Coronary calcification segmentation in intravascular OCT images using deep learning: application to calcification scoring

Yazan Gharaibeh
David Prabhu
Chaitanya Kolluru
Juhwan Lee
Vladislav Zimin
Hiram Bezerra
David Wilson

SPIE.

Coronary calcification segmentation in intravascular OCT images using deep learning: application to calcification scoring

Yazan Gharaibeh,[a] David Prabhu,[a] Chaitanya Kolluru,[a] Juhwan Lee,[a] Vladislav Zimin,[b] Hiram Bezerra,[b] and David Wilson[a,c,*]

[a]Case Western Reserve University, Department of Biomedical Engineering, Cleveland, Ohio, United States
[b]University Hospitals Cleveland Medical Center, Harrington Heart and Vascular Institute, Cardiovascular Imaging Core Laboratory, Cleveland, Ohio, United States
[c]Case Western Reserve University, Department of Radiology, Cleveland, Ohio, United States

Abstract. Major calcifications are of great concern when performing percutaneous coronary interventions because they inhibit proper stent deployment. We created a comprehensive software to segment calcifications in intravascular optical coherence tomography (IVOCT) images and to calculate their impact using the stent-deployment calcification score, as reported by Fujino et al. We segmented the vascular lumen and calcifications using the pretrained SegNet, convolutional neural network, which was refined for our task. We cleaned segmentation results using conditional random field processing. We evaluated the method on manually annotated IVOCT volumes of interest (VOIs) without lesions and with calcifications, lipidous, or mixed lesions. The dataset included 48 VOIs taken from 34 clinical pullbacks, giving a total of 2640 *in vivo* images. Annotations were determined from consensus between two expert analysts. Keeping VOIs intact, we performed 10-fold cross-validation over all data. Following segmentation noise cleaning, we obtained sensitivities of 0.85 ± 0.04, 0.99 ± 0.01, and 0.97 ± 0.01 for calcified, lumen, and other tissue classes, respectively. From segmented regions, we automatically determined calcification depth, angle, and thickness attributes. Bland–Altman analysis suggested strong correlation between manually and automatically obtained lumen and calcification attributes. Agreement between manually and automatically obtained stent-deployment calcification scores was good (four of five lesions gave exact agreement). Results are encouraging and suggest our classification approach could be applied clinically for assessment and treatment planning of coronary calcification lesions. © *The Authors. Published by SPIE under a Creative Commons Attribution 4.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.* [DOI: 10.1117/1.JMI.6.4.045002]

1 Introduction

Major calcifications are of great concern when performing percutaneous coronary intervention (PCI) because they can hinder stent deployment. Approximately 700,000 PCIs are performed each year, and many involve the use of stents to open up obstructed coronary arteries.[1] Calcified plaques are found in 17% to 35% of patients undergoing PCI.[2–4] Calcifications can lead to stent underexpansion and strut malapposition, which in turn can lead to increased risk of thrombosis and in-stent restenosis.[5–10] A cardiologist has several options when confronting a calcified lesion: high balloon pressures (up to 30 atm) to fracture the calcification, scoring balloon, Shockwave™ intravascular lithotripsy (IVL), rotational atherectomy, etc. In some cases, the lesion may not be treatable.

Intravascular optical coherence tomography (IVOCT) has significant advantages for characterizing coronary calcification, as compared to other imaging modalities commonly used by interventional cardiologists. Although clinicians routinely use x-ray angiography for treatment planning to describe the vessel lumen, angiography does not provide specific information regarding vascular wall composition except in the case of severely calcified lesions.[11] Intravascular ultrasound (IVUS) can identify the location of coronary calcification, but cannot assess the thickness because the radiofrequency signal is reflected from the calcium tissue interface giving an acoustic shadow.[12] IVOCT, however, provides the location and often the thickness of a calcification.[13] IVUS has better penetration depth (IVUS: 5 to 10 mm; IVOCT: 1 to 2 mm)[14,15] and does not require blood clearing for imaging. However, IVOCT has superior resolution (axial: 15 to 20 μm; lateral: 20 to 40 μm) as compared to IVUS (axial: 150 to 200 μm; lateral: 200 to 300 μm).[16,17]

Currently, the need for specialized training, uncertain interpretation, and image overload (>500 images in a pullback) have suggested a need for automated analysis of IVOCT images. There are multiple reports of automated IVOCT image analysis. Ughi et al.[18] applied machine learning to perform pixelwise classification of fibrous, lipid, and calcified plaque. Athanasiou et al.[19] segmented calcification and then classified lipid, fibrous, and mixed tissues using 17 features with k-means and postanalysis. Zhou et al.[20] developed a classification and segmentation method using texture features described by the Fourier transform and discrete wavelet transform to classify adventitia, calcification, lipid, and mixed tissue. Our group developed machine learning[21] and deep learning[22,23] methods to automatically classify plaque regions. Rico-Jimenez et al.[24] used linear

*Address all correspondence to David Wilson, E-mail: david.wilson@cwru.edu

discriminant analysis to identify normal and fibrolipidic A-lines. Yong et al.[25] proposed a linear regression convolutional neural network to automatically segment the vessel lumen. Abdolmanafi et al.[26] used deep learning to identify layers within the coronary artery wall and to identify Kawasaki disease.[27] Recently, Gessert et al.[28] used convolutional neural networks to identify IVOCT frames that contain plaque.

Although all of the aforementioned studies were promising, some limitations exist. (1) Many studies have a limited number of images, limiting the ability to generalize. (2) Experimental design for some studies used the same lesion for training, validation, and testing. This will cause the model to overfit. (3) One study did only lumen segmentation without any plaque characterization. (4) Many studies did slice-level or region of interest classification. It was unclear how this information could be used clinically. In our study, we did pixelwise segmentation and used the results to calculate the stent deployment calcification score that defines lesions that would benefit from plaque modification prior to stent implantation. (5) It is unclear if that all reports use a sufficiently large base of support (receptive field) in the image to capture *a priori* knowledge of calcified plaque distribution [e.g., calcified lesions have an "orientation" roughly parallel to the lumen in the (r, θ) representation].

In this paper, we focus on the important problem of segmenting calcifications in IVOCT images and assessing their impact on stent deployment. We build on previous studies and use deep learning to perform semantic segmentation of the lumen and calcification within IVOCT images. We use a large manually segmented training set with voxels labeled as lumen, calcification, and other. We use conditional random fields (CRF) to clean noisy segmentation results. Rather than simply reporting DICE or voxel sensitivity/specificity, as done in most previous publications, we report comparisons of automated versus manual assessments of clinically relevant calcification attributes. These include calcification depth, angle, and thickness. In addition,

to assess calcification impact on stent deployment, we evaluated a previously reported, stent deployment calcification score,[13] as computed from our automatically segmented calcifications. To our knowledge, this is the first publication focusing on segmentation and on clinically important analyses of calcified plaques.

2 Image Processing and Analysis

2.1 *Preprocessing and Data Sets Augmentation*

Preprocessing steps are applied to the raw IVOCT images obtained in the polar (r, θ) domain. Data values are log transformed to convert multiplicative speckle noise into an additive form. Image speckle noise is reduced by filtering with a normalized Gaussian kernel (standard deviation 2.5 pixels in a 7×7 footprint).[18] Optionally, IVOCT (r, θ) images are scan converted to create (x, y) images. We evaluate both (r, θ) and (x, y) data representations for segmentation of IVOCT data. Images in the (r, θ) representation are 960×480 pixels (5.2 μm by 0.75 deg). For (x, y) representations, images were 700×700 pixels (14.3 μm).

