Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Neurobiology

Object shape and surface properties are jointly encoded in mid-level ventral visual cortex

Anitha Pasupathy[1,2], Taekjun Kim[1,2] and Dina V Popovkina[3]

Recognizing a myriad visual objects rapidly is a hallmark of the primate visual system. Traditional theories of object recognition have focused on how crucial form features, for example, the orientation of edges, may be extracted in early visual cortex and utilized to recognize objects. An alternative view argues that much of early and mid-level visual processing focuses on encoding surface characteristics, for example, texture. Neurophysiological evidence from primate area V4 supports a third alternative — the joint, but independent, encoding of form and texture — that would be advantageous for segmenting objects from the background in natural scenes and for object recognition that is independent of surface texture. Future studies that leverage deep convolutional network models, especially focusing on network failures to match biology and behavior, can advance our insights into how such a joint representation of form and surface properties might emerge in visual cortex.

Addresses
[1] Department of Biological Structure, University of Washington, Seattle, WA, United States
[2] Washington National Primate Research Center, University of Washington, Seattle, WA, United States
[3] Department of Psychology, University of Washington, Seattle, WA 98195, United States

Corresponding author: Pasupathy, Anitha (pasupat@uw.edu)

Current Opinion in Neurobiology 2019, **58**:199–208

This review comes from a themed issue on **Computational neuroscience**

Edited by **Brent Doi ron** and **Máté Lengyel**

For a complete overview see the Issue and the Editorial

Available online 4th October 2019

https://doi.org/10.1016/j.conb.2019.09.009

Introduction

Primates are remarkable at recognizing a vast number of objects effortlessly—a capacity that is unmet by even the most advanced computer recognition systems despite dramatic recent advances in deep learning [1–3]. In the primate, object recognition is based on information processing along the ventral visual pathway, which runs from area V1, to V2, V4 and subregions of the inferotemporal cortex [4]. Past studies have demonstrated selectivity for local orientation and spatial frequency in V1 [5–7] and for more complex stimulus attributes in V2, V4 and IT cortex [8,9]. But there is still much debate about what these observed selectivities mean in terms of the underlying bases and how they might contribute to object recognition. Given that natural visual objects have both shape, defined by their bounding contours, and surface attributes, including color and texture, do neurons in the mid-level processing stages encode these stimulus attributes separately or jointly? And, how is such a representation, either joint or separate, advantageous for the purpose of object recognition?

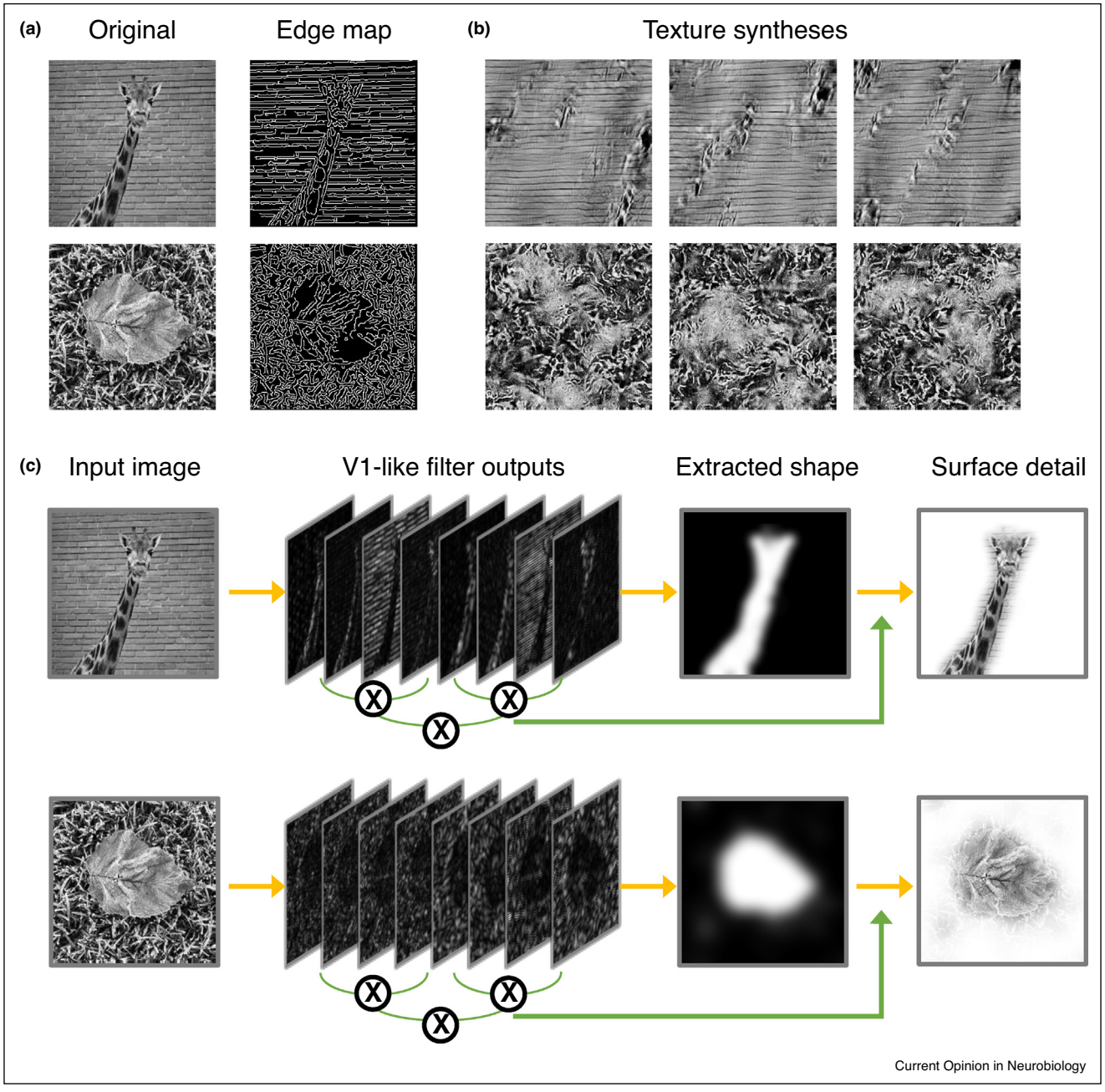
Shape-based approach to object recognition

