DELINEATION OF CORONARY STENTS IN INTRAVASCULAR ULTRASOUND PULLBACKS

Tobias Wissel[a+], Katharina A. Riedl[c+], Klaus Schaefers[b], Hannes Nickisch[a], Fabian J. Brunner[c], Nikolas Schnellbaecher[a], Stefan Blankenberg[c,d], Moritz Seiffert[c,d], Michael Grass[a]

[a]Philips Research – Hamburg, Germany; [b]Philips Research – Eindhoven, The Netherlands; [c]Department of Cardiology, University Heart & Vascular Center Hamburg, Hamburg, Germany; [d]German Center for Cardiovascular Research (DZHK), Partner Site Hamburg/Lübeck/Kiel, Germany; [+]Both authors contributed equally to this work

ABSTRACT

Ischemic heart disease remains one of the leading causes of death worldwide. Percutaneous coronary interventions (PCIs) for implanting coronary stents are preferred for patients with acute myocardial infarction but may also be performed in patients with chronic coronary syndromes to improve symptoms and outcome. During the PCI, the assessment of stent apposition, evaluation of in-stent restenosis or guidance for complex stenting of bifurcation lesions may be improved by intravascular imaging such as intravascular ultrasound (IVUS). However, advanced interpretation of the image often requires expertise and training.

To approach this issue, we introduce an automatic delineation of stent struts within the IVUS pullback. We propose a cascaded segmentation based on data-driven learning with a neural encoder-decoder architecture. The learning process uses 80 IVUS sequences from 28 patients which were acquired and partially annotated by the Department of Cardiology, University Heart & Vascular Center Hamburg, Germany. The annotations include 1108, 555 and 355 frames with delineated lumen, stent and calcium as well as 13696 and 10689 frame-wise stent and no-stent indications.

The network was pre-trained on lumen segmentation and refined to first identify stent frames using an encoder network and subsequently segment the struts with a decoder. Quantitative evaluation using 3-fold cross-validation revealed 88.3% precision, 92.4% recall and 0.824 Dice for the encoder and 67.0%, 60.3% and 0.611 for the decoder.

We conclude that the encoder successfully leverages the larger number of high-level annotations to reject non-stent frames avoiding unnecessary false positives for the decoder trained on much less, but fine-granular annotations.

Keywords: Intravascular ultrasound, stent, convolutional neural networks, segmentation, detection, encoder-decoder

1. INTRODUCTION

Ischemic heart disease remains one of the leading causes of death worldwide [1]. Percutaneous coronary interventions (PCIs) with balloon dilatation and implantation of coronary stents constitute the preferred strategy in the majority of patients with acute myocardial infarction but may also be performed in patients suffering from chronic coronary syndrome to improve symptoms and outcome. During the PCI, assessment of the stent apposition, evaluation of in-stent restenosis or guidance for complex stenting of bifurcation lesions may be improved by intravascular imaging such as intravascular ultrasound (IVUS). However, advanced interpretation of the image often requires expertise and training due to ambiguous appearances of relevant structures in an IVUS image as well as typical image properties such as speckle noise and acoustic shadowing [2].

