Overview of Left Ventricular Segmentation in Ultrasound Images

Bogdan Milićević[1,2,*], Miljan Milošević[2,3,4], Mina Vasković Jovanović[1], Vladimir Milovanović[1], Nenad Filipović[1,2], Miloš Kojić[2,5,6]

[1] University of Kragujevac, Faculty of Engineering, Sestre Janjić 6, 34000 Kragujevac, Serbia; e-mail: bogdan.milicevic@uni.kg.ac.rs, mina.vaskovic.jovanovic@kg.ac.rs, vlada@kg.ac.rs, fica@kg.ac.rs

[2] Bioengineering Research and Development Center (BioIRC), Prvoslava Stojanovića 6, 34000 Kragujevac, Serbia; e-mail: miljan.m@kg.ac.rs, mkojic42@gmail.com

[3] University of Kragujevac, Institute for Information Technologies, Jovana Cvijića bb, 34000 Kragujevac, Serbia

[4] Belgrade Metropolitan University, Tadeuša Košćuška 63, 11000 Belgrade, Serbia

[5] Serbian Academy of Sciences and Arts, Knez Mihailova 35, 11000 Belgrade, Serbia

[6] Houston Methodist Research Institute, The Department of Nanomedicine, 6670 Bertner Ave., R7 117, Houston, TX 77030

DOI: 10.46793/ICCBI23.359M

Abstract: Due to its great temporal resolution and quick acquisition periods, two-dimensional echocardiography, or shorter 2D echo, is the most used non-invasive approach for assessing heart disease. It offers a grayscale image that anatomical details can be extracted from to evaluate heart functioning. The initial stage in quantifying cardiac function in 2D echo is the segmentation of the left ventricular (LV) walls. The primary boundary identification methods used for 2D echo at the moment are semi-automatic or manual delineation carried out by professionals. However, manual or semi-automatic approaches take a lot of time and are subjective, which makes them vulnerable to both intra- and inter-observer variability. Many researchers have tried to automate the process of left ventricle segmentation. The extensive use of deep learning algorithms has lately changed medical image analysis. The revolution has primarily been powered by supervised machine learning with convolutional neural networks. In this paper, we will provide a short overview of some of the popular deep-learning techniques for left ventricular segmentation in two-dimensional echocardiography.

Keywords: echocardiography, left ventricle, image segmentation, convolutional neural networks

1. Introduction

The most popular non-invasive imaging technique for the detection of abnormalities within the heart is a two-dimensional echocardiogram (2D echo), which has a quick acquisition time, a low cost, and a high temporal resolution [1]. The initial stage in calculating clinical characteristics is boundary detection of the left ventricle (LV) in 2D echo, also known as image segmentation. LV segmentation in 2D echo is now mostly done semi-manually. More work must be done to segment the LV wall totally automatically. Deep learning techniques can aid in automating the annotation process, but their effectiveness is constrained by the quantity and caliber of labeled training data, which can be hard to come by since annotating the raw data is laborious. Self-supervised learning can use unlabeled data that does not require input from clinical professionals. Contrastive pretraining on unlabeled data demonstrated considerable increases in performance for comparable tasks [2].

For biomedical image segmentation, UNet [3] and DeepLabV3 [4] are often used, and they have shown great performance in numerous segmentation issues. UNet is a fully convolutional network with a U-shaped design that comprises of a contracting path (encoder) and an expanding path (decoder). With skip connections between corresponding layers in the contracting and expanding paths, features are retrieved by the contracting path and then incrementally upsampled by the expanding path. To handle the problem of segmenting objects at multiple scales, DeepLabV3 was designed using modules that employ atrous convolution in cascade or in parallel to capture multi-scale context by adopting multiple atrous rates.

2. Methods

In this section, we will give a short description of commonly used neural network architectures for left ventricle image segmentation: UNet, SegAN, and DeepLabV3.

UNet comprises of an expanded path on the right and a contracting path on the left [3]. The contracting path adheres to the standard convolutional network architecture. It entails applying two 3x3 convolutions (unpadded convolutions) repeatedly, followed by a rectified linear unit (ReLU), a 2x2 max pooling operation, and a stride 2 downsampling operation for each convolution. We quadruple the number of feature channels with each downsampling step. Every step in the expansive path entails upsampling the feature map, followed by a 2x2 convolution (also known as an "up-convolution") that cuts the number of feature channels in half, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3x3 convolutions, each followed by a ReLU. The cropping is necessary due to the loss of border pixels in every convolution. At the final layer, a 1x1 convolution is used to map each 64-component feature vector to the desired number of classes. In total, the network has 23 convolutional layers. To allow a seamless tiling of the output segmentation map, it is important to select the input tile size such that all 2x2 max-pooling operations are applied to a layer with an even x- and y-size.

The segAN, a neural network architecture for medical image segmentation is based on a generative adversarial model [5]. A generator (a network that learns how to sample from the sample's underlying distribution) and discriminator (a network that aids in enhancing the quality of the data points sampled by the generator) make up a generative adversarial network model (GAN). In the segAN design, the segmentor (which also serves as a generator) uses an encoder-decoder architecture to produce anticipated segmentation images from original images. The segmentor and the ground truth pictures serve as the discriminator's two inputs, which computes the loss between the two input images. During the training, the segmentor aims to minimize the loss, whereas the discriminator aims to maximize the same loss.

