TOMOGRAPHY AND DEEP LEARNING

- Giovanni Chierchia, ESIEE Paris, CNRS, Université Paris-Est
- Hugues Talbot, CentraleSupélec, Inria, Université Paris-Saclay

ESIEE PARIS

ICTMS 2019

22-26 July 2019 | Cairns, Australia

UNIVERSITÉ PARIS-EST

CentraleSupélec

Inria informatics mathematics

université PARIS-SACLAY

OUTLINE

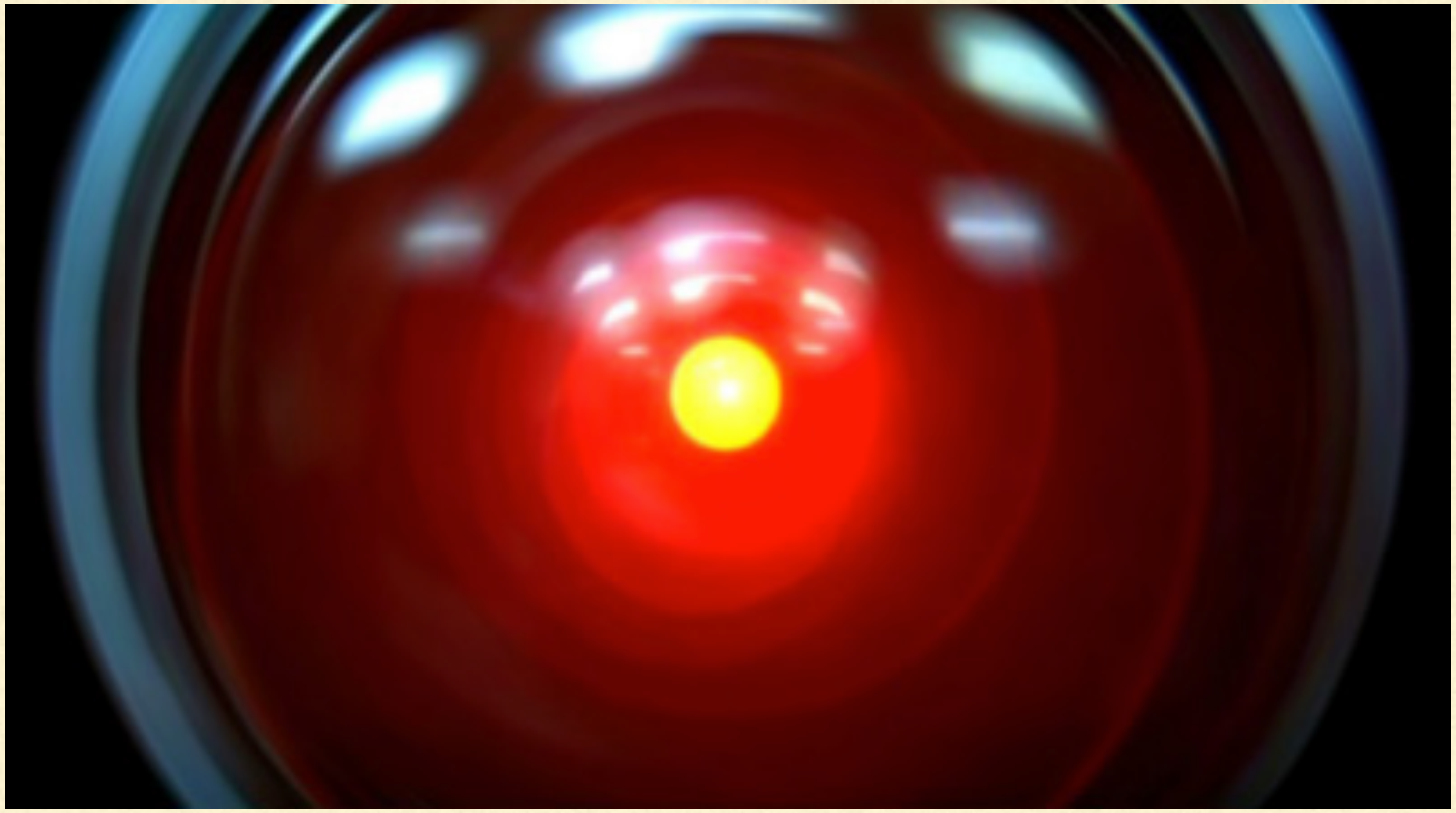
1. Introduction to machine learning
    - MV vs IA ; Regression / Classification
    - Deep Learning
2. Inverse problems and tomography
3. Using DL tools for Tomography
    - Projection and Back-projection operators on GPU using DL frameworks
    - Regularisation, sparse reconstruction, denoising, segmentation using DL frameworks.

CentraleSupélec

INTRODUCTION TO MACHINE LEARNING:

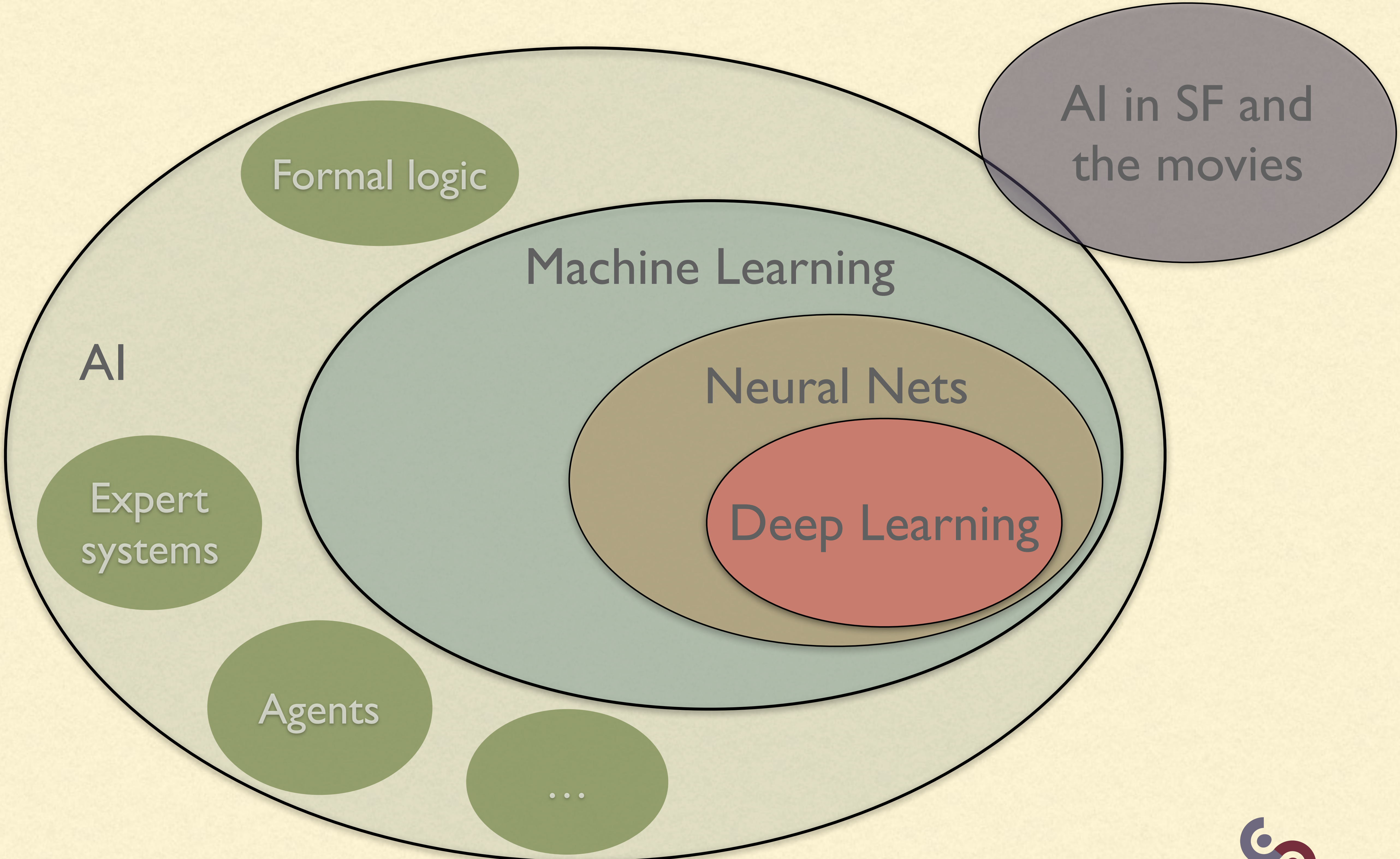
WHAT IS DEEP LEARNING?

FIRST: WHAT IS ARTIFICIAL INTELLIGENCE?

