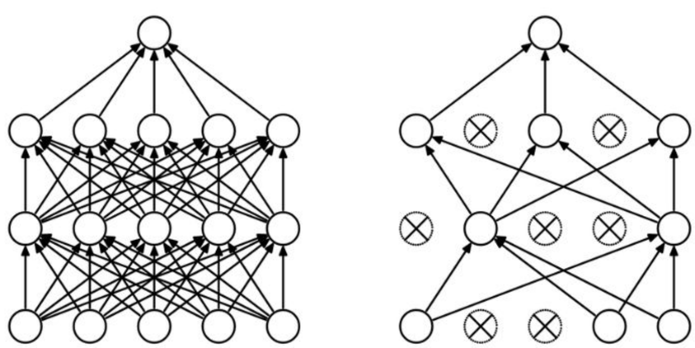
Computationally-Efficient Deep Learning for 3D Medical Imaging

Contact: Aiham Taleb (aiham.taleb@hpi.de)

Convolutional Neural Networks (CNNs) have advanced the state-of-the-art (SOTA) in many application fields, especially in visual tasks such as image recognition and semantic segmentation. However, such advancements in prediction performance incur additional computational costs. This is due to the growing model sizes (in number of parameters) and the added computational complexity related to that, e.g. the models that achieve SOTA on the imagenet benchmark have ~480M parameters [9]. Training such large models is a long process with large hardware requirements. Such model complexity issues become more pronounced when operating on 3D images, e.g. scans in the medical domain.

As a result, it has become vital to search for more efficient solutions for such applications. Here, by efficiency we mean model efficiency, i.e. where the model reaches satisfactory results with minimal computational requirements. To achieve model efficiency, there are many approaches in literature. One common approach to reduce the computational requirements is to replace the Convolutional blocks in CNNs with other, more efficient alternatives. Examples include: Shift blocks [2], Squeeze-and-excitation blocks [1], Self Attention layers [3], and Fire modules [4]. Another approach to reduce the model complexity is by utilizing model compression techniques, such as by pruning [4,5], and by sparsification [6]. One can also exploit data-specific properties, such as the temporal locality in 3D data, to improve the model-efficiency in CNNs [7]. Finally, design choices can influence the size of the CNN architecture being used [4,8], and hence can reduce the computational requirements significantly.



In this thesis, you will work on this line of research, and you can choose the theme of your thesis from the following:

- Applying efficient deep learning architectures to several 3D datasets and measuring the savings in model complexity. It is not mandatory to choose efficient techniques from above, these are only examples.
- Implementing new efficient techniques that may reduce the model complexity.

Your profile

- Master's student in Computer Science (ITSE), Digital Health (DH), Data Engineering (DE), and all related programs.
- Knowledge in areas of Deep Learning (ideally attended deep learning courses offered by the chair)
- Good programming skills (e.g. Python)
- Experience in Deep Learning frameworks (e.g. Tensorflow or PyTorch) is a great plus
- Quick learner and willing to share knowledge
- Good English language skills

If you find this topic interesting, please contact us.

References

[1] Squeeze and excitation: https://arxiv.org/abs/1709.01507
[2] Shifts: https://openaccess.thecvf.com/content_cvpr_2018/html/Wu_Shift_A_Zero_CVPR_2018_paper.html
[3] Self-Attention: https://arxiv.org/abs/1906.05909
[4] SqueezeNet: smaller CNNs with Fire-modules: https://arxiv.org/abs/1602.07360
[5] Model Compression through pruning: https://arxiv.org/abs/1510.00149
[6] Model Compression through sparsification: https://arxiv.org/abs/1803.03635
[7] Leveraging temporal locality: https://dl.acm.org/doi/abs/10.1145/3307650.3322260?casa_token=PEhPrSNF_WgAAAAA:SHI68NRKOVB5rmIJRbCyde_SIya0JLBZkoN7rfFQCmBspGz8fYYOT3JDz09t0DXFMa1C_zSaEg5C
[8] Various efficient networks: https://arxiv.org/abs/1904.02422
[9] https://paperswithcode.com/sota/image-classification-on-imagenet