HHS Public Access

Author manuscript

Neuropsychologia. Author manuscript; available in PMC 2023 January 14.

Published in final edited form as:

Neuropsychologia. 2022 January 07; 164: 108092. doi:10.1016/j.neuropsychologia.2021.108092.

Skeletal representations of shape in the human visual cortex

Vladislav Ayzenberg[a,*], **Frederik S. Kamps**[b], **Daniel D. Dilks**[c], **Stella F. Lourenco**[c,**]

[a]Department of Psychology, Carnegie Mellon University, USA

[b]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, USA

[c]Department of Psychology, Emory University, USA

Abstract

Shape perception is crucial for object recognition. However, it remains unknown exactly how shape information is represented and used by the visual system. Here, we tested the hypothesis that the visual system represents object shape via a skeletal structure. 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. 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 shape processing in human vision, as well as the computational properties of V3 and LO. We discuss how these regions may support two putative roles of shape skeletons: namely, perceptual organization and object recognition.

Keywords

fMRI; Perceptual organization; Object recognition; Medial axis; V3; Lateral occipital cortex (LO)

1. Introduction

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.”

*Corresponding author: vayzenbe@andrew.cmu.edu (V. Ayzenberg). **Corresponding author: stella.lourenco@emory.edu (S.F. Lourenco).

Declaration of competing interest

The authors declare no conflicts of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuropsychologia.2021.108092.

Credit Statement

V.A. performed the experiments and analyzed the data. All authors conceived and designed the experiments, interpreted the experimental results, and approved the submission.

Although recent behavioral studies suggest that humans also represent shape skeletons (Ayzenberg and Lourenco, 2019; Firestone and Scholl, 2014), few studies have explored how they are represented neurally.

Shape skeletons are models of structure based on the medial axis of an object (Blum and Nagel, 1978). They provide a quantitative description of the spatial arrangement of object contours and component parts via internal symmetry axes (see Fig. 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 and Singh, 2006; Kimia, 2003; Wilder et al., 2019) and to identify objects across viewpoints and category exemplars (Sebastian et al., 2004; Trinh and 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 and Scholl, 2015; Kovács et al., 1998; Psotka, 1978), even in the presence of border perturbations and illusory contours (Ayzenberg et al., 2019). Other research has shown that skeletal models are predictive of human object recognition (Destler et al., 2019; Lowet et al., 2018; Wilder et al., 2011), even when controlling for other models of vision (Ayzenberg and Lourenco, 2019). Thus, shape skeletons potentially play an important role in shape perception and object recognition, but the neural correlates of shape skeletons remain poorly understood.

In one study using fMRI with humans, Lescroart and Biederman (2012) decoded skeletal structures from objects in V3 and LO, regions involved in constructing shape percepts (Caplovitz and Peter, 2010; Mannion et al., 2010) and recognizing objects (Grill-Spector et al., 2001). However, this study did not measure skeletal coding directly, leaving it unknown whether another representation of structure could account for their results, nor did it compare skeletons to other models of vision, a crucial analysis because changes to object skeletons also induce changes along other visual dimensions. Thus, the question remains: how are shape skeletons represented in human cortex?

To address this question, we created a novel set of objects that allowed us to systematically vary object skeletons and directly measure skeletal coding. We then examined the unique contributions of skeletal information to neural responses across the visual hierarchy (V1–V4, LO, pFs). More specifically, we used representational similarity analysis (RSA) to test whether a model of skeletal similarity predicted the response patterns in these regions while controlling for other models of visual similarity that do not represent the shape skeleton, but approximate other aspects of visual processing. In particular, we included models that approximate early- (i.e., Gabor-jet; Margalit, Biederman, Herald, Yue, & von der Malsburg, 2016), mid- (i.e., GIST and HMAX; Oliva and Torralba, 2006; Serre et al., 2007), and high-level (i.e., AlexNet; Krizhevsky, Sutskever and Hinton, 2012) visual processing. We also examined the robustness of skeletal representations in each region by testing whether a model of skeletal similarity generalizes across a change to the object's component parts, which alters the non-accidental and image-level properties of the objects. Together, these comparisons and manipulations allowed us to directly measure skeletal coding in regions typically associated with perceptual organization and object recognition while ruling out alternative explanations.

2. Materials and methods

2.1. 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. The sample size was selected on the basis of prior work, which has typically tested between 15 and 25 participants (Dwivedi et al., 2021; Kamps et al., 2019; Magri et al., 2021). A power-analysis confirmed that 20 participants was sufficient to achieve reliable similarity metrics for each pair of objects (r = 0.60, 1 − β = 0.85; Bonett, 2002). Experimental procedures were approved by Emory University's Institutional Review Board (IRB). All experiments were performed in accordance with the relevant guidelines and regulations of the IRB.

3. Stimuli

Twelve novel objects were selected from the stimulus set created by Ayzenberg and Lourenco (2019; see Fig. 2A). The selected object set was composed of six distinct skeletons and two surface forms. Objects were procedurally generated using the Python API for Blender (Blender Foundation). Each skeleton was comprised of three segments created from Bezier curves of a random size and curvature scaled between 0.05 and 0.25 virtual Blender units (vu). The first axis segment was oriented forward towards the 'camera'. The second and third segments were oriented perpendicular to the first segment and attached to the first segment or second segment at a random point along their length. Each skeleton was rendered with one of two surface forms, which changed the contours and component parts of the object without altering the underlying skeleton. Surface forms were created by applying a circular bevel to the object's skeleton along with one of five taper properties that determined the shape of the surface form. Finally, the overall size of the object was normalized to 0.25 vu.

Skeletal similarity was calculated as the mean Euclidean distance between each point on one skeleton structure with the closest point on the second skeleton structure following maximal alignment. Maximal alignment was achieved by overlaying each structure by its center of mass and then iteratively rotating each object in the picture plane orientation by 15° until the smallest distance between the structures was found.

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 and 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 Fig. 2B). This method of stimulus selection ensured that the stimulus set used in the present study contained objects with both similar and dissimilar skeletons. Two additional objects were used as targets for an orthogonal target-detection task; these objects were not included in subsequent analyses.

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 (see Supplemental Materials). Importantly, we ensured that the surface forms were as discriminable as the skeletal structures (t [78] = 1.47, p = .146; see Supplemental Materials). These surface forms were also chosen because they had qualitatively different component parts, as measured by human ratings of non-accidental properties (Amir et al., 2012; Biederman, 1987; see Supplemental Materials for details). Thus, the surface forms produced large changes to the objects as measured by participants' discrimination judgments and their ratings of the non-accidental properties.

The number of objects in the current study was determined to ensure a sufficient number of presentations per object and to maximize the signal-to-noise ratio. It also allowed us to implement a continuous carry-over design with third-order counterbalancing, thereby minimizing carry-over effects across trials (Aguirre, Mattar and Magis-Weinberg, 2011).

Finally, given that our objects were artificial and might appear to have especially salient skeletons, we tested whether they could be discriminated using visual properties other than the shape skeleton and whether they evoked visual processes similar to real-world objects. We found that all non-skeletal models (GBJ, GIST, HMAX, AlexNet) could accurately discriminate objects with different skeletons (80.2%–95.3% accuracy), suggesting that skeletons were not the only available visual property and our objects differed along other visual dimensions (see Supplemental Materials for more details). Next, we tested whether participants could discriminate these stimuli in a speeded context (100 ms presentation), a task thought to evoke 'core' object perception (Cadieu et al., 2014; DiCarlo et al., 2012; see Supplemental Materials for details). This experiment revealed that participants were significantly above chance at discriminating these objects (M = 89.9%, t (13) = 14.02, p < .001; see Supplemental Materials). Thus, although we used artificial 3D objects designed to vary in skeletal similarity, they could also be recognized using visual properties other than the skeleton and they evoke processes typical of core object perception.

