Neurosci. Bull.
https://doi.org/10.1007/s12264-021-00734-4

www.neurosci.cn
www.springer.com/12264

ORIGINAL ARTICLE

Topography of Visual Features in the Human Ventral Visual Pathway

Shijia Fan[1,2] · **Xiaosha Wang**[1,2] · **Xiaoying Wang**[1,2] · **Tao Wei**[1,2] · **Yanchao Bi**[1,2,3]

Received: 12 December 2020 / Accepted: 24 February 2021

Abstract Visual object recognition in humans and nonhuman primates is achieved by the ventral visual pathway (ventral occipital-temporal cortex, VOTC), which shows a well-documented object domain structure. An on-going question is what type of information is processed in the higher-order VOTC that underlies such observations, with recent evidence suggesting effects of certain visual features. Combining computational vision models, fMRI experiment using a parametric-modulation approach, and natural image statistics of common objects, we depicted the neural distribution of a comprehensive set of visual features in the VOTC, identifying voxel sensitivities with specific feature sets across geometry/shape, Fourier power, and color. The visual feature combination pattern in the VOTC is significantly explained by their relationships to different types of response-action computation (fight-or-flight, navigation, and manipulation), as derived from behavioral ratings and natural image statistics. These results offer a comprehensive visual feature map in the VOTC and a plausible theoretical explanation as a mapping onto different types of downstream response-action systems.

Keywords Ventral occipital temporal cortex · Computational vision model · Domain organization · Response mapping

Introduction

The ventral occipital-temporal cortex (VOTC), which underlies visual object recognition in humans and nonhuman primates, has a hierarchical architecture, from a retinotopic organization of simple features in the early visual cortex to a domain-based organization (e.g., animate *vs* inanimate; faces *vs* scenes) in higher-order visual cortical areas [1–5]. One of the key questions is the nature of representation in the higher-level VOTC [6–12]: exactly what kinds of information about or associated with these various object domains are represented here?

As a higher-order "visual" cortex, visual features are assumed to be the major candidate representation [13]. The effects of two specific types of visual feature that are associated with the object domains in the higher-order VOTC have been very recently demonstrated in humans and nonhuman primates – mid-level shapes and colors. For mid-level shapes, high rectilinearity, especially right-angles, is more prevalent in images of scenes and places and activates scene-preferring regions including the parahippocampal place area (PPA) and transverse occipital sulcus more strongly than curved lines in humans [14]. Low rectilinearity, or high curvature, tends to be associated with animate items [15, 16], and tends to activate regions close to the face patches in the macaque brain [17]. Different colors have also been shown to be associated with objects *versus* their backgrounds, and with animate *versus* inanimate objects [18]. Three VOTC patches have been identified to be sensitive to color in the macaque

---

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12264-021-00734-4.

✉ Yanchao Bi
ybi@bnu.edu.cn

[1] State Key Laboratory of Cognitive Neuroscience and Learning, and IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, China

[2] Beijing Key Laboratory of Brain Imaging and Connectomics, Beijing Normal University, Beijing 100875, China

[3] Chinese Institute for Brain Research, Beijing 102206, China

monkey and human [19–21] and a more anterior medial patch in the macaque brain showed both a yellow/red preference and face/body preference [21]. These studies focused on individual visual features, and it is unknown whether the specific effects of features are driven by other features that are correlated with them in various object contexts. Furthermore, the extent to which a single feature can explain the observed VOTC domain distribution is controversial [22]: the anatomical overlap between feature effects and domain effects is far from perfect [17], and the domain preferences are still present when visual shape is controlled [8, 10]. Our first aim, then, was to depict a comprehensive topographical map of visual features across the VOTC, taking into consideration their correlational nature in the context of common objects.

The harder question is, if there were a systematic pattern of various visual feature sensitivity across the VOTC, what factor drives this organization. That is, why does a certain region prefer a particular feature or set of features together, or why are various features preferred by the same, or different, brain region in a particular location? Note that “domain preference” in the VOTC describes the phenomenon and does not constitute a satisfying explanatory variable for the potential effects of features here, because what constitutes the “domain” information being represented is not explicit. A recent proposal is that the neuronal functional preference of VOTC voxels is constrained, at least partly, by the downstream nonvisual, response-action computations for objects such as fight-or-flight, navigation, and manipulation [12, 23–31]. This hypothesis predicts that the visual feature distribution pattern in the VOTC is driven by how they may be associated with various salient response-action systems in the real world, and was tested empirically as the second aim of our study.

To depict a comprehensive topographical map of visual features across the VOTC, we combined computational vision modelling with parametric modulation analysis of fMRI responses. The parametric modulation approach exploits the natural variation in salience of various visual features across object images (obtained from computational vision modelling) and identifies brain regions responsive to each feature or combination of features by computing the degree of association between brain responses and image feature weights. To test the potential hypothesis about the driving forces of the organization of visual features in the VOTC, we examined what prototypical visual features are associated with the major types of response-actions to objects (fight-or-flight, navigation, and manipulation) in the real world, using behavioral rating and natural image statistics in broader image sets, and then examined how well they align with the preferred visual feature sets across various VOTC regions. Finally, the extent to which the documented object domain pattern might reflect the identified visual feature maps was tested using an independent fMRI dataset.

Materials and Methods

Participants

Twenty-nine participants (age range: 18–32 years; male/female: 10/19) participated in the main fMRI experiment. An independent group of 19 participants (age range: 19–29 years; male/female: 8/11) participated in a feature-validation fMRI experiment. An independent group of 24 participants (age range: 18–29 years; male/female: 9/15) participated in the rating study. All participants had no history of neurological or psychiatric impairment, had normal or corrected-to-normal vision, were native Chinese speakers, and provided written informed consent. The main fMRI experiment and rating studies were approved by the Institutional Review Board of the State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University (ICBIR_A_0040_008). The feature-validation fMRI experiment was approved by the Institutional Review Board of the Department of Psychology, Peking University (#2015-05-04). All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

fMRI Stimuli

Stimuli in the main fMRI experiment consisted of 95 colorful real-world objects centered on a white background, including animals (32 images) and inanimate man-made artifacts (28 large artifacts, 35 small manipulable artifacts; Fig. 1A). Images of animals included mammals, birds, reptiles, and insects. Images of large artifacts included buildings, furniture, appliances, communal facilities, and large transport. Images of small manipulable artifacts included common household tools, kitchen utensils, stationery, medical instruments, and accessories. These images were obtained from the Internet and resized to 400 × 400 pixels (10.55° × 10.55° of visual angle).

We also collected a validation dataset that included domain-functional localizer runs and runs of visual features that were presented out of the object contexts (see details in supplementary methods and results).

