Supplementary Material: Learning generative texture models with extended FoEs

Nicolas Heess and Christopher K.I. Williams

School of Informatics, University of Edinburgh, Edinburgh, UK

Geoffrey E. Hinton

Department of Computer Science, University of Toronto, Toronto, Canada

July 21, 2009

A Unimodality of Student-t FoE

Consider the energy of the Student-t *PoE*:

$$E(\boldsymbol{x}) = \sum_j v_j \log \underbrace{\left\{1 + \frac{1}{2}(\boldsymbol{w}_j^T \boldsymbol{x})^2\right\}}_{z_j(\boldsymbol{x})} \tag{1}$$

$$\frac{\partial E}{\partial x_i} = \sum_j \frac{v_j w_{ji}(\boldsymbol{w}_j^T \boldsymbol{x})}{z_j(\boldsymbol{x})} \tag{2}$$

$$= \frac{\sum_j \overbrace{\prod_{j' \neq j} z_{j'}(\boldsymbol{x})}^{b_j(\boldsymbol{x})} v_j w_{ji}(\boldsymbol{w}_j^T \boldsymbol{x})}{\underbrace{\prod_k z_k(\boldsymbol{x})}_{a(\boldsymbol{x})}} \tag{3}$$

$$= \sum_j \underbrace{\frac{b_j(\boldsymbol{x}) v_j}{a(\boldsymbol{x})}}_{c_j(\boldsymbol{x})} w_{ji} \boldsymbol{w}_j^T \boldsymbol{x} \tag{4}$$

$$= \left[\sum_j c_j(\boldsymbol{x}) w_{ji} \boldsymbol{w}_j\right]^T \boldsymbol{x} \tag{5}$$

$$\nabla_{\boldsymbol{x}} E = W C(\boldsymbol{x}) W^T \boldsymbol{x} \tag{6}$$

Here: $v_j > 0$; $b_j(\boldsymbol{x}) > 0$; $a_j(\boldsymbol{x}) > 0$ and therefore also $c_j(\boldsymbol{x}) > 0$. Also, $C = diag(c_1(\boldsymbol{x}), \ldots, c_N(\boldsymbol{x}))$ where M is the number of experts, $j = 1 \ldots M$. W is the matrix with the filters in its columns, i.e. the $D \times M$ dimensional matrix $W = (\boldsymbol{w}_1 \ldots \boldsymbol{w}_M)$. At the minimum we require $\nabla_{\boldsymbol{x}} E = \mathbf{0}$ and thus $\mathbf{0} = W C(\boldsymbol{x}) W^T \boldsymbol{x}$. Obviously, $\boldsymbol{x} = \mathbf{0}$ is one solution. Because $c_j(\boldsymbol{x}) > 0 \;\; \forall j$ (in fact, c_j is monotonously increasing with the norm of $\boldsymbol{x}$) and because the matrix C is diagonal and therefore only scales the filter vectors in W, this is also the only solution as long as WW^T has full rank, i.e. W contains at least one subset of D linearly independent vectors $\boldsymbol{w}_j$.

For the FoE W consists of the M concatenated convolution matrices corresponding the different M experts (filters).

B Experiments with 1D patterns

B.1 Data

In order to gain more insight into the nature of the representations learned by the different models we investigate their ability to model a set of periodic 1D patterns, including simple sine waves, square wave patterns, a linear combination of two phase-locked sine waves of different frequencies, and a series of pulses (delta peaks). Examples of these patterns are shown in Fig. 1. 1D Patterns were typically 30 or 32 pixels wide, we varied the period of the patterns but show results only for one particular set of parameters. In all cases we added some IID Gaussian noise ($\sigma = 0.05$) to the training patterns. The patterns were chosen so as to distinguish between the simple GFoE and a more powerful model: The GFoE should be able to model the sine wave, but we expected it to fail, for instance, for the two phase locked sine waves, as it should be incapable of modeling the specific phase relationship between the two sinusoids. The interesting question was then whether one of the other two models would be able to account for this relationship.

