Comparison of deep transfer learning strategies for digital pathology - 19

Romain Mormont, Pierre Geurts, Raphaël Marée

r.mormont@uliege.be, p.geurts@uliege.be, raphael.maree@uliege.be

Montefiore Institute, University of Liège, Belgium

Abstract

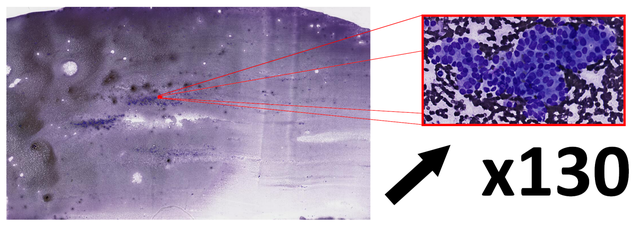
In this paper, we study deep transfer learning as a way of overcoming object recognition challenges encountered in the field of digital pathology. Through several experiments, we investigate various uses of pre-trained neural network architectures and different combination schemes with random forests for feature selection. Our experiments on eight classification datasets show that densely connected and residual networks consistently yield best performances across strategies. It also appears that network fine-tuning and using inner layers features are the best performing strategies, with the former yielding slightly superior results.

Digital pathology

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

Digital slides:

- big data: several multi-gigapixel slides per patient and case
- high variability: content, staining, acquisition,...
- data scarcity: annotating data is expensive and tedious



Need for efficient and versatile computer vision methods that can cope with data scarcity !

Deep transfer learning

Deep learning has a lot of potential for digital pathology but requires...

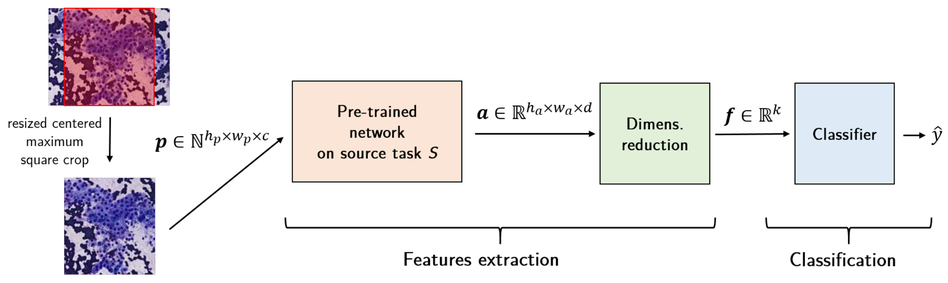
- lots of data
- expensive computing resources (*i.e.* GPUs)

Deep transfer learning alleviates those requirements:

- Either no network training or training converges faster
- Need less data and fewer resources

There are mainly two ways of using pre-trained networks [7]

1. using pre-trained features off-the-shelf (OTS)
   - with another classifier such as SVM or random forests
   - can run on CPU
   - requires less data then training from scratch
2. fine-tuning the networks
   - predicting with the fine-tuned network or using other classifiers on the fine-tuned features
   - need for a GPU but models converge faster
   - requires less data then training from scratch



Because of data scarcity, deep transfer learning is a promising approach for digital pathology.

Experiments

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 ?
- ...

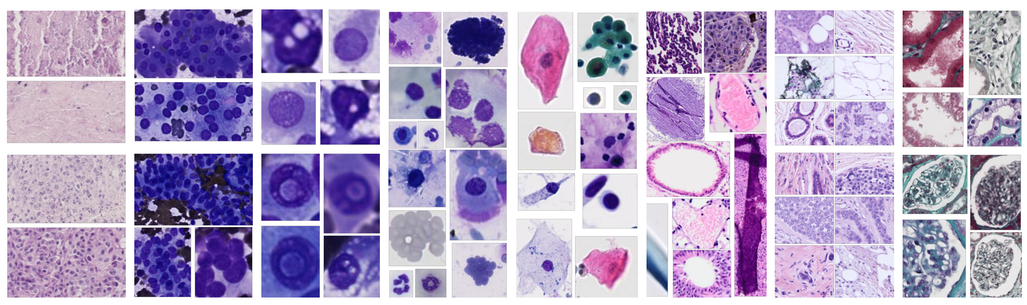
To answer those questions, we have carried out several experiments:

