CHAPTER

5 Automatic Interpretation of Carotid Intima–Media Thickness Videos Using Convolutional Neural Networks

Nima Tajbakhsh*, Jae Y. Shin*, R. Todd Hurst†, Christopher B. Kendall†, Jianming Liang*

*Arizona State University, Scottsdale, AZ, United States** *Mayo Clinic, Scottsdale, AZ, United States†*

CHAPTER OUTLINE

CHAPTER POINTS

- A framework based on CNNs that automates the entire CIMT interpretation process
- A novel method that selects the end-diastolic frames from an ultrasound video

Deep Learning for Medical Image Analysis. DOI: 10.1016/B978-0-12-810408-8.00007-9

- A new contextually-constrained method that localizes an ROI in an end-diastolic frame
- A framework that enables accurate interface segmentation in an ROI
- The suggested CNN-based system outperforms our previous handcrafted approach for CIMT interpretation

5.1 INTRODUCTION

Cardiovascular disease (CVD) is the number one killer in the United States [1]: every 40 seconds, one person in the United States dies of CVD [35]. More striking, nearly one-half of these deaths occur suddenly as the initial manifestation, and one-third of them occur in patients younger than 65 years [35,42]. Nevertheless, CVD is largely preventable [1]. However, the key is to identify at-risk persons before coronary events occur [42], so that scientifically proven and efficacious preventive care can be prescribed appropriately.

Carotid intima–media thickness (CIMT) is a noninvasive ultrasonography method that has proven to be valuable for predicting individual CVD risk [42]. It quantifies subclinical atherosclerosis, adds predictive value to traditional risk factors (e.g., the Framingham risk score [9]), and has several advantages over computed tomography (CT) coronary artery calcium score by being safer (no radiation exposure), more sensitive in a young population, and more accessible to the primary care setting. However, as illustrated in Fig. 5.1, interpretation of CIMT videos involves 3 manual operations: (i) selection of 3 end-diastolic ultrasonographic frames (EUFs) in each video; (ii) localization of a region of interest (ROI) in each selected EUF; and (iii) identification of the intima–media boundaries within each ROI to measure CIMT. These 3 operations, and in particular the CIMT measurement, are not only tedious and laborious, but also subjective to large interoperator variability [29] if guidelines are not properly followed. Therefore, many semi-automated computer-aided measurement systems have been suggested for CIMT (e.g., [26,7,37,47,16,3]). In this study, we present a *fully-automated* system to accelerate CIMT video interpretation while still offering a highly user-friendly interface that allows sonographers to offer their indispensable expertise.

The first appearance of CNNs dates back to 1989 when they were used for handwritten ZIP code recognition [21]. However, they had not demonstrated superior performance in image analysis until the ImageNet competition in 2012, a success that has triggered a revolution in computer vision through the efficient use of graphics processing units (GPUs) and other new technologies. This breakthrough widely inspired research on exploring the use of CNNs as a promising alternative to the existing handcrafted approaches. Recently, CNNs have been successfully used in medical image analysis, resulting in a new generation of computer-aided detection and diagnosis systems [5,41,8,46,43] with superior accuracy over the existing solutions. In this study, we show that with proper preprocessing and post-processing, our

CNN-based approach can substantially outperform our previous carefully engineered method for CIMT image interpretation, making the following specific contributions:

- A framework based on CNNs that automates the entire CIMT interpretation process. This is in contrast to the prior works where only the very last step of the CIMT interpretation process was automated. The performance of the suggested system significantly outperforms our previous handcrafted approach, which, to our knowledge, is the only system in the literature that aimed to automate all the above three tasks.
- A novel frame selection method based on the electrocardiogram signals displayed at the bottom of ultrasound frames. The suggested method utilizes effective preprocessing of patches and post processing of CNN outputs, enabling a significant increase in the performance of a baseline CNN.
- A new method that localizes the ROI for CIMT interpretation. The suggested method combines the discriminative power of a CNN with a contextual constraint to accurately localize the ROIs in the selected frames. We demonstrate that the suggested contextually-constrained CNN outperforms the performance of a baseline CNN.
- A framework that combines CNNs with active contour models for accurate boundary segmentation. Specifically, given a localized ROI, the CNN initializes two open snakes, which further deform to acquire the shapes of intima–media boundaries. We show that the segmentation accuracy of the suggested method is greater than our previous handcrafted approach.

5.2 RELATED WORK

Given the clinical significance of CIMT as an early and reliable indicator of cardiovascular risk, several methods have been developed for CIMT image interpretation. The CIMT is defined as the distance between the lumen–intima and media–adventitia interfaces at the far wall of the carotid artery (Fig. 5.1). Therefore, to measure CIMT, the lumen–intima and media–adventitia interfaces must be identified. Hence, earlier approaches were focused on analyzing the intensity profile and distribution, computing the gradient [38,44,11], or combining various edge properties through dynamic programming [24,6,39]. The recent approaches [26,7,37,47,16,3] are mostly based on active contours (a.k.a., snakes) or their variations [18]. Some of these approaches require user interaction; other approaches aim for complete automation through integration of various image processing algorithms, such as Hough transform [34] and dynamic programming [39]. Most recently, a committee of standard multilayer perceptrons was used in [31], and a single standard, multilayer perceptron with an autoencoder was used in [30] for CIMT image interpretation, but both methods did not outperform a snake-based approach from the same research group [3,4]. A more complete survey of methods for automatic CIMT measurements can be found in the review studies conducted by Molinari et al. [32] and Loizou et al. [25].

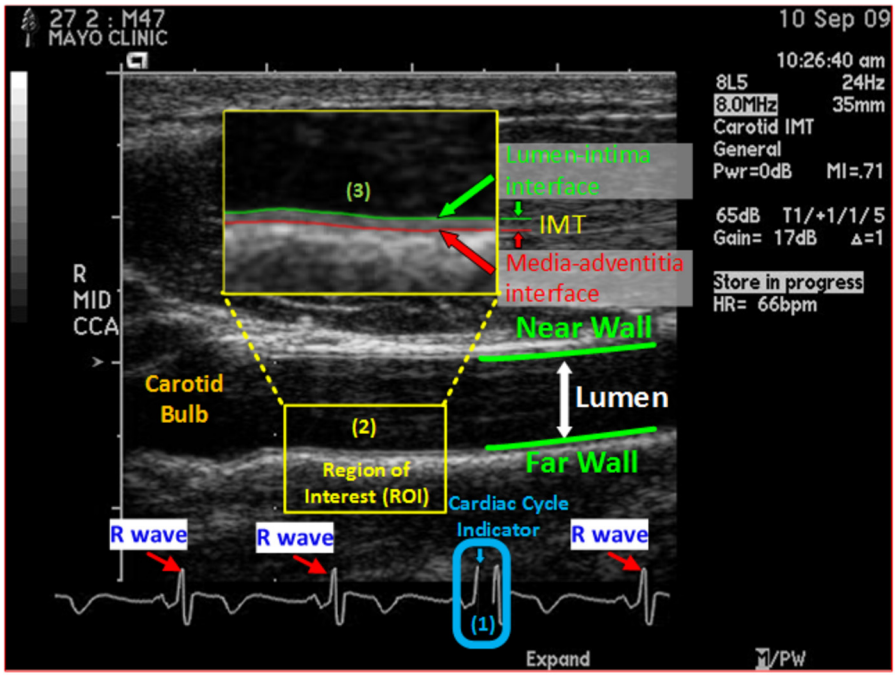

FIGURE 5.1

Longitudinal view of the carotid artery in an ultrasonographic B-scan image. CIMT is defined as the distance between the lumen–intima interface and the media–adventitia interface, measured approximately 1 cm from the carotid bulb on the far wall of the common carotid artery at the end of the diastole. Therefore, interpretation of a CIMT video involves 3 operations: (i) selecting 3 EUFs in each video (the cardiac cycle indicator shows to where in the cardiac cycle the current frame corresponds); (ii) localizing an ROI approximately 1 cm distal from the carotid bulb in the selected EUF; and (iii) measuring the CIMT within the localized ROI. In this paper, we aim to automate these 3 operations simultaneously through the use of CNNs.

However, nearly all the methods explained above are focused on only the third operation, CIMT measurement, and ignore the 2 preceding operations of frame selection and ROI localization. To our knowledge, the only system that simultaneously automates the 3 operations is our prior work [40], an extension of [48], which automatically selects the EUF, localizes the ROI in each selected EUF, and provides the CIMT measurement in the selected ROI. However, as with other works, this method is based on handcrafted algorithms, which often lack the desired robustness for routine clinical use [40]. Our new proposed method aims to overcome this weakness through CNNs. As demonstrated in Sections 5.5 and 5.6, our new system outperforms our previous handcrafted method in all aspects, including frame selection, ROI localization, and CIMT measurements.

