Automatic quality assessment of apical four-chamber echocardiograms using deep convolutional neural networks

12 authors, including:

Amir Hossein Abdi
University of British Columbia - Vancouver
39 PUBLICATIONS   **271** CITATIONS

Christina Luong
University of British Columbia - Vancouver
61 PUBLICATIONS   **499** CITATIONS

Teresa S M Tsang
University of British Columbia - Vancouver
243 PUBLICATIONS   **14,107** CITATIONS

Saman Nouranian
University of British Columbia - Vancouver
22 PUBLICATIONS   **198** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project   Echo Quality Assessment View project

Project   Mayo Echo Book View project

Automatic quality assessment of apical four-chamber echocardiograms using deep convolutional neural networks

Amir H. Abdi[a], Christina Luong[b], Teresa Tsang[b], Gregory Allan[a], Saman Nouranian[a], John Jue[b], Dale Hawley[c], Sarah Fleming[b], Ken Gin[b], Jody Swift[b], Robert Rohling[a], and Purang Abolmaesumi[a]

[a]Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, Canada
[b]Cardiology Lab, Vancouver General Hospital, Vancouver, Canada
[c]Vancouver Coastal Health Authority, Vancouver, Canada

ABSTRACT

Echocardiography (echo) is the most common test for diagnosis and management of patients with cardiac conditions. While most medical imaging modalities benefit from a relatively automated procedure, this is not the case for echo and the quality of the final echo view depends on the competency and experience of the sonographer. It is not uncommon that the sonographer does not have adequate experience to adjust the transducer and acquire a high quality echo, which may further affect the clinical diagnosis. In this work, we aim to aid the operator during image acquisition by automatically assessing the quality of the echo and generating the Automatic Echo Score (AES). This quality assessment method is based on a deep convolutional neural network, trained in an end-to-end fashion on a large dataset of apical four-chamber (A4C) echo images. For this project, an expert cardiologist went through 2,904 A4C images obtained from independent studies and assessed their condition based on a 6-scale grading system. The scores assigned by the expert ranged from 0 to 5. The distribution of scores among the 6 levels were almost uniform. The network was then trained on 80% of the data (2,345 samples). The average absolute error of the trained model in calculating the AES was 0.87 ± 0.72. The computation time of the GPU implementation of the neural network was estimated at 5 ms per frame, which is sufficient for real-time deployment.

Keywords: Convolutional neural network, Deep learning, Quality assessment, Echocardiography, Apical four-chamber

1. INTRODUCTION

Heart failure is one of the primary causes of death worldwide, giving more value to the early detection of cardiac problems. Echocardiography is the most common diagnostic test used in management and follow-up of patients with suspected or known heart problems. It can provide the doctor with helpful information, including the size and shape of the heart, pumping capacity, and extent of tissue damages.[1, 2]

Echocardiograms are obtained from various planes or acoustic windows, called echo views, which visualize different heart structures. The standard echo views are categorized into four groups, parasternal, apical, suprasternal notch, and subcostal.[3] To acquire a good quality echo of a certain view, the transducer should be positioned so that its beam sections through certain cardiac structures. Echo acquisition is relatively a manual procedure and it is the sonographer's job to find the correct acoustic window. An echo with suboptimal quality may affect the accuracy of measurements and even result in the misdiagnosis and misclassification of the patient in terms of the final treatment.

There has been some efforts in helping the sonographer during image acquisition. Some studies have tried to alert the operator on presence of shadows and aperture obstructions in the echo window via analyzing the power
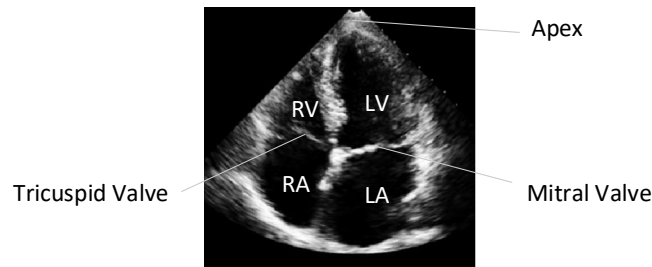

Figure 1. A typical apical four-chamber echocardiogram, depicting left ventricle (LV), right ventricle (RV), left atrium (LA), right atrium (RA) and the atrioventricular valves.

spectrum of the signal.[4, 5] However, these methods are blind to the anatomical structures on the echo image and cannot go beyond obstruction detection to determine the quality of a given echo.

