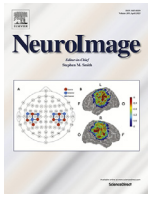
NeuroImage 269 (2023) 119935

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage



A stimulus-driven approach reveals vertical luminance gradient as a stimulus feature that drives human cortical scene selectivity

Annie Cheng[a,b], Zirui Chen[a,c], Daniel D. Dilks[a,*]

[a] Department of Psychology, Emory University, Atlanta, GA, USA
[b] Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA
[c] Department of Cognitive Science, Johns Hopkins University, Baltimore, MD, USA

ARTICLE INFO

Keywords:
Parahippocampal place area
Occipital place area
Scene perception
Scene selectivity
High-level vision
Lateral occipital complex
fMRI

ABSTRACT

Human neuroimaging studies have revealed a dedicated cortical system for visual scene processing. But what is a "scene"? Here, we use a stimulus-driven approach to identify a stimulus feature that selectively drives cortical scene processing. Specifically, using fMRI data from BOLD5000, we examined the images that elicited the greatest response in the cortical scene processing system, and found that there is a common "vertical luminance gradient" (VLG), with the top half of a scene image brighter than the bottom half; moreover, across the entire set of images, VLG systematically increases with the neural response in the scene-selective regions (Study 1). Thus, we hypothesized that VLG is a stimulus feature that selectively engages cortical scene processing, and directly tested the role of VLG in driving cortical scene selectivity using tightly controlled VLG stimuli (Study 2). Consistent with our hypothesis, we found that the scene-selective cortical regions—but not an object-selective region or early visual cortex—responded significantly more to images of VLG over control stimuli with minimal VLG. Interestingly, such selectivity was also found for images with an "inverted" VLG, resembling the luminance gradient in night scenes. Finally, we also tested the behavioral relevance of VLG for visual scene recognition (Study 3); we found that participants even categorized tightly controlled stimuli of both upright and inverted VLG to be a place more than an object, indicating that VLG is also used for behavioral scene recognition. Taken together, these results reveal that VLG is a stimulus feature that selectively engages cortical scene processing, and provide evidence for a recent proposal that visual scenes can be characterized by a set of common and unique visual features.

1. Introduction

Human neuroimaging studies have revealed a set of three cortical regions selectively involved in visual scene processing: the parahippocampal place area (PPA; Epstein and Kanwisher, 1998), the occipital place area (OPA; Dilks et al., 2013), and the retrosplenial complex (RSC; Maguire, 2001). However, despite a growing understanding of the neural mechanisms underlying visual scene processing (for review, see Dilks et al., 2021; Groen et al., 2017; Malcolm et al., 2016), a fundamental question remains: While these three cortical regions are known to be scene selective (i.e., responding two to four times more to images of scenes than to image of objects or faces), what precisely is a "scene" (versus an object or face), and thereby selectively engages cortical scene processing in the first place?

Over the past decade, several studies have attempted to identify the stimulus features that selectively engage cortical scene processing. Specifically, a collection of studies found that the scene-selective cortical regions show a preferential response to certain low-level features that are commonly found in visual scene stimuli – that is, high spatial frequency, rectilinearity, and cardinal orientations (Rajimehr et al., 2011; Kaufmann et al. 2014; Nasr et al., 2012, 2014). However, one recent study (Cheng et al., 2021) pointed out that, since these features are also commonly found in non-scene stimuli, especially objects, they cannot reliably enable the human brain to differentiate scene from non-scene stimuli, and proposed that cortical scene selectivity is rather driven by visual features that are not only common, but also unique to visual scenes. Indeed, they found that "concavity" (portraying inside) is one such stimulus feature, with the cortical scene processing system even responding significantly more to "concave" objects (e.g., the inside of a microwave) over "convex" objects (e.g., the outside of a microwave). However, the cortical scene processing system has also been shown to respond selectively to images of the exterior of buildings that are convex, and landscapes with no apparent cues of concavity (Epstein and Kanwisher, 1998). Thus, beyond concavity, there must exist other visual cues that also drive cortical scene processing, but what are they?

* Corresponding author.
E-mail address: dilks@emory.edu (D.D. Dilks).

https://doi.org/10.1016/j.neuroimage.2023.119935.
Received 7 October 2022; Received in revised form 19 January 2023; Accepted 7 February 2023
Available online 9 February 2023.

