Comparing the brain's representation of shape to that of a deep convolutional neural network

Dean A. Pospisil
Dept. of Biological Structure
University of Washington
Seattle, WA 98195
+1 206 221 2017
deanp3@uw.edu

Anitha Pasupathy
Dept. of Biological Structure
University of Washington
Seattle, WA 98195
+1 206 235 0678
pasupat@uw.edu

Wyeth Bair
Dept. of Biological Structure
University of Washington
Seattle, WA 98195
+1 206 221 8241
wyeth0@uw.edu

ABSTRACT

Hierarchical neural nets are currently the highest performing general purpose image recognition computer algorithms. Their design is loosely inspired by the neural architecture of the ventral visual pathway in the primate brain, which is believed to underlie the perception of form and the ability to recognize objects. The exact tuning of artificial neural units within an HNN, however, is not prescribed from known biology, but is trained using a performance-based learning algorithm. In evaluating whether HNNs are ripe for further bio-inspired performance improvements, it is of interest to test whether the response properties in the intermediate layers of the HNN approximate those of the ventral visual stream. We therefore characterized units within a popular HNN with a set of visual stimuli that has been employed by neurophysiologists to successfully characterize the shape-tuning properties of neurons in the intermediate visual cortical area V4 of the ventral stream. We found that the tuning and fits of a small but significant number of units in the HNN were strikingly similar to those of some V4 neurons for our simple set of test shapes. There tended to be more such units in the deeper layers of the HNN. We discuss implications of our results to the encoding of curvature in the primate brain and propose ways to further characterize V4-like shape tuning in HNNs.

Categories and Subject Descriptors

I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis – *object recognition, shape.*
I.2.6 [**Artificial Intelligence**]: Learning – *connectionism and neural nets.*
I.2.10 [**Artificial Intelligence**]: Vision and scene understanding – *shape.*
I.5.1 [**Pattern Recognition**]: Models – *neural nets.*
I.5.4 [**Pattern Recognition**]: Applications – *computer vision.*
C.1.3 [**Processor architectures**]: *Other architecture styles – neural nets.*

General Terms

Algorithms, Performance, Design, Experimentation, Theory.

Keywords

Deep networks, convolutional networks, hierarchical neural networks, visual cortex, shape representation, object recognition, curvature.

1. INTRODUCTION

The recognition of visual objects remains one of the most difficult and potentially rewarding challenges faced by computer vision. Human performance on this task has yet to be matched by any computer algorithm. It is perhaps unsurprising then that the best performing artificial vision algorithms [12], hierarchical neural nets (HNN), are built upon an architecture that is inspired by the hierarchical structure of the form processing pathway of the primate cerebral cortex [7, 10, 17].

While the overall architecture of an HNN is explicitly bio-inspired, the response properties, or "tuning," of single units within the network are not constrained to match neurobiology. Rather, single-unit properties are determined by a performance-based learning algorithm that operates iteratively across many pre-classified training images, tuning the parameters of the network to decrease the error between the network output and the target classification. In short, the net is tuned to approximate an idealized function that takes an image of an object and returns its label.

The input layers of these HNNs will often have response properties qualitatively similar to those of primary visual cortex (V1), i.e. selectivity for orientation and frequency [12] (see below in Figure 4). Some dissimilarities include a much larger number of subunits than the usual ~2-3 seen in V1 receptive fields (RFs), and deviations from the ratio of RF width perpendicular to orientation preference to its length ~1.5:1 in V1. This raises the tentative but exciting possibility that HNNs are approximating some aspects of the nervous system's own solution to the problem of object recognition.

In advancing this idea it is then of interest to determine whether HNNs also share higher level response properties with the ventral visual stream. This line of inquiry has been pursued by Yamins et al. [25] who found that responses in multi-unit V4 recordings to naturalistic images could be predicted by linear combinations of responses of units within a performance-optimized HNN. Here we take a more direct approach by comparing the physiological properties of single artificial units in the network to the well-characterized ability of V4 neurons to encode boundary shape [17, 18, 19, 20], and we attempt to estimate the fraction of neurons in the HNN that might be considered by a neurophysiologist to be V4-like. This is important for two reasons: i) if performance-

BICT 2015, December 03-05, New York City, United States

DOI 10.4108/eai.3-12-2015.2262486

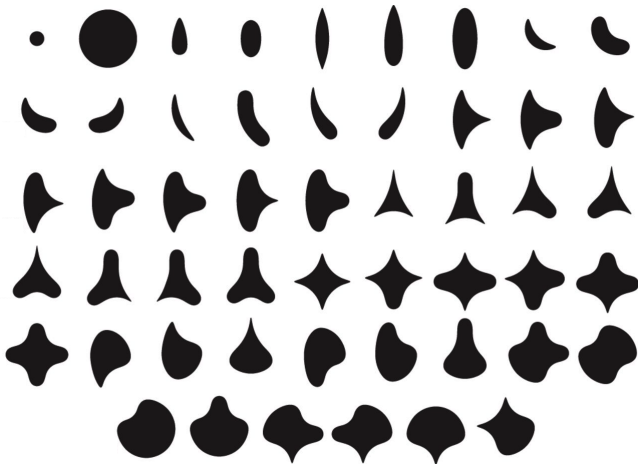

Figure 1. This set of 51 shapes was developed by Pasupathy and Connor to study V4 neurons and is used here to study units in an artificial neural network.

optimized systems arrive at similar tuning properties as V4 neurons, this would reinforce the belief that V4 tuning properties are optimized for recognition; ii) analyzing the connectivity patterns of any V4-like HNN units might provide insights into how V4 RF's are built.

