Computer Modeling in Engineering & Sciences
DOI:10.32604/cmes.2020.09718
Article

Tech Science Press

Machine Learning Model Comparison for Automatic Segmentation of Intracoronary Optical Coherence Tomography and Plaque Cap Thickness Quantification

Caining Zhang[1], Xiaopeng Guo[2], Xiaoya Guo[3], David Molony[4], Huaguang Li[2], Habib Samady[4], Don P. Giddens[4,5], Lambros Athanasiou[6], Dalin Tang[1*,7], Rencan Nie[2,*] and Jinde Cao[8]

[1]School of Biological Science and Medical Engineering, Southeast University, Nanjing, 210096, China
[2]School of Information Science and Engineering, Yunnan University, Kunming, 650091, China
[3]School of Science, Nanjing University of Posts and Telecommunications, Nanjing, 210023, China
[4]Department of Medicine, Emory University School of Medicine, Atlanta, GA, 30307, USA
[5]The Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, 30332, USA
[6]Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA
[7]Mathematical Sciences Department, Worcester Polytechnic Institute, Worcester, MA 01609, USA
[8]School of Mathematics, Southeast University, Nanjing, 210096, China
*Corresponding Authors: Dalin Tang, Southeast University. Email: dtang@wpi.edu; Rencan Nie. Email: rcnie@ynu.edu.cn
Received: 19 January 2020; Accepted: 06 March 2020

Abstract: Optical coherence tomography (OCT) is a new intravascular imaging technique with high resolution and could provide accurate morphological information for plaques in coronary arteries. However, its segmentation is still commonly performed manually by experts which is time-consuming. The aim of this study was to develop automatic techniques to characterize plaque components and quantify plaque cap thickness using 3 machine learning methods including convolutional neural network (CNN) with U-Net architecture, CNN with Fully convolutional DenseNet (FC-DenseNet) architecture and support vector machine (SVM). *In vivo* OCT and intravascular ultrasound (IVUS) images were acquired from two patients at Emory University with informed consent obtained. Eighteen OCT image slices which included lipid core and with acceptable image quality were selected for our study. Manual segmentation from imaging experts was used as the gold standard for model training and validation. Since OCT has limited penetration, virtual histology IVUS was combined with OCT data to improve reliability. A 3-fold cross-validation method was used for model training and validation. The overall tissue classification accuracy for the 18 slices studied (total classification database sample size was 8580096 pixels) was 96.36% and 92.72% for U-Net and FC-DenseNet, respectively. The best average prediction accuracy for lipid was 91.29% based on SVM, compared to 82.84% and 78.91% from U-Net and FC-DenseNet, respectively. The overall average accuracy (Acc) differentiating lipid and fibrous tissue were 95.58%, 92.33% and 81.84% for U-Net, FC-DenseNet and SVM, respectively. The average errors of U-Net, FC-DenseNet and SVM from the 18 slices for cap thickness quantification were 8.83%, 10.71% and 15.85%. The average relative errors of minimum cap thickness from 18 slices of U-Net, FC-DenseNet and SVM were 17.46%, 13.06%

and 22.20%, respectively. To conclude, CNN-based segmentation methods can better characterize plaque compositions and quantify plaque cap thickness on OCT images and are more likely to be used in the clinical arena. Large-scale studies are needed to further develop the methods and validate our findings.

Keywords: Image segmentation; plaque; cap thickness; OCT; CNN; SVM

1 Introduction

Atherosclerotic cardiovascular diseases cause millions of sudden deaths annually, with coronary heart disease accounting for the majority [1], and plaque rupture plays a major role in most acute coronary syndromes [2]. Various methods have been developed to quantify atherosclerosis, including invasive methods such as intravascular ultrasound (IVUS) and X-ray angiography [3], and non-invasive methods such as magnetic resonance imaging (MRI) [4,5]. For clinical prevention of coronary artery disease, the most difficult problem is to identify vulnerable patients before drastic cardiovascular events occur. Some clinical and biomechanical studies have shown that plaque cap thickness, lipid core size and plaque burden can be used as key predictors for plaque stability [6]. However, since the widely accepted threshold value of plaque cap thickness for vulnerable plaque is around 65 μm, current clinical imaging technologies including IVUS and MRI are not able to detect due to their limited imaging resolution [2]. As an emerging imaging modality, optical coherence tomography (OCT) has a high resolution of approximately 10 μm. With its high resolution, OCT can provide precise details on the morphology of coronary plaques, especially plaque cap thickness, which is the most closely watched feature for vulnerable plaque management. At present, clinical classification of OCT images is mainly accomplished by physicians and radiologists with visual inspections, which is time-consuming, lack of accuracy, and easily lead to observer variability [7]. Automatic classification and identification of vulnerable plaques are highly significant for cardiovascular research.

Over the past years, many scholars have dedicated themselves to developing automatic segmentation methods for OCT images and proposed several methodologies. Van Soest et al. [8] studied and analyzed the optical attenuation coefficient of tissue in intravascular OCT images, and proposed a method for automatic classification of atherosclerotic plaque components. Going one step further, Shalev et al. [9] and Guo et al. [10] applied support vector machine (SVM) to identify plaques, via machine learning method. Athanasiou et al. based on random forests method to detect vessel lumen boundaries, identify plaque region and detect four tissue types to achieve automatic segmentation [11]. More recently, Athanasiou et al., Abdolmanafi et al., Zhang et al. and He et al. applied convolutional neural network (CNN)-based deep learning models to plaque characterization and identification and have made noticeable contributions [12–15]. With all of those developments, most of these methods were still limited by intricate image preprocessing and large calculation cost. Features such as cap thickness which are important for plaque vulnerability assessment were generally not quantified and discussed.

CNN has outstanding ability of extracting abstracted features directly from the original input data and has achieved remarkable success in the field of computer vision, image classification, object detection, object recognition, instance segmentation and semantic segmentation. Fully convolutional network (FCN) [16] was first proposed and applied to semantic segmentation, which replaced fully connected layers with convolutional layers, selected skip architecture to incorporate semantic features and detailed features and employed deconvolution to improve the accuracy of segmentation results. After that, more semantic segmentation architecture were proposed by several investigators. SegNet [17] performed pooling indices in the decoder to upsample the feature map and maintain the integrity of high-frequency details in the segmentation. Based on FCN, the contracting path that captures context and the symmetric expanding path that enables precise localization make up the U-Net architecture [18], which could use high resolution information obtained by the shallow layer supply the missing detailed spatial information during the up-sampling process to obtain better

segmentation results from very few images and have fewer trainable parameters than SegNet. As well known, medical images are generally difficult to obtain and need more precision segmentation results than natural images. As such, U-Net was widely used in medical image segmentation and achieved widespread acceptance. Extending the powerful DenseNet [19] classification network and using the upsampling path to restore the full input resolution, fully convolutional DenseNet (FC-DenseNet) [20] was formed, which has a unique advantage in considering both network depth and a small number of parameters compared with other popular networks. And it does not require additional post-processing and pre-training. The connections of FC-DenseNet could combine information of multiple scales, which benefits the extraction of abundant features of images and makes it achieve state-of-art urban scene segmentation performance.