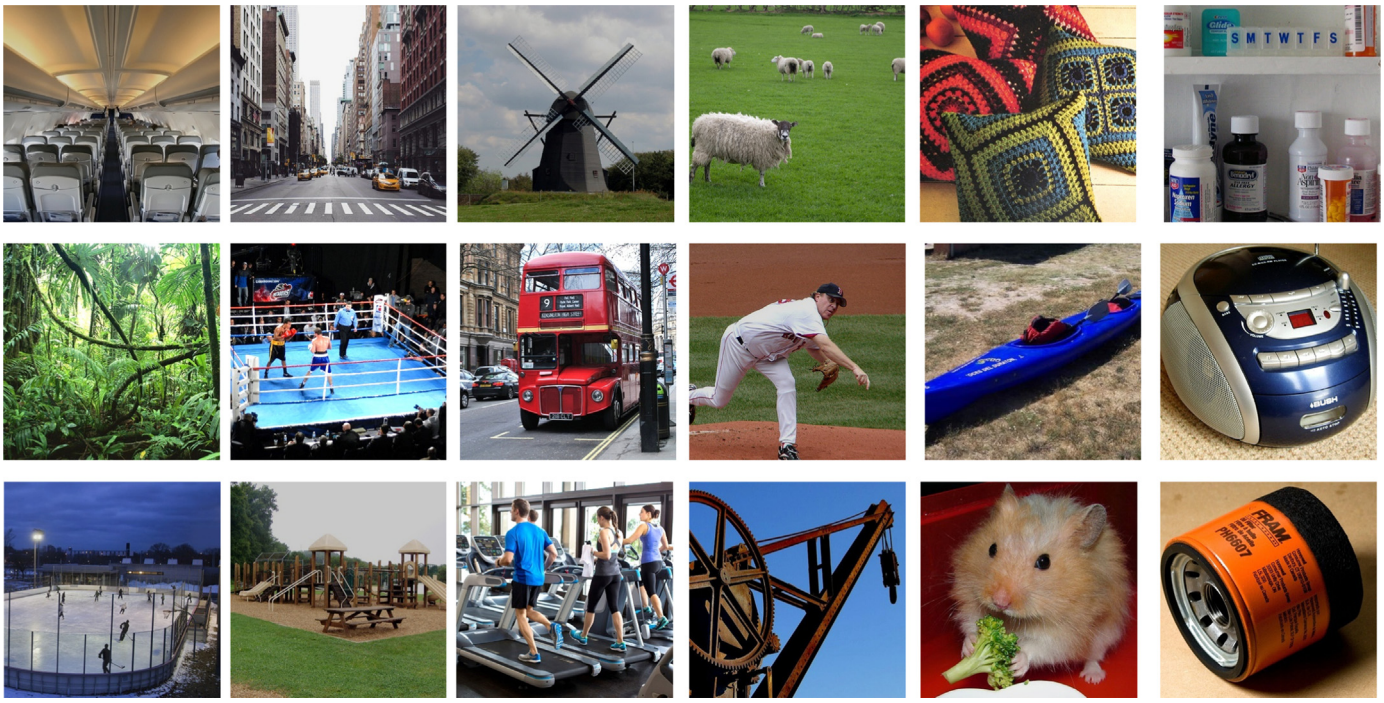


Fig. 1. Example stimuli from BOLD5000.

One challenge in answering this question is that visual scene stimuli are highly variable; thus, there is a vast number of stimulus features that could be potentially relevant, making it nearly impossible to generate an a priori hypothesis regarding what the precise stimulus features are that drive cortical scene processing. To resolve this challenge, in Study 1, we used an alternative, stimulus-driven approach to identify a candidate feature. Specifically, we reasoned that if there exists a stimulus feature that selectively drives cortical scene processing, then it should be identifiable in the visual stimuli that elicited a selective response in the scene-selective regions. As such, we made use of BOLD5000 (Chang et al., 2019)—an existing fMRI database of participants looking at approximately five thousand highly variable images—and examined whether there is a recurring feature in the visual stimuli that elicited a selective response in the scene-selective cortical regions. To anticipate, we found that among the images that elicited the strongest response in the scene-selective cortical regions, there is a common "vertical luminance gradient" (VLG) – with the upper half of a scene image significantly brighter than the lower half. Moreover, VLG systematically and selectively increases with the neural response in the scene-selective regions, but not in an object-selective region nor in early visual cortex. Thus, in Study 2, taking a hypothesis-driven approach, we hypothesized that VLG is a stimulus feature that drives cortical scene selectivity, and directly tested whether the scene-selective regions indeed show a selective response to tightly-controlled images of VLG. Finally, in Study 3, we further explored the behavioral relevance of VLG for visual scene recognition and hypothesized that VLG is a visual feature that humans use for behavioral scene recognition.

2. Study 1

2.1. Materials and methods

Visual stimuli. The entire set of visual stimuli used in the original BOLD5000 experiment (Chang et al., 2019)—including categories and stimuli drawn from three computer vision datasets (Deng et al., 2009; Lin et al., 2014; Xiao et al., 2010)—were included. Visual stimuli included 4916 unique images, with a diverse range of images, including 1000 images from 250 scene categories that varied in terms of indoor versus outdoor, and manmade versus natural scenes, non-exclusively; 1916 images of singular objects; and 2000 images of multiple objects that varied in manmade, natural, animate and inanimate objects, non-exclusively (see Fig. 1 for example stimuli).

Analysis of image statistics. To analyze the luminance statistics of the stimuli, we converted the original JPG images, which were encoded in standard Red Green Blue (sRGB) color space, into CIELAB color space in which the intensity of each pixel is expressed in three values: L^* (luminance), a^* (green-red opponency), and b^* (blue-yellow opponency). We chose to analyze the image statistics in the CIELAB color space since the feature dimensions in CIELAB more closely resemble the human perceptual experience than sRGB (Oliva and Schyns, 2000), and that the CIELAB format codes for the luminance and color, or chroma, properties of the stimuli separately, which is particularly relevant for this study. After the above transformation, we split each image into upper and lower halves along the middle of the image, averaged the luminance value (L^*) of the pixels within each of the upper and lower halves, and calculated the difference (upper–lower) to quantify VLG.

fMRI data analysis. We utilized BOLD5000 (Chang et al., 2019), an open fMRI dataset, which consisted of the BOLD response of four participants looking at 4916 unique images in a slow event-related design, for analysis. Specifically, the BOLD response during peak activity (both TR 3 and 4 for participant CSI1–3 and TR 3 only for participant CSI4), which were made directly available on the website, were used for analysis. The BOLD response was measured as the residuals from a general linear model in which nine nuisance variables (i.e., six motion parameters, the average signal inside the cerebral spinal fluid mask and inside the white matter mask, separately, and global signal within the whole-brain mask; all extracted from the fMRIprep pipeline; Esteban et al., 2019) were regressed out of the fMRI time series, and were demeaned across all image presentations. For images that were shown more than once to the participants, we only included the neural response to the first presentation of that image for analysis. Regions of Interests (ROIs) included the three known scene-selective cortical regions: PPA (Epstein and Kanwisher, 1998), OPA (Dilks et al., 2013) and RSC (Maguire, 2001). We also examined the neural response in an object-selective region (lateral occipital complex, LOC) and in early visual cortex (EarlyVis) as control regions. These ROIs were localized by Chang et al. (2019) using data

from an independent localizer. The scene-selective regions were defined using the contrast of scenes minus objects and scrambled images, and LOC was defined by the contrast of objects minus scrambled images. EarlyVis was defined as the cluster most confined to the calcarine sulcus using the contrast of scrambled images minus fixation baseline. All ROIs were defined with a threshold of $p<0.0001$ (or smaller, family-wise error corrected).

