Channel capacity in psychovisual deep-nets: Gaussianization versus Kozachenko-Leonenko

Jesus Malo[1]

Image Processing Lab, Universitat de Valencia, Spain,
jesus.malo@uv.es,
WWW home page: http://isp.uv.es/

Abstract. In this work we quantify how neural networks designed from biology using no statistical training have a remarkable performance in information theoretic terms. Specifically, we address the question of the amount of information that can be extracted about the images from the different layers of psychophysically-tuned deep networks. We show that analytical approaches are not possible, and we propose the use of two empirical estimators of capacity: the classical Kozachenko-Lonenko estimator and a recent estimator based on Gaussianization.
Results show that networks purely based on visual psychophysics are extremely efficient in two aspects: (1) the internal representation of these networks duplicates the amount of information that can be extracted about the images with regard to the amount of information that could be obtained from the input representation assuming sensors of the same quality, and (2) the capacity of internal representation follows the PDF of natural scenes over the chromatic and achromatic dimensions of the stimulus space.
This remarkable adaptation to the natural environment is an example of how imitation of biological vision may inspire architectures and save training effort in artificial vision.

Keywords: Spatio-chromatic information, Psychophysically-tuned neural networks, Entropy Gaussianization, Kozachenko-Leonenko estimator.

1 Introduction

The early stages of deep-nets trained to solve visual classification problems develop units that resemble the sensors found in biological vision systems [1]. One could save substantial effort figuring out the proper architecture and training by using the existing models of early human vision which already have the linear+nonlinear architecture of deep-nets.

In this work[1] we show a specific example of the above by quantifying the performance of a biological network in accurate information-theory units.

In particular, we use the linear+nonlinear architecture which is standard in visual neuroscience through associations of filterbanks and the Divisive Normalization inhibitory interaction [2, 3]. The model we consider here consists of a

[1] Partially funded by DPI2017-89867-C2-2-R and GrisoliaP/2019/035.

series of layers that reproduce the following perceptual facts: spectral integration at linear LMS sensors [4], nonlinear normalization of LMS signals using Von-Kries adaptation [5], linear transform to chromatic opponent ATD channels [6], saturation of the achromatic, red-green, and yellow-blue signals [7], linear bank of local-oriented filters and achromatic/chromatic contrast sensitivity weights [8], nonlinear Divisive Normalization of the spatio-chromatic local-frequency filters [3, 9], related to classical neural field models [10].

Following the tradition of use of Divisive Normalization to improve JPEG and MPEG image coders [11, 12] and image quality metrics [13, 14], the current state-of-the-art in these image processing problems is achieved using similar linear+nonlinear architectures [15, 16]. The novelty in the current approaches is that they use the biological model as a starting point and refine it using the automatic differentiation techniques fitting psychophysical databases.

2 Measuring channel capacity in biological networks

In a sensory system where the input, $\boldsymbol{r}$, undergoes certain deterministic transform, S, but the sensors are noisy:

$$\boldsymbol{r} \xrightarrow{S} \boldsymbol{x} = s(\boldsymbol{r}) + \boldsymbol{n} \tag{1}$$

the transmitted information about $\boldsymbol{r}$ available from the response $\boldsymbol{x}$ (i.e. the mutual information, $I(\boldsymbol{r}, \boldsymbol{x})$) is also called the *capacity* of the channel S [17].

In this work we are interested in comparing the performance of the system at different locations of the space of images with the distribution of natural scenes over that space. Moreover, this comparison should be done for different layers of the visual pathway to analyze their relative contribution to the information transmission. Ideally, we'd like to describe the trends of the performance from the analytical description of the response. However, that is not straightforward.

Note that the amount of transmitted information can be written in terms of the entropies, h, of the input and the noise, and of the Jacobian of the transform; or alternatively in terms of the total correlation, T, *aka* multi-information [18], that represents the redundancy within the considered vector (or layer). Specifically, in [19] I derive these two equations that are relevant to discuss capacity:

$$I(\boldsymbol{r}, \boldsymbol{x}) = h(\boldsymbol{r}) + E_{\boldsymbol{r}}\Big\{ \log_2 |\nabla_{\boldsymbol{r}} S| \Big\} - h(\boldsymbol{n}) + E_{\boldsymbol{n}}\Big\{ D_{\mathrm{KL}}\Big(p(s(\boldsymbol{r}))\Big| p(s(\boldsymbol{r}) + \boldsymbol{n}) \Big) \Big\} \tag{2}$$

$$I(\boldsymbol{r}, \boldsymbol{x}) = \sum_i h(x_i) - T(\boldsymbol{x}) - h(\boldsymbol{n}) \tag{3}$$

where $E_{\boldsymbol{v}}\{\cdot\}$ stands for expected value over the random variable $\boldsymbol{v}$, and $D_{\mathrm{KL}}(p|q)$ stands for the Kullback-Leibler divergence between the probabilities p and q.

