Robustness Evaluation of Pathology Foundation Models

Type: Msc Thesis (CIT / MH)
Starting date: ab sofort
Supervisor: Prof. Peter Schüffler (https://schuefflerlab.org)
Contact: peter.schueffler@tum.de

Summary

Pathology AI foundation models are trained on massive proprietary datasets. They serve as highly specific feature generators from histology images that can be used in multiple downstream tasks such as cancer detection, segmentation, subtyping, and prognosis estimation. Foundation models can empirically be evaluated and compared on standardized frameworks such as eva[1] PathPench[2], or PathoBench[3]. However, their robustness to image perturbations and data set shift is unclear. In this project we build a benchmark framework which considers augmented variance of data sets to estimate the robustness of foundation models

Daten

We will use public datasets and foundation models (see tasks).

Tasks

T1: Literature research and review / testing of a suitable benchmal method (e.g., eva[1] PathPench[2], or PathoBench[3]). The method must allow for the use of custom data sets.

T2: Selection of appopropriate public benchmark data sets (e.g. patchCamelyon[1]), and appropriate augmentation methods (color augmentation[4,5], compression levels[6], rotations/flipping). Visualization of the corresponding artifacts.

T3: Testing of different foundation (e.g. UNI[7], Virchow2[8], cTransPath[9], TUM) models according to the original and augmented data sets. Evaluation of the different models for their stability across augmentations.

An internal cluster (DGX H100) and the LRZ are available for training. Coding is usually done in Python (ScanPy, SquidPy, PyTorch).

References

1. kaiko.ai, Gatopoulos I, Känzig N, Moser R, Otálora S. eva: Evaluation framework for pathology foundation models. In: ; 2024. Accessed May 30, 2025. https://openreview.net/forum?id=FNBQOPj18N

2. Sun Y, Wu H, Zhu C, et al. PathBench: Advancing the Benchmark of Large Multimodal Models for Pathology Image Understanding at Patch and Whole Slide Level. *IEEE Trans Med Imaging*. Published online 2025:1-1. doi:10.1109/tmi.2025.3584857

3. Zhang A, Jaume G, Vaidya A, Ding T, Mahmood F. Accelerating Data Processing and Benchmarking of AI Models for Pathology. Published online 2025. doi:10.48550/ARXIV.2502.06750

4. Faryna K, Van Der Laak J, Litjens G. Automatic data augmentation to improve generalization of deep learning in H&E stained histopathology. *Computers in Biology and Medicine*. 2024;170:108018. doi:10.1016/j.compbiomed.2024.108018

[1] https://patchcamelyon.grand-challenge.org/

5. Cubuk ED, Zoph B, Mane D, Vasudevan V, Le QV. AutoAugment: Learning Augmentation Strategies From Data. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE; 2019:113-123. doi:10.1109/CVPR.2019.00020

6. Fischer M, Neher P, Schüffler P, et al. Unlocking the potential of digital pathology: Novel baselines for compression. *Journal of Pathology Informatics*. Published online January 2025:100421. doi:10.1016/j.jpi.2025.100421

7. Chen RJ, Ding T, Lu MY, et al. A General-Purpose Self-Supervised Model for Computational Pathology. Published online August 29, 2023. Accessed October 30, 2023. http://arxiv.org/abs/2308.15474

8. Zimmermann E, Vorontsov E, Viret J, et al. Virchow2: Scaling Self-Supervised Mixed Magnification Models in Pathology. Published online August 14, 2024. Accessed September 5, 2024. http://arxiv.org/abs/2408.00738

9. Wang X, Yang S, Zhang J, et al. TransPath: Transformer-Based Self-supervised Learning for Histopathological Image Classification. In: De Bruijne M, Cattin PC, Cotin S, et al., eds. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*. Vol 12908. Lecture Notes in Computer Science. Springer International Publishing; 2021:186-195. doi:10.1007/978-3-030-87237-3_18