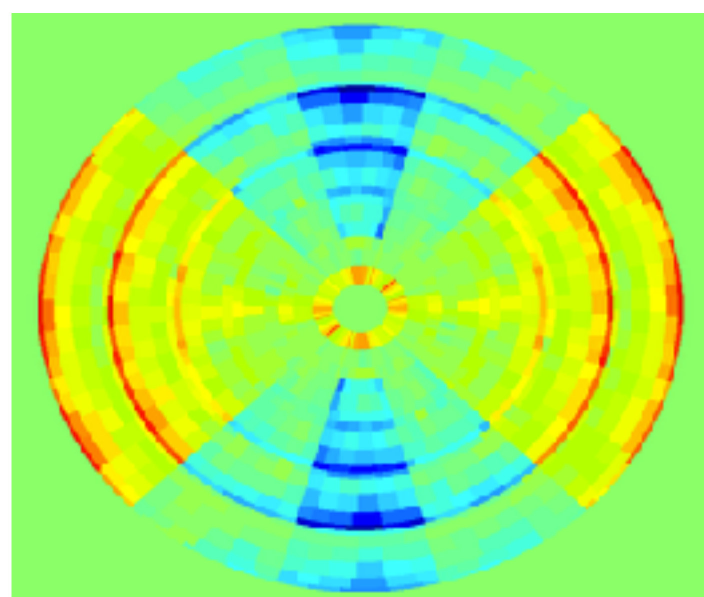
Organising Deep Networks

Edouard Oyallon



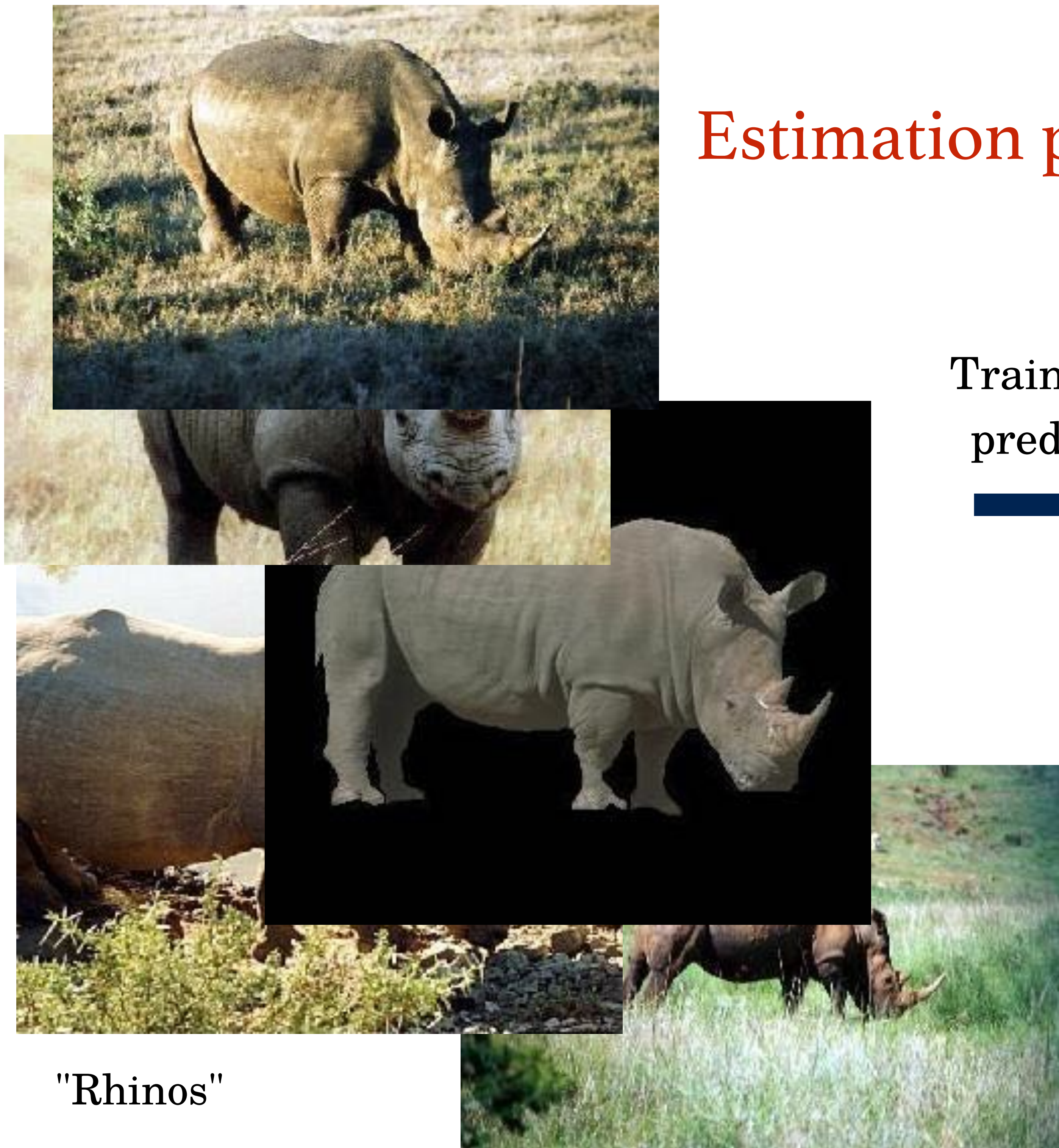
advisor: Stéphane Mallat

following the works of Laurent Sifre, Joan Bruna, …

collaborators: Eugene Belilovsky, Sergey Zagoruyko, Jörn Jacobsen, …

High Dimensional classification

$$(x_i, y_i) \in \mathbb{R}^{224^2} \times \{1, ..., 1000\}, i < 10^6 \longrightarrow \hat{y}(x)?$$



"Rhinos"