Eq. 3 identifies univariate and multivariate strategies for information maximization. When trying to assess the performance of a sensory system, reduction of the multivariate total correlation, $T(\boldsymbol{x})$, seems the relevant term to look at because univariate entropy maximization (the term depending on $h(x_i)$) can always be performed after joint PDF factorization through a set of (easy-to-do) univariate equalizations, and the noise of the sensors is a restriction that cannot be changed.

Then, the *reduction in redundancy*, $\Delta T(\boldsymbol{r},\boldsymbol{x}) = T(\boldsymbol{r}) - T(\boldsymbol{x})$, is a possible measure of performance: *the system is efficient in regions of the space of images where ΔT is big.* Interestingly, this performance measure, ΔT, can be written in terms of univariate quantities and the Jacobian of the mapping S [18]. Generalizing the expression given in [18] to response models that do nor preserve the dimensionality we have:

$$\Delta T(\boldsymbol{r},\boldsymbol{x}) = \Delta h_m(\boldsymbol{r},\boldsymbol{x}) + \frac{1}{2} E_{\boldsymbol{r}}\Big\{ \log_2 |\nabla_{\boldsymbol{r}} S^\top \cdot \nabla_{\boldsymbol{r}} S| \Big\} \tag{4}$$

Eq. 4 is good for our purposes for two reasons: (1) *in case the marginal difference, Δh_m, is approximately constant over the space of interest*, the performance is totally driven by the Jacobian of the response, so it can be theoretically studied from the model, and (2) even if Δh_m is not constant, the expression is still useful to get robust estimates of ΔT because the multivariate contribution may be get analytically from the Jacobian of the model and the rest reduces to a set of univariate entropy estimations (which do not involve multivariate PDF estimations). In the Results section 3, estimates of ΔT using Eq. 4 are referred to as *theoretical estimation* (as opposed to model-agnostic empirical estimates purely based on samples) because of this second reason.

In previous works, Eq. 4 has been used to describe the communication performance of Divisive Normalization [3] and Wilson-Cowan interaction [20] on achromatic scenes exclusively from the analytical expressions of the corresponding Jacobian. In both cases, these studies used Eq. 4 to analyze the performance at a single layer, and Δh_m was explicitly shown to be constant over the considered domain. Therefore the considerations on the analytical Jacobian certainly explained the behavior of the system.

However, Δh_m may not be constant in general, and hence, the trends obtained from the Jacobian of the model can be counteracted by the variation of Δh_m. Similar considerations can be made with Eq. 2: we also find this transform-dependent term, $\nabla_{\boldsymbol{r}} S$, whose behavior can be successfully analyzed over the considered image space [3, 20]; however, there is no guarantee that the other terms are constant and can be disregarded in the analysis, particularly dealing with comparisons between multiple layers. Moreover, the situation seems worse in Eq. 2 because the terms that should be constant are multivariate in nature, and hence in principle, more difficult to estimate.

Therefore, since the intuition from the analytical response (or from ΔT) is conclusive only in restricted situations, there is a need for empirical methods to estimate the capacity directly from sets of stimuli and the responses they elicit.

Here we use a recently proposed estimation of capacity [21] based on a Gaussianization technique, the so-called *Rotation-Based Iterative Gaussianization* (RBIG) [22], that reduces the problematic (multivariate) PDF estimation problem involved in naive estimation of I to a set of easy (univariate) marginal

PDF estimations. We compare the accuracy of RBIG with theoretical results of ΔT and with classical Kozachenko-Leonenko estimator [23], and variations [24].

3 Experiments and Results

In the experiments, $19 \cdot 10^6$ image patches from the IPL color-calibrated database [25] were characterized according to their chromatic contrast, achromatic contrast and mean luminance, and were injected through the perceptual network to get the responses and compute the information theoretic measures[2].