To address these challenges, we introduce an automatic delineation of stent struts within the IVUS pullback. The delineation aiming for a segmentation of visible stent struts on pixel level and use single frame inputs only.
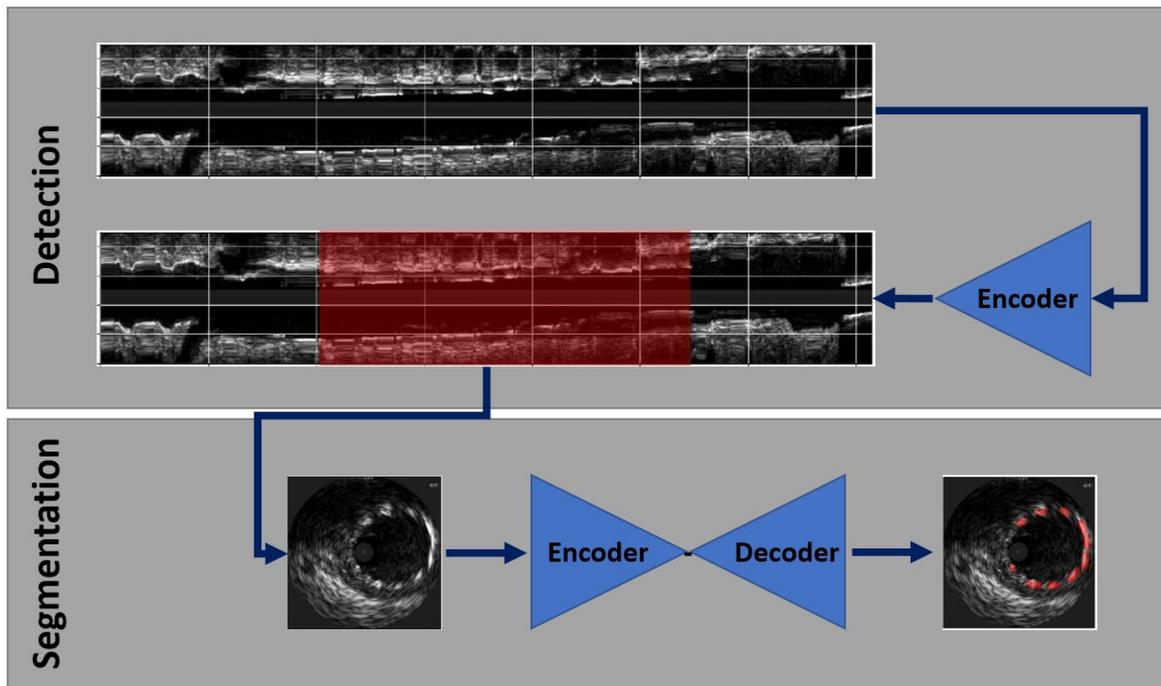
Figure 1: Illustration of the cascaded detector - segmenter approach. Given a manual IVUS pullback, an encoder network first acts as a per-frame detector for images which contain stent (easier task trained on many frames). The intervals identified as relevant are then passed to a segmenter network (encoder-decoder) to locate stent struts within the image. This approach aims at avoiding high false positive rates on frames which do not contain a stent in first place.

The latter is particularly desirable as manual pullbacks dominate clinical practice and an equidistant frame sampling in pullback direction is hence not guaranteed. Together with existing approaches for segmenting the lumen area as well as the media/adventitia outline [3], coronary interventions can be supported by valuable information such as semantic highlighting or derived quantitative measures such as extent of stent apposition in an automatic manner.

The next section first describes the IVUS dataset used for this work and then provides details on the learning approach utilized for stent detection and segmentation. Finally, detection as well as segmentation results are presented and discussed in a qualitative and quantitative manner.

2. MATERIAL AND METHOS

2.1 Dataset

The learning process uses 80 IVUS sequences of varying length from 28 patients. The sequences were acquired with a 20 MHz Eagle Eye phased array transducer (Philips Healthcare, San Diego, USA) and were not ECG-gated. Each frame has a size of 500 x 500 pixels with a resolution of 0.02 mm per pixel. All frames were used in cartesian coordinates.

The sequences were acquired at the Department of Cardiology, University Heart & Vascular Center Hamburg, Hamburg, Germany from patients suffering from coronary artery disease. They contain a variety of different morphological appearances. These include calcified and non-calcified plaques, bifurcations, neighboring vessels as well as devices such as stents, guidewires or the catheter.

The data was annotated by an experience clinical expert in two different ways. First, intervals on a per-frame level were marked on the pullbacks. Each interval was defined by a start frame and an end frame index and could be of type "stent" or "no-stent". In total 13696 stent and 10689 no-stent frames fell into the annotated intervals.

In addition, pixel-level annotations were carried out on selected frames. Frames were selected to cover as much variability in the frame appearance as possible. For labels "stent" and "calcium" 397 and 236 label bitmasks were annotated, respectively. In each of these frames the annotation was exhaustive, i.e. if one label was annotated it was made sure that the others were also covered in case, they were also present. A few weeks after the first annotation, the expert was asked to repeat the pixel-level annotation the target label "stent" on 28 randomly chosen frames from the above set. A comparison of both annotations revealed an intra-observer repeatability as quantified with a Dice coefficient of 75.47.

Finally, the blood-filled area within the vessel wall, the lumen, was delineated in 1108 frames.

