ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Bigger networks need bigger data, but what if data quantity impairs data quality?

PhD AI Seminars At BeCentral

Monday, December 9 · 16:00 – 18:00
Cantersteen 12 - 1000 Brussels - Belgium

Olivier Debeir [ULB]

ECOLE POLYTECHNIQUE DE BRUXELLES

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Bigger networks need bigger data, but what if data quantity impairs data quality?

Abstract:

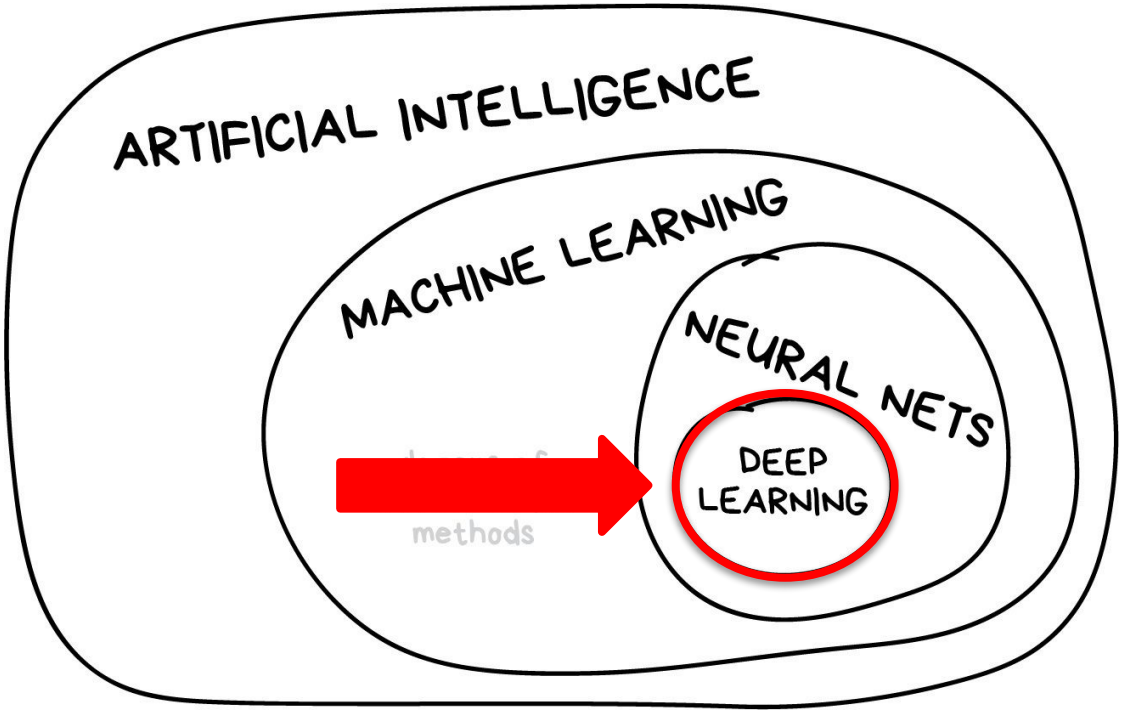
Deep-learning approaches had shown its high efficacy for a wide range of applications, in particular in image processing where until recently human vision was always more effective than machine approaches. This is maybe changing, or at least for some (not so) limited applications.

These techniques benefit on one hand from the cheap available computing power and on the other hand from huge image databases that are collected everywhere.

Though, to achieve a good performance, these deep learning-based algorithms need a huge amount of supervised data. In the medical domain, supervision may be costly in terms of expertise and time needed, resulting in incomplete or partially wrong supervision.

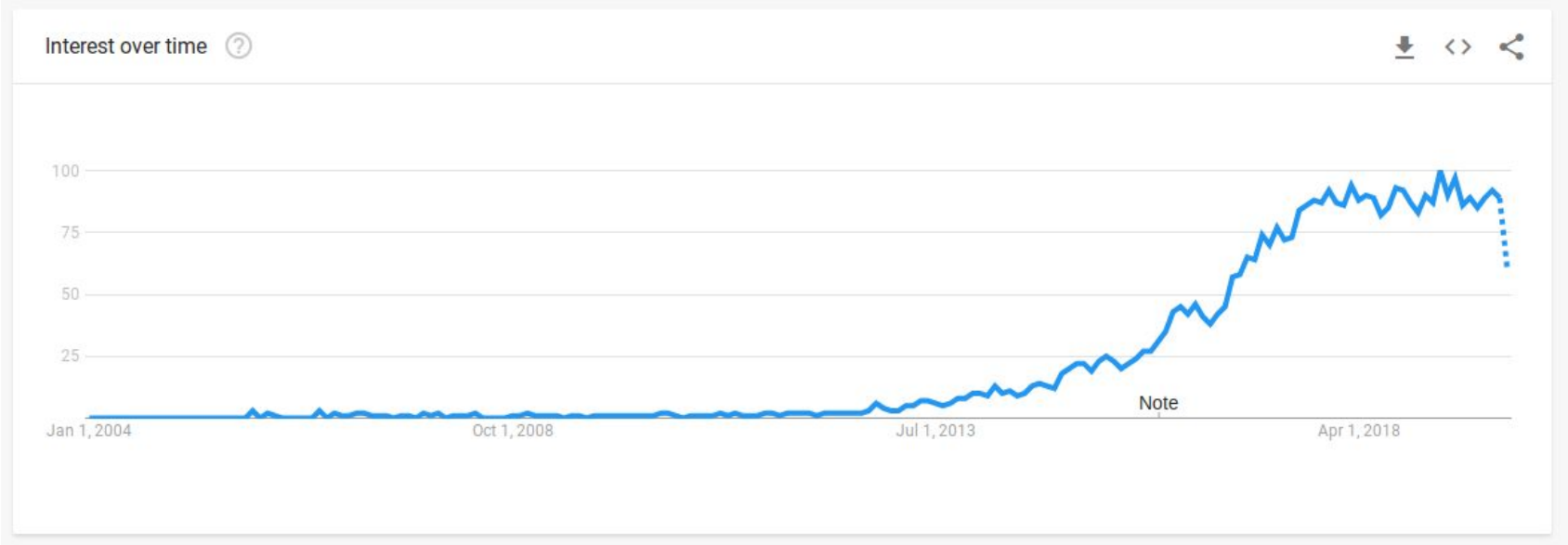
We, therefore, discuss here the trade-off that exists between the quality of the available training data and their size. We will show on a practical example how incomplete or wrong supervision can degrade network performances and what correcting strategies can be implemented.

Context



[vas3k.com/blog/machine_learning/]

Deep learning



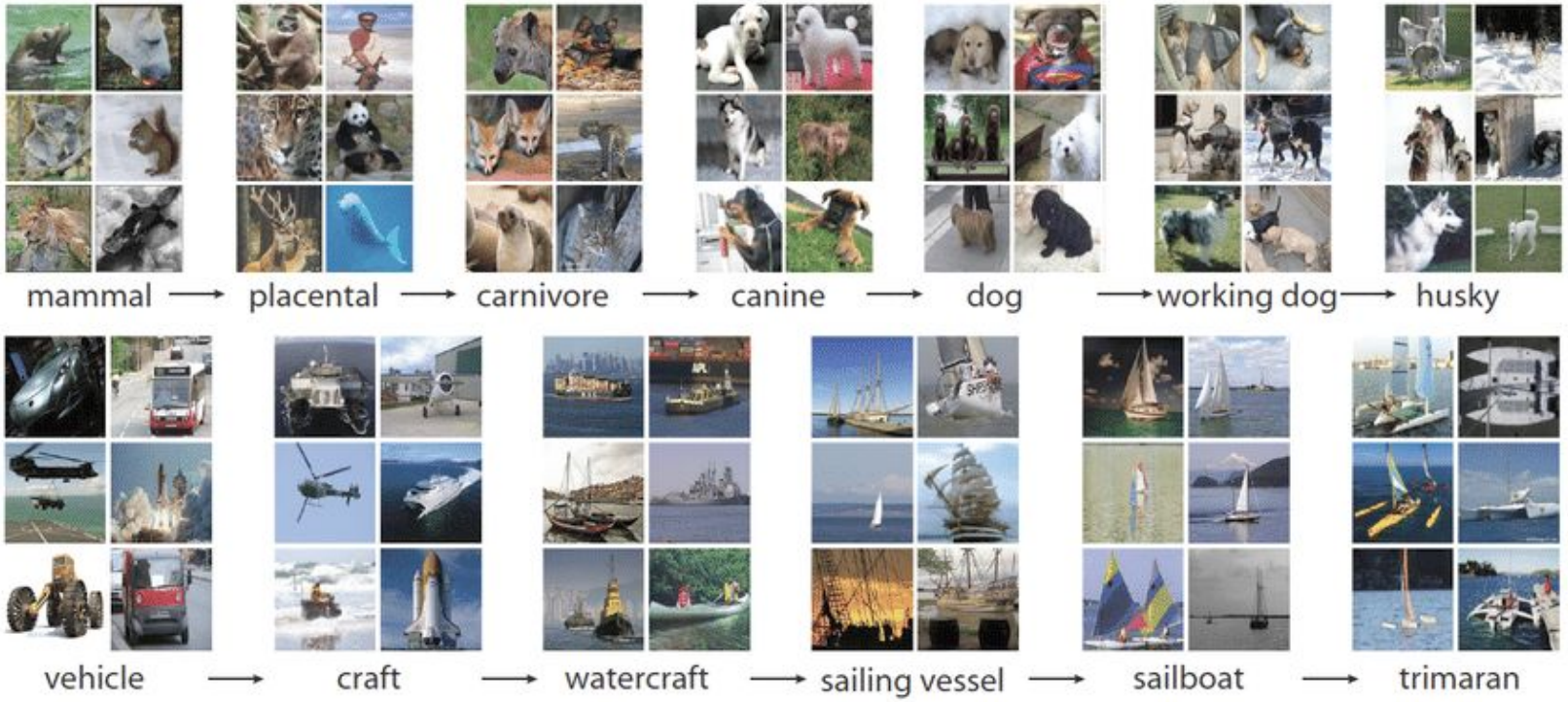
[Google Trends]

ImageNet (2009)
Large Scale Visual Recognition Challenge (ILSVRC) - (2010-2017)

Millions of images ...

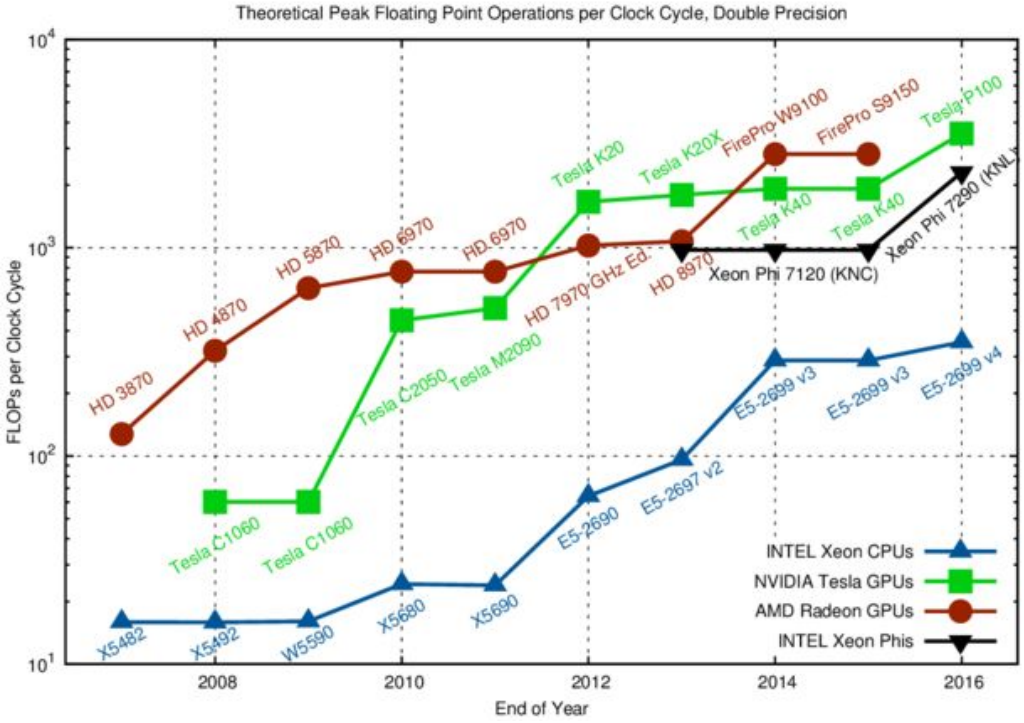
… annotated by humans using crowdsourcing

ImageNet - class examples



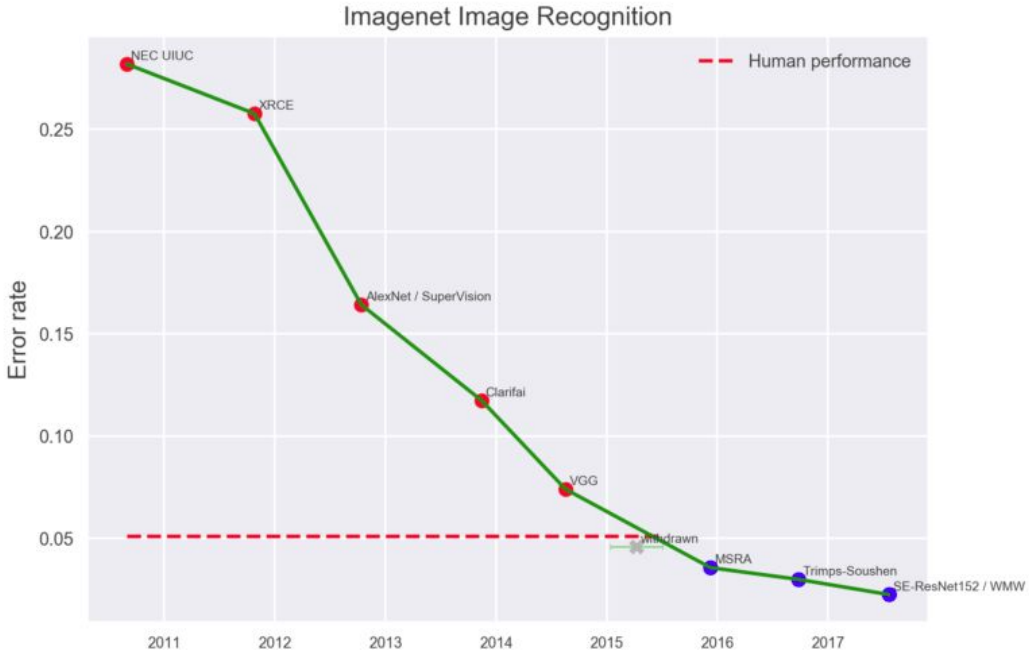
[Ye 2018]

… At the same time ...
the processing power….



[karlrupp.net]

Imagenet performances

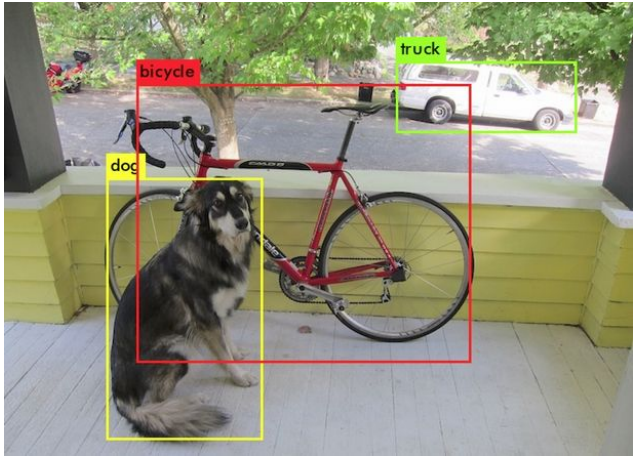


[Measuring the Progress of AI Research by EFF (CC BY-SA)]

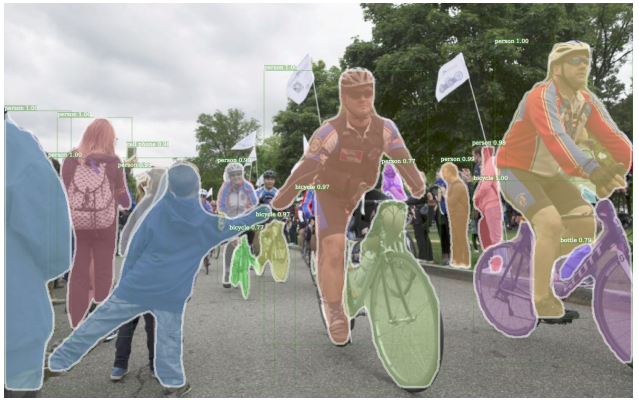
Some famous examples...

YOLO

Trained on Open Images
Dataset (~9 million images)



Mask R-CNN



What is an artificial Neural Networks ?

