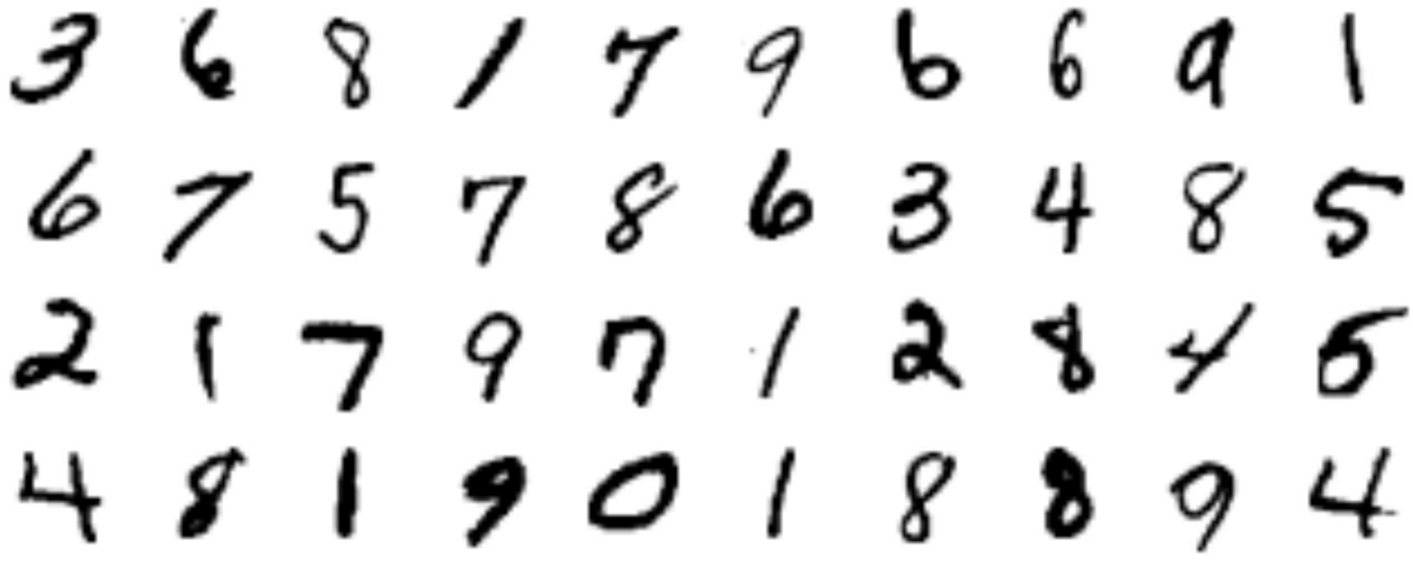
Image classification with Scattering and Dictionary learning

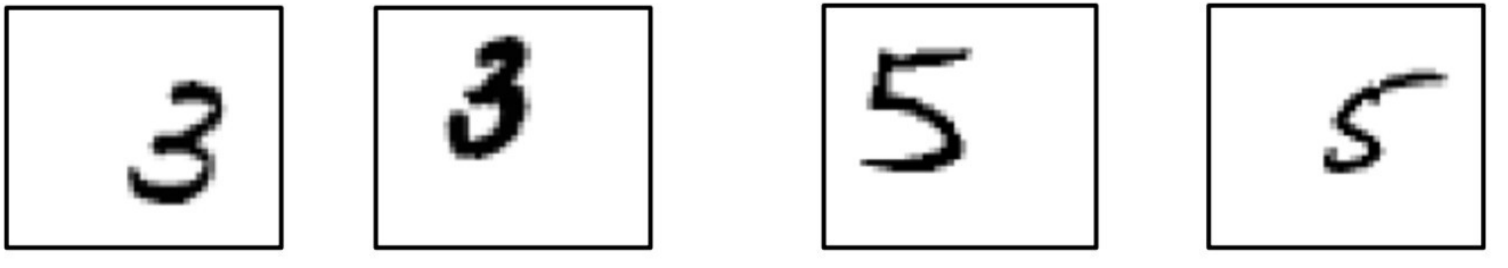
- https://arxiv.org/abs/1910.03561
- J. Zarka, L. Thiry, T. Angles, S. Mallat
- Pytorch code soon published on https://github.com/kymatio/kymatio

