Mixing hierarchical edge detection and medial axis models of object perception

Diverse models of cortical perception
- Multiple potential frameworks have been proposed to characterize cortical object perception
- Prominent models capturing holistic shapes and local-features show promise accounting for cortical responses to visual objects (Leeds 2013; Hung 2012; Yamins 2014)
- We show mixing holistic Shock graph and local-features SIFT models sometimes – but not always – provides a better account of cortical activity than does either model alone

Methods
- Participants shown photos of 60 real-world objects, 6 x each, passive viewing
- BOLD signals recorded with slow event-related design (2 sec TR, partial coverage) for 5 subjects

Representational dissimilarity analysis
- Representational dissimilarity: use pairwise distance matrix to show how stimuli are grouped by each neural and computational representation

Joint SIFT/Shock graph model
Investigate contributions of each model to explain neural groupings (neural distance matrices) with linear, bilinear, and quadratic terms

Find $c_{ij} \ldots c_{kj}$ to best-fit
$$\text{neural}(i,j) = \text{combine}(i,j) = c_1 + c_2 \text{SIFT}(i,j) + c_3 \text{Shock}(i,j) + c_4 \text{SIFT}(i,j)^2 + c_5 \text{Shock}(i,j)^2$$

for each pair of stimuli $(stim_1, stim_2)$

Find $r_{\text{combine}} = \text{corr(neural, combine)}$

Correlation increases from joint model
- LOCizer increases
- SIFT only
- Shock graph
- Joint region

Pairwise distance matrix

0.7 0.8 0.7 0.1 0.8 0.1 0

Neural: 123-voxel sphere of responses at each location

Computational models:
- SIFT – bag of multi-scale Gabor-based visual words (similarities to Convolutional Neural Nets)
- Shock graph – enhanced medial axis

Neural vs model comparison: Spearman correlation between elements of neural and model distance matrices (Leeds 2013)

Distribution of centers for regions of correlation increase
- Bilateral
- Posterior ventral (+ dorsal)
- Inferior occipital
- Fusiform
- Lateral occipital
- Lingual gyrus
- Superior parietal
- Anterior dorsal
- Prefrontal cortex
- Smaller regions

Joint coefficients for increased correlation
- Typically 1-2 common coefficients per combination-model term
- Typically ~1 standard deviation away from 0
- Typical scale (~0 to 10 or ~0.01 to 0.01) and values per combination-model term across subjects

Discussion
- Non-linear combination of diverse models better accounts for voxel responses in subset of mid- and high-level visual areas
- When combination provides true improvement, best fit combination coefficients fairly consistent across subjects and regions
- Future directions:
  - Expand to Convolutional Neural Networks
  - Expand to further non-linear model combinations

References

Acknowledgments
Fordham University Faculty Research Grant; Pennsylvania Department of Health’s Commonwealth with Universal Research Enhancement Program, NIH 1R15DA031320-01A1, and the Temporal Dynamics of Learning Center at UCD (NSF Science of Learning Center #SBE-1241756).