Capturing Nonlinear Visual Structure by Recursive ICA

Honghao Shan, Garrison W. Cottrell

University of California, San Diego

The efficient coding hypothesis has been a popular theory in explaining the functional role of low-level vision. In its spirit, the theory suggests that low-level vision captures the statistical structure of the visual stimuli with minimum energy cost. When independent component analysis (ICA), a linear implementation of the efficient coding theory, is applied to natural image patches, the resulting basis functions resemble the V1 simple cells’ receptive fields [1, 2].

Recently many efforts have been put into building a hierarchical statistical model which captures nonlinear structures not captured by ICA. Such a model might provide a functional explanation of higher visual layers in the human visual pathway. Since higher visual layers share similar anatomical structures as the primary visual cortex, we hypothesize that they might also work under similar computational principles as V1. Based on this hypothesis, we derived a hierarchical model, the Recursive ICA (RICA) model [3], which captures nonlinear visual structures of natural images. This model contains a number of layers. ICA is at work on each layer and coordinate-wise nonlinear activation functions transform the outputs of the lower layer as the inputs of the higher layer. The higher layer features have larger receptive fields than the lower layer features, and capture more complex visual structures.

We have applied the RICA model on different natural image datasets. The visual structures captured are qualitatively similar. Specifically, the second layer of the RICA model exhibits properties similar to what have been observed in V2 neurons. About 70% of the second layer features correspond to angles, curves and noncartesian gratings. The receptive fields of the rest 30% features consist of sub-regions tuned to different orientations of bars/edges. We applied the learned visual structures for different pattern recognition tasks, and the resulting classifier outperformed most state-of-the-art computer vision techniques.

Acknowledgments
This work was supported by NIMH grant MH57075 to GWC, and NSF Science of Learning Center grant to the Temporal Dynamics of Learning Center #SBE-0542013 (GWC, PI).

References