2.2 Learning Approach

We propose a cascaded segmentation approach based on data-driven learning with a neural encoder-decoder architecture based on VGG-like layering similar to Noh et al. [4]:

First, an encoder model is trained to classify frames into the categories {stent, no-stent}. Based on this identification of stent frames a separate encoder-decoder model then segments the struts within each frame (see Fig. 1). The latter model is first pre-trained on lumen segmentation – a task with pixel-wise ground truth and a higher number of annotated frames – to initialize the model with weights tailored to the IVUS imaging domain. The model was subsequently refined by re-training it to segment stent struts and calcium deposits. The latter introduces additional label information into the training to mitigate ambiguities between stent and calcium appearances.

The encoder model for stent detection as well as the encoder-decoder model for segmentation were trained using the Adam optimizer [5] with cross-entropy loss for the former and generalized dice loss [6] for the latter. The encoder part of the encoder-decoder model had an identical architecture compared to the encoder model for detection. Weights of the encoder parts for each of the two tasks were however trained from scratch (or from a pre-trained lumen segmentation model) as a simply freezing the trained detector weights for the segmentation task as well as simultaneously training the segmentation network on interval and bitmask labels was both found to be of inferior performance.

Both encoders consisted of five sub-blocks, each containing a convolutional layer (twice 5 x 5, twice 4 x 4 & once 3 x 3 kernels, both stride 1), batch normalization, leaky ReLU activation and an average pooling (2 x 2 kernel, stride 2) layer. The data was down-sampled to size 224 x 224 with a single channel and passed to the first block, which increased the number of channels to 16. Every other of the following blocks increased the number of channels by factor 2 yielding 64 at the bottleneck. The encoder output was then modeled with a three-unit fully connected layer plus softmax (for classes stent, no-stent and none, where the latter was never used during training and only a placeholder for the deployment on full pullbacks, where many frames had not been annotated (see Fig. 3), i.e. there is no ground truth for their content).

The decoder was designed equivalent to the encoder replacing the pooling with un-pooling layers and decreasing the channel numbers by factor two instead of increasing them every other block. At the decoder output the number of channels was reduced to three (for classes stent, calcium and none) before entering the softmax. Shortcut layers connected the out- and inputs of the corresponding encoder and decoder blocks, where the data passing through the shortcuts was concatenated to the inputs of the decoder blocks.

3. RESULTS AND DISCUSSION

Quantitative evaluation was done using 3-fold cross-validation with splits on patient level to avoid frames from the same patient in both training and test set. Pullback frames without any annotation were discarded from all trainings (label "none"), since no information about their content was given. They were however used during deployment even though they did not contribute to the scores.

