Hindawi Scientific Programming Volume 2020, Article ID 7025403, 6 pages https://doi.org/10.1155/2020/7025403



# Research Article

# Deep RetinaNet for Dynamic Left Ventricle Detection in Multiview Echocardiography Classification

Meijun Yang,<sup>1</sup> Xiaoyan Xiao,<sup>2</sup> Zhi Liu, <sup>1,3</sup> Longkun Sun,<sup>4</sup> Wei Guo,<sup>3</sup> Lizhen Cui,<sup>3</sup> Dianmin Sun,<sup>5</sup> Pengfei Zhang, <sup>4</sup> and Guang Yang,<sup>6</sup>

Correspondence should be addressed to Zhi Liu; liuzhi@sdu.edu.cn and Pengfei Zhang; pengf-zhang@163.com

Received 18 February 2020; Revised 31 May 2020; Accepted 29 June 2020; Published 1 August 2020

Academic Editor: Chenxi Huang

Copyright © 2020 Meijun Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. Currently, echocardiography has become an essential technology for the diagnosis of cardiovascular diseases. Accurate classification of apical two-chamber (A2C), apical three-chamber (A3C), and apical four-chamber (A4C) views and the precise detection of the left ventricle can significantly reduce the workload of clinicians and improve the reproducibility of left ventricle segmentation. In addition, left ventricle detection is significant for the three-dimensional reconstruction of the heart chambers. Method. RetinaNet is a one-stage object detection algorithm that can achieve high accuracy and efficiency at the same time. RetinaNet is mainly composed of the residual network (ResNet), the feature pyramid network (FPN), and two fully convolutional networks (FCNs); one FCN is for the classification task, and the other is for the border regression task. Results. In this paper, we use the classification subnetwork to classify A2C, A3C, and A4C images and use the regression subnetworks to detect the left ventricle simultaneously. We display not only the position of the left ventricle on the test image but also the view category on the image, which will facilitate the diagnosis. We used the mean intersection-over-union (mIOU) as an index to measure the performance of left ventricle detection and the accuracy as an index to measure the effect of the classification of the three different views. Our study shows that both classification and detection effects are noteworthy. The classification accuracy rates of A2C, A3C, and A4C are 1.000, 0.935, and 0.989, respectively. The mIOU values of A2C, A3C, and A4C are 0.858, 0.794, and 0.838, respectively.

#### 1. Introduction

Heart disease is a common circulatory disease that not only seriously affects the function of the cardiovascular system but also causes certain damage to the respiratory system [1]. In severe cases, heart failure can endanger life [2]. Therefore, it is important to use more advanced methods to observe cardiac symptoms or exercise [3]. In the clinical diagnosis of heart disease, echocardiography is the most commonly used tool [4]. Echocardiography is a noninvasive technique for

examining the anatomy and functional status of the heart and large blood vessels using ultrasound [5]. It uses pulsed ultrasound to measure the periodic activities of the underlying walls, ventricles, and valves through the chest wall and soft tissues [6].

The left ventricle is the focal part of the heart, and the symptoms of the left ventricle are an important basis for the diagnosis of heart disease [7]. Therefore, accurate information regarding the left ventricle extracted from echocardiography is crucial for further clinical procedures and

<sup>&</sup>lt;sup>1</sup>School of Information Science and Engineering, Shandong University, Qingdao 266237, China

<sup>&</sup>lt;sup>2</sup>Department of Nephrology, Qilu Hospital, Cheeloo College of Medicine, Shandong University, No. 107 Wenhuaxi Road, Jinan 250012, China

<sup>&</sup>lt;sup>3</sup>Joint SDU-NTU Centre for Artificial Intelligence Research (C-FAIR), Shandong University, Jinan 250101, China

<sup>&</sup>lt;sup>4</sup>Department of Cardiology, Qilu Hospital, Cheeloo College of Medicine, Shandong University, No. 107 Wenhuaxi Road, Jinan, Shandong Province, China

<sup>&</sup>lt;sup>5</sup>Department of Thoracic Surgery, Shandong Cancer Hospital and Institute,

Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan 250117, China

<sup>&</sup>lt;sup>6</sup>National Heart and Lung Institute, Imperial College London, SW7 2AZ, London, UK

prognosis. To extract the information of the left ventricle, the first step is to accurately detect its position. Echocardiography contains multiple views, and each view contains multiple anatomical parts. For example, the apical two-chamber (A2C), apical three-chamber (A3C), and apical four-chamber (A4C) views of the echocardiography all contain the left ventricle, left atrium, or more anatomical parts. However, the morphology of the left ventricle is different in different views, so we need to classify these three views and detect the left ventricle. This detection will greatly reduce the time it takes for doctors to find useful information from numerous echocardiographic information.