Recognition of lipid in OCT images can be classified as semantic segmentation, which is to assign each pixel in the input image a semantic class in order to get a pixel-wise dense classification. U-Net architecture due to its fast speed attribute and FC-DenseNet architecture because of its deep and powerful infrastructure have been proved to have promising future in the field of segmentation in recent research [18,19]. In this paper, three methods for automatically segmenting intracoronary OCT images were introduced to quantify plaque components and cap thickness. Two of the three methods were based on CNN (U-Net and FC-DenseNet) and the other was based on SVM. Manual segmentation from imaging experts was used as the gold standard for model training and validation. Since OCT has limited penetration, virtual histology IVUS was combined with OCT data to improve reliability. Results from the three segmentation methods were compared using the same dataset.

2 Data and Methods

In this section, we will introduce data acquisition procedures first, and then introduce two types of supervised learning methods, CNN-based methods (including U-Net and FC-DenseNet) which are semantic segmentation methods and SVM-based method which is an ordinary machine learning method.

2.1 Data Acquisition and Gold Standard

In vivo OCT and IVUS images were acquired from two patients with informed consent obtained. Data acquisition procedures were previously reported and are omitted here to avoid repetition [14,22]. It should be noted that Virtual-Histology IVUS (VH-IVUS) provided four classified tissue components: fibrous, fibro-fatty, lipid and calcification (Ca) [20]. VH-IVUS segmented tissue components were used as references by radiologists in their manual segmentation to produce gold standard for our automatic segmentation [21]. Fig. 1 gave matched sample IVUS and OCT slices with overlapping contours. Different colors were used to distinguish four plaque tissue types. Dynamic angiography data were recorded which provided positions of catheters and vessel segments for co-registration of OCT and IVUS images.

Eighteen matched VH-IVUS and OCT images with good image quality were selected to verify and compare the three proposed segmentation methods in this study. Manual segmentation of OCT images was done by experts using VH-IVUS segmented tissue components as references [23]. Since OCT has limited penetration depth and often cannot “see” the out-border of lipid cores, we focused on quantification of cap thickness. For each slice, to make the region of interest (ROI), an outline (called the out-boundary of the ROI) was generated by pushing the lumen border (the inner boundary of the ROI) outward 1 mm from the center of the catheter. Then ROI is defined as the area bounded by its inner-boundary (lumen border) and the out-boundary with artery branches and artifacts removed (see Fig. 1c). Combining expert manual segmentation images and ROI, the gold standard was developed (see Fig. 1d). Fig. 1 gave 6 matched sample OCT slices with ROI and segmented gold standard contours. Calcification was neglected since the lipid core and cap thickness were the selected foci in this study.

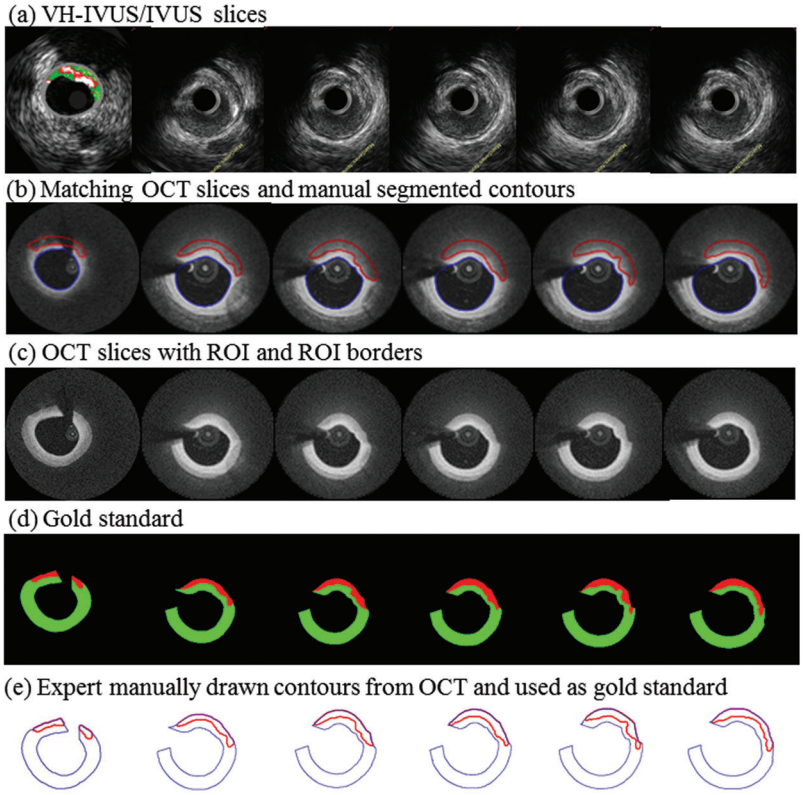

Figure 1: Six matching sample VH-IVUS/OCT images with overlapping contours and expert manually drawn contours used as gold standard. Colors for VH-IVUS: *Red*—lipid; *White* -Ca; *Dark Green*—Fibrous; *Light Green*—Fibro-Fatty. Colors for gold standard: *Red*—lipid tissue, *Green*—fibrous tissue; *Black*—background. Color for contour lines: *Red*—lipid; *Blue*—lumen and out-boundary. (a) VH-IVUS/IVUS slices. (b) Matching OCT slices and manual segmented contours. (c) OCT slices with ROI and ROI borders. (d) Gold standard. (e) Expert manually drawn contours from OCT and used as gold standard

2.2 Convolutional Neural Network: U-Net Architecture and Fully Convolutional DenseNet (FC-DenseNet) Architecture

2.2.1 Image Processing and Data Augmentation

As shown in Fig. 2, in order to remove redundant image information, the original OCT data (1024 × 1024 pixels or 1905 × 1905 pixels) of two patients were cropped (784 × 608 pixels) and used as the input image of the network to save the model training time.

