Addressing data scarcity with deep transfer learning and self-training in digital pathology

Romain Mormont

Montefiore Institute, University of Liège, Belgium

October 21, 2022

PhD defense

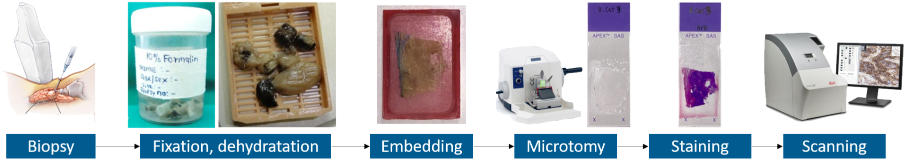
Introduction

"*Digital pathology incorporates the acquisition, management, sharing and* **interpretation** *of pathology information — including slides and data — in a digital environment*"

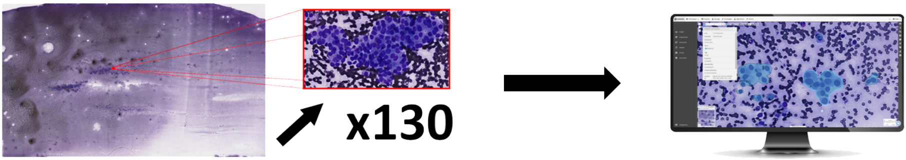
left: institutions.ville-geneve.ch, *center:* verywellhealth.com, *right:* healthcare-in-europe.com

From the body to the computer

A pathology workflow: from a biopsy...



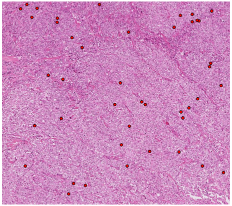
... to a whole-slide image and computer-assisted pathology.



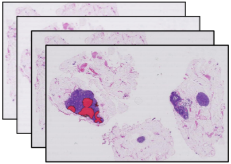
Left image: 163840 x 95744 pixels^2, 2.3 Gb.

Why computer-aided pathology

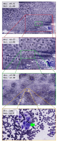
Some pathology analysis tasks are **tedious**, **time-consuming** and/or **costly**.
Computational methods has the potential of making them faster, more reliable therefore **improving costs and patient outcome**.



(a) Counting



(b) Multi-slide analysis



(c) Needle in a haystack

The field concerned with these methods is called **computational pathology**.

Computational pathology is challenging

Few of the challenges:

- high variability: content, staining, acquisition,...
- data scarcity: annotating data is expensive and tedious
- big data: up to millions of biological objects per multi-gigapixel image

Machine learning methods are excellent candidates for tackling these !

More on data scarcity

Causes:

- highly specialized annotators required
- simplification of the underlying medical problems
- privacy concerns
- deep learning methods are data hungry

Consequences:

- small datasets (compared to natural image domain)
- lack of variety on the problem targets
- weakly- or sparsely-labeled datasets

Working around data scarcity

Making more annotations within the same budget:

- AI-assisted annotation
- appropriate UI/UX tools
- citizen science/crowdsourcing

Using proper computational methods:

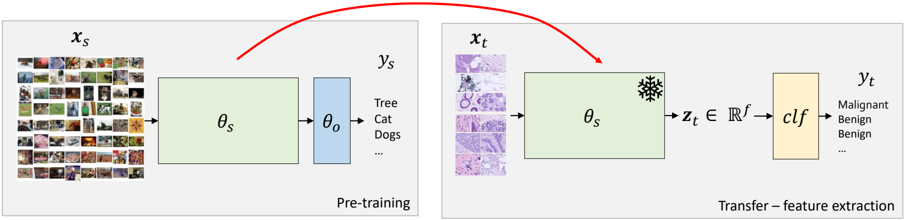
- **transfer learning**
- **self-training**
- weakly-supervised learning
- self-supervised learning

Contributions

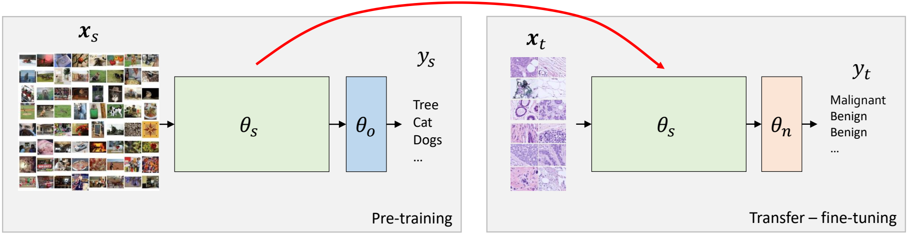
All contributions address digital patholog:

- Comparison of deep transfer learning strategies from ImageNet
- Multi-task pre-training of deep neural networks on pathology tasks
- Self-training for segmentation from sparsely-labeled data

Deep transfer learning



Feature extraction



Fine tuning

Deep transfer learning in pathology: how to ?

Goal: devising guidelines and best practices for deep transfer learning in digital pathology:

- Fine-tuning *vs.* OTS features: which one works better ?
- Which network works better ?
- Where to extract OTS features ?
- ...

Deep transfer learning: how to ?

