Mask R-CNN for Segmentation of Left Ventricle

Muhammad Ali Shoaib[1], Khin Wee Lai[2](✉), Azira Khalil[3], and Joon Huang Chuah[1]

[1] Department of Electrical Engineering, University of Malaya, Kuala Lumpur, Malaysia
shoaib.te@gmail.com, jhchuah@um.edu.my

[2] Department of Biomedical Engineering, University of Malaya, Kuala Lumpur, Malaysia
lai.khinwee@um.edu.my

[3] Faculty of Science and Technology, International Islamic University of Malaysia,
Negeri Sembilan, Malaysia
azira@usim.edu.my

Abstract. Globally, cardiovascular diseases (CVDs) remain the major cause of death among citizens. With echocardiography, doctors are able to diagnose and determine vital parameters for the evaluation of these diseases. Segmentation of left ventricular (LV) from echocardiography is a significant tool for cardiovascular medical analysis. Besides calculating important clinical indices (e.g. ejection fraction), segmentation also can be useful for the investigation of the basic structure of ventricle. Automatic segmentation of the LV has become a valuable means in echocardiography as we can achieve fast and accurate results and a large number of cases can be handled with limited availability of experts. The Convolutional Neural Networks (CNN) have shown outstanding outcomes for image classification, detection, and segmentation in numerous fields. Recently Mask Regions Convolutional Neural Network (Mask R-CNN) has emerged as a very good segmentation model. In this work, Mask R-CNN is proposed for the segmentation of LV. The Mask R-CNN model is first fine-tuned with Common Object in Context (COCO) weights and then the model is trained with our own data. The model first finds out the region of interest (ROI) in the image that contains the desired object i.e. LV. In the ROI, the model segment LV by generating the mask around it. The results demonstrated by the proposed method segments the LV accurately and efficiently with limited training data.

Keywords: Deep learning · Segmentation · Medical images · Left ventricle

1 Introduction

Cardiovascular diseases (CVDs) are one of the leading causes of deaths in developing countries. The World Health Origination (WHO) estimated that annually one-third of deaths in the world occur due to CVDs [1]. Heart diseases are caused by different reasons but mainly associated with diminished LV function. The LV segmentation is important for the assessment of LV function as it describes the ventricular volume, ejection fraction, wall motion irregularities, and myocardial thickness [2].

To analyze the heart and its LV, echocardiography is widely used technique. Being non-invasive, low-cost and non-ionization radiation, echocardiography makes its place

F. Ibrahim et al. (Eds.): ICIBEL 2019, IFMBE Proceedings 81, pp. 14–22, 2021.
https://doi.org/10.1007/978-3-030-65092-6_2

as the most frequently used technique for myocardial analysis. Doctors interpret these images manually which highly depends upon the expertise of clinicians or doctors [3]. For LV assessment mostly, the segmentation is performed manually. Manual segmentation is more time-consuming, labor-intensive and more-often the expert's proficiency affected by their work overload. It is highly beneficial to develop an automatic system for segmentation of LV.

In this study, we are focusing on developing an automatic LV segmentation tool based on CNN. CNN is a useful neural network used for image processing. In CNN few coefficients are used to extract the information from image compared to a simple neural network. In convolutional layer, the same coefficients are used across the different location of the image so CNN requires less memory. CNN's have had massive success in segmentation problems [3, 4]. For segmentation, CNN architectures are used without fully connected layers. This allows generating the segmentation maps for images of any size. Mask R-CNN, a model of CNN, is designed to perform segmentation task on natural images. As compared to existing work, we proposed the usage of Mask R-CNN for the segmentation of LV. Mask R-CNN has been used for natural image segmentation but its application in medicine is very new. Our results show that we can apply the Mask R-CNN for the segmentation of LV and it gave very good results.

The rest of the paper is organized as follows. Section 2 describes the existing literature and related work, Sect. 3 provides the methodologies adopted and development of the model. Section 4 discusses the results and finally, conclusions are drawn in Sect. 5.

2 Literature Review

There are different methods proposed by researchers for the automated segmentation of LV such as deformable models, statistical models, and machine learning models. For example, the authors in [5–7] proposed the deformable model approach for segmentation of LV. Deformable models require the initial position and shape of the model to be very close to the structure of desire object in the image. As good initialization is needed for deformable models, this makes automatic segmentation limited. Presently, the common initialization method of LV segmentation is manual or semiautomatic. Therefore, precise and automatic initialization technique is crucial for fully automatic LV segmentation.

