Version date: 10/7/19

A dual role for shape skeletons in human vision: perceptual organization and object recognition

Vladislav Ayzenberg[1], Frederik S. Kamps[2], Daniel D. Dilks[1], Stella F. Lourenco[1]
[1]Department of Psychology, Emory University
[2]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

Corresponding authors:
Vladislav Ayzenberg: vayzenb@emory.edu
Stella F. Lourenco: stella.lourenco@emory.edu

Abstract

Shape perception is crucial for object recognition. However, it remains unknown exactly how shape information is represented, and, consequently, used by the visual system. Here, we hypothesized that the visual system incorporates "shape skeletons" to *both* (1) perceptually organize contours and component parts into a shape percept, and (2) compare shapes to recognize objects. Using functional magnetic resonance imaging (fMRI) and representational similarity analysis (RSA), we found that a model of skeletal similarity explained significant unique variance in the response profiles of V3 and LO, regions known to be involved in perceptual organization and object recognition, respectively. Moreover, the skeletal model remained predictive in these regions even when controlling for other models of visual similarity that approximate low- to high-level visual features (i.e., Gabor-jet, GIST, HMAX, and AlexNet), and across different surface forms, a manipulation that altered object contours while preserving the underlying skeleton. Together, these findings shed light on the functional roles of shape skeletons in human vision, as well as the computational properties of V3 and LO.

Key words: fMRI, perceptual organization, object recognition, medial axis, V3, LO

A central goal of vision science is to understand how the human visual system represents the shapes of objects and how shape is ultimately used to recognize objects. Research from computer vision has suggested that shape representations can be created and then compared using computational models based on the medial axis, also known as the "shape skeleton." Although recent behavioral studies suggest that humans also represent shape skeletons (Ayzenberg & Lourenco, 2019; Firestone & Scholl, 2014), it remains unknown whether they contribute to perceptual organization, object recognition, or both. Here we provide important neural evidence that shape skeletons contribute to both functions.

Shape skeletons are models of structure based on the medial axis of an object (Blum & Nagel, 1978). They provide a quantitative description of the spatial arrangement of object contours and component parts via internal symmetry axes (see Figure 1). Computer vision research has shown that such a description can be used to determine an object's shape from noisy or incomplete contour information (Feldman & Singh, 2006; Kimia, 2003), and to identify objects across viewpoint and category exemplars (Sebastian, Klein, & Kimia, 2004; Trinh & Kimia, 2011). Indeed, incorporating a skeletal model into off-the-shelf convolutional neural networks (CNNs) significantly improves their performance on visual perception tasks (Rezanejad et al., 2019). Similarly, behavioral research with humans has shown that participants extract the skeleton of 2D shapes (Firestone & Scholl, 2015; Kovács, Fehér, & Julesz, 1998; Psotka, 1978), even in the presence of border perturbations and illusory contours (Ayzenberg, Chen, Yousif, & Lourenco, 2019). Other research has shown that skeletal models are predictive of human object recognition (Destler, Singh, & Feldman, 2019; Lowet, Firestone, & Scholl, 2018; Wilder, Feldman, & Singh, 2011), even when controlling for other models of vision (Ayzenberg & Lourenco, 2019). Although these few studies, using vastly different methods, have provided independent evidence that skeletons are used for either perceptual organization or object recognition, no study has directly tested whether they are used for *both* functions. In the present study, we hypothesized a dual role for shape skeletons in human vision.

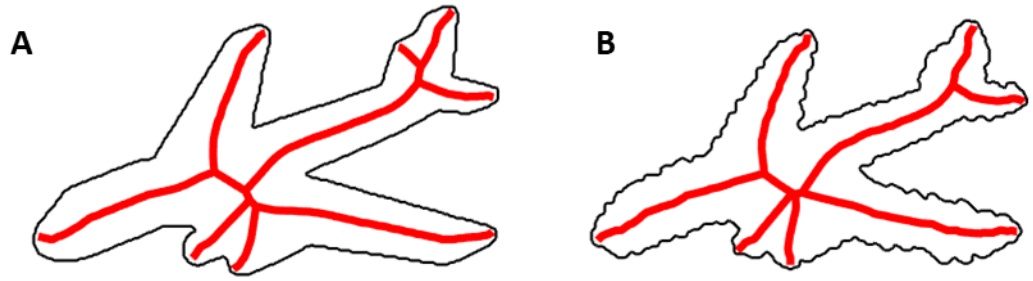

Figure 1. An illustration of the shape skeleton for a 2D airplane with (B) and without (A) perturbed contours. A strength of a skeletal model is that it can describe an object's shape structure across variations in contour. Skeletons computed using the ShapeToolbox (Feldman & Singh, 2006).

One method to address whether shape skeletons are used for both perceptual organization and object recognition is to test whether regions of the brain involved in these processes also represent the shape skeleton, without an explicit task. If shape skeletons are used to generate shape percepts, then they would be represented in early visual regions, such as V2-V4, which have been implicated in perceptual organization (Ardila, Mihalas, von der Heydt, & Niebur, 2012; Qiu & von der Heydt, 2007; Zhou, Friedman, & von der Heydt, 2000). If shape skeletons are also used for recognizing objects, then they should also be represented in object-selective visual regions, such as the lateral occipital cortex (LO) or posterior fusiform (pFs; Freud, Culham, Plaut, & Behrmann, 2017; Grill-Spector, Kourtzi, & Kanwisher, 2001; Grill-Spector, Kushnir, Edelman, Itzchak, & Malach, 1998), which have been implicated in object recognition.

We used representational similarity analysis (RSA) to test whether a model of skeletal similarity was predictive of the response patterns in early- (V2-V4) and higher- (LO, pFs) level visual areas. We also examined the specificity of the shape skeleton in these regions by controlling for other models of visual similarity that approximate early- (i.e., Gabor-jet; Margalit, Biederman, Herald, Yue, & von der Malsburg, 2016), mid- (i.e., GIST, and HMAX; Oliva & Torralba, 2006; Serre, Oliva, & Poggio, 2007), and high- (i.e., AlexNet-fc6; Krizhevsky, Sutskever, & Hinton, 2012) level visual processing. Finally, we ensured that the representation of shape skeletons could not be accounted for by lower-level shape properties (i.e., contours) by directly manipulating the object's contours while keeping the skeleton intact.