HNNs are notoriously difficult to interpret [6, 13, 16, 26], much like the nervous system, because of their heterogenous properties and compounded nonlinearities. Part of our motivation is to test the feasibility of applying an experimental neuroscience approach to interpret an object-recognition HNN. One widely-used and successful neurophysiological approach is to characterize heterogeneous response properties with respect to well defined and intuitively motivated response subspaces [22]. This approach necessarily captures less variance, but by design gains greater insight into the variance that it has selected to explain. Even if such an approach fails, it could hold important lessons for neurophysiologists regarding the limitations of their own traditional methods in the face of highly intricate, distributed systems.

In the primate, the ventral visual pathway is important for the processing of visual form and color information [7, 23]. Area V4 is an intermediate stage along this pathway: it receives feedforward input from cortical areas V1 and V2 and sends outputs to, among other areas, the inferotemporal (IT) cortex, the last stage of processing along this pathway. In addition to selectivity for local orientation and spatial frequency [5] as observed in V1, V4 neurons are selective for more complex stimulus features [8, 10, 18]. Specifically, responses of many neurons in area V4 are dictated by the boundary curvature of the shape at a specific location relative to object center [19]. For example, some neurons may respond preferentially to shapes with a sharp convex projection to the top, others to a concavity to the left, etc. Together these neurons can provide a complete and accurate representation of isolated shapes in terms of their boundary characteristics [19, 20]. Curvature-tuned V4 neurons are also more sensitive to visual form under partial occlusion conditions and may be better suited to contribute to their discrimination [11]. Consistent with these findings, shape theory and computer vision also suggest that boundary curvature is a useful parameter for robust recognition of occluded objects [1, 2, 3, 14, 24]. We therefore decided to ask whether there are HNN units that are also tuned to boundary curvature.

To address the above question, we adopted the approach of Pasupathy and Connor [19] who used a set of 51 systematically-designed simple closed shapes (Figure 1) presented at different rotations to study the responses of isolated V4 neurons. Stimulus size, position, luminance and color were customized for each neuron based on preliminary tests to maximize the signal-to-noise ratio for boundary curvature sensitivities. Their results from an example neuron are shown in Figure 2A. A variety of shapes evoked strong responses (see scale bar) from this neuron; all of these shapes shared a concavity to the right. To quantify this preference, Pasupathy and Connor parameterized each shape in terms of its boundary curvature relative to object-centered angular position (Figure 3, see Section 2.3) and used a 2D Gaussian function in angular position × curvature (APC) space to describe the selectivity. Responses were well fit by a 2D Gaussian model

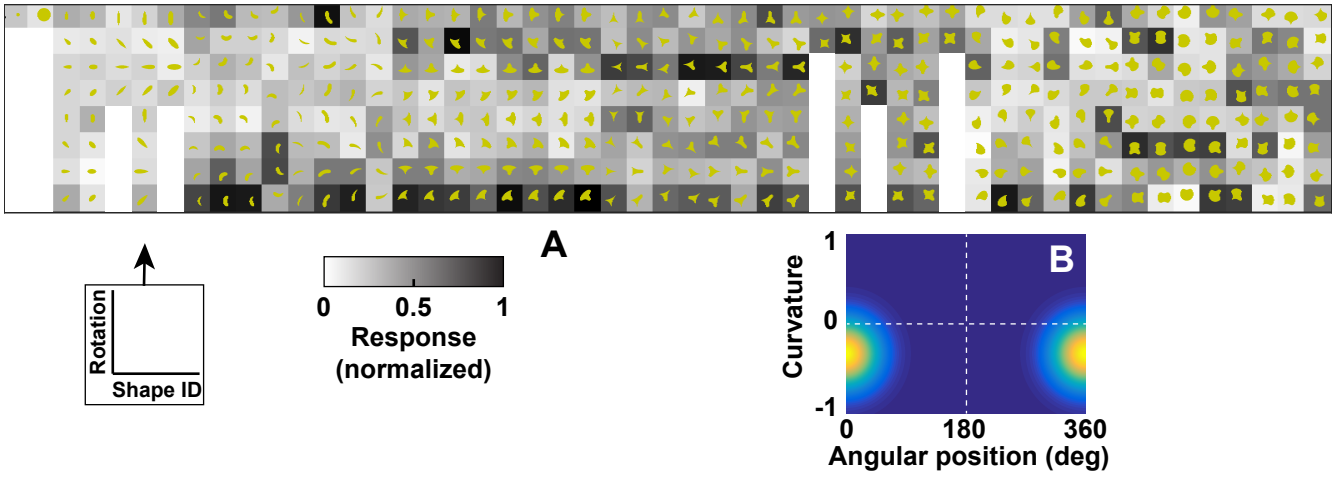

Figure 2. Example V4 neuron tuned for boundary curvature and angular position. (A) The mean firing rate (normalized to 1) in response to each shape stimulus is shown by the grayscale of the background (black indicates largest response) on which each shape is depicted. Not all shapes have 8 distinct rotations because of rotational symmetry. Shape ID changes along the horizontal axis, shape rotation changes along the vertical axis. (B) The two-dimensional Gaussian fit of the angular-position curvature (APC) model to the responses shown in A. The Gaussian is narrow in both the angular position and curvature axis, indicating that the neuron prefers concave features (negative curvature values) to the right (0 deg). The correlation value for this fit is r = 0.80.