Digits classification

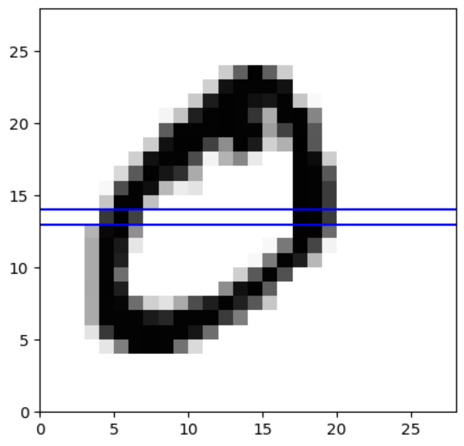
- MNIST database

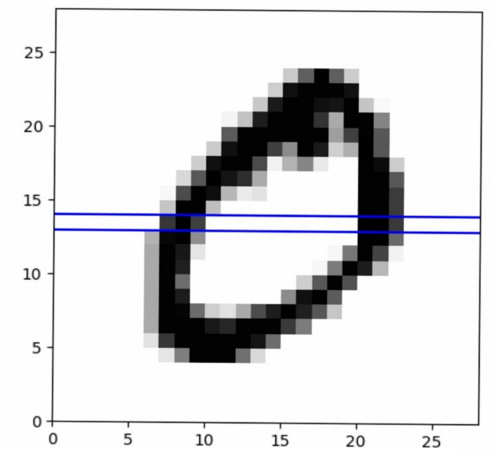


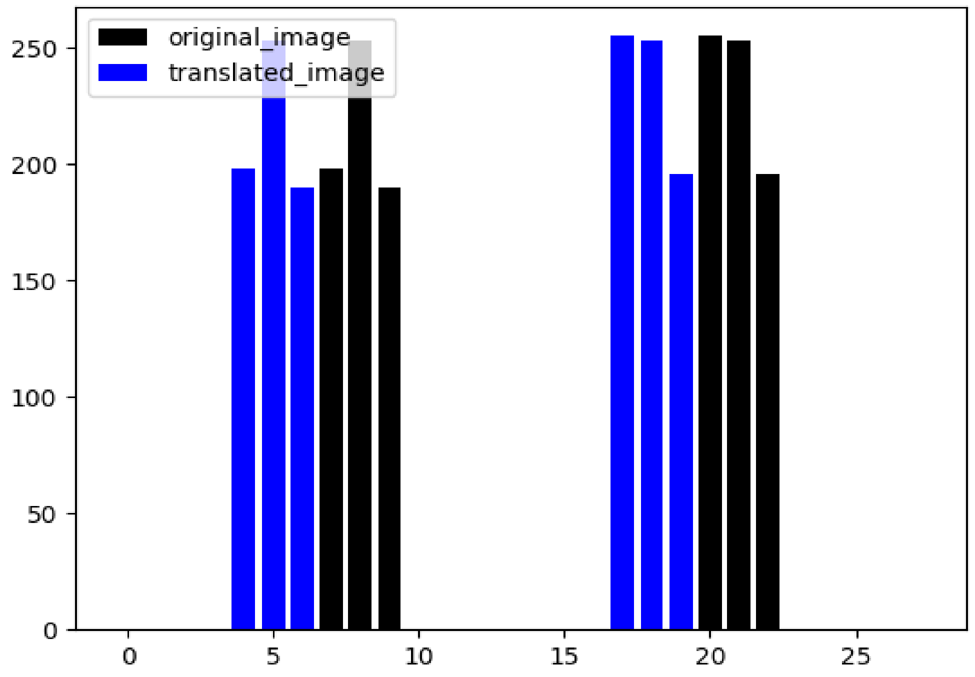
- Invariance to translations, stability to deformations



