A Hybrid Wavelet Convolution Network with Sparse-Coding for Image Super-Resolution

Xing Gao, Hongkai Xiong

Department of Electronic Engineering
Shanghai Jiao Tong University
Sep. 28th, 2016
**Sparse representation:**

Describe a signal as a linear combination of a few atoms from a pre-specified dictionary.

\[ \mathbf{x} \approx \mathbf{D} \alpha, \]

s.t. \( \|\alpha\|_0 \ll m \)

where signal \( \mathbf{x} \in \mathbb{R}^m \), dictionary \( \mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_k] \in \mathbb{R}^{m \times k} \), \( \alpha \in \mathbb{R}^k \) is sparse.
Background——Sparse Representation

◆ The quest for a dictionary:

A good choice of the dictionary is crucial for the success of the sparse representation model.

Analytic dictionaries: DCT, Wavelet, etc.
- *Mathematical simplicity & Multi-scale & fast implementation;*
- *Limited expressiveness, simplistic for complex natural phenomena.*

Learned dictionaries: MOD, K-SVD, etc.
- *Adaptive to specific data, better performance;*
- *Complex, costly to deploy unknown inner structure.*

How to incorporate two types of dictionaries efficiently?
Background——Deep Learning

- Stacked convolution, pooling, and fully-connected layers
- **Superiority**
  - Hierarchical structure features with various abstract levels
  - **State-of-the-art** in most of computer vision fields: classification, recognition, etc.
- **Inferiority**
  - Over 100M parameters
  - Highly nonlinearity: the property and optimal structure **not well understood**

Challenges

- **Sparse representation:**
  - Dictionary: Fast implementation & Adaptivity
- **Deep learning:**
  - Optimize structure (reduce parameters) & Improve performance

A Hybrid Wavelet Convolution Network (HWCN)
Contributions

1. **Sparse representation:**
   
   Provide an *end-to-end* and *adaptive* framework to implement sparse coding and reconstruction.

2. **Deep learning:**
   
   Design a *novel structure* of deep networks, composed of predefined and learned convolution kernels, and it is *easier* to train. ← fewer parameters

3. **Image super-resolution:**
   
   Provide another approach to achieve *state-of-the-art* performance and also enhance current performance.
The Architecture of the HWCN

- A hybrid wavelet convolution network =

Scattering convolution network + Convolutional neural network

- Directional complex wavelets → Sparse coding candidates → Sparse code → HR image

Reconstruct

Select and weight
The Architecture of the HWCN

Input
(low-resolution image)

Sparse code candidates

Sparse code

Convolutional neural network

Output
(high-resolution image)
Scattering convolution network:

A revised version of invariant scattering convolution networks introduced by S. Mallat et al. in 2012.

- Convolutional filter banks are defined by two-dimensional complex directional wavelets:

\[ D = \{ \psi_{\lambda}(u) : 2^{-2j}\psi(2^{-j}r^{-1}u) \}_{\lambda \in \Lambda}, \]

with

\[ \Lambda = \{ \lambda = 2^{-j}r | j = 0, 1, \cdots, J - 1; r = 0, 2\pi/L, \cdots, 2(L - 1)\pi/L \} \]

- Each wavelet in D defines a filter of each node, and each node performs convolution and modulus operations.

- From the root node to each node forms a path:

\[ p = [\lambda_1, \lambda_2, \cdots, \lambda_m] \]

which produces a scattering map

\[ S_x[p](u) = ||x * \psi_{\lambda_1} | * \psi_{\lambda_2} | \cdots | \psi_{\lambda_m} |. \]
The Scattering Convolution Component

Scattering convolution network:

Frequency domain analysis

Convolutional filter banks:

\[ D = \{ \psi_\lambda(u) : 2^{-2j} \psi(2^{-j} r^{-1} u) \}_{\lambda \in \Lambda}, \]

Fourier transform

Real parts of filter banks

Imaginary parts of filter banks
The Scattering Convolution Component

*Scattering convolution network:*

- **Optimize structure**

- Frequency decreasing path:
  \[ |\lambda_{i+1}| < |\lambda_i| \]

- Modulus operation removes oscillations, shift signal to lower frequency, major energy of signal along frequency decreasing paths, **pruning**
The Convolution Neural Component

$S_x$-125 $\rightarrow$ Scattering maps $\rightarrow$ Conv9-50 $\rightarrow$ ReLU $\rightarrow$ Conv5-1 $\rightarrow$ $F_x$ $\rightarrow$ HR image

$l_0 - penalty$ $\rightarrow$ Sparsity $\rightarrow$ Reconstruction
Analysis

◆ Relationship to CNNs

Feature maps extracted by the first layer of CNN [1].

