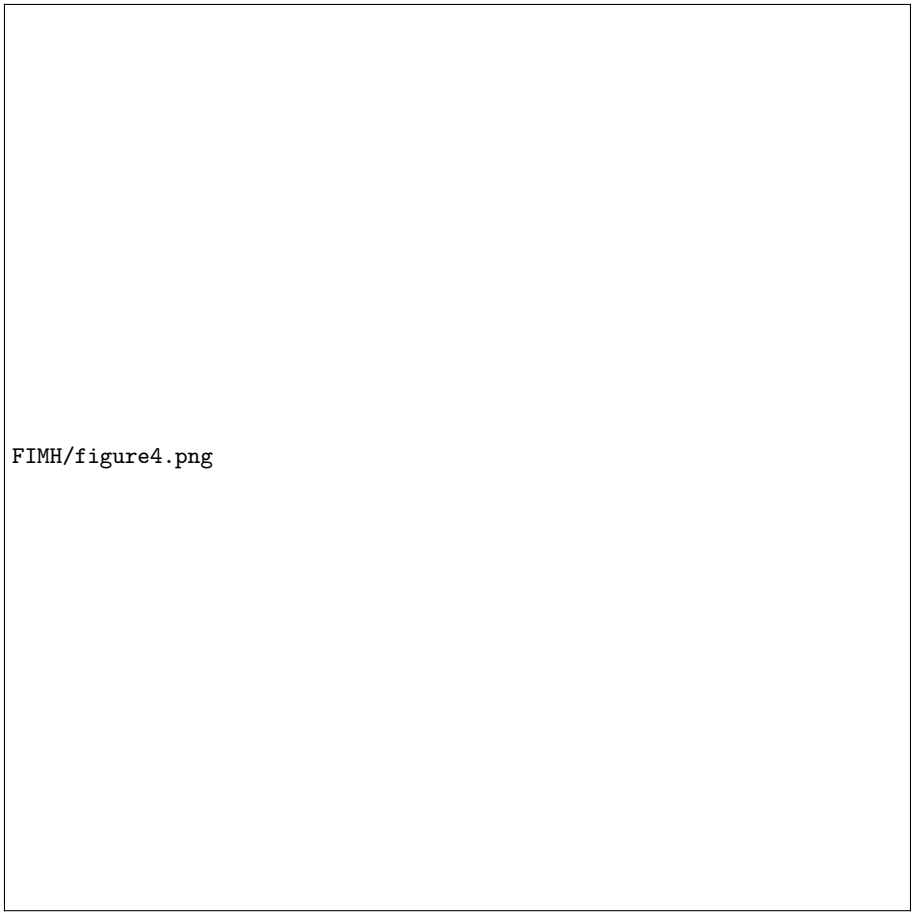

Fig. 4. A typical example of the segmentation results randomly selected from our MVM-CMR datasets with the global longitudinal, radial and circumferential velocity curves and peak velocities per slice of the example frames (Two detailed examples in the Supplementary Material). For the segmentation: Blue—true positive of our automated segmentation; Yellow—false positive; Red— false negative; Blue and Yellow regions—automated segmentation results. For the global velocity curves: Blue/Orange curves—derived from automated and manual segmentations.

3. Robin Simpson, Jennifer Keegan, and David Firmin. Efficient and reproducible high resolution spiral myocardial phase velocity mapping of the entire cardiac cycle. *Journal of Cardiovascular Magnetic Resonance*, 15(1):1–14, 2013.
4. Xiahai Zhuang, Lei Li, Christian Payer, Darko Štern, Martin Urschler, Mattias P Heinrich, Julien Oster, Chunliang Wang, Örjan Smedby, Cheng Bian, et al. Evaluation of algorithms for multi-modality whole heart segmentation: an open-access grand challenge. *Medical Image Analysis*, 58:101537, 2019.
5. Wenjia Bai, Matthew Sinclair, Giacomo Tarroni, Ozan Oktay, Martin Rajchl, Ghislain Vaillant, Aaron M Lee, Nay Aung, Elena Lukaschuk, Mihir M Sanghvi, et al.

Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *Journal of Cardiovascular Magnetic Resonance*, 20(1):65, 2018.
6. Guang Yang, Jun Chen, Zhifan Gao, Shuo Li, Hao Ni, Elsa Angelini, Tom Wong, Raad Mohiaddin, Eva Nyktari, Ricardo Wage, et al. Simultaneous left atrium anatomy and scar segmentations via deep learning in multiview information with attention. *Future Generation Computer Systems*, 107:215–228, 2020.
7. Lei Li, Fuping Wu, Guang Yang, Lingchao Xu, Tom Wong, Raad Mohiaddin, David Firmin, Jennifer Keegan, and Xiahai Zhuang. Atrial scar quantification via multi-scale CNN in the graph-cuts framework. *Medical Image Analysis*, 60:101595, 2020.
8. 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.
9. 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, 2015.
10. Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 565–571. IEEE, 2016.
11. Md Atiqur Rahman and Yang Wang. Optimizing intersection-over-union in deep neural networks for image segmentation. In *International Symposium on Visual Computing*, pages 234–244. Springer, 2016.