Learning generative texture models with extended Fields-of-Experts

Nicolas Heess, Chris Williams and Geoffrey Hinton

BMVC 2009
Overview and Motivation

- Generative models of natural image structure
- Investigation of a particular class of MRFs: Field-of-Experts (FoE; Roth & Black, 2005)
  - continuous-valued, high-order MRF
  - fully parametric
  - all parameters can be learned from data
- Test case: Modeling image texture
  - texture is an important aspect of natural images
    - images as compositions of multiple texture regions
  - suitable for understanding the “generative power” of a probabilistic model

[www.cgtextures.com]
Example texture
This talk in a nutshell

Example texture

Sample from standard FoE model

(trained on texture)
This talk in a nutshell

Example texture

Sample from standard FoE model (trained on texture)

Sample from extended FoE model (trained on texture)
Outline

- Field of Experts (FoE)
- Extended Field-of-Experts model
- Experiments: texture synthesis
- Experiments: texture inpainting
- Discussion
The Field-of-Experts model (Roth & Black, 2005)

- **Field of Experts**: High order MRF with potentials defined in terms of the responses of linear filters.
- PDF for a FoE with a single expert:

\[
p(x) = \frac{1}{Z} \prod_{i=1}^{N} \Phi(w^T x_{(i)})
\]

- clique centered at each pixel \(i = 1 \ldots N\);
- \(x_{(i)}\): image patch centered at pixel \(i\)
- \(w\): filter
- \(\Phi(y)\): expert nonlinearity
- For multiple experts (\(M\): # of experts):

\[
p(x) = \frac{1}{Z} \prod_{i=1}^{N} \prod_{j=1}^{M} \Phi_j(w_j^T x_{(i)})
\]
Field-of-Experts model (cont’d)

PDF defined by the FoE:

\[
p_{FoE}(\mathbf{x}; \Theta) = \frac{1}{Z} \prod_{i=1}^{N} \prod_{j=1}^{M} \Phi \left( \mathbf{w}_j^T \mathbf{x}(i); \theta_j \right)
\]

\(\mathbf{x}:\) image; \(i: \) index pixels; \(j: \) index experts; \(\mathbf{w}_j: \) filter; \(\Phi: \) expert nonlinearity (potential function)

“Standard” FoE uses (simplified) Student-t potentials:

\[
\Phi_{FoE}(\mathbf{y}; \nu) = \left(1 + \frac{1}{2} \mathbf{y}^2\right)^{-\nu}
\]

Thus:

\[
p_{FoE}(\mathbf{x}) = \frac{1}{Z} \exp(-E(\mathbf{x}))
\]

\[
E_{FoE}(\mathbf{x}) = \sum_i \sum_j \nu_j \log \left\{ 1 + \frac{1}{2} \left( \mathbf{w}_j^T \mathbf{x}(i) \right)^2 \right\}
\]

Note: \(p_{FoE}(\mathbf{x})\) is unimodal for standard FoE model
"Standard" FoE uses (simplified) Student-t potentials:

$$\Phi_{FoE}(y; \nu) = \left(1 + \frac{1}{2}y^2\right)^{-\nu}$$

$\nu > 0$: expert parameter

Extended FoE with bimodal potentials (BiFoE):

$$\Phi_{BiFoE}(y; a, b, \nu) = \left\{1 + \frac{1}{2} \left[(y + b)^2 + a\right]^2\right\}^{-\nu}$$

$$E_{BiFoE}(x) = \sum_i \sum_j \nu_j \log \left\{1 + \frac{1}{2} \left[\left(w_j^T x(i) + b_j\right)^2 + a_j\right]^2\right\}$$

$\nu > 0$: expert parameter; $a$: mode distance; $b$: center position
Why does \( \Phi \) matter?
Why does $\Phi$ matter?

Standard FoE

potential function of single expert

global probability distribution defined by multiple experts

$(M=1$ 1D experts$)$
Why does $\Phi$ matter?

Standard FoE

potential function of single expert

global probability distribution defined by multiple experts

$(M=2 \text{ 1D experts})$
Why does $\Phi$ matter?

Standard FoE

potential function of single expert

$\begin{align*}
x_2 \\
x_1
\end{align*}$

(M=3 1D experts)

global probability distribution defined by multiple experts
Why does $\Phi$ matter?

The global density defined by the standard FoE is unimodal. The BiFoE allows for considerably more flexibility for shaping the density function.
Setup of experiments - Data

Brodatz and synthetic textures

D6: woven aluminium wire
D21: french canvas
D35: oriental straw cloth
D77: cotton canvas
D4: pressed cork
D103: loose burlap
circle textons
cross textons
Setup of experiments - Evaluation

- Tasks:
  - Texture synthesis
  - Texture inpainting

- Baseline Model: Gaussian FoE (GFoE)

\[
\Phi_{\text{GFoE}}(y) = \exp(-(y - b)^2)
\]

\[
E(x) = \frac{1}{2} \sum_i \sum_j \left( w_j^T x(i) + b_j \right)^2
\]

\[x \sim N(\mu, \Sigma)\] and \(\mu\) and \(\Sigma\) can be computed explicitly.