Materials and Methods

Participants

Twenty participants (M_{age} = 19.29 years, range = 20 – 36 years; 8 females) were recruited from the Emory University community. All participants gave written informed consent to participate and had normal or corrected-to-normal vision.

Stimuli

Twelve novel objects were selected from the stimulus set created by Ayzenberg and Lourenco (2019; see Figure 2A). The selected object set was composed of six distinct skeletons and two surface forms. The six skeletons were chosen by first conducting a *k*-means cluster analysis (k = 3) on skeletal similarity data for 30 unique objects (for details, see Ayzenberg & Lourenco, 2019). We selected six objects whose within- and between-cluster skeletal similarities were matched (2 per cluster). That is, the two objects from the same cluster were approximately as similar to one another as the two objects within the other clusters; objects in different clusters had comparable levels of dissimilarity to one another (see Figure 2B). This method of stimulus selection ensured that the stimulus set used in the present study contained objects with both similar and dissimilar skeletons. Each skeleton was also rendered with one of two surface forms, which changed the contours and component parts of the object without altering the underlying skeleton. To provide the strongest test of a skeletal model, we chose the two surface forms (out of five) that a separate group of participants judged to be most dissimilar (Ayzenberg & Lourenco, 2019). Importantly, the surface forms also had qualitatively different component parts (Biederman, 1987) and differed in their image-level properties (Margalit et al., 2016). Two additional objects were used as targets for an orthogonal target-detection task; these objects were not included in subsequent analyses.

Stimulus presentation was controlled by a MacBook Pro running the Psychophysics Toolbox package (Brainard, 1997) in MATLAB (MathWorks). Images were projected onto a screen and viewed through a mirror mounted on the head coil.

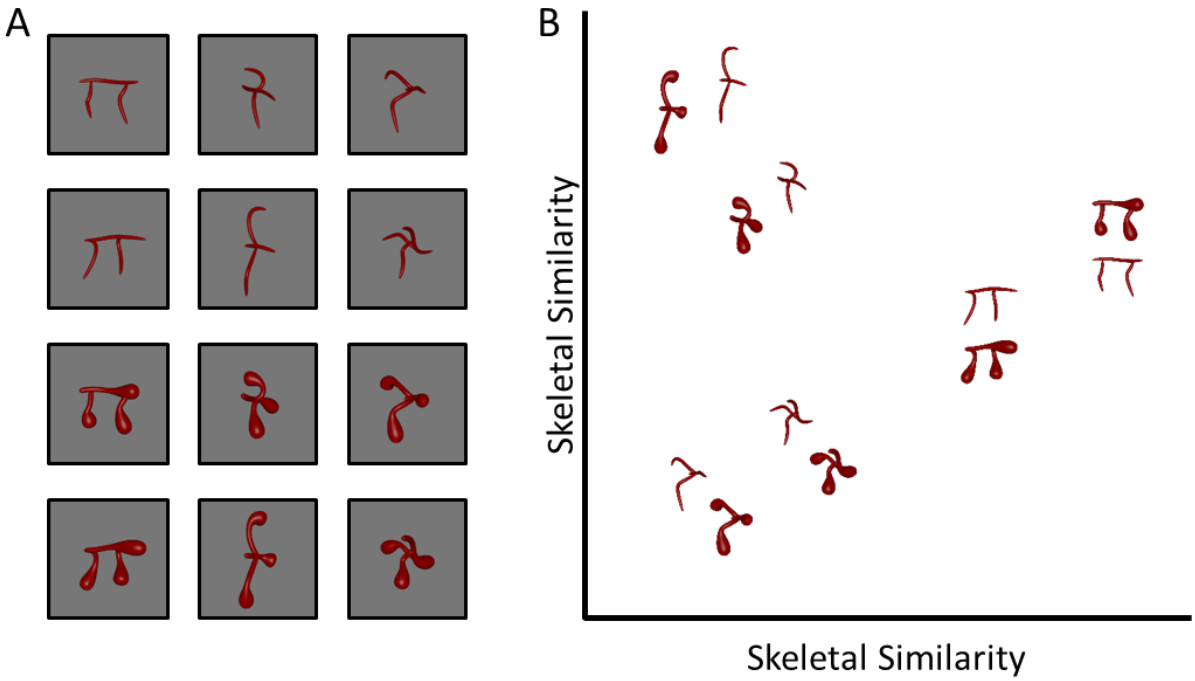

Figure 2. Stimuli used in the current experiment and a multi-dimensional scaling (MDS) plot illustrating the skeletal similarity between objects. (A) Six objects with unique skeletal structures were generated. Each object was rendered with two surface forms so as to change the objects' component parts without disrupting the skeleton. (B) To ensure that the stimulus set contained objects with both similar and dissimilar skeletons, objects were selected in pairs such that within- and between-pair skeletal similarity were approximately matched across objects.

Experimental design

First, we used a region of interest (ROI) approach, in which we independently localized our ROIs (localizer runs). Second, we used an independent set of data (experimental runs) to conduct representational similarity analyses in each ROI.

Localizer runs. We used a block design for the localizer runs. Participants viewed images of faces, bodies, objects, scenes, and scrambled objects, as previously described (Dilks, Julian, Kubilius, Spelke, & Kanwisher, 2011). Each participant completed three localizer runs, comprised of four blocks per stimulus category, each 400 s. Block order in each run was randomized. Each block contained 20 images randomly drawn from the same category. Each image was presented for 300 ms, followed by a 500 ms interstimulus interval (ISI), for a total of 16 s per block. We also included five 16 s fixation blocks: one at the beginning, three in the middle interleaved between each set of stimulus block, and one at the end of each run. To maintain attention, participants performed an orthogonal one-back task, responding to the repetition of an image on consecutive presentations.

