# Ultrasound segmentation of rat hearts using convolution neural networks

James D. Dormer, Rongrong Guo, Ming Shen, Rong Jiang, Mary B. Wagner, et al.

James D. Dormer, Rongrong Guo, Ming Shen, Rong Jiang, Mary B. Wagner, Baowei Fei, "Ultrasound segmentation of rat hearts using convolution neural networks," Proc. SPIE 10580, Medical Imaging 2018: Ultrasonic Imaging and Tomography, 105801A (6 March 2018); doi: 10.1117/12.2293558

SPIE.

Event: SPIE Medical Imaging, 2018, Houston, Texas, United States

# Ultrasound Segmentation of Rat Hearts Using Convolution Neural Networks

James D. Dormer [1], Rongrong Guo [1], Ming Shen [2], Rong Jiang [2], Mary B. Wagner [2], Baowei Fei [1,3,4 *]

[1]Department of Radiology and Imaging Sciences, Emory University, Atlanta, GA
[2]Department of Pediatrics, Emory University, Atlanta, GA
[3]Department of Biomedical Engineering, Emory University and Georgia Institute of Technology, Atlanta, GA
[4]Winship Cancer Institute of Emory University, Atlanta, GA
*E-mail: bfei@emory.edu; web: https://fei-lab.org

## ABSTRACT

Ultrasound is widely used for diagnosing cardiovascular diseases. However, estimates such as left ventricle volume currently require manual segmentation, which can be time consuming. In addition, cardiac ultrasound is often complicated by imaging artifacts such as shadowing and mirror images, making it difficult for simple intensity-based automated segmentation methods. In this work, we use convolutional neural networks (CNNs) to segment ultrasound images of rat hearts embedded in agar phantoms into four classes: background, myocardium, left ventricle cavity, and right ventricle cavity. We also explore how the inclusion of a single diseased heart changes the results in a small dataset. We found an average overall segmentation accuracy of 70.0% ± 7.3% when combining the healthy and diseased data, compared to 72.4% ± 6.6% for just the healthy hearts. This work suggests that including diseased hearts with healthy hearts in training data could improve segmentation results, while testing a diseased heart with a model trained on healthy hearts can produce accurate segmentation results for some classes but not others. More data are needed in order to improve the accuracy of the CNN based segmentation.

**Keywords:** Cardiac ultrasound, Image segmentation, Heart disease, Convolutional neural networks, Ultrasound, Cardiovascular disease, Myocardium segmentation

## 1. INTRODUCTION

Each year, cardiovascular diseases lead to more deaths in the United States than any other disease[1]. And each year, millions of ultrasound imaging procedures are performed in the pursuit of better cardiovascular care[2]. Ultrasound imaging is commonly used in a clinical setting to investigate and diagnose cardiovascular health because of the low cost, minimal risk, and short acquisition times[3]. The technology has even improved to the point where real-time monitoring devices are emerging onto the market[4]. However, ultrasound images are prone to a variety of artifacts, related to both the hardware used and the physiology of the patient[5, 6]. This can make it difficult to accurately segment cardiac volumes on ultrasound manually by trained radiologists or relying on simple intensity thresholding methods[7, 8]. Previous groups have developed computerized cardiac segmentation methods on ultrasound images using a variety of approaches, including Bayesian probability, sparse matrix transforms, and FSL[9-11]. An automated method that can accurately identify the chambers of the heart and the myocardial boundary could provide real-time measurement estimates for cardiac function, enabling clinical staff to more easily monitor patients during procedures or status following surgery. Myocardial identification can also be useful to quickly create masks to use during cardiac registration between different imaging modalities[12, 13].

One promising solution for automated segmentation is the use of convolutional neural networks (CNN). CNN's have recently emerged as a powerful tool for medical segmentation due to their flexibility[14-16]. However, accurate segmentation results are reliant on well-trained models, which can require many samples from hundreds or thousands of different patients, depending on how similar the object is between patients. In the case of hearts, the structure and size

Medical Imaging 2018: Ultrasonic Imaging and Tomography, edited by Neb Duric, Brett C. Byram, Proc. of SPIE Vol. 10580, 105801A · © 2018 SPIE
CCC code: 1605-7422/18/$18 · doi: 10.1117/12.2293558

can vary among healthy subjects, and even more so when including those with cardiovascular disease. Shape distortions due to cardiovascular disease can vary from slight myocardial wall thickening due to infarction or more extreme ventricular distensions, such as that seen in congenital heart disease[17].

