Cascaded learning in intravascular ultrasound: coronary stent delineation in manual pullbacks

Tobias Wissel[1*], Katharina A. Riedl[3], Klaus Schaefers[2], Hannes Nickisch[1], Fabian J. Brunner[3], Nikolas D. Schnellbaecher[1], Stefan Blankenberg[3,4], Moritz Seiffert[3,4], Michael Grass[1]

[1]Philips Research – Hamburg, Germany; [2]Philips Research – Eindhoven, The Netherlands
[3]Department of Cardiology, University Heart & Vascular Center Hamburg, Hamburg, Germany
[4]German Center for Cardiovascular Research (DZHK), Partner Site Hamburg/Lübeck/Kiel, Germany

Abstract

Purpose: Implanting stents to re-open stenotic lesions during percutaneous coronary interventions is considered a standard treatment for the acute or chronic coronary syndrome. Intravascular ultrasound (IVUS) can be used to guide and assess the technical success of these interventions. Automatically segmenting stent struts in IVUS sequences improves workflow efficiency but is non-trivial due to a challenging image appearance entailing manifold ambiguities with other structures. Manual, ungated IVUS pullbacks constitute a challenge in this context. We propose a fully data-driven strategy to first longitudinally detect and subsequently segment stent struts in IVUS frames.

Approach: A cascaded deep learning approach is presented. It first trains an encoder model to classify frames as "stent", "no stent", or "no use". A segmentation model then delineates stent struts on a pixel level only in frames with stent label. The first stage of the cascade acts as a gateway to reduce the risk for false positives in the second stage, the segmentation, trained on a smaller and difficult-to-annotate dataset. Training of the classification and segmentation model was based on 49,888 and 1,826 frames of 74 sequences from 35 patients, respectively.

Results: The longitudinal classification yielded Dice scores of 92.96%, 82,35%, and 94,03% for the classes "stent", "no stent" and "no use". The segmentation achieved a Dice score of 65.1% on the stent ground truth (intra-observer performance: 75,5%) and 43.5% on all frames (including frames without stent, with guidewires, calcium or without clinical use). The latter improved to 49.5% when gating the frames by the classification decision and further increased to 57.4% with a heuristic on the plausible stent strut area.

Conclusions: A data-driven strategy for segmenting stents in ungated, manual pullbacks was presented – the most common and practical scenario in the time-critical clinical workflow. We demonstrated a mitigated risk for ambiguities and false positive predictions.

Keywords: intravascular ultrasound (IVUS), coronary, stent, segmentation, detection.

***First Author, E-mail: tobias.wissel@philips.com**

1 Introduction

Coronary artery disease remains one of the leading causes of death worldwide, accounting for more than 9 million deaths alone in 2016 according to the World Health Organization (WHO) [1]. The disease is caused by atherosclerosis – an accumulation of plaques in the intima of the arterial wall which decrease the effective vessel diameter and thus form a stenosis. Stenotic vessels impede the flow of oxygenated blood into the cardiac muscle causing chest pain (angina pectoris) and ultimately provoking myocardial infarction.

Percutaneous coronary interventions (PCIs) with balloon dilatation and implantation of coronary stents constitute the preferred strategy in most patients with acute myocardial infarction but may also be performed in patients suffering from chronic coronary syndromes to improve symptoms and outcome. Here, intravascular imaging can support several parts of the clinical workflow: Recent studies suggest that treatment planning for complex lesions can significantly benefit from stenting criteria based on intravascular ultrasound [2] and that IVUS-guided procedures improve long-term clinical outcome in patients with acute myocardial infarction [3]. During the PCI, the assessment of stent malapposition, stent underexpansion, stent strut fractures, post-dilation decisions, its placement with respect to other anatomical structures such as bifurcations or the guidance for complex stenting procedures at bifurcation lesions may be improved by intravascular imaging [4]. In particular, the former two are seen as risk factors for in-stent restenosis or thrombosis [5, 6]. In addition, complex PCIs treating bifurcation lesions require elaborated procedures such as the culotte or other techniques. Here, typically two guidewires are inserted into both branches, one is jailed during stent deployment in the first branch and later used as a guide for re-wiring the second branch through the wire cells of the first stent (e.g., by proximal or distal cell technique) [7]. Only after successful rewiring, the second branch can also be stented, which is greatly facilitated by intra-vascular imaging and knowing the location of the first stent and its cells.

ACC/AHA and ESC/EACTS guidelines mention intravascular ultrasound (IVUS) as a complement to intravascular optical coherence tomography (IVOCT). Both imaging modalities can potentially mitigate the limitations of X-ray angiography in interventional guidance, but also have different strengths and shortcomings. While IVOCT exhibits a higher cross-sectional resolution and typically better image contrast, it also has a limited radial field-of-view (FoV) which – even though it captures the vascular lumen border – often reaches not far beyond, in particular for larger vessels. Adequate IVOCT imaging also requires saline flushing of the lumen, which is not the case for IVUS. Depending on the transducer frequency, the latter can achieve several different, but compared to IVOCT inferior, cross-sectional resolutions. Despite the characteristic speckle noise patterns, IVUS is well suited to also evaluate plaques and vessel wall compositions as it acquires larger FoVs.

Advanced image interpretation often requires expertise and training [8, 9]. This establishes the clinical need for making IVUS imaging more accessible and easier to use while minimizing, for instance, procedure prolongation [4]. Both imaging techniques provide sequences of 2D tomographic cross-sectional images when manually or automatically pulled along the artery. The tube-like metal mesh of a stent therefore appears as a collection of spots typically located at the inner circular border or sometimes within the intima. The number and shape of the visible spots depends, for instance, on the inclination of the imaging plane. In IVUS, stent struts often even appear merged into bright circular ring segments due to dense fibrous or micro-calcified tissue close to the stent [26].

Motivated by the clinical needs above, research groups successfully detected stent struts in IVOCT images [20, 21, 22]. Here algorithms such as the Tsantis et al. [22] can usually rely on the excellent image quality, the higher image resolution as well as contrast inside the vascular lumen up to the lumen boundary in IVOCT images. In their work they make use of characteristic intensity distributions and wavelet-based matching, which would be rather challenging on IVUS images with an inhomogeneous texture, speckle noise and many spot-like ambiguities.

Less groups targeted the more challenging task of extracting stents from IVUS images. Contour-based predictions as proposed by Dijkstra [23] or Kitahara [24] and colleagues worked well, but either contain semi-automatic steps, require well visible stent struts or neglect challenging frames with acoustic shadowing, bifurcations or generally more than 90° signal attenuation in order to work robustly. In clinical practice, frames of an IVUS pullback often contain artifacts caused, for instance, by heart beat related pulsation of the surrounding anatomy, longitudinal swinging of the transducer or simply rapid pulls, transducer-wall or transducer-device interactions during manual pullbacks. Apart from ambiguities due to similar appearance, these artifacts and the low resolution paired with speckle noise entail a high risk of false positives for fully automatic algorithms. This has also been recognized by the research group around Balocco and Ciompi et al. who provided a comprehensive approach to the problem of stent detection in automatic IVUS pullback data sets [26].

They present a pipeline of three steps validated on automatic pullbacks which first performs an image-based gating on the sequence [27], then detects stent struts and stent shape in cross-sectional

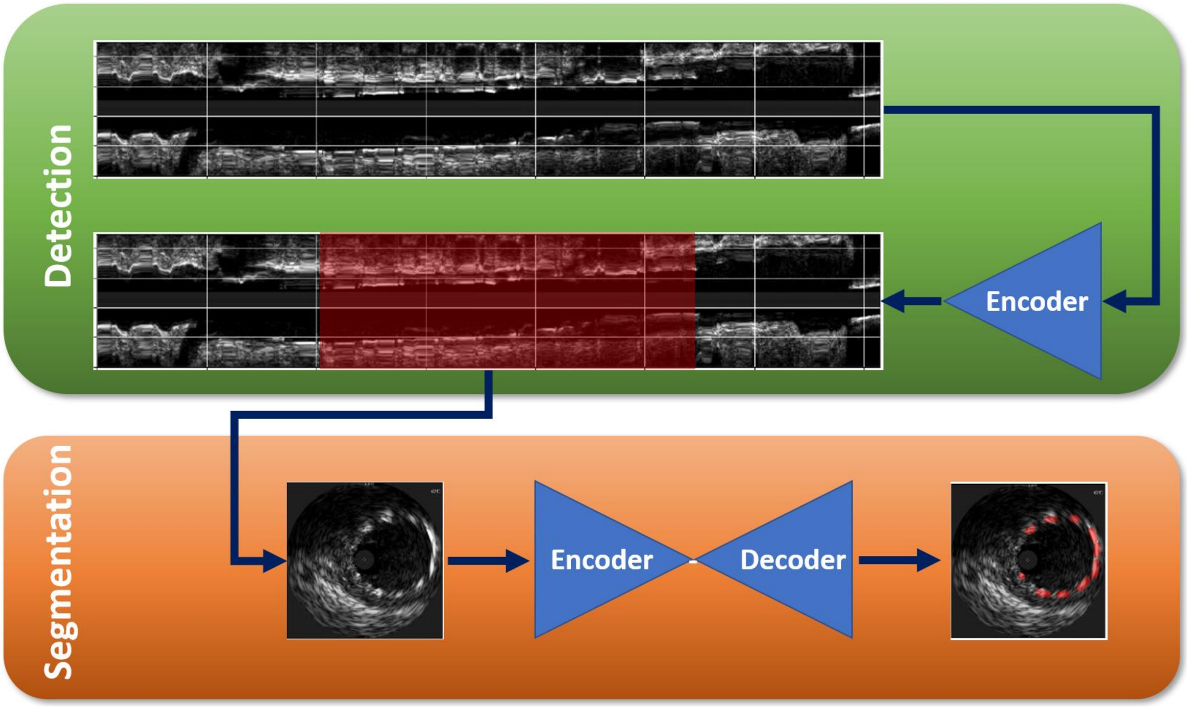

Figure 1: Illustration of the cascaded concept. Frames of the manual pullback are first analyzed by an encoder network, which decides for one of three classes per frame: stent, no stent or no use. Only stent frames are then passed on to the encoder-decoder to segment the stent struts. Apart from the favorable training setup, this is also targeting a reduction of false positive predictions on frames that do not show a stent anyway.

images and finally longitudinally localizes the positioned stent in pullback direction. They note that the stent strut detection as proposed by Ciompi et al. [28, 29] is an essential part of the pipeline. First, a 2-stage multi-class AdaBoost classifier generates pixel-wise label maps from handcrafted appearance features. Based on heuristics including knowledge about the luminal area, the most plausible elliptical stent shape is fitted to the strut class before a stent prototype filter confirms an output set of likely struts locations.

