The role of nonlinearities in hierarchical feed-forward models for pattern recognition

SPATIAL COGNITION

A5-[ActionSpace]

Sven Eberhardt[1], Tobias Kluth[2], Manfred Fahle[1], Christoph Zetzsche[2]

[1]Institute for Human Neurobiology, Universität Bremen [2]Cognitive Neuroinformatics, Universität Bremen

Introduction

Hierarchical feed-forward models for pattern recognition have been proven to be plausible in explaining receptive field properties of neurons in the ventral stream of primate visual cortex. In particular, the HMAX-model presented by Serre et al, 2007[1] can account for psychophysical results in rapid object categorization experiments by using alternations of matching and pooling layers.

Here we show a generalized form of the HMAX model in which we systemically vary the nonlinearities at each layer in the hierarchy between specificity and invariance using a control parameter. We also apply two popular training methods, Slow Feature Analysis[2] and Principal Component Analysis, to determine the weights for receptive fields of the neurons.

Control parameter

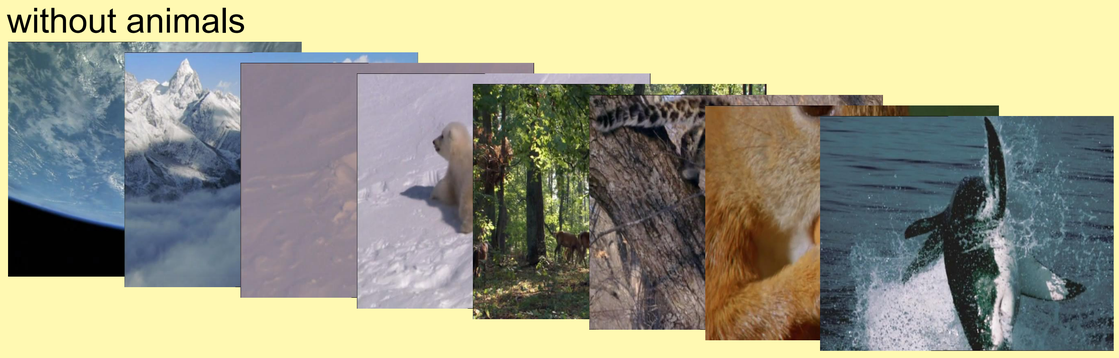
Here we introduce a control parameter for the nonlinearity which effects the grade of nonlinearity in the sense of being selective or invariant with respect to the different input dimensions. The function $f_k(x)$ is defined as follows:

$$f_k(x) = \begin{cases} n^{-e^k}\left(\sum_{i=1}^{n} |x_i|^{e^k}\right)^{\frac{1}{e^k}}, & \text{if } k \geq 0 \\ \max_i(|x_i|) - n^{-e^{|k|}}\left(\sum_{i=1}^{n} (\max_i(|x_i|) - |x_i|)^{e^{|k|}}\right)^{\frac{1}{e^{|k|}}}, & \text{if } k < 0 \end{cases}$$

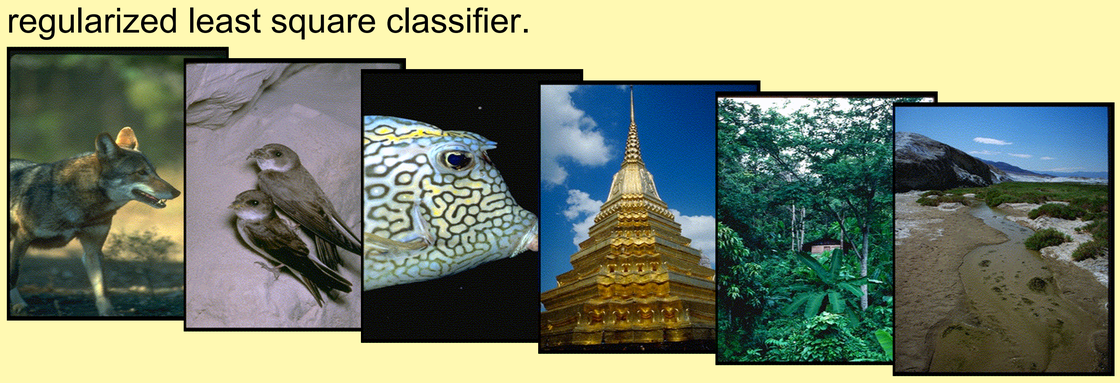
where k is the control parameter, x is the input vector including different dimensions, and n is the number of dimensions. Low values of k (left) produce specific, i.e. template matching, functions and high values prouce invariant, i.e. pooling, functions.