5.3 CIMT PROTOCOL

Each of the CIMT examinations was performed with high-resolution B-mode ultrasonography using a 15 MHz linear array transducer with fundamental frequency only (Acuson Sequoia, Mountain View, CA, USA). The carotid screening protocol begins with scanning up from the lower neck in the transverse manner to the carotid artery and then further to the carotid bulb and to internal and external carotid arteries. The probe is then turned to obtain the longitudinal view of the common carotid artery. The sonographer optimizes the 2-dimensional images of the lumen–intima and media–adventitia interfaces at the level of the common carotid artery by adjusting overall gain, time gain, compensation, and focus position. After the parameters are adjusted, the sonographer captures 2 CIMT videos focused on the common carotid artery from 2 optimal angles of incidence. The same procedure is repeated for the other side of neck, resulting in a total of 4 CIMT videos for each patient. The videos are recorded at 24 frames/second and consist of 640×480 ultrasound images. The pixel spacing is 0.09 mm/pixel along both x and y directions.

5.4 METHOD

We propose a solution based on CNNs for automating all 3 tasks in CIMT interpretation. Herein, we first briefly explain CNNs and then detail each stage of the suggested system.

5.4.1 CONVOLUTIONAL NEURAL NETWORKS (CNNS)

CNNs are deep learning machines that were originally proposed by LeCun in 1989 for handwritten ZIP code recognition [21]. Although CNNs were a breakthrough, they did not achieve much popularity at the time, mainly because of slow central processing units (CPUs) and limited memory resources. However, the advent of powerful GPUs in scientific computing, together with effective regularization techniques [12,13,19,45,22], has once again revived the CNNs, allowing researchers to break records in many computer vision and medical image analysis challenges. The main power of CNNs stems from a deep learning architecture that allows for learning a hierarchical set of features. One major advantage of CNNs is that they are end-to-end learning machines where the input images are directly mapped to the target labels or target bounding box coordinates. This direct mapping eliminates the need for designing suboptimal handcrafted features, which often provide a noisy image representation with insufficient discriminative power.

CNNs are an extension of multilayer perceptrons, in which fully connected hidden layers are replaced with convolutional layers. In a convolutional layer, each unit is connected to only a small subset of spatially connected units in the previous layer and the weights of such connections are shared among all units in the convolution layer. Weight sharing dramatically decreases the network's width (i.e., number of

parameters in each layer), enabling the design of deeper architectures. To reduce computational complexity and achieve a hierarchical image representation, each sequence of convolution layers is followed by a pooling layer, a workflow reminiscent of simple and complex cells in the primary visual cortex [14]. The pooling layer subsamples the feature maps of the preceding convolutional layer by computing the average response or selecting the maximum response in small overlapping or non-overlapping neighborhoods. A CNN typically consists of several pairs of convolutional and pooling layers, followed by a number of consecutive 1×1 convolutional layers (a.k.a., fully connected layers), and finally a softmax layer or a regression layer to generate the desired outputs.

If D denotes a set of training images, W denotes a matrix containing the weights of the convolutional layers, and $f_W(D^{(i)})$ denotes the loss for the ith training image, the loss over the entire training set is computed with the following equation:

$$\mathcal{L}(W) = \frac{1}{|D|} \sum_{i}^{|D|} f_W(X^{(i)}). \tag{5.1}$$

Gradient descent [2] is commonly used for minimizing the above loss function with respect to the unknown weights W. However, the modern, massively parallelized implementations of CNNs are limited by the amount of memory on GPUs; therefore, the loss function cannot be evaluated on the basis of the entire training set D at once. Instead, the loss function is approximated with the loss over the mini-batches of training images of size $N \ll |D|$. A common choice of the mini-batch size is 128, which is a reasonable trade-off between loss noise suppression and memory management. Given the size of mini-batches, the loss function can be approximated as $\mathcal{L}(W) \approx \frac{1}{N} \sum_{i=1}^{N} f_W(X^{(i)})$ and the weights of the network iteratively updated with the following equations:

$$\begin{aligned} \gamma_t &= \gamma^{\lfloor \frac{tN}{|D|} \rfloor}, \\ V_{t+1} &= \mu V_t - \gamma_t \alpha \Delta L(W_t), \\ W_{t+1} &= W_t + V_{t+1} \end{aligned} \tag{5.2}$$

where α is the learning rate, μis the momentum that indicates the contribution of the previous weight update in the current iteration, and γ is the scheduling rate that decreases the learning rate α at the end of each epoch.

5.4.2 FRAME SELECTION

For a CIMT video, the first step in cardiovascular risk assessment is to select 3 EUFs. The CIMT test is routinely performed with electrocardiogram (ECG) test, and the operator normally selects these frames on the basis of the ECG signal that is printed at the bottom of ultrasonographic frames. Each frame in the CIMT video corresponds to a particular location in the printed ECG signal. To establish this correspondence, as

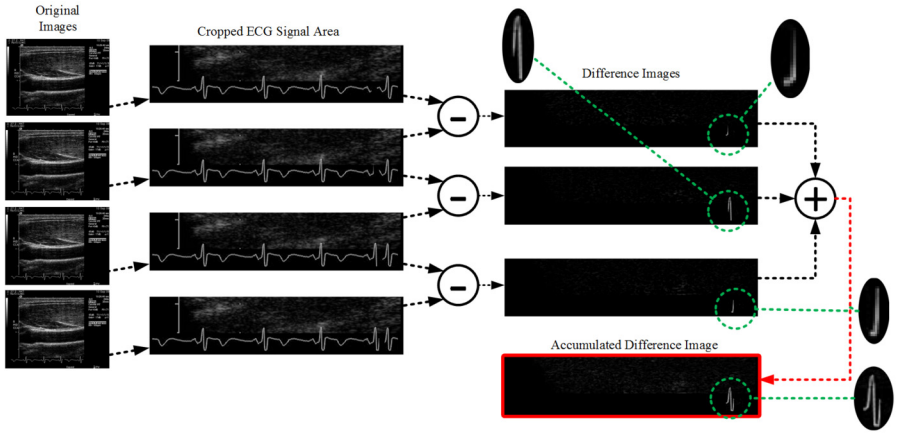

FIGURE 5.2

An accumulated difference image is generated by adding up 3 neighboring difference images.

shown in Fig. 5.1, a black line indicator is displayed on the ECG signal, indicating to where in the cardiac cycle the current frame corresponds. In routine clinical practice, the operator selects the EUFs so that the corresponding black line indicator coincides with the R wave in the QRS complex of the printed ECG signals.

We aim to automate the frame selection process by automatically identifying the frames that correspond to the R wave in the QRS complex of a printed ECG signal. Our idea is to first reconstruct the segment of the ECG signal that is masked by a black line indicator in the current frame and then determine whether the restored wavelet[1] resembles the appearance of an R wave or not. For this purpose, we introduce the notion of accumulated difference images that allows us to capture the missing wavelets and then use a CNN to classify these captured wavelets into R wave or non-R wave categories.

Let I^t denote an image subregion selected from the bottom 20% of an ultrasonographic frame so that it contains the displayed ECG signal. We first construct a set of difference images d^t by subtracting every consecutive pair of images, $d^t = |I^t - I^{t+1}|$, and then form accumulated difference images by adding up every 3 neighboring difference images, $D^t = \sum_{i=0}^{2} d^{t-i}$. Accumulated difference image D^t can capture the segment of the ECG signal that is masked by the black line indicator at frame t. Fig. 5.2 illustrates how an accumulated difference image is generated.

Second, we determine the location of the restored wavelet in each accumulated difference image. For this purpose, we find the weighted centroid $c = [c_x, c_y]$ of each accumulated difference image D^t as follows:

$$c = \frac{1}{Z_t} \sum_{p \in D^t} D^t(p_x, p_y) \times p$$

where $p = [p_x, p_y]$ is a pixel in the accumulated difference image and $Z_t = \sum_{p \in D^t} D^t(p_x, p_y)$ is a normalization factor that ensures the weighted centroid stays within the image boundary. After centroids are identified, we extract patches of size 32×32 around the centroid locations. Specifically, we extract patches with up to 2 pixel translations from each centroid. However, we do not perform data augmentation by scaling the patches because doing so would inject label noise in the training set. For instance, a small restored wavelet may take the appearance of an R wave after expanding or an R wave may look like a non-R wave after shrinking. We also do not perform rotation-based patch augmentation because we do not expect the restored wavelets to appear with rotation in the test image patches. After collection, the patches are binarized with Otsu's method [36]. In Section 5.5.1, we discuss the choice of binarization method through an extensive set of experiments. Each binary patch is then labeled as positive if it corresponds to an EUF (or an R wave); otherwise, it is labeled as negative. Basically, given a patch, we initially determine the accumulated difference image from which the patch is extracted. We then trace back to the underlying difference images and check whether they are related to the EUF or not. After the patches are labeled, we form a stratified set with 96,000 patches to train a CNN for distinguishing between R waves and non-R waves.

