2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)

BNU-Net: a Novel Deep Learning Approach for LV MRI Analysis in Short-Axis MRI

Wenhui Chu †, Giovanni Molina †, Nikhil V. Navkar ††, Christoph F. Eick §,
Aaron T. Becker ¶, Panagiotis Tsiamyrtzis *, Nikolaos V. Tsekos †
†*MRI Lab, Dept. of Computer Science, University of Houston, Houston, USA.*
wchu@uh.edu, gemolinaramos@uh.edu, nvtsekos@central.uh.edu
††*Dept. of Surgery, Hamad Medical Corporation, Doha, Qatar.*
nnavkar@hamad.qa
§ *DAIS Lab, Dept. of Computer Science, University of Houston, Houston, USA.*
ceick@uh.edu
¶*Dept. of Electrical and Computer Engineering, University of Houston, Houston, USA.*
atbecker@uh.edu
**Dept. of Statistics, Athens University of Economics and Business, Greece.*
pt@aueb.gr

***Abstract*—This work presents a novel deep learning architecture called BNU-Net for the purpose of cardiac segmentation based on short-axis MRI images. Its name is derived from the Batch Normalized (BN) U-Net architecture for medical image segmentation. New generations of deep neural networks (NN) are called convolutional NN (CNN). CNNs like U-Net have been widely used for image classification tasks. CNNs are supervised training models which are trained to learn hierarchies of features automatically and robustly perform classification. Our architecture consists of an encoding path for feature extraction and a decoding path that enables precise localization. We compare this approach with a parallel approach named U-Net. Both BNU-Net and U-Net are cardiac segmentation approaches: while BNU-Net employs batch normalization to the results of each convolutional layer and applies an exponential linear unit (ELU) approach that operates as activation function, U-Net does not apply batch normalization and is based on Rectified Linear Units (ReLU). The presented work (i) facilitates various image preprocessing techniques, which includes affine transformations and elastic deformations, and (ii) segments the preprocessed images using the new deep learning architecture. We evaluate our approach on a dataset containing 805 MRI images from 45 patients. The experimental results reveal that our approach accomplishes comparable or better performance than other state-of-the-art approaches in terms of the Dice coefficient and the average perpendicular distance.**

***Index Terms*—Magnetic Resonance Imaging; Batch Normalization; Exponential Linear Units**

I. Introduction

In the United States, cardiovascular disease is the primary cause of death for both males and females [1]. One of the parameters that cardiologists examine in the diagnosis of heart disease is the amount of blood ejected by the left ventricle [2]. Physicians use Magnetic Resonance Imaging (MRI) scans to obtain relevant images of the cardiac areas to assess the structural and functional features for cardiovascular diagnosis and disease management in a non-invasive manner [2]. Main indicators of cardiovascular disease are the left ventricle (LV) end-systolic volume (ESV), the end-diastolic volume (EDV), and the ejection fraction (EF) [3]. The segmented contour of the left ventricle has been crucial in determining ESV and EDV. Therefore, coherent and precise segmentation of the LV from MRI images is critical to the precision of identification of ESV, EDV, and EF, which is crucial to determine cardiac disease in a non-invasive manner. Currently, it takes physicians several minutes to diagnose a patients condition and the obtained results are not easily reproducible [4]. Therefore, the development of automated cardiac segmentation methods based on magnetic resonance imaging datasets is a crucial step to facilitate this cumbersome diagnosis process [5]. Using an improved automatic process to determine heart parameters and function can lead to a quicker, coherent diagnosis and generate a repeatable diagnostic process. While research over the past decade has addressed some of the above-mentioned technical difficulties in achieving tangible progress on automatically generated ventricle segmentation from short-axis MRI, the corresponding automated segmentation contours still have to be significantly improved for clinical use [4]. In addition, assessment of prior studies on small benchmark datasets, which may not reflect real-world variability in image resolution and heart physiological and functional features across locations, institutions, and populations, are restricted in scope [6]. This paper demonstrates an enhanced technique for automatic left ventricle segmentation in MRI pictures.

