Learning generative texture models with extended Fields-of-Experts

Nicolas Heess[1]
n.m.o.heess@sms.ed.ac.uk
Christopher K.I. Williams[1]
c.k.i.williams@inf.ed.ac.uk
Geoffrey E. Hinton[2]
hinton@cs.toronto.edu

[1] University of Edinburgh
School of Informatics
Edinburgh, UK

[2] Department of Computer Science
University of Toronto,
Toronto, Canada

Much effort has been devoted to the development of prior models of generic image structure. Such models are important for many image processing and synthesis tasks, and as a building block of more comprehensive probabilistic models of natural scenes. One successful example is the Field-of-Experts (FoE) framework, recently proposed by Roth and Black [3]. The FoE defines a probability distribution over images in the form of a homogeneous high-order Markov random field (MRF). This MRF-based model is translation invariant and can be applied to images of arbitrary size. The model is parametric and all parameters can be learned from training data, thus it can be directly adapted to the statistics of natural images.

Natural images are, however, extremely complex, containing different regions with very different visual characteristics. Attempting to learn these different characteristics with a single, generic model will most likely lead to the model learning only the most generic properties – such as piecewise smoothness in the case of the FoE [4, 6]. As an alternative approach we therefore propose to focus on models that are good at capturing *specific* structure in natural images and use these models as building blocks of more comprehensive, hierarchical models that can then account for more complex properties of natural images. For example, a number of texture models can be composed together to model an image comprised of multiple regions. Suitable "component models" are clearly an important prerequisite for such hierarchical models. Unfortunately, however, many of the most powerful methods for generating specific structure that have been proposed in the past (e.g. [1, 5]), are not formulated as probabilistic models and it is therefore not clear how they could be used in this context. In this work we first show that the FoE in its original formulation is a very limited model even of specific image structure, but that an extended formulation of the model gives rise to a powerful generative model of textures.

The FoE is a homogeneous high-order Markov random field (MRF) with clique potentials defined in terms of the responses of these linear filters

$$p(\boldsymbol{x}) = \frac{1}{Z(\Theta)} \prod_{i=1}^{N} \prod_{j=1}^{M} \Phi(\boldsymbol{w}_j^T \boldsymbol{x}_{(i)}; \boldsymbol{\alpha}_j), \tag{1}$$

where the index i runs over the pixels in the image, and $\boldsymbol{x}_{(i)}$ is the image patch of the size of filter $\boldsymbol{w}_j$ centered at pixel i. Roth & Black choose $\Phi(y;\boldsymbol{\alpha})$ to be the one dimensional Student-t potential, i.e. $\Phi(y;\nu) = (1 + \frac{1}{2}y^2)^{-\nu}$ (with $\nu > 0$) so that (1) can be written in terms of the energy as:

$$p_{FoE}(\boldsymbol{x}) = \frac{1}{Z(\Theta)} \exp(-E_{FoE}(\boldsymbol{x})) \tag{2}$$

$$E_{FoE}(\boldsymbol{x}) = \sum_i \sum_j \nu_j \log \left\{ 1 + \frac{1}{2} \left(\boldsymbol{w}_j^T \boldsymbol{x}_{(i)} \right)^2 \right\}. \tag{3}$$

This formulation defines a probabilistic model of images of arbitrary size all parameters of which can be learned from data. The choice of the Student-t potential is motivated by the properties of natural images and previous work on probabilistic models of natural image patches (e.g. [2, 7]), yet, the choice of a zero-centered Student-t potential forces the response marginals of the filters that are being learned to be centered at zero and monotonically decaying. The distribution defined by the FoE is unimodal and it is thus very restrictive. We propose an extension of the original FoE model, the bimodal FoE (BiFoE), which allows for bimodal expert functions $\Phi_{Bi}(y;\nu,a,b) = \left\{1 + \frac{1}{2}\left[(y+b)^2 + a\right]^2\right\}^{-\nu}$. This choice of Φ gives rise to the following energy:

$$E_{Bi}(\boldsymbol{x}) = \sum_i \sum_j \nu_j \log \left\{ 1 + \frac{1}{2} \left[\left(\boldsymbol{w}_j^T \boldsymbol{x}_{(i)} + b_j \right)^2 + a_j \right]^2 \right\}.$$

