Density Modeling of Images with Generalized Divisive Normalization

Johannes Ballé
Valero Laparra
Eero P. Simoncelli

New York University
Howard Hughes Medical Institute
Why unsupervised learning?

find structure in unlabeled data

understand sensory representation

figure: Hubel, 1995
Density estimation (parametric density)

\[ p_x(x) = \frac{1}{Z(\theta)} \exp\left(-f(x; \theta)\right) \]
Density estimation (parametric density)

\[ p_x(x) = \frac{1}{Z(\theta)} \exp(-f(x; \theta)) \]

\[ Z(\theta) = \int \exp(-f(x; \theta)) \, dx \]
Density estimation (parametric density)

\[ p_x(x) = \frac{1}{Z(\theta)} \exp\left(-f(x; \theta)\right) \]

\[ Z(\theta) = \int \exp\left(-f(x; \theta)\right) \, dx \]

tractable?
Density estimation (parametric transformation)

\[ x \sim p_x \]
Density estimation (parametric transformation)

\[ x \sim p_x \]

\[ g(x; \theta) \]

\[ y \sim \mathcal{N} \]

Friedman, 1984
Chen & Gopinath, 2001
Lyu & Simoncelli, 2009
Laparra et al., 2010
Dinh et al., 2015
Density estimation (parametric transformation)

\[ x \sim p_x \quad \xrightarrow{\text{Gaussianization}} \quad g(x; \theta) \quad \xrightarrow{} \quad y \sim \mathcal{N} \]

- Friedman, 1984
- Chen & Gopinath, 2001
- Lyu & Simoncelli, 2009
- Laparra et al., 2010
- Dinh et al., 2015
Density estimation (parametric transformation)

\[ x \sim p_x \]

\[ g(x; \theta) \]

\[ y \sim \mathcal{N} \]

"inferred" density:

\[ p_x(x) = \left| \frac{\partial g(x; \theta)}{\partial x} \right| \mathcal{N}(g(x; \theta)) \]

Friedman, 1984
Chen & Gopinath, 2001
Lyu & Simoncelli, 2009
Laparra et al., 2010
Dinh et al., 2015
Parameter estimation

\[ p_x(x) = \left| \frac{\partial g(x; \theta)}{\partial x} \right| \mathcal{N}(g(x; \theta)) \]
Parameter estimation

\[-\log p_x(x) = -\log \left| \frac{\partial g(x; \theta)}{\partial x} \right| - \log \mathcal{N}(g(x; \theta))\]
Parameter estimation

\[ x \rightarrow g(x; \theta) \rightarrow y \]

\[-\log p_x(x) = -\log \left| \frac{\partial g(x; \theta)}{\partial x} \right| - \frac{1}{2} \left\| g(x; \theta) \right\|_2^2 + C\]
Parameter estimation

\[
-x \log p_x(x) = - \log \left| \frac{\partial g(x; \theta)}{\partial x} \right| - \frac{1}{2} \left\| g(x; \theta) \right\|_2^2 + C
\]

minimize wrt. \( \theta \) using stochastic gradient descent
Marginal distribution of linear filter responses

Burt & Adelson, 1981
Field, 1987
Mallat, 1989

image ©CC-BY-NC 2.0 acevvvedo@flickr
Marginal distribution of linear filter responses

Burt & Adelson, 1981
Field, 1987
Mallat, 1989

image ©CC-BY-NC 2.0 acevvedo@flickr
\[ y = \frac{c}{1 + \exp(ax + b)} \]
\[ y = \frac{c}{1 + \exp(a x + b)} \]
\[ y = \frac{c}{1 + \exp(ax + b)} \]

\[ p_x(x) = \frac{\partial y}{\partial x} \mathcal{N}(y) \]
\[ y = \frac{c}{1 + \exp(ax + b)} \]

\[ p_x(x) = \frac{\partial y}{\partial x} \mathcal{N}(y) \]
\[ y = \frac{x}{(\beta + \gamma |x|)^\varepsilon} \]
\[ y = \frac{x}{(\beta + \gamma |x|)^\varepsilon} \]
\[
y = \frac{x}{(\beta + \gamma |x|)^\varepsilon}
\]

\[
p_x(x) = \frac{\partial y}{\partial x} \mathcal{N}(y)
\]
\[ y = \frac{x}{(\beta + \gamma |x|)^\varepsilon} \]

\[ p_x(x) = \frac{\partial y}{\partial x} \mathcal{N}(y) \]
Marginal distribution of linear filter responses
Joint distribution of linear filter responses

countour lines of joint density
\[ y_0 = \frac{x_0}{(\beta_0 + y_0|x_0|\alpha_0)^{\varepsilon_0}} \]

\[ y_1 = \frac{x_1}{(\beta_1 + y_1|x_1|\alpha_1)^{\varepsilon_1}} \]
Contour lines, Gaussianized responses
Contour lines, Gaussianized responses
Contour lines, Gaussianized responses