Fig. 5.3 shows our frame selection system for a test video. We compute an accumulated difference image for each frame in the video. We then extract image patches from the weighted centroids of the accumulated difference images. The probability of each frame being the EUF is measured as the average probabilities assigned by the CNN to the corresponding patches. By concatenating the resulting probabilities for all frames in the video, we obtain a probability signal whose local maxima indicate the locations of the EUFs. However, the generated probability signals often exhibit abrupt changes, which can cause too many local maxima along the signal. We therefore first smooth the probability signal using a Gaussian function and then find the EUFs by locating the local maxima of the smoothed signals. In Fig. 5.3, for illustration purposes, we also show the reconstructed ECG signal, which is computed as the average of the accumulated difference images, $\frac{1}{N}\sum_{t=1}^{N} D^t$ with N being the number of frames in the video. As seen, the probability of a frame being the EUF reaches a maximum around the R waves of the QRS complexes (as desired) and then smoothly decays as it distances from the R waves. By mapping the locations of the local maxima to the frame numbers, we can identify the EUFs in the test video.

5.4.3 ROI LOCALIZATION

Accurate localization of the ROI is challenging because, as seen in Fig. 5.1, no notable differences can be observed in image appearance among the ROIs on the far wall of the carotid artery. To overcome this challenge, we use the location of the carotid bulb as a contextual constraint. We choose this constraint for 2 reasons: (i) the carotid bulb appears as a distinct dark area in the ultrasonographic frame and thus can be uniquely identified; and (ii) according to the consensus statement of American Society of Electrocardiography for cardiovascular risk assessment [42], the ROI should

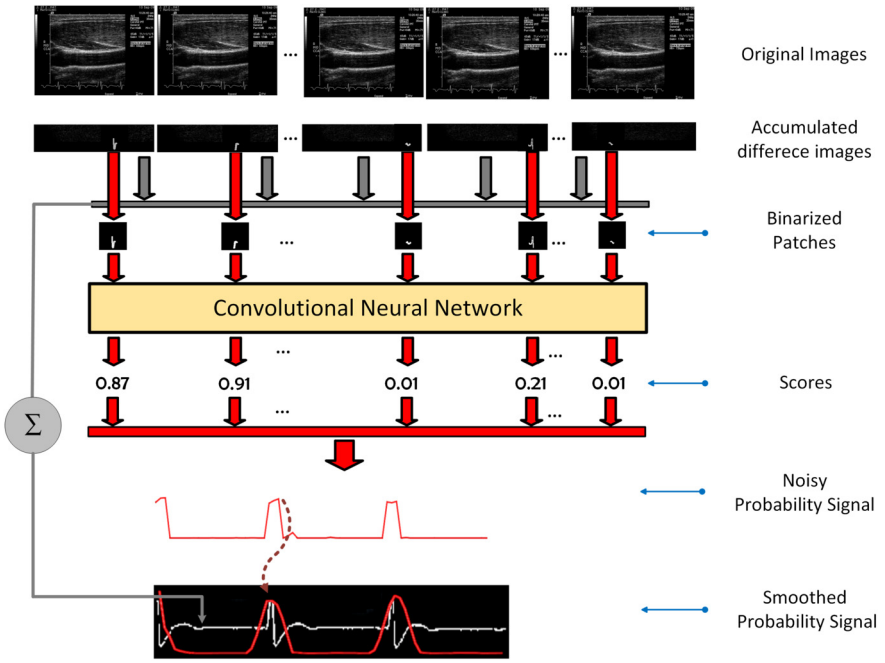

FIGURE 5.3

The test stage of our automatic frame selection scheme.

be placed approximately 1 cm from the carotid bulb on the far wall of the common carotid artery. The former motivates the use of the carotid bulb location as a constraint from a technical point of view, and the latter justifies this constraint from a clinical standpoint.

We incorporate this constraint in the suggested system by training a CNN for 3-class classification that simultaneously localizes both ROI and carotid bulb and then refines the estimated location of the ROI given the location of the carotid bulb. Fig. 5.4 shows how the image patches are extracted from a training frame. We perform data augmentation by extracting the training patches within a circle around the locations of the carotid bulbs and the ROIs. The background patches are extracted from a grid of points sufficiently far from the locations of the carotid bulbs and the ROIs. Of note, the described translation-based data augmentation is sufficient for this application because our database provides a relatively large number of training EUFs, from which a large set of training patches can be collected. After the patches are collected, we form a stratified training set with approximately 410,000 patches to train a CNN for constrained ROI localization.

In referring to Fig. 5.5, the trained CNN is applied during the test stage to all the pixels in the EUF, generating 2 confidence maps with the same size as the EUF. The

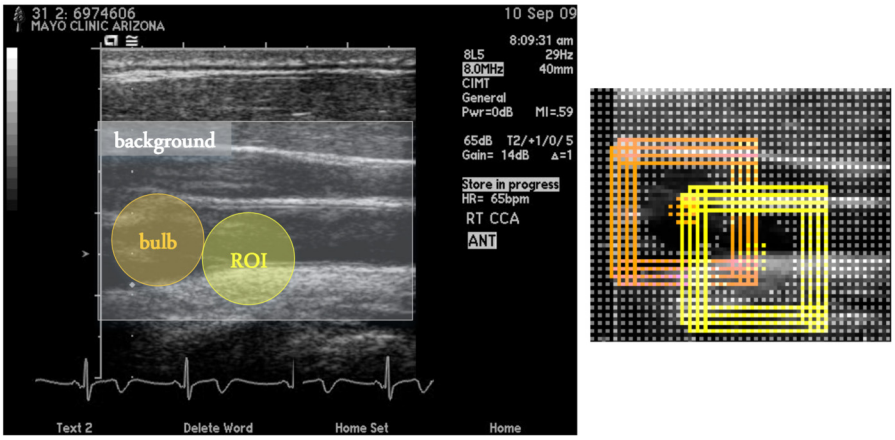

FIGURE 5.4

For constrained ROI localization, we use a CNN for 3-class classification whose training image patches are extracted from a grid of points on the background and around the ROI and the carotid bulb locations.

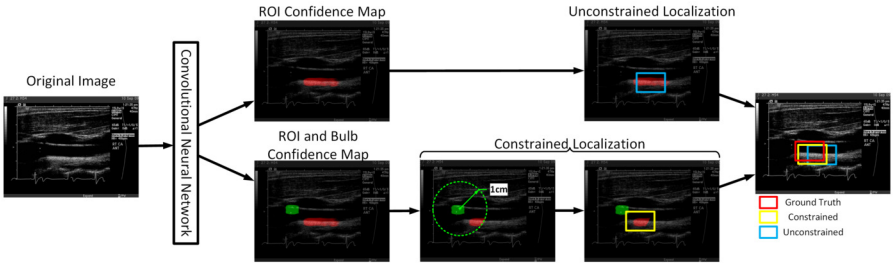

FIGURE 5.5

The test stage of our ROI localization method. In the unconstrained scenario, we use only the ROI confidence map, which results in a relatively large localization error. In the constrained mode, given the estimated location of the carotid bulb, we localize the ROI more accurately.

first confidence map shows the probability of a pixel being the carotid bulb, and the second confidence map shows the probability of a pixel being the ROI. One way to localize the ROI is to find the center of the largest connected component within the ROI confidence map without considering the detected location of the carotid bulb. However, this naive approach may fail to accurately localize the ROI. For instance, a long-tail connected component along the far wall of the carotid artery may cause substantial ROI localization error. To compound the problem, the largest connected component of the ROI confidence map may appear far from the actual location of the ROI, resulting in a complete detection failure. To overcome these limitations,

we constrain the ROI location l_{roi} by the location of the carotid bulb l_{cb}. For this purpose, we determine the location of the carotid bulb as the centroid of the largest connected component within the first confidence map and then localize the ROI using the following formula:

$$l_{roi} = \frac{\sum_{p \in C^*} M(p) \cdot p \cdot I(p)}{\sum_{p \in C^*} M(p) \cdot I(p)} \tag{5.3}$$

where M denotes the confidence map of being the ROI, C^* is the largest connected component in M that is nearest to the carotid bulb, and $I(p)$ is an indicator function for pixel $p = [p_x, p_y]$ that is defined as

$$I(p) = \begin{cases} 1, & \text{if } \|p - l_{cb}\| < 1\,\text{cm}, \\ 0, & \text{otherwise.} \end{cases}$$

Basically, the indicator function excludes the pixels located farther than 1 cm from the carotid bulb location. This choice of the distance threshold is motivated by the fact that the ROI is located within 1 cm to the right of the carotid bulb. Fig. 5.5 illustrates how the location of the carotid bulb as a contextual clue improves the accuracy of ROI localization.

5.4.4 INTIMA–MEDIA THICKNESS MEASUREMENT

To automatically measure intima–media thickness, the lumen–intima and media–adventitia interfaces of the carotid artery need to be detected within the ROI. Although the lumen–intima interface is relatively easy to detect, the detection of the media–adventitia interface is challenging because of the faint image gradients around its boundary. We formulate this interface segmentation problem as a 3-class classification task where the goal is to classify each pixel within the ROI into 3 categories: (i) a pixel on the lumen–intima interface, (ii) a pixel on the media–adventitia interface, and (iii) a background pixel.