- Comparing OTS features with fine-tuning
- Comparing networks: ResNet50 [4], DenseNet201 [5], VGG16/19 [9], InceptionResNetV2 [10],...
- Several classifiers for OTS features: SVM [1], extra-trees (ET) [2],...
- Evaluating features extracted at different depth of different networks

Datasets

Dataset	Domain	Cls	Total	
			Images	Slides
Necrosis (N)	Histo	2	882	13
ProlifPattern (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

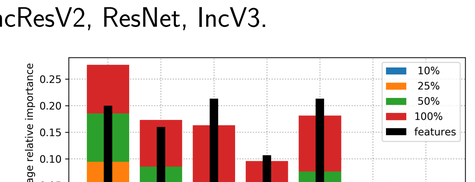
Our study uses eight image classification datasets collected over the years by biomedical researchers and pathologists using the Cytomine [8] web application. These contain tissues and cells from human or animal organs (thyroid, kidney, breast, lung,...).



(a) Necrosis (b) Prolif (c) CellIncl (d) Mouse (e) Human (f) Lung (g) Breast (h) Glomeruli

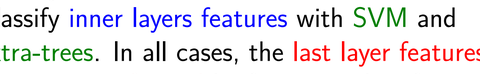
Results

Last layer features



Classify last layer OTS features with SVM, single layer perceptron and extra-trees. Older networks like VGGs and InceptionV3 are not competitive with more recent networks. SVM is better at classifying features. Best results with ResNet and SVM.

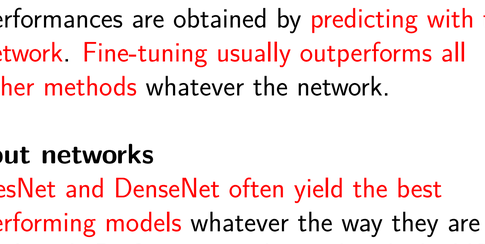
Feature selection with extra-trees



Feature selection with recursive feature elimination (RFE) [3] and extra-trees as features rankers. This experiment shows that most features are actually uninformative or redundant as RFE excluded on average 92.5% of the features. Our analysis also suggests that the best features are task-dependant.