The encoder training revealed 88.3% precision, 92.4% recall and 0.824 Dice for the identification of stent frames (more detailed scores are shown in Fig. 2). These promising scores suggest an ideal gating unit for the application of the decoder. Fig. 3 provides an example pullback aligned with the encoder decisions.
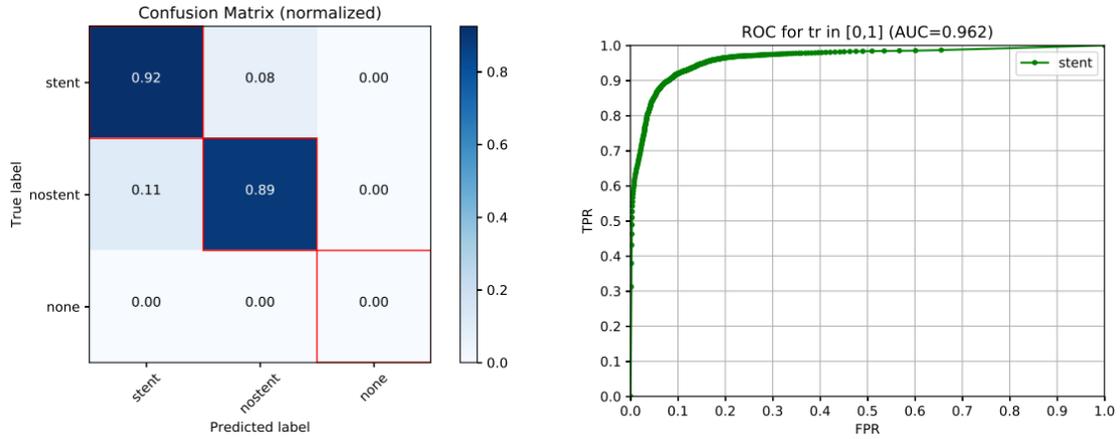
Figure 2: Encoder Evaluation. Left: Confusion matrix based on the pooled results of all test folds. Rows are normalized to one and indicate how many frames of each ground truth label were classified into which target class by the model. Right: Receiver-operator curve for different thresholdings on the probability outputs for the stent class.
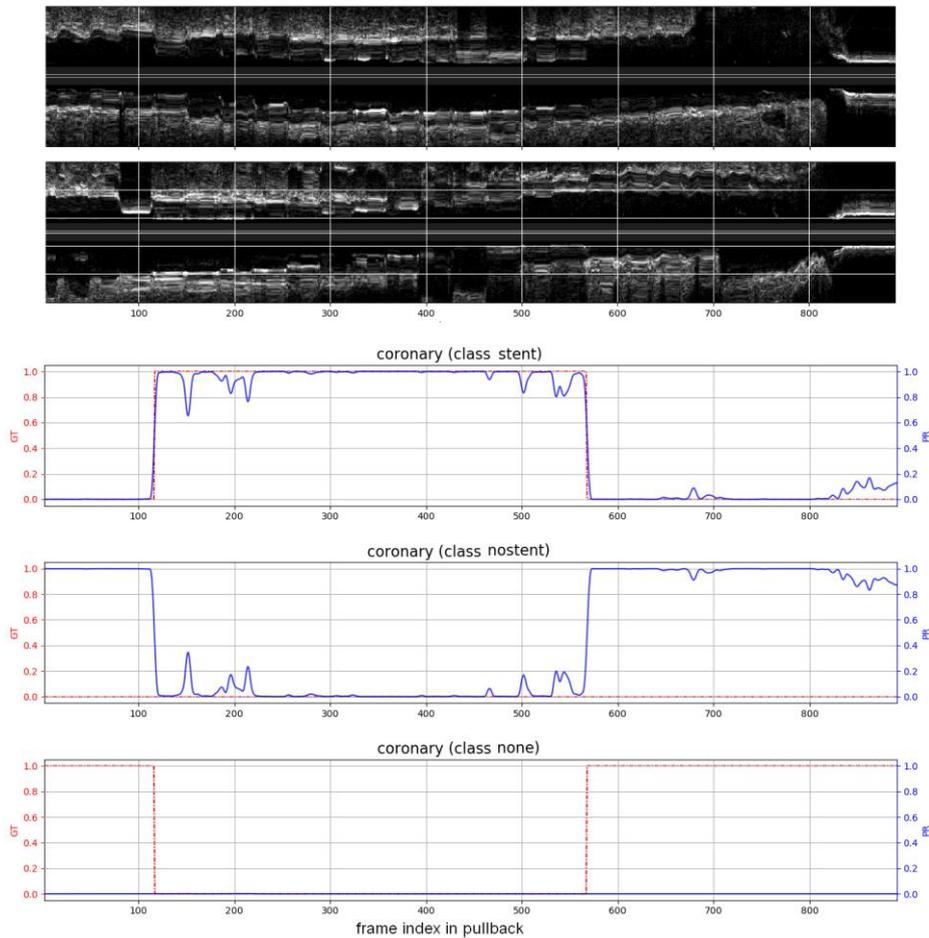
Figure 3: Encoder predictions for an example pullback. Longitudinal views of the pullback are shown on top and the prediction (blue) and ground truth (red) for each of the three classes is shown below (values as probabilities in [0, 1]). The last row called labeled as class none indicates intervals (in red) which were not annotated. As this class did never appear during training, the model predicts it during deployment