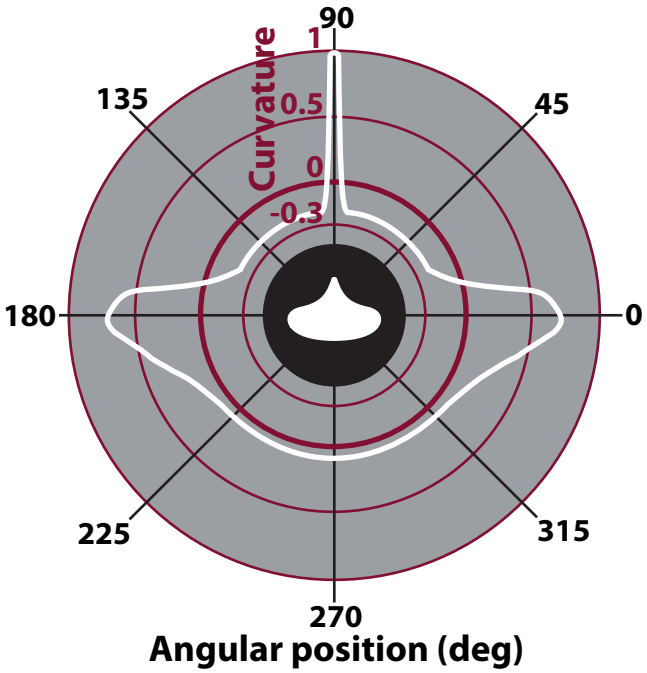

Figure 3. Boundary curvature vs. angular position. The white line plots the curvature as a function of angular position for the boundary of a simpe shape (white shape at center). At 90 deg (vertical), the sharp point on the shape has a large positive curvature value, whereas the concavities at 45 and 135 deg have negative curvatures.

with a peak close to 0 deg and at curvature close to -0.5 (Figure 2B).

Out of the 109 V4 neurons in the original dataset [19], responses of 49 neurons were well predicted by the best-fitting APC model; i.e. goodness of fit given by the correlation, r, between the observed and predicted responses was > 0.5. To determine whether units in an HNN were V4-like, we quantified the responses of each HNN unit to the shape stimuli used by Pasupathy and Connor. We then identified the best-fitting APC model for each unit and quantified the goodness of fit. As with the electrophysiology, if the goodness of fit for a unit was > 0.5, and the fitting parameters for the best model fell within a range of typical parameters seen for V4 neurons, then the HNN unit in question could reasonably be deemed V4-like in its responses.

2. METHODS

2.1 Architecture of the neural network

In this study, we investigate units within the artificial neural network known as CaffeNet, which is a replication of AlexNet [12] implemented in Caffe (http://caffe.berkeleyvision.org), an open source package for the training and use of convolutional neural nets (CNNs) on GPUs. AlexNet is an image recognition HNN that won the ImageNet LSVRC-2010 competition. Details of its training and construction can be found in the original paper [12]. In brief, AlexNet has 8 layers that include 5 convolutional layers followed by 3 fully connected layers. The first layer consists of 11 x 11 pixel linear filters (Figures 4 and 5), and the last layer is a categorical probability distribution, where the output of each unit corresponds to the likelihood that the image contains a particular object, e.g. unit 23 in layer 8 represents an eagle. The final layer is decidedly not an intermediate image representation, thus we do not present results here on the final layer. The weights

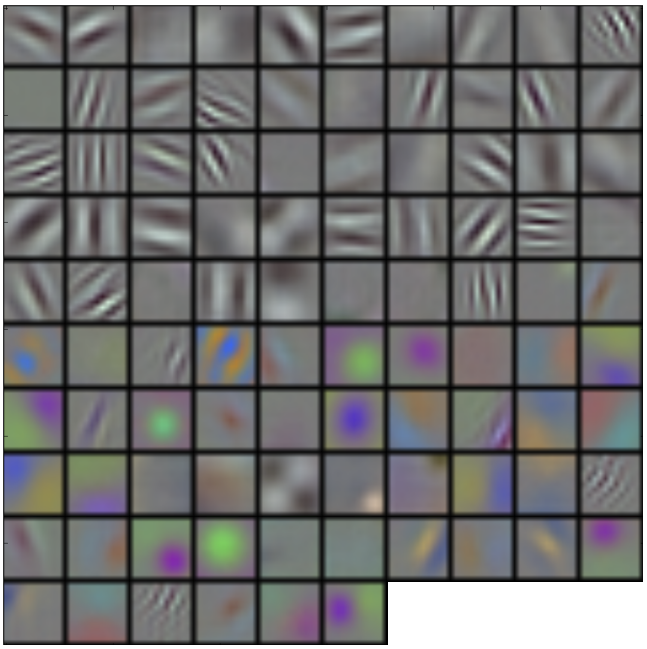

Figure 4. The 96 kernels (11 x 11 pix, by 3 colors) of the 1st layer of the AlexNet model tested here. Similar to V1 receptive fields, these kernels are band-limited in spatial frequency and orientation.

between units were determined by stochastic gradient descent over the labeled images in ILSVCR '12 (1.2 million images, with a thousand labeled categories).

In total there are 622,312 units in AlexNet, but this large number reflects the fact that, within each convolutional layer, each kernel is replicated many times to cover the two-dimensions of the image plane. We therefore examined only the subset of units with unique tuning characteristics in the convolutional layers, ignoring the spatial translations because the kernels, and thus functions, are identical. In particular, we chose only one column of kernels running through the center of each convolutional layer. The geometries (fan-in and maximum RF size) for the convolutional layers are shown in Figure 5. Thus, for the 5 convolutional layers we characterized respectively 96, 256, 384, 384, 256 units. We also analyzed all units in the fully connected layers 6 and 7; in total we examined 9,568 units.