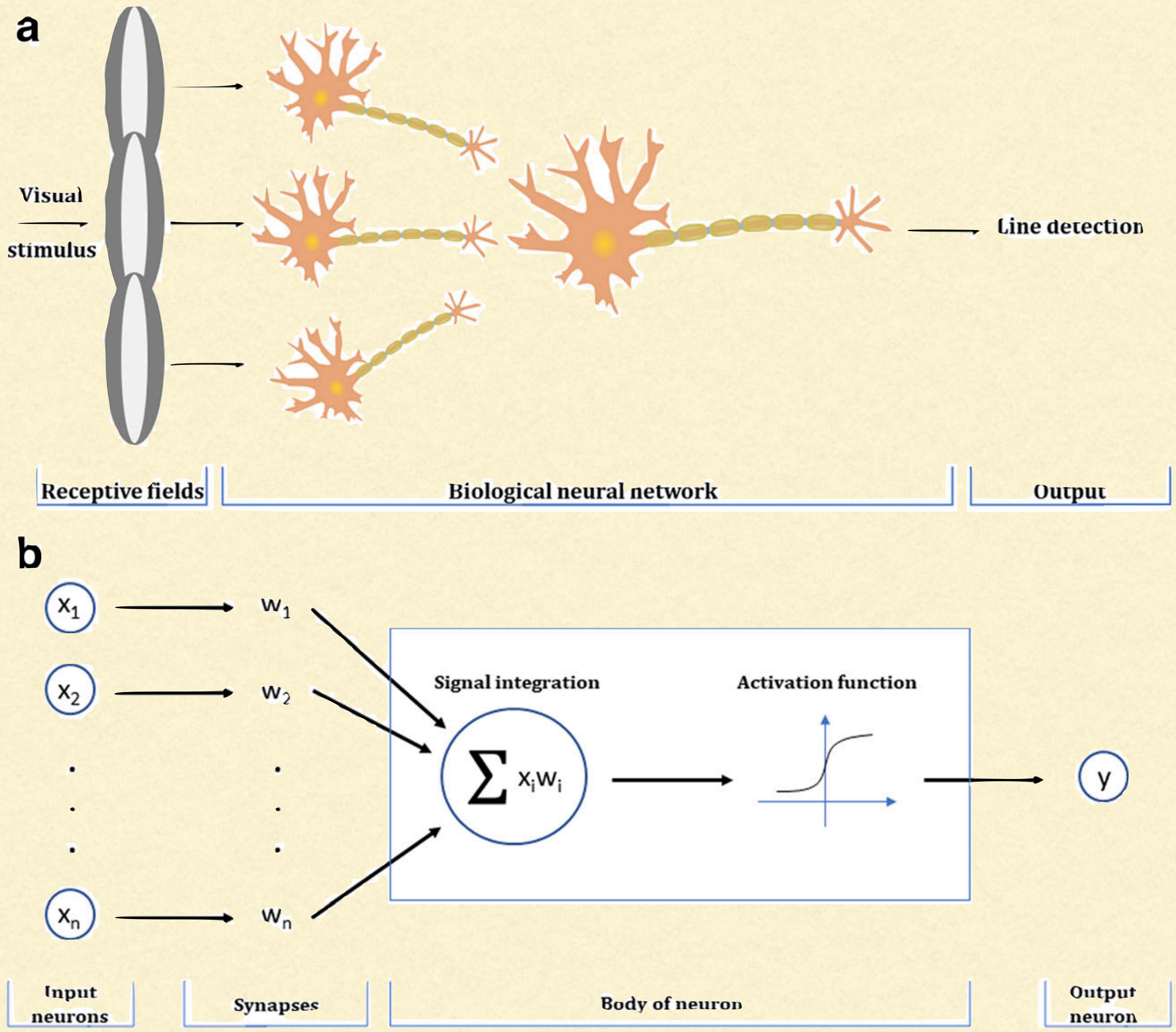


- AI is not like HAL in the "2001 A Space Odyssey" movie.

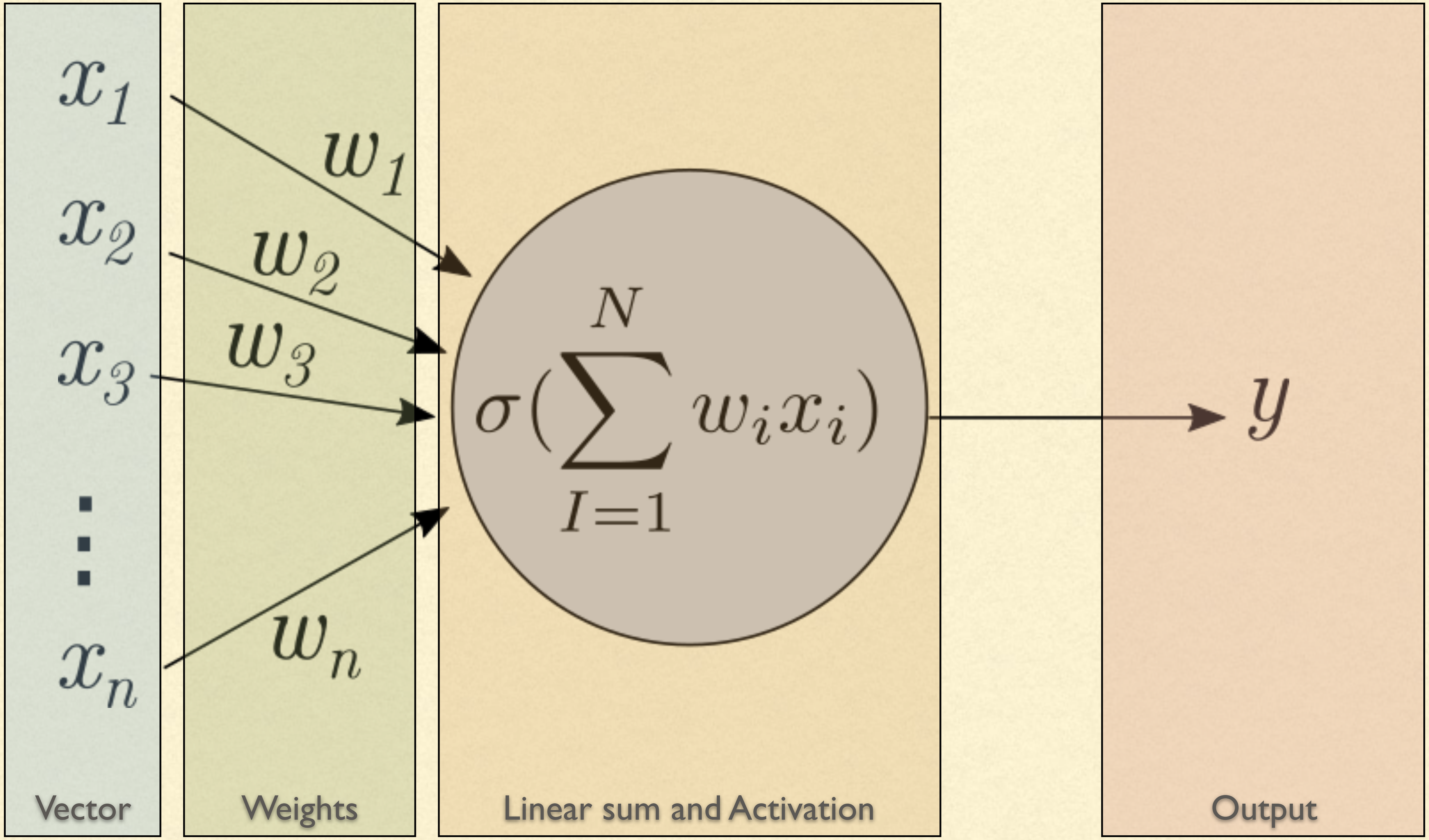
THE AI RUSSIAN DOLLS



NEURAL NETWORKS ARE BIOLOGICALLY-INSPIRED

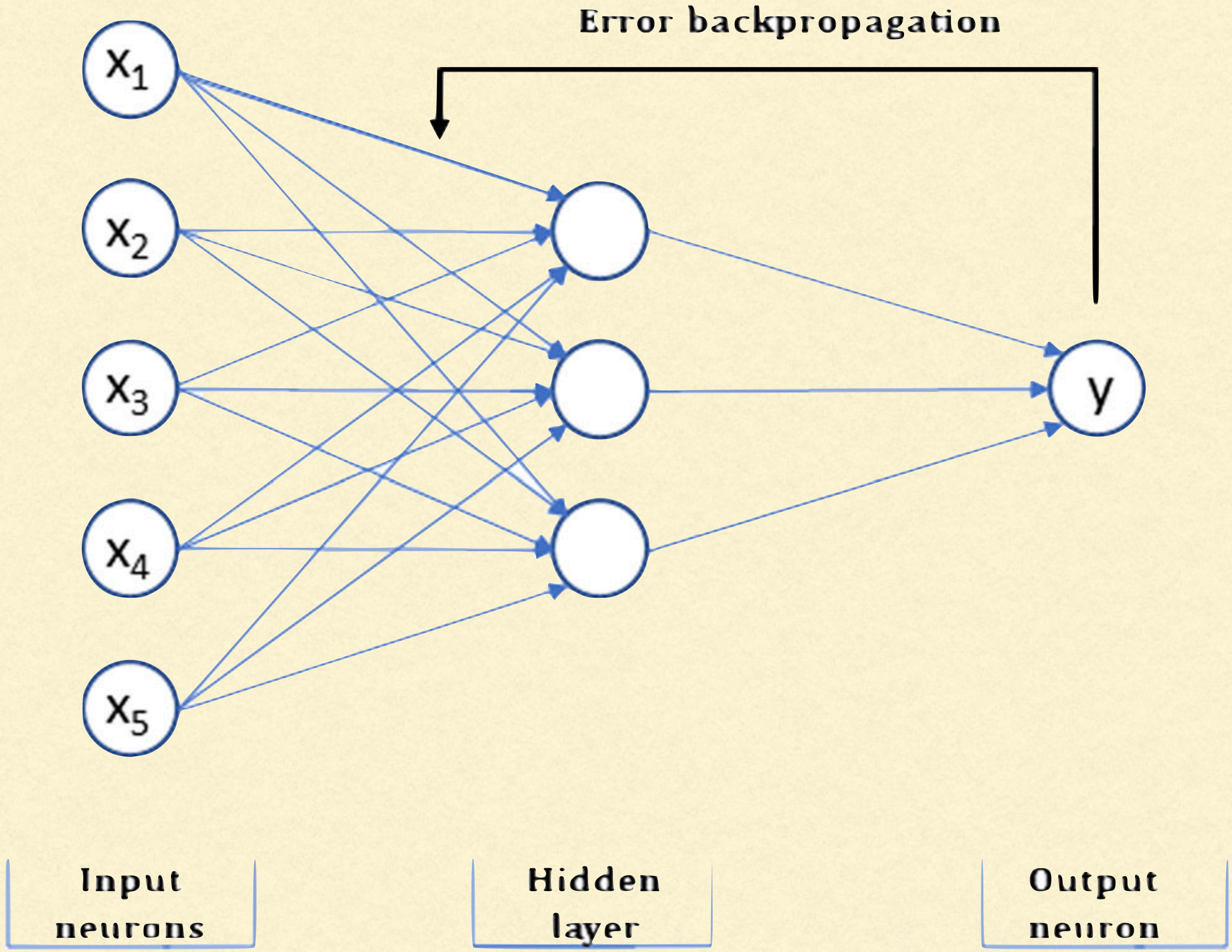


WHAT ARE NEURAL NETS ?