In this paper, we first present the method to augment the original images and ground truth deformation. We then present a novel CNN architecture, which we call the BNU-Net ("Batch-Normlization-U-Net") that leverages the power of a fully convolutional neural network. It is designed to process 2D MRI images as inputs and produce a labeled slice as output. Finally, we validate our method by comparing with a state-of-the-art network on a large database of 805 cardiac images from the Sunnybrook dataset. We demonstrate fast results for calculating MRI test volumes and we provide an experimental evaluation that directly compares with other methods using the same test data.

II. METHODOLOGY

A. The Structure of Network

A traditional learning process [7] faces numerous issues that limit the efficacy of the automatic diagnosis of the images. The first issue in training a fully convolutional network (FCN) for medical segmentation is that they are computationally expensive because of the number of FLOPs required to process an image; this is why an encoder architecture is preferred [6]. The second issue of FCN is internal covariate shift where the training kernel is encumbered by the distribution of change of input features [8] and also results in unsatisfactory learning speed. Therefore, we use *batch normalization* to address the problem of the slow processing of the training data. In addition, a regular FCN does not provide a well-defined segmentation, which means there is no obviously defined boundary between the pictures, and this causes ESV and EDV measurements to be inaccurate [6].

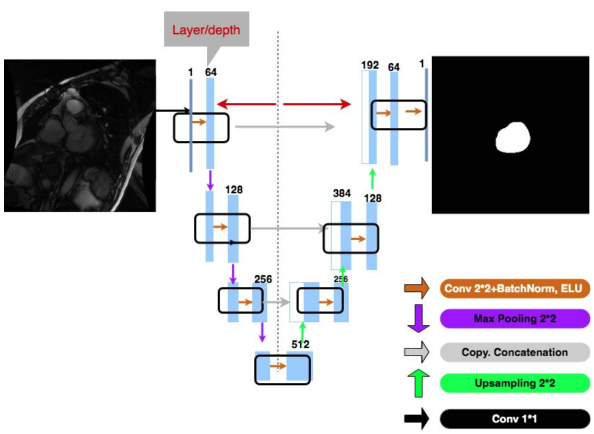

Fig. 1. Architecture of the BNU-Net convolutional network. (a) The contraction path is responsible for feature extraction. (b) Batch normalization is performed after each convolution in the convolutional layer.

In this paper, we present the feasibility of deep learning approaches for a fully convolutional neural network architecture which we call BNU-Net. Its architecture is shown in Fig. 1. This work modified and improved the U-Net architecture [9] to make it work with fewer training images and improve performance. BNU-Net has 11 convolution layers, 4 layers on the contraction path and 7 layers on the expansion path. In the contractive path, each convolutional block corresponds to 2 convolutional layers followed by 2×2 max pooling layers. One of the key features of this architecture is the use of concatenation path connections between down-sampling and up-sampling layers for the purpose of fusing local and global information. The expanding path is followed by a series of

TABLE I
COMPARISON BETWEEN OUR PROPOSED BNU-NET MODEL WITH ELU AND BNU-NET WITH RELU

	BNU-Net with ELU	BNU-Net with ReLU
Dice mean	**0.92**	0.90
Dice std	**0.04**	0.06
Sensitivity	**0.96**	0.95

convolutional filters and concatenations of feature maps from the contracting path at each stage for more precise localization as shown by the gray arrows in the figure. As in the U-Net architecture, the network used consists of two paths: the contracting path and the expansive path. There are six repeated applications of 3×3 convolutions, each followed by a rectified linear (ReLU) activation function in the contracting path. The above architecture is different from the U-Net based approach that contains four layers with batch normalization [10], and exponential linear in the contracting path.