Datasets

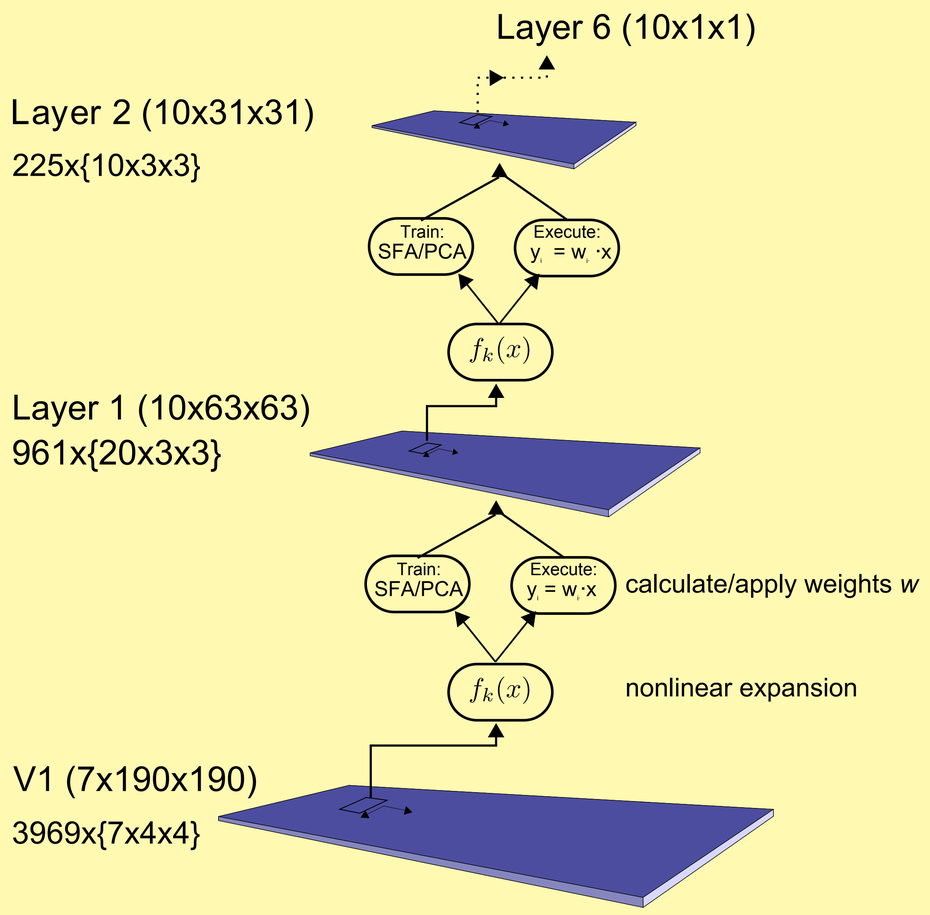
Training of model weights happened on a full 44 minute video of the BBC planet earth collection resampled to 10 frames per second. This results in 26207 image frames containing outdoor scenes both with and without animals



For categorization testing we used the animal/non-animal image database also used by VanRullen et al[3]. Randomized subsets of 400 training and 100 testing images were taken and fed into a linear regularized least square classifier.



Model

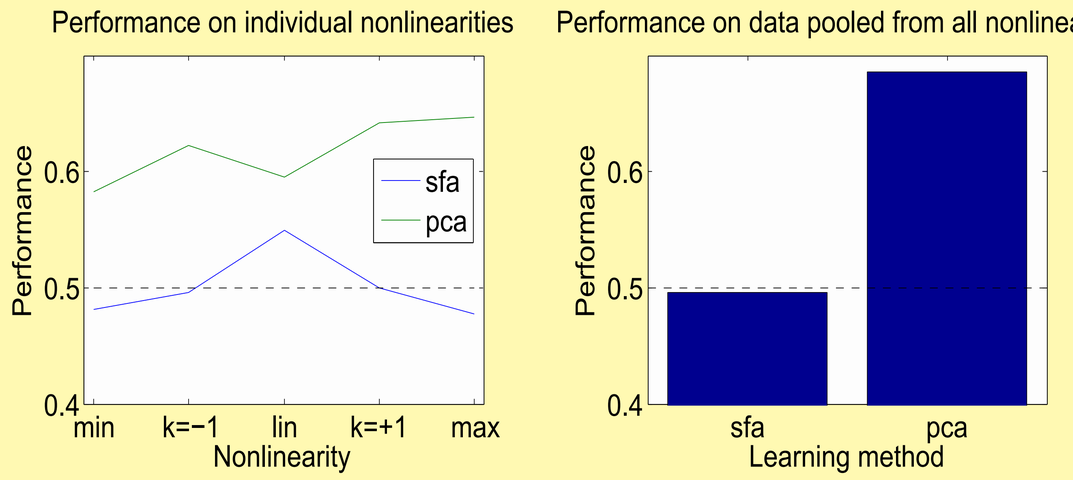


Setup of the hierarchical multi-layer system:

- **V1:** We used the responses of different Gabor filters (4 orientations) plus 3 relative color channels as the input of our model.
- **Layers 1-6:** We used the 10 slowest features (SFA) or 10 principal components at ech local position in the layer.

Each receptive fields of a cell in the layer above has three dimensions. The first dimension is the feature dimension or the orientations/colors in V1. The second and third dimensions are describing the patch size. For each layer the receptive field is illustrated as number of different patches times $\{n_{feature}, n_x, n_y\}$.

Classification results



The classification task fails on SFA trained features, but works on PCA despite using only 10 features!
There's a weak dependence on nonlinearity. But: Combined features of all nonlinearities increase performance again. This leads us to believe that different nonlinerities are required to encode different features.

References

[1] Serre T, Oliva A, Poggio T (2007). A feedforward architecture accounts for rapid categorization. *PNAS* 104, no. 15: 6424-6429.
[2] Wiskott L, Sejnowski T (2002). Slow feature analysis: Unsupervised learning of invariances. *Neural Computation* 14(4): 715-770
[3] VanRullen R, Thorpe S J (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. *Perception* 30: 655-668

Universität Bremen

DFG Deutsche Forschungsgemeinschaft

UNI FREIBURG