First, we computed redundancy reduction in the inner visual representation because we have a theoretical reference, Eq. 4, we can compare with. We computed ΔT with three estimators: the one based on Gaussianization [22], the classical Kozachenko-Leonenko estimator [23], and an improved version of Kozachenko-Leonenko [24]. There results are shown in Fig. 1.

Then, we analyze the transmitted information in two ways: (1) we compare the amount of information that can be extracted from images at the cortical representation with the PDF of natural images, see results in Fig. 2; and (2) we plot

[2] Model at http://isp.uv.es/code/visioncolor/vistamodels.html, data at http://isp.uv.es/data_calibrated.html, and Gaussianization estimator at http://isp.uv.es/rbig.html

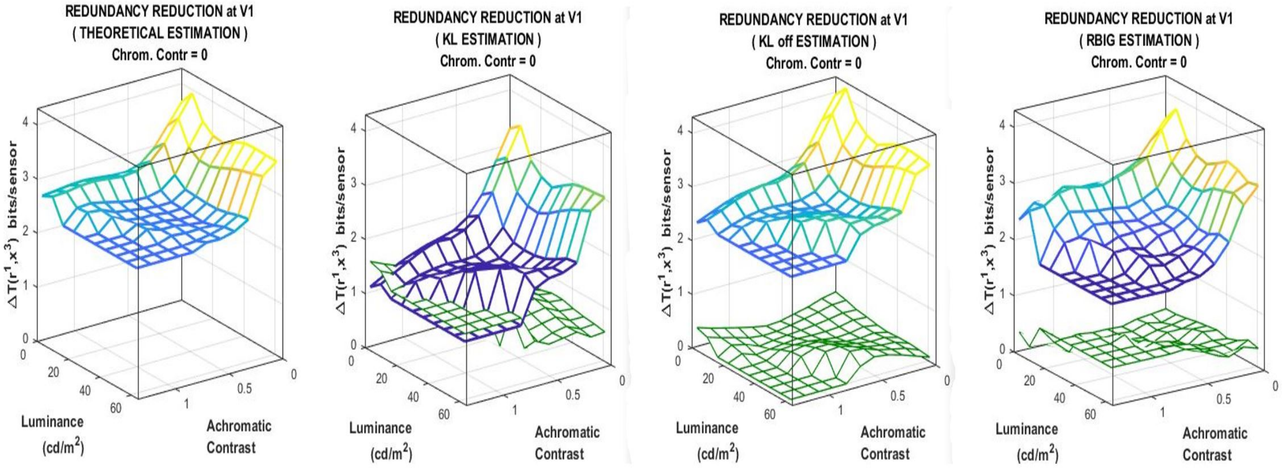

Fig. 1. Agreement between estimations of redundancy reduction (ΔT, in bits) between the LMS input and the internal representation in V1, estimated via the theoretical approach of Eq. 4, computed as in [3] (left), the classical Kozachenko-Leonenko estimator (second plot), and the offset corrected Kozachenko-Leonenko (third plot), and the RBIG estimator (right). The green surfaces represent the absolute difference between the theoretical estimation and the different estimates. These results imply that the estimation of I (for which there is no theoretical reference to compare with) can be trusted for the RBIG [21] and the modified Kozachenko-Leonenko estimator [24], but not for the classical Kozachenko-Leonenko estimator [23].