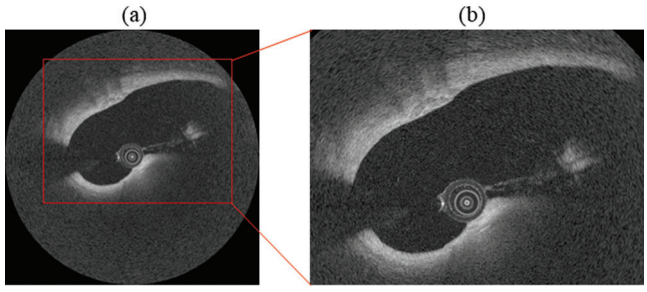

Figure 2: Image processing for CNN. (a) Original image (the red rectangular marked input image); (b) Magnified input image of the network

Previous research has shown that data augmentation can prevent neural network from overfitting and improve model performance in imbalanced class problems [24–26]. In addition, the quality of deep learning model training results is closely related to the size of the training data. To increase the data size for our modeling training, we augmented existing datasets to train more powerful deep convolutional neural network models. In this study, the data augmentation techniques included rotation (rotation angles: θ = 90°, 180°, and 270°), flipping (horizontal and vertical) and brightness transformation (gamma adjustment applied, the parameter values used: 0.5, 0.7, 0.9, 1.1, 1.3, 1.5). After augmentation, a total of 1512 slices were available for our use in machine learning method training and testing.

2.2.2 Network Architecture: U-Net

For U-Net architecture, it consists of a contracting path and an expansive path. As for the architecture employed in this work (see Fig. 3), the contracting path includes the repeated application of two 3 × 3 convolutions, each followed by scaled exponential linear units (SeLU) and a 2 × 2 max pooling operation with stride 2 for downsampling. In each downsampling step, the number of feature channels is doubled. Each step in the expansive path consists of an upsampling of the feature map followed by a 2 × 2 convolution a concatenation with the correspondingly cropped feature map from the contracting path, and two 3 × 3 convolutions, each followed by a SeLU. And the cropping is applied. Finally, a 1 × 1 convolution is used to map each feature vector to three classes.

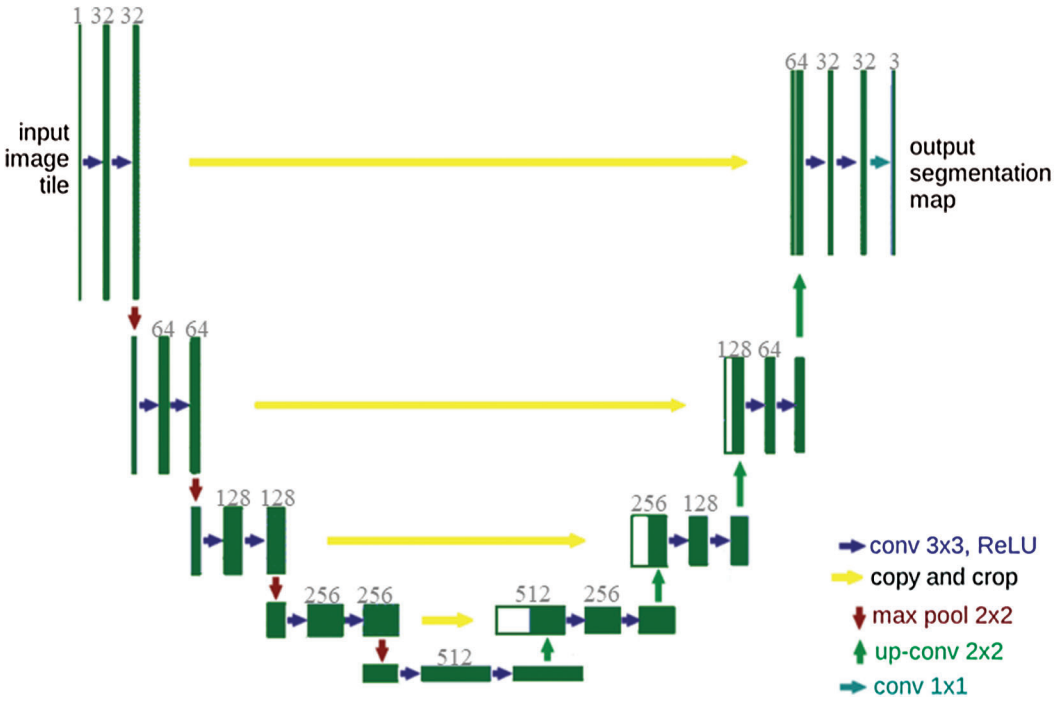

Figure 3: U-Net architecture employed in this work. Each dark green box corresponds to a multi-channel feature map and the number of channels is noted on top of the box. White boxes represent copied feature maps. The arrows denote the different operations

2.2.3 Network Architecture: FC-DenseNet

FC-DenseNet architecture includes dense block layer, transition down layer and transition up layer. For each dense block layer, it consists of batch normalization (BN), then ReLU, a 3 × 3 same convolution (no resolution loss) and dropout with probability $p = 0.2$. The growth rate of the layer is set to k = 16. Similarly,

for each transition down layer, it is composed of BN, then ReLU, a 1 × 1 convolution, dropout with $p = 0.2$ and a non-overlapping max pooling of size 2 × 2. For transition up layer, it consists of a 3 × 3 transposed convolution with stride 2. The architecture used in our work is built from 57 convolutional layers: firstly, one input layer, then 20 in the downsampling path, 5 in the bottleneck and 20 in the upsampling path. Five Transition Down (TD) and 5 Transition Up (TU), each one containing a convolution, is also applied in this architecture. Finally, 1 × 1 convolution followed by a softmax non-linearity is used to provide the per class distribution at each pixel. In Fig. 4, all layers in architecture were summarized.

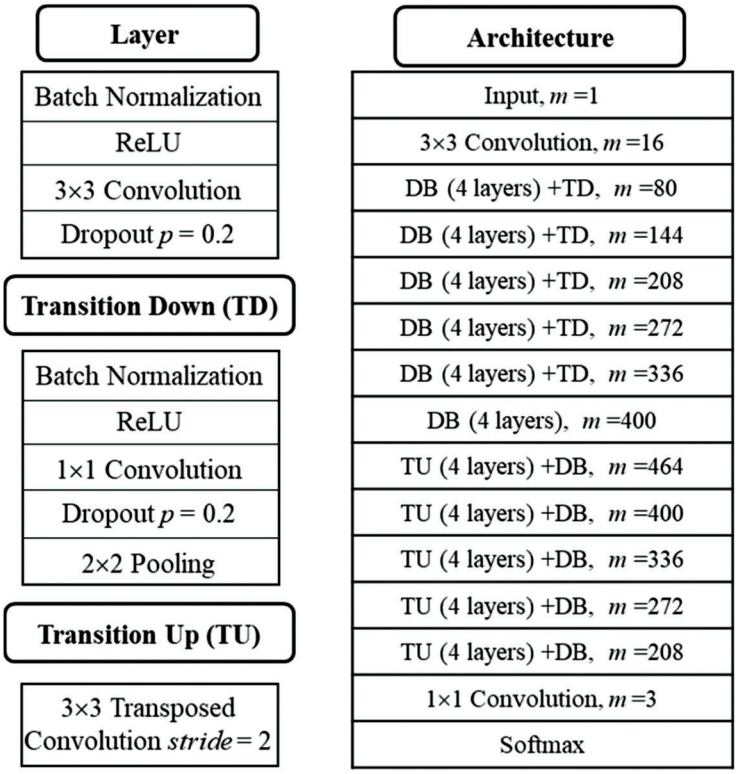

Figure 4: Building blocks of FC-DenseNet and the architecture details of FC-DenseNet57 model used in this work