In their most recent work, Balocco et al. define a likelihood function on the strut label mask, which they successfully convert into a stent indicator variable along the vessel using the SAX algorithm for solving the more challenging problem of stent localization [30]. As part of their conclusion, they acknowledge that the approach could benefit from fully data-driven deep learning techniques.

They note that a thorough learning strategy is needed to deal with the otherwise substantial requirement of carefully annotated data to solve this complex task with its high risk for false positive predictions.

In this work, we present such a strategy. We propose a cascaded deep learning approach as illustrated in Figure 1. The cascade reverses the pipeline from Balocco et al. [26] in that it first solves the task of longitudinal detection with an encoder network, before an encoder-decoder network segments the stent struts only in the detected stent ranges. This way, the encoder can be trained on a huge amount of efficiently, frame-wise annotated data to solve a simpler task, while the segmenter can be trained on a smaller amount of pixel-level annotated data to solve a more challenging task. The latter type of data is often tedious to annotate, prone to label noise and hard to get in larger quantities.

In contrast to previous work, the proposed concept has the ambition to work on single frames from manual, ungated pullbacks as this is closest to time-critical clinical practice. We further recognize that these practical scenarios impose a problem where – in contrast to other typical segmentation tasks on clinical data – many images do not contain the actual target structure or instead contain objects of similar appearance such as guidewires or calcium deposits. Due to the rather frequent and sometimes rapid back and forth transducer motion, manual pullbacks also exhibit a particular risk for artifacts and clinically irrelevant frames. Overall, these conditions require special attention dedicated to false positive predictions during algorithm design. Therefore, the encoder will also be trained against a “no use” class to serve as a better gateway for frames sent to the segmenter and to reduce the risk for false positives.

Finally, our contributions can complement algorithmic features from earlier work and can be used – together with existing solutions for lumen and vessel wall segmentation [31, 32, 33] – to assess stent malapposition [26] or facilitate the rewiring during the treatment of bifurcation lesions. For the two use cases, malapposition and rewiring – two different aspects of stent detection and segmentation – are emphasized and will be addressed by dedicated metrics in this work: While the first use case requires precise knowledge about the radial position of the stent mesh, the second one rather asks for the angular occupation to identify mesh cells enabling guidewire passage.

2 Material and Methods

2.1 Data and Annotations

The IVUS data were acquired at the University Heart and Vascular Center (UHC) Hamburg Eppendorf using the Core or Core Mobile precision guided therapy system (Philips Healthcare, San Diego, USA). The pullbacks were conducted manually within the coronary vessels without ECG gating. The acquisitions from pre- or post-interventional stenting procedures used a 20 MHz phased array Eagle Eye Platinum probe (Philips Healthcare, San Diego, USA). Frames were given in Cartesian coordinates of size 500 × 500 pixels (isotropic resolution of 0.02mm, i.e. a field-of-view (FoV) of 10mm). They were downsampled to 224 × 224 pixels for the predictive task with a preserved FoV. All annotations were obtained in form of a consensus between two clinical experts routinely experienced with IVUS during stenting procedures. The study was approved by the corresponding ethics committee and institutional review board (IRB).

The clinical experts partially annotated 74 sequences from 35 patients using two types of annotations: (1) Intervals indicating the presence of a certain label in frames along the pullback yielding binary, frame-wise annotations per label, and (2) pixel-wise annotations done with a brush tool covering relevant areas in a 2D frame where a label is located.

In the first category, 20,020 frames (83 intervals) of stent, 17,532 frames (76 intervals) without stent and 12,336 frames (91 intervals) of no clinical use were annotated. The latter contained, for instance, frames with artifacts due to rapid pulling, frames where the transducer was covered by the catheter or ostial frames where the intima moved out of the FoV.

In the second category, label masks for 827 frames containing stent struts, for 619 calcium frames and for 390 guidewire frames where annotated. In addition, 120 stent frames from another 19 pullbacks were annotated twice for assessing the intra-observer variability in the ground truth for the main predictive task.

The intra-observer variability for the task of stent segmentation is quantified by a Dice coefficient of 75.47%. This coefficient measures the agreement (in terms of true/false positive/negative pixels) between the first and second annotation of the expert. Only frames with a first annotation (which definitely contain a stent) were presented to the expert in the second round.

The data was split into five cross-validation (CV) folds for later evaluation. The partitioning was carried out on patient level, i.e. all pullbacks and therefore all frames that came from the same patients were always assigned to the same fold, such that a model tested on a particular fold would be unbiased and would not have seen data from the test patients during training.

Furthermore, the data was partitioned into five folds such that the number of frames per fold were roughly the same and labels per class were as balanced as possibly under these restrictions.

2.2 Approach and Implementation

The cascade makes use of two independently trained networks. During deployment, they subsequently act upon the incoming pullback frames where the decision of the first part, the stent detection, is passed forward to support the pixel-wise prediction of the second part, the stent segmentation. This aims at achieving a twofold goal: (1) reducing the false positive rates on pixel

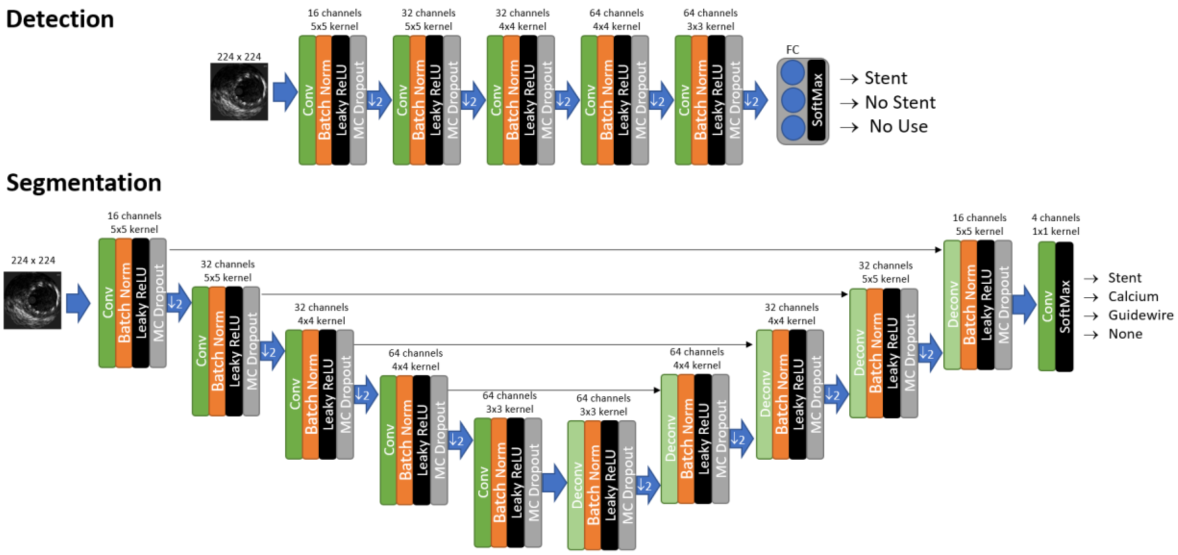

Figure 2: Network architectures for stent detection (top, encoder network) and stent segmentation (bottom, encoder-decoder network with skip connections between both parts).

level on frames that potentially do not contain a stent, and (2) allow for faster processing as the segmentation network would only act on positive stent predictions preselected by the detector. To implement this cascade concept, architectures belonging to the encoder-decoder family [12] as well as the architecturally different and popular DeepLabV3 as an alternative option for the more difficult segmentation task were used. The latter makes use of atrous spatial pyramid pooling (ASPP) with dilated convolutions on top of a ResNet backbone [10].

The encoder-decoder family comprises several possibilities of variations, which differ by distinct architectural features and give rise to well-known variants such as U-Net, DeconvNet or SegNet for a segmentation task [11, 12, 13]. We therefore used the training set of the first cross-validation fold to grid-search the benefit of such typical design features. These included: choice of the activation function, normalization layer, different pooling/un-pooling variants (including max-pooling, average pooling or strided (de)convolutions), usage of skip connections, Monte-Carlo drop-out layers, padded/unpadded convolutions, or residual and squeeze&excite blocks in the convolutional segments. We further optimized typical parameters such as the number of feature maps per segment or the kernel size of the convolutional filters. The search was carried out for the segmentation as well as detection task, whereas the latter only used the encoder.

The DeepLabV3 was used with a ResNet-50 backbone and explored in two versions: (1) pre-trained weights on ImageNet, or (2) random weight initialization.

The subsequent paragraphs will describe our final choices from these two architecture families. Decisions were made for the best performing design choice and in case of equal performance for the simpler variants following the concept of Occam's razor by not multiplying choices beyond necessity.

2.2.1 Detector

The detector as illustrated at the top of Figure 2 is a five-block encoder network, where each block consists of a convolution, batch normalization, leaky ReLU and Monte-Carlo dropout layer. The one-hot network output in terms of four mutually exclusive classes is realized by a fully connected (fc) layer. The detector was trained for 80 epochs with a learning rate of 0.01 and by augmenting the available number of frames with frame-level annotations by the same number of augmented frames. Data augmentation used random rotation, scaling and axis flips as well as deformations based on trigonometric displacement fields. Classes were equally weighted in a cross-entropy loss term.

2.2.2 Segmenter

The segmentation network candidate from the encoder-decoder family consisted of 2×5 blocks for the encoder and decoder respectively, which are connected via skip connections. The blocks are equivalent to the detector network but did not make use of any weight sharing. The network outputs probability maps for four mutually exclusive classes (stent, guidewire, calcium, none).

The encoder-decoder segmenter was trained for 120 epochs with a learning rate of 0.005 and by augmenting the available number of frames with pixel-level annotations by twice the amount of augmented frames. Reaching beyond point estimates for the respective predictions, Monte-Carlo dropout was used (50 samples during deployment) to minimize the effect of individual trainings on the performance and to increase repeatability over trainings [17]. Besides providing a measure of predictive uncertainty, the dropout was also found to improve the segmentation quality when using its mean predictions.

The DeepLabV3 candidate was explored in its original form, pre-trained on ImageNet and refined on the IVUS data for 60 epochs as well as with randomly initialized weights and trained from scratch for 120 epochs. The initial learning rate was taken from the original article.

All networks were trained using Adam [15], used a weight decay regularization ($\lambda = 0.0001$) and batch normalization pooled across all pixels in a frame, and were initialized with the He method if not stated otherwise [16]. The learning rate followed a schedule of hyperbolic decay such that half the learning rate is reached after half the number of specified epochs. Training was done using the generalized Dice loss (GDL) [14] where the weight of the stent class was doubled compared to the other, auxiliary classes. Encoder-decoder implementations were done in CNTK [18] while the DeepLabV3 variant also made use of PyTorch [19].

