A mathematical model of real-world object shape predicts human perceptual judgments

Caterina Magri (cmagri@fas.harvard.edu)
Psychology Department, Harvard University, 33 Kirkland street
Cambridge, MA 02138 United States

Andrew Marantan (andrew.marantan@gmail.com)
Physics Department, 17 Oxford Street
Cambridge, MA 02138 United States

L Mahadevan (lmahadev@g.harvard.edu)
Paulson School of Engineering and Applied Sciences, Department of Physics,
Department of Organismic and Evolutionary Biology, Center for Brain Science, 29 Oxford Street
Cambridge, MA 02138 United States

Talia Konkle (tkonkle@fas.harvard.edu)
Psychology Department, Harvard University, 33 Kirkland street
Cambridge, MA 02138 United States

Abstract:

How do we represent the shape of different real-world objects? Modern approaches to explore this question with deep neural networks are highly efficient but as of yet not clearly interpretable. In this paper, we examined the Normalized Contour Curvature model (NCC), which represents a shape as an interpretable histogram over curvature values (from very concave, to straight, to very convex). To explore the shape-space produced by this model, we submitted the feature profile of thousands of objects to a principal component analysis, revealing that 4 axes summarized the space. To compare this model to behavior we tested both perceived curvature (E1) and overall shape similarity (E2). Behavioral judgments of the perceived curvature of an object were well predicted by this model, largely isolated to the second PC loading. Behavioral measures of overall shape similarity were also predicted reasonably well using the four PCs, and approached the performance of deep neural networks (E3). The success of this model implies that perceptual shape-space can be summarized with a relatively small number of dimensions through an interpretable feature space, where the major axes are meaningfully related to perceived shape and curvature of inanimate objects.

Keywords: curvature; shape; object recognition

Introduction

Objects in the world come in all different shapes and sizes, and understanding how our brain represents them is a complex problem (DiCarlo et al., 2012). Modern approaches with deep neural networks (DNNs) have been very efficient in capturing shape and predicting neural responses to real-world objects (Yamins & DiCarlo, 2016). However, the model fits are currently harder to interpret, because of the hundreds of units included within each layer.

Here, we consider a new Normalized Contour Curvature model (NCC), and examine how well this model predicts behavioral judgments of shape and curvature relative to deep neural networks. The NCC model takes as input a picture of a real-world object, then uses pixel intensity to compute a series of contours over a slightly blurred image, then computes the curvature along each point along these contours, and finally aggregates these values into a histogram. The output is a probability distribution of curvature values, from very convex, to straight, to very concave (see Figure 1). An advantage of this model is that the extracted features are easily interpretable. Further, this model is naturally rotation-, translation-, and scale-invariant.

Methods and Results

NCC Shape-Space

To understand the kind of objects that are similar and different under the NCC model, we aimed to characterize the "shape-space" of this model using a

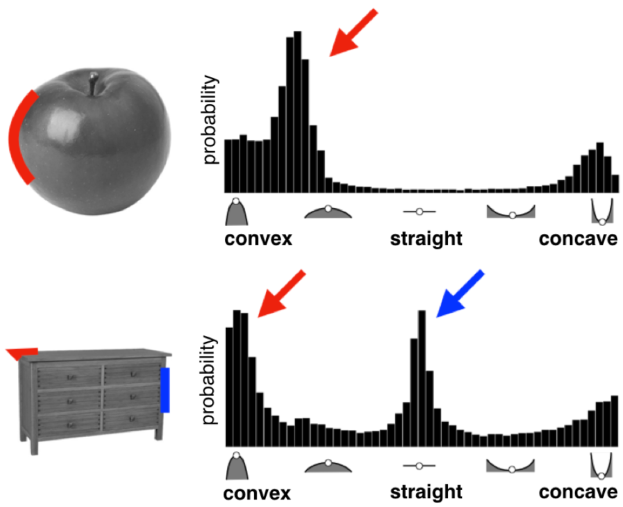

Figure 1: Histograms resulting from the Normalized Contour Curvature model applied to two objects.

wide range of inanimate objects. To do so, NCC features were extracted from pictures of ~7k inanimate objects to the model, and then submitted to a principal component analysis. A parallel analysis (Horn, 1965) revealed that 4 principal components (PCs) explained variance significantly better than a random scrambling of the features (see Figure 2).

The first PC (29% variance explained) put objects with narrow convexities such as dumbbells at the positive pole, and objects with wide convexities such as bowling balls at the negative pole. The second PC (21% of variance explained) pitted objects with convexities and concavities against objects with straight contour lines, similar to intuitive judgments of curvature. The third PC (14% of variance explained) put objects with internal patterns and line junctions at one pole, and objects with elongated shapes at the other pole. For the fourth PC (9% variance explained) a clear qualitative divide wasn't as easily identifiable.

Experiment 1: Curvature Judgments

First, we examined whether this space could predict perceived curvature ratings along a curvy-to-boxy scale, as this perceptual axis has been show to account for a notable amount of variance in neural response patterns along the ventral stream (Long et al., 2017).

