Multiomics Data Science
Research Group

DeepHisto:
first results and discussing potential collaboration in brain histopathology

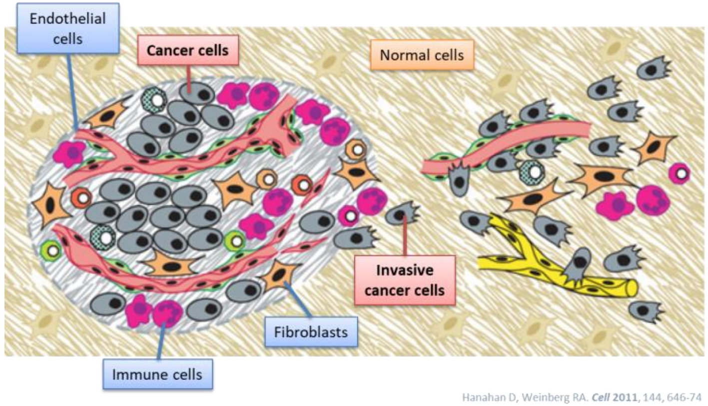
**Sang Yoon KIM, Ann-Christin HAU,
Michel MITTELBRONN, Petr NAZAROV**

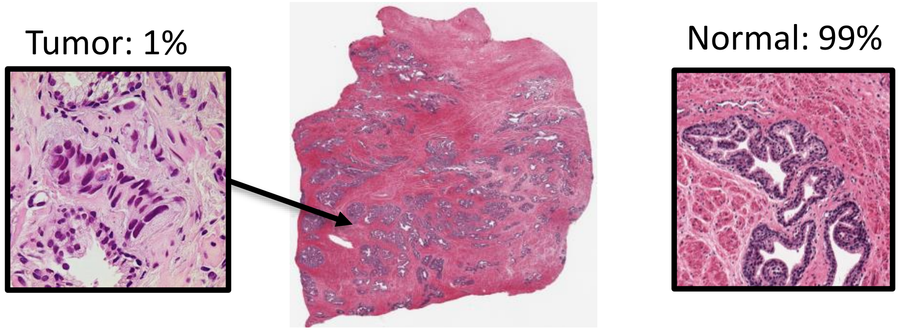
Department of Oncology, LIH
National Center of Pathology, LNS

petr.nazarov@lih.lu

Online meeting

Background

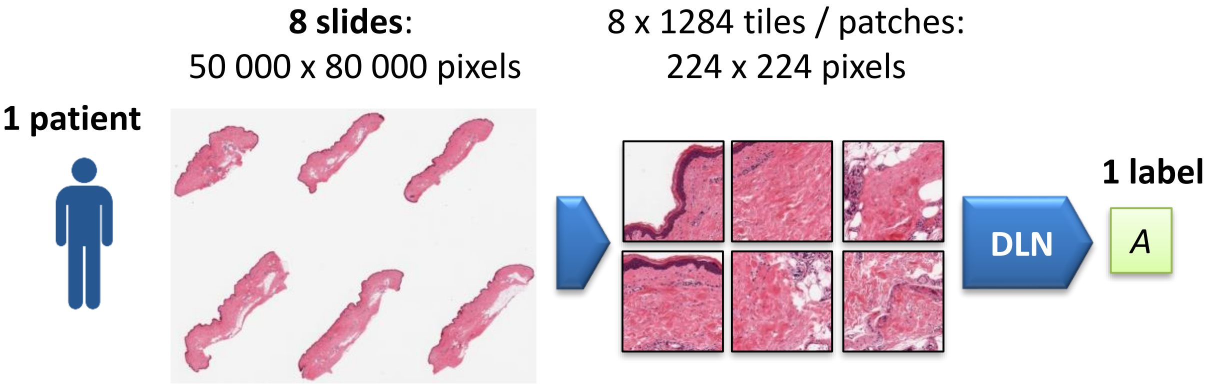




- Native heterogeneity of tissues
- Inter/intra tumor heterogeneity

Issues in histopathological image analysis:

- Tedious analysis
- In some cancers (e.g. prostate) < 1% of the image is cancer-related
- Standard approaches require supervised "pixel-wise" labelling – almost unrealistic



DLN – deep learning network (model)

Deep Learning Approach is Widely Used

LUXEMBOURG INSTITUTE OF HEALTH

Clinical-grade computational pathology using weakly supervised deep learning on whole slide images

Gabriele Campanella[1,2], Matthew G. Hanna[1], Luke Geneslaw[1], Allen Miraflor[1], Vitor Werneck Krauss Silva[1], Klaus J. Busam[1], Edi Brogi[1], Victor E. Reuter[1], David S. Klimstra[1] and Thomas J. Fuchs[1,2*]

Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images

Joel Saltz,[1,*] Rajarsi Gupta,[1,4] Le Hou,[2] Tahsin Kurc,[1] Pankaj Singh,[3] Vu Nguyen,[2] Dimitris Samaras,[2] Kenneth R. Shroyer,[4] Tianhao Zhao,[4] Rebecca Batiste,[4] John Van Arnam,[5] The Cancer Genome Atlas Research Network, Ilya Shmulevich,[6] Arvind U.K. Rao,[3,7] Alexander J. Lazar,[8] Ashish Sharma,[9] and Vésteinn Thorsson[6,10,*]

Article

AI-based pathology predicts origins for cancers of unknown primary

https://doi.org/10.1038/s41586-021-03512-4

Received: 27 June 2020

Ming Y. Lu[1,2,3], Tiffany Y. Chen[1,2,5], Drew F. K. Williamson[1,2,5], Melissa Zhao[1], Maha Shady[1,2,3,4], Jana Lipkova[1,2,3] & Faisal Mahmood[1,2,3]

A deep learning model to predict RNA-Seq expression of tumours from whole slide images

Benoît Schmauch[1], Alberto Romagnoni[1,4], Elodie Pronier[1,4], Charlie Saillard[1], Pascale Maillé[2,3], Julien Calderaro[2,3], Aurélie Kamoun[1], Meriem Sefta[1], Sylvain Toldo[1], Mikhail Zaslavskiy[1], Thomas Clozel[1], Matahi Moarii[1], Pierre Courtiol[1,5] & Gilles Wainrib[1,5]

ARTICLE

Check for update

https://doi.org/10.1038/s41467-021-21727-x

OPEN

Joint analysis of expression levels and histological images identifies genes associated with tissue morphology

Jordan T. Ash[1,4], Gregory Darnell[2,4], Daniel Munro[2,4] & Barbara E. Engelhardt[1,3]

Interesting ideas:

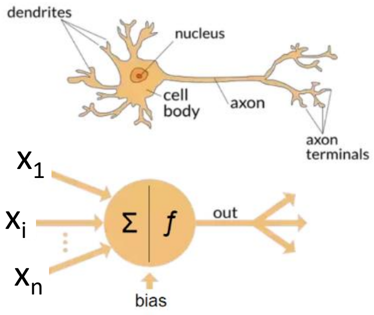
- weakly supervised learning
- attention-based learning
- *predicting molecular signatures*

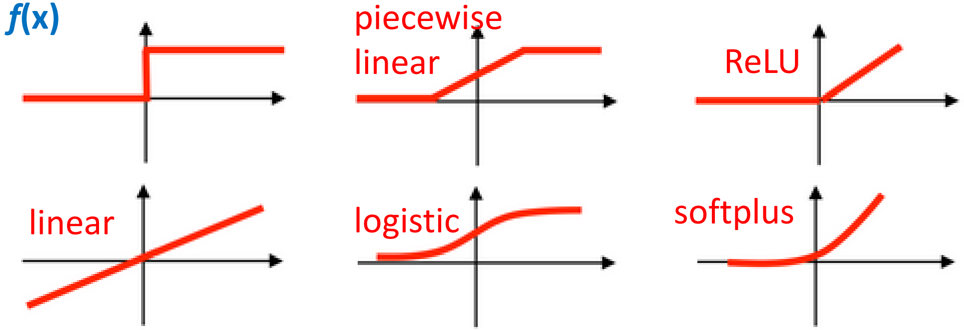
We started a small project DeepHisto in deep-learning applied to neuro-oncology