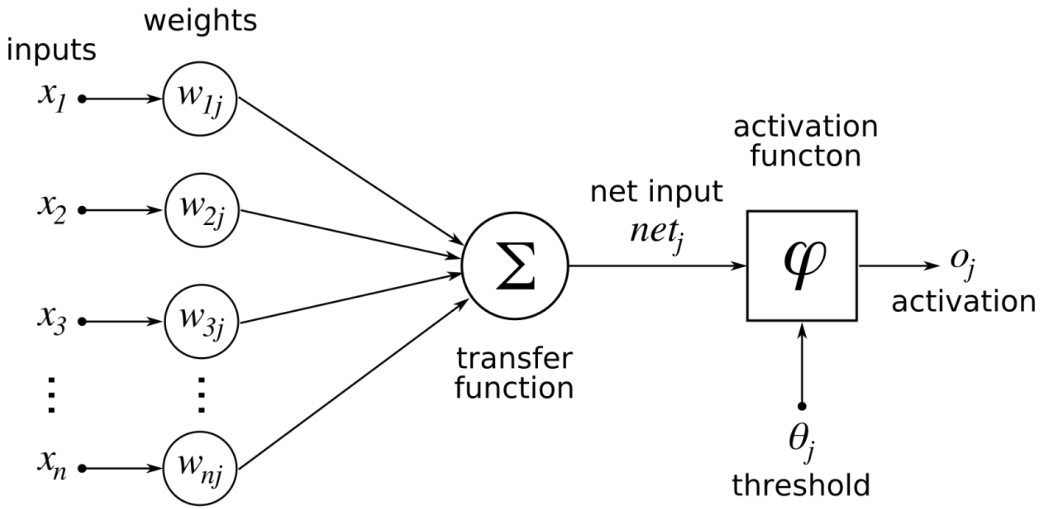
(simplified) History of the artificial NN

1940s-1950s : Perceptron (McCulloch & Pitts, Rosenblatt…)

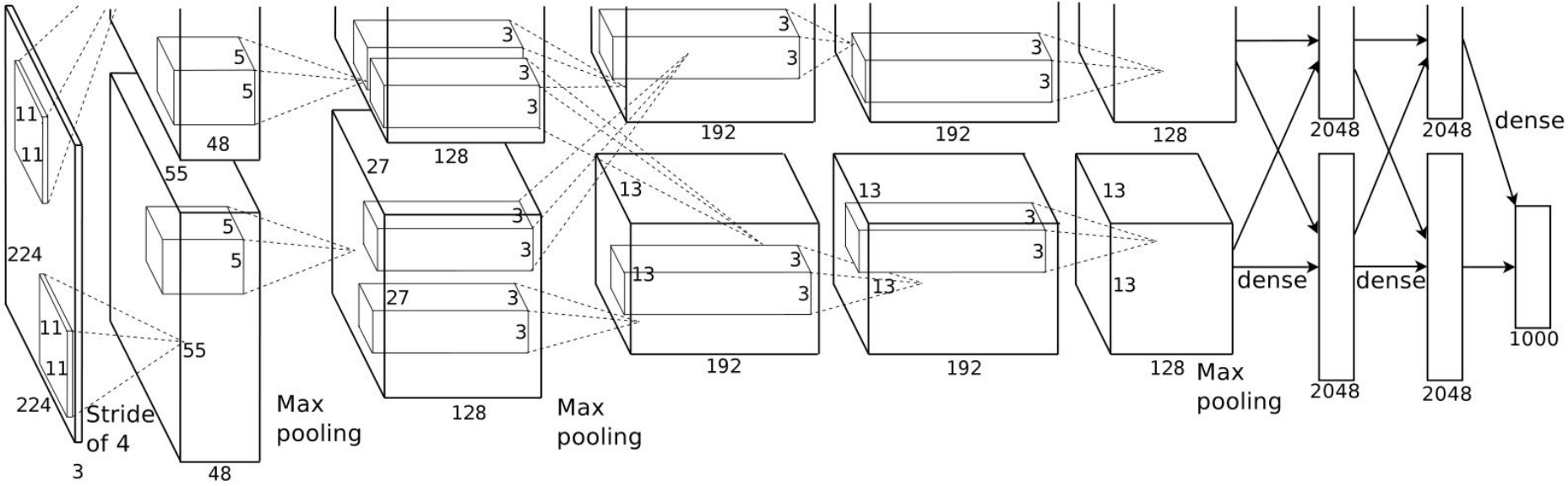
1950s-1960s : Multi-Layer Perceptrons (Ivakhnenko…)

…

1980s-1990s : Convolutional networks (Fukushima, LeCun…), backpropagation (Werbos, LeCun…)

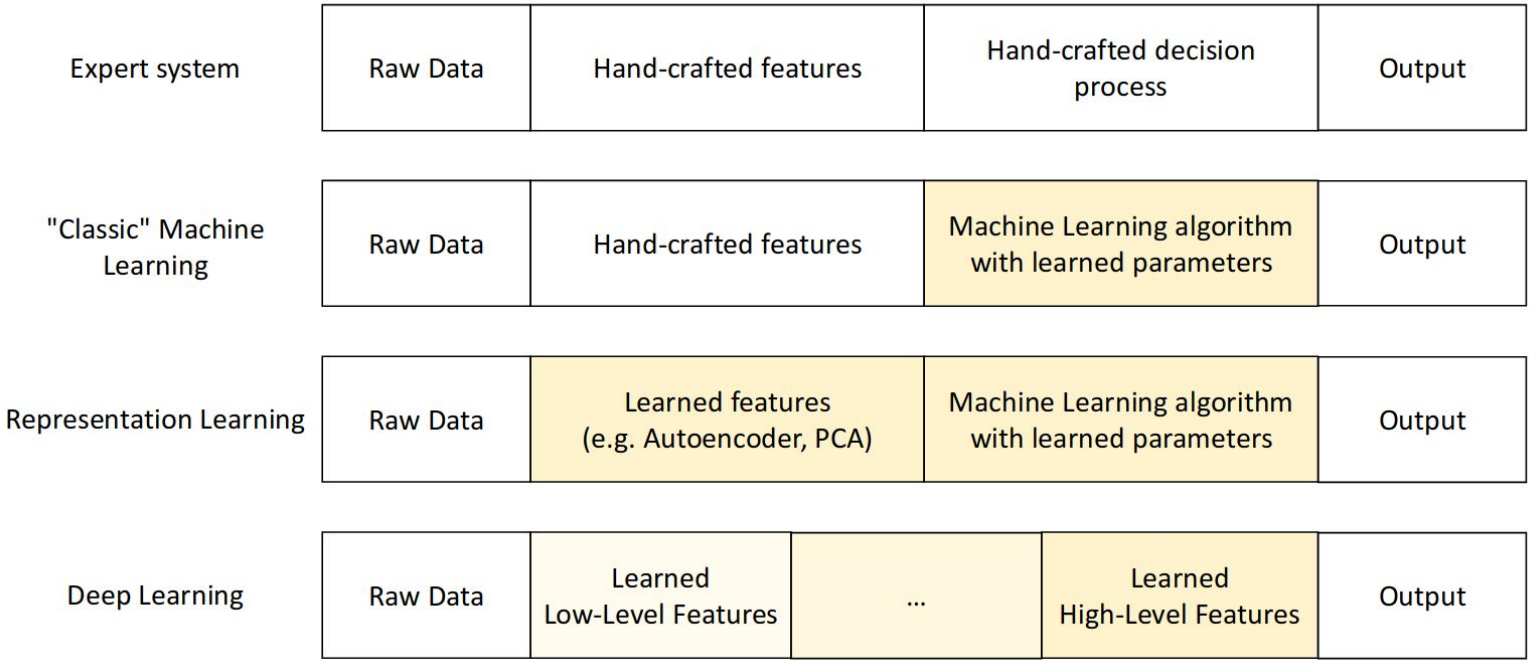


What is a “deep” artificial Neural Networks ?

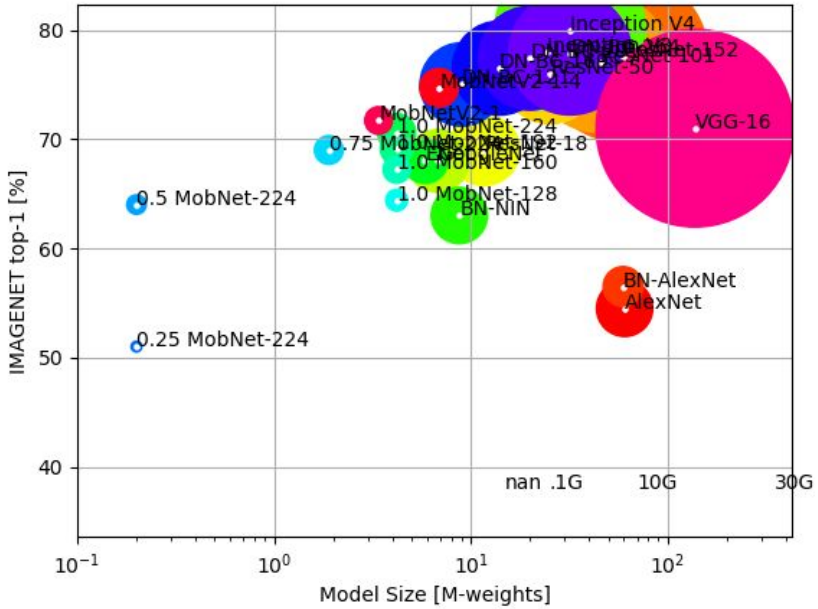


[A. Krizhevsky et al., 2012]

Why Deep ?


Expert system
Raw Data
Hand-crafted features
Hand-crafted decision process
Output
"Classic" Machine Learning
Raw Data
Hand-crafted features
Machine Learning algorithm with learned parameters
Output
Representation Learning
Raw Data
Learned features (e.g. Autoencoder, PCA)
Machine Learning algorithm with learned parameters
Output
Deep Learning
Raw Data
Learned Low-Level Features
...
Learned High-Level Features
Output

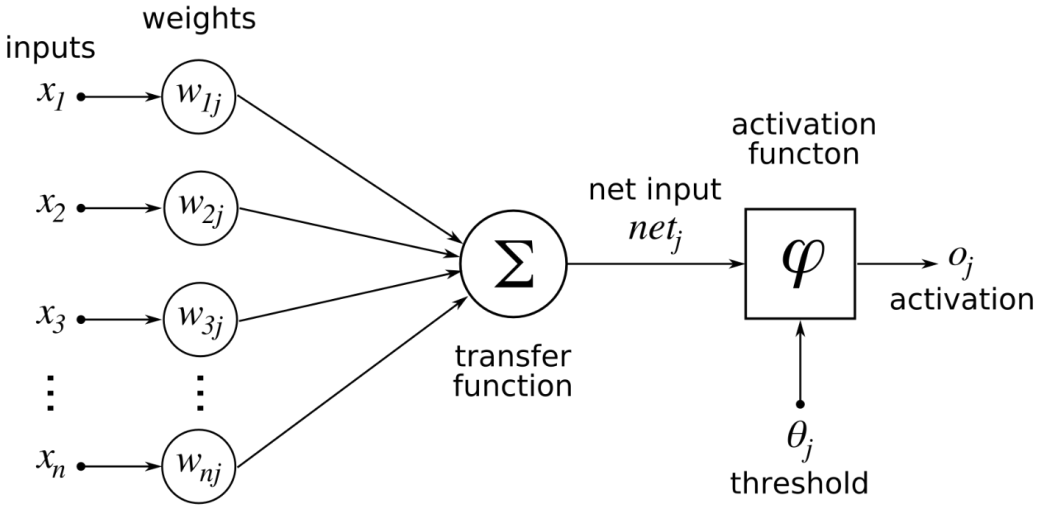
Network size



IMAGENET top-1 accuracy vs #weights, blob size is the #flops

[github.com/BertMoons]

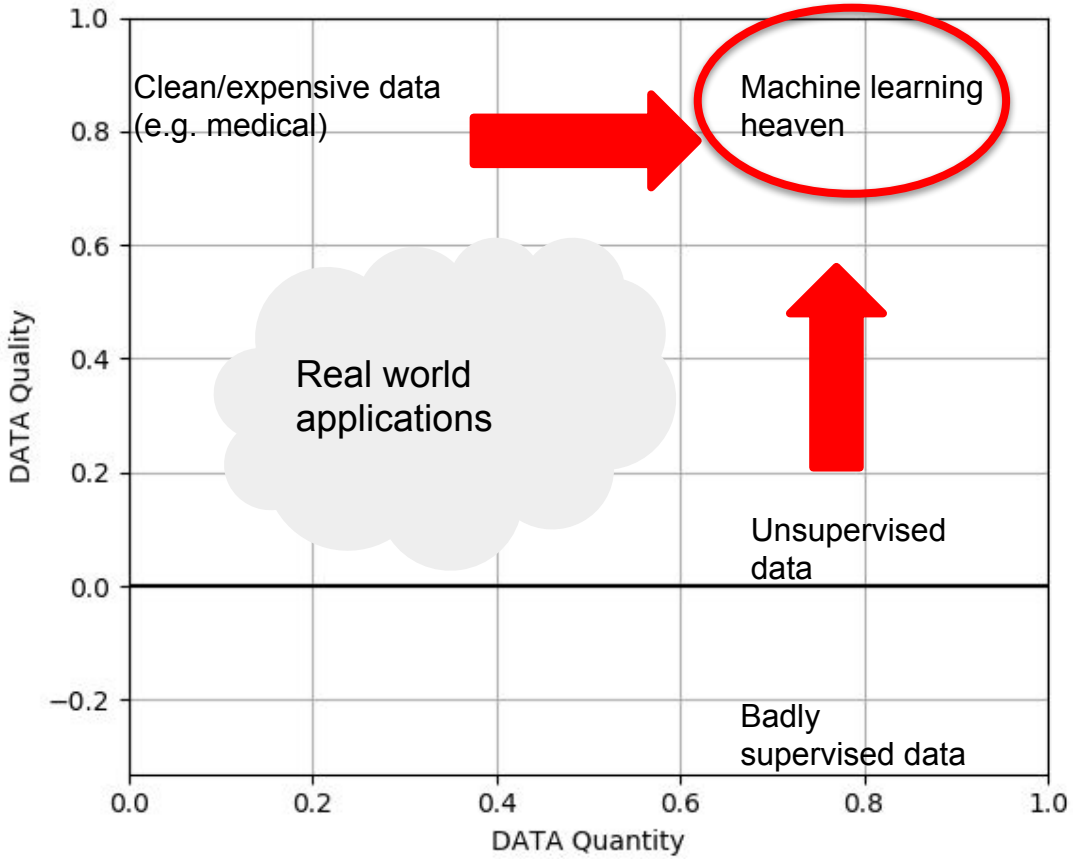
The cost of training a NN :



The rule of the thumb for a NN training is to have **10-20 training samples for each parameter** (weights) …

… training data can become an issue ...

Quality of the data vs Quantity of data

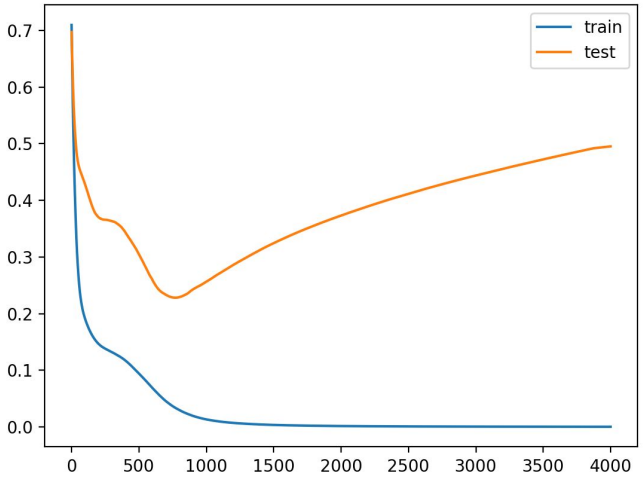


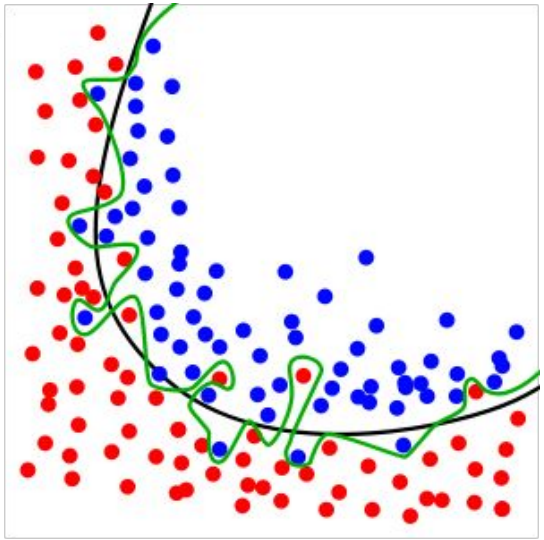
Why dataset size and quality are important ?

- Number of parameters
- Overfitting
- Risk of bias
- Class hidden in metadata
  - In a recent challenge, cancer image are encoded on 16 bits while benign on 8 bits …
  - In an other challenge, test set is extracted from the same patient (mitose detection) …
- Noise in the features
  - Missing data
  - errors
- Noise in the supervision (class)
  - Inter-expert discordance ?
  - Expert fatigue ...

Overfitting

- Too many parameters,
- Too few training samples
- Not enough diversity among traini samples

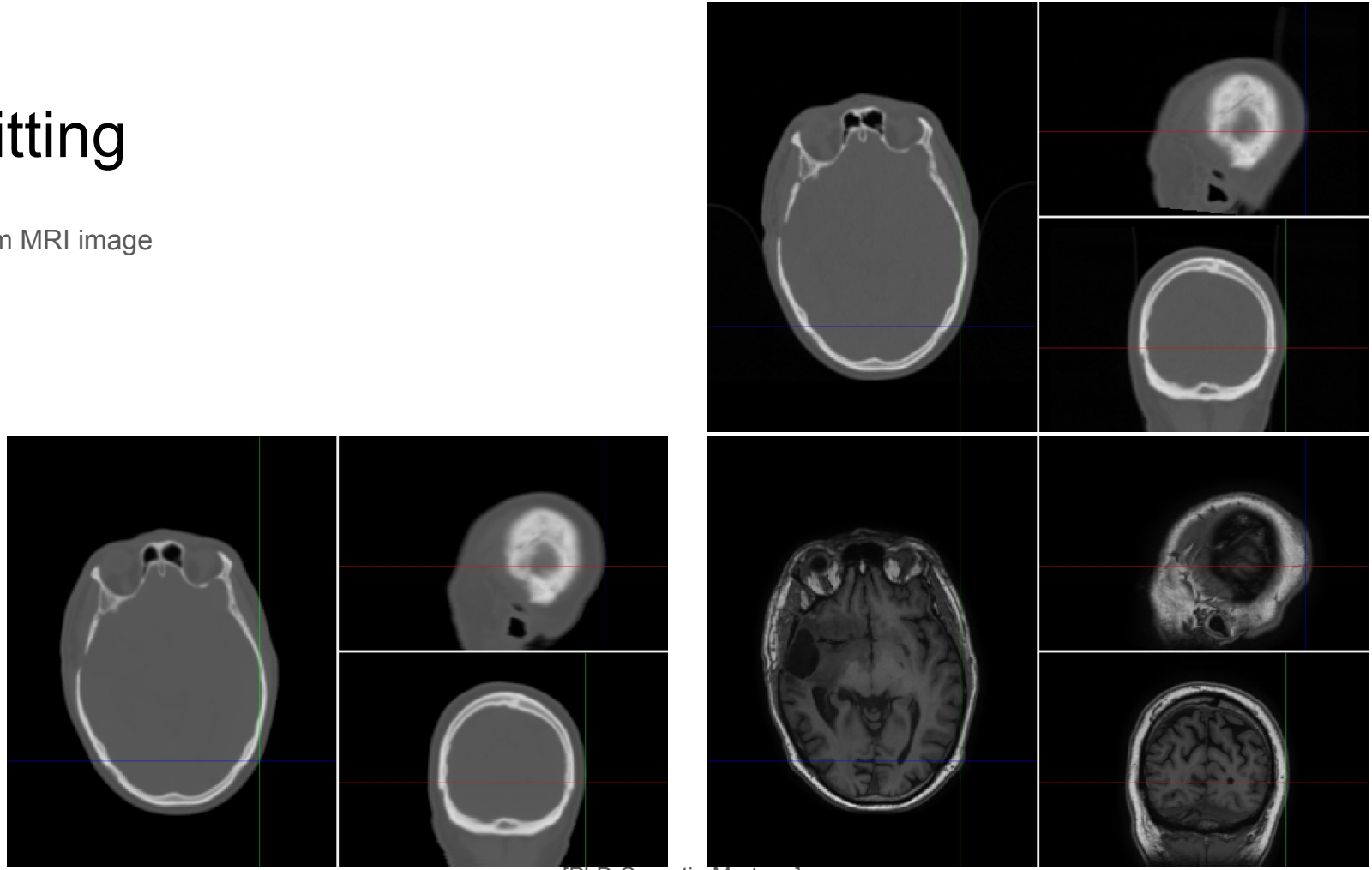




[wiki commons]

Overfitting

Pseudo CT from MRI image



[PhD Corentin Martens]

Why numerous data of high quality is rare ?

