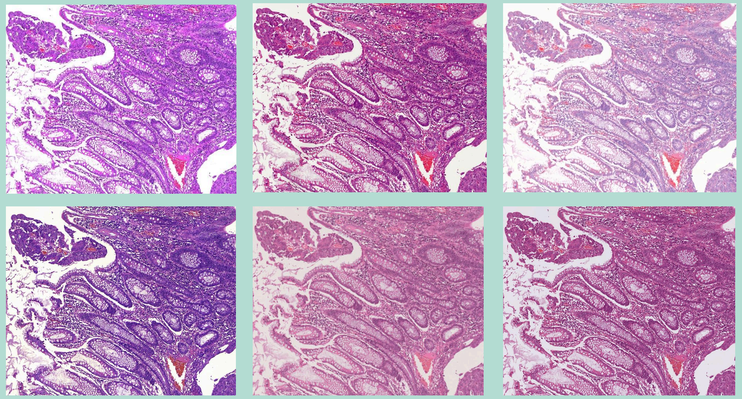
Comparison of CNN models on a multi-scanner database in colon cancer histology

Petr Kuritcyn[1], Michaela Benz[1], Jakob Dexl[1], Volker Bruns[1], Arndt Hartmann[2], Carol Geppert[2]

[1] Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany [2] Institute of Pathology, University Hospital Erlangen

Motivation

One of the most important challenges for computer-aided analysis in digital pathology is the development of robust deep neural networks, which can cope with variations in color and resolution of digitized whole-slide images (WSIs).



Color variations in WSIs from different scanners (upper row, left to right: MIDI (original), M8, iSTIX, lower row: Scube, S210, S360 .

Goal

Comparison of different state-of-the-art CNNs on a multi-scanner database in terms of their robustness and inference time.

Methods

Augmentation during training aids a model to generalize better to heterogeneous data. We introduced color variance in the training data using a combination of hue, saturation and H&E color augmentations. Additionally, we added a blur augmentation to counter the presence of out of focus regions in some WSIs. For all models, the Adam optimizer with a learning rate of 0.001 and an exponential decay was used. Each network was trained three times and test results were averaged. Inference tests were done with a NVIDIA GeForce GTX 1060 using the TensorFlow 2.3 C API with a batch size of 30 and averaging over 5275 batches.

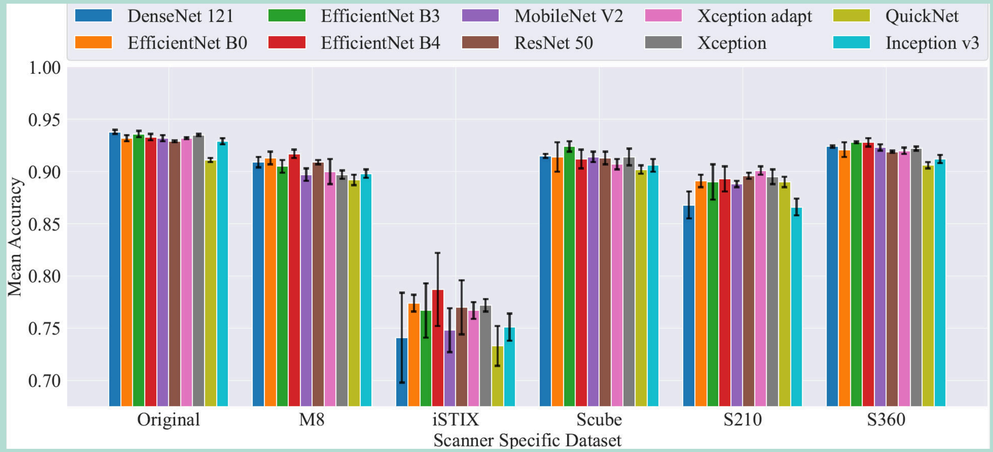
Multi-scanner database

Scanner	Resolution in µm/pixel	#Patches for test
MIDI	0.22	1,381,316
iSTIX	0.17	2,123,364
Scube	0.27	857,511
M8	0.35	514,397
S210	0.22	1,424,716
S360	0.23	1,298,056

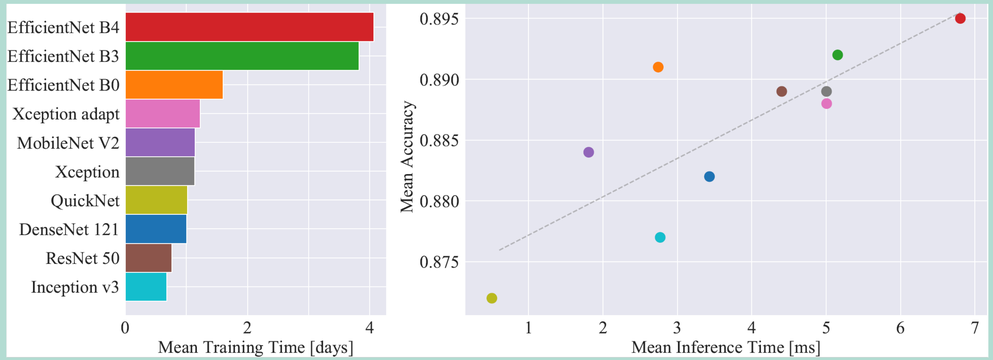
The dataset comprises 161 hematoxylin and eosin (HE) stained colon sections from the Institute of Pathology at the University Hospital Erlangen. The training database comprises 2,173,515 patches from 92 slides. The validation set contains 719,010 patches from a disjoint set of 30 slides. The remaining 39 glass slides were additionally digitized with four other automated scanners and with a manual microscope using the real-time stitching software iSTIX.

Results

Most of classifiers achieved recognition rate over 90% on five out of six datasets, with the exception of iSTIX dataset. Better performance showed more complex classifiers EfficientNet B3/B4, Xception, Xception adapt, ResNet 50.



Average classification accuracy on the different scanner test datasets. All models were trained on the 3DHISTECH MIDI scanner (Original).



Mean training time until early stopping (left). Mean classification accuracy over all datasets plotted against the mean inference time per image patch (right).

IN COOPERATION WITH

Universitätsklinikum Erlangen

ADA LOVELACE CENTER

Acknowledgements

This work was supported by the Bavarian Ministry of Economic Affairs, Regional Development and Energy through the **Center for Analytics – Data – Applications (ADA-Center)** within the framework of „BAYERN DIGITAL II“ (20-3410-2-9-8).

This work is partly funded by the **Federal Ministry of Education and Research** under the project reference numbers 16FMD01K, 16FMD02 and 16FMD03.

Fraunhofer IIS