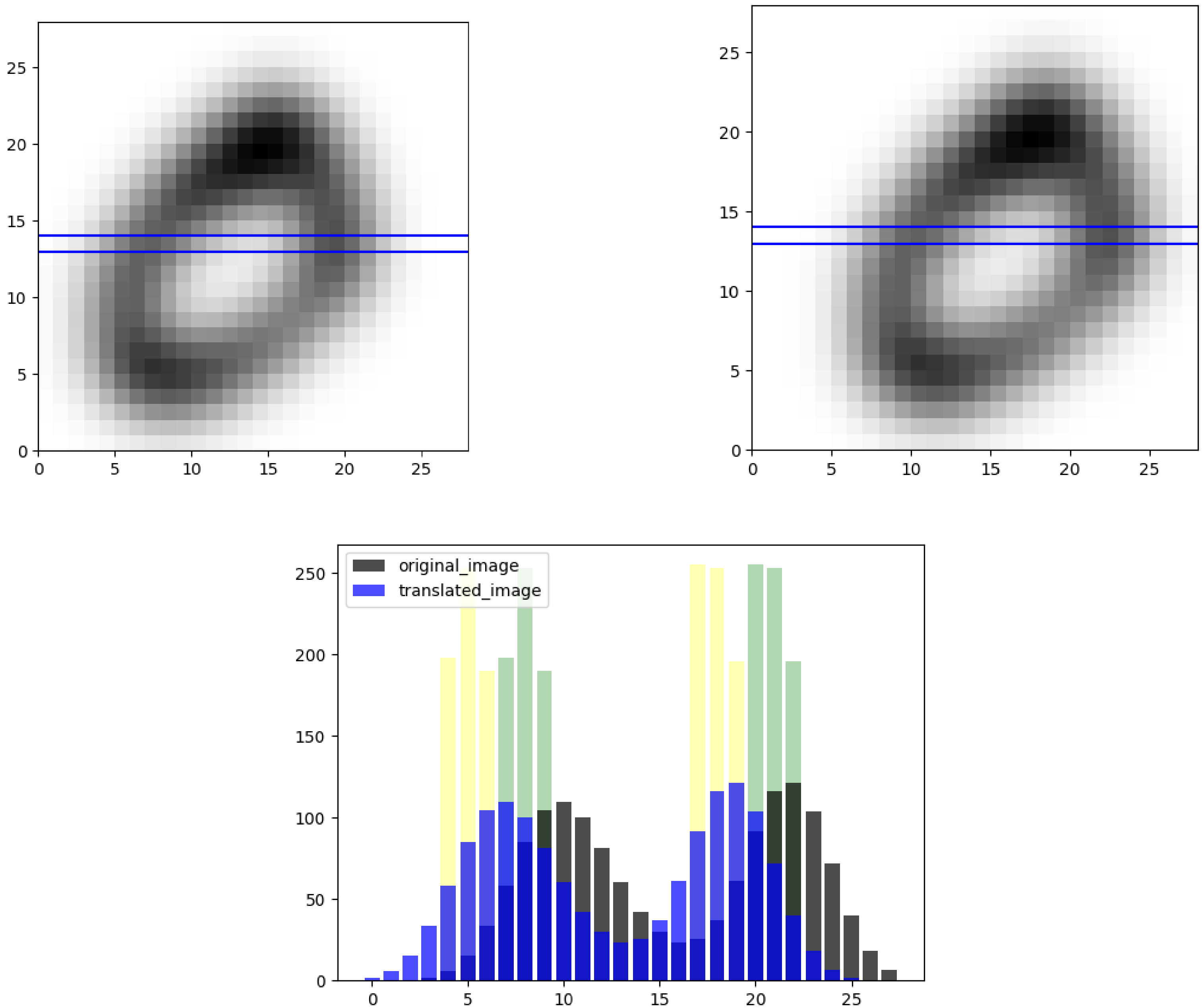
l_2 metric Instability to translations



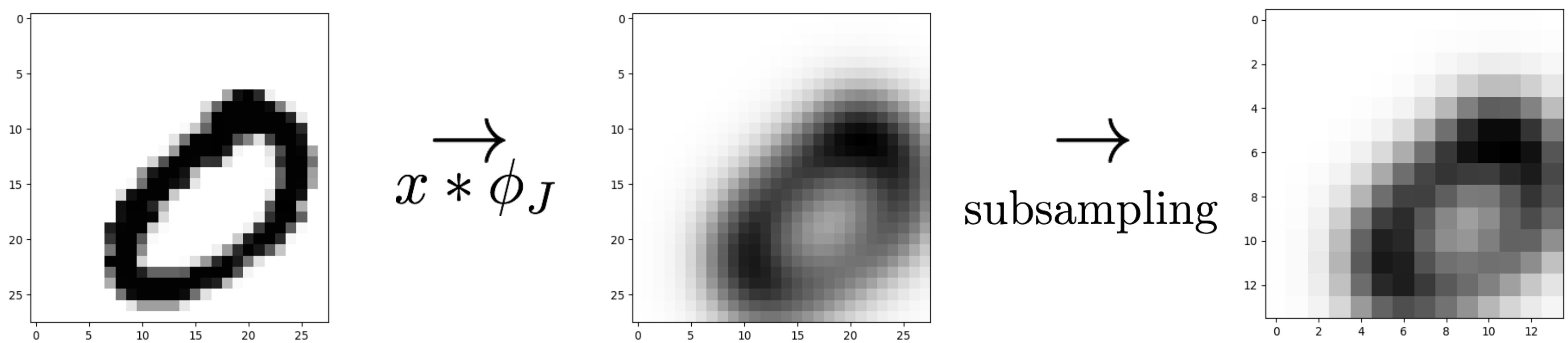




Averaging



Stability to geometric transformations

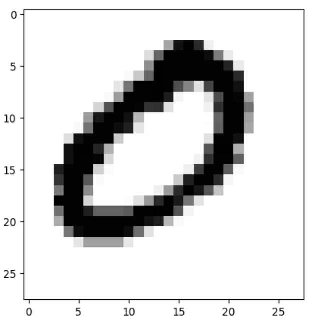


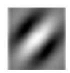
Convolution with Gaussian kernel ϕ_J :

- stable to geometric deformations
- dimensionality reduction via subsampling
- lots of details are lost

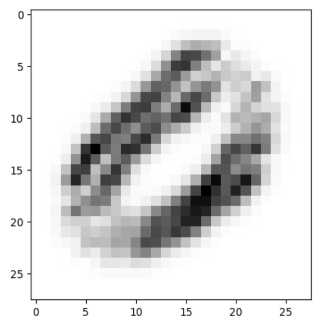
Preserving signal information

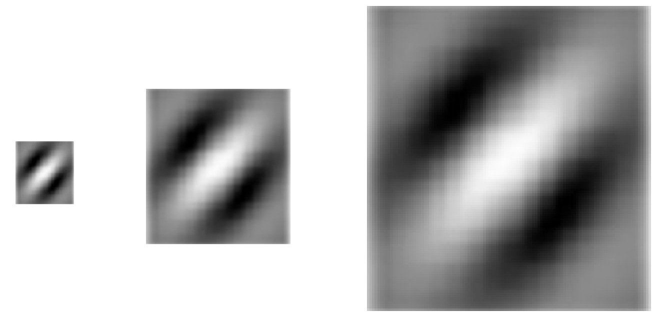
Recover information lost in averaging





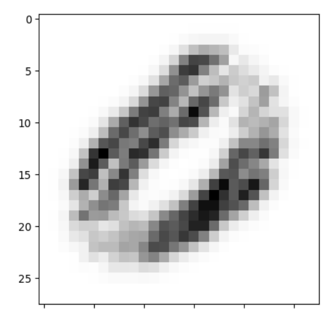
$$\longrightarrow \quad |x * \psi_{j,\theta}|$$



