Invariant Scattering Convolution Networks

Joan Bruna and Stephane Mallat
Submitted to PAMI, Feb. 2012

Presented by Bo Chen

Other important related papers:
Agenda

- Translation invariance, deformation stability and high frequency preservation
- Fourier, SIFT and wavelet
- Wavelet scattering convolution networks
- Scattering properties
- Classification experimental results
- Conclusions
Notations

1. $\mathbb{Z}^d$ and $\mathbb{R}^d$ denote the set of integers and real vectors respectively.

2. $\mathbf{L}^2(\mathbb{R}^d)$ denotes the vector space of measurable, square-integrable $d$-dimensional functions $f(x)$

3. For $f(x) \in \mathbf{L}^2(\mathbb{R}^d)$ and $g(x) \in \mathbf{L}^2(\mathbb{R}^d)$, the inner product of $f(x)$ with $g(x)$ is written as $\langle g(u), f(u) \rangle = \int_{-\infty}^{+\infty} g(u)f(u)du$

4. The norm of $f(x)$ in $\mathbf{L}^2(\mathbb{R}^d)$ is $\|f\|^2 = \int_{-\infty}^{+\infty} |f(u)|^2du$ and the norm in $\mathbf{L}^1(\mathbb{R}^d)$: $\|f\|_1 = \int_{-\infty}^{+\infty} |f(u)|du$

5. the convolution of two functions: $f \star g(x) = (f(u) \star g(u))(x) = \int_{-\infty}^{+\infty} f(u)g(x-u)du$
Translation Invariance, Deformation Stability and High Frequency Preservation

• Translation Invariance:
  An operator $\Phi$ is invariant to global translations if $\Phi(L_c f(x)) = f(x - c)$ for all $c \in \mathbb{R}^d$ and $f \in \mathbf{L}^2(\mathbb{R}^d)$ if $\Phi(L_c f) = \Phi(f)$.

• Deformation:
  $L_\tau f(x) = f(x - \tau(x))$

  Stability in $\mathbf{L}^2(\mathbb{R}^d)$: $\forall (f, h) \in \mathbf{L}^2(\mathbb{R}^d)^2$, $\|\Phi(f) - \Phi(h)\|_\mathcal{H} \leq \|f - h\|$

  Lipschitz continuous to the action of diffeomorphisms:
  $\|\Phi(L_\tau f) - \Phi(f)\| \leq C \|f\| (\sup_{x \in \mathbb{R}^d} |\nabla \tau(x)|)$, where $\|f\|^2 = \int |f(x)|^2 dx$

  $\nabla \tau(x)$ is the deformation gradient tensor matrix.
  $|\nabla \tau(x)|$ measures the deformation amplitude at $x$. not deformation invariant, but continuous to deformations!

• High Frequency Preservation:
  It’s good to discriminate different types signals.
Fourier Transform

The Fourier transform of \( f(x) \in L^2(\mathbb{R}^d) \):

\[
\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(x) e^{-i\omega x} \, dx
\]

- FFT Basis functions: sinusoids
- FFT can only offer frequency information
- Loses time (location) coordinate completely
- Analyses the \textit{whole} signal
- A Fourier modulus is translation invariant but unstable with respect to deformations at high frequencies.

\[
\|\tilde{L}_\tau f - \hat{f}\| \sim s|\xi|\|\theta\| = \|\nabla\tau\|_\infty|\xi|\|f\|
\]

Arbitrarily large at a high frequency!

The central frequency of \( f(x) \)
1. SIFT computes the local sum of image gradient amplitudes among image gradients having nearly the same direction, in a histogram having 8 different direction bins.

2. SIFT does not have enough frequency and directional resolution to discriminate complex directional structures due to the ‘averaging pooling’.
Wavelet

“The wavelet transform is a tool that cuts up data, functions or operators into different frequency components, and then studies each component with a resolution matched to its scale” ----Dr. Ingrid Daubechies, Duke U

- A wavelet is a localized waveform and is stable to deformation. A wavelet transform computes convolutions with wavelets, which is translation covariant, not invariant.