Estimation problem

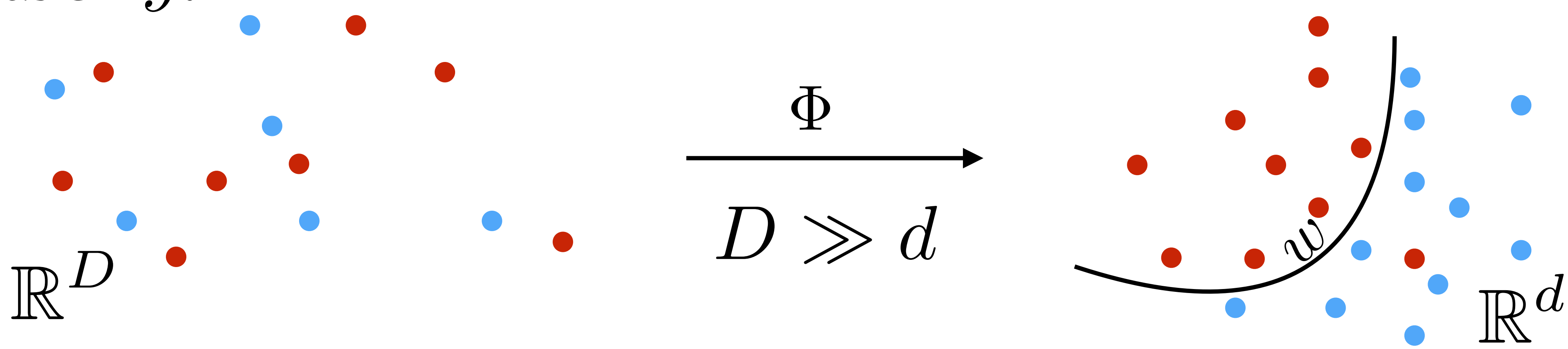
Training set to predict labels

"Rhino"

Not a "rhino"

Fighting the curse of dimensionality

- **Objective:** building a representation Φx of x such that a simple (say euclidean) classifier $\hat{y}$ can estimate the label y:

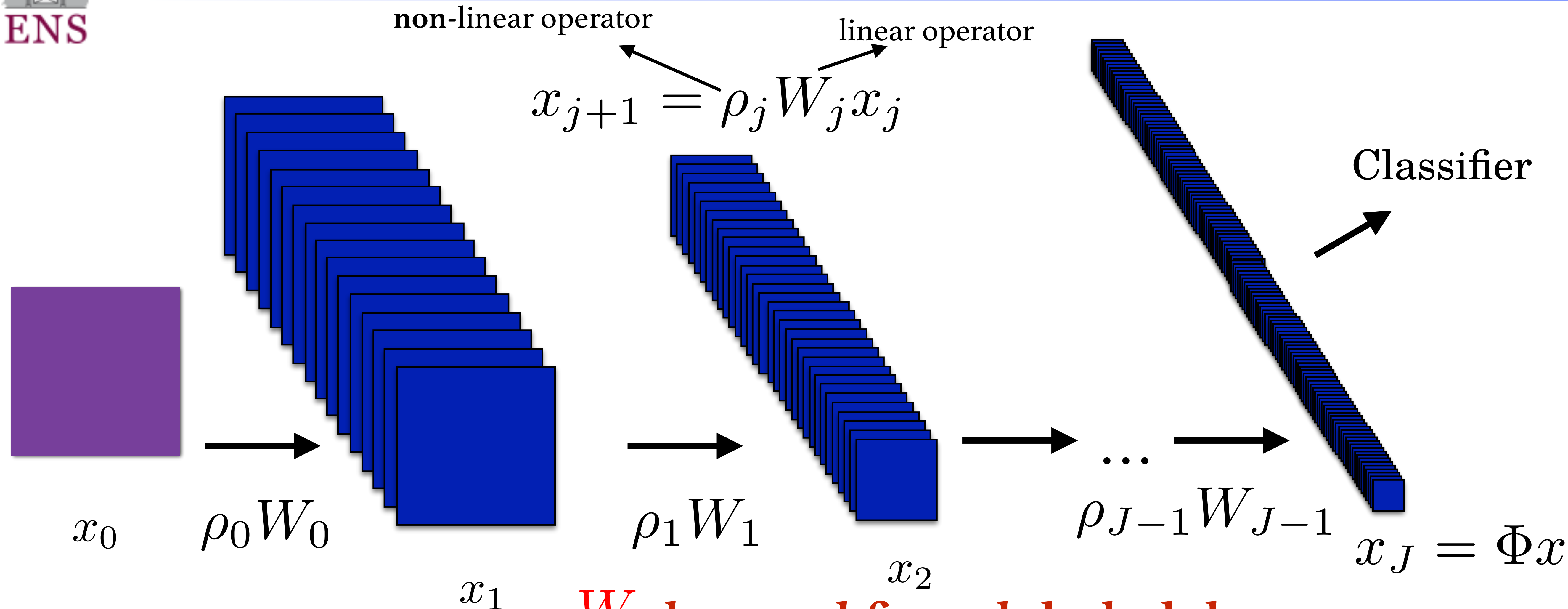


- Designing Φ consist of building an approximation of a low dimensional space which is regular with respect to the class:

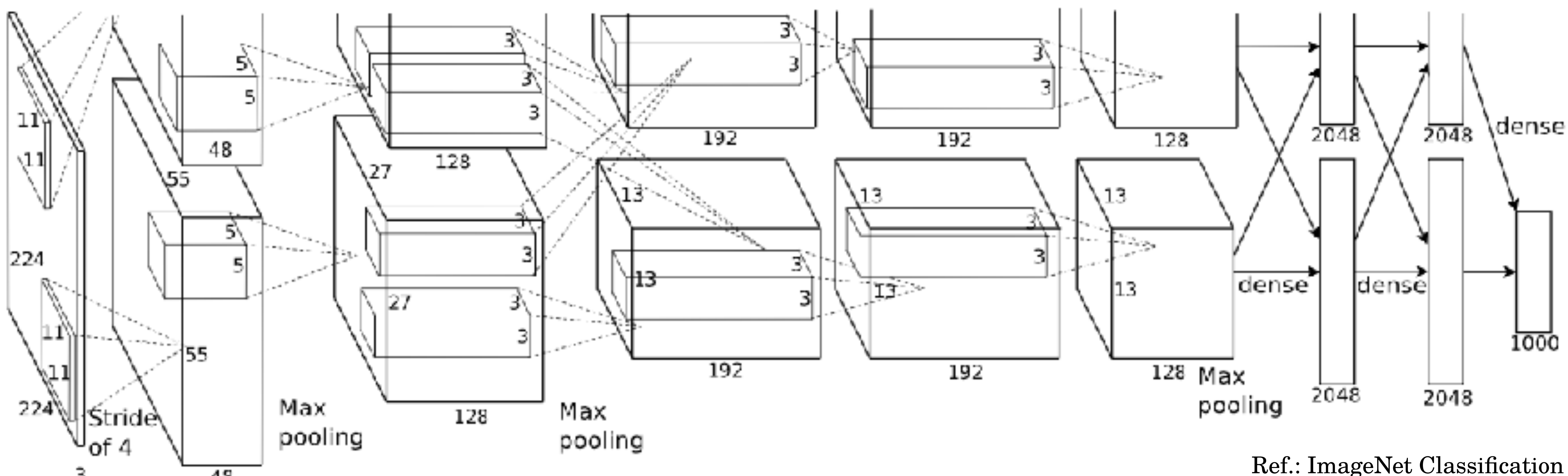
$$\|\Phi x - \Phi x'\| \lll 1 \Rightarrow \hat{y}(x) = \hat{y}(x')$$