Because the training process is affected by the parameters of each input layer, small changes of network parameters will affect the network greatly. The batch normalization [10] is a mechanism that aims to make the training of neural networks more stable to a given network layer. One of the motivations for the development of batch normalization was to allow each layer of the network to train more efficiently. It reduced problems when the input values change and reduced internal covariate shift. We saw that when the inputs of the network were transformed such that the first two moments (mean and variance) were respectively set to zero and unit values, the network converges faster and makes the optimization landscape smoother.

Different from the original U-Net which used ReLU [4], we integrated exponential linear units to make the mean activations closer to zero, which helps improve the efficiency of the data compiling and calculations process [5]. we applied ReLU and ELU separately in BNU-Net and the segmentation evaluation metrics as shown in table I. The improved method also incorporates *cropping 2D*, which is a cropping layer used to crop feature maps and concatenate multiple feature maps from the contraction path. Cropping is essential as each convolution will blur the pixels of the border of MRI image, impacting the accuracy of the estimation.

B. Dataset

We evaluated our approach using the dataset from the MICCAI 2009 challenge on automated left ventricle segmentation; it was downloaded from the Imaging Research Centre for Cardiovascular Interventions at Sunnybrook Health Sciences Centre [11]. The Sunnybrook dataset comprises cine MRI from 45 patients (a total of 805 images) suffering from different cardiac conditions: heart failure with infarction (12 cases), heart failure without infarction (12 cases), hypertrophy (12 cases) and healthy patients (9 cases). Each time series consists of 6 to 12 2D cine stacks with 8 mm slice thickness and 1.3

mm to 1.4 mm in-plane resolution [11]. For the purpose of the research, the MRI image files were originally split into training, validation, and testing sets in the ratio of 15 : 15 : 15. Each patient has 12 to 28 images. The database also provides MRI ground truth medical images manually segmented by an expert.

III. RESULTS

A. Experimental Studies

The normal unsupervised learning process [12] requires a large amount of the data to yield a satisfactory classification. This is impractical given the limited amount of the data available. We present a neural network and training approach which depends heavily on information increase for more efficient use of the dataset noted. Specifically, the improved methods of processing use elastic deformations and random affine transformations to do data augmentation. Elastic Transform was first proposed by Patrice et al. [13] in 2003. It was first applied in the MNIST handwritten digit recognition data set. The implementation performs on-the-fly elastic transformations for efficient data augmentation during training. We also perform affine transformation (rotation, scaling, translation) to augment the training set and mitigate overfitting and improve generalization. Figure 2 shows examples of using elastic transform and random affine transformation.

We applied elastic deformations to the available training images in our data set. The elastic deformation enables the neural network to grasp invariance to this kind of deformations without seeing these changes throughout the annotated corpus of the picture. Such a process is especially essential in biomedical segmentation as deformation is one of the most prevalent tissue variations, which allows effective simulation of realistic deformations.

B. Experimental Metrics

To measure the performance of the BNU-Net, we provide an overview of the main metrics reported in the literature for comparative purposes.

Let A and G be the automatically segmented and ground truth (manual) region/contours, respectively. Let a and g be the predicted (automated) and ground truth (manual) contours delineating the object class in short-axis MRI, respectively.

1) The Dice Metric: a measure of contour overlap between automatically and manually segmentation, and is defined as:

$$Dice = \frac{2(A \cap G)}{A + G} \tag{1}$$

where $A \cup G$ denotes the intersection between A and G, and $A \cap G$ means the union between A and G. The Dice index varies from zero to one. Zero indicates a total mismatch with the ground truth and one indicates there is perfect match.

2) Average perpendicular distance (APD) measures the distance in mm between contours a and g, averaged over all contour points. A low value implies that the two contours match closely.