2.2 Visual stimulus set

To compare the responses of HNN units to V4 neurons, we used the stimuli developed by Pasupathy and Connor [19]. Briefly, they used 51 shape stimuli (Figure 1) that were parametrically constructed by combining 4-8 concave or convex contour segments (e.g., Figure 3). Each shape in Figure 1 was rotated in 45 deg increments to create rotational variation; discarding duplications because of rotational symmetry resulted in a stimulus set that included 370 unique shapes. To match the format of the training data of AlexNet, centered stimuli were converted to 224 x 224 jpegs, then had the mean of the 2012 ImageNet (www.image-net.org) data set subtracted from them, before being passed through AlexNet. Lower level units in AlexNet are influenced by smaller swaths of the original input image (Figure 5, bottom), analogous to receptive field sizes decreasing at earlier regions of the ventral stream. AlexNet units were tested with 4 sets of shape stimuli that were sized so that the diameter of the large circle (stimulus 2 in Figure 1) fit within the receptive fields of the 2nd,

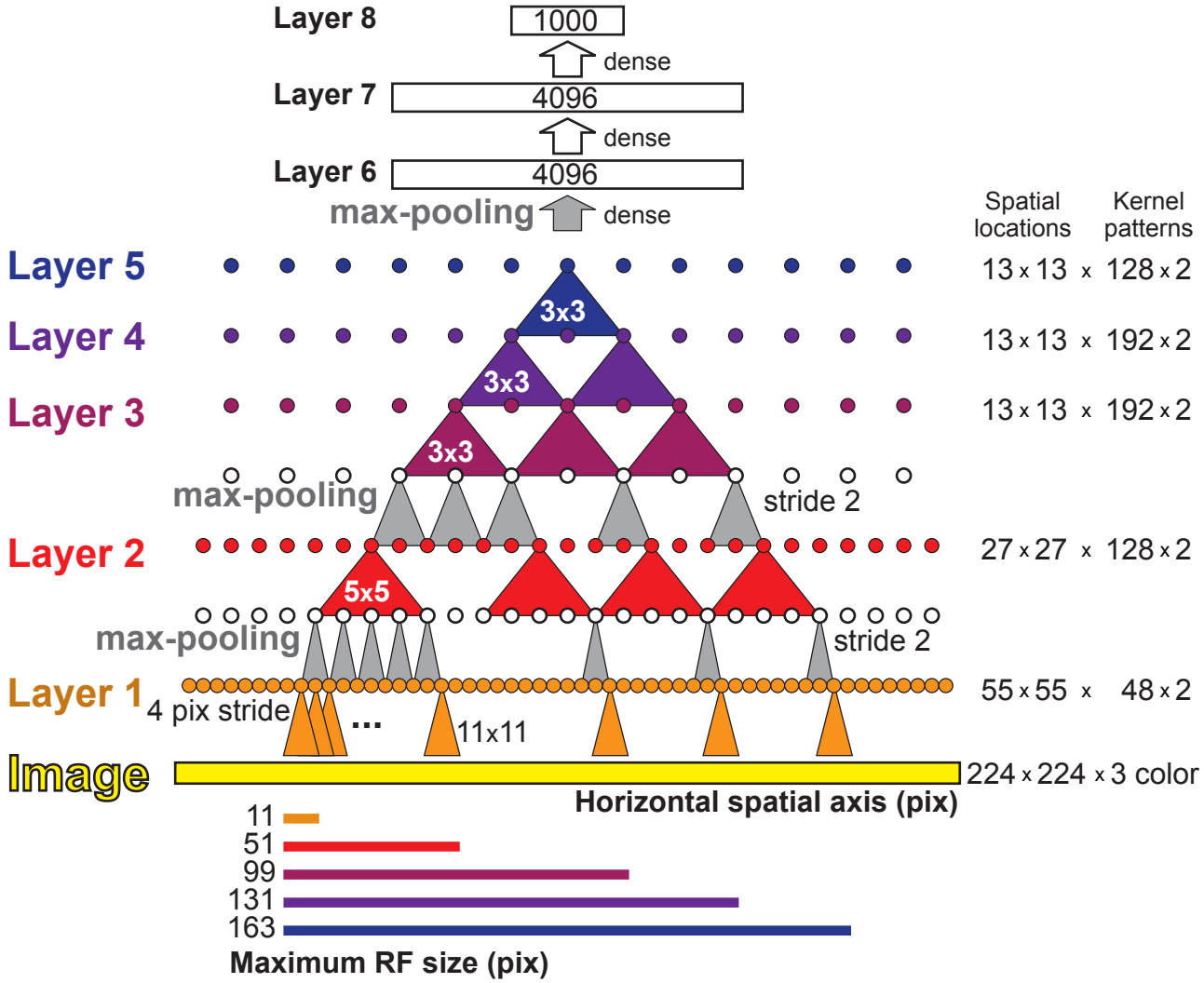

Figure 5. Fan-in and maximum RF sizes for the five convolutional layers in Alex Net. Only one spatial dimension is shown. The image (yellow) is 224 x 224 x 3 pix, where 3 provides for RGB color values. The first convolutional layer, Layer 1 (orange), has kernels that are 11 x 11 pix and occur every 4 pixels (stride of 4). There are 55 x 55 spatial locations and at each location there are 96 kernels, which are divided into two sets of 48. Outputs of Layer 1 are pooled (average of 3, stride of 2, gray triangles), and then Layer 2 (red) operates on the pooled values with kernels of 5 x 5 pix. Layer 2 has 27 x 27 spatial locations and 256 kernels at each location, which are divided into two groups of 128. Layer 2 outputs are pooled (gray triangles) and Layer 3 (red-violet) integrates over 3 x 3 kernels in a 13 x 13 grid, with 384 kernels at each location. Layer 4 (purple) and Layer 5 (blue) also use 3 x 3 kernels on a 13 x 13 grid. Layer 5 has only 256 kernels at each location. Kernels were divided into two groups in the original model to facilitate parallel computation. The lines at the bottom show the maximal RF size at each layer, given the kernel sizes and pooling steps. For clarity, the tree beneath one central unit in Layer 5 is show, and not all kernels are shown in each layer. In addition to the 5 convolutional layers, three top layers are shown that are densely connected to their respective inputs. Thus, RFs in these layers can span the entire input image.

