Image classification with Scattering and Dictionary learning

- 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
Convolution with Gaussian kernel $\phi_J$:

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

Recover information lost in averaging

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

Stability to geometric transformations

subsampling
Scattering transform

Mallat (2011), Mallat, Bruna (2012)

Theorem

$$\| Sx_\tau - Sx \| \leq K \| x \| \| \nabla \tau \|_\infty$$
## Scattering vs Deep ConvNets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Scattering Transform</th>
<th>AlexNet</th>
<th>ResNet</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNIST</td>
<td>&gt;99 %</td>
<td>&gt;99 %</td>
<td>&gt;99 %</td>
</tr>
<tr>
<td>28² digit images 10 classes</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![Handwritten MNIST digits]
# Scattering vs Deep ConvNets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Scattering Transform</th>
<th>AlexNet</th>
<th>ResNet</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIFAR-10 32² object images 10 classes</td>
<td>84.7 %</td>
<td>89.1 %</td>
<td>95.5 %</td>
</tr>
</tbody>
</table>

![Image with examples of CIFAR-10 images]
### Scattering vs Deep ConvNets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Scattering Transform</th>
<th>AlexNet</th>
<th>ResNet</th>
</tr>
</thead>
<tbody>
<tr>
<td>ImageNet</td>
<td>61.4 %</td>
<td>79.1 %</td>
<td>94.2 %</td>
</tr>
<tr>
<td>224² object images 1000 classes</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![Fish and Shark Images]
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
- Scattering transform
- 3 layers 1x1 convolution with ReLU and BatchNorm
- MLP Classifier

Performance on Imagenet 2012 :
- 57.8 % top1
- 79.6 % top5
$I_1$ sparse coding of Scattering

Convolution $1 \times 1$

Non-negative sparse coding of Scattering

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

ISTA-like algorithm

$$\alpha_{n+1} = \text{ReLU} \left( (I - \frac{1}{L} D^T D) \alpha_n + \frac{1}{L} D^T Sx - \frac{\lambda}{L} \right)$$
Supervised dictionary learning + LISTA

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

\textbf{SDL}

\[
\min_{D,\alpha,\theta} C(y, f(\alpha, Sx, \theta)) + \lambda_0 \|D\alpha - Sx\|_2^2 + \lambda_1 \|\alpha\|_1
\]

\textbf{LISTA}

\[
\alpha_0 = 0, \quad \alpha_{n+1} = \text{st}_{\lambda_{n+1}} (VSx + W\alpha_n), \quad n = 0 \ldots N - 1
\]

\textbf{Performance}

64\% top1, 86\% top5

\textbf{\textit{l}_1 \text{ Convergence}}

\[
\frac{\mathcal{L}(\alpha_N)}{\mathcal{L}(\alpha_*)} = 4.3
\]
Task Driven dictionary learning + ISTC

Mairal et al. (2011), ours

**TDDL**

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

**ISTC**

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

**Theorem**

- $s$ support size of $\alpha_*$
- $\mu = \max_{m \neq m'} \langle D_m, D_{m'} \rangle$

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

$$\|\alpha_n - \alpha_*\|_\infty \leq K \gamma^n$$
Scattering + ISTC classification

Architecture

\[ \alpha_0 = 0, \quad \alpha_{n+1} = \text{st}_{\lambda_{\infty}, \gamma'} \left( (I - W^T D)\alpha_n + W^T Sx \right), \quad \langle W_m, D_m \rangle = 1 \]

ALISTA

<table>
<thead>
<tr>
<th>I$_1$ algo</th>
<th>ISTC $\alpha$</th>
<th>ISTC $D\alpha$</th>
<th>ALISTA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 1</td>
<td>58.0</td>
<td>55.4</td>
<td>61.9</td>
</tr>
<tr>
<td>Top 5</td>
<td>80.9</td>
<td>78.3</td>
<td>84.3</td>
</tr>
</tbody>
</table>
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 $\lambda_*$, reconstruction error $\|Sx - D\alpha\|/\|Sx\| = 0.5$
• Hard to reconstruct the original image from $\alpha$
• 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
  $\rightarrow$ what’s in LISTA with SDL?
Questions ?