Experimental runs. We used a continuous carry-over design for the experimental runs, wherein participants viewed images of each novel object. Each run was 360 s long. Using a de Bruijn sequence (Aguirre, Mattar, & Magis-Weinberg, 2011), we applied third-level counterbalancing on the image presentation order, which minimized any carry-over effects across the experiment. Each participant completed nine experimental runs. Each image was presented for 600 ms, followed by a 200 ms ISI. Each run began and ended with 6 s of fixation. To maintain attention, participants

performed an orthogonal target-detection task. At the beginning of each experimental run, participants were shown one of two objects (not included in subsequent analyses) and were instructed to press a response button any time the target object appeared within the image stream.

MRI scan parameters

Scanning was done on a 3T Siemens Trio scanner at the Facility for Education and Research in Neuroscience (FERN) at Emory University. Functional images were acquired using a 32-channel head matrix coil and a gradient echo single-shot echoplanar imaging sequence. Thirty slices were acquired for both localizer and experimental runs. For all runs: repetition time = 2 s; echo time = 30 ms; flip angle = 90°; voxel size = 1.8 × 1.8 × 1.8 mm with a 0.2 mm interslice gap. Slices were oriented approximately between perpendicular and parallel to the calcarine sulcus, covering the occipital and temporal lobes. Whole-brain, high-resolution T1-weighted anatomical images (repetition time = 1900 ms; echo time = 2.27 ms; inversion time = 900 ms; voxel size = 1 × 1 ×1 mm) were also acquired for each participant for registration of the functional images. Analyses of the fMRI data were conducted using FSL software (Smith et al., 2004) and custom MATLAB code.

Data Analysis

Images were skull-stripped (Smith, 2002) and registered to participants' T1 weighted anatomical image (Jenkinson et al., 2002). Prior to statistical analyses, images were motion corrected, de-trended, and intensity normalized. Localizer, but not experimental, data were spatially smoothed (6 mm kernel). All data were fit with a general linear model consisting of covariates that were convolved with a double-gamma function to approximate the hemodynamic response function.

To investigate whether skeletal descriptions of shape play a role in the creation of shape percepts, we defined visual regions V2, V3, and V4, bilaterally, using probabilistic parcels (Wang, Mruczek, Arcaro, & Kastner, 2014). As a control region, we also defined V1 bilaterally from the same set of probabilistic parcels. Each parcel was registered from MNI standard space to participants' individual anatomical space.

To investigate whether skeletal descriptions of shape also play a role in object recognition, we functionally defined object-selective regions LO and pFs bilaterally in each individual as the voxels that responded more to images of intact objects than scrambled objects ($p < 10^{-4}$, uncorrected; Grill-Spector et al., 1998). Furthermore, to test the specificity of skeletal representations in object-selective regions, rather than higher-level visual regions more generally, we also defined the extrastriate body area (EBA; Downing, Jiang, Shuman, & Kanwisher, 2001) and fusiform body area (FBA; Peelen & Downing, 2005), as the voxels that responded more to images of bodies than objects ($p < 10^{-4}$, uncorrected).

Analyses were conducted using the top 2000 voxels (1.8 × 1.8 × 1.8 mm) from each ROI (in each hemisphere) when available. For regions comprised of fewer than 2000 voxels, all voxels in the ROI were used. To ensure that results were not related to the size of the ROI, we also conducted our primary analyses using 100, 500, and 1000 voxels. The same qualitative results were found for all ROI sizes. For each functionally defined ROI, we selected voxels that exhibited the greatest selectivity to the category of interest from the localizer runs (e.g., the 2000 most object-selective voxels in right LO). For the probabilistically-defined ROIs, we selected voxels with the greatest probability value (e.g., the 2000 voxels most likely to describe right V1). ROIs were analyzed by combining left and right hemispheric ROIs (4000 voxels total).

To investigate whether a model of skeletal similarity explained unique variance in each ROI, we used RSA (Kriegeskorte, Mur, & Bandettini, 2008). For each participant, parameter estimates for each stimulus (relative to fixation) were extracted for each voxel in an ROI. Responses to the stimuli

in each voxel were then normalized by subtracting the mean response across all stimuli. A 12 × 12 symmetric neural representational dissimilarity matrix (RDM) was created for each ROI and participant by correlating (1-Pearson correlation) the voxel-wise responses for each stimulus with every other stimulus in a pairwise fashion. Neural RDMs were then Fisher transformed and averaged across participants separately for each ROI. Only the upper triangle of the resulting matrix (excluding the diagonal) was used in the following analyses. Neural RDMs were then compared to RDMs created from a model of skeletal similarity, as well as other models of visual similarity (GBJ, GIST, HMAX, and AlexNet-fc6). For a detailed description of each model, see Ayzenberg and Lourenco (2019).

Results

How are shape skeletons represented in the visual system?

We first tested whether skeletal similarity was predictive of the multivariate response pattern in each ROI by correlating the neural RDMs from each ROI with an RDM computed from a model of skeletal similarity. Significant correlations were found for V1-V4, and LO, *rs* = 0.35 – 0.67, R^2 = 12.5 – 50.1 (*ps* < .001; significance determined via permutation test with 10,000 permutations; see Figure 3). Skeletal similarity was not predictive of the response pattern in pFs, EBA, or FBA (*ps* > .23), revealing specificity in the predictive power of the skeletal model (see Table 1).

