Learning the Lumen Border using a Convolutional Neural Networks classifier

José Marone*[1], Simone Balocco[2,3], Marc Bolaños[2,3], José Massa[1], and Petia Radeva[2,3]

[1] INTIA, Dept. Computación y Sistemas, UNCPBA, Buenos Aires, Argentina
[2] Dept. Matematica i Informatica, UB, Barcelona, Spain
[3] Computer Vision Center, Bellaterra, Spain

Abstract. IntraVascular UltraSound (IVUS) is a technique allowing the diagnosis of coronary plaque. An accurate (semi-)automatic assessment of the luminal contours could speed up the diagnosis. In most of the approaches, the information on the vessel shape is obtained combining a supervised learning step with a local refinement algorithm. In this paper, we explore for the first time, the use of a Convolutional Neural Networks (CNN) architecture that on one hand is able to extract the optimal image features and at the same time can serve as a supervised classifier to detect the lumen border in IVUS images. The main limitation of CNN, relies on the fact that this technique requires a large amount of training data due to the huge amount of parameters that it has. To solve this issue, we introduce a patch classification approach to generate an extended training-set from a few annotated images. An accuracy of 93% and *F-score* of 71% was obtained with this technique, even when it was applied to challenging frames containig calcified plaques, stents and catheter shadows.

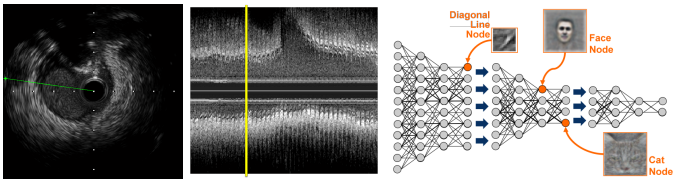
1 Introduction

Atherosclerosis is a progressive disease affecting arterial blood vessels caused by an inflammatory response of the wall followed by the accumulation of fat and tissues on the vascular membrane. Arteriosclerosis is particularly dangerous because, if not treated, it can totally obstruct the artery and lead to stroke. IntraVascular UltraSound (IVUS) is a clinically available imaging technique providing real-time hi-resolution cross-sectional sequences of images of the coronary artery in patients. It is an essential mean to quantify and characterize coronary plaque, for diagnostic purposes and for guiding percutaneous coronary intervention (PCI).

The procedure for the acquisition of an IVUS sequence consists of inserting an ultrasound emitter, carried by a catheter, into the arterial vessel and dragging the probe from the distal to the proximal position. This procedure generates an orderer set of high amount of frames denoted as pullback. The standard IVUS image is a 360-degree tomographic cross-sectional view of the

* contact author marone@exa.unicen.edu.ar

vessel walls, denoted as *short-axis view*, which allows an accurate assessment of vessel morphology. Given an angular position on the *short-axis view* (indicated in Figure 1-a by a green line), the corresponding *longitudinal view* can be generated by considering the gray-level values of the sequence along the diameter at the chosen angle (Figure 1-b). This longitudinal image depicts the morphology of the vessel section according to the selected orientation.



(a) (b) (c)

Fig. 1. Short-axis view of a vessel in cartesian coordinates (a) and longitudinal view of the same pullback (b). General Scheme of a CNN classifier (c).

The IVUS sequence can be represented as $I(x, y; t)$, where x and y are the spatial coordinates of the image, and t is the index of a frame in a pullback.

1.1 Related Work

Since a typical pullback contains more than 3000 IVUS frames, an accurate (semi-)automatic assessment of lumen and media contours is highly desirable to reduce the workload of the medical doctor, and to speed up the diagnosis.

Manual lumen and media segmentation is a laborious task that suffers from inter- and intra-observer variabilities due to the high amount of noise and artifacts present on the IVUS images. Consequently, much work has been performed on (semi-)automated IVUS image processing as illustrated by Katouzian and Balocco in recent reviews [9, 1]. Observing such surveys, in most of the approaches, the information on the vessel shape is obtained in two phases: a) pre-processing supervised learning step is initially applied in order to roughly identify the region of the image containing the contour, b) an active contours technique is then initialized using the classification results in order to locally segment the membrane border.