Two-dimensional multiscale directional wavelet:

$$\psi_{2^j r}(u) = 2^{2j} \psi(2^j r^{-1} u)$$, where $r$ denotes the discrete rotations.

If the Fourier transform $\hat{\psi}(\omega)$ is centered at a frequency $\eta$ then $\hat{\psi}_{2^j r}(\omega) = \hat{\psi}(2^{-j} r^{-1} \omega)$ has a support centered at $2^j r \eta$, with a bandwidth proportional to $2^j$. To simplify notations, we denote $\lambda = 2^j r \in \Lambda = G \times \mathbb{Z}$, and $|\lambda| = 2^j$.

The scale plays the same role with the frequency. The finer the scale, the higher the frequency.
Multiresolution Wavelet Analysis (1)

- A sequence of embedded approximation subsets of $L^2(\mathbb{R})$:
  \[
  \{0\} \leftarrow \ldots V_{j-1} \subset V_j \subset V_{j+1} \ldots \rightarrow L^2(\mathbb{R}^d)
  \]
  with:
  
  \[f(u) \in V_j \iff f(2u) \in V_{j+1}\]
  
  \[f(u) \in V_0 \iff f(u-k) \in V_0, \quad k \in \mathbb{Z}\]
  
  \[\left(\sqrt{2^{-j}} \phi_{2j}(u - 2^{-j}k)\right)_{k \in \mathbb{Z}}\] forms an orthonormal basis of $V_{2j}$

- And a sequence of orthogonal complements, details’ subspaces:
  
  $O_{2j}$ such that $V_{2j+1} = V_{2j} \oplus O_{2j}$

  Orthonormal wavelet:
  
  \[\left(\sqrt{2^{-j}} \psi_{2j}(u - 2^{-j}k)\right)_{k \in \mathbb{Z}}\] forms an orthonormal basis of $O_{2j}$

The detail signal at $2^j$ is the difference of information between the approximation of function $f(u)$ at resolutions $2^{j+1}$ and $2^j$.

- $\phi$ is the scaling function and a low pass filter.
Multiresolution Wavelet Analysis (2)

Fast algorithms using filter banks

2D Orthogonal wavelet transform
Wavelet Transform

- As opposed to wavelet bases, a Littlewood-Paley wavelet transform is a redundant representation which computes convolution values without subsampling:

\[ \forall x \in \mathbb{R}^d \quad W[\lambda]f(x) = f \ast \psi_{\lambda}(x) = \int f(u) \psi_{\lambda}(x - u) \, du \]

- Its Fourier transform is

\[ \hat{W[\lambda]f}(\omega) = \hat{f}(\omega) \hat{\psi}_{\lambda}(\omega) = \hat{f}(\omega) \hat{\psi}(\lambda^{-1}\omega) \]

If \( f \) is real then \( \hat{f}(-\omega) = \hat{f}^*(\omega) \) and if \( \hat{\psi}(\omega) \) is real then \( W[-\lambda]f = W[\lambda]f^* \).

It is sufficient to compute \( W[2^jr]f \) for “positive” rotations \( r \in G^+ \).

- A wavelet transform at a scale \( 2^j \) only keeps wavelets of frequencies \( 2^j > 2^{-J} \). The low frequencies which are not covered by these wavelets are provided by an averaging over a spatial domain proportional to \( 2^J \):

\[ A_J f = f \ast \phi_{2J} \quad \text{with} \quad \phi_{2J}(x) = 2^{-dJ} \phi(2^{-J}x) \]

- If \( f \) is real then the wavelet transform \( W_J f = \left\{ A_J f, (W[\lambda]f)_{\lambda \in \Lambda_J} \right\} \) is indexed by \( \Lambda_J = \left\{ \lambda = 2^j r : r \in G^+, 2^j > 2^{-J} \right\} \). Its norm is

\[ \|W_J f\|^2 = \|A_J f\|^2 + \sum_{\lambda \in \Lambda_J} \|W[\lambda]f\|^2 \]
Scattering Wavelets (1)