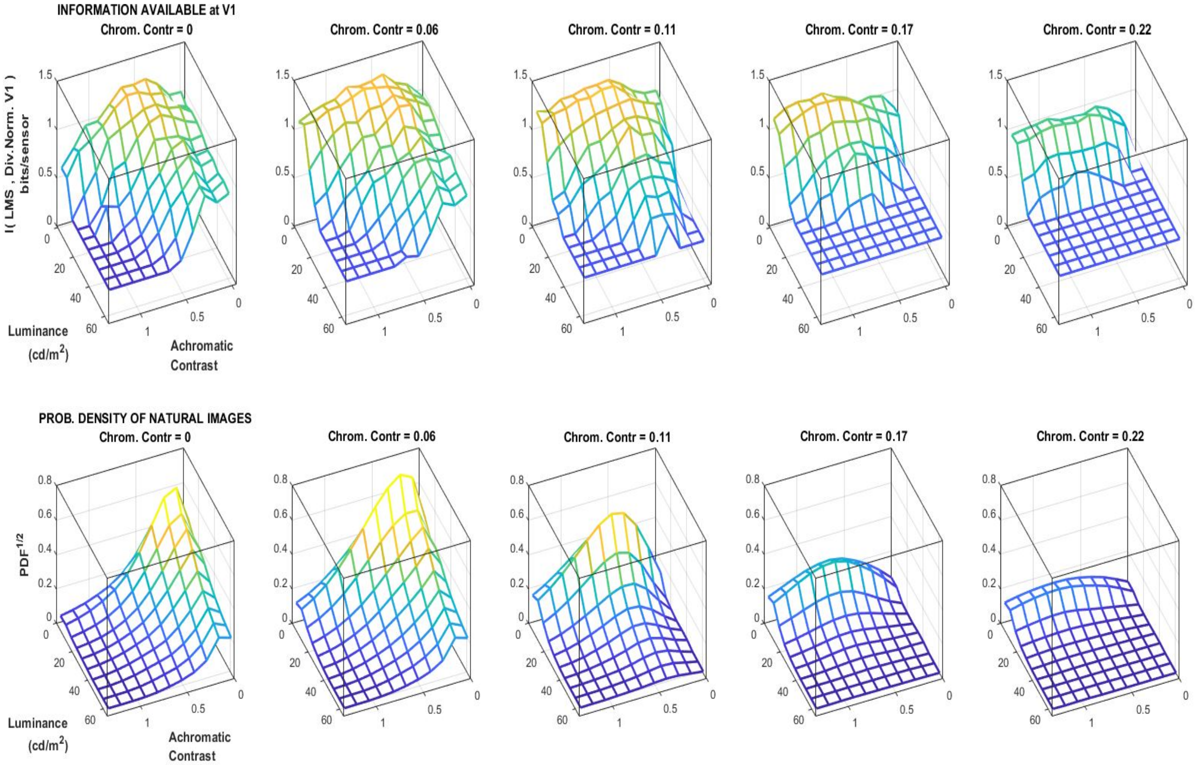

Fig. 2. Information available at the cortical representation after Divisive Normalization at different regions of the image space (top) compared to the PDF of natural images (bottom). Note that images with smooth achromatic variation are more frequent than sharp chromatic patterns, and the cortical representation captures more information exactly in those regions.

the amount of information available from different layers of the psychophysical network assuming sensors of the same signal-to-noise ratio at every layer.

On the one hand, Fig. 2 shows that, despite using no statistical training, the perceptual network is more efficient in the more populated regions of the image space. And, on the other hand, Fig. 3 shows how the series of transforms along the neural pathway progressively increase the information about the input which is available from the corresponding layer. In this regard, note that spatial transforms have a bigger contribution to the increase in available information than chromatic transforms.

4 Discussion

Results imply that (1) the biggest contribution to improve transmission is the analysis of opponent images through local-oriented filters and Divisive Normal-