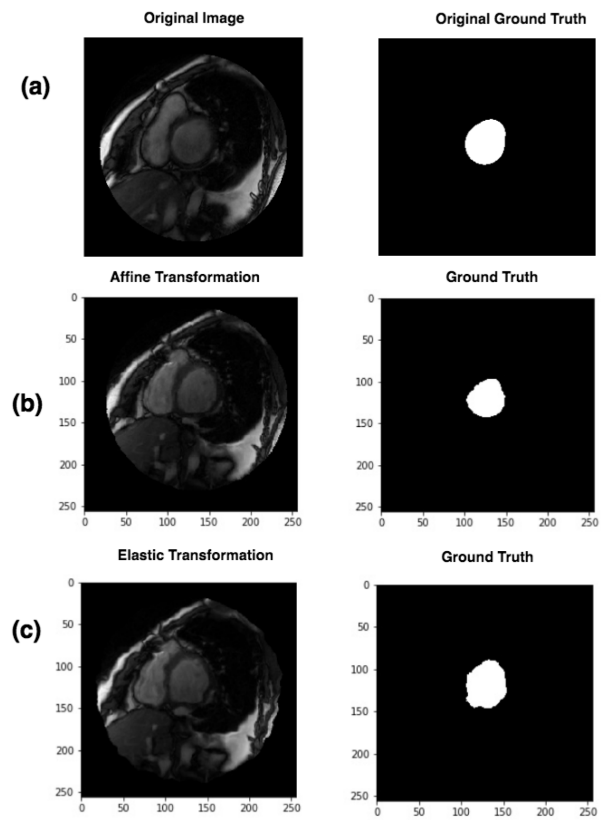

Fig. 2. The contour comparison based on MRI input. (a) Sunnybrook Image and corresponding label. (b) Perform affine transformations (translation=0.03, rotation=4.6 and scaling=[0.98,1.02]) (c) Perform elastic transformation to augment the training set, in which alpha=[28, 30], sigma=[3.5, 4.0]

3) In addition to the two similarity measures above, two additional metrics used to evaluate the performance of the segmentation are the sensitivity and the specificity. They are computed using the formulas:

$$Sensitivity = \frac{T_1}{B_1} \tag{2}$$

$$Specificity = \frac{T_0}{B_0} \tag{3}$$

where T_1 and T_0 are the total number of correctly predicted object and background pixels, respectively. The total number of object and background pixels are denoted by B_1 and B_0, respectively.

C. Experimental Results

Table II summarizes the results of BNU-Net with and without data augmentation and compares them to U-Net results. From the graph, we can see that with and/or without data augmentation, the BNU-Net achieved the best dice score and sensitivity score. We found that the left ventricle recognition is improved when using data augmentation.

TABLE II
OUTPUT OF THE MODEL AND EFFICIENCY METRICS RESULTS

	Unet with data augmentation	Unet without data augmentation	BNU-Net with data augmentation	BNU-Net without data augmentation
Dice mean	0.88	0.87	**0.93**	0.92
Dice std	0.09	0.11	**0.03**	0.04
Sensitivity	0.96	0.95	**0.97**	0.96