Processing medical images through deep learning has become a popular topic. Based on the development of convolutional neural networks (CNNs), a variety of object detection algorithms have been proposed [8]. At present, the deep learning methods in the object detection field can be roughly divided into two categories: two-stage algorithms and one-stage algorithms. For example, two-stage algorithms include Region-CNN (R-CNN), Fast R-CNN, and Spatial Pyramid Pooling Convolutional Networks (SPP-Net). One-stage algorithms mainly include You Only Look Once (YOLO), Single Shot Multibox Detector (SSD), and other methods [9]. In general, two-stage algorithms are superior in detection accuracy and positioning accuracy, and one-stage algorithms are more efficient. To the best of our knowledge, there are no previous studies considering the use of deep learning-based object detection to identify the type of views and location of the left ventricle. RetinaNet [10] is a one-stage detector that can achieve the accuracy of a twostage detector without affecting the algorithm efficiency. The innovation of RetinaNet is the introduction of focal loss, which overcomes the class imbalance problem of object detection. In this paper, to identify the A2C, A3C, and A4C views and locate the left ventricle accurately, we developed RetinaNet for multiview echocardiography. It is significant for the three-dimensional reconstruction of the heart chambers [11]. In addition, extracting the located left ventricle and then segmenting the left ventricle will improve the segmentation accuracy.

Currently, there are many classic object detection algorithms. Girshick et al. proposed the R-CNN [12], which is a milestone of deep learning-based object detection. Instead of selecting a sliding window to traverse over a picture, it used the strategy of selecting region candidate boxes, which might contain the objects to be detected. The R-CNN architecture consisted of 5 convolutional layers and 2 fully connected layers. In addition to the R-CNN, SPP-Net [13], Fast R-CNN [14], Faster R-CNN [15], R-FCN [16], and Mask R-CNN [17] improved the basis of the R-CNN to achieve better performance. The YOLO [18] algorithm proposed by Redom et al. in 2015 indicated that the object detection algorithm can be roughly divided into two categories: twostage algorithms and one-stage algorithms. The major difference between the YOLO algorithm and the two-stage algorithms represented by the R-CNN series is that the YOLO algorithm discarded the candidate box extraction branch. The YOLO algorithm directly performed feature extraction, candidate frame regression, and classification in

the same branchless convolution network, which made the network structure simple, and the detection speed was nearly 10 times faster than that of the Faster R-CNN. YOLOv2 [19], YOLOv3 [20], SSD [21], and RetinaNet are also one-stage algorithms.

There have been many studies to classify, detect, and segment cardiac images. Luo et al. [22] applied an eight-layer convolutional neural network to detect and locate the location of interest, which is a bounding box of the short-axis MRI image containing the left and right ventricles. This approach yielded better right ventricle segmentation. Vigneault et al. [23] designed so-called  $\Omega$ -Net for multiview cardiac MR detection, orientation, and segmentation. Li et al. [24] applied an 11-layer convolutional neural network to automatically detect the bounding box of the myocardium from the myocardial echocardiography (MCE) images. Nizar et al. [25] proposed the use of a machine learning technique that automatically detects and localizes the aortic valve in echocardiography images. The detection used AlexNet, and the detection accuracy was 99.87% for the aortic valve.

Although previous studies have obtained promising results, these studies have only applied several layers of simple convolutional neural networks to detect a certain part of the heart and then performed segmentation based on this detection. The segmentation may not be still accurate or reproducible. Through applying the classic detection algorithms mentioned above, we can achieve more functions that have significant research value in the field of medical image analysis.

Khamis et al. [26] introduced a multistage classification algorithm that employed spatiotemporal feature extraction (cuboid detector) and supervised dictionary learning (LC-KSVD) approaches to classify the A2C, A4C, and apical long-axis (ALX) views. The recognition accuracies achieved were 97%, 91%, and 97% for the A2C, A4C, and ALX views, respectively. Madani et al. [27] trained a convolutional neural network to classify 15 standard views, and the accuracy achieved was 91.7%. Although the classification results shown in these articles were satisfactory, these studies simply identified types of views that were relatively simple and therefore have less clinical impact.

