Lennart Bargsten*, Katharina A. Riedl, Tobias Wissel, Fabian J. Brunner, Klaus Schaefers, Michael Grass, Stefan Blankenberg, Moritz Seiffert, and Alexander Schlaefer

Deep learning for calcium segmentation in intravascular ultrasound images

Abstract: Knowing the shape of vascular calcifications is crucial for appropriate planning and conductance of percutaneous coronary interventions. The clinical workflow can therefore benefit from automatic segmentation of calcified plaques in intravascular ultrasound (IVUS) images. To solve segmentation problems with convolutional neural networks (CNNs), large datasets are usually required. However, datasets are often rather small in the medical domain. Hence, developing and investigating methods for increasing CNN performance on small datasets can help on the way towards clinically relevant results. We compared two state-of-the-art CNN architectures for segmentation, U-Net and DeepLabV3, and investigated how incorporating auxiliary image data with vessel wall and lumen annotations improves the calcium segmentation performance by using these either for pre-training or multi-task training. DeepLabV3 outperforms U-Net with up to 6.3 % by means of the Dice coefficient and 36.5 % by means of the average Hausdorff distance. Using auxiliary data improves the segmentation performance in both cases, whereas the multi-task approach outperforms the pre-training approach. The improvements of the multi-task approach in contrast to not using auxiliary data at all is 5.7 % for the Dice coefficient and 42.9 % for the average Hausdorff distance. Automatic segmentation of calcified plaques in IVUS images is a demanding task due to their relatively small size compared to the image dimensions and due to visual ambiguities with other image structures. We showed that this problem can generally be tackled by CNNs. Furthermore, we were able to improve the performance by a multi-task learning approach with auxiliary segmentation data.

Keywords: Multi-task learning, Small dataset, Coronary artery, Vessel, Convolutional neural network.

https://doi.org/10.1515/cdbme-2021-1021

***Corresponding author: Lennart Bargsten:** Hamburg University of Technology, Institute of Medical Technology and Intelligent Systems, Hamburg, Germany, E-mail: lennart.bargsten@tuhh.de
Alexander Schlaefer: Hamburg University of Technology, Institute of Medical Technology and Intelligent Systems, Hamburg, Germany
Katharina A. Riedl, Fabian J. Brunner, Stefan Blankenberg, Moritz Seiffert: Department of Cardiology, University Heart & Vascular Center Hamburg, Hamburg, Germany
Tobias Wissel, Michael Grass: Philips Research - Hamburg, Germany
Klaus Schaefers: Philips Research - Eindhoven, The Netherlands

1 Introduction

Intravascular ultrasound (IVUS) is an important and frequently used imaging modality in cardiac catheter laboratories. It enables assessing the morphology of vessel structures like lumen, vessel wall and plaque distribution. Intracoronary imaging has been proven to have a positive impact on planning and performing image guided percutaneous coronary interventions [13].

If calcified plaques are present, their volume and shape have to be determined to decide on appropriate treatment. Usually, the physician estimates these morphological features by delineating the plaque in multiple IVUS images. This is a rather time-consuming task and the results depend heavily on the experience of the physician. The automated segmentation of calcifications can therefore help streamlining the clinical workflow. So far, this task has been tackled by methods like energy-based contour models, shape-based techniques or thresholding [1, 14, 18]. In recent years, deep learning methods like convolutional neural networks (CNNs) have gained much importance due to their superior performance on automated analysis of medical images [5, 6, 10, 19]. Segmentation of lumen and vessel wall in IVUS images has also been tackled by CNNs in recent work [2, 11, 16, 17]. However, calcium segmentation with CNNs has not been studied so far.