We use a 3-way CNN to segment the lumen–intima and media–adventitia interfaces. To train the CNN, we collect image patches from the lumen–intima interface and media–adventitia interface, as well as from other random locations far from the desired interfaces. Fig. 5.6 shows how the training patches are collected from 1 ROI. For the positive patches, we choose not to perform data augmentation. This decision is made because 92×60 ROIs allow us to collect a large number of patches around the lumen–intima and media–adventitia interfaces. Furthermore, given the relatively small distance between the 2 interfaces, translation-based data augmentation would inject a large amount of label noise in the training set, which would negatively impact the convergence and the overall performance of the CNN. When the patches are collected, we form a stratified training set with approximately 380,000 patches to train a 3-way CNN for interface segmentation.

Fig. 5.7 illustrates how our system measures intima–media thickness in a test ROI. The trained CNN is applied to a given test ROI in a convolutional manner, generating

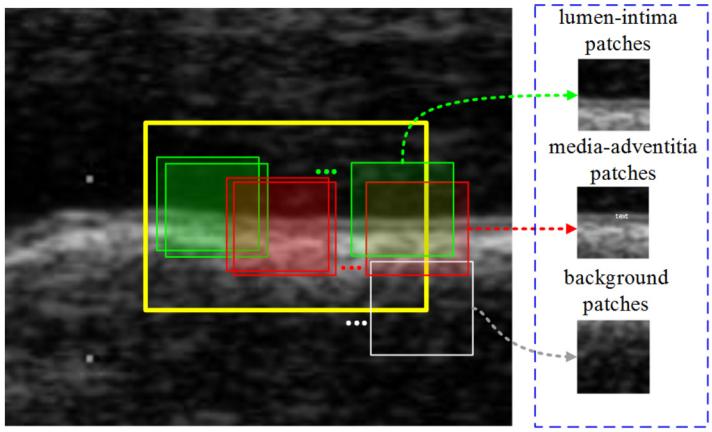

FIGURE 5.6

For lumen–intima and media–adventitia interface segmentation, we use a CNN whose training image patches are extracted from the background and around the lumen–intima and media–adventitia interfaces.

FIGURE 5.7

The test stage of lumen–intima and media–adventitia interface detection. (A) A test region of interest. (B) The trained CNN generates a confidence map where the green and red colors indicate the likelihood of lumen–intima interface and media–adventitia interface, respectively. (C) The thick probability band around each interface is thinned by selecting the largest probability for each interface in each column. (D) The step-like boundaries are refined through 2 open snakes. (E) The ground truth made as the consensus of two experts.

2 confidence maps with the same size as the ROI. The first confidence map shows the probability of a pixel being on the lumen–intima interface; the second confidence map shows the probability of a pixel being on the media–adventitia interface. We have shown the 2 confidence maps in Fig. 5.7 where the green and red colors indicate the likelihood of being the lumen–intima interface and the media–adventitia interface, respectively. A relatively thick high-probability band is apparent along each interface, which hinders the accurate measurement of intima–media thickness. To thin the detected interfaces, we scan the confidence map column by column, searching for the rows with the maximum response for each of the 2 interfaces. By doing so, we obtain a 1-pixel-thick boundary with a step-like shape around each interface as shown in Fig. 5.7C. To further refine the boundaries, we use 2 active contour models (a.k.a.,

Table 5.1 The CNN architecture used in our experiments. Of note, C is the number of classes, which is 2 for frame selection and 3 for both ROI localization and intima–media thickness measurements

Layer	Type	Input	Kernel	Stride	Pad	Output
0	input	32×32	N/A	N/A	N/A	32×32
1	convolution	32×32	5×5	1	0	$64 \times 28 \times 28$
2	max pooling	$64 \times 28 \times 28$	3×3	2	0	$64 \times 14 \times 14$
1	convolution	$64 \times 14 \times 14$	5×5	1	0	$64 \times 10 \times 10$
2	max pooling	$64 \times 10 \times 10$	3×3	2	0	$64 \times 5 \times 5$
2	fully connected	$64 \times 5 \times 5$	5×5	1	0	250×1
2	fully connected	250×1	1×1	1	0	$C \times 1$

snakes) [23], one for the lumen–intima interface and one for the media–adventitia interface. The open snakes are initialized with the current step-like boundaries and then deform solely based on the probability maps generated by the CNN rather than the original image content. Fig. 5.7D shows the converged snakes for the test ROI. We determine intima–media thickness as the average of vertical distance between the 2 open snakes.

5.5 EXPERIMENTS

We use UFL MCAEL CIMT research database [15] with an average population age of 27.0 and standard deviation of 2.8 years. We select 23 patients from this database using systematic random sampling, resulting in a total of 92 CIMT videos (4 videos per patient). The number of frames in each video ranges between 49 to 119. Each video covers at least 3 cardiac cycles and thus a minimum of 3 EUFs. The ground truth for each video consists of the locations of EUFs, the locations of ROIs, and the segmentation of lumen–intima and media–adventitia interfaces. The ground truth is created by a sonographer using a spline-based graphical user interface system and is then refined as the consensus of the first sonographer and a new expert. For consistency among the 3 detection tasks, we use the same training set for tuning the parameters and the same test set (no overlap with training) for evaluating the detection accuracy of each task. For this purpose, we randomly divide the CIMT videos at the patient level into training and test sets. The training set contains 48 CIMT videos of 12 patients with a total of 4456 frames; the test set contains 44 CIMT videos of 11 patients with a total of 3565 frames. For each task, we perform leave-1-patient-out cross-validation on the basis of the *training patients* to tune the parameters and then we evaluate the performance of the tuned system using the test patients.

For all 3 detection tasks, we use a widened version of LeNet [20], that is, we have used the same number of layers but employed a larger number of kernels in the convolutional layers. This is because wider CNNs tend to perform better than narrower CNNs [10]. As summarized in Table 5.1, the selected architecture has 2 convolutional

layers with rectified linear unit activation functions, 2 subsampling layers, and 2 fully connected layers. We also append a softmax layer to the last fully connected layer to generate probabilistic confidence values for each class. Since the selected architecture is designed for input patches of size 32×32, we resize the collected patches to 32×32 before the training process. For the CNNs used in our experiments, we use a learning rate of $\alpha = 0.001$, a momentum of $\mu = 0.9$, and a constant scheduling rate of $\gamma = 0.95$. For our experiments, we also use an open source GPU implementation of CNNs.[2]

5.5.1 PRE- AND POST-PROCESSING FOR FRAME SELECTION

We have experimentally determined that binarized image patches improve the quality of convergence and the accuracy of frame selection. Furthermore, we have observed that the standard deviation of the Gaussian function used for smoothing the probability signals can also substantially influence frame selection accuracy. Therefore, we have conducted leave-1-patient-out cross-validation based on the training patients to find the best binarization method and the optimal standard deviation of the Gaussian function. For binarization, we have considered a fixed set of thresholds and an adaptive thresholding scheme with Otsu's method. For smoothing, we have considered a Gaussian function with different standard deviation (σ_g), as well as the scenario where no smoothing is applied. For each configuration of parameters, we have done a free-response receiver operating characteristic (FROC) analysis. We consider a selected frame a true positive when it is found within 1 frame from the expert-annotated EUF; otherwise, it is considered a false positive.

Fig. 5.8 shows the FROC curves for the training patients using leave-1-patient-out cross-validation. As seen, no smoothing or a small degree of Gaussian smoothing leads to relatively low frame selection accuracy. This is because a trivial level of smoothing may not properly handle the fluctuations in the probability signals, causing a large number of false positives around an EUF. Yet, a large degree of smoothing may decrease the sensitivity of frame selection because the locations of the local maxima may be found more than 1 frame away from the expert-annotated EUFs. We therefore use a Gaussian function with $\sigma_g = 1.5$ for smoothing the probability signals. Our results also indicate that the adaptive thresholding method and a fixed threshold of 0.2 achieve the highest frame selection accuracy. However, we choose to use adaptive thresholding because it decreases the parameters of our system by one and that it performs more consistently at different levels of Gaussian smoothing. Fig. 5.9 shows the FROC curve of our system for the test patients, using the tuned parameters. For comparison, we also show the operating point of our prior work [40], which is outperformed by the suggested system.

