Density Modeling of Images Using a Generalized Normalization Transform

Johannes Ballé, Valero Laparra, Eero P. Simoncelli

Howard Hughes Medical Institute
New York University

http://arxiv.org/abs/1511.06281

Summary

Gaussianization is a methodology for density estimation and unsupervised learning of a representation.

We introduce GDN, a joint nonlinearity applied across subbands, inspired by nonlinearities of biological neurons. It generalizes sigmoids used in ANNs and is capable of Gaussianizing image data.

One stage of GDN is more efficient than many stages of linear filters/marginal nonlinearities.

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

Density estimation by Gaussianization

Parametric density model

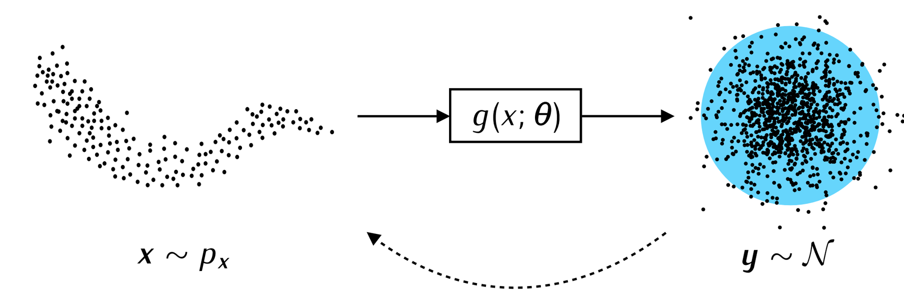
$$p_x(x) = \frac{1}{Z(\theta)} \exp\big(-f(x;\theta)\big)$$

$$Z(\theta) = \int \exp\big(-f(x;\theta)\big)\,dx$$

tractable?

the textbook way
(many well-known densities, maximum entropy ...)

Parametric transformation (Gaussianization)



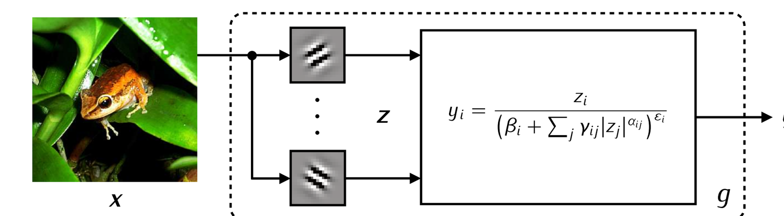
"inferred" density:

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

Friedman, 1984
Chen & Gopinath, 2001
Lyu & Simoncelli, 2009
Laparra et al., 2010

Generalized divisive normalization (GDN)

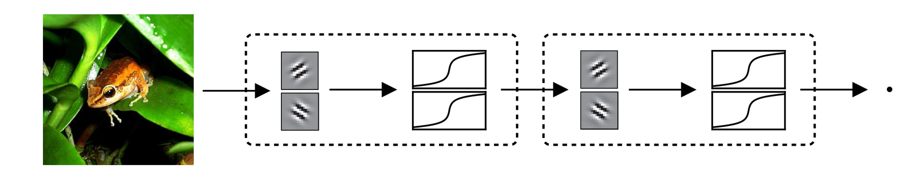
We introduce a transformation particularly suited to the Gaussianization of natural image densities.



Special cases/related models:

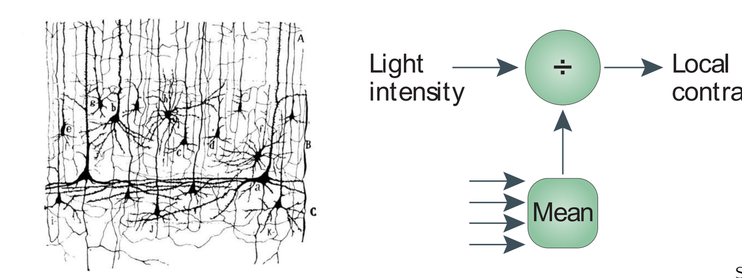
- Independent Component Analysis, Cardoso, 2003
- Independent Subspace Analysis, Hyvärinen & Hoyer, 2000
- Weighted normalization model, Schwartz & Simoncelli, 2001
- Topographic ICA, Hyvärinen et al., 2001
- Radial Gaussianization, Lyu & Simoncelli, 2009
- L_p-nested symmetric distributions, Sinz & Bethge, 2010
- "Two-layer model", Köster & Hyvärinen, 2010

Previous work: iterated marginal Gaussianization
(g is a composition of linear transformations and marginal nonlinearities akin to ANNs)



Chen & Gopinath, 2001
Laparra et al., 2010

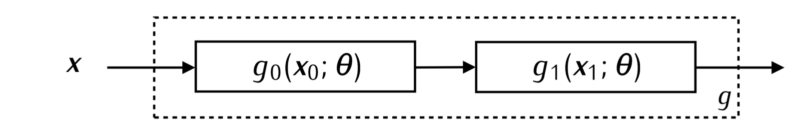
Here, we take inspiration from biology for more efficient Gaussianization.
(g is a a linear transformation followed by a form of divisive normalization, a joint nonlinearity)



Heeger, 1992
Schwartz & Simoncelli, 2001
figures: Cajal; Carandini & Heeger, 2012

Fitting the parameters

We use a maximum likelihood approach to fit the parameters.
If each component of g is complete and invertible, the chain rule simplifies the computation significantly.

