Wavelet Scattering Transforms

Haixia Liu

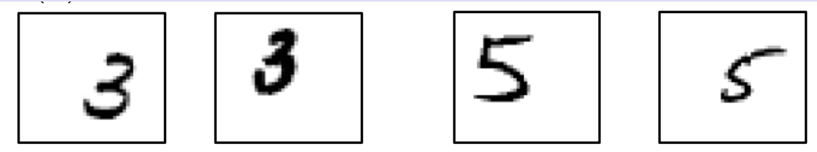
Department of Mathematics
The Hong Kong University of Science and Technology

February 6, 2018

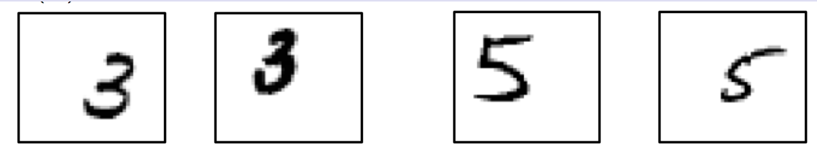
Outline

1. Problem
   - Dataset
   - Problem
   - two subproblems
   - outline of image classification problem
2. Wavelet Scattering Transform
   - Review of Multiscale Wavelet Transform
   - Why Wavelets?
   - Wavelet Convolutional Networks
3. Digit Classification: MNIST by Joan Bruna *et al.*
4. MATLAB code of Wavelet convolutional Networks

Digit classification

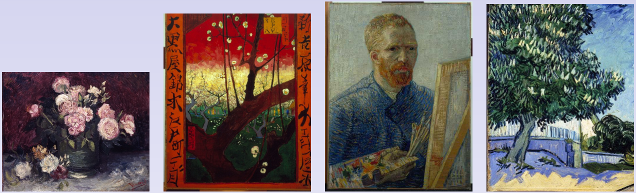


Digit classification



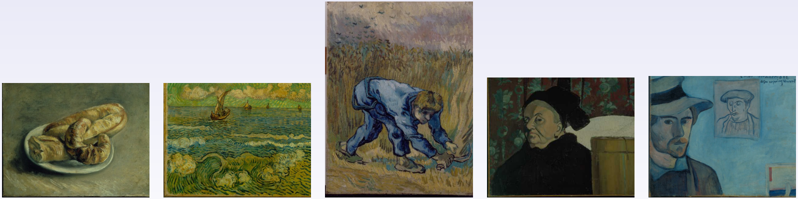
- Translation
- Deformation

Dataset



(a) f249 (b) f371 (c) f522 (d) f752

Figure: van Gogh's paintings.



(a) f253a (b) f418 (c) f687 (d) s205 (e) s206v

Figure: Forgeries.

The Problem

- 79 paintings authenticated by experts
- 64 genuine paintings and 15 forgeries
- Forgeries are ‘quite’ genuine with 6 historically wrongly attributed to van Gogh
- High-resolution professional images provided by van Gogh Museum and Kröller-Müller Museum
- Design an algorithm to determine if a painting is from van Gogh or NOT

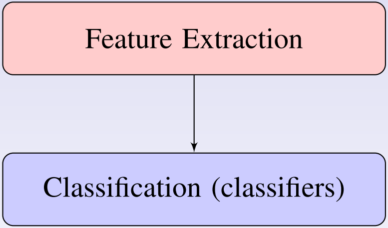
Image classification can be contributed to the following two subproblems:

- Feature extraction (image processing),
  - Fourier Transform,
  - Wavelet,
  - EMD,
  - Tight frame
  - ...

Image classification can be contributed to the following two subproblems:

- Feature extraction (image processing),
  - Fourier Transform,
  - Wavelet,
  - EMD,
  - Tight frame
  - ...
- Clustering or classification (data analysis).
  - SVM,
  - HMM,
  - ...

Image Classification



Aims

AIM: Classify correctly although translation and deformation, i.e.,

- Globally invariant to the translation group
- Locally invariant to small deformation