Statistical models are built on the statistical figures from big labeled data. The statistical figures from the labeled data are modeled using parameters mostly based on contour borders and image textures information in the image. In the recent past, the active appearance model (AAM) and active shape model (ASM) have been used for the LV segmentation of echocardiography [8, 9]. In these approach initialization and assumption of shape model restrict automatic segmentation.

Unlike statistical and deformable models, machine learning approaches are not depending on the initialization and, the assumption of shape and appearance. In machine learning, the deep learning and more specifically convolution neural networks have attained great segmentation results in natural images [10]. Due to outstanding achievement in natural image segmentation, some recent works have been done on the application of LV segmentation using CNN. However, the main edge in natural image segmentation is the availability of a large amount of data while for the LV segmentation training dataset is limited. Therefore, limited researchers try to apply deep learning for the LV segmentation task.

Some research combined deep learning method and deformable model to segment LV on cardiac images. In these works, deep learning methods were employed to detect and categories the ROI of LV, and then another postprocessing method was used to make a final segmentation of LV. In [11], Luo et.al used CNN with the deformable model to segment LV from 3D echocardiography. CNN is used initially to find out ROI and then used Gradient Vector Flow (GVF) snack deformable model for the segmentation. As for deformable models, good initialization needed so this is achieved by using stack autoencoder technique. Other researchers have applied deep learning on 3D echocardiography along with the deformable model. Fully Convolutional Network (FCN) is applied for coarse segmentation 2D and deformable model is used for fine segmentation [12].

As labeling of data is a very time-consuming task so researchers also used some machine learning algorithm to label the data and train the network. This pre-trained network is used on manually annotated data. In [13], U-net architecture is used for segmentation of LV. Instead of manual annotation, LV is modeled as cubic Hermite spline methods and transformation of points to fit the spline on LV is done using a Kalman filter. They pre-train the network using labeled data which is annotated using Kalman filter and then use fine-tuning using manually annotated data. This method can reduce the amount of manual labeling, but here again, the overall method is depending upon another method i.e. conventional machine learning.

3 Methodology

In this research, we utilize a deep learning method for segmentation of LV, unlike other previous researches which have used other methods like deformable or simple machine learning with deep learning. Mask R-CNN architecture is a promising approach for image segmentation.

3.1 Dataset and Annotation

Echocardiography data of thirty patients is collected from the institution specialized in cardiovascular diseases. In this research, the apical 4 chamber (A4C) view is used for the analysis of LV. For training the neural network, we obtain echocardiography videos of twenty-five patients. Performance of the trained model was analyzed by using the test data of five different patients not used in training. LV is labeled using the Visual Geometry Group (VGG) annotation tool. VGG Image Annotator (VIA) is an open source annotation tool and used to describe and label a region in an image. We classified into two classes in the images i.e. background and left ventricle. The professionals from the medical field authenticate the labeled images.

3.2 Neural Network

The data set available for training included data of 25 patients since the data was not sufficient for training, thus it was compensated using transfer learning. First, we trained the model with pre-trained COCO weights. After that, we used our own data for training the model. Currently, the Mask R-CNN have been mostly used for the segmentation of natural images and have shown very good results. In larger part, Mask R-CNN has been driven by powerful CNN architectures, such as the Faster R-CNN [14] used for object detection and FCN [10] used for semantic segmentation.

In faster R-CNN, Region Proposal Network (RPN) a fully convolutional network is used to extract region proposals. Thus, RPN proposes regions with objects for further classification. The second stage, which is in core Fast R-CNN extracts features from each proposed region and do the classification and bounding-box regression.