Traditional theories of visual object recognition have largely focused on how shape cues, for example, the orientation of object boundaries and their surfaces, can contribute to their recognition. For example, in Marr's approach [10], the detection of local edge and surface orientations contributes to the construction of 3D models in terms of surface and volumetric primitives which can then be matched to stored templates of an object. Such shape-based recognition algorithms are intuitively appealing—they match how one might sketch an object—and several lines of neurophysiological evidence appear to support this hypothesis. For example, V1 neurons tuned to local orientation may be thought of as encoding local edges in a visual scene [5], thereby providing the building blocks for developing a shape-based code. In mid-level visual processing stages, in V2 and V4, past studies have demonstrated selectivity for border ownership [11] and 2D boundary curvature [12], respectively. The hierarchical max (HMax) model of object recognition, developed by Poggio *et al.* [13–15], is based on this strategy. In this model, oriented signals from Gabor filters, akin to V1 simple cells, are pooled to generate selectivity for phase-insensitive local orientation. Various combinations of these signals then produce shape templates for recognition. These boundary shape-based models tackle the problem of recognition of isolated objects, in which case detecting edge and surface orientation from an image can effectively accomplish the task. In natural vision, however, objects are embedded in a field of texture and are partially occluded by other objects, so detecting object boundaries from an edge map followed by recognizing component objects may be untenable (Figure 1a).

Encoding visual scenes as textures

In most natural environments, the visual scene is composed of a few objects and a sea of texture, for example, sky, forest, road, beach; even the surfaces of objects are

Figure 1



Candidate models for object recognition.
(a) Deficiencies of a shape-based approach. Example natural scenes (left), and the component edges in the scenes (right) as computed by a Canny edge detector. These examples show that object shape segmentation based on edge-detection becomes very difficult when region contrast information (color, luminance, or texture) is not available. **(b)** Deficiencies of a texture-based approach. Each row shows multiple synthesized images using the Portilla–Simoncelli algorithm [20]. They have the same higher-order texture statistics as the original images in (a), but global form is destroyed. **(c)** Framework for a new joint encoding model. The input image is first processed to extract V1-like features. Here the image is passed through a bank of 4 orientation $\times$ 2 spatial frequency filters, and the output is shown. Component objects in the image are extracted by identifying uniform image regions with K-means clustering (set to $K = 2$). Detailed processing of surface texture also starts from V1 outputs and runs in parallel. As proposed by Okazawa *et al.* [19••], the texture information is encoded by computing correlations in activity among neighboring neurons. This computation may be mediated by slow lateral cortical connections rather than fast feedforward connections.

composed of textures [16]. In terms of making decisions and navigating the world, the success of our visual system depends on parsing objects and interpreting scenes based on visual texture, in addition to recognizing the parsed objects. In this context, V1 selectivity for local orientation and spatial frequency may be better interpreted as encoding the texture rather than shape edges or boundaries, cleverly paraphrased as encoding the 'stuff' rather than 'things' in an image [17]. More recent studies have demonstrated that neurons in V2 [18•]—and V4 [19••] show selective responses to homogeneous textures, which can be explained on the basis of selectivity for sparse combinations of higher-order summary statistics within the receptive field (RF) of the neuron in question [20]. This supports the idea that the early and mid-level processing stages emphasize the encoding of images as textures [21,22], an argument also supported by the demonstration in human psychophysical studies that subjects cannot differentiate between image patches in the near periphery if they are matched in terms of higher order texture statistics [23•, but see Ref. 24••]. In this construct, where neuronal selectivity is simply based on computations of summary statistics within a receptive field region, bottom-up shape selectivity demonstrated in many previous studies could arise as a simple by-product of a statistical texture representation of natural scenes [21,22] allowing for no mechanistic basis for a distinction between shape and texture selectivity at least through mid-level processing. In other words, because different shapes can be described with different sets of higher order texture statistics, selectivity for combinations of higher order statistics could underlie the observed shape selectivity. But because the set of summary statistics associated with a shape changes dramatically when different textures are painted on its surface, and multiple images can have the same summary statistics despite vastly different global form (Figure 1b), such a summary statistics-based encoding scheme cannot underlie our ability to perceive and recognize shape regardless of their surface properties. Other, more complex texture-based models [25–28], which compute different texture statistics in different RF subunits, could differentiate between shape and texture and could support texture-invariant shape recognition.