During training, data are augmented to provide more examples and to change locations of calcifications so as to improve spatial invariance of methods. For anatomical (x, y) images, we rotate the images with an angle picked randomly between -180 deg to $+180$ deg. To augment (r, θ) data, we concatenate all the (r, θ) images to form one large 2-D array, where the r direction corresponds to tissue depth and the θ corresponds to catheter rotation, which rotates from 0 deg to 360 deg for each image. By changing an offset angular shift, we can resample new 360 deg (r, θ) images. In practice, we shifted the starting A-line five times by increments of 100 A-lines. Data augmentation steps for the (r, θ) representations are shown in Fig. 1. Note that all images in this report are shown after log conversion for improved visualization.
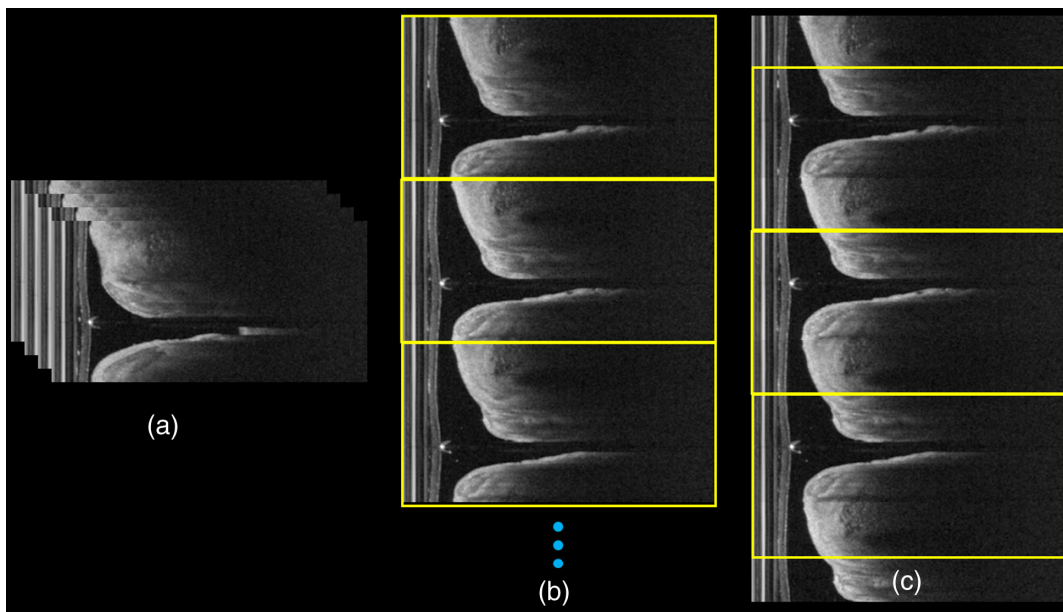

Fig. 1 Augmentation of IVOCT data. (a) The original spiral data set is rather arbitrarily split into (r, θ) image frames ($W \times H$: 960×480 pixels). (b) We concatenated images to form one large 2-D array. (c) Following an offset of pixel rows (e.g., 100 rows as shown here), we extract new image frames. Tissue structures will now appear in different portions of the image in these augmented images, reducing any dependence of θ location in the training set.

2.2 Deep Learning Model Architecture and Implementation Details

We choose SegNet[29] as our network architecture (Fig. 2). SegNet is an end-to-end hour-glass-shaped encoder–decoder convolutional neural network, which was pretrained on the CamVid dataset.[30] Each encoder/decoder convolution set consists of a convolution layer, a batch normalization layer,[31] and a rectified linear unit (ReLU) layer.[32] All convolution layers were set to have the following hyperparameters: filter size of 3, a stride of 1, and zero padding of size 1. These parameters were empirically selected using onefold of our training data as described in Sec. 3.2. This filter size was chosen to detect small features, including the edges of calcified plaques. The depth of the network was 5. In our implementation, we performed transfer learning with weighted initialized using VGG-16.

The base of support (or receptive field) for each layer can be found as

$$r_{out} = r_{in} + (k - 1) * j_{in}, \tag{1}$$

where r_{out} is the receptive field size for the current layer; r_{in} is the receptive field size for the previous layer; k is the convolution kernel size; and j_{in} is the jump, or distance between two consecutive features. The receptive field size for the deepest layer was 212×212.[33,34]

We process the data by using a batch size of 2. We implement batch normalization layer to normalize each input channel across a minibatch. This is done as

$$x_{new} = \frac{x - \mu}{\sqrt{\sigma^2 + \varepsilon}}, \tag{2}$$

where x is the input, μ is the mean, σ^2 is the variance, and ε corresponds to Epsilon. The use of Epsilon improves numerical stability when the minibatch variance is very small. The batch normalization layer further shifts and scales the activations as

$$y = \alpha x_{new} + \beta, \tag{3}$$

where the offset β and scale factor α are learnable parameters that are updated during network training. This shifting and scaling of the activations are done to account for the possibility that inputs with zero mean and unit variance are not optimal for the layer that follows the batch normalization layer.[31]

Finally, in our implementation, convolutional and batch normalization layers are followed by a ReLU and a max pooling layer. A ReLU layer performs a threshold operation to each element, where any input value less than zero is set to zero[32]

$$f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}. \tag{4}$$

A max pooling layer is inserted at the end of each encoder step.[35] All max pooling layers had a pool size of 2 pixels and stride of 2 pixels. Max pooling channels transfer the maximum responses and their indices from the encoder to the decoder to identify corresponding locations when upsampling. The model produces pixelwise probability scores for each class label (lumen, calcification, or other) with the same size and resolution as the input image.

2.3 Segmentation Refinement Strategy

We use CRF as a postprocessing step to refine the results from the deep learning model. A method to integrate network outputs to a fully connected CRF is described previously.[36] The deep learning model gives a score (vector of class probabilities) at each pixel. The CRF uses these values, pixel intensities, and corresponding spatial location information to generate crisp class labels. This process results in images with reduced noise as compared to simply performing a classwise median filter operation over the image. The goal is to reduce noise by generating a new labeling that favors assigning the same label to pixels that are closer to each other spatially using the scores generated by the neural network. For IVOCT images, the appearance kernel is inspired by the observation that nearby pixels with similar intensity are likely to be in the same class.

