

Automatically Segmenting the Left Atrium from Cardiac Images Using Successive 3D U-Nets and a Contour Loss

Shuman Jia, Antoine Despinasse, Zihao Wang, Hervé Delingette, Xavier Pennec, Pierre Jaïs, Hubert Cochet, Maxime Sermesant

► **To cite this version:**

Shuman Jia, Antoine Despinasse, Zihao Wang, Hervé Delingette, Xavier Pennec, et al.. Automatically Segmenting the Left Atrium from Cardiac Images Using Successive 3D U-Nets and a Contour Loss. STACOM: Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges, Sep 2018, Granada, Spain. pp.221-229, 10.1007/978-3-030-12029-0_24. hal-01860285v4

HAL Id: hal-01860285

<https://hal.inria.fr/hal-01860285v4>

Submitted on 1 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Automatically Segmenting the Left Atrium from Cardiac Images Using Successive 3D U-Nets and a Contour Loss

Shuman Jia¹, Antoine Despinasse¹, Zihao Wang¹, Hervé Delingette¹, Xavier Pennec¹, Pierre Jais², Hubert Cochet², and Maxime Sermesant¹

¹ Université Côte d’Azur, Epione Research Group, Inria, Sophia Antipolis, France

² IHU Liryc, University of Bordeaux, Pessac, France

Abstract. Radiological imaging offers effective measurement of anatomy, which is useful in disease diagnosis and assessment. Previous study [1] has shown that the left atrial wall remodeling can provide information to predict treatment outcome in atrial fibrillation. Nevertheless, the segmentation of the left atrial structures from medical images is still very time-consuming. Current advances in neural network may help creating automatic segmentation models that reduce the workload for clinicians. In this preliminary study, we propose automated, two-stage, three-dimensional U-Nets with convolutional neural network, for the challenging task of left atrial segmentation. Unlike previous two-dimensional image segmentation methods, we use 3D U-Nets to obtain the heart cavity directly in 3D. The dual 3D U-Net structure consists of, a first U-Net to coarsely segment and locate the left atrium, and a second U-Net to accurately segment the left atrium under higher resolution. In addition, we introduce a Contour loss based on additional distance information to adjust the final segmentation. We randomly split the data into training datasets (80 subjects) and validation datasets (20 subjects) to train multiple models, with different augmentation setting. Experiments show that the average Dice coefficients for validation datasets are around 0.91 - 0.92, the sensitivity around 0.90-0.94 and the specificity 0.99. Compared with traditional Dice loss, models trained with Contour loss in general offer smaller Hausdorff distance with similar Dice coefficient, and have less connected components in predictions. Finally, we integrate several trained models in an ensemble prediction to segment testing datasets.

Keywords: 3D U-Net, segmentation, left atrium, loss function, Contour loss, distance map, ensemble prediction

1 Introduction

Atrial fibrillation (AF) is the most frequently encountered arrhythmia in clinical practice, especially in aged population [2,3]. It is characterized by uncoordinated electrical activation and disorganized contraction of the atria. This condition is associated with life-threatening consequences, such as heart failure, stroke and

vascular cerebral accident. AF also leads to increased public resource utilization and expense on health care.

With evolving imaging technologies, the analysis of cardiovascular diseases and computer-aided interventions has been developing rapidly. Imaging of the heart is routinely performed in some hospital centers when managing AF and prior to atrial ablation therapy, an invasive treatment to establish trans-mural lesions and block the propagation of arrhythmia. Automated segmentation from cardiac images will benefit the studies of left atrial anatomy, tissues and structures, and provide tools for AF patient management and ablation guidance.

In recent years, with the continuous development of deep learning, neural network models have shown significant advantages in different visual and image processing problems [4]. Automatic segmentation of 3D volumes from medical images by deep neural network also attracts increasing attention in the research community of medical image analysis [5,6].

In this study, we utilize 3D U-Nets with convolutional neural network (CNN), which shows clear advantages compared with traditional feature extraction algorithms [7,8]. Based on that, Ronneberger et al. proposed the original U-Net structure [9] with CNN. Traditional 2D U-Net has achieved good results in the field of medical image segmentation [9,10]. However, it performs convolution on the 2D slices of the images and cannot capture the spatial relationship between slices. Its 3D extension [11], expands the filter operator into 3D space. This extracts image features in 3D, and hence takes into account the spatial continuity between slices in medical imaging. This may better reflect shape features of the corresponding anatomy, enabling full use of the spatial information in 3D images.

