Hierarchical feed-forward visual models and recurrent semantic models predict fMRI pattern-information in the ventral object processing stream

Barry Devereux, Alex Clarke & Lorraine K. Tyler
Centre for Speech, Language and the Brain, Dept. of Psychology, University of Cambridge
barry@csbl.psychol.cam.ac.uk  csbl.psychol.cam.ac.uk  @csblab

Introduction

- How does visual object processing activate meaningful conceptual representations?
- Recent work in machine learning has shown that deep convolutional neural networks (DCNNs) can capture high-level visual information in images (Krizhevsky et al., 2012, Szegedy et al., 2015, Zeiler & Fergus, 2013)
- These neurobiologically-inspired feed-forward models can explain activation patterns in the visual object processing stream:
  - Lower layers of the DCNN better match earlier stages of the stream
  - Higher layers of the DCNN better match later stages of the stream (Clarke et al., 2014, Gaff & Gereman, 2015, Khalighi-Razavi & Kriegeskorte, 2014)
- However, models are designed to distinguish between objects for the purpose of determining the correct object label; unlike humans, they do not activate conceptual information
- How visual processing activates and interacts with semantic processing in the brain remains unclear

We combine a DCNN model of visual processing with an attractor network model of distributed feature-based conceptual semantics

- In activating the appropriate target semantic representation for object images, the combined network:
  - Models statistical regularities between high-level visual properties and semantic features, and
  - Models statistical regularities between semantic features
  - Earlier activation of shared, categorical information; later activation of specific concept

fMRI data

- 16 participants (age: 19-29) named images of 131 object concepts. 6 runs, random order on each run (Clarke & Tyler, 2014)
- GLM on unsmoothed native-space data to obtain β-image for each of the 131 Items (Kriegeskorte et al., 2008)
- Image patterns used as input to representational similarity analysis (RSA; Kriegeskorte et al., 2008)

GLM analysis

- Multi-voxel searchlight mapping
  - Spearman's rho (Kriegeskorte et al., 2008) as well as anatomical ROIs along the ventral stream

Computational modelling & Analysis Methods

Input image

Deep neural network (feedforward pretrained)

- 7 layers of Caffe implementation of Krizhevsky et al. (2012) DCNN, trained on ImageNet (tsx @ 2014)
- We replace the final output label layer (no semantics) with a distributed attractor network (semantics)
  - To reduce weight parameters for semantic training, we reduce dimensionality of penultimate layer (fc7) using singular value decomposition (SVD)
  - SVD preserves almost all information about object pairwise similarity in fc7 (Spearman's rho = 0.981)
- Later layers of DCNN capture high-level visual information (Zeiler & Fergus, 2013)
  - e.g. Images with strongest activation on nodes 2810 and 7040 of layer conv5

Semantic attractor network (SAN)

- Semantic layer consisting of 2528 recurrently connected semantic feature nodes
- Model trained to activate the distributed semantic feature representation for each image
  - Features from the C117b norms (Devereux et al., 2014)
  - Trained with backprop through time, 25 iterations
  - Based on word meaning model of Cisek et al (2016)
  - Model is sensitive to conceptual structure of features and type of feature:
    - Shared semantic features (shared category info) & visual semantic features activate faster than distinctive and non-visual features

RSAs

- Which stages of the DN+SAN model account for activation patterns along the ventral stream?
- RSA: compare dissimilarity structure of activation patterns in each network layer to the dissimilarity structure of the fMRI activation patterns (Kriegeskorte et al, 2008, Zeiler & Fergus, 2013)
- Multi-voxel searchlight mapping
  - Spearman's rho for each combination of model layer & ROI

Results

- Lateral gradient of visual information throughout the ventral object processing stream
- Posterior temporal gradient of visual-to-semantic processing in ventral stream
- Anterior-posterior gradient of visual-to-semantic processing in ventral stream
- Mechanisms of the combined vis+sem model
- Posterior ventral temporal activation best explained by early stages of the semantic network, where shared, general superordinate category information is dominant
- Full activation of detailed semantics of specific target concepts best accounted for activation in bilateral perirhinal cortex
- Anterior-posterior gradient of visual-to-semantic processing in ventral stream

Representational Similarity Analysis (RSA)

- Posterior ventral temporal activation best explained by early stages of the semantic network, where shared, general superordinate category information is dominant
- Full activation of detailed semantics of specific target concepts best accounted for activation in bilateral perirhinal cortex
- Layers of the visual DCNN account for visual regions but are not sufficient to account for anterior temporal activation
- Semantic distinctions are not adequately represented in a DCNN model designed to produce object labels (see also Khalighi-Razavi & Kriegeskorte, 2014)
- Integrating visual and semantic representations can account for fMRI pattern-information throughout the ventral object processing stream

Conclusions

- Posterior ventral temporal activation best explained by early stages of the semantic network, where shared, general superordinate category
  - Information is dominant
- Full activation of detailed semantics of specific target concepts best accounted for activation in bilateral perirhinal cortex
- Layers of the visual DCNN account for visual regions but are not sufficient to account for anterior temporal activation
- Semantic distinctions are not adequately represented in a DCNN model designed to produce object labels (see also Khalighi-Razavi & Kriegeskorte, 2014)
- Integrating visual and semantic representations can account for fMRI pattern-information throughout the ventral object processing stream