Other methods aimed for the expected anatomical structures and evaluated the quality based on the goodness-of-fit of a predetermined template on the image.[6, 7] Due to the intrinsic nature of the echocardiography imaging, records from different patients may not follow a defined template. However, the mentioned methods rely solely on the low-level intensity-based features. Meaning, they do not capture the large range of variations present inside each echo view. Moreover, they are sensitive to the speckle noise, which is naturally present on echo images. Consequently, these template matching methods do not perform well in this domain.

To learn the complexities of an echo view, a model with a large learning capacity is essential. In the past few years, deep convolutional neural networks (CNNs) have become the state-of-the-art method for visual recognition tasks.[8, 9] CNNs extract hierarchical discriminative features and their capacity can be tuned through their depth and breadth.[10] Due to their sparse connections, CNNs have fewer weights and parameters to adjust compared to other deep architectures; thus, they are easier to train and less prone to over-fit on the training data. A typical CNN is structured in two parts. The first part extracts the features via convolutional layers, while reducing the spatial variance of feature maps via pooling layers. The second part uses the extracted features for discriminative or generative purposes.

In this research, a deep generative model is proposed to learn the appropriate features from a fairly large dataset of echo images. The trained model will then automatically evaluate the quality of a given echo frame, in real time. The experiments in this study only focus on the apical four-chamber view; but our approach is general and can be extended towards other views.

2. METHODS AND MATERIAL

In this study, a deep learning model is trained to calculate a quantitative metric of the image quality, the Automatic Echo Score (AES). The model is trained on a dataset of echo images, annotated by an expert cardiologist based on their quality.

2.1 Dataset and Manual Quality Assessment

For this project, echo views were fetched from the Vancouver General Hospital echo database with ethics approval from the Clinical Medical Research Ethics Board of the Vancouver Coastal Health (VCH) and consultation with the VCH Information Privacy Office. No new patients were scanned for this study. At VGH, these echo images are acquired mostly by echo-technicians.

For each patient, echo images are obtained from different standard views. This research only focuses on the A4C view, which is one of the most challenging views to obtain for novice sonographers (Fig. 1). The A4C view is mainly used to evaluate the cardiac chamber volumes and their contractility. The cross-longitudinal sections of both ventricles and atria, along with the mitral and tricuspid valves, and the heart apex are visualized in the A4C view.[11]

The end-systolic frame of each echo cine was extracted and then examined by an expert cardiologist. The clinician assigned an integer value of 0 to 5 to each image based on its quality. We ended up with 2,904 scored
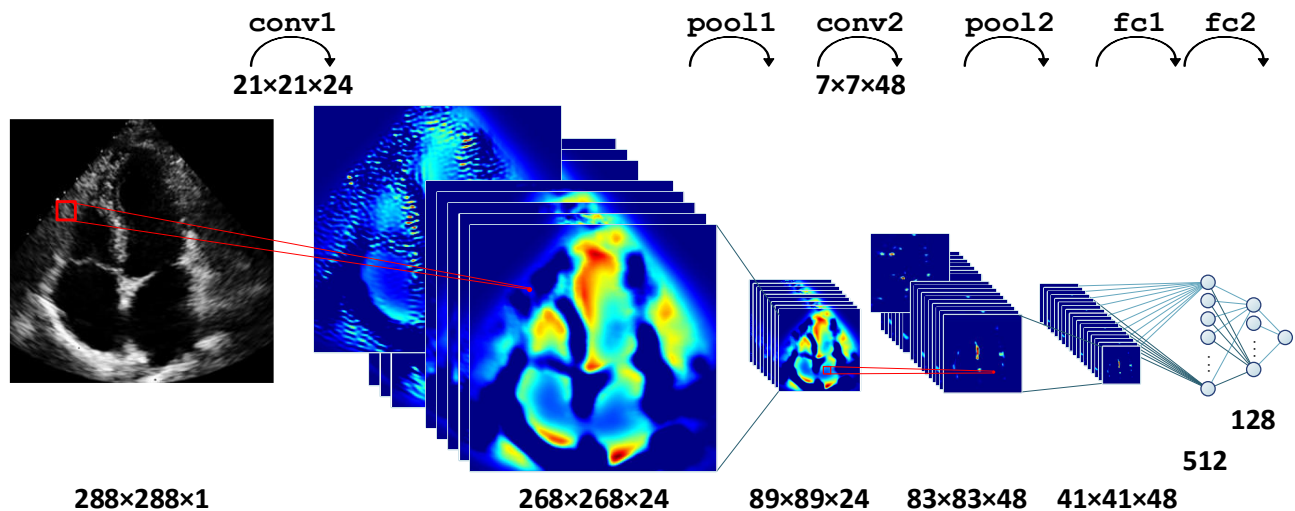
Figure 2. Network architecture, consisting of two convolutional layers (`conv1`, `conv2`), two max-pooling layers (`pool1`, `pool2`) and two fully connected layers (`fc1`, `fc2`). Size of data-layers (bottom) and convolutional kernels (top) are visible.