3rd, 4th, and 5th layer. Responses from units were stored for each presented image.

2.3 Angular position x curvature (APC) model

To quantify shape responses in terms of boundary curvature, Pasupathy and Connor represented each stimulus in terms of 4-8 points in a two-dimensional angular position x contour curvature (APC) plane. Unit responses were fit with the product of twoGaussians defined over the 2 axes of the APC plane (Figure 2B). More generally each stimulus is represented by P points in an n-dimensional stimulus space. X_{ip} represents the value of the i^{th} stimulus dimension for the p^{th} point. The response function along each dimension is fit by a one dimensional Gaussian with peak at μ_i and standard deviation ς_i, the overall response is fit by the product of the n Gaussians. The predicted response r is given by:

$$r = \max_p \left[k \prod_{i=1}^{n} e^{-(X_{ip}-\mu_i)^2/2\varsigma_i^2} \right]$$

where k represents the amplitude of the product of Gaussians. The predicted response is the maximum of the calculated response to the P stimulus points. In this case, n = 2, and the dimensions are angular position and curvature.

In the original Pasupathy and Connor paper, a gradient descent method, the Gauss-Newton algorithm with a grid of starting points across the APC plane was used to fit responses. The Pearson correlation coefficient was used as the measure of goodness of fit. In our case, given the large number of units requiring fitting, we instead used a rapid brute force fitting method with 16 potential values for each of the four free parameters being fit (mean and standard deviation for the curvature and angular position dimensions). This resulted in a total of 65,536 models tested for each unit's responses. Means of the APC models tested were linearly spaced within their respective ranges: [0,360] for angular position and [-1, 1] for curvature. Standard deviations were log spaced, and their endpoints were the minimum and maximum standard deviation values seen for gradient descent fits for the V4 neurons that had $r > 0.5$. We found that the brute force fits were generally very similar in goodness of fit to those from the gradient descent method, with a slight tendency for the brute force method to underperform.

To allow for direct comparison, V4 neurons from Pasupathy and Connor and units from AlexNet were fit using the same set of models with the brute force methods described above. The median r-value of the 109 V4 neuron fit by gradient descent was 0.48. We set an r-value of 0.5 as an initial threshold for a fit to be considered V4-like.

3. RESULTS

We measured the responses of units in the first seven layers of AlexNet to our shape stimuli. Our first observation was that many units across layers did not respond at all to any of the shape stimuli, and many others responded to only one or a few of the 370 stimuli. This is unlike V4 cells, which generally show greater dispersion in their responses across these stimuli [4]. Such sparse responding AlexNet units could be mistakenly classified as being V4-like because they can be over fit by the APC model and achieve a high r-value. However, such cases are typified by having very low standard deviations on the fitted parameters. Thus, to be considered V4-like, units had to have $r > 0.5$ and have model fits with standard deviations within the range observed for ~90% of the well-fit V4 neurons (49/109). Specifically, the SD of both the angular position and curvature parameters had to be greater than the lowest 5% of SDs for V4 units and less than the highest 5% of V4 SDs, thus eliminating atypical outlier fits. This both reduced the number of over-fit units, and served as a more conservative threshold for judging a unit as V4-like.

The fitting parameters of model units and macaque neurons that passed these requirements are plotted in Figure 6. There is clear overlap in the parameters of V4 neurons (Figure 6E-H) and AlexNet units (Figure 6A-D), but there are also clear differences. In the V4 data, there is an over-representation of sharp convexities indicated by a peak near 1 on the vertical axis of Figure 6G, whereas there is only a weak sign of such a peak for the AlexNet units (Figure 6C). Also, units in AlexNet show a strong bias to prefer features that occur at an angular position of 90 or 270 degrees (Figure 6A), but no such trend occurred for the V4 neurons (Figure 6E). This is not only true for the subset of V4 neurons examined here, but also for all 109 neurons in the original study [19]. Possible explanations for this difference are considered in the Discussion.

We next determined the percentage of V4-like units within each layer. In doing this, it is important to test various stimulus sizes to facilitate the comparison to V4 neurons. Under physiological conditions, stimuli were sized to fit within the V4 RF. If stimuli were too large, then none of the boundary would fall within the RF of the unit. In the cortex, there is often no clearly defined end