In this work, we explore using convolutional neural networks to segment ultrasound images of rat hearts embedded in phantoms. In addition, we include a rat with right ventricle heart failure to investigate how the segmentation results change when an extreme case is included.

## 2. METHODS

### *2.1 Sample Preparation and Imaging*

A summary of the processing workflow is shown in Figure 1. First, hearts were excised from five Sprague Dawley rat, one of which had undergone pulmonary artery banding in order to model heart failure[18], following Emory University IACUC-approved standards. Each was promptly perfused with PBS to remove blood from the heart chambers. Once thoroughly rinsed, the hearts were fixed in formalin overnight. Each heart was then rinsed with cold PBS and embedded in 2% agar phantoms, with each chamber being carefully filled with agar to ensure structural integrity and measurement repeatability during imaging. The rat hearts were imaged in B-mode using a 30 MHz probe on a Vevo 2100 ultrasound system (FUJIFILM VisualSonics, Inc., Toronto, Canada) in the short-axis view. Each image was saved and exported as a TIFF for processing.

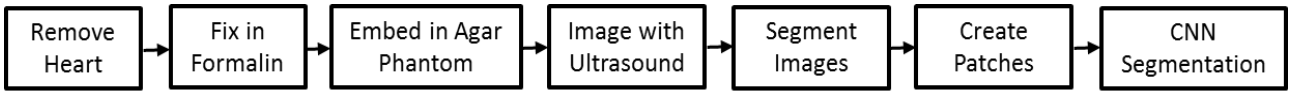


Figure 1. Complete workflow summary for the presented project. Following removal, the hearts were fixed and embedded into phantoms. Each phantom was then imaged with ultrasound, and the images segmented by a trained radiologist to create masks. The masks were used to create patches which were subsequently used in a CNN.

### *2.2 Data Processing*

The ultrasound images were loaded into Matlab (The MathWorks, Inc., Natick, MA), combined to form a 3D volume, and saved in the Analyze 7.5 data format. The ultrasound volumes were then opened in Analyze 10.0 (AnalyzeDirect, Overland Park, KS) where the myocardium (Myo), left ventricle cavity (LVC), and right ventricle cavity (RVC) were segmented by a radiologist to create binary masks. The hearts were also rotated to ensure similar orientations along the long axis. Patches were created in Matlab using the masks as labels, with each patch being centered on a pixel from the mask. Any pixel not part of an existing mask was used to create background (BG) patches.

### *2.3 Convolutional Neural Network*

The CNN dataset for each heart was generated by randomly selecting 5000 patches from each class to ensure no training bias from class imbalance in the network, for a total of 20,000 patches per heart. These patches were fed into TensorFlow[19] using the AdaDelta[20] optimizer for classification into one of the four classes. An example of the mask used to create the patches is displayed in Figure 2. The CNN (Figure 3) consisted of four convolutional layers followed by two fully connected layers. Convolutions were performed using the 'SAME' specification. After these fully connected layers, each patch is assigned to one of the four classes mentioned above. Training was performed using leave-one-out cross-validation. Parameters were adjusted to optimize the overall accuracy of the segmentation.