- **Necessary** dimensionality and variance reduction

Completely solved by the deep <u>blackbox</u>



W_i **learned from labeled data**



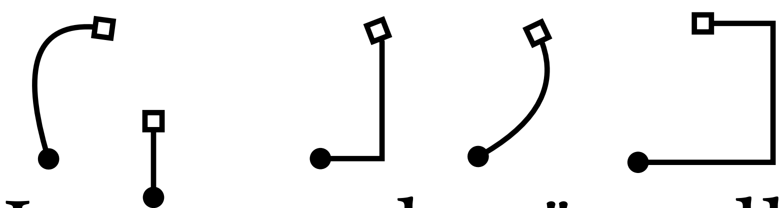
Solving it: DeepNetwork

Ref.: ImageNet Classification with Deep Convolutional Neural Networks. A Krizhevsky et al.

Why mathematics about deep learning are important

- **Pure black box.** Few mathematical results are available. Many rely on a "manifold hypothesis".

  Ex: stability to diffeomorphisms



- No **stability results.** It means that "small" variations of the inputs might have a large impact on the system. And this happens.

  Ref.: Intriguing properties of neural networks. C. Szegedy et al.

- **No generalisation result.** Rademacher complexity can not explain the generalization properties.

  Ref.: Understanding deep learning requires rethinking generalization C. Zhang et al.

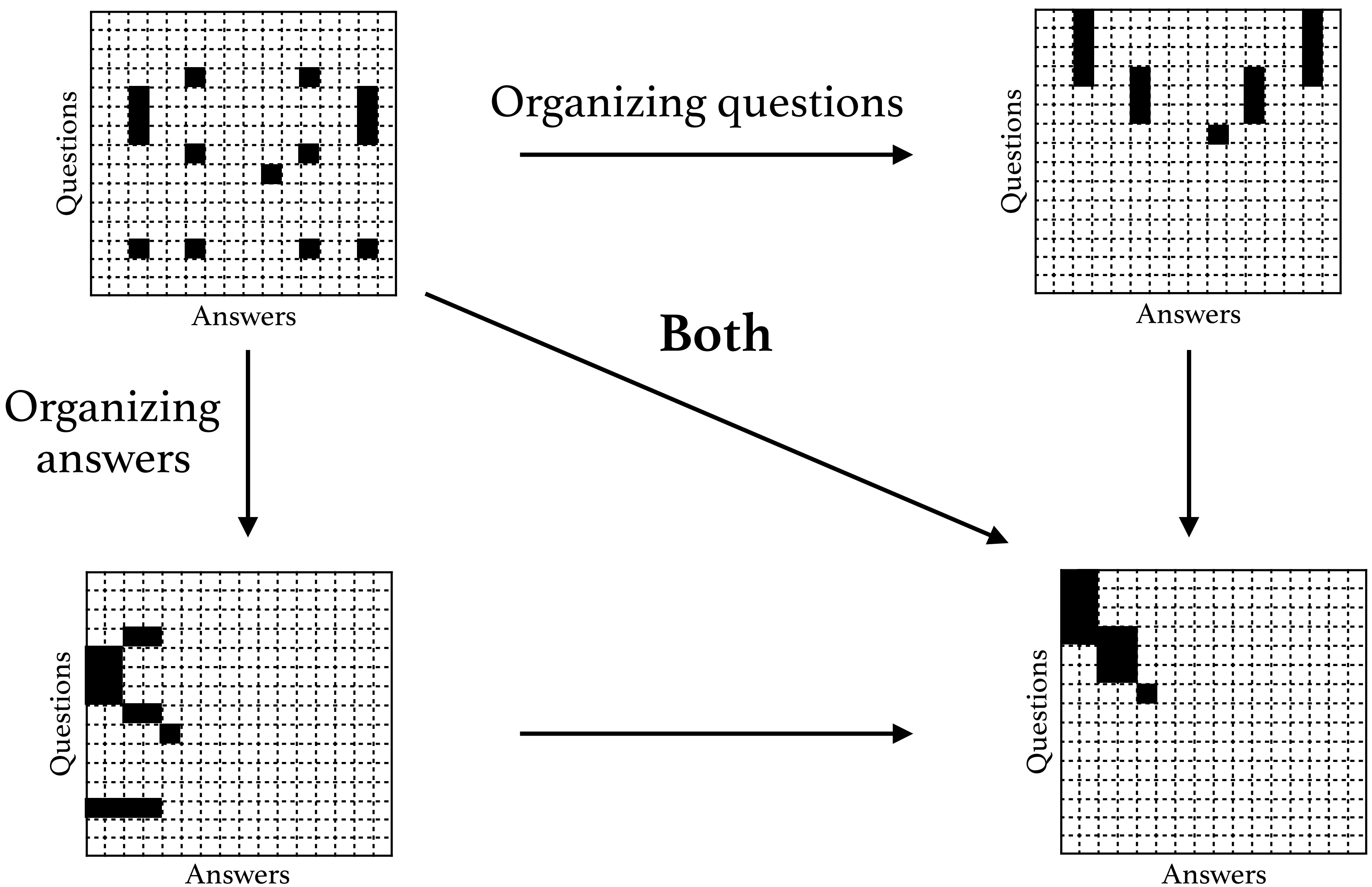
- Shall we learn **each layer** from scratch? (geometric priors?) The deep cascade makes features are hard to **interpret**

  Ref.: Deep Roto-Translation Scattering for Object Classification. EO and S Mallat

Organization is a key

- Consider a problem of questionnaires: people answer to 0 or 1 to some question. What does organizing mean?

Ref.: Harmonic Analysis of Digital Data Bases
Coifman R. et al.



In general, works tackle only one of the aspect

neighbours become meaningful: **local metrics**

Structuring the input with the Scattering Transform

- Scattering Transform S_J is a deep local descriptor of neighbourhood of amplitude 2^J, for images.