2.3 Evaluation

Typical evaluation metrics such as receiver-operator characteristics, Dice score, recall, precision, false positive rates (FPR) and area-under-the curve (AUC) were used to compare performances of the detector and segmenter.

As it was not feasible to single out individual stent struts in the majority IVUS frames (see discussion of stent appearances in the introduction), the segmentation ground truth was given in the form of general masks on pixel level. Although this could not give rise to a stent strut detection rate as in [31], we defined two dedicated metrics which are tailored to the two major motivations: (1) malapposition, and (2) rewiring.

2.3.1 Angular Dice Score (DCA)

Successful rewiring requires knowledge of angular occupation to identify cell passages through the stent mesh. To quantify angular overlap, we do not compute the Dice score between the predicted voxels p and the human annotation. Instead, we project both annotation and prediction onto a ring of constant radius around the IVUS catheter tip and compute the Dice score between the projections PR(p) and PR(a). The resulting metric is – like the plain Dice score -- symmetric, between 0 and 1 but focuses on angular overlap while ignoring discrepancies along the radius.

2.3.2 Symmetric Median Skeleton Distance (SMSD)

Confirmation of stent apposition predominantly requires radial accuracy of the stent segmentation to set it into relation to the intima contour of the vessel wall. To achieve this, we construct a metric that is not a relative measure of performance but quantifies accuracy in pixels, i.e. multiples of the image resolution (0.0446mm) and therefore de facto a metric length scale . We skeletonize the prediction and the annotation masks and compute the 50% Hausdorff distance between the two resulting skeletons. Formally, we compute the median Euclidean distance between all prediction skeleton points and their respective closest point on the annotation skeleton. To guarantee symmetry, the final SMSD is then the average between this median and the median from annotation to skeleton

The two proposed metrics DCA and SMSD measure complementary aspects of the overlap between prediction and annotation and hence provide a more detailed picture of the performance of our pipeline than the Dice score alone.

3 Results

3.1 Stent Detection

The detector part of the cascade was trained on frame-level annotations to predict intervals which contain a certain class. These classes were modeled to be mutually exclusive.

The resulting scores pooled from all test sets of the 5-fold cross-validation are shown in Figure 3. All scores were computed based on the mean probabilities of the Monte-Carlo output distributions. The normalized confusion matrix in Figure 3 shows a true positive rate above 90% for all classes on the diagonal (probability threshold 0.5). Highest accuracies are achieved on the “no use” task, i.e., when detecting frames with artifacts or frames of no clinical use. Lowest accuracy was obtained for the detection of areas without stent. The same effect can also be observed from the receiver-operator-characteristic (ROC) curves in Figure 4 (left) as well as from the Dice curves in Figure 4 (right). Both curves were generated by computing the scores based on a sweep across

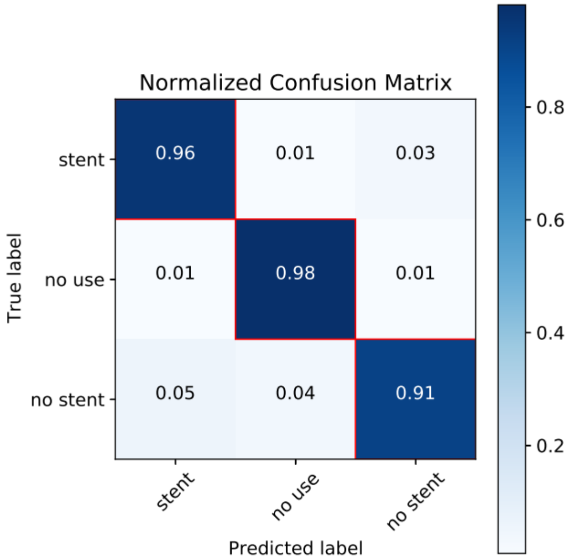

Figure 3: Confusion matrix for the stent detection network after pooling the results on all five test folds.

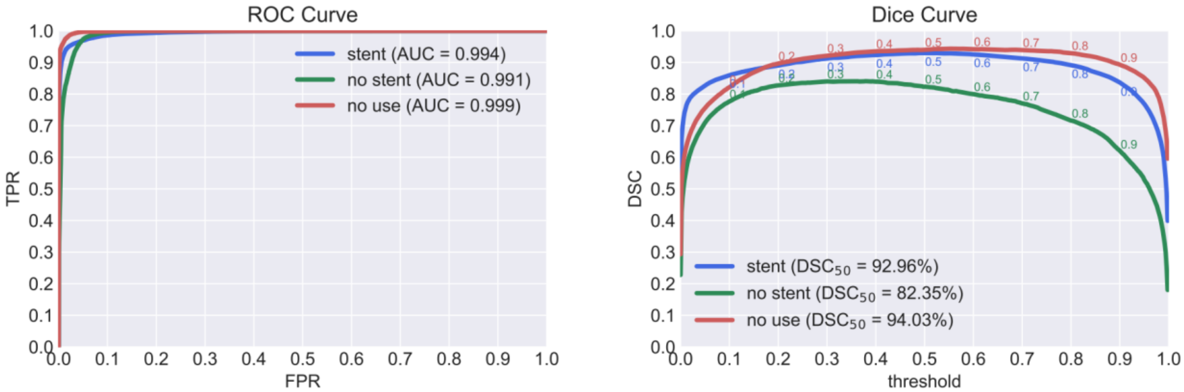

Figure 4: Left: ROC curve for varying thresholds on the output probability maps. ROC-AUC values are listed for a threshold of t = 0.5. Right: Dice curves showing the dependence of the Dice score on the chosen threshold. A good compromise is achieved when thresholding all classes at 0.5.

different output thresholds for the probability maps (t ∈ [0, 1.0]). The ROC curves yield high areas-under-the-curve (AUCs).

Figure 5 shows the mean and standard deviation predictions along an example pullback sequence for each of the first three classes longitudinally plotted.

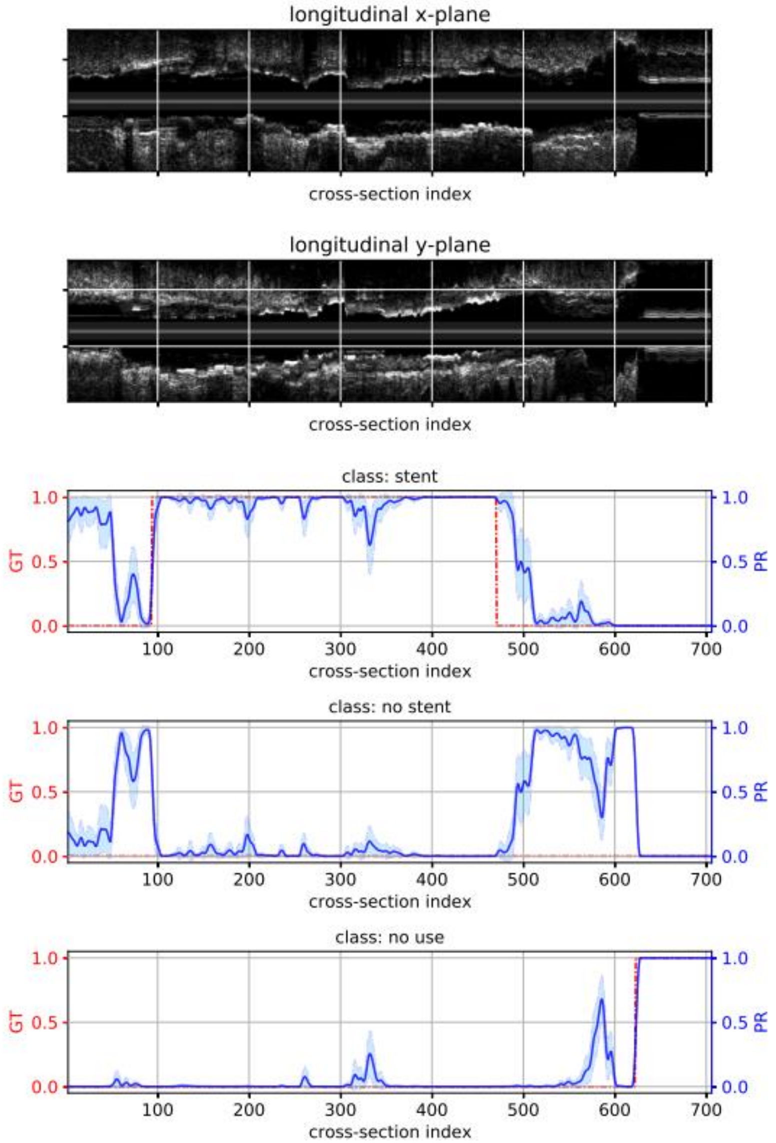

Figure 5: Example pullback along with encoder results. First two rows show orthogonal cuts through the pullback in longitudinal direction and the lower three graphs the ground truth (dash-dotted red) and predictions (solid blue line Monte-Carlo mean and blue shading Monte-Carlo estimate of the standard estimation) Regions where none of the three classes has a red ground truth value at 1 have not been annotated by the expert. Here, predictions cannot be compared with a target label. During the last frames the transducer was covered by the catheter, which is correctly recognized as "no use".

From a cross-sectional perspective, Figure 6 presents example frames along with the corresponding detector decision. The number of true positives, false positives and false negatives displayed does not reflect their corresponding proportions in the overall predictions.