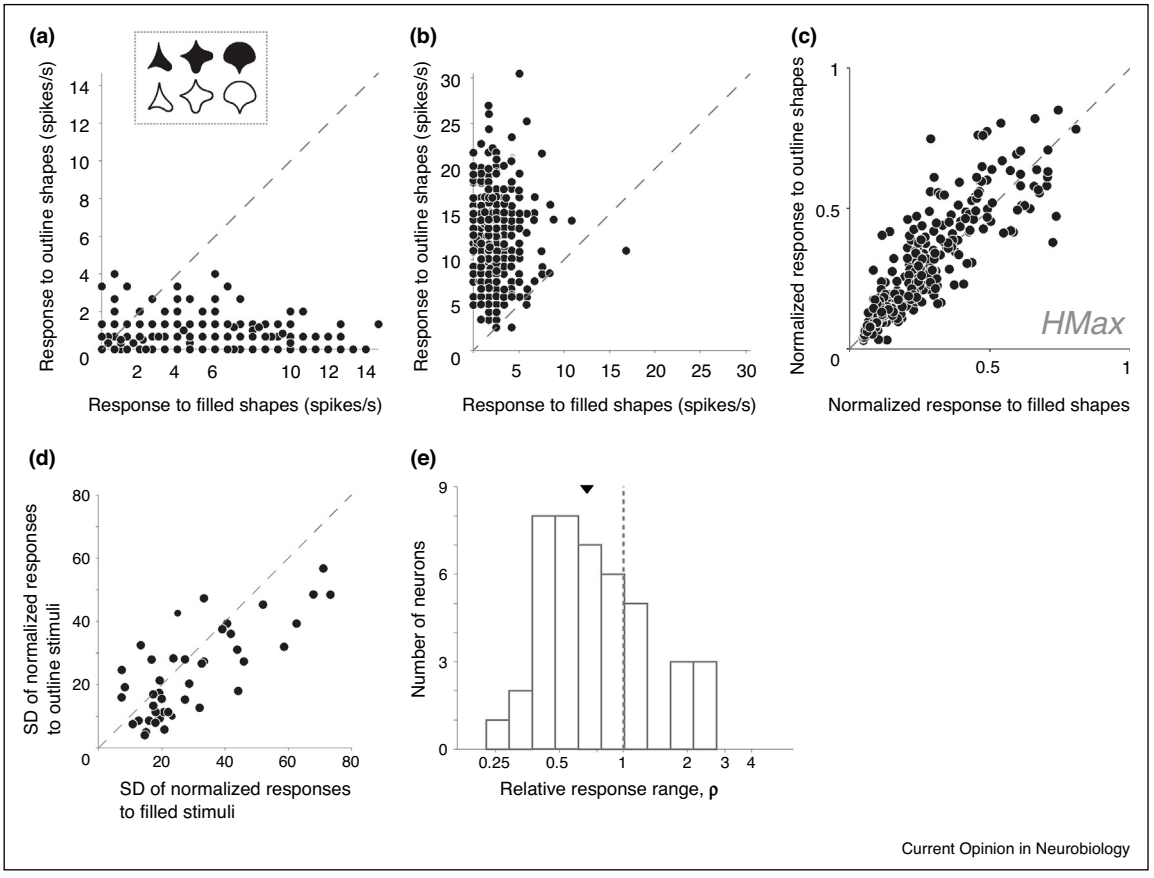
Joint code for shape and texture

The arguments above imply that object recognition is unlikely to be based on neurons that process information using a pure boundary-based or surface-based strategy. The third alternative is that object recognition may be subserved by neurons that jointly encode a shape boundary and the associated surface, that is, by neurons simultaneously tuned to shape and texture/color. Such neurons would encode only those contours that are associated with a local contrast in either texture, luminance and/or color across the boundary and not contours that lack such a contrast. This could provide an automatic means to segment object parts from background texture and internal contours, since the latter two will lack the contrast necessary to drive such neurons. Such a strategy could also explain why segmenting scenes and recognizing objects are difficult when luminance (or color) contrast is stripped from a natural scene (see Figure 1a). The balanced encoding of shape and texture has been proposed previously [25] and emerges as a representational characteristic in physiological and behavioral data-driven models [26–28]. These results are consistent with recent human psychophysical results suggesting that texture summary models may not be sufficient to explain peripheral appearance [24••].

Three recent neurophysiological studies from area V4, an important intermediate stage for form processing in the macaque monkey, provide evidence in support of this hypothesis. First, we recently demonstrated that for a majority of V4 neurons shape selectivity is not invariant of interior fill [29••]. In other words, a neuron may exhibit different levels of responses and shape selectivity when tested with outlines versus filled shapes (Figure 2a,b). Across the V4 population, we found that neurons were generally more responsive and shape-selective with filled shapes than with outlines (Figure 2d,e). Responses to outline shapes reflected a true loss of selectivity and not a simple gain change, suggesting that shape-selective signals are informed by the conformation of a contrast boundary, rather than a simple curved line. In striking contrast to the observed V4 data, the HMax model, which uses an edge orientation-based strategy to build shape selectivity [15], exhibits strong interior fill invariance in shape selectivity [29••] (Figure 2c). Popovkina *et al.* [29••] propose two modifications to the model that incorporate information about object surface: one relies on maintaining sensitivity of oriented units to polarity across an edge, while the other includes unoriented units sensitive to lower spatial frequencies. Both modifications are better able to capture the differential responses to filled and outline stimuli observed in the neuronal data.

Second, many V4 neurons are jointly tuned to both shape and texture attributes of visual stimuli [30••]. A recent study has demonstrated tuning for naturalistic texture in V4 neurons and developed models to explain such selectivity on the basis of higher order texture statistics [19••]. Older studies, In contrast, have demonstrated systematic tuning for shape stimuli [12,31,32]. But because V4 studies have typically focused either on shape or texture tuning, we do not know whether different subgroups of neurons are tuned to shapes versus texture, whether the same subset of neurons is tuned to both attributes, or whether tuning for texture in terms of higher order image statistics can explain observed V4 shape selectivity [22]. Recently, we studied the responses of V4 neurons to both shape and texture stimuli and found that overlapping subsets of V4 neurons likely contribute to shape and texture perception: while some neurons respond strongly and selectively to shapes but not textures, others respond preferentially to