- A wavelet transform commutes with translations, and is therefore not translation invariant. The main difficulty is to compute translation invariant coefficients, which remain stable to the action of diffeomorphisms, and retain high frequency information provided by wavelets.

\[
\text{If } \psi(x) = e^{i\eta \cdot x}\theta(x) \text{ then } \psi_\lambda(x) = e^{i\lambda \eta \cdot x}\theta_\lambda(x), \text{ and hence }
\]

\[
W[\lambda]f(x) = e^{i\lambda \eta \cdot x} \left( f^\lambda \ast \theta_\lambda(x) \right) \text{ with } f^\lambda(x) = e^{-i\lambda \eta \cdot x}f(x)
\]

**Note:** The convolution \( f^\lambda \ast \theta_\lambda \) is a low-frequency filtering because \( \hat{\theta}_\lambda(\omega) = \hat{\theta}(\lambda^{-1}\omega) \) covers a frequency ball centered at \( \omega = 0 \), of radius proportional to \( |\lambda| \).

The modulus maps \( W[\lambda]f \) into a lower frequency envelop:

\[
M[\lambda]W[\lambda]f = |W[\lambda]f| = |f^\lambda \ast \theta_\lambda|
\]

Let \( U[\lambda]f(x) = |f \ast \psi_\lambda(x)| dx \)

- The integration \( \int U[\lambda]f(x) \, dx = \int |f \ast \psi_\lambda(x)| \, dx \) is translation invariant but it removes all the high frequencies of \( |f \ast \psi_\lambda(x)| \). To recover these high frequencies, a scattering also computes the wavelet coefficients of each \( U[\lambda]f \): \( \{U[\lambda]f \ast \psi_\lambda'\}_\lambda \). Translation invariant coefficients are again obtained with a modulus \( U[\lambda']U[\lambda]f = |U[\lambda]f \ast \psi_\lambda'| \) and an integration \( \int U[\lambda']U[\lambda]f(x) \, dx \).
Scattering Wavelets (2)

- **Scattering propagator:**

**Definition 2.2** An ordered sequence \( p = (\lambda_1, ..., \lambda_m) \) with \( \lambda_k \in \Lambda_\infty = 2^\mathbb{Z} \times G^+ \) is called a path. The empty path is denoted \( p = \emptyset \). Let \( U[\lambda]f = |f \ast \psi_\lambda| \) for \( f \in L^2(\mathbb{R}^d) \). A scattering propagator is a path ordered product of non-commutative operators defined by

\[
U[p] = U[\lambda_m] \cdots U[\lambda_2] U[\lambda_1] \quad \text{with} \quad U[\emptyset] = \text{Id}.
\]

\[
U[p]f = |f \ast \psi_{\lambda_1} \ast \psi_{\lambda_2} \cdots \ast \psi_{\lambda_m}|
\]

Each \( U[\lambda] \) filters the frequency component in the band covered by \( \hat{\psi}_\lambda \), and maps it to lower frequencies with the modulus. The index sequence \( p = (\lambda_1, ..., \lambda_m) \) is thus a frequency path variable.

- A wavelet modulus propagator keeps the low-frequency averaging and computes the modulus of complex wavelet coefficients:

\[
U_J f(x) = \{ f \ast \phi_{2^j}(x), |f \ast \psi_\lambda(x)| \}_{\lambda \in \Lambda_J}
\]

**Note:** This transform has many similarities with the Fourier transform modulus, which is also translation invariant. However, a scattering is Lipschitz continuous to deformations as opposed to the Fourier transform modulus.
Scattering Wavelets (3)

- For classification, it is often better to compute localized descriptors which are invariant to translations smaller than a predefined scale $2^J$, while keeping the spatial variability at scales larger than $2^J$.