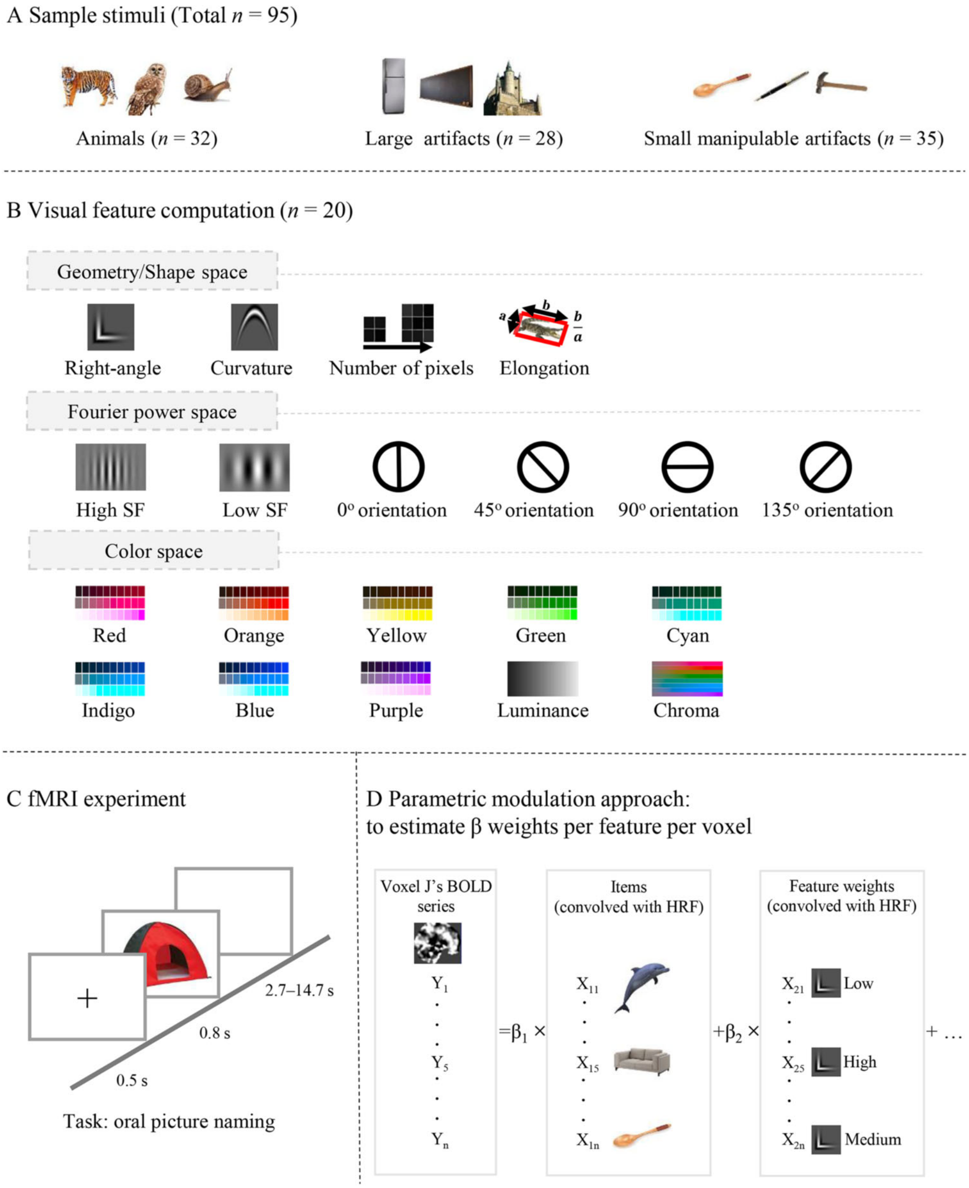

Fig. 1 Schematic overview of the methods in main fMRI experiment. **A** Sample stimuli. Images of 95 common objects (32 animate items and 63 inanimate items, including 28 large artifacts and 35 small manipulable artifacts) were used. **B** Visual feature construction from computational vision models. For each picture, computational vision models were used to obtain values for 20 visual features, including geometry/shape (based on modified Gabor filters), Fourier power features (using 2D fast Fourier transform), color (based on CIE L × C × H space) (see also Fig. S1). **C** fMRI experiment. In an event-related fMRI experiment, participants viewed and named these objects. **D** Parametric modulation analysis. Parametric modulation was used to estimate the degree of association between brain responses and visual feature weights across the whole VOTC.

Computation of Visual Feature Weights in Object Images

The weights of 20 visual features covering a broad range of shapes, spatial frequencies, orientations, and color properties were extracted using computational vision models for each of 95 object images (Fig. S1). Note that the feature set being considered was not aimed to be exhaustive or most optimal, which was extremely difficult due to the open nature of the feature space (e.g., see discussion in Kourtzi and Connor, 2011 [32]). Our approach here was to borrow the conventional relatively low-level visual features in computational vision practice, because (1) they naturally provide computable visual features that comprehensively describe a visual image; (2) they offer a more parsimonious explanation than more complex features; and (3) some of them have been shown to modulate responses in the VOTC (see Introduction).

Geometry/Shape Space

We examined four geometry/shape features: number of pixels, right-angle, curvature, and elongation. For number of pixels, a binary object mask (defined as pixels with grayscale values <240) was created and each pixel in the mask was counted. Overall right-angle and curvature information was measured largely following previous approaches with some modification [14, 17, 33]. Specifically, for right-angle, 64 right-angle Gabor filters (using an absolute function [14]) were constructed using 4 spatial scales (1/5, 1/9, 1/15, and 1/27 cycles per pixel) and 16 orientations (22.5°–360° in 22.5° steps). Images were converted to grayscale and edge maps were constructed using Canny edge detection at a threshold of 0.1 [34]. Each edge map was convolved with 64 Gabor filters of different spatial scales and orientations. This produced 64 Gabor coefficient images, which were then normalized by dividing by the mean magnitude of each Gabor filter. For each spatial scale, the largest magnitude across the 16 coefficient images of different orientations was extracted for each pixel to obtain a peak Gabor coefficient image, which was then averaged across all pixels of each image and Z-scored across the image set. The resulting Gabor coefficient values for each image were finally averaged across 4 spatial scales and Z-scored to provide a single value for each image to represent the amount of right-angle information in that image. For curvature, the same procedure was used using the bank of 320 curved Gabor filters {using a square root function [35], composed of 4 spatial scales, 16 orientations, and 5 levels of curvature (π/256, π/128, π/64, π/32, and π/16)}, to generate a single value for the amount of overall curvature information for each image. Elongation was measured as the aspect ratio of the rectangle that enclosed the object parallel to the object's longest axis.

Fourier Power Space

Images were converted to grayscale and submitted to a 2D fast Fourier transform [built-in MATLAB (MathWorks, Natick, USA) function fft2]. The high/low spatial frequency and 4 orientations (0°, 45°, 90°, and 135°) were measured based on previous approaches [36–38] to parameterize the energy variation in Fourier power space. The overall energy at high and low spatial frequency was calculated by averaging the energy of the high (>5 cycles/degree) and low (<1 cycle/degree) band for each image [36]. Note that we considered the two ends of the spatial frequency bands in the main analysis to align with the previous studies that tested the effects of spatial frequency in higher-order VOTC, which mainly focused on contrasting the effects of high and low bands of spatial frequency [36, 39, 40], except for Rajimehr *et al*., 2011 [36], who tested the effect of the middle band of spatial frequency (1–5 cycles/degree) and found that it had effects similar to the low band of spatial frequency (<1 cycle/degree) when contrasted with the high band of spatial frequency. In a validation analysis, we also included the middle band of spatial frequency as a further variable, and found that the middle band of spatial frequency-modulated response in the left medial fusiform gyrus, and its inclusion did not change the pattern of the other features in the full model (Fig. S2). For orientations, we selected four directions centered on vertical (0°), left oblique (45°), horizontal (90°), and right oblique (135°) with a bandwidth of 20° [38]. For each orientation range, the energy across spatial frequencies was averaged.

Color Space

Three main perceptual dimensions of color – luminance (L), chroma (C), and hue (H) – were quantified using Commission Internationale de l'Eclairage (CIE) L × C × H space following previous studies [18, 41]. Pixel colors in each image were converted from RGB space into CIE L × C × H space using the MATLAB "colorspace" package (https://www.mathworks.com/matlabcentral/fileexchange/28790-colorspace-transformations). The white point for the transformation of the image colors was set to D65 (the standard illumination for noon daylight). The luminance and chroma of each image were calculated by averaging these values across pixels within the object. The hue layer was divided into 8 bins of equal width, which started from the 338°–23° bin, in 45° steps, and roughly corresponded to red, orange, yellow, green, cyan, indigo, blue, and purple [41]. The number of pixels in each bin was then counted as the hue-specific measure; pixels that did not belong to objects or were ambiguous (defined as luminance or chroma values <10) were excluded.

Distribution of Visual Features Across and Within Object Domains

Distribution plots of Z-transformed visual feature weights across 95 images were plotted and the mean and standard deviation (SD) by domains were calculated. Welch's *t*-test (for unequal variances, in MATLAB) was used to test whether there were significant differences between animate and inanimate objects for each feature. When separating the inanimate objects further into large artifacts and small manipulable artifacts, one-way ANOVA [in SPSS Statistics Software version 26 (IBM, Armonk, New York, USA)] was used to test for significant differences among three domains for each feature, followed by Tukey's HSD *post hoc* comparison. Multiple comparisons across features were corrected using the false discovery rate (FDR).

fMRI Experiment Design

Participants were asked to name the 95 object images overtly and as quickly and accurately as possible. There were 6 runs, each lasting 528 s. Each image (0.5 s fixation followed by 0.8 s image) was presented once per run. Inter-trial intervals ranged from 2.7 s to 14.7 s. The order of stimuli and length of inter-trial interval were optimized using optseq2 (http://surfer.nmr.mgh.harvard.edu/optseq/).

The order of items was randomized across runs. Each run started and ended with 10 s of blank screen.

MRI Acquisition and Data Preprocessing

The main fMRI experiment was conducted at the Beijing Normal University Neuroimaging Center using a 3T Siemens Trio Tim scanner (Siemens, Erlangen, Germany). Functional data were collected using an echo-planar imaging sequence [33 axial slices, repetition time (TR) = 2000 ms, echo time (TE) = 30 ms, flip angle = 90°, matrix size = 64 × 64, voxel size = $3 \times 3 \times 3.5$ mm^3 with a gap of 0.7 mm]. T1-weighted anatomical images were acquired using a 3D MPRAGE sequence: 144 slices, TR = 2530 ms, TE = 3.39 ms, flip angle = 7°, matrix size = 256 × 256, voxel size = $1.33 \times 1 \times 1.33$ mm^3.

Functional images were preprocessed and analyzed using Statistical Parametric Mapping (SPM12, http://www.fil.ion.ucl.ac.uk/spm), Statistical Non-parametric Permutation Testing Mapping (SnPM13, http://warwick.ac.uk/snpm), and Data Processing & Analysis of Brain Imaging (DPABI) [42]. The first 5 volumes in each run of the main fMRI experiment and feature-validation experiment were discarded. Image preprocessing included slice-time correction, head-motion correction, normalization to the Montreal Neurological Institute (MNI) space using unified segmentation (resampling voxel size = $3 \times 3 \times 3$ mm^3 in the main fMRI experiment; $2 \times 2 \times 2$ mm^3 in the feature-validation experiment), and spatial smoothing with a Gaussian kernel of 6 mm full-width at half-maximum. Three participants in the main fMRI experiment were excluded from analyses due to excessive head motion (>3 mm maximum translation or 3° rotation).

