Exploring computational models of visual object perception

Darren A Seibert\textsuperscript{2,3}, Daniel D Leeds\textsuperscript{1}, John A Pyles\textsuperscript{1,4} and Michael J Tarr\textsuperscript{1,4}

\textsuperscript{1}Center for the Neural Basis of Cognition, Carnegie Mellon University (CMU), Pittsburgh, PA, \textsuperscript{2}Department of Biomedical Engineering, University of Houston, Houston, TX, \textsuperscript{3}UpMC Summer Program, CMU, \textsuperscript{4}Department of Psychology, CMU

Encoding and decoding ventral activity

- Models of perception in anterior stages of the visual stream are few in number and tests of these models' consistency with neural data have been limited
- Cadieu et al have demonstrated HMAX's ability to predict responses in V4
- We explore HMAX's ability to describe fMRI activity throughout the ventral stream

Experimental design

- Participants shown images of 60 objects, 6 x each
- BOLD signals recorded with slow event-related design (2 sec TR, partial coverage)

Measuring responses—Searchlight Projection

- Constructed "searchlight"—123 voxel sphere—centered at each voxel (Kriegeskorte et al., 2006)

The HMAX model:

Selectivity layer:
- Determine extent to which image patch contains an edge

Tolerance layer:
- Determine maximal response laterally and across scales

Fitting the 1st pair of layers

- Adjust model configuration to reduce error (greedy descent)
- Compute prediction error (mean squared difference)

HMAX

Fitting the 2nd pair of layers

- Determine feature selectivity for the second pair of layers using a greedy search algorithm (Cadieu et al., 2007)

Fitting results for the first pair of layers

- Voxels significantly predicted by C1 and fit to have $n=1$ for C2

Correlation between C2 and fMRI

- Preference for larger edges

Increasing scale size

Voxels significantly predicted by C1 and fit to have $n=1$ for C2

Second pair of layers

- Best model fMRI data
- Selecting for fewer C1 units
- Using 5x as much training available to us

References


Acknowledgments

Funded by NSF IGERT, R.K. Mellon Foundation, NIH EUREXA Award #R35013MR084955-01, and the Temporal Dynamic of Learning Center at UCSF (NSF Science of Learning Center SBE-0542013)