3.1. Experimental design

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

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 et al., 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 inter-stimulus 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 blocks, 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 et al., 2011), we applied third-order counterbalancing on the image presentation order, which minimized any carry-over effects between stimuli. Importantly, this design supports smaller inter-stimulus intervals (ISIs) between stimuli (Aguirre et al., 2011; Drucker and Aguirre, 2009; Hatfield et al., 2016) and allowed for a greater number of presentations per image. Each image was presented for 600 ms, followed by a 200 ms ISI, and shown 225 times across the entire session. 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 each time the target object appeared within the image stream. Each participant completed a total of nine experimental runs.

3.2. MRI scan parameters

Scanning was done on a 3 T 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 parallel to the anterior and posterior cingulate, 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.

3.3. Data analyses

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.

We defined regions V1–V4 bilaterally using probabilistic parcels created from a large sample of participants (see Wang et al., 2014 for details). Each parcel was registered from MNI standard space to participants' individual anatomical space.

We also functionally defined object-selective region LO, as well as 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 defined the extrastriate body area (EBA; Downing et al., 2001) and fusiform body area (FBA;Peelen and Downing, 2005), as the voxels that responded more to images of bodies than objects ($p < 10^{-4}$, uncorrected). However, because EBA shows a high degree of overlap with LO, we subtracted any EBA voxels that overlapped with LO for each participant. The same qualitative results were found when LO and EBA were not subtracted. The anatomical consistency of each participant's ROIs was confirmed using probabilistic parcels functionally defined using a large sample of participants (see Julian et al., 2012 for details). We further ensured that functional ROIs were reliable for each participant by computing the average distance of peak voxels in each run from the peak voxel in data averaged across runs. This analysis revealed significant overlap such that, on average, peak voxels varied by only 5.66 mm across runs (~3 voxels; SD = 4.85 mm).

To ensure that our results were not influenced by different numbers of voxels in each ROI, we conducted analyses using the top 2000 voxels (1.8 × 1.8 × 1.8 mm) from each ROI (in each hemisphere) when available. For regions composed of fewer than 2000 voxels, all voxels in the ROI were used (see Fig. 3). To further 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).

In subsequent analyses, we used RSA to investigate the extent to which a model of skeletal similarity explained unique variance in each ROI (Kriegeskorte et al., 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 (see Fig. 4A). Only the upper triangle of the resulting matrix (excluding the diagonal) was used in subsequent analyses. Although most dissimilarity measures produce similar results, we used Pearson correlation similarity because simulations have shown it to be more reliable than other similarity measures (Walther et al., 2016).

Neural RDMs were compared to RDMs created from a model of skeletal similarity, as well as other models of visual similarity (GBJ, GIST, HMAX, and AlexNet; see Fig. 4B). As described previously, skeletal similarity was calculated in 3D, object-centered, space as the mean Euclidean distance between each point on one skeleton and the closest point on the second skeleton following maximal alignment. Similarity for Gabor-jet, GIST, and AlexNet (each layer) was calculated by extracting feature vectors from each model and computing

the mean Euclidean distance between feature vectors. HMAX (C2-layer) similarity was calculated as the Pearson correlation between feature vectors. Because our primary analyses involve comparing the amount of unique variance explained by the skeletal model relative to the other models, we ensured that the skeletal model did not exhibit a high degree of multicollinearity with any other model, VIF = 2.65. Multicollinearity statistics for control models (GBJ, GIST, HMAX, AlexNet) were also within an acceptable range (VIFs <4.60; O'Brien, 2007).

4. 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 the skeletal model. Significant correlations were found for V1–V4, and LO, rs = 0.35–0.67, r^2 = 0.13–0.50 (ps < .004; significance determined via permutation test with 10,000 permutations; see Fig. 5). 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 for correlations between the ROIs and all other models).

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. 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 U GBJ U GIST U HMAX U AlexNet; see Fig. 4B). Only the AlexNet layer that best correlated with each ROI was included in the regression. These analyses revealed that the skeletal model explained unique variance in V3 (β = 0.38, p = .016) and LO (β = 0.51, p = .024), but not in the other regions (βs > 0.32, ps < .186; the results of the other models is described separately below). Moreover, to examine the specificity of the skeletal model to these regions, we swapped the independent and dependent variables and tested whether V3 and LO explained significant variance in the skeletal similarity between objects when controlling for the other regions. Separate regression analyses were conducted in which the predictors were either V3 and the other early visual regions (V1, V2, V4) or LO and the other high-level visual regions (pFs, EBA, FBA). The skeletal RDM was the dependent variable in both cases. These analyses revealed that V3 (β = 0.60, p = .040) and LO (β = 0.68, p = .002) explained unique variance in skeletal similarity, even when controlling for other early- and high-level visual regions, respectively.

We also conducted variance partitioning analyses (VPA) to explore how much unique variance was explained by the skeletal model in V3 and LO relative to the other models (Bonner and Epstein, 2018; Lescroart et al., 2015). VPA, also known as commonality analyses, estimates how much of the total variance explained (R^2) is unique to the different models (e.g., $r^2_{\text{skeleton}} = R^2 - r^2_{\text{GBJ U GIST U HMAX U AlexNet}}$). Each value is expressed as a percentage of the total variance explained by all of the models ($r^2_{skeleton} \div R^2$) × 100. These analyses revealed that the skeletal model uniquely accounted for 6.0% of the total explainable variance in V3 and 27.4% of the explainable variance in LO (see Fig. 6A and

B). Thus, shape skeletons account for significant unique variance in V3 and LO even when compared with other models of visual similarity.

Several follow-up analyses were conducted to ensure that the significant variance explained in V3 and LO by the skeletal model was not due to ROI or model selection decisions. First, as mentioned previously, we analyzed each ROI using 100, 500, and 1000 voxels and found qualitatively similar results (see Supplemental Data). We also evaluated different variations of our functionally-defined LO ROI. We found that the skeletal model continued to explain significant unique variance (relative to other models of visual similarity) when LO was split into LO1 (VPA = 7.7%, $\beta = 0.42$, $p = .006$) and LO2 (VPA = 13.0%, $\beta = 0.48$, $p = .009$), using probabilistic parcels (Wang et al., 2014), and when LO was functionally defined using an object > scene contrast ("LO_{scene}"; VPA = 28.3%, $\beta = 0.58$, $p = .008$).

Next, we ensured that our results were not specific to the choice of CNN (i.e., AlexNet). To this end, we replicated our analysis with another CNN, namely VGG-19 (Simonyan and Zisserman, 2014), using the layer best correlated with each ROI within the regression analysis. We found qualitatively similar results such that the skeletal model explained unique variance in V3 (VPA = 6.6%, $\beta = 0.39$, $p = .013$) and LO (VPA = 14%, $\beta = 0.34$, $p = .127$), though the effect in LO did not reach the criteria for statistical significance. However, the skeletal model explained significant unique variance in LO1, LO2, and LO_{scene} (VPAs >8%, $\beta s > 0.43$, $ps < .009$), suggesting consistency of results across different types of neural network models in LO.

Finally, we tested whether the success of the skeletal model might be explained by the fact that it has access to 3D object properties, whereas the other models do not. To this end, we reanalyzed our data using a 2D implementation of the skeletal model, which computed the medial axis from the object's silhouette (Rezanejad and Siddiqi, 2013); we found that even a 2D skeleton continued to explain unique variance in V3 (VPA = 10.7%, $\beta = 0.39$, $p < .001$) and LO (VPA = 28.42%, $\beta s = 0.39$, $ps = .020$), as well as LO1, LO2, and LO_{scene} (VPAs >11.7%, $\beta s > 0.34$, $ps < .043$), suggesting that access to 3D information cannot explain our results.

To further explore how skeletal structure is represented neurally, we conducted a whole-brain searchlight analysis examining where the 3D skeletal model explained unique variance (Kriegeskorte et al., 2006). For each participant, an 8 mm sphere was centered on every voxel within the slice prescription and a neural RDM was constructed. Each central voxel was assigned the standardized parameter estimate for the skeletal model ($\beta_{skeleton}$) from a regression analysis with each neural RDM as the dependent variable and the different models of visual similarity as predictors (Skeleton U GBJ U GIST U HMAX U AlexNet). The best correlated layer of AlexNet was used in each search iteration. The resulting map from each participant was then registered to a MNI space, averaged together, and normalized into a single FDR-corrected ($p < .05$) map. As can be seen in Fig. 7, the skeletal model explained unique variance in regions corresponding to V3 and LO (LO1/LO2), with a particularly large cluster centered around left V3 (see Supplemental Materials for searchlight results with other models). Together, these findings suggest that skeletal structure

is represented in early visual regions involved in creating shape percepts, such as V3, as well as higher-level regions involved in object recognition, such as LO.