2.3 Support Vector Machine

Local binary patterns (LBPs), gray level co-occurrence matrices (GLCMs) which contains contrast, correlation, energy and homogeneity, entropy and mean value were calculated as features in this study, which have proven to be the optimal feature combination after trials. A neighborhood window of 11 × 11 pixels was selected in the ROI to calculate all features. The parameters used to calculate the ten features of rotation-invariant LBP are $p = 8$ and R = 1. In the end, 28 features including 10 LBPs, 16 GLGMs, entropy and mean were selected and used in the classification algorithm. These selected features extracted from all pixels in the ROI were assembled into an n × m-dimensional data matrix to feed the SVM, where n equals the number of pixels and m is the length of the feature vector.

2.4 Three-Fold Model Training and Cross Validation Methods

2.4.1 Three-Fold Methods for CNN (U-Net and FC-DenseNet)

The original 18 OCT image datasets were randomly divided into three groups, and the three-fold cross validation was employed to perform model training and testing tasks. First, 2 groups (3 total possible combinations) were chosen as the training set. With augmentation, the training set had 1008 slices. Then

the remaining group served as the testing set. Due to the class imbalance in the proportion of lipid tissue compared to other classes, focus loss function was applied in the model between the gold standard and the output image of our model [27]. And for model selection, 1-standard error rule was performed on the validation data set [28]. The trained model was applied to the testing group to obtain prediction sensitivity, specificity and accuracy (to be defined later). This was done for all 3 grouping choices with each group serving as the testing set once. The random group dividing process was done 10 times and the results obtained were averaged as our final results. Further repeating did not noticeable improvements. All CNN models were performed in Pytorch, version 3.6. Figs. 5b and 5c show seven images of classification result based on U-Net and FC-DenseNet, respectively. The input images were divided into three categories, which are lipid tissue (LT), fibrous tissue (FT), and background (BG, except for ROI).

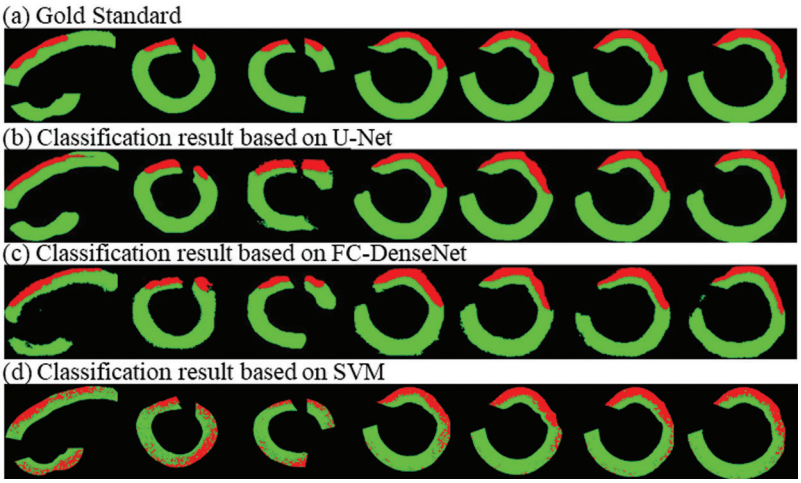

Figure 5: Magnified image classification result. Colors used in images: *Red*—lipid tissue, *Green*—fibrous tissue, *Black*—background. (a) Gold Standard. (b) Classification result based on U-Net. (c) Classification result based on FC-DenseNet. (d) Classification result based on SVM

2.4.2 Three-Fold Methods for SVM

In ROI, plaque cap can be determined once the lipid region is identified. As such, only two types of tissue (pixels) need to be characterized: lipid tissue and fibrous tissue. Fibro-fatty tissue that has no effect on the measurement of the plaque cap was considered as fibrous tissue. Taking the Gaussian Radial Basis function as the kernel function, the least square support vector machine (SL-SVM) classifier was selected to classify lipid tissue and fibrous tissue. As with CNN method, three-fold cross validation was also used in the SVM method. For the training process, 500 pixels were randomly chosen from two classes (lipid and fibrous) in the OCT images of each training group, that is, 12000 pixels ($12 \times 500 \times 2 = 12000$) were obtained from 12 training slices. All pixels in the ROI of the test group OCT images constitute the test set. Five trainings and tests were performed in this way. Fig. 5d shows seven images of classification result based on SVM.

2.5 Contours Extraction

Plaque cap thickness was defined as the distance between extracted lipid contours and lumen contour. In order to measure plaque cap thickness, the lipid contours were extracted using classification result based on U-Net, FC-DenseNet and LS-SVM, respectively. Some scattered pixels in the classification results from the 3 methods were filtered out and smoothing method was applied to obtain clear borders for lipid contours. Parametric active contour model was used for the lipid border detection. Fig. 6 shows contours of seven sample slices (same as Fig. 5) obtained from gold standard and three different classification results.

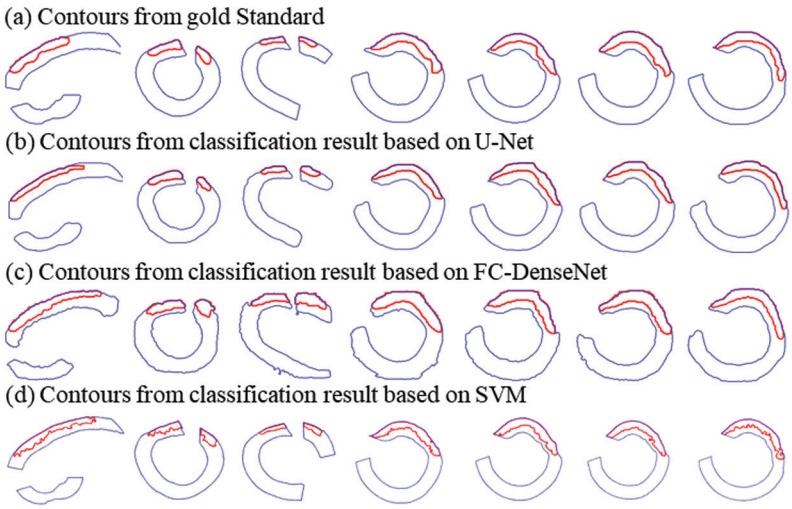

Figure 6: Lipid contours and borders. Color for contour lines: *Red*—lipid, *Blue*—borders. (a) Contours from gold Standard. (b) Contours from classification result based on U-Net. (c) Contours from classification result based on FC-DenseNet. (d) Contours from classification result based on SVM

2.6 Method Comparisons and Data Analysis

The results of segmentation methods were evaluated and compared using the OCT dataset and gold standard at pixel level. Accuracy (Acc), sensitivity (Sen) and specificity (Spe) are defined as follows:

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$Sen = \frac{TP}{TP + FN} \tag{2}$$

$$Spe = \frac{TN}{TN + FP} \tag{3}$$

where TP represents the number of true positive results, TN represents the number of true negative results, FP represents the number of false positive results, and FN represents the number of false negative results. Segmentation accuracy, sensitivity and specificity from our data set using the three methods are presented in the next section.