By analyzing the previous studies, we have found that the classification of multiview echocardiography has been studied, and the detection of a certain part of the heart has also been investigated. However, there is no previous study that directly detected the left ventricle with multiview classification. In this study, we developed a RetinaNet-based method for identifying A2C, A3C, and A4C views and detecting the left ventricle simultaneously from multiview echocardiography images. When a patient's echocardiographic image is put through our network, the network will automatically recognize the view types and detect the left ventricle. This detection will greatly reduce the time it takes for doctors to find useful information from numerous echocardiographic information. More importantly, this work is of great significance for the three-dimensional reconstruction of the heart chamber and left ventricular segmentation.

#### 2. Materials and Methods

2.1. Datasets and Clinical Background. Our echocardiography images were collected from two hospitals in China, i.e., Shandong Qilu Hospital and Shengjing Hospital, with different devices by Philips and GE. The temporal rate is 65-70 Hz among frames. We extracted the A4C images of seven patients, with a total of 1238 images. We also extracted the A2C images of seven patients, with a total of 1011 images. We collected the A3C images of nine patients, with a total of 404 images. Each patient's image set contains at least one temporally cropped sequence that captures one complete cardiac cycle from end-systole (ES) to end-diastole (ED). In clinical practice, A4C is one of the main standard views of cardiac function analysis. However, when encountering a complex condition, the clinician needs to analyze heart function from multiple views, so A2C and A3C views are also necessary for research. To train and validate the results of the model, we asked a professional radiologist who is experienced in echocardiography to help us mark the correct position of the left ventricle on the A2C, A3C, and A4C images. Example datasets and left ventricle annotations are shown in Figure 1.

To have sufficient datasets and avoid overfitting, we performed data augmentation operations on the raw datasets. The data augmentation we used mainly includes the following three procedures:

- (1) Random flip: flips the input images and the corresponding boxes with a probability of 0.5
- (2) Random crop: crops the given images to a random size and aspect ratio
- (3) Resize: resizes the input images to the given size

2.2. Network Architecture. RetinaNet mainly includes three subnetworks: a residual network (ResNet) [28], a feature pyramid network (FPN) [29], and two fully convolutional networks (FCNs) [30]. The RetinaNet network architecture is summarized in Figure 2.

ResNet: the main contribution of ResNet was the idea of residual learning, which allows the original input information to be directly transmitted to the following layer [31]. ResNet can use different network layers. The commonly used types of network layers are 50-layer, 101-layer, and 152-layer. We chose the 101-layer architecture with the best training performance [32]. We extracted the features of the echocardiography using ResNet and then put them forward to the next subnetwork.

FPN: FPN is a method for efficiently extracting the features of each dimension in a picture using a conventional CNN model. First, we used a single-dimensional image as the input to ResNet. Then, starting from the second layer of the convolutional network, the features of all the layers were selected by the FPN and then combined to form the final feature output combination.

FCN: the class subnet in the FCN performed the classification task. This subnet could identify which view the echocardiography image belongs to. The box subnet in the FCN performed the border regression task. Its role was to detect the position of the left ventricle in the echocardiography images and record the coordinates.

Focal loss: focal loss is an improved version of the cross-entropy loss, and the binary cross-entropy expression is as follows:

$$CE(p, y) = \begin{cases} -\log(p), & \text{if } y = 1, \\ -\log(1 - p), & \text{otherwise,} \end{cases}$$
 (1)

where  $y \in [\pm, 1]$  is the ground truth category and  $p \in [0, 1]$  is the predicted probability of the model for category y = 1.

$$p_t = \begin{cases} p, & \text{if } y = 1, \\ 1 - p, & \text{otherwise.} \end{cases}$$
 (2)

The above formula can be abbreviated as

$$CE(p, y) = CE(p_t) = -\log(p_t).$$
 (3)

To solve the problem of the data imbalance between the positive and negative samples, the original form is changed into the following form:

$$CE(p_t) = -\alpha_t \log(p_t), \tag{4}$$

among them,

$$\alpha_t = \begin{cases} \alpha, & \text{if} \quad y = 1, \\ 1 - \alpha, & \text{otherwise,} \end{cases}$$
 (5)

where  $\alpha \in [0, 1]$  is the weight factor. To solve the problem of the difficult sample, the focusing parameter  $\gamma$  is introduced to obtain the final form of the focal loss:

$$FL(p_t) = -\alpha_t (1 - p_t)^{\gamma} \log(p_t). \tag{6}$$

### 3. Results

We have a total of 1238 images of the A4C view, 1011 images of the A2C view, and 404 images of the A3C view. For each view, we divided the images into a training set, a validation set, and an independent testing set using the ratio of 7:1:2.