Feature maps extracted by the scattering part.

- **CNNs:**
  - Learned
  - Some ‘dead’ features

- **HWCN:**
  - Predefined and learned
  - Multiscale and multidirectional features
  - Features can be selected based on scale, direction, and frequency property

Analysis

◆ Relationship to typical dictionary based methods:

\[ x = D_l \alpha, \ y = D_h \alpha \]

- Dictionary based sparse coding methods:
  - Training dictionary:
  \[
  \min_{D_l, D_h} \frac{1}{2n} \sum_{i=1}^{n} \| D_h \alpha_i - y_i \|^2 \\
  \text{s.t.} \quad \alpha_i = \arg\min_z \| x_i - D_l \alpha_i \|^2 + \lambda \| z \|_1
  \]
  - Sparse coding:
  \[
  \alpha_i = \arg\min_z \| x_i - D_l \alpha_i \|^2 + \lambda \| z \|_1
  \]

- HWCN:
  - Scattering convolution network:
    • Predefined wavelet filter banks
    • Produce sparse code candidates
  - Convolution neural network:
    • Learned from data
    • Select and weight code candidates
    • Reconstruct from sparse code

Analytic dictionary ➔ Fast implementation
Learned dictionary ➔ Adaptivity

Fast implementation
Adaptivity
Application to Image Super-Resolution

◆ **Training set:**

91 natural images, the same as prior methods based on sparse representation [1~3].

◆ **Test sets:**

Set5 & Set14, the same as [3].

◆ **Loss function-MSE:**

\[
L(\theta) = \frac{1}{n} \sum_{i=1}^{n} \| F_{xi}(\theta) - y_i \| ^2,
\]

where \( F_{xi}, y_i \) represent reconstructed and ground truth HR image, respectively.

Table 1. The result of PSNR (dB) on the Set5 dataset.

<table>
<thead>
<tr>
<th>Set5</th>
<th>Scale</th>
<th>Bicubic</th>
<th>SC</th>
<th>K-SVD</th>
<th>NE+LLE</th>
<th>ANR</th>
<th>CNN</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>baby</td>
<td>3</td>
<td>33.870</td>
<td>34.258</td>
<td>35.038</td>
<td>35.018</td>
<td><strong>35.092</strong></td>
<td>34.968</td>
<td>34.903</td>
</tr>
<tr>
<td>bird</td>
<td>3</td>
<td>32.648</td>
<td>34.266</td>
<td>34.689</td>
<td>34.678</td>
<td>34.724</td>
<td>35.052</td>
<td><strong>35.504</strong></td>
</tr>
<tr>
<td>butterfly</td>
<td>3</td>
<td>24.064</td>
<td>25.675</td>
<td>26.000</td>
<td>25.809</td>
<td>25.964</td>
<td><strong>27.677</strong></td>
<td>27.648</td>
</tr>
<tr>
<td>head</td>
<td>3</td>
<td>32.824</td>
<td>33.150</td>
<td>33.508</td>
<td>33.556</td>
<td>33.590</td>
<td>33.499</td>
<td><strong>33.703</strong></td>
</tr>
<tr>
<td>woman</td>
<td>3</td>
<td>28.654</td>
<td>30.073</td>
<td>30.451</td>
<td>30.335</td>
<td>30.444</td>
<td>31.007</td>
<td><strong>31.260</strong></td>
</tr>
<tr>
<td>average</td>
<td>3</td>
<td>30.412</td>
<td>31.484</td>
<td>31.937</td>
<td>31.879</td>
<td>31.963</td>
<td>32.441</td>
<td><strong>32.604</strong></td>
</tr>
</tbody>
</table>
Application to Image Super-Resolution

Table 2. Average performance (PSNR dB) and time (s) on the Set14 dataset.