Overall, for each pixel, the CRF takes in probability estimates of each class, and the image pixel intensity, as input and outputs its final class ownership. Similar processing was performed when network training experiments were performed on the (r, θ) images as well. Details of this implementation are described in A.2 in the Supplementary Material.
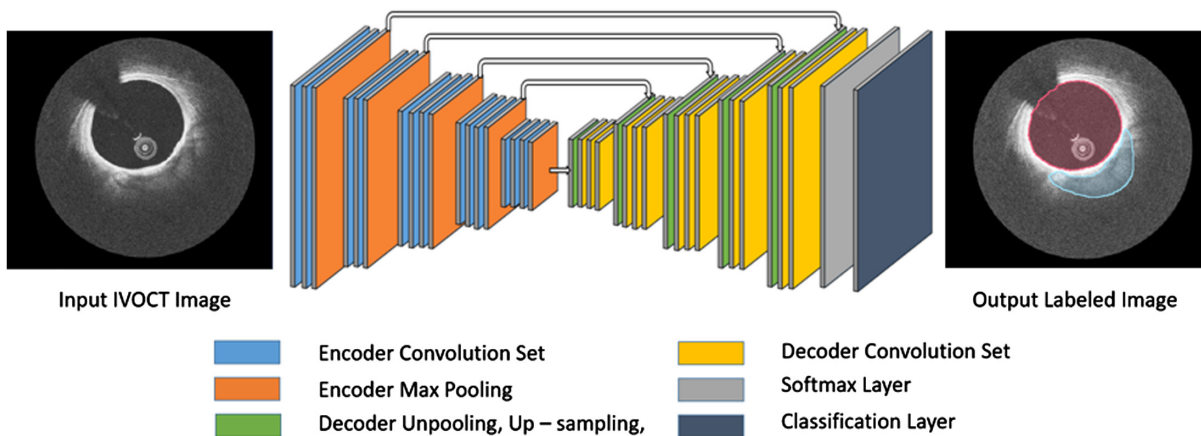

Fig. 2 Deep learning convolution neural network for semantic segmentation. Each convolution set consists of a convolution layer, batch normalization layer, and rectification layer. The arrows between the encoder and decoder layer are the pool indices channels. In the output labeled image, the shaded red area is the lumen and the blue one is the calcified plaque.

2.4 Computation of Calcification Attributes and Stent Deployment Calcification Score

We followed methods described previously[37,38] to calculate plaque average thickness, average depth, and angle automatically. Figure 3 summarizes the method of calcified plaque quantification. First, the centroid of the lumen was determined (indicated by O). Next, rays were computed, which initiate from the centroid of the lumen and traverse to the back edge of the calcification border. The average depth and thickness of the calcification are defined using the following equations:

$$\text{Depth} = \frac{1}{n} \sum_{i}^{n} D_i, \tag{5}$$

$$\text{Thickness} = \frac{1}{n} \sum_{i}^{n} T_i, \tag{6}$$

where n is the maximum number of nonoverlapping rays radiating from O spanning across the calcification. In our implementation, we used 360 rays, which were evenly spaced every 1 deg. Calcification arc is the angle between the rays at the boundary of the calcification. Plaque length is the total length (number of frames × frame interval) over which the calcification spans.

We used the method described by Fujino et al.[13] for determining the stent deployment calcification score. The idea of calcification scoring is to define lesions that would benefit from plaque modification prior to stent implantation. The method is a cumulative score based on calcification: length, maximum angle, and maximum thickness. As quoting from their manuscript: "we assigned 1 or 2 points to each of three conditions: 2 points for maximum calcium angle >180°, 1 point for maximum calcium thickness >0.5 mm, and 1 point for calcium length >5 mm."[13] In their study, they found that lesions with a

calcification score of 0 to 3 had "adequate stent expansion," whereas lesions with a score of 4 had "poor stent expansion."

3 Experimental Methods

3.1 Datasets and Labeling

The dataset included 48 VOIs taken from 34 clinical pullbacks, giving a total of 2640 *in vivo* images. The average number of images per VOI is 55 images. *In vivo* IVOCT pullbacks were obtained from the University Hospitals Cleveland Medical Center (UHCMC) imaging library.[39] The dataset has calcification lesions, lipidous lesions, and mixed lesions with both calcification and lipidous regions, sometimes in the same image. In addition, VOIs not containing a calcification were also included in the dataset. All pullbacks were imaged prior to any stent implantation.

The *in vivo* IVOCT images were acquired using a frequency-domain OCT system using Illumien Optis (St. Jude Medical, St. Paul, Minnesota). The system comprises of a tunable laser light source sweeping from 1250 to 1360 nm. The system was operated at a frame rate of 180 fps, at a pullback speed of 36 mm/s, and has an axial resolution around 20 μm. The pullbacks were analyzed by two expert readers in the Cartesian (x, y) view. Labels from (x, y) images were converted back to the polar (r, θ) system for polar data set training.

The two expert readers manually labeled the VOIs using definitions given in the consensus document.[37] Labels required consensus between the two readers. Calcifications are seen as a signal poor regions with sharply delineated front and/or back borders in IVOCT images. When a calcification was extremely thick and its back border was not clear due to attenuation, the maximum thickness was limited to 1 mm. An additional class "other" was used to include all pixels which could not be labeled into lumen or calcified plaque.
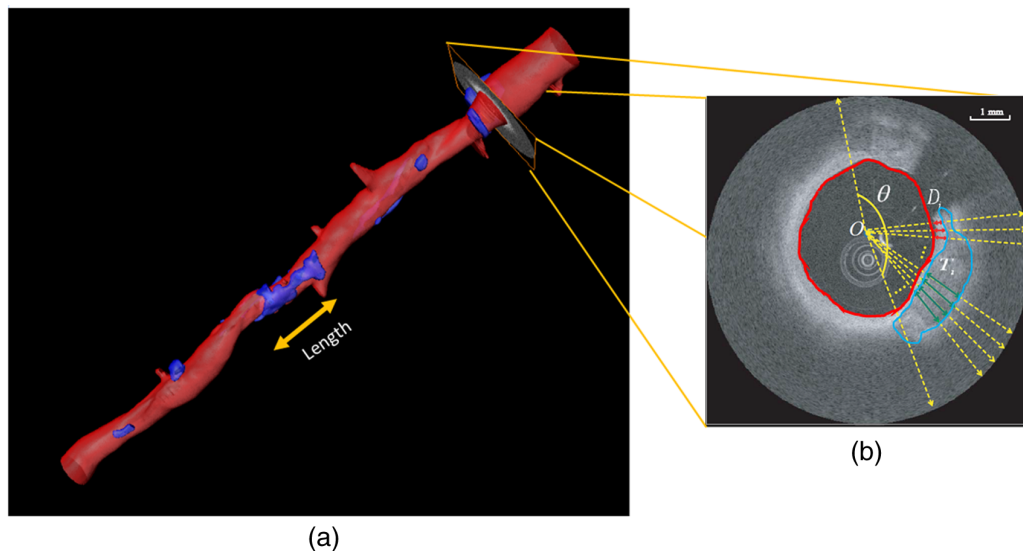

(a)

(b)

Fig. 3 Calcifications and their quantification. The 3-D rendering (a) includes multiple calcifications in blue. In an image slice (b), the calcification is tinted blue, radial lines from the lumen centroid (*O*) are shown as a function of angle (*θ*), and calcification thickness (*T*) and depth (*D*) are shown. The calcification arc is the angle between the rays at the boundary of the calcification. To compute the IVOCT-based calcification score for a specific lesion, we used three attributes: (1) maximum calcification length, (2) maximum thickness, and (3) maximum calcification angle. See text for full details on calculating the score (Sec. 2.4).

3.2 Network Training and Optimization

Our data was split into training, validation, and test, where VOIs were kept intact within a group. A tenfold cross-validation procedure was used to measure classifier performance and variation across data samples. For each fold, we assigned roughly 80% of the VOIs for training, 10% for validation (used to determine stopping criteria for training), and 10% for held out testing. The VOIs were rotated until all VOIs were in the test set once. Mean and standard error of sensitivities over the tenfolds are determined. As classes are not balanced regarding numbers of pixels, we use class weighting, as described by Eigen and Fergus.[40] Details of this are described in A.1 in the Supplementary Material.