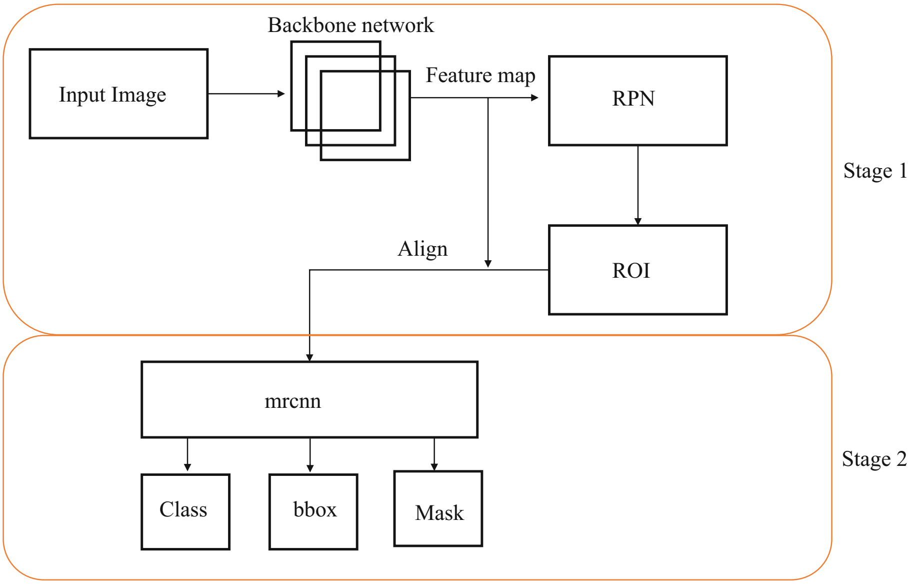


Fig. 1. Model Architecture

In the model of Mask R-CNN, we follow the same basic two-stage process of Faster R-CNN. In the first part backbone, neural network extracts the features from the image and passes to RPN and RPN extracts proposal regions. Here ROI-Align is used instead of ROI pooling to set the bounding boxes which could possibly contain the LV.

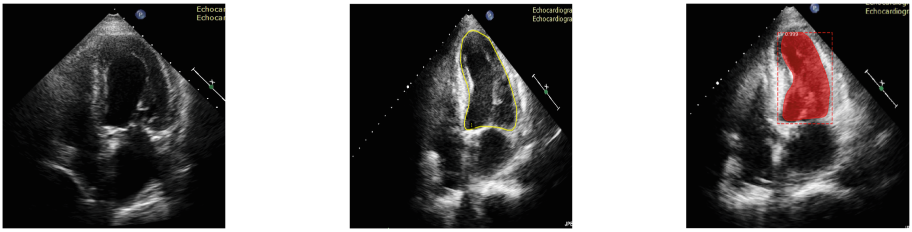
In the second stage, we not only predict the class and bounding box offset but also makes a binary mask for each ROI. Mask R-CNN works on the principle of faster R-CNN that applies bounding-box classification and regression in parallel. Fig. 1 shows the basic architecture of mask R-CNN. So, for each sample of ROI three losses are calculated: classification loss, bounding box offset loss, and mask loss.

$$L = L_{class} + L_{bbox} + L_{mask}$$

The neural network training was implemented using TensorFlow and Keras in NVIDIA DIGITS (GTX1080Ti) on an Intel Core i7. The learning rate of 0.01 and 50 epochs are used for training.

4 Results

In this paper, we present the initial results achieved by our proposed technique. Fig. 2(a) shows the one sample image of echocardiography images. In Fig. 2(b) LV boundary is drawn using the VGG annotator tool. These labeled images are used for training and testing. Fig. 2(c) is the output generated by the model. As Mask R-CNN has two stages, in the first part rectangular box is generated by RPN. The dotted line in the output figure shows the ROI generated by RPN. The second stage mask is generated within the ROI, the red labeled area is the segmented LV.



(a) (b) (c)

Fig. 2. (a) Echocardiography image (b) labeled image (c) segmented output

First, we will show some loss reduction during the training process. The losses during the training after each epoch have been analyzed. Tensorborad utility is used for analyzing the losses during training. Tensorboard is a remarkable utility which allows us to visualize data and how it behaves. The robustness of the neural network can be analyzed by the loss functions. The training process continued to 50 epochs. We use a smooth L1 loss, which is the absolute value between the prediction and ground truth. The reason is that L1 loss is less sensitive to outliers compared to losses like L2.

The model first calculates the Bounding box refinement loss. Figure 3 shows the bbox loss with the number of epochs on the x-axis. Loss decreases from 0.3505 to 0.0121 in 50 epochs.

The second loss is class loss. In this study, we have only two classes i.e. background and LV. The model calculates the class loss and Fig. 4 demonstrates the reduction in class loss with the number of epochs.