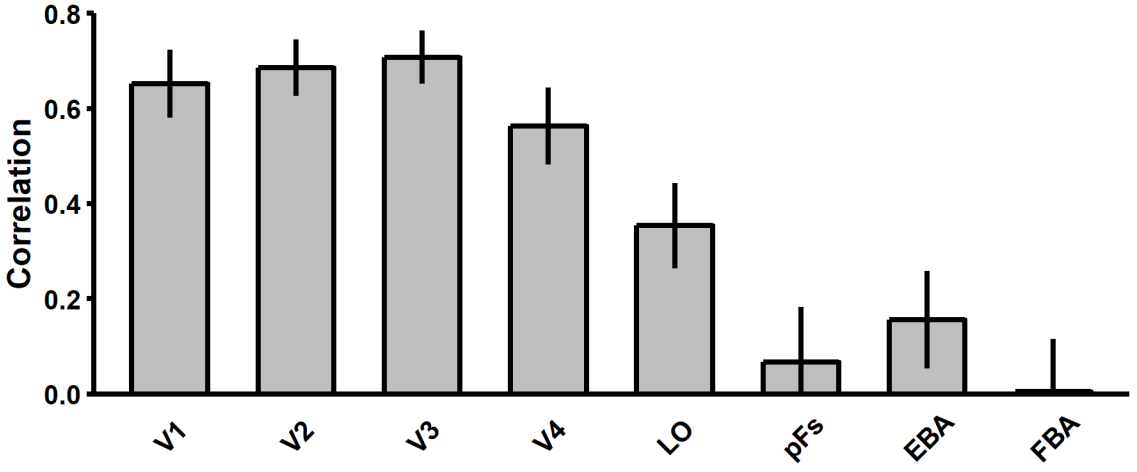

Figure 3. Bar plot displaying the correlations between the skeletal model and the multivariate response pattern in each ROI. A model of skeletal similarity was significantly correlated with response patterns in V1-V4 and LO. A skeletal model was not predictive of the response pattern in pFs, EBA, or FBA. Error bars represent bootstrapped SE.

Next, we tested whether skeletal similarity explained unique variance in each region, or whether these effects could be explained by another model of visual similarity (see Table 1 for correlations between the skeletal model and all other models). To test whether the skeletal model explained unique variance in each ROI, we conducted linear regression analyses with each neural RDM as the dependent variable, and the different models of visual similarity as predictors (Skeleton ∪ GBJ ∪ GIST ∪ HMAX ∪ AlexNet-fc6; see Figure 4A). These analyses revealed that the skeletal model explained unique variance in V3 (β = 0.46, p = .003) and LO (β = 0.49, p = .029), but not in the other regions (βs < 0.29, ps > .14).

We also conducted variance partitioning analyses to determine how much unique variance was explained by the skeletal model in V3 and LO (Bonner & Epstein, 2018; Lescroart, Stansbury, & Gallant, 2015). These analyses revealed that the skeletal model uniquely accounted for 9.0% of the total explainable variance in V3 and 25.5% of the explainable variance in LO (details about other models are provided below; see Figure 4B). Thus, shape skeletons account for significant unique variance in V3 and LO even when compared with other models of visual similarity. Together, these results are consistent with the hypothesized dual role of shape skeletons in visual processing: namely, for perceptual organization and object recognition.

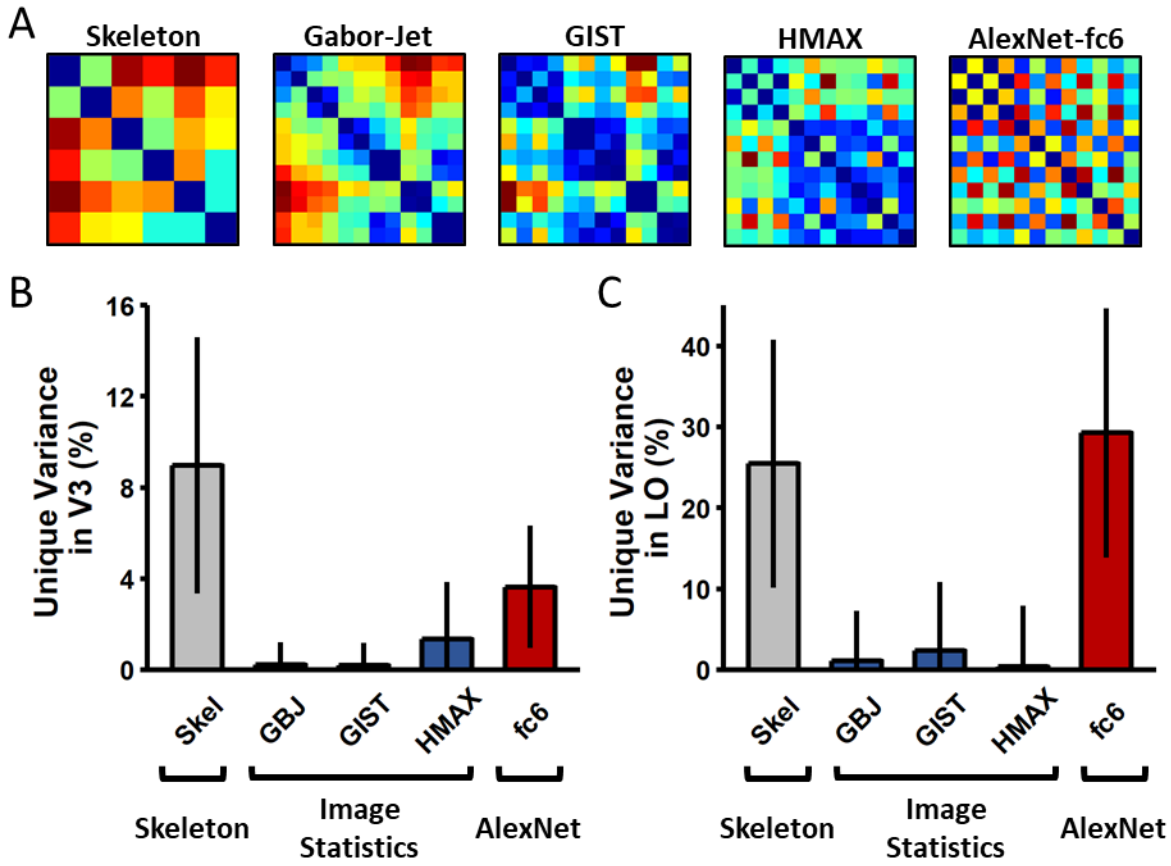

Figure 4. Variance partitioning results. (A) Dissimilarity matrices computed from models of skeletal similarity and other models of visual similarity. (B) Bar plot displaying the percentage of unique variance accounted for by each model in V3. (C) Bar plot displaying the percentage of unique variance accounted for by each model in LO. A model of skeletal similarity explained significant unique variance in both V3 and LO, but not other cortical regions. Error bars represent bootstrapped SE.

Table 1. Results of the correlation, regression, and variance partitioning analyses for each ROI and model. Correlation analyses were conducted by correlating (1-Pearson correlation) RDMs created from the neural data from each ROI with RDMs created from each model. Regression analyses were conducted for each neural RDM by entering each model RDM as a predictor into a linear regression model. The R^2 value indicates the total explained variance by all of the models. Variance partitioning analyses were conducted by iteratively regressing each neural RDM on RDMs from each model and combination of models, and then calculating the percentage of the total explained variance (R^2) uniquely explained by each model.