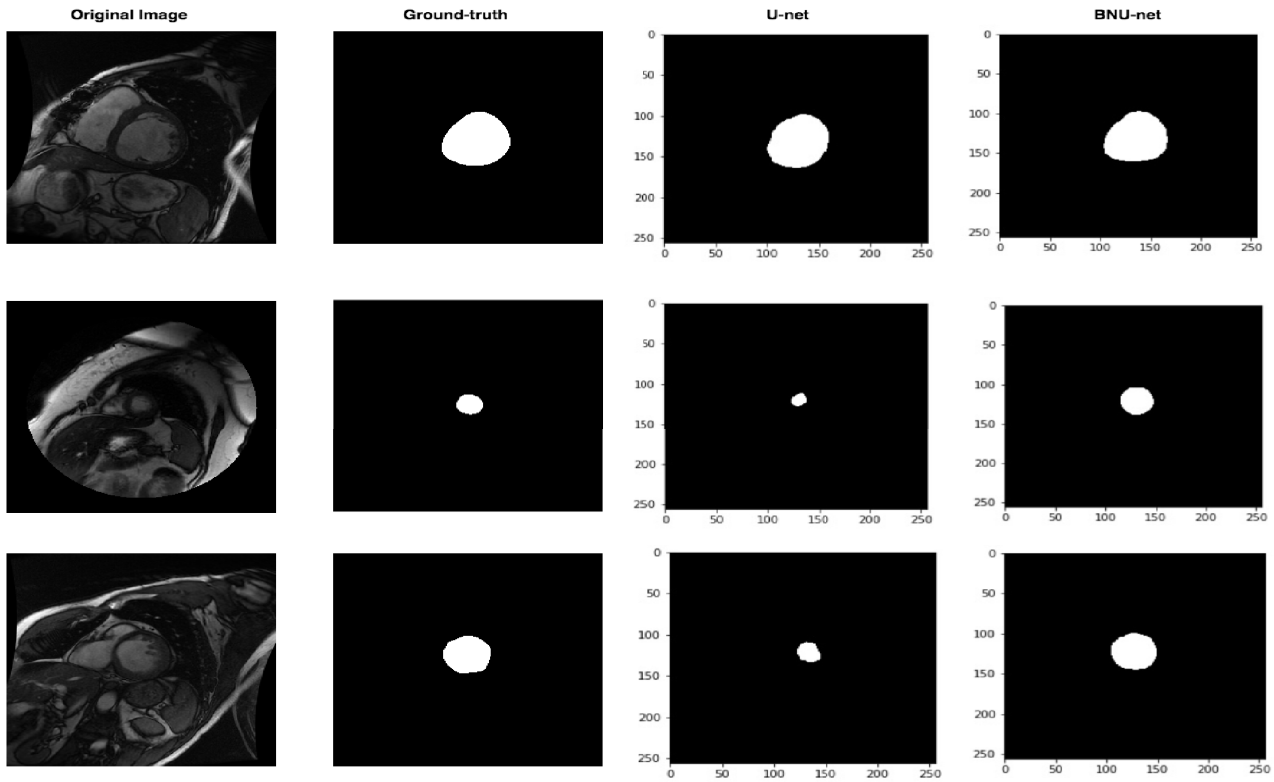

Fig. 3. Segmentation predictions on raw inputs from the Sunnybrook dataset.

TABLE III
SEGMENTATION RESULTS ON THE SUNNYBROOK DATASET, COMPARED TO THE PERFORMANCE FROM THE STATE OF THE ART METHODS

	Dice Mean	Dice Std	APD (mm)
BNU-Net with data augmentation	**0.93**	**0.03**	**1.94**
BNU-Net without data augmentation	0.91	0.04	2.06
Huang et al[2]	0.89	0.04	2.16
Ngo and Carneiro[14]	0.90	**0.03**	2.08
Hu et al[15]	0.89	**0.03**	2.24
Liu et al[16]	0.88	**0.03**	2.36

Figure 3 shows two examples of the output segmentation. We have demonstrated the efficacy and utility of the BNU-Net architecture for semantic segmentation in cardiac MRI. The quantitative evaluation of the BNU-Net is based on the the dice coefficient, and the average perpendicular distance.

This fast training time with limited resources, including the amount of the images and the processing power, has made it possible to apply this methodology to more settings. Among the methods in Sunnybrook dataset, four groups report their performances on the 45 patients cases that have the same data and split. The performances on these cases of BNU-Net and these methods are shown in Table III. We find that BNU-Net with data augmentation achieves best dice mean, dice std, and APD performance compared with other group's results.