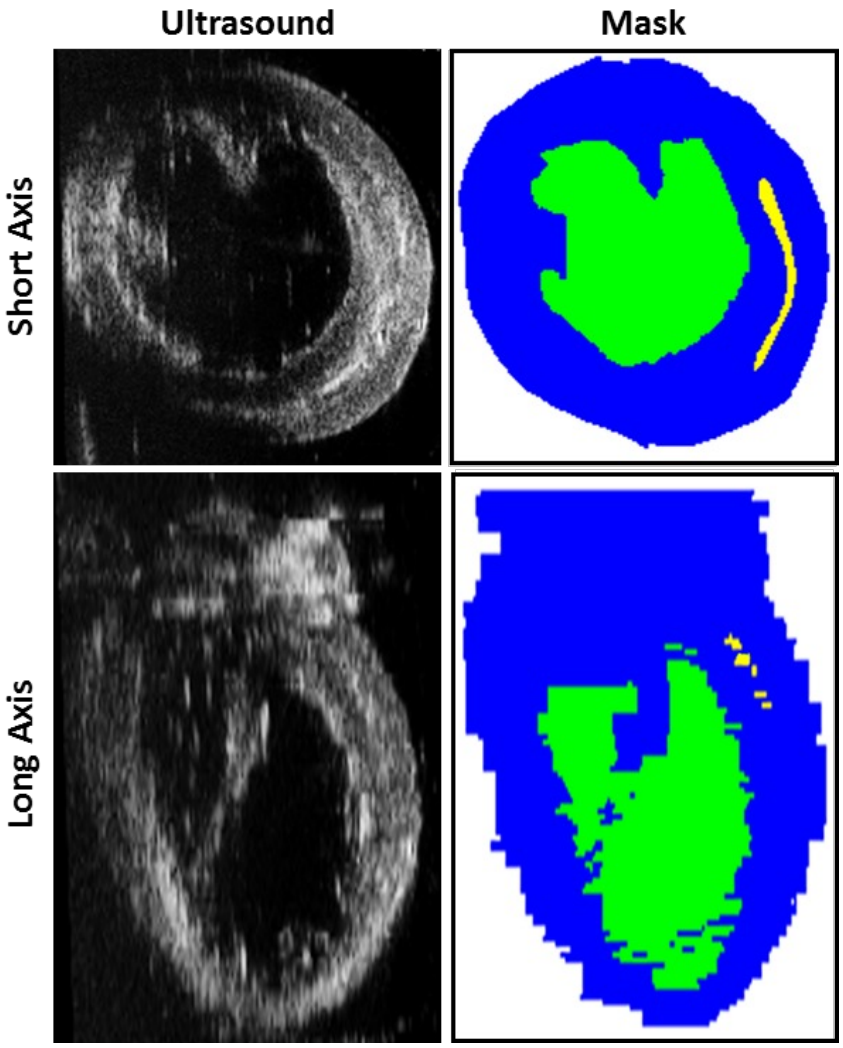


Figure 2. Gold standard segmentation examples. Segmentations were performed in the short-axis view. On the masks, blue represents the myocardium, green the interior of the LV, and yellow the interior of the right ventricle. Background was considered anything in the volume not segmented by the radiologist.

Three different scenarios were considered for the CNN. In Scenario 1, 20,000 patches from each of three healthy hearts, for a total of 60,000 patches, were used to train a CNN model. This model was then tested on 20,000 patches from the fourth healthy heart. Scenarios 2 and 3 were investigated using four healthy hearts and the diseased heart with standard leave-one-out cross-validation. This provided 80,000 training patches, 20,000 patches from each of the four hearts used for training, and 20,000 patches for validation. Scenario 2 used the result when the four healthy hearts were used as the training dataset and the diseased heart was left out for validation. Scenario 3 used the remaining results, where three healthy hearts and the diseased heart were used for training and the model was tested on the healthy heart left out. This produced a total of four results, one for each healthy heart.