Classical Artificial Neural Networks

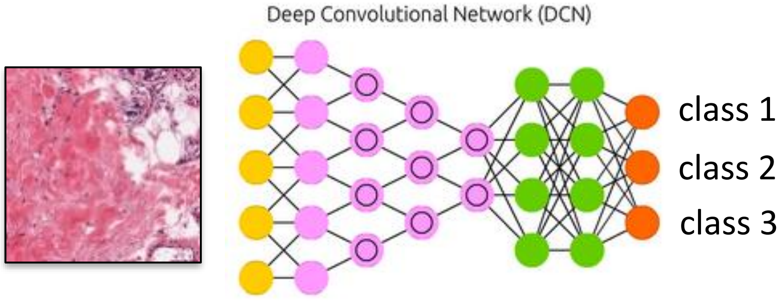
Classical artificial neuron

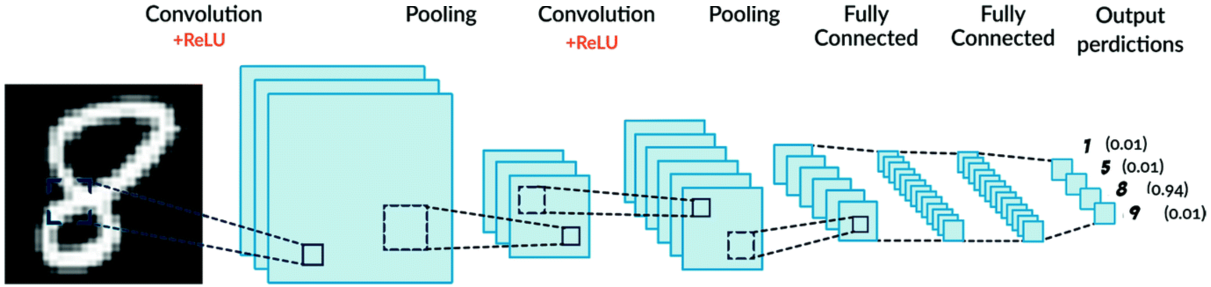
out = $\boldsymbol{f}\,(\,\Sigma w_i x_i + \text{bias})$



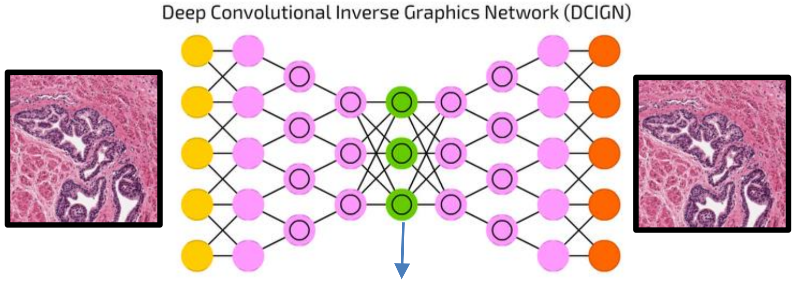


Convolutional networks





Autoencoders



Latent variables
(coded image)

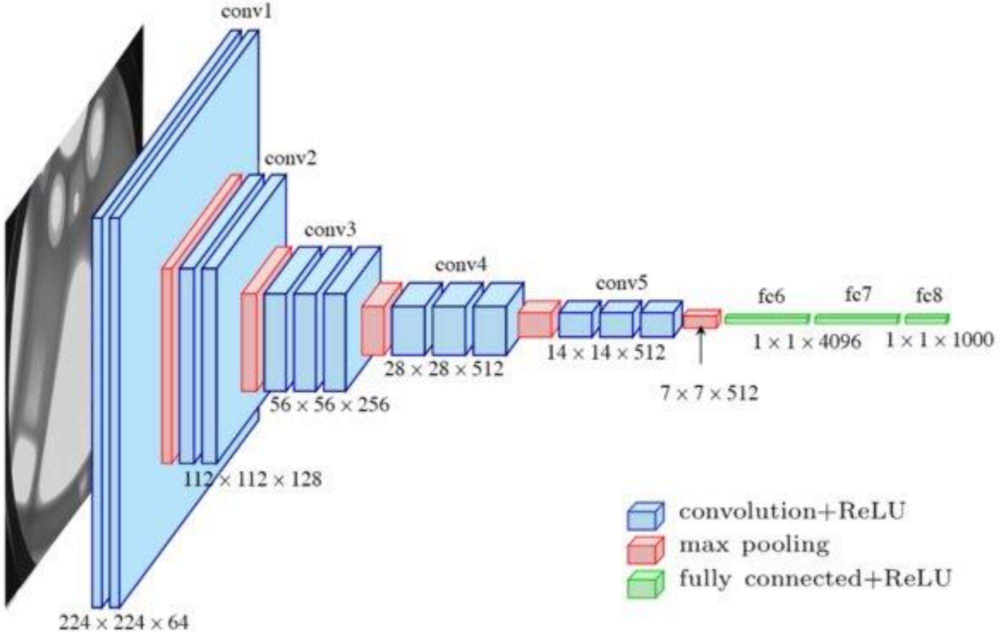
https://towardsdatascience.com/

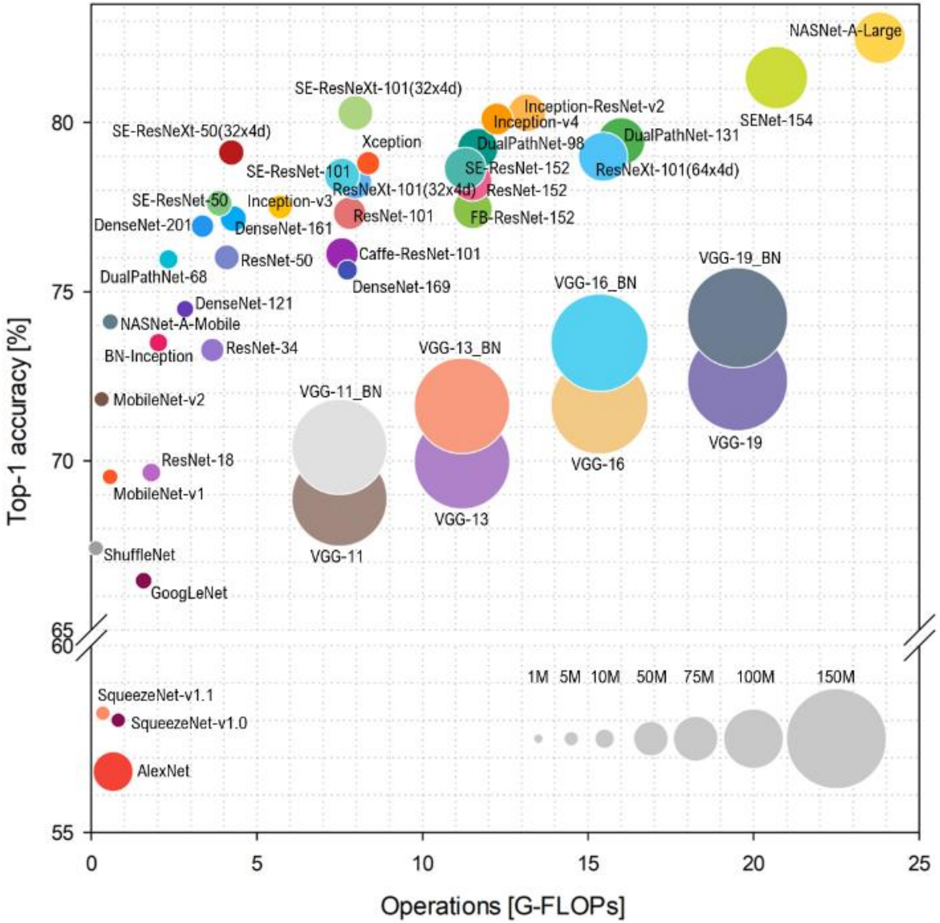
http://rasbt.github.io/mlxtend/user_guide/general_concepts/activation-functions/