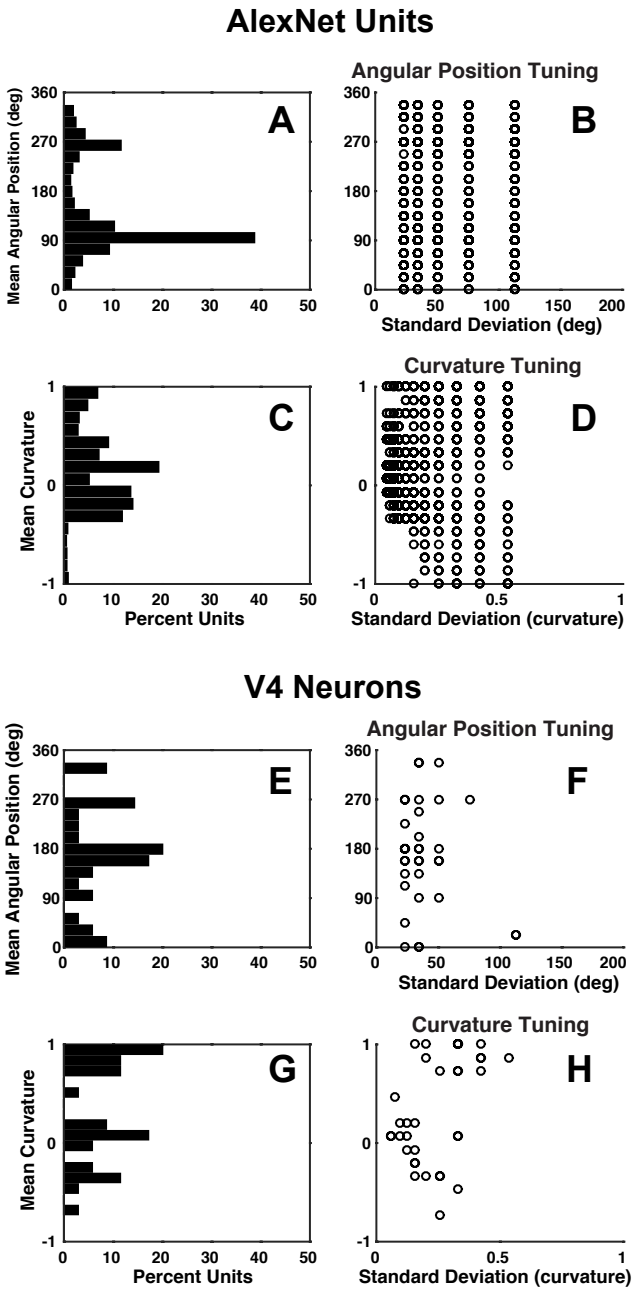

Figure 6. Distributions of parameters for fits to APC model. (A) The distribution of the mean of the Gaussian along the angular position axis for APC fits to all units in all layers. There is a strong tendency for the Gaussian to be centered around 90 deg or 270 deg. (B) Each point shows the mean vs. SD for a Gaussian fit for angular position for a single unit. Points fall on a regular grid because of the limited set of values tested (see Methods). (C) The distribution of the mean of the Gaussian in the curvature axis is plotted. (D) The mean vs. SD of the Gaussian fit on the curvature axis. (E-H) are similar to (A-D), except these parameters are for the fits to the macaque V4 neurons in the Pasupathy and Connor dataset that had r>0.5 and were within the 90% range (see Methods).

to the RF, which may have weaker and modulatory influences from the surround, but in AlexNet the feedforward architecture defines a hard limit to the RF. Beyond their hard RF limits, AlexNet units could not possibly encode information about the

boundary curvature. Figure 7 summarizes the percent of V4-like units found in each layer for four stimulus sizes. For stimuli of the largest size (163 pix diameter, yellow bars), very few V4-like neurons were encountered in the early layers, where RFs were small. For stimuli of the smallest size (51 pix, dark blue bars) there were about 10-20% of V4-like units across all layers. Thus, there was a general trend for units in higher layers to be V4-like for all stimulus sizes that we tested, while units in lower layers required smaller stimuli to display V4-like tuning. Interestingly, there appears to be an increase in V4-like units going from the convolutional layers to the fully-connected layers (5 to 6). Overall, the first fully-connected layer, Layer 6, had the highest mean concentration of V4-like units across stimulus sizes (14.59%, ~497 units).

Figure 8A shows the responses of an example of one of the best fit AlexNet units (from layer 7, unit 2689 r = 0.81) using the same plot format as for the example V4 neuron in Figure 2. The APC fit for this unit revealed broad tuning for moderate concavities pointed upwards (mean angular position = 90 deg, SD 75 deg, mean curvature = -0.33, SD = 0.33).

Given the large number of units that we are testing within AlexNext, it is reasonable to assume that V4-like units could occur by chance. To quantify what fraction of units might be classified as V4-like simply by chance, we shuffled the responses across shapes for each unit and then re-fit all models across all units. None of the units for any stimulus sizes with shuffled responses met our criteria to be considered V4-like.

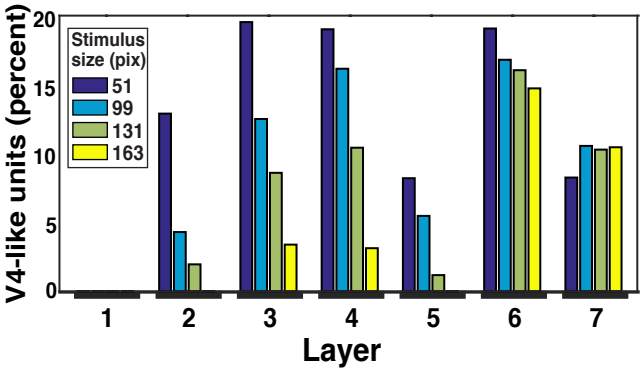

Figure 7. Distribution of V4-like units across layers. For each layer, the percent of units that met our criteria for V4-like behavior (defined in terms of the fit parameters and goodness of fit to the APC model - see Methods) is plotted for four different stimulus sizes (inset). A clear influence of stimulus size on the number of V4-like units is apparent. When the responses were scrambled for each unit across visual stimuli, no V4-like units were found. The total number of units that we characterized in each layer, 1 to 7, was: 96, 256, 384, 384, 256, 2048 and 2048, respectively.

4. DISCUSSION

We studied the responses of units in AlexNet using methods similar to those used by neurophysiologists to study neurons in mid-level visual areas of the ventral visual pathway of the macaque. While it was already known that the units in the first layer of AlexNet have properties that resemble the spatial frequency and orientation selectivity of V1 neurons, we have found the first evidence that a significant fraction of AlexNet units are tuned for the boundary curvature of simple shapes. This raises the possibility that units at mid-level representations within networks trained to classify images share computational similarities to V4 neurons in the primate ventral pathway.

