Chapter 8

Localization of Common Carotid Artery Transverse Section in B-mode Ultrasound Images using Faster RCNN: A Deep Learning Approach

Summary

Cardiologists can acquire important information related to patients' cardiac health using carotid artery stiffness, its lumen diameter (LD), and its carotid intima-media thickness (cIMT). The sonographers primarily concern about the location of the artery in B-mode ultrasound images. Localization using manual methods is tedious and time-consuming and also may lead to some errors. On the other hand, automated approaches are more objective and can provide the localization of the artery at near real time. Above arterial parameters may be determined after localization of the artery in real time. A novel method of localization of common carotid artery (CCA) transverse section is presented in this work. The method is known as fast region convolutional neural network (FRCNN)-based localization method and is designed using a stack of three layers viz. convolutional layers, fully connected layers, and pooling layers. These organized layers constitute a region proposal network (RPN) and an object class detection network (OCDN). We obtain an outcome as a bounding box along with a score of prediction around the cross-section of the CCA. B-mode ultrasound image database of CCA is split into training and testing set, to accomplish this, three partition methods K = 2, 5, and 10 are used in our work. The training is extended for 2000 epochs in order to achieve fine-tuned features from the convolutional neural network. After 2000 epochs, we obtain 95% validation accuracy; however, mean of the accuracies up to 2000 epochs is 89.36% for K = 10 partitions protocol (training 90%, testing 10%). Generated CNN model is tested on a different dataset of 433 images and the acquired accuracy is 87.99%. Thus, the proposed method including an advanced deep learning technique demonstrates promising localization for carotid artery transverse section.

8.1 Introduction

Cardiovascular diseases are a vital reason for death in the modern world [297]. One of the critical determination procedures is ultrasound image analysis of cardiovascular illness, as it offers numerous points of interest over other imaging modalities like CT, MRI, and PET [36, 83, 298, 299]. In any case, nature of the ultrasound image is typically poorer than that in different modalities. Confining significant organ or region of interest is an essential task in a diagnostic procedure identified with US image examination. In B-mode ultrasound images of the carotid artery transverse area, the sonographer primarily confines to the artery cross-section. Based on this, one can measure different parameters, for example, cIMT [300, 301], artery stiffness constant [302, 303], and LD [37], for further analysis. Many

algorithms have been deployed by the experts to process the ultrasound images of the carotid artery. Riha and Benes in [302] proposed the circle recognition in cross-section of the artery utilizing Hough transform for a noisy image; however, the utilization of median filtering and iterative thresholding upgrades the activity time and weights of the framework fundamentally for noisy information. Benes *et al.* in connected the genetic algorithm to circle out artery from BUS images [301]. Riha *et al.* added to a similar job in [300] with their framework utilizing the Viola-Jones detector. Their algorithm is trained with a set of labelled images using Adaboost classifier, Haar-like features, and Matthews's coefficient.

Utilizing the conventional strategies like histogram equalization, Canny edge detection, Yang *et al.* recognized the CCA and the lumen region in ultrasound images [304]. Yoon *et al.* have found application of CNN in the reconstruction of ultrasound images utilizing RF information from a receiver (check line) subsampling. A deep neural system proposed by them beats other best in class strategies significantly [304]. Multiresolution convolution neural system proposed by Vedula *et al.* beats the blurring, shading, and speckle noise impacts delivered by beamforming-based image reconstruction methods alongside the decreased, computational efforts [305]. In another work, Peridos *et al.* have used the stacked denoising autoencoder (SDA) based four-layer architecture. The first layer plays out the compression and rests of the layers perform reconstruction, along these lines reducing the noise effect in the image [306].