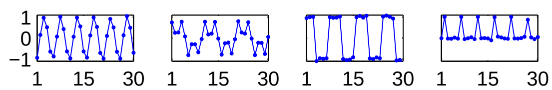


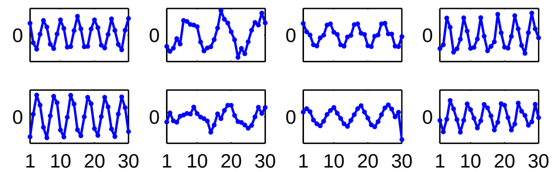
Figure 1: Examples of 1D patterns: sine wave ($f = 0.2$ cyc/pt), phase locked sine waves ($f_0 = 0.1$ cyc/pt, $f_1 = 0.3$ cyc/pt), square wave (period: 8), series of pulses (pulse distance: 5pt); $\sigma_{noise} = 0.05$

B.2 Results

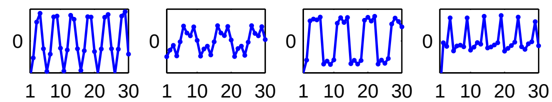
Below we show results for models with $M = 9$ 7-point filters unless noted otherwise.

GFoE / FoE: Fig. 2a shows representative examples of the patterns generated by the GFoE and FoE models trained on the four 1D patterns. It is clear that both models fail except for the sine wave. For the GFoE this is what we expected (see section 2.1 in the main paper) and the learned representation can be easily verified by computing the principal components of the covariance matrix: For the sine wave, there are two relevant components, corresponding to a sine/cosine pair at the appropriate frequency; for the phase locked sinusoids there are four components corresponding to two sine / cosine pairs at the frequencies of the component sinusoids but the model does not account for their specific phase relationship. Comparing the sets of samples generated from the FoE and the GFoE there is no obvious difference (note that Fig. 2 shows only one sample per model and pattern).

BiFoE: BiFoE models were reliably learned for all four 1D patterns, as is shown in Fig. 2b, and the parameters were often interpretable. This is illustrated in Fig. 3 which shows the experts (filters and potential functions) for a model of the square wave pattern. It is easy to see how the filters together with the corresponding potentials constrain the patterns generated by this model. Note that in order to obtain more easily interpretable results this model was trained with the minimum number of filters for the square wave, and with periodic boundary conditions. See figure legend for details.



(a)



(b)

Figure 2: (a) Samples drawn from GFoE (top) and FoE (bottom) models trained on the 1D patterns shown in Fig. 1. Left to right: sine wave, phase locked sine waves, square wave, pulses. (b) Samples drawn from the BiFoE models.

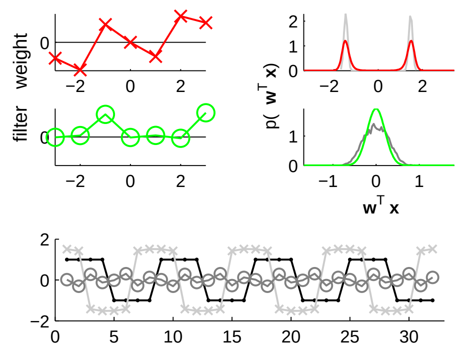


Figure 3: BiFoE model learned for a square wave pattern with period 8. $M = 2$ experts, filter width 7, trained with periodic boundary conditions. Top left: Filters $\boldsymbol{w}_1$, $\boldsymbol{w}_2$ of the two experts (shown in red/crosses and green/circles). Top right: Potential functions of the experts (also shown in red and green respectively) and corresponding response histograms of the two filters (light and dark gray). Bottom: The first filter responds with either -1.5 or 1.5 as it is shifted along the square wave. The second filter responds with (approximately) zero everywhere. Square wave: black; response of first filter: light gray/crosses; response of second filter: dark gray/circles. In an idealized (noise free) form the two filters are $\boldsymbol{w}_1 = (\; -c \quad -d \quad c \quad 0 \quad -c \quad d \quad c \;)^T$ and $\boldsymbol{w}_2 = (\; 0 \quad 0 \quad e \quad 0 \quad 0 \quad 0 \quad e \;)^T$ where $c, d, e \neq 0$ can be chosen arbitrarily. The two experts together constrain the patterns generated by the model to be (noisy) square waves.