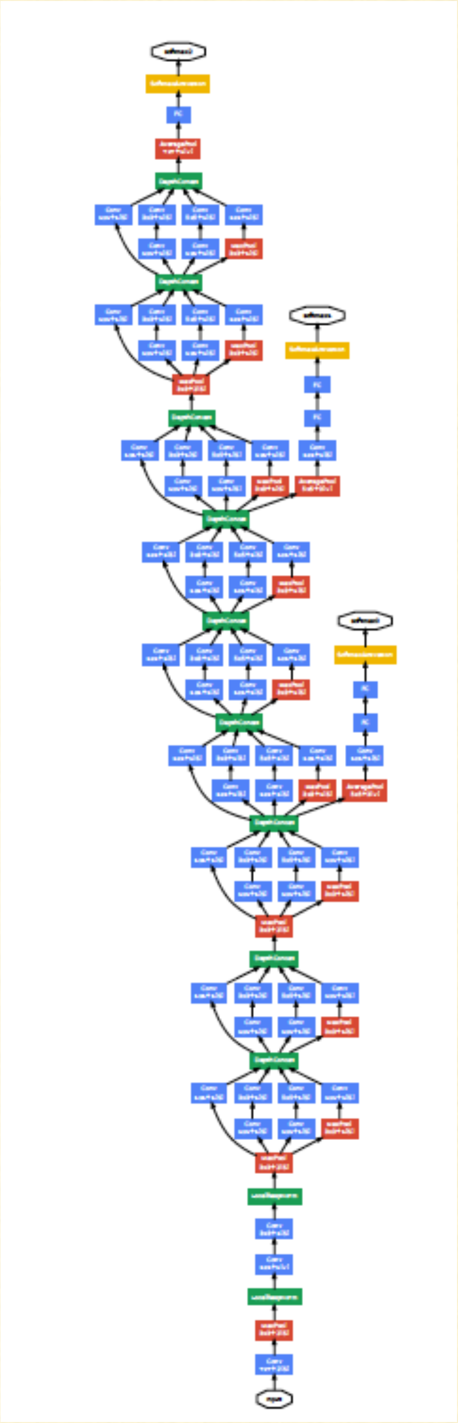
- This is a single artificial neuron : the Perceptron, exactly identical to *logistic regression*, which is a classification method.