		Model Correlations			Model Regression			Variance Partitioning
ROI	Model	r	r^2	p	R^2	β	p	Percentage
V1		-	-	-	0.52	-	-	-
	Skeleton	0.65	0.43	< .001	-	0.26	.138	3.52
	Gabor-Jet	0.66	0.44	< .001	-	0.43	.070	5.27
	GIST	0.56	0.31	< .001	-	-0.09	.618	0.39
	HMAX	0.46	0.21	< .001	-	0.14	.297	1.72
	AlexNet-fc6	0.38	0.15	.002	-	0.13	.366	1.29
V2		-	-	-	0.63	-	-	-
	Skeleton	0.69	0.47	< .001	-	0.23	.140	2.19
	Gabor-Jet	0.71	0.51	< .001	-	0.55	.010	7.06
	GIST	0.60	0.36	< .001	-	-0.18	.275	1.19
	HMAX	0.53	0.28	< .001	-	0.13	.265	1.24
	AlexNet-fc6	0.48	0.23	< .001	-	0.24	.049	3.98
V3		-	-	-	0.63	-	-	-
	Skeleton	0.71	0.50	< .001	-	0.46	.003	8.97
	Gabor-Jet	0.67	0.44	< .001	-	0.10	.637	0.22
	GIST	0.62	0.39	< .001	-	0.07	.666	0.18
	HMAX	0.56	0.32	< .001	-	0.14	.239	1.36
	AlexNet-fc6	0.51	0.26	< .001	-	0.23	.056	3.63
V4		-	-	-	0.51	-	-	-
	Skeleton	0.56	0.32	< .001	-	0.18	.321	1.58
	Gabor-Jet	0.60	0.36	< .001	-	0.41	.085	4.83
	GIST	0.55	0.30	< .001	-	-0.11	.558	0.55
	HMAX	0.50	0.25	< .001	-	0.06	.678	0.27
	AlexNet-fc6	0.54	0.29	< .001	-	0.38	.009	11.53
LO		-	-	-	0.25	-	-	-
	Skeleton	0.35	0.13	.002	-	0.49	.029	25.46
	Gabor-Jet	0.25	0.06	.022	-	-0.13	.644	1.10
	GIST	0.23	0.05	.036	-	-0.16	.498	2.37
	HMAX	0.33	0.11	.004	-	-0.05	.776	0.42
	AlexNet-fc6	0.39	0.15	< .001	-	0.41	.020	29.27
pFs		-	-	-	0.04	-	-	-
	Skeleton	0.07	0.00	.294	-	0.29	.240	52.33
	Gabor-Jet	-0.02	0.00	.563	-	-0.08	.799	2.44
	GIST	-0.09	0.01	.761	-	-0.17	.505	16.71
	HMAX	-0.07	0.01	.727	-	-0.10	.590	10.90
	AlexNet-fc6	-0.10	0.01	.782	-	0.01	.959	0.10
EBA		-	-	-	0.07	-	-	-
	Skeleton	0.16	0.02	.107	-	0.26	.280	25.29
	Gabor-Jet	0.07	0.01	.287	-	0.04	.900	0.34
	GIST	0.00	0.00	.515	-	-0.30	.233	30.91
	HMAX	0.16	0.03	.096	-	0.05	.790	1.52
	AlexNet-fc6	0.12	0.02	.159	-	0.15	.428	13.57

FBA		-	-	-	0.06	-	-	-
	Skeleton	0.00	0.00	.484	-	-0.19	.442	16.68
	Gabor-Jet	0.04	0.00	.393	-	0.48	.142	61.84
	GIST	-0.06	0.00	.695	-	-0.42	.105	75.66
	HMAX	0.09	0.01	.242	-	0.12	.515	11.93
	AlexNet-fc6	0.05	0.00	.351	-	0.05	.785	2.09

Does skeletal coding in V3 and LO generalize across changes in surface form?

As described previously, a strength of skeletal models is that they can be used to describe an object's shape across variations in contours or component parts. Thus, if V3 and LO indeed incorporate a skeletal model, then these regions should represent objects by their skeletons across changes in surface form (see Figure 2). To test this prediction, new dissimilarity vectors were created from neural and model RDMs by extracting similarity values from only those object pairs whose surface forms differed and then correlating them to one another.

As predicted, skeletal similarity was a significant predictor of both V3 ($r = 0.77$, $p < .001$) and LO ($r = 0.47$, $p < .001$), even though object pairs were comprised of different surface forms. Notably, the finding that both V3 and LO represent shape skeletons across changes in surface form provides further evidence that skeletal coding in these regions cannot be accounted for by low-level shape properties such as contours and component parts.

But might another model of visual similarity account for these results? Here we conducted a similar regression analysis as above (neural RDMs ~ f[Skeleton ∪ GBJ ∪ GIST ∪ HMAX ∪ AlexNet-fc6]), but now included subject as the random effect because fewer object pairs were involved. This analysis revealed that the skeletal model was predictive in both V3 ($\beta = 0.28$, $p < .001$) and LO ($\beta = 0.21$, $p < .001$). Thus, not only are V3 and LO sensitive to object skeletons, the skeletal representations in these regions are invariant to changes in surface form.

What role do other models of visual similarity play in the visual processing of objects?

Although the skeletal model was predictive of the response profiles of V3 and LO, even across different surface forms, one might ask whether the other visual models still play a role in the neural processing of objects. For example, previous research has shown that the other models of visual similarity used here account for unique variance in participants' object similarity judgments, even when controlling for a skeletal model (Ayzenberg & Lourenco, 2019). To explore this possibility further, we tested whether these other models explained significant unique variance in the ROIs. Linear regression analyses revealed that the Gabor-jet model, which approximates V2-like complex cells, accounted for significant unique variance in the response profile of V2 ($\beta = 0.55$, $p = .009$), but not other regions. We also found that AlexNet-fc6, a model consisting of non-linear features, explained increasingly more variance in increasingly higher-level visual regions (V2: $\beta = 0.24$, $p = .049$; V4: $\beta = 0.38$, $p = .009$; LO: $\beta = 0.41$, $p = .020$). None of the models were predictive of the response profiles of V1, pFs, EBA, or FBA ($ps > .070$; see Table 1). Thus, the predictive power of these models of visual processing is largely consistent with the hypothesized regions they are meant to approximate.