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 Fig. 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.

Skeletal similarity was a significant predictor of the response profile in 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 that conducted above (neural RDMs ~ f[Skeleton ∪ GBJ ∪ GIST ∪ HMAX ∪ AlexNet]) but now including subject as the random effect because fewer object pairs were involved. This analysis revealed that the skeletal model explained the greatest amount of variance in both V3 ($\beta = 0.24$, $p < .001$) and LO ($\beta = 0.28$, $p < .001$; see Supplemental Table 1 for variance explained by the other models). Thus, not only are V3 and LO sensitive to object skeletons, but these skeletal representations in these regions are also invariant to changes in surface form.

Are V3 and LO predictive of participants' similarity judgments of objects?

Previous research has shown that shape skeletons are predictive of human participants' behavioral judgments of object similarity (Ayzenberg and Lourenco, 2019; Destler et al., 2019; Lowet et al., 2018). Our neuroimaging results suggest that these judgments may be supported by areas V3 and LO. Here we directly tested this possibility by examining whether the response patterns of V3 and LO explain unique variance in humans' judgments of object similarity.

Behavioral RDMs for the present objects were generated using discrimination data from Ayzenberg and Lourenco (2019). Using linear regression analyses, we first tested whether a model of skeletal similarity explained unique variance in behavioral judgments, after controlling for other models of visual similarity (GBJ, GIST, HMAX, AlexNet). We found that the skeletal model explained the greatest amount of variance in participants' judgments (VPA = 20.5%, $\beta = 0.85$, $p < .001$), replicating Ayzenberg and Lourenco (2019; see Fig. 2C). Next, we tested whether the response profiles of V3 and LO were also predictive of the behavioral RDM. These analyses revealed significant correlations for both regions and participants' judgments: V3, $r = 0.81$, $p < .001$; LO, $r = 0.46$, $p < .001$.

In a final analysis, we examined the specificity of V3 and LO in explaining participants' behavioral judgments by testing whether another region could explain this effect. We tested whether V3 and LO explained unique variance in participants' judgments by conducting separate regression analyses in which the predictors were either V3 and the other early visual regions (V1, V2, V4) or LO and the other high-level visual regions (pFs, EBA, FBA). The behavioral RDM was the dependent variable in both cases. These analyses revealed that V3 (VPA = 10%, $\beta = 0.83$, $p = .002$) and LO (VPA = 70%, $\beta = 0.70$, $p < .001$) explained unique variance in participants' similarity judgments, even when controlling for other early- and high-level visual regions, respectively.

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 what role the other models of visual perception play in the neural processing of objects. For example, previous research has shown that other models of visual similarity account for unique variance in participants' object similarity judgments (Ayzenberg and Lourenco, 2019; Cadieu et al., 2014; Schrimpf et al., 2018). Linear regression analyses revealed that the Gabor-jet model, which approximates early visual neurons, accounted for unique variance in the response profile of V2 ($\beta = 0.44$, $p = .031$), but not other regions. Likewise, HMAX, which may approximate complex cells in extrastriate visual cortex, explained significant variance in V2 ($\beta = 0.24$, $p = .009$), V3 ($\beta = 0.26$, $p = .007$), and V4 ($\beta = 0.23$, $p = .027$). We also found that different layers of AlexNet explained significant variance in the ROIs—namely, AlexNet-conv5 explained significant variance in V2 ($\beta = 0.35$, $p = .024$) and V4 ($\beta = 0.46$, $p = .011$) and AlexNet-fc7 explained significant variance in LO ($\beta = 0.35$, $p = .020$). 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.

5. General discussion

In the present study, we examined how shape skeletons are represented neurally. We found that a model of skeletal similarity was predictive of the response pattern in V3 and LO. Moreover, and crucially, 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 for the processing of shape skeletons in V3 and LO.

The finding of skeletal processing in V3 is consistent with human neuroimaging studies showing its involvement in perceptual organization (Sasaki, 2007). Indeed, V3 has been consistently implicated in creating shape percepts (Caplovitz et al., 2008; McMains and Kastner, 2010; Montaser-Kouhsari et al., 2007) and is the earliest stage of the visual hierarchy where symmetry structure has been decoded (Keefe et al., 2018; Sasaki et al., 2005; Van Meel, Baeck, Gillebert, Wagemans and Op de Beeck, 2019). But how might shape skeletons arise in V3? One possibility is that shape skeletons reflect 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 the subsequent visual region (e.g., V3), which coordinate the firing of B-cells via top-down connections and help to specify the contours that belong to the same figure (von der Heydt, 2015; Zhou, Friedman, & von der Heydt, 2000). Interestingly, G-cells exhibit properties associated with shape skeletons. For example, G-cells specify the relations between contours, which may allow the visual system to determine an object's shape despite noisy or incomplete visual information (Craft, Schütze, Niebur, & von der Heydt, 2007; Martin & von der Heydt, 2015). Moreover, the response profile of G-cells within a shape corresponds to the points of the shape's skeleton (Craft et al., 2007), as would be expected if they implement a skeletal computation. Indeed, 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 of G-cells (Ardila, Mihalas, von der Heydt and Niebur, 2012).

Nevertheless, one might ask why we did not find evidence of skeletal representations in either V2 or V4, given that these regions are also frequently implicated in perceptual organization (Cox et al., 2013; McMains and Kastner, 2010; Zhou et al., 2000), particularly in electrophysiology studies with monkeys (von der Heydt, 2015). First, if shape skeletons reflect the response profile of G-cells, then they would not arise in V2, which is primarily comprised of B-cells. Moreover, G-cells are thought to arise in the visual region directly following V2 (Craft et al., 2007; Martin & von der Heydt, 2015) which, in humans, is V3 but, in monkeys, is often delineated as V4 (DiCarlo et al., 2012; Gross et al., 1993; Serre et al., 2007). Studies have shown that V3 in humans is proportionally much larger than in monkeys and there is debate regarding whether monkeys have a human-like V3 at all (Arcaro and Kastner, 2015; but, see Brewer et al., 2002). Most relevant here is the fact that few studies on perceptual organization with monkeys have recorded from V3. Instead, these studies primarily focus on V2 and V4 (Hegdé and Van Essen, 2006; Poort et al., 2012; Zhou et al., 2000). Given that we found evidence of skeletal representations in V3, an intriguing possibility is that V3 may be the locus of G-cells and that skeletal representations within V3 may be an emergent property of G-cell responses.

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 for object recognition (Chouinard et al., 2009; Grill-Spector et al., 2000). This region has been shown to be particularly sensitive to object-centered shape information and is tolerant to some 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 can be decoded from monkey IT across changes in both object orientation and surface form (Hung et al., 2012). Our findings 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 et al., 2006; de-Wit et al., 2013; Konen et al., 2011). Building on these studies, the present work provides direct evidence of skeletal representations in human LO and, crucially, demonstrates that such representations cannot be accounted for by other models of visual processing.

Interestingly, we did not find evidence of skeletal representations in another object-selective region, namely pFs. 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 and Op de Beeck, 2016; Dilks et al., 2011; 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 et al., 2007; Behrmann et al., 2016). Given that skeletal models are exclusively descriptions of shape, such that they do not take semantic content into account, it follows that we did not find evidence of shape skeletons in pFs. Nevertheless, this response profile could reflect the relatively artificial nature of our objects. Future research might test whether the shape skeleton model explains unique variance in the response profile of pFs when familiar objects are used.