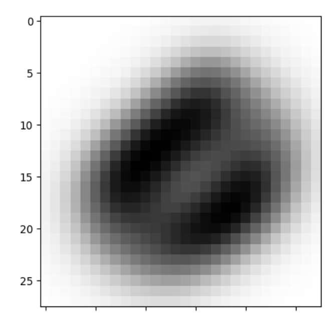


Gabor wavelets $\psi_{j,\theta}$

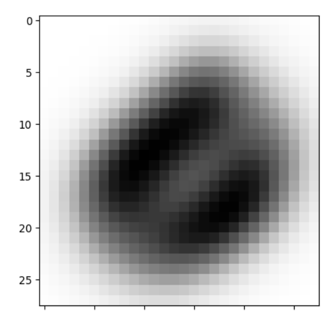
Stability to geometric transformations



$$\longrightarrow \quad |x * \psi_{j,\theta}| * \phi_J$$

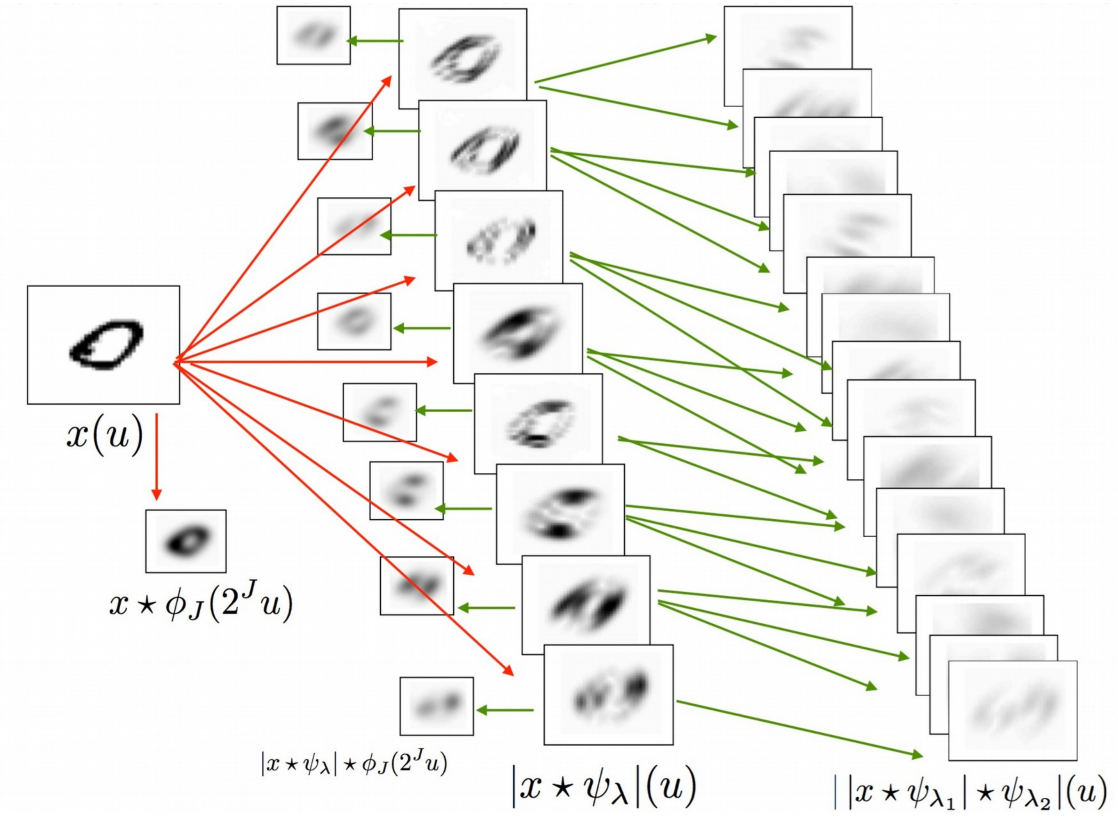


$\longrightarrow$ subsampling



Scattering transform

Mallat (2011), Mallat, Bruna (2012)