We used a Dell Inspiron 3670 workstation, which had an Intel Core i7-8700 CPU @ 3.20 GHz, 8 GB of memory, and an NVIDA GeForce 1050 Ti GPU for training and testing our developed model. Our work was implemented using the PyTorch platform.

The images are preprocessed so that only the cardiac images are selected and resized to 512 by 512 when put into the model. The gray value is normalized to [0, 1] by dividing by 256. We extract the features of the echocardiography using ResNet and then put them forward to the FPN to form the final features. The classifier subnet in the FCN identifies which view the echocardiography image belongs to. The box subnet in the FCN detects the position of the left ventricle in the echocardiography images and records the coordinates.

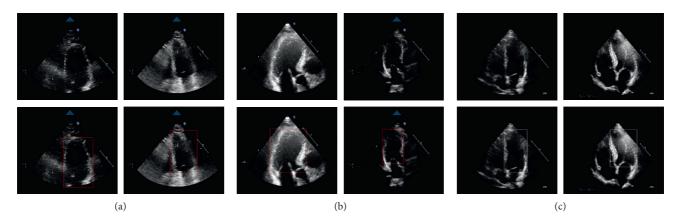


FIGURE 1: The first row shows the original echocardiography images, and the second row shows the images with left ventricle region annotations. (a) A2C images. (b) A3C images. (c) A4C images.

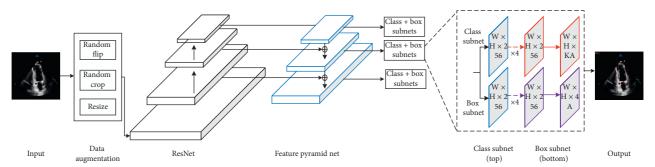


FIGURE 2: The RetinaNet network architecture consists of a ResNet, a FPN, and two FCN subnets.

TABLE 1: Results of the classification and mIOU of the three views.

	A2C	A3C	A4C
Accuracy	1.000	0.935	0.989
mIOU	0.858	0.794	0.838

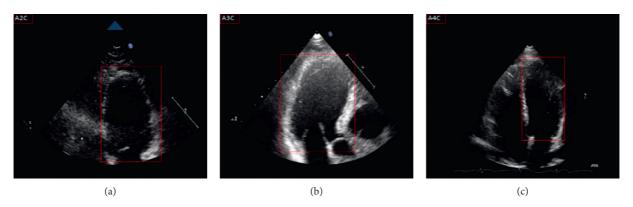


FIGURE 3: We input test images into our trained RetinaNet for independent testing and automatically identified A2C, A3C, and A4C images and detected the localization of the left ventricle. The type of view the network predicts is shown in the upper left corner with white text. The red rectangle in the middle represents the left ventricle position detected by our model. Left ventricle detection results in (a) A2C, (b) A3C, and (c) A4C.

The kernel size is 3, and the batch size is 16. We use the ReLU activation in all constitutional layers of the deep networks and the sigmoid function in the prediction layers. We use the SGD optimizer with a learning rate of 1e-3. In addition to displaying the position of the left ventricle on the test image, we also display the category of the view in the upper left corner of the test image, which will facilitate diagnosis.

We use the accuracy and the mean intersection-overunion (mIOU) to evaluate the performance of our model. The accuracy was defined as the ratio of the number of correctly predicted samples to the total number of predicted samples. In our experiments, IOU was expressed as the overlap of the candidate box generated by the network with the ground truth box, that is, the ratio of its intersections to the unions. The higher the correlation, the larger the value. The ideal situation is a complete overlap, which is indicated by a ratio of 1. The mIOU and accuracy results are shown in Table 1. The types of views we predicted through our model and the detected left ventricles are shown in Figure 3.