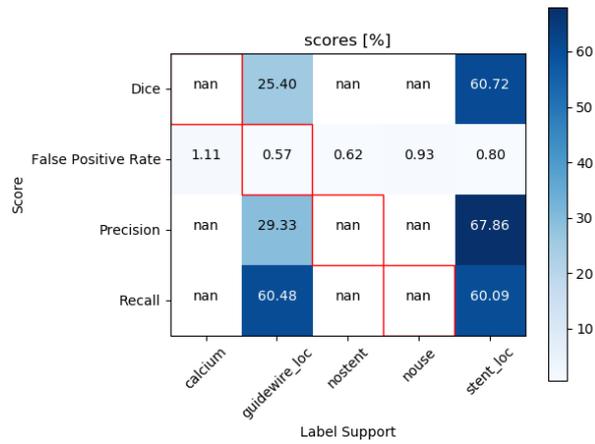
Figure 4: Decoder scores for all test sets pooled. Scores are displayed separately based on the frame content. Not-a-number (nan) notes indicate that the score could not be computed due to a missing overlap with stent strut labels. False positive rates must be interpreted by taking into account the high imbalance between stent and background pixels in a typical IVUS frame.
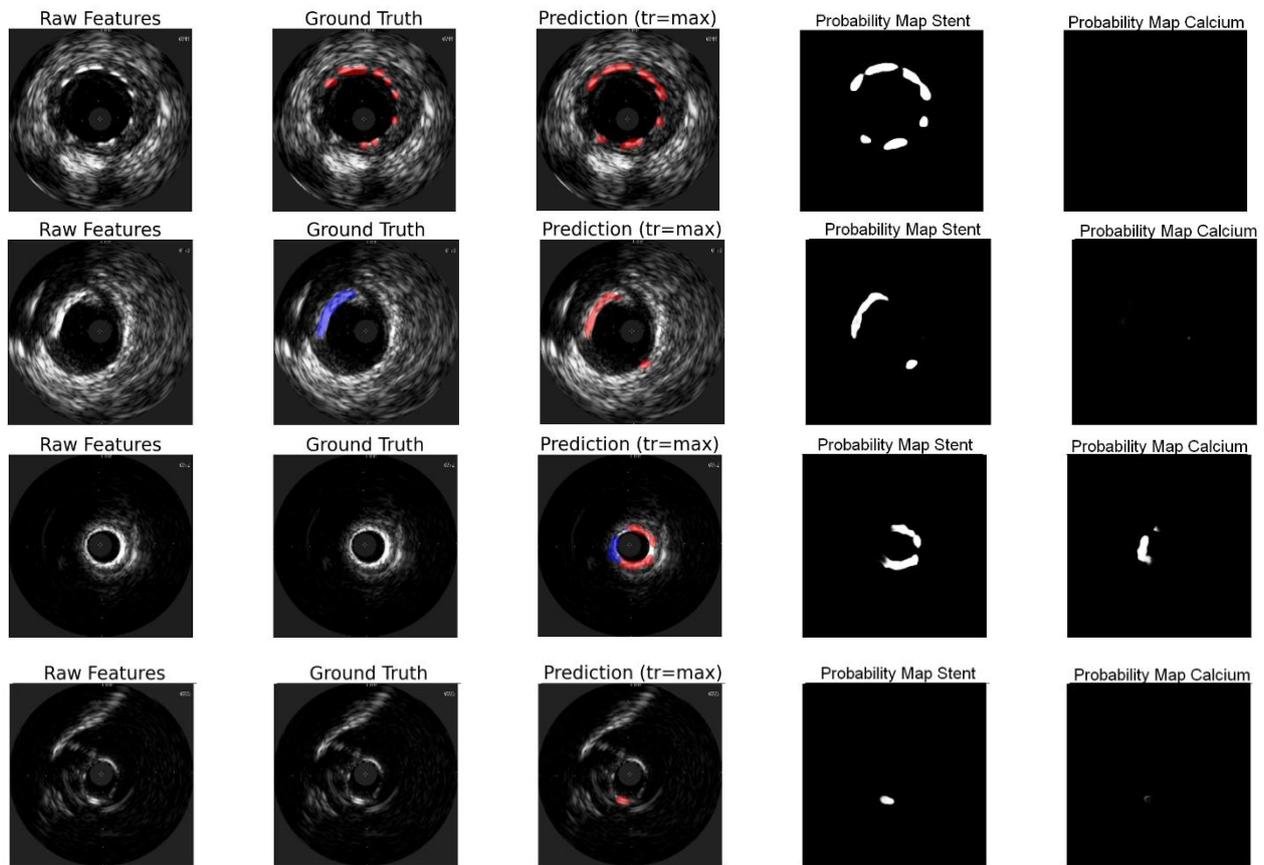
Figure 5: Example frames with overlays of the segmentation results. Top: Typical frame containing stent struts along the lumen border. Depending on how the stent is cut, struts show in irregular spacings along the lumen border. Second row: False positive segmentation of a calcium arc. Third row: False positives in a context where the transducer was covered by the catheter. Bottom: False positives while imaging the ostial part of the coronary artery. The latter three cases are ruled out in the cascaded approach where the encoder acts as a gate for the decoder predictions.