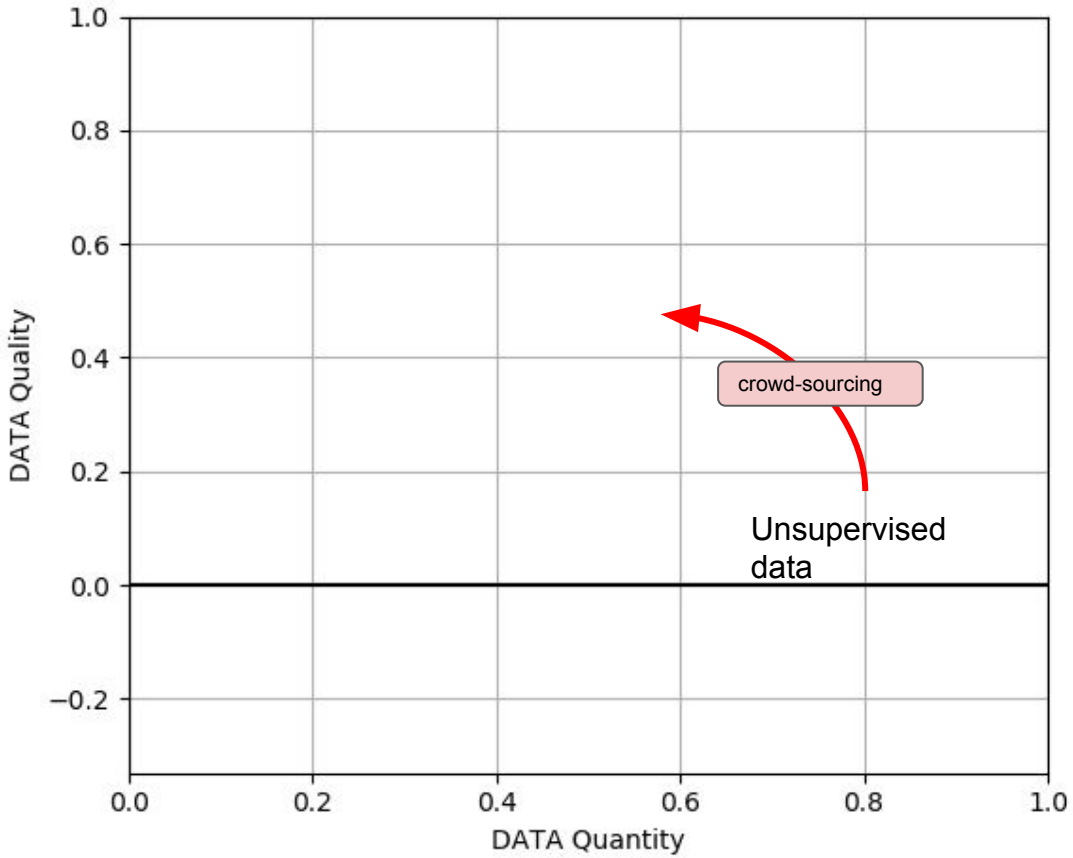
- Cost of data acquisition (destructive test)
- Rare events (limited time)
- Cost of data annotation (specialist)
- Data privacy (data difficult to share, GDPR)
- No clear consensus on the class definition (subjectivity)
- ...

Techniques

- Crowd-sourcing
- Data augmentation
- GAN data augmentation
- Variational auto-encoder
- Transfer learning
- Weakly supervised learning /Multiple Instance Learning
- Teacher learning
- Federate learning

Crowd-sourcing

Quality of the data vs Quantity of data



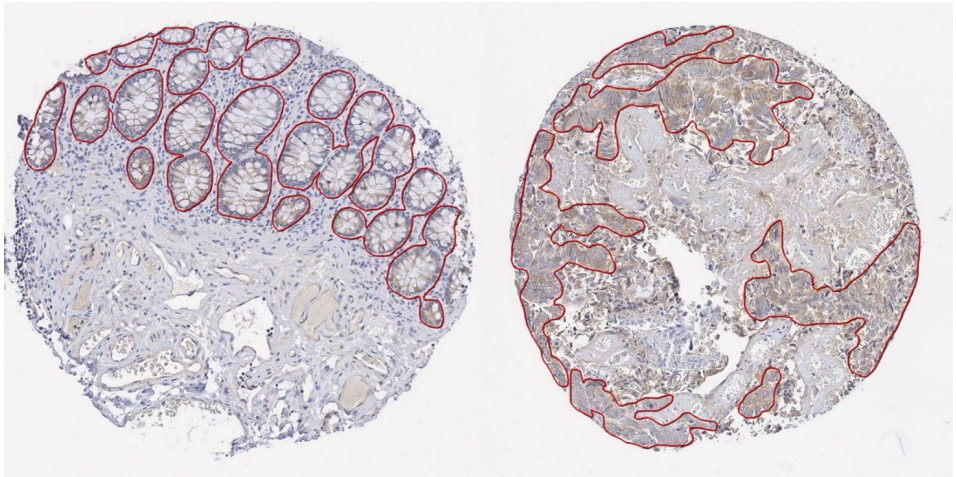
Data augmentation

Data augmentation
Image segmentation in digital pathology

- Huge datasets (large images)
- Time-consuming annotations
- Many objects of interests
- Diverse shapes and sizes

Aims:

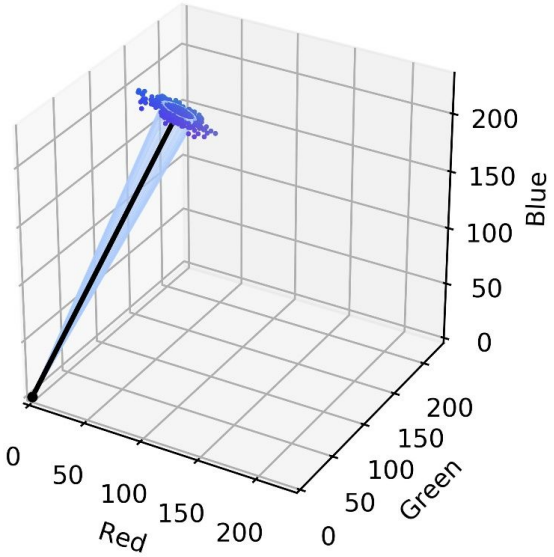
- extract as much (quantitative) information as possible from digital histological slides
- Object counting, density evaluation
- Morphological characterization: size, shape, …
- Quantification in specific compartments (e.g. tumor vs. stroma) of tissue-based biomarkers (IHC, CISH,…)

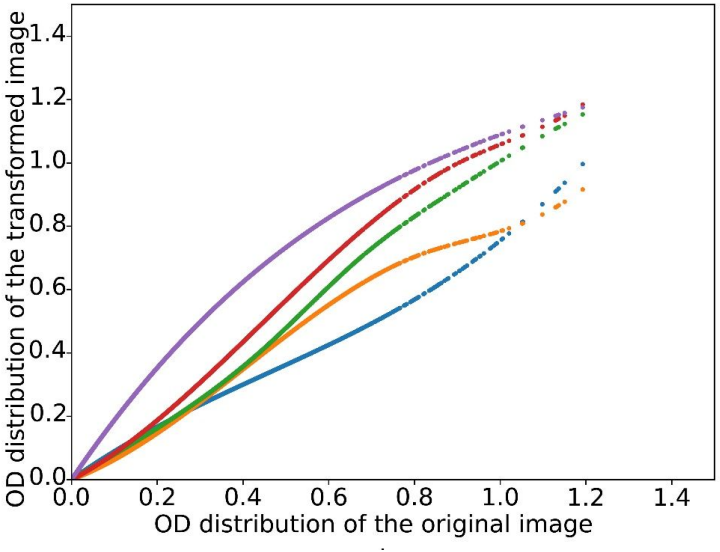


[Van Eycke YR et al., Med Image Analysis 2018]

Data augmentation
Image segmentation in digital pathology

Realistic image deformations shape and color

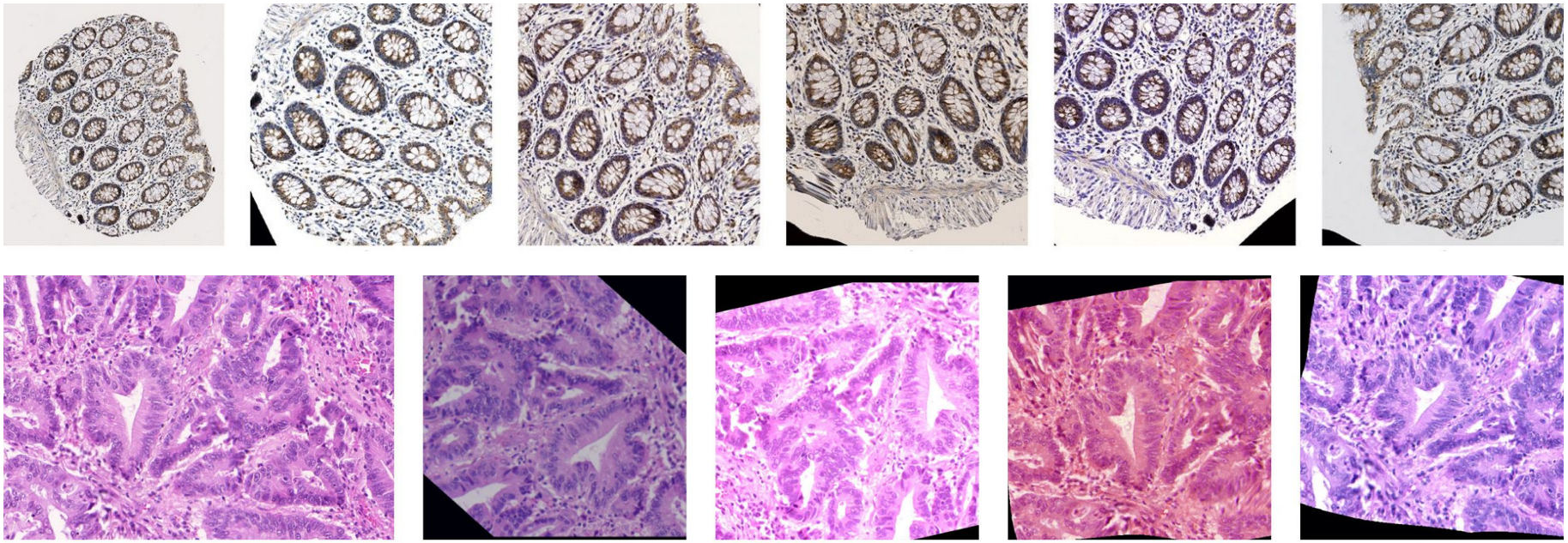




[Van Eycke YR et al., Med Image Analysis 2018]

Data augmentation
Image segmentation in digital pathology

Realistic image deformations shape and color



[Van Eycke YR et al., Med Image Analysis 2018]

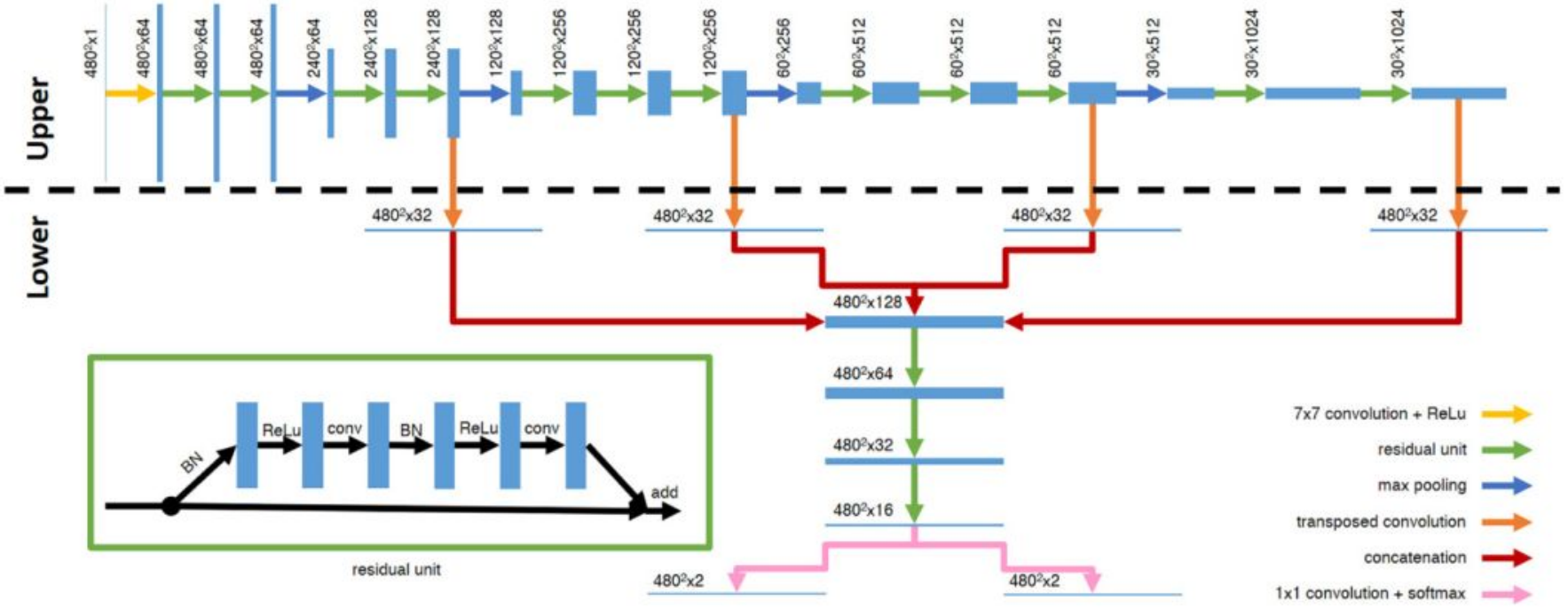
Data augmentation Image segmentation in digital pathology

Best participating teams	Data augmentation methods
GlaS challenge[a] (colon gland segmentation)	
CuMedVision	Transfer learning from natural images; affine and elastic (geometry) transforms
ExB	Affine and elastic (geometry) transforms; Gaussian blurring
Freiburg	Affine and elastic (geometry) transforms; random multiplications in HSV color space
Tupac 2016[b] (breast tumor proliferation assessment)	
Lunit inc.	Image translation; color, brightness, and contrast modifications
Contextvision	Affine (geometry) transforms
Sectra	No information
Radboud UMC	Affine and elastic (geometry) transforms; linear intensity transforms of the deconvoluted color channels; brightness, contrast, and saturation modifications; blurring and additive Gaussian noise
IBM Research	No information
Camlyon 2016[c] (detection of lymph node metastases)	
HMS and MIT	Image rotation; additive color noise
ExB	Image rotation and mirroring
Q.Wong	Image mirroring
Camelyon 2017[d] (detection of lymph node metastases)	
Shlee	Affine (geometry) transforms; contrast and HSV color space modifications
Ozymandias.watchman	Image flip and rotations; HSV color space modifications
Ericzz	Affine (geometry) transforms; linear transforms of the RGB color channels and HSV modifications

[a] *https://warwick.ac.uk/fac/sci/dcs/research/tia/glascontest*
[b] *http://tupac.tue-image.nl*
[c] *https://camelyon16.grand-challenge.org*
[d] *https://camelyon17.grand-challenge.org/*

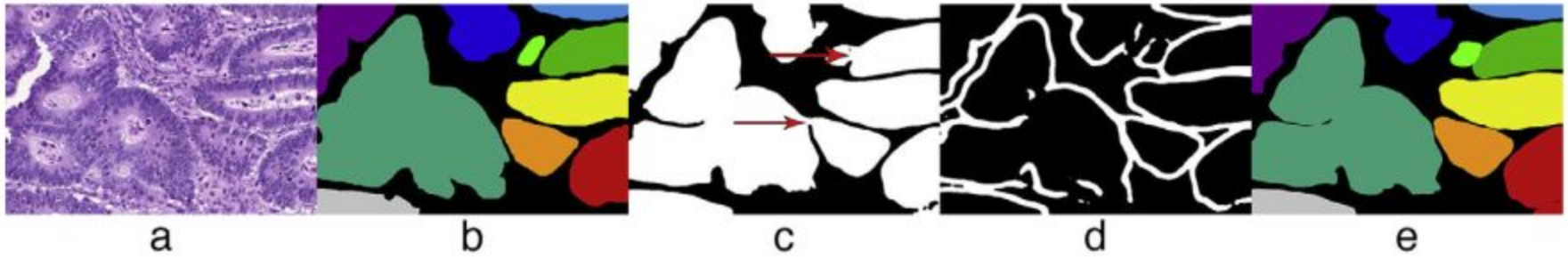
[Van Eycke YR et al., Frontiers in Medecine 2019]

Data augmentation
Image segmentation in digital pathology



[Van Eycke YR et al., Med Image Analysis 2018]

Data augmentation
Image segmentation in digital pathology



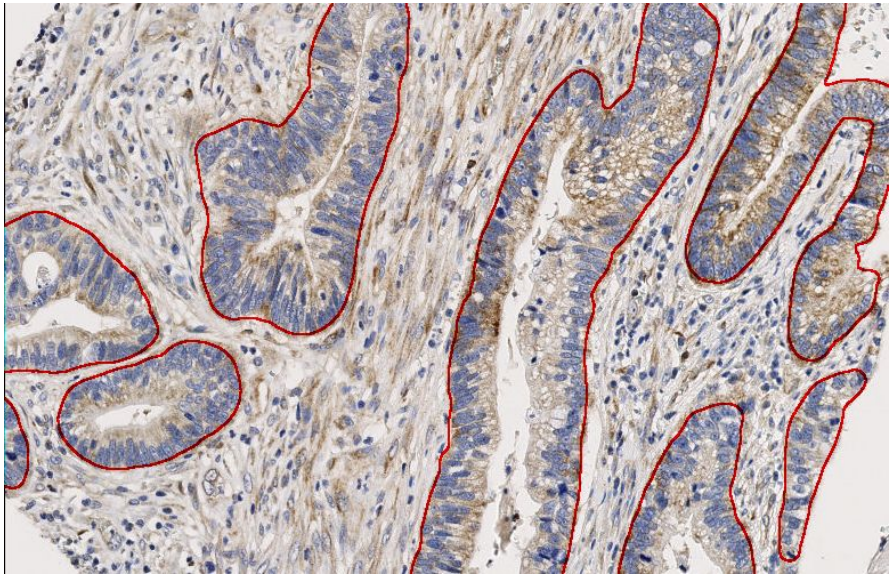
[Van Eycke YR et al., Med Image Analysis 2018]

Data augmentation
Image segmentation in digital pathology

Advantages:

able to produce very large datasets:

- tenths/hundreds of thousands generated from about 100 real images
- easy to implement (e.g. using Python library)
- enhanced robustness to the variations introduced in the data
- very efficient to train deep networks