samples, obtained from independent studies, with almost uniform distribution across the six quality levels. The following criteria were considered when assigning the scores: 0) Clear presence of the aortic valve and/or only the interatrial or interventricular septums are visible; 1) Edges are clear enough to measure dimensions of one or two chambers; 2) Edges are clear enough to measure dimensions of three chambers; 3) Clear four chambers, acceptable edges but the image is off-axis (crooked) or significantly foreshortened; 4) Clear four chambers, acceptable edges, proper axis, mildly foreshortened; 5) Clear four chambers, proper axis, good edges, not foreshortened.

2.2 Network Architecture

The proposed regression model follows the standard deep convolutional neural network architecture, with adequate parameters to learn complex feature representations of echo images from thousands of training samples. The design is composed of convolutional layers (`conv`), max-pooling layers (`pool`) and fully-connected layers (`fc`).

The convolutional layers model the spatial correlation in the image. The max-pooling layers reduce the spatial variance of the convolutional features, encourage generalization and allow faster convergence and selection of superior invariant features.[12] The fully-connected layers calculate the inner-product of their input with their associated weight parameters.

The network architecture is illustrated briefly in Fig. 2. It consists of two convolutional layers (`conv1, conv2`), each followed by Rectified Linear Units (ReLUs)[13] and a pooling layer (`pool1, pool2`). The first convolutional layer has 24 kernels of size 21×21, and the second convolutional layer has 48 kernels of size 7×7. The first and second pooling layers subsample data using a window of size 5×5 with a stride of 3 and 3×3 with a stride of 2, respectively.

The network combines local features learned by the convolutional layers into a fewer number of signals using two fully-connected layers (`fc1, fc2`) consisting of 512 and 128 neurons. In the end, a single fully-connected neuron computes the final output of the network. The loss function of this regression model is the ℓ_2 norm of error which is the Euclidean distance of the network's output to the manual quality score (Section 2.1) assigned by the expert.

