Supplementary Materials: A whole slide image managing library based on fastai for deep learning in the context of histopathology: Two use-cases explained

Christoph Neuner, Roland Coras, Ingmar Blümcke, Alexander Popp, Sven-Martin Schlaffer, Andre Wirries, Buchfelder Michael, Samir Jabari

# 

|  |  |  |
| --- | --- | --- |
| Dataset | gonadotroph | corticotroph |
| Patients | 229 | 181 |
| All WSI1 | 644 | 405 |
| H&E WSI | 229 | 202 |
| ROIs | 7,796 | 5,373 |
| Tiles | 206,517 | 63,893 |

1 H&E and immunostained

SUPPLEMENT 1 | Dataset First, we selected 410 patients and identified one case of pituitary adenoma per patient. Then we scanned H&E-stained slides with their positive immunostained slides respectively. After that, ROIs with matching hormone expression in the immunostained slides were extracted from the H&E slides manually. After that the ROIs were split into smaller 1024x1024 pixel tiles via a self-written algorithm.

|  |  |  |
| --- | --- | --- |
| Class distribution | cases | tiles |
| ACTH | 44.7% | 23.6% |
| silent ACTH | 9.5% | 9.7% |
| LH | 49.0% | 59.3% |
| FSH | 44.1% | 62.3% |
| LH or FSH | 55.3% | 76.4% |

SUPPLEMENT 2 | Class distribution E.g. 23.6% means, that 23.6% of all tiles had the label ACTH. It was a multilabel classification problem. Only cases, where in the end tiles had been extracted from, were used for that calculation.

(0): AdaptiveConcatPool2d(

(ap): AdaptiveAvgPool2d(output\_size=1)

(mp): AdaptiveMaxPool2d(output\_size=1)

)

(1): Flatten()

(2): BatchNorm1d(4096, eps=1e-05, momentum=0.1, affine=True, track\_running\_stats=True)

(3): Dropout(p=0.25, inplace=False)

(4): Linear(in\_features=4096, out\_features=512, bias=False)

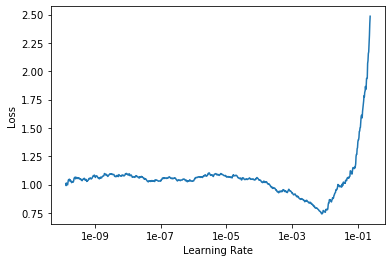
(5): ReLU(inplace=True)

(6): BatchNorm1d(512, eps=1e-05, momentum=0.1, affine=True, track\_running\_stats=True)

(7): Dropout(p=0.5, inplace=False)

(8): Linear(in\_features=512, out\_features=4/2, bias=False)

SUPPLEMENT 3 | Custom head (PyTorch). We used ResNext101/resnet50 as our final network architecture with weights pretrained on ImageNet. ImageNet has 1000 different classes. We only had 4/2. So, we had to change the head of the models, to only have 4/2 output probabilities. This is done conveniently by the fastai library. (4 outputs for the pituitary gland adenoma classifier and 2 outputs for the DNET-ganglioglioma classifier. The rest of the head was the same.)