- Other details:
  - Models with \(M = 9/15\) experts, filter size 7 × 7 pixels
  - Training on 500 25 × 25 pixels texture patches
Learning

- Learning of the parameters by gradient ascent in the log-likelihood:

\[
\frac{\partial}{\partial \theta_j} \mathcal{L}(X; \Theta) = -\left\langle \frac{\partial E_{FoE}(x; \Theta)}{\partial \theta_j} \right\rangle_X + \left\langle \frac{\partial E_{FoE}(x; \Theta)}{\partial \theta_j} \right\rangle_{p_{FoE}(x; \Theta)}
\]

- Roth & Black propose contrastive divergence for learning
  - insufficient in the case of the BiFoE

- Better: Approximating the model distribution using K persistent chains (Tieleman 2008)
  - Chains initialized at the beginning of learning
  - Alternating update of Markov chains and model parameters
  - “Persistence” seems to be essential for learning good BiFoE models!
Texture Synthesis

50 x 50 texture patches / samples from the models

Data

FoE

GFoE
Texture Synthesis

50 × 50 texture patches / samples from the models

Data

FoE

GFoE

BiFoE
Texture Synthesis

50 × 50 texture patches / samples from the models

Data

FoE

GFoE

BiFoE

![Normalized CC](image)

D6, D21, D53, D77
Texture Synthesis

50 × 50 texture patches / samples from the models

Data

FoE

GFoE

BiFoE
Texture Synthesis

50 × 50 texture patches / samples from the models

Data

FoE

GFoE

BiFoE

All BiFoE models learn experts with bimodal nonlinearities ($a_j < 0$)
Texture Inpainting

Results for $70 \times 70$ inpainting frames with $50 \times 50$ “unobserved” regions. Missing pixels sampled conditioned on observed pixels with HMC-MCMC.

<table>
<thead>
<tr>
<th>Original</th>
<th>Inpainting Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="D6" /></td>
<td><img src="image" alt="D6" /></td>
</tr>
<tr>
<td><img src="image" alt="D21" /></td>
<td><img src="image" alt="D21" /></td>
</tr>
<tr>
<td><img src="image" alt="D53" /></td>
<td><img src="image" alt="D53" /></td>
</tr>
<tr>
<td><img src="image" alt="D77" /></td>
<td><img src="image" alt="D77" /></td>
</tr>
</tbody>
</table>
Texture Inpainting

Results for $70 \times 70$ inpainting frames with $50 \times 50$ “unobserved” regions. Missing pixels sampled conditioned on observed pixels with HMC-MCMC.

<table>
<thead>
<tr>
<th>Original</th>
<th>Inpainting Frame</th>
<th>Efros &amp; Leung</th>
</tr>
</thead>
<tbody>
<tr>
<td>D6</td>
<td>D21</td>
<td>D53</td>
</tr>
<tr>
<td>D77</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Original

Inpainting Frame

Efros & Leung
Results for $70 \times 70$ inpainting frames with $50 \times 50$ “unobserved” regions. Missing pixels sampled conditioned on observed pixels with HMC-MCMC.

Average reconstruction quality in terms of normalized cross-correlation with ground truth ($\pm$ std.-dev)

<table>
<thead>
<tr>
<th></th>
<th>D6</th>
<th>D21</th>
<th>D53</th>
<th>D77</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efros &amp; Leung</td>
<td>0.8300 ± 0.0380</td>
<td>0.8330 ± 0.0351</td>
<td>0.8878 ± 0.0300</td>
<td>0.6325 ± 0.0490</td>
</tr>
<tr>
<td>BiFoE</td>
<td>0.8769 ± 0.0163</td>
<td>0.8653 ± 0.0244</td>
<td>0.9145 ± 0.0125</td>
<td>0.6567 ± 0.0205</td>
</tr>
</tbody>
</table>
“Standard FoE” is a limited model of textures
the bimodal potential gives rise to a considerably more powerful model
performance equivalent to non-parametric approach on textures considered
but description is more compact and in terms of a generative model
can be used as a component e.g. for a texture segmentation task in a fully generative setting

Results not inconsistent with the good performance of the standard FoE for image denoising / inpainting:
FoE trained on natural images seems to model mainly piecewise smoothness (Weiss & Freeman, 2007; Tappen 2007)
for simple image properties such as piecewise smoothness a unimodal PDF is sufficient