The stimulus set consisted of 72 images of inanimate object on a white background. Twenty raters rated each item on a curvy-to-boxy 5-point scale, for a total of 20 ratings per image. Next, we examined how well these behavioral curvature scores could be predicted by a weighted combination of the four shape-space PCs with a simple linear regression (9-fold cross-validation).

Overall, perceptual curvature judgments were predicted remarkably well (Pearsons's r = 0.69, inter-class correlation = 0.53), with strongest weights on the second PC. This result suggests that the shape-space produced by the NCC model naturally captures a perceived curvature axis as the second principle component.

Experiment 2: Overall Shape Similarity

We then explored how the NCC model would perform in predicting more generic shape similarity ratings, to test whether the four-dimensional space could well summarize the overall shape representation in behavioral judgments.

Participants (n = 20) were asked to arrange 72 images of inanimate objects in a circular arena based on shape similarity (Kriegeskorte and Mur, 2012), and to actively avoid organizing them based on other properties (e.g. color, context). As in Experiment 1, we adopted a modeling approach in which we tested how well behavioral ratings were predicted by a weighted

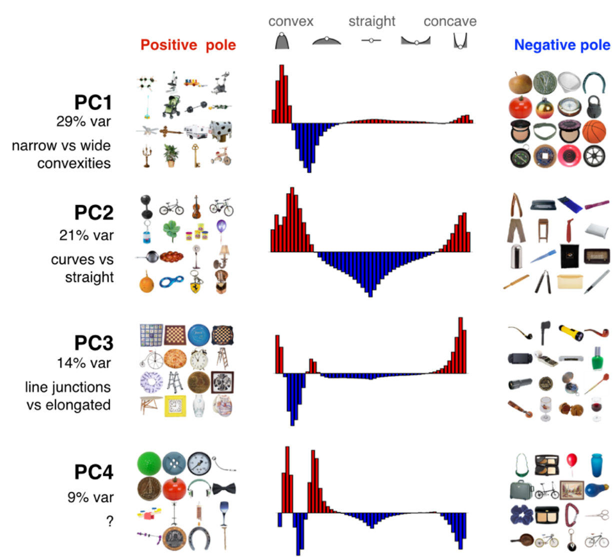

Figure 2: The four PCs resulting from the Principal Component Analysis.

combination of our four PCs (9-fold cross-validation). In this case, however, the behavioral ratings were in the form of a Euclidean distance matrix, with a distance value for each pair of objects. Thus, the PCs were transformed into Euclidean distance matrices, and to avoid negative beta weights we performed a non-negative least square regression (NNLS; Jozwik et al., 2016).

The NCC model predicted overall shape similarity judgments almost at the noise ceiling (Kendall's τ = 0.21, noise ceiling = 0.28-0.34). This result suggests that the model is able to predict part of the human shape representation; however there also seems to be some additional variance that cannot be accounted for by our model. With regard to the average weights assigned to the PCs, we observed an important role of all four PCs.

Experiment 3: DNN Performance in Overall Shape Similarity

To put the NCC model's performance into context, we compared its prediction in the overall shape similarity judgments with a pre-trained deep neural network's performance.

For each layer of Alexnet (5 convolutional and 3 fully connected layers), we measured activations to each of the 72 items, and produced a representational dissimilarity matrix, using a Euclidean distance metric. We then used the eight distance matrices as predictive variables for a model predicting the behavioral overall shape similarity data (9-fold cross-validation; NNLS regression, as in Experiment 2).

The DNN model predicted overall shape similarity judgments better than the NCC model, and within the noise ceiling of the behavioral data (Kendall's τ = 0.28, noise ceiling = 0.28-0.34). When looking at the layers that weighted most in this performance, conv3, fc6 and fc8 showed the strongest weight. Thus, we confirm that deep nets are best at maximizes predictive power, but what information might be contained in these layers is not easily interpretable.

Conclusions

The way in which objects' shape is represented in humans is still the subject of much exploration. In the past few years, modeling approaches with DNNs have produced exciting results; however, these approaches are still limited by our ability to interpret the hundreds of units within each layer. Simpler, quantitative models of shape might serve a complementary role to help us better understand and interpret these representations. The Normalized Contour Curvature model examined in the current work reveals that it is possible for the shape-space of objects to be parameterized by a relatively small number of dimensions.

References

DiCarlo, J.J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, 73(3), 415-434.

Horn, J.L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrica, 30*(2), 179–185.

Jozwik, K. M., Kriegeskorte, N., & Mur, M. (2016). Visual features as stepping stones toward semantics: Explaining object similarity in IT and perception with non-negative least squares. *Neuropsychologia*, *83*, 201-226.

Kriegeskorte, N., & Mur, M. (2012). Inverse MDS: Inferring dissimilarity structure from multiple item arrangements. *Frontiers in psychology*, *3*, 245.

Long, B., Yu, C. P., & Konkle, T. (2017). A mid-level organization of the ventral stream. *bioRxiv*, 213934.

Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience*, *19*(3), 356.