Statistical analyses were carried out within a functionally defined bilateral VOTC mask (containing 3915 voxels for 3-mm voxel size) constructed in a previous study [43], which was defined as brain regions activated by the contrast of all objects *versus* fixation in an object picture perception task in the VOTC. Activation maps for parametric modulation and contrasts between conditions (see below for details) were first created in individual participants and then submitted to group-level random-effects analyses using SnPM13. No variance smoothing was used and 5,000 permutations were performed. A conventional cluster extent-based inference threshold (voxel level at $P < 0.001$; cluster-level family-wise error (FWE) corrected $P < 0.05$ within the VOTC mask) was adopted unless stated explicitly otherwise.

Topography of Visual Features in the VOTC

To identify brain regions associated with each feature, parametric modulation was employed to investigate the correlations between activity levels and feature weights across the 95 stimulus images in the main fMRI experiment. For the full model that considered the correlations among multiple features, the variance inflation factor (VIF) for each feature was calculated using SPSS Statistics Software version 26 and features with a VIF >10 were excluded from analysis to reduce multicollinearity [44]. Then the preprocessed functional images from each participant were entered into a General Linear Model (GLM), which included the onsets of items as one regressor, the weights of all features for each image in the parametric modulation module, and 6 head-motion regressors for each run. A high-pass filter cutoff was 128 s. Contrast images for each feature *versus* baseline were then calculated and submitted for random-effects analyses. Because there was no *a priori* expectation that any brain region should become "less" active as the processing demands for a given feature increased, making the interpretation of negative correlations speculative, only positive modulations were reported. To obtain raw feature maps without considering correlations among features, we also conducted parametric modulation analyses for each feature by including one feature at a time in the GLM.

To have reference to landmarks showing well-documented object domain preferences, in the result visualization (Fig. 2) we marked the object-domain-preferring clusters for animals (lateral fusiform gyrus, latFG; bilateral), large artifacts (PPA; bilateral), and small manipulable artifacts (lateral occipital temporal cortex, LOTC; left). A GLM that included animals, large artifacts, small manipulable artifacts and 6 head-motion regressors was constructed. Contrast images of each object domain with the other two domains were calculated at the individual level and submitted to SnPM13 for random-effects analyses. The group-level activation maps obtained were thresholded at the cluster-level, FWE-corrected $P < 0.05$ within the VOTC mask with voxel-wise $P < 0.0001$ for animals and large artifacts, and voxel-wise $P < 0.01$ for small manipulable artifacts. The details of the identified regions were as follows: for animal >others, the bilateral latFG, 51 voxels; for large artifacts >others, the bilateral PPA, 464 voxels; and for small manipulable artifacts >others, the left LOTC, 93 voxels.

Factors Driving the Visual Feature Distribution Patterns in VOTC Voxels

After establishing the topography of visual features in the VOTC, here we tried to understand why visual feature sensitivity was distributed across VOTC voxels in the observed way. To test the feasibility of hypothesis of visual features mapping with response actions, we first determined what type of visual feature clustering pattern is

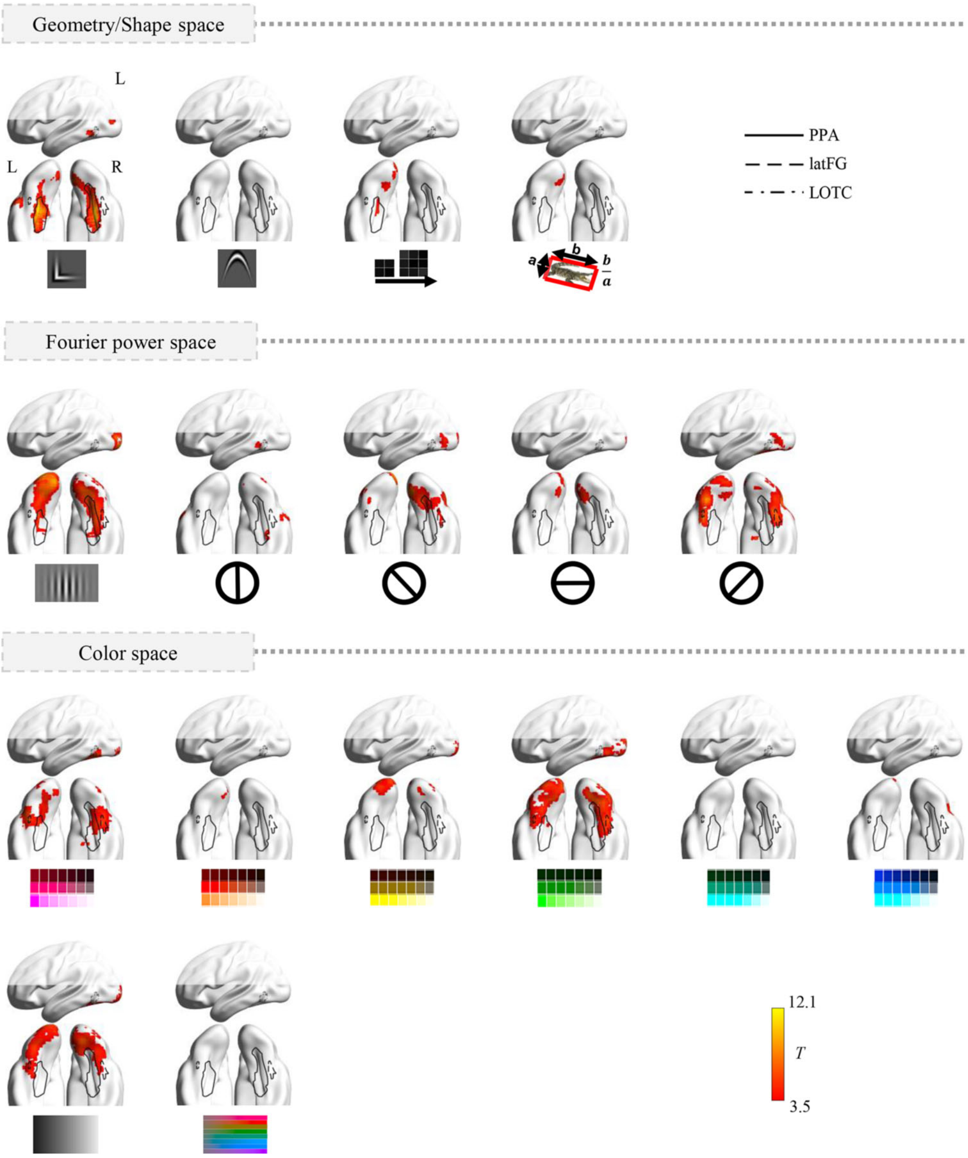

Fig. 2 Object visual feature topography in a full-model parametric modulation analysis. All visual feature weights were entered into the parametric modulation model for BOLD activity estimates, yielding an activation map for each visual feature in the VOTC mask. The maps are thresholded at cluster-level FWE-corrected $P < 0.05$ within the VOTC mask, with voxel-wise $P < 0.001$. The outlines show the object-domain-preferring clusters for animals (bilateral latFG), large artifacts (bilateral PPA), and small manipulable artifacts (left LOTC), localized by contrasting each object domain with the other two domains in the main fMRI experiment.

associated with the non-visual response-action properties by behavioral ratings and computations of natural images. A binary-labeled "domain" model was also tested as a reference. We then tested whether the visual feature combination patterns associated with response-action and binary domain categorization indeed aligned with the visual feature organization of the VOTC.

Prototypical Visual-feature Vectors for Response-actions and Domains

To gain an unbiased understanding of the feature distribution among objects, we built a larger object image dataset containing 672 images from three previous image sets [45–47] and the 95 images from our main fMRI experiment. We used these image sets because they had isolated objects presented on a white background. One object image was the same in our current experiment and in Downing et al. [45] and thus only one of these was included. There were 419 animals (mammals, marine creatures, birds, insects, fish, and reptiles) and 348 inanimate man-made artifacts (168 large artifacts and 180 small manipulable artifacts, including buildings, furniture, appliances, communal facilities, large transportation, common household tools, kitchen utensils, and accessories). All images were re-sized to 256 × 256 pixels with 72 dots per inch using Adobe Photoshop CS6 (Adobe, San Jose, USA). For each image, the feature weights were measured using computational vision models, as described above for the main fMRI experiment stimuli.

For response-driven prototypical visual-feature vectors, we examined three theorized response-action systems:

fight-or-flight, navigation, and manipulation [12, 23, 25, 48]. The relevance of the 767 images (the set described above) to each response-action system was rated by an independent group of participants (n = 24, age range 18–29 years, 15 females) on a 1–5 scale. For fight-or-flight, the participants were asked to rate "to what extent the object depicted in image would make you to show a stress-related response, e.g., run away, attack, or freeze." For navigation, the participants were asked to rate "to what extent the object depicted in image could offer spatial information to help you explore the environment." For manipulation, the participants were asked to rate "to what extent the object depicted in image can be grasped easily and used with one hand." The ratings were averaged across participants to get one relevance index for each image to each response. Then pairwise linear regression analyses were conducted between each response-action relevance type and each visual feature, resulting in 3 response-feature beta vectors.