Φ_{Bi} is bimodal for $a < 0$ and b determines the center of the potential. The BiFoE can be sampled from and learned in essentially the same way as the FoE but it allows for considerably more flexibility with respect to the shape of the response marginals of the filters that it learns and it can model globally highly multimodal distributions.

In the paper we describe experiments in which we learn FoE and BiFoE models of several Brodatz textures. We synthesize textures from the models and compare results quantitatively using a correlation score. While the FoE is unable to model these textures well (with results similar to a much simpler Gaussian MRF), the BiFoE produces samples that are very similar to the original textures. We further evaluate the BiFoE quantitatively on a texture inpainting task and find that for the textures considered its performance is comparable to the non-parametric method suggested in [1]. Unlike the latter, however, it defines a compact parametric model that can be used as component in more comprehensive probabilistic models, making it a promising building block of generative models of mid-level vision.

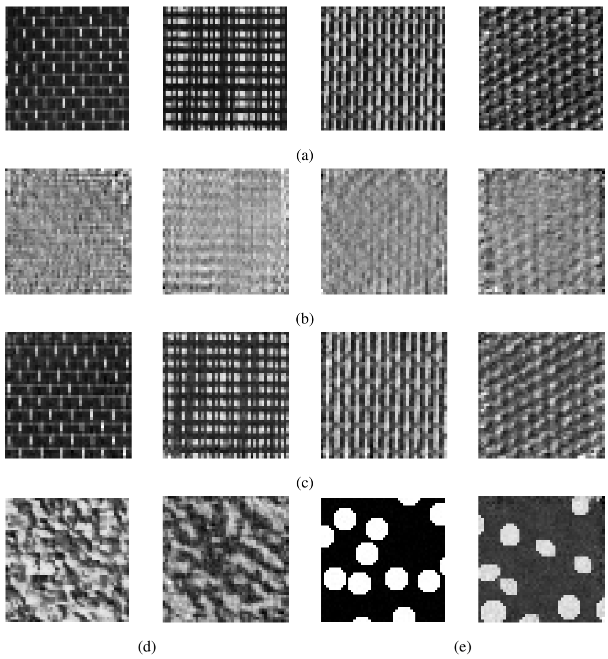

Figure 1: (a) Training data: 50×50 patches of some of the textures used in our experiments. (b) FoE samples of the textures in (a). (c) BiFoE samples. (d, e) Two more challenging textures: patches of original textures (left) and corresponding BiFoE samples (right).

[1] Alexei A. Efros and Thomas K. Leung. Texture Synthesis by Non-parametric Sampling. In *Proc ICCV*, pages 1033–1038, 1999.

[2] B.A. Olshausen and D.J. Field. Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1? *Vision Research*, 37:3311–3325(15), 1997.

[3] S. Roth and M.J. Black. Fields of Experts: a framework for learning image priors. In *Proc CVPR*, pages 860–867 vol. 2, 2005.

[4] M.F. Tappen. Utilizing Variational Optimization to Learn Markov Random Fields. In *Proc CVPR*, pages 1–8, 2007.

[5] L.-Y. Wei and M. Levoy. Fast texture synthesis using tree-structured vector quantization. In *SIGGRAPH*, pages 479–488, 2000.

[6] Y. Weiss and W.T. Freeman. What makes a good model of natural images? In *CVPR*, pages 1–8, 2007.

[7] M. Welling, G. Hinton, and S. Osindero. Learning Sparse Topographic Representations with Products of Student-t Distributions. In *NIPS 15*, pages 1359–1366. 2003.