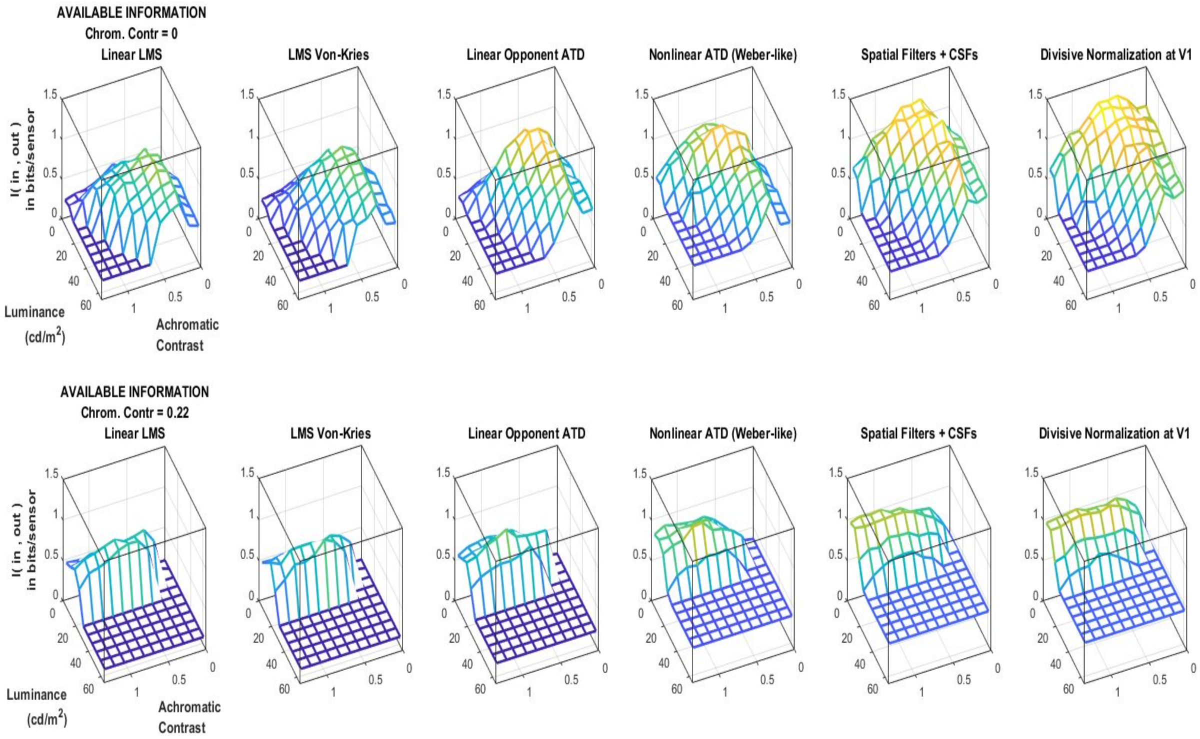

Fig. 3. Information about the scene available from different layers of the visual pathway. Results are shown over the achromatic contrast and luminance space for two fixed chromatic contrasts: the minimum (zero, on the top) and the maximum in our set (on the bottom). This result implies that using sensors of equivalent quality (5% of signal deviation), the cortical representation is more appropriate because it doubles the amount of information captured from the input.

ization (about 70%) as opposed to the 30% that comes from the previous chromatic transforms, (2) the capacity is remarkably well adapted to the PDF of natural images, and (3) the internal representation captures substantially more information about the images than the trivial representation at the photoreceptor domain. This efficiency is inspiring for artificial systems, particularly considering that no statistical training was required here.

Examples of the consequences in image processing include the generalization of the *Visual Information Fidelity* (VIF) [26] concept. VIF is an original approach to characterize the distortion introduced in an image which is based in comparing the information about the scene that *a human* could extract from the distorted image wrt the information that he/she could extract from the original image. Our results have two kinds of implications in VIF. First, one may improve the perceptual model and noise schemes in VIF because the non-parametric RBIG estimation is insensitive to the complexity of the model. Second, original

VIF made crude approximations on the PDF of the signals to apply analytical estimations of I, which may be too biased. Better measures of I not subject to approximated models could certainly improve the results.

5 Conclusions

In this work we quantified how neural networks designed from biological models and using no statistical training have a remarkable performance in information theoretic terms. Specifically, using two empirical estimators of mutual information [22, 23], we computed the transmission capacity at different layers of standard biological models [3, 19, 20]. From the technical point of view, we found that Gaussianization-based estimations of total correlation [22] are substantially more accurate than the original Kozachenko-Leonenko estimator [23], and its performance is similar to more recent (offset corrected) Kozachenko-Leonenko estimators [24]. Regarding the behavior of the considered visual network, we found three interesting results: (1) progressively deeper layers have bigger capacity (assuming the same quality of the sensors at every layer) indicating that biological transforms may be optimized to maximize transmitted information. (2) the internal representation of these networks duplicates the amount of information that can be extracted about the images with regard to the amount of information that could be obtained from the input representation, and (3) the capacity of internal representation follows the PDF of natural scenes over the chromatic and achromatic dimensions of the stimulus space.

This remarkable adaptation to the natural environment is an additional confirmation of the Efficient Coding Hypothesis [20, 27, 28], and an additional example of how imitation of biological vision may inspire architectures and save training effort in artificial vision.