2.3 Training

The regression model was trained using stochastic gradient descent (SGD) accompanied by some techniques to facilitate learning. We used small batches of size 16, a momentum of 0.95, weight decay of 0.02, with initial learning rate of 0.0002 which gradually dropped by a factor of 0.95 every 1000 iterations. The parameters of the convolutional and fully-connected layers were initialized randomly from a zero-mean Gaussian distribution. To
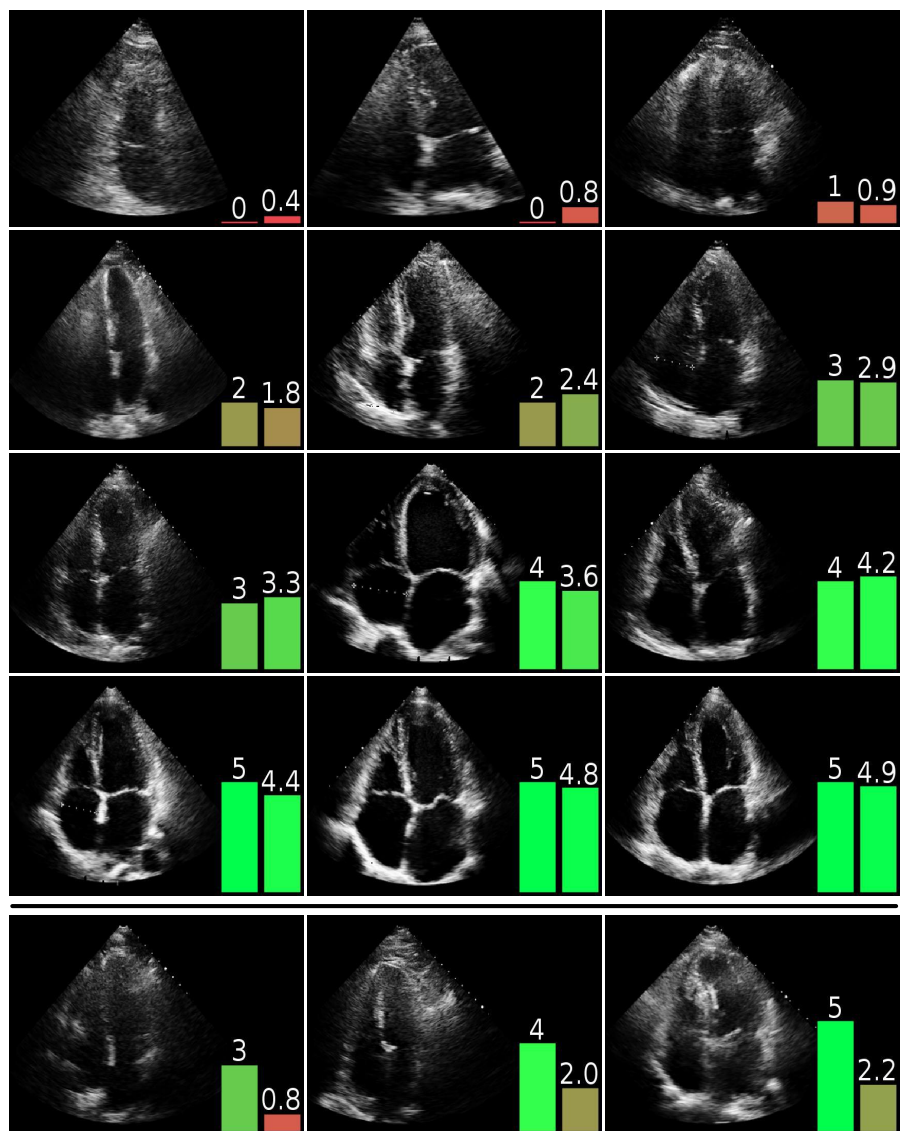
Figure 3. Some test samples along with their manually assigned quality scores (left bar) and DCNN-generated AES (right bar). The bottom row presents some outlier samples.

stabilize learning and prevent the model from overfitting on the training data, drop-out layers were also deployed to prevent co-adaptation of feature extractors and encourage neurons to follow the population behaviour.[14]

To add translational and rotational invariance during training, in each iteration, each sample was translated with a limited random number of pixels and rotated with 0, ±5 or ±10 degrees, comparable to augmenting the training set with 10,000 copies of each sample. This data augmentation technique also contributed to minimizing the possibility of overfitting. The expert cardiologist confirmed that a 10-degree rotation of the view does not affect image quality.

3. EXPERIMENTS AND RESULTS

The end-systolic A4C echo dataset was partitioned into 2,344 training-validation (80%) and 560 test samples (20%). Different network architectures and hyper-parameters were examined on the training-validation dataset, and the values which achieved the lowest absolute validation error was chosen as the final design (explained in Section 2.3). Ultimately, the final network was trained on the complete set of training-validation data and the
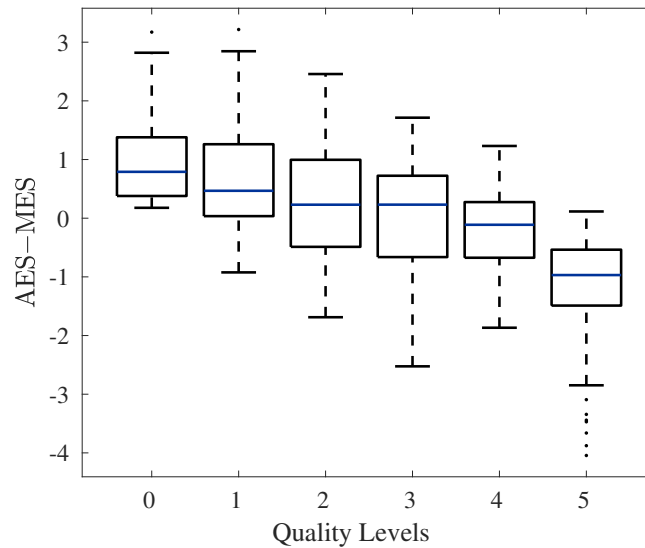
Figure 4. Distribution of error in each quality-level.