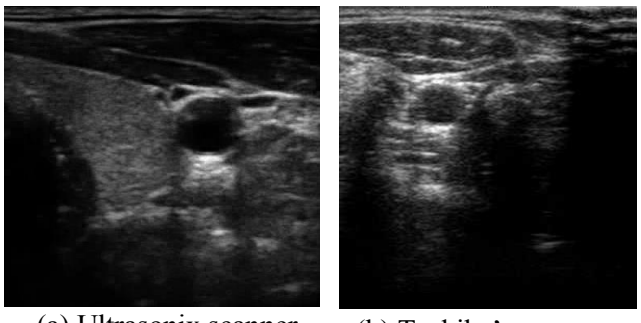
Remarkable outcomes obtained in an exploration performed by Tajbakhsh *et al.* fine-tuning the pre-trained deep neural network in medical image analysis [307]. Their outcomes recommended adjusting the pre-trained deep CNN rather than full training starting with scratch. Training from scratch is a tedious procedure and now and then experiences overfitting and convergence issues, likewise, it requires biomarker amid training which is possibly unavailable sometimes. Along these lines, fine-tuning the prepared CNN show is the better option for the diagnosis of medical images. Ma *et al.* have dealt with the identification of a thyroid tumour utilizing CNN-based strategy over ultrasound images. They utilized two systems CNN15 and CNN4 in their strategy independently and furthermore in a cascade, and compared their outcomes with state-of-the-art systems like VGG, ResNet, and so on [308]. Sudha *et al.* utilized a patch-based segmentation technique for the measurement of intima media thickness in 2640 patches of 220 frames from left and right CCA of 110 patients [88]. Detection of artery cross-section [300-302, 309, 310] resembles that of identifying different objects in an image; along these lines, one can recognize the artery cross-segment utilizing the techniques applied for object detection. One such popular approach is that of region convolution neural network (RCNN) which is accessible. This paper introduces a novel procedure based on fast RCNN for detection of carotid artery transverse segment in B-mode ultrasound (BUS) image. Our deep learning–based technique learns features, which are useful for localization of the image [311-314].

The proposed method gives better accuracy compared with other best in class strategies with the minimal loss of detection. The method is also robust to the variation in the source of data by sonographers. While setting, in situ parameters of an ultrasound scanner for example, gain, frequency , and position of the patients (supine, lateral, recumbent) [315] not affect this robust detection technique. As specified before, the proposed strategy utilizes a best in a class region-based deep neural system named FRCNN [311, 312, 314, 316]. Based on the training with labelled data, the FRCNN approach detects an ROI with the highest possibility of occurrence. The system picks up this ROI as the best bounding box around the artery cross-segment [317]. We provide the training image along with bounding box coordinate as input to a deep neural network, which learns suitable local features for achieving an optimal localization describing the training data. As part of our strategy, we utilize the best deep neural system with the ResNet 50 [313].

Data acquisition and demographics are discussed in Section 8.2. The strategy of the FRCNN framework is depicted in Section 8.3 where region proposal generation, losses, and other performance parameters and the training process are described. Section 8.4 portrays tests and consequences of the framework. Section 8.5 examines the general framework and comparisons.

8.2 Common Carotid Artery Database

A dataset of CCA transverse section from signal processing laboratory, Brno University of Technology, was used in this work [318]. The dataset include three sets of B-mode ultrasound images, which consists of 283 training images (388 × 400 pixels) and 538 validation images (388 × 400 pixels) from an Ultrasonix OP scanner with linear probe L14-5/38 (Ultrasonix Medical, Richmond, BC, Canada), and 433 test images (283 × 322 pixels) from Toshiba Nemio XG scanner with linear probe at a frequency of 7.5 MHz (Toshiba Medical Systems Shimoishigami Otawara-Shi Tochigi-Ken, Japan). . Figure 8.1 below shows raw images obtained from Ultrasonix scanner (left) and Toshiba's scanner (right).



(a) Ultrasonix scanner (b) Toshiba's scanner

Figure 8.1 Raw images of B-mode ultrasound of CCA transverse section obtained from (a) Ultrasonix scanner (b) Toshiba's scanner.

Another feature associated with the database is the canter coordinate (x, y), width and height of the carotid artery transverse section. We have utilized this information as ground truth biomarker to train our system. All patients were Caucasians from South Moravia in the Czech Republic. The images from the Toshiba probe (433 images) are comparatively noisy. We downloaded the database during April 2017 for research purpose only and have given due credits to the mainstay of corresponding dataset [318].

8.3 Methodology

Figure 8.2 demonstrates the schematic of the algorithm to localize the CCA transverse segment. We provide a B-mode ultrasound image of CCA as input to the FRCNN network. The FRCNN design and CCA localization approach comprises two modules as appeared in the block diagram.