#### 4. Discussion

There are still a few improvements that can be applied to our work. For example, there are limited image samples in our dataset. In the next phase of the study, we will extract the detected left ventricle and, on this basis, try to achieve better left ventricle segmentation using an unsupervised method. At the same time, we will continue to study the application of this work in the three-dimensional reconstruction of heart chambers.

#### 5. Conclusion

Left ventricle detection from multiview echocardiography images can help clinicians diagnose heart disease more comprehensively and accurately and, more importantly, is of great significance for the three-dimensional reconstruction of heart chambers. In addition, it can improve the accuracy of left ventricle segmentation. In this study, we propose to simultaneously use RetinaNet to identify A2C, A3C, and A4C images and detect the left ventricle in multiview echocardiography. The results have demonstrated that our proposed method can better classify the A2C, A3C, and A4C views and can also better detect the left ventricle from these views.

#### **Data Availability**

All the data used to support the findings of this study are available from the corresponding author upon request.

#### **Conflicts of Interest**

The authors declare that they have no conflicts of interest.

#### **Authors' Contributions**

Meijun Yang contributed to the experiments and writing of the paper. Xiaoyan Xiao provided the fundamental medical analysis for the experimental results. And Meijun Yang and Xiaoyan Xiao contributed equally to this paper. Zhi Liu and Pengfei Zhang organized this study. Longkun Sun and Dianmin Sun were in charge of data collection and building the dataset. Wei Guo and Lizhen Cui gave computing supports. Guang Yang contributed to writing.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant nos. 1192780063 and 91846205, the Key Research and Development Plan of Shandong Province under Grant nos. 2018YFJH0506 and 2019JZZY011007, and the Major Fundamental Research of the Natural Science Foundation of Shandong Province under Grant no. ZR2019ZD05.

#### References

- [1] A. Dewan and M. Sharma, "Prediction of heart disease using a hybrid technique in data mining classification," in *Proceedings of 2nd International Conference on Computing for Sustainable Global Development (INDIA.Com)*, IEEE, pp. 704–706, New Delhi, India, March, 2015.
- [2] C. Xu, L. Xu, Z. Gao et al., "Direct delineation of myocardial infarction without contrast agents using a joint motion feature learning architecture," *Medical Image Analysis*, vol. 50, pp. 82–94, 2018.
- [3] Z. Gao, H. Zhang, D. Wang et al., "Robust recovery of myocardial kinematics using dual *H* ∞ \$\mathbb{M} \text{\text{Nonterior}} \text{\text{Multimedia Tools and Applications}}, vol. 77, no. 17, pp. 23043–23071, 2018.
- [4] V. Apte, L. Tam, A. Han et al., "Evaluation of circumferential and longitudinal strain in a rabbit fetal heart model using 4d echocardiography," in *Proceedings of 39th Annual Northeast Bioengineering Conference*, pp. 23-24, IEEE, Syracuse, NY, USA, April 2013.
- [5] P. Joos, H. Liebgott, F. Varray et al., "High-frame-rate 3-D echocardiography based on motion compensation: an in vitro evaluation," in *Proceedings of IEEE International Ultrasonics Symposium (IUS)*, pp. 1–4, IEEE, Washington, DC, USA, September 2017.
- [6] R. Al Mukaddim, K. Samimi, A. Rodgers, T. A. Hacker, and T. Varghese, "Comparison of cardiac displacements in a murine model of myocardial ischemia using Cardiac Elastography and speckle tracking echocardiography," in *Pro*ceedings of IEEE International Ultrasonics Symposium (IUS), pp. 1–4, IEEE, Washington, DC, USA, September 2017.
- [7] A. Goyal, D. Bathla, P. Sharma, M. Sahay, and S. Shagun, "MRI image based patient specific computational model reconstruction of the left ventricle cavity and myocardium," in Proceedings of the International Conference on Computing, Communication and Automation (ICCCA), pp. 1065–1068, IEEE, Noida, India, April 2016.
- [8] M. Z. Alom, T. M. Taha, and C. Yakopcic, "The history began from alexnet: a comprehensive survey on deep learning approaches," 2018, https://arxiv.org/abs/1803.01164.
- [9] H. Zhang and X. Hong, "Recent progresses on object detection: a brief review," *Multimedia Tools and Applications*, vol. 78, no. 19, pp. 27809–27847, 2019.
- [10] T. Y. Lin, P. Goyal, R. Girshick, K. He, and D. Piotr, "Focal loss for dense object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2980–2988, IEEE, Venice, Italy, October 2017.