Although we found that HMAX was predictive of the response profile in V3 and that AlexNet-fc7 was also predictive in LO, the predictive value of a skeletal model in V3 and LO held even when controlling for other models. Not only are these other 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 and Bülthoff, 1998) and feature descriptions (i.e., AlexNet; Ullman, Assif, Fetaya and Harari, 2016; Yamins et al., 2014). Moreover, by changing the object's surface forms, we changed the non-accidental properties of the objects' component parts, thereby allowing for a test of component description theories (Biederman, 1987; Kayaert et al., 2003). These results directly address concerns of previous studies that skeletal coding could be accounted for by other models of vision (e.g., Lescroart and Biederman, 2012) and provide more direct evidence of skeletal coding in the neural processing of objects in V3 and LO.

Nevertheless, there are important caveats to the current findings. We used novel 3D objects because they allowed us to systematically control both the object skeletons and other visual features. Yet a potential concern with these stimuli is whether our results would generalize to real objects. Although we did not test this possibility directly, we predict generalizability given that participants readily discriminated these stimuli in speeded contexts, suggesting that perception of our stimuli evoke processes similar to those of real objects (DiCarlo et al., 2012). Indeed, two recent studies demonstrated unique selectivity to skeletal information in natural images within the ventral visual stream (Papale et al., 2019, 2020). Moreover, although one might argue that our stimuli made the skeleton especially salient, not typical of other objects, we would point out that other models of vision could readily discriminate

these objects and that the different surface forms directly mimic real world contexts in which different objects with the same skeletons vary in external features.

We would also acknowledge that the spatial and temporal resolution of fMRI places important qualifiers on these conclusions. First, although our results were consistent across different ROI sizes and localization approaches (see Supplemental Data), it is nevertheless possible that shape skeletons are represented in sub-populations of neurons within each region and that these regions have secondary functions. For instance, V3 and LO have been shown to be sensitive to other types of visual cues, including motion (Dupont et al., 1997; Felleman and Van Essen, 1987) and depth (Parker, 2007; Welchman, 2016). Moreover, our own results illustrate how multiple models explain unique variance within the same region. Second, although we found that skeletal sensitivity peaks in V3 and LO, our searchlight results nevertheless suggest that skeletal sensitivity may exist along a gradient within the visual hierarchy much like shape sensitivity more generally (Freud et al., 2017; Freud et al., 2019). Third, our data cannot address whether skeletal representations in these regions arise via feedforward or feedback processes. Indeed, feedback processes are known to be important for both perceptual organization (Mannion et al., 2010; Murray et al., 2002; Wokke et al., 2013) and invariant object recognition (Kar et al., 2019; Tang et al., 2018). A more complete understanding of V3 and LO, along with experiments designed to test the causal role of shape skeletons in human vision, will be needed to confirm the claims of the present research.

In conclusion, our work highlights the unique role that shape skeletons play in the neural processing of objects. These findings not only enhance our understanding of how objects may be 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 et al., 2018; Geirhos et al., 2018).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

This work was supported by a National Institutes of Health (NIH) institutional training grant (T32 HD071845) and a seed grant from the Facility for Education and Research in Neuroscience (FERN) awarded to VA.

Data availability

All data and stimuli are available at: https://osf.io/svz86/.

References

Aguirre GK, Mattar MG, Magis-Weinberg L, 2011. de Bruijn cycles for neural decoding. Neuroimage 56 (3), 1293–1300. 10.1016/j.neuroimage.2011.02.005. [PubMed: 21315160]

Amir O, Biederman I, Hayworth KJ, 2012. Sensitivity to nonaccidental properties across various shape dimensions. Vis. Res. 62, 35–43. 10.1016/j.visres.2012.03.020. [PubMed: 22491056]

Arcaro MJ, Kastner S, 2015. Topographic organization of areas V3 and V4 and its relation to supra-areal organization of the primate visual system. Vis. Neurosci. 32, E014. 10.1017/S0952523815000115. [PubMed: 26241035]

Ardila D, Mihalas S, von der Heydt R, Niebur E, 2012. Medial axis generation in a model of perceptual organization. Conf. Inf. Sci. Syst. 1–4.

Ayzenberg V, Chen Y, Yousif SR, Lourenco SF, 2019. Skeletal representations of shape in human vision: evidence for a pruned medial axis model. J. Vis. 19 (6), 1–21. 10.1167/19.6.6.

Ayzenberg V, Lourenco SF, 2019. Skeletal descriptions of shape provide unique perceptual information for object recognition. Sci. Rep. 9 (1), 1–13. 10.1038/s41598-019-45268-y. [PubMed: 30626917]

Baker N, Lu H, Erlikhman G, Kellman PJ, 2018. Deep convolutional networks do not classify based on global object shape. PLoS Comput. Biol. 14 (12), e1006613 10.1371/journal.pcbi.1006613. [PubMed: 30532273]

Barense MD, Gaffan D, Graham KS, 2007. The human medial temporal lobe processes online representations of complex objects. Neuropsychologia 45 (13), 2963–2974. [PubMed: 17658561]

Behrmann M, Lee ACH, Geskin JZ, Graham KS, Barense MD, 2016. Temporal lobe contribution to perceptual function: a tale of three patient groups. Neuropsychologia 90, 33–45. 10.1016/j.neuropsychologia.2016.05.002. [PubMed: 27150707]

Behrmann M, Peterson MA, Moscovitch M, Suzuki S, 2006. Independent representation of parts and the relations between them: evidence from integrative agnosia. J. Exp. Psychol. Hum. Percept. Perform. 32 (5), 1169–1184. [PubMed: 17002529]

Biederman I, 1987. Recognition-by-components: a theory of human image understanding. Psychol. Rev. 94 (2), 115–147. [PubMed: 3575582]

Blum H, Nagel RN, 1978. Shape description using weighted symmetric axis features. Pattern Recogn. 10 (3), 167–180. 10.1016/0031-3203(78)90025-0.

Bonett DG, 2002. Sample size requirements for estimating intraclass correlations with desired precision. Stat. Med. 21 (9), 1331–1335. [PubMed: 12111881]

Bonner MF, Epstein RA, 2018. Computational mechanisms underlying cortical responses to the affordance properties of visual scenes. PLoS Comput. Biol. 14 (4), e1006111 10.1371/journal.pcbi.1006111. [PubMed: 29684011]

Bracci S, Op de Beeck H, 2016. Dissociations and associations between shape and category representations in the two visual pathways. J. Neurosci. 36 (2), 432–444. [PubMed: 26758835]

Brewer AA, Press WA, Logothetis NK, Wandell BA, 2002. Visual areas in macaque cortex measured using functional magnetic resonance imaging. J. Neurosci. 22 (23), 10416–10426. [PubMed: 12451141]

Brincat SL, Connor CE, 2004. Underlying principles of visual shape selectivity in posterior inferotemporal cortex. Nat. Neurosci. 7, 880. 10.1038/nn1278. [PubMed: 15235606]

Brincat SL, Connor CE, 2006. Dynamic shape synthesis in posterior inferotemporal cortex. Neuron 49 (1), 17–24. 10.1016/j.neuron.2005.11.026. [PubMed: 16387636]

Cadieu CF, Hong H, Yamins DL, Pinto N, Ardila D, Solomon EA, DiCarlo JJ, 2014. Deep neural networks rival the representation of primate IT cortex for core visual object recognition. PLoS Comput. Biol. 10 (12), e1003963. [PubMed: 25521294]

Caplovitz GP, Barroso DJ, Hsieh PJ, Tse PU, 2008. fMRI reveals that non-local processing in ventral retinotopic cortex underlies perceptual grouping by temporal synchrony. Hum. Brain Mapp. 29 (6), 651–661. [PubMed: 17598165]

Caplovitz GP, Peter UT, 2010. Extrastriate cortical activity reflects segmentation of motion into independent sources. Neuropsychologia 48 (9), 2699–2708. [PubMed: 20478319]

Chouinard PA, Whitwell RL, Goodale MA, 2009. The lateral-occipital and the inferior-frontal cortex play different roles during the naming of visually presented objects. Hum. Brain Mapp. 30 (12), 3851–3864. 10.1002/hbm.20812. [PubMed: 19441022]

Cox MA, Schmid MC, Peters AJ, Saunders RC, Leopold DA, Maier A, 2013. Receptive field focus of visual area V4 neurons determines responses to illusory surfaces. Proc. Natl. Acad. Sci. Unit. States Am. 110 (42), 17095–17100. 10.1073/pnas.1310806110.