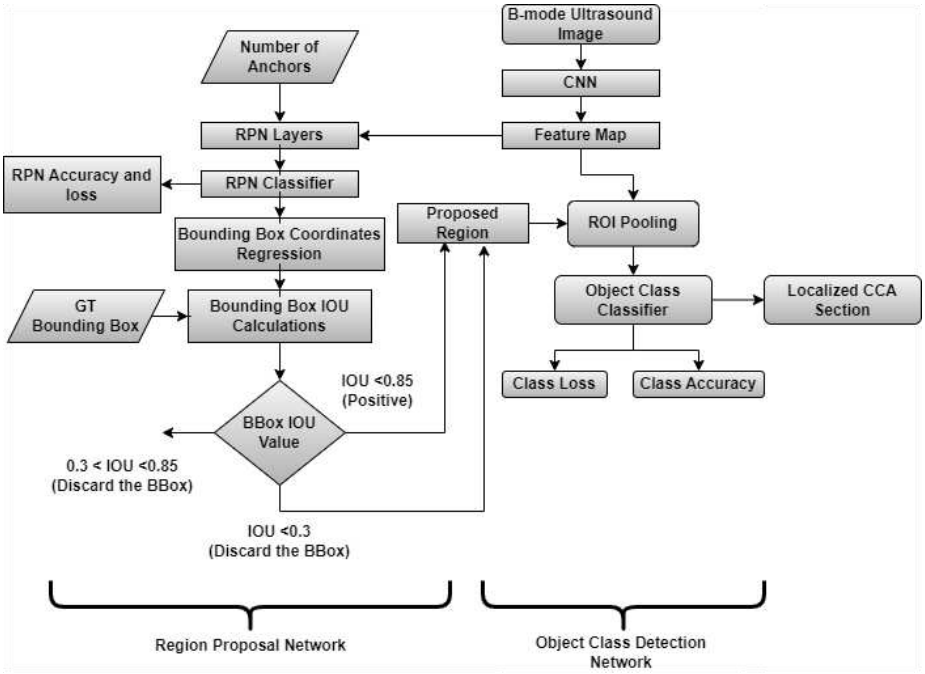


Figure 8.2 Flowchart of FRCNN explaining the localization of artery in BUS images.

(1) A region proposal network (RPN) (regression layer)
(2) Object class detection network (OCDN) (classification layer)

Figure 8.3 demonstrates a schematic diagram of an algorithm joining RPN and OCDN to shape FRCNN. The RPN directs the OCDN network to investigate an image for a specific region. In our model, we are utilizing the same neural system (ResNet) for RPN and OCDN. Amid testing (or validation) stage, input to RPN is CCA transverse area images and proposed rectangular region yielded with a likelihood score of

being the said object. RPNs predict the proposed regions with an extensive range of scales and aspect ratio.

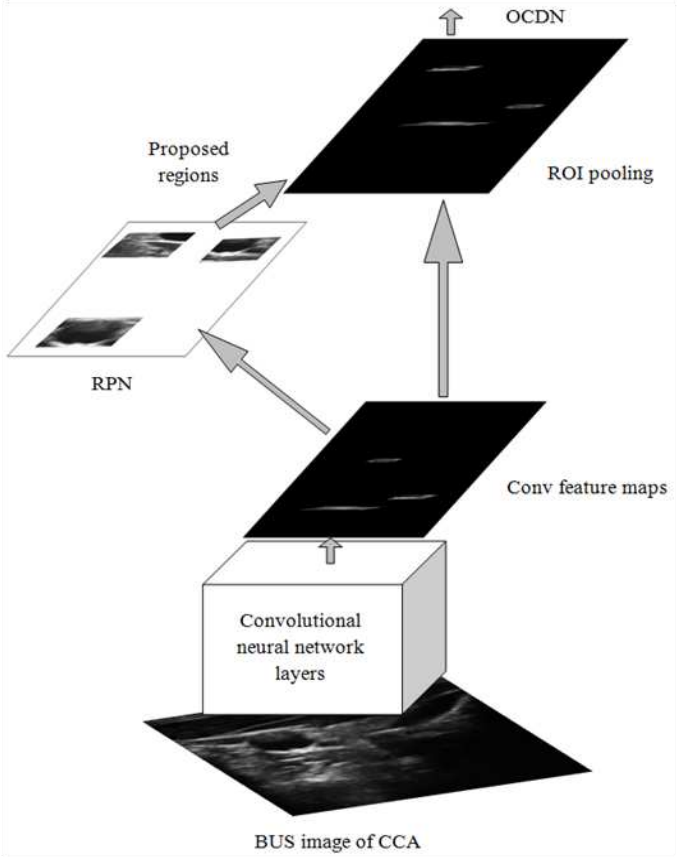


Figure 8.3 Schematic diagram of object class detection with RPN as a combined network: FRCNN.