General Discussion

In the current study, we tested the hypothesis that shape skeletons play a dual role in visual processing, both in the creation of shape percepts and object recognition. Consistent with this hypothesis, we found that a model of skeletal similarity was predictive of the response pattern in

both V3, a region implicated in perceptual organization, and LO, a region involved in object recognition. Moreover, and importantly, skeletal representations in these regions could not be explained by low-, mid-, or high-level image properties, as described by other computational models of vision, nor by representations based on contours or component parts (i.e., surface forms) of the objects. These results provide novel neural evidence that the human visual system represents shape skeletons and does so for both perceptual organization and object recognition.

The finding that V3 represents shape skeletons is consistent with the response profile of grouping cells (G-cells), which play an important role within neural models of perceptual organization. More specifically, these models suggest that perceptual organization is accomplished by border ownership cells (B-cells) in V2, which selectively respond to the contours of a figure (rather than the background), as well as G-cells in higher visual regions (V3-IT), which coordinate the firing of B-cells via top-down connections, and help specify the contours that belong to the same figure (von der Heydt, 2015; Zhou et al., 2000). Interestingly, G-cells exhibit many properties of shape skeletons. For example, G-cells specify the relations between contours, which may allow the visual system to determine an object's shape using noisy or incomplete visual information (Craft, Schütze, Niebur, & von der Heydt, 2007; Martin & von der Heydt, 2015). Likewise, the response profile of G-cells within a shape corresponds to the points of the shape's skeleton (Craft et al., 2007). Moreover, pruned shape skeletons, resembling those extracted from 2D shapes by human participants (Ayzenberg et al., 2019), can be generated using a model of perceptual organization that incorporates the response profile G-cells (Ardila et al., 2012). Importantly, these results are also consistent with human neuroimaging work showing that both symmetry structure (Keefe et al., 2018; Van Meel, Baeck, Gillebert, Wagemans, & Op de Beeck, 2019), as well as the spatial arrangement of object parts could be decoded in V3 (Lescroart & Biederman, 2013). Thus, our findings suggest that V3 may be the locus of G-cells, and that the skeletal representations within V3 may be an emergent property of G-cell responses, ultimately creating the shape percept.

We also found evidence of shape skeletons in LO, which is consistent with a role for skeletons in object recognition. Much work has illustrated the importance of LO in using shape information for object recognition (Chouinard, Whitwell, & Goodale, 2009; Grill-Spector, Kushnir, Hendler, & Malach, 2000). This region has been shown to be particularly sensitive to object-centered shape information and is tolerant to viewpoint changes and border perturbations (Grill-Spector et al., 2001; Grill-Spector et al., 1998). Our results suggest that LO may achieve such invariance by incorporating a skeletal description of shape, which provides a common format by which to compare shapes across variations in contours and component parts. Importantly, our results are consistent with electrophysiology work in monkeys in which the skeletal structure of 3D objects could be decoded from monkey IT across changes in both object orientation and surface form (Hung, Carlson, & Connor, 2012). They are also consistent with patient studies in which damage to LO results in a specific impairment perceiving the spatial relations of component parts, but not the parts themselves, as would be predicted by a skeletal model (Behrmann, Peterson, Moscovitch, & Suzuki, 2006; Konen, Behrmann, Nishimura, & Kastner, 2011). Building on these studies, the present work provides the first direct evidence of skeletal representations in human LO and, critically, demonstrates that such representations cannot be accounted for by other models of visual processing.

Interestingly, we found evidence of skeletal representations in LO, but not in another object-selective region, namely pFs. Though not predicted, this finding may reflect a division of labor between LO and pFs, following the posterior-to-anterior anatomical gradient of shape-to-category selectivity in the ventral stream (Bracci & Op de Beeck, 2016; Freud et al., 2017). More specifically, many studies have illustrated that shape selectivity peaks in posterior regions of the ventral stream and decreases in higher-level anterior regions (Brincat & Connor, 2004, 2006; Freud et al., 2017).

By contrast, sensitivity to semantic category-level information, and other non-shape visual information, progressively increases in anterior regions of the temporal lobe (Barense, Gaffan, & Graham, 2007; Behrmann, Lee, Geskin, Graham, & Barense, 2016). Given that skeletal models are exclusively descriptions of shape, such that they do not take semantic content, or other visual properties, into account, it follows that we did not find evidence of shape skeletons in pFs.

We found that a model of skeletal similarity was predictive of the response profile in V3 and LO even when controlling for low- (i.e., Gabor-jet), mid- (i.e., GIST, and HMAX), and high-level (i.e., AlexNet-fc6) models of visual processing, as well as across changes to the object's surface forms. Not only are these models representative of different levels of visual processing, but they also approximate different theories of object recognition, such as those based on image-level similarity (i.e., Gabor-jet and HMAX; Tarr & Bülthoff, 1998) and feature descriptions (i.e., AlexNet-fc6; Ullman, Assif, Fetaya, & Harari, 2016; Yamins et al., 2014). Moreover, by changing the object's surface forms, we changed the non-accidental properties of the object's component parts, thereby allowing for a test of component description theories (Biederman, 1987). That skeletal models explained unique variance even when controlling for these other properties suggests that shape skeletons may play a privileged role in the visual processing of objects and highlights their importance for theories of object recognition.

Yet our results also point to the contributions of two other models of vision to the neural processing of objects. We found that the Gabor-jet model was predictive of the response profile in V2 (earlier in the hierarchy than shape skeletons) and that AlexNet-fc6 was most predictive in LO (same region as skeletons). That these models may also play a role in object processing is perhaps unsurprising given that other visual properties are important for solving a range of object recognition tasks. For instance, lower-level visual properties and feature descriptions may be particularly important for subordinate- (Biederman, Subramaniam, Bar, Kalocsai, & Fiser, 1999; Davitt, Cristino, Wong, & Leek, 2014) or superordinate-level (Long, Störmer, & Alvarez, 2017; Long, Yu, & Konkle, 2018) categorization. Collectively, our findings illustrate how the visual system may incorporate multiple models in parallel to create holistic object representations.