Ref.: Group Invariant Scattering, Mallat S

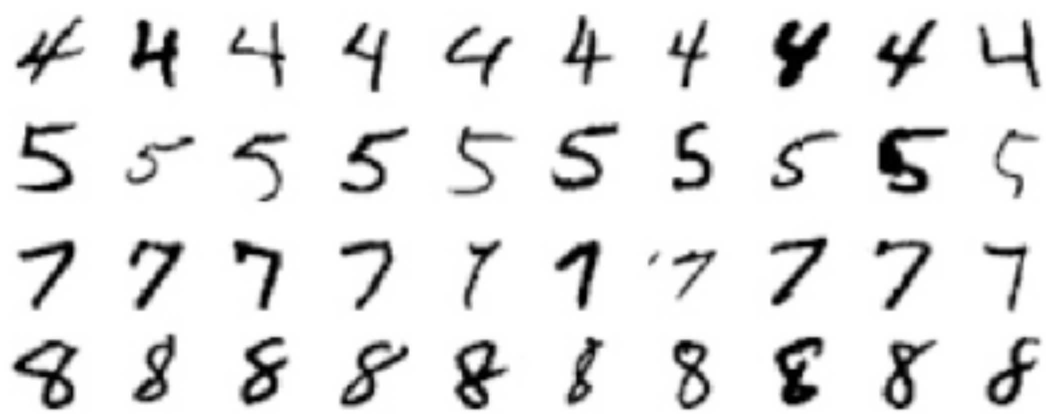
- It is a representation built via geometry with limited learning. (~SIFT)

Ref.: Invariant Convolutional Scattering Network, J. Bruna and S Mallat

- Successfully used in several applications:

All variabilities are known

  - Digits

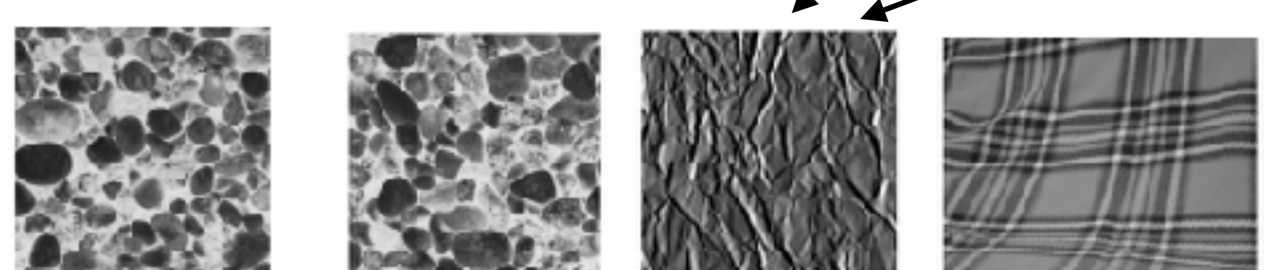


Small deformations +Translation

Ref.: Rotation, Scaling and Deformation Invariant Scattering for texture discrimination, Sifre L and Mallat S.

Rotation+Scale

  - Textures



Scattering on ImageNet: Geometry in CNNs

Ref.: Scaling the Scattering Transform: Deep Hybrid Networks
EO, E Belilovsky, S Zagoruyko

$$x \longrightarrow \boxed{S_J} \longrightarrow \boxed{ResNet} \longrightarrow$$

- Cascading a modern CNN leads to almost state-of-the-art result on Imagenet2012:

  1M images for training, 400k testing, 1000 classes

	Accuracy	Depth	#params
AlexNet	80.1	9	61M
ResNet	88.8	18	11.7M
Scat+ResNet	88.6	**10**	12.8M

- Demonstrates no loss of information + Less layers

Learning?

Benchmarking
Scattering + small data

$$x \longrightarrow \boxed{S_J} \longrightarrow \boxed{ResNet} \longrightarrow$$

Ref.: Scaling the Scattering Transform: Deep Hybrid Networks
EO, E Belilovsky, S Zagoruyko

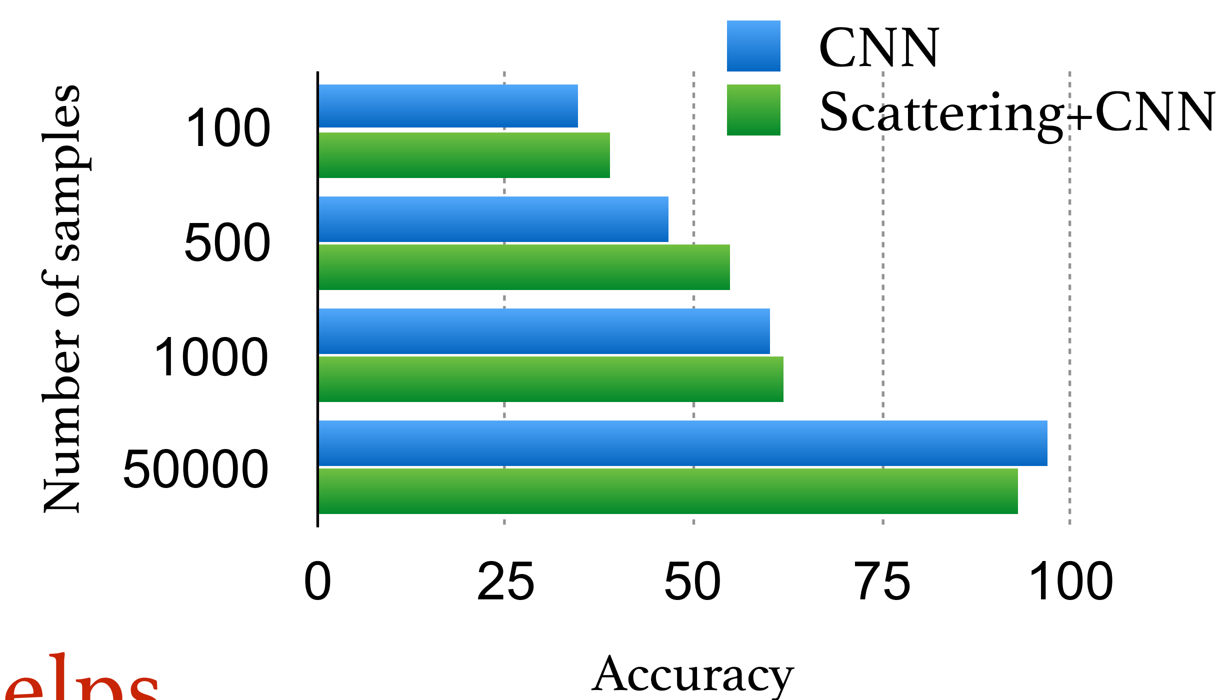
- Adding geometric prior regularises the CNN input, in the particular case of limited samples situations, **without reducing** the number of parameters.

- State-of-the-art results on STL10 and CIFAR10:

STL10: 5k training, 8k testing, 10 classes
+100k unlabeled(not used!!)