Deep Learning Networks

VGG16 model architecture

https://arxiv.org/pdf/1810.00736.pdf





DeepHisto: Initial Settings

Task

Get HE-stained slides, separate them into tiles and classify the tiles into 5 categories:

- **glioblastoma** (IDHwt)
- **astrocytoma** (IDHmut, non-codel)
- **oligodendroglioma** (IDHmut, codel)
- **necrosis**
- **normal tissue** (grey and white matters)

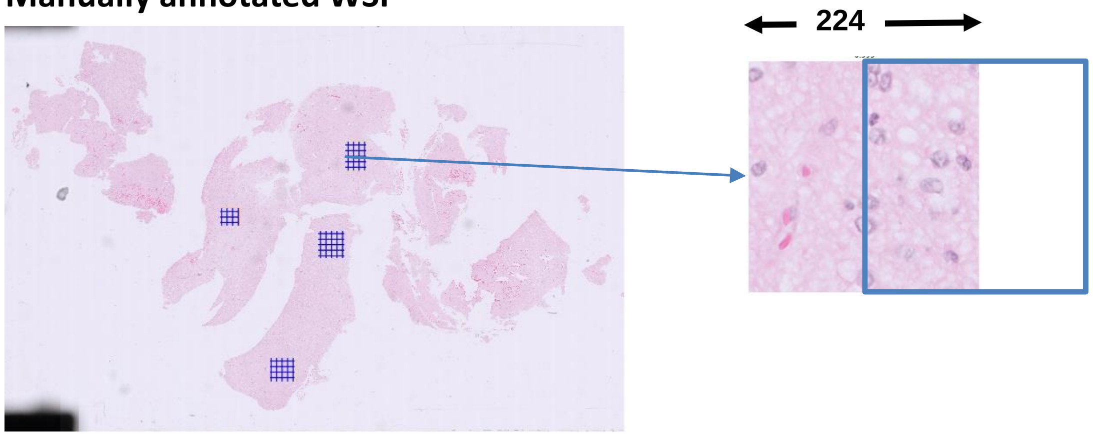
Cohort

58 slides, with labelled areas of "representative" pattern

- GBM: 20 WSIs (slides)
- Astro: 16 WSIs
- Oligo: 22 WSIs

In each group 2 WSI were preserved for testing

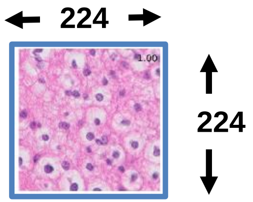
Manually annotated WSI



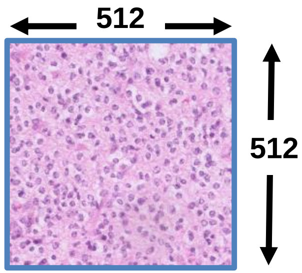
Overview of WSI Analysis

LUXEMBOURG INSTITUTE OF HEALTH

Version 1:

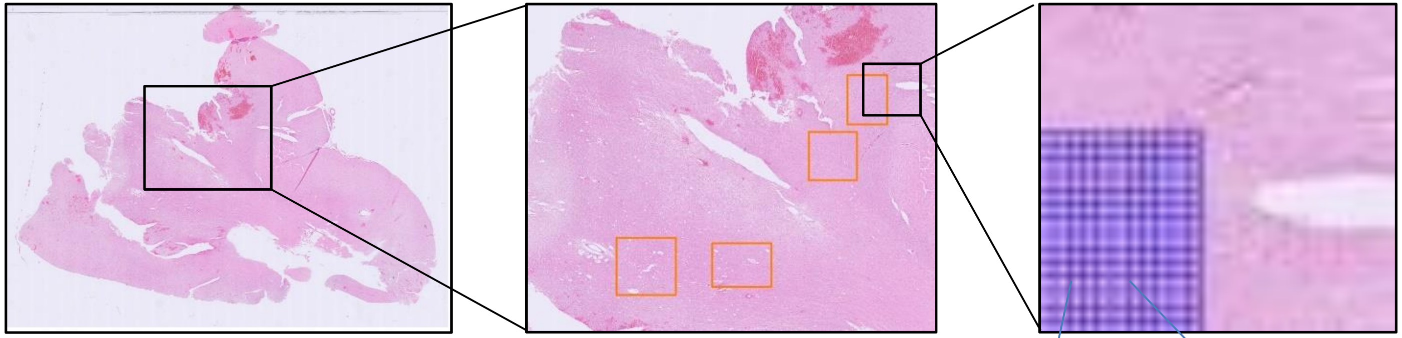


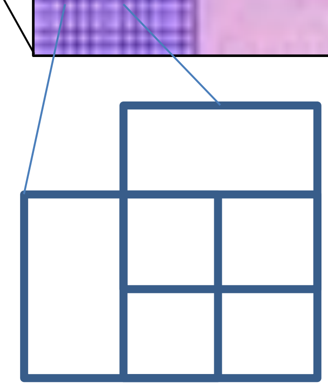
Version 2:



manual annotation

WSI patching





- **Training** on tiles from annotated areas
- "**Heatmap**" generation on entire WSI

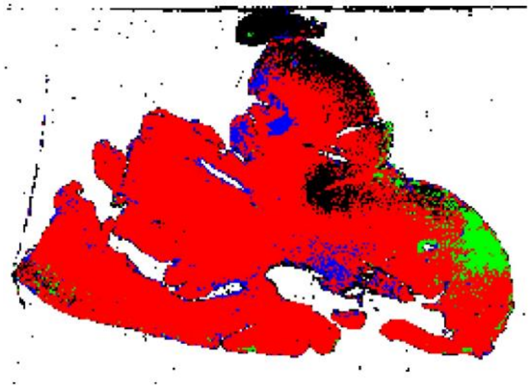
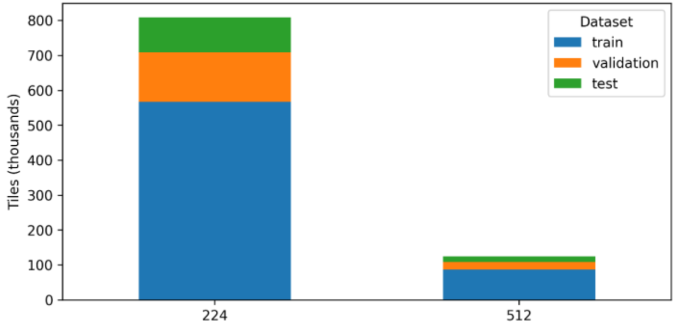


Image Augmentation

- Rotation and flip
- 50% overlap with the adjacent tiles
- 500K , 100K tiles for training and testing respectively