performance of the model was determined by the test set and reported as the accuracy of the proposed method. Performance of the trained model on the test data showed an average absolute error of 0.87 ± 0.72 against the manual scores. The distribution of error for each quality-level, calculated as $Error = AES - MES$, is depicted in the boxplot of Fig. 4. Some test samples along with their corresponding expert's score and their calculated AES are depicted in Fig. 3.

The designed network was deployed in the Caffe deep learning framework developed by the Berkeley Vision and Learning Center,[15] which has the advantage of an efficient GPU implementation. In our experiments on the Nvidia GeForce GTX 980 Ti GPU, the trained network calculated the AES for each frame of 288×288 pixels in 5 ms.

Although the network was only trained on the end-systolic frames, we ran an experiment to check the consistency of scores across the other frames of the echo cine. In this experiment, all the frames of an echo cine were sent to the same model which was only trained on the end-systolic frames. Since we did not have the expert's manual quality scores for these frames, we could not report the performance of our model across the cine. The deployed model, and its performance on some samples are presented in Video 1 (Fig. 5). In this experiment, we calculated the AES of the cine as the weighted average of the recent frames' AES while diminishing the weight by a factor of 0.9 for older frames.

4. DISCUSSION AND CONCLUSION

The accuracy of cardiac measurements on echocardiograms, which lead to correct diagnosis, depends on the quality of views obtained by the sonographers and can be influenced by various imaging factors. Suboptimal echo images can affect clinical measurements, which can result in misclassification of the patient in terms of needs for specific treatments. While experienced clinicians can find the correct acoustic window with no trouble, this is not always the case for less-experienced users. It has been reported that providing real-time feedback during image acquisition encourages sonographers to acquire better quality echo images.[6] Based on the above observation, a deep learning framework is proposed to automatically assess the quality of a given echo image and feedback the operator with a quality metric in real time. The GPU implementation of the model was capable of processing more than 100 frames of 288×288 pixels per second, which is faster than the native frame rate and suitable for real-time deployment. The average absolute error of the model was 0.87.

In the test set, there was only 40 samples, corresponding to 7 percent of the test data, for which the absolute error was more than 2 ($|AES - ManualScore| > 2$). The expert cardiologist examined these outliers and agreed that the AES score was more accurate for some, as she had only studied the end-systolic frames. She also
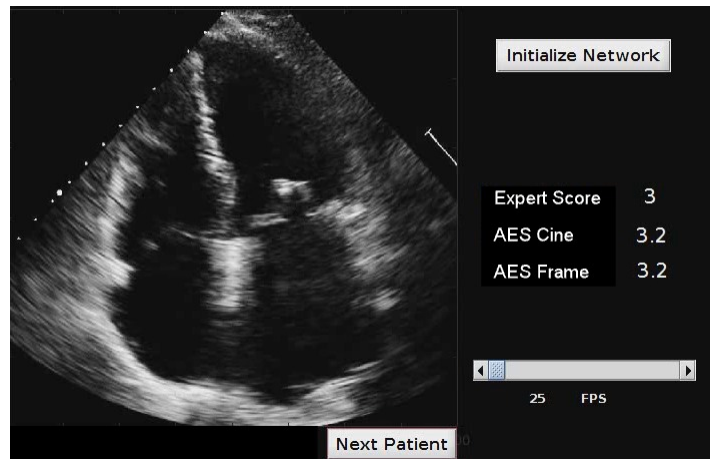
Figure 5. Video 1, Quality assessment of echo cine. Three values are displayed on the right: the Expert Score is the manual score assigned by the expert to the end-systolic frame of the cine. The AES Frame is the AES of the current frame calculated by the network. The AES Cine is the weighted average of the previous frames' AES.

http://dx.doi.org/10.1117/12.2254585.1

emphasized that she would have had assigned a different score if she had examined the whole cine instead of a single end-systolic frame.