Previously, Tran et al. used 3D CNNs to extract temporal and spatial features [12]. They experimented with different sets of data. Hou et al. used 3D CNN to detect and segment pedestrians in a video sequence [13]. The previous studies show that 3D CNN outperformed 2D CNN when dealing with sequences issues.

3D U-Net was used in [11] to realize semi-automatic segmentation of volumetric images. Oktay et al. segmented the ventricle from magnetic resonance (MR) images with 3D U-Net. They introduced an anatomical regularization factor into the model [14], while we choose to use loss function at pixel level.

In the following sections, we will present the two-stage network to segment the left atrium from MR images. The network consists of two successive 3D U-Nets. The first U-Net is used to locate the segmentation target. The second U-Net performs detailed segmentation from cropped region of interest. We introduce a new loss function, Contour loss for the second U-Net. Results will be shown in Section 3.

2 Method

2.1 Dual 3D U-Nets - cropping and segmenting

U-Net is a typical encoder-decoder neural network structure. The images are encoded by the CNN layers in the encoder. The output characteristics and the

feature maps at different feature levels of the encoder serve as input of the decoder. The decoder is an inverse layer-by-layer decoding process. Such a codec structure can effectively extract image features of different levels so as to analyze the images in each dimension. The 3D U-Net used in this paper is a specialization of 3D U-Net proposed by Çiçek et al [11]. The implementation of U-Net follows the work of Isensee et al [15]. We propose a successive dual 3D U-Net architecture, illustrated in Fig. 1.

The first 3D U-Net locates and coarsely extracts the region of interest. Its input is MR images normalized and re-sized to $[128, 128, 128]$. Its output is preliminary predicted masks of the left atrium. We keep the largest connected component in the masks, and compute the spatial location of the left atrium. Then, we crop the MR images and ground truth masks with a cuboid centered at the left atrium.

The second network performs a secondary processing of the cropped images using the full resolution. Because the higher is the resolution, the larger is the needed memory, we keep only the region around the left atrium, so as to preserve information that is essential for left atrial segmentation. But also, this allows to put a higher resolution on the region of interest with the same amount of memory resource. The input for the second U-Net is MR images cropped around the predicted left atrium without re-sampling of size $[224, 144, 96]$. Its output is our prediction for the left atrial segmentation. We train the second U-Net with two kinds of ground truths, binary segmentation masks M and euclidean distance maps $D(M)$, as shown in Fig 2. Here, we introduce a new loss function based on Contour distance.

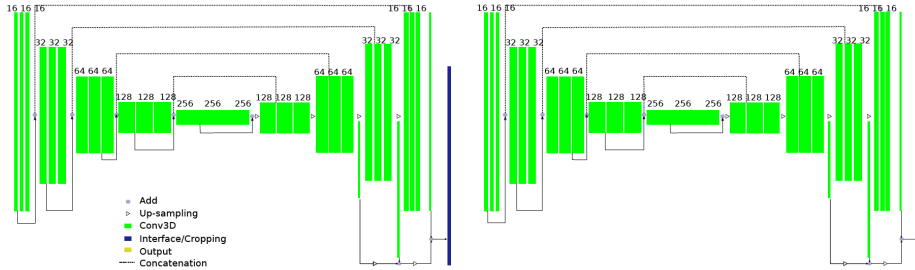


Fig. 1. Proposed Dual 3D U-Net Structure. Green blocks represent 3D features; Dark blue refers to interface operation to crop the region of interest based on first U-Net prediction.

2.2 Loss Functions

Using Dice coefficient as loss function can reach high accuracy. However, experiments show that, because the inside of the left atrial body accounts for most pixels, the network would stop to optimize when it finds satisfying segmentation

of the left atrial body. Instead of the volume inside, the contour is what we want to obtain accurately for the segmentation. It is challenging to segment accurately especially the region around the pulmonary veins and the appendage. Hence, we introduce the Contour distance into the loss function.