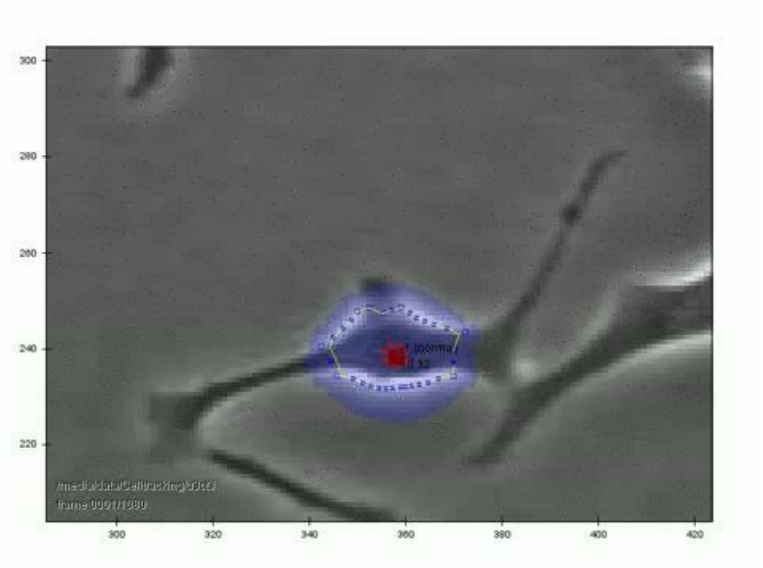
[Van Eycke YR et al., Med Image Analysis 2018]

Data augmentation
In-vitro cell detection and tracking

- Long sequences
- Time-consuming annotations
- Many objects of interests
- Diverse shapes and sizes

Aims:

- Detect automatically all the cells
- Object counting, density evaluation
- Event detection (cell division)



[Debeir O. et al., ISBI 2019 Proceedings]

Data augmentation
In-vitro cell detection and tracking

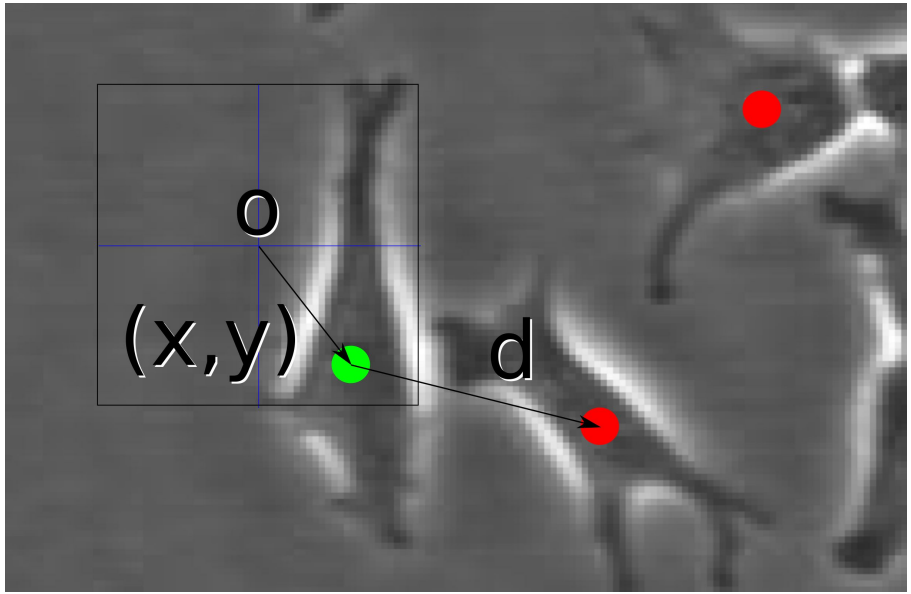
(x, y) vector to the closest tagged cell

d distance between the green cell and its closest neighbor

p probability of having a cell inside the neighborhood

(x, y, d, p) is the model describing the tile.

→ Deep regression

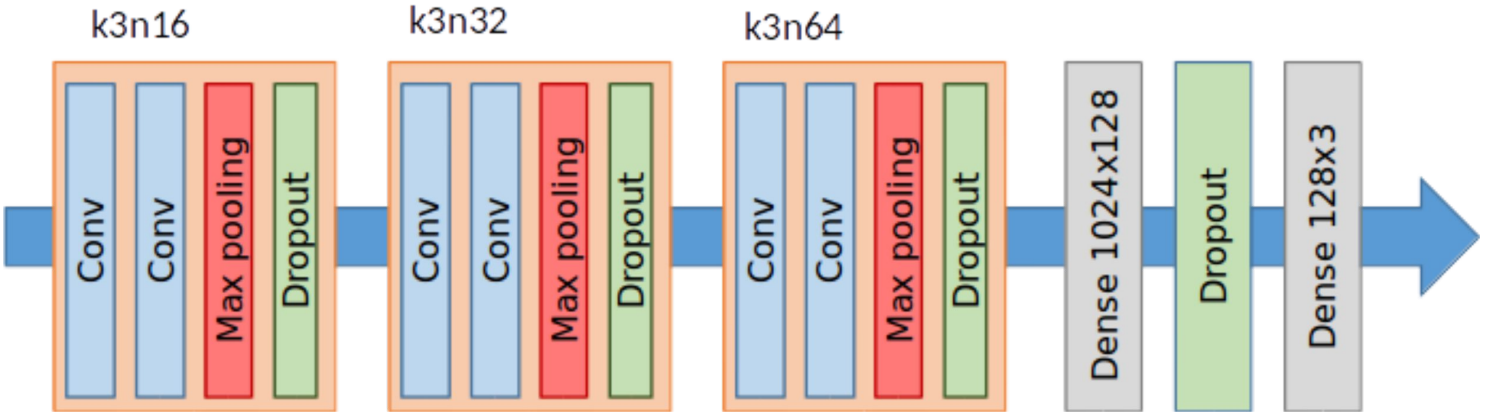


[Debeir O. et al., ISBI 2019 Proceedings]

Data augmentation
In-vitro cell detection and tracking

Predict (x,y,d,p) from 2D image time

203,379 parameters

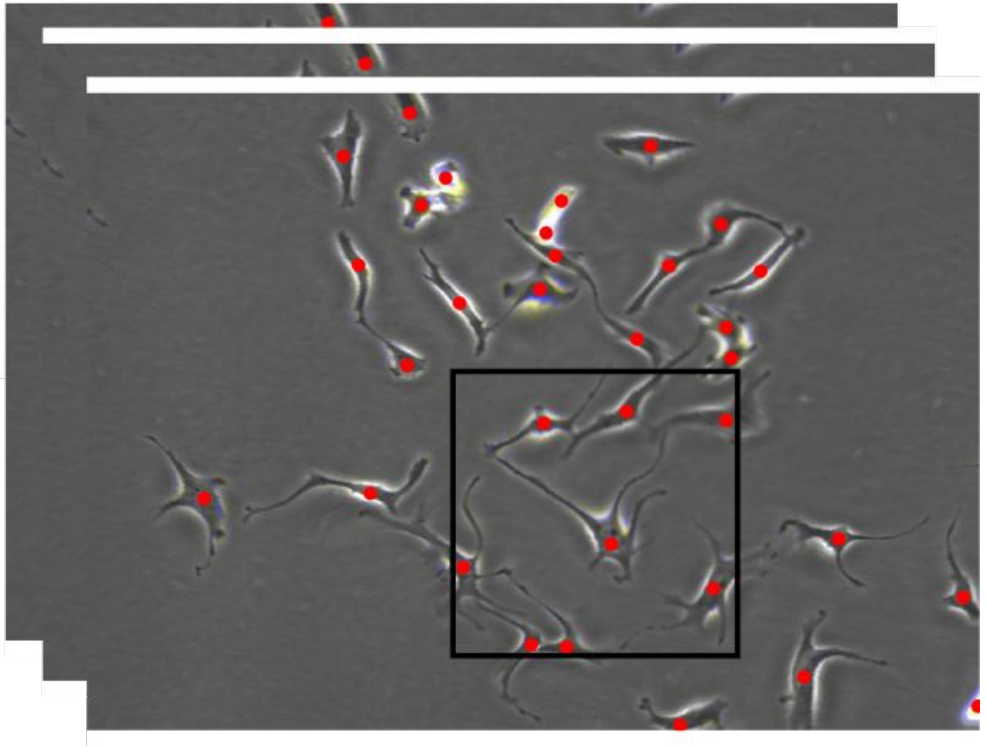


[Debeir O. et al., ISBI 2019 Proceedings]

Data augmentation
In-vitro cell detection and tracking

Training:

- **5** frames from a phase-contrast sequence of an in vitro U373 cell culture with a resolution of .92µm per pixel
- **261** cell centroids supervised



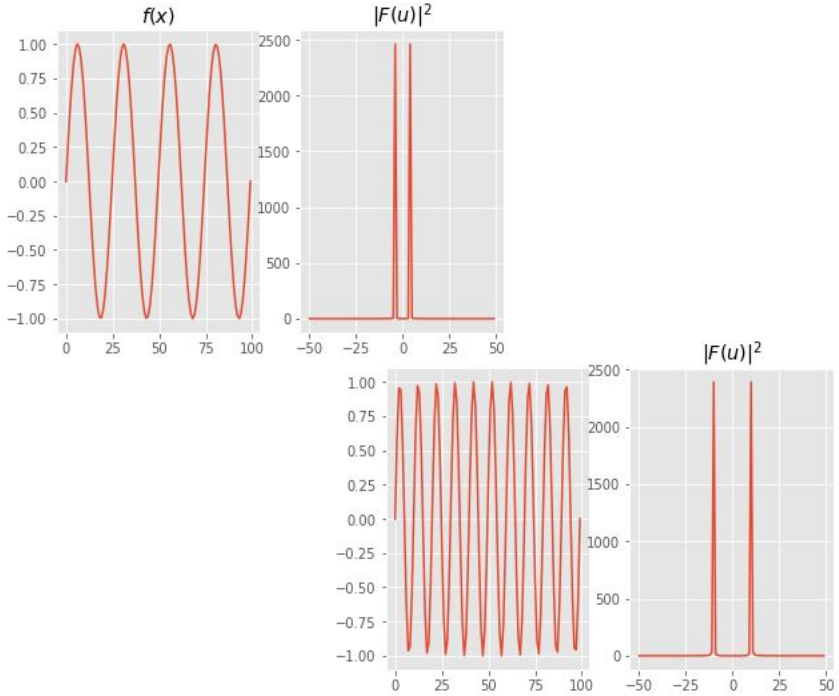
[Debeir O. et al., ISBI 2019 Proceedings]

Data augmentation
In-vitro cell detection and tracking

1) **Fourier geometrical distortion**

2) Random gamma brightness distortion

3) Random tiles 64x64



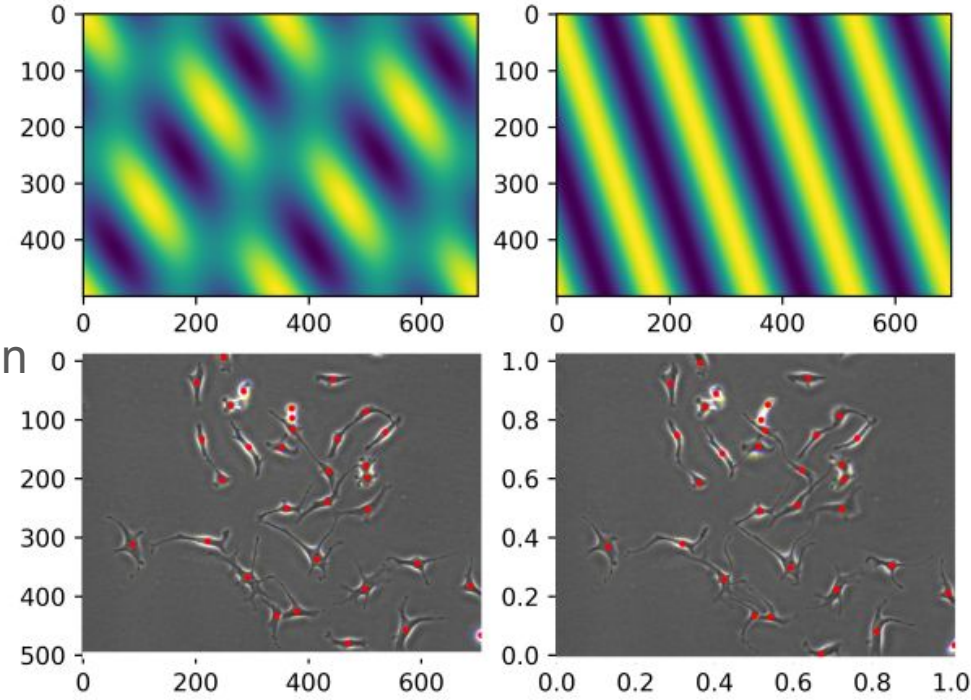
[Debeir O. et al., ISBI 2019 Proceedings]

Data augmentation
In-vitro cell detection and tracking

1) **Fourier geometrical distortion**

2) Random gamma brightness distortion

3) Random tiles 64x64



[Debeir O. et al., ISBI 2019 Proceedings]

Data augmentation
In-vitro cell detection and tracking

1) Fourier geometrical distortion

2) **Random gamma brightness distortion**

$$g_{out} = g_{in}^{\gamma} 255^{(1-\gamma)}, \begin{cases} \gamma = U(0,2) & \text{if } U(0,1) > .5 \\ \gamma = 1/U(.5,1) & \text{else} \end{cases}$$

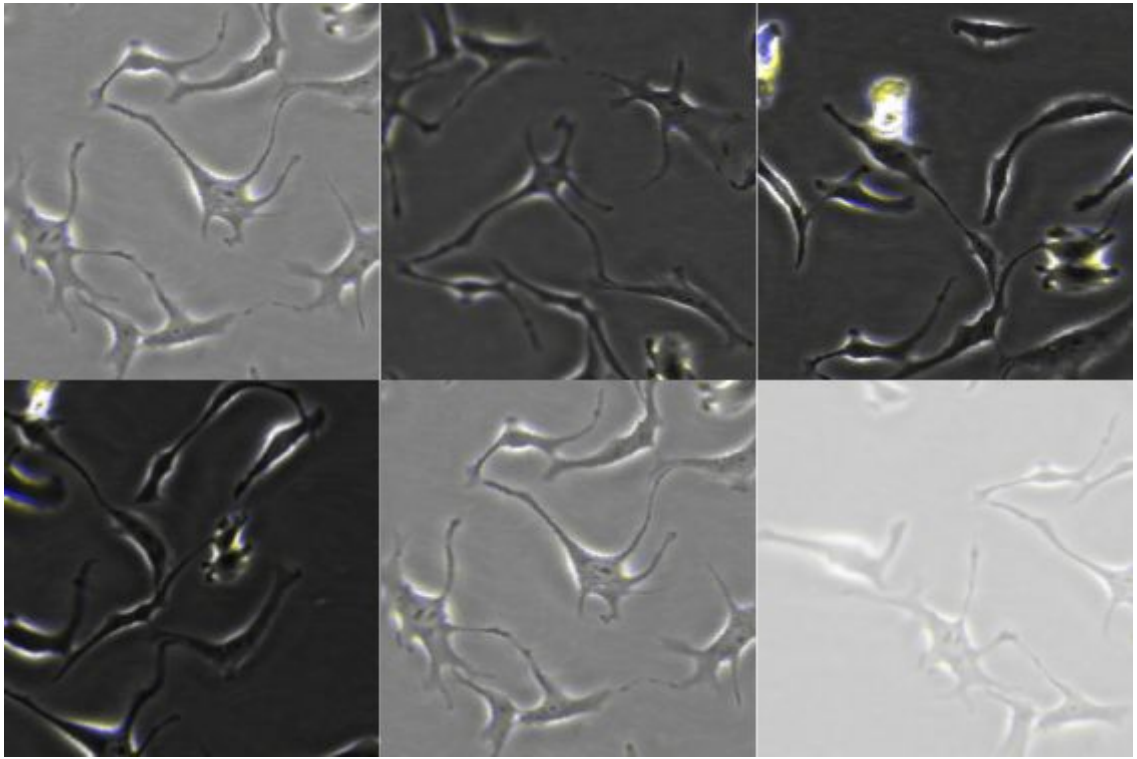
3) Random tiles 64x64

Data augmentation
In-vitro cell detection and tracking

1) Fourier geometrical distortion

2) **Random gamma brightness distortion**

3) Random tiles 64x64



[Debeir O. et al., ISBI 2019 Proceedings]