There are several issues associated with training. We optimize the categorical cross entropy error using the Adam optimizer[41] with weight decay of 10^{-3}. We avoid overfitting by adding a regularization term for the weights to the loss function. Training is stopped when the loss on the validation dataset does not improve by more than 0.01% for 10 consecutive epochs or when the network is trained for 120 epochs. In practice, the maximum number of epochs was rarely reached.

3.3 Software Implementation

Image preprocessing and deep learning models are implemented using MATLAB 2017b (MathWorks Inc., Natick, Massachusetts) environment. The execution of the network is performed on a Linux-based Intel Xeon Processors x86_64 (x86_64 indicates Intel Xeon 64-bit platform; architecture based on Intel 8086 CPU) with a CUDA-capable NVIDIA™ Tesla P100 16GB GPU.

4 Results

We now describe semantic segmentation results. In Fig. 4, segmentation of lumen and calcification are shown prior to CRF refinement. Both lumen and calcification regions show good agreement with GR labels. In Table 1, we compare segmentation performance when using the same labeled data arranged in (x, y) and (r, θ). Segmentation on the (r, θ) representation gave superior performance for all classes. Therefore, all figures and all remaining analyses are done using the (r, θ) data representation. We simply map results to (x, y) for easier visual interpretation. We found that refinement of segmentation results using CRF was a desirable step (Fig. 5). Deep learning segmentation after noise cleaning gave visually more accurate results in all test cases and enhanced performance (Table 2).

We determined that lumen segmentation via deep learning was superior to our earlier dynamic programming lumen segmentation approach.[38] Using the Wilcoxson signed-rank test, we determined statistically significant differences ($p < 0.05$) between the two methods. Some clear instances of improvement are shown in Fig. 6. In particular, the dynamic programming approach can fail in the presence of thrombus or very eccentric lumens.

We used automated semantic segmentations to compute calcification attributes (Fig. 7). We analyzed the agreement between automated and manual measurements, including lumen area [Fig. 7(a)], calcification angle [Fig. 7(b)], calcification thickness [Fig. 7(c)], and calcification depth [Fig. 7(d)]. We observed excellent agreement of lumen areas, except for
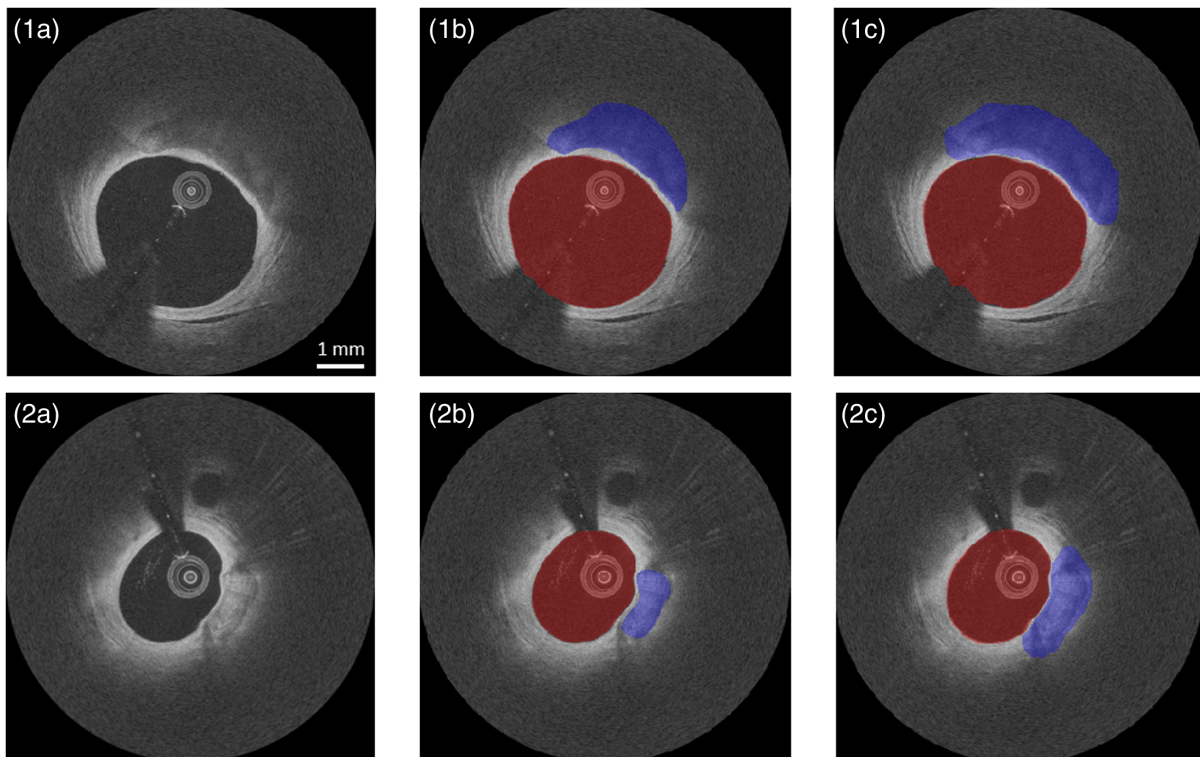

Fig. 4 Automated segmentation results on two calcified vessel images (1 and 2). Images are: (a) IVOCT image, (b) labeled image, and (c) automatic segmentation. Both examples show good agreement between manual and automated assessments. In all examples, lumen is shown in red and calcification in blue. The scale bar applies to all the images.

Table 1 Comparison of segmentation performance when using the same labeled data arranged in (x, y) and (r, θ). Confusion matrices show performance of classifier across all 10 folds of the training data. Numbers indicate the mean and standard deviation for segmentation sensitivity (in percentage) across all folds. All results are after using noise-cleaning strategy. For x, y data: mean values ± standard deviation for (sensitivity, specificity, and F1 score) for each class is: other: (0.95 ± 0.02, 0.96 ± 0.02, 0.97 ± 0.03), lumen: (0.98 ± 0.02, 0.98 ± 0.01, 0.90 ± 0.01), calcium: (0.82 ± 0.06, 0.97 ± 0.01, 0.42 ± 0.03). For (r, θ) data: mean values ± standard deviation for (sensitivity, specificity, and F1 score) for each class is: other: (0.97 ± 0.01, 0.98 ± 0.01, 0.98 ± 0.01), lumen: (0.99 ± 0.01, 0.99 ± 0.006, 0.99 ± 0.008), calcium: (0.85 ± 0.04, 0.99 ± 0.004, 0.73 ± 0.01). Overall, when analyzing sensitivity, specificity, and F1 score, the classifier trained on the (r, θ) data had better performance. Using the Wilcoxson signed-rank test, we determined statistically significant differences ($p < 0.01$) between the two methods for calcification F1 score.

x, y	Predicted "other"	Predicted "lumen"	Predicted "calcification"
True "other"	$95.18 \pm 2.83^{*}$	1.49 ± 1.23	3.33 ± 2.08
True "lumen"	1.55 ± 2.48	$98.03 \pm 2.42^{*}$	0.42 ± 0.54
True "calcification"	13.76 ± 6.96	3.345 ± 2.40	$82.89 \pm 6.81^{*}$

r, θ	Predicted "other"	Predicted "lumen"	Predicted "calcification"
True "other"	$97.62 \pm 1.47^{*}$	0.62 ± 0.55	1.75 ± 1.09
True "lumen"	0.56 ± 0.60	$99.42 \pm 1.05^{*}$	0.01 ± 0.02
True "calcification"	14.50 ± 7.33	0.22 ± 0.23	$85.27 \pm 4.82^{*}$