Wavelet Scattering Transform

Aims

AIM: Classify correctly although translation and deformation, i.e.,

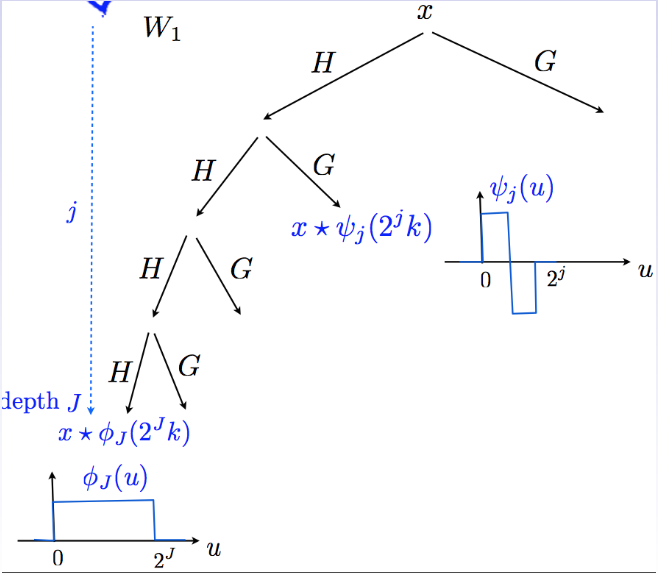
- Globally invariant to the translation group
- Locally invariant to small deformation

Wavelet Scattering Transform

Some advantages of Wavelet Scattering Transform:

- Share hierarchical structure of DNNs
- replace data-driven filters by wavelets
- have strong theoretical support
- better performance for small-sample data

Haar wavelet transform



Haar Filtering

$$\begin{tikzcd}[column sep=small]
& \{x(u)\}_{u \leq d} \arrow[dl, "H"'] \arrow[dr, "G"] & \\
\left\{ \frac{x(2u) + x(2u+1)}{\sqrt{2}} \right\}_{u \leq d/2} & & \left\{ \frac{x(2u) - x(2u+1)}{\sqrt{2}} \right\}_{u \leq d/2}
\end{tikzcd}$$

$$Hx(u) = x * h(2u) \text{ and } Gx(u) = x * g(2u)$$

where h is a low frequency and g is a high frequency.

Review of Multiscale Wavelet Transform

wavelet filters $\{\psi_\lambda\}_\lambda$

- Dilated Wavelets: $\psi_\lambda(t) = 2^j\psi(2^jt)$ with $\lambda = 2^j$.
- Multiscale and oritented wavelet filters

$$\psi_\lambda = 2^j\psi(2^j\theta x)$$

where $\theta \in \mathcal{R}(\mathbb{R}^2)$ be a rotation matrix and $\lambda = (2^j, \theta)$.

$$x * \psi_\lambda(\omega) = \int x(u)\psi_\lambda(\omega - u) \Rightarrow \widehat{x * \psi_\lambda}(\omega) = \hat{x} \cdot \widehat{\psi_\lambda}$$

- Wavelet transform:

$$Wx = \begin{bmatrix} x * \phi_{2^J(t)} \\ x * \psi_\lambda(t) \end{bmatrix}_{\lambda \leq 2^J}$$

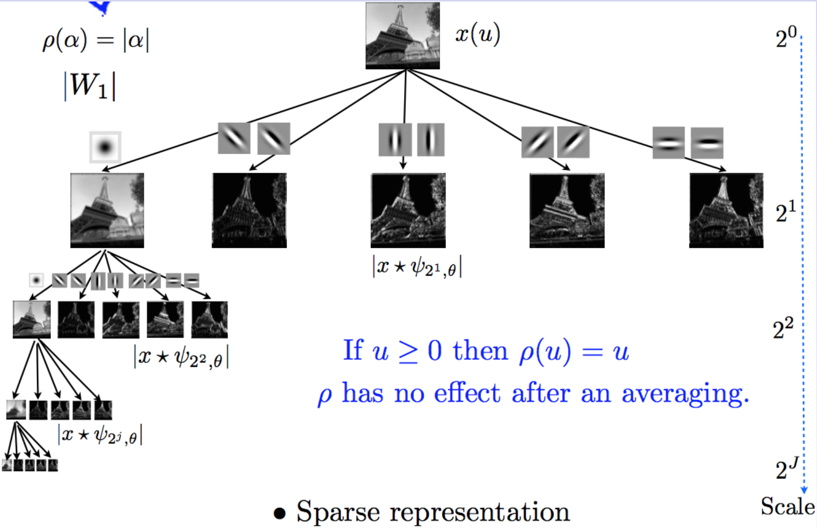
Advantages of Wavelets

- Wavelets separate multiscale information
- Wavelets provide sparse representation
- Wavelets are uniformly stable to deformations.
  If $\psi_{\lambda,\tau} = \psi_\lambda(t - \tau(t))$, then

$$\|\psi_\lambda - \psi_{\lambda,\tau}\| \leq C \sup_t |\nabla \tau|$$

- Modulus improves invariance
- Fourier transform on translated function, modulus lead to translation invariance

$$|W|x = \begin{bmatrix} x * \phi_{2^J(t)} \\ |x * \psi_\lambda(t)| \end{bmatrix}_{\lambda \leq 2^J}$$



If $u \geq 0$ then $\rho(u) = u$

ρ has no effect after an averaging.