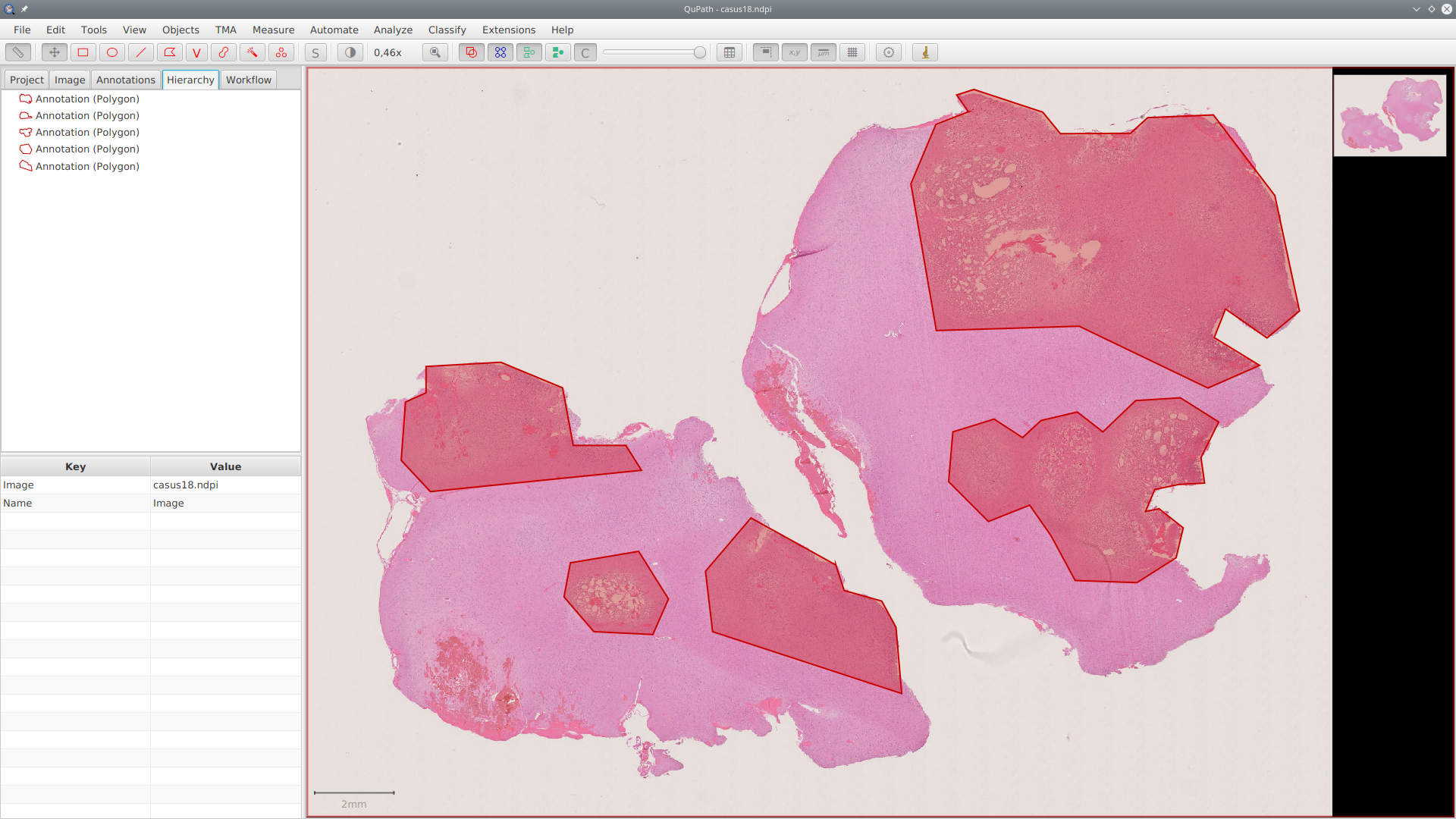
SUPPLEMENT 4 | Learning rate finder pituitary adenoma classifier. Ideally, a learning rate at the steepest part of the curve with a descending slope should be picked.Left: For training the head a maximum learning rate of 10-3 was picked.Right: For training the body maximum differential learning rates between 10-9 and 10-6 were picked.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Networks | ACTH | Silent ACTH | LH | FSH |
| ResNet50 | 0.91/0.83 | 0.85/0.74 | 0.84/0.84 | 0.84/0.87 |
| ResNet101 | 0.90/0.85 | 0.83/0.76 | 0.82/0.85 | 0.83/0.85 |
| ResNet152 | 0.89/0.84 | 0.84/0.75 | 0.86/0.82 | 0.85/0.86 |
| DenseNet121 | 0.87/0.82 | 0.83/0.71 | 0.81/0.80 | 0.82/0.83 |
| Xception | 0.85/0.87 | 0.95/0.81 | 0.85/0.90 | 0.75/0.81 |
| Inceptionv4 | 0.93/0.90 | 0.96/0.81 | 0.93/0.94 | 0.78/0.84 |
| se\_ResNext101\_32x4d | 0.96/0.87 | 0.93/0.81 | 0.95/0.94 | 0.89/0.87 |
| ResNext101\_32x8d | 0.93/0.96 | 0.96/0.96 | 0.91/0.90 | 0.87/0.90 |

SUPPLEMENT 5 | Evaluated Networks. We specified a validation and a test set. Then we trained this spectrum of network architectures and tried to optimize the accuracy on a case basis for the validation set. The final trained models were then evaluated on the test set. This table shows validation/test accuracy for the four classes. Our final choice fell on ResNext101, which was then trained and evaluated using 5-fold cross validation (see Supplement 6).