8.3.1 Generation of region proposal

The RPN basically includes a convolution neural system, which inside learns feature maps valuable for localizing proposals. The sorted out stacks of convolution and max-pooling layers together produce convolution feature maps. The produced region proposal incorporates features learned at these feature maps. Above, convolution feature map is inserted into a small network to generate region proposals. This network takes a 3 × 3 spatial sliding window from convolution feature and maps to a lower dimensional feature. Now, this low dimensional component is fed to two parallel layers to perform classification and regression to create the bounding boxes. Each spatial sliding window predicts multiple region proposals from convolutional feature map. These multiple region proposals cantered at sliding windows are known as anchors and are related with scale and aspect ratio of the sliding window. Figure 8.4 describes generation of k bounding boxes from convolutional feature map of a sliding window.

# of anchors (k) = length of scale parameter * length of aspect ratio parameter

$$k = 3 * 3 = 9$$

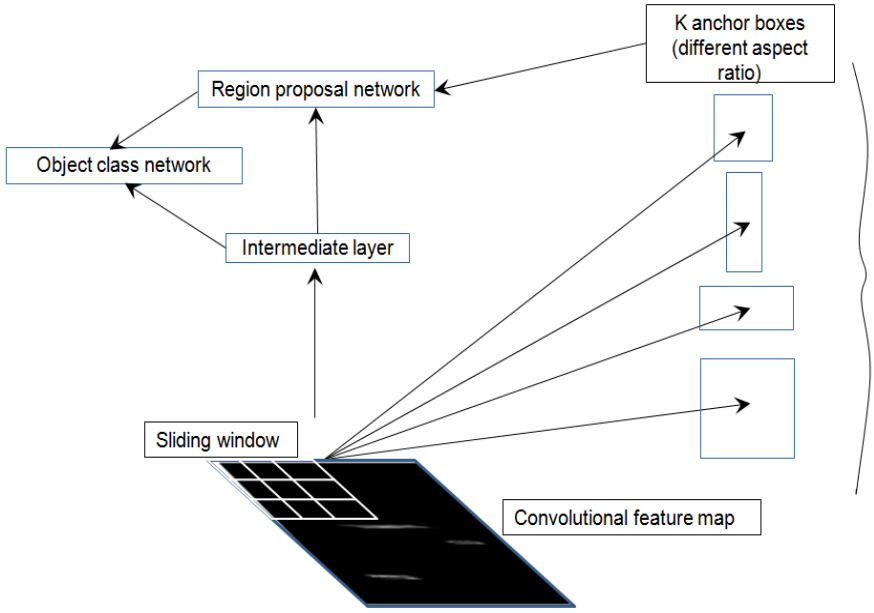


Figure 8.4 Generation of the anchor boxes from convolutional feature map for each sliding window.

A maximum of 9 bounding boxes is utilized for a selected scale and the aspect ratio of 3, respectively. These bounding box coordinates are fed to the box regression layer as indicated Figure 8.2. The box regression layer yields coordinates of "k" bounding box. Further, the regression layer also yields intersection over union (IOU) ascertained utilizing ground truth information. In light of evaluated IOU values, the bounding boxes were named as positive (IOU > 0.85) or negative (IOU < 0.3) proposed regions. Box classification layer fed with these proposals gives an expected likelihood of the object be the artery cross-segment or not. The most extreme number of anchors for an image is "w × h × k," where "w" and "h" are width and height of the images. Counting every term, an objective function characterized as the multitask loss function [312, 314] for our framework appears in Eq. (8.1).

$$L(p_k, t_k) = \frac{1}{Nbatch} \sum L(pk) + \frac{1}{Nanchor} \sum \lambda * L(tk) + \qquad (8.1)$$

Where, k is the index of the anchor, p_k is the predicted score of the anchor k for an object and t_k is the vector of 4 parameters of the bounding box. These loss functions are discussed in detail in the next section.

8.3.2 Losses and other performance indices

8.3.2.1 Loss functions of layers

Amid training, we evaluate IOU for each bounding box compared with the ground truth. Further, the anchors with IOU higher than 0.85 are doled out with the positive class mark (i.e. the likelihood of closeness to ground truth) and anchors under 0.3 are doled out negative class names. Anchors having IOUs somewhere in the range of 0.3 and 0.85 do not get any class mark so they do not take part in training. Equation (8.2) shows the loss function for an anchor.