The model segments the LV and generates the mask of it. The mask loss is a loss of mask generated by the model and the original boundary of LV. In FCN mask loss is combined loss of mask and class of object, while Mask R-CNN only calculate mask loss here as the class has been already identified. Figure 5 represents the mask loss of the model that is 0.457 after first epoch and 0.057 after 50 epochs.

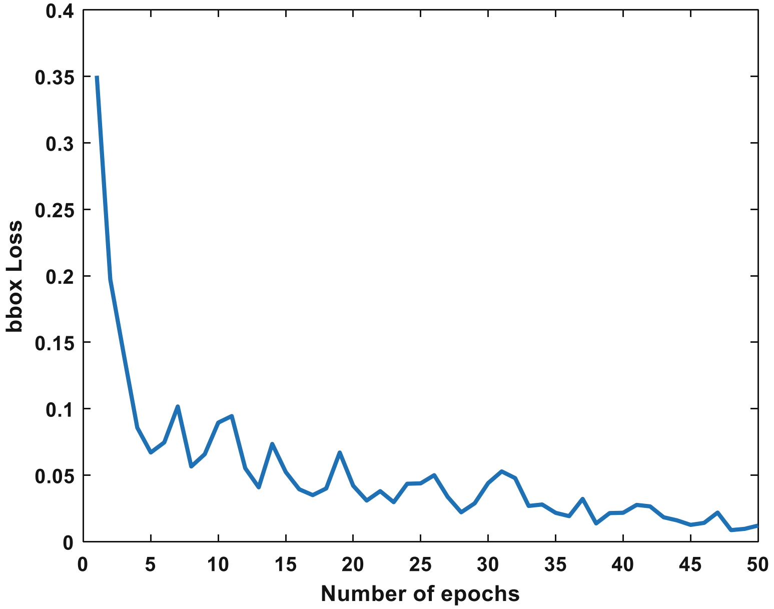


Fig. 3. Bounding box loss of the model

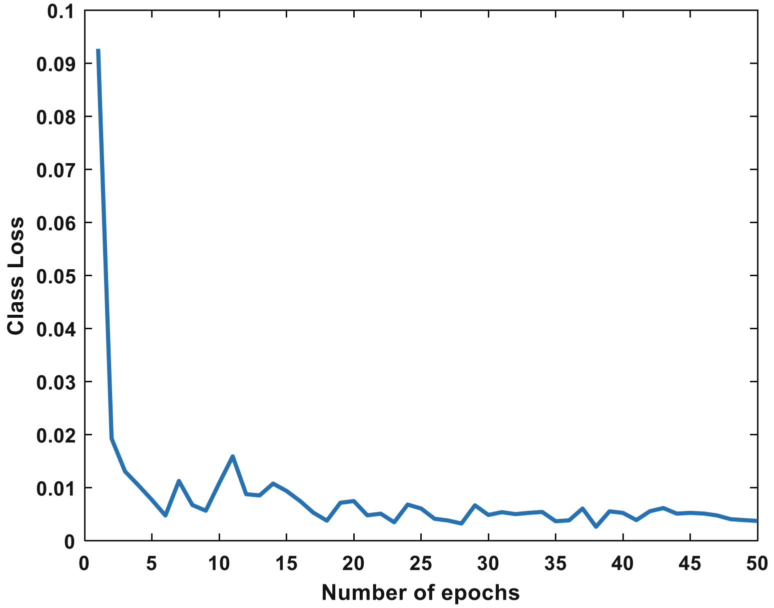


Fig. 4. Class loss of the model

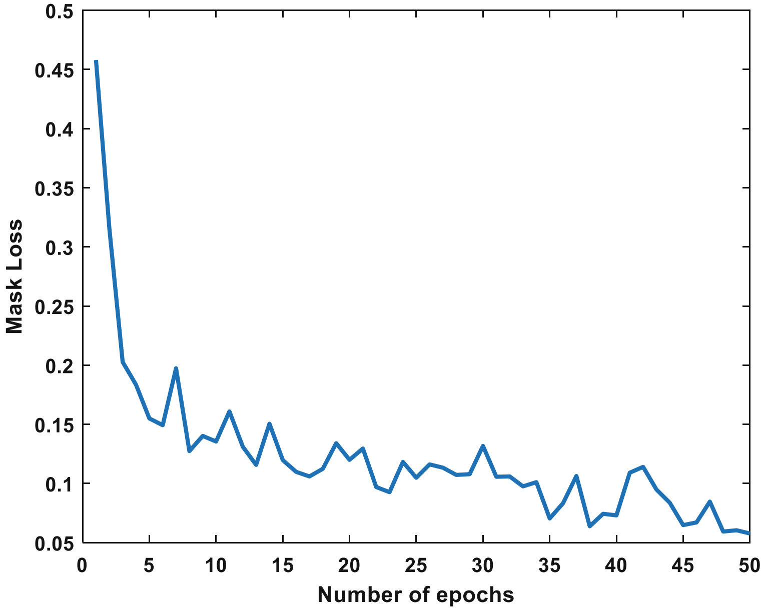


Fig. 5. Mask loss of the model

The overall loss function is calculated by adding all loss values like bbox loss, class loss and segmented mask loss. All these losses are shown in the above figures and overall loss is shown in Fig. 6.

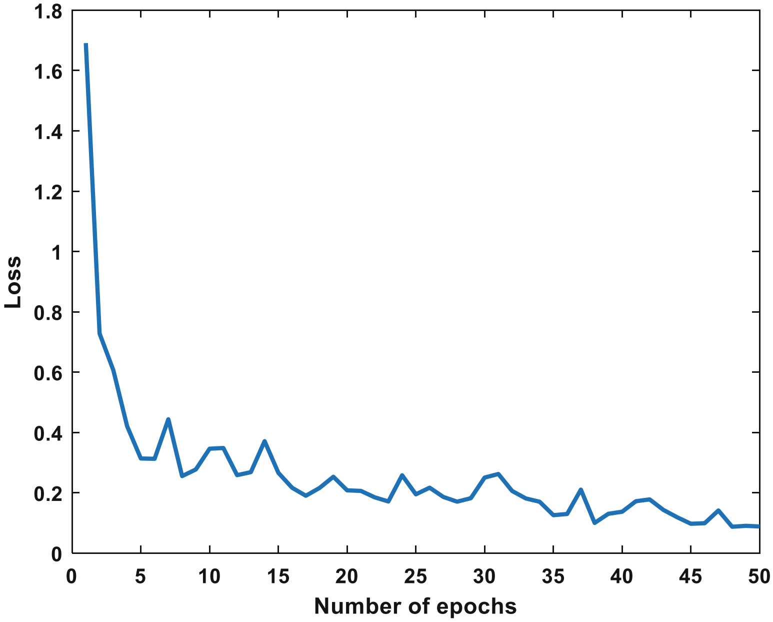