Figure 2



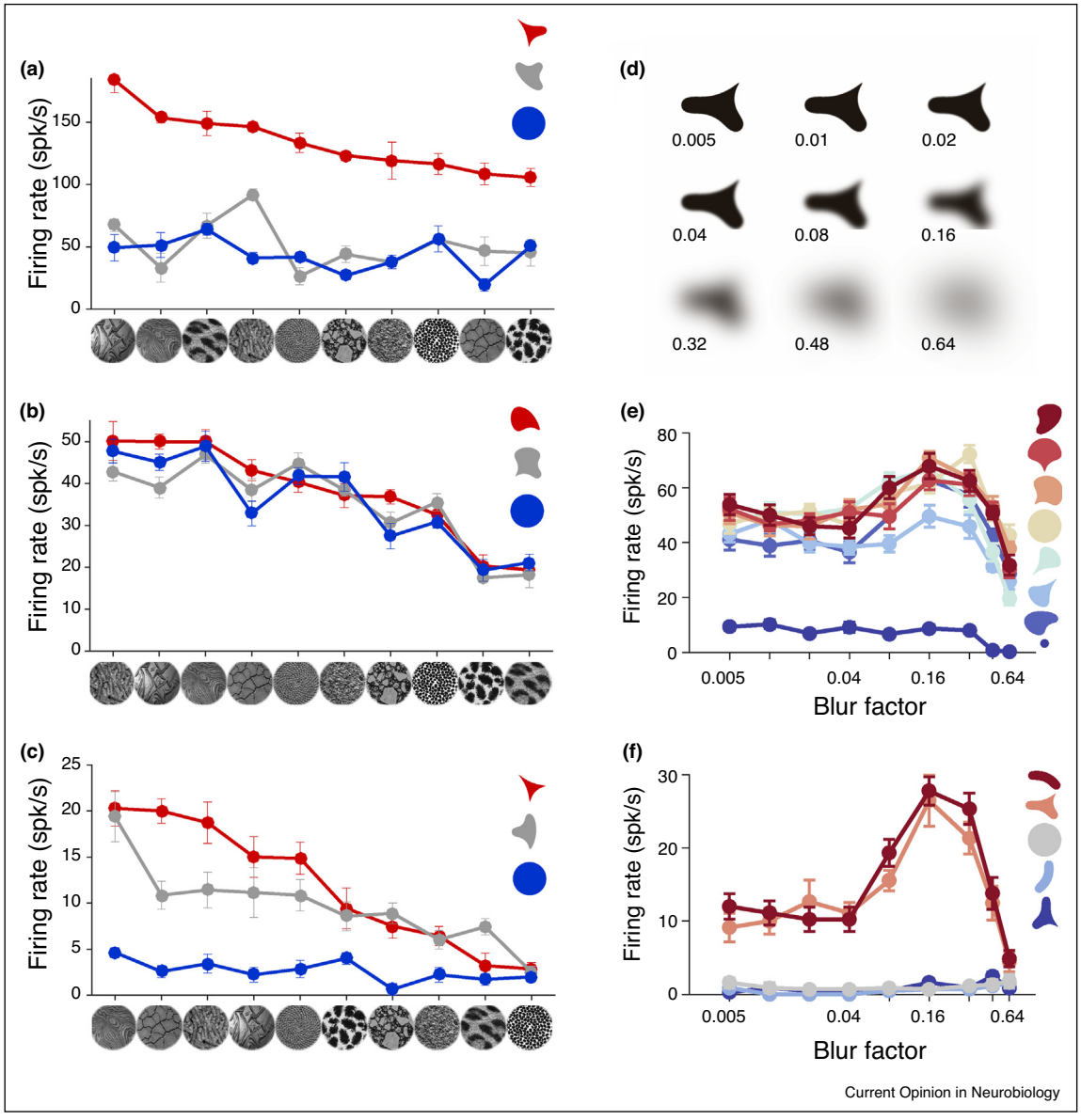
Responses to filled and outline stimuli in V4.
(a),**(b)** Results from two example V4 neurons (a) and (b) and from an instantiation of the Hmax model **(c)** are shown. Each point represents the mean response of a single neuron to a filled (X-axis) and an outline (Y-axis) stimulus with the same boundary shape (inset: stimulus examples). We recorded the responses of each neuron to 362 shapes. (a) An example neuron that evoked a larger range, and more shape-selective responses to filled shapes than outline shapes. (b) An example neuron that responded strongly and selectively to outlines, but not to filled stimuli. (c) A typical HMax model unit responds strongly and selectively to both filled and outline stimuli. Importantly, the responses are highly correlated. **(d)**,**(e)** Responsiveness to filled and outline stimuli in a population of 43 V4 neurons. (d) Standard deviation (SD) of normalized responses to outline stimuli plotted against SD of normalized responses to filled stimuli; each point represents one neuron. Most points lie below the diagonal indicating that dispersion of responses across shapes was greater for filled shapes. (e) Distribution of relative response range, ρ, in the same population of neurons. For each neuron, this metric compares the responsiveness to outline versus filled stimuli (see Ref. [29••] for further details). Dashed line: similar responsiveness to filled and outline stimuli; left of dashed line, more responsive to filled stimuli; right of dashed line, more responsive to outline stimuli. Triangle indicates median.

textures but not shapes, and still others are tuned to both (Figure 3a–c). Importantly, we found that tuning for shape and texture are separable, arguing against the possibility that tuning for higher order texture statistics can explain shape tuning. These results are consistent with previous studies demonstrating joint, separable tuning for luminance/color contrast and shape [33,34] and argue for an emergent object code that reflects joint tuning for shape and surface attributes of texture, luminance and color.

Third, many shape-selective V4 neurons also exhibit tuning for boundary blur, that is, the spatial gradient of image intensity across the edge (Figure 3d–f; [35••]). In natural vision, blurred boundaries may arise due to defocus, cast shadows or surface shading [36] and can contribute to segmentation, depth perception and scene understanding [37–39]. This provides further support for an object-code in V4 that includes information about boundary shape and the surface attributes of the image.

Overall, the above results suggest that shape selectivity in most V4 neurons likely arises by pooling both surface contrast and boundary contour information. Such a strategy would facilitate the segmentation of objects from natural scenes and can explain the difficulty in detecting shapes and recognizing them when surface color/

Figure 3



Responses to shapes, textures, and boundary blur in V4.
(a)–(c) Responses of example neurons to shapes (shown in each panel) with different textures painted on the surface. Line colors denote shape; textures are ordered along the X-axis in accordance with responses to the shape shown in red. For each neuron we chose a preferred shape (red) based on preliminary characterization, a less-preferred shape (gray) and a circle. Error bars indicate standard error of mean. (a) Example neuron with strong selectivity for shape but not texture. (b) Example neuron with strong selectivity for texture but not shape. All three line colors follow a similar trend. (c) Example neuron that is tuned to both shape and texture. In all three cases, tuning for shape and texture are separable and responses can be modeled as a product of tuning for shape × texture. **(d),(e)** Encoding of boundary blur in V4. (d) Stimuli with boundary blur. Blurring was achieved by applying a circular 2D Gaussian blur kernel to the shape. The numbers denote blur factor, which represents the standard deviation of the Gaussian kernel in units relative to the radius of the circle shape (see panel c). (e),**(f)**. Two example neurons that exhibit tuning for blur. Responses to preferred (more red) and non-preferred (more blue) shapes are plotted as a function of the blur factor. Both neurons are selective for intermediate levels of blur. Tuning for blur was independent of the shapes used for this and other neurons [35••]. Error bars indicate standard error of mean. This figure is adapted with permission from [35••] under the Creative Commons Attribution 4.0 International License.

luminance or texture information is excluded. Next, we discuss how these results relate to studies in human subjects and propose a framework for a new model of how shape selectivity might arise.

Evidence from human psychophysics and neuroimaging

To fully contextualize the evidence supporting joint coding of shape and surface properties in monkey V4,

it is important to understand the potential behavioral relevance of such a representation, and ultimately how these insights extend to object recognition processes in the human brain. Evidence from human psychophysics experiments supports the viewpoint that both shape and surface attributes contribute to object and scene recognition. For example, adding surface-based cues significantly improves the accuracy of boundary detection [40•]; human subjects use both surface and boundary-based segmentation strategies to detect objects [41]; and, both chromatic and achromatic boundary information contributes to object contour perception [42,43•].

There is still considerable debate regarding which human mid-level form processing stage best corresponds to monkey V4. In accordance with the current literature, human area V4 and lateral occipital cortex (LOC) may be most like monkey V4 in terms of neural computations and representations [44–46], and the transition from single feature coding to conjunction coding [47]. In these mid-level processing stages in the human, neuroimaging studies have typically found evidence for segregated as opposed to joint tuning. This includes evidence for strong color selectivity in regions poorly selective for shape [48], selectivity for object/scene shape but not texture in LOC [49–51], and for texture in posterior collateral sulcus [51–53]. While such segregated representations may be a feature of the human (as opposed to monkey) brain, it is also possible that observing segregated patches is a consequence of the analytical methods typically employed in neuroimaging studies: intermediates in a continuum of joint encoding could be missed by threshold response contrasts. For example, a texture versus shape response contrast would distinguish areas that respond more strongly to shape than texture (above some threshold), and vice versa, while failing to identify areas with joint selectivity for both attributes.