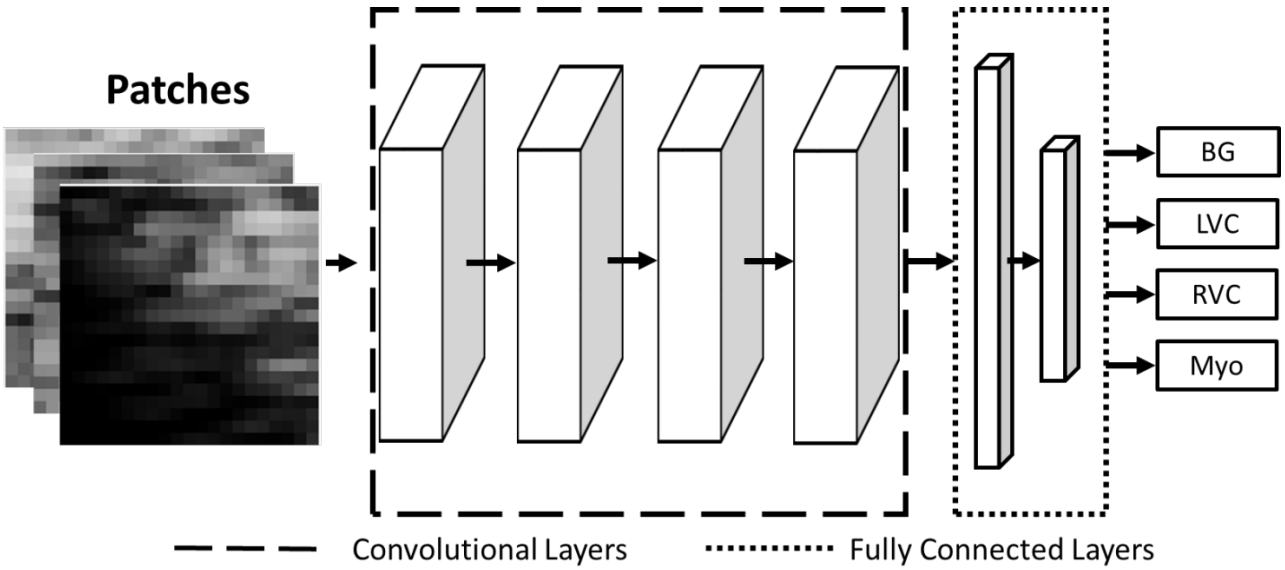


Figure 3. Diagram of the neural network used for the rat heart segmentation. The input data was in the form of 2D patches and fed into four convolutional layers followed by two fully connected layers. After the fully connected layers, a classification label was applied to each patch.

*2.4 Validation*

Validation was performed by comparing the CNN classification label to the actual label of the patch based on the gold standard from the radiologist. Accuracy for each class was calculated as the percent of correctly labeled patches over the total number of patches for that class. In addition to individual class accuracies (Myo, LVC, RVC, and BG), the overall accuracy was calculated as a sum of all the correct patches divided by the total number of patches used. Chamber accuracy was then calculated as the number of correct LVC and RVC patches over the total number of patches for both classes.

## 3. RESULTS

A visual comparison between a healthy and diseased rat heart is shown in Figure 4, with the image acquired using B-mode ultrasound in the short-axis orientation near the mid-ventricular level. The overall size of the diseased heart is not noticeably different than the healthy heart. However, inside the diseased heart, the left ventricle is noticeably shrunken and less circular than that of the healthy heart. In contrast, the right ventricle of the diseased heart has a larger volume and a more half-circle shape compared to the more crescent shape (white arrow) seen in the healthy heart. This is primarily due to the distortion of the ventricular septum, denoted by the red arrows in the figure, and not from any change in the right ventricle exterior myocardial wall.