SHALLOW NN: MULTI-LAYER PERCEPTRON

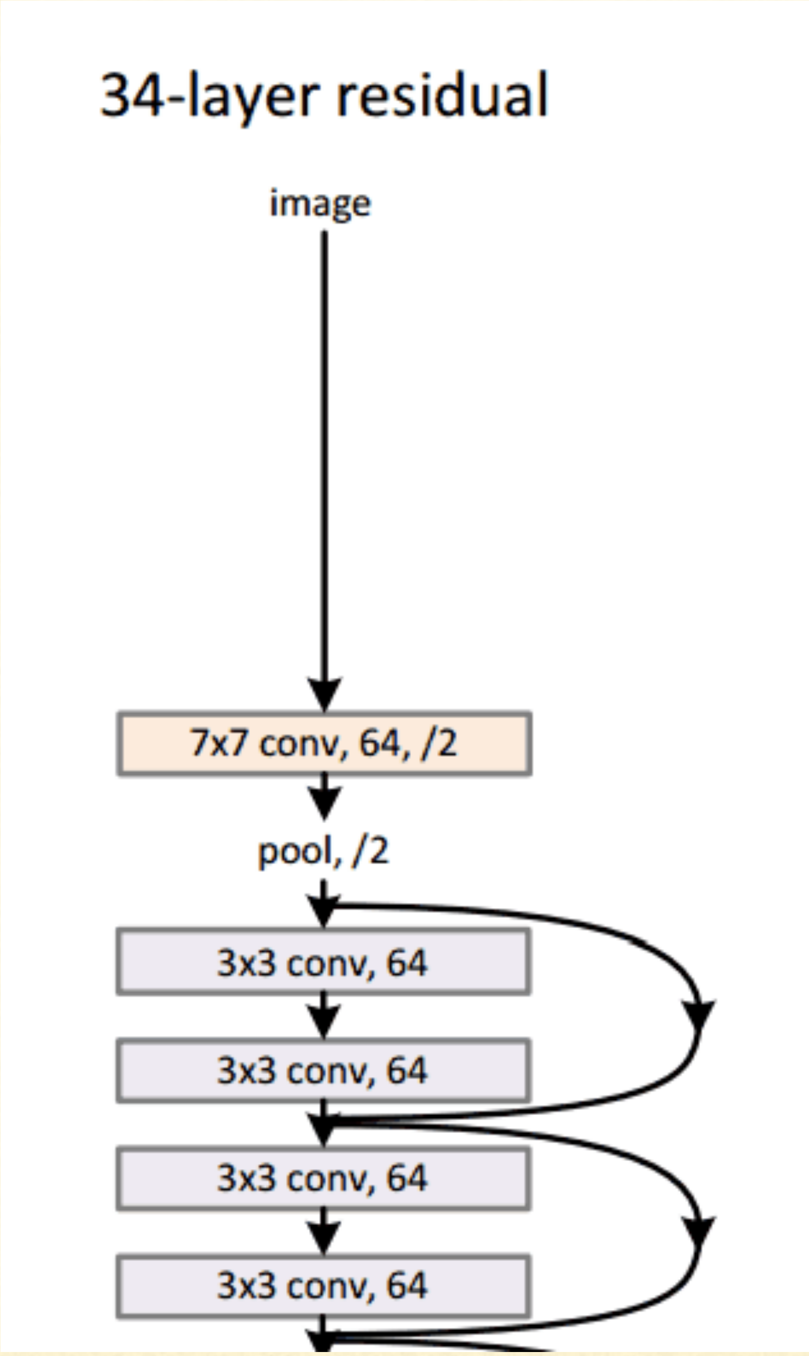


DEEP NN: MANY, MORE VARIED LAYERS

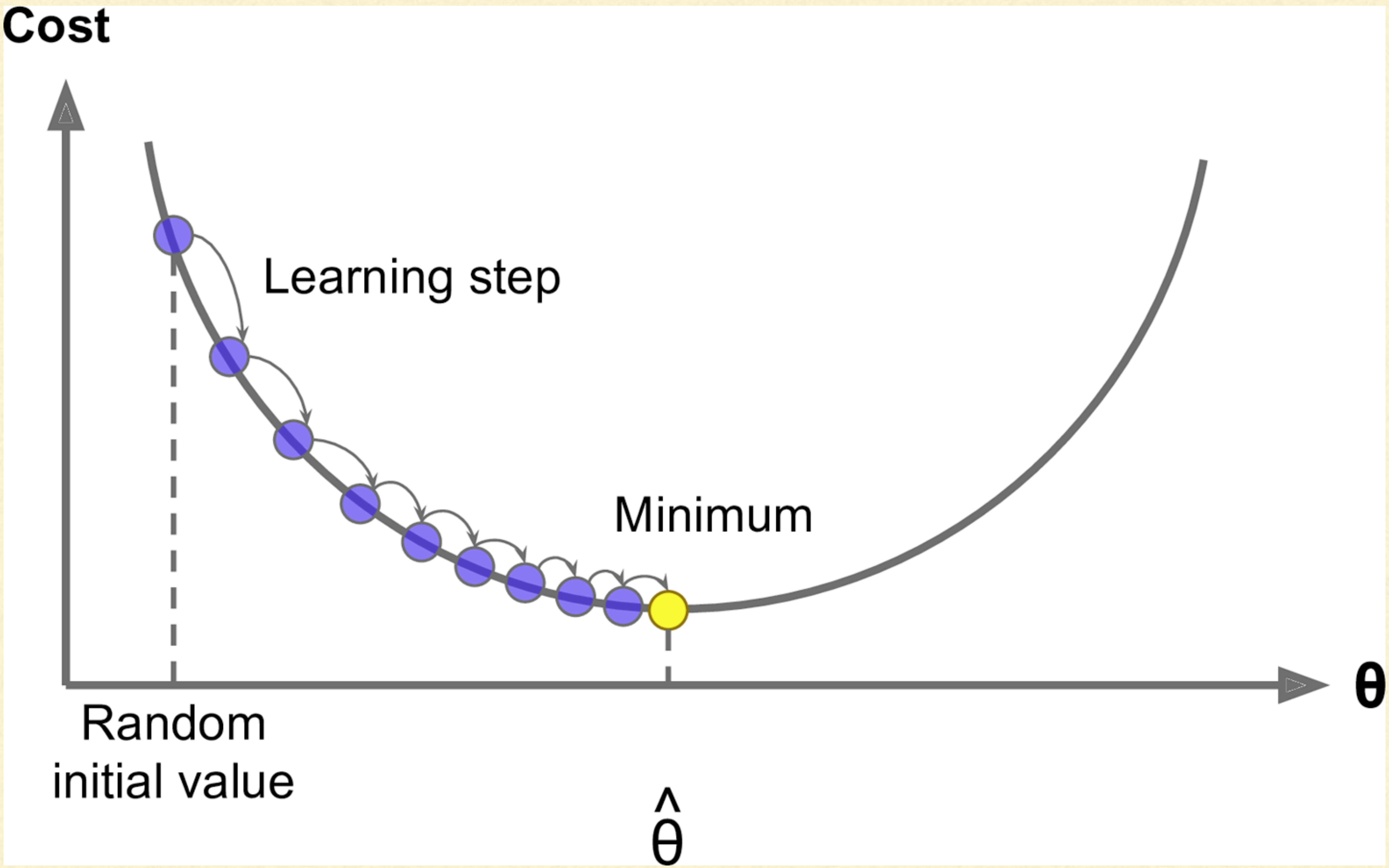
GoogLeNet
Szegedy et al
2015



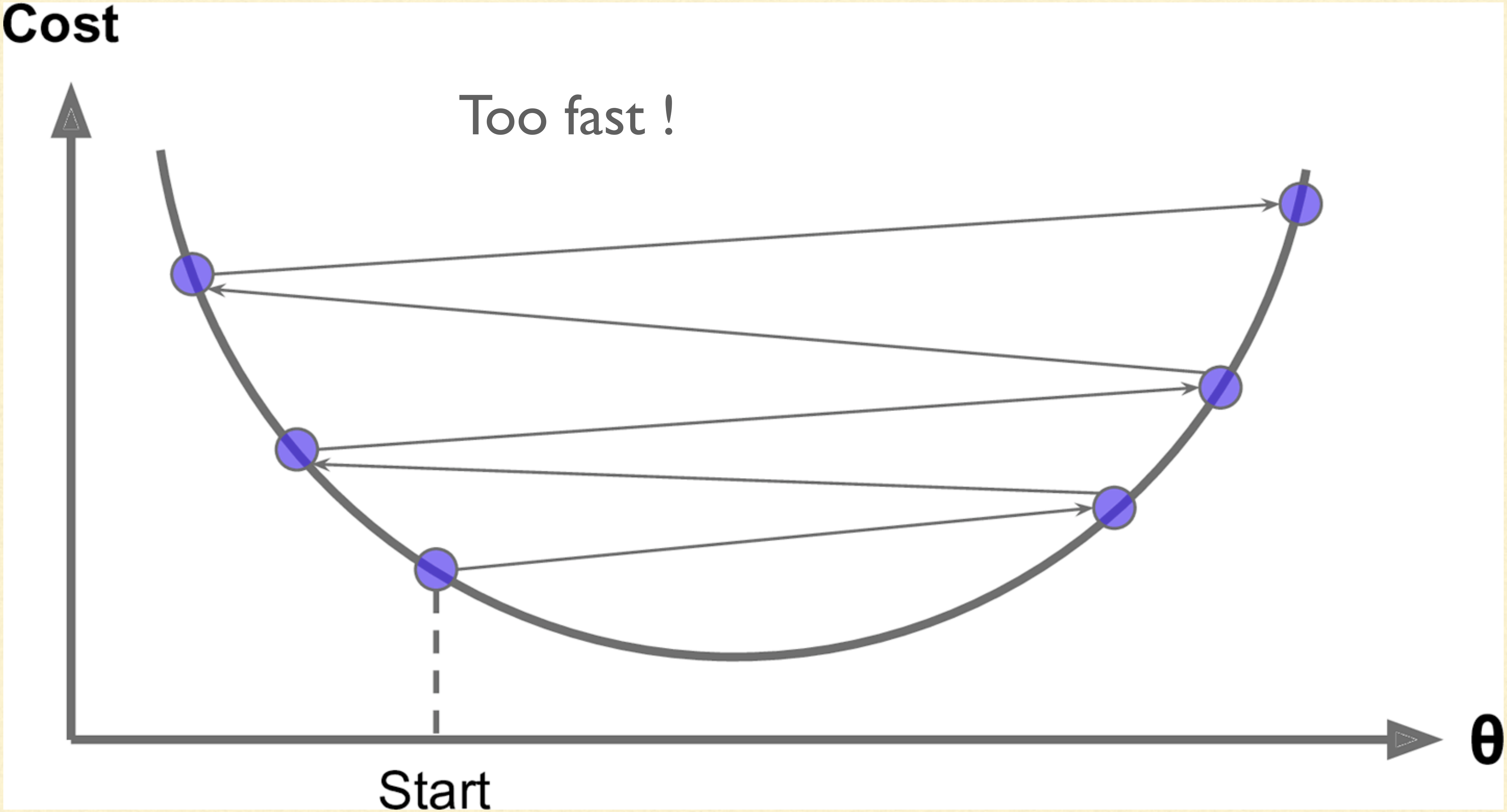
Microsoft ResNet
He et al 2016



OPTIMISATION TOOLS: SIMPLE CASE

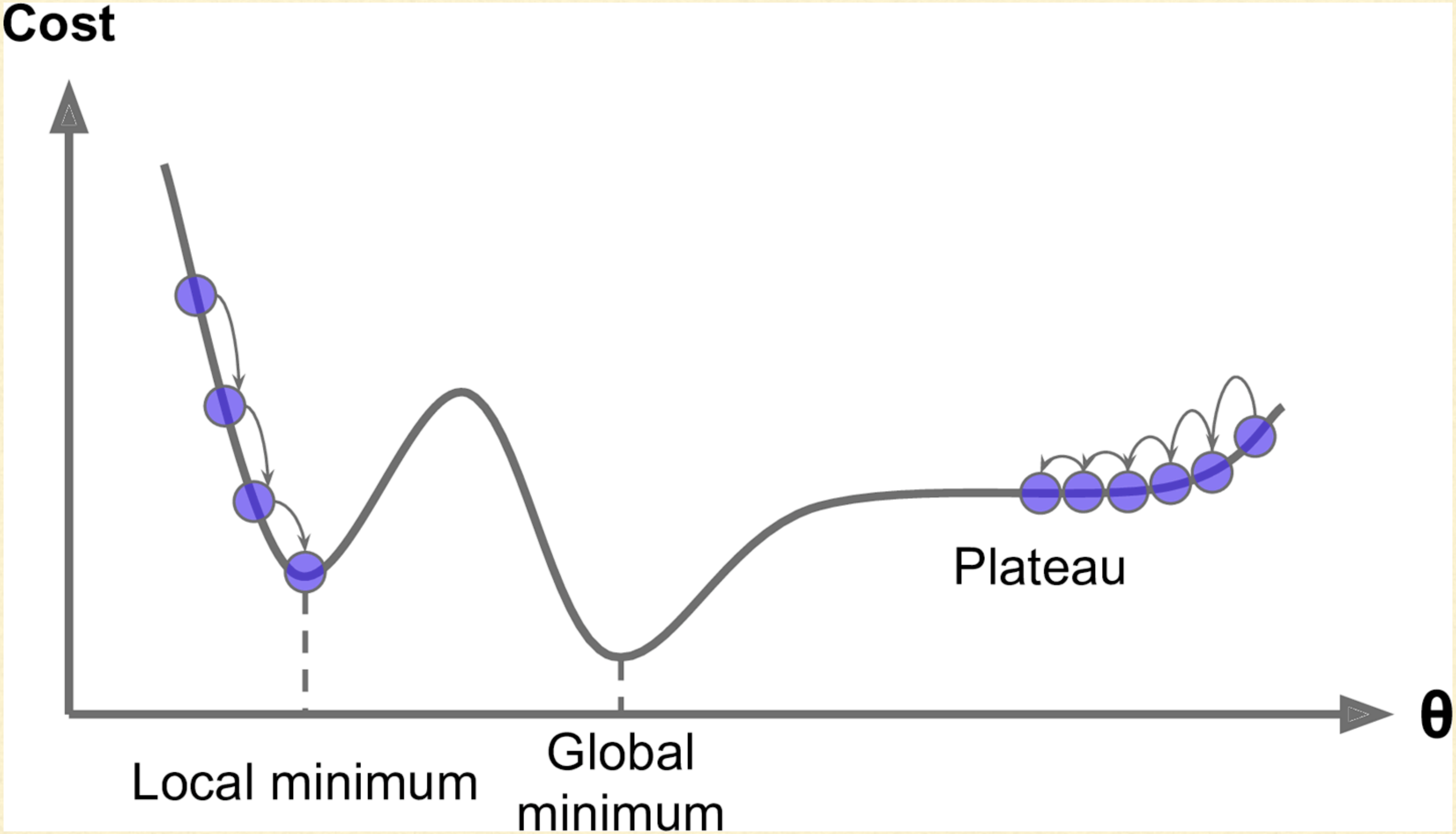


OPTIMISATION: SIMPLE CASE NOT SO SIMPLE

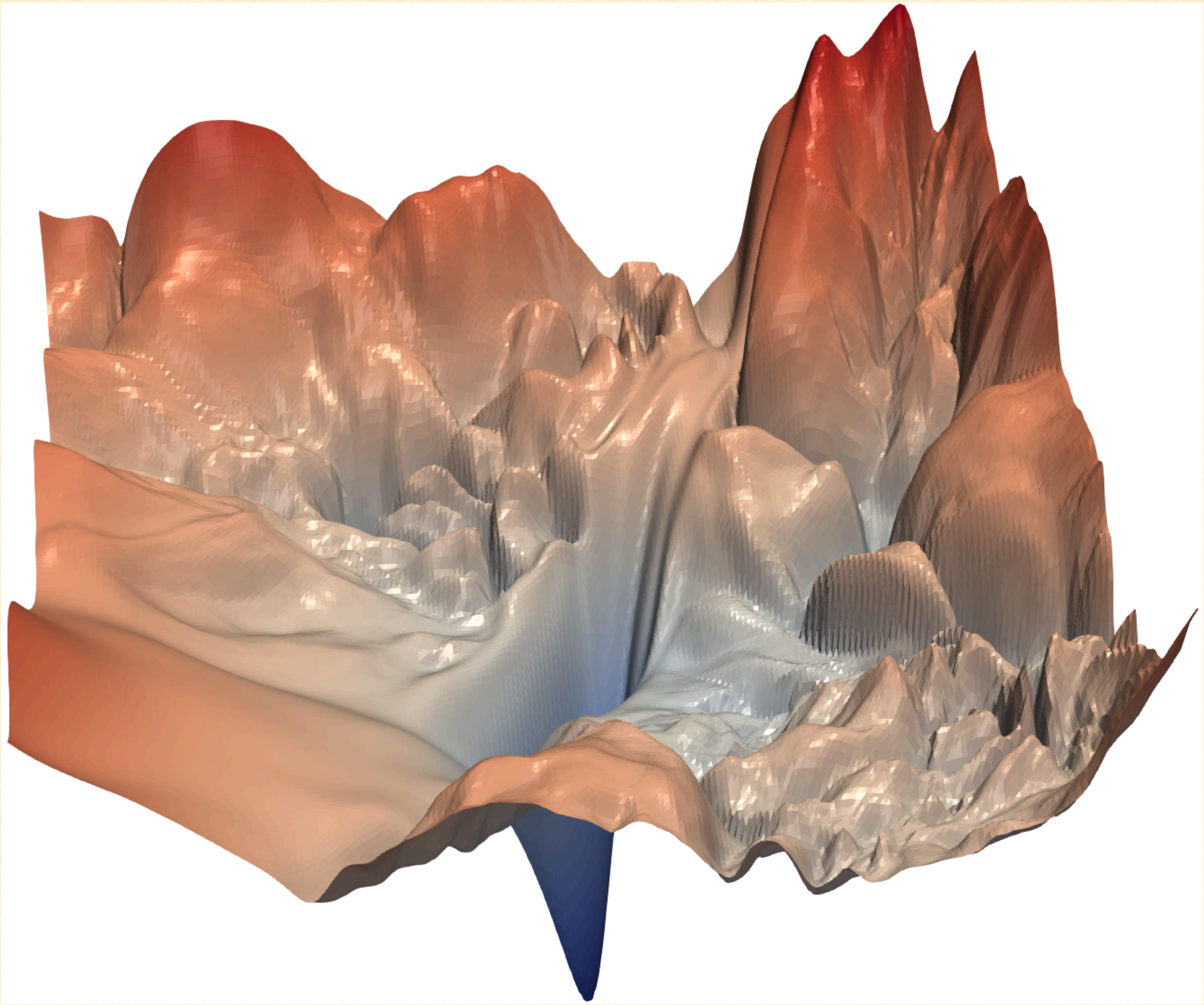


CentraleSupélec

OPTIMISATION: LESS SIMPLE CASE



OPTIMISATION: LESS SIMPLE CASE



Real surface of a ResNet-56 trained on CIFAR-10 [Li et al., 2018]

RELEVANT ELEMENTS OF DL FOR TOMOGRAPHY

- Deep learning composes:
  - Linear and non-linear computation layers;
  - with advanced optimisation methods;
  - with computations distributed on GPUs;
  - easily available, well maintained, open-source, packages.
- Questions: can we formulate tomography reconstruction in with DL packages, and is it interesting, useful, etc?
- Can we bring actual learning into the picture ?

TOMOGRAPHY POST-PROCESSING

DL CAN BE USED FOR DENOISING AND SEGMENTATION

- DL typically requires many *labeled* examples for segmentation, as well as many hours of experimentations, hyper parameter-tuning, etc[1].