For domain-driven prototypical visual-feature vectors, we constructed corresponding domain labels (e.g., for the animacy domain model: animate = 1, inanimate = 0; for the animal, large artifact and small artifact tripartite domain model: animals = 1, others = 0) and performed pairwise logistic regression analyses between each domain label and each visual feature, resulting in domain-feature beta vectors.

Effects of Response-driven and Domain-driven Models on VOTC Feature Topography

We examined whether the visual feature profiles of VOTC voxels could be driven by response- and/or domain-feature vectors based on natural image statistics. The visual feature profile of VOTC voxels was generated by extracting group-averaged parameter estimates of each feature in independent models, which resulted in a neural-feature beta vector for each voxel. The independent model results were used here because the beta values were more transparently interpretable and the results of the independent models were largely consistent with those of the full model. Then the neural-visual feature vectors were correlated with each of three response-feature vectors or the object-domain-feature vectors using Pearson correlation, resulting in three r maps for response-driven and tripartite-domain-driven models, and two r maps for the animacy-domain-driven model. The significance of the r maps above zero (one-tailed) was thresholded at cluster-level FWE-corrected P <0.05 within the VOTC mask with voxel-wise P <0.001. To compare the explanatory powers between response-driven and domain-driven hypotheses, the three r maps derived from response-feature vectors were collapsed by extracting the highest r-value in each voxel in the VOTC and the two (or three) r maps derived from domain-feature vectors were also collapsed using the same method. The resulting two maximum r maps were then Fisher-Z transformed and compared using the paired t-test across voxels. Because the primary focus of the current study was the higher-order VOTC, we also compared the explanatory power between the two hypotheses in a higher-order VOTC mask which included the region before −71 on the y-axis of the MNI coordinates in our VOTC mask (containing 2121 voxels at a 3-mm voxel size) [49] (when the higher-order VOTC was defined by excluding Broca's areas 17 and 18, the results were very similar).

Association Between the Effects of Visual Features and of Domains in the VOTC

To test whether the documented domain selectivity (e.g., animate *vs* inanimate) might reflect the effects of visual features, we determined whether the visual feature vector of a voxel could predict its domain preference strength across different datasets. We used an independent dataset, in which 31 healthy participants viewed blocks of gray-scale pictures of animals and artifacts, collapsing the two healthy control group data from our published studies that used identical data acquisition procedures [50, 51]. The data were preprocessed with the procedure described above and the preprocessed functional images of each participant were entered into a GLM that included onset regressors for animate and inanimate together with 6 head-motion regressors. Contrast images for animate *vs* inanimate items were then calculated and averaged across participants to construct the VOTC map of animacy-domain selectivity. In the linear regression model (in SPSS), the VOTC map of domain selectivity (from this independent dataset) was treated as the dependent variable and the 17 full-model visual feature maps (from the main fMRI experiment) were included as independent variables. The adjusted-R^2 of the regression model is reported.

To further test the effects of a tripartite structure [4], using the independent fMRI dataset, we constructed a GLM that included onset regressors for animals, large artifacts, and small manipulable artifacts as well as 6 head-motion regressors. Contrast images for each object domain with the other two object domains were calculated and averaged across participants to construct three VOTC maps of tripartite-domain selectivity. The associations between each of these maps and 17 full-model visual feature maps were then tested using the same procedure.

The effects of animacy structure and tripartite structure were also tested using similar procedures in the higher-order VOTC mask where the domain effects were mainly located.

Results

Twenty visual features covering a broad range of shape, spatial frequency, orientation, and color information were tested, and their weights were extracted for each of 95 object images using computational vision models (see Materials and Methods and Fig. S1) [14, 17, 33, 36, 38, 41, 52]. fMRI responses for these images were also obtained from 26 participants, and parametric modulation models were used to compute the effects of visual features across VOTC voxels, taking into consideration their inter-correlations (Fig. 1). Then an explicit theoretical hypothesis for VOTC computation (visual-feature for action-response mapping) was tested for explanatory power for the VOTC visual feature patterns. The relation between the feature effects and the domain effects was also examined.

Computation of Visual Feature Weights in Object Images

A set of 95 real object images (32 animate items and 63 inanimate artifacts, including 28 large artifacts and 35 small manipulable artifacts) were analyzed using computational vision models to obtain their properties for 20 visual features: in geometry/shape space these features were right-angle, curvature, number of pixels, and elongation; in Fourier power space, high/low spatial frequencies and four orientations (0°, 45°, 90°, and 135°); in color space, eight hues, luminance, and chroma. The descriptive statistics, including distribution plots for each feature across the whole image set, as well as the mean and SD by domains, are shown in Fig. S3. The Pearson correlations among features are shown in Fig. S4, left panel (note this correlation matrix was highly correlated ($r = 0.84$) with the correlation matrix derived from a broader image set, indicating adequate representativeness of the current image sample). As often reported, we found significant differences between animate items and inanimate artifacts (Welch t-test and FDR corrected $q < 0.05$) across three visual features: right-angle ($t_{(63.41)} = -3.96$, $P = 1.90 \times 10^{-4}$), elongation ($t_{(68.60)} = -3.97$, $P = 1.74 \times 10^{-4}$), and 135° orientation ($t_{(39.85)} = 3.12$, $P = 3.33 \times 10^{-3}$). When separating the inanimate objects further into large artifacts and small manipulable artifacts, more features exhibited significant between-domain differences (one-way ANOVA and FDR corrected $q < 0.05$): right-angle ($F_{(2,92)} = 6.77$, $P = 0.002$), number of pixels ($F_{(2,92)} = 16.37$, $P = 8.27 \times 10^{-7}$), and elongation ($F_{(2,92)} = 15.47$, $P = 1.61 \times 10^{-6}$) in geometry/shape space; low spatial frequency ($F_{(2,92)} = 6.59$, $P = 0.002$), 0° orientation ($F_{(2,92)} = 6.21$, $P = 0.003$), 90° orientation ($F_{(2,92)} = 5.08$, $P = 0.008$), and 135° orientation ($F_{(2,92)} = 8.06$, $P = 0.001$) in Fourier power space; orange ($F_{(2,92)} = 5.11$, $P = 0.008$) and yellow ($F_{(2,92)} = 5.43$, $P = 0.006$) in color space. The *post hoc* comparisons across domain pairs are shown in Table S1. Pairs of highly-correlated visual features (Pearson $r > 0.85$) were collapsed into one by taking the means (cyan/indigo, $r = 0.92$, red/purple, $r = 0.86$). To reduce the chance of multicollinearity, low spatial frequency was further excluded from the full parametric modulation model analysis because it had a VIF >10 [44] (VIF = 48.25; the VIFs of other features were within the range 1.26–5.41). Thus, 17 features were retained the subsequent parametric modulation analysis, with pairwise correlations in the range −0.56 to 0.64.

Topographic Map of Visual Features in the VOTC

For all the fMRI results below, we adopted a threshold of cluster-level FWE corrected $P < 0.05$ within the VOTC mask [43], with voxel-wise $P < 0.001$ unless explicitly stated otherwise.

The results of the full model analysis, where the 17 visual feature weights were entered into the parametric modulation model for BOLD (blood-oxygen-level-dependent) activity estimates, are shown in Fig. 2. In the higher-order VOTC, for geometry/shape-space features, right-angle modulated responses in the bilateral medial fusiform gyrus (medFG) and left LOTC; number of pixels modulated responses in the left medFG. For Fourier-power-space features, high spatial frequency modulated responses in the bilateral medFG; 0° orientation modulated responses in the right medFG and bilateral LOTC; oblique orientations (45° and 135°) modulated responses in the right latFG and 135° orientation additionally modulated responses in the left latFG. For color-space features, red/purple and green modulated broad regions in the bilateral FG; red/purple additionally modulated responses in the left LOTC; luminance modulated responses in the bilateral latFG.

Independent models, in which each feature was entered into the parametric modulation model separately without considering the correlations among features, were also used and are shown in Fig. S5. Here, more commonalities across features were found, most features showing regions largely consistent with those obtained in the full model above with effects covering broader regions in the higher-order VOTC. Five features showed differences between the two analyses: The effects of elongation (in the left LOTC), 90° orientation (in the bilateral medFG), and blue (in the bilateral medFG) were significant in the independent model but not in the full model, while the effects of 0° orientation (in the bilateral LOTC) and luminance (in the bilateral latFG) were significant in the full model but not in the independent model. These differences are likely due to their correlations with other features (Fig. S4).