In fact, there is some evidence suggesting that shape and surface properties of objects do jointly contribute to representations in object-selective LOC. Neural responses in LOC are selective for objects with smooth, compared to textured surfaces [54], and LOC is thought to initiate filling-in processes that produce illusory Kanizsa percepts [55]. A study examining whether stereotypy in object-color combinations (e.g. yellow bananas) influences responses to achromatic stimuli [56] found evidence for interaction between form and 'memory color' information in human visual cortex. Future targeted experimental and analytical approaches could clarify functional clustering and feature integration results in human cortex, and more directly address the existence of joint processing of object shape and surface. A joint code for edges and surfaces could support the integration of features in mid-level areas such as V4 or LOC, and enable efficient recognition and segmentation.

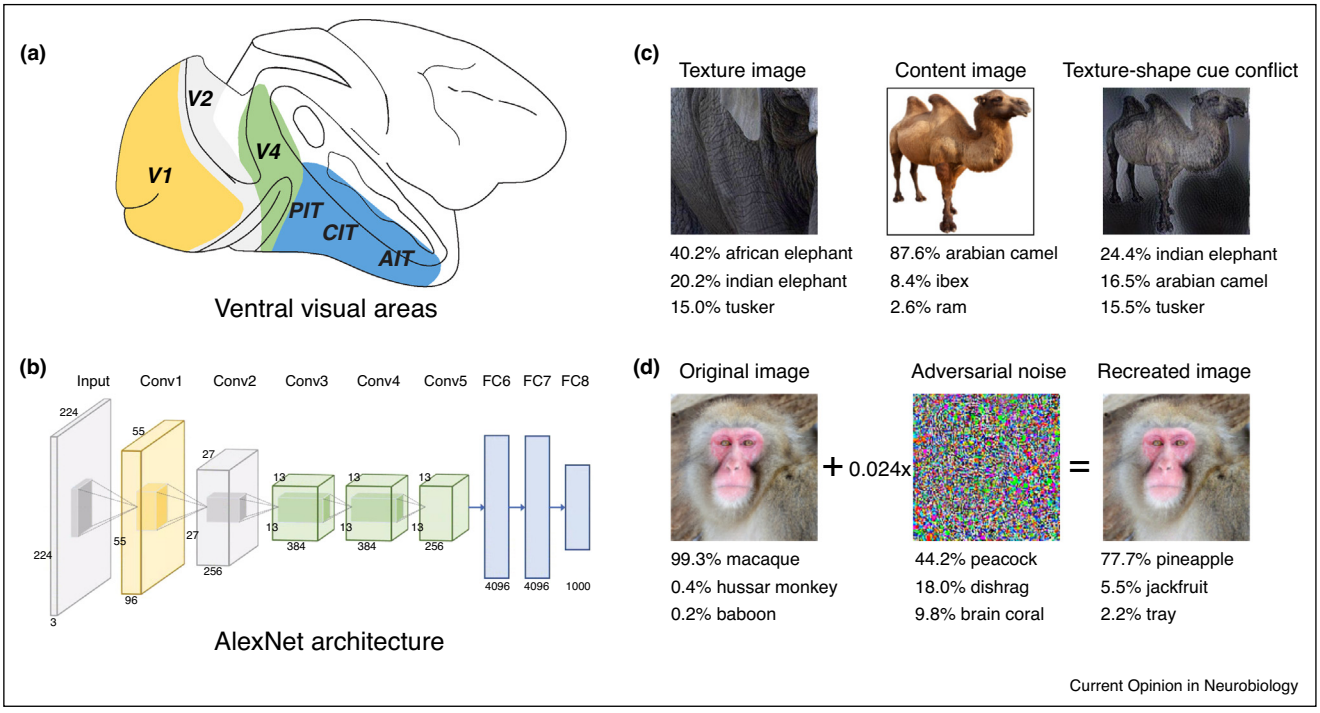
How might joint encoding arise in primate cortex?

Current models of V4 fail to explain how joint encoding of shape and texture tuning might arise in V4. As explained above, the HMax model with its focus on boundary-based shape attributes is poor at capturing tuning for texture [29••,30••]. In contrast, the higher order texture statistics model of Okazawa *et al.* [19••] can capture tuning for texture but it fails to generate the texture-invariant shape tuning observed in V4 (Figure 3). Nandy *et al.* [57] have proposed that V4 tuning for curved contours could arise from pooling heterogeneous orientation signals from earlier visual areas. However, such a model cannot produce translation invariance in shape tuning, a fundamental property of many V4 neurons [12,32,58]. The spectral receptive field model of Gallant *et al.* [59] can achieve translation invariance but not object-centered position-specific tuning for boundary curvature [60]. A new model is therefore needed and we propose a framework in Figure 1c. The first stage of this model includes a bank of orientation × spatial frequency × spatial position filters, analogous to V1. The goal of the next stage is to encode the shape of segmented image regions with similar surface characteristics of luminance, color and texture. In this framework, we achieve such segmentation using K-means clustering, but how this would be implemented in the primate brain is unknown; it could amount to finding contiguous image regions with similar output across the V1 filter bank. Because shape selectivity emerges rapidly in V4 [61], this segmentation and shape encoding process will need to be based on fast computations that rely on feedforward connections. Texture selectivity in V4, in contrast, emerges more slowly [30••] and is likely based on lateral processes. We reiterate that the schematic in Figure 1c is simply a framework that will need to be revised on the basis of future experiments and results.