In order to perform sufficiently well, CNNs have to be trained with large amounts of annotated data. Creating annotations is a rather time-consuming task and has to be performed by clinical experts to ensure high quality. Hence, annotations are usually scarce in the medical domain. One possibility to increase the model performance despite the lack of data is transfer learning. Here, a CNN is pre-trained with

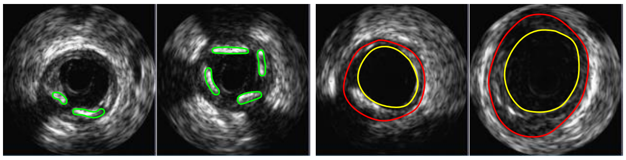

Figure 1: Exemplary images from our dataset and corresponding annotations. The left side shows images with calcium segmentation contours. The right side shows images with lumen (yellow) and vessel wall (red) segmentation contours.

another dataset such that the model weights for the fine-tuning process are initialized advantageously. Another method is multi-task learning, where the CNN learns multiple tasks in parallel. Due to possible commonalities across tasks and the enlargement of the dataset, the network is likely to learn more meaningful features. Furthermore, it is initially unknown which network architectures are beneficial for the underlying task. Therefore, we compared two different network architectures: U-Net [12] and DeepLabV3 [4]. The dataset for our study comprises images with annotated calcifications as well as auxiliary images with annotated vessel wall and lumen.

Our contribution is twofold. First, we show that segmentation of calcifications can successfully be performed by CNNs. Second, we investigated how incorporating auxiliary segmentation data of lumen and vessel wall improves calcium segmentation performance when using this data either for pre-training or for multi-task training.

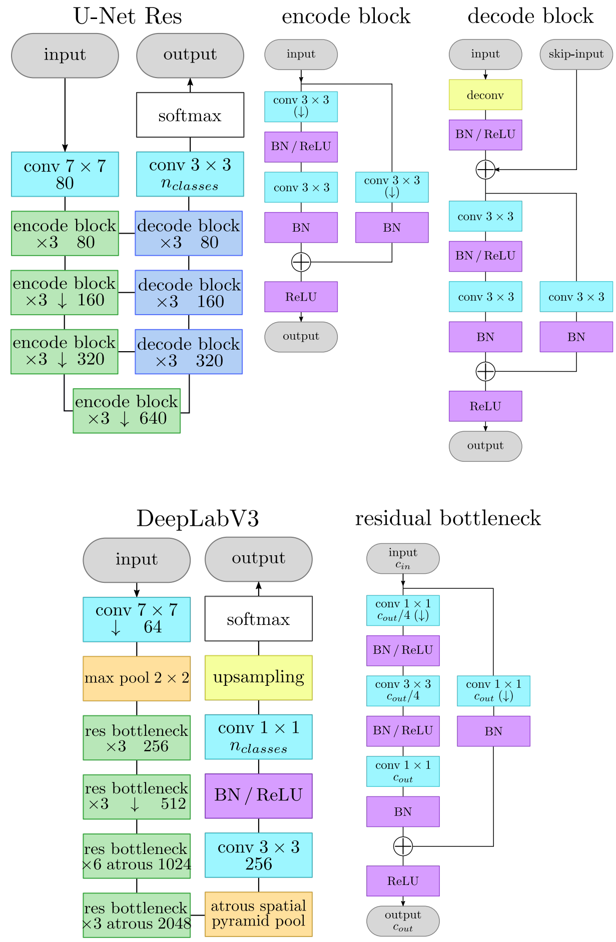

Figure 2: Sketches of CNN architectures used in this work. The first number in encode, decode and residual bottleneck block indicates repetitions. The second number indicates the number of output feature maps. Arrows pointing downwards indicate downsampling via strided convolutions.

2 Methods and material

2.1 Dataset

Our dataset comprises several IVUS sequences of 23 cases in total with 620 annotated frames. 210 frames were annotated for calcifications and 410 frames were annotated for vessel wall and lumen. All annotations were made by an experienced cardiologist. Throughout the rest of this work, the calcification class is referred to as "calcium" and the vessel wall and lumen classes together are referred to as “wall-lumen”. There are no frames which exhibit both wall-lumen segmentation masks and calcium segmentation masks.