Factors Driving the Feature Distribution Patterns in VOTC Voxels

We have described the distributional topography of a comprehensive set of visual features in the VOTC, identifying voxels by their sensitivities to specific feature sets. The follow-up question is why visual features are distributed in this way. Why does a certain region prefer a particular set of features together (e.g., right-angle, high spatial frequency, and 0° orientation in the medFG), or why is a particular feature set preferred by the same brain region in a particular location? We tested the explicit hypothesis that the neuronal functional preference of VOTC voxels for certain visual features is constrained, at least partly, by downstream computations [12, 23–31]. That is, the pattern of visual features associated with various types of response-actions in the real world constrains the visual feature distribution pattern in the VOTC. We first obtained the type of visual feature clustering patterns associated with the nonvisual response-action properties by behavioral ratings and computations of natural images (procedure in Fig. 3A; results in the radial maps in Fig. 3B, left panels). A binary-labeled "animacy domain" model (e.g., animate = 1, inanimate = 0) was also tested as a reference (Fig. S6, left panels). We then tested whether the patterns of visual features associated with response- and animacy-domain categorization indeed align with the visual feature topography of the VOTC (Figs 3B and S6).

Relationship Between Observed VOTC Voxel Feature Vectors and Response-Feature Vectors

We obtained prototypical visual feature vectors associated with hypothesized response-action systems (i.e., response-feature vectors) based on subjective rating and natural image statistics to approximate the feature-response association profile in the real world. For this purpose, we used a broader image set containing 767 images, which included the 95 from the current fMRI experiment and 672 (isolated objects with clear domain membership on a white background) selected from three previous studies [45–47]. Three types of response-actions were considered: fight-or-flight, navigation, and manipulation [12, 23, 25, 48]. We first asked an independent group of participants to rate each object image on its relevance to each response. Linear regression was conducted between the rated value for each response-action system and each of the 18 visual feature weights; the resulting beta values were used as the prototypical response-feature vector for each response-action system in natural object images. The results (FDR corrected $q < 0.05$) are shown in Fig. 3B, left panels (see Table S2 for beta and P-values).

We then tested whether and in what manner the visual feature sensitivity profiles of the VOTC voxels align with these response-associated feature vectors. For each VOTC voxel we correlated its neural-feature vector with each of the three prototypical response-feature vectors, resulting in 3 VOTC correlation maps (Fig. 3B). Note that the independent parametric modulation model results were used (with 18 visual features), because the beta weights were more transparently interpretable and the results were also largely consistent with those of the full model (17 features, with "low spatial frequency" excluded due to a high VIF). The navigation-response-feature vector was significantly correlated with the neural-feature vector in clusters located in the bilateral medFG, left middle occipital gyrus, and right lingual gyrus. The manipulation-response-feature vector was significantly correlated with the neural-feature vector of VOTC voxels in clusters located in the left LOTC and left lingual gyrus. The fight/flight-response-feature vector showed no significant correlations at the standard threshold. When we lowered the threshold (voxel-wise $P < 0.01$, cluster size >10), it was correlated with the neural-feature vector of VOTC voxels in the bilateral lateral occipital cortex, right latFG, and bilateral occipital pole.

Relationship Between Observed VOTC Voxel Feature Vectors and Animacy-domain-feature Vectors

Although the specific computation goals served by the "domain" organization (i.e., driving factor) are not articulated, we nonetheless generated prototypical visual feature vectors associated with the widely-proposed categorization (i.e., domain-feature vectors) based on natural image statistics to approximate the feature-domain association profile in the real world. Binary domain labels were constructed. Logistic regression was conducted between each binary domain label and each of the visual feature weights across the 767 images noted above, and the resulting beta values were taken as a prototypical domain-feature vector for each domain. This vector reflected the visual feature patterns that best distinguished animate and inanimate items in natural object images. The results (FDR corrected $q < 0.05$) are shown in Fig. S6, left panels; see Table S2 for beta and P-values). We then tested whether the visual feature sensitivity profiles of the VOTC voxels reflect these animacy-domain-feature vectors. For each VOTC voxel, we correlated its neural-feature vector (obtained using the independent parametric modulation model; Fig. S5) and each of the two prototypical domain-feature vectors, resulting in a correlation map for each domain (Fig. S6). The animate-domain-feature vector was

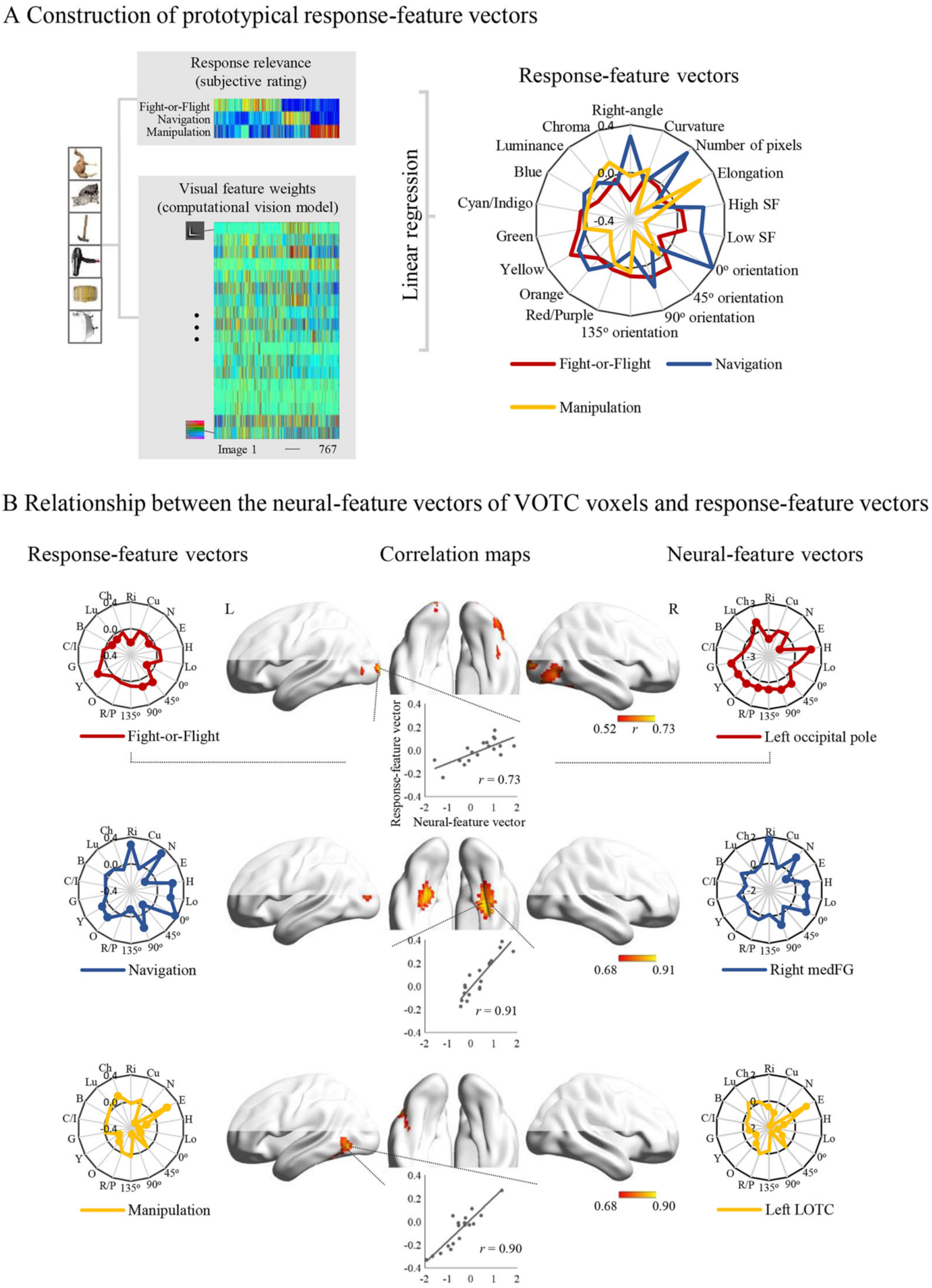

Fig. 3 Relationship between the response-driven model and the visual feature topography of the VOTC. **A** Construction scheme of prototypical visual-feature vectors for the three response-action systems. In an image set of 767 images, visual feature weights for each image were obtained using computational vision models. We examined 3 theorized response systems (fight-or-flight, navigation, and manipulation) by asking 24 participants to rate how strongly each object is associated with each of the three response-action systems. Linear regressions were conducted between each response vector and each visual feature weight, resulting in 3 response-feature beta vectors. **B** Left panels, "prototypical" visual feature vectors associated with each response-action system (fight-or-flight, navigation, and manipulation). Dots indicate that beta values were significant at FDR-corrected $q < 0.05$ for 54 comparisons. Middle panels, Pearson correlation maps between each of these "prototypical" response-driven-feature vectors and the neural-feature vectors of VOTC voxels obtained from the fMRI parametric modulation analyses. The correlation maps are thresholded at cluster-level FWE-corrected $P < 0.05$, voxel-wise $P < 0.001$ for the navigation-driven and manipulation-driven vectors, and voxel-wise $P < 0.01$, cluster size >10 for the fight/flight-driven vector. Scatter plots show the correlations for the peak voxels. Right panels, peak neural-feature vectors of voxels.

significantly correlated with the neural-feature vector of the VOTC voxels in one cluster located in the right lateral occipital cortex. The inanimate-domain-feature vector was significantly correlated with the neural-feature vector of the VOTC voxels in three clusters located in the bilateral medFG and left LOTC. These results suggest that the feature-sensitivity patterns of VOTC voxels are associated with the natural image statistics of two major object domains.