C Comparison with non-parametric texture synthesis methods on texture inpainting task

Figure 4 shows further results for the texture inpainting task.

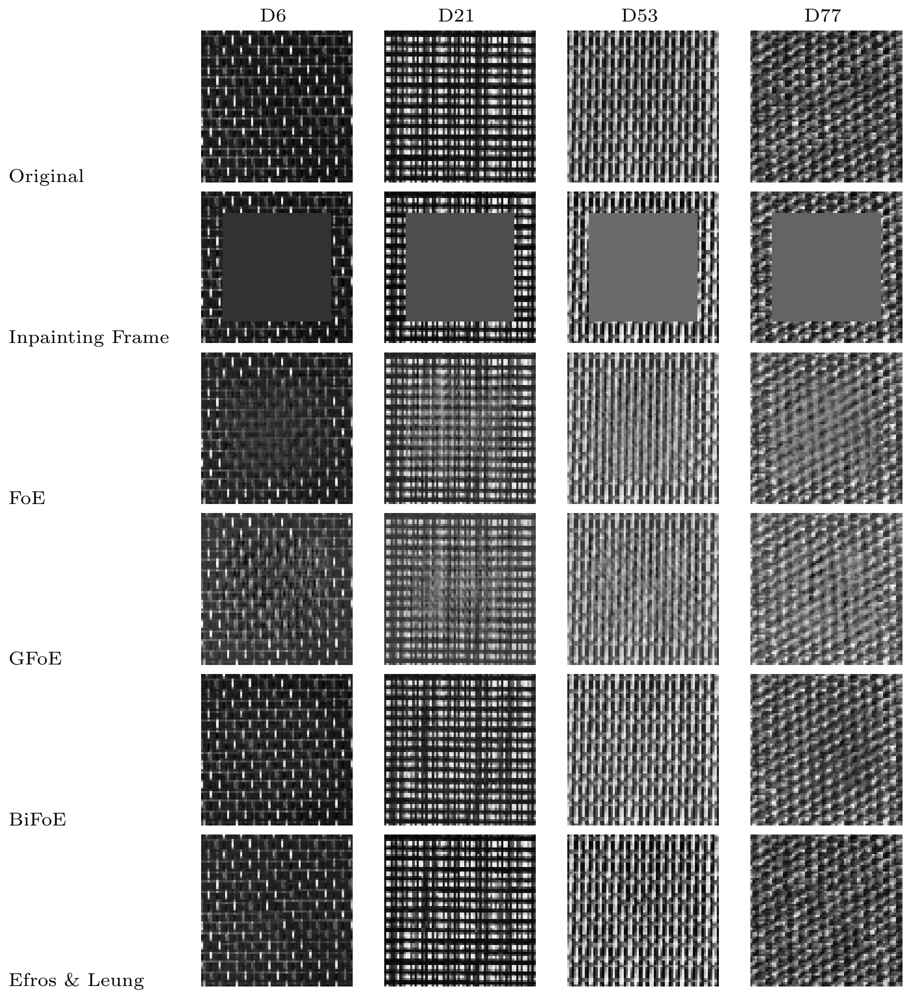


Figure 4: Inpainting results for all regular textures
Top to bottom: Original image; inpainting frame; image completed by FoE; image completed with GFoE; image completed BiFoE; image completed by Efros & Leung's method