Adversarial Training for Patient-Independent Feature Learning with IVOCT Data for Plaque Classification

Nils Gessert[*]
nils.gessert@tuhh.de

Markus Heyder[*]
markus.heyder@tuhh.de

Sarah Latus[*]
sarah.latus@tuhh.de

David M. Leistner[†]
david-manuel.leistner@charite.de

Youssef S. Abdelwahed[†]
youssef.abdelwahed@charite.de

Matthias Lutz[‡]
matthias.lutz@uksh.de

Alexander Schlaefer[*]
schlaefer@tuhh.de

Abstract

Deep learning methods have shown impressive results for a variety of medical problems over the last few years. However, datasets tend to be small due to time-consuming annotation. As datasets with different patients are often very heterogeneous generalization to new patients can be difficult. This is complicated further if large differences in image acquisition can occur, which is common during intravascular optical coherence tomography for coronary plaque imaging. We address this problem with an adversarial training strategy where we force a part of a deep neural network to learn features that are independent of patient- or acquisition-specific characteristics. We compare our regularization method to typical data augmentation strategies and show that our approach improves performance for a small medical dataset.

1 Introduction

Deep learning methods have been successful for a variety of medical image-based learning problems, such as segmentation or disease classification [1]. Due to time-consuming annotation, e.g., by a trained expert, medical datasets are often small compared to typical datasets in the natural image domain [2]. Often, medical datasets are heterogeneous due to differences between patients which makes generalization to new patients challenging. Variations in data acquisition can complicate this even further. As these variations are independent of the underlying disease, a deep learning model should learn invariance towards these properties, given a large enough dataset. However, for small datasets, a model might quickly overfit to patient- or acquisition-specific features, which requires additional regularization or data augmentation.

This problem is present in particular for intravascular optical coherence tomography (IVOCT) imaging. The technique is used for imaging the inside of coronary arterial walls in order to identify plaque deposits that might lead to atherosclerosis. Deep learning-based plaque classification methods have been proposed, however, their performance is limited [3]. Thus, data augmentation and transfer learning have been used for IVOCT-based classification [4] which significantly improved performance.

---

[*]Hamburg University of Technology, Am Schwarzenberg-Campus 3, 21073 Hamburg, Germany.
[†]Charité – Universitätsmedizin Berlin, Hindenburgdamm 30, 12203 Berlin, Germany
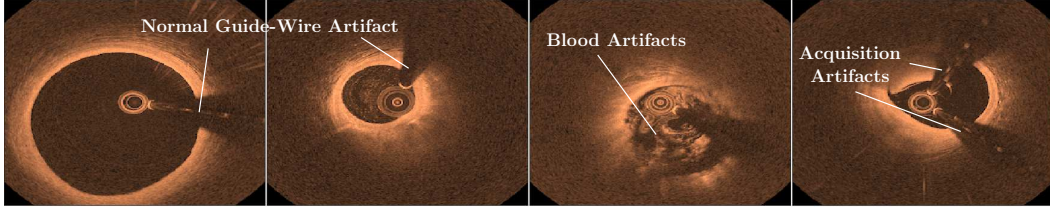[‡]Universitätsklinikum Schleswig-Holstein, Arnold-Heller-Straße 3, 24105 Kiel, Germany

Figure 1: Exemplary IVOCT images from different patients in cartesian representation. Note the differences in shape and acquisition artifacts.
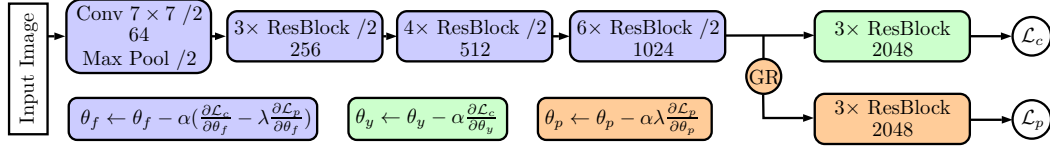

Figure 2: Sketch of the used architecture. The number in each block denotes the number of output feature maps. /2 denotes a spatial reduction by 2 within the block. GR denotes the gradient reversal layer. θ denotes the learnable weights in each part of the model. α denotes the learning rate. Note that the gradient updates are symbolic and the actual updates are performed with the Adam algorithm.

We propose a new regularization method that enforces patient-independent feature learning for small medical datasets. The adversarial training strategy is inspired by domain adversarial training [5]. We adopt the proposed structure and use it for a very different purpose. We show its effectiveness compared to standard data augmentation methods on a small dataset for IVOCT-based plaque classification.