$\mathcal{N}$

$y_0$  $y_1$
Improved Gaussianization

1. Iterated marginal Gaussianization

Chen & Gopinath, 2001
Laparra et al., 2010
Improved Gaussianization

1. Iterated marginal Gaussianization

2. Joint Gaussianization (inspired by biology)
\[ y_0 = \frac{x_0}{(\beta_0 + \gamma_0 x_0^{\alpha_0})^{\varepsilon_0}} \]

\[ y_1 = \frac{x_1}{(\beta_1 + \gamma_1 x_1^{\alpha_1})^{\varepsilon_1}} \]
\[ y_0 = \frac{x_0}{(\beta_0 + y_0|x_0|^{\alpha_0})^{\epsilon_0}} \]

\[ y_1 = \frac{x_1}{(\beta_1 + y_1|x_1|^{\alpha_1})^{\epsilon_1}} \]
\begin{align*}
y_0 &= \frac{x_0}{(\beta_0 + \gamma_{01} |x_1|^{\alpha_{01}} + \gamma_{00} |x_0|^{\alpha_{00}})^{\varepsilon_0}} \\
y_1 &= \frac{x_1}{(\beta_1 + \gamma_{10} |x_0|^{\alpha_{10}} + \gamma_{11} |x_1|^{\alpha_{11}})^{\varepsilon_1}}
\end{align*}
Contour lines, Gaussianized responses
Variety of shapes, joint density of filter responses

elliptical

?

marginally independent

Lyu & Simoncelli, 2009
Sinz et al., 2009
Contour lines, linear filter responses
Contour lines, linear filter responses
Contour lines, linear filter responses

---

model

histogram estimate
Contour lines, linear filter responses
Contour lines, linear filter responses
Generalized divisive normalization (GDN)

\[ y_i = \frac{z_i}{(\beta_i + \sum_j \gamma_{ij} |z_j|^{\alpha_{ij}})^{\epsilon_i}} \]

Special cases/related models:

- Independent Component Analysis, Cardoso, 2003
- Independent Subspace Analysis, Hyvärinen & Hoyer, 2000
- Weighted normalization model, Schwartz & Simoncelli, 2001
- Topographic ICA, Hyvärinen et al., 2001
- Radial Gaussianization, Lyu & Simoncelli, 2009
- \( L_p \)-nested symmetric distributions, Sinz & Bethge, 2010
- “Two-layer model”, Köster & Hyvärinen, 2010
Parameter estimation (multiple layers)

\[ - \log p_x(x) = - \log \left| \frac{\partial g(x; \theta)}{\partial x} \right| - \frac{1}{2} \| g(x; \theta) \|_2^2 + C \]

minimize wrt. \( \theta \) using stochastic gradient descent
Parameter estimation (multiple layers)

\[ - \log p_x(x) = - \log \left| \frac{\partial g(x; \theta)}{\partial x} \right| - \frac{1}{2} \| g(x; \theta) \|_2^2 + C \]

\[ - \log \left| \frac{\partial g_0(x_0; \theta)}{\partial x_0} \right| - \log \left| \frac{\partial g_1(x_1; \theta)}{\partial x_1} \right| - \ldots \]

minimize wrt. \( \theta \) using stochastic gradient descent
One layer of joint GDN > many layers of marginal GDN
What are the perceptual properties of the representation?

figure: Hubel, 1995
What are the perceptual properties of the representation?

figure: Hubel, 1995
What are the perceptual properties of the representation?

figure: Hubel, 1995
What are the perceptual properties of the representation?

figure: Hubel, 1995
original

increasing Euclidean distance in pixel representation
increasing Euclidean distance in Gaussianized representation
Pixel representation

Euclidean distance (pixel representation)

human-reported distortion

$\rho = 0.39$

data: TID 2008
Multi-scale GDN representation

Euclidean distance (Gaussianized representation)

human-reported distortion

$\rho = 0.84$

data: TID 2008
Multi-scale GDN representation

Euclidean distance (Gaussianized representation)

human-reported distortion

\[ \rho = 0.84 \]

SSIM: \[ \rho = 0.74 \]

SSIM: Wang et al., 2004

data: TID 2008
• Gaussianization: Methodology for density estimation and unsupervised learning of a representation

• GDN: joint nonlinearity applied across feature maps
  – inspired by nonlinearities of biological neurons
  – generalizes sigmoids used in ANNs
  – capable of Gaussianizing image data

• one layer of GDN > many layers of marginal nonlinearities

• accounts for human judgements of image quality (more so than SSIM, the de facto standard)

Thank you!