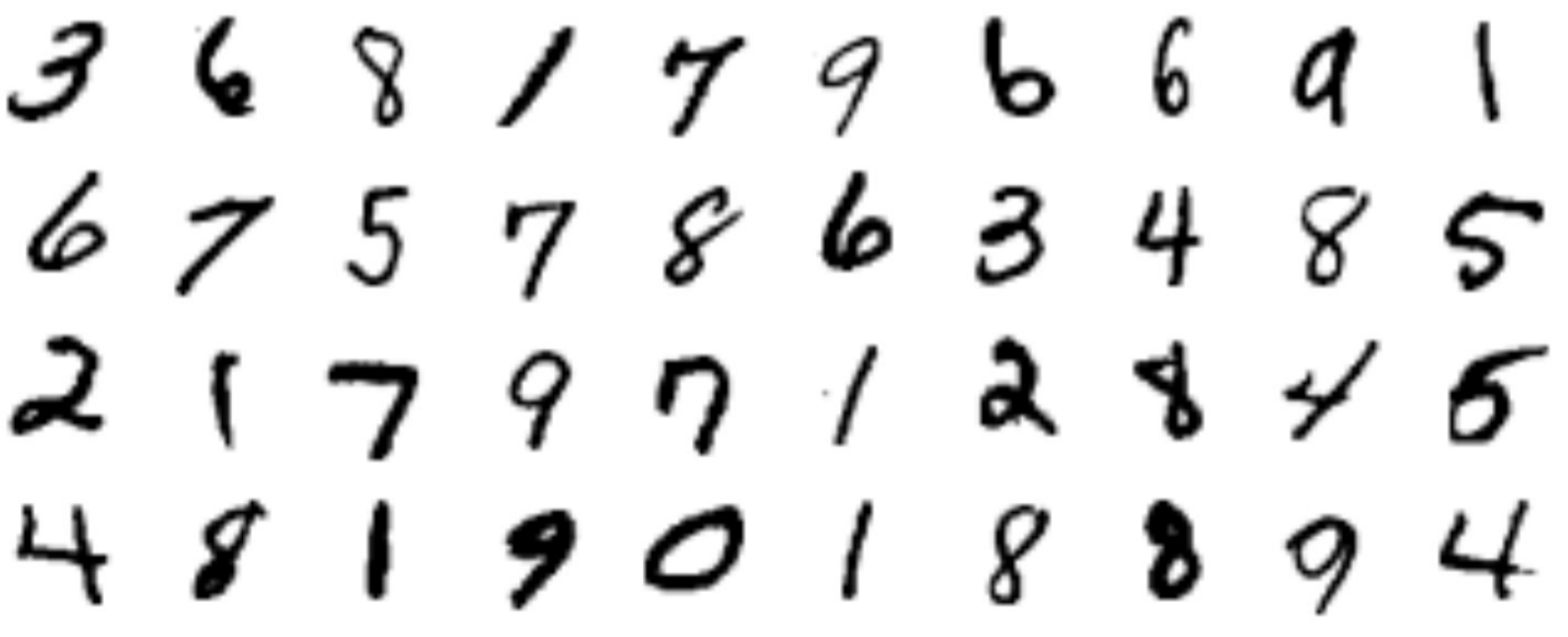


<u>Theorem</u>

$$\|Sx_\tau - Sx\| \leq K\|x\|\|\nabla\tau\|_\infty$$

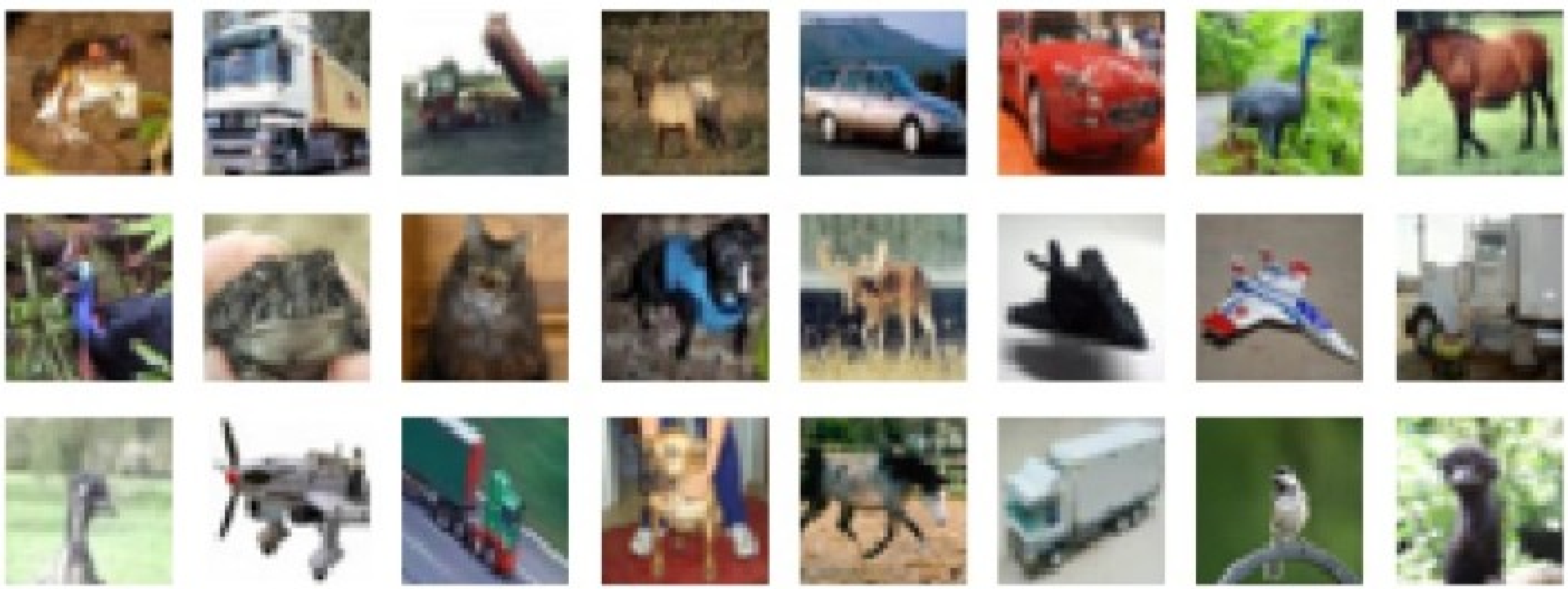
Scattering vs Deep ConvNets

Dataset	Scattering Transform	AlexNet	ResNet
MNIST 28^2 digit images 10 classes	>99 %	>99 %	>99 %



Scattering vs Deep ConvNets

Dataset	Scattering Transform	AlexNet	ResNet
CIFAR-10 32^2 object images 10 classes	84.7 %	89.1 %	95.5 %

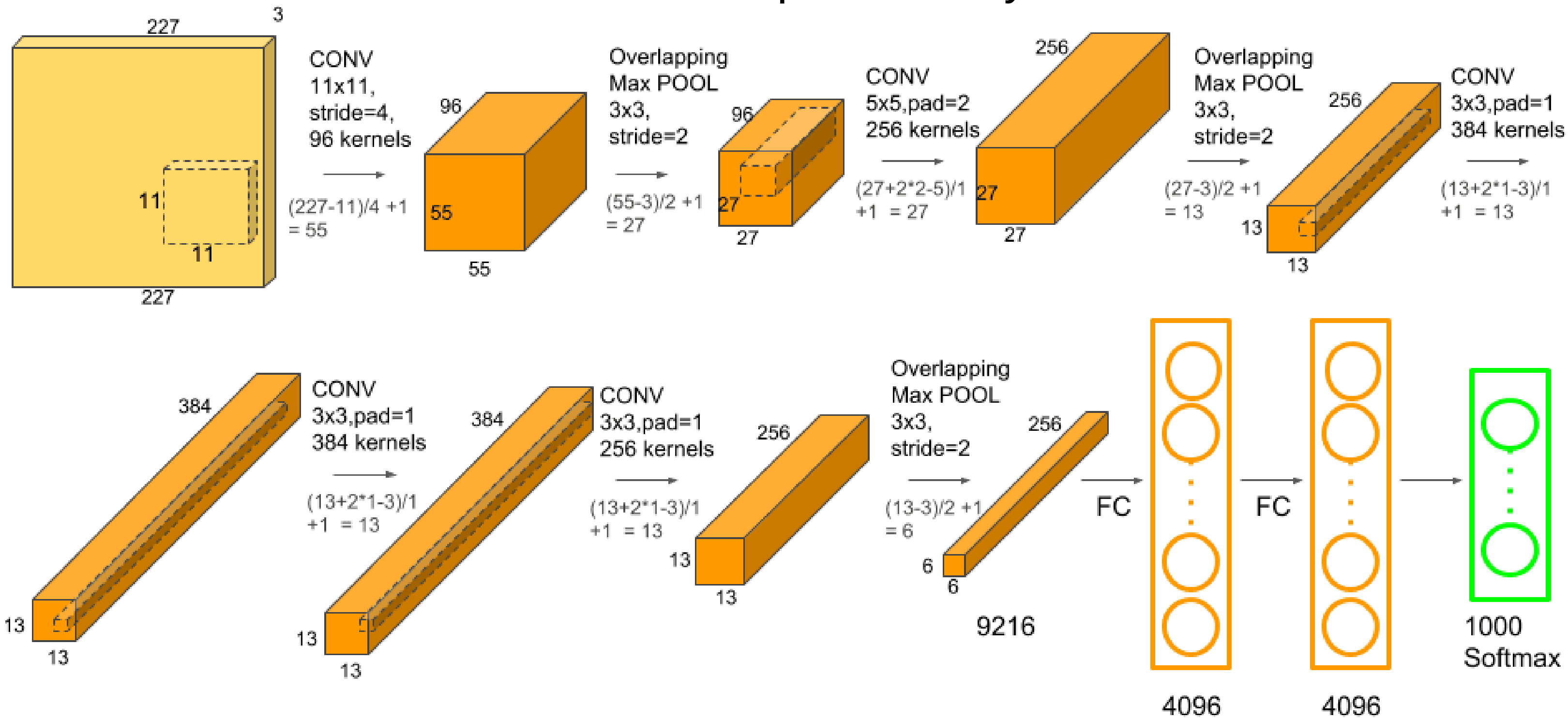


Scattering vs Deep ConvNets

Dataset	Scattering Transform	AlexNet	ResNet
ImageNet 224^2 object images 1000 classes	61.4 %	79.1 %	94.2 %

