General Shape Features Allow for Categorization of Written Symbols Across Font Variation

Daniel Janini (daniel_janini@g.harvard.edu)
Department of Psychology, Harvard University, 33 Kirkland St.
Cambridge, MA 02138 USA

Talia Konkle (tkonkle@fas.harvard.edu)
Department of Psychology, Harvard University, 33 Kirkland St.
Cambridge, MA 02138 USA

Abstract:

With extensive experience, humans become experts at recognizing and reading letters and digits. Does the ability to categorize these symbols require a specialized visual feature space, or can this capacity be supported to some extent by a more general feature space also used to represent other visual categories like objects? To examine this question, we tested whether multiple models of general shape features could categorize written symbols across large variations in font. Moderate to robust categorization accuracy was accomplished using deep convolutional neural networks trained to do object categorization, as well as in simpler models like Gist and Normalized Contour Curvature. These models also showed moderate correlations to human classification behavior. Broadly, these results are in line with the possibility that the visual system processes written symbols by leveraging features in place for recognizing real-world objects, rather than primarily relying on symbol-specific feature tuning.

Keywords: CNN; AlexNet; Gist; Curvature; Letters; Digits; Categorization

Introduction

Ventral visual cortex contains a number of brain regions whose responses are highly selective to specific categories: faces, bodies, scenes, and letter-strings (Op de Beeck, Haushofer, & Kanwisher, 2008). Letter strings are interesting relative to the other categories because the invention of writing systems is too recent to have had an evolutionary impact on innate brain organization. Thus, neural regions that process these symbols may be "recycled" from evolutionarily older neural maps (Dehaene & Cohen, 2007).

Within a cortical recycling framework, however, there is a range of possibilities for the manner of recycling. On one hand, extensive plasticity in the feature tuning may result in a highly-specialized feature space. Alternatively, this region might represent letters by primarily leveraging existing shape features, perhaps with only a small degree of fine-tuning for letter shapes. Here, we explore the viability of the latter hypothesis. Specifically, we examined whether a variety of general shape feature models can categorize written symbols and tested the extent to which they predict human categorization behavior.

Methods

Three image sets were created consisting of different typeset symbols: (i) digits 0-9, (ii) all 26 lower-case letters in the Roman Alphabet, and (iii) all 26 upper-case letters in the Roman Alphabet. Each alphanumeric symbol was formatted in 180 fonts.

Next, the representation of each image set was computed from 10 different visual features models: Pixel values, Gist, Normalized Contour Curvature (NCC), and each of the seven layers of a deep convolutional neural network trained to do object categorization (AlexNet). The pixel model simply considered each pixel as an independent feature dimension. The gist model was originally developed to quantify the spatial layout of scene images (Oliva & Torralba, 2006), but here it served as a model of the global structure of each image. Normalized contour curvature is a measure of the probability distribution of concave and convex contours in an object and is naturally rotation- and translation-invariant representation (Mahadevan & Marantan, *in preparation*). Finally, we considered each layer of pre-trained AlexNet as a separate feature model. For convolutional layers, activations were summed across space for each feature channel.

Results

First, we assessed how well each feature representation could distinguish symbols over variations in font. The results are shown in Figure 1. All models performed above chance, with the lowest performance by the pixel model, moderate performance by the NCC model and earliest layer of AlexNet, and near ceiling performance for the Gist features and all subsequent layers of AlexNet.

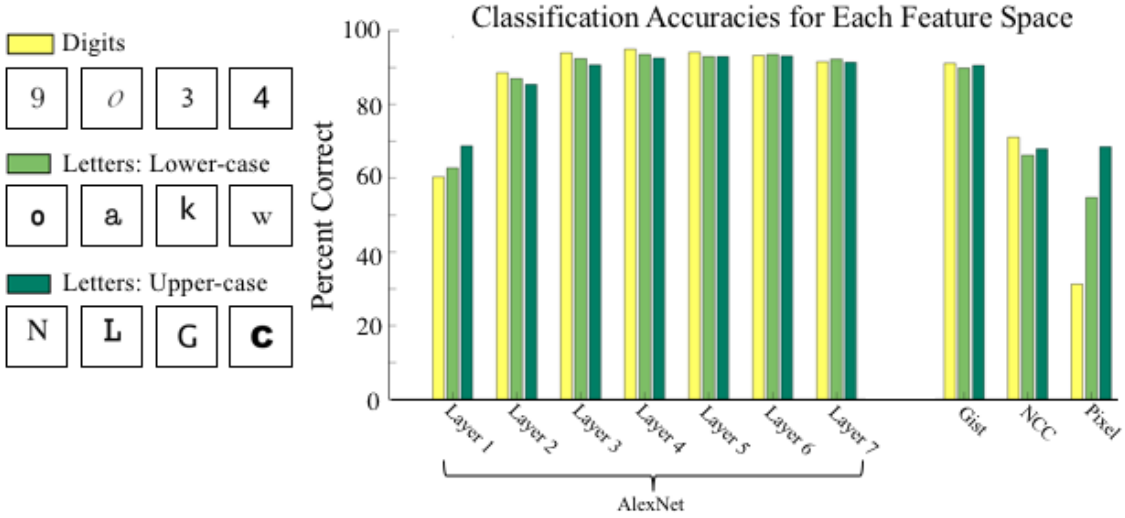

Figure 1. Classification accuracies for each feature space and for each image set.