Our findings support the hypothesis that both biological neurons and artificial units in a HNN approximate a similar representation of the boundaries of objects. It should be noted that AlexNet was explicitly designed and trained solely for object recognition performance not physiological plausibility. Such different instantiations exhibiting the same properties can be taken as evidence to support the notion that there are natural solutions to the problem of object identification and the commonalities between these networks reflect them. Because some V4 units and HNNs share tuning similarity, and HNNs were optimized for recognition, our findings reinforce the belief that the V4 tuning properties are optimized for recognition. Our study has not provided direct evidence that the units responding to curvature are in fact relevant to object recognition in AlexNet, but a prior study

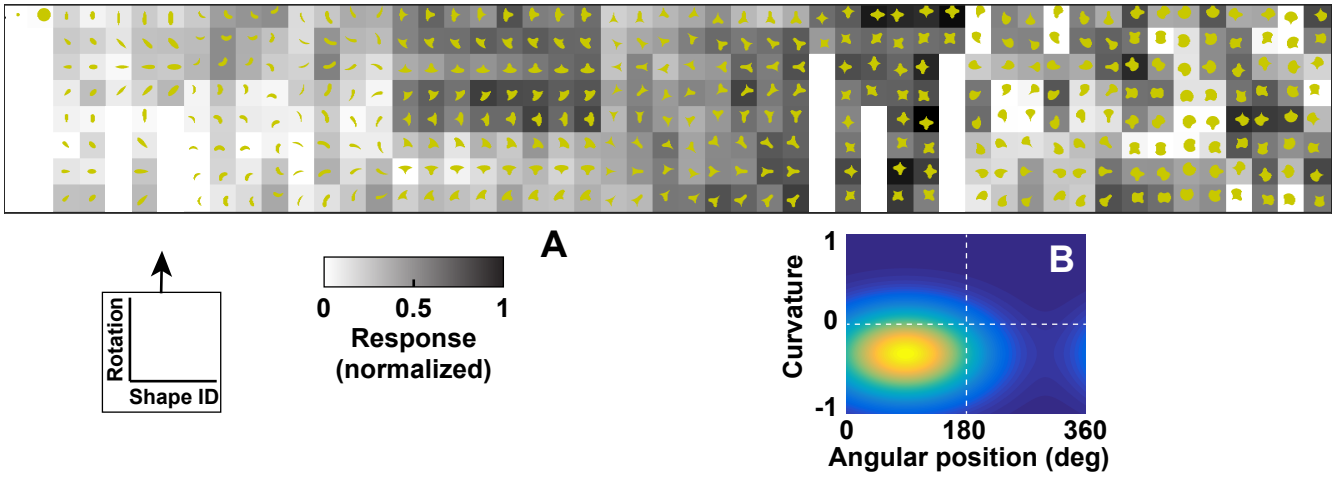

Figure 8. Example AlexNet unit tuned for boundary curvature and angular position. (A) The unit's response (maximum normalized to 1) to each shape stimulus is shown by the grayscale of the background (black indicates largest response) on which each shape is depicted. Shape ID changes along the horizontal axis, shape rotation changes along the vertical axis. (B) The two-dimensional Gaussian fit of the angular-position curvature (APC) model to the responses shown in A. The location and scale of the Gaussian indicates that this unit is tuned for broad concavities on the upper regions of the shape. This is somewhat difficult to see in the (A). The correlation value for this fit is r = 0.81.

has shown this type of representation as effective in identifying digits and silhouettes [15]. The importance of curvature for object recognition is also supported by ideas set forth by shape theory [1, 2, 3, 14, 24].

While our results support the use of the angular position and curvature plane for the characterization of mid-level visual units, it is important to note that only a small fraction (10-20%) of AlexNet neurons appear to be curvature-tuned. Interestingly, it is also the case in V4 that many neurons are not well-described by the APC model: from one estimate 24/62 V4 neurons had $r > 0.5$ [11]. This suggests that it may be important to test AlexNet with additional stimulus sets to see if it has units that correspond to other attributes (e.g., color, texture, medial axis symmetry) that have been observed in the brain.

We also found several features of the AlexNet units that were unlike V4 responses. First, many units gave zero response for all or nearly all stimuli. Perhaps adding background texture or noise to the visual stimuli, to increase responses in early layers and overcome rectification, could mitigate this behavior, which makes it difficult to characterize the units. Also, AlexNet units show a strong tendency to be selective for boundary features that were located at 90 or 270 deg (to the top or bottom) relative to the center of mass of each shape. The fact that the parameters from the V4 fits, which were computed under identical conditions as those for AlexNet, did not show any sign of this 90/270 deg bias suggests that the bias is not simply an artifact of the shape set or caused by the fitting procedure. In addition, we computed the parameter distributions for our scrambled response control data, and found that the orientation distribution was completely flat (not shown). We are left to speculate that the bias might result from the over-representation of cardinal axes in the natural scenes on which the network was trained [12]. However, it is possible that it is an artifact of the digitization of images and kernels within the model - unlike AlexNet, the macaque visual system does represent images in a Cartesian coordinate system. More testing is required to determine the origin of this bias. Finally, there are interesting non-monotonic patterns in the number of V4-like neurons found as layer number increases. In particular, the number of V4-like neurons appears to peak in the middle convolutional layers (Layer 3 and 4, Figure 7) and decline by Layer 5. Then, at Layer 6, there is a dramatic increase in percent of V4-like units. More testing will be required to understand how these changes relate to the sequence of computations in AlexNet.

The difficulty of recording from a neuron for an extended period of time necessitates using relatively small sets of stimuli to characterize these neurons. Thus, APC fits of the V4 neurons under study here did not test for numerous other properties that have been established as relevant to V4 in previous work. The response profile of V4 neurons across stimuli have been shown to be largely translation invariant within their receptive fields and robust to occlusions. A clear and important next step to the line of research presented here would be to test if HNN units share these properties. For the stimuli that we have used, the difference between the orientation of a boundary element (which direction it points) and angular position with respect to object center (where it sits on the object boundary) are highly correlated, but stimuli designed to disambiguate these options have provided evidence that V4 neurons use an object centered reference frame [19]. Testing whether HNN unit responses are object centered would be another important step. Identifying where HNN unit response properties fall short of those of real neurons could reveal opportunities for improvement in the representations learned by HNNs. Training on carefully parameterized stimuli that embody these representation could be crucial to embedding the optimal representations in object recognition HNNs.