All images were captured in a non-gated fashion with a 20 MHz Eagle Eye Platinum phased array transducer (Philips Healthcare, San Diego, USA) and have a size of 500×500 pixels. Figure 1 shows exemplary images and corresponding annotations.

2.2 CNN architectures

We compared two state-of-the-art CNN architectures for segmentation: U-Net [12] with residual blocks [8], which we call U-Net Res throughout the rest of this work, and DeepLabV3 [4] with a ResNet50 [8] backbone. See Figure 2 for sketches of the networks and their building blocks. U-Net Res consists of three residual blocks per downscaling and upscaling stage. Downscaling is performed three times with strided convolutions whereas upscaling is performed with transposed convolutions (also called deconvolutions). The DeepLabV3 network uses a ResNet50 backbone with atrous (or dilated) convolutions and atrous spatial pyramid pooling [3]. Both networks have approximately 40 M parameters and perform three downsamplings in a forward pass. Hence, we can assume that the capacities of both networks are comparable.

Table 1: Calcium segmentation results of different CNN models by means of Dice coefficient and average Hausdorff distance. Bold values indicate best results.

model	training scheme	Dice coefficient [%]	average Hausdorff distance [mm]
baseline	calcium only	57.30 ± 0.87	0.653 ± 0.055
	wall-lumen pre-training	58.38 ± 2.05	0.526 ± 0.081
	wall-lumen multi-task training	60.59 ± 1.81	0.451 ± 0.058
DeepLabV3	calcium only	60.89 ± 1.59	0.501 ± 0.093
	wall-lumen pre-training	61.61 ± 1.20	0.494 ± 0.097
	wall-lumen multi-task training	64.32 ± 1.09	$\mathbf{0.286 \pm 0.044}$
DeepLabV3 COCO	calcium only	61.79 ± 1.78	0.504 ± 0.093
	wall-lumen pre-training	61.91 ± 1.22	0.488 ± 0.046
	wall-lumen multi-task training	$\mathbf{64.51 \pm 0.70}$	0.304 ± 0.055

2.3 CNN Training and evaluation

For both CNN models, we investigated the following training schemes:

1. Training on calcium labels only,
2. Pre-training with wall-lumen labels and subsequent fine-tuning with calcium labels,
3. Multi-task training with calcium and wall-lumen labels in parallel.

In addition, we applied all these training schemes to a DeepLabV3 network which was pre-trained on the COCO dataset [9]. This dataset comprises more than 200 k labelled natural images of 171 different classes.

The calcium dataset was split into a training set comprising 153 images of 10 different cases and a test set comprising 57 images of 6 different cases. The wall-lumen dataset was split into a training set comprising 295 images of 16 different cases and a test set comprising 115 images of 6 different cases. The 6 cases of both test sets are identical for preventing data leakage. All images were resized to 256×256 pixels. During training, an online data-augmentation pipeline was used to randomly flip and rotate the training images. We employed the generalized Dice loss [15], which is a differentiable version of the Dice coefficient with different weights for every segmentation class taking class imbalance into account. The weights are proportional to the inverse ground truth segmentation areas of the individual classes. For optimization we used Adam with a learning rate of $\ell = 2 \cdot 10^{-4}$.

We split the training set into three folds for performing three-fold cross-validation. Random network initializations usually lead to a larger variance in results when training with smaller datasets. To average these out, we repeated three-fold cross-validation ten times and reported the resulting mean and standard deviation on the independent test set. This procedure was performed for each network model and training scheme. We used the Dice coefficient and the average Hausdorff distance [7] as metrics for segmentation performance.

3 Results and discussion

Table 1 shows the calcium segmentation results divided according to the individual CNN models and training schemes. See also Figure 3 for exemplary images. It can be seen, that DeepLabV3 outperforms U-Net Res in all cases. A possible reason could be the use of atrous convolutions and atrous spatial pyramid pooling which are able to capture multi-scale information quite efficiently [4]. This could be beneficial for the task of calcium segmentation because calcified plaques appear in very different sizes and shapes.