(i) Classification layer loss

$$L\left(p_k\right) = \sum_k L_{cls\_layer}\left(p_k, p_k^*\right) \tag{8.2}$$

Where L_{cls_layer} is binary cross entropy loss (log loss) for positive class and is given by equation 8.3

$$L_{cls\_layer}\left(p_k, p_k^*\right) = \log p_u \tag{8.3}$$

(ii) RPN regression loss:

$$L\left(t_k\right) = \sum_k p_k^* L_{reg\_layer}\left(t_k, t_k^*\right) \tag{8.4}$$

$$L_{reg\_layer} = \sum_{k \in (x,y,w,h)} smooth\ L1\left(pred_k^u - true_k\right) \tag{8.5}$$

Where,

$$smooth\ \mathrm{L1} = \begin{cases} 0.5\,x^2 & if\ abs\,(x) < 1 \\ abs(x) - 0.5 & otherwise \end{cases} \tag{8.6}$$

As stated earlier, the classification layer in this work outputs binary class label of artery cross-section, thus it has two outputs (i.e. 1 or 0) whereas regression layer outputs the coordinates of bounding box around cross-section of artery. For a test image, p_k is predicted score of kth anchor for being an artery cross section. The ground truth label p_k^* is assigned with binary value (1 or 0) for artery cross-section (1 for positive and 0 for negative). Thus regression layer loss exists only for positive anchor ($p_k^* = 1$) whereas for negative anchors ($p_k^* = 0$) the loss ceases to zero.

8.3.2.2 Accuracy

Accuracy is a performance index of the classification system, in general, defined in percentage prediction of the classification system.

$$\text{Accuracy} = \frac{\#\text{ of correct predictions}}{Total\ \#\ of predictions}$$

Equation (8.7) shows accuracy defined as positive and negative predictions of the classification model for a binary classification:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{8.7}$$

Where TP =True Positives, TN =True Negatives, FP = False Positives, and FN = False Negatives.

8.3.3 Model Pre-training

We train both networks (RPN and OCDN) end to end by using backpropagation and Adam optimizer (with a learning rate of 0.00001). The image-centric technique is applied here to train the network. Each image generates mini batches of size 10 and each batch consists of anchors of positive and negative class values. The loss functions given by equations (8.1)-(8.6) can be optimized for the given number of anchors (positive and negative) in a mini batch. However, possibly all anchors are negatively biased, thus to avoid this situation, we randomly select 256 anchors across an image to compute the loss function of a batch. Both positive- and negative-sampled anchors equally contribute to loss function calculation. If number of either class of anchors (positive or negative) drops below 128 in a mini-batch, the other class pads it.

8.3.4 Joint training of the network

In joint training, we enter the proposed regions and the object classes to the network. The joint network in back propagation process updates the weights for shared layers from combined RPN loss and the object detection loss.

8.4 Experimental protocols

We follow three partition protocols for the training and validation of the whole network (K-fold cross-validation where K = 2, 5, 10). For K = 2 protocol, we divided the whole training dataset of 821 images from Ultrasonix scanner into 50% training (410 images) and 50% validation set (411 images). Thus, the network trains with 50% training set and then validated on remaining 50% validation dataset. In K = 5 partition protocol, we divide dataset into 80% training (656 images) and 20% validation (165 images) set. Similarly, in K = 10 partition protocol, dataset was divided into 90% training (738 images) and 10% validation (83 images) set. We observe a different aspect of increasing training dataset in the result. The weights of neural network are updated after each epoch, which comprises one forward and one backward pass. All the three protocols are carried out for 2000 epochs in order to observe the effect of updated convolutional neural network. We also have a test dataset of 433 patients from Toshiba's scanner which we applied to the trained CNN model of 2000 epochs, for K = 2, 5, and 10 partition protocol.

8.5 Results

As explained in methodology, the highest scoring positive bounding box ideally formed around the cross-section of a carotid artery. Figure 8.5a shows the bounding box with the highest IOU along with a score of 99%. We have obtained similar, single, correct bounding boxes in most cases indicated in quantitative results later. However, in a limited number of cases, some error conditions also exist at the output. These include multiple bounding boxes achieving a high score, as shown in Figure 8.5b, no detections if none of the bounding boxes qualifies score of IOU = 0.85, as shown in Figure 8.5c, and incorrect detections as shown in Fig. 5d, where some other part of the artery treated as a false cross-section.