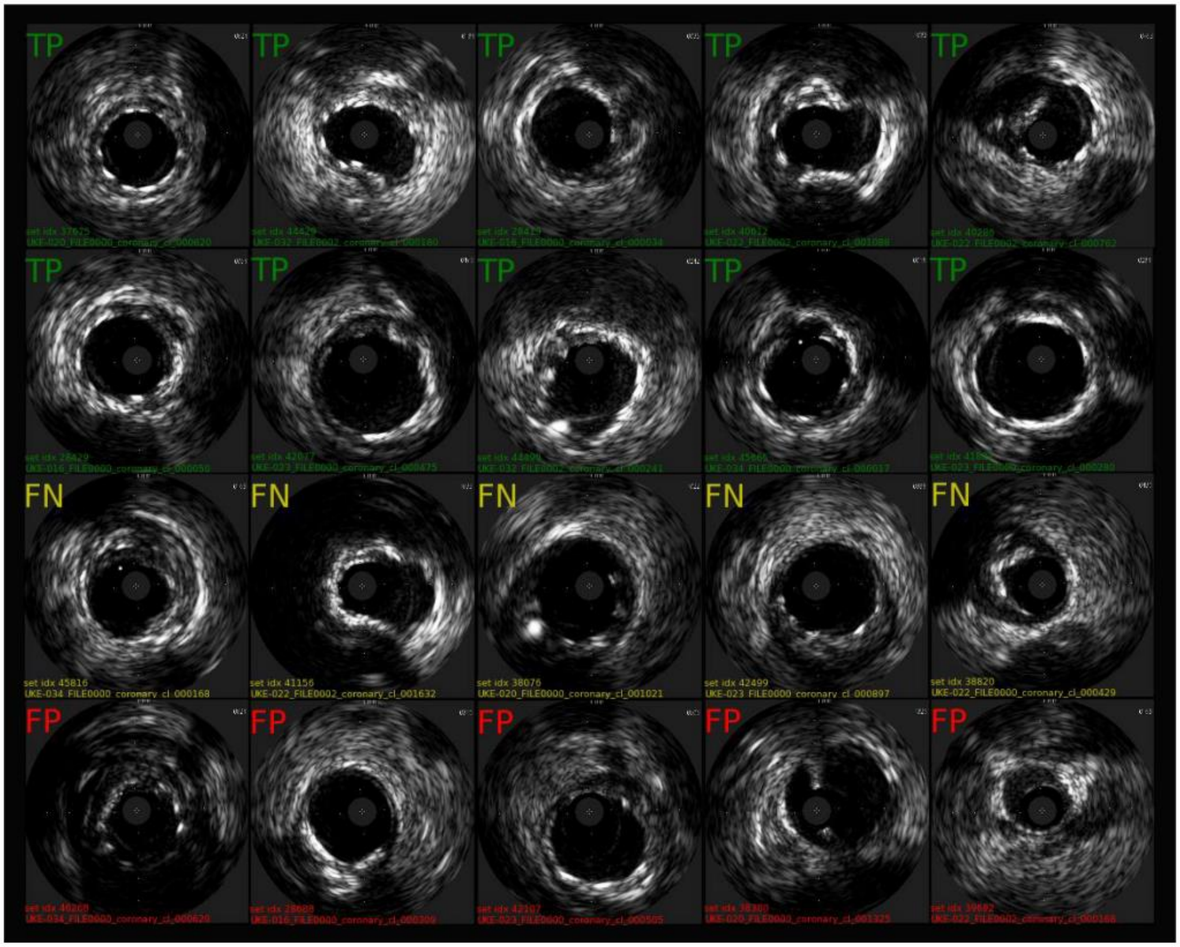
TP
TP
TP
TP
TP
TP
TP
TP
TP
TP
FN
FN
FN
FN
FN
set idx 45816
UKE-034_FILE0000_coronary_cl_000168
set idx 41156
UKE-022_FILE0002_coronary_cl_001632
set idx 38076
UKE-020_FILE0000_coronary_cl_001021
set idx 42499
UKE-023_FILE0000_coronary_cl_000897
set idx 38820
UKE-022_FILE0000_coronary_cl_000429
FP
FP
FP
FP
FP

Figure 6: Example frames labeled with their corresponding predictions from the stent output of the detector network: true positive decisions (top two rows, typical to challenging examples from left to right), false negative decisions (detector missed ground truth annotations, third row), and false positive detections (bottom row, wrong predictions without ground truth label).

In the first two rows very typical stent examples can be seen on the left, where bright speckle spots – indicating metal struts – are visible along the outer lumen contour. Further to the right, frames are getting more challenging: stent struts can only be seen partially along the lumen border (sometimes only two or three of them), calcium deposits are also present in the frame or smaller calcifications are even attached to the stent making it more difficult to distinguish single struts and merging them into arc-shaped structures.

Visual inspection of some of the false positive examples (first and third example in that row) suggests that although ground truth was not labeling them as stent frames, they still seem to clearly show stent struts. In that sense they are only tagged as false positives because they are evaluated against remaining inconsistencies in the ground truth. The other false positives again fall into the ambiguous borderline category described above, where a retrospective decision even for an expert is challenging.

3.2 Stent Segmentation

The second stage of the cascade, the segmentation network, was implemented using an encoder-decoder and a DeepLabV3 architecture trained on pixel-level annotations. The intra-observer variability for the task of stent segmentation yielded a Dice coefficient of 75.47% indicating the extent of agreement between the first and second annotation of the experts. For the second annotation only frames from the first annotation (which definitely contain a stent) were presented to the expert. The Dice metric does therefore not capture any errors the expert would make when re-identifying these frames as stent frames or when ignoring frames without a stent (but maybe calcifications or other ambiguities).

Given this reference context we evaluated the automatic segmentation as follows: The network was exclusively trained on frames of the overall dataset, which are accompanied by pixel-level annotations of the involved classes (stent, calcium, guidewire). All other frames were ignored as previous scouting experiments did not provide any beneficial evidence for the involvement of other

Table 1:Comparison of metrics scoring the quality of stent segmentation on two different data supports: (1) all frames with available ground truth, (2) only frames containing stent pixel-level ground truth. Scores are Dice coefficient, false positive rate (FPR), angular Dice coefficient (DCA) and symmetric median skeleton distance (SMSD) in pixels, i.e. multiples of the image resolution 0.0446mm. Best scores are highlighted in bold.

All	**Dice [%]**	**FPR [%]**	**DCA [%]**	**SMSD [px]**
Encoder-Decoder	43.47	**0.38**	50.29	5.17
DeepLabV3, 120ep, no pretrain	**46.65**	0.71	**54.54**	5.12
DeepLabV3, 60ep, pretrain	45.42	0.9	53.83	**5.06**
Stent Loc				
Encoder-Decoder	**65.05**	**0.51**	**75.20**	**2.77**
DeepLabV3, 120ep, no pretrain	62.50	1.09	72.95	3.47
DeepLabV3, 60ep, pretrain	61.96	1.37	73.69	3.23

additional labels (such as “no stent”). At each epoch a new balanced and shuffled set of frames from this pre-selection was used for training.

In the following we tag this set of frames with the identifier “all” in contrast to the identifier “stent loc”, which refers to the set of frames on which for which a stent ground truth was available on pixel level.

In a first step, we evaluated the performances of the encoder-decoder candidate as well as the DeepLabV3 architecture on both sets. The results in Table 1 show similar scores with either approach. While the encoder-decoder performed slightly better on the stent frames, DeepLabV3 had partly better scores when evaluated on all frames. Note, that the encoder-decoder variant was insensitive to varying thresholds on its probability outputs and therefore used a simple argmax, while the probability threshold on the stent class was optimized for DeepLabV3.

Overall, we consider differences in segmentation quality only as minor and hence proceed with the encoder-decoder variant for further evaluations. With this decision we follow Occam's razor by selecting the simpler of two equally performing approaches. Furthermore, we consider the slightly better performance of the encoder-decoder network on the stent frames as more relevant because the stent detector part of the later cascade will prevent the segmentation network from being applied to all frames. Finally, even on all frames it showed the lowest FPR establishing excellent conditions for the cascade.

In a second step, evaluation and testing of stent segmentation performance then had different foci of interest:

1. Stent segmentation was evaluated on all frames carrying a stent pixel-level label (equivalent to the intra-observer scenario, and similar to evaluations in other studies on segmentation tasks).
2. Stent segmentation was evaluated on all frames of the above-mentioned pre-selection (including all frames which did not contain a stent, but only one of the other labels).
3. Stent segmentation was evaluated based on all frames that passed a gating criterion (either a simple heuristic, or a positive stent detector decision, or both).

The three scenarios above can be interpreted as follows. The first scenario would be the ideal case where it was possible to perfectly identify which frames contain a stent and therefore avoiding any false positives on frames which do not show a stent anyway. This constitutes an upper bound for scores which can be achieved with the other two scenarios. The second scenario is the naive solution: The segmentation network needs to identify frames without a stent on its own and avoid highlighting any pixels in them as stent. Finally, the last scenario is the proposed one, where a mechanism is used to sort out frames at which the segmentation network does not need to look, because they contain no stent. This has the advantage of not only minimizing the risk for unnecessary false positives, but it also reduces computation cost as the segmentation network is only deployed if really required.

3.2.1 Simple Gating Heuristic

A simple heuristic can be constructed under two hypotheses. First, the majority of false positive predictions on frames that do not show a stent are of rather small size and it is rare that the network would claim large areas as stent although they are not. Second, frames which do show a stent typically cover a certain minimal area as there should be a minimal number of struts visible in the image, e.g. at least more than three [40].

Under these circumstances a threshold on size of the segmented area can be applied (in mm² or as the image size is always the same in number of pixels). If the number of pixels segmented as stent in a frame falls below the threshold, then the frame will be cleared from them.

A histogram on areas covered by the strut ground truth masks reveals that areas below 500 pixels are very unlikely. In frames with 224×224 pixels, 125 pixels (0.249 mm²) roughly correspond to the area of one stent strut. We therefore explore thresholds of either 400 or 500 pixels only.

3.2.2 Stent Detector Gating.

In line with the initial proposal for the cascaded approach, we investigate the benefit of using the detector network from Section 3.1 for gating the segmentation task. The segmentation network will only act upon frames which have a positive decision for stent from the detector network.

The detector network was only applied on unseen frames: As the trainings for detector and segmentation networks used the same cross-validation split on patient level, we always applied the corresponding unbiased detector network to the segmentation test fold at hand.

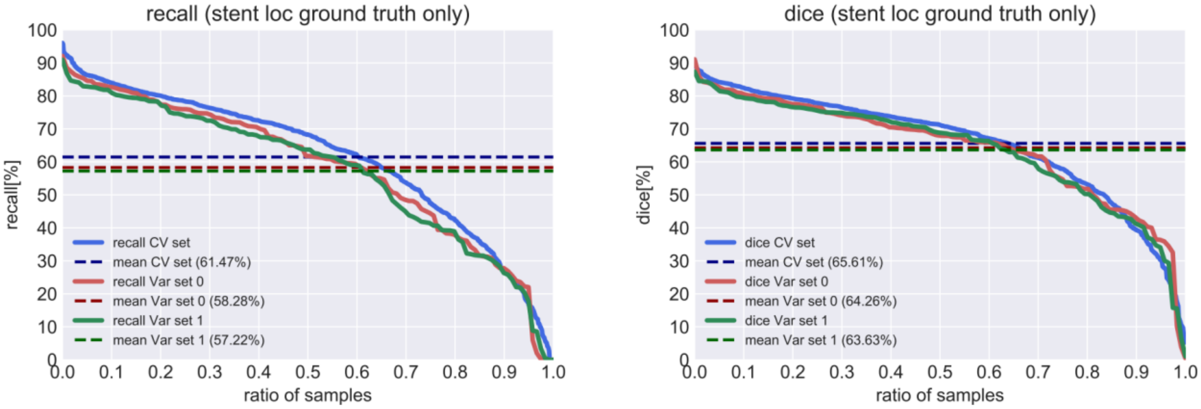

Figure 8: Sorted rank plots for recall and Dice scores computed per-frame. Dashed lines indicate average metric on the intra-observer variance set. Left: Recall ranking for all frames with annotated ground truth. Right: Dice score ranking for all frames with annotated ground truth.