Craft E, Schütze H, Niebur E, von der Heydt R, 2007. A neural model of figure–ground organization. J. Neurophysiol. 97 (6), 4310–4326. 10.1152/jn.00203.2007. [PubMed: 17442769]

de-Wit LH, Kubilius J, de Beeck HPO, Wagemans J, 2013. Configural Gestalts remain nothing more than the sum of their parts in visual agnosia. i-Perception 4 (8), 493–497. [PubMed: 25165506]

Destler N, Singh M, Feldman J, 2019. Shape discrimination along morph-spaces. Vis. Res. 158, 189–199. [PubMed: 30878276]

DiCarlo James J., Rust Zoccolan, D., Nicole C, 2012. How does the brain solve visual object recognition? Neuron 73 (3), 415–434. 10.1016/j.neuron.2012.01.010. [PubMed: 22325196]

Dilks DD, Julian JB, Kubilius J, Spelke ES, Kanwisher N, 2011. Mirror-image sensitivity and invariance in object and scene processing pathways. J. Neurosci. 31 (31), 11305–11312. [PubMed: 21813690]

Downing PE, Jiang Y, Shuman M, Kanwisher N, 2001. A cortical area selective for visual processing of the human body. Science 293 (5539), 2470–2473. 10.1126/science.1063414. [PubMed: 11577239]

Drucker DM, Aguirre GK, 2009. Different spatial scales of shape similarity representation in lateral and ventral LOC. Cerebr. Cortex 19 (10), 2269–2280. 10.1093/cercor/bhn244.

Dupont P, De Bruyn B, Vandenberghe R, Rosier A-M, Michiels J, Marchal G, Orban G, 1997. The kinetic occipital region in human visual cortex. Cerebr. Cortex 7 (3), 283–292.

Dwivedi K, Bonner MF, Cichy RM, Roig G, 2021. Unveiling functions of the visual cortex using task-specific deep neural networks. PLoS Comput. Biol. 17 (8), e1009267. [PubMed: 34388161]

Feldman J, Singh M, 2006. Bayesian estimation of the shape skeleton. Proc. Natl. Acad. Sci. Unit. States Am. 103 (47), 18014–18019.

Felleman DJ, Van Essen DC, 1987. Receptive field properties of neurons in area V3 of macaque monkey extrastriate cortex. J. Neurophysiol. 57 (4), 889–920. [PubMed: 3585463]

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. J. Vis. 15 (12) 10.1167/15.12.1125,1125-1125.

Firestone C, Scholl BJ, 2014. “Please tap the shape, anywhere you like” shape skeletons in human vision revealed by an exceedingly simple measure. Psychol. Sci. 25 (2), 377–386. [PubMed: 24406395]

Freud E, Culham JC, Plaut DC, Behrmann M, 2017. The large-scale organization of shape processing in the ventral and dorsal pathways. eLife 6, e27576. [PubMed: 28980938]

Freud E, Plaut DC, Behrmann M, 2019. Protracted developmental trajectory of shape processing along the two visual pathways. J. Cognit. Neurosci. 31 (10), 1589–1597. [PubMed: 31180266]

Geirhos R, Rubisch P, Michaelis C, Bethge M, Wichmann FA, 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. Vis. Res. 41 (10), 1409–1422. [PubMed: 11322983]

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. [PubMed: 9697863]

Grill-Spector K, Kushnir T, Hendler T, Malach R, 2000. The dynamics of object-selective activation correlate with recognition performance in humans. Nat. Neurosci. 3 (8), 837–843. [PubMed: 10903579]

Gross CG, Rodman HR, Cochin PM, Colombot MW, 1993. Inferior temporal cortex as a pattern recognition device. In: Paper Presented at the Computational Learning & Cognition: Proceedings of the Third NEC Research Symposium.

Hatfield M, McCloskey M, Park S, 2016. Neural representation of object orientation: a dissociation between MVPA and Repetition Suppression. Neuroimage 139, 136–148. [PubMed: 27236084]

Hegdé J, Van Essen DC, 2006. A comparative study of shape representation in macaque visual areas V2 and V4. Cerebr. Cortex 17 (5), 1100–1116. 10.1093/cercor/bhl020.

Hung C-C, Carlson ET, Connor CE, 2012. Medial axis shape coding in macaque inferotemporal cortex. Neuron 74 (6), 1099–1113. [PubMed: 22726839]

Julian JB, Fedorenko E, Webster J, Kanwisher N, 2012. An algorithmic method for functionally defining regions of interest in the ventral visual pathway. Neuroimage 60 (4), 2357–2364. 10.1016/j.neuroimage.2012.02.055. [PubMed: 22398396]

Kamps FS, Morris EJ, Dilks DD, 2019. A face is more than just the eyes, nose, and mouth: fMRI evidence that face-selective cortex represents external features. Neuroimage 184, 90–100. [PubMed: 30217542]

Kar K, Kubilius J, Schmidt K, Issa EB, DiCarlo JJ, 2019. Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. Nat. Neurosci. 22 (6), 974–983. 10.1038/s41593-019-0392-5. [PubMed: 31036945]

Kayaert G, Biederman I, Vogels R, 2003. Shape tuning in macaque inferior temporal cortex. J. Neurosci. 23 (7), 3016–3027. 10.1523/jneurosci.23-07-03016.2003. [PubMed: 12684489]

Keefe BD, Gouws AD, Sheldon AA, Vernon RJ, Lawrence SJ, McKeefry DJ, Morland AB, 2018. Emergence of symmetry selectivity in the visual areas of the human brain: fMRI responses to symmetry presented in both frontoparallel and slanted planes. Hum. Brain Mapp. 39 (10), 3813–3826. [PubMed: 29968956]

Kimia BB, 2003. On the role of medial geometry in human vision. J. Physiol. Paris 97 (2), 155–190. [PubMed: 14766140]

Konen Christina S., Behrmann M, Nishimura M, Kastner S, 2011. The functional neuroanatomy of object agnosia: a case study. Neuron 71 (1), 49–60. [PubMed: 21745637]

Kovács I, Fehér Á, Julesz B, 1998. Medial-point description of shape: a representation for action coding and its psychophysical correlates. Vis. Res. 38 (15), 2323–2333. [PubMed: 9798002]

Kriegeskorte N, Goebel R, Bandettini P, 2006. Information-based functional brain mapping. Proc. Natl. Acad. Sci. Unit. States Am. 103 (10), 3863–3868.

Kriegeskorte N, Mur M, Bandettini P, 2008. Representational similarity analysis - connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2 (4) 10.3389/neuro.06.004.2008.

Krizhevsky A, Sutskever I, Hinton GE, 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 1097–1105.

Lescroart MD, Biederman I, 2012. Cortical representation of medial axis structure. Cerebr. Cortex 23 (3), 629–637.

Lescroart MD, Stansbury DE, Gallant JL, 2015. Fourier power, subjective distance, and object categories all provide plausible models of BOLD responses in scene-selective visual areas. Front. Comput. Neurosci. 9 (135), 1–20. 10.3389/fncom.2015.00135. [PubMed: 25767445]

Lowet AS, Firestone C, Scholl BJ, 2018. Seeing structure: shape skeletons modulate perceived similarity. Atten. Percept. Psychophys. 80 (5), 1278–1289. 10.3758/s13414-017-1457-8. [PubMed: 29546555]

Magri C, Konkle T, Caramazza A, 2021. The contribution of object size, manipulability, and stability on neural responses to inanimate objects. Neuroimage 237, 118098. 10.1016/j.neuroimage.2021.118098. [PubMed: 33940141]

Mannion DJ, McDonald JS, Clifford CW, 2010. The influence of global form on local orientation anisotropies in human visual cortex. Neuroimage 52 (2), 600–605. [PubMed: 20434564]

Margalit E, Biederman I, Herald SB, Yue X, von der Malsburg C, 2016. An applet for the Gabor similarity scaling of the differences between complex stimuli. Atten. Percept. Psychophys. 78 (8), 2298–2306. 10.3758/s13414-016-1191-7. [PubMed: 27557818]

Martin AB, von der Heydt R, 2015. Spike synchrony reveals emergence of proto-objects in visual cortex. J. Neurosci. 35 (17), 6860–6870. [PubMed: 25926461]

McMains SA, Kastner S, 2010. Defining the units of competition: influences of perceptual organization on competitive interactions in human visual cortex. J. Cognit. Neurosci. 22 (11), 2417–2426. [PubMed: 19925189]

Montaser-Kouhsari L, Landy MS, Heeger DJ, Larsson J, 2007. Orientation-selective adaptation to illusory contours in human visual cortex. J. Neurosci. 27 (9), 2186–2195. [PubMed: 17329415]

Murray SO, Kersten D, Olshausen BA, Schrater P, Woods DL, 2002. Shape perception reduces activity in human primary visual cortex. Proc. Natl. Acad. Sci. Unit. States Am. 99 (23), 15164–15169.