8.5.3 Experiment 3 (2000 epochs)

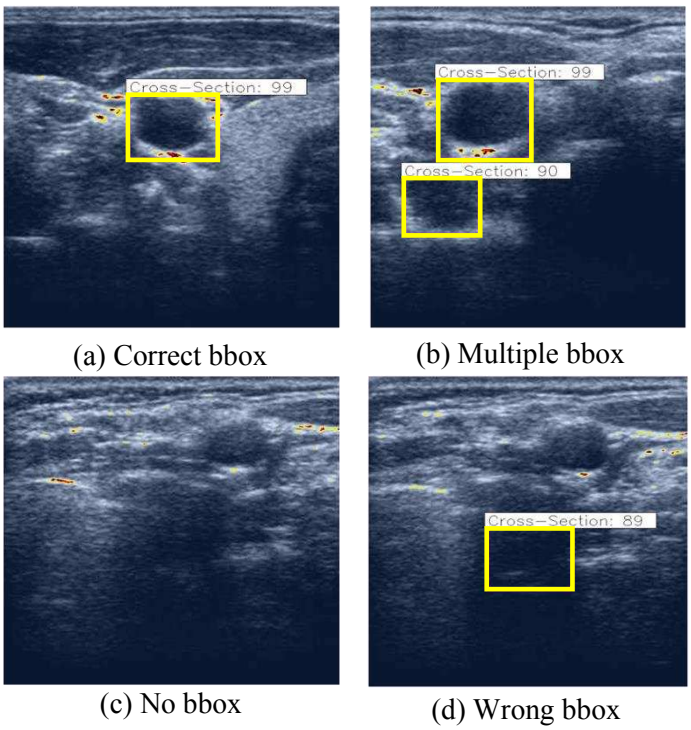
Now, we perform our experiment for 2000 epochs. In this examination, we watch the mean of all validation accuracies up to 2000 epochs are 89.91%, 89.71%, and 89.36% for K = 2, 5, and 10 segment conventions respectively, as appear in Table 8.1. The losses shown in column 6 and 7 of Table 8.1 are RPN and classifier loss for 2000 epochs. At the point when we provide test data from Toshiba's scanner to the trained system, the outcomes demonstrate nearly equal accuracy to approval results. Our framework accurately recognizes the cross-segment in 87.76% images out of 433 images, expecting just a single cross-segment in the image for K = 2.While, same for K = 5 and K = 10 partition, the system accurately relates 84.53% and 87.99% images to the cross-segment. Table 8.2 talks about the test outcomes. We perform all of the analysis in the Intel Xeon processor with 4 GB NVIDIA Quadro K2200 GPU empowered machine and with Python 2.7 and Ubuntu 16.04 operating system framework.

Table 8.1 Validation data results for a mean of 30, 200, 2000 epochs respectively.

	Total patients	Training images	Validation images	Validation accuracy (2000 epochs)	RPN loss (2000 epochs)	Classifier loss (2000 epochs)
K=2	821	410	411	89.91%	0.040	0.239
K=5	821	656	165	89.71%	0.041	0.243
K=10	821	738	83	89.36%	0.045	0.250

Table 8.2 Test data results of Toshiba's scanner data set of 433 images.

Sr. #	Total Training images	No bounding box	Correct Bounding box	More than 1 bounding box	Wrong bounding box	% Accuracy (True detection)	% False Detection
K-2	433	30	380	22	3	87.76	5.77
K-5	433	54	366	13	0	84.53	3.0
K-10	433	44	381	8	0	87.99	1.85



(a) Correct bbox (b) Multiple bbox

(c) No bbox (d) Wrong bbox

Figure 8.5 FRCNN output of combined layer using Toshiba's scanner test data showing (a) correct bbox (b) multiple boxes (c) no bbox (d) wrong bbox, around the cross-section of the carotid artery with classification score.