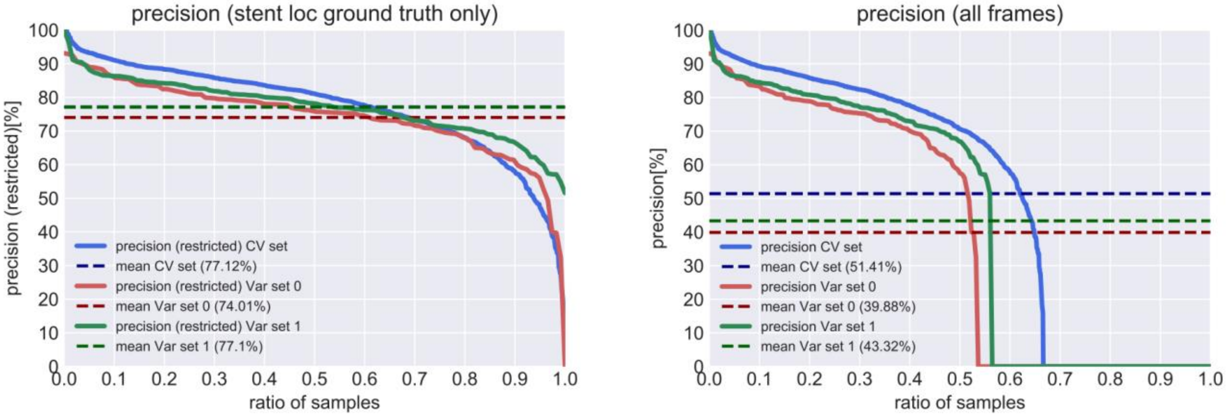

Figure 7: Sorted rank plots for the precision score computed per-frame. Dashed lines indicate average metric on the intra-observer variance set. Left: Precision ranking for all samples with annotated ground truth. Right: Precision ranking for all frames containing automatic segmentations.

Scores can be computed (1) either on the complete set of frames after the gating correction of the detector took place, or (2) scores can be computed only on all frames with a positive detector decision.

Both scores are expected to be similar, but not equal as the detector network still has a low risk for false negatives outside the set of frames with positive detector decision.

3.2.3 Overall Scores

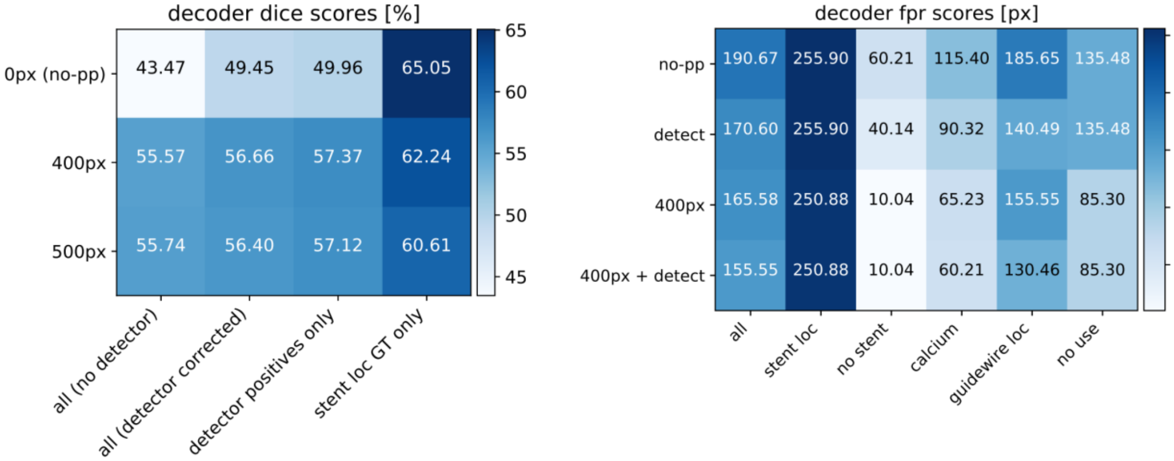

Figure 9: Score matrices for the segmentation network. Scores are presented for different frame supports and post processing (pp) steps: no post-processing (no-pp), segmentation rejection based on a "< N pixel" threshold (Npx) and based on detector decisions (detect). Left: Dice scores for three different post-processing scenarios (vertical) on different frame supports (horizontal). The effect of the detector is evaluated on all frames (segmentations are corrected based on detector decision) and on frames with positive detector decision for stent only. Right: False positive rates (fpr) (listed as average number of pixels in a 224×224 frame) for four different post-processing scenarios (vertical) on different frame supports (horizontal). A false positive rate of 0.25% or 125 pixels (0.249 mm²) roughly corresponds to the area of one stent strut as pictured by the imaging modality.

The first three graphs in Figure 8 and Figure 7 show the scores only computed on all frames that have a pixel-level stent ground truth. Scores (recall, Dice, and precision) have been computed per frame and ranked from the highest to the lowest value. Dashed horizontal lines indicate the average scores on test sets from all folds joined as well as the average scores on the held-out intra-observer set (both first as well as second annotation run of the expert).

Generally, the average scores for all sets are very similar and low scores are only found on a small set of frames (steep fall on the right margin of the Dice and precision plots). The fourth plot shows the precision evaluated on all frames. A precision can be computed on all frames on which the number of predicted stent pixels is different from zero. It can be observed that there is a larger number of frames with zero precision, i.e. frames which did not have a ground truth annotation but were still assigned stent pixels by the segmentation network. Further investigations were carried out to evaluate the characteristics of these predictions (e.g. whether they contain large or small areas of false positives).

3.2.4 Dice Scores and False Positive Rates

Figure 9 (left) presents Dice scores in a matrix for different frame supports (sets of frames on which the score was computed) and different post-processing procedures. In the very right column, the upper bound Dice score is shown. This bound is achieved when only evaluating on frames with stent ground truth. After applying the thresholding heuristic at 500 px or 400 px a small decrease can be observed. While the threshold can correct a moderate number of probable false positive segmentations, it also bears the risk to erroneously reject actual true positives in rare cases.

On the very left column, we list the Dice scores when computing them on all frames. Scores are generally smaller than for the previous case due to frames of zero precision (cf. Figure 7 right). After applying the heuristic, more than 10% improvement in Dice can be observed. This indicates that indeed most frames with zero precision are caused by only small false positively assigned areas in the frame.

When using the detector-based gating, an increase in Dice score can be observed, too and when applying both at the same time, an additional benefit of the detector given the already applied heuristic can be seen. This is true for both scores, those computed on detector frames with positive decision for stent as well as on all frames after detector gating has been applied. The latter is slightly lower, which reflects the small number of false negatives from the detector network.

The Dice scores also indicate that a more conservative threshold of 400 px is more suitable than the higher 500 px one. Overall, the segmentation achieves 86.2% of the human (intra-operator) performance when only applied to frames with a stent ground truth (same scenario the human expert faced) and 75.1% of the human performance on all frames when used within the cascade.

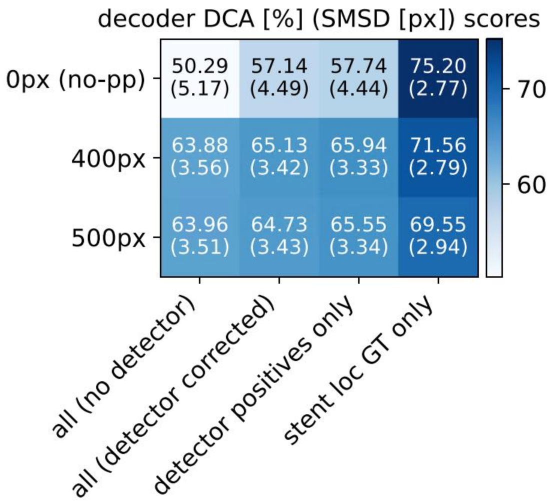

Figure 10: Angular Dice (DCA) and symmetric median skeleton distance (SMSD, in brackets) scores of the segmentation network. The former is reported in percent and the latter in pixel, i.e. multiples of the image resolution 0.0446mm. Equivalent to the Dice scores in Figure 9 results are shown without post-processing (no-pp) and after applying the 400px or 500px heuristic. Again, we evaluate on all frames, stent-only frames as well as on the two cascading scenarios

Figure 9 (right) provides more details on specific false positive rates. The matrix illustrates average false positive rates for different post-processing scenarios (no post-processing, the thresholding heuristic, detector gating or both) and frame supports (arranged according to the expert label they carry and irrespective of whether they have been part of the cross validation sets or not). Again, the combination of detector and heuristic post-processing yields best results.

3.2.5 Angular Dice and Median Skeleton Distance

Equivalent to Figure 9, Figure 10 reports the results for the angular Dice (DCA) and the symmetric median skeleton distance. The overall performance of the segmentation network in terms of these scores confirms the trend observations made on the Dice score earlier. The angular Dice exceeds the full Dice by approximately 8-10% indicating a good angular coverage of the stent ground truth.

The median skeleton distance yielded values in the lower single digit regime which also confirms a segmentation which is well aligned with the ground truth in radial direction. Both scores are constructed to be insensitive with respect to inaccuracies in the thickness of the stent segmentation.

3.2.6 Example Segmentations

Example frames with their corresponding segmentations can be seen in Figure 11. Both images illustrate random examples drawn equally spaced from the score rankings discussed earlier (Dice score in the top and precision in the bottom image). Green areas indicate true positive segmentations (network was correct), red indicates false positive areas (network predicted stent on pixels where ground truth said the opposite), and yellow indicates false negatives (stent areas missed by the column-wise (from left-to-right and top-to-bottom). Scores for the examples decrease from left to right due to the way they are sampled from the ranking plots. Therefore, the first frames give an impression on very good segmentations, while frames on the right indicate some of the poorer frames where the stent is visible in terms of small bright struts along the outer lumen contour.