- However, for denoising, this is relatively easy: provide a number of images of clean output (many photons) representative of your data, use a denoising auto-encoder architecture, and train with artificially noisy input.

- The network can learn from few images. It can be resolution-independent.

[1] You need a graduate student for this…

EXAMPLE DENOISING ARCHITECTURE

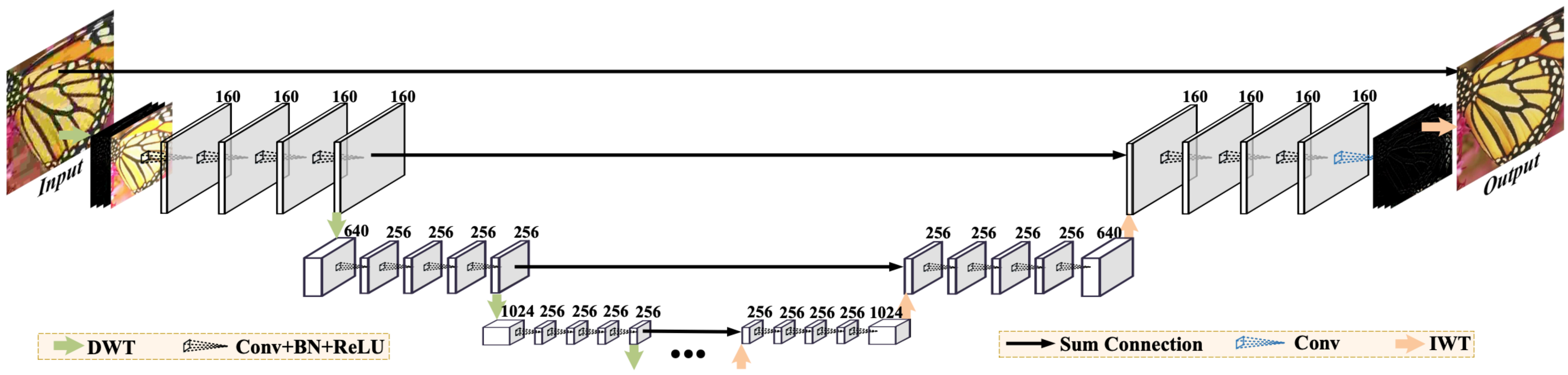
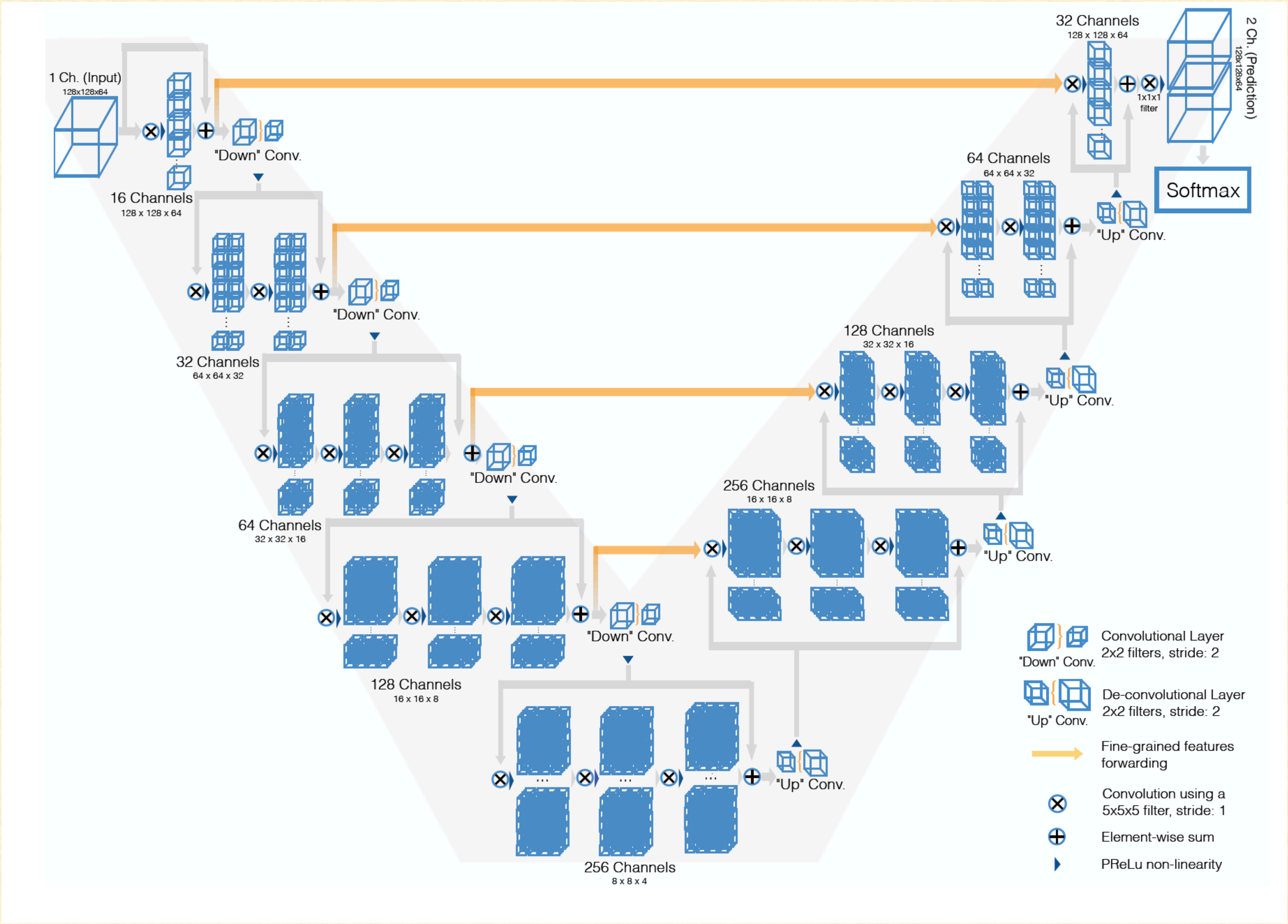


Figure 3. Multi-level wavelet-CNN architecture. It consists two parts: the contracting and expanding subnetworks. Each solid box corresponds to a multi-channel feature map. And the number of channels is annotated on the top of the box. The network depth is 24. Moreover, our MWCNN can be further extended to higher level (e.g., ≥ 4) by duplicating the configuration of the 3rd level subnetwork.

- This is a multi-resolution denoising auto-encoder with wavelet features
- Reproducible research: https://github.com/wenbihan/reproducible-image-denoising-state-of-the-art

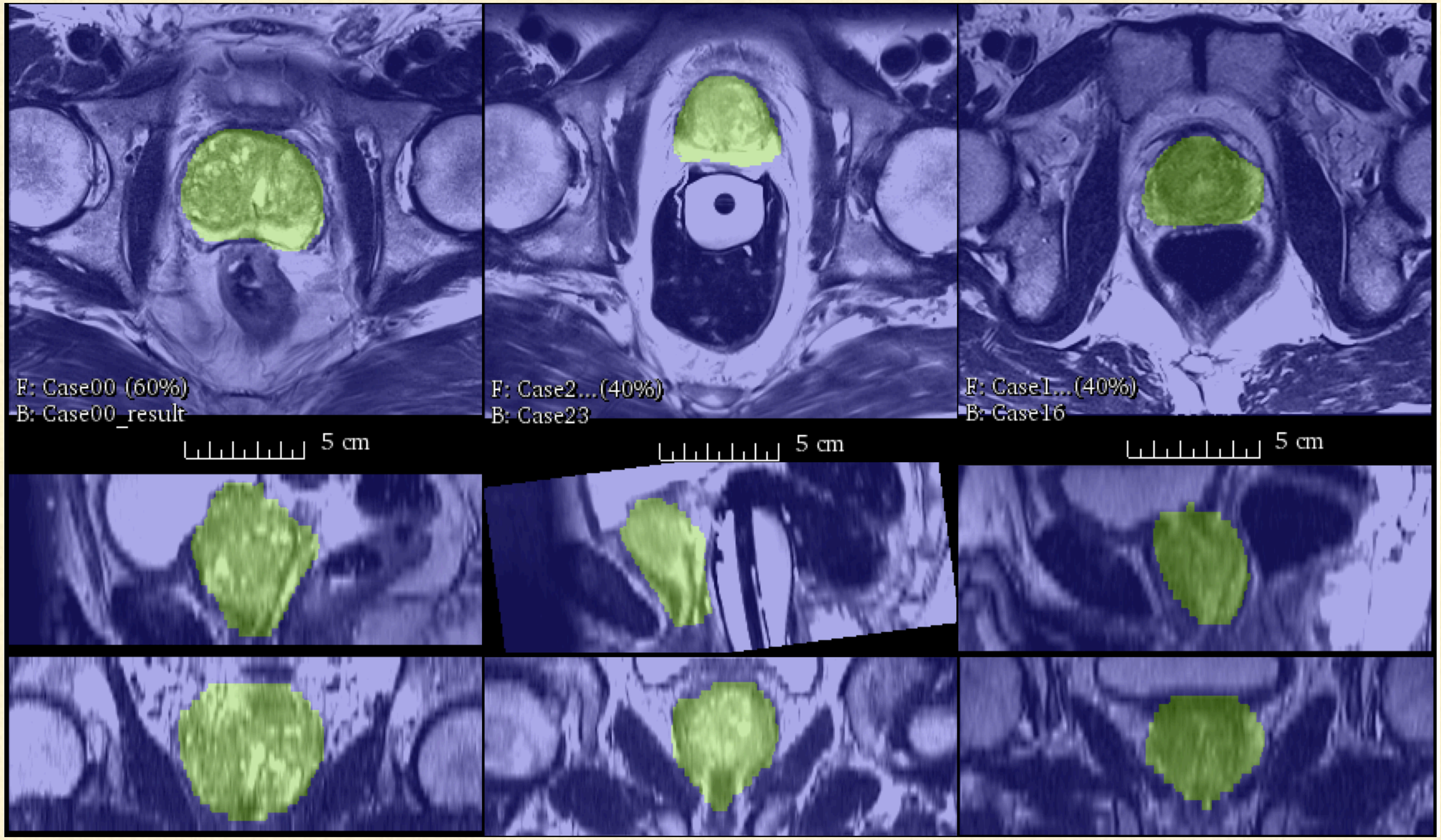
CentraleSupélec

SEGMENTATION IS THE SAME



- This is the V-NET architecture

EXAMPLE OF 3D V-NET SEGMENTATION



- This and many other architectures: https://niftynet.io/ and on GitHub: https://github.com/NifTK/NiftyNet

CLASSICAL AND VARIATIONAL TOMOGRAPHY

FILTERED BACK-PROJECTION IN MATRIX FORM

- Reconstruction is classically performed by FBP

- The FBP is a composition of linear operators. It can be carried out by matrix multiplications:

$$\mathbf{I} = \mathbf{T}^{\top}(\mathbf{T}\mathbf{T}^{\top})^{-1}\,.\,\mathrm{S}$$
$$\mathbf{I} = \mathbf{T}^{\top}(\mathbf{F}^{\mathrm{H}}\mathbf{K}\mathbf{F})\,.\,\mathrm{S}\ ,$$

where $\mathbf{F}$ is the Fourier transform and $\mathbf{F}^{\mathrm{H}}$ it adjoint, the inverse Fourier transform.