Comparison of Response-driven and Animacy-domain-driven Hypotheses

We directly compared the explanatory power of these two types of feature vector to see if, by being more specific, the response-driven model captures finer properties of the VOTC visual feature topography. To do this, we first generated a response-driven maximum r map by selecting the highest r value for each voxel out of the three response-driven r maps shown in Fig. 3B, and generated the animacy-domain-driven maximum r map in the same way using the two maps shown in Fig. S6. Then the two max r maps were Fisher-Z transformed and compared using the paired t-test. The results showed that the response-driven map had significantly higher r-values than the animacy-domain-driven map (global mean $r \pm$ SD: 0.57 ± 0.27 *vs* 0.39 ± 0.23, $t_{(3914)} = 34.87$, $P = 2.87 \times 10^{-232}$). The same analysis was performed within the higher-order VOTC (region anterior to y = −71 on the MNI coordinates in the VOTC mask [49]) and the results were similar (global mean of $r \pm$ SD: 0.58 ± 0.33 *vs* 0.37 ± 0.24, $t_{(2120)} = 25.54$, $P = 1.14 \times 10^{-125}$).

Inanimate artifacts have been further divided into large artifacts and small manipulable artifacts in recent studies that showed a tripartite structure of large artifacts, animals, and small manipulable artifacts in the VOTC, spanning from the medial fusiform/parahippocampal gyrus to the LOTC [4, 43, 50]. We also tested whether the response-driven model has greater explanatory power than the tripartite-domain-driven model because they strongly correspond (fight-or-flight responses with animals; navigation with large artifacts; and manipulation with small manipulable artifacts). Procedures similar to the previous analysis were repeated and the results showed that the response-driven map still had significantly higher r values than the tripartite-domain-driven map (mean ± SD: 0.57 ± 0.27 *vs* 0.54 ± 0.28, $t_{(3914)} = 12.21$, $P = 1.07 \times 10^{-33}$; see Fig. S7 for results of prototypical tripartite-domain-feature vectors and correlation maps). When the analysis was restricted in the higher-order VOTC, the results also held: mean ± SD: 0.58 ± 0.33 *vs* 0.57 ± 0.33, $t_{(2120)} = 3.02$, $P = 0.003$.

Association Between Visual Feature Effects and Domain Effects in the VOTC

Having established the topography of visual features in the VOTC and tested the driving variables for such distributions, here we assessed to what extent the well-established object-domain observations (i.e., animacy and size) can be accounted for by the underlying feature representations.

A multiple linear regression model was constructed to predict a voxel's selectivity strength for object domains (obtained from an independent fMRI dataset) using its visual feature sensitivity patterns, across all VOTC voxels. That is, the 17-feature sensitivity maps in the VOTC from the full parametric modulation model were taken as the independent variables. The dependent variable was the VOTC animacy-domain-selectivity strength map obtained from an independent dataset (contrasting animate items with inanimate items; see details in [50, 51]). The results (Fig. 4A) showed high explanatory power of the linear regression model: adjusted-$R^2 = 0.815$. Using the animacy-domain-selectivity strength map computed from the main fMRI experimental data with the identical contrast (i.e., within-subject analysis) yielded an adjusted-R^2 of 0.959.

To predict the tripartite-structure (animals, large artifacts, small manipulable artifacts), the dependent variable was obtained by contrasting the beta values of each domain with the mean of the other two. Again, using the independent dataset, a voxel's visual feature vector highly significantly predicted its selectivity strength (Fig. 4B) for animals (adjusted-$R^2 = 0.816$), for large artifacts (adjusted-$R^2 = 0.772$), and for small manipulable artifacts (adjusted-$R^2 = 0.694$). The results were higher using data from the same main fMRI experiment (for animals, adjusted-$R^2 = 0.957$; for large artifacts selectivity, adjusted-$R^2 = 0.946$; for small manipulable artifacts selectivity, adjusted-$R^2 = 0.973$). When the analysis was restricted to the higher-order VOTC, all results remained similar (see adjusted-R^2 in Table S3).

Discussion

Combining computational vision models, a parametric modulation analysis of fMRI data, and natural image statistics, we depicted the distributional topography of a comprehensive set of visual features (geometry/shape, Fourier power, and color) in the VOTC, identifying the sensitivities of voxels to specific feature sets. We demonstrated that the relationship with salient response actions in the real world offers one possible explanation of why visual features are distributed this way in the VOTC.

In contrast to recent studies that focused on one or two specific visual features or unarticulated deep neural

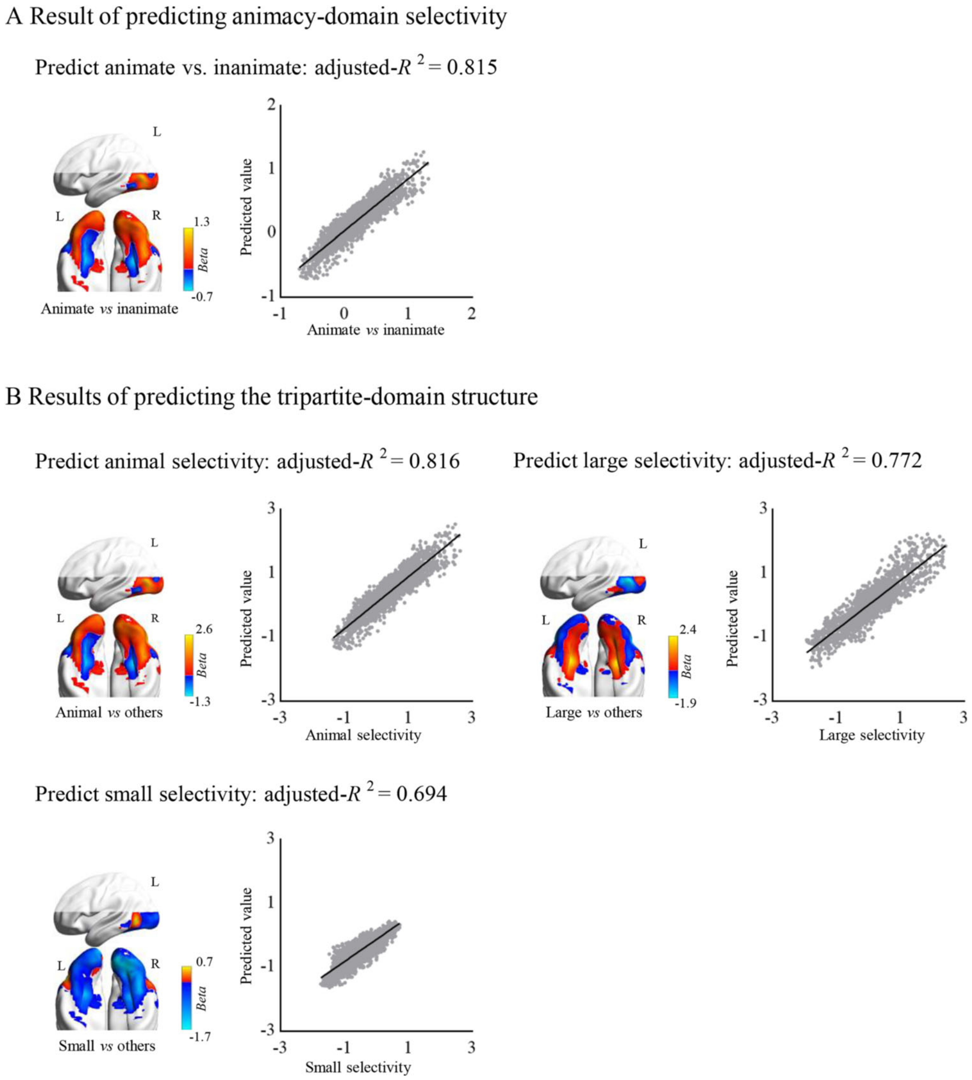

Fig. 4 The association between visual feature topography and object domain effects. **A** Result of using visual feature vectors of the VOTC to predict animacy-domain selectivity: a multiple linear regression model was constructed to predict domain selectivity strength for animate/inanimate domains, using the beta values of 17 visual features as predictors, across all VOTC voxels. The brain map is the unthresholded animate *vs* inanimate activation map, showing the group-averaged selectivity strength (beta values of animate items – inanimate items for all VOTC voxels). The scatter plot shows the correlation between predicted animacy-domain-selectivity strength using VOTC visual-feature maps and the observed domain-selectivity strength across all VOTC voxels. **B** Results of using visual feature vectors to predict the tripartite-domain structure: three multiple linear regression models were constructed to predict domain-selectivity strength for animals, large artifacts, or small manipulable artifacts, using the beta values of 17 visual features as predictors, across all VOTC voxels. The brain maps are the unthresholded activation maps for animals, large artifacts, or small manipulable artifacts, showing the group-averaged selectivity strength (beta values of one domain – those of the other two) for all VOTC voxels. The scatter plots show correlations between predicted domain-selectivity strength using VOTC visual-feature maps and the observed domain-selectivity strength across all VOTC voxels.

network-derived hidden spaces [13, 53], our approach tested a much more comprehensive set of visual features and the correlations among them, and showed highly significant explanatory power for the well-documented object domain structure. Previous studies have shown associations between certain visual features and domain preferences: A preference for rectilinearity, high spatial frequency, and cardinal orientation features has been reported in regions preferring scenes/large objects [14, 36, 37, 54] and a preference for high curvature, low spatial frequency, and red/yellow hues in regions preferring faces [17, 21, 36, 55]. However, these visual features do not explain the domain observations satisfactorily, in terms of the selectivity strengths [22], the anatomical overlap [17], and the domain preference effects that are still present when visual shape is controlled [10]. Here, by incorporating the combinational effects of multiple visual features together, we showed remarkably high explanatory power of visual features for domain-preference: the visual-feature-preference vectors of voxels accounted for >80% of the variance in the selectivity of the VOTC for the animate/inanimate domain selectivity, and >69% of the variance in selectivity for animals, large artifacts or small manipulable artifacts. Our results not only provide a computational model that theoretically may predict the VOTC neural activity pattern for objects based on their visual feature properties, including those along fuzzy domain boundaries, but also offer positive evidence for a plausible, specific representation theory of the VOTC that can explain the domain-like phenomenon what the VOTC represents (at least) is visual features.