<table>
<thead>
<tr>
<th>Set14</th>
<th>Scale</th>
<th>Bicubic</th>
<th>SC</th>
<th>K-SVD</th>
<th>NE+LLE</th>
<th>ANR</th>
<th>CNN</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg time</td>
<td>3</td>
<td>-</td>
<td>80.8284</td>
<td>3.8231</td>
<td>4.9020</td>
<td><strong>0.7295</strong></td>
<td>2.1742</td>
<td>3.6492</td>
</tr>
</tbody>
</table>
Application to Image Super-Resolution

Original/PSNR

Bicubic/31.618dB

SC/32.601dB

K-SVD/32.945dB

NE+LLE/32.958dB

ANR/33.027dB

CNN/33.344dB

Proposed/33.472dB
Conclusion

◆ The hybrid wavelet convolution network

= scattering convolution network + convolutional neural network

= analytic dictionary + learned dictionary

Sparse & Multiscale Adaptive

= predefined convolution kernel + trainable convolutional kernel

Prior knowledge & Generalization Flexible & Powerful

◆ Follow-up

• Further optimize the structure of HWCN: contourlet?

• Application in other fields: image classification, recognition.
Q & A

Thanks
Table 3. The result of PSNR (dB) on the Set14 dataset.

<table>
<thead>
<tr>
<th>Set14</th>
<th>Scale</th>
<th>Bicubic</th>
<th>SC</th>
<th>K-SVD</th>
<th>NE+LLE</th>
<th>ANR</th>
<th>CNN</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>baboon</td>
<td>3</td>
<td>23.210</td>
<td>23.482</td>
<td>23.523</td>
<td>23.558</td>
<td>23.569</td>
<td>23.605</td>
<td><strong>23.617</strong></td>
</tr>
<tr>
<td>coastguard</td>
<td>3</td>
<td>26.715</td>
<td>27.054</td>
<td>27.168</td>
<td>27.146</td>
<td>27.174</td>
<td>27.196</td>
<td><strong>27.208</strong></td>
</tr>
<tr>
<td>comic</td>
<td>3</td>
<td>23.045</td>
<td>23.844</td>
<td>23.882</td>
<td>23.901</td>
<td>23.966</td>
<td>24.307</td>
<td><strong>24.405</strong></td>
</tr>
<tr>
<td>face</td>
<td>3</td>
<td>32.759</td>
<td>33.086</td>
<td>33.475</td>
<td>33.517</td>
<td>33.566</td>
<td>33.525</td>
<td><strong>33.710</strong></td>
</tr>
<tr>
<td>foreman</td>
<td>3</td>
<td>31.667</td>
<td>33.336</td>
<td>33.743</td>
<td>33.805</td>
<td>33.850</td>
<td>34.339</td>
<td><strong>34.759</strong></td>
</tr>
<tr>
<td>lenna</td>
<td>3</td>
<td>31.618</td>
<td>32.601</td>
<td>32.945</td>
<td>32.958</td>
<td>33.027</td>
<td>33.344</td>
<td><strong>33.472</strong></td>
</tr>
<tr>
<td>man</td>
<td>3</td>
<td>26.978</td>
<td>27.738</td>
<td>27.873</td>
<td>27.839</td>
<td>27.895</td>
<td>28.150</td>
<td><strong>28.302</strong></td>
</tr>
<tr>
<td>monarch</td>
<td>3</td>
<td>29.348</td>
<td>30.646</td>
<td>31.029</td>
<td>30.872</td>
<td>31.016</td>
<td>32.316</td>
<td><strong>32.349</strong></td>
</tr>
<tr>
<td>pepper</td>
<td>3</td>
<td>32.435</td>
<td>33.390</td>
<td>34.130</td>
<td>33.879</td>
<td>33.905</td>
<td>34.447</td>
<td><strong>34.720</strong></td>
</tr>
</tbody>
</table>