- These matrices can be implemented as layers in DL architectures.

EXAMPLE FRAMEWORK: PYRO-NN

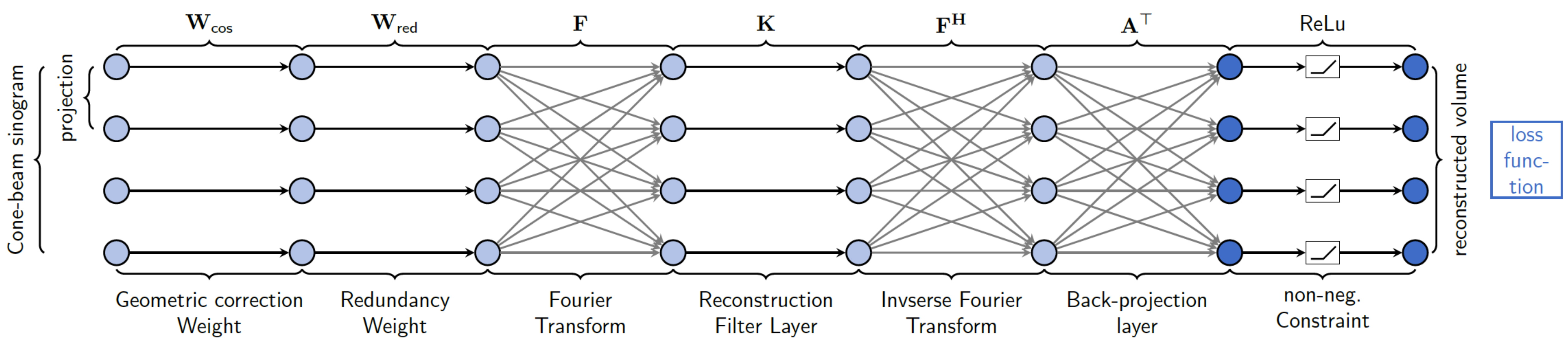


FIG. 3. FDK-Reconstruction Network. Light blue nodes represent the projection domain, while dark blue nodes stands for volume domain.

- 2D and 3D reconstruction operators implemented in TF 1.x

CentraleSupélec

TOMOGRAPHY AS AN INVERSE PROBLEM

- Computing the projection is a well-posed, direct : a solution exist, is unique, and is regularly continuously variable: small changes in the input cause small changes in the output.
- Tomography is an example of an *inverse problem*. The solution may not be unique, and small changes in the input may induce large changes in the back-projection.
- In reality we observe $\mathrm{S} = \mathbf{T}\mathrm{I} + \eta$, where η is some noise (Poisson, Gauss, …) plus some degradations (beam hardening, blur due to motion, sensor artifacts…)
- The adjoint operator yielding $\mathrm{J} = \mathbf{T}^{\top}\mathrm{S}$ is the back-projection operator, not the inverse.

CentraleSupélec

OPTIMISATION FORMULATION

- The direct observation is, for an additive noise $\mathrm{S} = \mathbf{T}\,.\,\mathrm{I} + \eta$
- In this case, it makes sense to seek to minimize $\|\eta\|_2^2$, I.e

  $$\mathrm{I}^\star = \mathrm{argmin}_\mathrm{I} \|\mathbf{T}\mathrm{I} - \mathrm{S}\|_2^2$$

- This is the least-square estimate. It corresponds to minimizing the negative log-likelihood of the distribution of η.

  $$\eta \sim \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \frac{\|\mathbf{T}\mathrm{I} - \mathrm{S}\|_2^2}{2\sigma^2}$$

- $\mathrm{I}^\star$ is the Maximum Likelihood estimator (MLE) for this problem.

ILL-POSED PROBLEMS

- Least-square problems are easy to solve:
$\|\mathbf{T}\mathrm{I} - \mathrm{S}\|^2 = \mathrm{I}^\top\mathbf{T}^\top\mathbf{T}\mathrm{I} - 2\mathrm{S}^\top\mathbf{T}\mathrm{I} + \mathrm{S}^\top\mathrm{S}$

$$\frac{\partial\|\mathbf{T}\mathrm{I} - \mathrm{S}\,|^2}{\partial \mathrm{I}} = 2(\mathbf{T}^\top\mathbf{T}\mathrm{I} - \mathbf{T}^\top\mathrm{S}) = 0$$

yields $\mathrm{I}^\star = (\mathbf{T}^\top\mathbf{T})^{-1}\,.\,\mathbf{T}^\top\mathrm{S}$, the *normal equation.* Not *quite* the min-norm solution of the FBP.

- Generally, $\mathbf{T}$ is ill-conditioned, and the inverse of $\mathbf{T}^\top\mathbf{T}$ is not sparse.

- This is an example of an ill-posed problem: the solution is sensitive to noise.

TIKHONOV REGULARIZATION

- Tikhonov (1942) proposed a regularization, by using a prior term:

  $I^\star = \text{argmin}_I \; \|\mathbf{T}\mathrm{I} - \mathrm{S}\|_2^2 + \|\Gamma \mathrm{I}\|_2^2,$

  where Γ is a linear operator.

- In practice, $\Gamma = \mathrm{Id}$; $\Gamma = \nabla$; or $\Gamma = \Delta$.

- This regularization is a *prior* on the distribution of I, under the Bayesian interpretation of the problem.

  $$p(\theta | x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

- Tikhonov regularization is an example of Maximum A Posteriori (MAP) estimation.

MAP GENERALIZATION

- MAP generalizes a lot of well-known solutions:
  - Tikhonov regularization with $\Gamma = \text{Id}$ is equivalent to Wiener filtering.
  - Using the Poisson likelihood instead of least-squares yields the Lucy-Richardson algorithm.
- Image restoration "reduces" to:
  - Finding a good prior (model) for the noise
  - Finding a good prior for the observation model
  - Optimizing the resulting formulation.

PENALIZED FORMULATION

- Formally:

  $$I^\star = \text{argmin}_I \Phi(\mathbf{T}\mathrm{I} - \mathrm{S}) + \lambda R(\mathrm{I})$$

- Φ is called the fidelity function
  R is the regularization
  λ is a Lagrange parameter.

- There has been an enormous amount of work in this area in signal/image processing since the 1990s.

- This is called the variational approach to inverse problem solving.

EXAMPLES OF REGULARIZATIONS

TOTAL VARIATION

- A seminal article is the Rudin-Osher-Fatemi article (1992).
- They proposed as regularization:

  $R = ||\nabla \mathrm{I}||_2$,

  I.e. the ℓ_1 norm of the Euclidean norm of the gradient.
- Compared with Tikhonov, only a $||.||^2$ is missing.
- This is based on the observation that the gradients of natural images follow an exponential distribution: $\nabla I \sim \exp(-x)$, and not a normal distribution.

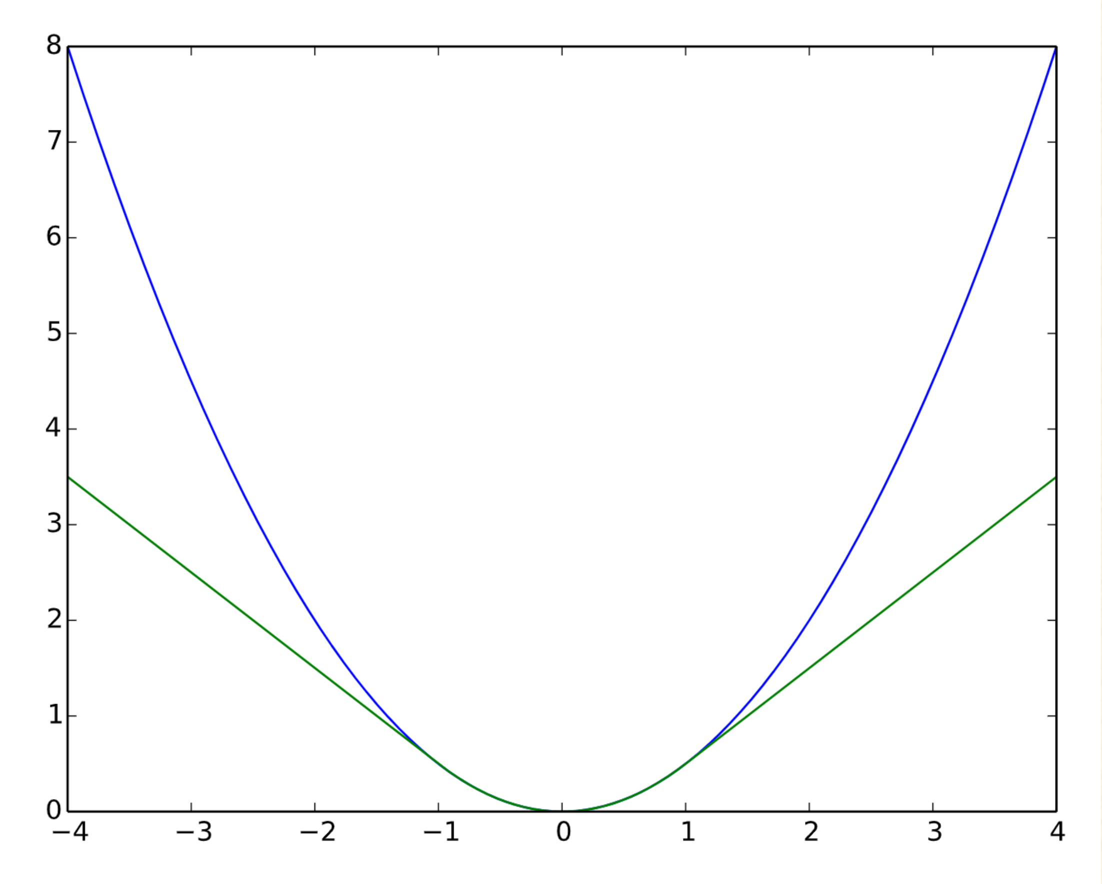
CentraleSupélec

NON-DIFFERENTIABILITY

- TV is non-differentiable. In the 1992 article, ROF used an approximation, the Huber norm:

$$L_\delta(a) = \begin{cases} \frac{1}{2}a^2 & \text{for } |a| \leq \delta \\ \delta(|a| - \frac{1}{2}\delta) & \text{otherwise} \end{cases}$$

It look quadratic near zero and like | . | far from zero.