[11] L. Xu, X. Huang, J. Ma et al., "Value of three-dimensional strain parameters for predicting left ventricular remodeling after ST-elevation myocardial infarction," *The International Journal of Cardiovascular Imaging*, vol. 33, no. 5, pp. 663–673, 2017.

- [12] R. Girshick, J. Donahue, and T. Darrell, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, IEEE, Columbus, OH, USA, June 2014.
- [13] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.
- [14] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1440–1448, IEEE, Santiago, Chile, December 2015.
- [15] S. Ren, K. He, R. Girshick, and J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal networks: Advances in Neural Information Processing Systems, Neural Information Processing Systems (NIPS), Montreal, Quebec, Canada, 2015.
- [16] J. Dai, Y. Li, K. He, and J. Sun, Object Detection via Region-Based Fully Convolutional Networks: Advances in Neural Information Processing Systems, Neural Information Processing Systems (NIPS), Barcelona, Spain, 2016.
- [17] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proceedings of the IEEE International Conference* on Computer Vision, pp. 2961–2969, IEEE, Venice, Italy, October 2017.
- [18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings* of the IEEE conference on computer vision and pattern recognition, pp. 779–788, IEEE, Las Vegas, NV, USA, June 2016.
- [19] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7263–7271, IEEE, Honolulu, HI, USA, July 2017.
- [20] J. Redmon and A. Farhadi, "Yolov3: an incremental improvement," 2018, https://arxiv.org/abs/1804.02767.
- [21] W. Liu, D. Anguelov, D. Erhan et al., "SSD: single shot multibox detector," in *European Conference on Computer Vision*, pp. 21–37, Springer, Cham, Switzerland, 2016.
- [22] G. Luo, R. An, K. Wang, S. Dong, and D. Zhang, "A deep learning network for right ventricle segmentation in shortaxis MRI," in *Proceedings of Computing in Cardiology Con*ference (CinC), pp. 485–488, IEEE, Vancouver, BC, Canada, September 2016.
- [23] D. M. Vigneault, W. Xie, C. Y. Ho, D. A. Bluemke, and J. A. Noble, "Ω-net (omega-net): fully automatic, multi-view cardiac MR detection, orientation, and segmentation with deep neural networks," *Medical Image Analysis*, vol. 48, pp. 95–106, 2018.
- [24] Y. Li, C. P. Ho, M. Toulemonde, N. Chahal, R. Senior, and M.-X. Tang, "Fully automatic myocardial segmentation of contrast echocardiography sequence using random forests guided by shape model," *IEEE Transactions on Medical Im*aging, vol. 37, no. 5, pp. 1081–1091, 2017.
- [25] M. H. A. Nizar, A. Khalil, C. K. Chan, N. P. Utama, and K. W. Lai, "Pilot study on machine learning for aortic valve detection in echocardiography images," *Journal of Medical Imaging and Health Informatics*, vol. 9, no. 1, pp. 9–14, 2019.
- [26] H. Khamis, G. Zurakhov, V. Azar, A. Raz, Z. Friedman, and D. Adam, "Automatic apical view classification of

- echocardiograms using a discriminative learning dictionary," *Medical Image Analysis*, vol. 36, pp. 15–21, 2017.
- [27] A. Madani, R. Arnaout, M. Mofrad, and R. Arnaout, "Fast and accurate view classification of echocardiograms using deep learning," *NPJ Digital Medicine*, vol. 1, no. 1, pp. 1–8, 2018.
- [28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference* on Computer Vision and Pattern Recognition, pp. 770–778, IEEE, Las Vegas, NV, USA, June 2016.
- [29] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2117–2125, IEEE, Honolulu, HI, USA, July 2017.
- [30] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, IEEE, Boston, MA, USA, June 2015.
- [31] Z. Gao, Y. Li, Y. Sun et al., "Motion tracking of the carotid artery wall from ultrasound image sequences: a nonlinear state-space approach," *IEEE Transactions on Medical Imaging*, vol. 37, no. 1, pp. 273–283, 2017.
- [32] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "Automatic ship detection based on RetinaNet using multi-resolution gaofen-3 imagery," *Remote Sensing*, vol. 11, no. 5, p. 531, 2019.