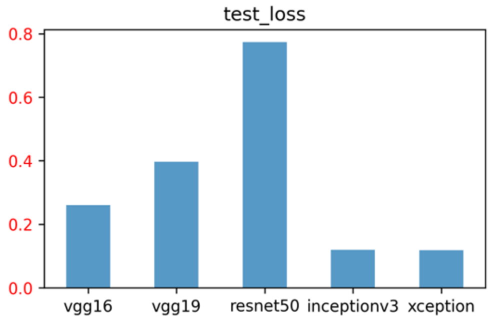
Training

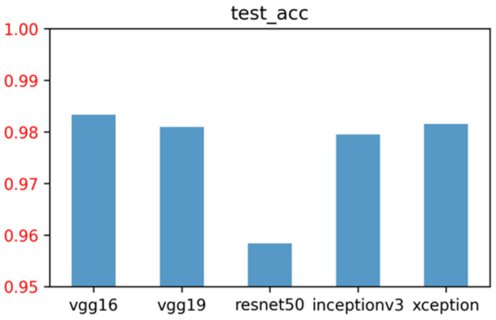


Larger tiles reduce the sample number, so the data sets are different to each other, so we are not able to compared these results directly

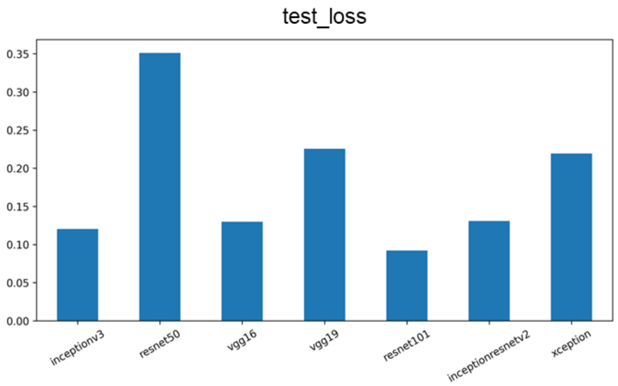
For both versions (224x224 and 512x512) **VGG16 performed very well!**

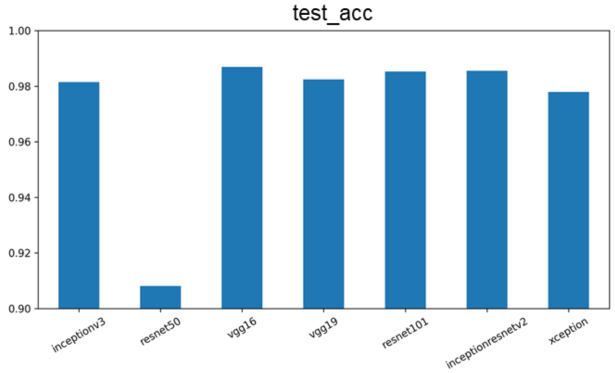
224x224





512x512





Results for VGG16 (test tiles)

We started with the "classical" pre-trained **VGG16**. Results were not bad compared to more advanced models, with top accuracy for 2 versions 0.984 <-> 0.986

N=100K **224x224**

Predictions

0.984	GBM	Astro	Oligo	Necro	Normal
GBM	6733	2	48	57	0
Astro	0	17160	0	0	0
Oligo	94	140	47847	2	293
Necro	17	0	1	4782	0
Normal	2	25	963	1	22265

Labels

N=15K **512x512**

Predictions

0.986	GBM	Astro	Oligo	Necro	Normal
GBM	1028	2	0	2	0
Astro	0	2752	0	0	0
Oligo	30	0	7658	2	54
Necro	0	0	0	624	0
Normal	0	0	121	1	3255

Labels

Balanced:

Predictions

0.985	GBM	Astro	Oligo	Necro	Normal
GBM	0.984	0.000	0.007	0.008	0.000
Astro	0.000	1.000	0.000	0.000	0.000
Oligo	0.002	0.003	0.989	0.000	0.006
Necro	0.004	0.000	0.000	0.996	0.000
Normal	0.000	0.001	0.041	0.000	0.957

Labels

Balanced:

Predictions

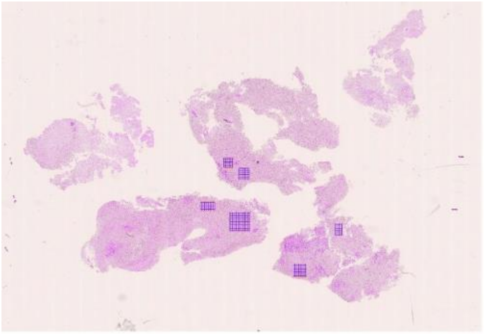
0.990	GBM	Astro	Oligo	Necro	Normal
GBM	0.996	0.002	0.000	0.002	0.000
Astro	0.000	1.000	0.000	0.000	0.000
Oligo	0.004	0.000	0.989	0.000	0.007
Necro	0.000	0.000	0.000	1.000	0.000
Normal	0.000	0.000	0.036	0.000	0.964