Exploring deep networks for models of form processing

In terms of identifying good models for form processing in the primate ventral pathway, much remains to be accomplished. Deep convolutional neural networks (DCNs)—the current best image recognition algorithms—may serve as a crucial guide and tool in this endeavor. There are high-level similarities between DCNs and primate brains in terms of architecture and increasing representational complexity with depth [62], and there have been efforts to relate DCN layers to areas along the ventral visual pathway [63]. But, there are also striking differences: DCNs lack feedback which is abundant in biology, and DCNs are easily fooled (Figure 4). With these limitations in mind (see Ref. [64]), comparing DCNs to the primate brain in terms of both physiological properties and psychophysics can be highly insightful. For example, identifying DCN units that match neurons in the brain in terms of responses to visual stimuli could

Figure 4



Deep convolutional neural nets and the primate ventral visual pathway.
(a) Schematic of a macaque brain (side view) indicating ventral visual areas V1, V2, V4, and inferotemporal cortex (IT; PIT: posterior IT; CIT: central IT; AIT: anterior IT). **(b)** Schematic of AlexNet, an example deep convolutional network. Color coding identifies stages of the network and primate brain areas in (a) hypothesized to be analogous based on published results [V1: [65]; V4: [66•]; IT: [63]). **(c)**,**(d)**. Examples of images leading to classification failure in deep nets. (c). A conflict between texture and shape in the input image, generated by style transfer between the first two images [68••,69], produces texture-based classification of a camel as an elephant. (d) Adversarial noise [70] produces misclassification of a modified input image that is indistinguishable from the original by human observers. In this case, a macaque monkey is misclassified as a pineapple with considerable certainty despite minimal difference between the original and modified image.

provide insights into how those properties arise. A recent study, which compared V4 responses to simple shapes to those of units in the mid-level stages of AlexNet (a deep network trained to categorize images, [65]) identified the current best models for V4-like boundary curvature selectivity [66•]. Such DCN models of V4 are image-computable, that is, they can predict responses to any arbitrary image. Thus, in a closed loop with primate physiology experiments, they can be validated by testing response predictions to novel synthetic stimuli [e.g. see Ref. 67•]. While such endeavors highlight the similarities between the two structures, deeper edification may come from leveraging their differences. For example, comparing recognition performance in human subjects and DCNs has revealed the striking texture-bias in the representations of the latter and how an object-based representation may be beneficial in the face of distortions and degradations [68••]. Further studies comparing physiological responses, behavior, and DCNs, to shape and texture stimuli may provide insights into how joint encoding of shapes and surfaces emerge in the primate brain.

Conclusion

Recent studies are beginning to shed light on how shapes and surfaces are encoded in mid-level ventral visual cortex. Results reviewed above support the emergent view that mid-level stages jointly encode shape and surface attributes and such a code may be advantageous for successful object segmentation in natural scenes. Further studies are needed to extend our understanding to more naturalistic viewing conditions, to understand how objects may be segmented from the background, how such objects may be encoded, and how object encoding may be degraded in clutter, providing much needed neurophysiological basis for the phenomena of visual crowding [71]. To further understand how these properties arise in cortex, data-driven models that leverage the power of DCNs could provide key insights. Most of all, it would be crucial to bring to bear all available physiological and psychophysical data as constraints to identify the best candidate models for how the brain builds a complex object categorization system from simple earlier representations.

Conflict of interest statement

Nothing declared.

Acknowledgements

The authors would like to thank Dr. Timothy Oleskiw for providing images for Figure 3d–f. This work was supported by National Eye Institute Grants R01 EY018839 and EY029997 to A.P. and the National Institutes of Health/Office of Research Infrastructure Programs Grant P51 OD010425 to the Washington National Primate Research Center.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Nguyen A, Yosinski J, Clune J: **Deep neural networks are easily fooled: high confidence predictions for unrecognizable images**. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2014:427-436.

2. Pepik B, Benenson R, Ritschel T, Schiele B: **What is holding back convnets for detection?** *DAGM 2015: Pattern Recognition*. Springer; 2015:517-528.

3. Marcus G: *Deep Learning: A Critical Appraisal*. 2018. arXiv preprint arXiv:1801.00631.

4. Felleman DJ, Van Essen DC: **Distributed hierarchical processing in the primate cerebral cortex**. *Cereb Cortex* 1991, **1**:1-47.

5. Hubel DH, Wiesel TN: **Receptive fields and functional architecture of monkey striate cortex**. *J Physiol* 1968, **195**:215-243.

6. Movshon JA, Thompson ID, Tolhurst DJ: **Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex**. *J Physiol* 1978, **283**:101-120.

7. Albrecht DG, De Valois RL, Thorell LG: **Visual cortical neurons: are bars or gratings the optimal stimuli?** *Science (80-)* 1980, **207**:88-90.

8. Kravitz DJ, Saleem KS, Baker CI, Ungerleider LG, Mishkin M: **The ventral visual pathway: an expanded neural framework for the processing of object quality**. *Trends Cogn Sci* 2013, **17**:26-49.

9. Wilson HR, Wilkinson F: **From orientations to objects: configural processing in the ventral stream**. *J Vis* 2015, **15**:4.

10. Marr D, Vision: *Chapter 1: The Philosophy and the Approach*. W.H. Freeman and Co.; 1982.

11. Zhou H, Friedman HS, von der Heydt R: **Coding of border ownership in monkey visual cortex**. *J Neurosci* 2000, **20**:6594-6611.

12. Pasupathy A, Connor CE: **Shape representation in area V4: position-specific tuning for boundary conformation**. *J Neurophysiol* 2001, **86**:2505-2519.

13. Riesenhuber M, Poggio T: **Hierarchical models of object recognition in cortex**. *Nat Neurosci* 1999, **2**:1019-1025.

14. Serre T, Wolf L, Poggio T: **Object recognition with features inspired by visual cortex**. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. 2005:994-1000.

15. Cadieu C, Kouh M, Pasupathy A, Connor CE, Riesenhuber M, Poggio T: **A model of V4 shape selectivity and invariance**. *J Neurophysiol* 2007, **98**:1733-1750.

16. Adelson EH: **On seeing stuff: the perception of materials by humans and machines**. In *Proc. SPIE 4299, Human Vision and Electronic Imaging VI*. Edited by Rogowitz BE, Pappas TN. International Society for Optics and Photonics; 2001:1-12.

17. Adelson EH, Bergen JR: **The plenoptic function and the elements of early vision**. In *Computational Models of Visual Processing*. Edited by Landy M, Movshon JA. MIT Press; 1991:3-20.

18. • Freeman J, Ziemba CM, Heeger DJ, Simoncelli EP, Movshon JA: **A functional and perceptual signature of the second visual area in primates**. *Nat Neurosci* 2013, **16**:974-981.
Using functional magnetic resonance imaging (fMRI) studies in human subjects and neurophysiological studies in monkeys, this paper reveals a functional role of V2 in the representation of natural image structure. Macaque V2, but not V1, neurons respond more strongly to stimuli with the higher-order statistical dependencies found in natural texture images. fMRI measurements in humans revealed differences between V1 and V2 that paralleled the neuronal measurements.