The distance maps $D(M)$ of the ground truth segmentation M illustrate how far is each pixel from the contour of the left atrium. We compute the distance for pixels inside and outside the left atrium, based on the euclidean distance transform algorithm implemented in `scipy`. The definition of a Hausdorff distance is symmetric between two point sets. But to make it easy to be implemented in neural networks, we do not compute the distance map of the changing prediction P in the training process, and use an unitary distance:

$$loss_{contour} = \sum_{p \in contour(P)} \min_{m \in M} \|p - m\|_2 = \sum (D(M) \circ contour(P)), \quad (1)$$

where \circ performs element-wise multiplication. P is the prediction of the U-Net, after sigmoid activation of the output layer. To compute a contour of predicted left atrium, we can apply 3D Sobel filters on P and add the absolute value of results in three directions:

$$contour(P) = |P * s_x| + |P * s_y| + |P * s_z|, \quad (2)$$

where $*$ denotes 3D convolution operation and $+$ denotes pixel-wise addition. s_x , s_y and s_z are 3D SobelFeldman kernels in x-, y- and z-direction, for example:

$$s_z = \left[\begin{bmatrix} +1 & +2 & +1 \\ +2 & +4 & +2 \\ +1 & +2 & +1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} -1 & -2 & -1 \\ -2 & -4 & -2 \\ -1 & -2 & -1 \end{bmatrix} \right].$$

In the optimization process, the Contour loss decreases when the contour of prediction is nearer that of the ground truth. However, if $D(M)$ and $contour(P)$ are both always positive, a bad global minimal exists: let prediction P remain constant so that $contour(P) \sim 0$. To avoid this, we add a *drain* on distance maps. For example, we set $drain = -1.8$, and therefore $D(M) + drain$ has negative value on the contour of M , since all pixels on the contour of M have a euclidean distance $\leq \sqrt{3}$. This creates a drain channel for the model to continue the optimization process towards negative loss value.

$$loss_{contour} = \sum ((D(M) + drain) \circ contour(P)). \quad (3)$$

The loss function is differentiable and converging in 3D U-Net training trials.³

2.3 Ensemble Prediction

Contour loss provides a spatial distance information for the segmentation, while Dice coefficient measures the volumes inside the contour. We combine the two loss functions in a ensemble prediction model.

We visualize the process in Fig. 2.

³ The code is available on https://gitlab.inria.fr/sjia/unet3d_contour.

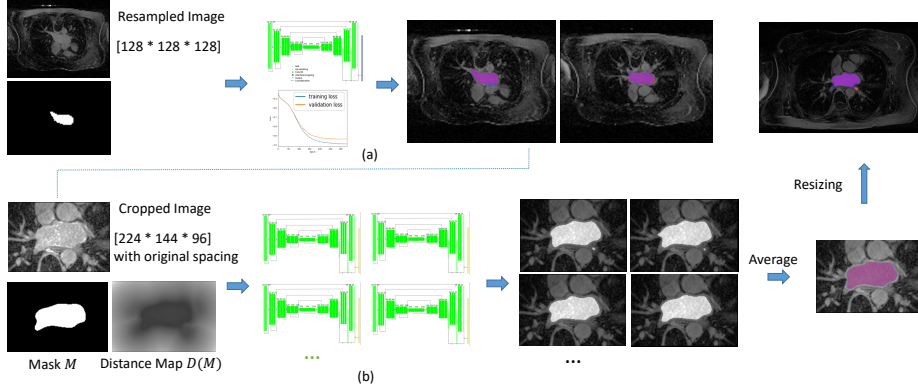


Fig. 2. The framework of successive U-Nets training. (a) The first U-Net - cropping; (b) the second U-Net - segmenting, with ensemble prediction models. We show here axial slices of MR images, overlapped with manual segmentation of the left atrium in blue, our segmentation in red, intersection of the two in purple.

Fixed Experimental Setting In this study, we set batch size to 1. Training-validation split ratio equals to 0.8. We perform normalization of image intensity. The initial learning rate of our neural network is $5e^{-4}$. Learning rate will be reduced to half after 10 epochs if the validation loss is not improving. The early convergence is defined as no improvement after 50 epochs. Number of base filters is 16. Maximum number of epochs is 500, and 200 steps per epoch. Using large initial learning rate reduces the time to find the minimum and may also overstep the local minima to converge closer to the global minimum of the loss function. However, the accuracy of segmentation also relies on when we stop the training to avoid over-fitting.