One of the most promising techniques for supervised learning are the Convolutional Neural Network (CNN), which are mathematical models consisting of multiple layers of small neuron collections. Every neuron in a convolutional layer represents the response of a filter applied to the previous layer. The job of this neuron is to pass this response through some non-linearity. The area of the previous layer that this filter is applied to, is called the *receptive field* of that neuron. CNNs were initially introduced in a 1980 paper by Kunihiko Fukushima [4], later LeCun [11], successfully applied these concepts to computer vision problems. In recent years, the rise of efficient GPU computing made possible to

train larger networks increasing the number of layers [7]. One major advantage of CNNs is the use of shared weights in convolutional layers, which means that the same filter is applied for different pixels in the layer, this both reduces required memory size and improves performance [10]. Another advantage is that the network is responsible for learning the filters that in traditional algorithms were hand-engineered, which means that the network has the ability to learn by itself which are the most discriminative shapes, colours and textures in the image for the problem at hand.

The ability of a CNN to recognize complex patterns varies depending on the number of layers. For instance, the first convolution layer extracts the low-level features, like edges, lines and corners, while the deeper the layers, the higher-level is the information they provide (see Figure 1-c). In each layer, the parameters of every convolution kernel is trained by the back-propagation algorithm. In order to reduce the dimensionality problem, pooling layers compute the maximum or average value of a particular feature over a region of the image, hence it chooses only the best features and eliminates the others.

The main limitation of using CNN, is that this technique requires a large amount of training data due to the huge amount of parameters that it has. A common solution is to train the network on a large data-set, once the network parameters converged, an additional training step is performed using the in-domain data to fine-tune the network weights. This allows convolutional networks to be successfully applied to problems with small training sets.

In medical images applications, several approaches of CNN use have been proposed [5], being a good example the one from Cernazanu-Glavan and Holban [3], who designed a convolutional network able to segment bone struture in X-ray images. Their approach consists in a CNN with three convolutional, three max-pooling and two fully-connected layers that offer a binary classification. In [2], Brebisson and Montana apply a similar approach, but on MRI brain images, classifying each voxel to its corresponding anatomical region from multiple 2D patches, a 3D patch and a centroid distances vector. Recently, Havaei et. al. [6] presented a fully automatic brain tumor segmentation method based on CNNs, high performance is achieved with the help of a novel two-pathway architecture (which can model both the local details and global context). Up to our knowledge, no work has been applied on IVUS segmentation until now.

1.2 Motivation

In this paper, we explore the use of a CNN architecture to detect different patterns in IVUS images. We focus on the analysis of a novel machine learning method aimed at classifying the pixel containing the lumen membrane. The main advantage of CNN with respect to the other classifiers is that it is able to extract contextual information around of a target pixel. We decided to leave for a future study the complete lumen segmentation and the comparison versus the other state of the art segmentation methods.

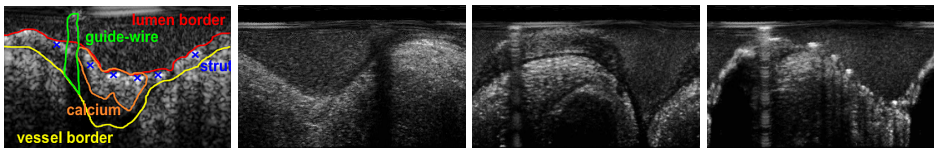
Although some recent CNN architectures provide the ability to directly segment complete images [13], they need thousands of annotated samples (i.e. images with their corresponding desired segmentations) to be trained. Instead,

by using a patch classification approach, we are able to generate an extended training-set from a few annotated images thus allowing the process of CNN training.

2 Method

2.1 Materials

One hundred images have been selected from 7 IVUS pullbacks of different patients. The Imaging System used for the acquisition is an iLab IVUS (Boston Scientific, Freemont), equipped with a 40 MHz catheter Atlantis SR 40 Pro. The IVUS images were selected in order to represent typical IVUS configurations, such as, healthy vessels, eccentric (calcified, fibrotic and lipidic) plaques, catheter shadows, and frames containing a stent, as illustrated in Figure 2.