IV. DISCUSSION AND CONCLUSION

Today, cardiac segmentation of two-dimensional medical images is vital for medical image analysis. During the last decade, machine learning and deep learning has been applied to understand images and segmentation, and has proven versatile compared to traditional methods. Its practical implementation in the clinical realm has been limited since well-labelled medical data is harder to obtain. U-Net was originally proposed for general biomedical image segmentation [7]. In this paper, we proposed a novel method called BNU-Net model for the segmentation of the left ventricle. The model consists of two main paths which are the encoding path and the decoding (up-sampling) path. In the encoding path, successive layers consist of the convolutional filters followed by an activation function (ELU) which is used to learn a representation of the input image. To solve the internal covariate shift and help the network train faster and achieve higher accuracy, one important modification in the model is that we applied batch normalization after the convolutional filters. We use batch normalization throughout the BNU-Net, which helps bypass sharp local minima and correct activations to be zero-mean and of unit standard deviation. It allow us to prevent small changes in layer parameters from amplifying through the deep network [9]. In addition, we are less careful about initialization. A series of 2×2 max pooling layers were used to down-sample the output and capture the features crossing different layers. The decode path consists of the up-sampling of the channels followed by convolutional blocks and ELU. One of the important steps in up-sampling is the concatenated connection from the encode path, which is used to propagate context information to detect the fine, higher resolution features.

We applied data augmentation when training our model by applying affine transformations and elastic deformations to the Sunnybrook biomedical segmentation dataset. It achieves outstanding performance, only needs a few annotated images, and achieved high training speed on each epoch (11 seconds) with a batch size of 16 on a NVIDIA GeForce Titan X Pascal GPU. At test time, our model segments the whole series of images in less than 12 seconds. We also used the smaller images size of 176×176 pixels, which yielded slightly better performance results on the testing set and faster training time. With 1 GPU and 200 epochs, 256×256 images took 36 minutes to complete training, while 176×176 images took 30 minutes.

The paper introduced a deep learning approach for detecting and segmenting left ventricals with fast training speed. We presented the BNU-Net architecture for constructing, training, and testing with batch-normalized. It is based on the premise that covariate shift, which is known to complicate the training of machine learning systems, also applies to sub-networks and layers, and removing this shift from internal activations of the network may aid in training. By combining multiple models trained with Sunnybrook dataset, we performed better segmentation than other networks. Moreover, we identified the BNU-Net network with batch normalization that the inputs stabled and shifted to maintain the network expressivity. A real-time implementation of the proposed method is sufficiently fast to be used for intraoperative registration of preoperative cardiac anatomy [17].

ACKNOWLEDGMENT

This work was supported by the National Science Foundation award CNS-1646566. All opinions, findings, conclusions or recommendations expressed in this work are those of the authors and do not necessarily reflect the views of our sponsors.

REFERENCES

[1] Centers For Disease Control and Prevention, "Heart Disease Statistics and Maps," [Online]. Available: https://www.cdc.gov/heartdisease/facts.html

[2] Huang, S., Liu J., Lee L.C., Venkatesh S., Teo L., Au C., Nowinski W. "An image-based comprehensive approach for automatic segmentation of left ventricle from cardiac short axis cine MR images." Journal of digital imaging, vol 24, no 4, pp. 598608.2011. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3138938/

[3] Kerkhof, P. L. M. "Characterizing Heart Failure in the Ventricular Volume Domain." Clinical Medicine Insights: Cardiology. 2015. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4345934/

[4] W. Wang, Y. Wang, Y. Wu, T. Lin, S. Li and B. Chen, "Quantification of Full Left Ventricular Metrics via Deep Regression Learning With Contour-Guidance," in IEEE Access, vol. 7, pp. 47918-47928, 2019. doi: 10.1109/ACCESS.2019.2907564. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8674807

[5] J. Cong, Y. Zheng, W. Xue, B. Cao and S. Li, "MA-Shape: Modality Adaptation Shape Regression for Left Ventricle Segmentation on Mixed MR and CT Images," in IEEE Access, vol. 7, pp. 16584-16593, 2019. doi: 10.1109/ACCESS.2019.2892965. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8611328