- Sparse representation

Scattering Coefficients

- first-layer scattering coefficients

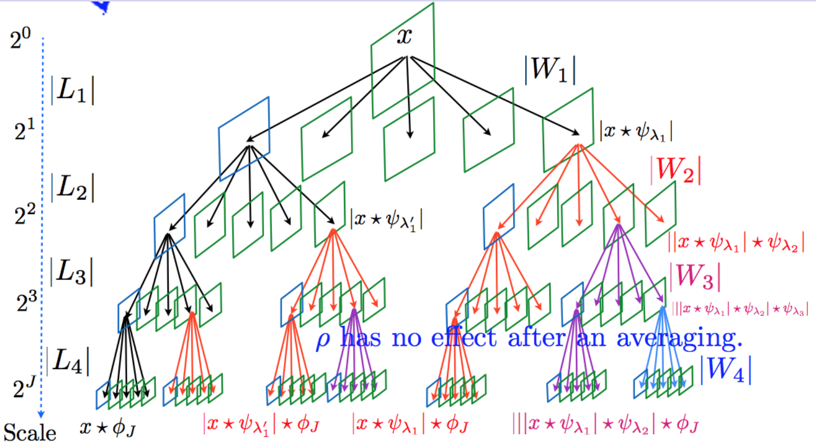
$$S_{1,J}((\lambda_1),x) = |X * \psi_{\lambda_1}| * \phi_J(x)$$

- second-layer scattering coefficients

$$S_{2,J}((\lambda_1,\lambda_2),x) = ||X * \psi_{\lambda_1}| * \psi_{\lambda_2}| * \phi_J(x)$$

- *m*-th layer scattering coefficients

$$S_{2,J}((\lambda_1,\lambda_2,\cdots,\lambda_m),x) = ||X * \psi_{\lambda_1}|\cdots * \psi_{\lambda_m}| * \phi_J(x)$$



$$S_4 x = |L_4|\,|L_3|\,|L_2|\,|L_1| x = |W_4|\,|W_3|\,|W_2|\,|W_1|\, x$$

Renormalization

$$\tilde{S}_{1,J}((\lambda_1)) = S_{1,J}((\lambda_1))$$

and

$$\tilde{S}_{2,J}((\lambda_1, \lambda_2)) = \frac{S_{2,J}((\lambda_1, \lambda_2))}{S_{1,J}((\lambda_1))}$$

Paper *Deep Scattering Spectrum* points out second coefficients can be decorrelated to increase their invariance through a renormalization.

Features based on Scattering Coefficients

One choice is to take spatial averages of scattering coefficients

$$\bar{S}_{m,J} = \sum_x \tilde{S}_{m,J}((\lambda_1,\cdots,\lambda_m),x).$$

- dimension reduction
- destroy the spatial information contained in scattering coefficients

Classifiers

There are a lot of classifiers can be used if features are extracted

- Logistic regression
- Random forest
- SVM
- LDA
- Sparse SVM
- Sparse LDA
- and so on $\cdots$

Numerical results

Training size	x PCA	x SVM	Wind. Four. PCA	Wind. Four. SVM	Scat. $m_{\max}=1$ PCA	Scat. $m_{\max}=1$ SVM	Scat. $m_{\max}=2$ PCA	Scat. $m_{\max}=2$ SVM	Conv. Net.
300	14.5	15.4	7.35	7.4	5.7	8	**4.7**	5.6	7.18
1000	7.2	8.2	3.74	3.74	2.35	4	**2.3**	2.6	3.21
2000	5.8	6.5	2.99	2.9	1.7	2.6	**1.3**	1.8	2.53
5000	4.9	4	2.34	2.2	1.6	1.6	**1.03**	1.4	1.52
10000	4.55	3.11	2.24	1.65	1.5	1.23	0.88	1	**0.85**
20000	4.25	2.2	1.92	1.15	1.4	0.96	0.79	**0.58**	0.76
40000	4.1	1.7	1.85	0.9	1.36	0.75	0.74	**0.53**	0.65
60000	4.3	1.4	1.80	0.8	1.34	0.62	0.7	**0.43**	0.53

Figure: Results from paper *Invariant Scattering Convolution Networks*

Software

Code can be downloaded from
`http://www.di.ens.fr/data/software/`.

Thank you!!!