Scaling the Scattering Transform: Deep Hybrid Networks

Edouard Oyallon,[1] *Eugene Belilovsky*,[2] *Sergey Zagoruyko*[3]

(1) ENS/PSL Research University (2) KU Leuven/Université Paris-Saclay/INRIA

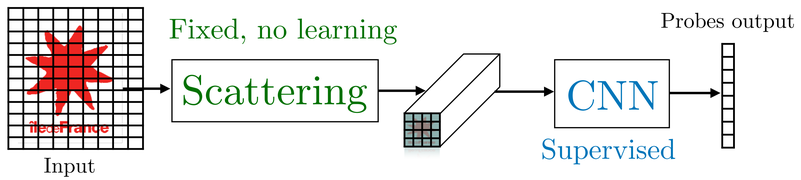
(3) École des Ponts ParisTech/Université Paris-Est

On the necessity of learning early layers

- Can we exploit physical laws and avoid learning some invariances?

Example: getting invariants along: translation, rotation, scales

- Do the first layers need to be learned?



- Can we learn from a descriptor and not the pixel level?

Short review of the Scattering Transform

- Cascade of complex wavelets transforms followed by moduli and finally a linear averaging parametrized by its invariance scale J:

- Additive, diffeomorphisms and translation stabilities,...

Nice mathematical foundations!

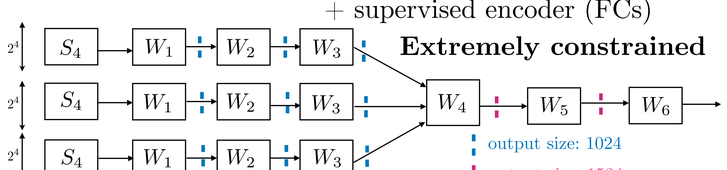
- Characteristics for an RGB input batch for the scale $J = 2$:

	Input	Output	Timing (TitanX)	Timing (CPU)
$J = 2$	32x32x3x128	8x8x243x128	0.03s	2.5s
$J = 2$	128x128x3x128	32x32x243x128	0.26s	16s
$J = 2$	256x256x3x128	64x64x243x128	0.71s	160s

Shared Local Encoder (SLE) for Imagenet

- Scat. + 1x1 convolution $\Longleftrightarrow$ Non-overlapping Scat. + supervised encoder (FCs)

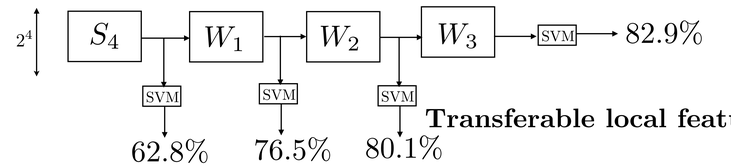
Extremely constrained



- AlexNet performances:

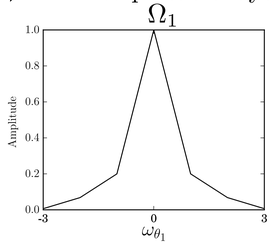
		Top 1	Top 5
Predefined	**Scat. + SLE**	**57.0**	79.6
Unsupervised	SIFT + FV + FCs	55.6	78.4
Supervised	SIFT + FV + SVM	54.3	74.3
	AlexNet	56.9	**80.1**

- Caltech101 generalization:



Transferable local features!

- An interpretable layer: (example is given only for the first order S_1 scattering)



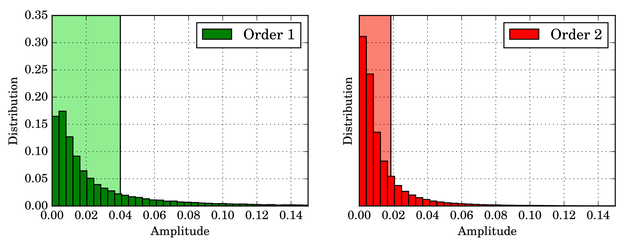
$$W_1 S_1 x(\lambda) = \sum_{j_1,\theta_1} w_{(j_1,\theta_1),\lambda} S_1 x(j_1,\theta_1)$$

$\hat{w}$, Fourier transform along the angle θ_1

$$\Omega_1(\omega_{\theta_1}) = \sum_{j_1,\lambda} |\hat{w}_{(j_1,\omega_{\theta_1}),\lambda}|^2 \Longleftrightarrow \text{Energy propagated by a given angular frequency}$$

Builds invariance along rotations

- Sparsification(threshold) in the angular Fourier bases of W_1:



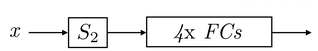
Threshold level 80%:
2% acc. loss on Imagenet

Fourier is a natural basis!

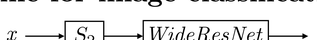
ResNet and Scattering on CIFAR10 and Imagenet

- CIFAR10:

Good prior baseline for image classification

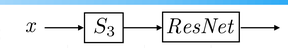


Unsupervised	Acc.
Scat. + FCs	**84.7**
Roto-scat + Gaussian SVM	82.3
ExemplarCNN	84.3
DCGAN	82.8



Supervised	Depth
Scat. + WRN	93.1
Highway Network	92.4
All-CNN	92.8
WRN 16 - 8	95.7
WRN 28 - 10	**96.0**

- ImageNet2012: Scat.+ResNet10:

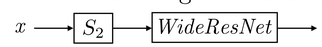


No information loss w.r.t. classification + less layers

	Top 1	Top 5	#params
Scat. + Resnet-10	68.7	88.6	12.8M
VGG-16	68.5	88.7	138M
ResNet-18	68.9	88.8	11.7M
ResNet-200	**78.3**	**94.2**	64.7M

ResNet and Scattering: Limited sample situations

- We only optimize the learning rate schedule in this case.
- CIFAR10:



#train	100	500	1000	Full
WRN 16-8	34.7±0.8	46.5±1.4	60.0±1.8	**95.7**
VGG 16	25.5±2.7	46.2±2.6	56.0±1.0	92.6
Scat. + WRN	**38.9±1.2**	**54.7±0.6**	**62.0±1.1**	93.1

- STL10 (train: 10 folds of 500, test: 10k) 96x96 color images:

	Acc.
Scat. + WRN	**76.0±0.6**
CNN	70.1±0.6
Exemplar CNN	75.4±0.3
Stacked AE	74.3
Hierchical Matching Pursuit	64.5±1
Convolutions K-means	60.1±1

Incorporating geometrical invariants improves limited settings

Software

- We provide a python implementation on GPUs which mixes CuPy and PyTorch.
- **Soon**: gradients, fast reconstruction, GANs, cache system...

Conclusion

- Scattering does not lose important information and is an initialisation that improves in the context of limited samples.
- No-mixing of spatial information permits to obtain good performances. Can we incorporate more structure?
- Can we incorporate knowledge of groups beyond euclidean ones?

Contacts

- Emails: edouard.oyallon@ens.fr, eugene.belilovsky@inria.fr
- Website + codes: https://edouardoyallon.github.io/