(a) (b) (c) (d)

Fig. 2. IVUS image samples in polar view. In (a), main areas of interest, (b) normal, (c) calcified plaque, (d) guide-wire shadow and stent struts

The IVUS frames were converted in polar coordinates and having height and width of 255 x 360 pixels. In general, the exploration depth of the IVUS image chosen by the physician led to polar images in which the lumen area occupies only the upper part of the frame while the vessel border occupies the remaining bottom area.

2.2 CNN Architecture

In order to identify the lumen border contour, a pixel-wise classification was performed. Instead of classifying all the image context at once, the classifier was optimized to be able to identify specific image details. In our experiment, the network has been trained to discriminate between patches belonging to the lumen border or not.

Considering the nature of the IVUS images, encoded in grayscale values and having very basic texture patterns, the strategy proposed by [3] was chosen. We used the LeNet architecture [12] as a base model. This network consists of two convolutional layers and is only able to learn small and local patterns (suitable for the original digit classification problem). In order to provide the ability to capture local and global patterns, which is necessary for IVUS images, a third convolutional layer with the same parameters as the second one was introduced. An overview of our CNN architecture is presented in Figure 3.

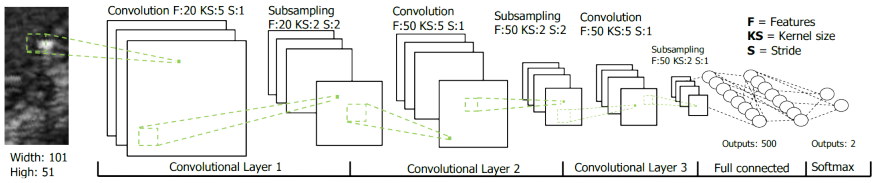


Fig. 3. Proposed CNN Architecture

2.3 Training-set patches extraction

On each IVUS image, a curve defining the interface between the lumen and the vessel has been manually annotated by an expert, along with the catheter shadow area (Figure 4-a). The region of the lumen curve, hidden behind the catheter shadow, has been removed, since in that region the border outline is ambiguous. The obtained contour was enlarged by morphological dilation to 5 pixels wide in order to cover all the thickness of the visible lumen membrane. In such way, the region of the image used as Ground Truth mask shown in Figure 4-b (red curve) has been obtained.

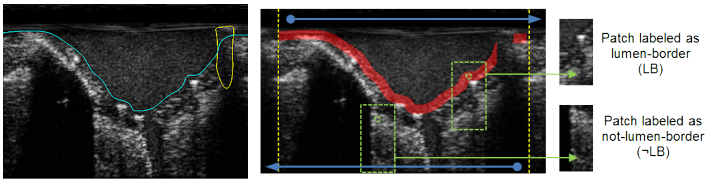
In order to extract training patches from the pixel lying close to the image borders, each IVUS frame has been extended on each side by adding a strip composed of pixels from the opposite side of the image, mirrored with respect to the image axis as shown by the yellow dotted lines in Figure 4-b. For the upper edge, a black patch was added since the catheter area does not contain any relevant textural information.

Patches with sizes of 101x51 pixeles (0.65 mm x 0.32 mm) were extracted. The set of patches is obtained following a sliding window strategy, i.e. scanning the entire image from the top left to the bottom right pixel and cropping the area surrounding $I_{x,y}$ named as *target pixel* (*tp*). An uneven size of the patch is chosen: the bounding box of the patch crop is horizontally centered with respect to the *tp*, but vertically located at 25% of the crop height. The size of the patch has been chosen in such way that the CNN will be able to learn that the lumen is always present above of its visible membrane and to cover the surrounding tissues such as media membrane and plaque. Each patch is finally labeled as lumen border depending if *tp* belongs to the lumen-border (LB) mask or not (¬LB), (Figure 4-b).