Although we only experimented on the A4C view, the proposed deep learning approach makes no a priori assumptions on this specific echo view; nevertheless, it learns the appropriate hierarchical features directly from the training samples. As a result, the proposed method is not affected by the tissue-dependent speckle patterns of the ultrasound and filtering the training or test images to reduce the speckle noise did not increase the performance of our method.

Although the results for the proposed deep CNN model are promising, some challenges remain unmet. In future works, we are changing our manual scoring workflow by defining new scoring criteria and presenting the entire cine to the expert for scoring. Future steps include covering other standard echo views and extending the framework. Moreover, the model should be trained on all frames inside an echo cine, rather than a single end-systolic frame, while it is expected to detect and treat each frame differently. For the first step, we plan to use multi-domain deep architectures which share the weights among the initial layers and diverge on the final few layers.[16] With this architecture, the model learns distinctive features shared among all domains; thus, reduces the total training time and requires much less training samples per domain.

We believe, by integrating our system into the echo acquisition workflow and providing real-time feedback to the sonographers, we can increase the overall quality of the acquired echo images.

ACKNOWLEDGMENTS

REFERENCES

[1] Curtis, J. P., Sokol, S. I., Wang, Y., Rathore, S. S., Ko, D. T., Jadbabaie, F., Portnay, E. L., Marshalko, S. J., Radford, M. J., and Krumholz, H. M., "The association of left ventricular ejection fraction, mortality, and cause of death in stable outpatients with heart failure.," *Journal of the American College of Cardiology* **42**(4), 736–742 (2003).

[2] Ciampi, Q. and Villari, B., "Role of echocardiography in diagnosis and risk stratification in heart failure with left ventricular systolic dysfunction.," *Cardiovascular ultrasound* **5**, 34 (2007).

[3] Mann, D. L., Zipes, D. P., Libby, P., Bonow, R. O., and Braunwald, E., "Echocardiography," in [*Braunwald's heart disease : a textbook of cardiovascular medicine*], ch. 14, Elsevier Saunders (2015).

[4] Huang, S.-W., Radulescu, E., Wang, S., Thiele, K., Prater, D., Maxwell, D., Rafter, P., Dupuy, C., Drysdale, J., and Erkamp, R., "Detection and display of acoustic window for guiding and training cardiac ultrasound users," *Progress in Biomedical Optics and Imaging - Proceedings of SPIE* **9040**, 904014 (2014).

[5] Løvstakken, L., Ordernd, F., and Torp, H., "Real-time indication of acoustic window for phased-array transducers in ultrasound imaging," *Proceedings of IEEE Ultrasonics Symposium* , 1549–1552 (2007).

[6] Snare, S. R., Torp, H., Orderud, F., and Haugen, B. O., "Real-time scan assistant for echocardiography," *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control* **59**(3), 583–589 (2012).

[7] Pavani, S. K., Subramanian, N., Das Gupta, M., Annangi, P., Govind, S. C., and Young, B., "Quality metric for Parasternal Long AXis B-mode echocardiograms," *MICCAI 2015* **15**(Pt 2), 478–485 (2012).

[8] LeCun, Y., Bengio, Y., and Hinton, G., "Deep learning," *Nature* **521**, 436+ (sep 2015).

[9] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., and Wang, G., "Recent Advances in Convolutional Neural Networks," *arXiv* , 1–14 (2015).

[10] Krizhevsky, A., Sutskever, I., and Hinton, G. E., "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems* (2012).

[11] Solomon, S. D., [*Essential Echocardiography*], Humana Press Inc. (2007).

[12] Scherer, D., Müller, A., and Behnke, S., "Evaluation of pooling operations in convolutional architectures for object recognition," *Lecture Notes in Computer Science* **6354 LNCS**(PART 3), 92–101 (2010).

[13] Nair, V. and Hinton, G. E., "Rectified Linear Units Improve Restricted Boltzmann Machines," *27th International Conference on Machine Learning* , 807–814 (2010).

[14] Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R., "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv: 1207.0580* , 1–18 (2012).

[15] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., and Darrell, T., "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093* (2014).

[16] Chen, H., Zheng, Y., Park, J.-H., Heng, P.-A., and Zhou, S. K., "Iterative multi-domain regularized deep learning for anatomical structure detection and segmentation from ultrasound images," *MICCAI 2016* , 487–495 (2016).