Labels

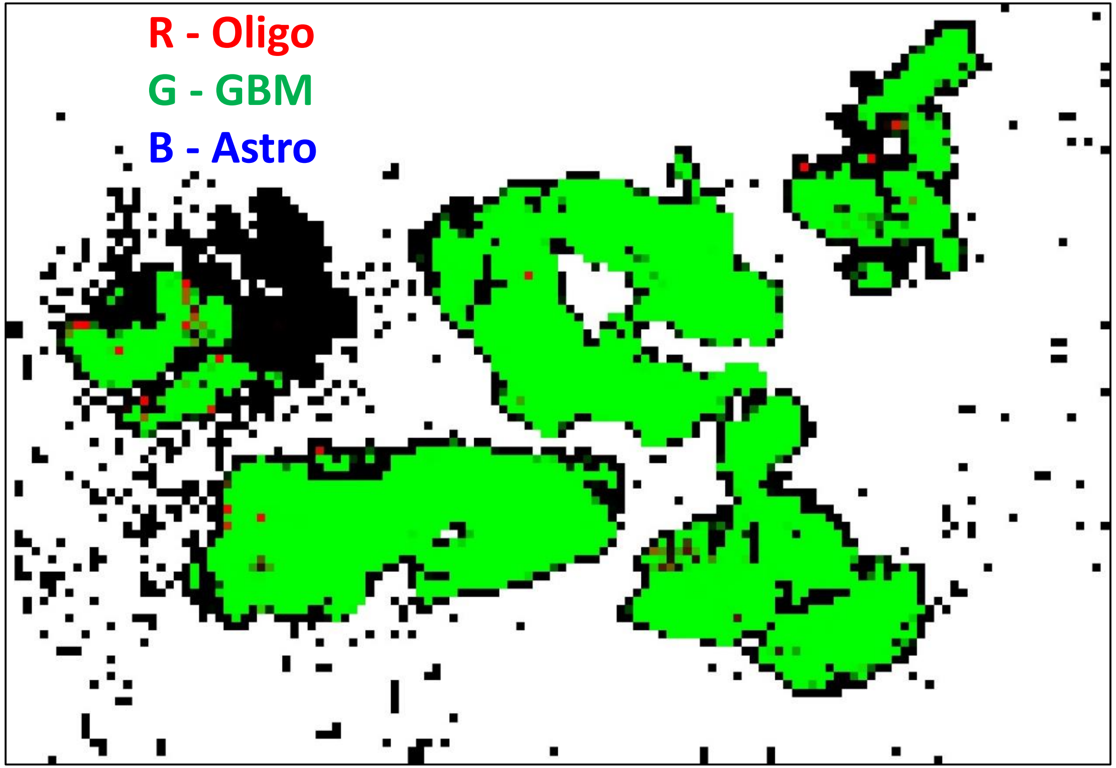
Heatmap: GBM case

This example: a training-set slide from an **glioblastoma** patient

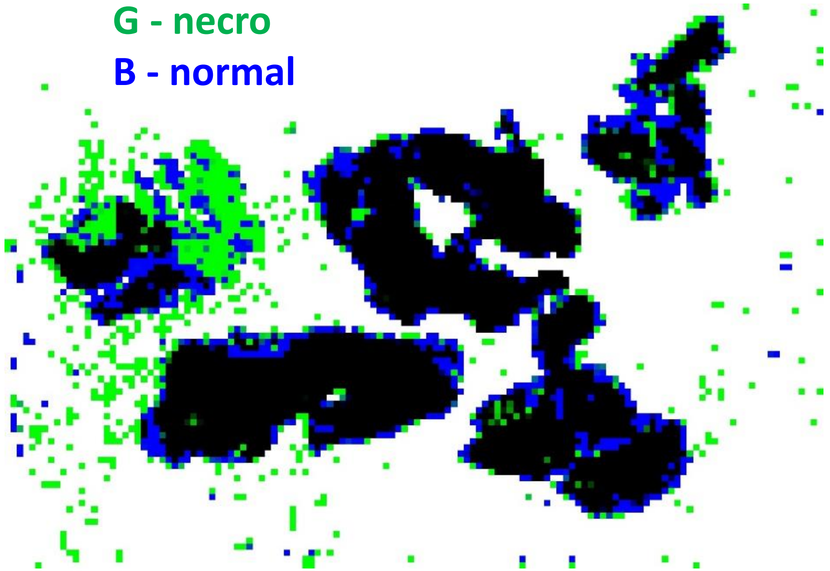
GBM



Tumor classes highlighted



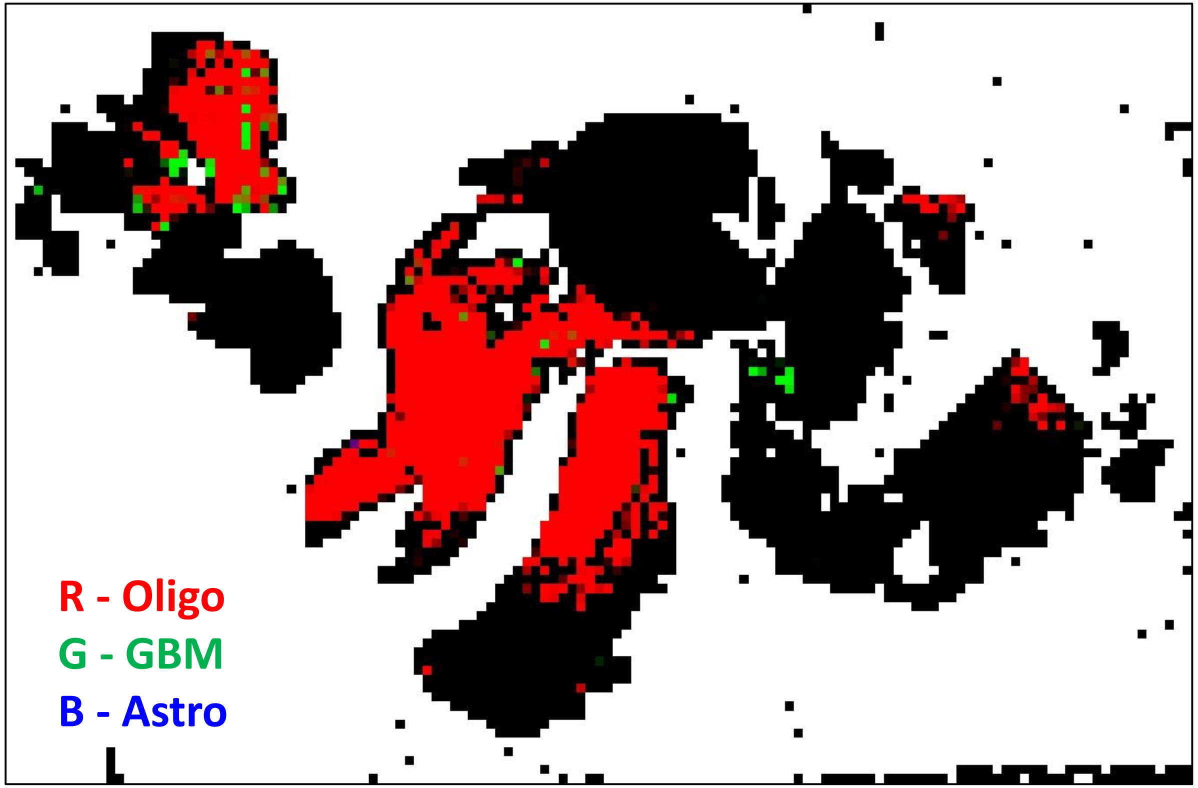
Non-tumor classes



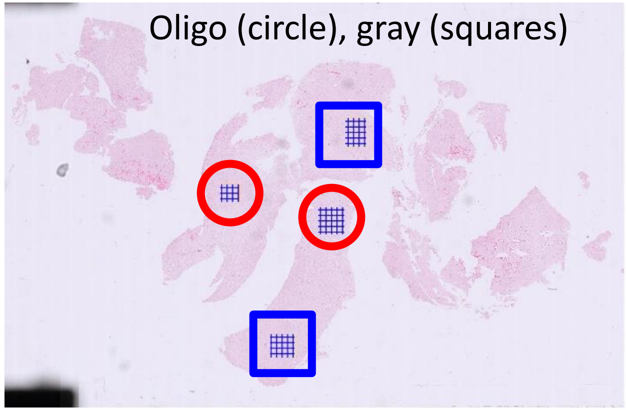
Heatmap: test Oligo case

This example: a test-set slide from an **oligodendroglioma** patient (never seen by the network)