The multi-task learning approach led to the best segmentation performances for all models ($p \leq 0.01$ for Welch's t-test in all cases). In contrast to pre-training with the wall-lumen dataset, a parallel training with calcium and wall-lumen labels is more likely to result in meaningful features extracted by the network. In the case of pre-training, the learned features tend to be forgotten during fine-tuning with the calcium labels and thus the networks lose generalizability.

DeepLabV3 models with COCO-pre-training do not have a significant advantage in comparison to the respective models without COCO-pre-training ($p \geq 0.16$ for Welch's t-test in all cases). This is a rather surprising result because models pre-trained on large amounts of natural images usually outperform models without pre-training, even when dealing with medical images. We have two possible explanations. First, the ultrasound image data distribution deviates too much from the COCO image distribution. Second, shifting and scaling of input image pixel values, which is needed when using the

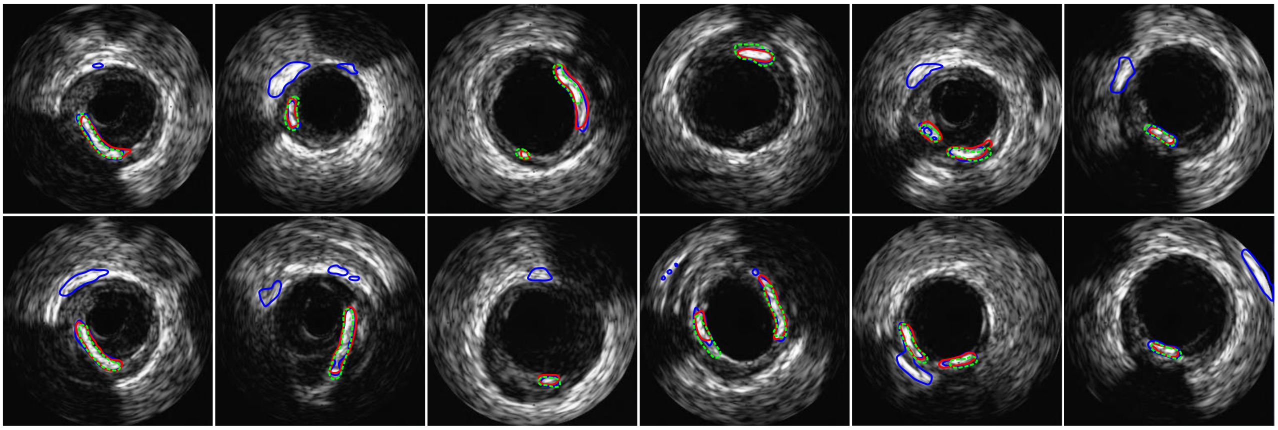

Figure 3: Exemplary calcium segmentation results. Green dashed lines depict ground truth. Blue lines depict results by DeepLabV3 trained with calcium data only. Red lines depict results by DeepLabV3 with multi-task training (proposed).