ULB

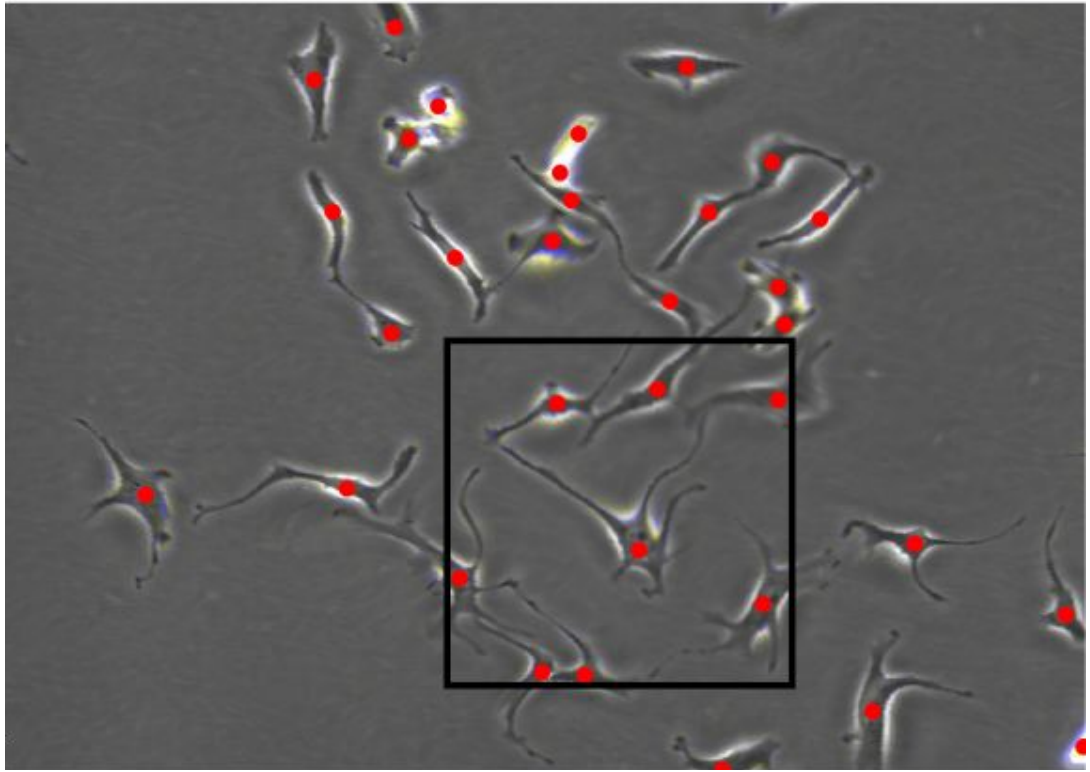
UNIVERSITÉ LIBRE DE BRUXELLES

Data augmentation

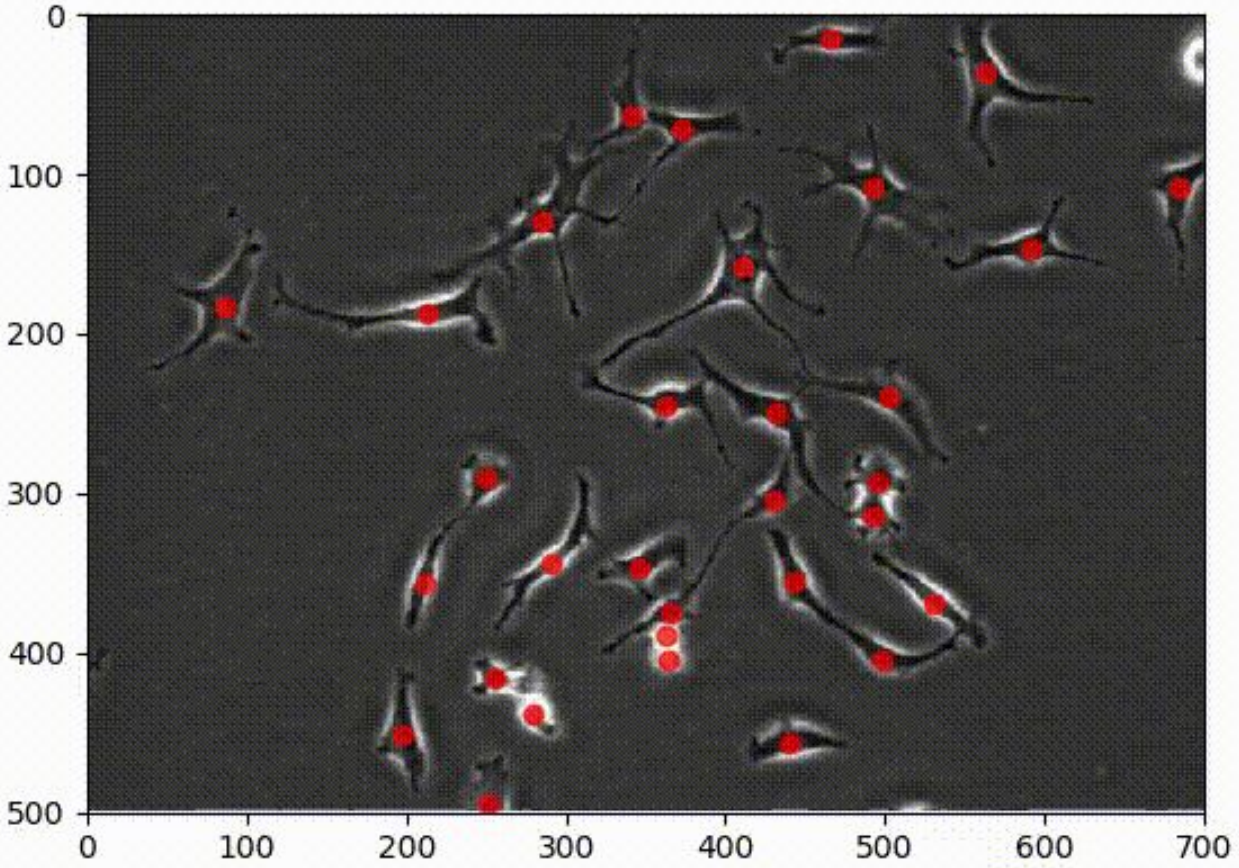
1) Fourier geometrical distortion

2) Random gamma brightness distortion

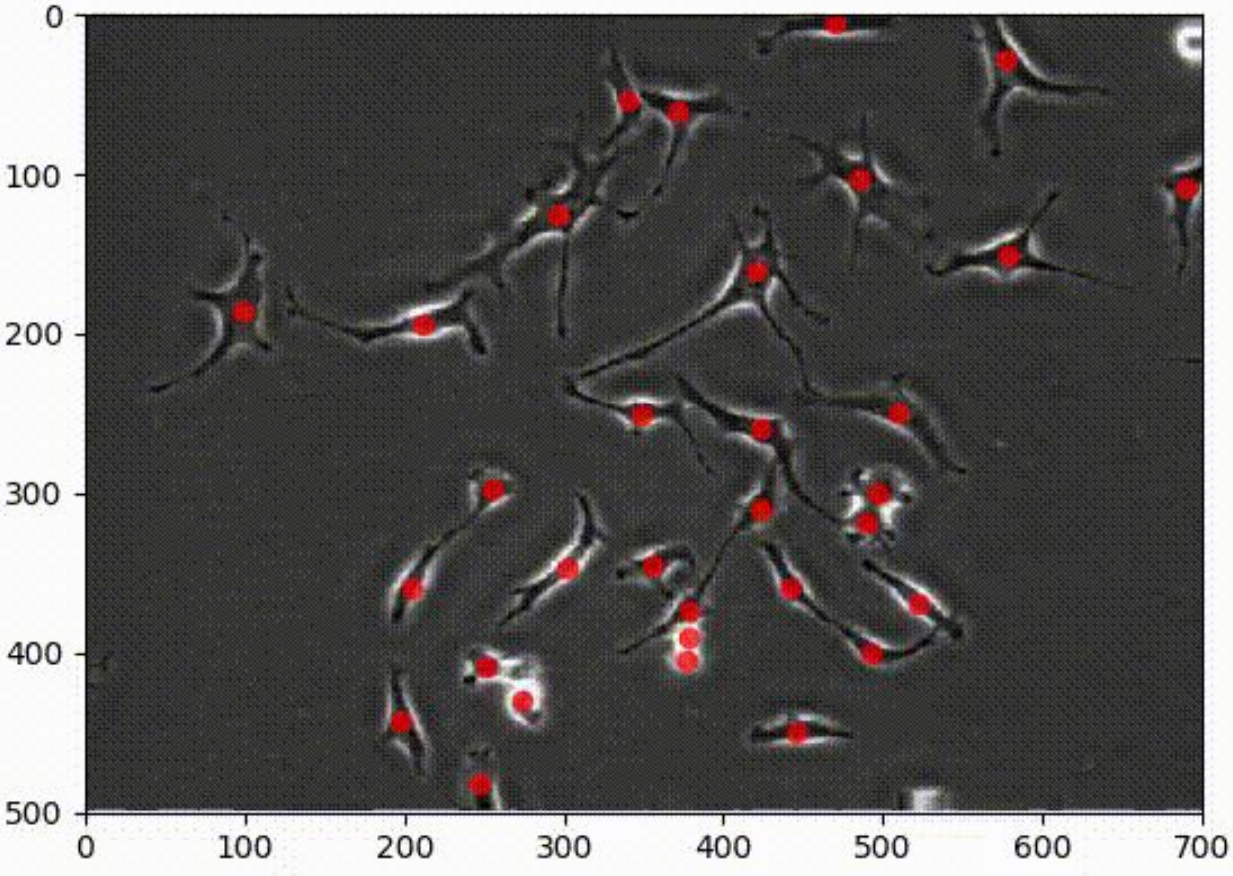
3) **Random tiles 64x64**



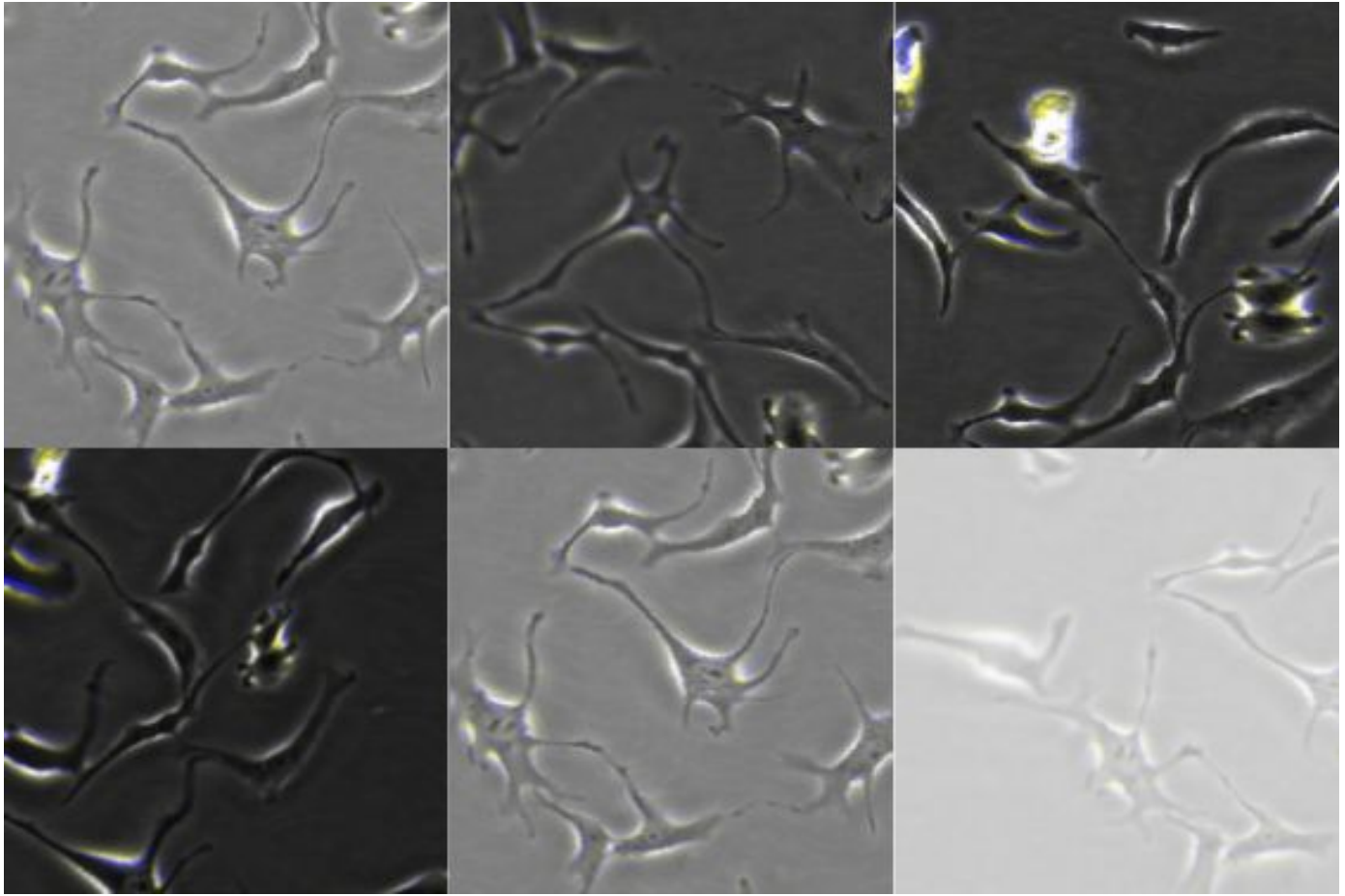
[Debeir O. et al., ISBI 2019 Proceedings]



[Debeir O. et al., ISBI 2019 Proceedings]



[Debeir O. et al., ISBI 2019 Proceedings]

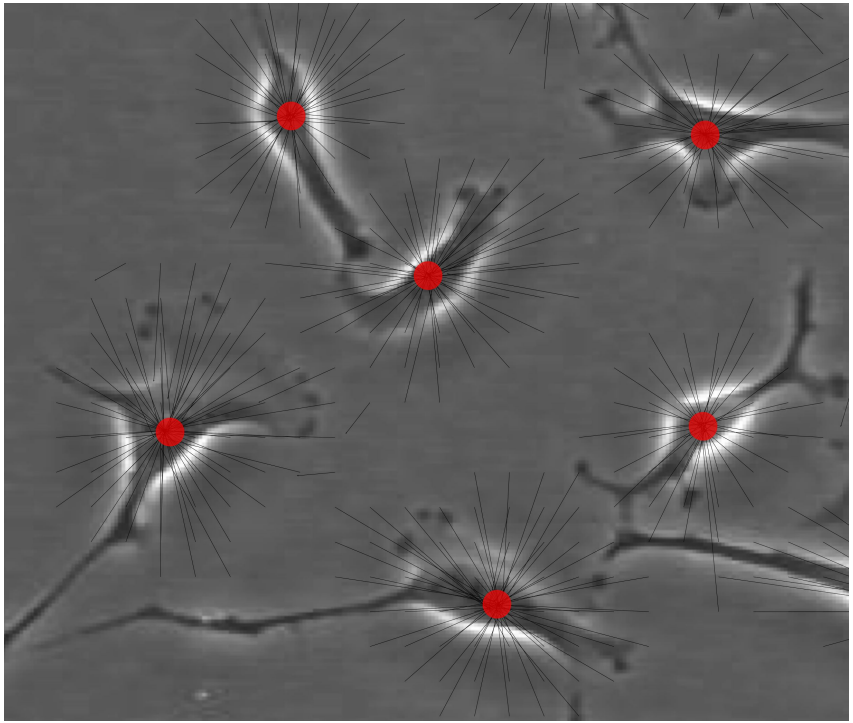


[Debeir O. et al., ISBI 2019 Proceedings]

Data augmentation
In-vitro cell detection and tracking

Result:

- a total of **160k** training samples are generated
- 2000 epoch using Adam optimizer.

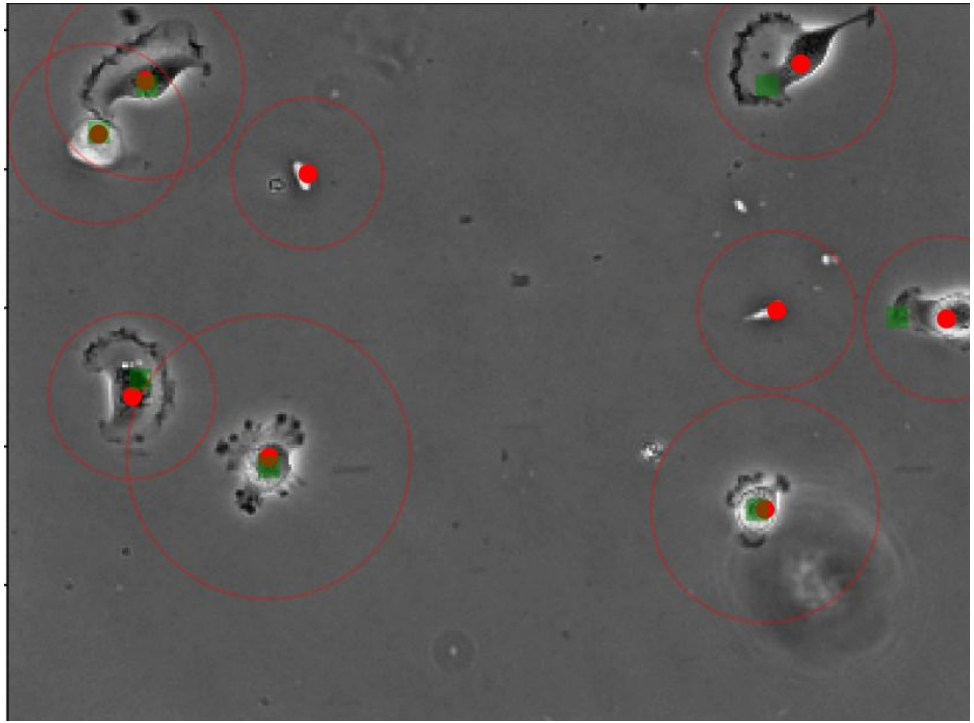


[Debeir O. et al., ISBI 2019 Proceedings]

Data augmentation
In-vitro cell detection and tracking

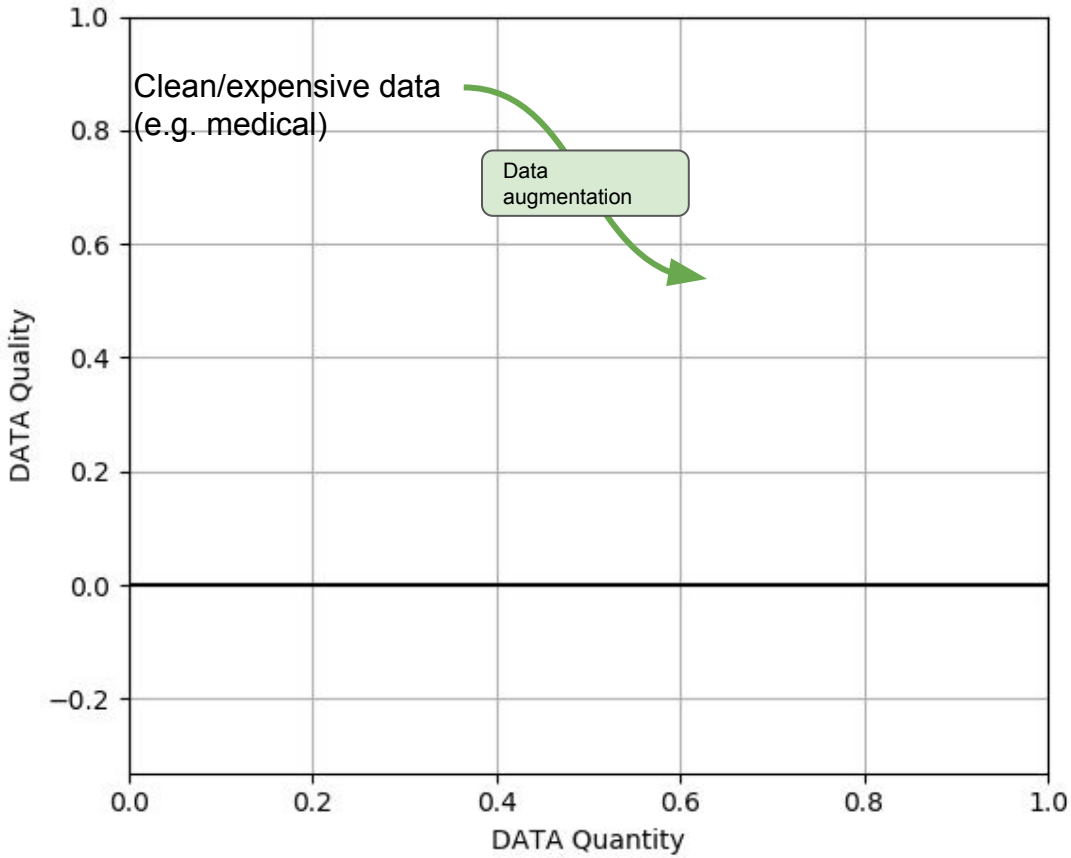
Unseen test sequence:

- V. Ulman et al.,“An objective comparison of cell-tracking algorithms,” Nature methods,vol. 14, no. 12, pp. 1141–1152, 2017.
- detection rate of 98:75% (7:1% multiple detections)
- false positive rate of 18:8%.



