Compressing the Input for CNN with the First Order Scattering Transform

Edouard Oyallon, Eugene Belilovsky, Sergey Zagoruyko, Michal Valko
(1) CentraleSupélec (2) INRIA (3) Mila

Reduction the input size

- CNNs for images are typically fed with large images that have some redundant structures. Can we exploit this for reducing the input size?
- We propose to introduce a representation which:
  - Reduces the spatial resolution and dimensionality
  - Preserves the input and is predefined, for natural images

Gabor wavelets and modulus

- We consider the Gabor wavelets, that have a good trade-off between space and frequency localization.
- The enveloppe is more invariant to translations: ideal for .

First order Scattering Transform

\[ x \xrightarrow{W} A_{J} \]

- The first order scattering is the succession of a wavelet transform, a point modulus and a spatial averaging.
- It is similar to a SIFT with appropriate wavelets.

\[ Sx = \{ |x \ast \psi_{j,0}| \ast \phi_{j}, x \ast \phi_{j} \}_{0 \leq J} \]

It compresses the input image:

\[
\frac{\#Sx}{\#x} = \frac{(1 + \#J)}{2^J}
\]

<table>
<thead>
<tr>
<th>J</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compression ratio</td>
<td>2.2</td>
<td>1.1</td>
<td>0.39</td>
<td>0.13</td>
</tr>
</tbody>
</table>

Information preservation

- We propose a simple algorithm for reconstructing order via MSE minimization:
  \[ x = \arg \inf_{\tilde{x}} \| S\tilde{x} - Sy \| \]
- We observe that the first order Scattering does not lead to a significant loss when reconstructing:

\[ J = 3 \quad J = 4 \]

Classification performances

- We empirically observe that the loss of image details is due to the windowed averaging.

Detection performances

- We ablated the initial layers of a pre-trained ScatResNet.
- Our detection experiments demonstrate the spatial localisation of image details is preserved.

Pascal VOC7

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Speed (64 images)</th>
<th>Max Im</th>
<th>Speed (4 images)</th>
<th>Max Im (Coco)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faster-RCNN Order 1 + ScatResNet-50</td>
<td>128M</td>
<td>85.6</td>
<td>68.7</td>
<td>74.5</td>
</tr>
<tr>
<td>Faster – RCNN ResNet-50 (ours)</td>
<td>111M</td>
<td>87.7</td>
<td>67.7</td>
<td>56.2</td>
</tr>
<tr>
<td>Faster – RCNN ResNet-101 (ours)</td>
<td>120M</td>
<td>90.2</td>
<td>77.9</td>
<td>77.9</td>
</tr>
<tr>
<td>Faster-RCNN ResNet-50 VGG-16</td>
<td>120M</td>
<td>90.2</td>
<td>77.9</td>
<td>77.9</td>
</tr>
</tbody>
</table>

COCO

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Speed (4 images)</th>
<th>Max Im</th>
<th>Speed (4 images)</th>
<th>Max Im (Coco)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faster-RCNN Order 1 + ScatResNet-50</td>
<td>0.072</td>
<td>175</td>
<td>0.073</td>
<td>9</td>
</tr>
<tr>
<td>Faster – RCNN ResNet-50 (ours)</td>
<td>0.095</td>
<td>120</td>
<td>0.104</td>
<td>7</td>
</tr>
<tr>
<td>Faster – RCNN ResNet-101 (ours)</td>
<td>0.158</td>
<td>70</td>
<td>0.182</td>
<td>2</td>
</tr>
</tbody>
</table>

Conclusion

- Compress inputs and obtain a limited loss for supervised tasks
- Allows several memory and computation savings without learning.
- We applied no learning as the signals are natural images: can we learn better filters than wavelets for reducing a signal?

Related works

- Towards Image Understanding from Deep Compression Without Decoding, Tofasos et al., ICLR 2018
- Faster Neural Networks Straight from JPEG, Guo et al., ICLR workshop 2018