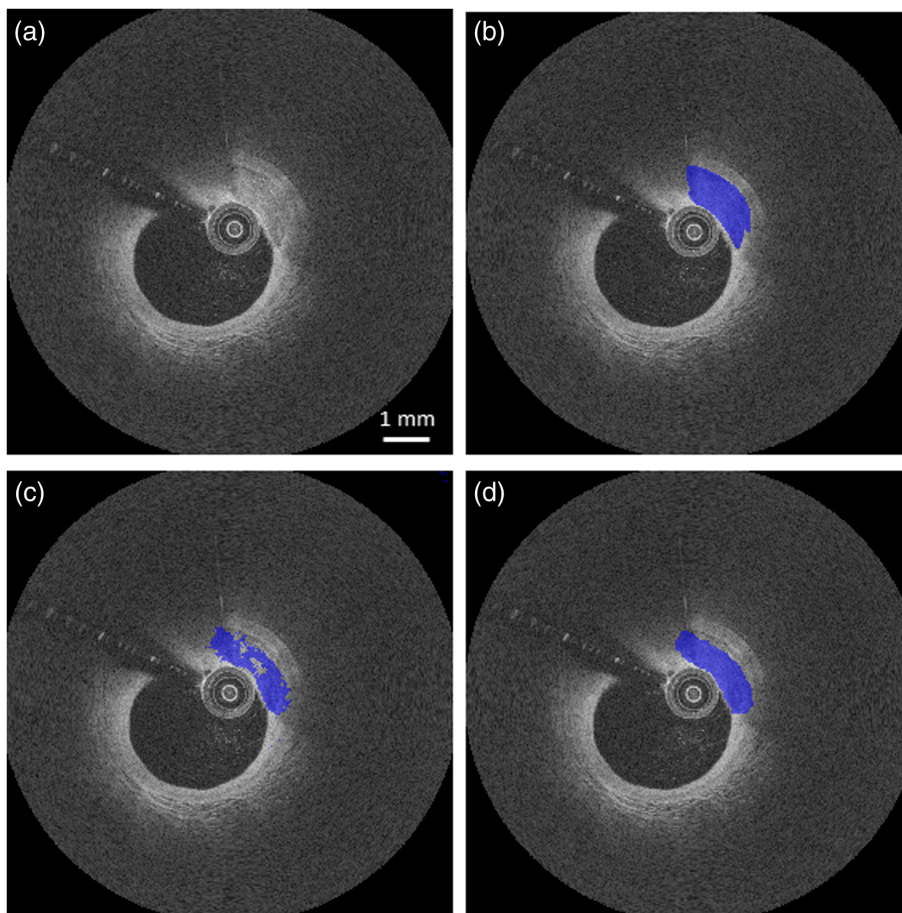*Statistical significant differences ($p < 0.01$).

Fig. 5 CRF refinement of segmentation improves segmentation performance. (a) IVOCT, (b) ground truth labels, (c) initial segmentation results from the deep learning model, and (d) output after CRF processing. CRF smoothed segmentation results and removed isolated islands. The scale bar applies to all the images.

Table 2 Sensitivity and Dice coefficient calculated (A) before and (B) after segmentation noise cleaning using CRF for all classes for (r, θ) dataset. Improvement was not only observed visually but also numerically, as Dice coefficient for calcifications was improved from 0.42 to 0.76 with noise cleaning as in Table 2. CRF noise cleaning improved performance, and Wilcoxon signed-rank test suggested a significant difference ($p < 0.005$) for calcifications.

	Sensitivity	Dice coefficient
A		
Other	0.94 ± 0.02	0.97 ± 0.01
Lumen	0.98 ± 0.02	0.98 ± 0.02
Calcifications	0.81 ± 0.1	0.42 ± 0.04
B		
Other	0.97 ± 0.01	0.98 ± 0.006
Lumen	0.99 ± 0.01	0.98 ± 0.01
Calcifications	0.85 ± 0.04	0.76 ± 0.03

mismatch in images containing side branches. Calcification angle, thickness, and depth had good agreement between manual and automated measurements across the range of calcifications observed. Mean values of agreement were -0.54 mm^2 (95% CI, -1.2 to 0.14 mm^2); -7 deg (95% CI, -52 deg to 37 deg); -0.13 mm (95% CI, -0.52 to 0.25 mm); and 0.05 mm (95% CI, -0.15 to 0.25 mm) for lumen area, calcification angle, calcification thickness, and calcification depth, respectively.

Finally, we used automated semantic segmentation to compute the stent deployment calcification score as described in Sec. 2.4 (Table 3). We assessed five representative lesions. We found strong agreement between manual and automated assessments for four out of five cases. The case that had the least agreement between manual and automated assessment is shown in Fig. 8. What makes this case challenging is the calcification is separated by the guidewire shadow. Manual analysts defined this lesion as two calcifications; automated results showed this as one.

5 Discussion

We developed an automated method for calcification analysis, which included methods for semantic segmentation using deep learning, for calculation of calcification attributes, and for calculation of a previously developed stent-deployment calcification score. We used SegNet [with transfer learning using the pretrained VGG-16 weights and with receptive field of (212×212) that enable substantial contextual information to be included for determining areas containing calcifications] and trained/tested on 48 VOIs (2640 IVOCT images). The dataset contained a variety of lesion types, including: calcifications, lipidous, and mixed segments with both calcifications and lipidous
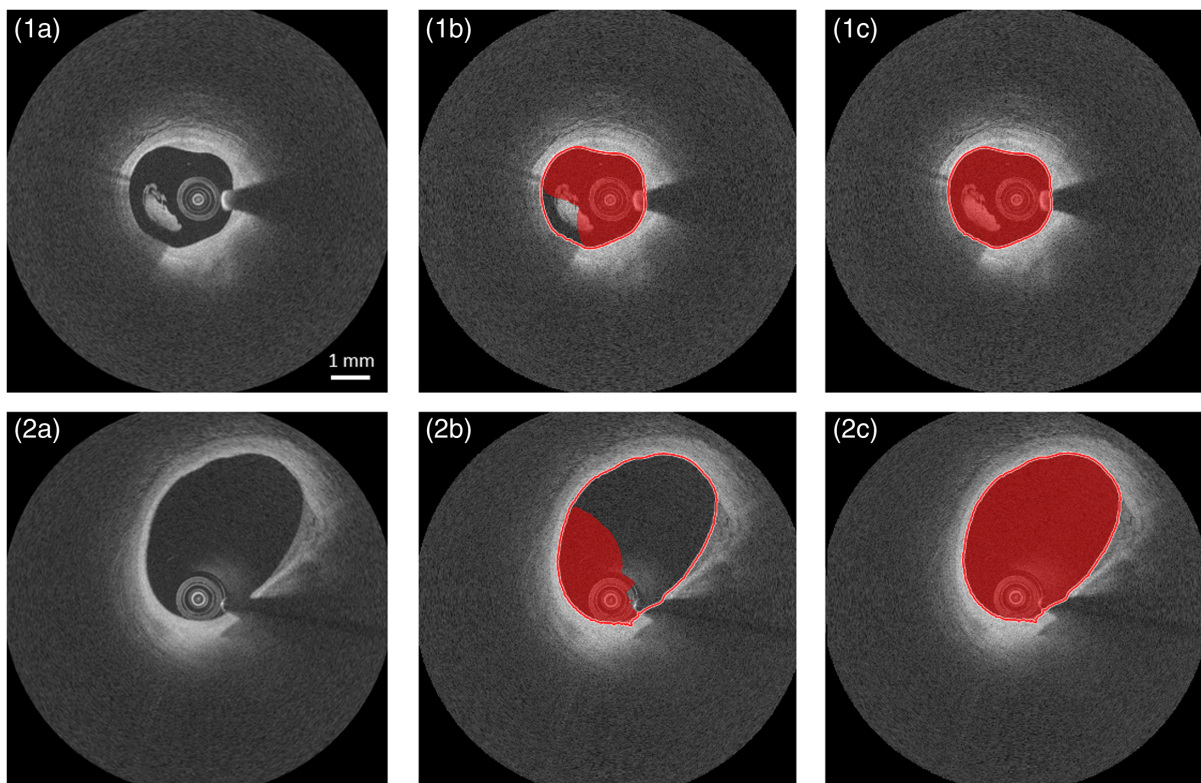