19. •• Okazawa G, Tajima S, Komatsu H: **Image statistics underlying natural texture selectivity of neurons in macaque V4**. *Proc Natl Acad Sci U S A* 2015, **112**:E351-E360.
In this paper the authors used efficient sampling techniques and texture synthesis to characterize responses of individual V4 in a high-dimensional texture space. Their results demonstrate that many neurons in macaque V4 selectively encode sparse combinations of higher-order image statistics to represent natural textures.

20. Portilla J, Simoncelli EP: **A parametric texture model based on joint statistics of complex wavelet coefficients**. *Int J Comput Vis* 2000, **40**:49-70.

21. Movshon JA, Simoncelli EP: **Representation of naturalistic image structure in the primate visual cortex**. *Cold Spring Harb Symp Quant Biol* 2014, **79**:115-122.

22. Ziemba CM, Freeman J: **Representing "stuff" in visual cortex**. *Proc Natl Acad Sci U S A* 2015, **112**:942-943.

23. • Freeman J, Simoncelli EP: **Metamers of the ventral stream**. *Nat Neurosci* 2011, **14**:1195-1201.
Using a combination of modelling and psychophysics, this paper demonstrates that human subjects cannot distinguish between image patches in the near-periphery with different global form but matched higher-order image statistics. Measurements in human subjects suggest that a texture-based representation at the level of V2 may explain observed psychophysical results and can provide a neurophysiological basis for visual crowding.

24. •• Wallis TS, Funke CM, Ecker AS, Gatys LA, Wichmann FA, Bethge M: **Image content is more important than Bouma's Law for scene metamers**. *eLife* 2019, **8**:e42512.
This paper rigorously tests the hypothesis that the visual system averages information over small regions (about the size of V2 receptive field) in the peripheral visual field. On the basis of psychophysical experiments the authors report that human observers are highly sensitive to differences between natural scene images and synthesized images with matched texture statistics. This suggests that the visual system incorporates global properties of the scene rather than solely using textures to represent information in the peripheral visual field.

25. Landy M, Graham N: **Visual perception of texture**. In *The Visual Neurosciences*. Edited by Chalupa LM, Werner JS. MIT Press; 2004:1106-1118.

26. Yu Y, Schmid AM, Victor JD: **Visual processing of informative multipoint correlations arises primarily in V2**. *eLife* 2015, **4**:e06604.

27. Rowekamp RJ, Sharpee TO: **Cross-orientation suppression in visual area V2**. *Nat Commun* 2017, **8**:15739.

28. DiMattina C, Baker CL: **Modeling second-order boundary perception: a machine learning approach**. *PLOS Comput Biol* 2019, **15**:e1006829.

29. •• Popovkina D, Bair W, Pasupathy A: **Modeling diverse responses to filled and outline shapes in macaque V4**. *J Neurophysiol* 2019, **121(3)**:1059-1077.
This paper demonstrates that responses of shape-selective V4 neurons are dictated both by the shape of the object boundary and the object's interior fill. Some neurons are shape-selective only for outline shapes, some only for filled shapes and a minority for both classes of stimuli. This observation violates the prediction of the prominent HMax model for object recognition, where shape selectivity is informed just by integrating boundary orientation information. Two new models are proposed that can capture fill-dependent shape selectivity observed in V4. One model takes advantage of sensitivity to polarity across an oriented edge already present in early layer units, and extends it to units in later layers. A

second model introduces unoriented early layer units sensitive to lower spatial frequency content.

30. •• Kim T, Bair W, Pasupathy A: **Neural coding for shape and texture in macaque area V4**. *J Neurosci* 2019, **39**:4760-4774.
Using a systematically designed set of shape, texture and combination stimuli, this study provides evidence that shape and texture properties are encoded by independent mechanisms at the level of V4. Some neurons specialize in shape processing whereas others specialize in processing texture. In most neurons that lie between the ends of this continuum, tuning for shape and texture are largely separable. Furthermore, tuning for texture emerges significantly later than for shape in individual neurons.

31. Kobatake E, Tanaka K: **Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex**. *J Neurophysiol* 1994, **71**:856-867.

32. Gallant JL, Braun J, Van Essen DC: **Selectivity for polar, hyperbolic, and cartesian gratings in macaque visual cortex**. *Science (80-)* 1993, **259**:100-103.

33. Bushnell BN, Pasupathy A: **Shape encoding consistency across colors in primate V4**. *J Neurophysiol* 2012, **108**:1299-1308.

34. McMahon DBT, Olson CR: **Linearly additive shape and color signals in monkey inferotemporal cortex**. *J Neurophysiol* 2009, **101**:1867-1875.

35. •• Oleskiw TD, Nowack A, Pasupathy A: **Joint coding of shape and blur in area V4**. *Nat Commun* 2018, **9**:466.
This paper demonstrates that many neurons in macaque V4 exhibit tuning for both object shape and the blur of the boundary. Specifically, V4 neurons tuned to intermediate blur levels respond best when high spatial frequency content is removed by blurring the boundary. V4 responses are well-described by a model wherein blur selectivity is cast as a distinct neural process that modulates the gain of shape-selective V4 neurons, supporting the hypothesis that shape and blur are fundamental features of a sufficient neural code for natural image representation in V4.

36. Elder JH: **Are edges incomplete?** *Int J Comput Vis* 1999, **34**:97-122.

37. Rensink RA, Cavanagh P: **The influence of cast shadows on visual search**. *Perception* 2004, **33**:1339-1358.

38. Held RT, Cooper EA, Banks MS: **Blur and disparity are complementary cues to depth**. *Curr Biol* 2012, **22**:426-431.

39. Burge J, Geisler WS: **Optimal disparity estimation in natural stereo images**. *J Vis* 2014, **14** 1–1.

40. • Mély DA, Kim J, McGill M, Guo Y, Serre T: **A systematic comparison between visual cues for boundary detection**. *Vision Res* 2016, **120**:93-107.
This study investigates how early visual processes inform boundary detection in natural scenes. The authors used a set of color binocular video sequences annotated with ground-truth contours and machine learning classifiers to assess which of color, luminance, motion and stereo cues were the most diagnostic for boundary detection. Color and luminance were found to be the most diagnostic and combining all cues yielded a significant improvement in accuracy beyond that of any cue in isolation. Furthermore, region-based methods were found to be superior to edge-based approaches.

41. Machilsen B, Wagemans J: **Integration of contour and surface information in shape detection**. *Vision Res* 2011, **51**:179-186.

42. Hansen T, Gegenfurtner KR: **Independence of color and luminance edges in natural scenes**. *Vis Neurosci* 2009, **26**:35.