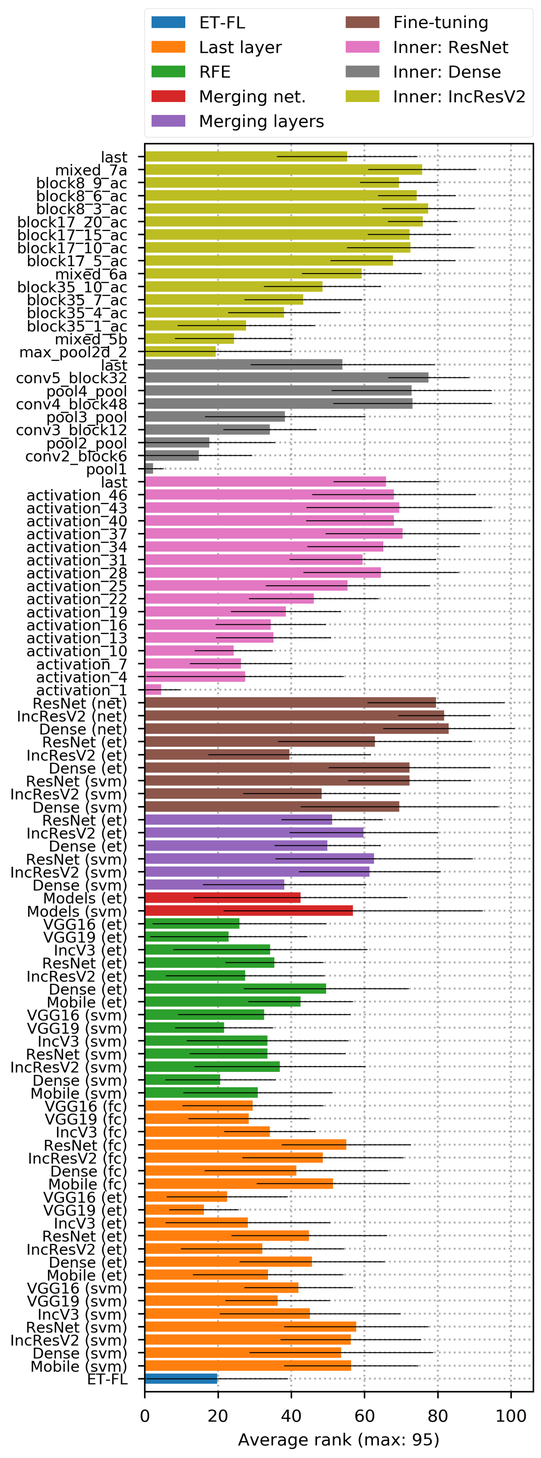
Merging features across networks

Merging last layer OTS features from all studied networks. An analysis of the features importances given by ET shows that DenseNet features bring 25% of the information on average. The next most informative networks features are from IncResV2, ResNet, IncV3.



Inner layers features

Classify inner layers features with SVM and extra-trees. In all cases, the last layer features are always outperformed by features taken from an inner layer of the network. However, the optimal layer is always located rather at the end of the network, while the first layers are clearly never competitive.



Fine-tuning

Fine-tuning networks, then either predicting with the fine-tuned networks directly or using fine-tuned features with SVM or ET. Best performances are obtained by predicting with the network. Fine-tuning usually outperforms all other methods whatever the network.

About networks

ResNet and DenseNet often yield the best performing models whatever the way they are exploited. Performances obtained with the VGG networks are below those of the others. This is corroborated by the recent findings of [6].

Conclusion and future works

We have carried out several experiments with 8 histopathology image datasets to devise guidelines for deep transfer learning in digital pathology.

Main takeaways:

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

In the future, we want to study further deep transfer learning and to collect and merge larger annotated biomedical datasets to train networks using a source dataset closer to our target tasks.

Bibliography

[1] Rong-En Fan et al. "LIBLINEAR: A library for large linear classification". In: *Journal of machine learning research* 9.Aug (2008), pp. 1871–1874.

[2] Pierre Geurts, Damien Ernst, and Louis Wehenkel. "Extremely randomized trees". In: *Machine learning* 63.1 (2006), pp. 3–42.

[3] Isabelle Guyon et al. "Gene selection for cancer classification using support vector machines". In: *Machine learning* 46.1-3 (2002), pp. 389–422.

[4] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[5] Gao Huang et al. "Densely connected convolutional networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. Vol. 1. 2. 2017, p. 3.

[6] Simon Kornblith, Jonathon Shlens, and Quoc V Le. "Do Better ImageNet Models Transfer Better?" In: *arXiv preprint arXiv:1805.08974* (2018).

[7] Geert Litjens et al. "A survey on deep learning in medical image analysis". In: *Medical image analysis* 42 (2017), pp. 60–88.

[8] Raphaël Marée et al. "Collaborative analysis of multi-gigapixel imaging data using Cytomine". In: *Bioinformatics* 32.9 (2016), pp. 1395–1401.

[9] Karen Simonyan and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556* (2014).

[10] Christian Szegedy et al. "Inception-v4, inception-resnet and the impact of residual connections on learning.". In: *AAAI*. Vol. 4. 2017, p. 12.