Generic Deep Networks with Wavelet Scattering

Edouard Oyallon, Stéphane Mallat and Laurent Sifre
DATA, Département Informatique, Ecole Normale Supérieure

DATA

Scattering network as Deep architecture

- We build a 2 layers network without training and which achieves similar performances with a convolutional network pretrained on ImageNet (Alex CNN [1]).
- Via groups acting on images, scattering network creates a representation Φ invariants to:
  - rotation
  - translation.
- Other properties:
  - discriminability of colors
  - stability to small deformations [2].

Deep scattering representation

- A scattering transform is the cascading of linear wavelet transform W_n and modulus non-linearities $|.|$:

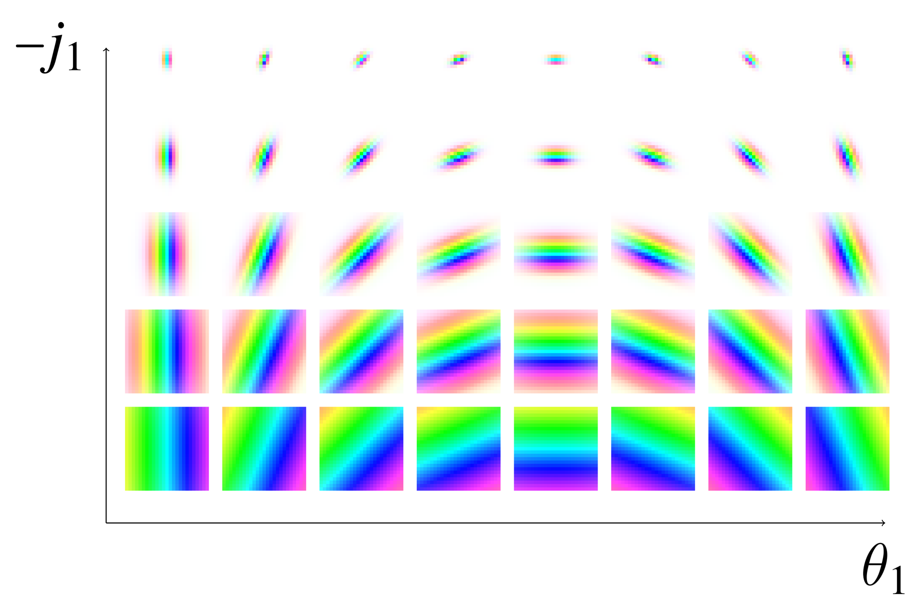
$$x \text{---} \boxed{|W_1|} \text{---} U_1 x \text{---} \boxed{|W_2|} \text{---} U_2 x \text{---} \dots$$

Pooling: $U_0 = x \to S_0 x$, $U_1 x \to S_1 x$, $U_2 x \to S_2 x$

Pooling is Average-Pooling (Avg) or the Max-Pooling (Max), defined on blocks of size 2^J.

- The first linear operator is a convolutional wavelet transform along space:

$$U_1 x(u, \theta_1, j_1) = |x \star \psi^1_{\theta_1, j_1}|(u)$$

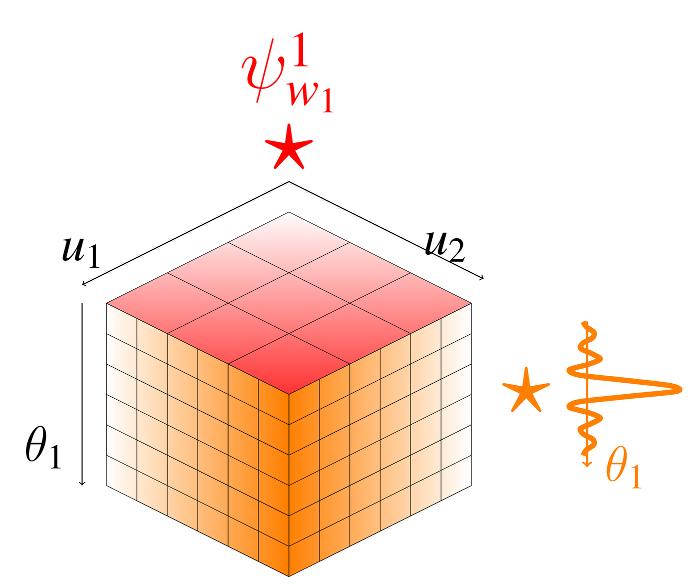


Complex wavelets. Phase is given by color, amplitude by contrast.

- The second linear operator is a wavelet transform along angles and space applied on U_1 and performed with a separable convolution $\circledast$:

$$U_2 x = |U_1 x \circledast \psi^2_{\theta_2, j_2, k_2}|$$

$$\psi^2_{\theta_2, j_2, k_2}(u, \theta_1) = \psi^1_{\theta_2, j_2}(u)\, \psi^b_{k_2}(\theta_1)$$



