Master's Thesis on Early Stopping without Pixel-wise Annotations for Semantic Segmentation with Scarce Data

Advisor: Simon Reiß (simon.reiss@kit.edu)

1 Motivation and relevance

Deployment of semantic segmentation systems is often slowed down by the time-consuming task of data annotation. Especially when expensive experts need to provide pixel-wise annotations as, *e.g.* in medical image segmentation tasks. Finding strategies to get more out of a few labeled examples can help immensely when training segmentation models and bringing them to real applications.

2 Coarse topic and starting point

In network training, a good parameter configuration for the segmentation network, *i.e.* before phasing into overfitting is determined by evaluating the model that is currently trained periodically on a validation set [2]. To get the best network configuration, one can select the model that performed best on this validation set. This is especially important for extremely small datasets.
For semantic segmentation, this validation set is, a collection of pixel-wise annotated images. These annotations are very costly and thus, having to set up a training and validation set adds to the costs in development of a segmentation algorithm in practise. A question that comes to mind is whether we can find a criterion for performing early stopping that is not dependent on a fully pixel-wise annotated validation set but rather a more cost efficient alternative which still quantifies generalization capabilities.
In semi-weakly supervised semantic segmentation [3] we have access to either bounding boxes, scribbles, image-level labels or potentially unlabeled data. The questions I want to answer together with you is: Can we design a fitting validation criterion for early stopping based on these weak annotations? How well do early stopping approaches that omit validation samples completely [1] work in scarce data domains?

3 Requirements

English language skills and good programming skills in Python are required. Some basic knowledge about the machine learning framework `https://pytorch.org/` is recommended. You should have a solid understanding of deep learning models, e.g., by having attended lectures like {Deep Learning for Computer Vision, Computer Vision for Human Computer Interaction, Neural Networks, Machine Learning}. Take a quick glance at the referenced papers, if the topic sounds exciting, interesting and you feel like it's a big enough challenge for you, feel free to contact me.

References

[1] Ciprian A Corneanu, Sergio Escalera, and Aleix M Martinez. Computing the testing error without a testing set. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2677–2685, 2020.

[2] Lutz Prechelt. Early stopping-but when? In *Neural Networks: Tricks of the trade*, pages 55–69. Springer, 1998.

[3] Simon Reiß, Constantin Seibold, Alexander Freytag, Erik Rodner, and Rainer Stiefelhagen. Every annotation counts: Multi-label deep supervision for medical image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9532–9542, 2021.