In conclusion, our work highlights the unique contributions of shape skeletons to the neural processing of objects. A model of skeletal similarity was predictive of the response profile in both V3 and LO, consistent with a dual role for shape skeletons in supporting the creation of shape percepts and object recognition. Importantly, these findings not only enhance our understanding for how shapes and objects are represented during visual processing, but they also shed light on the computations implemented in V3 and LO. Lastly, our results underscore the importance of incorporating shape information, and skeletons in particular, into models of object recognition, which currently are not implemented by most state-of-the-art CNNs (Baker, Lu, Erlikhman, & Kellman, 2018; Geirhos et al., 2018).

References

Aguirre, G. K., Mattar, M. G., & Magis-Weinberg, L. (2011). de Bruijn cycles for neural decoding. *Neuroimage, 56*(3), 1293-1300. doi:https://doi.org/10.1016/j.neuroimage.2011.02.005

Ardila, D., Mihalas, S., von der Heydt, R., & Niebur, E. (2012). Medial axis generation in a model of perceptual organization. *Conference on Information Sciences and Systems (CISS)*, 1-4.

Ayzenberg, V., Chen, Y., Yousif, S. R., & Lourenco, S. F. (2019). Skeletal representations of shape in human vision: Evidence for a pruned medial axis model. *Journal of Vision, 19*(6), 1-21. doi:10.1167/19.6.6

Ayzenberg, V., & Lourenco, S. F. (2019). Skeletal descriptions of shape provide unique perceptual information for object recognition. *Scientific reports, 9*(1), 1-13. doi:10.1038/s41598-019-45268-y

Baker, N., Lu, H., Erlikhman, G., & Kellman, P. J. (2018). Deep convolutional networks do not classify based on global object shape. *PLOS Computational Biology, 14*(12), e1006613. doi:10.1371/journal.pcbi.1006613

Barense, M. D., Gaffan, D., & Graham, K. S. (2007). The human medial temporal lobe processes online representations of complex objects. *Neuropsychologia, 45*(13), 2963-2974.

Behrmann, M., Lee, A. C. H., Geskin, J. Z., Graham, K. S., & Barense, M. D. (2016). Temporal lobe contribution to perceptual function: A tale of three patient groups. *Neuropsychologia, 90*, 33-45. doi:https://doi.org/10.1016/j.neuropsychologia.2016.05.002

Behrmann, M., Peterson, M. A., Moscovitch, M., & Suzuki, S. (2006). Independent representation of parts and the relations between them: evidence from integrative agnosia. *Journal of Experimental Psychology: Human Perception and Performance, 32*(5), 1169-1184.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review, 94*(2), 115-147.

Biederman, I., Subramaniam, S., Bar, M., Kalocsai, P., & Fiser, J. (1999). Subordinate-level object classification reexamined. *Psychological research, 62*(2), 131-153. doi:10.1007/s004260050047

Blum, H., & Nagel, R. N. (1978). Shape description using weighted symmetric axis features. *Pattern Recognition, 10*(3), 167-180. doi:http://dx.doi.org/10.1016/0031-3203(78)90025-0

Bonner, M. F., & Epstein, R. A. (2018). Computational mechanisms underlying cortical responses to the affordance properties of visual scenes. *PLOS Computational Biology, 14*(4), e1006111. doi:10.1371/journal.pcbi.1006111

Bracci, S., & Op de Beeck, H. (2016). Dissociations and associations between shape and category representations in the two visual pathways. *Journal of Neuroscience, 36*(2), 432-444.

Brincat, S. L., & Connor, C. E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature Neuroscience, 7*, 880. doi:10.1038/nn1278

https://www.nature.com/articles/nn1278#supplementary-information

Brincat, S. L., & Connor, C. E. (2006). Dynamic Shape Synthesis in Posterior Inferotemporal Cortex. *Neuron, 49*(1), 17-24. doi:https://doi.org/10.1016/j.neuron.2005.11.026

Chouinard, P. A., Whitwell, R. L., & Goodale, M. A. (2009). The lateral-occipital and the inferior-frontal cortex play different roles during the naming of visually presented objects. *Human Brain Mapping, 30*(12), 3851-3864. doi:doi:10.1002/hbm.20812

Craft, E., Schütze, H., Niebur, E., & von der Heydt, R. (2007). A Neural Model of Figure–Ground Organization. *Journal of Neurophysiology, 97*(6), 4310-4326. doi:10.1152/jn.00203.2007

Davitt, L. I., Cristino, F., Wong, A. C. N., & Leek, E. C. (2014). Shape information mediating basic- and subordinate-level object recognition revealed by analyses of eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 40*(2), 451-456. doi:10.1037/a0034983

Destler, N., Singh, M., & Feldman, J. (2019). Shape discrimination along morph-spaces. *Vision Research, 158*, 189-199.

Dilks, D. D., Julian, J. B., Kubilius, J., Spelke, E. S., & Kanwisher, N. (2011). Mirror-image sensitivity and invariance in object and scene processing pathways. *The Journal of Neuroscience, 31*(31), 11305-11312.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A Cortical Area Selective for Visual Processing of the Human Body. *Science, 293*(5539), 2470-2473. doi:10.1126/science.1063414

Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. *Proceedings of the National Academy of Sciences, 103*(47), 18014-18019.

Firestone, C., & Scholl, B. (2015). Can you simultaneously represent a figure as both an object and an open contour? Hybrid shape representations revealed by the "tap-the-shape" task. *Journal of Vision, 15*(12), 1125-1125. doi:10.1167/15.12.1125

Firestone, C., & Scholl, B. J. (2014). "Please tap the shape, anywhere you like" shape skeletons in human vision revealed by an exceedingly simple measure. *Psychological Science, 25*(2), 377-386.

Freud, E., Culham, J. C., Plaut, D. C., & Behrmann, M. (2017). The large-scale organization of shape processing in the ventral and dorsal pathways. *Elife, 6*, e27576.

Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., & Brendel, W. (2018). ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv*.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research, 41*(10), 1409-1422.

Grill-Spector, K., Kushnir, T., Edelman, S., Itzchak, Y., & Malach, R. (1998). Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron, 21*(1), 191-202.

Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nature Neuroscience, 3*(8), 837-843.

Hung, C.-C., Carlson, E. T., & Connor, C. E. (2012). Medial axis shape coding in macaque inferotemporal cortex. *Neuron, 74*(6), 1099-1113.

Keefe, B. D., Gouws, A. D., Sheldon, A. A., Vernon, R. J., Lawrence, S. J., McKeefry, D. J., . . . Morland, A. B. (2018). Emergence of symmetry selectivity in the visual areas of the human brain: fMRI responses to symmetry presented in both frontoparallel and slanted planes. *Human Brain Mapping, 39*(10), 3813-3826.

Kimia, B. B. (2003). On the role of medial geometry in human vision. *Journal of Physiology-Paris, 97*(2), 155-190.

Konen, Christina S., Behrmann, M., Nishimura, M., & Kastner, S. (2011). The functional neuroanatomy of object agnosia: A case study. *Neuron, 71*(1), 49-60.

Kovács, I., Fehér, Á., & Julesz, B. (1998). Medial-point description of shape: A representation for action coding and its psychophysical correlates. *Vision Research, 38*(15), 2323-2333.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience, 2*(4). doi:10.3389/neuro.06.004.2008

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 1097-1105.

Lescroart, M. D., & Biederman, I. (2013). Cortical representation of medial axis structure. *Cerebral Cortex, 23*(3), 629-637.

Lescroart, M. D., Stansbury, D. E., & Gallant, J. L. (2015). Fourier power, subjective distance, and object categories all provide plausible models of BOLD responses in scene-selective visual

areas. *Frontiers in Computational Neuroscience, 9*(135), 1-20. doi:10.3389/fncom.2015.00135

Long, B., Störmer, V. S., & Alvarez, G. A. (2017). Mid-level perceptual features contain early cues to animacy. *Journal of Vision, 17*(6), 1-20. doi:10.1167/17.6.20

Long, B., Yu, C.-P., & Konkle, T. (2018). Mid-level visual features underlie the high-level categorical organization of the ventral stream. *Proceedings of the National Academy of Sciences, 115*(38), E9015-E9024. doi:10.1073/pnas.1719616115

Lowet, A. S., Firestone, C., & Scholl, B. J. (2018). Seeing structure: Shape skeletons modulate perceived similarity. *Attention, Perception, & Psychophysics, 80*(5), 1278-1289. doi:10.3758/s13414-017-1457-8

Margalit, E., Biederman, I., Herald, S. B., Yue, X., & von der Malsburg, C. (2016). An applet for the Gabor similarity scaling of the differences between complex stimuli. *Attention, Perception, & Psychophysics, 78*(8), 2298-2306. doi:10.3758/s13414-016-1191-7

Martin, A. B., & von der Heydt, R. (2015). Spike synchrony reveals emergence of proto-objects in visual cortex. *Journal of Neuroscience, 35*(17), 6860-6870.

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.

Peelen, M. V., & Downing, P. E. (2005). Selectivity for the Human Body in the Fusiform Gyrus. *Journal of Neurophysiology, 93*(1), 603-608. doi:10.1152/jn.00513.2004

Psotka, J. (1978). Perceptual processes that may create stick figures and balance. *Journal of Experimental Psychology: Human Perception and Performance, 4*(1), 101-111.

Qiu, F. T., & von der Heydt, R. (2007). Neural representation of transparent overlay. *Nature Neuroscience, 10*, 283. doi:10.1038/nn1853

https://www.nature.com/articles/nn1853#supplementary-information

Rezanejad, M., Downs, G., Wilder, J., Walther, D. B., Jepson, A., Dickinson, S., & Siddiqi, K. (2019). *Scene Categorization from Contours: Medial Axis Based Salience Measures.* Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Sebastian, T. B., Klein, P. N., & Kimia, B. B. (2004). Recognition of shapes by editing their shock graphs. *IEEE Transactions on pattern Analysis and Machine Intelligence, 26*(5), 550-571.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences, 104*(15), 6424-6429. doi:10.1073/pnas.0700622104

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., . . . Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage, 23*, S208-S219. doi:https://doi.org/10.1016/j.neuroimage.2004.07.051

Tarr, M. J., & Bülthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition, 67*(1), 1-20. doi:http://dx.doi.org/10.1016/S0010-0277(98)00026-2

Trinh, N. H., & Kimia, B. B. (2011). Skeleton search: Category-specific object recognition and segmentation using a skeletal shape model. *International Journal of Computer Vision, 94*(2), 215-240.

Ullman, S., Assif, L., Fetaya, E., & Harari, D. (2016). Atoms of recognition in human and computer vision. *Proceedings of the National Academy of Sciences, 113*(10), 2744-2749.

Van Meel, C., Baeck, A., Gillebert, C. R., Wagemans, J., & Op de Beeck, H. P. (2019). The representation of symmetry in multi-voxel response patterns and functional connectivity throughout the ventral visual stream. *Neuroimage, 191*, 216-224.

von der Heydt, R. (2015). Figure–ground organization and the emergence of proto-objects in the visual cortex. *Frontiers in Psychology, 6*(1695). doi:10.3389/fpsyg.2015.01695

Wang, L., Mruczek, R. E. B., Arcaro, M. J., & Kastner, S. (2014). Probabilistic Maps of Visual Topography in Human Cortex. *Cerebral Cortex, 25*(10), 3911-3931. doi:10.1093/cercor/bhu277

Wilder, J., Feldman, J., & Singh, M. (2011). Superordinate shape classification using natural shape statistics. *Cognition, 119*(3), 325-340. doi:10.1016/j.cognition.2011.01.009

Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences, 111*(23), 8619-8624.

Zhou, H., Friedman, H. S., & von der Heydt, R. (2000). Coding of Border Ownership in Monkey Visual Cortex. *The Journal of Neuroscience, 20*(17), 6594-6611. doi:10.1523/jneurosci.20-17-06594.2000