CentraleSupélec

TV AND PROXIMITY OPERATORS

- An algorithm to optimize the real non-differentiable TV was proposed by Chambolle (2004)

- TV optimisation is an example of a *proximal* operator

$$\text{prox}_f(v) = \text{argmin}_x \left(f(x) + \frac{1}{2}\|x - v\|_2^2\right)$$

- Proximal operators are a generalization of projection operators for convex functions f.

$$p = \text{prox}_f(x) \Leftrightarrow x - p \in \partial f(p),$$

where ∂f is the *subgradient.*

CentraleSupélec

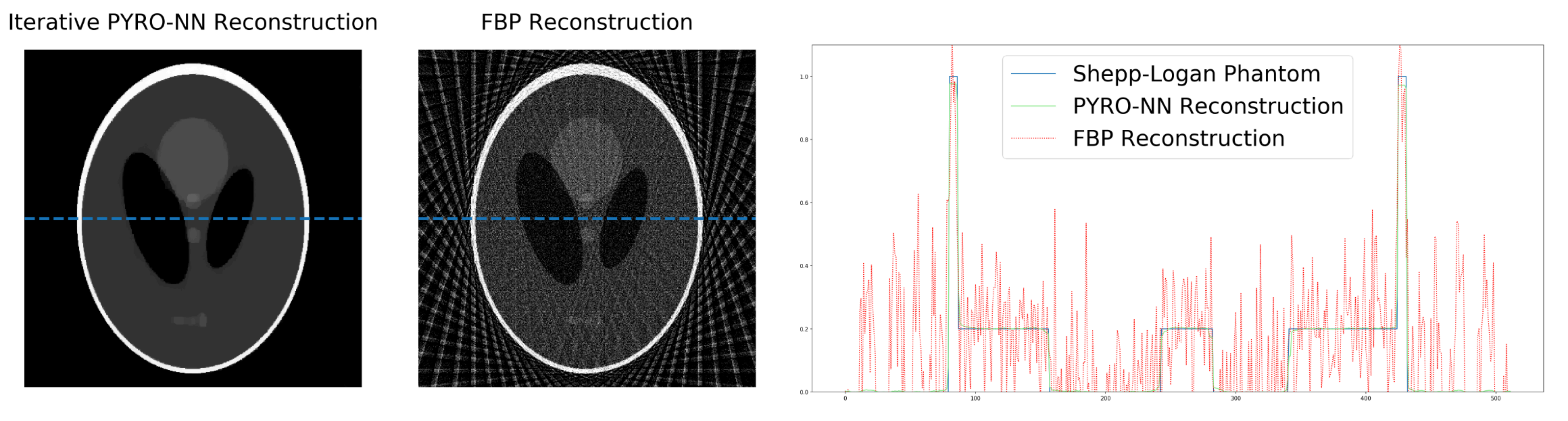
ALGORITHMS

- Iterative algorithms for solving proximity (or proximal) based problems have been proposed since the 1960s and the work of Moreau (1965) and Rockafeller (1970).

- The simplest algorithms are the proximal point algorithm, which can optimise a non-differential term ; and the forward-backward, which can optimize a sum of a non-differentiable and differentiable terms (i.e what we usually need for image processing).

$$\forall \ell \in \mathbb{N},\; x_{\ell+1} = x_\ell - \gamma_\ell (t'_\ell + \nabla g(x_\ell)), t'_\ell \in \partial f(x_{\ell+1})$$
$$\Leftrightarrow\; x_{\ell+1} = \mathrm{prox}_{\gamma_\ell f}(x_\ell - \gamma_\ell \nabla g(x_\ell))$$

EXAMPLE RECONSTRUCTION



- FBP and iterative reconstruction in Pyro-NN.

REAL EXAMPLE IN NEUTRON TOMOGRAPHY

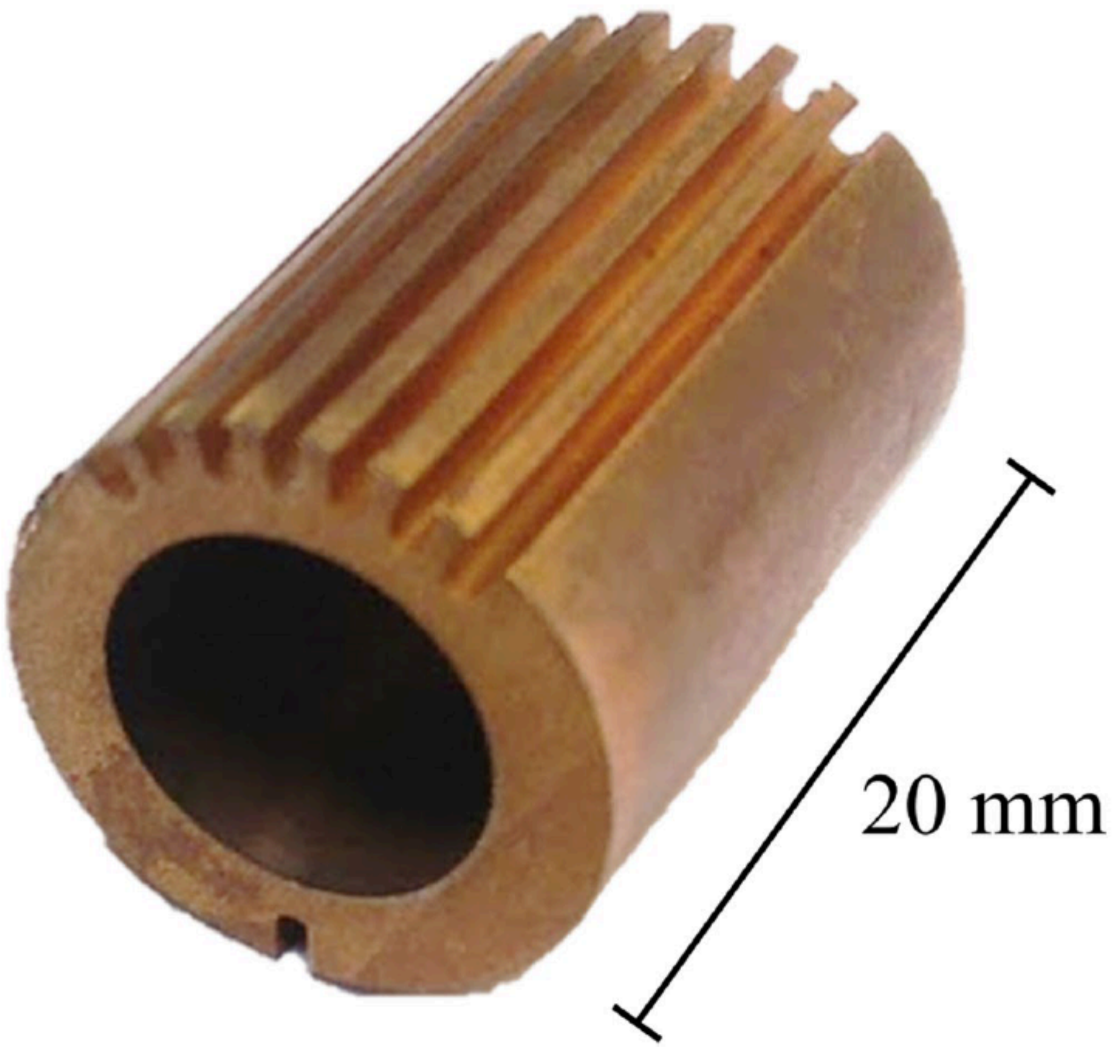
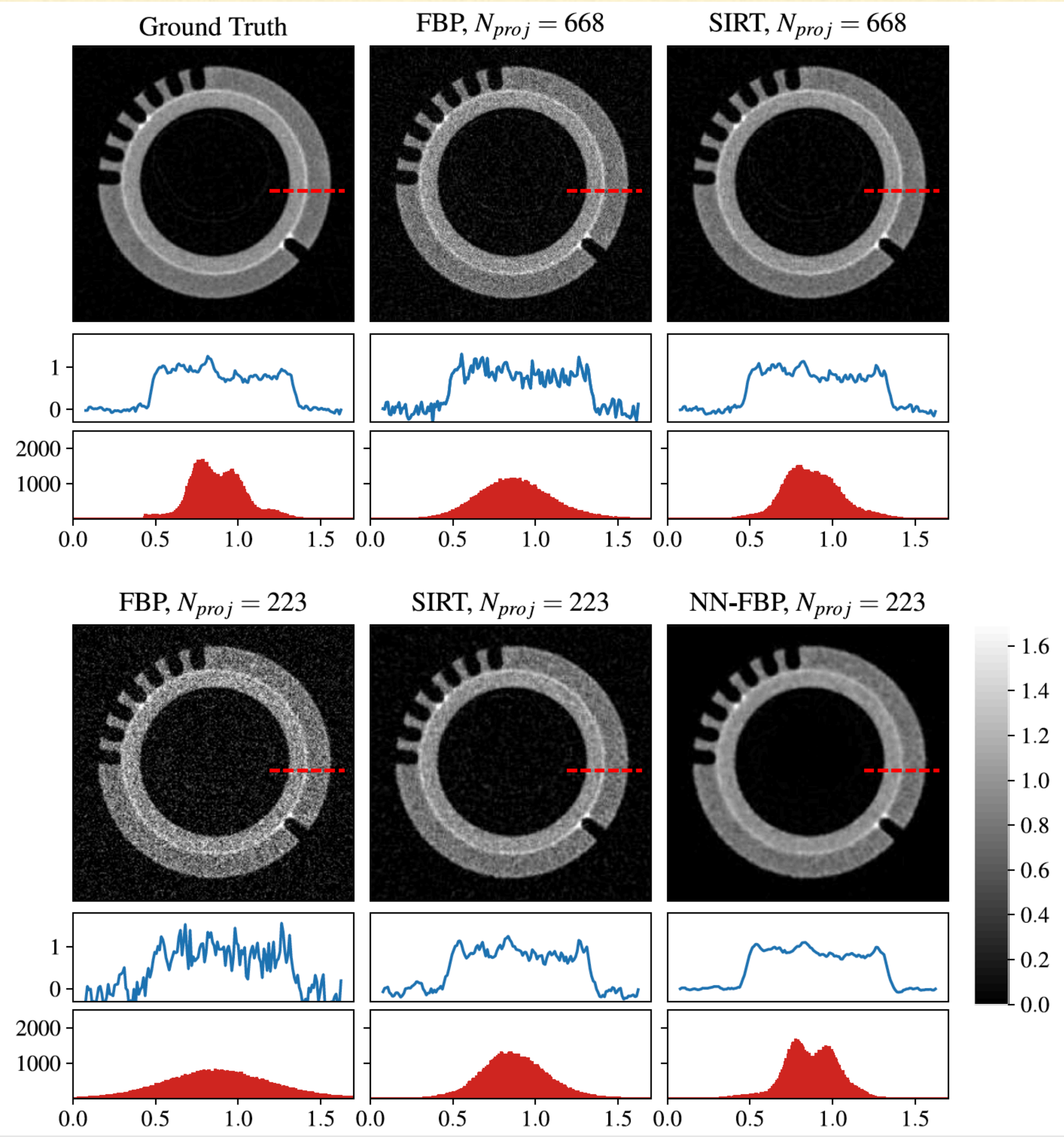


Figure 1. The sample inspected using NT. The Cu-CuCrZr pipe is the central section of the Culham Centre for Fusion Energy thermal break concept monoblock.