O'Brien RM, 2007. A caution regarding rules of thumb for variance inflation factors. Qual. Quantity 41 (5), 673–690.

Oliva A, Torralba A, 2006. Building the gist of a scene: the role of global image features in recognition. Prog. Brain Res. 155, 23–36. [PubMed: 17027377]

Papale P, Betta M, Handjaras G, Malfatti G, Cecchetti L, Rampinini A, Leo A, 2019. Common spatiotemporal processing of visual features shapes object representation. Sci. Rep. 9 (1), 7601. 10.1038/s41598-019-43956-3. [PubMed: 31110195]

Papale P, Leo A, Handjaras G, Cecchetti L, Pietrini P, Ricciardi E, 2020. Shape coding in occipito-temporal cortex relies on object silhouette, curvature, and medial axis. J. Neurophysiol. 124 (6), 1560–1570. 10.1152/jn.00212.2020. [PubMed: 33052726]

Parker AJ, 2007. Binocular depth perception and the cerebral cortex. Nat. Rev. Neurosci. 8 (5), 379–391. 10.1038/nrn2131. [PubMed: 17453018]

Peelen MV, Downing PE, 2005. Selectivity for the human body in the fusiform gyrus. J. Neurophysiol. 93 (1), 603–608. 10.1152/jn.00513.2004. [PubMed: 15295012]

Poort J, Raudies F, Wannig A, Lamme VA, Neumann H, Roelfsema PR, 2012. The role of attention in figure-ground segregation in areas V1 and V4 of the visual cortex. Neuron 75 (1), 143–156. [PubMed: 22794268]

Psotka J, 1978. Perceptual processes that may create stick figures and balance. J. Exp. Psychol. Hum. Percept. Perform. 4 (1), 101–111. [PubMed: 627839]

Rezanejad M, Downs G, Wilder J, Walther DB, Jepson A, Dickinson S, Siddiqi K, 2019. Scene categorization from contours: medial Axis based salience measures. In: Paper Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Rezanejad M, Siddiqi K, 2013. Flux graphs for 2D shape analysis. In: Shape Perception in Human and Computer Vision. Springer, pp. 41–54.

Sasaki Y, 2007. Processing local signals into global patterns. Curr. Opin. Neurobiol. 17 (2), 132–139. [PubMed: 17369036]

Sasaki Y, Vanduffel W, Knutsen T, Tyler C, Tootell R, 2005. Symmetry activates extrastriate visual cortex in human and nonhuman primates. Proc. Natl. Acad. Sci. U. S. A 102 (8), 3159–3163. 10.1073/pnas.0500319102. [PubMed: 15710884]

Schrimpf M, Kubilius J, Hong H, Majaj NJ, Rajalingham R, Issa EB, DiCarlo JJ, 2018. Brain-score: which artificial neural network for object recognition is most brain-like? bioRxiv. 10.1101/407007.

Sebastian TB, Klein PN, Kimia BB, 2004. Recognition of shapes by editing their shock graphs. IEEE Trans. Pattern Anal. Mach. Intell. 26 (5), 550–571. [PubMed: 15460278]

Serre T, Oliva A, Poggio T, 2007. A feedforward architecture accounts for rapid categorization. Proc. Natl. Acad. Sci. Unit. States Am. 104 (15), 6424–6429. 10.1073/pnas.0700622104.

Simonyan K, Zisserman A, 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv Preprint arXiv:1409.1556.

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Matthews PM, 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23, S208–S219. 10.1016/j.neuroimage.2004.07.051. [PubMed: 15501092]

Tang H, Schrimpf M, Lotter W, Moerman C, Paredes A, Ortega Caro J, Kreiman G, 2018. Recurrent computations for visual pattern completion. Proc. Natl. Acad. Sci. Unit. States Am. 115 (35), 8835–8840. 10.1073/pnas.1719397115.

Tarr MJ, Bülthoff HH, 1998. Image-based object recognition in man, monkey and machine. Cognition 67 (1), 1–20. 10.1016/S0010-0277(98)00026-2. [PubMed: 9735534]

Trinh NH, Kimia BB, 2011. Skeleton search: category-specific object recognition and segmentation using a skeletal shape model. Int. J. Comput. Vis. 94 (2), 215–240.

Ullman S, Assif L, Fetaya E, Harari D, 2016. Atoms of recognition in human and computer vision. Proc. Natl. Acad. Sci. Unit. States Am. 113 (10), 2744–2749.

Van Meel C, Baeck A, Gillebert CR, Wagemans J, Op de Beeck HP, 2019. The representation of symmetry in multi-voxel response patterns and functional connectivity throughout the ventral visual stream. Neuroimage 191, 216–224. [PubMed: 30771448]

von der Heydt R, 2015. Figure–ground organization and the emergence of proto-objects in the visual cortex. Front. Psychol. 6, 1695. 10.3389/fpsyg.2015.01695. [PubMed: 26579062]

Walther A, Nili H, Ejaz N, Alink A, Kriegeskorte N, Diedrichsen J, 2016. Reliability of dissimilarity measures for multi-voxel pattern analysis. Neuroimage 137, 188–200. [PubMed: 26707889]

Wang L, Mruczek REB, Arcaro MJ, Kastner S, 2014. Probabilistic maps of visual topography in human cortex. Cerebr. Cortex 25 (10), 3911–3931. 10.1093/cercor/bhu277.

Welchman AE, 2016. The human brain in depth: how we see in 3D. Ann. Rev. Vision Sci. 2, 345–376.

Wilder J, Feldman J, Singh M, 2011. Superordinate shape classification using natural shape statistics. Cognition 119 (3), 325–340. 10.1016/j.cognition.2011.01.009. [PubMed: 21440250]

Wilder J, Rezanejad M, Dickinson S, Siddiqi K, Jepson A, Walther DB, 2019. Local contour symmetry facilitates scene categorization. Cognition 182, 307–317. 10.1016/j.cognition.2018.09.014. [PubMed: 30415132]

Wokke ME, Vandenbroucke AR, Scholte HS, Lamme VA, 2013. Confuse your illusion: feedback to early visual cortex contributes to perceptual completion. Psychol. Sci. 24 (1), 63–71. [PubMed: 23228938]

Xu Y, Vaziri-Pashkam M, 2021. Limits to visual representational correspondence between convolutional neural networks and the human brain. Nat. Commun. 12 (1), 1–16. [PubMed: 33397941]

Yamins DL, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ, 2014. Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proc. Natl. Acad. Sci. Unit. States Am. 111 (23), 8619–8624.

Zhou H, Friedman HS, von der Heydt R, 2000. Coding of border ownership in monkey visual cortex. J. Neurosci. 20 (17), 6594–6611. 10.1523/jneurosci.20-17-06594.2000. [PubMed: 10964965]

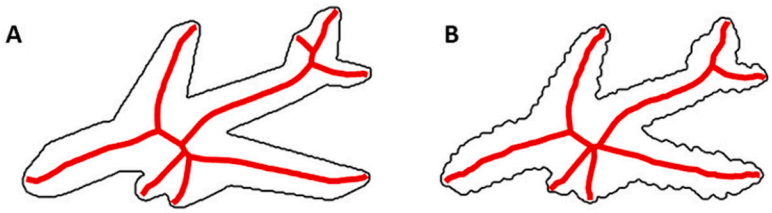

Fig. 1.