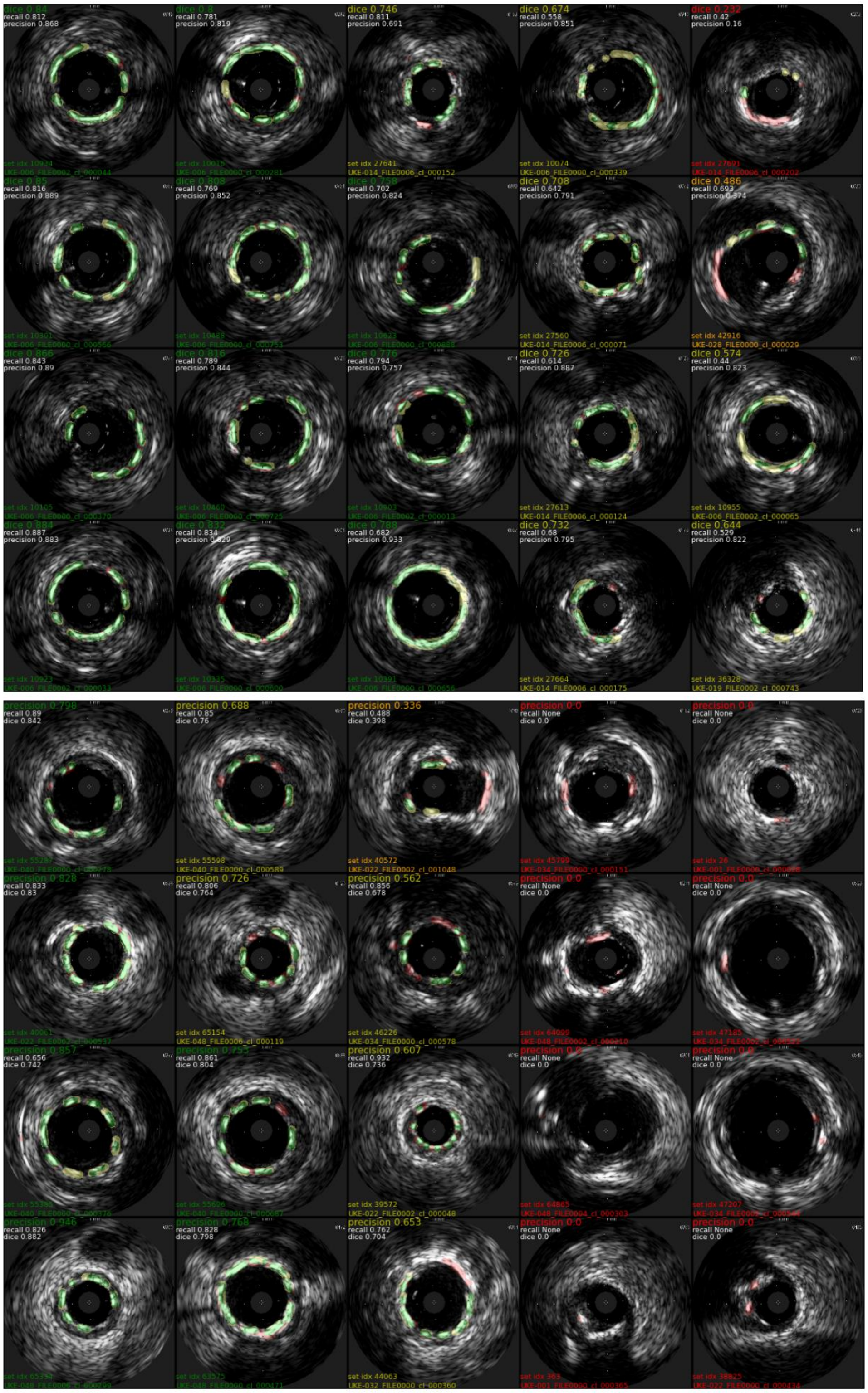

Figure 11: Example frames drawn in steps of constant frame proportions from the score rankings. Therefore, the plots provide a representative selection of frames covering the full range of scores achieved per frame (scores decrease from left to right). Green: true positive pixels, Yellow: false negative pixels, Red: false positive pixels. Top: Frames drawn at equal spacings from the Dice rankings (Figure 8, right). Bottom: Frames drawn at equal spacings from the precision rankings (Figure 7, right).

4 Discussion

4.1 Stent Detection

With a true positive rate beyond 95% stent detection performance was found to be excellent and hardly justifies the usage of more complex and heavy-weight architectures. Increasing complexity by elevating the number of feature maps, has even been found to have a negative effect on the performance.

Possible reasons for the slightly weaker results on the “no stent” class are twofold. On the one hand, it is the class with the least training examples available. This could introduce a bias with respect to the other two classes even though samples are balanced out prospectively by the minibatch sampling. However, the pool of input image variation from which the balanced set can be sampled in each epoch is still higher for the two larger classes.

Moreover, frames without a stent were maybe the most difficult to annotate. While for artifact or stent regions the expert would always choose the indicative example intervals, it is more difficult to define the typical “no stent” interval. Such parts of the pullback can still contain a lot of variation such as guidewires, calcium, bifurcations etc. Furthermore, intervals between clear stent intervals, for instance, could still accidentally contain a (for the human eye) poorly visible stent at the pullback transition into the stent.

The former also appears to be one of the main reasons for false negative stent predictions: Most of the examples in row three of Figure 6 exhibit thickened bright outlines at the lumen border which are easily confused between calcification, stent or calcified stent – even for a human expert. Only rarely a few single struts stand out clearly. Finally, the bottom row of the same Figure illustrates label noise originating from stent frames that were occasionally missed by the expert during the time-consuming annotation process. While most of these frames only contain a few single struts, they still impact the performance measures of the detector in terms of false positive predictions.

4.2 Stent Segmentation

Despite the very promising segmentation quality as evaluated by visual inspection (see for instance Figure 11), the absolute Dice values may seem only moderate. This is a well-known phenomenon which is frequently observed with the segmentation of smaller structures as opposed to segmenting larger organs from images. One example is given by lymph node segmentations from CT images where Dice scores of around 50% are typically achieved [40]. Like stent struts, these structures only occupy smaller areas with a low number of pixels. The label noise is typically higher as they are more difficult to recognize or to precisely annotate by the clinical expert. Apart from overlooking some of them, the annotation brush as used in this work easily misses or erroneously annotates a few pixels. Furthermore, the transition between stent strut and surrounding areas is not always clear due to the speckle noise and point spread function of the image acquisition which operate on a similar scale. Finally, the amount of data and the extent of the annotation burden naturally contribute to this effect and the resulting human inconsistencies. The segmentation network will find a systematic way of labelling stent struts, but inaccuracies of a few pixels will significantly impact the Dice score. The Dice achieved in the intra-observer study confirms this effect without involvement of the network.

Dice scores of the encoder-decoder candidate were found to be like those of the DeepLabV3 candidate. As a possible reason we remark that due to the number of convolutional segments and larger kernel sizes the effective field size of the optimized encoder-decoder network covers almost the full field-of-view. We found this to be an important aspect as only the global view onto the image will provide information about systematic patterns in the composition of a typical IVUS frame and the typical location of a stent within it. The very low number of false positives found further away from the lumen border supports this hypothesis. At the same time the large effective field size prevents the network from getting lost in the interpretations on finer scales comprising ambiguous speckles. Finally, the ASPP blocks with their dilated convolutions in DeepLabV3 may play a similar role for the other candidate. It however does not suffice for a unique advantage as the encoder-decoder candidate can compensate for this with its effective field size. This is also support by the scores achieved. Finally, the lower complexity encoder-decoder network also serves as a regularization against an over-interpretation of the image (see the lower FPR). This has particular importance for predictive tasks on IVUS data on which the risk of misinterpreting speckle as the target seems to be higher than missing a uniquely identifying feature in the appearance of stent struts.

In relative comparisons, the Dice scores are highest when only considering frames that definitely show stent struts. Any segmentation decision on the other frames of the set can only degrade this score and depart from this upper bound or in the best case keep it unchanged. The latter can only be achieved if the segmentation network will not decide for stent in any pixel of these frames. To avoid this decision in first place, the cascade proposes to shift it to either the introduced heuristic, the detector network or both. Indeed, best performances are achieved when combining the 400px heuristic with the detector network. The positive impact of the former confirms that the decrease of the Dice in the larger frame sets mainly originates from a small number of falsely labelled pixels such as bright ambiguously looking speckles.

Yet, the false positive rate is highest in frames containing a stent label. This can be expected as the typical speckle appearance of IVUS frames makes it difficult to decide where a strut starts and where it ends. We have also seen in Figure 6 that for parts of the bright outer lumen contour it is sometimes difficult to judge whether it still is a (e.g. micro-calcified) stent. This also leads to the moderate Dice score between the two expert annotations.

In general, the low average rates for the FPR again confirm that the zero precision frames in Figure 7 are mainly caused by small erroneously segmented areas. The rates are naturally lowest on "no stent" frames, and higher on calcified frames, frames of no clinical use (e.g. with artifacts) or frames showing guidewires.

Once calcifications enter the frame or start to mix with the stent struts, the task gets more difficult for both expert and network. In the bottom image of Figure 11, the very right columns present zero precision frames. Again, we observe only small areas – mainly single, sometimes stretched, bright spots which are mistaken as stent by the network. Some of them are again debatable from a ground truth perspective (such as first row, second from the left).

Finally, the DCA and SMSD confirm the trends discussed above. Moreover, their motivation is twofold: First, they are tailored to the two main clinical use cases identified earlier: the detection of stent malapposition which requires accurately locating the stent contour in radial direction, and re-wiring branches in bifurcation lesions which require knowing which angular segments are occupied by the stent mesh and which parts could allow guidewire passage. The former is targeted by SMSD, which evaluates the median distance between the ground truth and segmentation skeletons of the stent masks. We compute the median as opposed to larger percentiles or the maximum (as done in Hausdorff distances) to focus on the radial errors as much as possible. We expect higher percentiles to be more influenced by relative angular shifts between ground truth and segmentation. This is however already captured by the angular Dice (DCA), which is more dedicated to the re-wiring use case.

As a second motivation, both scores aim at decreasing the impact originating from the "thickness" of the annotated or segmented stents. In addition to the influential factors discussed above, also the brush size favored by the clinical expert may have introduced a bias into the annotations, which should not enter the evaluation. The existence of this bias seems to be confirmed by an angular Dice which is up to 10% higher than the original Dice score while observing a small SMSD at the same time.