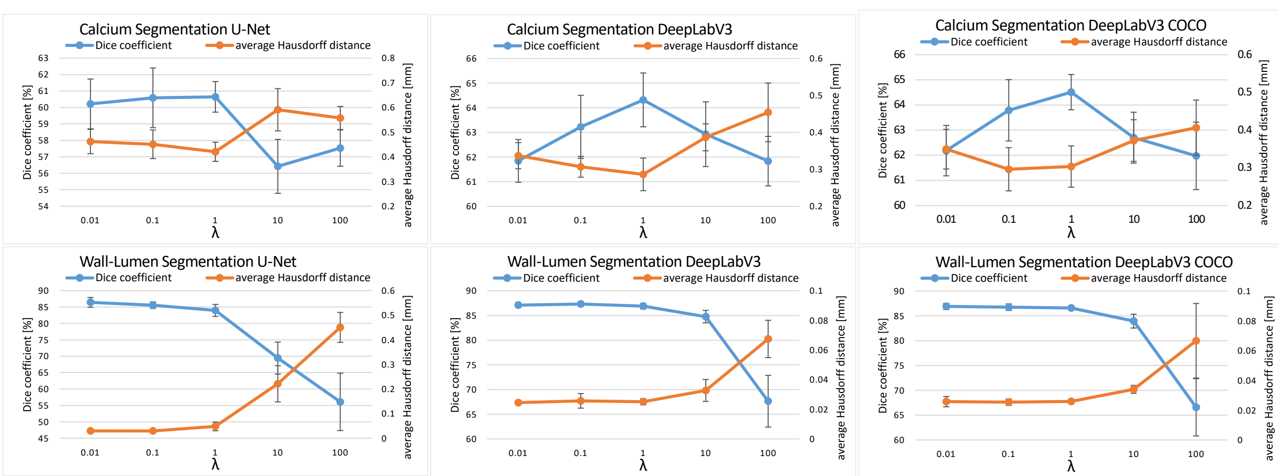

Figure 4: Results for different weighting ratios λ of the multi-task loss functions for calcium segmentation and wall-lumen segmentation (mean of vessel wall and lumen metrics). Note the different scales of the vertical axes. Error bars indicate two standard deviations.

COCO-pre-trained model, transforms the ultrasound image distribution into a rather unfavorable representation.

In the case of multi-task learning, we get an additional hyperparameter $\lambda = w^{\text{loss}}_{\text{calcium}}/w^{\text{loss}}_{\text{calcium}}$ which determines the weighting ratio of the individual loss functions for calcium segmentation and wall-lumen segmentation. Figure 4 shows how the variation of λ affects the performances of both segmentation tasks. For wall-lumen segmentation, metrics are calculated as the mean of individual vessel wall and lumen segmentation metrics. As expected, the performance of wall-lumen segmentation decreases for larger values of λ and saturates for $\lambda \leq 1$. The performance of calcium segmentation has its peak around $\lambda \approx 1$ in all cases. The calcium segmentation performance decreases for large values of λ. This is likely due to the gradients of the wall-lumen segmentation loss having less influence on the extraction of meaningful image features. Interestingly, the calcium segmentation performance of U-Net Res does not drop very much for small values of λ. It seems that in this case U-Net Res is capable of generating meaningful features for both tasks although the gradients of the calcium segmentation loss are comparatively small.

Finally, we want to report some technical metrics. All models were trained on an NVIDIA Titan RTX GPU. With a batch size of 12, inference time of DeepLabV3 was approximately 97 ms whereas the model occupied 5.4 GB of memory. U-Net only achieved an inference time of 268 ms and occupied 15.8 GB of memory. If we simulate deploying both models in clinical practice with a batch size of 1, DeepLabV3 achieves an inference time of 30 ms and occupies 2.1 GB of memory. U-Net achieves 26 ms of inference time and occupies 3.0 GB of memory. This means that both networks are basically able to provide results in real time.

4 Conclusion

In this work, we demonstrated that CNNs are capable of segmenting calcifications in intravascular ultrasound (IVUS) images. DeepLabV3 outperformed U-Net Res in all training schemes, although both had approximately the same capacity. We showed that incorporating auxiliary lumen and vessel wall segmentation data by means of multi-task training increased the performance significantly. Using the auxiliary data for pre-training only led to minor improvements.

Due to the small size of calcifications in contrast to the whole IVUS image, even small wrongly classified areas lead to large performance decreases. In order to achieve a plaque quantification accuracy which is sufficient for clinical applications, future research should focus on increasing the robustness of corresponding deep learning methods. This could be accomplished by larger datasets or the incorporation of further auxiliary tasks like plaque classification.