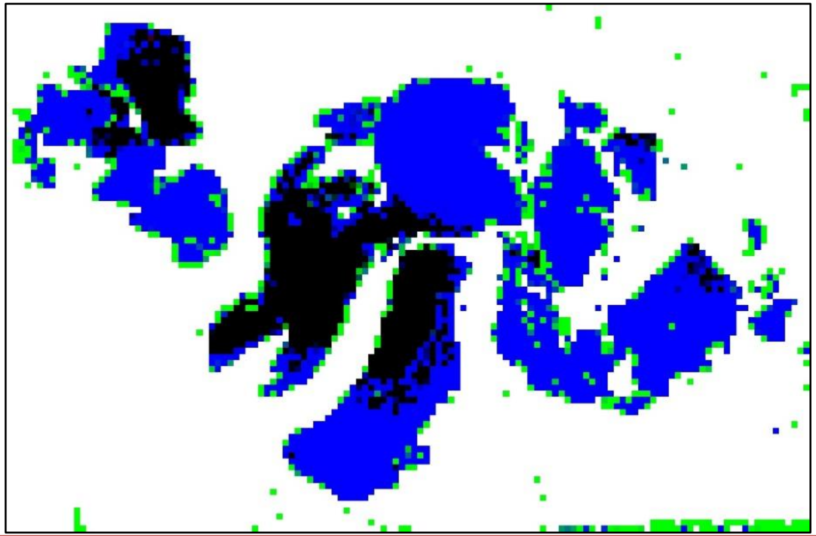
Tumor classes highlighted



R - Oligo
G - GBM
B - Astro



Non-tumor



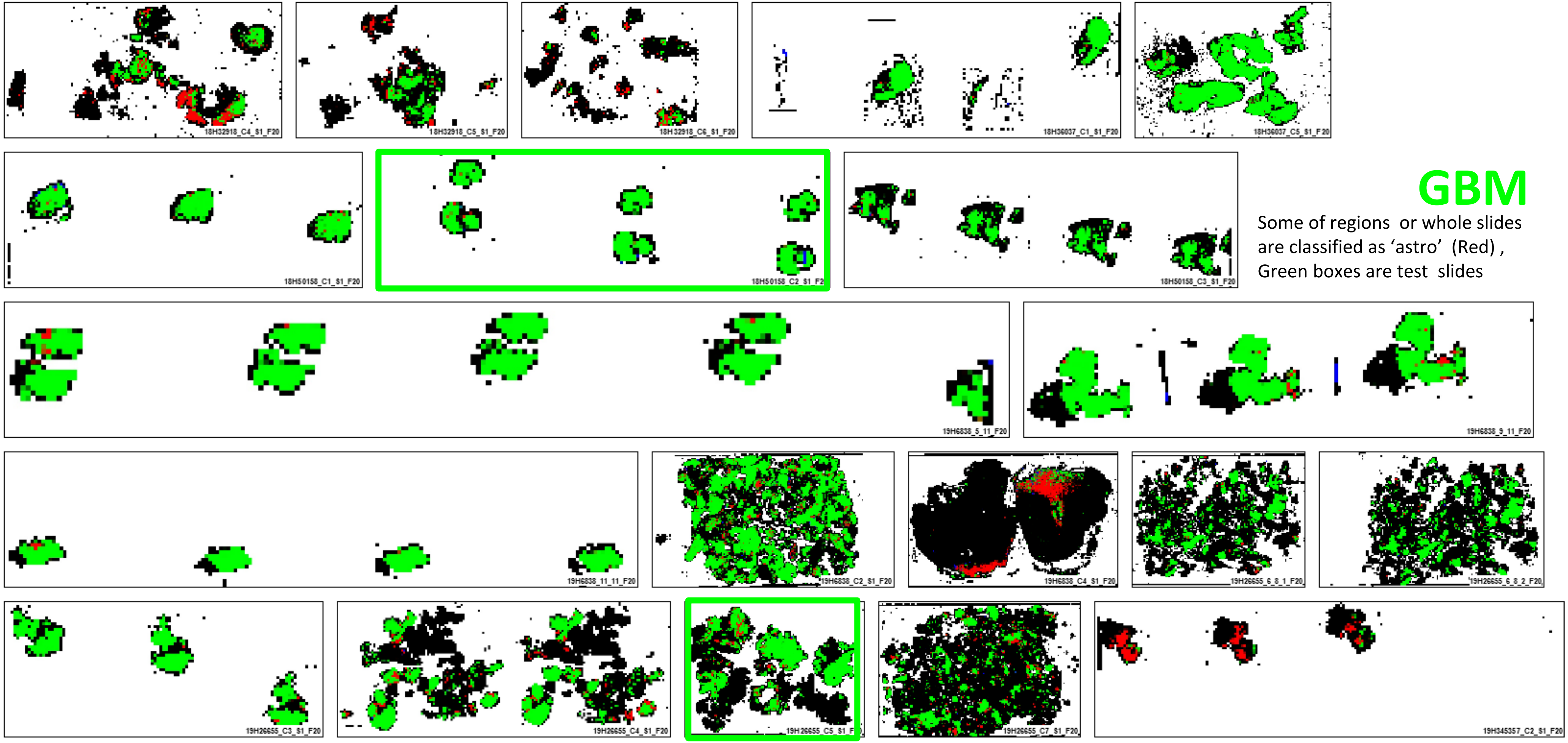
Astro | GBM | Oligo

GBM

Some of regions or whole slides are classified as 'astro' (Red) , Green boxes are test slides

Astro | GBM | Oligo

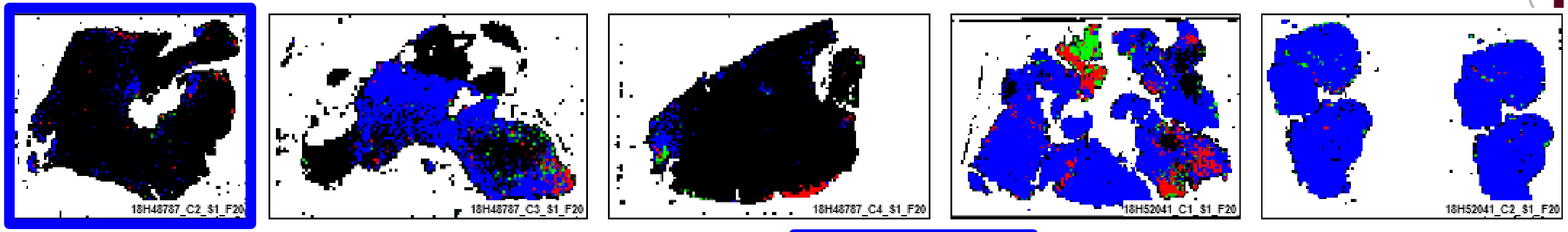
LUXEMBOURG INSTITUTE OF HEALTH

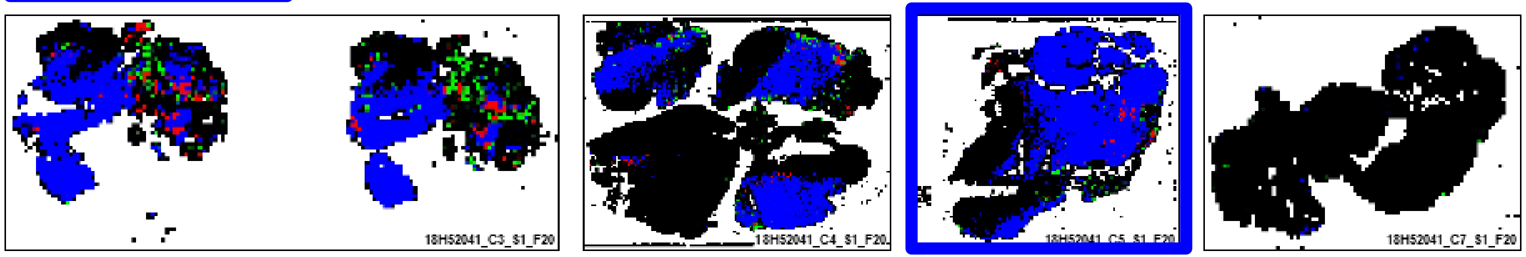


GBM

Some of regions or whole slides are classified as 'astro' (Red) ,
Green boxes are test slides

