Generating the Features for Category Representation using a Deep Convolutional Neural Network
Chen-Ping Yu (cxy7452@gmail.com), Justin Maxfield, and Gregory J. Zelinsky
Dept of Computer Science, Dept of Psychology, Stony Brook University

Introduction

- **Question:** The search for object categories, categorical search, is characterized by a subordinate-level advantage in target guidance (e.g., time to target fixation) and a basic-level advantage in target verification. What are the underlying features that make these effects possible?
- **Exp 1:** Previous work ([16]) modeled categorical guidance and verification using three levels of a hierarchy (subordinate, basic, and superordinate) using category-consistent features built from SIFT and Bag of Words. How would fixation prediction using these biologically-uninspired features compare to those learned by a CNN of the ventral visual stream?
- **Exp 2:** How do fixation predictions using our Ventral-Stream Network (VsNet) model compare to previous work used a Bag-of-Word model against biologically-inspired versions of AlexNet and HMAX ([4] Deep-HMAX: [6]).

Categorical Search & Category-Consistent Features (CCFs)

- **Hierarchical categories:** We collected a 68-category hierarchical dataset with 4, 4, and 4 subordinate, basic, and superordinate categories, respectively.
- **Dataset:** 500 training and 50 validation images per category, for a total of 26,400 images over 48 subordinate categories (basics and superordinates are derived from subordinates).
- **Behavioral Data:** Time to target (guidance) and subsequent verification times were measured using a categorical search task (N=28).
- **Category-Consistent Features (CCFs):** Visual features learned by an unsupervised generative model that represent an object category, defined as common features that occur frequently and reliably.
- **SIFT BoW-CCF:** previous work used a Bag-of-Word model with SIFT features ([16]) as the base feature pool (1094-d) features with high signal-to-noise ratio (SNR) were identified as the CCFs.

This model predicted guidance and verification across hierarchical level, but not at the level of individual categories!

Deep Convolutional Neural Network Models (DCNNs)

- **CNN-CCFs:** CNNs learn hierarchical features directly from images. Can the CCFs extracted from CNNs do better in predicting guidance and verification performance? Yes, at even the level of individual object categories!
- **Biological-Plausibility:** Most CNNs are learned to predict classification accuracy, but they have also been used to predict behavioral and neural data ([17]). We evaluated our VsNet model against biologically-inspired versions of AlexNet and HMAX ([4] Deep-HMAX: [6]).

Experiments and Results

- **1:** Pre-train the networks using ImageNet for classification.
- **2:** Fine-tune the networks with our dataset, using multi-task learning:
  - **CNN-CCFs:** For each object category, extract the neurons that are highly and reliably activated by its exemplars. (Thresholded Signal-to-Noise Rate).
  - **More CCFs = higher specificity = faster time-to-target [16].

Conclusion

- **Brain-inspired networks are better at object category classification.**
- **They also better predict guidance of attention to targets compared to AlexNet.**
- **Our VsNet is the first deep learning model that is brain-inspired. Future DCNNs should continue to exploit neural constraints so as to better predict behavior.

References & Acknowledgments