AlexNet

Krizhevsky et al. 2012
79.1 % top5 accuracy

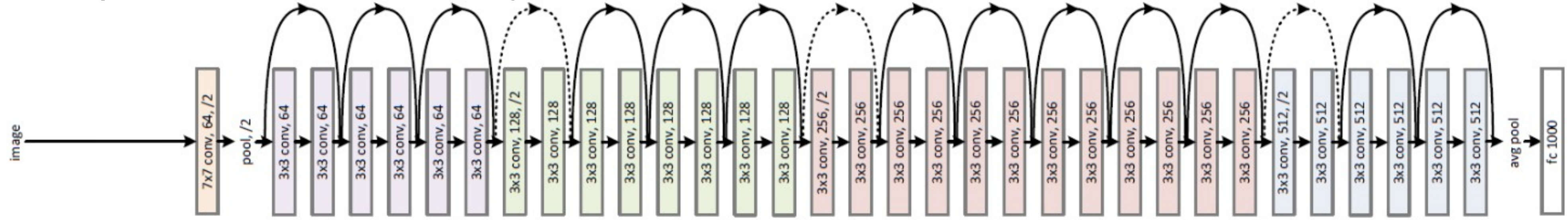


First layer learned filters

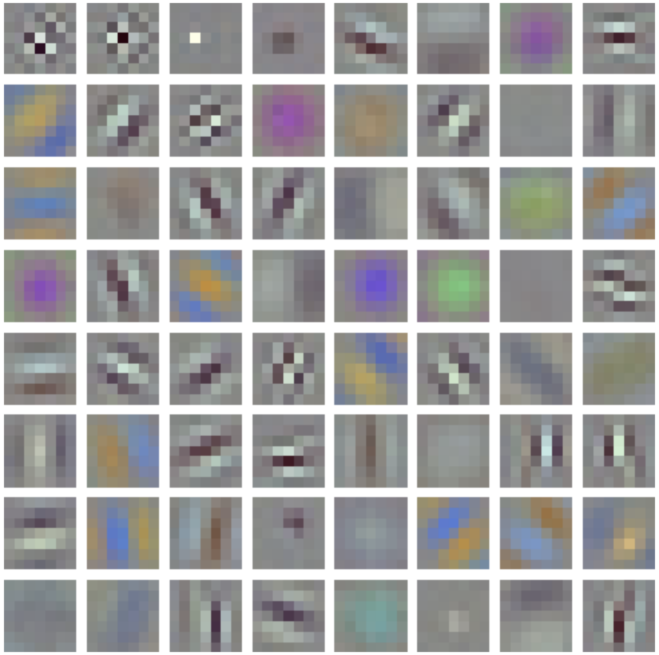
ResNet

He et al. 2016
94.2 % top5 accuracy

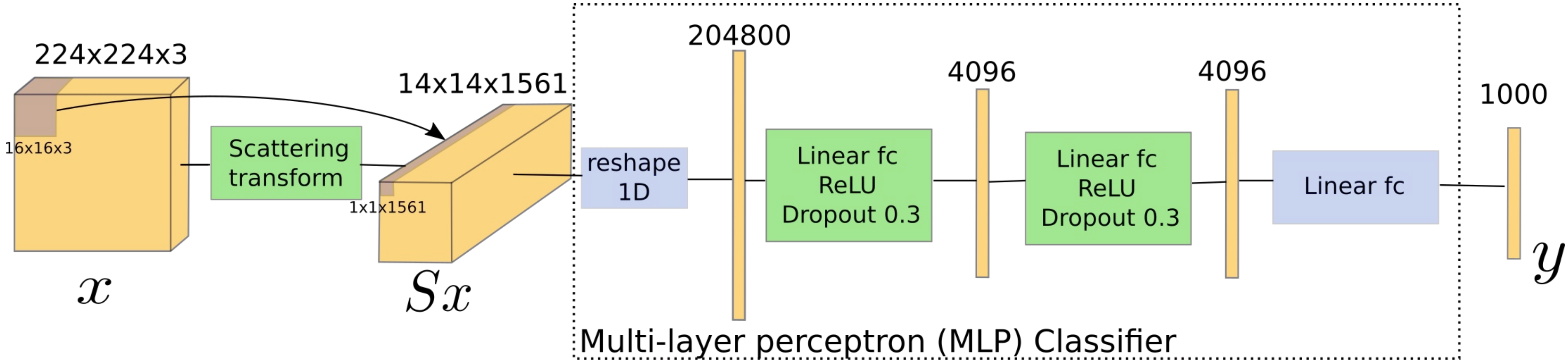
- skip connections
- up to 152 convolutional layers



First layer learned filters



Scattering : ImageNet classification



- Scattering transform has 16x16 receptive field
- Same MLP Classifier as AlexNet
- 38.1 % top1, 61.5 % top5

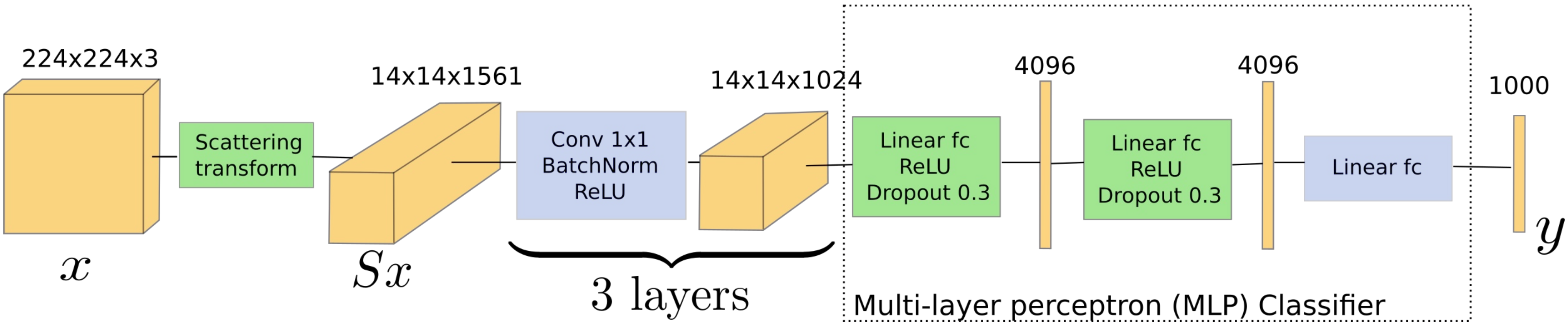
Research directions

What are in the convolutional layers of a Deep Networks ?
→ Visualizing and Understanding Convolutional Networks, Zeiler, Fergus 2014

What's needed to fill the gap between Scattering and DeepNets ?
→ Scaling the Scattering Transform: Deep Hybrid Networks, Oyallon et al. 2017

Shared Local Encoder

Oyallon et al. (2017)