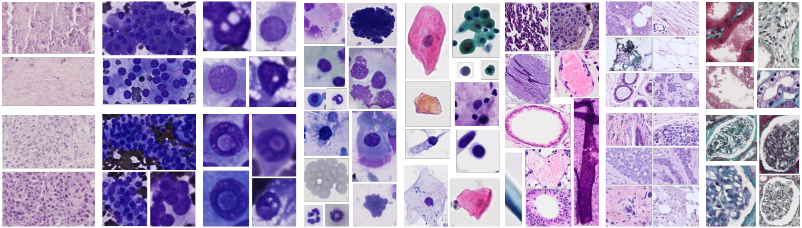
We have carried out several experiments with ImageNet as source task:

- OTS vs. fine-tuning
- Networks: ResNet50, DenseNet201, VGG16/19, InceptionResNetV2,...
- Features classifiers: SVM , extra-trees (ET),...
- OTS features extraction at increasing depth
- ...

Datasets

8 image classification datasets.

Dataset	Domain	Cls	Total	
			Images	Slides
Necrosis (N)	Histo	2	882	13
ProliferativePattern (P)	Cyto	2	1857	36
CellInclusion (C)	Cyto	2	3638	45
MouseLba (M)	Cyto	8	4284	20
HumanLba (H)	Cyto	9	5420	64
Lung (L)	Histo	10	6331	882
Breast (B)	Histo	2	23032	34
Glomeruli (G)	Histo	2	29213	205



Results

Fine-tuning is the best performing strategy

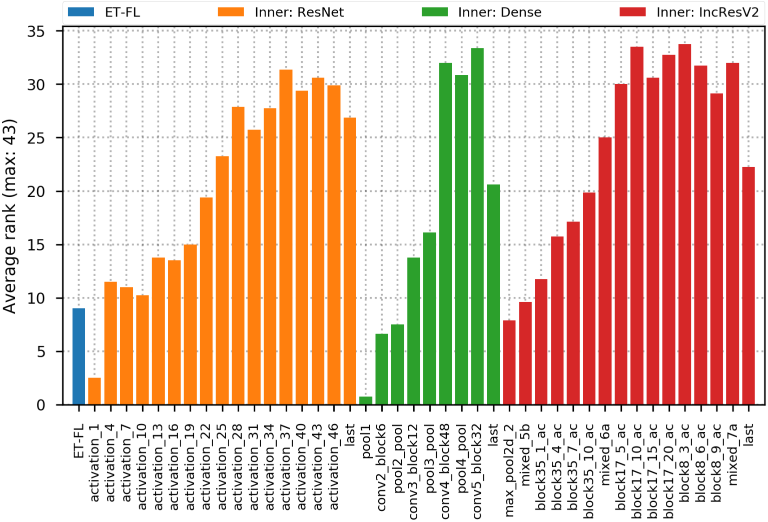
Fine-tuning is often the best performing method
... but OTS features are close on most problems and less computationally expensive !

	Datasets							
Strategy	**Cell**	**Prolif**	**Glom**	**Necro**	**Breast**	**Mouse**	**Lung**	**Human**
Baseline (ET-FL)	0.9250	0.8268	0.9551	0.9805	0.9345	0.7568	0.8547	0.6960
Last layer	0.9822	0.8893	0.9938	*0.9982*	0.9603	0.7996	0.9133	0.7820
Feat. select.	0.9676	0.8861	0.9843	**0.9994**	0.9597	0.7438	0.8941	0.7703
Merg. networks	*0.9897*	**0.8984**	0.9948	0.9864	0.9549	*0.8169*	0.9155	0.7928
Merg. layers	0.9808	0.8906	0.9944	0.9964	0.9639	0.7941	0.9268	0.7977
Inner ResNet	0.9748	*0.8959*	0.9949	0.9964	0.9664	0.8131	*0.9291*	*0.8113*
Inner DenseNet	0.9862	**0.8984**	*0.9962*	0.9917	0.9699	0.8012	0.9268	0.7967
Inner IncResV2	0.9873	0.8948	*0.9962*	*0.9982*	*0.9720*	0.8137	0.9234	0.7713
Fine-tuning	**0.9926**	0.8797	**0.9977**	0.9970	**0.9873**	**0.8727**	**0.9405**	**0.8641**
Metric	Roc AUC					Accuracy (multi-class)		

Table: Best in **bold**, second best in *italic*

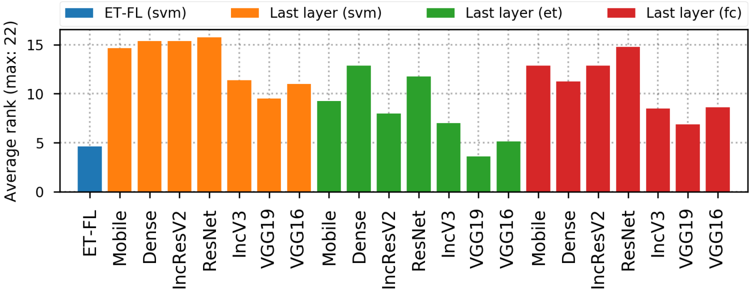
Results

When working with OTS features, use some inner layer features



Results

More recent networks like DenseNet or ResNet work better



See also: Kornblith, S., Shlens, J., & Le, Q. V. (2018). *Do Better ImageNet Models Transfer Better?*. arXiv preprint arXiv:1805.08974.

Conclusion

Main takeaways:

- Fine tuning is the best performing method
- OTS features often close to fine-tuning and less computationally expensive
- Prefer inner layers OTS features to last layers OTS features
- Use more recent networks such as DenseNet and ResNet

Thank you !

Meet me at poster **FP249** for more information !

Thank you !

Meet me at poster **FP249** for more information !