Slow features between encode high-level concepts on HMAX outputs

SPATIAL COGNITION

A5-[ActionSpace]

Sven Eberhardt, Tobias Kluth, Christoph Zetzsche, Kerstin Schill

Introduction

Biologically motivated models for visual object recognition often assume a processing hierarchy that starts from detecting simple features such as edges and corners on lower levels and combines those to complex high-level features such as objects or faces on higher levels. In particular, the HMax model (as popularized by Serre et al., 2007) proposes a feed-forward architecture that combines matching of increasingly complex features on higher levels with increasing invariance to scale and position by a structure of alternating layers. Features of the model have been used successfully to perform classification tasks such as animal detection or object classification. However, although wiring of the lower levels can be sampled from natural images, high-level concepts such as animals or object categories are usually introduced by means of supervised learning employing labeled examples. Here, we show that Slow Feature Analysis (SFA, Berkes and Wiskott, 2005) can be used to learn these concepts on high-level HMax outputs in an unsupervised manner.

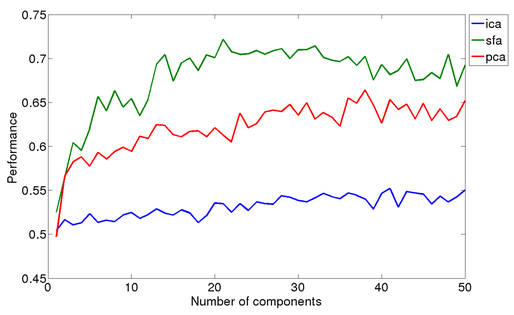
Unsupervised learning

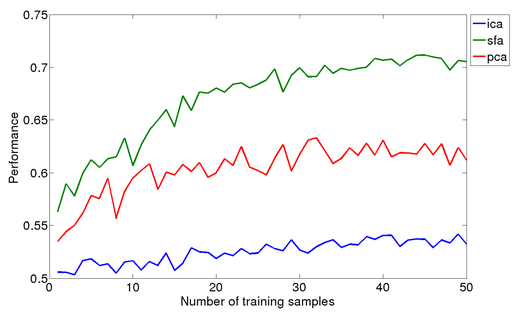
We used three different types of unsupervised learning techniques to obtain thr top layer transformation.

- **PCA**: The features are linearly transformed into the principal components, i.e. the components which are decorrelated and have maximum variance.
- **ICA**: First the the features are whitened, i.e. a PCA is performed and the inverse of the eigenvalue matrix is multiplicated. Second the features are transformed into statistically independet components.
- **SFA**: First the features are whitened. Second the features are transformed into slow components, i.e. the components which minimize the variance of the time-derivative of the input feature.

Classification

We test classification performance on the animal/nonanimal dataset as used be Serre et al.[ref] We vary both the number of (PCA-/ICA-/SFA-)components and the number of training samples.





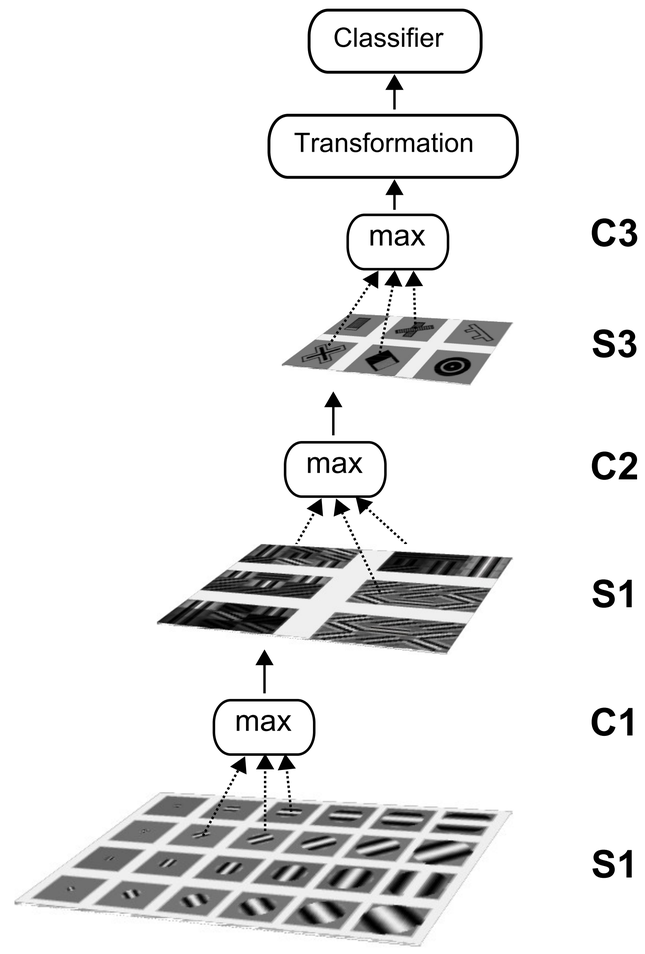
The results show that

- Component weights determined by SFA lead to higher performance than principal and independent components.
- This result is stable for varying training set sizes
- The strongest effect happens for the first 10-30 components, hinting that the most slowly varying features are relevant for animal detection

We conclude that slow feature analysis can be used to extract feature weights for high level object classification tasks. However, the results are preliminary and some control studies need to be made. Most importantly:

- We do not know if weights determined by SFA actually encode animal features, or just features good for object classification tasks in general
- We do not know if the results are due to using a video containing animals, or if any input video would have resulted in this finding
- We do not know if features are actually high level patterns such as animal contours, or if low level cues such as statistics over gabor filter orientation are sufficient

Model



The setup of the hierarchical multi-layer system is as follows:

- **HMAX**: We used the responses of the complex layer C3 as the input for the linear transformation.
- **Top layer**: We used different numbers of different features (trained by PCA,SFA, or ICA)

The different layers consist of 6 scales. The size of the biggest scale in each layer is shown in the sketch of the model. The transformed features, i.e. the output features of the model, are directed to a linear classifier.

Training

To train the components, we use a 256x256 central patch of part one of the BBC Planet Earth video collection. The video is 44 minutes long and resampled at 10 frames per second resulting in 26,207 input frames. It contains a large number of outdoor scenes both with and without animals moving over the input region at several scales. Scenes are typically between 10 and 200 frames long and show animals from arctic, prairie, rainforest and underwater regions.

The images are fed into the HMax model and 2000 C3 features are extracted for each frame, producing a 2000x26207 output vector. On this vector, weights for PCA components, ICA components and SFA features are computed. The weights for each method are then applied to the HMax model outputs of the classification image dataset to yield input for the classifier.

References

1. Serre T, Oliva A, Poggio T (2007). A feedforward architecture accounts for rapid categorization. PNAS 15(105): 6424-6429.
2. Hyvrinen A, Oja E (2000). Independent component analysis: algorithms and applications. Neural Networks 4-5(13): 411-430.
3. Wiskott L, Sejnowski T (2002). Slow feature analysis: Unsupervised learning of invariances. Neural Computation 14(4): 715-770.

Universität Bremen

DFG Deutsche Forschungsgemeinschaft

UNI FREIBURG