	Accuracy
Scattering+CNN	**76**
Deep	70
Unsupervised	75

Cifar10, 10 classes



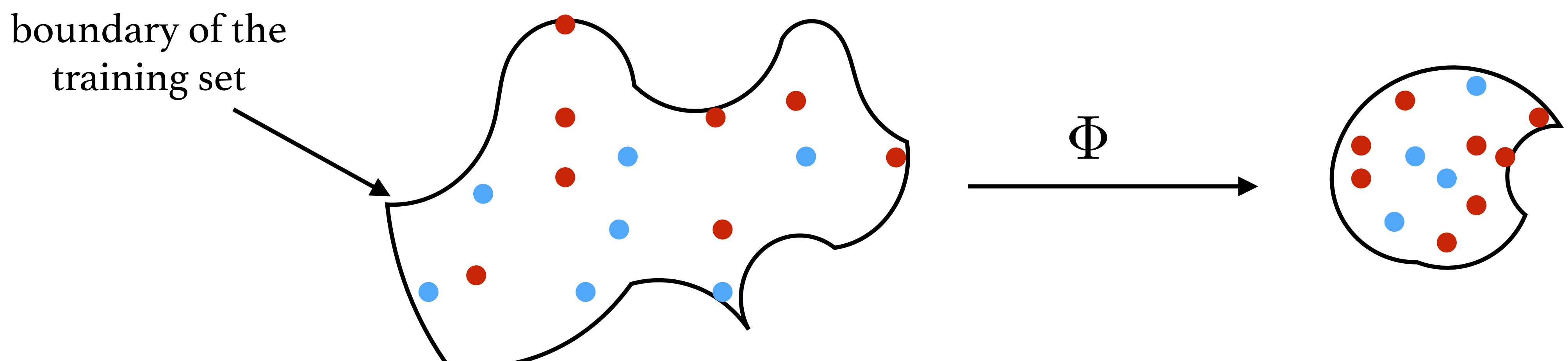
Geometry helps

Necessary mechanism: Separation - Contraction

- In high dimension, typical distances are huge, thus an appropriate representation must contract the space:

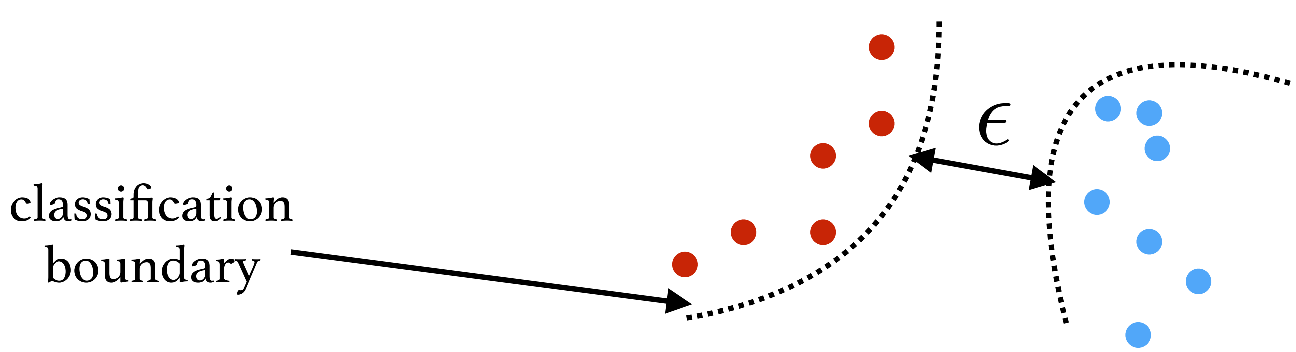
$$\|\Phi x - \Phi x'\| \leq \|x - x'\|$$

Ref.: Understanding deep convolutional networks S Mallat



- While avoiding the different classes to collapse:

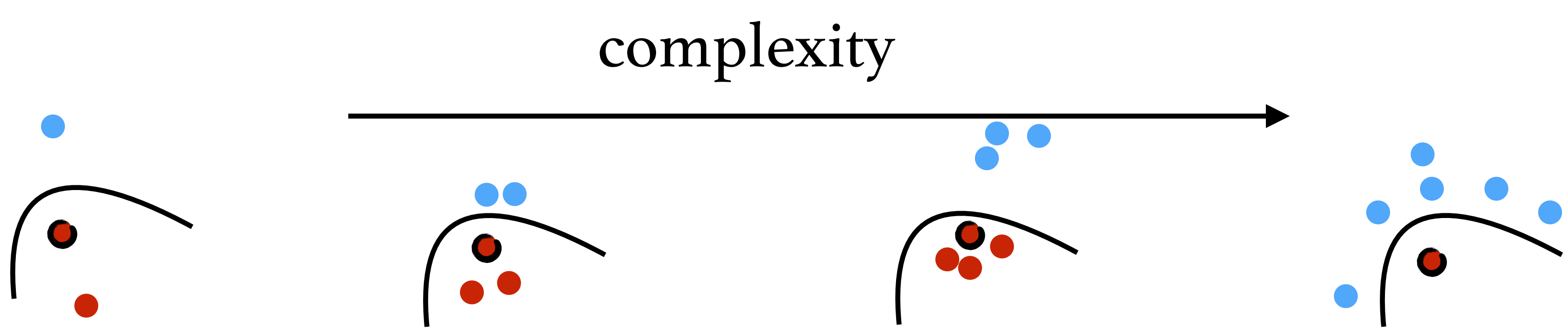
$$\exists \epsilon > 0, y(x) \neq y(x') \Rightarrow \|\Phi x - \Phi x'\| \geq \epsilon$$