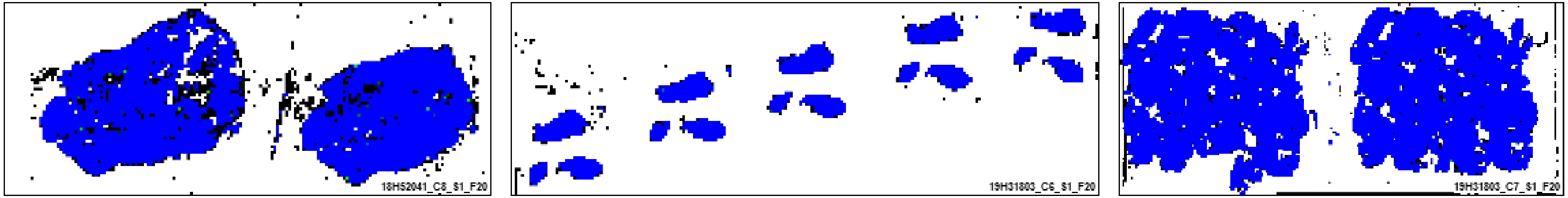
Astro | GBM | Oligo

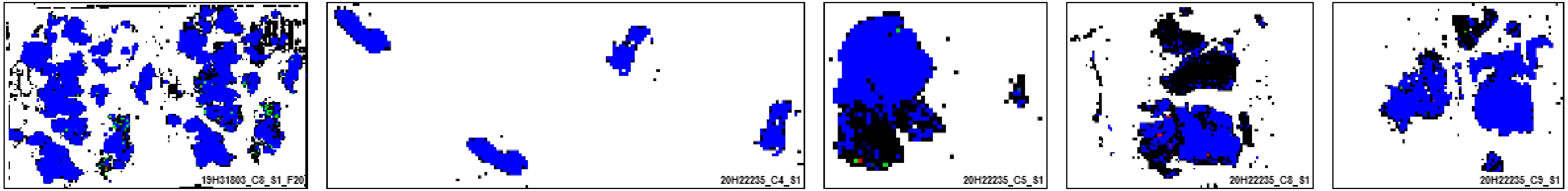


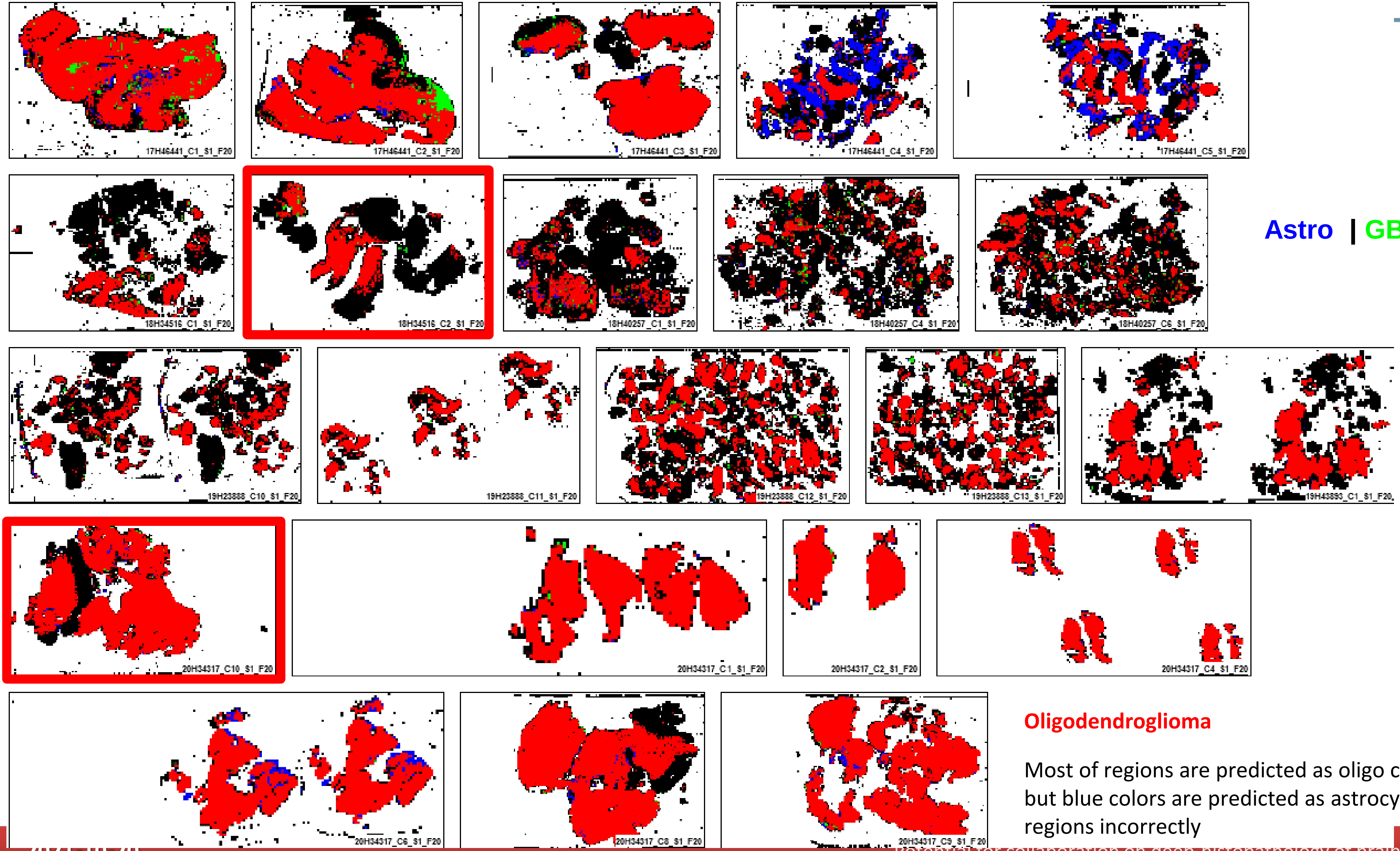


Astrocytoma

Black regions are mainly predicted as gray matter





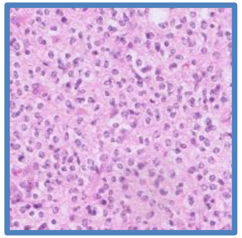


Astro | GBM | Oligo

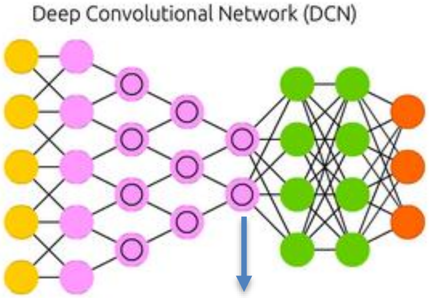
Oligodendroglioma

Most of regions are predicted as oligo correctly , but blue colors are predicted as astrocytoma regions incorrectly

DLN for Feature Extraction



512 x 512
= **262 144**
pixels

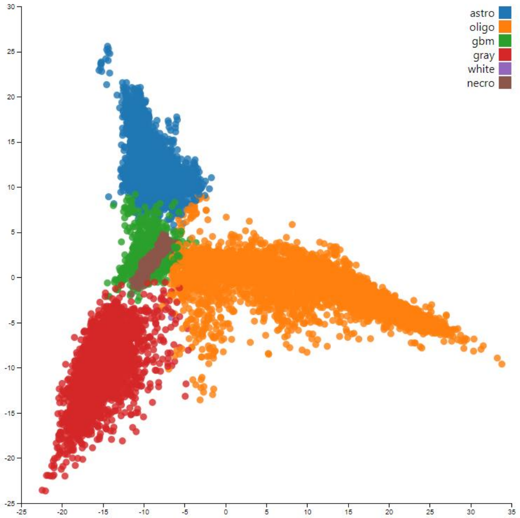


512 "latent" features

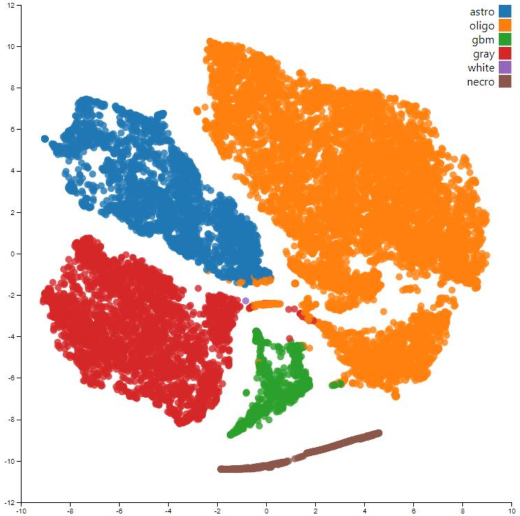
These plots are also available online:
https://lih-biomod.lu/deep/512/

⇒ features are extracted correctly!
⇒ classifier is not optimal!