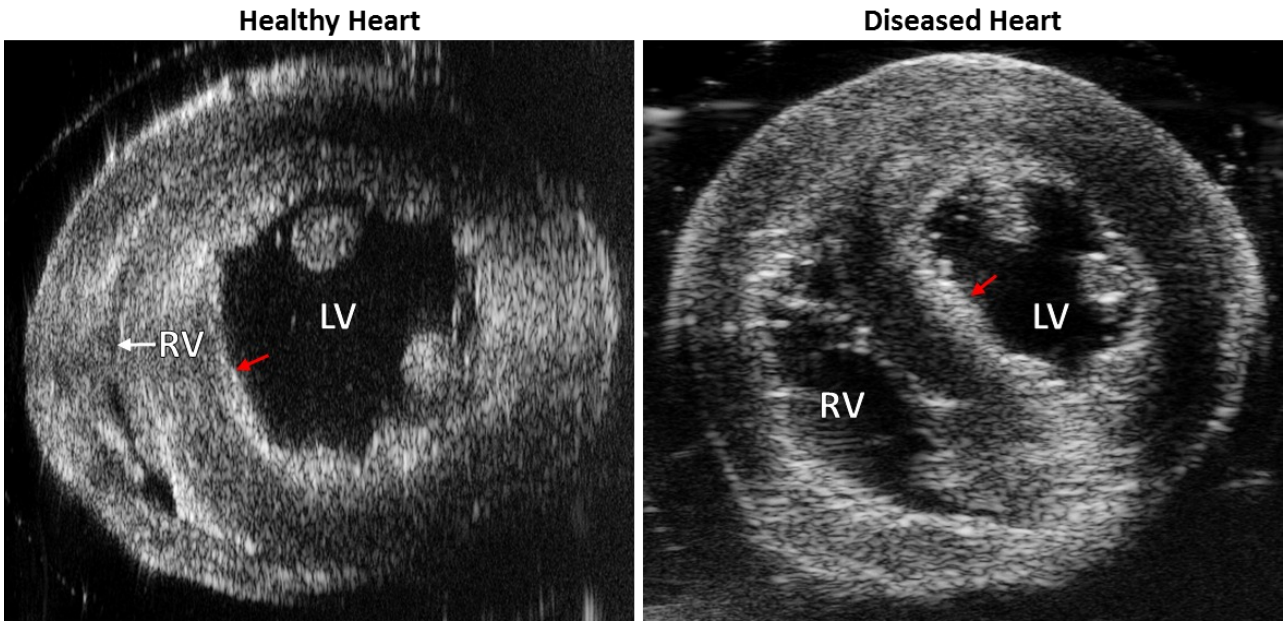


Figure 4. A visual comparison between a healthy rat heart (left) and the diseased heart (right), which had undergone pulmonary artery banding. The images were acquired using B-mode ultrasound near the midpoint of the ventricles. There is a significant difference in left and right ventricle size between the two hearts. Additionally, the shape of the ventricular septum (red arrows) shows more curvature in the healthy heart than the diseased heart.

The CNN parameters were optimized using the overall average accuracy of Scenario 3: three healthy and one diseased hearts are used as the training data and one additional healthy heart is used as the testing data. Training was performed for 20 epochs, with each epoch representing a round of training on all the training data. The testing data was evaluated at the end of each epoch to prevent any training bias. The result with the highest overall accuracy for each rat heart in the testing dataset was chosen as the final result during the leave-one-out cross-validation. Patches were loaded with a batch size of 5 and the order of patches was randomly shuffled after each epoch. Patch sizes of 11 x 11 and 21 x 21 were tested, with 21x21 found to produce better results. Convolutional kernel sizes of 3 x 3 and 5 x 5 were tested, with the 5 x 5 kernel size being selected for the final results. Filter sizes for the four convolutional layers were optimal at 200, 200, 600, and 600, while the fully connected layers contained 256 and 128 neurons after testing various combinations. Drop-out rates from 0.5 to 1.0 were investigated, with 1.0 having the best results. The optimal learning rate, AdaDelta parameter $\rho$, and AdaDelta parameter $\varepsilon$ were found to be 0.001, 0.98, and 1.0 x $10^8$, respectively. Once the optimal parameters were determined, they were used for the remaining scenarios.

Neural network segmentation classification results are shown in Table 1 for Scenario 1, which performed the leave-one-out cross-validation using the data of four healthy rat hearts, i.e.,the data of three hearts for training and the data of one additional heart for testing. Accuracies for each segmentation class, the overall accuracy, and accuracy of the chambers, are shown, as well as the epoch number which produced the best accuracy for each heart. Averaged results from all four hearts are shown at the bottom of the table. The overall average accuracy was found to be 72.4% ± 6.6%, with an average chamber cavity accuracy of 70.6% ± 13.1%. The overall accuracy for the background patches was relatively high at 78.8% ± 4.6%, with the value for each heart over 70%. RVC patches had the lowest average accuracy at 67.3% ± 21.1% followed by the Myo patches at 69.6% ± 8.7%. This was due to the poor performance of healthy heart #1 in these two categories.

