



Native Vision Transformers for Variable-Resolution Chest X-Ray Classification

Thesis FA/MA/BA
Supervisor Jakob Hofmann
Examiner Prof. Dr.-Ing. Bin Yang

Date
22. April 2026

Motivation

Vision Transformers (ViTs) have become the de facto standard for medical image analysis, yet they impose a rigid constraint: all input images must be resized to a fixed resolution before being split into equal-sized patches. In clinical practice, however, chest X-rays arrive in variable native resolutions depending on the acquisition device, patient positioning, and hospital infrastructure. Current practice discards this information by forcibly resizing all images, potentially destroying fine-grained details in high-resolution scans or wasting computational resources on artificially upsampled low-resolution images.

This thesis explores native-resolution Vision Transformers that can process chest X-rays at their original resolutions without preprocessing. Building on the NaViT approach, which uses the native-resolution, the student will investigate how variable-resolution processing affects classification accuracy on large-scale medical datasets.

Objectives

- Implement a native-resolution Vision Transformer architecture based on the NaViT method, adapted for single-image chest X-ray classification tasks.
- Train and evaluate the model on the MIMIC-CXR dataset, handling its heterogeneous native resolutions and comparing performance against standard fixed-resolution ViT baselines.
- Analyze the resolution-accuracy trade-off: investigate whether processing images at native resolution improves detection of subtle pathologies compared to resizing, and measure computational savings from avoiding unnecessary upsampling.

Prerequisites

- Solid background in machine learning, deep learning, and computer vision
- Good programming skills in Python
- Experience with PyTorch
- Confidence in reading and understanding recent research papers

If this topic has sparked your interest, write me an email and we can discuss the proposal in more detail. Please include your current transcript and CV.

References

- [1] A. E. W. Johnson, T. J. Pollard, S. J. Berkowitz, N. R. Greenbaum, M. P. Lungren, C.-y. Deng, R. G. Mark and S. Horng, "MIMIC-CXR: A large publicly available database of labeled chest radiographs," *arXiv preprint arXiv:1901.07042*, 2019. [Online]. Available: <https://arxiv.org/abs/1901.07042>
- [2] M. Dehghani, B. Bas, A. Gritsenko, L. Beyer, X. Zhai, A. Arnab, A. Mustafa, F. Motamed, J. Puigcerver and M. Minderer, "Patch n' pack: NaViT, a vision transformer for any aspect ratio and resolution," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2023. [Online]. Available: <https://arxiv.org/abs/2307.06304>

Jakob Hofmann
jakob.hofmann@iss.uni-stuttgart.de