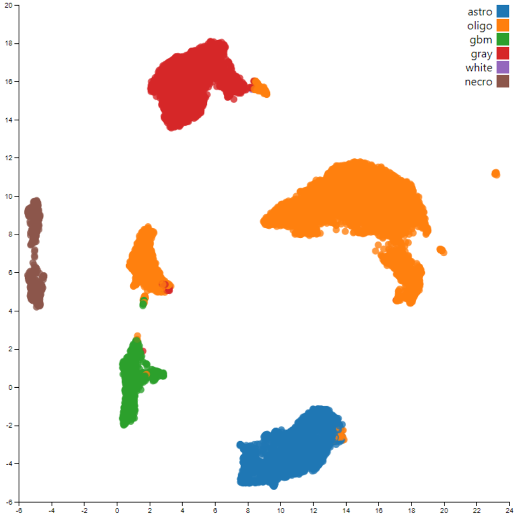
PCA



t-SNE

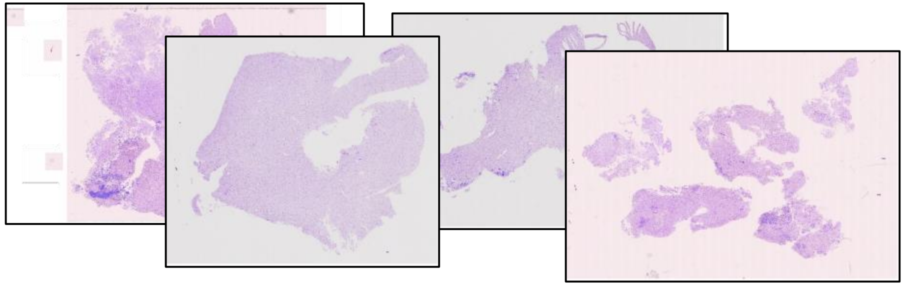


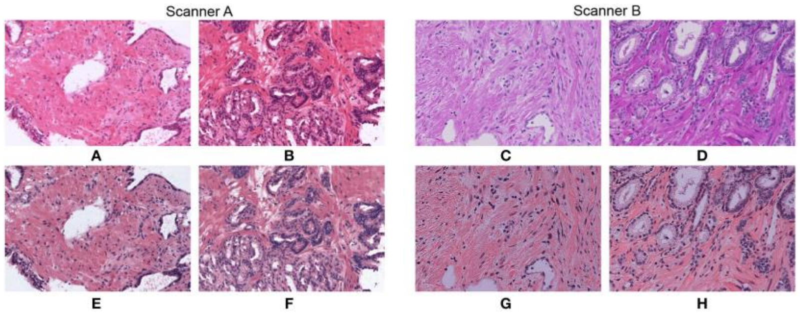
UMAP

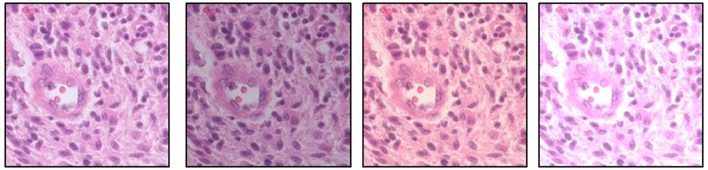


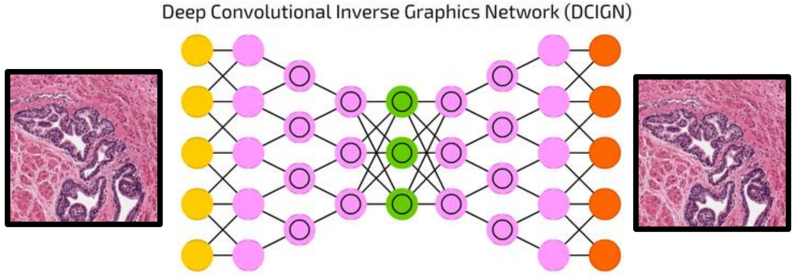
Further Improvements: Classical

- get more (annotated) slides
- add tile / WSI color normalization
- add color-based data augmentation
- replace VGG16 by an autoencoder and improve classification based on the latent features





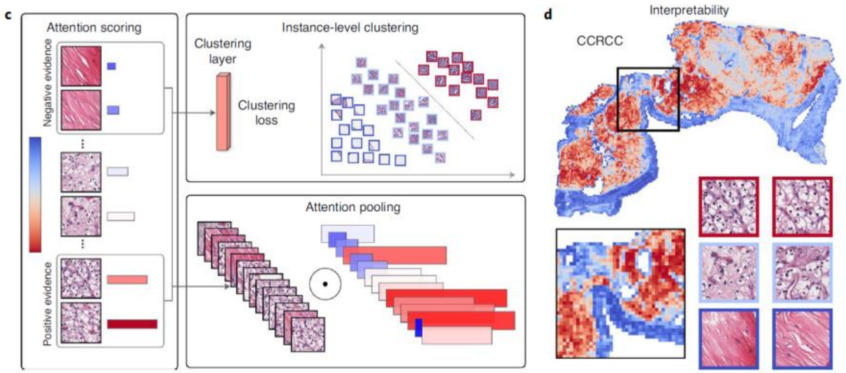


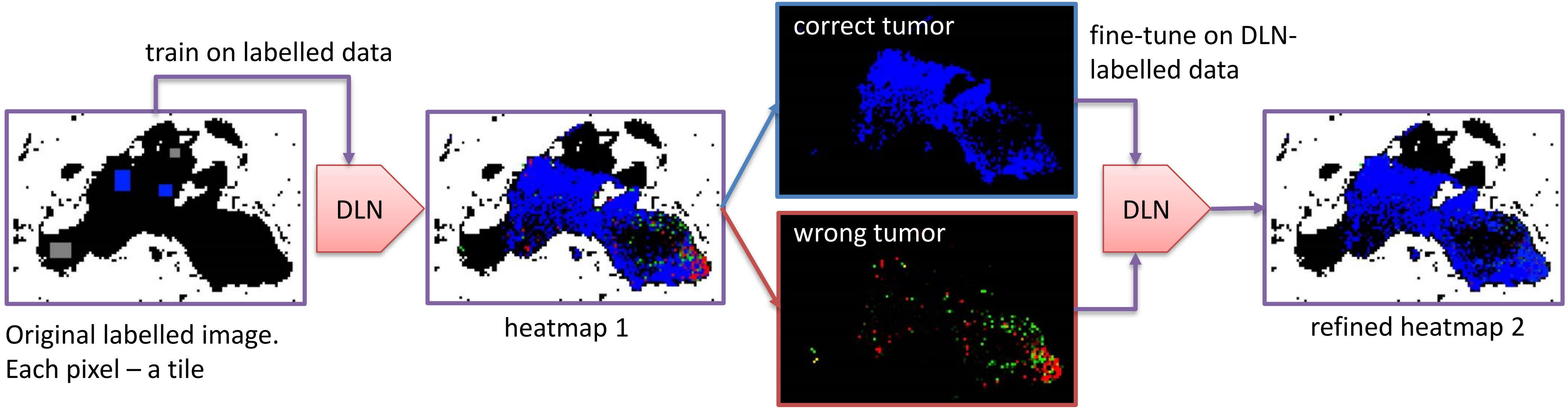


works on unlabeled data!

Further Improvements: Modern

- weakly supervised (multiple instance learning)
- attention-based training
- use 2-cascade training





Can we get more slides?

Shall we try different scanners and thickness?

Discussion

Is it interesting?

What was developed in your groups?

Anything we can contribute?

Can we incorporate more brain tumors?