Fig. 6 Comparison of lumen segmentation obtained with deep learning as compared to our earlier dynamic programming solution. (a) IVOCT Image, (b) automatic segmentation using dynamic programming, and (c) segmentation using the deep learning model. The bright red contour is the ground truth label. The deep learning method gives a much better result in these two cases. In particular, the dynamic programming approach can fail in the presence of thrombus in the lumen. Similar results were obtained in other problematic images. The scale bar applies to all the images.
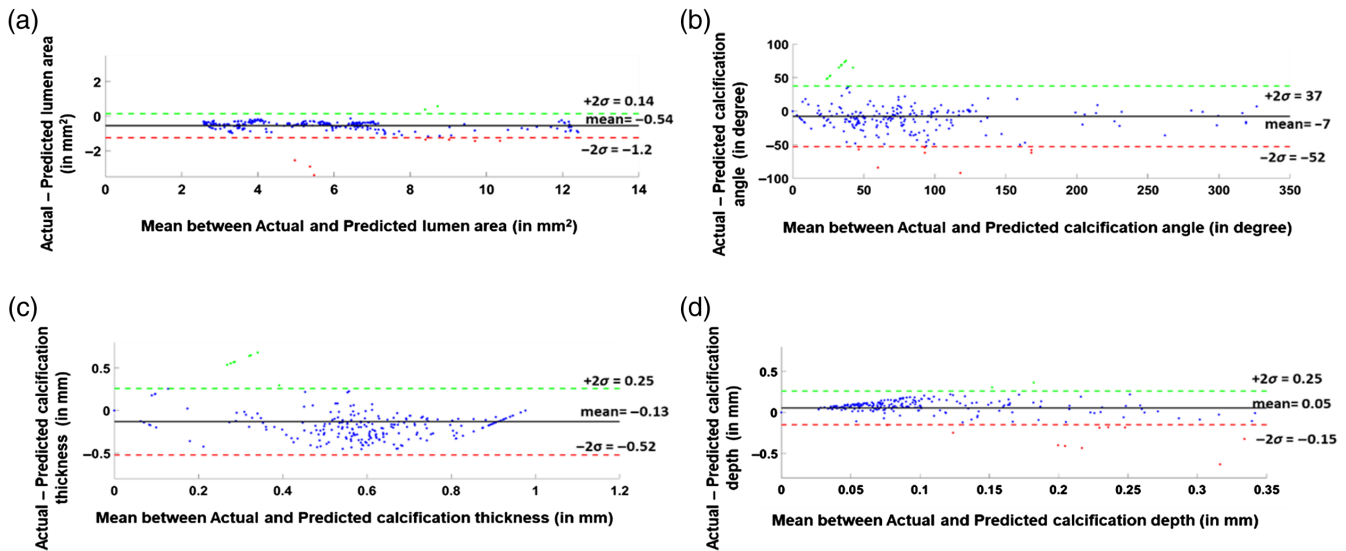
Fig. 7 Comparisons between manual and automated calcification measurements. Graphs are: (a) lumen area, (b) calcification arc angle, (c) calcification thickness, and (d) calcification depth. In these Bland–Altman plots, each plotted value is (manual-predicted). Hence, a negative mean indicates that the predicted value on average is biased larger than the manually obtained value. Bland–Altman analysis demonstrates a strong correlation between the manual and automated assessments. The mean value of agreement was -0.54 mm^2 (95% CI, -1.2 to 0.14 mm^2), -7 deg (95% CI, 52 deg to 37 deg), -0.13 mm (95% CI, -0.52 to 0.25 mm), and 0.05 mm (95% CI, -0.15 to 0.25 mm) for lumen area, calcification angle, calcification thickness, and calcification depth, respectively.

Table 3 IVOCT-based calcification scoring for representative lesions. We used the calcification scoring system developed by Fujino et al.[13] on five held out lesions and compared manual and automated measurements. Scores are based on lesion length, maximum thickness, and maximum angle. Score is cumulative sum of the following metrics: two points for maximum angle >180 deg, one point for maximum thickness >0.5 mm, and one point for length >5 mm. The idea of calcium scoring is to define the lesion that would benefit from plaque modification prior to stent implantation. Lesions with calcium score of 0 to 3 had excellent stent expansion, whereas the lesions with a score of 4 had poor stent expansion. Scores for each attribute as shown as follows: attribute value (score). The calcification scores are identical between manual and predicted results for the first four lesions. Lesion 5 is a challenging case and is shown in Fig. 8.

Lesion	Name	Frames	Length (mm)	Ground truth Maximum calcium angle (deg)	Ground truth Maximum calcium thickness (mm)	Score	Prediction Maximum calcium angle (deg)	Prediction Maximum calcium thickness (mm)	Score
1	Ca1	19	1.9 (0)	45 (0)	0.52 (1)	1	89 (0)	0.91 (1)	1
2	Ca2	20	2 (0)	68 (0)	0.67 (1)	1	124 (0)	0.91 (1)	1
3	Ca3	73	7.3 (1)	330 (2)	0.60 (1)	4	328 (2)	0.82 (1)	4
4	Ca4	32	3.2 (0)	132 (0)	1.1 (1)	1	123 (0)	1.4 (1)	1
5	Ca5	123	12.3 (1)	146 (0)	0.88 (1)	2	227 (2)	1.0 (1)	4

regions, as well as segments devoid of these characteristics. Having a variety of disease states is the key for any robust learning system. In a remaining dataset held out from any optimization, we automatically computed the stent-deployment calcification score and obtained very good agreement with manual determinations. This suggests that our methods (with optional manual corrections, as argued below) could predict stent treatment outcomes from present IVOCT images and could help determine, which lesions would benefit from present lesion preparation (e.g., atherectomy).