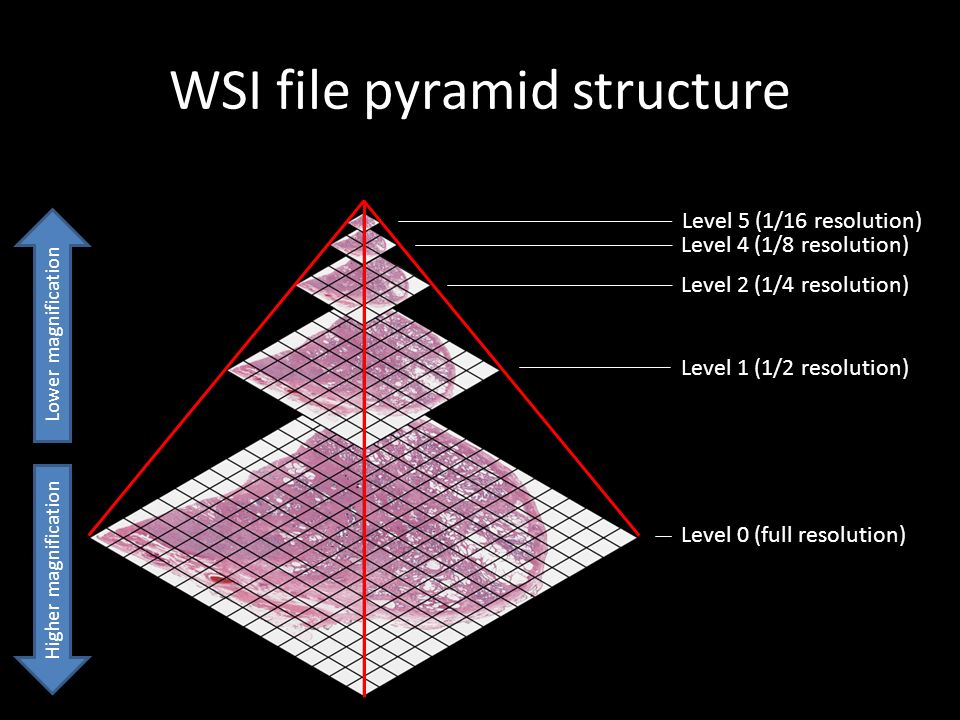
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | ACTH | Silent ACTH | LH | FSH | LH or FSH |
| Validation set 1 | 0.94 | 0.97 | 0.98 | 0.95 | 0.98 |
| Validation set 2 | 0.99 | 0.90 | 0.95 | 0.93 | 0.98 |
| Validation set 3 | 0.97 | 0.89 | 1.00 | 0.94 | 1.00 |
| Validation set 4 | 0.94 | 0.98 | 0.98 | 0.94 | 1.00 |
| Validation set 5 | 0.99 | 0.98 | 0.97 | 0.89 | 0.99 |

SUPPLEMENT 6 | AUCs of the ROC-curves for the five validation sets of 5-fold cross-validation. We trained and evaluated ResNext101 using 5-fold cross validation. We let each model predict its validation set respectively and calculated the labels on a case basis to finally compute AUCs for each class.

**SUPPLEMENT 7 | QuPath** Defining regions of interest in QuPath

SUPPLEMENT **8 | ROIs with overlaid grids** Over each ROI a grid with all possible tile locations is drawn.

SUPPLEMENT **9 | Tissue filtering** Separate tissue from background.

SUPPLEMENT **10 | WSI file pyramid structure**