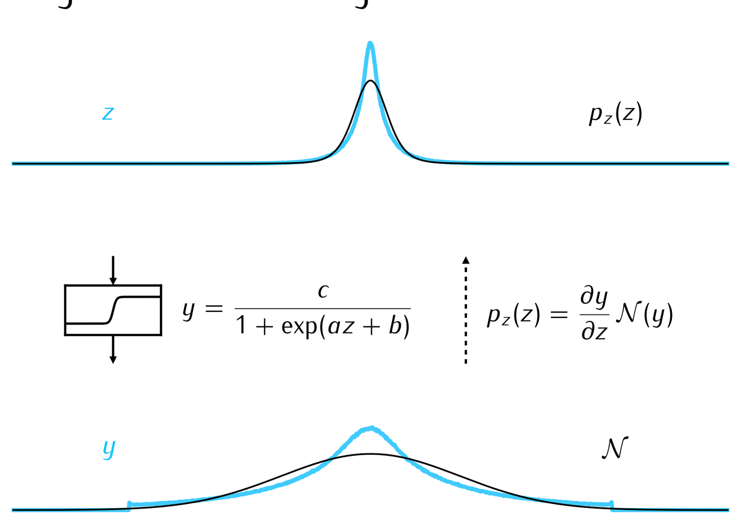


$$-\log p_x(x) = \underbrace{-\log\left|\frac{\partial g(x;\theta)}{\partial x}\right|}_{-\log\left|\frac{\partial g_0(x_0;\theta)}{\partial x_0}\right| - \log\left|\frac{\partial g_1(x_1;\theta)}{\partial x_1}\right| - \dots} - \frac{1}{2}\|g(x;\theta)\|_2^2 + C$$

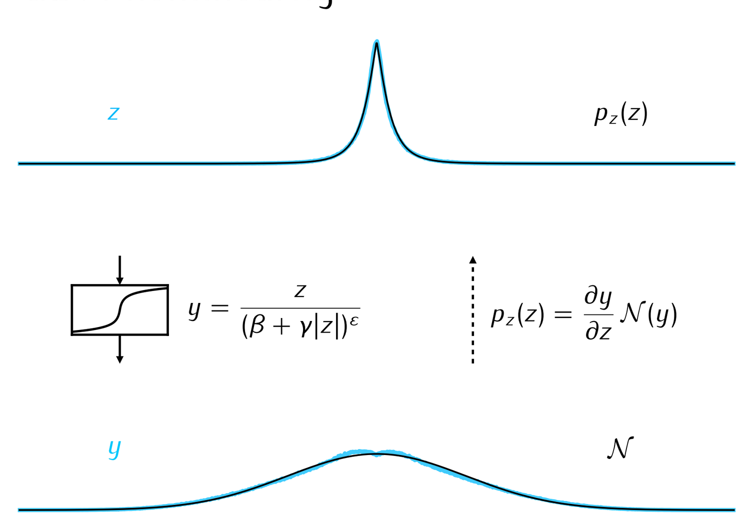
minimize wrt. θ using stochastic gradient descent