To identify a recurring feature in the stimuli that selectively engaged the scene-selective regions, we rank ordered the stimuli by the corresponding mean voxel-wise BOLD response in each ROI, sorted the stimuli into 5 separate bins by the ranks (with 1000 images per bin), and averaged across the pixel intensity values of the stimuli within each bin to wash out the unique features of each image. We then examined 1) whether any visual features remain in the mean image of the stimuli that elicited the greatest response in the scene-selective cortical regions, and 2) whether any visual features systematically change across these mean images. Finally, to test for the relationship between VLG and the neural response of an ROI at the level of individual images, we also correlated VLG with the BOLD response across the entire image set.

2.2. Results

2.2.1. Validation of cortical scene selectivity in the BOLD5000 dataset

We first validated whether the scene-selective regions demonstrated selectivity for scene over object stimuli, and whether an object-selective region—LOC—demonstrated a selectivity for object over scene stimuli. To do so, we directly compared the neural response to scene versus singular object stimuli in these ROIs. A 4 (ROI: PPA, OPA, RSC, LOC) × 2 (Category: Scene, Object) mixed-effect repeated-measures ANOVA revealed a significant ROI × Category interaction ($F_{(3,8742)}=1497.36$, $p<0.001$, $\eta_p{}^2=0.34$), with post-hoc comparisons revealing a significantly greater response to Scene over Object stimuli in all three scene-selective cortical regions (all ps <0.001), and a significantly greater response to Object over Scene stimuli in LOC ($p<0.001$). Together, these results confirmed that the scene-selective regions indeed showed a selective response to scene over object stimuli, and LOC showed a selective response to object over scene stimuli, consistent with the known selectivity of these ROIs (Epstein and Kanwisher, 1998; Kamps et al., 2016; Cheng et al., 2021).

2.2.2. VLG is common across the visual stimuli that selectively drive cortical scene processing

Having validated that the BOLD5000 dataset captured the known cortical scene selectivity, we next examined whether there is a stimulus feature common across visual stimuli that elicited the greatest response in the scene-selective regions. To do so, we first sorted the stimuli into 5 separate bins by the ranks of the corresponding BOLD response within each ROI (see Fig. 2Ai-2Ei for an illustration of every thousandth stimuli, ranked from the lowest to highest BOLD response), and averaged across the thousand images within each bin to wash away the unique features