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 and Singh, 2006).

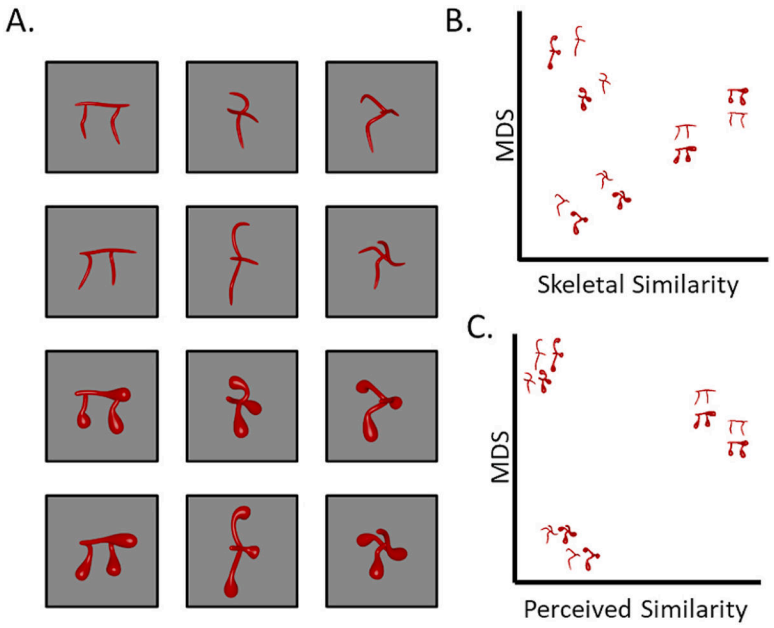

Fig. 2.

Experimental stimuli and multi-dimensional scaling (MDS) plots illustrating the similarities between objects as computed by a skeletal model and as judged by human participants. (A) Six objects with unique skeletal structures were generated. Each object was rendered with two surface forms to change the objects' component parts without disrupting the skeleton. (B) Objects were selected in pairs using the skeletal model, such that within- and between-pair skeletal similarity were approximately matched across objects. (C) Human similarity judgments closely align with skeletal similarity. Surface form similarity is not illustrated in the MDS plot.

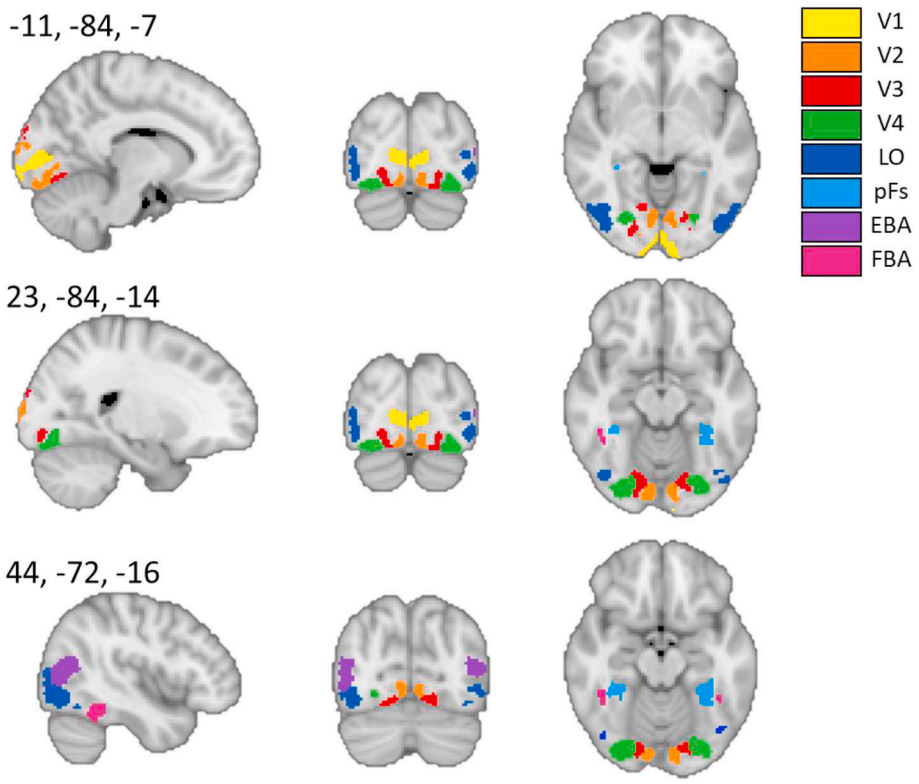

Fig. 3.

Group averaged ROIs displayed in MNI standard space. Each color corresponds to a different ROI. Early visual cortex ROIs (V1–V4) were defined using probabilistic maps. Higher-level visual regions (LO, pFs, EBA, FBA) were functionally defined in each participant using an independent localizer. Numerical values correspond to the MNI coordinates (x, y, z) of each example slice. Primary data analyses were conducted in each participant's native space. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

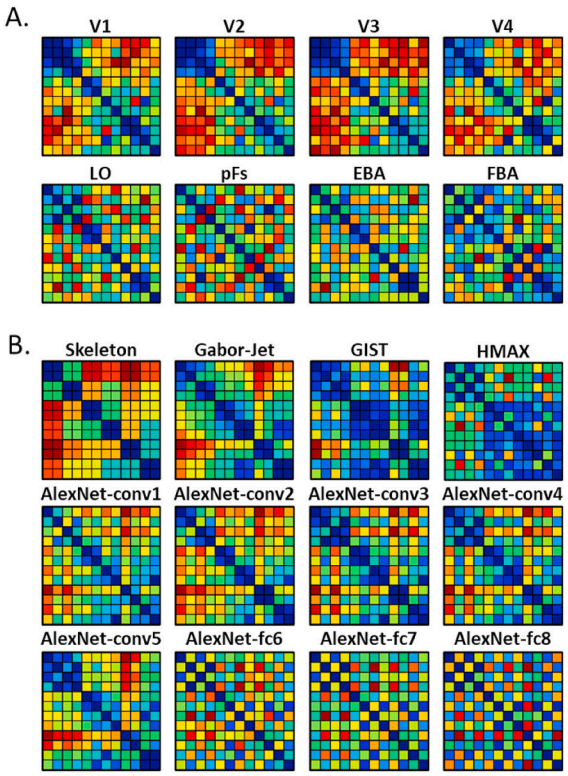

Fig. 4.

Representational dissimilarity matrices (RDMs) for (A) primary regions of interest (ROIs), as well as (B) models and model layers.

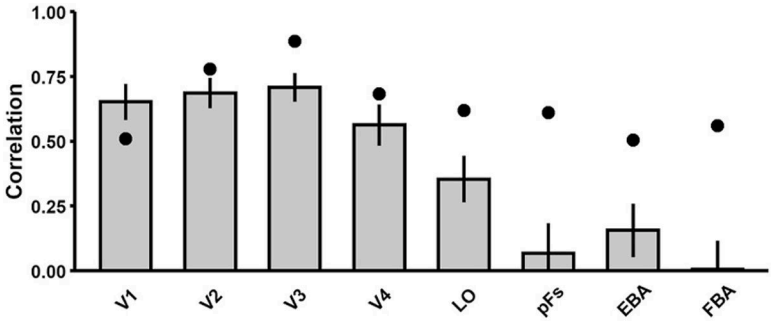

Fig. 5.

Bar plot displaying the correlations between the skeletal model and the multivariate response pattern in each ROI. The model of skeletal similarity was significantly correlated with response patterns in V1–V4 and LO. The skeletal model was not predictive of the response pattern in pFs, EBA, or FBA. Error bars were calculated by resampling the data 10,000 times (with replacement) and computing a bootstrapped SE. Blacks circles indicate the noise ceiling for each ROI (see also Table 1). Note that the model correlation in V1 exceeds the noise ceiling. Although rare, model fits can exceed noise ceilings, particularly for early visual regions where there may be more variability across participants (c.f. Xu and Vaziri-Pashkam, 2021).