[Debeir O. et al., ISBI 2019 Proceedings]

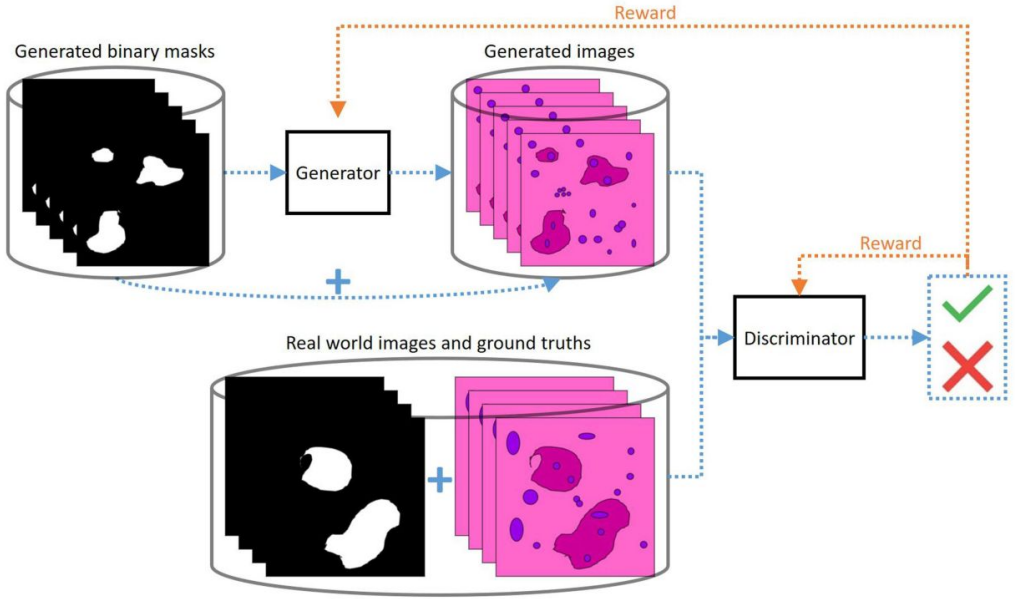
Quality of the data vs Quantity of data



GAN data augmentation

GAN data augmentation

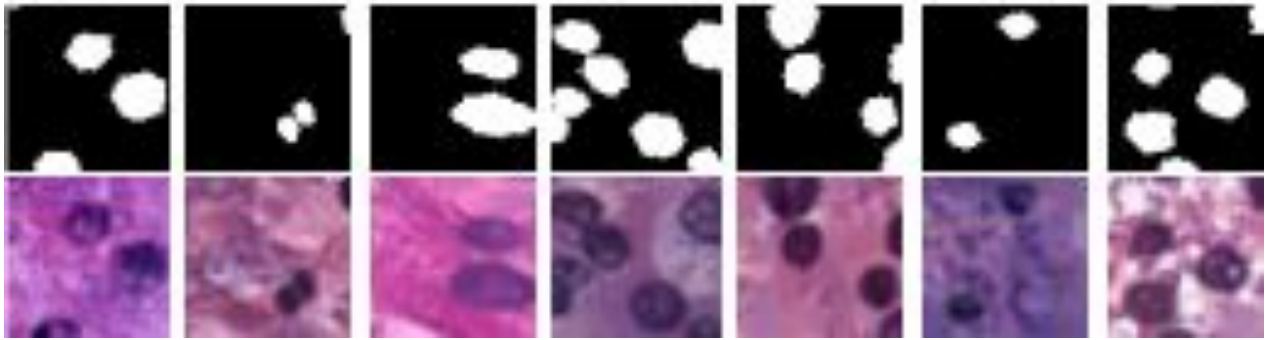
- The **Genarator** try to generate images as realistic as possible from a computer generated tissue mask
- The **Discriminator** receives as input a binary image and a tissue image, artificial or real, of which it must determine the origin.



[Van Eycke YR et al., Frontiers in Medecine 2019]

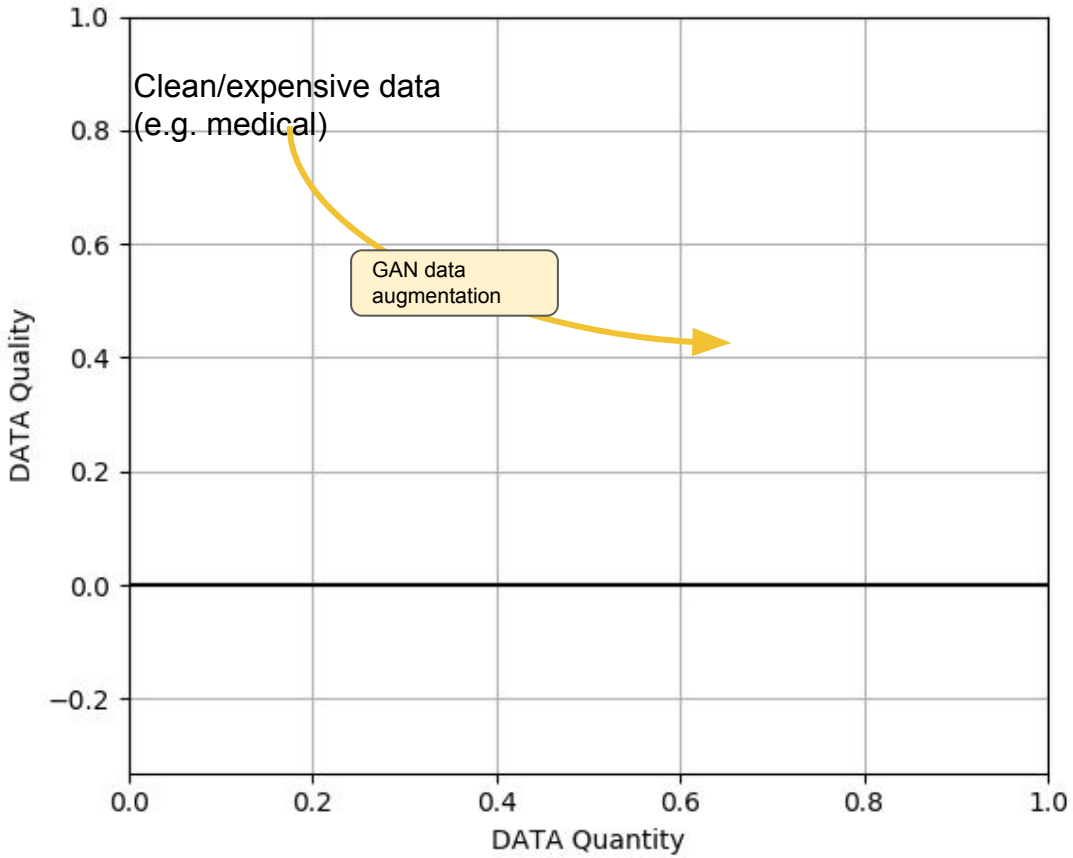
GAN data augmentation

- Synthetic patches for training in nucleus segmentation



[Hou L et al. CVPR 2019]

Quality of the data vs Quantity of data



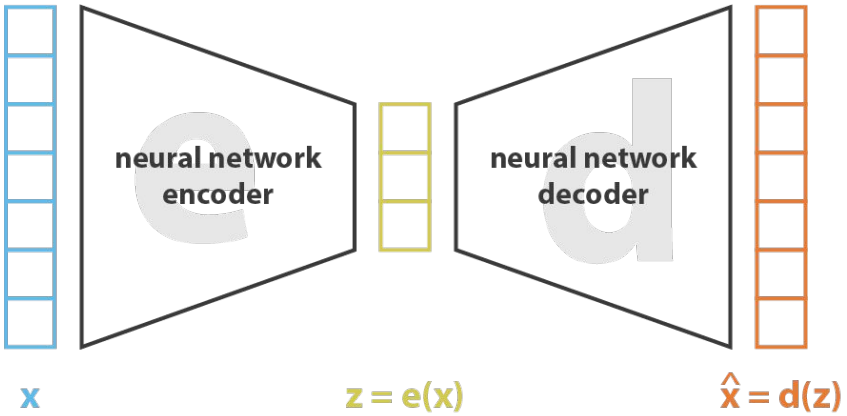
Variational auto-encoder

Variational auto-encoder

Auto-encoder allow to compress data into some ‘latent’ space of lower dimensions

Using the decoder as a generator of data is possible

But latent space is not well organised



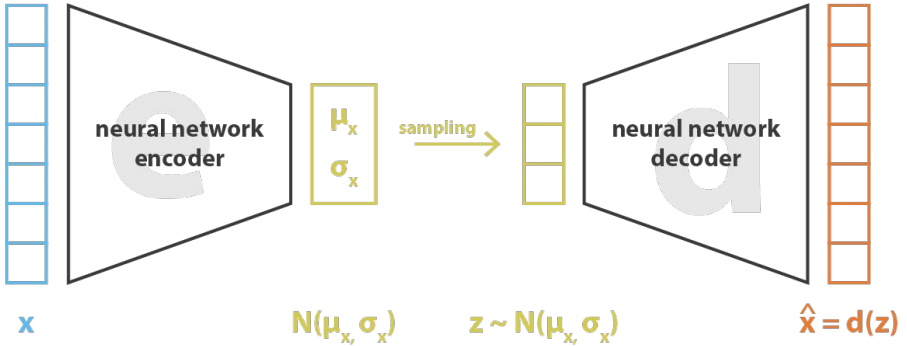
$$\text{loss} = \| x - \hat{x} \|^2 = \| x - d(z) \|^2 = \| x - d(e(x)) \|^2$$

[Joseph Rocca 2019]

Variational auto-encoder

Variational auto-encoder

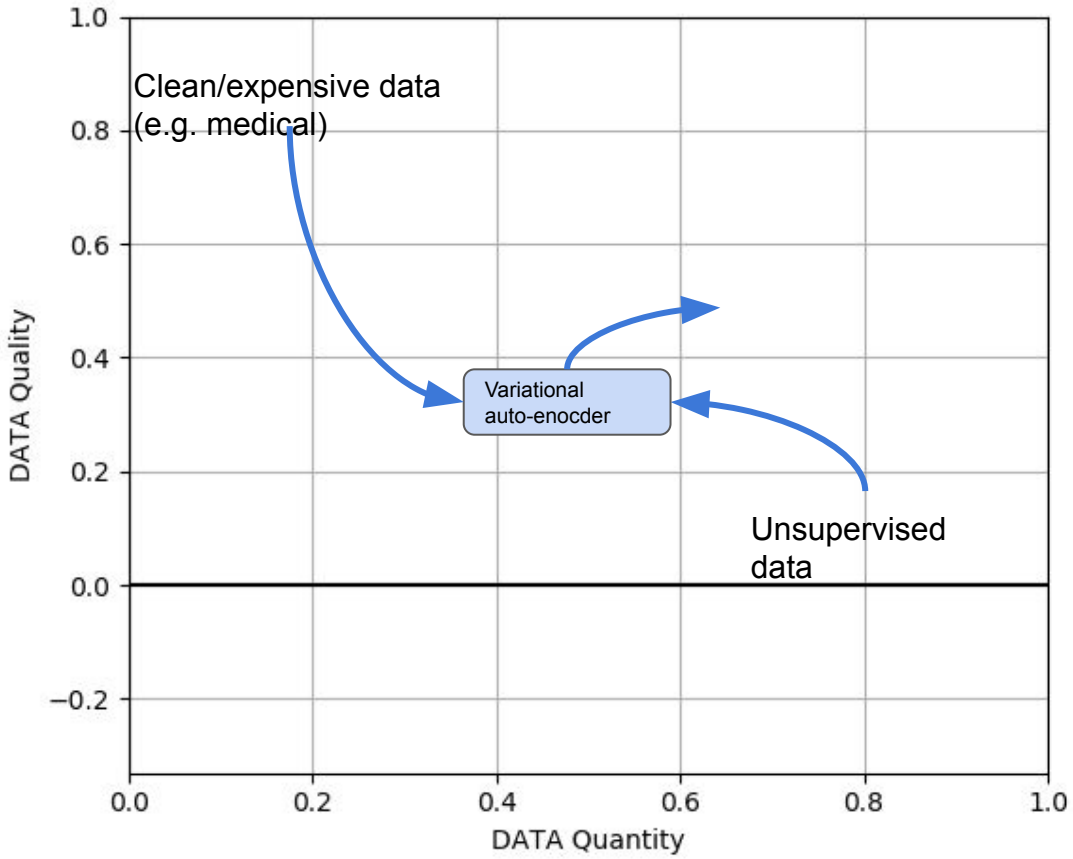
Regularisation of the training such that the latent space is fit for better generation



$$\text{loss} = ||x - \hat{x}||^2 + KL[N(\mu_x, \sigma_x), N(0, I)] = ||x - d(z)||^2 + KL[N(\mu_x, \sigma_x), N(0, I)]$$

[Joseph Rocca 2019]

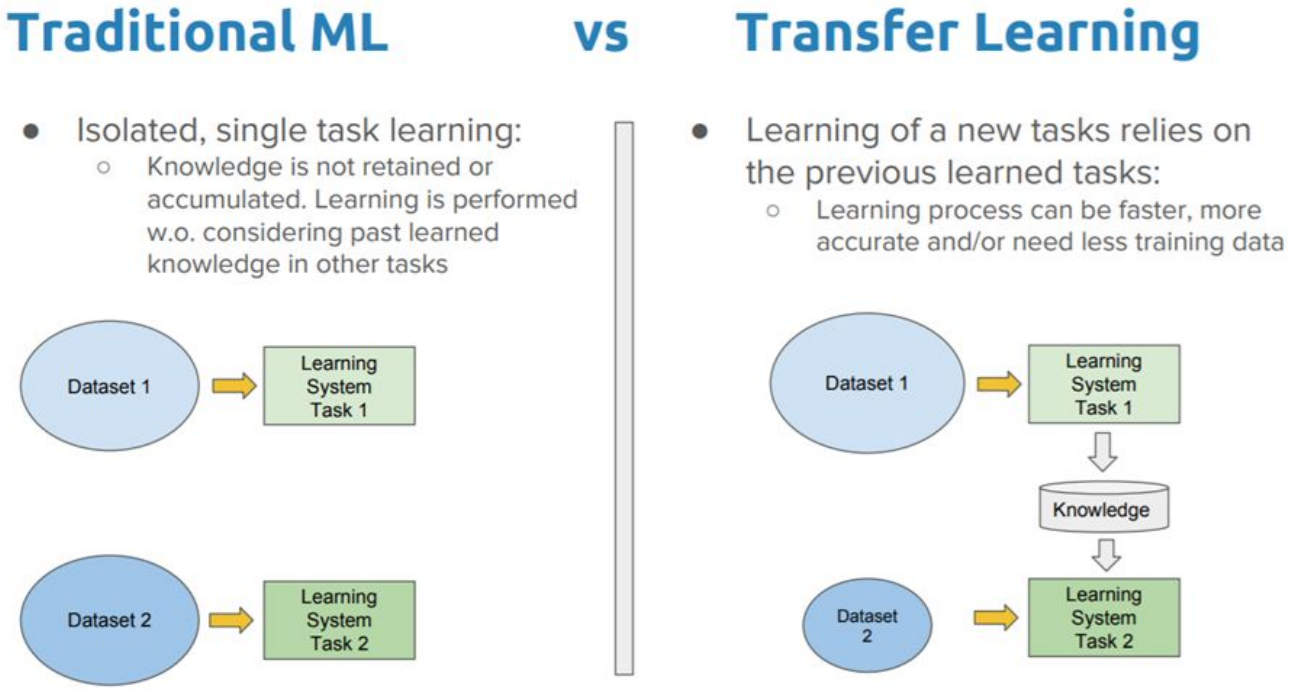
Quality of the data vs Quantity of data



Transfer learning

Transfer learning

General principle: to utilize knowledge acquired for one task to solve related ones

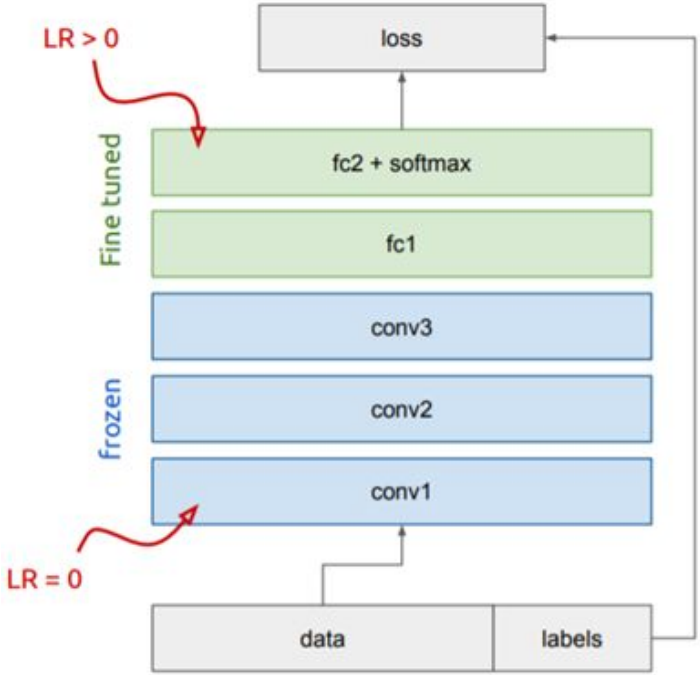


[DJ Sarkar, towardsdatascience.com, November 2018]

Transfer learning

Application to image segmentation using deep neural networks (DNN):

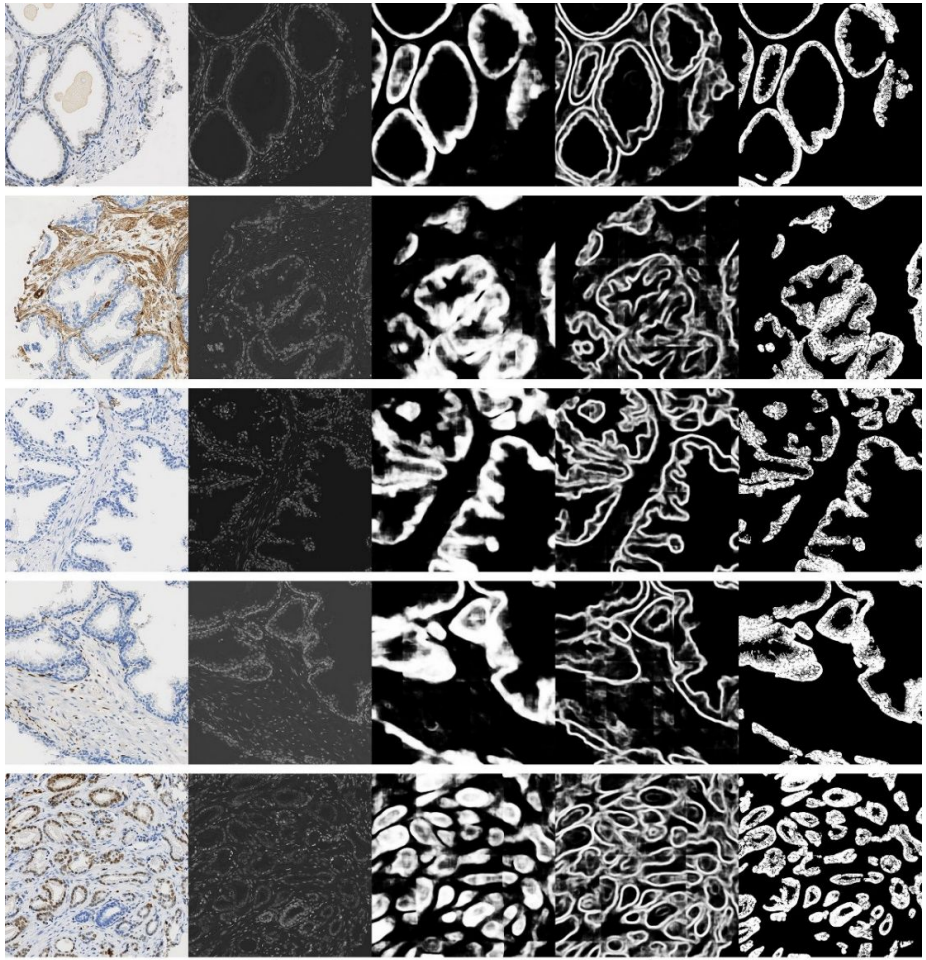
- the first layers = generic (low-level) feature extractor (edges, shapes, corners, colors, …)
- can be shared across tasks
- the last layers: task-specific should be fine-tuned on a new task
- OR: the whole NN can be fine-tuned if enough data are available (e.g. by using data augmentation and GAN)