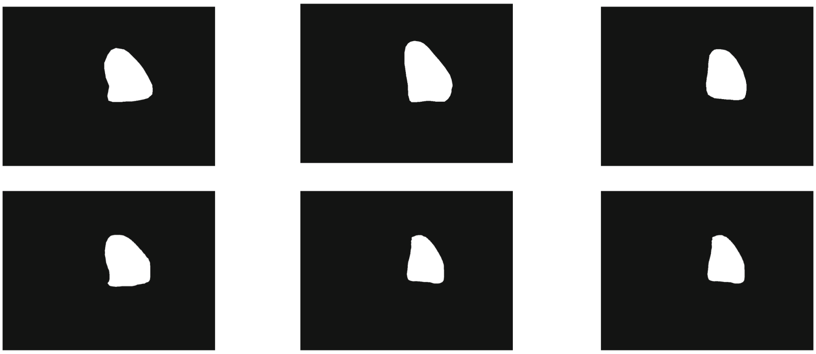
Fig. 6. The overall loss of the model

We also evaluate the segmented accuracy by using the Dice Similarity coefficient (DSC) [15, 16] It measures the overlap region between segmented and ground truth image using the following formula.

$$DSC = \frac{2|A \cup B|}{|A| + |B|}$$

DSC is equaled to the twice the number of pixels common in both ground truth and segmented binary masks divided by a total number of pixels in both marks.

We evaluate the model by calculating the DSC of all images of five patients. The data of these patients were reserved for the test purpose only and was not used for training, so all the data was unseen for the model before. The average value of DSC was 0.8940 ± 0.0365. Three examples of ground truth and segmented binary masks with DSC values are shown in Fig. 7.