We use computation cluster with GPU capacity 6.1 for training. The first U-Net takes around 200 - 300 epochs to reach convergence. We extract the largest connected components in predictions of the first U-Net to crop the original MRI images around the left atrium, without re-sampling, of size $[224, 144, 96]$. The second U-Net takes around 150 - 200 epochs to reach convergence. Then the prediction results of the second U-Net are re-sized without re-sampling to the original image size.

Varying Experimental Setting For the second U-Net, we change on the one hand, the options for augmentation: horizontal/vertical flip set to True or False; rotation range set to 0, 7, 10 degree; width/height shift range set to 0.0 - 0.1; zoom in or not, zoom range set to 0.1, 0.2. On the other hand, we alter the loss function option, Dice coefficient and Contour loss. We choose multiple trained U-Net models with above experimental settings for ensemble prediction. We also train twice with some parameters but with different validation splitting. We make the final decision of segmentation based on the average of all predictions,

similar to letting multiple agents vote for each pixel if it belongs to left atrium or not, in majority voting system.

3 Evaluation on Clinical Data

3.1 Materials

A total of 100 3D GE-MRIs from patients with AF are provided by the STACOM 2018 Atrial Segmentation Challenge. The original resolution of the data is $0.625 \times 0.625 \times 0.625 \text{ mm}^3$. 3D MR images were acquired using a clinical whole-body MRI scanner and the corresponding ground truths of the left atrial masks were manually segmented by experts in the field.

3.2 Comparison of Loss Functions

We assess the segmentation results of individual models, trained with different experimental setting, as described in Sec. 2.3, to compare the prediction performance using Dice loss and Contour loss.

Evaluation Metrics The evaluation metrics are Dice coefficient, Hausdorff distance (HD) and confusion matrix, as shown in Table 1. Multiple evaluation metrics provide us different views to assess the performance of models.

The Dice index for predicted segmentation in validation datasets attained 0.91 - 0.92. The sensitivity of prediction was around 0.90 - 0.94 and the specificity 0.99. The proposed method closely segmented the atrial body, with both loss functions, compared with manual segmentation. Different from traditional Dice loss, models trained with Contour loss in general offered smaller Hausdorff distance with similar Dice coefficient.

Visualization We visualize the predicted segmentation results of validation datasets in Fig. 3 in a 3D view. Case 1 and Case 2 are selected to represent respectively, good scenario and bad scenario. For the two loss functions, differences lay in the boundary, the region close to the pulmonary veins and septum. With Dice loss function, there were more details and sharp edges, and therefore more disconnected spots not belonging to the left atrium. With the Contour loss function, the smoothness of the contour, and shape consistency were better maintained.

Connected Components The left atrium should be a single connected component in binary segmentation mask. We present in Table 2 the number of connected components in raw predictions given by U-Nets and compare the two loss functions. Using Dice coefficient loss alone produced more disconnect components not belonging to the left atrial structures.

Table 1. Evaluation of validation datasets segmentations using two loss functions: Dice coefficient loss (top) and Contour loss (bottom).