3 Results

3.1 Accuracy of Tissue Region Detection Using Two CNN Methods

Two CNN network-based methods took the full image as input (not ROI region like SVM method), so it is necessary to test their ability of extracting lipid and fibrous tissue. In this evaluation, lipid tissue and fibrous tissue were considered as positive patterns respectively (prediction target) while background was regarded as negative pattern. Prediction sensitivity (Sen), specificity (Spe) and accuracy (Acc) of tissue region using two methods are shown in Tab. 1. The overall average prediction accuracy from the 18 slices (total pixel sample size: 8580096) was 96.36% and 92.72% for U-Net and FC-DenseNet, respectively. The best accuracy was 97.45% for U-Net and 94.74% for FC-DenseNet. Even for the worst case, the prediction accuracy was 94.35% by U-Net and 86.45% by FC-DenseNet. Obviously, U-Net provided more accurate results than that from FC-DenseNet. FC-DenseNet had considerable poorer sensitivity (84.65% *vs.* 94.02%).

Table 1: Accuracy of tissue region detection based on U-net and FC-DenseNet

	U-Net			FC-DenseNet		
Slices	Sen	Spe	Acc	Sen	Spe	Acc
1	91.89	98.04	96.54	76.68	95.95	91.25
2	84.33	99.42	95.70	87.11	95.56	93.68
3	86.70	98.37	95.57	88.13	95.56	93.82
4	93.69	96.94	96.22	80.32	97.81	93.13
5	93.55	96.94	96.12	91.82	94.85	94.04
6	87.07	99.20	96.31	80.45	98.60	93.58
7	95.70	96.85	96.58	77.31	95.11	90.72
8	97.47	96.85	96.17	88.75	92.81	91.83
9	84.13	96.96	94.35	87.64	96.49	94.74
10	97.51	96.81	97.00	92.65	94.80	94.23
11	98.21	95.21	96.01	85.85	96.48	93.59
12	98.55	95.55	96.35	87.99	92.74	91.40
13	95.38	96.85	96.46	74.44	97.28	91.79
14	98.21	96.65	97.07	90.28	94.23	93.29
15	97.74	97.34	97.45	94.57	93.47	93.69
16	97.85	96.99	97.23	87.67	95.24	93.23
17	97.81	95.99	96.50	67.33	93.66	86.45
18	96.57	97.00	96.88	84.79	98.32	94.45
Average	**94.02**	**97.05**	**96.36**	**84.65**	**95.50**	**92.72**

3.2 Accuracy of Lipid and Overall Classification

The performed methods are able to extract lipid and fibrous tissue directly from OCT images, and sort each pixel into its corresponding category. Tab. 2 records the prediction accuracy of the lipid and the prediction accuracy of the input images (overall prediction accuracy). Tab. 2 shows that for the prediction accuracy of lipid, the best result was given by SVM method, which is 91.29%. However, SVM method gave the worst overall prediction accuracy (a total of 2156747 pixels) 81.84%. CNN scored much higher on the overall accuracy rate, 95.40% for U-Net and 91.14% for FC-DenseNet, which were 16.57% and 11.36% higher than that from SVM, respectively. Compared to other objects, the percentage of lipid pixels in the OCT image is much lower, causing class imbalance problems. Although the focal loss function was applied to improve it, the lipid accuracy score of CNN was lower than that of SVM.

3.3 Evaluation of the Discrimination Between Lipid Tissue (LT) and Fibrous Tissue (FT)

Considered LT and FT as positive and negative respectively to further evaluate the difference between them. Tab. 3 shows the results of differentiating LT and FT. The overall average accuracy (*Acc*) were 95.58%, 92.33% and 81.84% for U-Net, FC-DenseNet and SVM, respectively. Method based on FC-DenseNet depicted the best average sensitivity (*Sen)* and U-Net provided the best average specificity (*Spe)*. Overall, both CNN and SVM provided good discrimination between LT and FT.

Table 2: Classification accuracy of objects using CNN and SVM

	U-Net		FC-DenseNet		SVM	
Slices	Lipid	Overall	Lipid	Overall	Lipid	Overall
1	68.26	95.35	65.69	88.79	93.12	75.05
2	61.22	94.74	81.24	91.27	89.16	68.90
3	70.60	94.63	85.27	92.66	76.41	66.43
4	83.41	95.68	77.12	92.05	80.80	67.75
5	91.21	95.58	90.94	92.73	96.71	70.33
6	78.54	95.88	80.82	92.64	90.36	75.56
7	86.83	95.96	65.32	88.32	88.44	69.66
8	82.71	95.60	79.58	89.73	80.15	83.41
9	98.30	93.60	86.16	93.98	91.80	83.97
10	93.92	95.82	87.37	92.89	96.48	87.84
11	83.65	94.67	81.70	92.08	96.97	85.22
12	81.75	95.22	84.03	90.01	96.62	91.33
13	90.20	95.08	62.31	88.91	98.30	91.97
14	79.02	95.74	82.55	91.53	94.41	91.57
15	84.17	96.31	86.18	92.14	96.37	91.29
16	88.91	96.01	81.40	91.86	95.93	89.62
17	78.08	95.21	57.40	85.16	90.45	91.36
18	90.41	96.06	85.29	93.75	90.65	91.91
Average	**82.84**	**95.40**	**78.91**	**91.14**	**91.29**	**81.84**

3.4 Cap Thickness Quantification: Mean and Minimum Cap Thickness

Tabs. 4 and 5 gave the mean and minimum cap thickness for the 18 OCT slices obtained from the three segmentation methods (U-Net, FC-DenseNet, SVM). Using the gold standard as a measure, U-Net, FC-DenseNet, generally gave better results than that from SVM. The average of mean cap thickness from gold standard and three methods were 0.739, 0.740, 0.703 and 0.633, respectively. The average errors of U-Net, FC-DenseNet and SVM from the 18 slices were 8.83%, 10.71% and 15.85%, respectively. Average of minimum cap thickness from gold standard and three methods were 0.513, 0.556, 0.489 and 0.433, respectively. The average relative error of minimum cap thickness from 18 slices of U-Net, FC-DenseNet and SVM were 17.46%, 13.06% and 22.20%, respectively. These results indicated that U-Net method provided the best mean cap thickness results. FC-DenseNet method provided the best minimum cap thickness results, and SVM method performed poorly.