5.5.2 CONSTRAINED ROI LOCALIZATION

We conduct a leave-1-patient-out cross-validation study based on the training patients to find the optimal size of the training patches. Fig. 5.10 shows the box plots of

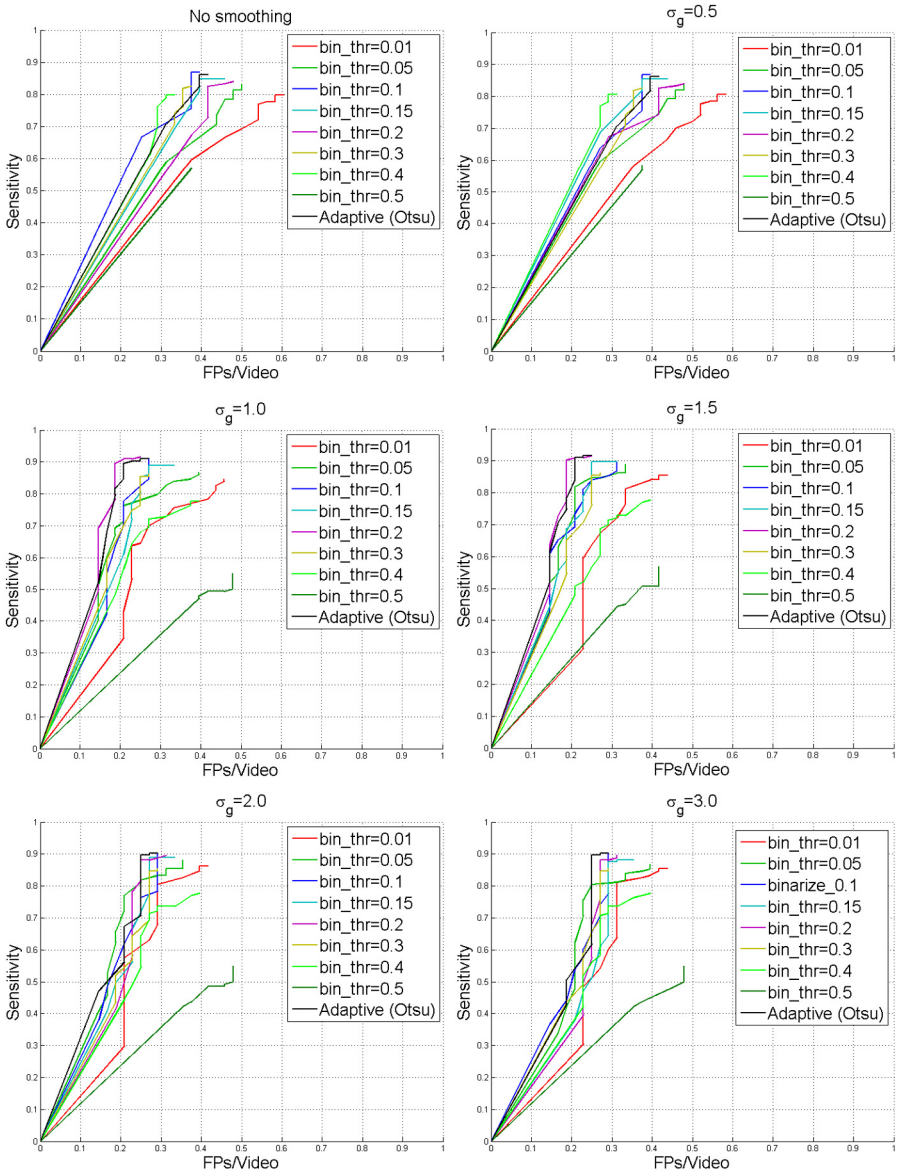

FIGURE 5.8

FROC curves of our system for automatic frame selection. Each plot shows FROC curves for different binarization thresholds and different levels of Gaussian smoothing. The results are obtained using leave-1-patient-out cross-validation based on the training patients.

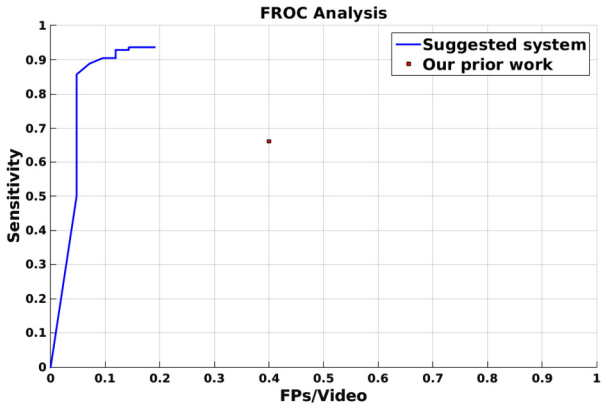

FIGURE 5.9

FROC curve of our frame selection system for the test patients with use of tuned parameters. For comparison, we also show the operating point of our prior work [40], which is outperformed by the suggested system.

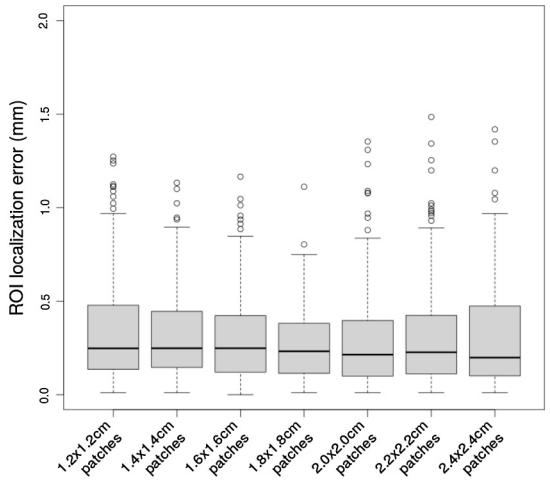

FIGURE 5.10

ROI localization error of our system for different sizes of patches. The results are obtained using leave-1-patient-out cross-validation based on the training patients.

ROI localization error for different sizes of patches. In our analyses, we measure the localization error as the Euclidean distance between the estimated ROI location and the one provided by the expert. According to our cross-validation results, the use of 1.8×1.8 cm patches achieves the most stable performance, yielding low ROI localization error with only a few outliers.

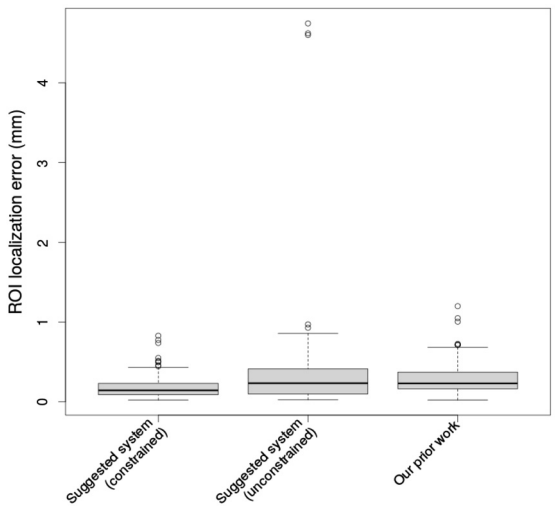

FIGURE 5.11

ROI localization error for the test patients. Our method in the constrained mode outperforms both the unconstrained counterpart and our prior work [40].

Fig. 5.11 shows the ROI localization error of our system for the test patients, using the optimal size of training patches. To demonstrate the effectiveness of our constrained ROI localization method, we include the performance of the unconstrained counterpart. In the constrained mode, we use Eq. (5.3) for ROI localization; in the unconstrained mode, we localize the ROI as the center of the largest connected component in the corresponding confidence map without considering the location of the carotid bulb. Our method achieves an average localization error of 0.19 and 0.35 mm in the constrained and unconstrained modes, respectively. The decrease in localization error is statistically significant ($p < 0.01$). Furthermore, as shown in Fig. 5.11, our method in the unconstrained mode has resulted in 3 complete localization failures (outliers), which have been corrected in the constrained mode. Compared with our prior work [40], our system in the constrained mode shows a decrease of 0.1 mm in ROI localization error, which is statistically significant ($p < 0.00001$).

5.5.3 INTIMA–MEDIA THICKNESS MEASUREMENT

We conducted a leave-1-patient-out cross-validation study based on the training patients to find the optimal size of the training patches. The results are depicted in Fig. 5.12, where each box plot shows the combined localization error of lumen–intima and media–adventitia interfaces for a different size of patches. In our analyses, we determine the localization error as the average of absolute vertical distances between our detected boundaries and the expert-annotated boundaries for the interfaces. Although our system shows a high degree of robustness against different sizes of

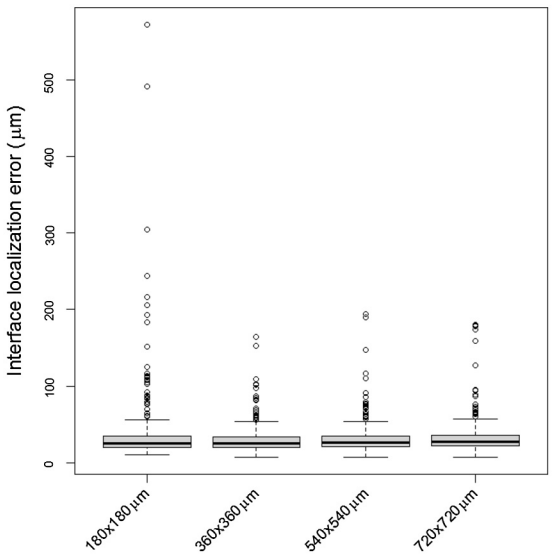

FIGURE 5.12

Combined interface localization error for different sizes of patches. The results are produced through a leave-1-patient-out cross-validation study based on the training patients.

input patches, the use of patches of size 360×360 μm achieves slightly lower localization error and fewer outliers. Furthermore, this choice of patches yields greater computational efficiency than the larger counterpart patches.