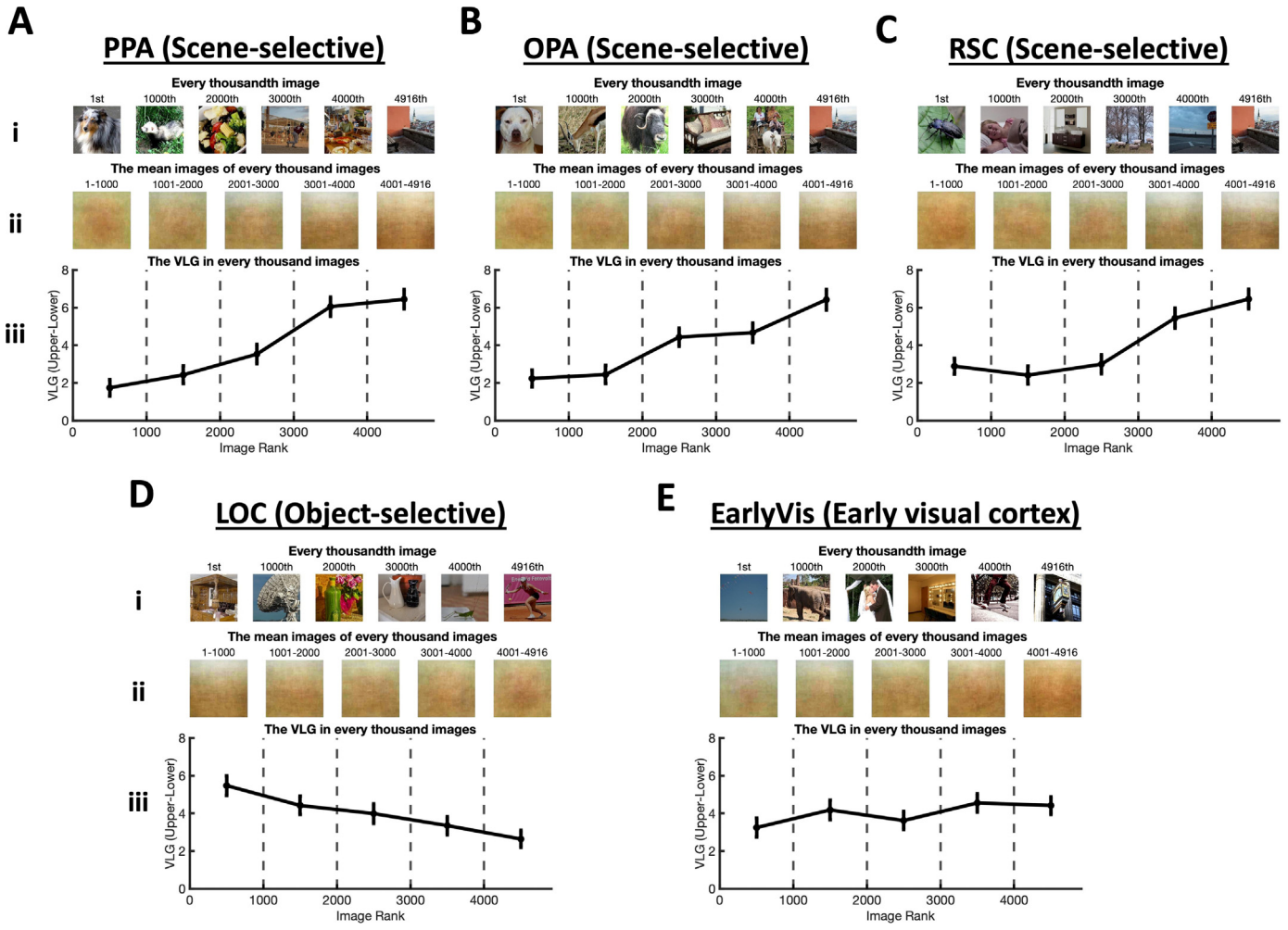


Fig. 2. **i**, Every thousandth stimuli as ranked by the BOLD response (lowest to highest, from left to right) of **(A)** PPA, (**B**) OPA, and (**C**) RSC, (**D**) LOC, and (**E**) EarlyVis. **ii,** The mean images of every thousand images, as binned by the ranks of the corresponding BOLD response of an image in an ROI (also lowest to highest, from left to right). **iii,** The VLG in every thousand images within each ROI. VLG systematically increases as the BOLD response in the scene-selective regions increases, but not in LOC nor EarlyVis. Error bars represent ± 1 standard error of the mean.

of each image. Next, in an exploratory analysis, we examined whether any visual features remain in the mean images of the stimuli that elicited the greatest response in the scene-selective cortical regions.

Intriguingly, as seen in the rightmost image in Fig. 2Aii-2Cii, the mean images of the stimuli that elicited the greatest response across all three scene-selective regions are quite similar. Specifically, they share a common "vertical luminance gradient" (VLG), with the top half of the image brighter than the bottom half. Importantly, VLG is not observable in the mean images of the stimuli that elicited the lowest response (i.e., the leftmost) across all three scene-selective regions, and VLG becomes more salient as the neural response of the scene-selective regions increases across the mean images. To directly test this observation, we next quantified the VLG of the stimuli in each of the five bins by splitting each stimulus image from BOLD5000 into upper and lower halves along the middle of the image, calculating the difference of luminance intensity (using the L^* value from CIELAB color space) between the pixels in the two halves, and then examining the mean differences across these bins. Consistent with our observation, we found an increase of VLG as the neural response increases across the bins (Fig. 2Aiii-2Ciii). Finally, we directly tested for a correlation between the amount of VLG of an image and the neural response of the scene-selective cortical regions at the level of individual images. Consistently, we found a significant correlation between the amount of VLG in an image with the neural response in the scene-selective cortical regions (PPA: $r = 0.12$, $p<0.001$; OPA: $r = 0.10$, $p<0.001$; RSC: $r = 0.09$, $p<0.001$), indicating VLG increases with cortical scene selectivity.

But does VLG increase with the neural response of the scene-selective cortical regions only and not in other brain regions? To test this possibility, we examined whether VLG systematically increases with the neural response of LOC and EarlyVis. Unlike in the scene-selective regions, VLG *decreases* as the response in LOC increases, whereas EarlyVis shows no visible linear trend of VLG. Furthermore, when we directly tested for a correlation between VLG and the neural response of these regions, we found a significant but *negative* correlation ($r=-0.07$, $p<0.001$) in LOC, and no significant correlation between VLG and the EarlyVis response ($r = 0.03$, $p = 0.07$). Thus, VLG indeed selectively increases with the neural response in scene-selective cortical regions.

2.2.3. But what about night scenes in which the luminance gradient is reversed?

Across most environments in our everyday lives, light tends to come from above, whether it is from the sun in outdoor scenes, or overhead artificial lighting in indoor environments. As such, across most scenes, the upper half of a scene—which often contains these "above" sources of illumination—is brighter than the lower half of a scene. Importantly, since non-scene objects and faces have a smaller surface area, they tend to capture far less—if any—luminance changes; thus, VLG is a stimulus feature commonly and uniquely found in visual scenes, and its presence in visual stimuli can be a reliable indicator of a scene.

This idea, however, also leads to a curious question: How then do humans recognize outdoor, night scenes in which the lower half of a scene (i.e., the ground surface) is often brighter than the upper half (i.e., the pitch-black sky), thus resulting in a reverse, dark-to-light luminance gradient? One possibility is that cortical scene selectivity may be driven by not only a light-to-dark luminance gradient, but also a luminance gradient in the reverse direction (i.e., dark-to-light). To explore this possibility, we identified the limited, but nevertheless existing night scene stimuli (16 in total; see Fig. 3A for examples) in BOLD5000 and examined the corresponding neural response in the scene-selective regions (Fig. 3B). We found that the scene-selective regions showed a comparable response between the night scenes and "day" scenes, and both more than the response to object stimuli. Furthermore, when we tested for a correlation between the absolute value of VLG of the BOLD5000 stimuli and the corresponding neural response in the scene-selective regions to account for both light-to-dark and dark-to-light VLG, we found comparable correlations (PPA: $r = 0.10$, $p<0.001$; OPA: $r = 0.09$, $p<0.001$; RSC: $r = 0.13$, $p<0.001$). As such, it is likely that cortical scene selectivity is driven by VLG in both directions. Note, however, since there were relatively few night scene stimuli presented in BOLD5000, the effect of a dark-to-light gradient was washed away when we averaged the stimuli by the rank of the neural responses in our analysis. Thus, in Study 2, in addition to directly testing the role of VLG in driving cortical scene selectivity (discussed next), we also probed the effect of light-to-dark versus dark-to-light VLG in driving cortical scene selectivity.