8.6. Discussion

8.6.1 Learning ability of the system

Keeping the framework's learning capacity in mind and making it ready, to sum up with the data size, we decided to pick the three partition conventions. At first, for K = 2 partition convention, we opted to train the framework with just 50% of training data and created a training model. Now, we connected this model to Toshiba's test data to check generalization of the framework and we get 380 pictures with a single correct bounding box (ideal case), 22 pictures with multiple bounding boxes (amongst which one is correct), 30 pictures with no bounding box, and 3 wrong recognition. As opposed to this, when we increase the training size with K = 5 partition convention (80% training), we obtain 366 images with a single correct bounding box, there are 54 cases of no detection, 13 images within excess of one or more bounding boxes, yet 0 wrong detections. In this manner, a trade-off exists between the two cases. Be that as it may, with training size, 90%, (K = 10 parcel convention) the framework moves towards a greatly improved summed up network model yielding 381 images with single correct bounding box, 44 instances of no identifications, yet just 8 instances of in excess of one bounding box, and zero wrong detection. If we dispose of the images with no bounding box from test data set from K = 10 partition convention

considering noisy images, the effective accuracy for K = 10 is 97.94% (disposing of such cases effectively sums to the manual observation of a small number of frames.)

8.6.2 Benchmarking

A few different researchers have displayed techniques for location of the artery in ultrasound images. A list of correlation between existing techniques and the proposed system appears in Table 8.3. We discovered four attributes to find correlations between the current strategies and proposed technique viz. algorithm, data size, processing time, and the execution parameters (accuracy, success rate, true/false rate, and so on.). However, recent findings of deep learning in the field of ultrasound imaging are extremely uncommon, its scope is not limited to only pattern recognition and classification. The model presented by Riha and Benes in [302] accomplished 79.2% accuracy for a noisy image. An automated presented by Benes *et al.* in [301] uses grammar-guided genetic programming method to detect the artery from BUS images. The accomplished accuracy is 4% more prominent than the past strategy yet the framework is intricate and slower contrasted with the past technique. The outcome of the algorithm proposed by Riha *et al.* in [300] is an improved true positive rate of 86.12% with a base false positive rate of 4.16%. Yang *et al.* in [13] achieved a figure of merit value just 0.705 and mean absolute error of 0.47 ± 0.13 for 180 images of 10 patients. In spite of the above frameworks, our framework utilizes best in class object identification strategy. The accomplished accuracy is close to 89% for K = 10 partition convention with zero false location. We observe the following strength and weakness in our system.

(1) The proposed work combines the use of automated localization of the cross section in B-mode ultrasound images of CCA.

(2) In this work, we have demonstrated the use of deep neural network to extract the feature from the ultrasound images and updated model is used further for validation.

(3) Generated model can be used online for validation on new images and also can be integrated with hardware for detection of the artery cross-section.

(4) Testing time is very low (of the order of seconds) and is a direct function of high-performance computing (HPC) resources. With the availability of HPC, validation time can be reduced significantly. However, availability of HPC is a major setback of this system. We accomplish the real-time assessment of CCA ultrasound image for identification of artery cross segment. With the expanding number of epochs, deep learned features enhance the decision-making capability of the machine.

8.7 Conclusion

We have presented an FRCNN system as the combination of RPN and OCDN for localization of carotid artery transverse section in BUS images which shows fast and efficient performance. The arteries are located in ultrasound images by generating a bounding box around them. Localization of the artery

leads to measurement of other arterial parameters such as cIMT, LD, and stiffness. With the interface of GPU, experiments consume very low time. Also, the method presents a generalized model which is free from the source of ultrasound image data. The results are very reliable and can be used to design the prototype for clinical setups also; however, the demand for high-performance computing will always be a constraint.

Table 8.3 Comparison of existing literature with the current study based on some parameters.

Sr #	Name of author	Task	Algorithm	Data Size (Training/Validation/Testing	Detection Time	Performance Parameter
1	Benes *et al.* 2014 [301]	Automatic Localization of CCA Transverse section	Genetic algorithm	16 Training 52 Testing (video Sequences)	96 Hours (training) 15 min testing	Accuracy = 82.7%
2	Riha *et al.* 2010 [302]	CCA detection in medical video sequences	Circle Hough Transformation and Bayes Classifier	250 Video sequences	3.2 sec/50 frames	Accuracy = 79.2%
3	Riha *et al.* 2013 [300]	Localization of CCA using Viola Jones detector	Viola Jones detector	Set 1 = 283 Set 2 = 538 Set 3 = 433 Images	0.013 sec/image	TPR = 86.14* FPR = 4.16*
4	Our method	Localization of CCA using Deep Neural Network	Deep Neural Network (ResNet)	Set 1 = (283 + 538) Set 2 = 433 Images	1.7 sec to 3.2 sec/image	Success rate = 87.99 False rate = 1.84