DSC= 0.8983 DSC= 0.8874 DSC= 0.8470

Fig. 7. Three samples of ground truth (top) and segmented binary masks (bottom)

5 Conclusion

This paper proposed a fully automatic method for LV segmentation using Mask R-CNN model. Our trained convolutional neural network first correctly detects the ROI and then generate the mask precisely. In case of lacking training data, we successfully applied to transfer learning by first training the network with COCO weights. DSC value shows that our results are very promising and encouraging. These experimental results proposed that the Mask R-CNN model on the area of nature image can be effectively transferred to the field of echocardiography images segmentation.

As future work, the authors plan to evaluate the model on other data sets to test the robustness and generality of the proposed approach. Effect of increasing the training data on the accuracy and losses of the model will be analyzed. Also, the evaluation of the model using different evaluation matrices and clinical indices will be done in the future.

Acknowledgement. This work was supported by Fundamental Research Grant Scheme (FRGS) FP092-2018A, Ministry of Education, Malaysia.

References

1. Saldivar, F., et al.: The worldwide environment of cardiovascular disease: prevalence, diagnosis, therapy, and policy issues. J. Am. Coll. Cardiol. **60**(25), S1–S49 (2012)
2. Sinusas, A.J., et al.: Contour tracking in echocardiographic sequences via sparse representation and dictionary learning. Med. Image Anal. **18**(2), 253–271 (2013)
3. Hanif, M., Nizar, A., Khalil, A., Chan, C.K., Utama, N.P., Lai, K.W.: Pilot study on machine learning for aortic valve detection in echocardiography images **8**(xx), 1–6 (2018)
4. Ibtehaz, N., Rahman, M.S.: MultiResUNet: rethinking the U-Net architecture for multimodal biomedical image segmentation, pp. 1–25 (2019)
5. Zhang, Y., Chandler, D.M., Mou, X.: Quality assessment of screen content images via convolutional-neural-network-based synthetic/natural segmentation. IEEE Trans. Image Process. **27**(10), 5113–5128 (2018)
6. de Alexandria, A.R., Cortez, P.C., Bessa, J.A., da Silva Félix, J.H., de Abreu, J.S., de Albuquerque, V.H.C.: PSnakes: a new radial active contour model and its application in the segmentation of the left ventricle from echocardiographic images. Comput. Methods Programs Biomed., **116**(3), 260–273 (2014)
7. Dietenbeck, T., et al.: Whole myocardium tracking in 2D-echocardiography in multiple orientations using a motion constrained level-set. Med. Image Anal. **18**(3), 500–514 (2014)
8. Barbosa, D., Friboulet, D., Jan, D., Bernard, O.: Fast tracking of the left ventricle using global anatomical affine optical flow and local recursive block matching. Midas J.**10** (2014)
9. D'hooge, J., et al.: Real-time 3D interactive segmentation of echocardiographic data through user-based deformation of B-spline explicit active surfaces. Comput. Med. Imaging Graph.**38**(1), 57–67 (2013)
10. Shelhamer, E., Long, J., Darrell, T.: Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(4), 640–651 (2017)
11. Luo, G., Zhang, H., Sun, G., Wang, K., Dong, S.: "A Left ventricular segmentation method on 3D echocardiography using deep learning and snake. 2016 Comput. Cardiol. Conf. **43**, 473–476 (2017)
12. Dong, S., Luo, G., Wang, K., Cao, S., Li, Q., Zhang, H.: A combined fully convolutional networks and deformable model for automatic left ventricle segmentation based on 3D echocardiography, vol. 2018 (2018)
13. Smistad, E., Ostvik, A., Haugen, B.O., Lovstakken, L.: 2D left ventricle segmentation using deep learning. In: IEEE International Ultrason. Symposium IUS, pp. 4–7 (2017)
14. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. **39**(6), 1137–1149 (2017)
15. Khalil, A., Faisal, A., Ng, S.-C., Liew, Y.M., Lai, K.W.: Multimodality registration of two-dimensional echocardiography and cardiac CT for mitral valve diagnosis and surgical planning. J. Med. Imaging **4**(03), 1 (2017)
16. Khalil, A., Faisal, A., Lai, K.W., Ng, S.C., Liew, Y.M.: 2D to 3D fusion of echocardiography and cardiac CT for TAVR and TAVI image guidance. Med. Biol. Eng. Comput. **55**(8), 1317–1326 (2017)