Transfer learning

Example:

a network trained for **colonic gland** segmentation is then applied on **prostate tissue** and can be improved / fined-tuned with a minimum number of examples



[Van Eycke YR et al., Med Image Analysis 2018]

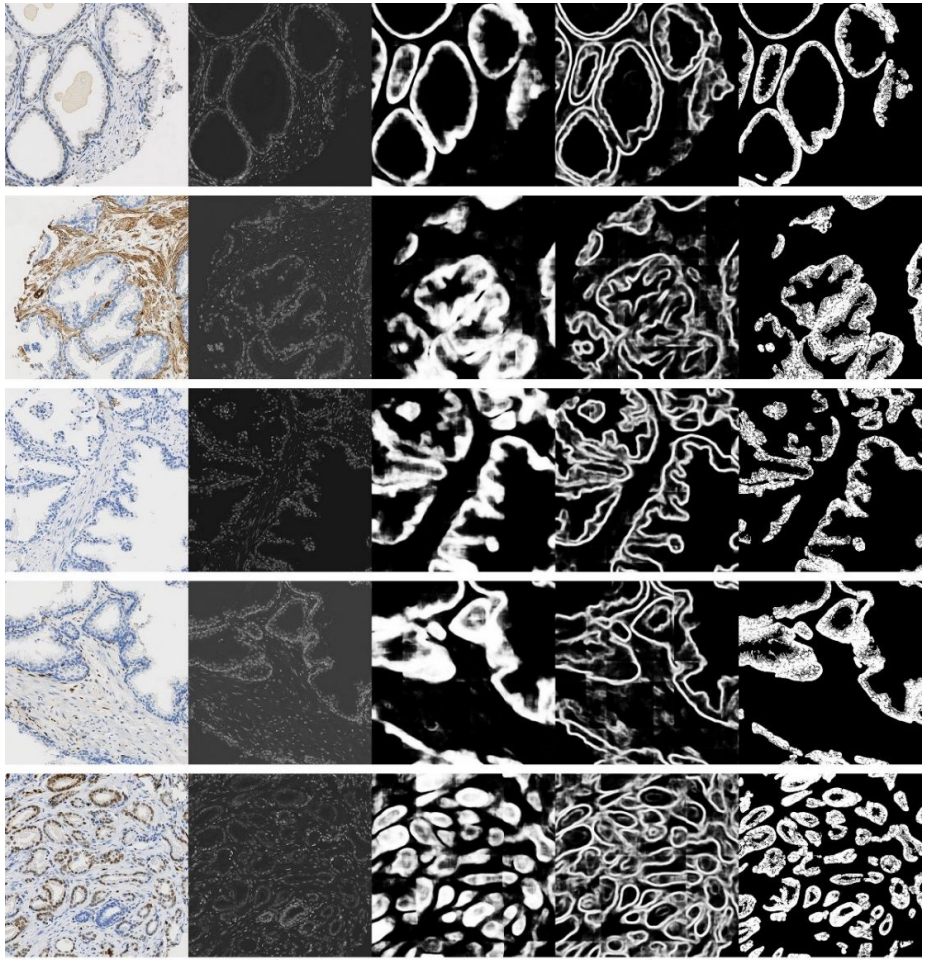
Transfer learning

Advantages:

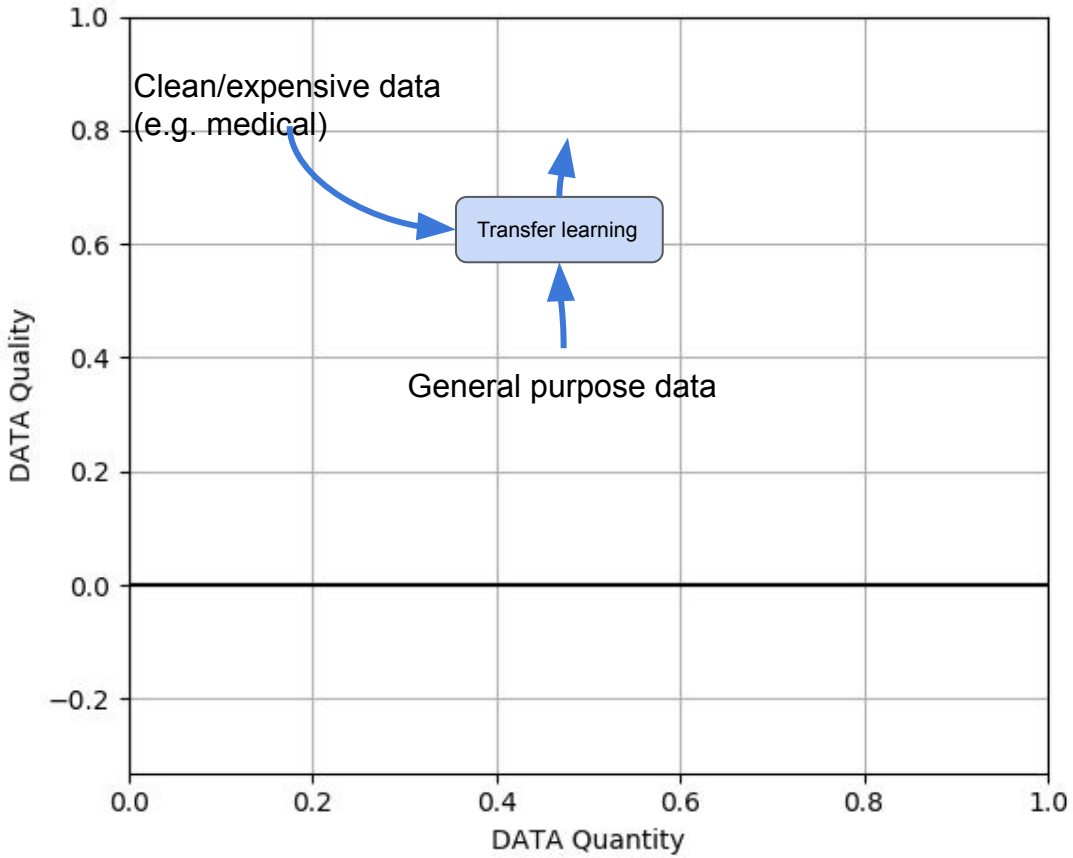
- public availability of pre-trained DNN (usually on large public databases of natural images) histological databases with supervision (e.g. via public challenges organized during conferences)

- reduce significantly both the training time and the number of examples specific to the final task needed to fine-tune the DNN

[Van Eycke YR et al., Med Image Analysis 2018]



Quality of the data vs Quantity of data

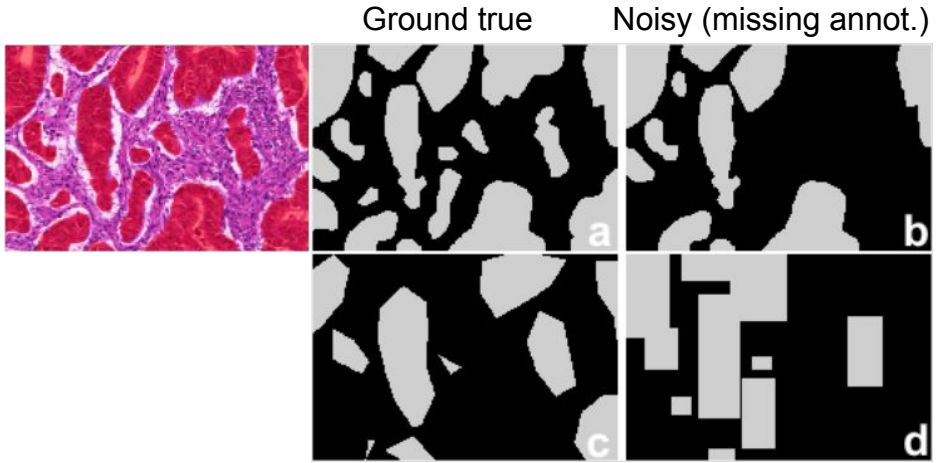


Weakly supervised learning

Weakly supervised learning

General principles:

- develop the ability to use imperfect or imprecise supervision
- See also multiple instances learning (MIL)



[Foucart A et al., ISBI 2019 Proceedings]

Weakly supervised learning

Related learning concepts :

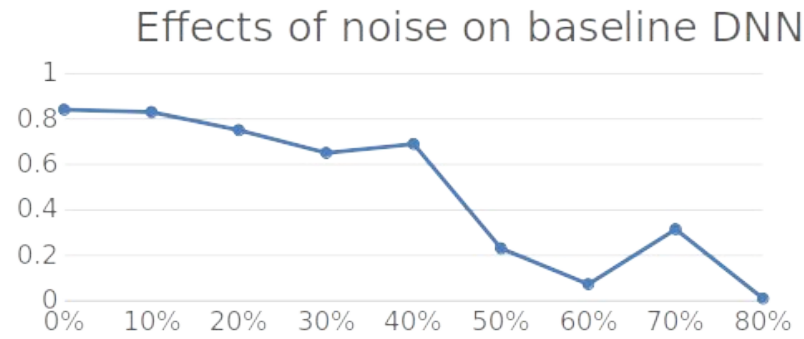
- unsupervised learning for datasets without any label: estimate data distribution properties, extraction of features explaining data (e.g., using auto-encoder)
- semi-supervised learning for datasets only partially supervised: combines unsupervised and supervised learning
- multiple instance learning: only bag of instances (e.g. image patches in place of pixels) are labeled, from which algorithms try to predict instance labels (such as pixel labels to segment images)
- learning with label noise: for datasets with label errors (e.g. unlabeled “positive” pixels)

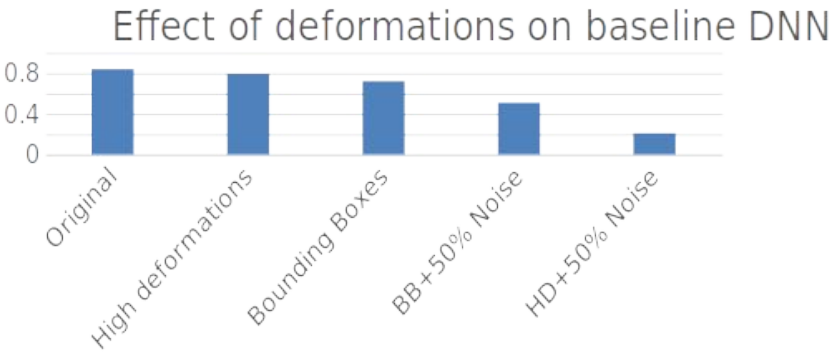
[Foucart A et al., ISBI 2019 Proceedings]

Weakly supervised learning

Impact of imperfect, imprecise and noisy (erroneous) annotations on a baseline DNN (per-pixel F1 score):

- Robustness to some amount of label noise and deformation or inaccuracy of contours





[Foucart A et al., ISBI 2019 Proceedings]

Weakly supervised learning strategies

- Only positive
  - Keep only fully annotated tiles, here patches which contain at least 80 pixels belonging to the “gland” class
- Semi-Supervised
  - An auto-encoder is trained on the full dataset without supervision
  - The weights in the first residual units are used as the initialization for the training on the supervised set.
  - (+) trained with only the positive examples
- Weak
  - Imprecise supervision
  - adding a global pooling layer after the segmentation.
- Soft Weak
  - Imprecise pixel-level annotations, combined with the image-level labels, both in the cost function
- Noisy Soft Weak
  - The dataset is highly asymmetrical
  - treat unlabeled examples as both a positive and a negative by randomly choosing its label
- + Combinations of the previous strategies...

[Foucart A et al., ISBI 2019 Proceedings]

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

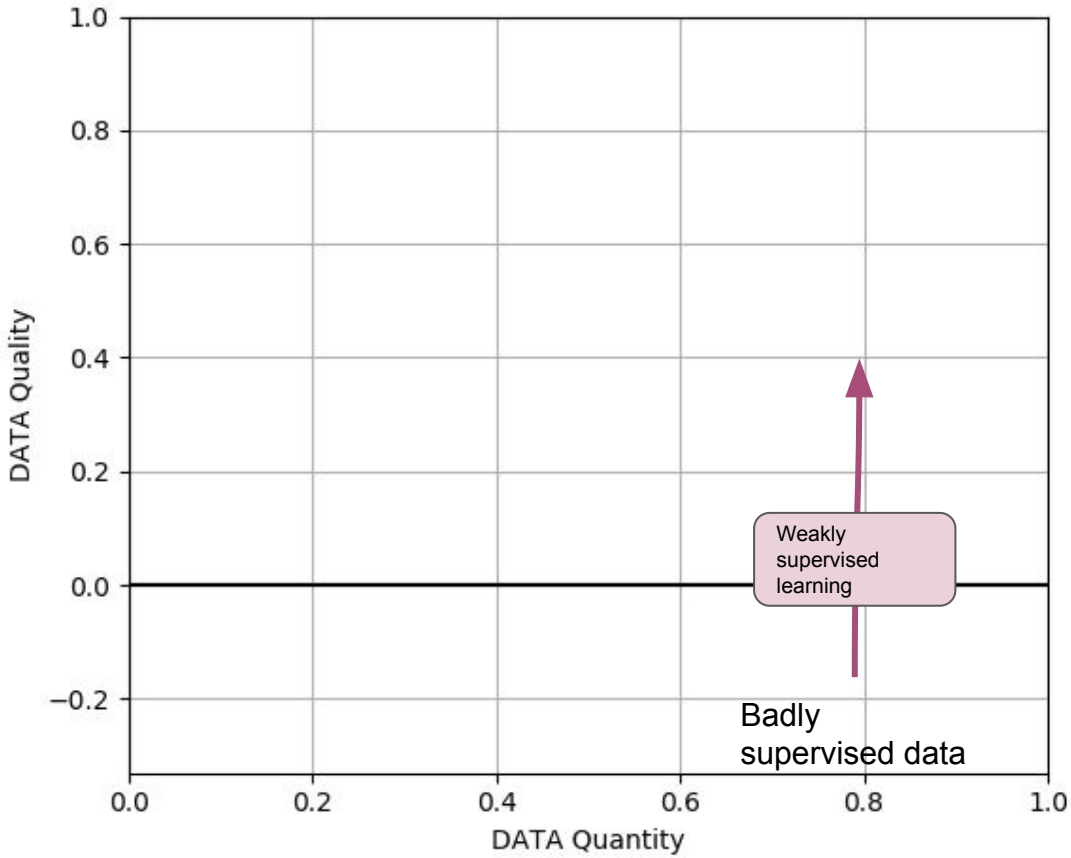
Weakly supervised learning

Recovery of accuracy on strongly degraded datasets with some learning strategies

Network	F_1 (mean on all test images) Noisy	BB	N+BB	N+HD	Stat. Score
Baseline	0.231	0.724	0.511	0.212	-21
Only positive	**0.768**	0.730	**0.697**	**0.660**	**14**
SS	0.467	**0.756**	0.522	0.207	-8
SS+	**0.729**	0.740	**0.730**	0.428	**12**
Weak	0.659	0.211	0.647	**0.648**	-8
SoftWeak	**0.724**	0.741	0.683	0.018	-4
Noisy SW	0.547	**0.756**	0.656	0.252	-1
SS+ SW	**0.735**	0.737	**0.711**	**0.671**	**15**
SS Noisy SW	0.592	**0.738**	0.613	0.364	0

[Foucart A et al., ISBI 2019 Proceedings]

Quality of the data vs Quantity of data



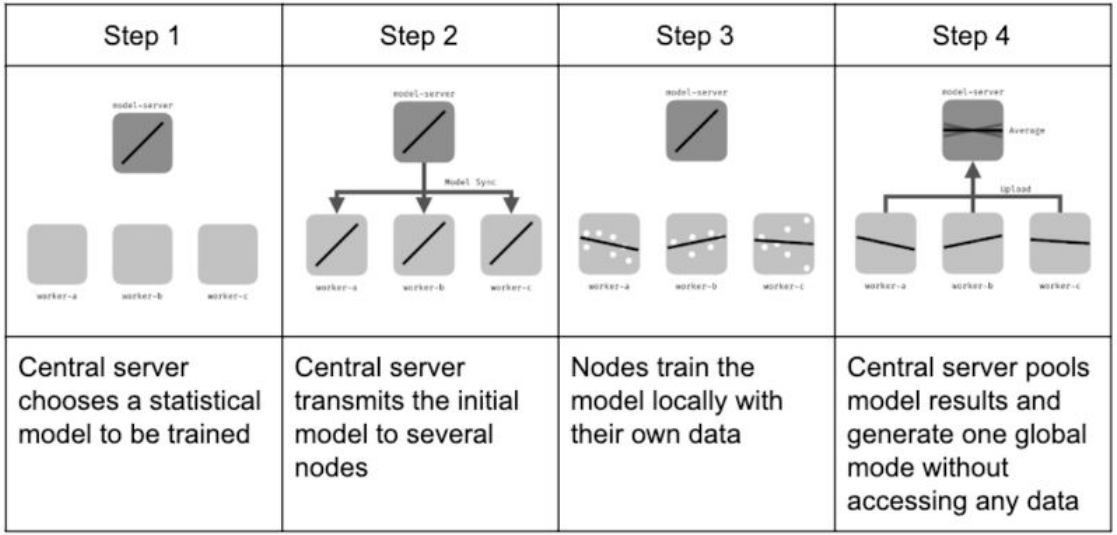
Federate learning

Federate learning

- Having numerous data may be difficult
  - Grouping data from several sources

But:

- Sharing data may be difficult
  - property
  - secret
  - GDPR



[Wikimedia]