4.3 Cascading Concept for a Data-Driven Approach to Stent Segmentation

In the cascaded concept, the encoder successfully served as a gateway for triggering the segmentation network and can thus mitigate the high risk for false positives. This demonstrates the value and superiority of the cascade approach with respect to a conventional pure segmentation. The segmentation network faces the challenge of solving the harder task of pixel-wise stent strut localization while learning from a very limited set of difficult-to-annotate frames. A risk for false positive segmentations then naturally stems from manifold ambiguities in the IVUS image which are difficult to resolve without a more global view on the data – even for the human expert. The disproportionally higher number of frames without a stent compared to those that in fact contain a stent contributes to this. On frames with a stent ground truth the segmentation network achieved 86.2% of the human expert performance. This constitutes the fairest comparison because the expert faced only frames during intra-observer variability analysis of which it was known that they show stent struts. Yet, the detector-segmenter cascade applied to on all frames still reached 75.1% of this previous human performance.

This aspect gains even higher relevance when considering that all predictions operate on a per-frame basis as all pullbacks were ungated and conducted manually unlike the automatic pullbacks from previous work [26, 29]. These are carried out at constant directed speed and allow the algorithm to exploit longitudinal context. This was not the case here, i.e. the transducer could, for instance, remain at a certain location for some time, go back and forth, be pulled at varying speeds, or even get caught in calcified twists and be subject to rapid jumps. In addition, anatomical and device-related ambiguities with similarly bright contrast have also been identified earlier: Liu at al. [34], for instance, report the visibility of the pericardial border close to the vessel, guidewires or varying contrast due to non-orthogonal reflections of the acoustic wave as misleading factors for calcium assessment. These and others certainly also play a role in automatic stent detection as struts typically have a similar localization in the vessel, exhibit similar contrast and texture features and may finally be entangled with other structures such as spotty calcifications or dense fibers [26].

5 Conclusions and Future Directions

In our work, we successfully demonstrated a data-driven deep learning strategy for segmenting stent struts in IVUS frames. We used a cascaded approach which reverses the order of in-frame strut segmentation and longitudinal stent detection as it was proposed in previous work on the same topic which still made use of a handcrafted processing chain [26, 40]. Using a cascade over a conventional frame-wise segmentation along the pullback yielded a successful strategy for opening up the clinical problem to modern data-driven learning. The results confirm our rationale:

Training an encoder network for the simpler task of frame classification on a large set of easy-to-annotate frames yields promising results which can be leveraged to facilitate strut segmentation of a subsequent encoder-decoder segmentation network. Being able to address the problem of stent segmentation by modern learning techniques in such a way will substantially facilitate future development cycles reacting, for instance, on covariate shift stemming from different IVUS transducer types, transducer frequencies, varying post-processing steps in the image acquisition chain or the application to different vasculature. Furthermore, our work sets the stage for easily integrating further predictions in terms of multi-task learning using the same strategy but pushing the exploitation of synergies between related tasks.

We also demonstrated that additional regularization and post-processing can further increase consistency in the segmentation output. Here, we applied a simple heuristic based on the total area covered by the strut prediction in a frame. Although it was not our focus here, future work can make use of more elaborate heuristics such as number and size of connected components in the thresholded output probability mask, or the position of the segmented strut areas with respect to other identified anatomical structures such as the outer lumen contour (internal elastic lamina) or plaques in intima as done in other studies [26, 29].

In this context also other valuable findings from previous studies can complement our work. Inspired by recent successes of combining radiomics and data-driven approaches, the handcrafted features identified by Ciompi et al. [29] can be integrated and possibly further increase robustness of our approach. In addition, the SAX algorithm proposed by Balocco et al. [30] can be investigated as a replacement for the simple thresholding strategy which we used for the encoder part. As the encoder, however, also classifies frames of no use and without a stent in a mutually exclusive manner, the feasibility of a multivariate variant needs to be considered carefully. Our more intuitive likelihood function compared to [30] at the encoder output (defined in [0, 1]) may already sufficiently solve the problem of mapping output probabilities to discrete class decisions. Our Dice curves confirm minimal dependence on the choice of the output threshold.

Finally, future work aims at improving the robustness of our approach e.g. exploiting prior knowledge about the data domain [36] and common error sources [35] or synthetic data generation approaches [37]. Furthermore, while our approach already solves other problems such as the identification of clinically useful frames, e.g. for interventional or retrospective navigation, further predictive models can be combined with the cascade. A natural goal is an extension by existing lumen/vessel wall segmentation models [35] such that stent malapposition can be automatically assessed or a combination with bifurcation and side branch detection to facilitate rewiring by intelligent guidance in complex PCIs.

Disclosures

Dr. Seiffert reports non-financial support from Boston Scientific as well as grants and personal fees from Philips outside the submitted work.

Acknowledgments

This work is supported by European Regional Development Fund (ERDF) and the Free and Hanseatic City of Hamburg in the IFB-Program PROFI Transfer Plus under grant MALEKA.

References

1. Alexandra N. Nowbar, Mauro Gitto, James P. Howard, Darrel P. Francis, Rasha Al-Lamee, 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. 2019; 12:e005375, 2019
2. Zhang J, Gao X, Kan J, Ge Z, Han L, Lu S, Tian N, Lin S, Lu Q, Wu X, Li Q, Liu Z, Chen Y, Qian X, Wang J, Chai D, Chen C, Li X, Gogas BD, Pan T, Shan S, Ye F, Chen SL. Intravascular Ultrasound Versus Angiography-Guided Drug-Eluting Stent Implantation: The ULTIMATE Trial. J Am Coll Cardiol. 2018 Dec 18;72(24):3126-3137. doi: 10.1016/j.jacc.2018.09.013. Epub 2018 Sep 24. PMID: 30261237.
3. Ik Jun Choi, Sungmin Lim, Eun Ho Choo, Byung-Hee Hwang, Chan Joon Kim, Mahn-Won Park, Jong-Min Lee, Chul Soo Park, Hee Yeol Kim, Ki-Dong Yoo, Doo Soo Jeon, Ho Joong Youn, Wook-Sung Chung, Min Chul Kim, Myung Ho Jeong, Youngkeun Ahn, & Kiyuk Chang (2021). Impact of Intravascular Ultrasound on Long-Term Clinical Outcomes in Patients With Acute Myocardial Infarction. *JACC: Cardiovascular Interventions, 14(22), 2431-2443.*
4. Gao XF, Ge Z, Kong XQ, Kan J, Han L, Lu S, Tian NL, Lin S, Lu QH, Wang XY, Li QH, Liu ZZ, Chen Y, Qian XS, Wang J, Chai DY, Chen CH, Pan T, Ye F, Zhang JJ, Chen SL; ULTIMATE

Investigators. 3-Year Outcomes of the ULTIMATE Trial Comparing Intravascular Ultrasound Versus Angiography-Guided Drug-Eluting Stent Implantation. JACC Cardiovasc Interv. 2021 Feb 8;14(3):247-257. doi: 10.1016/j.jcin.2020.10.001. Epub 2020 Oct 29. PMID: 33541535.

5. Trabattoni D, Bartorelli AL. IVUS in bifurcation stenting: what have we learned? EuroIntervention. 2010; 6(suppl J): J88 – J93. https://doi.org/10.4244/EIJV6SUPJA14.

6. Yoon HJ, Hur SH. Optimization of stent deployment by intravascular ultrasound. Korean J Intern Med. 2012; 27 (1):30 – 38. https://doi.org/10.3904/kjim.2012.27.1.30.

7. Fadi J. Sawaya, Thierry Lefèvre, Bernard Chevalier, Phillipe Garot, Thomas Hovasse, Marie-Claude Morice, Tanveer Rab, & Yves Louvard (2016). Contemporary Approach to Coronary Bifurcation Lesion Treatment. *JACC: Cardiovascular Interventions, 9(18), 1861-1878.*

8. Neumann F-J, Sousa-Uva M, Ahlsson A, Alfonso F, Banning AP, Benedetto U, et al., 2018 ESC/EACTS Guidelines on myocardial revascularization. European heart journal. 2019;40(2):87-165

9. ACC/AHA/SCAI Writing Committee. American college of cardiology/american heart association task force on practice guidelines. Journal of the American College of Cardiology, 47(1):e1–e121, 2006.

10. Liang-Chieh Chen and George Papandreou and Florian Schroff and Hartwig Adam, 2017. “Rethinking Atrous Convolution for Semantic Image Segmentation”. CoRR, abs/1706.05587.

11. Ronneberger, Olaf; Fischer, Philipp; Brox, Thomas, 2015. "U-Net: Convolutional Networks for Biomedical Image Segmentation". arXiv:1505.04597

12. 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, 2015, 1520-1528

13. Vijay Badrinarayanan and Alex Kendall and Roberto Cipolla, 2015. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. CoRR, abs/1511.00561.

14. Carole H. Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M. Jorge Cardoso. 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, pages 240–248, 2017.A. Harris et al., “Free-space optical wavelength diversity scheme for fog mitigation in a ground-to-unmanned-aerial-vehicle communications link,” *Opt. Eng.* **45**(8), 086001 (2006) [doi:10.1117/1.2338565].

15. Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In International Conference on Learning Representations (ICLR), 2015.
16. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. 2015.
17. Yarin Gal, Riashat Islam, & Zoubin Ghahramani (2016). Deep Bayesian Active Learning with Image Data. In *Bayesian Deep Learning workshop, NIPS*.
18. Microsoft Corporation. Microsoft cognitive toolkit, release 2.3, 2017.
19. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., … Chintala, S., 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In Advances in Neural Information Processing Systems 32 (pp. 8024–8035). Curran Associates
20. Zhao Wang, Jenkins, M. W., Linderman, G. C., Bezerra, H. G., Fujino, Y., Costa, M. A., Wilson, D. L., & Rollins, A. M. (2015). 3-D Stent Detection in Intravascular OCT Using a Bayesian Network and Graph Search. IEEE transactions on medical imaging, 34(7), 1549–1561
21. Hong Lu, Madhusudhana Gargesha, Zhao Wang, Daniel Chamie, Guilherme F. Attizzani, Tomoaki Kanaya, Soumya Ray, Marco A. Costa, Andrew M. Rollins, Hiram G. Bezerra, and David L. Wilson, "Automatic stent detection in intravascular OCT images using bagged decision trees," Biomed. Opt. Express 3, 2809-2824 (2012)
22. Tsantis S, Kagadis GC, Katsanos K, Karnabatidis D, Bourantas G, Nikiforidis GC. Automatic vessel lumen segmentation and stent strut detection in intravascular optical coherence tomography. Med Phys. 2012 Jan;39(1):503-13. doi: 10.1118/1.3673067. PMID: 22225321.

23. Jouke Dijkstra, Gerhard Koning, Joan C Tuinenburg, Pranobe V Oemrawsingh, & Johan H.C Reiber (2003). Automatic stent border detection in intravascular ultrasound images. International Congress Series, 1256, 1111-1116.