1-D: marginal density of linear filter responses

Logistic nonlinearity

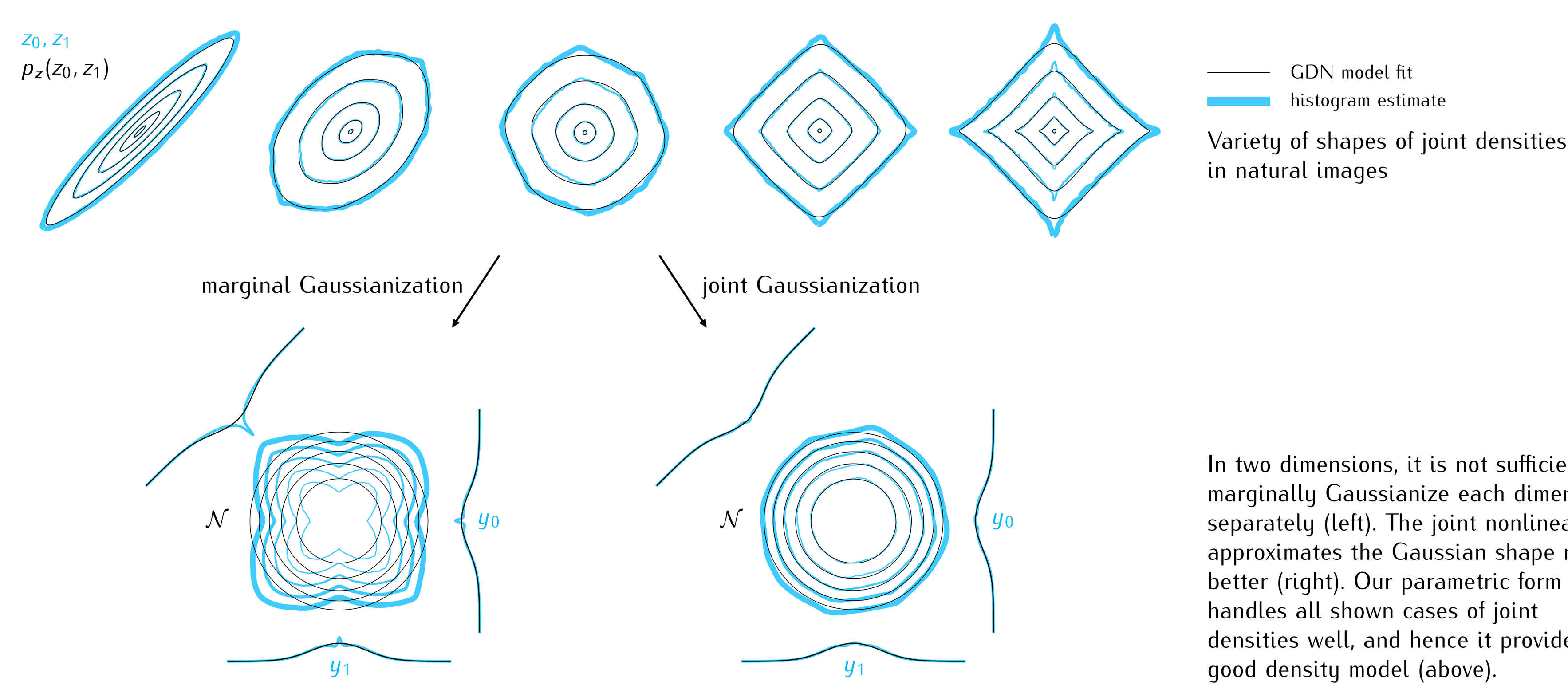


GDN nonlinearity



To Gaussianize marginal densities of linear filter responses to natural images, the transformation needs to take the shape of a sigmoidal. Our parametric form provides a better fit than typical sigmoidal forms known from ANNs. It also fits better than parametric density models such as the generalized Gaussian family (not shown).

2-D: joint density of linear filter responses

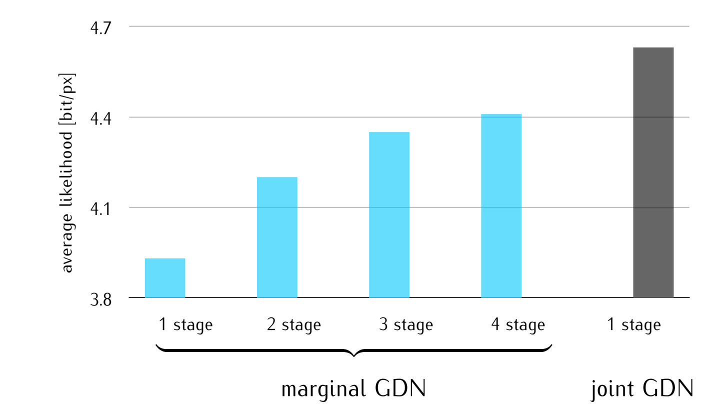


Variety of shapes of joint densities in natural images

In two dimensions, it is not sufficient to marginally Gaussianize each dimension separately (left). The joint nonlinearity approximates the Gaussian shape much better (right). Our parametric form handles all shown cases of joint densities well, and hence it provides a good density model (above).

