presented at the annual
Bernstein Conference
BCCN 2009 in Frankfurt.

# Hierarchical Models of Natural Images

Reshad Hosseini and Matthias Bethge
*Max Planck Institute for Biological Cybernetics, Spemannstr. 41, 72076 Tübingen*

Here, we study two different approaches to estimate the multi-information of natural images. In both cases, we begin with a whitening step. Then, in the first approach, we use a hierarchical multi-layer ICA model [1] which is an efficient variant of projection pursuit density estimation. Projection pursuit [2] is a nonparametric density estimation technique with universal approximation properties. That is, it can be proven to converge to the true distribution in the limit of infinite amount of data and layers.
For the second approach, we suggest a new model which consists of two layers only and has much less degrees of freedom than the multi-layer ICA model. In the first layer we apply symmetric whitening followed by radial Gaussianization [2,3] which transforms the norm of the image patches such that the distribution over the norm of the image patches matches the radial distribution of a multivariate Gaussian. In the next step, we apply ICA. The first step can be considered as a contrast gain control mechanism and the second one yields edge filters similar to those in primary visual cortex.
By evaluating quantitatively the redundancy reduction achieved with the two approaches, we find that the second procedure fits the distribution significantly better than the first one. On the van Hateren data set (400.000 image patches of size 12x12) with log-intensity scale, the redundancy reduction in the multi-layer ICA model yields 0.162, 0.081, 0.034, 0.021, 0.013, 0.009, 0.006, 0.004, 0.003, 0.002 bits/pixel after the first, second, third, fourth, …, tenth layer, respectively. For the training set size used, the performance decreases after the tenth layer). In contrast, we find a redundancy reduction of 0.342 bits/pixel after the first layer and 0.073 bits/pixel after the second layer for the second approach.
In conclusion, the universal approximation property of the deep hierarchical architecture in the first approach does not pay off for the task of density estimation in case of natural images.

[1] Chen and Gopinath. 2001. *Proc. NIPS*, vol. 13, pp. 423–429.
[2] Friedman J. et al. 1984. *J. Amer. Statist. Assoc.*, vol. 71, pp. 599–608.
[3] Lyu S. and Simoncelli E. P. 2008. *Proc. NIPS*, vol. 21, pp.1009–1016.
[4] Sinz F. H. and Bethge M. 2008. *MPI Technical Report*

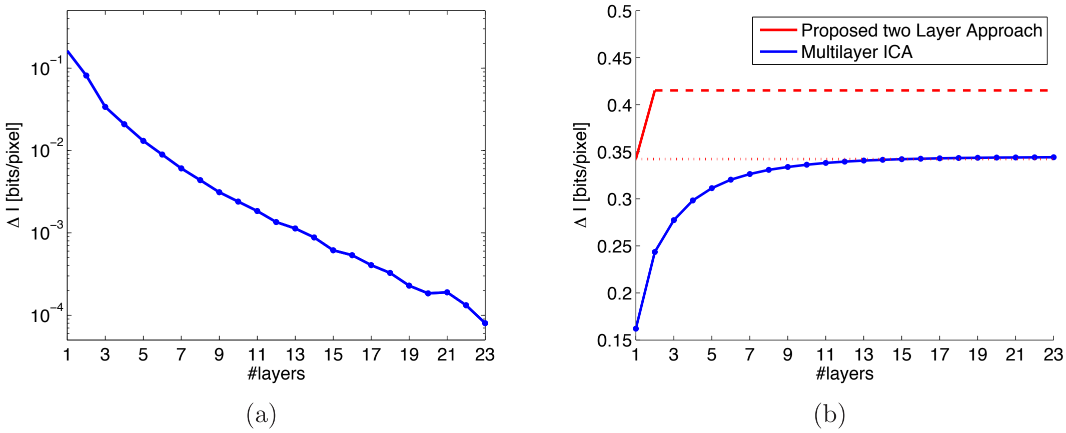



Figure 1: Subfigure (a): The amount of improvement of the mutual information in the multilayer ICA method for different layers. Blue curve in subfigure (b): Cumulative curve of the same improvement. Red curve in the subfigure (b): Cumulative mutual information improvement for our two layer approach.

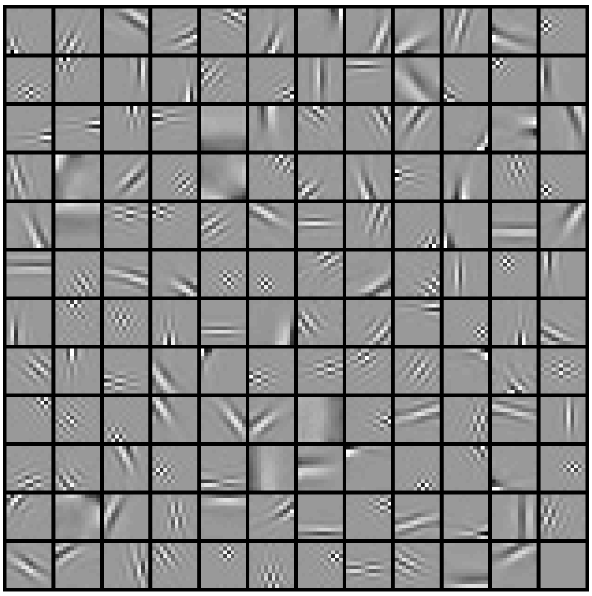


Figure 2: The filters that we obtain in the second layer of our proposed method.