Why does the VOTC have this specific type of visual feature topography then? We provided evidence that is consistent with the recent proposal that the neuronal functionality of VOTC voxels is constrained, at least partly, by the association pattern with downstream nonvisual, action computations such as fight-or-flight, navigation, and manipulation [12, 23–31]. Prototypical visual feature sets that were associated with the three types of response actions, obtained through rating and natural image statistics, indeed aligned with the preferred visual feature combination patterns in different patches of the VOTC: the fight/flight-response vector (the 3 highest loadings were in yellow, 45° orientation, and 90° orientation; and the 3 lowest loadings were in right-angle, 0° orientation, and blue) was associated with the right latFG, bilateral lateral occipital cortex and bilateral occipital pole; the navigation-response vector (the 3 highest loadings in 0° orientation, number of pixels, and right-angle; and the 3 lowest loadings in elongation, 135° orientation, and 45° orientation) was associated with the bilateral medFG, left middle occipital gyrus, and right lingual gyrus; and the manipulation-response vector (the 3 highest loadings in elongation, chroma, and luminance; and the 3 lowest loadings in number of pixels, 90° orientation, and low spatial frequency) with the left LOTC and left lingual gyrus. That is, how sensitivity to visual features is organized in VOTC neurons is aligned with how visual features map with down-stream action responses. Those visual features (combinations) that tend to indicate and be associated with a certain action response (e.g., manipulation) are preferentially processed and represented, together, in regions that are optimally connected with the corresponding action systems [12, 23–31].

It should be emphasized that we interpret our results as showing that the representation in the VOTC is of visual features, organized in a way that allows them to optimize mapping with (i.e., driven by) the response-action programs, and not the action programs themselves. This is in line with the vast literature demonstrating that the VOTC is important for visual processing and that damaging dorsal regions, and not the VOTC, leads to action deficits. Also worth noting is that these response models are clearly associated with the object domains that have been used to label the VOTC selectivity [2, 4]: fight/fight responses with animals; navigation responses with large objects; and manipulation responses with small objects. We treat this “domain” structure as a result to be explained rather than an explanatory theory, because it is descriptive, vaguely defined, and does not offer a hypothesis about exactly what information is represented here. The “visual-feature-driven-by-action-mapping” account not only explains this result, but also makes predictions that are consistent with a series of results comparing the feature *vs* domain effects in the literature: objects that do not have the prototypical shape of a domain (e.g., a cup shaped like a cow) are processed by the VOTC more similarly to items sharing its surface shape (e.g., animal in this case) and not to those in the same domain (regular cups) [56]; the animate-preferring areas are modulated by how “typical” (human-like) animals are [57]; features without domain contexts may still be able to produce effects [14, 17, 37]. Our supplementary analyses of the main fMRI experiment and analyses of the feature-validation fMRI experiment provided further support for this last point (Figs S8, S9): The feature effects were largely present when regressing out domain structure; The effect of right-angle in the bilateral medFG (aligned with the PPA) was present when the features were shown in isolation without object contexts and/or other features, and even during the presentation of objects from non-preferred domains (i.e., when objects were small manipulable artifacts or animals). Interestingly, the effects of other features such as hue and orientation were only found when presented within objects, and disappeared when shown in isolation, indicating that they are processed in combination with other visual features and/or object contexts in the VOTC [19].

There are two caveats to consider. One is that the visual features we tested were based on knowledge and algorithms from computational vision practice and the relatively low-level visual features that have been considered in VOTC research. There is always a possibility that other relevant types of visual feature were missed, and that the algorithm choice was not optimal. For instance, the current curvature computation considers 5 arbitrarily-selected concavity features, and its effects on the VOTC based on this computation were not significant yet were visible when using a direct contrast (top 25% amount of curvature – top 25% amount of right-angle; Fig. S10), more in line with studies using subjective curvature ratings, which may reflect a composite index of various types of curvature [13]. There are almost infinite potential (unarticulated) mid-level or high-level visual features (e.g., see discussion in Kourtzi and Connor, 2011 [32]) that are untested, such as circles, texture patterns, or eye-like or mouth-like patterns [58, 59]. In this context, the deep convolutional neural network (DCNN), in which the features extracted by various layers have been well studied and visualized in the computer vision field, offers a special opportunity. In a typical multi-layer DCNN (for instance, AlexNet), the first or second convolution layers might extract some simple edge/line features of various orientations and scales, while the third/fourth/fifth convolution layers model more complicated visual features, such as corners, circles/ellipses, or even sub-components shared by many objects [60]. These features are promising candidates for future studies. Importantly, however, our result that the feature combination model highly significantly predicts the domain-preference strength in VOTC voxels indicates the power of the included features. Furthermore, the theoretical framework we developed based on the principle of observed feature organization (i.e., response-constraint) may lead to a more productive approach to identifying effective features (e.g., specific features that may be associated with a particular response), and constrain the type of DCNN model to be adopted (e.g., training for object classification or responses). Another caveat is that we only examined the major common objects domains, and did not test other classical domains for the VOTC: scenes and faces. The current framework makes the same predictions about preferences for these two types of images, which remain to be empirically tested.

To conclude, we found that there are systematic patterns of various visual feature sensitivity across the VOTC, offering a comprehensive visual feature topography map. Such visual feature topography is aligned with how features map onto different types of response actions. The object-domain-related results can be largely explained by voxel sensitivity patterns to the visual features. These findings led us to propose a visual-feature based representation in the VOTC, driven by their association with nonvisual-action computations (stored elsewhere).

Acknowledgements This work was supported by the National Natural Science Foundation of China (31671128, 31925020, 31700999, 31700943, and 31500882), the Changjiang Scholar Professorship Award (T2016031), and Fundamental Research Funds for the Central Universities (2017EYT35). We thank Shiguang Shan for discussions about computational vision models, and Joshua B. Julian for generously sharing codes for the right-angle and curvature computations.

Conflict of interest The authors claim that there are no conflicts of interest.

References

1. Felleman DJ, van Essen DC. Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1991, 1: 1–47.
2. Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, *et al*. Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron 2008, 60: 1126–1141.
3. Kanwisher N. Functional specificity in the human brain: A window into the functional architecture of the mind. Proc Natl Acad Sci U S A 2010, 107: 11163–11170.
4. Konkle T, Caramazza A. Tripartite organization of the ventral stream by animacy and object size. J Neurosci 2013, 33: 10235–10242.
5. Grill-Spector K, Weiner KS. The functional architecture of the ventral temporal cortex and its role in categorization. Nat Rev Neurosci 2014, 15: 536–548.
6. Levy I, Hasson U, Avidan G, Hendler T, Malach R. Center-periphery organization of human object areas. Nat Neurosci 2001, 4: 533–539.
7. Hasson U, Levy I, Behrmann M, Hendler T, Malach R. Eccentricity bias as an organizing principle for human high-order object areas. Neuron 2002, 34: 479–490.
8. Bracci S, Op de Beeck H. Dissociations and associations between shape and category representations in the two visual pathways. J Neurosci 2016, 36: 432–444.
9. Kaiser D, Azzalini DC, Peelen MV. Shape-independent object category responses revealed by MEG and fMRI decoding. J Neurophysiol 2016, 115: 2246–2250.
10. Proklova D, Kaiser D, Peelen MV. Disentangling representations of object shape and object category in human visual cortex: The animate-inanimate distinction. J Cogn Neurosci 2016, 28: 680–692.
11. Bracci S, Ritchie JB, de Beeck HO. On the partnership between neural representations of object categories and visual features in the ventral visual pathway. Neuropsychologia 2017, 105: 153–164.
12. Peelen MV, Downing PE. Category selectivity in human visual cortex: Beyond visual object recognition. Neuropsychologia 2017, 105: 177–183.
13. Long B, Yu CP, Konkle T. Mid-level visual features underlie the high-level categorical organization of the ventral stream. Proc Natl Acad Sci USA 2018, 115: E9015–E9024.
14. Nasr S, Echavarria CE, Tootell RB. Thinking outside the box: Rectilinear shapes selectively activate scene-selective cortex. J Neurosci 2014, 34: 6721–6735.
15. Perrinet LU, Bednar JA. Edge co-occurrences can account for rapid categorization of natural *versus* animal images. Sci Rep 2015, 5: 11400.
16. Long B, Störmer VS, Alvarez GA. Mid-level perceptual features contain early cues to animacy. J Vis 2017, 17: 20.
17. Yue X, Pourladian IS, Tootell RB, Ungerleider LG. Curvature-processing network in macaque visual cortex. PNAS 2014, 111: E3467–E3475.
18. Rosenthal I, Ratnasingam S, Haile T, Eastman S, Fuller-Deets J, Conway BR. Color statistics of objects, and color tuning of object cortex in macaque monkey. J Vis 2018, 18: 1.
19. Lafer-Sousa R, Conway BR, Kanwisher NG. Color-biased regions of the ventral visual pathway lie between face- and place-selective regions in humans, as in macaques. J Neurosci 2016, 36: 1682–1697.
20. Lafer-Sousa R, Conway BR. Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. Nat Neurosci 2013, 16: 1870–1878.