Fig. 5.13 shows the interface localization error of our system with and without snakes on the test patients. For a more accurate analysis, we break down the overall localization error into the localization error of lumen–intima interface and the localization error of the media–adventitia interface. We have also included the localization error of our prior work [40] for each of the 2 interfaces. As shown, our system yields a median error of 2.68 μm for the lumen–intima interface and a median error of 3.49 μm for the media–adventitia interface, which is significantly lower than a CNN without snakes and our prior work (paired t-test, $p < 0.0001$). The higher error for the media–adventitia interface is due to lower contrast and faint gradients around its boundary.

We further analyzed agreement between our system and the expert for the assessment of intima–media thickness. To this end, we use the Bland–Altman plot, which is a well-established technique to measure agreement among different observers. The Bland–Altman plot for the test patients is shown in Fig. 5.14, where each circle represents a pair of thickness measurements, one from our method and one from the expert. The majority of circles are within 2 standard deviations from the mean error, which suggests a large agreement between the automatically computed thickness measurements and those of the expert. Furthermore, Pearson product–moment corre-

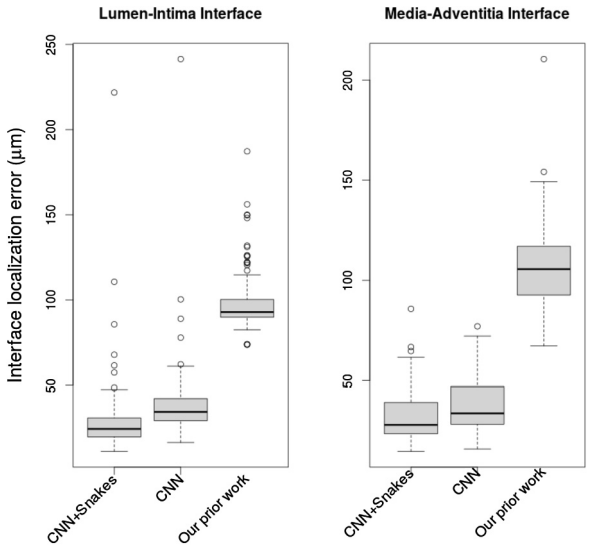

FIGURE 5.13

Localization error of the lumen–intima and media–adventitia interfaces for the suggested system (CNN and CNN+Snakes) and our prior work [40]. The results are obtained for the test patients.

lation coefficient for the average and difference measurements is -0.097, indicating that the agreement between our method and the expert does not depend on intima–media thickness.

5.5.4 END-TO-END CIMT MEASUREMENT

In the previous analyses, we evaluated each stage of the suggested system independently. Herein, we analyze the performance of our system as a whole. Specifically, the input to our system is an ultrasound video and the output is the measured CIMT as the average of measurements made in the EUFs of the input video. This analysis mimics the actual clinical scenario, permitting us to investigate how the error of each stage of the suggested system accumulates through the pipeline. For this analysis, we asked 3 new experts to measure CIMT for the 44 videos from the 11 test patients. The difference in CIMT measurements between readers 1 and 2 was 16 ± 12 μm, between readers 1 and 3 was 22 ± 15 μm, and between reader 2 and 3 was 28 ± 18 μm. We obtained an interclass correlation coefficient of 0.986 (95% CI 0.977–0.991), suggesting an excellent level of agreement among the 3 experts. The ANOVA test also yielded p-values around 0.50, suggesting that there is a lack of evidence to show a difference among the CIMT measurements made by our system and those made by the 3 experts. Our results suggest that the suggested system is robust against variability in CIMT measurements made by different experts.

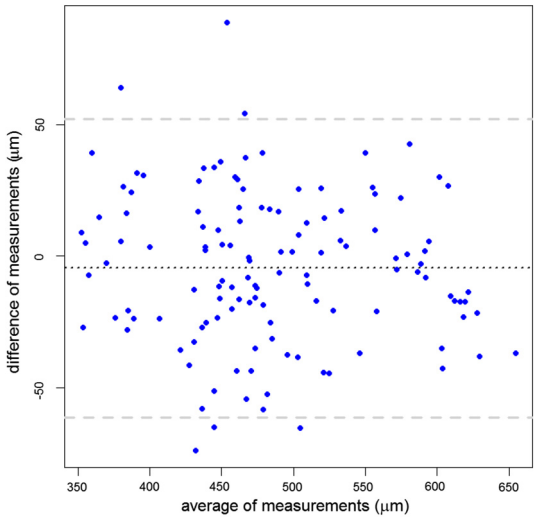

FIGURE 5.14

The Bland–Altman plot shows high agreement between our system and the expert for the assessment of intima–media thickness. Each circle in the plot represents a pair of thickness measurements from our method and the expert for a test region of interest. The plot shows a total of 126 circles corresponding to 44 test videos.

5.6 DISCUSSION

Our computer-aided measurement system consisted of 3 components, each designed according to the guidelines suggested for a quality CIMT exam [42]. Specifically, we used the ECG signal for EUF selection because the guidelines for the CIMT exams recommend the use of echocardiography signals for accurate and consistent selection of EUFs. We utilized the location of carotid bulb for more accurate ROI localization, because the guidelines recommend the measurement of CIMT approximately 1 cm from the carotid bulb. Finally, our method for CIMT measurement implicitly assumed the horizontal appearance of the common carotid artery in the ultrasound images, which was in accordance with the guidelines for a quality CIMT exam. However, our method could also accommodate a range of incidence angles by rotating the ultrasound images according to the orientation of the common carotid artery.

Throughout the experiment section, we demonstrated that our CNN-based system outperformed our previous handcrafted approach. Here, we discuss the reasons behind the observed superiority. Our previous approach relies on altitude of the ECG signal to find the R waves of the QRS complexes, but the CNN-based approach explicitly learns the appearance of the QRS complex. The former is affected by the variability of peak altitude across different vendors but the latter is robust against such variations in the displayed ECG signals. Our previous approach identifies the

bulb area (ROI) by computing curvature along the boundary of the carotid artery. However, curvature calculation requires a clean segmentation of the carotid artery, which is not a trivial task in low-contrast ultrasound images. On the other hand, the CNN-based approach explicitly learns the appearance of the ROI, eliminating the need for artery segmentation. For CIMT measurements, our previous method initializes the open snakes based on image gradient, which is often affected by spurious edges and low contrast areas in the ultrasound frames. However, our CNN-based approach initializes the snakes closer to the desired boundaries by recognizing the upper and lower interfaces of the carotid arteries.

In Section 5.5.1, we investigated how the choice of patch binarization and degree of Gaussian smoothing affect the accuracy of frame selection. Herein, we discuss our findings and provide insights about our choices. We choose to binarize the patches because the binarization reduces appearance variability and suppresses the low-magnitude noise content in the patches. Without patch binarization, a large amount of variability can be expected in the appearance of wavelets that, as shown in Fig. 5.8, can deteriorate the performance of the subsequent CNN. The choice of the binarization threshold is another important factor. The use of a high threshold leads to the partial appearance of the wavelets in the resulting binary patches, but a low threshold can intensify noise content in the images. It turns out that Otsu's adaptive threshold strikes a balance between suppressing the noise content and keeping the shapes of the restored wavelets intact in all collected patches. For patches with intensity values between 0 and 1, the adaptive thresholds have a mean of 0.15 and standard deviation of 0.05. The wide range of adaptive thresholds explains why a constant threshold may not perform as desirably.

Gaussian smoothing of the probability signals is also essential for accurate frame selection because the high-frequency fluctuations of raw probability signals may complicate the localization of EUFs. These high-frequency changes are caused when the weighted centroid deviates from the center of the restored wavelet. This type of error can manifest as a sudden change in the CNN output and, as a result, in the corresponding probability signal. The second cause of high-frequency changes is the inherited high variance of CNNs. Use of ensemble of CNNs and data augmentation can alleviate this problem at a substantial computation cost. Alternatively, we choose to mitigate these undesirable fluctuations using Gaussian smoothing, which allows for both time and computational efficiency.

As described in Section 5.4.3, we constrain our ROI localization method by the location of the carotid bulb because the bulb area appears as a relatively distinct dark area in the ultrasonographic frame. The distinct appearance of the carotid bulb is also confirmed by our experiments, where we obtain the average bulb localization error of 0.26 mm for the test patients and with only 1 failure case, which is more favorable than the average unconstrained ROI localization error of 0.38 mm with 3 failure cases. Therefore, the localization of the bulb area can be done more reliably than the localization of the ROI, which motivates the use of the bulb location as a guide for more accurate ROI localization. Alternatively, we could train a regression CNN where each pixel in the image directly votes for the location of the ROI. However, this

approach may be hindered by the lack of stable anatomical structures in ultrasonography images. We will explore the use of a regression CNN for ROI localization in our future work.

In Fig. 5.7, we presented how the use of open snakes as a post-processing scheme further improved the interface localization accuracy of the CNN. Herein, we emphasize that the two open snakes employed in our system do not see the actual images during the deformation process, but rather the probability maps that are produced by the CNN. By using snakes in the CNN probability maps, we can smooth the initial step-like boundary while maximizing the probability by which each snake point lies on the desired interface. Therefore, the snake deformation process strikes a balance between boundary smoothness and adherence to CNN classification scores. A by-product of our approach is that the snake deformation process will not be directly affected by inherently large amount of noise and artifacts that are present in ultrasound images. This indeed distinguishes our approach from the literature where the snake deformation process is mainly governed by the image content. In Fig. 5.13, we demonstrated the superiority of our hybrid CNN–Snake approach over our prior work [40], wherein image gradient information together with vertical intensity scanning are used to initialize 2 image-based open snakes.