The available IVUS images were grouped in three non-overlapping sets: training (TrS), validation (VS) and test (TeS), composed of 18, 12 and 70 images, respectively. In such way considering 55.800 patches per image, we obtain around 1.000.000 and 500.000 patches, for the TrS and VS data-sets, respectively.

2.4 Numerical implementation

The Caffe [8] framework was used to train the neural network. To speed up the training time and to keep the memory usage in optimum levels a GPU GeForce GTX 780 was used. The network training was parametrized with a learning rate (lr) of 10^{-6} a momentum of $1 - lr$ and a weight decay of $5 \cdot 10^{-4}$ and



(a) (b)

Fig. 4. In (a), IVUS image annotations, indicating lumen border and guide shade are shown. In (b), two exemplar patches (squared rectangles), when the *tp* (green circle) falls into the lumen-membrane area (in red), it is labeled as lumen-border. Otherwise it is annotated as not-lumen-border. The blue arrows indicate that the leftmost and rightmost texture stripes have been copied and mirrored.

parallelized in batches of 50 patches. The criterion to select an optimal network was a trade-off between a high *accuracy* and low *loss* value, tested every 500 training iterations. The optimal training configuration was reached as a snapshot at iteration 14000 after only 30 minutes, the *accuracy* and *loss* on VS was 94% and 0.15, respectively.

3 Results

Classification performance in several IVUS cases (including stent, plaque or guide shadow frame) was qualitative illustrated in Figure 5. Additionally, the results of the classification obtained over the test-set (TeS) were reported in Table 1.

3.1 Qualitative results

In order to fairly illustrate the classification performance, the results were grouped in three categories, namely "good", "average", and "poor" classification results, corresponding to frames having a $F - score$ ranging between 0 - 0.5 (poor), 0.5 - 0.75 (average) and 0.75 - 1 (good), respectively. The three most representative images of each group are presented in Figure 5.

In most of the frames the result of the classification follows the ground-truth curve. Although the appearance of the lumen and media membranes might be similar in some cases, the pixels belonging to the media incorrectly classified as lumen are few as can be seen in Figure 5-f. The classification of the lumen border is poor only when the lumen-border is blurry (Figure 5-g and 5-h).

The robustness of the approach can be appreciated in several challenging clinical cases such as frames containing stent or calcium (Figure 5-b and 5-c). The low performance obtained in the upper area of the image observed in Figures 5-g, 5-h and 5-i), is due to the proximity of the lumen to the catheter guide. Such area was manually removed from the images used in the training set, and it has not been learned by the classifier.

3.2 Quantitative results

A total of 4.10^6 pixels composing the test-set (TeS) were classified in 17 minutes (15 seconds per image) approximately. The performance of CNN classification is shown in Table 1.