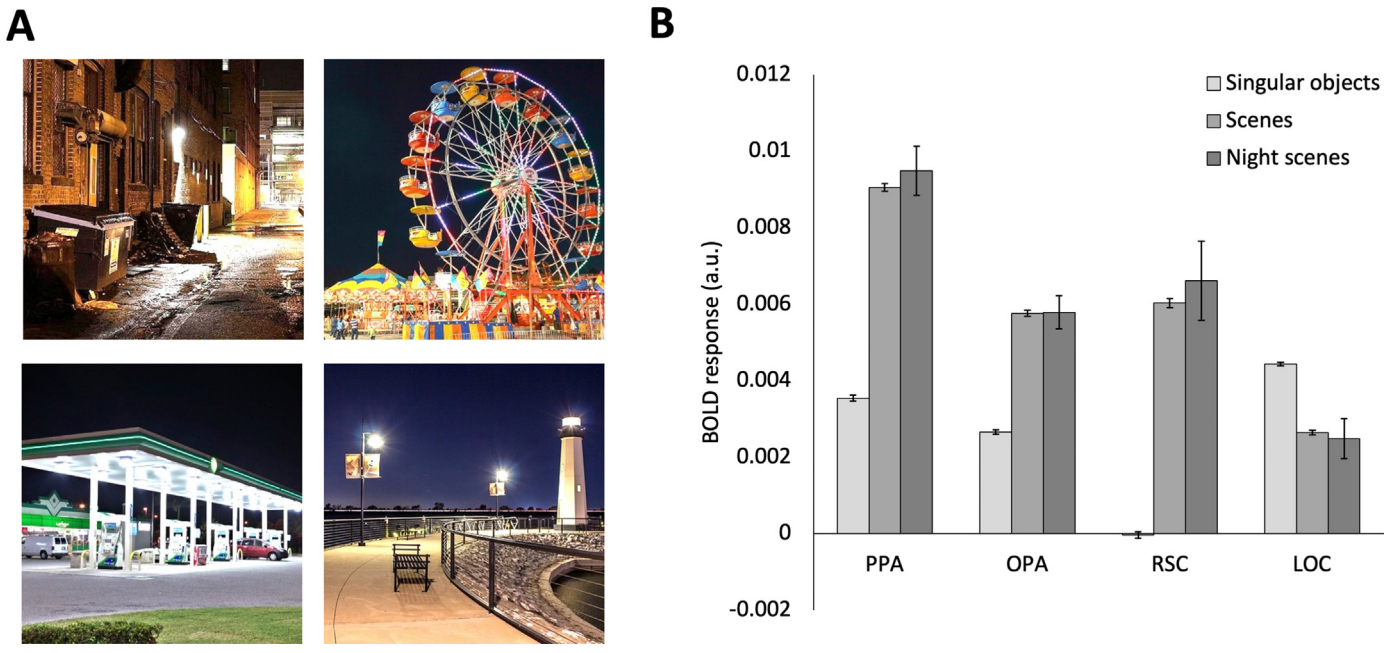


Fig. 3. **A**, Example stimuli of night scenes from BOLD5000. **B,** The BOLD response of PPA, OPA, RSC and LOC to stimuli of singular objects, scenes and night scenes in BOLD5000, respectively. Error bars represent ± 1 standard error of the mean.

3. Study 2

In Study 1, we found that VLG may be a stimulus feature that drives cortical scene selectivity. One alternative hypothesis, however, is that—given the visual stimuli tested in BOLD5000 are naturalistic, complex stimuli that are highly variable in not only VLG, but also many other visual features—the response in scene-selective cortical regions might not be driven by VLG per se, but by other confounding visual features that covary with VLG. Thus, to directly test for the effect of VLG in driving cortical scene selectivity, we created tightly controlled stimuli of VLG that are impoverished of other visual features, together with a set of Control stimuli with minimal VLG (which mimic the mean images of the stimuli that elicited the lowest response in the scene-selective cortical regions in Study 1, for comparison). We predicted that 1) if VLG (i.e., both light-to-dark and dark-to-light, as found in Study 1) indeed selectively drives cortical scene selectivity, then the scene-selective cortical regions will show a greater response to images with a strong VLG over the Control stimuli with minimal VLG, and 2) if cortical scene selectivity is driven only by luminance gradient along the *vertical* dimension, then Upright and Inverted VLG—but not Rotated VLG (in which the luminance gradient varies along the horizontal dimension)—will selectively drive the neural response in the scene-selective regions.

3.1. Materials and methods

Participants. Twenty participants (Age:21–40; 12 females) were recruited from the Emory University community, and no participants were excluded. All participants gave informed consent and had normal or corrected-to-normal vision.

Visual stimuli. To directly test for the effect of VLG in driving cortical scene selectivity, we created twelve artificial images of VLG that contain a strong difference in the luminance value between the upper versus lower halves of an image, and are highly impoverished with respect to non-VLG stimulus features (Fig. 5A). To create these VLG stimuli, we made use of highly variable scene images that were made available by Konkle et al. (2010), and then averaged across these images to wash away the unique visual features of each scene image. Sixty-eight unique scene images were used to create each VLG image. The VLG images were converted into grayscale to further control for non-VLG features, and we enhanced the contrast of these images to amplify the VLG. A paired *t*-test revealed a significant difference in the luminance value between the upper versus lower halves of the stimuli ($t_{(11)}$=12.48, p<0.001). In addition to the VLG stimuli, we also created a set of Control stimuli with minimal VLG, which mimic the mean images of the stimuli that elicited the weakest response in the scene-selective cortical regions in Study 1, as a control comparison. To do so, we used the same averaging procedure to average across highly variable object images made available by Hebart et al. (2019). A minimum of thirty-five unique object images were used to create each Control image. A paired *t*-test revealed a significant difference in the luminance value between the upper versus lower halves of the Control stimuli ($t_{(11)}$=5.47, p<0.001). Crucially, however, a 2 (Condition: VLG, Control) × 2 (Half: Upper, Lower) mixed-effect repeated measures ANOVA revealed a significant Condition × Half interaction ($F_{(1,22)}$=30.36, p<0.001, η_p^2=0.58), indicating a greater difference in the luminance value between the upper versus lower halves of the VLG stimuli than the Control stimuli. Furthermore, we also tightly controlled for 1) the amount of high spatial frequency (HSF) information (Rejimehr et al., 2011; Berman et al., 2017; Bainbridge and Oliva 2015); a two-sample *t*-test confirmed no significant difference in the HSF information between the VLG versus the Control stimuli ($t_{(22)}$=−0.99, p = 0.33), and 2) the amount of rectilinearity (Nasr et al., 2014; Bryan et al., 2016); a two-sample *t*-test confirmed no significant difference between the amount of rectilinearity between the VLG and the Control stimuli ($t_{(22)}$=1.19, p = 0.25). After creating the Upright stimuli in both conditions, we then turned the images upside down for the Inverted condition, which tested the effect of a reverse VLG in driving cortical scene selectivity. In addition, we also created the rotated (90° clockwise) version of the same set of stimuli to test whether cortical scene selectivity for luminance gradient is specific to the vertical dimension, or general across all orientations.