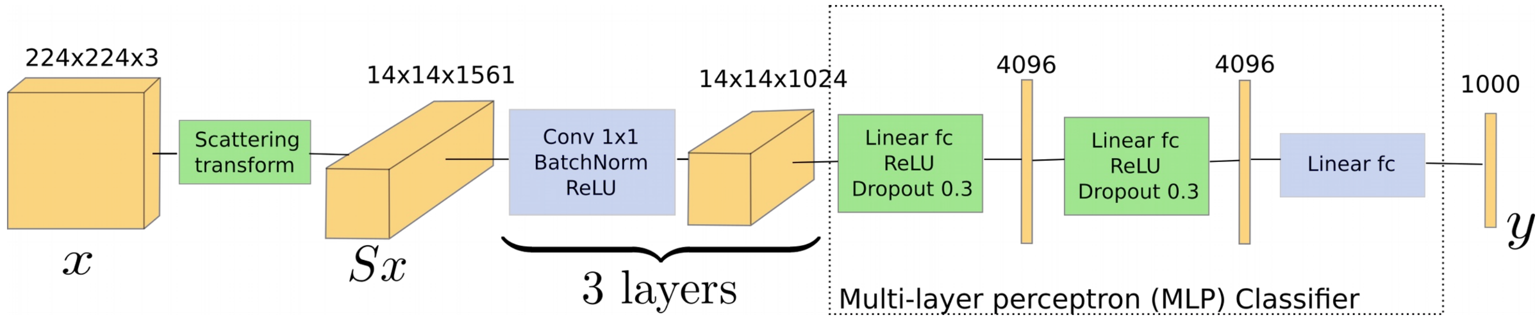


- Scattering transform
- 3 layers 1x1 convolution with ReLU and BatchNorm
- MLP Classifier

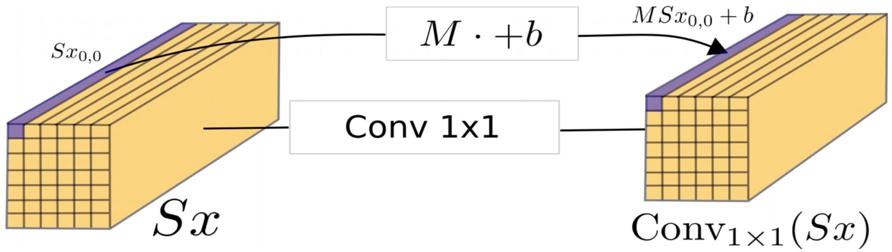
Performance on Imagenet 2012 :

- 57.8 % top1
- 79.6 % top5

l_1 sparse coding of Scattering



Convolution 1x1



Non negative sparse coding of Scattering

$$\alpha_*(D,\lambda,Sx) = \underset{\alpha\geq 0}{\operatorname{argmin}}\ \mathcal{L}(\alpha),\ \ \mathcal{L}(\alpha) = \|D\alpha - Sx\|_2^2 + \lambda\|\alpha\|_1$$

ISTA-like algorithm

$$\alpha_{n+1} = \mathrm{ReLU}\left((Id - \frac{1}{L}D^T D)\alpha_n + \frac{1}{L}D^T Sx - \frac{\lambda}{L}\right)$$

Supervised dictionary learning + LISTA

Mairal et al. (2008), Gregor and Lecun (2011)

SDL

$$\min_{D,\alpha,\theta} \mathcal{C}\left(y, f(\alpha, Sx, \theta)\right) + \lambda_0 \|D\alpha - Sx\|_2^2 + \lambda_1 \|\alpha\|_1$$

LISTA

$$\alpha_0 = 0,\ \alpha_{n+1} = \mathrm{st}_{\lambda_{n+1}}\left(VSx + W\alpha_n\right), n = 0 \ldots N-1$$

Performance

64% top1, 86% top5

l1 Convergence

$$\frac{\mathcal{L}(\alpha_N)}{\mathcal{L}(\alpha_*)} = 4.3$$

Task Driven dictionary learning + ISTC

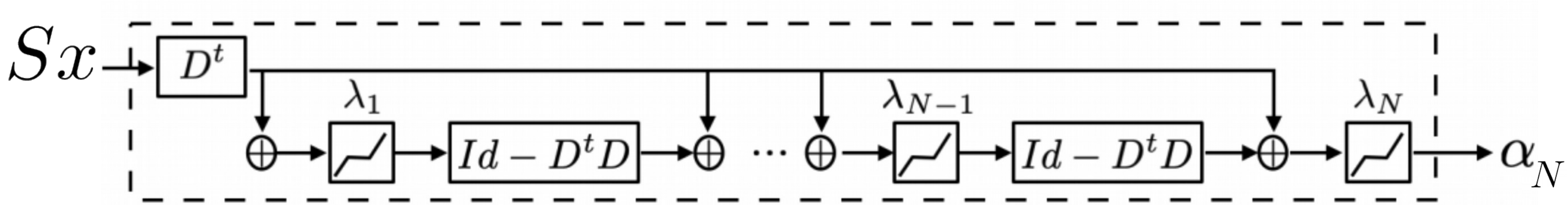
Mairal et al. (2011), ours

TDDL

$$\min_{D,\theta,\lambda} \mathcal{C}\left(y, f(\alpha_*(D,\lambda,Sx),\theta)\right)$$

ISTC

$$\alpha_0 = 0,\ \alpha_{n+1} = \mathrm{st}_{\lambda_\infty \gamma^n}\left((Id - D^T D)\alpha_n + D^T Sx\right),\ \lambda_\infty \gamma^N = \lambda$$