In Section 5.5.3, we showed a high level of agreement between our system and the expert for the assessment of intima–media thickness. The suggested system achieves a mean absolute error of 23.4 μm with a standard deviation of 17.3 μm for intima–media thickness measurements. However, this level of measurement error cannot hurt the interpretation of the vascular age, because a minimum difference of 400 μm exists between the average intima–media thickness of the healthy and high-risk populations (600 μm for healthy population and $\geq$1000 μm for high-risk population) [17]. To further put the performance of our system into perspective, in Table 5.2, we show the accuracy of intima–media thickness measurements produced by our system and those of the other automatic methods recently suggested in the literature. Of note, the studies listed in Table 5.2 are evaluated using different non-public databases, and thus the tabulated results are not directly comparable.

We used a LeNet-like CNN architecture in our study, but it does not limit the suggested framework to this architecture. In fact, we have experimented with deeper CNN architectures such as AlexNet [19] in both training and fine-tuning modes; however, we did not observe a substantial performance gain. This result was probably because the higher-level semantic features detected by the deeper networks are not relevant to the tasks in our CIMT applications. Meanwhile, the concomitant computational cost of deep architectures may hinder the applicability of our system, because it lowers the speed, a key usability factor of our system. We also do not envision that a shallower architecture can offer the performance required for clinical practice. This is because a network shallower than the LeNet has only 1 convolutional layer and thus is limited to learning primitive edge-like features. Detecting the carotid bulb and the ROI and segmenting intima–media boundaries are relatively challenging tasks, requiring more than primitive edge-like features. Similarly, for frame selection, classifying the restored wavelets into R wave and non-R wave categories is similar to digit

Table 5.2 CIMT error ($\mu \pm \sigma$) for our system and the other state-of-the-art methods. Note that the listed studies are evaluated using different non-public databases, and thus the tabulated results are not directly comparable

Author	Thickness error (μm)	Image size	Scan resolution (mm/pixel)	Cohort
Current work	23.4 ± 17.3	640×480	0.09	23 asymptomatic patients
Bastida [4]	13.8 ± 31.9	600×800	0.029 to 0.081	27 mostly healthy patients
Ilea [16]	80 ± 40	600×800	0.029 to 0.095	23 asymptomatic patients
Loizou [27]	30 ± 30	576×768	0.060	20 symptomatic patients
Molinari [33]	43 ± 93	576×768	0.060	150 asymptomatic patients

recognition, for which LeNet is a common choice of architecture. Therefore, LeNet-like CNN architecture seems to represent an optimal balance between efficiency and accuracy for the CIMT video interpretation.

On a desktop computer with a 3.6-GHz quad core Intel and a GTX 970 GPU, our system detects each EUF in 2.9 seconds, localizes each ROI in 12.1 seconds, and measures intima–media thickness in 8.2 seconds. Although the current speed is suitable for offline processing of the CIMT videos, further performance speedup is required for an interactive use in clinical practice. We note that use of CNNs does not hinder the interactive use of our system; rather, extracting a large number of patches from a dense set of locations in the ultrasonography images causes a computational bottleneck. Therefore, substantial performance speedup can be achieved by using fully convolutional networks [28], which eliminate the need for computationally expensive image patch extraction. Further performance speedup can also be obtained using more dedicated graphics cards.

We also note that throughout this chapter, all performance evaluations were performed without involving any user interactions. However, our goal is not to exclude the user (sonographer) from the loop, but rather to relieve the sonographer from the 3 tedious, laborious, and time-consuming operations. We accomplish this relief by automating the operations while still offering a highly user-friendly interface to bring the sonographer's indispensable expertise onto CIMT interpretation through refining the automatic results easily at the end of each of the automated operations. For instance, the user can simply press an arrow key to move 1 frame forward or backward in case the EUF is mislocalized. From our experience, the automatically localized ROI is acceptable even with a small distance from the ground truth location, but the user still can easily drag the ROI and move it around as desired. Finally, the user can adjust the CIMT measurement by changing "movable" hard constraints [23] on the snakes.

5.7 CONCLUSION

In this study, we presented a computer-aided measurement system to fully automate and accelerate CIMT video interpretation. Specifically, we suggested a computer-aided CIMT measurement system with 3 components: (i) automatic frame selection in CIMT videos, (ii) automatic ROI localization within the selected frames, and (iii) automatic intima–media boundary segmentation within the localized ROIs. We based each of these components on a CNN with a LeNet-like architecture and then boosted the performance of the CNNs with effective pre- and post-processing techniques. For frame selection, we demonstrated how patch binarization as a pre-processing step and smoothing the probability signals as a post-processing step improve the results generated by the CNN. For ROI localization, we experimentally proved that the location of the carotid bulb as a constraint in a post-processing setting, substantially improves ROI localization accuracy. For intima–media boundary segmentation, we used open snakes as a post-processing step to further improve the segmentation accuracy. We compared the results produced by the suggested system with those of our prior work, demonstrating more accurate frame selection, ROI localization, and CIMT measurements. This superior performance is attributed to the effective use of CNNs coupled with pre- and post-processing steps, uniquely designed for each of these 3 tasks.

ACKNOWLEDGEMENT

This research has been partially supported by the Mayo Clinic Discovery Translation Program.

REFERENCES

1. World Health Organization, Global atlas on cardiovascular disease prevention and control, September 19, 2011, available at www.who.int/cardiovascular_diseases/publications.
2. Shun-ichi Amari, Backpropagation and stochastic gradient descent method, Neurocomputing 5 (4–5) (1993) 185–196.
3. M. Consuelo Bastida-Jumilla, Rosa M. Menchón-Lara, Juan Morales-Sánchez, Rafael Verdú-Monedero, Jorge Larrey-Ruiz, José Luis Sancho-Gómez, Segmentation of the common carotid artery walls based on a frequency implementation of active contours, J. Digit. Imaging 26 (1) (2013) 129–139.
4. M.C. Bastida-Jumilla, R.M. Menchón-Lara, J. Morales-Sánchez, R. Verdú-Monedero, J. Larrey-Ruiz, J.L. Sancho-Gómez, Frequency-domain active contours solution to evaluate intima–media thickness of the common carotid artery, Biomed. Signal Process. Control 16 (2015) 68–79.
5. Hao Chen, Qi Dou, Dong Ni, Jie-Zhi Cheng, Jing Qin, Shengli Li, Pheng-Ann Heng, Automatic fetal ultrasound standard plane detection using knowledge transferred recurrent neural networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 507–514.

6. Da-Chuan Cheng, Xiaoyi Jiang, Detections of arterial wall in sonographic artery images using dual dynamic programming, IEEE Trans. Inf. Technol. Biomed. 12 (6) (2008) 792–799.
7. Silvia Delsanto, Filippo Molinari, Pierangela Giustetto, William Liboni, Sergio Badalamenti, Jasjit S. Suri, Characterization of a completely user-independent algorithm for carotid artery segmentation in 2-D ultrasound images, IEEE Trans. Instrum. Meas. 56 (4) (2007) 1265–1274.
8. Q. Dou, H. Chen, L. Yu, L. Zhao, J. Qin, D. Wang, V.C. Mok, L. Shi, P.A. Heng, Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks, IEEE Trans. Med. Imaging 35 (5) (2016) 1182–1195.
9. Ralph B. D'Agostino, Ramachandran S. Vasan, Michael J. Pencina, Philip A. Wolf, Mark Cobain, Joseph M. Massaro, William B. Kannel, General cardiovascular risk profile for use in primary care the Framingham heart study, Circulation 117 (6) (2008) 743–753.
10. David Eigen, Jason Rolfe, Rob Fergus, Yann LeCun, Understanding deep architectures using a recursive convolutional network, arXiv:1312.1847, 2013.
11. Francesco Faita, Vincenzo Gemignani, Elisabetta Bianchini, Chiara Giannarelli, Lorenzo Ghiadoni, Marcello Demi, Real-time measurement system for evaluation of the carotid intima–media thickness with a robust edge operator, J. Ultrasound Med. 27 (9) (2008) 1353–1361.
12. Ian J. Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron Courville, Yoshua Bengio, Maxout networks, arXiv:1302.4389, 2013.
13. Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, Ruslan R. Salakhutdinov, Improving neural networks by preventing co-adaptation of feature detectors, arXiv:1207.0580, 2012.
14. David H. Hubel, Torsten N. Wiesel, Receptive fields of single neurones in the cat's striate cortex, J. Physiol. 148 (3) (1959) 574–591.
15. R. Todd Hurst, Robert F. Burke, Erik Wissner, Arthur Roberts, Christopher B. Kendall, Steven J. Lester, Virend Somers, Martin E. Goldman, Qing Wu, Bijoy Khandheria, Incidence of subclinical atherosclerosis as a marker of cardiovascular risk in retired professional football players, Am. J. Cardiol. 105 (8) (2010) 1107–1111.
16. E. Dana Ilea, Caoimhe Duffy, Liam Kavanagh, Alice Stanton, Paul F. Whelan, Fully automated segmentation and tracking of the intima media thickness in ultrasound video sequences of the common carotid artery, IEEE Trans. Ultrason. Ferroelectr. Freq. Control 60 (1) (2013).
17. Douglas S. Jacoby, Emile R. Mohler III, Daniel J. Rader, Noninvasive atherosclerosis imaging for predicting cardiovascular events and assessing therapeutic interventions, Curr. Atheroscler. Rep. 6 (1) (2004) 20–26, http://dx.doi.org/10.1007/s11883-004-0112-8.
18. Michael Kass, Andrew Witkin, Demetri Terzopoulos, Snakes: active contour models, Int. J. Comput. Vis. 1 (4) (1988) 321–331.
19. Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, ImageNet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.
20. B. Boser, Y. LeCun, John S. Denker, D. Henderson, Richard E. Howard, W. Hubbard, Lawrence D. Jackel, Handwritten digit recognition with a back-propagation network, in: Advances in Neural Information Processing Systems, 1990.
21. Yann LeCun, Bernhard Boser, John S. Denker, Donnie Henderson, Richard E. Howard, Wayne Hubbard, Lawrence D. Jackel, Backpropagation applied to handwritten zip code recognition, Neural Comput. 1 (4) (1989) 541–551.