**Definition 2.4** Let $J \in \mathbb{Z}$ and $\mathcal{P}_J$ be the set of finite paths $p = (\lambda_1, \ldots, \lambda_m)$ with $\lambda_k \in \Lambda_J$ and hence $|\lambda_k| = 2^{j_k} > 2^{-J}$. A windowed scattering transform is defined for all $p \in \mathcal{P}_J$ by

$$S_J[p]f(x) = U[p]f \star \phi_{2J}(x) = \int U[p]f(u) \phi_{2J}(x - u) \, du$$

- The convolution with $\phi_{2J}(x) = 2^{-dJ} \phi(2^{-J} x)$ localizes the scattering transform over spatial domains of size proportional to $2^J$:

$$S_J[p]f(x) = | | f \star \psi_{\lambda_1} | \star \psi_{\lambda_2} | \cdots | \star \psi_{\lambda_m} | \star \phi_{2J}(x)$$

- It defines an infinite family of functions indexed by $\mathcal{P}_J$, denoted

$$S_J[\mathcal{P}_J]f = \{ S_J[p]f \}_{p \in \mathcal{P}_J}$$
Scattering Convolution Network (1)

- Recursively apply $U_J$ to each $U[p]f$

$$U_J U[p]f = \{ U[p]f \ast \phi_{2^J}(x), |U[p]f \ast \psi_{\lambda}(x)| \}$$

and $S_J[p]f = U[p]f \ast \phi_{2^J}(x)$, it results that

$$U_J U[\Lambda^m_J]f = \{ U_J U[p]f \}_{p \in \Lambda^m_J} = \{ S_J[\Lambda^m_J]f, S_J[\Lambda^{m+1}_J]f \}$$

This implies that $S_J[p]f$ can be computed along paths of length $m \leq m_{max}$ by first calculating $U_J f = \{ S_J[\phi]f, U[\Lambda^1_J]f \}$ and iteratively applying $U_J$ to each $U[\Lambda^m_J]f$ for increasing $m \leq m_{max}$. Let $\Lambda^m_J$ be the set of all paths $p = (\lambda_1, ..., \lambda_m)$ of length $m$

- For appropriate wavelets, first order coefficients are equivalent to SIFT coefficients.
Scattering Convolution Network (2)

\[ S_J[\emptyset]f = f \ast \phi_J \]

\[ S_J[\lambda_1]f \]

\[ S_J[\lambda_1, \lambda_2]f \]

\[ U[\lambda_1]f \]

\[ U[\lambda_1, \lambda_2]f \]

\[ \Omega[2^{j_1}r_1, 2^{j_2}r_2] \]

\( m = 0 \)

\( m = 1 \)

\( m = 2 \)

\( m = 3 \)
The Comparison to Existing Deep Convolution Networks

1. A scattering network outputs coefficients $S_J[p].f$ at all layers $m \leq m_{\text{max}}$, and not just at the last layer $m_{\text{max}}$

2. The filters are not learned from data but are predefined wavelets, which are stable with respect to deformations and provide sparse image representations.

3. The averaging by $\phi_{2J}$ at the output is also a pooling operator. The high frequencies lost by the averaging are recovered as wavelet coefficients in the next layers.

4. It loses the phase of these wavelet coefficients. This phase may however be recovered from the modulus thanks to the wavelet transform redundancy.
Scattering Properties

- Both of propagator and scattering transform are nonexpansive

\[ \| U_J f - U_J h \|^2 = \| A_J f - A_J h \|^2 + \sum_{\lambda \in \Lambda_J} \| W[\lambda] f - W[\lambda] h \|^2 \leq \| W_J f - W_J h \|^2 \leq \| f - h \|^2 \]

\( \forall (f, h) \in L^2(\mathbb{R}^d)^2 \), \( \| S_J[\mathcal{P}_J] f - S_J[\mathcal{P}_J] h \| \leq \| f - h \| \)

- Energy conservation

\[ \| S_J f \|^2 = \sum_{m=0}^{\infty} \| S_J[\Lambda_J^m] f \|^2 = \sum_{m=0}^{\infty} \sum_{p \in \Lambda_J^m} \| S_J[p] f \|^2 = \| f \|^2 \]

\( \forall f \in L^2(\mathbb{R}^d) \), \( \lim_{m \to \infty} \| U[\Lambda_J^m] f \|^2 = \lim_{m \to \infty} \sum_{n=m}^{\infty} \| S_J[\Lambda_J^n] f \|^2 = 0 \)

- Modulus projects wavelet coefficients to lower frequencies.