- From [Micieli et al., 2019]

NEUTRON 3D RECONSTRUCTION



- Acquired at the IMAT beamline
- Ground truth with 1335 projections and 400 iterations of SIRT.
- Comparisons: 668 and 223 projections
- NN-FBP adapted from [Pelt and Batenburg, 2013]

CentraleSupélec

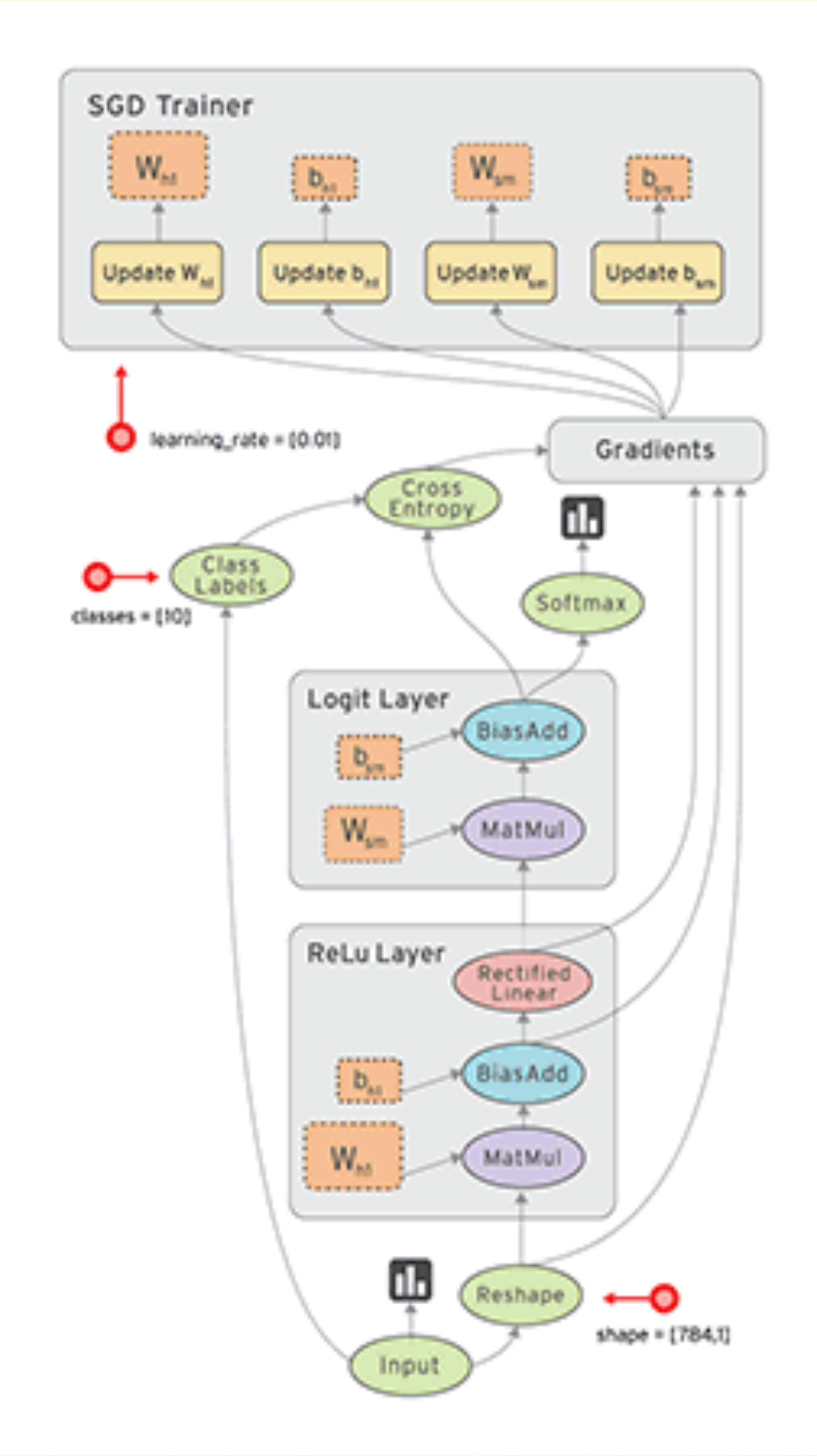
DEEP LEARNING FRAMEWORKS FOR TOMOGRAPHY

DEEP LEARNING FRAMEWORKS

- Deep learning frameworks have existed for nearly 20 years.
- They have become popular and well-maintained in the last 10 years
- They are implemented in C++ / CUDA with a Python front-end
- Most popular:
  - Tensorflow + Keras
  - Pytorch

TENSORFLOW

- Pure tensorflow uses a graph computation model
- Define a graph, then run it : unusual for most scientists, semi-natural for CS people.
- Very powerful, pushed and developed by Google.
- Quick introduction in Tutorial
- Keras a useful front end, not good for extensions.



PYTORCH

- PyTorch is an evolution of Torch, mostly used by FAIR (Facebook AI Research center). Initiated partly at UCL with Yann LeCun.
- Nicer and very good auto-differentiate (autograd).
- Extensive tutorial here:
- Main site https://pytorch.org

PYTORCH VERY INTUITIVE

```
# gradient descent
history = []
for t in range(epochs+1):

    # data fidelity
    fidelity = (kernel @ target.reshape(-1,1) - signal)**2

    # regularization
    regularization = tv_prior(target)

    # loss function
    loss = fidelity.mean() + reg * regularization.mean()

    # gradient step
    loss.backward()
    optimizer.step()
    optimizer.zero_grad()

    # projection step
    target.data.clamp_(min=0)
```

CONCLUSION

WHAT CAN DEEP-LEARNING DO FOR TOMOGRAPHY?

- 1- Learning aspects:
  - Denoising, segmentation
  - Learn a regularisation operator
- 2- Provide fast forward and backward projection operators
  - Open-source, varied geometries, 2D or 3D, on GPU,
- 3- Provide excellent, general purpose, distributed, stochastic optimisation methods implemented on GPU.
- Altogether: a new software platform to develop advanced reconstruction algorithms: you "only" need to provide the forward model.

BIBLIOGRAPHY

[Chambolle, 2004] Chambolle, A. (2004). An algorithm for total variation minimization and applications. J. Math. Imaging Vis., 20(1–2):89–97.
[Chouzenoux et al., 2015] Chouzenoux, E., Jezierska, A., Pesquet, J.-C., and Talbot, H. (2015). A convex approach for image restoration with exact poisson–gaussian likelihood. SIAM Journal on Imaging Sciences, 8(4):2662–2682.
[Couprie et al., 2011] Couprie, C., Talbot, H., Pesquet, J.-C., Najman, L., and Grady, L. (2011). Dual constrained TV-based regularization. In Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, pages 945–948. IEEE.
[Gonzalez and Woods, 2001] Gonzalez, R. C. and Woods, R. E. (2001). Digital image processing. Inter- science, NY.
[Grady and Polimeni, 2010] Grady, L. and Polimeni, J. (2010). Discrete Calculus: Applied Analysis on Graphs for Computational Science. Springer Publishing Company, Incorporated.
[Goodfellow et al., 2016] Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep learning. MIT press.
[Li et al., 2018] Li, H., Xu, Z., Taylor, G., Studer, C., and Goldstein, T. (2018). Visualizing the loss landscape of neural nets. In Advances in Neural Information Processing Systems, pages 6389–6399.
[Liu et al., 2018] Liu, P., Zhang, H., Zhang, K., Lin, L., and Zuo, W. (2018). Multi-level wavelet-cnn for image restoration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 773–782.
[Mallat, 2009] Mallat, S. (2009). A wavelet tour of signal processing. Academic press, 3rd edition.
[Micieli et al., 2019] Micieli, D., Minniti, T., Evans, L. M., and Gorini, G. (2019). Accelerating neutron to-
mography experiments through artificial neural network based reconstruction. Scientific reports, 9(1):2450.
[Milletari et al., 2016] Milletari, F., Navab, N., and Ahmadi, S.-A. (2016). V-net: Fully convolutional neural networks for volumetric medical image segmentation. In 2016 Fourth International Conference on 3D Vision (3DV), pages 565–571. IEEE.
[Moreau, 1965] Moreau, J. (1965). Proximit´e et dualit´e dans un espace hilbertien. Bull. Soc. Math. France, 93(2):273–299.
[Pelt and Batenburg, 2013] Pelt, D. M. and Batenburg, K. J. (2013). Fast tomographic reconstruction from limited data using artificial neural networks. IEEE Transactions on Image Processing, 22(12):5238–5251.
[Rockafellar, 1970] Rockafellar, R. T. (1970). Convex analysis. Princeton Univ Press. Reprinted 1997. [Rudin et al., 1992] Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise
removal algorithms. Phys. D, 60(1-4):259–268.
[Syben et al., 2019] Syben, C., Michen, M., Stimpel, B., Seitz, S., Ploner, S., and Maier, A. K. (2019).
Pyro-nn: Python reconstruction operators in neural networks. arXiv preprint arXiv:1904.13342.
[Tikhonov and Arsenin, 1977] Tikhonov, A. N. and Arsenin, V. I. (1977). Solutions of ill-posed problems, volume 14. Vh Winston.[Wiener, 1949] Wiener, N. (1949). Extrapolation, interpolation and smoothing of stationary time series-with engineering applications MIT Press.