4 Discussion

4.1 Signification

Plaque cap thickness, lipid core size and plaque burden have been identified as key predictors of plaque vulnerability and likelihood of rupture. Segmentation and classification of plaque components of OCT images is able to provide precise plaque morphology with cap thickness and lipid core size which are the most closely watched features in patient management. Image-based biomechanical plaque models may be

Table 3: Evaluation of the discrimination between LT and FT

	U-Net			FC-DenseNet			SVM		
Slices	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe	Acc
1	74.08	98.07	94.69	93.15	85.87	86.91	71.33	88.43	75.05
2	71.06	99.00	95.38	65.35	95.86	87.57	64.44	83.75	68.90
3	81.45	98.43	95.53	74.81	96.57	94.31	61.30	64.33	66.43
4	91.99	99.16	97.42	98.02	94.44	95.00	62.55	69.62	67.75
5	95.61	97.72	97.62	95.12	94.62	94.70	67.86	87.11	70.33
6	89.77	98.47	97.89	95.15	95.91	95.78	73.13	79.19	75.56
7	88.40	98.28	97.21	91.27	86.75	87.43	67.86	80.79	69.66
8	84.62	98.84	97.04	90.95	90.20	90.24	82.30	66.25	83.41
9	99.56	95.16	95.66	72.17	99.19	95.57	81.85	84.32	83.97
10	96.68	95.25	95.49	95.47	94.42	94.57	87.47	96.15	87.84
11	88.97	95.81	94.86	92.22	93.91	93.56	85.10	96.46	85.22
12	83.93	98.62	95.66	82.16	98.30	94.41	91.01	96.07	91.33
13	92.18	95.18	94.59	98.81	81.30	83.87	91.06	97.53	91.97
14	81.23	98.73	94.99	89.41	91.99	91.80	91.11	93.33	91.57
15	86.19	98.40	95.74	99.67	91.10	91.92	90.63	95.58	91.29
16	90.79	96.76	95.50	90.68	95.13	94.14	89.33	95.25	89.62
17	84.23	98.23	95.32	84.83	96.67	92.99	90.59	88.91	91.36
18	92.12	98.42	97.03	90.56	98.78	97.11	91.47	90.40	91.91
Ave.	**87.38**	**97.70**	**95.98**	**88.88**	**93.39**	**92.33**	**80.02**	**86.30**	**81.84**

used to research and quantify plaque mechanical stress and strain conditions, identify possible mechanical factors associated with plaque rupture, assess plaque rupture risk and predict rupture location. According to the research result proposed by Li et al. [29], in a 40% degree stenotic artery, a reduction of the cap thickness from 0.4 mm to 0.2 mm could cause an increase in plaque stress from 141 kPa to 409 kPa. Hence, the accuracy and efficiency of OCT image segmentation and quantification of cap thickness play a significant role in clinical diagnosis, image analysis and vulnerable plaque research.

4.2 Comparison of the Three Methods

The proposed methods are able to segment the lipid component automatically with high accuracy. It is worth mentioning that the CNN method using the U-Net or FC-DenseNet model is an "end-to-end" method, and it is not necessary to detect the lumen and outer boundary before classification, which makes it relative to previous segmentation methods has certain advantages. In addition, the proposed CNN method does not require a large amount of image pre-processing and post-processing, thereby significantly improving the segmentation efficiency. In CNN methods, U-Net model performed better in accurately identifying lipid components, whereas FC-DenseNet model had fewer parameters and required less storage memory.

Obviously, in multiple indicators, CNN-based methods are better than SVM-based method. For example, CNN-based methods have the advantage of unified implementation of feature extraction and classification in a supervised mode. However, in SVM-based method, feature extraction and classification

Table 4: Mean cap thickness for 18 slices from four methods

Mean Cap Thickness (mm)							
Slice	Gold Standard	U-Net	Error	FC	Error	SVM	Error
1	0.770	0.713	7.39%	0.725	5.87%	0.619	19.55%
2	0.903	0.829	8.19%	0.795	12.00%	0.637	29.43%
3	0.660	0.751	13.76%	0.639	3.12%	0.601	9.01%
4	0.697	0.645	7.44%	0.624	10.44%	0.554	20.51%
5	0.611	0.731	19.61%	0.514	15.88%	0.576	5.73%
6	0.621	0.657	5.80%	0.555	10.58%	0.572	7.94%
7	0.634	0.706	11.43%	0.496	21.74%	0.540	14.78%
8	0.581	0.607	4.43%	0.789	35.82%	0.617	6.19%
9	0.650	0.685	5.38%	0.721	10.95%	0.647	0.48%
10	0.884	0.969	9.67%	0.811	8.30%	0.722	18.38%
11	1.048	0.979	6.58%	0.959	8.51%	0.727	30.62%
12	0.988	0.979	0.83%	0.925	6.39%	0.807	18.26%
13	0.499	0.550	10.25%	0.565	13.14%	0.624	24.96%
14	0.796	0.759	4.56%	0.710	10.73%	0.448	43.65%
15	0.674	0.737	9.28%	0.650	3.60%	0.681	1.05%
16	0.736	0.642	12.76%	0.711	3.39%	0.590	19.85%
17	0.747	0.704	5.81%	0.718	3.94%	0.696	6.81%
18	0.808	0.681	15.70%	0.740	8.44%	0.743	8.07%
Ave.	**0.739**	**0.740**	**8.83%**	**0.703**	**10.71%**	**0.633**	**15.85%**

have to be performed separately. Moreover, SVM cannot extract shallow and deep features and combine them like as what CNN methods could do. Setting ROI is a necessary image preprocessing step for SVM methods. Another limitation of SVM methods is that parameters of SVM cannot be learned automatically.

4.3 Computing Time and Time Complexities

Machine learning algorithms are very time consuming and normally a GPU computer is needed. For the two CNN-based methods, all models were implemented in Pytorch, version 1.1.0 (GPU computer with Titan XP 12G, Intel i7 8700k, 6 core/12 threads). For U-Net, the training time for 1 epoch was 9–10 minutes (training 1002 slices) and the total training time was 15–17 hours for 100 epochs, which can ensure that model is trained to converge . In terms of testing time, which closely related to time complexity, the testing time for one original image was 1–2 seconds, and the total testing time for the testing set (6 original images) was 6–12 seconds. The training and testing time of FC-DenseNet were slightly longer than U-Net. The training time of FC-DenseNet for 1 epoch was 10–11 minutes and the total training time is 16–18 hours. And for testing time, the testing time for 1 image is 1.5–2.5 seconds, and the overall testing time is 9–15 seconds. SVM-based method took much less time than CNN-based methods to train and test. The SVM-based method was performed in MATLAB R2015a (Dell Precision T5810 E5-1607V4 32G RAM) and the total time including feature extraction, training and testing was about 5–6 hours.