Experimental design. We used a region of interest (ROI) approach in which we localized the cortical regions of interest with the Localizer runs, and then used an independent set of Experimental runs to investigate the responses of these regions when viewing blocks of images from the stimulus categories of interest. Our ROIs included PPA, OPA and RSC. We also examined the neural response of an object-selective region (lateral occipital cortex, LO) and primary visual cortex (V1) as control regions.

For the Localizer runs, we used a blocked design in which participants viewed images of faces, objects, scenes, and scrambled objects. Each Localizer run was 336 s long. There were four blocks per stimulus category within each run, and 20 images from the same category within each block. Each image was presented for 300 ms, followed by a 500 ms ISI for a total of 16 s per block. Image order within each block was randomized. The order of the blocks in each run was palindromic, and the order of the blocks in the first half of the palindromic sequence was pseudo-randomized across runs. Five 16 s fixation blocks were included: one at the beginning, three in the middle interleaved between each set of stimulus blocks, and one at the end of each run. Participants performed a one-back repetition detection task, responding every time the same image was presented twice in a row.

For the Experimental runs, we used a block design in which participants viewed blocks of images from each condition of interest (see the Visual Stimuli section). Seventeen participants completed nine Experimental runs; two participants completed six Experimental runs; one participant completed eight Experimental runs. Each run was 368 s long. There were three blocks per condition of interest within each run, and 12 images from the same condition within each block. Each image was presented for 300 ms, followed by a 700 ms ISI for a total of 12 s per block. Image order within each block, and the order of blocks in each run were randomized. Each block was preceded by an 8 s fixation block. Participants performed a one-back repetition detection task, responding every time the same image was presented twice in a row.

MRI scan parameters. Scanning was done on a 3T Siemens Trio scanner at the Facility for Education and Research in Neuroscience (FERN) at Emory University (Atlanta, GA). Functional images were acquired using a 32-channel head matrix coil and a gradient echo single-shot echo planar imaging sequence. Thirty-two slices were acquired for all runs: repetition time=2 s; echo time=30 ms; flip angle = 90°; voxel size = 3.0 × 3.0 × 3.0 mm; and slices were oriented approximately between perpendicular and parallel to the calcarine sulcus, covering the occipital as well as the posterior portion of temporal lobes. Whole-brain, high-resolution T1-weighted anatomical images were also acquired with 1 × 1 × 1 mm voxels.

Data analysis. fMRI data were processed in FSL software (Smith et al., 2004) and the FreeSurfer Functional Analysis Stream (FS-FAST). Data were analyzed in each participant's native space. Pre-processing included skull-stripping (Smith, 2002), linear-trend removal, and three-dimensional motion correction using FSL's MCFLIRT tool. Data were then fit using a double gamma function, and spatially smoothed with a 5-mm kernel.

After preprocessing, the ROIs were bilaterally defined in each participant using data from the Localizer runs. PPA, OPA and RSC were defined as those regions that responded more strongly to scenes than objects (p<10^{-4}, uncorrected; Fig. 4), whereas LO was defined as those regions that responded more strongly to objects than scrambled objects, following the conventional method of Epstein and Kanwisher (1998) and Grill-Spector et al. (1998). To define V1, we used a probabilistic atlas made available by Wang et al. (2015) and registered the ROIs from standard MNI space to each subject's native space using the FSL linear registration tool.

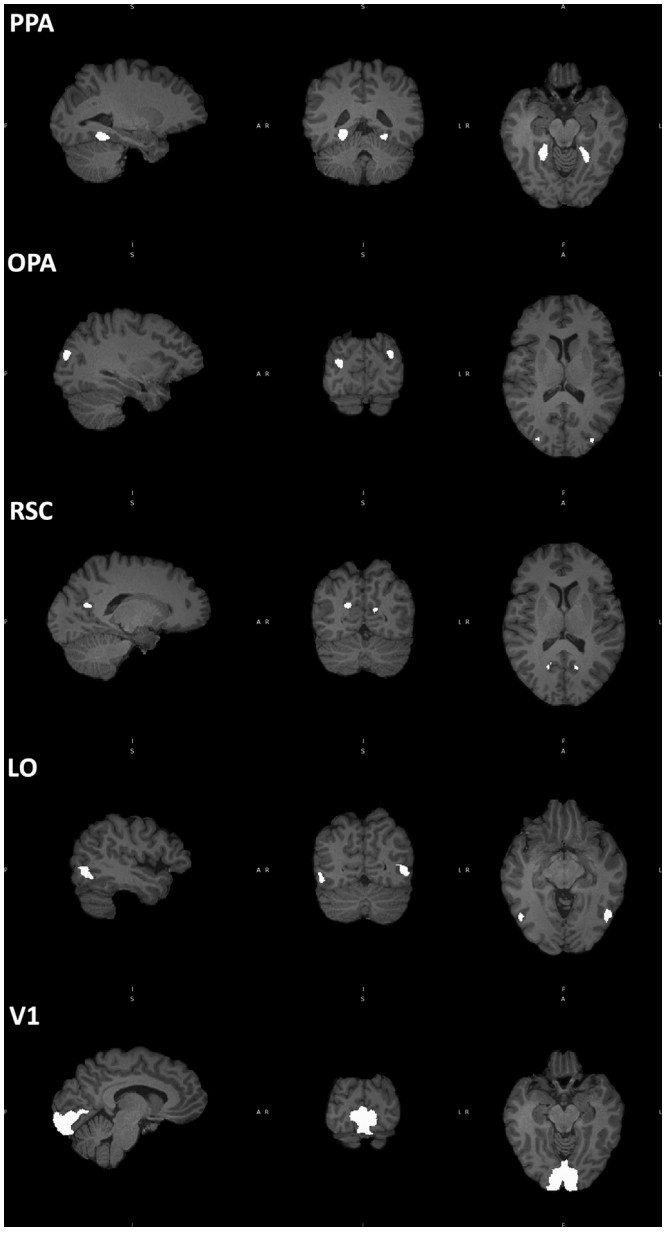


Fig. 4. Regions of interest (PPA, OPA, RSC, LO, V1) from an example participant.