Convolution of $U_1 x$ using separable wavelets

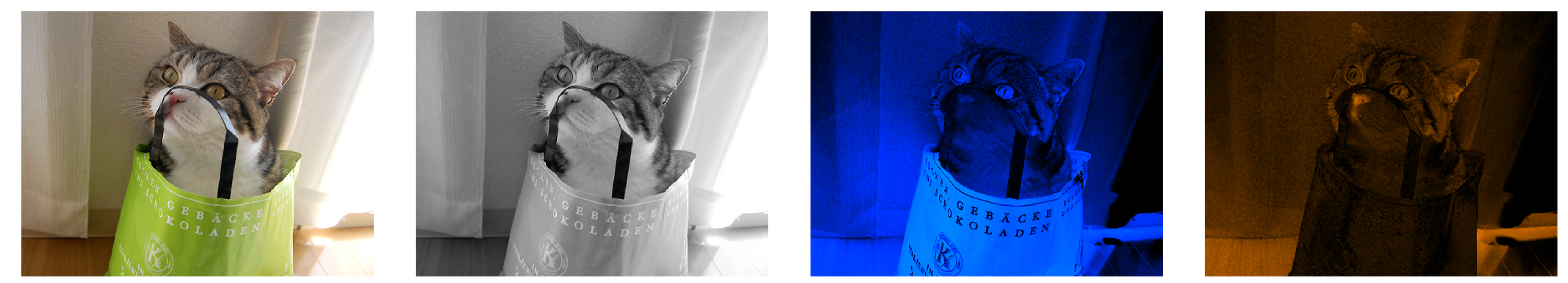
- Scattering coefficients are then

$$Sx = \{S_0 x, S_1 x, S_2 x\}$$

Color discriminability

Image x is separated into 3 color channels, x_Y, x_U, x_V. The final image representation is the aggregation of the scattering coefficients of each channels:

$$\Phi x = \{Sx_Y, Sx_U, Sx_V\}$$



Cat · Y channel · U channel · V channel

Classifier

$$\{x_1, \dots, x_M\} \text{---} \boxed{\Phi} \text{---} \{\Phi x_1, \dots, \Phi x_N\} \text{---} \boxed{\text{Standardization}} \text{---} \boxed{\text{Linear SVM}}$$

- Computation of the representations.
- Standardization: normalization of the mean and variance.
- Fed to a linear kernel SVM.

Numerical results

5 splits on Caltech-101 and Caltech-256.
Image inputs: 256×256, $J = 6$, 8 angles, final descriptor size is 1.1×10^5.

Samples from Caltech-101 and Caltech-256

Caltech-101 (101 classes, 10^4 images)

Architecture	Layers	Accuracy
Alex CNN	1	44.8 ± 0.8
Scattering, Avg	1	54.6 ± 1.2
Scattering, Max	1	55.0 ± 0.6
LLC	2	73.4
Alex CNN	2	66.2 ± 0.5
Scattering, Avg	2	68.9 ± 0.5
Scattering,Max	2	68.7 ± 0.5
Alex CNN	7	85.5 ± 0.4

Caltech-256 (256 classes, 3.10^4 images)

Architecture	Layers	Accuracy
Alex CNN	1	24.6 ± 0.4
Scattering, Avg	1	23.5 ± 0.5
Scattering, Max	1	25.6 ± 0.2
LLC	2	47.7
Alex CNN	2	39.6 ± 0.3
Scattering, Avg	2	39.0 ± 0.5
Scattering, Max	2	37.2 ± 0.5
Alex CNN	7	72.6 ± 0.2

Comparison with other architecture

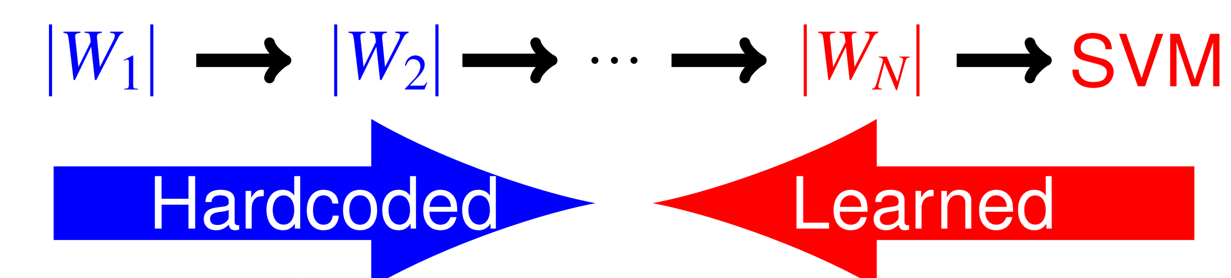
- LLC[3] is a two layers architecture with SIFT + unsupervised dictionary learning (specific to the dataset).
- Scattering performs similarly to Alex CNN on 2 layers [4].

Main differences with Alex CNN

- **No learning step**
- **Avg≈Max**
- No contrast normalization
- Complex wavelets instead of real filters
- Modulus (l^2-pooling) instead of ReLu
- Separable filters (tensor structure).

Open questions

- Predefined VS learned



Until which depth $n \leq N$ can we avoid learning?

- Max Pooling VS Avg Pooling

Conclusion & future work

- Scattering network provides an efficient initialization of the first two layers of a network.
- Optimizing scale invariance.
- Designing a third layer?

Contacts

- Website of the team: **http://www.di.ens.fr/data/**
- Edouard Oyallon, edouard.oyallon@ens.fr

References

[1] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25*, pages 1106–1114, 2012.

[2] S. Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10):1331–1398, 2012.

[3] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-constrained linear coding for image classification. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3360–3367. IEEE, 2010.

[4] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional neural networks. *arXiv preprint arXiv:1311.2901*, 2013.