21. Chang L, Bao P, Tsao DY. The representation of colored objects in macaque color patches. Nat Commun 2017, 8: 2064.
22. Bryan PB, Julian JB, Epstein RA. Rectilinear edge selectivity is insufficient to explain the category selectivity of the parahippocampal place area. Front Hum Neurosci 2016, 10: 137.
23. Mahon BZ, Caramazza A. What drives the organization of object knowledge in the brain? Trends Cogn Sci 2011, 15: 97–103.
24. Abboud S, Maidenbaum S, Dehaene S, Amedi A. A number-form area in the blind. Nat Commun 2015, 6: 6026.
25. Bi YC, Wang XY, Caramazza A. Object domain and modality in the ventral visual pathway. Trends Cogn Sci 2016, 20: 282–290.
26. Osher DE, Saxe RR, Koldewyn K, Gabrieli JD, Kanwisher N, Saygin ZM. Structural connectivity fingerprints predict cortical selectivity for multiple visual categories across cortex. Cereb Cortex 2016, 26: 1668–1683.
27. Bouhali F, Thiebaut de Schotten M, Pinel P, Poupon C, Mangin JF, Dehaene S, *et al*. Anatomical connections of the visual word form area. J Neurosci 2014, 34: 15402–15414.
28. Mahon BZ, Milleville SC, Negri GA, Rumiati RI, Caramazza A, Martin A. Action-related properties shape object representations in the ventral stream. Neuron 2007, 55: 507–520.
29. Chen QJ, Garcea FE, Almeida J, Mahon BZ. Connectivity-based constraints on category-specificity in the ventral object processing pathway. Neuropsychologia 2017, 105: 184–196.
30. Stevens WD, Tessler MH, Peng CS, Martin A. Functional connectivity constrains the category-related organization of human ventral occipitotemporal cortex. Hum Brain Mapp 2015, 36: 2187–2206.
31. Hutchison RM, Culham JC, Everling S, Flanagan JR, Gallivan JP. Distinct and distributed functional connectivity patterns across cortex reflect the domain-specific constraints of object, face, scene, body, and tool category-selective modules in the ventral visual pathway. Neuroimage 2014, 96: 216–236.
32. Kourtzi Z, Connor CE. Neural representations for object perception: Structure, category, and adaptive coding. Annu Rev Neurosci 2011, 34: 45–67.
33. Zachariou V, Del Giacco AC, Ungerleider LG, Yue XM. Bottom-up processing of curvilinear visual features is sufficient for animate/inanimate object categorization. J Vis 2018, 18: 3.
34. Canny J. A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 1986, PAMI-8: 679–698.
35. Krüger N. Learning object representations using a priori constraints within ORASSYLL. Neural Comput 2001, 13: 389–410.
36. Rajimehr R, Devaney KJ, Bilenko NY, Young JC, Tootell RB. The "parahippocampal place area" responds preferentially to high spatial frequencies in humans and monkeys. PLoS Biol 2011, 9: e1000608.
37. Nasr S, Tootell RB. A cardinal orientation bias in scene-selective visual cortex. J Neurosci 2012, 32: 14921–14926.
38. Goffaux V, Duecker F, Hausfeld L, Schiltz C, Goebel R. Horizontal tuning for faces originates in high-level Fusiform Face Area. Neuropsychologia 2016, 81: 1–11.
39. Canário N, Jorge L, Loureiro Silva MF, Alberto Soares M, Castelo-Branco M. Distinct preference for spatial frequency content in ventral stream regions underlying the recognition of scenes, faces, bodies and other objects. Neuropsychologia 2016, 87: 110–119.
40. Berman D, Golomb JD, Walther DB. Scene content is predominantly conveyed by high spatial frequencies in scene-selective visual cortex. PLoS One 2017, 12: e0189828. https://doi.org/10.1371/journal.pone.0189828.
41. Brouwer GJ, Heeger DJ. Decoding and reconstructing color from responses in human visual cortex. J Neurosci 2009, 29: 13992–14003.
42. Yan CG, Wang XD, Zuo XN, Zang YF. DPABI: data processing & analysis for (resting-state) brain imaging. Neuroinformatics 2016, 14: 339–351.
43. Wang XY, Peelen MV, Han ZZ, He CX, Caramazza A, Bi YC. How visual is the visual cortex? comparing connectional and functional fingerprints between congenitally blind and sighted individuals. J Neurosci 2015, 35: 12545–12559.
44. Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate data analysis. Pearson New International Edition. Edinburgh Gate Harlow: Pearson Education Limited, 2014: 200–201.
45. Downing PE, Chan AWY, Peelen MV, Dodds CM, Kanwisher N. Domain specificity in visual cortex. Cereb Cortex 2006, 16: 1453–1461.
46. Moreno-Martínez FJ, Montoro PR. An ecological alternative to Snodgrass & Vanderwart: 360 high quality colour images with norms for seven psycholinguistic variables. PLoS One 2012, 7: e37527. https://doi.org/10.1371/journal.pone.0037527.
47. Brodeur MB, Guérard K, Bouras M. Bank of Standardized Stimuli (BOSS) phase II: 930 new normative photos. PLoS One 2014, 9: e106953. https://doi.org/10.1371/journal.pone.0106953.
48. Tang N, Ding YF, Zhang W, Hu J, Xu XH. Stay active to cope with fear: A cortico-intrathalamic pathway for conditioned flight behavior. Neurosci Bull 2019, 35: 1116–1119.
49. Thorat S, Proklova D, Peelen MV. The nature of the animacy organization in human ventral temporal cortex. Elife 2019, 8: e47142.
50. He CX, Peelen MV, Han ZZ, Lin N, Caramazza A, Bi YC. Selectivity for large nonmanipulable objects in scene-selective visual cortex does not require visual experience. Neuroimage 2013, 79: 1–9.
51. Wang XY, He CX, Peelen MV, Zhong SY, Gong GL, Caramazza A, *et al*. Domain selectivity in the parahippocampal gyrus is predicted by the same structural connectivity patterns in blind and sighted individuals. J Neurosci 2017, 37: 4705–4716.
52. Troiani V, Stigliani A, Smith ME, Epstein RA. Multiple object properties drive scene-selective regions. Cereb Cortex 2014, 24: 883–897.
53. Bao PL, She L, McGill M, Tsao DY. A map of object space in primate inferotemporal cortex. Nature 2020, 583: 103–108.
54. Lescroart MD, Stansbury DE, Gallant JL. Fourier power, subjective distance, and object categories all provide plausible models of BOLD responses in scene-selective visual areas. Front Comput Neurosci 2015, 9: 135.
55. Caldara R, Seghier ML, Rossion B, Lazeyras F, Michel C, Hauert CA. The fusiform face area is tuned for curvilinear patterns with more high-contrasted elements in the upper part. Neuroimage 2006, 31: 313–319.
56. Bracci S, Ritchie JB, Kalfas I, Op de Beeck HP. The ventral visual pathway represents animal appearance over animacy, unlike human behavior and deep neural networks. J Neurosci 2019, 39: 6513–6525.
57. Sha L, Haxby JV, Abdi H, Guntupalli JS, Oosterhof NN, Halchenko YO, *et al*. The animacy continuum in the human ventral vision pathway. J Cogn Neurosci 2015, 27: 665–678.
58. Long B, Konkle T. The role of textural statistics *vs* outer contours in deep CNN and neural responses to objects. 2018 Conference on Cognitive Computational Neuroscience, Philadelphia, Pennsylvania, USA: Cognitive Computational Neuroscience, 2018. https://doi.org/10.32470/CCN.2018.1118-0.
59. Harris A, Aguirre GK. Neural tuning for face wholes and parts in human fusiform gyrus revealed by FMRI adaptation. J Neurophysiol 2010, 104: 336–345.
60. Zeiler MD, Fergus R. Visualizing and Understanding Convolutional Networks. Computer Vision – ECCV 2014, Cham: Springer International Publishing, 2014: 818–833.