A deep learning-based model of normal histology

Tobias Sing[1]*, Imtiaz Hossain[1]*, Holger Hoefling[1]*, Arno Doelemeyer[1], Chandrassegar Saravanan[2], Alessandro Piaia[1], Kuno Wuersch[1], Vincent Romanet[1], Wolfgang Zipfel[1], Mike Steeves[2], Oliver Turner[3]*, Pierre Moulin[1]*

[1]Novartis Institutes for BioMedical Research, Basel, Switzerland; [2]Novartis Institutes for BioMedical Research, Cambridge, MA, USA; [3]Novartis Institutes for BioMedical Research, East Hanover, NJ, USA

*These authors contributed equally

Background

Several deep learning models, trained on pathologies in various tissues, have been reported in the literature. However, these models are all specific to a certain tissue/context. A layer describing the diversity of *normal* tissues could be an ingredient of a more generic approach.

Objective

To establish a model of normal, non-diseased tissue, as a foundation for future models of histopathology

Methods

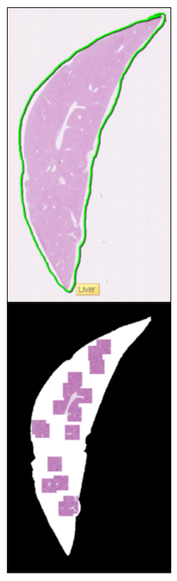
Histological slides; tissue taxonomies

On 449 slides scanned at 40x (0.25 microns/pixel), representing various tissues from normal rats (*Rattus norvegicus*), tissue regions were outlined and annotated by pathologists according to two taxonomies of different detail ("fine", "coarse").

Adipose tissue, brown (BF)	Adipose tissue, white (WF)	Lung (LU)	Slide Background (BKG)	Tongue (TO)	Urinary bladder (UB)	Large intestine (LA)	Small intestine (SI)	Stomach (ST)	Lymph node (LN)	Spleen (SP)	Thymus (TH)	Heart (HT)	Muscle, skeletal (SM)	Gland, adrenal (AG)	Gland, parathyroid (PTG)	Gland, thyroid (TYD)	Liver (LI)	Gland, harderian (HG)	Gland, mammary (MG)	Gland, salivary (SG)	Pancreas (PAN)	Kidney (KD)	Skin (SKN)
Alveolar (ALV)			BKG	Com plex (CPX)	Hollow (HOO)				Lymphoid (LYM)			Musculo-skeletal (MST)		Parenchyma (PAR)									Tegum entary (TEM)

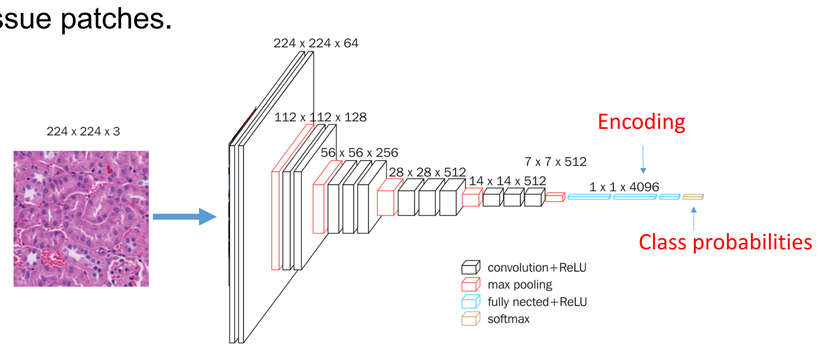
Generation of training and validation datasets

With typical resolutions of 120,000 x 80,000 pixels and file sizes of ~250 megabytes, histological slides scanned at high resolution are orders of magnitude larger than images commonly used for the development and benchmarking of novel machine learning methods. Therefore, popular approaches such as semantic segmentation of entire images are not directly applicable to histological images. Instead, we develop tissue type classifiers from small image patches (224 x 224 pixels), which are generated at various levels of resolution down-sampling (2x, 4x, 8x, 16x down-sampling). A total of ~800,000 patches were generated for this study. To avoid contaminating the validation set, patches from any given animal were always allocated either all to the training set, or all to the validation set.



Training of patch classifiers

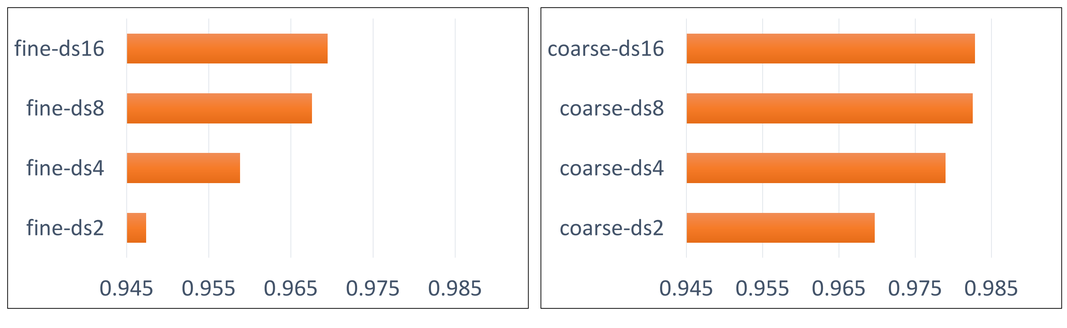
For each combination of tissue taxonomy and down-sampling level, a VGG-16 neural network [1] was trained using a transfer learning approach, with weights pre-initialized via the ImageNet dataset, followed by fine-tuning on our training dataset of tissue patches.