2 Methods

2.1 Dataset & Learning Problem

We create a new dataset of IVOCT images for plaque classification; as to our knowledge, none are publicly available. The data is acquired in-vivo with a St. Jude Medical Ilumien OPTIS. We use the acquired data in cartesian representation, i.e., the images represent 2D slices of a coronary artery. In Figure 1 we show example images from different patients. Three expert cardiologists with daily routine in IVOCT usage provide the binary ground-truth labels. Each 2D image is assigned a label, "plaque" or "no plaque". In total, the dataset consists of 2600 2D images from 36 patients. For testing, 700 images from 8 patients are split off from that set. In order to show our method's effectiveness with small datasets, we also use a reduced training set size of 20 patients.

2.2 Adversarial Training Strategy

Our base model is a convolutional neural network (CNN), Resnet50 with post-norm structure [6]. Similar to [4], the model is pretrained on the ImageNet dataset. Following [5], we introduce an additional path, close to the output of the architecture, see Figure 2. The second path is also initialized with the pretrained weights. The main path leads to a binary output that classifies the image with respect to the plaque type present. The second path leads to a classification layer with N outputs, classifying the patient dataset where the image originated from . N is the number of patients in the training set. Using a gradient reversal layer [5], the loss function $\mathcal{L}(\theta_f, \theta_y, \theta_p) = \mathcal{L}_c(\theta_f, \theta_y) - \lambda \mathcal{L}_p(\theta_f, \theta_p)$ is minimized which leads to the gradient updates depicted in Figure 2. $\mathcal{L}$ is the cross-entropy loss function and λ is a positive scalar which defines the trade-off between the disease and the patient classification loss. We choose $\lambda = 0.5$ for our experiments. The two paths compete in an adversarial way where the main (blue) path attempts to learn features which make it impossible to classify from which patient the input image originated. Thus, the main path learns weights θ_f that maximize the loss $\mathcal{L}_p$ while the second path learns weights θ_p that minimize the loss $\mathcal{L}_p$, using the features produced by the main path.

	N	AUG	ADV	**AUG+ADV**	AUG 20	**ADV 20**	AUG+ADV 20
Accuracy	0.776	0.833	0.873	**0.883**	0.726	**0.816**	0.805
Sensitivity	0.856	**0.903**	0.857	0.896	0.740	0.832	**0.859**
Specificity	0.698	0.755	**0.898**	0.872	0.675	**0.800**	0.752

Table 1: Comparison of different training scenarios. `AUG` denotes data augmentation being applied. `ADV` denotes our adversarial training strategy. `N` denotes training without either. 20 denotes training with a reduced training set size with 20 patients.

We compare this approach to typical data augmentation and regularization. We employ dropout with a keep probability of $p = 0.8$. Moreover, we apply random cropping with a crop size of 270×270 from the original image with size 300×300. Additionally, random rotations with $\alpha \in [0°, 360°]$ and random flipping are applied. We implement our models in Tensorflow [7].

3   Results

The results for various training scenarios are shown in Table 1. Our method shows better results when being applied on top of data augmentation and also without data augmentation. Moreover, the difference in performance becomes even larger when training with a reduced training set.

4   Discussion & Conclusion

We propose a new regularization method for deep learning problems with small datasets. The method uses an adversarial training strategy where the network is forced to learn features that are invariant towards patient-specific characteristics. Our results in Table 1 show that our method outperforms classic data augmentation techniques for our dataset. When applied on top of data augmentation, accuracy is increased slightly more. However, it is notable that additional data augmentation has little effect when using the adversarial training strategy. This implies that our method already enforces invariance towards properties that are usually covered by data augmentation. When reducing the training set size, the improvement becomes even more pronounced. This implies that our method prevents overfitting to patient-specific features more effectively than standard augmentation methods. For further research, our method could be tested with different datasets and more CNN models.

References

[1] Greenspan, H., van Ginneken, B., Summers, R.M. (2016) Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. IEEE Transactions on Medical Imaging **35**(5), 1153–1159

[2] Shin, H.C., Roth, H.R., Gao, M., Le Lu, Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M. (2016) Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Transactions on Medical Imaging **35**(5), 1285–1298

[3] Abdolmanafi, A., Duong, L., Dahdah, N., Cheriet, F. (2017) Deep feature learning for automatic tissue classification of coronary artery using optical coherence tomography. Biomedical Optics Express **8**(2), 1203–1220

[4] Gessert, N., Heyder, M., Latus, S., Lutz, M., Schlaefer, A. (2018) Plaque Classification in Coronary Arteries from IVOCT Images Using Convolutional Neural Networks and Transfer Learning. In: Proceedings, supplement of the International Journal of CARS'2018, Accepted for Publication

[5] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V. (2016) Domain-adversarial training of neural networks. The Journal of Machine Learning Research **17**(1), 2096–2030

[6] He, K., Zhang, X., Ren, S., Sun, J. (2016) Identity mappings in deep residual networks. In: ECCV, pp. 630–645

[7] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M. (2016) Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467