PPA, OPA, RSC, LO and V1 were defined in at least one hemisphere of all subjects. For each ROI of each participant, the average response across voxels for each condition was extracted and converted to percent signal change (PSC) relative to fixation, and repeated-measures ANOVAs were performed. Finally, a 3 (ROI: PPA, OPA, RSC) × 2 (Hemisphere: Left, Right) × 2 (Condition: VLG, Control) × 2 (Orientation: Upright, Inverted) repeated-measures ANOVA revealed no significant ROI × Hemisphere × Condition × Orientation interaction ($F_{(2,36)}$=0.85, p = 0.44, η_p^2=0.05); thus, data from the left and right hemispheres of the same ROI were collapsed.

3.2. Results

We predicted that 1) if VLG (i.e., both light-to-dark and dark-to-light, as found in Study 1) indeed selectively drives cortical scene selectivity, then the scene-selective cortical regions will show a greater response to images with a strong VLG over the Control stimuli with minimal VLG, and 2) if cortical scene selectivity is driven only by luminance gradient along the *vertical* dimension, then Upright and Inverted VLG—but not Rotated VLG (in which the luminance gradient varies along the horizontal dimension)—will selectively drive the neural response in the scene-selective regions.

To test Prediction 1, we first examined whether the scene-selective cortical regions show a greater response to the VLG over the Control stimuli (Fig. 5B). Indeed, a 3 (ROI: PPA, OPA, RSC) × 2 (Condition: VLG, Control) × 2 (Orientation: Upright, Inverted) repeated-measures ANOVA revealed a significant main effect of Condition ($F_{(1,19)}$=65.82, p<0.001, η_p^2=0.78), with an overall greater response for the VLG over the Control conditions, consistent with our hypothesis. Moreover, post-hoc comparisons revealed a significantly greater response to the VLG over the Control conditions across all three scene-selective cortical regions (all ps <0.001), confirming a common selectivity for VLG. Next, we examined whether the scene-selective regions show a similar response to Upright and Inverted VLG. We found no significant ROI × Condition × Orientation interaction ($F_{(2,38)}$=2.74, p = 0.08, η_p^2=0.13), with post-hoc comparisons revealing no significant difference between Upright and Inverted VLG in all three regions (all ps >0.07). Together, these results suggest that all three scene-selective regions showed a significantly greater response to the Upright and Inverted VLG conditions, compared to the Control conditions, consistent with our prediction that both light-to-dark and dark-to-light VLG drives cortical scene processing.

One might notice that RSC response for the VLG stimuli was below baseline (relative to fixation), why might this be the case? Previous studies have suggested RSC is involved in more navigation- and memory-related processing of visual scenes (Aguirre et al., 1998; Maguire, 2001; Marchette et al., 2014; Park and Chun, 2009; Persichetti and Dilks, 2019; Baldassano et al., 2016; Silson et al., 2019), and thus responds only to more naturalistic and real-world relevant scene stimuli (Choo and Walther, 2016; Cheng et al., 2021). Since the stimuli in Study 2 are tightly controlled and highly impoverished, they might not be the optimal stimuli to drive RSC response, thus resulting in a relatively low level of response. Note, however, despite an overall low level of response, we nevertheless observed RSC showing a similar response pattern for VLG over Control stimuli, just like PPA and OPA, presenting evidence that is consistent with our hypothesis.

But is the selective response to VLG specific to the scene-selective cortical regions, or a more general preference across high-level visual cortex, perhaps as a result of the VLG conditions being somehow more interesting and thus engaging more attention? If so, then we would expect other regions in high-level visual cortex (e.g., LO) to also demonstrate the same response pattern. A 4 (ROI: PPA, OPA, RSC, LO) × 2 (Condition: VLG, Control) × 2 (Orientation: Upright, Inverted) repeated-measures ANOVA revealed a significant ROI × Condition interaction ($F_{(3,57)}$=60.03, p<0.001, η_p^2=0.76), with LO showing a selectively greater response to the Control over the VLG conditions (post-hoc comparison, p<0.001), unlike the scene-selective regions. Thus, the selective response to VLG is not general across high-level visual cortex.

Could the selective response to VLG simply be driven by low-level visual information directly inherited from early visual cortex? If so, then we would expect V1 to demonstrate the same response pattern as those in the scene-selective regions. A 4 (ROI: PPA, OPA, RSC, V1) × 2 (Condition: VLG, Control) × 2 (Orientation: Upright, Inverted) repeated-measures ANOVA revealed a significant ROI × Condition interaction ($F_{(3,57)}$=37.27, p<0.001, η_p^2=0.66), with V1 showing a selectively greater response to Control over VLG conditions (post-hoc comparison, p = 0.002), unlike the scene-selective regions. Thus, the selective response to VLG is unlikely to be simply driven by the low-level visual information directly inherited from early visual cortex.

Finally, to test Prediction 2 – that is, whether cortical scene selectivity is specific to a luminance gradient along the vertical dimension only – we examined whether the scene-selective regions show a selective response to Upright and Inverted, but not Rotated, VLG (Fig. 6).