References

1. A. Krizhevsky, I. Sutskever, and G.E. Hinton. Imagenet classification with deep convolutional neural networks. In *25th Neur. Inf. Proc. Syst.*, NIPS'12, pages 1097–1105, USA, 2012. Curran Associates Inc.
2. M. Carandini and D. J Heeger. Normalization as a canonical neural computation. *Nature Rev. Neurosci.*, 13(1):51–62, 2012.
3. M. Martinez, P. Cyriac, T. Batard, M. Bertalmío, and J. Malo. Derivatives and inverse of cascaded L+NL neural models. *PLOS ONE*, 13(10):1–49, 10 2018.
4. A. Stockman and D.H. Brainard. *OSA Handbook of Optics (3rd. Ed.)*, chapter Color vision mechanisms, pages 147–152. McGraw-Hill, NY, 2010.
5. M.D. Fairchild. *Color Appearance Models*. The Wiley-IS&T Series in Imaging Science and Technology. Wiley, 2013.
6. L.M. Hurvich and D. Jameson. An opponent-process theory of color vision. *Psychol. Rev.*, 64(6):384–404, 1957.
7. J. Krauskopf and K. Gegenfurtner. Color discrimination and adaptation. *Vision Res.*, 32(11):2165 – 2175, 1992.

8. K. T. Mullen. The CSF of human colour vision to red-green and yellow-blue chromatic gratings. *J. Physiol.*, 359:381–400, 1985.
9. A. B. Watson and J. A. Solomon. Model of visual contrast gain control and pattern masking. *JOSA A*, 14(9):2379–2391, 1997.
10. J Malo, JJ Esteve-Taboada, and M Bertalmío. Divisive normalization from Wilson-Cowan dynamics. *ArXiv: Quant. Biol. https://arxiv.org/abs/1906.08246*, 2019.
11. J. Malo, J. Gutiérrez, I. Epifanio, F. J Ferri, and José M Artigas. Perceptual feedback in multigrid motion estimation using an improved dct quantization. *IEEE Trans. Im. Proc.*, 10(10):1411–1427, 2001.
12. J. Malo, I. Epifanio, R. Navarro, and EP. Simoncelli. Nonlinear image representation for efficient perceptual coding. *IEEE Trans. Im. Proc.*, 15(1):68–80, 2006.
13. A. B Watson and J. Malo. Video quality measures based on the standard spatial observer. In *IEEE Int. Conf. Im. Proc.*, volume 3, pages III–41, 2002.
14. V. Laparra, J. Muñoz-Marí, and J. Malo. Divisive normalization image quality metric revisited. *JOSA A*, 27(4):852–864, 2010.
15. Johannes Ballé, Valero Laparra, and Eero P. Simoncelli. End-to-end optimized image compression. In *5th Int. Conf. Learn. Repres., ICLR 2017*, 2017.
16. V. Laparra, A. Berardino, J. Balle, and E.P. Simoncelli. Perceptually optimized image rendering. *JOSA A*, 34(9):1511–1525, 2017.
17. T. M. Cover and J. A. Thomas. *Elements of Information Theory, 2nd Edition*. Wiley-Interscience, 2 edition, July 2006.
18. M. Studeny and J. Vejnarova. *The Multi-information function as a tool for measuring stochastic dependence*, pages 261–298. Kluwer, January 1998.
19. J. Malo. Spatio-chromatic information available from different neural layers via gaussianization. *ArXiv: Quant. Biol. https://arxiv.org/abs/1910.01559*, 2019.
20. A. Gomez-Villa, M. Bertalmio, and J. Malo. Visual information flow in Wilson-Cowan networks. *ArXiv: Quant. Biol. https://arxiv.org/abs/1907.13046*, 2019.
21. J.E. Johnson, V. Laparra, R. Santos, G. Camps, and J. Malo. Information theory in density destructors. In *7th ICML 2019, Workshop Invertible Norm. Flows*, 2019.
22. V. Laparra, G. Camps-Valls, and J. Malo. Iterative gaussianization: from ICA to random rotations. *IEEE Trans. Neural Networks*, 22(4):537–549, 2011.
23. L.F. Kozachenko and N.N. Leonenko. Sample estimate of the entropy of a random vector. *Probl. Inf. Trans.*, 23:95–101, 09 1987.
24. I. Marin and DH. Foster. Estimating information from image colors: Application to digital cameras and natural scenes. *IEEE Trans. PAMI*, 35(1):78–91, 2013.
25. V. Laparra, S. Jiménez, G. Camps-Valls, and J. Malo. Nonlinearities and adaptation of color vision from sequential principal curves analysis. *Neural Computation*, 24(10):2751–2788, 2012.
26. H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, Feb 2006.
27. H. Barlow. Redundancy reduction revisited. *Network: Comp. Neur. Syst.*, 12(3):241–253, 2001.
28. J. Malo and V. Laparra. Psychophysically tuned divisive normalization approximately factorizes the pdf of natural images. *Neural computation*, 22(12):3179–3206, 2010.