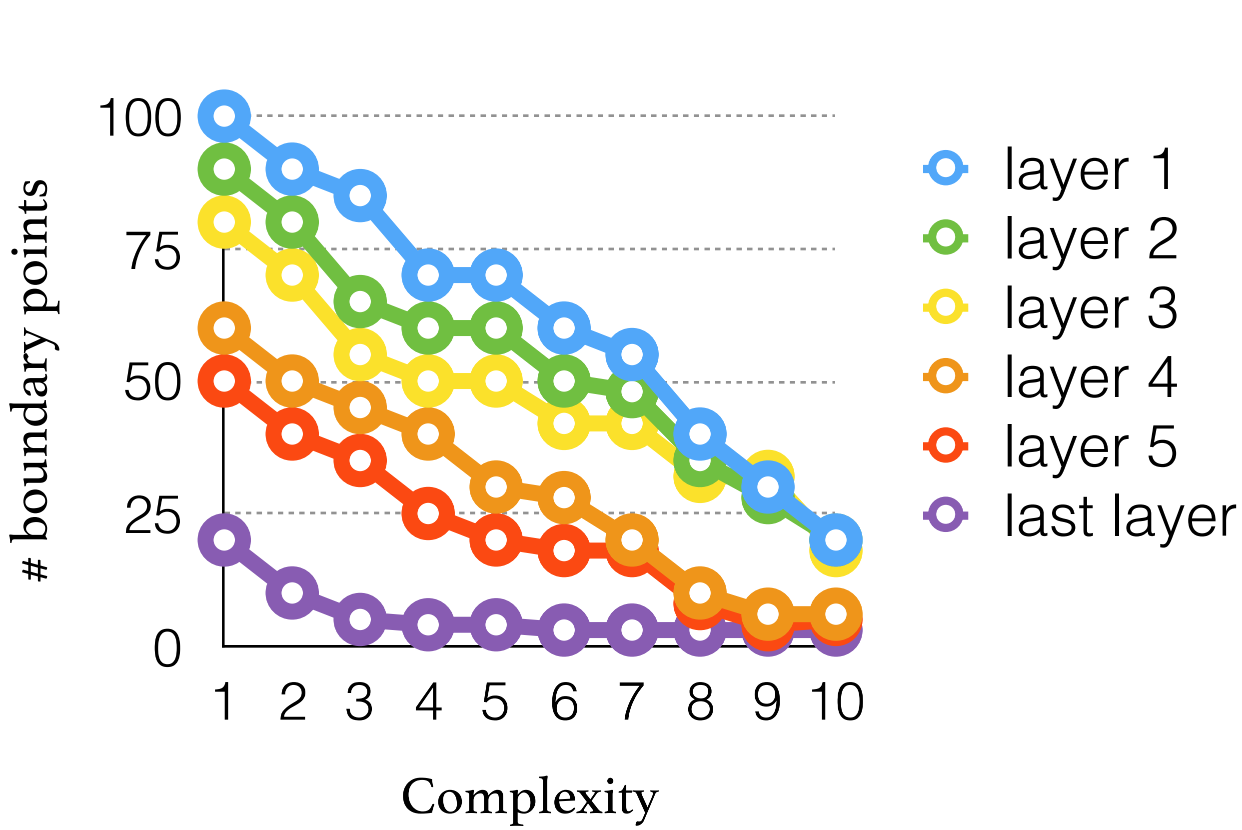
Complexity measure

- Measuring the complexity of the classification boundary (estimating the local dimensionality is **hard**)

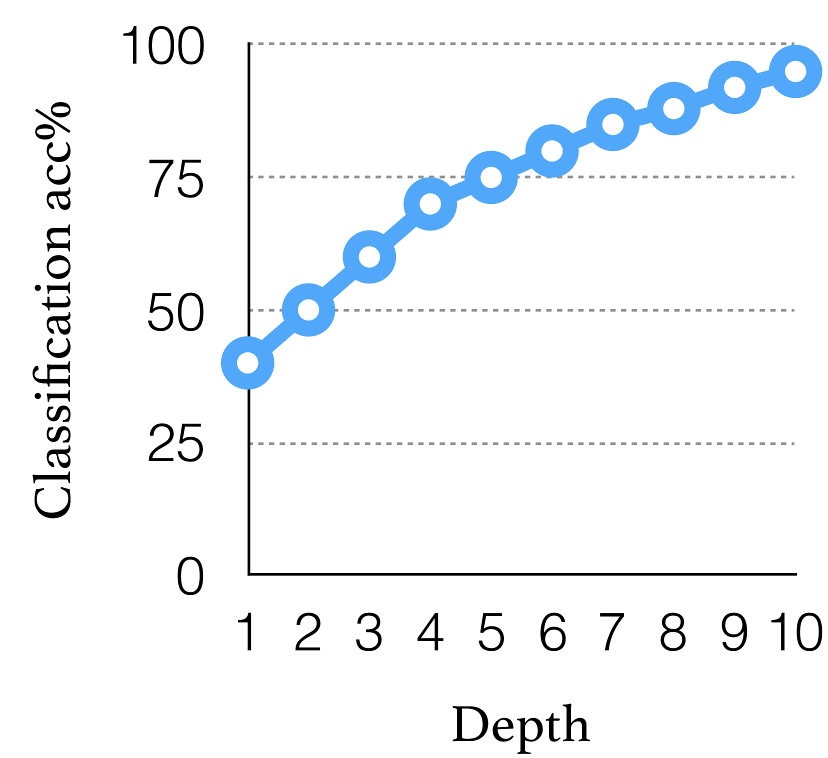
Ref.: Building a Regular Decision Boundary with Deep Networks EO



- Progressive contraction of the space, at each layer:



Explains the improvement



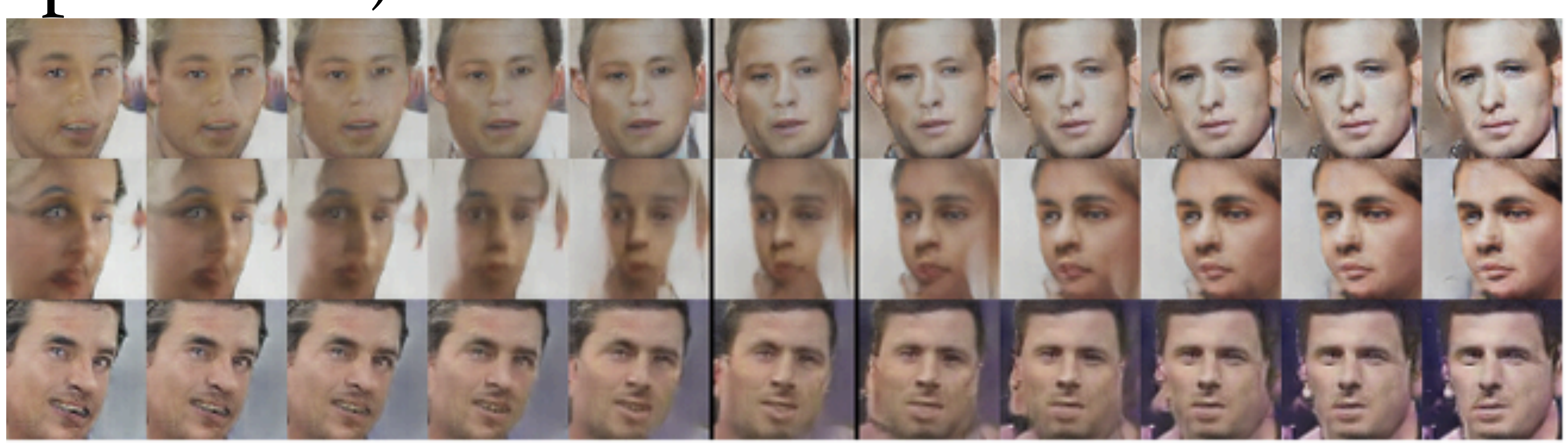
What variabilities are reduced?