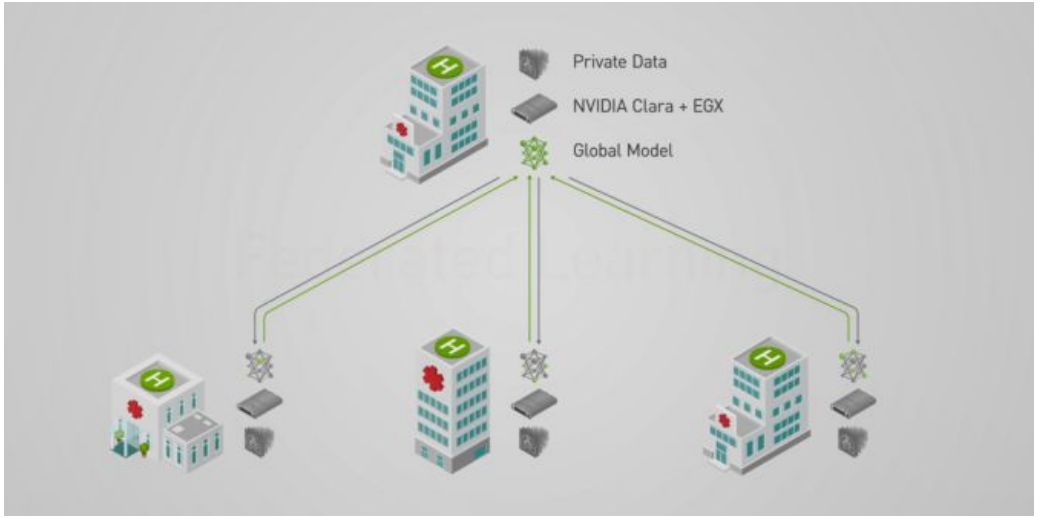
ULB

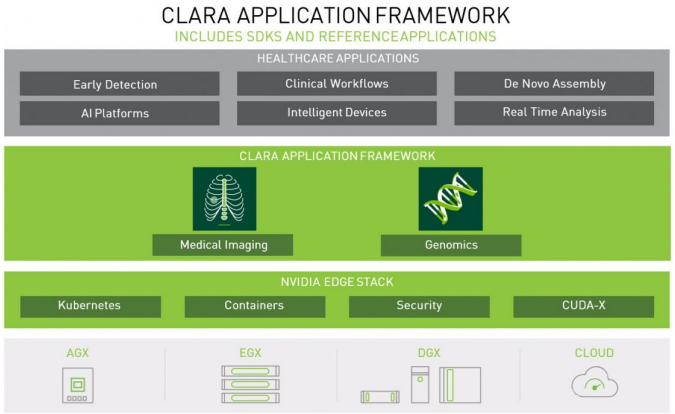
UNIVERSITÉ LIBRE DE BRUXELLES

Federate learning

- Federated learning: From Clara platform, train models by preserving data privacy
- Hyperfine: iot + deep learning. Portable Point of care MRI with included GPU to make real time predictions
- Clara Nvidia: Microservice applications to deliver Machine learning into healthcare environment for images and genomics

[NVidia.com]





Federate learning

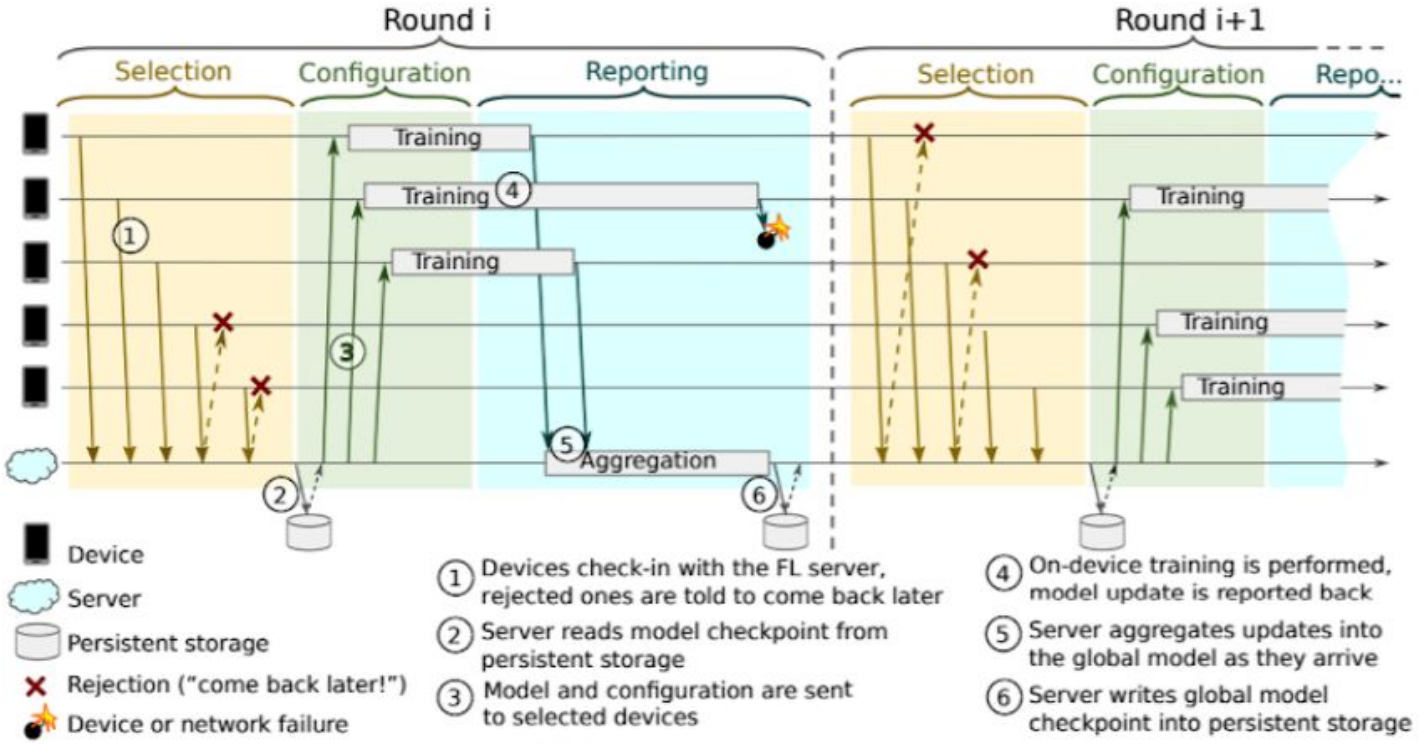
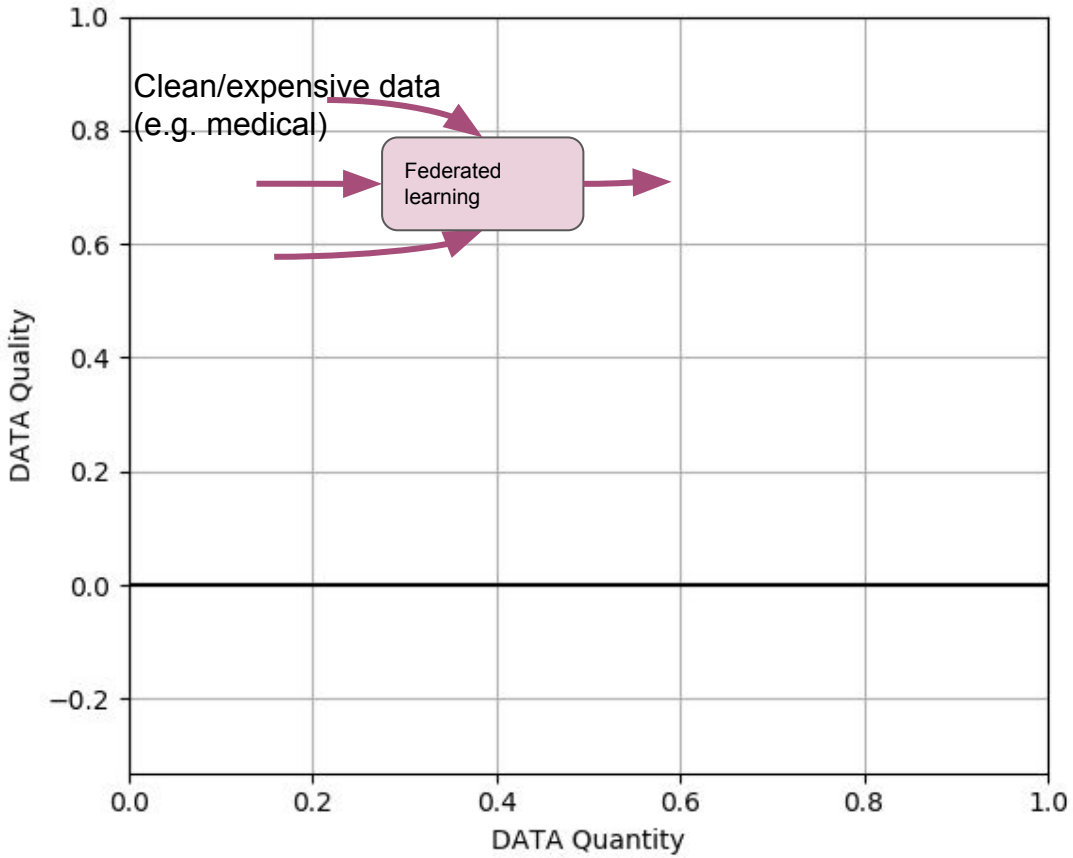


Figure 1: Federated Learning Protocol

[NVidia.com]

Quality of the data vs Quantity of data

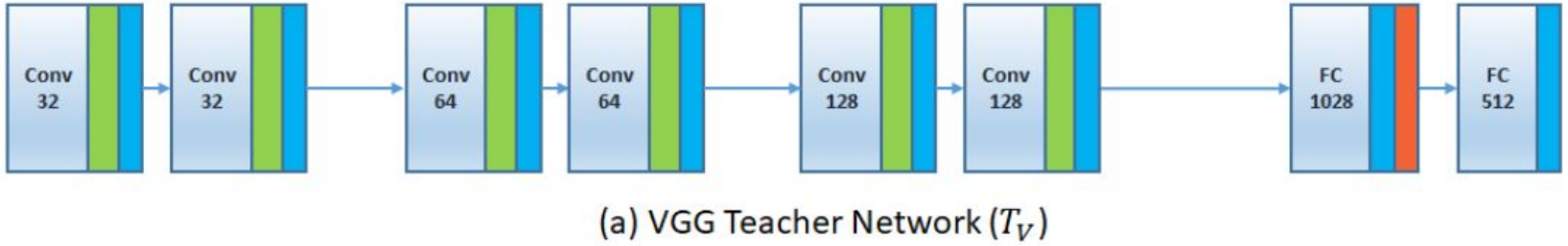


Teacher learning

Teacher learning

- Use a Clean data set to train a first network
- Predict a big unsupervised dataset bigger dataset with a confidence estimation
- Keep only the higher confidence data to train a second network
- Fine tune with transfer learning using clean dataset
- See also “Knowledge Distillation”

Teacher learning

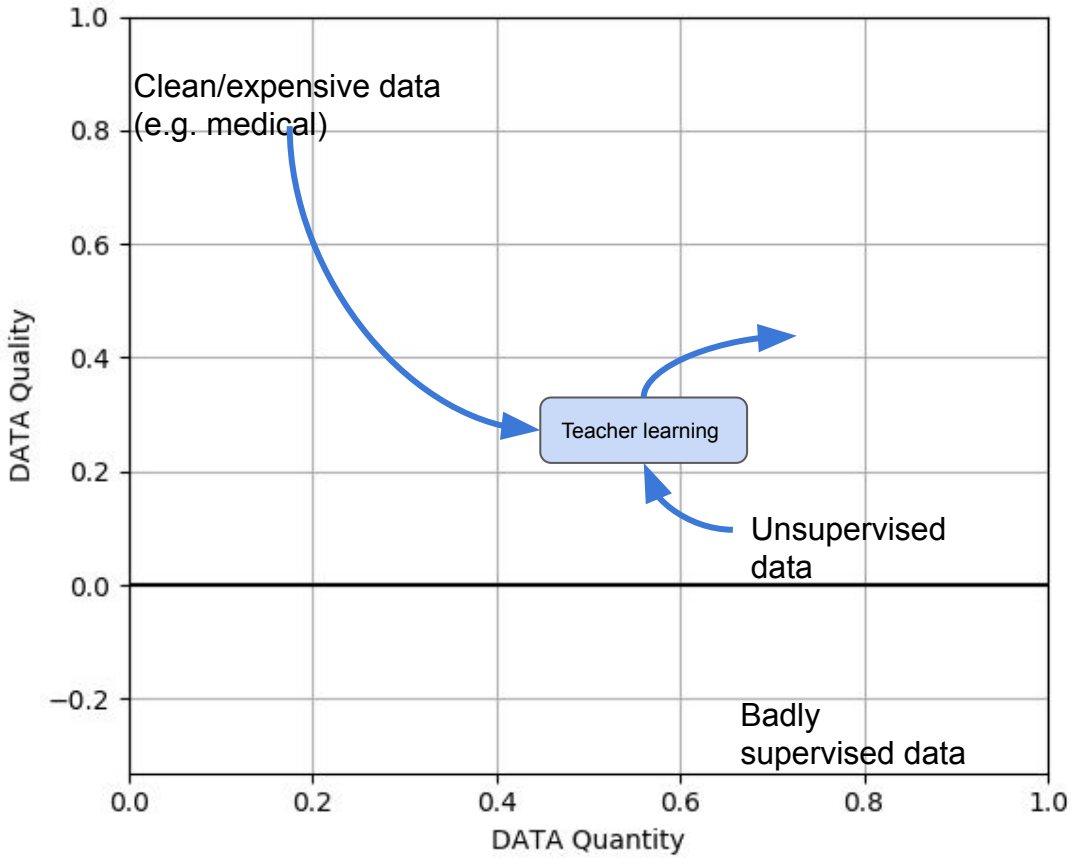


(a) VGG Teacher Network (T_V)

(b) VGG Student Network (S_V)

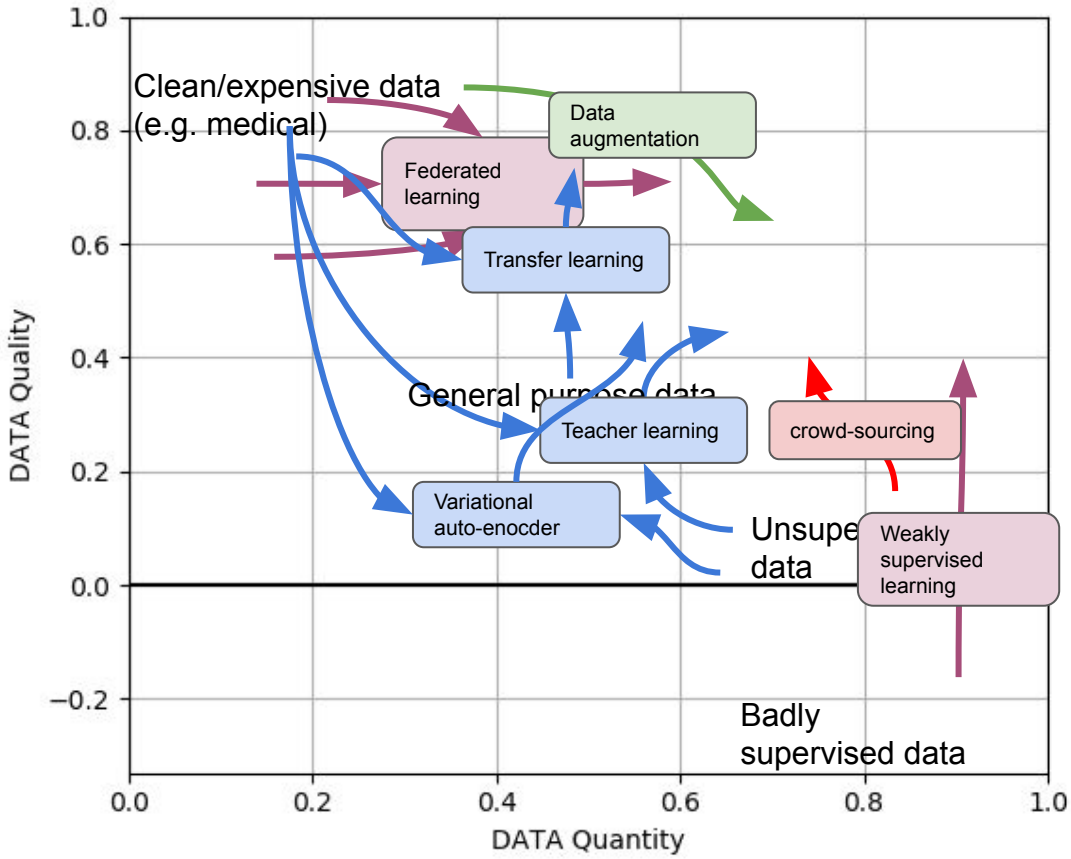
[Vaseli, H.,et al. In Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling (Vol. 10951, p. 109510F). International Society for Optics and Photonics.]

Quality of the data vs Quantity of data



Conclusion

Conclusion


Clean/expensive data
(e.g. medical)
Data augmentation
Federated learning
Transfer learning
General purpose data
Teacher learning
crowd-sourcing
Variational auto-enocder
Unsupe
data
Weakly supervised learning
Badly supervised data
DATA Quality
DATA Quantity
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0.0
0.2
0.4
0.6
0.8
1.0

Conclusion

Clean your data

Do data augmentation

Be careful, the performance slope is not always monotonic, more data does not mean always better performances

Be critical, avoid the silver bullet anti-pattern

Initialization	mIOU
ImageNet	73.6
300M	75.3
ImageNet+300M	**76.5**

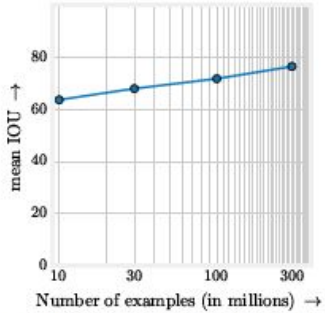


Figure 6. Semantic segmentation performance on Pascal VOC 2012 val set. (left) Quantitative performance of different initializations; (right) Impact of data size on performance.

[Chen Sun, et al.2017 arXiv:1707.02968

]

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

References

- https://towardsdatascience.com/review-senet-squeeze-and-excitation-network-winner-of-ilsvrc-2017-image-classification-a887b98b2883
- https://www.karlrupp.net/2013/06/cpu-gpu-and-mic-hardware-characteristics-over-time/
- J. Deng, W. Dong, R. Socher, L. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, 2009, pp. 248-255.
- A. Krizhevsky et al., ImageNet Classification with Deep Convolutional Neural Networks. In NIPS, 2012
- https://deepsense.ai/human-log-loss-for-image-classification/
- https://github.com/BertMoons/Comparing-CNN-Architectures
- DJ Sarkar, towardsdatascience.com, November 2018
- Hou L et al. CVPR 2019
- Foucart A et al., ISBI 2019 Proceedings
- Van Eycke YR et al., Med Image Analysis 2018
- Debeir et al. ISBI 2019 Proceedings
- https://github.com/facebookresearch/Detectron
- https://vas3k.com/blog/machine_learning/
- https://www.johner-institute.com/articles/software-iec-62304/artificial-intelligence/
- https://stanfordmlgroup.github.io/projects/chexnext/
- https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73 (Joseph Rocca 2019)
- Bin Kong et al. MICAAI 2018
- Vaseli, H.,et al. In Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling (Vol. 10951, p. 109510F). International Society for Optics and Photonics.
- Chen Sun, et al.2017 arXiv:1707.02968