All intervals with a ground truth label called "none" (last row) contain frames which were not annotated and where the content of the frames is unknown in terms of available ground truth.

The decoder training resulted in 68.2% precision, 60.4% recall and 0.611 Dice for the decoder. Detailed scores separated by frame category are illustrated in Fig. 4. The matrix shows that due to ambiguities, false positive predictions are still highest for frames including calcium or artifacts (frames of no clinical use labeled "nouse", since the transducer was close to the ostium, within the catheter sheath or exhibits severe motion artifacts) – while false positives are lowest in the no-stent regime showing non-stented, healthy or stenotic vessel parts.

All segmentation scores must be seen in the context of intra-observer variability. Tests for stent strut annotation reproducibility yielded and overlap score of 0.754 Dice for the expert. This limit is typical for smaller structures in a noisy environment such as it is the case for IVUS imaging.

Finally, Fig. 5 illustrates qualitative examples of typical frame categories. While stent struts are labelled quite reliably by the decoder, the risk for false positives in frames with calcium or artifacts is still there (but not always the case). Joining the encoder decision in a cascaded approach with the decoder removes most of these false positives, because none of the frames would be labelled to contain a stent.

4. CONCLUSIONS

We conclude that the encoder successfully leverages the larger number of high-level annotations to reject non-stent frames avoiding unnecessary false positives for the decoder trained on much less, but fine-granular annotations. The decoder performance on frames only containing a stent is promising to approach the aforementioned clinical challenges.

ACKNOWLEDGEMENTS

REFERENCES

[1] Nowbar, A. N., Gitto, G., Howard, J. P., Francis, D. P., Al-Lamee, R., "Mortality From Ischemic Heart Disease, Analysis of Data From the World Health Organization and Coronary Artery Disease Risk Factors From NCD Risk Factor Collaboration," Circulation: Cardiovascular Quality and Outcomes, 12:e005375 (2019).

[2] Neumann, F.-J., Sousa-Uva, M., Ahlsson, A., Alfonso, F., Banning, A. P., Benedetto, U., Byrne, R. A., Collet, J. P., Falk, V., Head, S. J., Jüni, P., Kastrati, A., Koller, A., Kristensen, S. D., Niebauer, J., Richter, D. J., Seferovic, P. M., Sibbing, D., Stefanini, G. G., Windecker, S., Yadav, R., Zembala, M. O., "2018 ESC/EACTS Guidelines on myocardial revascularization," European heart journal, 40(2), 87-165 (2019)

[3] Balocco, S., Gatta, C., Ciompi, F., Wahle, A., Radeva, P., Carlier, S., Unal, G., Sanidas, E., Mauri, J., Carillo, X., Kovarnik, T., Wang, C. W., Chen, H. C., Exarchos, T. P., Fotiadis, D. I., Destrempes, F., Cloutier, G., Pujol, O., Alberti, M., Mendizabal-Ruiz, E. G., Rivera, M., Aksoy, T., Downe, R. W., Kakadiaris, I. A., "Standardized evaluation methodology and reference database for evaluating IVUS image segmentation," Comput Med Imaging Graph, 38(2), 70-90 (2014).

[4] Noh, H., Hong, S., Han, B, "Learning Deconvolution Network for Semantic Segmentation," Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), IEEE Computer Society, 1520-1528, (2015).

[5] Kingma, D. P. and Ba, J., "Adam: A method for stochastic optimization," in [3rd International Conference on Learning Representations, ICLR 2015, Conference Track Proceedings], (2015).

[6] Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., and Jorge Cardoso, M., "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in [Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support], 240–248 (2017).