[6] Margeta, J., Geremia E., Criminisi A., Ayache N, "Layered spatio temporal forests for left ventricle segmentation from 4D cardiac MRI data." MICCAI workshop: Statistical Atlases and Computational Models of the Heart (STACOM). 2012. [Online]. Available: https://www.microsoft.com/en-us/research/publication/layered-spatio-temporal-forests-for-left-ventricle-segmentation-from-4d-cardiac-mri-data/

[7] Shengfeng Liu, Yi Wang, Xin Yang, Baiying Lei, Li Liu, Shawn Xiang Li, Dong Ni, Tianfu Wang, "Deep Learning in Medical Ultrasound Analysis: A Review,Engineering," Vol 5, Issue 2, 2019, pp 261-275, ISSN 2095-8099, https://doi.org/10.1016/j.eng.2018.11.020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2095809918301887

[8] X.Zhou and G.Yang, "Normalization in Training U-Net for 2-D Biomedical Semantic Segmentation," in IEEE Robotics and Au- tomation Letters, vol. 4, no. 2, pp. 1792-1799, April 2019. doi: 10.1109/LRA.2019.2896518. [Online]. Available: https://arxiv.org/abs/1809.03783

[9] Ronneberger, Olaf et al. "U-Net: Convolutional Networks for Biomedical Image Segmentation. ArXiv abs/1505.04597 (2015). [Online]. Available: https://arxiv.org/abs/1505.04597

[10] Ioffe, Sergey and Szegedy, Christian. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift.." Paper presented at the meeting of the ICML, 2015. [Online]. Available: https://dl.acm.org/citation.cfm?id=3045118.3045167

[11] Phi Vu Tran, "A Fully Convolutional Neural Network for Cardiac Segmentation in Short-Axis MRI," 2016. [Online]. Available: https://arxiv.org/abs/1604.00494

[12] Croitoru, Ioana and Bogolin, Simion-Vlad and Leordeanu, Marius. "Unsupervised Learning of Foreground Object Segmentation." vol 127, no 9, pp.1279-1302 https://doi.org/10.1007/s11263-019-01183-3 [Online]. Available: https://link.springer.com/article/

[13] P.Y.Simard, D.Steinkraus and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," Seventh International Conference on Document Analysis and Recognition, Proceedings., Edinburgh, UK, 2003, pp. 958-963. doi: 10.1109/ICDAR.2003.1227801. [Online]. Available: https://ieeexplore.ieee.org/document/1227801

[14] Ngo, T.A., G. Carneiro. "Left ventricle segmentation from cardiac MRI combining level set methods with deep belief networks." 20th IEEE International Conference on Image Processing. pp. 695-699, 2013. [Online]. Available: https://ieeexplore.ieee.org/document/6738143

[15] Hu, H., Liu H., Gao Z., Huang L. "Hybrid segmentation of left ventricle in cardiac MRI using gaussian-mixture model and region restricted dynamic programming." vol. 31, 5 2013, pp. 575-584. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0730725X12003670

[16] Liu, H., Hu H., Xu X., Song E., 2012. "Automatic Left Ventricle Segmentation in Cardiac MRI Using Topological Stable-State Thresholding and Region Restricted Dynamic Programming." Academic Radiology, vol 19, no. 6, pp. 723 - 731. 2012. [Online]. Available: https://doi.org/10.1016/j.acra.2012.02.011

[17] X. Gao, N. V. Navkar, D. J. Shah, N. V. Tsekos and Z. Deng, "Intraoperative registration of preoperative 4D cardiac anatomy with real-time MR images," 2012 IEEE 12th International Conference on Bioinformatics Bioengineering (BIBE), Larnaca, 2012, pp. 583-588. doi: 10.1109/BIBE.2012.6399737 [Online]. Available: https://ieeexplore.ieee.org/document/6399737