When we compared segmentation performance using (r, θ) and (x, y) representations of the data, we found that (r, θ) gave a better sensitivity, specificity, and $F1$ across all classes. There are multiple potential reasons. First, data are originally acquired in the (r, θ) domain. To create the (x, y) representation, data must be geometrically transformed leading to increased interpolation as one goes out from the catheter center. Potentially, this interpolation effect could negatively affect the success of local kernels. Second, the (r, θ) data representation was amenable to an elegant data augmentation scheme as described in Sec. 2.1, allowing us to create heavily augmented data. Third, we were able to process the (r, θ) images at full resolution, but had to resize the (x, y) images in order to train the SegNet model. This could have affected the ability of the CNN to recognize
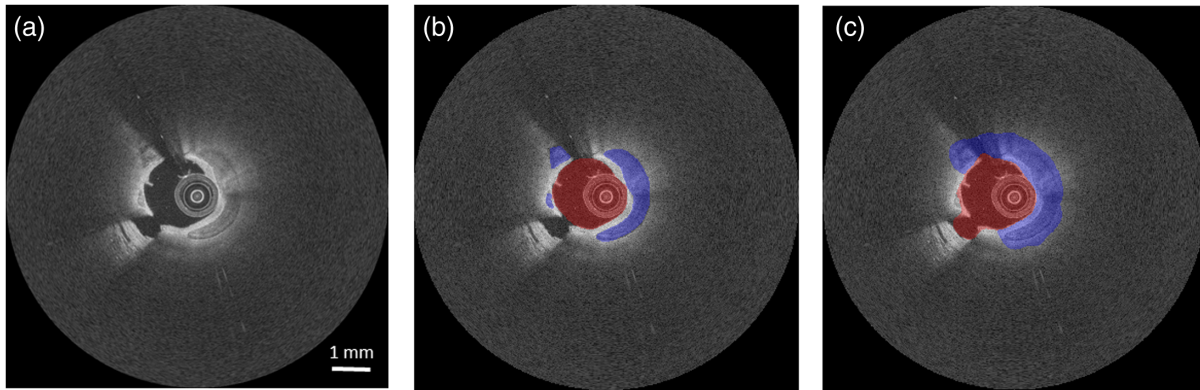
Fig. 8 Challenging case for calcification scoring. (a) Original image, (b) manual annotations, and (c) automated results. Lumen is shown in red, and calcification is shown in blue. In this case, we see a calcification separated by a guidewire shadow. Expert analysts chose not to segment the region behind the guidewire, while automated method calls this region as calcification. It is possible that calcification exists behind the shadow, but pathology would be needed for a definitive answer, but could not be acquired in this clinical case. The scale bar applies to all the images.

features such as the sharp edges at calcifications. Fourth, in the (r, θ) domain, calcified lesions have one "orientation" with the leading and trailing edges roughly parallel to the lumen. In the case of the (x, y) representation, lesions are at all possible orientations in the image array. Even though we augmented data by rotating the (x, y) images, the similar look of lesions in (r, θ) may have comparatively enhanced learning.

We found it beneficial to implement CRF for refinement of initial segmentation results. We applied CRF to the vector of class probabilities and the input image intensity, at each pixel location. This enhanced the final segmentation and improved the performance of the downstream analysis. As shown in Fig. 5, CRF smooths the segmentation results and prevents isolated spots of calcification from appearing in our results. This causes a visual improvement in our results, and this improvement is reflected numerically by the increase in sensitivities and Dice coefficient following CRF implementation.

Our approach has advantages for segmenting the lumen as compared to previous methods such as dynamic programming[38] (Fig. 6). The presence of image artifacts (e.g., thrombus or improper blood clearing during image acquisition) as well as very eccentric lumens create challenges to lumen segmentation algorithms that use edges, such as our previous dynamic programing approach. Our deep learning approach takes contextual area information into account, which reduces the impact of these artifacts on determining the lumen border.

We were able to quantify calcification attributes based on the automated segmentations, including lumen area, calcification arc, thickness, and depth (Fig. 7). For lumen area, the automated measurements were excellent (good precision and bias) as compared to manual assessments. Most errors were in regions with side branches, which are ambiguous for analysts to label. Automated measurements of calcification arc also had strong agreement with manual assessments. Segmentation errors are mostly related to calcification deposits that have small arc angles (>40 deg), which have less impact on clinical decision-making. We had high correlation with manual analysis with large arc angles (>200 deg), which is encouraging, as these large calcifications are more likely candidates for plaque modification prior to stenting. Calcification thickness measurements had good agreement between manual and automated assessments, although our algorithm had a tendency to overestimate

calcification thickness. Our algorithm tends to agree with manual determination of the calcification front border but has less agreement with the back border. This is due to the IVOCT signal having limited depth penetration, making determination of the calcification back border difficult, even for manual assessments. Finally, the calcification depth had a strong correlation between automated and manual measurements. We observe a trend that errors tend to increase with larger depths. One reason is that calcification depth is based on both the lumen and calcification segmentation, so errors in lumen segmentation (observed in larger lumens) could propagate to the calcification depth measurement.

Ultimately, we want to be able to use calcification segmentations to provide information to cardiologists concerning the need for employing calcification modification strategies (e.g., atherectomy or IVL as with Shockwave[TM]). Visualization of segmented calcification is one approach, but another is calculation of the stent-deployment calcification score. Automatically obtained scores were identical to manually obtained ones in four out of five cases. The score defines lesions that would benefit from plaque modification prior to stent implantation. The method is a cumulative score based on calcification attributes (i.e., maximum angle). Lesions with calcification score of 0 to 3 had "adequate stent expansion," whereas lesions with a score of 4 had "poor stent expansion." The case with disagreement is shown in Fig. 8. This case is challenging because the calcification is separated by the IVOCT guidewire shadow. Analysts chose not to label this region, but our automated method bridged the guidewire region, calling it as one continuous calcification. It is highly likely that calcifications occur behind the guidewire in this lesion, but we can only be certain if histology is acquired from this sample.[42] Based on the scoring system presented in Table 3, if this region was calcifications, lesion preparation would be necessary for treatment. Thus, interpreting what is behind the guidewire would alter clinical decision-making. Although automated stent deployment calcification score is promising, if this were to be implemented clinically, one would likely want to allow operator editing of calcifications, particularly at locations important to the score (e.g., the image having the maximum arc angle). Using today's GPU hardware (NVIDIA GTX 1080 Ti), it is possible to perform calcification semantic segmentation in under 1 s per frame.

This suggests that live-time use in the clinic would be possible, especially if the operator identified volumes of interest (VOI) for analysis.

There are potential modifications to our study. Developing our segmentation method required the manual labeling of thousands of IVOCT images. It is possible that some of our labels could be wrong (e.g., Fig. 8), and that analysts might change their mind after viewing automated results. Thus, we could implement an active learning scheme where analysts could do a second pass of the dataset to possibly modify the labels after viewing automated results. In this study, 48 VOI from 34 pullbacks were used. It is possible that the use of more cases could improve generalizability. In addition, it would be interesting to include labeled lipidous regions. Finally, adding additional 3-D information might help make some determinations.

6 Conclusion

Coronary calcifications are a major determinant of the success of coronary stenting. We developed an automatic method for semantic segmentation of calcifications in IVOCT images using deep learning. Results can be applied to determine calcification attributes, and for computation of an IVOCT-based calcification score, which can help predict stent treatment outcome for target lesions.