22. Yann LeCun, Yoshua Bengio, Geoffrey Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
23. Jianming Liang, Tim McInerney, Demetri Terzopoulos, United snakes, Med. Image Anal. 10 (2) (2006) 215–233.
24. Quan Liang, Inger Wendelhag, John Wikstrand, Tomas Gustavsson, A multiscale dynamic programming procedure for boundary detection in ultrasonic artery images, IEEE Trans. Med. Imaging 19 (2) (2000) 127–142.
25. Christos P. Loizou, A review of ultrasound common carotid artery image and video segmentation techniques, Med. Biol. Eng. Comput. 52 (12) (2014) 1073–1093.
26. Christos P. Loizou, Constantinos S. Pattichis, Marios Pantziaris, Andrew Nicolaides, An integrated system for the segmentation of atherosclerotic carotid plaque, IEEE Trans. Inf. Technol. Biomed. 11 (6) (2007) 661–667.
27. Christos P. Loizou, Takis Kasparis, Christina Spyrou, Marios Pantziaris, Integrated system for the complete segmentation of the common carotid artery bifurcation in ultrasound images, in: Harris Papadopoulos, Andreas S. Andreou, Lazaros Iliadis, Ilias Maglogiannis (Eds.), Artificial Intelligence Applications and Innovations, in: IFIP Adv. Inf. Commun. Technol., vol. 412, Springer, Berlin, Heidelberg, ISBN 978-3-642-41141-0, 2013, pp. 292–301.
28. Jonathan Long, Evan Shelhamer, Trevor Darrell, Fully convolutional networks for semantic segmentation, arXiv:1411.4038, 2014.
29. Kristen M. Meiburger, Filippo Molinari, Justin Wong, Luis Aguilar, Diego Gallo, David A. Steinman, Umberto Morbiducci, Validation of the carotid intima–media thickness variability: can manual segmentations be trusted as ground truth?, Ultrasound Med. Biol. 42 (7) (2016) 1598–1611.
30. Rosa-María Menchón-Lara, José-Luis Sancho-Gómez, Fully automatic segmentation of ultrasound common carotid artery images based on machine learning, Neurocomputing 151 (2015) 161–167.
31. Rosa-María Menchón-Lara, María-Consuelo Bastida-Jumilla, Antonio González-López, José Luis Sancho-Gómez, Automatic evaluation of carotid intima–media thickness in ultrasounds using machine learning, in: Natural and Artificial Computation in Engineering and Medical Applications, Springer, 2013, pp. 241–249.
32. Filippo Molinari, Guang Zeng, Jasjit S. Suri, A state of the art review on intima–media thickness (IMT) measurement and wall segmentation techniques for carotid ultrasound, Comput. Methods Programs Biomed. 100 (3) (2010) 201–221.
33. Filippo Molinari, Kristen M. Meiburger, Guang Zeng, Andrew Nicolaides, Jasjit S. Suri, CAUDLES-EF: carotid automated ultrasound double line extraction system using edge flow, J. Digit. Imaging 24 (6) (2011) 1059–1077, http://dx.doi.org/10.1007/s10278-011-9375-0.
34. Filippo Molinari, Kristen M. Meiburger, Luca Saba, Guang Zeng, U. Rajendra Acharya, Mario Ledda, Andrew Nicolaides, Jasjit S. Suri, Fully automated dual-snake formulation for carotid intima–media thickness measurement a new approach, J. Ultrasound Med. 31 (7) (2012) 1123–1136.
35. Dariush Mozaffarian, Emelia J. Benjamin, Alan S. Go, Donna K. Arnett, Michael J. Blaha, Mary Cushman, Sarah de Ferranti, Jean-Pierre Despres, Heather J. Fullerton, Virginia J. Howard, et al., Heart disease and stroke statistics – 2015 update: a report from the American Heart Association, Circulation 131 (4) (2015) e29.
36. Nobuyuki Otsu, A threshold selection method from gray-level histograms, Automatica 11 (285–296) (1975) 23–27.

37. Styliani Petroudi, Christos Loizou, Marios Pantziaris, Constantinos Pattichis, Segmentation of the common carotid intima–media complex in ultrasound images using active contours, IEEE Trans. Biomed. Eng. 59 (11) (2012) 3060–3069.
38. P. Pignoli, T. Longo, Evaluation of atherosclerosis with b-mode ultrasound imaging, J. Nucl. Med. Allied Sci. 32 (3) (1987) 166–173.
39. Alessandro C. Rossi, Peter J. Brands, Arnold P.G. Hoeks, Automatic localization of intimal and adventitial carotid artery layers with noninvasive ultrasound: a novel algorithm providing scan quality control, Ultrasound Med. Biol. 36 (3) (2010) 467–479.
40. Haripriya Sharma, Ramsri G. Golla, Yu Zhang, Christopher B. Kendall, R. Todd Hurst, Nima Tajbakhsh, Jianming Liang, ECG-based frame selection and curvature-based ROI detection for measuring carotid intima–media thickness, in: SPIE Medical Imaging, International Society for Optics and Photonics, 2014, 904016.
41. Hoo-Chang Shin, Holger R. Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, Ronald M. Summers, Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning, IEEE Trans. Med. Imaging 35 (5) (2016) 1285–1298.
42. J.H. Stein, C.E. Korcarz, R.T. Hurst, E. Lonn, C.B. Kendall, E.R. Mohler, S.S. Najjar, C.M. Rembold, W.S. Post, American Society of Echocardiography carotid intima–media thickness task force. Use of carotid ultrasound to identify subclinical vascular disease and evaluate cardiovascular disease risk: a consensus statement from the American Society of Echocardiography carotid intima–media thickness task force. Endorsed by the Society for Vascular Medicine, J. Am. Soc. Echocardiogr. 21 (2) (2008) 93–111.
43. Nima Tajbakhsh, Jae Y. Shin, Suryakanth R. Gurudu, R. Todd Hurst, Christopher B. Kendall, Michael B. Gotway, Jianming Liang, Convolutional neural networks for medical image analysis: full training or fine tuning?, IEEE Trans. Med. Imaging 35 (5) (2016) 1299–1312.
44. Pierre-Jean Touboul, Patrizio Prati, Pierre-Yves Scarabin, Valérie Adrai, Emmanuel Thibout, Pierre Ducimetière, Use of monitoring software to improve the measurement of carotid wall thickness by b-mode imaging, J. Hypertens. 10 (1992) S37–S42.
45. Li Wan, Matthew Zeiler, Sixin Zhang, Yann L. Cun, Rob Fergus, Regularization of neural networks using DropConnect, in: Proceedings of the 30th International Conference on Machine Learning (ICML-13), 2013, pp. 1058–1066.
46. Jun Xu, Xiaofei Luo, Guanhao Wang, Hannah Gilmore, Anant Madabhushi, A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images, Neurocomputing 191 (2016) 214–223.
47. Xiangyang Xu, Yuan Zhou, Xinyao Cheng, Enmin Song, Guokuan Li, Ultrasound intima–media segmentation using hough transform and dual snake model, Comput. Med. Imaging Graph. 36 (3) (2012) 248–258.
48. Xiangjun Zhu, Christopher B. Kendall, R. Todd Hurst, Jianming Liang, A user friendly system for ultrasound carotid intima–media thickness image interpretation, in: SPIE Medical Imaging, International Society for Optics and Photonics, 2011, 79681G.

NOTES

1. This should not to be confused with the wavelet transform in signal processing. A wavelet, in this context, refers to a small part of the ECG signal.
2. https://github.com/sdemyanov/ConvNet.