Identifying the variabilities?

- Several works showed a DeepNet exhibits some covariance: Ref.: Understanding deep features with computer-generated imagery, M Aubry, B Russel

(c) Object color (d) Background color (a) Lighting (b) Scale

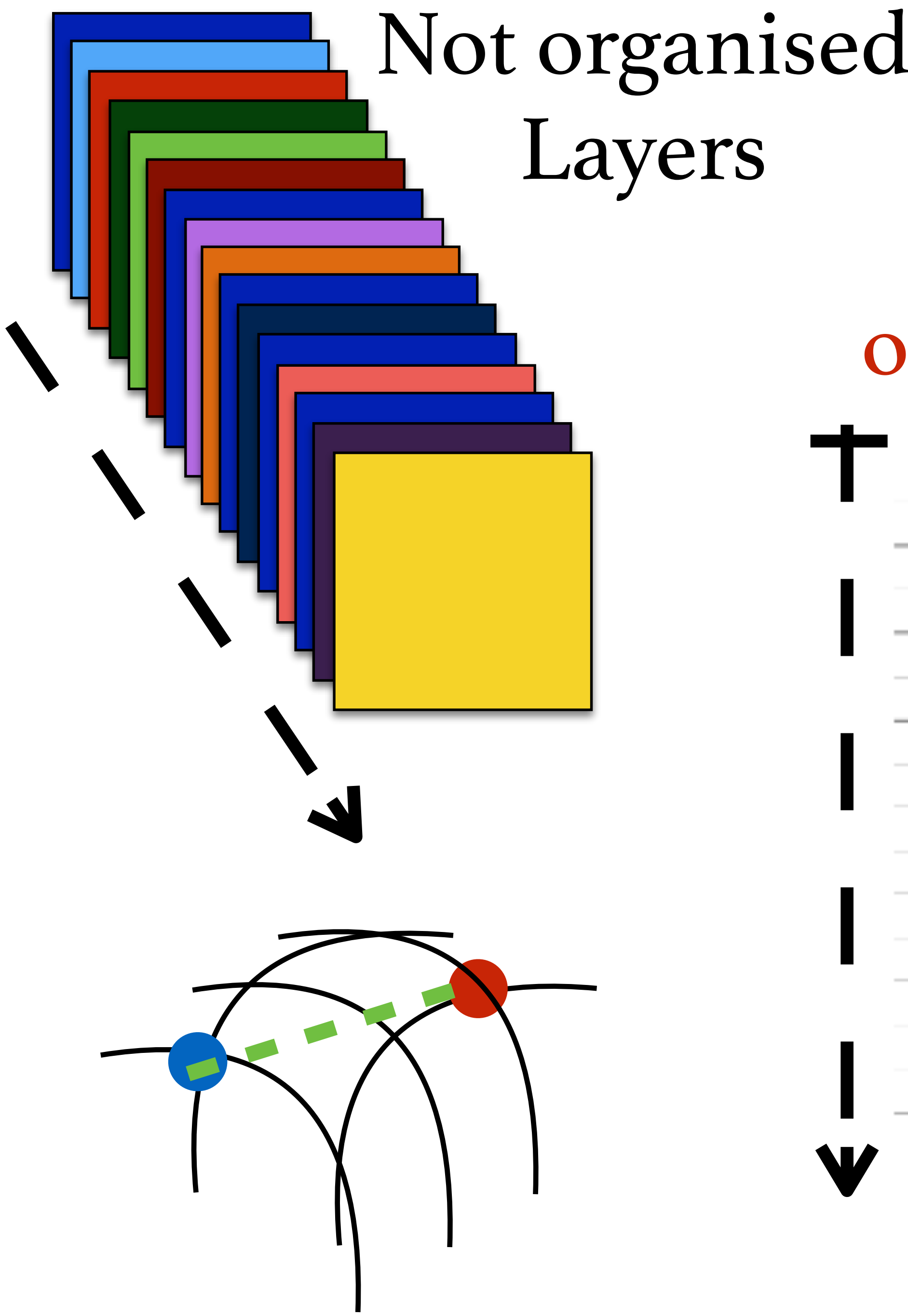
- Manifold of faces at a certain depth (e.g. good interpolations):



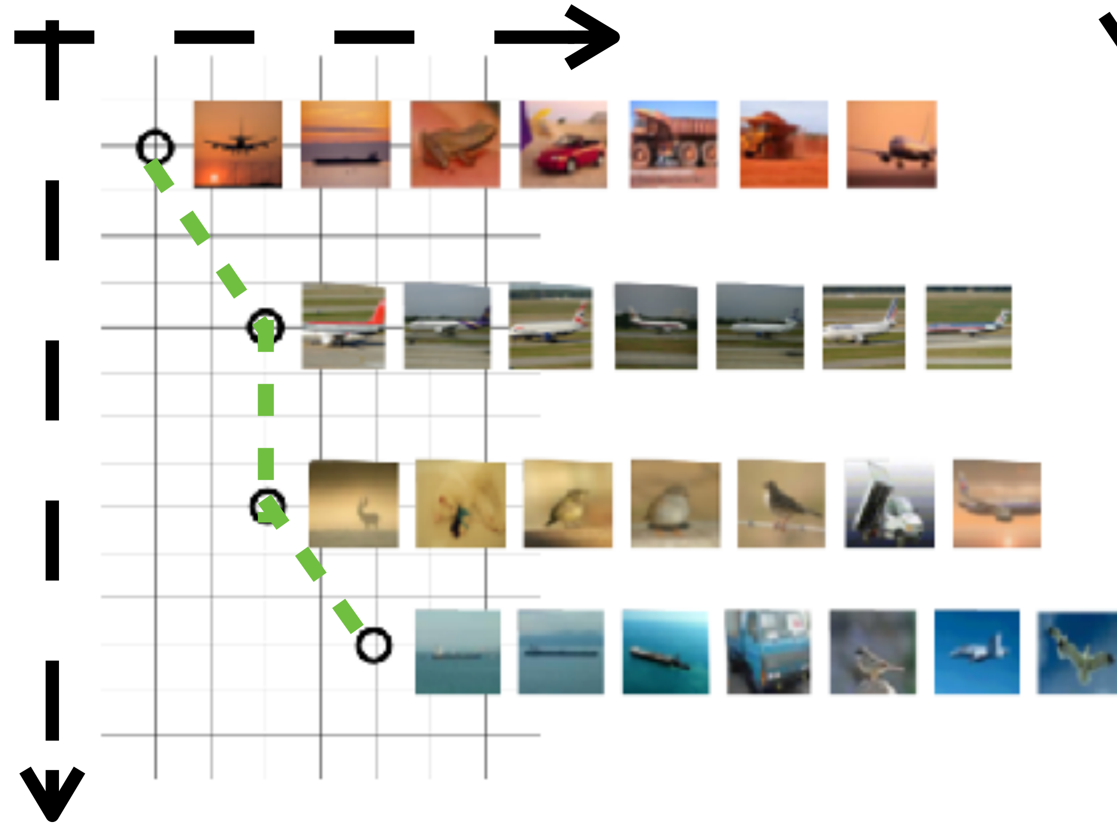
Ref.: Unsupervised Representation Learning with Deep Convolutional GAN, Radford, Metz & Chintalah