If \( \lambda = 2^j r \) then \( |f \ast \psi_\lambda(x)| \ast \psi_\lambda \) for \( \lambda' = 2^{j'} r' \) is non-negligible only if \( \psi_{\lambda'} \) is located at low frequencies and hence if \( 2^{j'} < 2^j \).

- Stability to deformation

\[ \| S_J(L_\tau f) - S_J f \| \leq C m_{max} \| f \| \| \nabla \tau \|_\infty \]
Two images having same first order scattering coefficients, but the top image is piecewise regular and hence has wavelet coefficients which are much more sparse than the uniform texture at the bottom. As a result the top image has second order scattering coefficients of larger amplitude than at the bottom.
Energy Concentration on Limited Depth

This table gives the percentage of scattering energy \( \| S_J (\Lambda^m_J x) \|^2 / \| x \|^2 \) captured by frequency-decreasing paths of length \( m \), as a function of \( J \). These are averaged values computed over normalized images with \( \int x(u) du = 0 \) and \( \| x \| = 1 \), in the Caltech-101 database. The scattering is computed with cubic spline wavelets.

<table>
<thead>
<tr>
<th>( J )</th>
<th>( m = 0 )</th>
<th>( m = 1 )</th>
<th>( m = 2 )</th>
<th>( m = 3 )</th>
<th>( m = 4 )</th>
<th>( m \leq 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>95.1</td>
<td>4.86</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>99.96</td>
</tr>
<tr>
<td>2</td>
<td>87.56</td>
<td>11.97</td>
<td>0.35</td>
<td>-</td>
<td>-</td>
<td>99.89</td>
</tr>
<tr>
<td>3</td>
<td>76.29</td>
<td>21.92</td>
<td>1.54</td>
<td>0.02</td>
<td>-</td>
<td>99.78</td>
</tr>
<tr>
<td>4</td>
<td>61.52</td>
<td>33.87</td>
<td>4.05</td>
<td>0.16</td>
<td>0</td>
<td>99.61</td>
</tr>
<tr>
<td>5</td>
<td>44.6</td>
<td>45.26</td>
<td>8.9</td>
<td>0.61</td>
<td>0.01</td>
<td>99.37</td>
</tr>
<tr>
<td>6</td>
<td>26.15</td>
<td>57.02</td>
<td>14.4</td>
<td>1.54</td>
<td>0.07</td>
<td>99.1</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>73.37</td>
<td>21.98</td>
<td>3.56</td>
<td>0.25</td>
<td>98.91</td>
</tr>
</tbody>
</table>
Scattering Stationary Processes

- The role of $L^2(\mathbb{R}^d)$ norm is replaced by the mean square norm $E(|X(x)|^2)^{1/2}$ on stationary stochastic process, which does not depend upon $x$ and is thus denoted $E(|X|^2)^{1/2}$.
- Convolutions as well as a modulus preserve stationarity. If $X(x)$ is stationary, it results that $U[p]X(x)$ is also stationary and its expected value thus does not depend upon $x$.

**Definition 4.1** The expected scattering transform of a stationary process $X$ is defined for all $p = (\lambda_1, ..., \lambda_m) \in \mathcal{P}_\infty$ by

$$\overline{S}X(p) = E(U[p]X) = E(|X \ast \psi_{\lambda_1} \ast ... \ast \psi_{\lambda_m}|).$$

Fig. 5. Two different textures having the same Fourier power spectrum. (a) Textures $X(u)$. Top: Brodatz texture. Bottom: Gaussian process. (b) Same estimated power spectrum $\hat{R}X(\omega)$. (c) Nearly same scattering coefficients $S_J[p]X$ for $m = 1$ and $2^J$ equal to the image width. (d) Different scattering coefficients $S_J[p]X$ for $m = 2$. 
Classification on MNIST

Fig. 7. (a): Image $X(u)$ of a digit ‘3’. (b): Array of scattering vectors $S_J[p]X(u)$, for $m = 1$ and $u$ sampled at intervals $2^J = 8$. (c): Scattering vectors $S_J[p]X(u)$, for $m = 2$.