24. Kitahara H, Kobayashi Y, Yock PG, Fitzgerald PJ, Honda Y. Deep learning-based intravascular ultrasound segmentation for the assessment of coronary artery disease. Int J Cardiol. 2021 Jun 15;333:55-59. doi: 10.1016/j.ijcard.2021.03.020. Epub 2021 Mar 16. PMID: 33741429.

25. Iuga, AI., Carolus, H., Höink, A.J. *et al.* Automated detection and segmentation of thoracic lymph nodes from CT using 3D foveal fully convolutional neural networks. *BMC Med Imaging* **21,** 69 (2021). https://doi.org/10.1186/s12880-021-00599-z

26. Simone Balocco, Francesco Ciompi, Juan Rigla, Xavier Carrillo, Josepa Mauri, Petia Radeva, Chapter 10 - Computer-Aided Detection of Intracoronary Stent Location and Extension in Intravascular Ultrasound Sequences, Editor(s): Simone Balocco, Intravascular Ultrasound, Elsevier, 2020, Pages 159-183, ISBN 9780128188330, https://doi.org/10.1016/B978-0-12-818833-0.00010-2.

27. Gatta C, Balocco S, Ciompi F, Hemetsberger R, Rodriguez Leor O, Radeva P. Real-time gating of IVUS sequences based on motion blur analysis: method and quantitative validation. Med Image Comput Comput Assist Interv. 2010;13(Pt 2):59-67. doi: 10.1007/978-3-642-15745-5_8. PMID: 20879299.

28. Francesco Ciompi, Rui Hua, Simone Balocco, Marina Albert, iOriol Pujol, Carles Caus, Josepa Mauri, Petia Radeva (2013) Learning to Detect Stent Struts in Intravascular Ultrasound. In: Sanches J.M., Micó L., Cardoso J.S. (eds) Pattern Recognition and Image Analysis. IbPRIA 2013. Lecture Notes in Computer Science, vol 7887. Springer, Berlin, Heidelberg

29. Ciompi, F., Balocco, S., Rigla, J., Carrillo, X., Mauri, J. and Radeva, P. (2016), Computer-aided detection of intracoronary stent in intravascular ultrasound sequences. Med. Phys., 43: 5616-5625. https://doi.org/10.1118/1.4962927

30. Balocco, S., Ciompi, F., Rigla, J., Carrillo, X., Mauri, J. and Radeva, P. (2019), Assessment of intracoronary stent location and extension in intravascular ultrasound sequences. Med. Phys., 46: 484-493. https://doi.org/10.1002/mp.13273

31. 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, Computerized Medical Imaging and Graphics , 2014, 38, 70 - 90

32. Katouzian, A, Angelini, ED, Carlier, SG, Suri, JS, Navab, N, Laine, AF (2012). A state-of-the-art review on segmentation algorithms in intravascular ultrasound (IVUS) images. IEEE Trans Inf Technol Biomed, 16, 5:823-34.

33. Bargsten L, Schlaefer A. SpeckleGAN: a generative adversarial network with an adaptive speckle layer to augment limited training data for ultrasound image processing. Int J Comput Assist Radiol Surg. 2020 Sep;15(9):1427-1436. doi: 10.1007/s11548-020-02203-1. Epub 2020 Jun 18. PMID: 32556953; PMCID: PMC7419454

34. Shengnan Liu, Tara Neleman, Eline M.J. Hartman, Jurgen M.R. Ligthart, Karen T. Witberg, Antonius F.W. van der Steen, Jolanda J. Wentzel, Joost Daemen, & Gijs van Soest (2020). Automated Quantitative Assessment of Coronary Calcification Using Intravascular Ultrasound. Ultrasound in Medicine & Biology, 46(10), 2801-2809.

35. Bargsten, L., Riedl, K. A., Wissel, T., Brunner, F. J., Schaefers, K., Sprenger, J., Grass, M., Seiffert, M., Blankenberg, S., & Schlaefer, A. (2021). Tailored methods for segmentation of intravascular ultrasound images via convolutional neural networks. In N. V. Ruiter & B. C. Byram (Eds.), Medical Imaging 2021: Ultrasonic Imaging and Tomography. SPIE. https://doi.org/10.1117/12.2580720

36. Bargsten, L., Riedl, K. A., Wissel, T., Brunner, F. J., Schaefers, K., Grass, M., Blankenberg, S., Seiffert, M & Schlaefer, A. (2021). Attention via Scattering Transforms for Segmentation of Small Intravascular Ultrasound Data Sets. In Proceedings of Machine Learning Research 1–14, 2021.

37. Bargsten, L., Schlaefer, A. SpeckleGAN: a generative adversarial network with an adaptive speckle layer to augment limited training data for ultrasound image processing. Int J CARS 15, 1427–1436 (2020). https://doi.org/10.1007/s11548-020-02203-1

Tobias Wissel is a research scientist at Philips Research Hamburg. He received his PhD in robotics and cognitive systems from the University of Lübeck. His research interests include signal and medical image processing as well as machine learning for diagnostics and image-guided therapy with a focus on cardiovascular diseases.

Katharina A. Riedl is a physician at the Department of Cardiology at the University Heart & Vascular Center Hamburg. She received her MD in ultra-high field MRI from the University of Würzburg. Her clinical and research interest include intravascular imaging, cardiovascular magnetic resonance imaging and coronary computed tomography with focus on coronary artery disease.

Klaus Schaefers has studied information management and worked in several research and development positions in organizations such as IBM and Fraunhofer. Since 2016 Klaus is working with Philips Research as a software architect, where he focusses his work on big data and AI related projects.

Hannes Nickisch is a senior research scientist at Philips Research Hamburg. He obtained a PhD from Technical University Berlin and Max Planck Institute in probabilistic modeling. His research interests include medical image analysis and biophysical simulations – in particular for cardiovascular applications – as well as probabilistic machine learning methods.

Fabian J Brunner is a cardiologist at the Department of Cardiology at the University Heart & Vascular Center Hamburg, Germany. His clinical and research interest include the prevention of coronary artery disease as well as its interventional treatment.

Nikolas D. Schnellbächer is a research scientist at Philips Research Hamburg. He obtained a PhD in Physics from Heidelberg University.

His research interests cover statistical physics, signal processing and machine learning and their application for image processing and image reconstruction for medical imaging.

Stefan Blankenberg is a full Professor of Medicine. He is Director of the Clinic for Cardiology and Medical Director at the University Heart & Vascular Center Hamburg.

Moritz Seiffert is head senior physician at the Department of Cardiology at the University Heart & Vascular Center Hamburg. His clinical focus and research interests include complex coronary interventions and percutaneous treatment of valvular heart disease.

Michael Grass is a Principal Scientist at Philips Research Hamburg and a lecturer on Medical Imaging Systems at the Hamburg University of Technology, Germany. He holds a PhD in physics and a Dr. habil. and Privatdozent in engineering for his work on Medical Imaging Systems. Research interests cover diagnostic and interventional medical imaging, tomographic image reconstruction, and artificial intelligence in medical imaging.

List of Captions

Figure 1: Illustration of the cascaded concept. Frames of the manual pullback are first analyzed by an encoder network, which decides for one of three classes per frame: stent, no stent or no use. Only stent frames are then passed on to the encoder-decoder to segment the stent struts. Apart from the favorable training setup, this is also targeting a reduction of false positive predictions on frames that do not show a stent anyway.

Figure 2: Network architectures for stent detection (top, encoder network) and stent segmentation (bottom, encoder-decoder network with skip connections between both parts).

Figure 3: Confusion matrix for the stent detection network after pooling the results on all five test folds.

Figure 4: Left: ROC curve for varying thresholds on the output probability maps. ROC-AUC values are listed for a threshold of t = 0.5. Right: Dice curves showing the dependence of the Dice score on the chosen threshold. A good compromise is achieved when thresholding all classes at 0.5.

Figure 5: Example pullback along with encoder results. First two rows show orthogonal cuts through the pullback in longitudinal direction and the lower three graphs the ground truth (dash-dotted red) and predictions (solid blue line Monte-Carlo mean and blue shading Monte-Carlo estimate of the standard estimation) Regions where none of the three classes has a red ground truth value at 1 have not been annotated by the expert. Here, predictions cannot be compared with a target label. During the last frames the transducer was covered by the catheter, which is correctly recognized as "no use".

Figure 6: Example frames labeled with their corresponding predictions from the stent output of the detector network: true positive decisions (top two rows, typical to challenging examples from left to right), false negative decisions (detector missed ground truth annotations, third row), and false positive detections (bottom row, wrong predictions without ground truth label).

Figure 7: Sorted rank plots for the precision score computed per-frame. Dashed lines indicate average metric on the intra-observer variance set. Left: Precision ranking for all samples with annotated ground truth. Right: Precision ranking for all frames containing automatic segmentations.

Figure 8: Sorted rank plots for recall and Dice scores computed per-frame. Dashed lines indicate average metric on the intra-observer variance set. Left: Recall ranking for all frames with annotated ground truth. Right: Dice score ranking for all frames with annotated ground truth.

Figure 9: Score matrices for the segmentation network. Scores are presented for different frame supports and post processing (pp) steps: no post-processing (no-pp), segmentation rejection based on a "< N pixel" threshold (Npx) and based on detector decisions (detect). Left: Dice scores for three different post-processing scenarios (vertical) on different frame supports (horizontal). The effect of the detector is evaluated on all frames (segmentations are corrected based on detector decision) and on frames with positive detector decision for stent only. Right: False positive rates (fpr) (listed as average number of pixels in a 224×224 frame) for four different post-processing scenarios (vertical) on different frame supports (horizontal). A false positive rate of 0.25% or 125 pixels (0.249 mm^2) roughly corresponds to the area of one stent strut as pictured by the imaging modality.

Figure 10: Angular dice (DCA) and symmetric median skeleton distance (SMSD, in brackets) scores of the segmentation network. The former is reported in percent and the latter in pixel, i.e. multiples of the image resolution 0.0446mm. Equivalent to the dice scores in Figure 9 results are shown without post-processing (no-pp) and after applying the 400px or 500px heuristic. Again, we evaluate on all frames, stent-only frames as well as on the two cascading scenarios

Figure 11: Example frames drawn in steps of constant frame proportions from the score rankings. Therefore, the plots provide a representative selection of frames covering the full range of scores achieved per frame (scores decrease from left to right). Green: true positive pixels, Yellow: false negative pixels, Red: false positive pixels. Top: Frames drawn at equal spacings from the Dice rankings (Figure 7, right). Bottom: Frames drawn at equal spacings from the precision rankings (Figure 8, right).