	Confusion Matrix		Dice	Contour distance (pixel)		
	Sensitivity	Specificity		Average HD	HD	
Model (Dice loss)	1	0.93 ± 0.03	0.99 ± 0.00	0.910 ± 0.032	1.71 ± 0.53	34.29 ± 15.00
	2	0.94 ± 0.04	0.99 ± 0.00	0.913 ± 0.029	1.63 ± 0.57	27.39 ± 12.19
	3	0.91 ± 0.05	0.99 ± 0.00	0.914 ± 0.023	1.62 ± 0.42	30.04 ± 11.76
	4	0.92 ± 0.04	0.99 ± 0.00	0.920 ± 0.030	1.41 ± 0.51	23.50 ± 12.34
	5	0.91 ± 0.05	0.99 ± 0.00	0.918 ± 0.019	1.47 ± 0.47	21.66 ± 7.75
	6	0.91 ± 0.04	0.99 ± 0.00	0.921 ± 0.024	1.45 ± 0.48	26.52 ± 16.14
	7	0.92 ± 0.03	0.99 ± 0.00	0.923 ± 0.027	1.40 ± 0.67	23.38 ± 12.03
	8	0.93 ± 0.04	0.99 ± 0.00	0.927 ± 0.024	1.28 ± 0.46	22.48 ± 16.35
Model (Contour loss)	1	0.90 ± 0.05	0.99 ± 0.00	0.911 ± 0.020	1.60 ± 0.41	19.80 ± 6.82
	2	0.92 ± 0.04	0.99 ± 0.00	0.906 ± 0.045	1.60 ± 0.58	18.88 ± 7.13
	3	0.91 ± 0.04	0.99 ± 0.00	0.917 ± 0.025	1.48 ± 0.51	21.57 ± 8.13
	4	0.92 ± 0.04	1.00 ± 0.00	0.917 ± 0.024	1.40 ± 0.33	20.21 ± 7.53
	5	0.90 ± 0.04	1.00 ± 0.00	0.919 ± 0.024	1.53 ± 0.51	24.81 ± 6.57
	6	0.89 ± 0.03	1.00 ± 0.00	0.921 ± 0.019	1.58 ± 0.59	24.04 ± 11.72
	7	0.91 ± 0.04	0.99 ± 0.00	0.914 ± 0.021	1.56 ± 0.47	22.65 ± 7.52
	8	0.92 ± 0.04	1.00 ± 0.00	0.917 ± 0.030	1.41 ± 0.45	19.66 ± 4.52

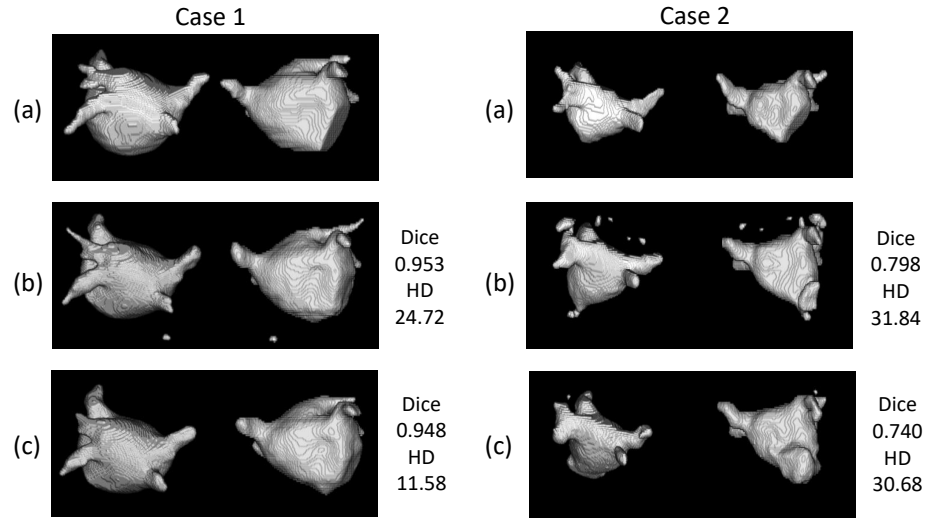


Fig. 3. Visualization of good and bad validation datasets segmentations, with their Dice, Hausdorff distance (HD) with respect to the ground truth. (a) manually segmented; (b) predicted with Dice coefficient loss; (c) predicted with Contour loss.

Table 2. Number of connected components in predicted segmentations.

	Mean \pm σ	Maximum
Model (Dice loss)	2.85 ± 2.83	20
Model (Contour loss)	1.28 ± 0.65	5

Ensemble Prediction To reduce the irregular bumps and disconnected components, we choose in total 11 trained U-Net models with both loss functions, to perform an ensemble prediction for testing datasets. We add the probabilistic segmentations of all U-Nets and threshold their sum ≥ 5.5 .

With Dice loss, more details can be captured. With Contour loss, the shapes look more regular and smoother globally. The difficult regions to segment remain the septum and especially the appendage for both loss functions. The manual segmentations are usually performed slice-by-slice, there exists sudden discontinuity between two axial slices. While our segmentation is based on 3D operators, the segmented region is continuous between slices, which accounts for part of mismatch between manual segmentation and network segmentation that cannot be avoided.

4 Conclusion

In this paper, we proposed a deep neural network with dual 3D U-Net structure, to segment the left atrium from MR images. To take into consideration the shape characteristics of the left atrium, we proposed to include distance information and created a Contour loss function. Using multiple trained models in an ensemble prediction can improve the performance, reducing the impact of accidental factors in neural network training.