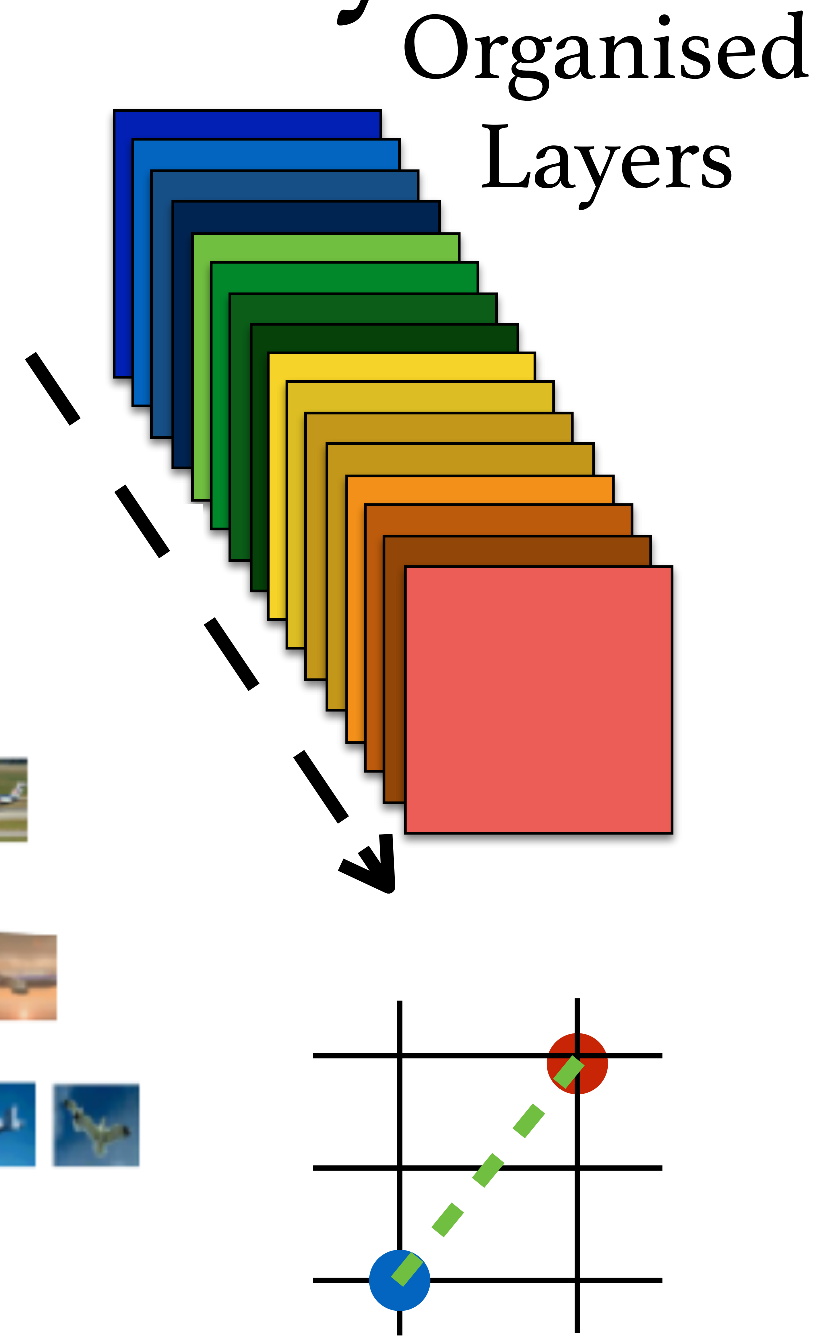
- It is **hard** to enumerate them…

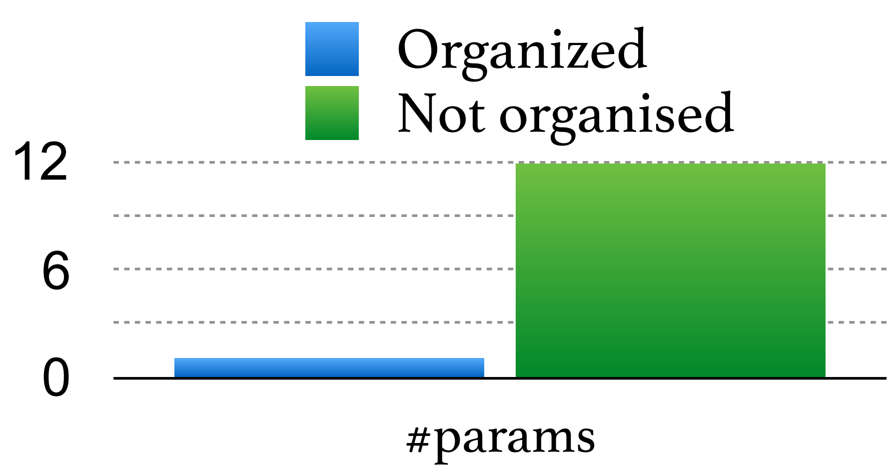
Flattening the variability



Defining an order
on layers of neurons







Ref.: Multiscale Hierarchical Convolutional Networks
J Jacobsen, EO, S Mallat, AWM Smeulders

Conclusion

- Stability, generalisation results, interpretability are important aspects…
- Check the website of the team DATA: http://www.di.ens.fr/data/
- Check my webpage for softwares and papers: http://www.di.ens.fr/~oyallon/

Stéphane Mallat

Eugene Belilovsky

Sergey Zagoruyko

Jörn Jacobsen

Thank you!