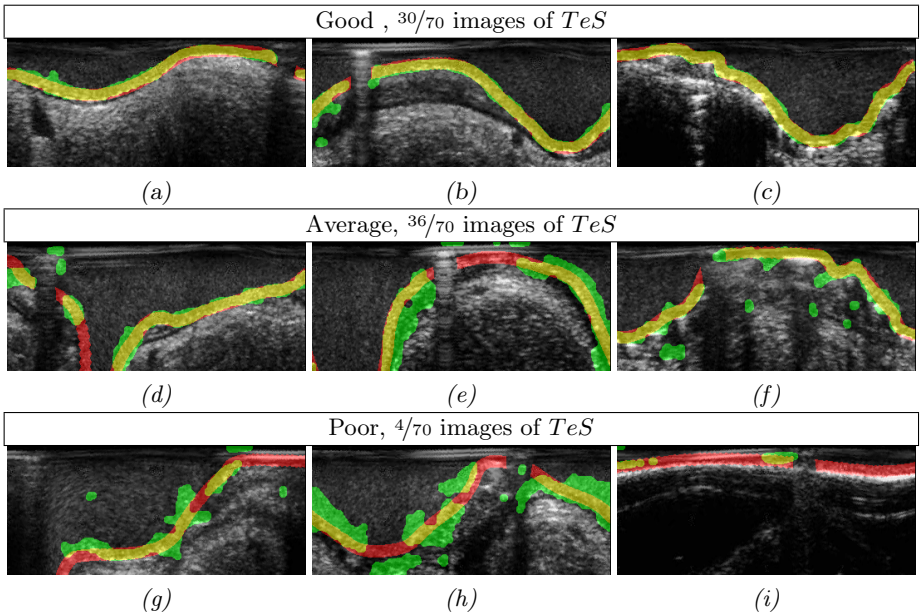


Fig. 5. Samples of qualitative results on TeS. In each image three curves can be seen, the red is the ground-truth traced by the expert, the green is the classification result achieved by the CNN, the overlapping area is represented in yellow.

	Accuracy	Sensitivity	Specificity	Precision	F-score
All the test-set	0.93	0.80	0.95	0.64	0.71
Normal (16 frames)	0.93	0.72	0.95	0.61	0.66
Presence of Stent (30 frames)	0.93	0.83	0.95	0.64	0.73
Presence of Calcium (26 frames)	0.93	0.80	0.95	0.63	0.70

Table 1. Quantitative results on TeS

It is interesting to note that when the image contains stents and calcified plaques, the classification indicators are higher compared to normal images. Although this seems counterintuitive, this can be explained by the fact that CNN classifies pixels using contextual information, by learning the presence of defined patterns such as the presence of plaque or stent below the membrane.

4 Conclusions

A new approach aimed at classifying the pixel containing luminal border for successive lumen segmentation has been presented. The main strength of our approach consists in using a convolutional neural network as automatic pixel classifier on IVUS images.

In contrast with previous approaches, in which the performances of the algorithm decreased when challenging frames were analyzed, the results of the

classifier were acceptable (*F-score* of 71%) on all the analyzed frames categories (images containing stent, calcium plaques and guide shadows). Future work will be addressed towards identifying a lumen border contour, allowing the comparison against the state of the art segmentation techniques [1].

References

1. Balocco, S., et al.: Standardized evaluation methodology and reference database for evaluating ivus image segmentation. Comput Med Imaging Graph 38(2), 70–90 (Mar 2014), http://dx.doi.org/10.1016/j.compmedimag.2013.07.001
2. de Brebisson, A., Montana, G.: Deep neural networks for anatomical brain segmentation. arXiv preprint arXiv:1502.02445 (2015)
3. Cernazanu-Glavan, C., Holban, S.: Segmentation of bone structure in x-ray images using convolutional neural network. Adv. Electr. Comput. Eng 13(1), 87–94 (2013)
4. Fukushima, K.: Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics 36(4), 193–202 (1980), http://dx.doi.org/10.1007/BF00344251
5. Greenspan, H., van Ginneken, B., Summers, R.M.: Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. IEEE Transactions on Medical Imaging 35(5), 1153–1159 (May 2016)
6. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A.C., Bengio, Y., Pal, C., Jodoin, P., Larochelle, H.: Brain tumor segmentation with deep neural networks. CoRR abs/1505.03540 (2015), http://arxiv.org/abs/1505.03540
7. Hinton, G., Osindero, S., Teh, Y.W.: A fast learning algorithm for deep belief nets. Neural computation 18(7), 1527–1554 (2006)
8. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. arXiv preprint arXiv:1408.5093 (2014)
9. Katouzian, A., Angelini, E.D., Carlier, S.G., Suri, J.S., Navab, N., Laine, A.F.: A state-of-the-art review on segmentation algorithms in intravascular ultrasound (ivus) images. IEEE Trans Inf Technol Biomed 16(5), 823–34– (2012)
10. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097–1105 (2012)
11. LeCun, Y., Bengio, Y.: The handbook of brain theory and neural networks. chap. Convolutional Networks for Images, Speech, and Time Series, pp. 255–258. MIT Press, Cambridge, MA, USA (1998), http://dl.acm.org/citation.cfm?id=303568.303704
12. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE 86(11), 2278–2324 (1998)
13. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3431–3440 (2015)