### MNIST classification results

<table>
<thead>
<tr>
<th>Training size</th>
<th>PCA</th>
<th>SVM</th>
<th>PCA</th>
<th>SVM</th>
<th>PCA</th>
<th>SVM</th>
<th>PCA</th>
<th>SVM</th>
<th>PCA</th>
<th>SVM</th>
<th>PCA</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>14.5</td>
<td>15.4</td>
<td>7.35</td>
<td>7.4</td>
<td>5.7</td>
<td>8</td>
<td>4.7</td>
<td>5.6</td>
<td>7.18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>7.2</td>
<td>8.2</td>
<td>3.74</td>
<td>3.74</td>
<td>2.35</td>
<td>4</td>
<td>2.3</td>
<td>2.6</td>
<td>3.21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>5.8</td>
<td>6.5</td>
<td>2.99</td>
<td>2.9</td>
<td>1.7</td>
<td>2.6</td>
<td>1.3</td>
<td>1.8</td>
<td>2.53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5000</td>
<td>4.9</td>
<td>4</td>
<td>2.34</td>
<td>2.2</td>
<td>1.6</td>
<td>1.6</td>
<td>1.03</td>
<td>1.4</td>
<td>1.52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10000</td>
<td>4.55</td>
<td>3.11</td>
<td>2.24</td>
<td>1.65</td>
<td>1.5</td>
<td>1.23</td>
<td>0.88</td>
<td>1</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20000</td>
<td>4.25</td>
<td>2.2</td>
<td>1.92</td>
<td>1.15</td>
<td>1.4</td>
<td>0.96</td>
<td>0.79</td>
<td>0.58</td>
<td>0.76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40000</td>
<td>4.1</td>
<td>1.7</td>
<td>1.85</td>
<td>0.9</td>
<td>1.36</td>
<td>0.75</td>
<td>0.74</td>
<td>0.53</td>
<td>0.65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60000</td>
<td>4.3</td>
<td>1.4</td>
<td>1.80</td>
<td>0.8</td>
<td>1.34</td>
<td>0.62</td>
<td>0.7</td>
<td>0.43</td>
<td>0.53</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Percentage of errors on an MNIST rotated dataset

<table>
<thead>
<tr>
<th>Scat. $m_{\text{max}} = 1$ PCA</th>
<th>Scat. $m_{\text{max}} = 2$ PCA</th>
<th>Conv. Net.</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>4.4</td>
<td>8.8</td>
</tr>
</tbody>
</table>

### Percentage of errors on scaled and/or rotated MNIST

<table>
<thead>
<tr>
<th>Transformed Images</th>
<th>Scat. $m_{\text{max}} = 1$ PCA</th>
<th>Scat. $m_{\text{max}} = 2$ PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without</td>
<td>1.6</td>
<td>0.8</td>
</tr>
<tr>
<td>Rotation</td>
<td>6.7</td>
<td>3.3</td>
</tr>
<tr>
<td>Scaling</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Rot. + Scal.</td>
<td>12</td>
<td>5.5</td>
</tr>
</tbody>
</table>
Texture Discrimination

Classes: 61
Image: 200x200
Training: 46 each class

Percentage of errors on CURET for different training sizes.

<table>
<thead>
<tr>
<th>Training size</th>
<th>X PCA</th>
<th>Four. Spectr. PCA</th>
<th>Scat. $m_{\text{max}} = 1$ PCA</th>
<th>Scat. $m_{\text{max}} = 2$ PCA</th>
<th>Textons SVM [14]</th>
<th>MRF [36]</th>
</tr>
</thead>
<tbody>
<tr>
<td>46</td>
<td>17</td>
<td>1</td>
<td>0.5</td>
<td>0.2</td>
<td>1.53</td>
<td>2.4</td>
</tr>
</tbody>
</table>
Conclusions

- A wavelet scattering transform
- Good properties: translation invariant, stable to deformation and high frequency preservation
- Realization via deep convolution network architecture.
- State-of-the-art classification results are obtained for handwritten digit recognition and texture discrimination, with an SVM or a PCA classifier.