Table 1. Results from the segmentation of rat hearts using a CNN when the diseased heart is excluded from model training and testing.

| Verification Dataset | Best Epoch | Accuracy (%) | | | | | |
|---|---|---|---|---|---|---|---|
| | | BG | Myo | LVC | RVC | Chambers | Overall |
| Healthy #1 | 2 | 85.4 | 59.3 | 69.9 | 32.9 | 51.4 | 61.9 |
| Healthy #2 | 7 | 75.3 | 71.3 | 75.5 | 85.2 | 80.4 | 76.8 |
| Healthy #3 | 9 | 80.8 | 65.2 | 85.8 | 83.9 | 84.9 | 78.9 |
| Healthy #4 | 11 | 73.7 | 82.7 | 63.9 | 67.5 | 65.7 | 72.0 |
| **Average** | - | **78.8 ± 4.6** | **69.6 ± 8.7** | **73.8 ± 8.1** | **67.3 ± 21.1** | **70.6 ± 13.1** | **72.4 ± 6.6** |

**BG**: Background, **Myo**: Myocardium, **LVC**: Left Ventricle Cavity, **RVC**: Right Ventricle Cavity, **Chambers**: Combined LVC and RVC accuracies

Results from Scenarios 2 and 3 are shown in Table 2. The epoch with the best result for each heart was chosen and the various accuracies for that epoch are shown in the table, with the averages displayed at the bottom of the table. Scenario 2 had an overall accuracy of 57.2%, mostly due to poor performance on RVC patches. Accuracies for BG and LVC patches were high, with values of 80.6% and 86.2%, respectively. Scenario 3 had an average overall accuracy of 73.2% ± 4.0%, which was the highest average overall accuracy of all three scenarios investigated. Compared to Scenario 1 in Table 1, the RVC accuracy for the healthy heart #1 was much higher in Scenario 3, with an accuracy of 93.3%. However, Myo accuracy decreased by 13.9% to 45.4% for the same heart.

## 4. DISCUSSION

When trained on three healthy hearts and tested on a fourth (Scenario 1, Table 1), the average background patch accuracy was higher compared to the average chamber cavity and myocardium accuracies. The overall accuracy for this scenario was adversely affected by poor results from healthy heart #1, which had low accuracies for the right ventricle cavity and myocardium. This also contributed to the large standard deviations for the average RVC and chambers accuracies. This could indicate the presence of features unique to the healthy heart #1 dataset in the Myo and RVC patches.

The accuracy when testing on the diseased heart after training on four healthy hearts (Scenario 2) was poor, due almost exclusively to the lack of training on RVC patches. The diseased heart performed well on the LVC and BG patches, but struggled with the Myo patches and failed completely with the RVC. This is due to the absence of RVC training samples which depict such a large chamber. On a healthy rat, the RVC is a thin crescent compared to that of the diseased heart (Figure 4). The close proximity of the right ventricle walls in the healthy hearts provides useful classification features for the CNN. This was not the case in the diseased heart, causing the RVC to often be misclassified as the LVC or BG.

Table 2. Results from the segmentation of rat hearts using a CNN
when 4 healthy and 1 diseased heart are used for leave-one-out cross-validation.

| Verification Dataset | Best Epoch | Accuracy (%) | | | | | |
|---|---|---|---|---|---|---|---|
| | | BG | Myo | LVC | RVC | Chambers | Overall |
| Diseased | 4 | 80.6 | 61.8 | 86.2 | 0.0 | 43.1 | 57.2 |
| Healthy #1 | 3 | 78.1 | 45.4 | 73.1 | 93.3 | 72.5 | 72.5 |
| Healthy #2 | 10 | 76.5 | 79.1 | 67.3 | 76.4 | 74.8 | 74.8 |
| Healthy #3 | 6 | 76.3 | 58.1 | 89.9 | 88.4 | 78.8 | 78.2 |
| Healthy #4 | 3 | 71.1 | 74.0 | 56.9 | 67.3 | 66.1 | 67.3 |
| **Average: All** | - | **76.5 ± 3.1** | **63.7 ± 11.9** | **74.7 ± 12.1** | **65.1 ± 33.8** | **69.9 ± 16.3** | **70.0 ± 7.3** |
| **Average: Healthy** | - | **75.5 ± 2.6** | **64.2 ± 13.3** | **71.8 ± 12.0** | **81.4 ± 10.2** | **76.6 ± 10.4** | **73.2 ± 4.0** |