Based on previous studies on cardiac cavities segmentation, our method accurately located the region of interest and provided good segmentations of the left atrium. Experiments show that the proposed method well captured the anatomy of the atrial volume in 3D space from MR images. The new loss function achieved a fine-tuning of contour distance and good shape consistency.

The automated segmentation model can in return reduce manual work load for clinicians and has promising applications in clinical practice. Potential future work includes integrating the segmentation model into a clinic-oriented AF management pipeline.

Acknowledgments. Part of the research was funded by the Agence Nationale de la Recherche (ANR)/ERA CoSysMed SysAFib project. This work was supported by the grant AAP Sant 06 2017-260 DGA-DSH, and by the Inria Sophia Antipolis - Mditerrane, NEF computation cluster. The author would like thank the work of relevant engineers and scholars.

References

1. Christopher McGann, Nazem Akoum, Amit Patel, Eugene Kholmovski, Patricia Revelo, Kavitha Damal, Brent Wilson, Josh Cates, Alexis Harrison, Ravi Ranjan, et al. Atrial fibrillation ablation outcome is predicted by left atrial remodeling on mri. *Circulation: Arrhythmia and Electrophysiology*, 7(1):23–30, 2014.
2. Massimo Zoni-Berisso, Fabrizio Lercari, Tiziana Carazza, Stefano Domenicucci, et al. Epidemiology of atrial fibrillation: European perspective. *Clin Epidemiol*, 6(213):e220, 2014.
3. Carlos A Morillo, Amitava Banerjee, Pablo Perel, David Wood, and Xavier Jouven. Atrial fibrillation: the current epidemic. *Journal of geriatric cardiology: JGC*, 14(3):195, 2017.
4. Carlo N De Cecco, Giuseppe Muscogiuri, Akos Varga-Szemes, and U Joseph Schoepf. Cutting edge clinical applications in cardiovascular magnetic resonance. *World Journal of Radiology*, 9(1):1, 2017.
5. Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A W M van der Laak, Bram van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, dec 2017.
6. Jin Liu, Yi Pan, Min Li, Ziyue Chen, Lu Tang, Chengqian Lu, and Jianxin Wang. Applications of deep learning to MRI images: A survey. *Big Data Mining and Analytics*, 1(1):1–18, 2018.
7. Matthew D. Zeiler and Rob Fergus. Visualizing and Understanding Convolutional Networks arXiv:1311.2901v3 [cs.CV] 28 Nov 2013. *Computer VisionECCV 2014*, 8689:818–833, 2014.
8. Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation Learning: A Review and New Perspectives. (1993):1–30, 2012.
9. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. pages 1–8, 2015.
10. Xiaomeng Li, Hao Chen, Xiaojuan Qi, Qi Dou, Chi-Wing Fu, and Pheng Ann Heng. H-DenseUNet: Hybrid Densely Connected UNet for Liver and Liver Tumor Segmentation from CT Volumes. (1):1–10, 2017.
11. Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. 3D U-net: Learning dense volumetric segmentation from sparse annotation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9901 LNCS:424–432, 2016.
12. D Tran, L D Bourdev, R Fergus, L Torresani, and M Paluri. Learning Spatiotemporal Features with 3D Convolutional Networks. *CoRR*, abs/1412.0, 2015.
13. Rui Hou, Chen Chen, and Mubarak Shah. An End-to-end 3D Convolutional Neural Network for Action Detection and Segmentation in Videos. 14(8):1–15, 2017.
14. Ozan Oktay, Enzo Ferrante, Konstantinos Kamnitsas, Mattias Heinrich, Wenjia Bai, Jose Caballero, Stuart Cook, Antonio De Marvao, Timothy Dawes, Declan O Regan, Bernhard Kainz, Ben Glocker, Daniel Rueckert, and C V Dec. Anatomically Constrained Neural Networks (ACNN): Application to Cardiac Image Enhancement and Segmentation. pages 1–13, 2017.
15. Fabian Isensee, Philipp Kickingereder, Wolfgang Wick, Martin Bendszus, and Klaus H Maier-Hein. Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. In *International MICCAI Brainlesion Workshop*, pages 287–297. Springer, 2017.