Table 5: Minimum cap thickness for 18 slices from four methods

Min Cap Thickness (mm)							
Slice	Gold Standard	U-Net	Error	FC	Error	SVM	Error
1	0.512	0.552	7.71%	0.538	4.97%	0.481	6.03%
2	0.475	0.464	2.27%	0.625	31.59%	0.458	3.53%
3	0.322	0.446	38.34%	0.400	24.10%	0.378	17.39%
4	0.389	0.410	5.50%	0.330	15.12%	0.292	24.91%
5	0.311	0.524	68.44%	0.354	13.73%	0.332	6.84%
6	0.372	0.420	12.92%	0.361	3.09%	0.347	6.85%
7	0.389	0.527	35.48%	0.369	5.15%	0.280	27.97%
8	0.443	0.430	2.87%	0.438	1.03%	0.434	2.03%
9	0.519	0.536	3.35%	0.567	9.37%	0.449	13.34%
10	0.761	0.899	18.10%	0.737	3.24%	0.532	30.06%
11	0.814	0.781	4.00%	0.608	25.33%	0.535	34.30%
12	0.752	0.809	7.54%	0.737	1.96%	0.532	29.19%
13	0.256	0.362	41.55%	0.266	3.96%	0.458	79.07%
14	0.553	0.574	3.83%	0.596	7.76%	0.377	31.86%
15	0.500	0.633	26.41%	0.312	37.62%	0.541	8.04%
16	0.652	0.534	18.11%	0.517	20.69%	0.417	36.05%
17	0.628	0.563	10.34%	0.509	18.97%	0.485	22.78%
18	0.586	0.542	7.43%	0.543	7.32%	0.472	19.38%
Ave.	**0.513**	**0.556**	**17.46%**	**0.489**	**13.06%**	**0.433**	**22.20%**

4.4 Large Relative Errors for Minimum Cap Thickness

In Tab. 5, the relative error reached to 68% for U-net method (Slice 5) and 79% for SVM method (Slice 13). One may be wondering how we could justify our methods are proper for quantifying the Min Cap thickness. The answer to this concern is actually non-trivial.

One reason that minimum cap thickness may have larger errors is that minimum cap thickness locations of a given slice may be different for different methods (see Fig. 5). The contours and the cap thickness derived from those contours given by Fig. 6 demonstrated that clearly. Fig. 6 also explains why mean cap thickness had smaller errors.

Errors reported in Tab. 5 are relative errors which may be magnified by the small minimum cap thickness of the slices from the gold standard. The absolute error of Slice 5 by U-Net was 0.213 mm. The absolute error of Slice 13 by SVM was 0.202 mm. Both were smaller than VH-IVUS resolution limit.

One of the biggest challenges for segmentation of *in vivo* human coronary plaque images and quantifying thin cap thickness is lack of real gold standard, i.e., histopathological data of the plaques from which images were taken. In fact, thin cap thickness is the #1 closely watched risk indicator for vulnerable plaque identification. A threshold cap thickness of 65 μm is currently accepted as identification of a vulnerable plaque. VH-IVUS technology was developed and validated by limited

histological data, and then was accepted as virtual histology. However, IVUS resolution was limited to about 200–300 μm which is not enough to capture plaques with cap thickness around 65 μm. That means any differences smaller than 200–300 μm provided by VH-IVUS could be considered random errors. OCT has resolution of 5–10 μm and could be the long-waited tool to identify vulnerable plaques. However, its segmentation remains to be the current challenge. In this paper, VH-IVUS was used as references by skilled radiologists to generate segmentation gold standard for OCT images. As it can be seen from Fig. 1, manual segmentation of OCT images is very challenging and could be subjected human errors and influences from VH-IVUS. That being said, this is still the best researchers could do at present time since human coronary plaque tissues corresponding to *in vivo* images are not available to generate histopathological data to serve as real gold standard.

When our models are trained by large data set, we may have reasons to hope that cap thickness determined by machine learning methods could be meaningful and reliable. Using histopathological data to train our models could also provide stronger results. Manual segmentation by more experts could also reduce the error of the gold standard itself.

4.5 Limitations

In this study, a more accurate gold standard which combined both OCT and VH-IVUS images was proposed, however, the lack of histology is a disadvantage. Histological data remain the current gold standard in tissue detection. Histology could provide more effective and additional information which could be applied to improve and optimize the reliability and applicability of our method. Moreover, since the limitation of a small sample size of OCT dataset, more data would be desirable to improve the accuracy and reliability of methods performed in our study.

4.6 Future Work

Automatic segmentation of medical images, in particular, OCT images of in vivo human coronary plaques, is of utmost importance in diagnosis, prevention, prediction, treatment and management of related cardiovascular diseases including heart attack and stroke. We will continue to search for better segmentation methods, develop automatic algorithms, and utilize the segmentation results in mechanical model construction for plaque progression and rupture predictions. Other deep learning methods including FCN, R-FCN, GAN will be considered in our future effort. As the U-Net and FC-DenseNet architectures have mutual advantages in different levels of feature extraction and combination, we plan to combine the two architectures (Dense-Unet) and look forward to better results. In addition, in view of the limitations of a large amount of noise and small amount of data in medical images, a direction worthy of research is to pre-train deep learning models on ex vivo images or non-medical images where true gold standard (histological data) could be available. Based on the segmentation results from those efforts, computational mechanical vascular models based on OCT segmentation images will be established and used for plaque stress analysis and prediction of plate growth and rupture. Those models would be the first models where truly reliable cap thickness could be used to generate accurate and reliable plaque stress/strain data and make plaque growth and vulnerability predictions which were not possible with IVUS 200–300 μm resolution limitations.