The final fully connected layer of the VGG-16 network constitutes a learned representation (encoding) of a given image patch. Patterns formed by the entirety of image patches in the encoding space were analyzed using the t-SNE dimensionality reduction method.

Results

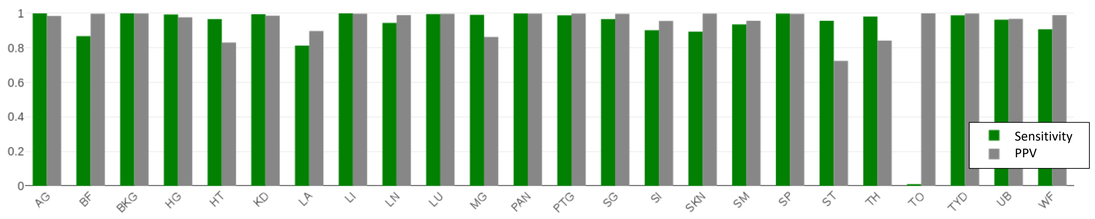
Model accuracy ranged from 94.7% to 98.3% on the validation datasets. Increasing accuracy was observed with increasing down-sampling level (2x, 4x, 8x, 16x down-sampling) and less detailed taxonomy ("fine" vs. "coarse").



Results (cont.)

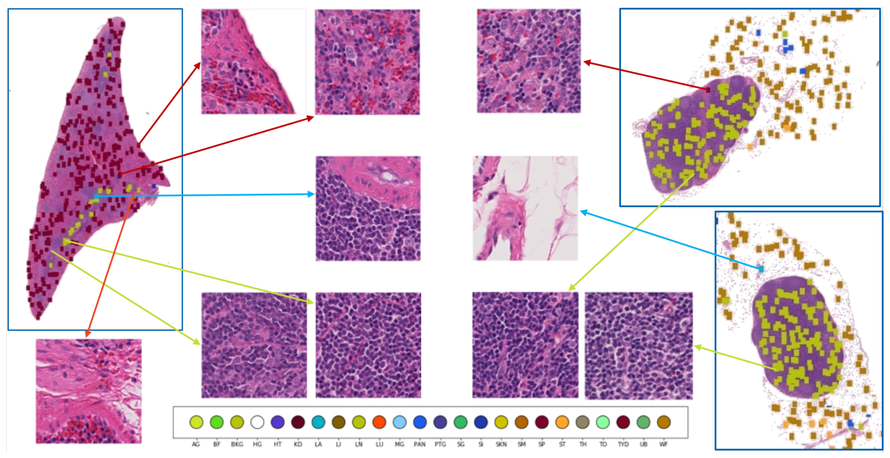
Class-specific sensitivity & positive predictive value (PPV)

Most tissue types can be predicted at high sensitivity and PPV. The example below shows tissue-specific sensitivity and PPV for the model: ["fine" taxonomy, 8x down-sampling]. Tongue is a difficult tissue on all four down-sampling levels because it appears as a mix of tissues. The characteristics of tongue tissue will likely become apparent only at even higher down-sampling levels.



Assessment of misclassified patches by expert pathologists

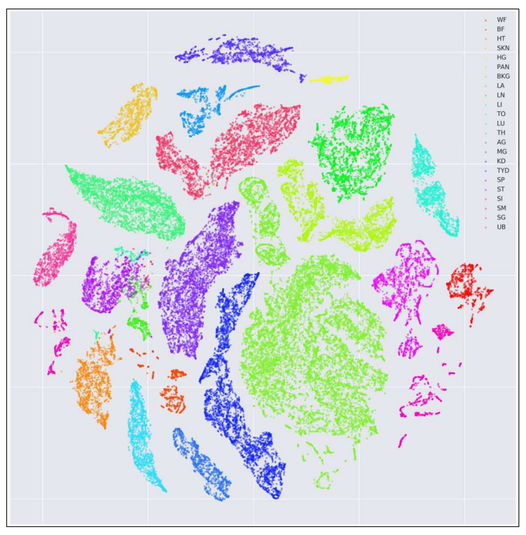
Unequivocal tissue identification of misclassified patches was often impossible for expert pathologists as well, because some patches may contain structures that are common to several tissues. This observation is demonstrated by the following example comparing patches from spleen (left) and lymph node (right):



- Bottom row: Spleen patches "misclassified" as lymph node (left). Comparison with patches from a lymph node (right) shows that these spleen patches consist entirely of lymphoid tissue (as a component of spleen).
- Middle row: Spleen and lymph node patches "misclassified" as mammary gland (MG). Both patches contain artery parts, which are abundant in MG.
- Top row: A lymph node patch "misclassified" as spleen. The patch contains a considerable number of macrophages with red blood cells, which make it look like part of spleen.

t-SNE visualization of the encoding space

The t-SNE visualization of the encoding space (here colored by the true tissue type) revealed pronounced, mostly non-overlapping clusters corresponding to individual tissue classes.



Conclusions

Accurate tissue type classification from small image patches can be achieved using "off-the-shelf" neural network architectures. Segmentation of tissue regions on a whole slide will be possible at even higher accuracy, as misclassifications made on some patches can be "outvoted" by classifications on adjacent or overlapping patches at various resolution down-sampling levels.

References

[1] Simonyan & Zisserman (ICLR 2015): Very Deep Convolutional Networks for Large-Scale Image Recognition

Acknowledgments

We gratefully acknowledge the consultancy and advice of Thomas Fuchs (Memorial Sloan Kettering Cancer Center / Paige.AI) and Geert Litjens (Radboud University Medical Center, Nijmegen, Netherlands).