References

1. L. K. Kim et al., "Rate of percutaneous coronary intervention for the management of acute coronary syndromes and stable coronary artery disease in the United States (2007 to 2011)," *Am. J. Cardiol.* **114**(7), 1003–1010 (2014).
2. I. Moussa et al., "Impact of coronary culprit lesion calcium in patients undergoing paclitaxel-eluting stent implantation (a TAXUS-IV sub study)," *Am. J. Cardiol.* **96**(9), 1242–1247 (2005).
3. M. Farag et al., "Treatment of calcified coronary artery lesions," *Expert Rev. Cardiovasc. Ther.* **14**(6), 683–690 (2016).
4. R. Kawaguchi et al., "Impact of lesion calcification on clinical and angiographic outcome after sirolimus-eluting stent implantation in real-world patients," *Cardiovasc. Revasc. Med.* **9**(1), 2–8 (2008).
5. G. D. Dangas et al., "In-stent restenosis in the drug-eluting stent era," *J. Am. Coll. Cardiol.* **56**(23), 1897–1907 (2010).
6. K. Fujii et al., "Stent underexpansion and residual reference segment stenosis are related to stent thrombosis after sirolimus-eluting stent implantation: an intravascular ultrasound study," *J. Am. Coll. Cardiol.* **45**(7), 995–998 (2005).
7. G. F. Attizzani et al., "Mechanisms, pathophysiology, and clinical aspects of incomplete stent apposition," *J. Am. Coll. Cardiol.* **63**(14), 1355–1367 (2014).
8. H. Doi et al., "Impact of post-intervention minimal stent area on 9-month follow-up patency of paclitaxel-eluting stents: an integrated intravascular ultrasound analysis from the TAXUS IV, V, and VI and TAXUS ATLAS workhorse, long lesion, and direct stent trials," *JACC Cardiovasc. Interventions* **2**(12), 1269–1275 (2009).
9. M.-K. Hong et al., "Intravascular ultrasound predictors of angiographic restenosis after sirolimus-eluting stent implantation," *Eur. Heart J.* **27**(11), 1305–1310 (2006).
10. N. G. Uren et al., "Predictors and outcomes of stent thrombosis: an intravascular ultrasound registry," *Eur. Heart J.* **23**(2), 124–132 (2002).
11. G. S. Mintz et al., "Unstable angina/myocardial infarction/atherosclerosis: patterns of calcification in coronary artery disease a statistical analysis of intravascular ultrasound and coronary angiography in 1155 lesions," *Circulation* **91**(7), 1959–1965 (1995).
12. G. S. Mintz et al., "American College of Cardiology clinical expert consensus document on standards for acquisition, measurement and reporting of intravascular ultrasound studies (IVUS): a report of the American College of Cardiology task force on clinical expert consensus documents developed in collaboration with the European Society of Cardiology endorsed by the Society of Cardiac Angiography and Interventions," *J. Am. Coll. Cardiol.* **37**(5), 1478–1492 (2001).
13. A. Fujino et al., "A new optical coherence tomography-based calcium scoring system to predict stent underexpansion," *EuroIntervention* **13**(18), 2182–2189 (2018).
14. T. Ma et al., "Multi-frequency intravascular ultrasound (IVUS) imaging," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **62**(1), 97–107 (2015).
15. X. Li et al., "Integrated IVUS-OCT imaging for atherosclerotic plaque characterization," *IEEE J. Sel. Top. Quantum Electron.* **20**(2), 196–203 (2014).
16. F. Prati et al., "Expert review document part 2: methodology, terminology and clinical applications of optical coherence tomography for the assessment of interventional procedures," *Eur. Heart J.* **33**(20), 2513–2520 (2012).
17. H. G. Bezerra et al., "Intracoronary optical coherence tomography: a comprehensive review clinical and research applications," *JACC Cardiovasc. Interventions* **2**(11), 1035–1046 (2009).
18. G. J. Ughi et al., "Automated tissue characterization of in vivo atherosclerotic plaques by intravascular optical coherence tomography images," *Biomed. Opt. Express* **4**(7), 1014–1030 (2013).
19. L. S. Athanasiou et al., "Methodology for fully automated segmentation and plaque characterization in intracoronary optical coherence tomography images," *J. Biomed. Opt.* **19**(2), 026009 (2014).
20. P. Zhou et al., "Automatic classification of atherosclerotic tissue in intravascular optical coherence tomography images," *J. Opt. Soc. Am. A* **34**(7), 1152–1159 (2017).
21. D. Prabhu et al., "Automated A-line coronary plaque classification of intravascular optical coherence tomography images using handcrafted features and large datasets," *J. Biomed. Opt.* **24**(10), 106002 (2019).
22. C. Kolluru et al., "Deep neural networks for A-line-based plaque classification in coronary intravascular optical coherence tomography images," *J. Med. Imaging* **5**(4), 044504 (2018).
23. Y. Gharaibeh et al., "Deep learning segmentation of coronary calcified plaque from intravascular optical coherence tomography (IVOCT) images with application to finite element modeling of stent deployment," *Proc. SPIE* **10951**, 109511C (2019).
24. J. J. Rico-Jimenez et al., "Automatic classification of atherosclerotic plaques imaged with intravascular OCT," *Biomed. Opt. Express* **7**(10), 4069–4085 (2016).
25. Y. L. Yong et al., "Linear-regression convolutional neural network for fully automated coronary lumen segmentation in intravascular optical coherence tomography," *J. Biomed. Opt.* **22**(12), 126005 (2017).
26. A. Abdolmanafi et al., "Deep feature learning for automatic tissue classification of coronary artery using optical coherence tomography," *Biomed. Opt. Express* **8**(2), 1203–1220 (2017).

27. A. Abdolmanafi et al., "Characterization of coronary artery pathological formations from OCT imaging using deep learning," *Biomed. Opt. Express* **9**(10), 4936–4960 (2018).
28. N. Gessert et al., "Automatic plaque detection in IVOCT pullbacks using convolutional neural networks," *IEEE Trans. Med. Imaging* **38**, 426–434 (2019).
29. V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(12), 2481–2495 (2017).
30. M. Cordts et al., "The cityscapes dataset for semantic urban scene understanding," in *IEEE Conf. Comput. Vision and Pattern Recognit.*, IEEE, Las Vegas, Nevada, pp. 3213–3223 (2016).
31. S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Int. Conf. Mach. Learn. (ICML)*, F. Bach and D. Blei, Eds., Vol. **37**, pp. 448–456 (2015).
32. V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, Omnipress, pp. 807–814 (2010).
33. W. Yu et al., "Visualizing and comparing AlexNet and VGG using deconvolutional layers," in *Proc. 33 rd Int. Conf. Mach. Learn.* (2016).
34. B. Shuai, T. Liu, and G. Wang, "Improving fully convolution network for semantic segmentation," arXiv:1611.08986 [cs] (2016).
35. J. Nagi et al., "Max-pooling convolutional neural networks for vision-based hand gesture recognition," in *IEEE Int. Conf. Signal and Image Process. Appl.*, pp. 342–347 (2011).
36. K. Kamnitsas et al., "Efficient multi-scale 3-D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.* **36**, 61–78 (2017).
37. G. J. Tearney et al., "Consensus standards for acquisition, measurement, and reporting of intravascular optical coherence tomography studies: a report from the international working group for intravascular optical coherence tomography standardization and validation," *J. Am. Coll. Cardiol.* **59**(12), 1058–1072 (2012).
38. Z. Wang et al., "Semiautomatic segmentation and quantification of calcified plaques in intracoronary optical coherence tomography images," *J. Biomed. Opt.* **15**(6), 061711 (2010).
39. G. T. Stefano et al., "Unrestricted utilization of frequency domain optical coherence tomography in coronary interventions," *Int. J. Cardiovasc. Imaging* **29**(4), 741–752 (2013).
40. D. Eigen and R. Fergus, "Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture," in *Proc. IEEE Int. Conf. Comput. Vision* (2015).
41. D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," arXiv:1412.6980 [cs] (2014).
42. D. Prabhu et al., "Three-dimensional registration of intravascular optical coherence tomography and cryo-image volumes for microscopic-resolution validation," *J. Med. Imaging* **3**(2), 026004 (2016).

Biographies of the authors are not available.