n-D: natural image densities

Average likelihoods



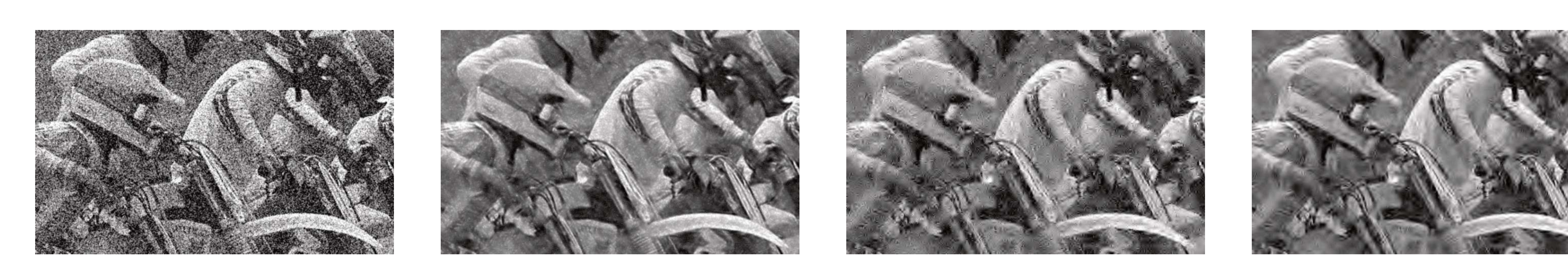
	63 dim. [nat]	64 dim. [bit/px]
GMM (200 comp.)	153.7	3.36
RIDE (1 layer)	150.7	3.29
EoMCGSM (128 comp.)	**158.1**	**3.48**
GDN (1 layer)	151.5	3.47

small image patches, variable preprocessing
need better datasets for comparison to other models!

from Theis & Bethge, 2015

1 stage of joint GDN > many stages of marginal GDN

Denoising experiment



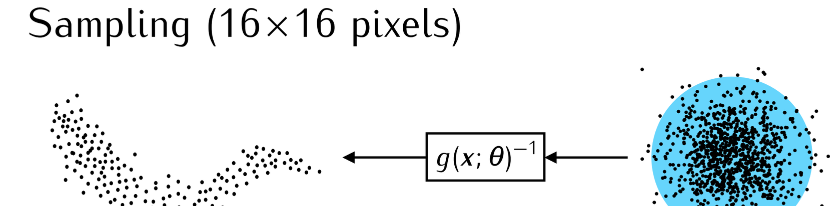
noisy | GDN: PSNR 22.6, SSIM 0.78 | marginal: PSNR 20.6, SSIM 0.68 | GSM: PSNR 22.4, SSIM 0.75

$$\hat{x} = \tilde{x} + \sigma^2 \frac{\partial}{\partial \tilde{x}} \log p_{\tilde{x}}(\tilde{x})$$

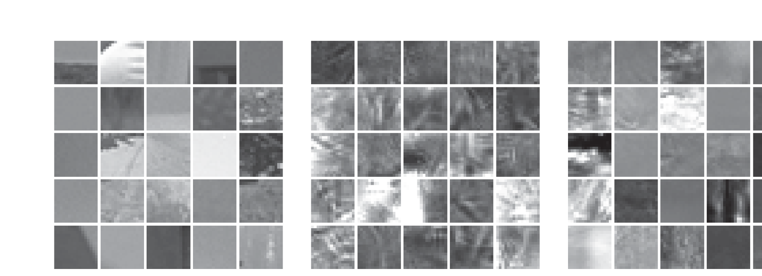
We used the empirical Bayes solution of Miyasawa (1961), which expresses the least-squares optimal solution directly as a function of the noisy data distribution.

For comparison, we show results for two denoising methods operating on orthogonal wavelets: one assuming a marginal model (Figueiredo & Nowak, 2001) and the other assuming a Gaussian scale mixture (GSM) model with elliptically symmetric densities (Portilla et al., 2003).

Sampling (16×16 pixels)



Since each component of the transformation is invertible by construction, sampling from the model is simple.



training samples | marginal samples | joint samples

Perceptual properties of normalized representation

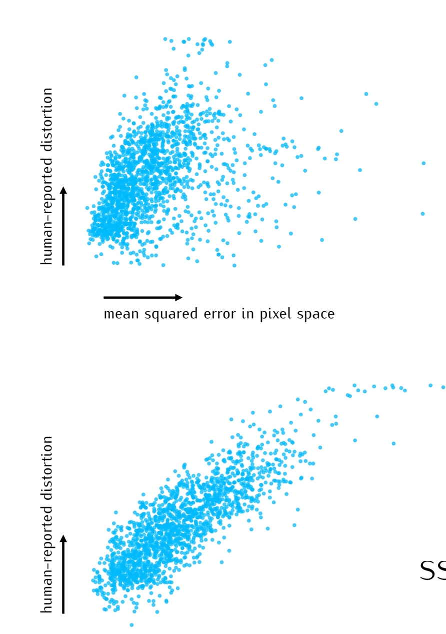
figure: Hubel, 1995

Do distances in Gaussianized representation correspond to *perceived* differences of images?

We made a multi-scale version of the GDN model and evaluated the Pearson correlation of distances in the Gaussianized representation with perceptual judgements of image distortions (provided in the TID 2008 database). The correlation exceeds the correlation reported for SSIM, the de facto standard in image quality assessment.

original — increasing Euclidean distance in pixel representation →

original — increasing Euclidean distance in Gaussianized representation →



data: TID 2008
SSIM: Wang et al., 2004