Author Statement

Research funding: This work was partially funded by the European Regional Development Fund (ERDF) and the Free and Hanseatic City of Hamburg in the Hamburgische Investitions- und Förderbank (IFB)-Program PROFI Transfer Plus under grant MALEKA.

Compliance with ethical standards: The local institutional review board approved our retrospective single-center study and waived the requirement for informed consent.

References

[1] Araki T, Ikeda N, Dey N, Acharjee S, Molinari F, Saba L, et al. Shape-based approach for coronary calcium lesion volume measurement on intravascual ultrasound imaging and its association with carotid intima-media thickness. Journal of Ultrasound in Medicine 2015;34:469–482

[2] Bargsten L, Riedl KA, Wissel T, Brunner FJ, Schaefers K, Sprenger J, et al. Tailored methods for segmentation of intravascular ultrasound images via convolutional neural networks. In: Medical Imaging: Ultrasonic Imaging and Tomography. 2021;11602:1–7

[3] Chen L, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution and fully connected crfs. IEEE Transactions on Pattern Analysis and Machine Intelligence 2018;40:834–848

[4] Chen LC, Papandreou G, Schroff F, Adam H. Rethinking atrous convolutions for semantic image segmentation. ArXiv 2017:1706.05587

[5] Chihati A, Daceb D. A review of recent progress in deep learning-based methods for mri brain tumor segmentation. In: 11th International Conference on Information and Communication Systems. 2020:149–154

[6] Domingues I, Pereira G, Martins P, Duarte H, Santos J, Abreu PH. Using deep learning techniques in medical imaging: a systematic review of applications on ct and pet. Artificial Intelligence Review 2020;53:4093–4160

[7] Dubuisson MP, Jain A. A modified hausdorff distance for object matching. In: Proceedings of the 12th International Conference on Pattern Recognition. 1994:566–568

[8] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition. 2016:770–778

[9] Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL. Microsoft coco: common objects in context. In: Computer Vision – ECCV. 2014:740–755

[10] Liu S, Wang Y, Yang X, Lei B, Liu L, Li SX, et al. Deep learning in medical ultrasound analysis: A review. Engineering 2019;5:261–275

[11] Nandamuri S, China D, Mitra P, Sheet D. Sumnet: fully convolutional model for fast segmentation of anatomical structures in ultrasound volumes. In: IEEE 16th International Symposium on Biomedical Imaging. 2019:1729–1732

[12] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention. 2015:234–241

[13] Räber L, Mintz GS, Koskinas KC, Johnson TW, Holm NR, Onuma Y, et al. Clinical use of intracoronary imaging. Part 1: guidance and optimization of coronary interventions. An expert consensus document of the European Association of Percutaneous Cardiovascular Interventions. European Heart Journal 2018;39:3281–3300

[14] Santos Filho E, Saijo Y, Tanaka A, Yoshizawa M. Detection and quantification of calcifications in intravascular ultrasound images by automatic thresholding. Ultrasound in Medicine and Biology 2008;34:160–165

[15] Sudre CH, Li W, Vercauteren T, Ourselin S, 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. 2017:240–248

[16] Xia M, Yan W, Huang Y, Guo Y, Thou G, Wang Y. Extracting membrane borders in ivus images using a multi-scale feature aggregated u-net. In: 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2020:1650–1653

[17] Yang J, Faraji M, Basu A. Robust segmentation of arterial wall in intravascular ultrasound images using dual path u-net. Ultrasonics 2019;96:24–33

[18] Zhang Q, Wang Y, Wang W, Ma J, Qian J, Ge J, Automatic segmentation of calcifications in intravascular ultrasound images using snakes and contourlet transform. Ultrasound in Medicine and Biology 2010;36:111–129

[19] Zhou SK, Greenspan H, Davatzikos C, Duncan JS, van Ginneken B, Madabhushi A, et al. A review of deep learning in medical imaging: imaging traits, technology trends, case studies with progress highlights and future promises. Proceedings of the IEEE 2021;109:820–838