This initial result demonstrates that multiple models of general shape features are sufficient for the categorization of typeset symbols. However, this result does not indicate whether these models are succeeding at the symbol classification task by different means. To assess this, we compared the representational geometries of the ten models through the construction of representational dissimilarity matrices (RDMs).

Within each model's feature space, correlation distances were calculated for each pair of images. Then the average distances were calculated for each symbol category to construct a 10x10 RDM for the digit symbols, and separate 26x26 RDMs for the lower-case and upper-case letters. We correlated RDMs between models to determine which models had similar representational spaces.

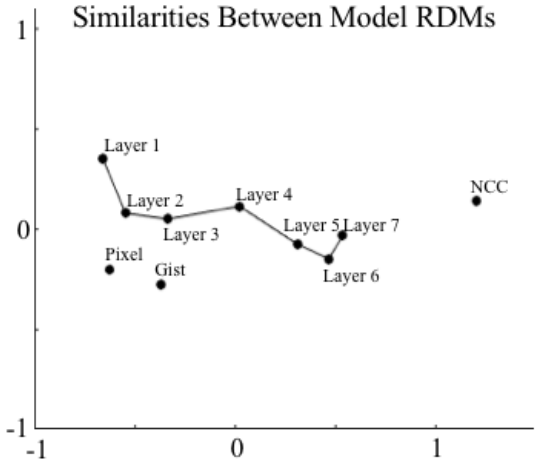

Figure 2. Similarities between model RDMs for lower-case letters visualized using multidimensional scaling

The similarity between model geometries for lower-case letters is visualized using multidimensional scaling in Figure 2. Each layer of AlexNet has a slightly different representational geometry of letters, with the Pixel and Gist models most similar to early layers, and the NCC model more similar to the later layers. These same relationships were found for digits and upper-case letters.

Next, we determined which models could predict human categorization of lower-case letters. We obtained discrimination times from Courrieu, Farioli, & Grainger (2004), in which participants judged whether pairs of letters were the same or different as quickly as possible. Each model's RDM for lower-case letters was correlated to the behavioral RDM constructed from the reaction times (Figure 3).

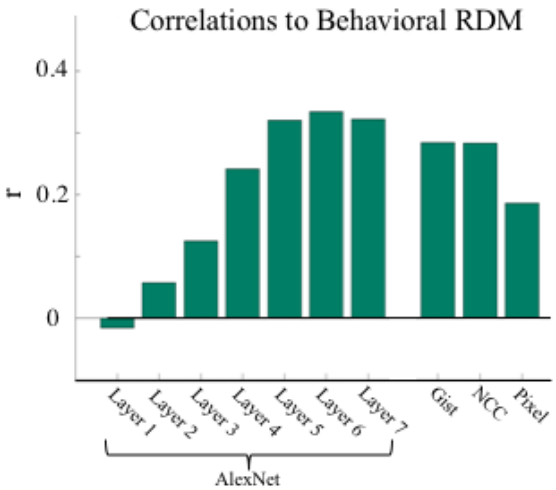

Figure 3. Correlations between each model's RDM for lower-case letters and human classification times (Courrieu, Farioli, & Grainger, 2004)

While early layers of AlexNet allowed for classification of letters, only later layers predicted behavior. The representational geometries of Gist and NCC also made moderate predictions of human behavior. These results indicate that the representational space underlying behavioral letter discrimination can be moderately predicted by feature spaces that were not directly tuned to represent and distinguish letters.

However, it is worth noting a few caveats to the behavior-to-model relationship. First, none of these models made excellent predictions of human behavior, though it is difficult to determine what would constitute a high correlation as we were not able to calculate the noise ceiling from this pre-existing dataset. Second, the behavior was only performed for lower-case letters in one font. It will be necessary to obtain behavioral data using more varied stimuli to further assess how well these general shape models can match behavior.

Conclusions

Classification accuracies indicated that multiple general shape spaces are sufficient for fairly accurate categorization of digits and letters across variation in font. Moreover, the comparisons between model RDMs and behavior indicate that later layers of AlexNet, Gist, and NCC could predict behavioral letter similarity. Thus, these results indicate that it is not necessary to employ a highly specialized feature space to categorize written symbols. Broadly, these results have implications for the nature of feature tuning in letter-specific regions, which may largely reflect pre-existing general shape spaces rather than novel features uniquely tuned for categorizing alphanumeric symbols.

Acknowledgments

Funding Sources: Star Family Challenge Grant to TK. National Defense Science and Engineering Graduate Fellowship to DJ.

References

Courrieu, P., Farioli, F., Grainger, J. (2004). Inverse discrimination time as a perceptual distance for alphabetic characters. *Visual Cognition, 11(7),* 901-919.

Dehaene, S., Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron, 56,* 384-398.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: the role of global image features in recognition. *Progress in Brain Research, 155*, 23-36.

Op de Beeck, H.P., Haushofer, J., Kanwisher, N. (2008). Interpreting fMRI data: maps, modules and dimensions. *Nature Reviews Neuroscience, 9*, 123–135