# Hierachical Models of Natural Images

Reshad Hosseini and Matthias Bethge

[1]Max Planck Institute for Biological Cybernetics,Tübingen

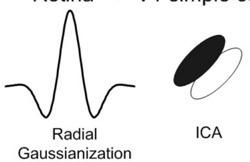
## Background/Motivation

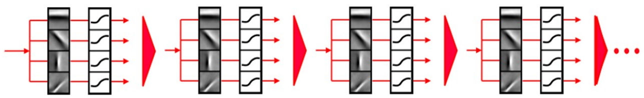
ICA → $L_p$ radial factorization [Sinz & Bethge] (State-of-the-art in redundancy reduction)

ICA → Projection Pursuit density estimation [Friedman]

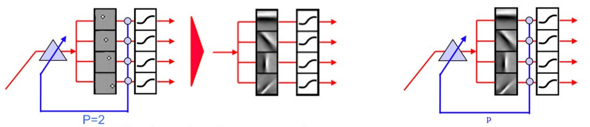
$L_p$ radial factorization → Physiologically-inspired two-layer model
Retina + V1 simple cells

Cascadable Neural Architechture for redundancy reduction
Multilayer ICA [Chen & Gopinath]



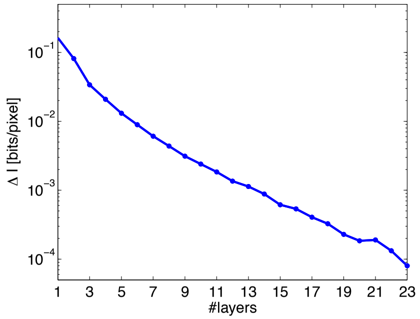



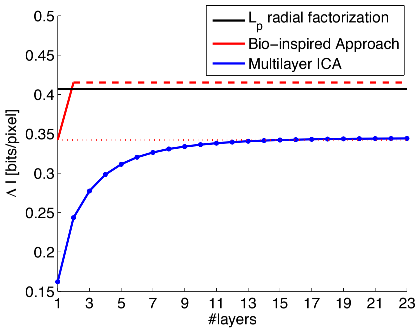
Muli-layer ICA

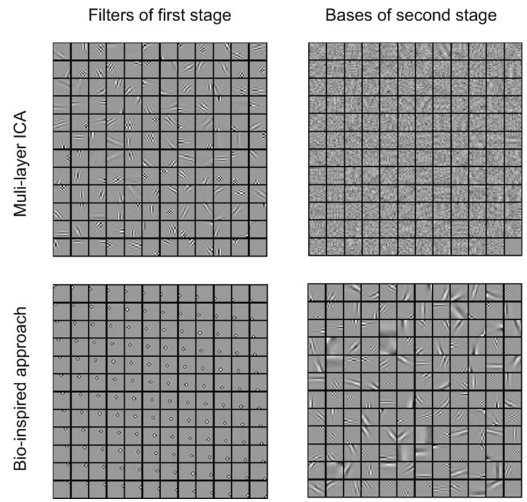



Bio-inspired approach
$L_p$ radial factorization

## Results




## Conclusion

- The universal approximation property of the deep hierarchical architecture in the first approach does not pay off for the task of density estimation in case of natural images.
- Finding effective neural architechture (i.e. nonlinearity) is critical.

## Method

### Quantitative evaluation

- Multi-information:

$$I[X_1,...,X_k] = \sum_{i=1}^{k} \mathrm{H}(X_i) - \mathrm{H}(\mathbf{X}_{1:k})$$

- Averaged Log-loss measure:

$$\hat{\mathrm{H}}(\mathbf{X}_{1:k}) = -\langle \log \hat{p}_{\mathbf{X}_{1:k}}(\mathbf{x})\rangle_{\mathbf{X}_{1:k}} \approx -\frac{1}{m}\sum_{i=1}^{m} \log \hat{p}_{\mathbf{X}_{1:k}}(\mathbf{x}_i)$$

$$\hat{\mathrm{H}}(\mathbf{X}_{1:k}) = \mathrm{H}(\mathbf{X}_{1:k}) + \mathrm{D}_{KL}(\hat{p}_{\mathbf{X}_{1:k}} || p_{\mathbf{X}_{1:k}})$$

$$\mathrm{ALL} = \sum_{i=1}^{k} \mathrm{H}(X_i) - \hat{\mathrm{H}}(\mathbf{X}_{1:k})$$

### Independent component Analysis

- Given the ensemble of data $X$
- Finding a linear transformation $A$

$$Y = BX$$
$$A = \min_B I[Y_1,...,Y_k]$$

- The matrix has the following factorization:

$$A = Q\Lambda$$

$Q$ is orthogonal and $\Lambda$ is positive definite

- $\Lambda$ is called whitenning matrix. Let $\Sigma$ be the covariance of the data, then

$$\Lambda = \Sigma^{-1/2}$$

## References

[1] Scott Shaobing Chen and Ramesh A. Gopinath. Gaussianization. *Proc. NIPS*, 13:423–429, 2001.

[2] Jerome H. Friedman, Werner Stuetzle, and Anne Schroeder. Projection pursuit density estimation. *J. Amer. Statist. Assoc.*, 71:599–608, 1984.

[3] Siwei Lyu and Eero P. Simoncelli. Reducing statistical dependencies in natural signals using radial gaussianization. *Proc. NIPS*, 21:1009–1016, 2008.

[4] Fabian H. Sinz and Matthias Bethge. How much can orientation selectivity and contrast gain control reduce the redundancies in natural images. *MPI Techniacal Report*, (169), 2008.