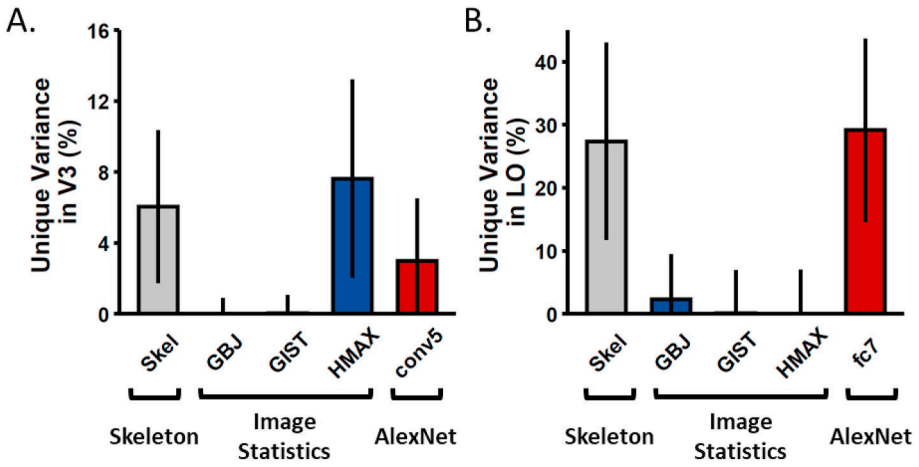

Fig. 6.

Variance partitioning results. Bar plot displaying the percentage of unique variance accounted for by each model in (A) V3 and (B) LO. A model of skeletal similarity explained unique variance in both V3 and LO, but not in other cortical regions. For AlexNet, the best correlated layer for each ROI was used. Error bars represent bootstrapped SE.

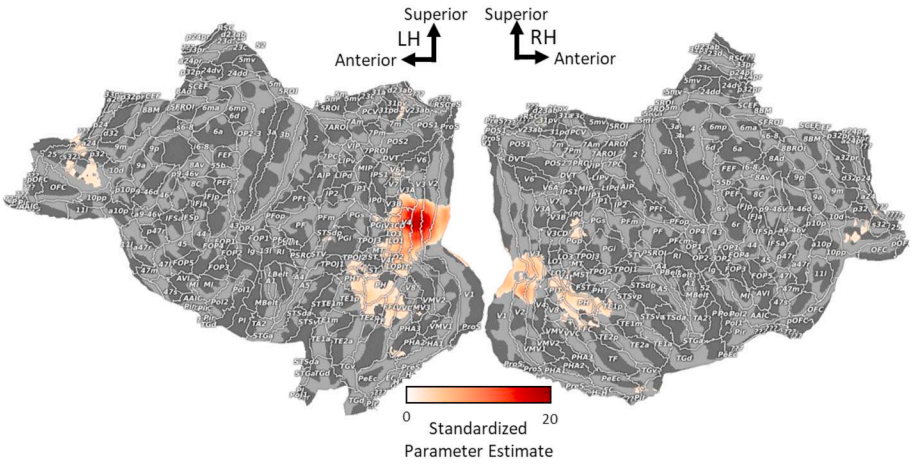

Fig. 7.

Unique variance explained by the skeletal model displayed on a whole-brain flat-map. Colored regions indicate significantly positive clusters (FDR corrected, $p < .05$).

Neuropsychologia. Author manuscript; available in PMC 2023 January 14.

Table 1

Results of the correlation, regression, and variance partitioning analyses for each ROI and each 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. A reliability noise ceiling was calculated by iteratively splitting the participant data into two random sets and then correlating them with one another (10,000 iterations). Regression analyses were conducted for each neural RDM by entering each model RDM as a predictor into a linear regression model. R^2 values indicate 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 the combination of models, and then, calculating the percentage of the total explained variance (R^2) uniquely explained by each model.

			Model Correlations			Model Regression			
ROI	**Model**	**Noise Ceiling**	***r***	***r*** ^{2}	***p***	***R*** ^{2}	***β***	***p***	**VPA %**
V1		0.51				0.52			
	Skeleton	-	0.65	0.43	< .001	-	0.16	0.368	0.01
	Gabor-Jet	-	0.66	0.44	< .001	-	0.34	0.160	0.03
	GIST	-	0.56	0.31	< .001	-	−0.07	0.702	0.00
	HMAX	-	0.46	0.21	< .001	-	0.19	0.070	0.05
	AlexNet-conv2	-	0.66	0.44	< .001	-	0.20	0.263	0.02
V2		0.78	-	-	-	0.64	-	-	-
	Skeleton	-	0.69	0.47	< .001	-	0.13	0.397	0.01
	Gabor-Jet	-	0.71	0.51	< .001	-	0.45	0.032	0.05
	GIST	-	0.60	0.36	< .001	-	−0.25	0.143	0.02
	HMAX	-	0.53	0.28	< .001	-	0.24	0.009	0.07
	AlexNet-conv5	-	0.71	0.50	< .001	-	0.35	0.025	0.05
V3		0.89	-	-	-	0.63	-	-	-
	Skeleton	-	0.71	0.50	< .001	-	0.38	0.016	0.06
	Gabor-Jet	-	0.67	0.44	< .001	-	0.01	0.946	0.00
	GIST	-	0.62	0.39	< .001	-	0.03	0.850	0.00
	HMAX	-	0.56	0.32	< .001	-	0.26	0.007	0.08
	AlexNet-conv5	-	0.70	0.48	< .001	-	0.27	0.085	0.03
V4		0.68	-	-	-	0.51	-	-	-
	Skeleton	-	0.56	0.32	< .001	-	0.03	0.861	0.00
	Gabor-Jet	-	0.60	0.36	< .001	-	0.27	0.257	0.02

			Model Correlations			Model Regression			
ROI	**Model**	**Noise Ceiling**	***r***	r^2	***p***	R^2	***β***	***p***	**VPA %**
	GIST	-	0.55	0.30	< .001	-	−0.18	0.354	0.01
	HMAX	-	0.50	0.25	< .001	-	0.24	0.028	0.08
	AlexNet-conv5	-	0.67	0.44	< .001	-	0.47	0.011	0.11
LO		0.62	-	-	-	0.25	-	-	-
	Skeleton	-	0.35	0.13	.004	-	0.51	0.024	0.27
	Gabor-Jet	-	0.25	0.06	.040	-	−0.19	0.503	0.02
	GIST	-	0.23	0.05	.063	-	−0.03	0.884	0.00
	HMAX	-	0.33	0.11	.005	-	−0.01	0.942	0.00
	AlexNet-fc7	-	0.36	0.13	.003	-	0.35	0.020	0.29
pFs		0.61	-	-	-	0.06	-	-	-
	Skeleton	-	0.07	0.00	.594	-	0.32	0.186	0.43
	Gabor-Jet	-	−0.02	0.00	.882	-	−0.21	0.532	0.10
	GIST	-	−0.09	0.01	.478	-	−0.23	0.348	0.22
	HMAX	-	−0.07	0.01	.554	-	−0.15	0.312	0.25
	AlexNet-conv1	-	0.03	0.00	.804	-	0.24	0.241	0.34
EBA		0.50	-	-	-	0.10	-	-	-
	Skeleton	-	0.16	0.02	.209	-	0.32	0.194	0.25
	Gabor-Jet	-	0.07	0.01	.570	-	0.01	0.964	0.00
	GIST	-	0.00	0.00	.969	-	−0.28	0.234	0.21
	HMAX	-	0.16	0.03	.193	-	−0.02	0.894	0.00
	AlexNet-fc7	-	0.21	0.04	.094	-	0.26	0.108	0.39
FBA		0.56	-	-	-	0.07	-	-	-
	Skeleton	-	0.00	0.00	.968	-	−0.15	0.546	0.08
	Gabor-Jet	-	0.03	0.00	.782	-	0.47	0.147	0.48
	GIST	-	−0.06	0.00	.610	-	−0.42	0.081	0.70
	HMAX	-	0.09	0.01	.464	-	0.05	0.761	0.02
	AlexNet-fc7	-	0.14	0.02	.261	-	0.16	0.339	0.21