**BG**: Background, **Myo**: Myocardium, **LVC**: Left Ventricle Cavity, **RVC**: Right Ventricle Cavity, **Chambers**: Combined LVC and RVC accuracies

Training on the diseased heart and three healthy hearts (Scenario 3) produced the highest average overall accuracy of all three scenarios considered. However, while some accuracies did improve after including the diseased heart, others deteriorated when compared with Scenario 1. This suggests that the inclusion of the diseased heart only in the training data might provide improvements just by increasing the number of training samples and patches, even if some of the patches correspond to diseased tissue.

There were several limitations for the presented work. First, only a small number of rats were used, which can contribute to fewer useful features in the CNN. A training set with patches from more hearts could improve the presented results. Second, the hearts were only roughly aligned relative to one another. A more rigorous, deformable registration could improve the similarity between the samples, although the importance of this would decrease as the number of training samples increased. Finally, a patch size of 21 x 21 was chosen for the work, but a larger patch size could improve the accuracy by providing more potential features for the CNN to extract. This is especially important for the LVC and the BG, where areas near the myocardium would be similar. A larger patch size would enable the CNN to use more of the myocardium shape and location to determine the correct patch label for segmentation.

## 5. Conclusion

In this work, we proposed and implemented a convolutional neural network to segment the heart on ultrasound images.. Modest accuracies were obtained when a CNN model was trained on limited number of dataSlight improvements in overall average accuracies were noted when a diseased heart was included in the training data. This showed that the model was able to compensate for the inclusion of diseased heart samples in the model when testing on a healthy heart. More data are needed in order to improve the accuracy of the CNN-based segmentation approach.

## ACKNOWLEDGMENTS

This research is supported in part by NIH grants (CA176684, CA156775, and CA204254) and a pilot grant from the Children's Heart Research and Outcomes Center of Children's Healthcare of Atlanta. The content is solely the responsibility of the authors and does not necessarily reflect the official views of the National Institutes of Health.

## REFERENCES

[1] Benjamin, E. J., Blaha, M. J., Chiuve, S. E., Cushman, M., Das, S. R., Deo, R., de Ferranti, S. D., Floyd, J., Fornage, M., Gillespie, C., Isasi, C. R., Jiménez, M. C., Jordan, L. C., Judd, S. E., Lackland, D., Lichtman, J. H., Lisabeth, L., Liu, S., Longenecker, C. T., Mackey, R. H., Matsushita, K., Mozaffarian, D., Mussolino, M. E., Nasir, K., Neumar, R. W., Palaniappan, L., Pandey, D. K., Thiagarajan, R. R., Reeves, M. J., Ritchey, M., Rodriguez, C. J., Roth, G. A., Rosamond, W. D., Sasson, C., Towfighi, A., Tsao, C. W., Turner, M. B., Virani, S. S., Voeks, J. H., Willey, J. Z., Wilkins, J. T., Wu, J. H. Y., Alger, H. M., Wong, S. S., and Muntner, P., "Heart Disease and Stroke Statistics—2017 Update: A Report From the American Heart Association," Circulation, 135(10), e146 (2017).

[2] Pearlman, A. S., Ryan, T., Picard, M. H., and Douglas, P. S., "Evolving trends in the use of echocardiography: a study of Medicare beneficiaries," J Am Coll Cardiol, 49(23), 2283-91 (2007).

[3] Fenster, A., Downey, D. B., and Cardinal, H. N., "Three-dimensional ultrasound imaging," Physics in Medicine & Biology, 46(5), R67 (2001).

[4] Walsh, J. A., Topol, E. J., and Steinhubl, S. R., "Novel Wireless Devices for Cardiac Monitoring," Circulation, 130(7), 573 (2014).