An important next step in this line of research will be to use the curvature-tuned HNN units identified here to gain insights into how curvature-tuned V4 RFs might be built. Unlike for V1, we do not still have a good circuit model for V4 neurons. Current models [27, 28, 29] have limitations in their abilities to simultaneously capture the selectivity of V4 neurons with respect to object-centered encoding, contour tuning and translation invariance. Further study is required to determine if HNN units are in fact able to capture all of these key features of the V4 representation.

5. ACKNOWLEDGMENTS

This work was funded by a National Science Foundation (NSF) Graduate Research Fellowship (D.A.P.), a Google Faculty Research Award (W.B. and A.P.) and an NSF Collaborative Research in Computational Neuroscience Grant IIS-1309725 (W.B. and A.P.). We thank Blaise Aguera y Arcas for helpful suggestions and advice.

6. REFERENCES

1. Asada, H, Brady M (1984) The curvature primal sketch. In: MIT Artificial Intelligence Laboratory A.I. Memo 758 (MIT,Cambridge).
2. Attneave F (1954) Some informational aspects of visual perception. Psychol Rev. 61:183-193.
3. Besl J, Jain R (1985) Three-Dimensional Object Recognition. Computing Surveys 17:75-145.
4. Bushnell BN, Pasupathy A (2012) Shape encoding consistency across colors in primate V4. J Neurophysiol 108:1299-1308.
5. Desimone R and Schein SJ (1987) Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. J Neurophysiol 57:835-868.
6. Dosovitskiy A, Brox T (2015) Inverting convolutional networks with convolutional networks. arXiv:1506.02753v1 [cs.NE]. [http://arxiv.org/abs/1506.02753].
7. Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cerebral Cortex 1:1-47.
8. Gallant JL, Braun J, van Essen DC (1993) Selectivity for polar, hyperbolic, andcartesian gratings in macaque visual cortex. Science 259:100-103.
9. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T (2014) Caffe: Convolutional architecture for fast feature embedding. arXiv:1408.5093.
10. Kobatake E, Tanaka K (1994) Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. J Neurophysiol 71:856-867.
11. Kosai Y, El-Shamayleh Y, Fyall A, Pasupathy A (2014) The role of area V4 in the discrimination of partially occluded shapes. J Neurosci 34:8570-8584.

12. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems 25, Eds: Pereira F, Burges CJC, Bottou L, Weinberger KQ. pp 1097–1105. [http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf]
13. Mahendran A, Vedaldi A (2014) Understanding deep image representations by inverting them. arXiv:1412.0035v1 [cs.CV]. [http://arxiv.org/abs/1412.0035]
14. Marimont DHA (1984) Representation for image curves. AAAI Proceedings 84:237-242.
15. Murphy TM, Finkel LH (2007) Shape representation by a network of v4-like cells. Neural Networks 20:851–867. DOI= http://dx.doi.org/10.1016/j.neunet.2007.06.004.
16. Nguyen A, Yosinski J, Clune J (2015) Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images. In: Computer Vision and Pattern Recognition (CVPR'15), IEEE. [http://arxiv.org/pdf/1412.1897v4.pdf]
17. Pasupathy A (2006) Neural basis of shape representation in the primate brain. Progress in Brain Research 154:293-313.
18. Pasupathy A, Connor CE (1999) Responses to contour features in macaque area V4. J Neurophysiol 82:2490-2502.
19. Pasupathy A, Connor CE (2001) Shape representation in area V4: position-specific tuning for boundary conformation. J Neurophys 86:2505-2519. [http://jn.physiology.org/cgi/content/abstract/86/5/2505]
20. Pasupathy A, Connor CE (2002) Population coding of shape in area V4. Nature Neuroscience 5:1332-1338.
21. Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. Nat Neurosci 2:1019-1025. [http://dx.doi.org/10.1038/14819]
22. Rust NC, Movshon JA (2005) In praise of artifice. Nature Neuroscience 8:1647-1650.
23. Ungerleider LG, Mishkin M (1982) Two cortical visual systems. Analysis of Visual Behavior, eds, Ingle DJ, Goodale MA, Mansfield RJW (MIT Press, Cambridge, MA) pp 549-586.
24. Verri A, Yuille A (1986) Perspective projection invariance. In: MIT Artificial Intelligence Laboratory A.I. Memo 832 (MIT, Cambridge).
25. Yamins DLK, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ (2014) Performance-optimized hierarchical models predict neural responses in higher visual cortex. PNAS 111:8619-8624. [doi: 10.1073/pnas.1403112111]
26. Zeiler MD, Fergus R (2014) Visualizing and understanding convolutional networks. D. Fleet et al., eds. ECCV 2014, Part 1, LNCS 8689, pp 818-833.
27. Oleskiw TD, Pasupathy A, Bair W (2014) Spectral receptive fields do not explain tuning for boundary curvature in V4 neurons. J Neurophysiol 112:2114-2122.
28. Cadieu C, Kouh M, Pasupathy A, Connor CE, Riesenhuber M, Poggio T (2007) A model of V4 shape selectivity and invariance. J Neurophysiol 98:1733-1750.
29. Rodriguez-Sanchez AJ, Tsotsos JK (2012) The roles of endstopped and curvature tuned computations in a hierarchical representation of 2D shape. PLoS One DOI: 10.1371/journal.pone.0042058