43. • Hansen T, Gegenfurtner KR: **Color contributes to object-contour perception in natural scenes**. *J Vis* 2017, **17**:14.
This paper demonstrates that chromatic information contributes to object-contour perception. To assess the relative role of chromatic and achromatic edges, the authors investigated how well human-marked contours can be predicted from achromatic and chromatic edge contrasts. Prediction improved dramatically when chromatic edge information was used in addition to achromatic edge information. The gains in the opposite direction were less dramatic.

44. Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB: **Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex**. *Proc Natl Acad Sci U S A* 1995, **92**:8135-8139.

45. Kourtzi Z, Kanwisher N: **Representation of perceived object shape by the human lateral occipital complex**. *Science (80-)* 2001, **293**.

46. Grill-Spector K, Weiner KS: **The functional architecture of the ventral temporal cortex and its role in categorization**. *Nat Rev Neurosci* 2014, **15**:536-548.

47. Cowell RA, Leger KR, Serences JT: **Feature-coding transitions to conjunction-coding with progression through human visual cortex**. *J Neurophysiol* 2017, **118**:3194-3214.

48. Lafer-Sousa R, Conway BR, Kanwisher NG: **Color-biased regions of the ventral visual pathway lie between face- and place-selective regions in humans, as in macaques**. *J Neurosci* 2016, **36**:1682-1697.

49. Kourtzi Z, Kanwisher N: **Cortical regions involved in perceiving object shape**. *J Neurosci* 2000, **20**:3310-3318.

50. Grill-Spector K, Kourtzi Z, Kanwisher N: **The lateral occipital complex and its role in object recognition**. *Vision Res* 2001, **41**:1409-1422.

51. Cant JS, Xu Y: **The contribution of object shape and surface properties to object ensemble representation in anterior-medial ventral visual cortex**. *J Cogn Neurosci* 2017, **29**:398-412.

52. Cavina-Pratesi C, Kentridge RW, Heywood CA, Milner AD: **Separate processing of texture and form in the ventral stream: evidence from fMRI and visual agnosia**. *Cereb Cortex* 2010, **20**:433-446.

53. Cavina-Pratesi C, Kentridge RW, Heywood CA, Milner AD: **Separate channels for processing form, texture, and color: evidence from fMRI adaptation and visual object agnosia**. *Cereb Cortex* 2010, **20**:2319-2332.

54. Echavarria C, Nasr S, Tootell R: **Smooth versus textured surfaces: feature-based category selectivity in human visual cortex**. *eNeuro* 2016, **3** ENEURO.0051-16.2016.

55. Wokke ME, Vandenbroucke ARE, Scholte HS, Lamme VAF: **Confuse your illusion**. *Psychol Sci* 2013, **24**:63-71.

56. Bannert MM, Bartels A: **Decoding the yellow of a gray banana**. *Curr Biol* 2013, **23**:2268-2272.

57. Nandy AS, Sharpee TO, Reynolds JH, Mitchell JF: **The fine structure of shape tuning in area V4**. *Neuron* 2013, **78**:1102-1115.

58. El-Shamayleh Y, Pasupathy A: **Contour curvature as an invariant code for objects in visual area V4**. *J Neurosci* 2016, **36**.

59. David SV, Hayden BY, Gallant JL: **Spectral receptive field properties explain shape selectivity in area V4**. *J Neurophysiol* 2006, **96**.

60. Oleskiw TD, Pasupathy A, Bair W: **Spectral receptive fields do not explain tuning for boundary curvature in V4**. *J Neurophysiol* 2014, **112**.

61. Bushnell BN, Harding PJ, Kosai Y, Pasupathy A: **Partial occlusion modulates contour-based shape encoding in primate area V4**. *J Neurosci* 2011, **31**:4012-4024.

62. Güçlü U, van Gerven MAJ: **Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream**. *J Neurosci* 2015, **35**:10005-10014.

63. Yamins DLK, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ: **Performance-optimized hierarchical models predict neural responses in higher visual cortex**. *Proc Natl Acad Sci U S A* 2014, **111**:8619-8624.

64. Yuille AL, Liu C: *Deep Nets: What Have They Ever Done for Vision?* 2018. arXiv preprint arXiv:1805.04025.

65. Krizhevsky A, Sutskever I, Hinton GE: **ImageNet classification with deep convolutional neural networks**. *Adv Neural Inf Process Syst* 2012, **25**:1097-1105.

66. • Pospisil DA, Pasupathy A, Bair W: **"Artiphysiology" reveals V4-like shape tuning in a deep network trained for image classification**. *eLife* 2018, **7**.

Taking the approach of an electrophysiologist, this paper asks whether individual units in a deep convolutional neural network (AlexNet) exhibit translation-invariant boundary curvature selectivity similar to that observed in V4 neurons. The authors find that some units, especially in the middle layers of AlexNet, exhibit properties consistent with selectivity for object boundaries and thus, provide the best image-computable model for object-centered boundary curvature selectivity. This raises the possibility that single-unit selectivity in artificial networks will become a guide for understanding sensory cortex.

67. • Bashivan P, Kar K, DiCarlo JJ: **Neural population control via deep image synthesis**. *Science* 2019, **364**:eaav9436.

In this paper, the authors used artificial neural network-driven image synthesis methods to create synthetic images to push spiking activity in V4 to higher levels than observed with natural stimuli or parametrically designed artificial stimuli.

68. •• Geirhos R, Rubisch P, Michaelis C, Bethge M, Wichmann FA, Brendel W: *Imagenet-trained Cnns Are Biased towards Texture; Increasing Shape Bias Improves Accuracy and Robustness*. 2018. arXiv preprint arXiv:1811.12231.

The authors found that ImageNet-trained CNN models rely heavily on local texture cues and underutilize shape information to classify images than humans do. Therefore, their classification accuracy is easily deteriorated by pixel-level image manipulations, even when they are imperceptible to humans. They proposed that texture-bias can be corrected (and/or human-like shape-based representation can be learned) by training CNNs with Stylized-ImageNet, in which object-related textures are replaced randomly. This substantially improves accuracy and robustness against a wide range of image distortions.

69. Gatys LA, Ecker AS, Bethge M: **Image style transfer using convolutional neural networks**. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2016:2414-2423.

70. Goodfellow IJ, Shlens J, Szegedy C: *Explaining and Harnessing Adversarial Examples*. 2014. arXiv preprint arXiv:1412.6572.

71. Whitney D, Levi DM: **Visual crowding: a fundamental limit on conscious perception and object recognition**. *Trends Cogn Sci* 2011, **15**:160-168.