[5] Prabhu, S. J., Kanal, K., Bhargava, P., Vaidya, S., and Dighe, M. K., "Ultrasound Artifacts: Classification, Applied Physics With Illustrations, and Imaging Appearances," Ultrasound Quarterly, 30(2), 145-157 (2014).

[6] Pustavoitau, A., Blaivas, M., Brown, S. M., Gutierrez, C., Kirkpatrick, A. W., Kohl, B. A., Oren-Grinberg, A., and Frankel, H. L., "Recommendations for Achieving and Maintaining Competence and Credentialing in Critical Care Ultrasound with Focused Cardiac Ultrasound and Advanced Critical Care Echocardiography," Documents/Critical% 20care% 20Ultrasound. pdf> Accessed Oct, 27, (2016).

[7] Sonka, M., Zhang, X., Siebes, M., Bissing, M. S., DeJong, S. C., Collins, S. M., and McKay, C. R., "Segmentation of intravascular ultrasound images: A knowledge-based approach," IEEE Transactions on Medical Imaging, 14(4), 719-732 (1995).

[8] Xiao, G., Brady, M., Noble, J. A., and Zhang, Y., "Segmentation of ultrasound B-mode images with intensity inhomogeneity correction," IEEE Transactions on Medical Imaging, 21(1), 48-57 (2002).

[9] Hansson, M., Brandt, S. S., Lindstrom, J., Gudmundsson, P., Jujic, A., Malmgren, A., and Cheng, Y. J., "Segmentation of B-mode cardiac ultrasound data by Bayesian Probability Maps," Medical Image Analysis, 18(7), 1184-1199 (2014).

[10] Qin, X., Cong, Z., and Fei, B., "Automatic segmentation of right ventricular ultrasound images using sparse matrix transform and level set," Physics in Medicine and Biology 58(21), 7609-24 (2013).

[11] Dormer, J., Qin, X., Shen, M., Wang, S., Zhang, X., Jiang, R., Wagner, M. B., and Fei, B., "Determining cardiac fiber orientation using FSL and registered ultrasound/DTI volumes." 979015-979015-7.

[12] Qin, X. L., and Fei, B. W., "DTI template-based estimation of cardiac fiber orientations from 3D ultrasound," Medical Physics, 42(6), 2915-2924 (2015).

[13] Dormer, J. D., Meng, Y. G., Zhang, X. D., Jiang, R., Wagner, M. B., and Fei, B. W., "Estimating cardiac fiber orientations in pig hearts using registered ultrasound and MR image volumes," Medical Imaging 2017: Ultrasonic Imaging and Tomography, 10139, (2017).

[14] Ronneberger, O., Fischer, P., and Brox, T., [U-Net: Convolutional Networks for Biomedical Image Segmentation] Springer International Publishing, Cham(2015).

[15] Zhang, W., Li, R., Deng, H., Wang, L., Lin, W., Ji, S., and Shen, D., "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," NeuroImage, 108(Supplement C), 214-224 (2015).

[16] Moeskops, P., Viergever, M. A., Mendrik, A. M., Vries, L. S. d., Benders, M. J. N. L., and Išgum, I., "Automatic Segmentation of MR Brain Images With a Convolutional Neural Network," IEEE Transactions on Medical Imaging, 35(5), 1252-1261 (2016).
[17] Davlouros, P. A., Niwa, K., Webb, G., and Gatzoulis, M. A., "The right ventricle in congenital heart disease," Heart, 92(Suppl 1), i27-i38 (2006).
[18] Hirata, M., Ousaka, D., Arai, S., Okuyama, M., Tarui, S., Kobayashi, J., Kasahara, S., and Sano, S., "Novel Model of Pulmonary Artery Banding Leading to Right Heart Failure in Rats," BioMed Research International, 2015, 10 (2015).
[19] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X., "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," arXiv preprint arXiv:1603.04467, (2015).
[20] Zeiler, M. D., "ADADELTA: an adaptive learning rate method," arXiv preprint arXiv:1212.5701, (2012).