Theorem

$\cdot\ s$ support size of $\alpha*$

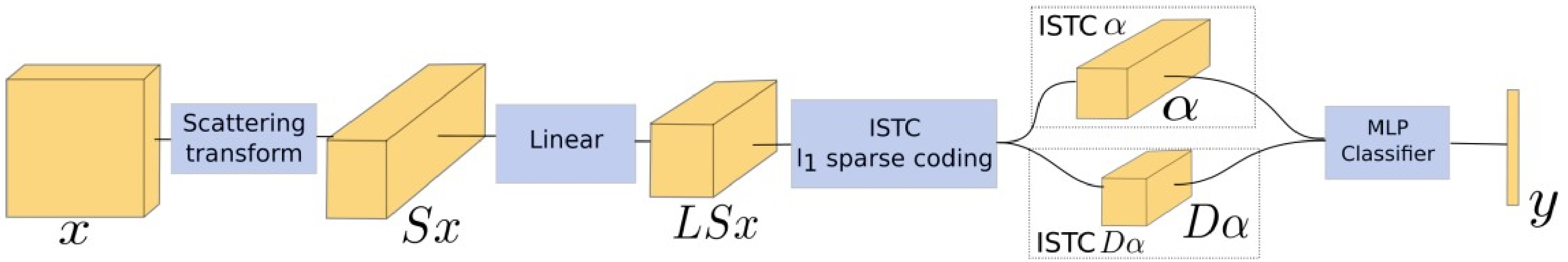
$\cdot\ \mu = \max_{m \neq m'} \langle D_m, D_{m'} \rangle$

If $s\mu \leq 1/2$ and $2s\mu < \gamma < 1$:

$$\|\alpha_n - \alpha_*\|_\infty \leq K\gamma^n$$

Scattering + ISTC classification

Architecture



ALISTA

$$\alpha_0 = 0,\ \alpha_{n+1} = \mathrm{st}_{\lambda_\infty \gamma^n}\left((Id - W^T D)\alpha_n + W^T Sx\right),\ \langle W_m, D_m \rangle = 1$$

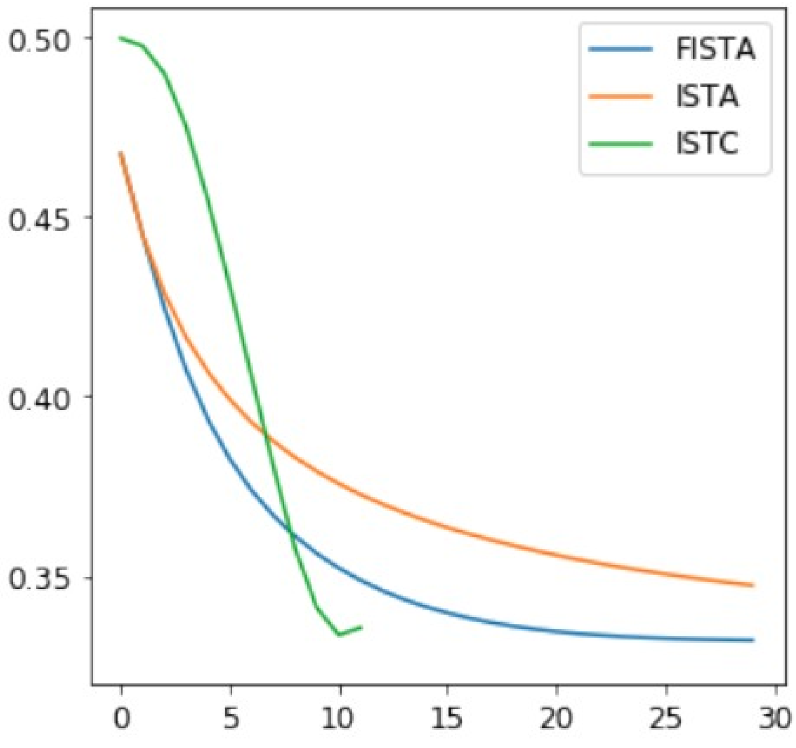
Results

l_1 algo	ISTC α	ISTC $D\alpha$	ALISTA
Top 1	58.0	55.4	61.9
Top 5	80.9	78.3	84.3

Convergence analysis

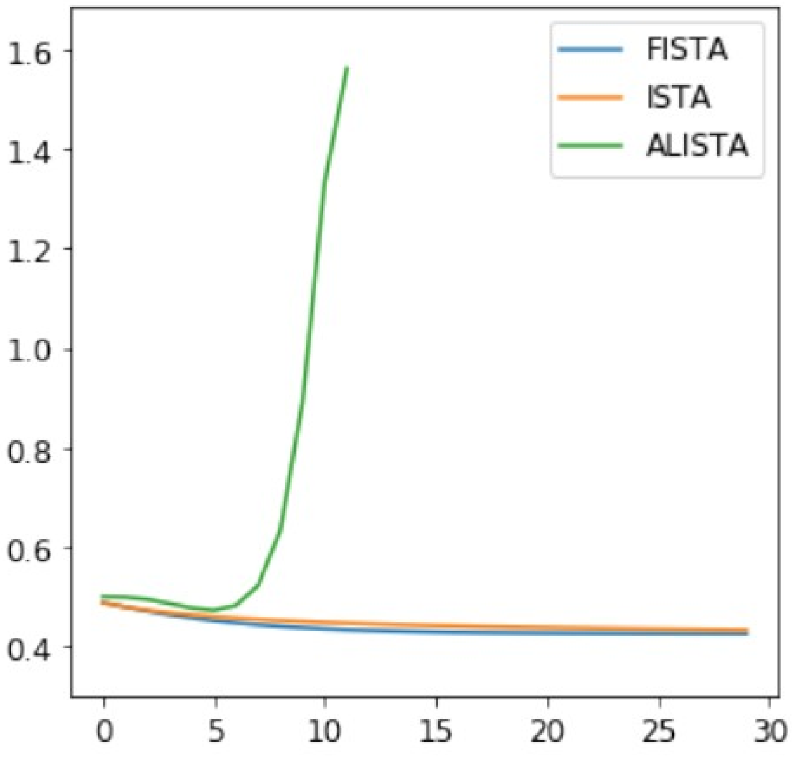
ISTC

$$\frac{\mathcal{L}(\alpha_N)}{\mathcal{L}(\alpha_*)} = 1.01$$



ALISTA

$$\frac{\mathcal{L}(\alpha_N)}{\mathcal{L}(\alpha_*)} = 3.8$$



Comments

- Improvement of 20% over Scattering alone
- Large factor λ_*, reconstruction error $\|Sx - D\alpha\| / \|Sx\| = 0.5$
- Hard to reconstruct the original image from α
- Classification works with the reconstruction $D\alpha$
  55 % top1, 78 % top5
- Atoms D_m are in Scattering space, can not be visualised like usual dictionary atoms
- ISTC can be replaced by FISTA (requires classifier retraining)
- Still far from ResNet performance
  → what's in LISTA with SDL ?

Questions ?