DeepLabV3 offers an Atrous Spatial Pyramid Pooling (ASPP) module [4] that makes use of atrous (dilated) convolutions at various speeds to address the issue of object scale fluctuations as well as enlarge the receptive field while maintaining the spatial dimensions of the feature maps. The ASPP consists of a series of parallel, dilated convolutions with varying rates, followed by a concatenation of the outputs of those convolutions. Before upsampling back to the original resolution, the ASPP module processes features from the encoder. In [2], DeepLabV3 was chosen to train on a large publicly available EchoNet-Dynamic dataset [6], containing 20,060 annotated images from 10,030 patients. Random samples of echocardiographic images and corresponding masks from the EchoNet-Dynamic dataset are shown in Figure 1.

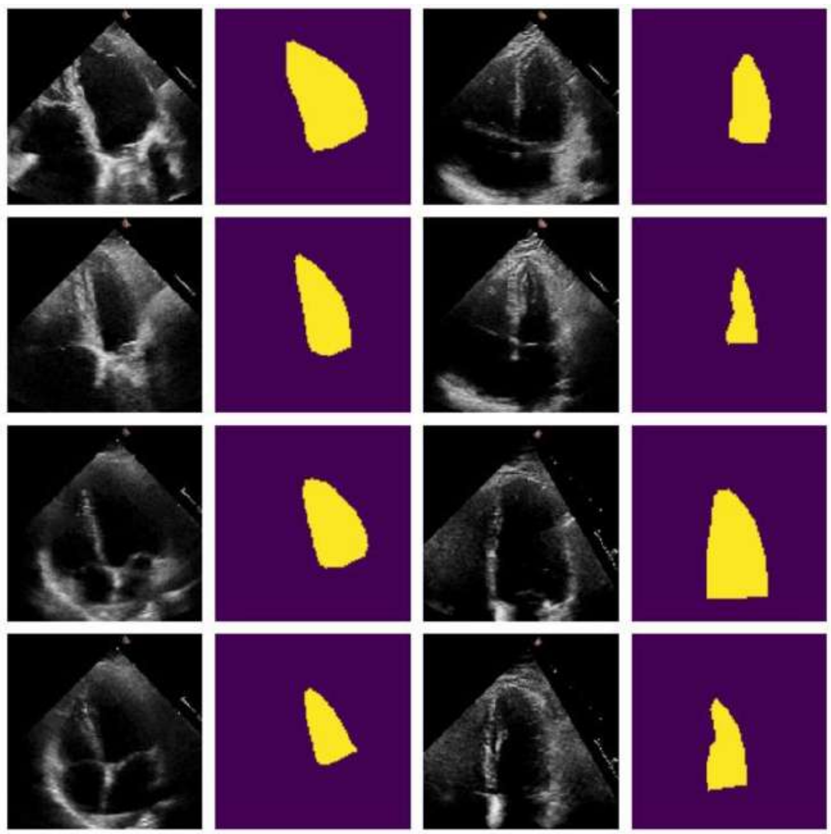

Figure 1. Random samples of echocardiographic images and corresponding masks from the EchoNet-Dynamic dataset.

3. Conclusions

Different CNN models, UNet, segAN, and DeepLabV3, have been employed to segment the LV from 2D echo images automatically by researchers. The assessment metrics and segmented images show that all of these methods achieve high performance on LV segmentation. While contrastive learning is an open research problem, it can be concluded that it could lead to an improvement in cardiac ultrasound segmentation, especially when annotated data for the downstream task is limited.

Acknowledgment

This research was supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No 952603 (http://sgabu.eu/). This article reflects only the author's view. The Commission is not responsible for any use that may be made of the information it contains. Research was also supported by the SILICOFCM project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 777204. This article reflects only the authors' views. The European Commission is not responsible for any use that may be made of the information the article contains. The research was also funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia, contract numbers [451-03-68/2022-14/200107 (Faculty of Engineering, University of Kragujevac) and 451-03-68/2022-14/200378 (Institute for Information Technologies Kragujevac, University of Kragujevac)].

References

[1] T. Kim, M. Hedayat, V. V. Vaitkus, M. Belohlavek, V. Krishnamurthy, and I. Borazjani, "*Automatic segmentation of the left ventricle in echocardiographic images using convolutional neural networks*," Quantitative Imaging in Medicine and Surgery, vol. 11, no. 5. AME Publishing Company, pp. 1763–1781, May 2021. doi: 10.21037/qims-20-745.

[2] M. Saeed, R. Muhtaseb, and M. Yaqub, "*Contrastive Pretraining for Echocardiography Segmentation with Limited Data*" arXiv, 2022. doi: 10.48550/ARXIV.2201.07219.

[3] O. Ronneberger, P. Fischer, and T. Brox, "*U-Net: Convolutional Networks for Biomedical Image Segmentation*" arXiv, 2015. doi: 10.48550/ARXIV.1505.04597.

[4] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "*Rethinking Atrous Convolution for Semantic Image Segmentation*" arXiv, 2017. doi: 10.48550/ARXIV.1706.05587.

[5] Y. Xue, T. Xu, H. Zhang, L. R. Long, and X. Huang, "*SegAN: Adversarial Network with Multi-scale L1 Loss for Medical Image Segmentation*," Neuroinformatics, vol. 16, no. 3–4. Springer Science and Business Media LLC, pp. 383–392, May 03, 2018. doi: 10.1007/s12021-018-9377-x.

[6] D. Ouyang et al., "*Video-based AI for beat-to-beat assessment of cardiac function*," Nature, vol. 580, no. 7802. Springer Science and Business Media LLC, pp. 252–256, Mar. 25, 2020. doi: 10.1038/s41586-020-2145-8.