Funding Statement: This research was supported in part by National Sciences Foundation of China grants 11972117 and 11672001 and a Jiangsu Province Science and Technology Agency grant BE2016785.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Weintraub, W. S., Daniels, S. R., Burke, L. E., Franklin, B. A., Goff Jr, D. C. et al. (2011). Value of primordial and primary prevention for cardiovascular disease. a policy statement from the American Heart Association. *Circulation, 124(8),* 967–990. DOI 10.1161/CIR.0b013e3182285a81.
2. Virmani, R., Kolodgie, F. D., Burke, A. P., Farb, A., Schwartz, S. M. (2000). Lessons from sudden coronary death. *Arteriosclerosis, Thrombosis, and Vascular Biology, 20(5),* 1262–1275. DOI 10.1161/01.ATV.20.5.1262.
3. Wahle, A., Lopez, J. J., Olszewski, M. E., Vigmostad, S. C., Chandran, K. B. et al. (2006). Plaque development, vessel curvature, and wall shear stress in coronary arteries assessed by X-ray angiography and intravascular ultrasound. *Medical Image Analysis, 10(4),* 615–631. DOI 10.1016/j.media.2006.03.002.
4. Schaar, J. A., Mastik, F., Regar, E., den Uil, C. A., Gijsen, F. J. et al. (2007). Current diagnostic modalities for vulnerable plaque detection. *Current Pharmaceutical Design, 13(10),* 995–1001. DOI 10.2174/138161207780487511.
5. Nair, A., Kuban, B. D., Tuzcu, E. M., Schoenhagen, P., Nissen, S. E. et al. (2002). Coronary plaque classification with intravascular ultrasound radiofrequency data analysis. *Circulation, 106(17),* 2200–2206. DOI 10.1161/01.CIR.0000035654.18341.5E.
6. Finet, G., Ohayon, J., Rioufol, G. (2004). Biomechanical interaction between cap thickness, lipid core composition and blood pressure in vulnerable coronary plaque: impact on stability or instability. *Coronary Artery Disease, 15(1),* 13–20. DOI 10.1097/00019501-200402000-00003.
7. Kume, T., Akasaka, T., Kawamoto, T., Watanabe, N., Toyota, E. et al. (2006). Assessment of coronary arterial plaque by optical coherence tomography. *American Journal of Cardiology, 97(8),* 1172–1175. DOI 10.1016/j.amjcard.2005.11.035.
8. Van Soest, G., Goderie, T. P., Regar, E., Koljenovic, S., van Leenders, A. G. J. et al. (2010). Atherosclerotic tissue characterization *in vivo* by optical coherence tomography attenuation imaging. *Journal of Biomedical Optics, 15(1),* 011105. DOI 10.1117/1.3280271.
9. Shalev, R., Prabhu, D., Tanaka, K., Rollins, A. M., Costa, M. et al. (2015). Intravascular optical coherence tomography image analysis method. *41st Annual Northeast Biomedical Engineering Conference IEEE*, pp. 1–2.
10. Guo, X., Tang, D., Molony, D., Yang, C., Samady, H. et al. (2019). A machine learning-based method for intracoronary OCT segmentation and vulnerable coronary plaque cap thickness quantification. *International Journal of Computational Methods, 16(3),* 1842008. DOI 10.1142/S0219876218420082.
11. Athanasiou, L. S., Bourantas, C. V., Rigas, G., Sakellarios, A. I., Exarchos, T. P. et al. (2014). Methodology for fully automated segmentation and plaque characterization in intracoronary optical coherence tomography images. *Journal of Biomedical Optics, 19(2),* 026009. DOI 10.1117/1.JBO.19.2.026009.
12. Abdolmanafi, A., Duong, L., Dahdah, N., Adib, I. R., Cheriet, F. (2018). Characterization of coronary artery pathological formations from OCT imaging using deep learning. *Biomedical Optics Express, 9(10),* 4936–4960. DOI 10.1364/BOE.9.004936.
13. He, S., Zheng, J., Maehara, A., Mintz, G., Tang, D. et al. (2018). Convolutional neural network based automatic plaque characterization for intracoronary optical coherence tomography images. *Medical Imaging 2018. Image Processing, 10574,* 1057432.
14. Zhang, C., Li, H., Guo, X., Molony, D., Guo, X. et al. (2019). Convolution neural networks and support vector machines for automatic segmentation of intracoronary optical coherence tomography. *Molecular & Cellular Biomechanics, 16(2),* 153–161. DOI 10.32604/mcb.2019.06873.
15. Athanasiou, L. S., Olender, M. L., José, M., Ben-Assa, E., Edelman, E. R. (2019). A deep learning approach to classify atherosclerosis using intracoronary optical coherence tomography. *Medical Imaging 2019. Computer-Aided Diagnosis, 10950,* 109500N.
16. Long, J., Shelhamer, E., Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440.
17. Badrinarayanan, V., Kendall, A., Cipolla, R. (2017). SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12),* 2481–2495. DOI 10.1109/TPAMI.2016.2644615.

18. Ronneberger, O., Fischer, P., Brox, T. (2015). U-net: convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham, Springer, 234–241.
19. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii*, 4700–4708.
20. Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., Bengio, Y. (2017). The one hundred layers tiramisu. Fully convolutional densenets for semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii*, 11–19.
21. Kubo, T., Maehara, A., Mintz, G. S., Doi, H., Tsujita, K. et al. (2010). The dynamic nature of coronary artery lesion morphology assessed by serial virtual histology intravascular ultrasound tissue characterization. *Journal of the American College of Cardiology, 55(15),* 1590–1597. DOI 10.1016/j.jacc.2009.07.078.
22. Guo, X., Zhu, J., Maehara, A., Monoly, D., Samady, H. et al. (2017). Quantify patient-specific coronary material property and its impact on stress/strain calculations using *in vivo* IVUS data and 3D FSI models: a pilot study. *Biomechanics and Modeling in Mechanobiology, 16(1),* 333–344. DOI 10.1007/s10237-016-0820-3.
23. Räber, L., Heo, J. H., Radu, M. D., Garcia-Garcia, H. M., Stefanini, G. G. et al. (2012). Offline fusion of co-registered intravascular ultrasound and frequency domain optical coherence tomography images for the analysis of human atherosclerotic plaques. *EuroIntervention, 8(1),* 98–108. DOI 10.4244/EIJV8I1A16.
24. Chawla, N. V., Bowyer, K. W., Hall, L. O., Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16,* 321–357. DOI 10.1613/jair.953.
25. Simard, P. Y., Steinkraus, D., Platt, J. C. (2003). Best practices for convolutional neural networks applied to visual document analysis. *Proceedings of the Seventh International Conference on Document Analysis and Recognition, 2,* 958–962.
26. Cireşan, D. C., Meier, U., Gambardella, L. M., Schmidhuber, J. (2010). Deep, big, simple neural nets for handwritten digit recognition. *Neural Computation, 22(12),* 3207–3220. DOI 10.1162/NECO_a_00052.
27. Lin, T. Y., Goyal, P., Girshick, R., He, K., Dollár, P. (2017). Focal loss for dense object detection. *Transactions on Pattern Analysis and Machine Intelligence, 42(2),* 318–327.
28. Friedman, J., Hastie, T., Tibshirani, R. (2001). *The elements of statistical learning,* 1(10)*,* Springer Series in Statistics, New York.
29. Li, Z. Y., Howarth, S. P., Tang, T., Gillard, J. H. (2006). How critical is fibrous cap thickness to carotid plaque stability? A flow-plaque interaction model. *Stroke, 37(5),* 1195–1199. DOI 10.1161/01.STR.0000217331.61083.3b.