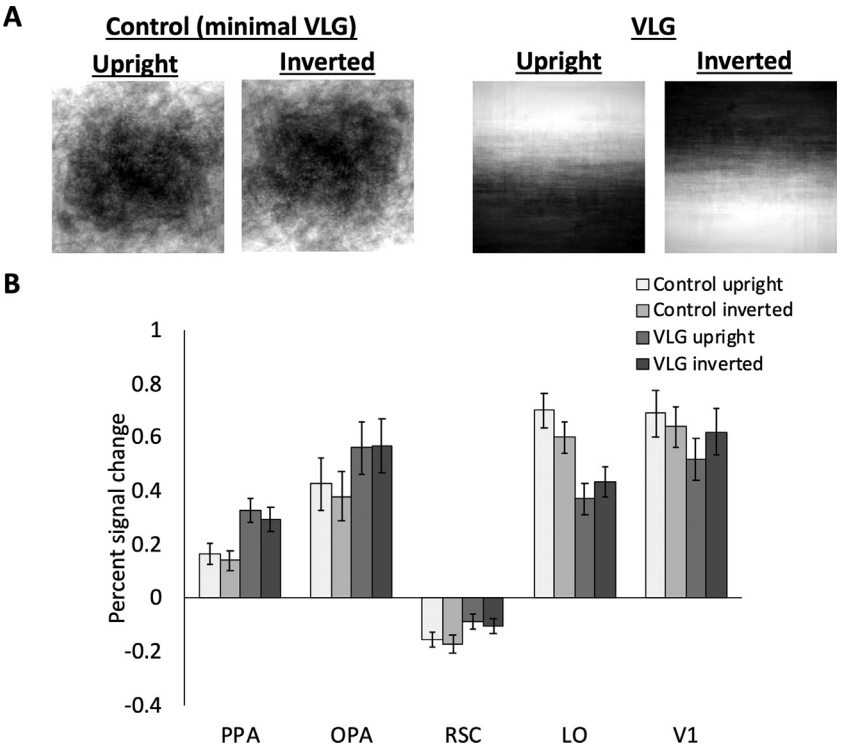


Fig. 5. A, Example stimuli in Study 2. **B.** Neural response of the ROIs. The scene-selective regions show a significantly greater response to the VLG over the Control stimuli, whereas LO and V1 do not. Error bars represent ± 1 standard error of the mean.

Surprisingly, a 3 (ROI: PPA, OPA, RSC) × 3 (Orientation: Upright, Inverted, Rotated VLG) repeated-measures ANOVA revealed a significant ROI × Orientation interaction ($F_{(4,76)}$=7.02, p<0.001, η_p^2=0.27), with post-hoc comparisons revealing a similar response across all Orientations in PPA (all *ps* >0.07), and a significantly higher response for Rotated over Upright and Inverted VLG in OPA and RSC (OPA: both *ps* <0.01; RSC: both *ps* <0.03) – presenting evidence that seemingly contradicts our hypothesis. But does the Rotated VLG stimuli *specifically* drive cortical scene selectivity, or might it drive the neural response in high-level visual cortex more generally (e.g., LO)? A 4 (ROI: PPA, OPA, RSC, LO) × 3 (Orientation: Upright, Inverted, Rotated) repeated-measures ANOVA revealed a significant ROI × Orientation interaction ($F_{(6,114)}$=19.97, p<0.001, η_p^2=0.51), with post-hoc comparisons revealing a significantly greater response for Rotated over Upright and Inverted VLG in LO (both *ps* <0.001), relative to PPA, OPA and RSC. Together, these results reveal that the Rotated VLG stimuli do *not* specifically drive cortical scene selectivity, and—in fact—may even specifically drive cortical object selectivity.

Nevertheless, given the high response to the Rotated VLG stimuli in the scene-selective regions, might it still be a diagnostic feature of scene stimuli? To further probe this possibility, we returned to our BOLD5000 data. Specifically, we measured the horizontal luminance gradient in the BOLD5000 stimuli by the absolute value of the luminance difference between the left and right halves of an image, and tested whether horizontal luminance gradient correlated with the neural response of the scene-selective cortical regions. We found no significant, positive correlation between horizontal luminance gradient and human cortical scene selectivity (PPA: r=−0.04, p = 0.01; OPA: r=−0.02, p = 0.25; RSC: r=−0.03, p = 0.02), indicating that horizontal luminance gradient does not drive cortical scene selectivity. By contrast, we did find a positive, significant correlation between horizontal luminance gradient and the neural response in LO (r = 0.06, p<0.001), consistent with the just discussed finding, further raising the intriguing possibility that horizontal luminance gradient may enable the visual system to differentiate object from scene stimuli.

Why might we observe a high—despite not selective—response to the Rotated VLG in the scene-selective regions? One plausible explanation is that the Rotated VLG stimuli may have resembled "walls", which has been shown to drive the neural response in the scene-selective regions to some extent (Epstein and Kanwisher, 1998; Kamps et al., 2016). This possibility is consistent with a previous study in which scene-selective cortical regions showed a stronger response to visual stimuli of fragmented wall surfaces relative to everyday objects, and that the stimuli of fragmented wall surfaces elicited a stronger response in LOC relative to visual stimuli of empty indoor rooms (Kamps et al., 2016). Note, however, since the Rotated VLG stimuli did not *selectively* drive neural response in scene-selective regions, nor did we find horizontal luminance gradient driving neural response in the scene-selective regions among the complex, naturalistic stimuli in BOLD5000, horizontal luminance gradient is unlikely a feature that humans use to recognize scene from non-scene stimuli. Nevertheless, in Study 3, in addition to testing the behavioral relevance of VLG for visual scene recognition, we also directly tested horizontal luminance gradient to further ensure that humans indeed do *not* use horizontal luminance gradient for visual scene recognition.

4. Study 3

In Studies 1 and 2, we found that VLG selectively drives cortical scene processing, in both naturalistic, complex stimuli and artificial, tightly-controlled stimuli. In Study 3, we then asked whether VLG is actually used by humans for behavioral scene recognition. We hypothesized that if VLG is a visual feature that is behaviorally relevant for humans to recognize scene from non-scene stimuli, then participants will categorize even highly impoverished stimuli of VLG as a "place".

4.1. Materials and methods

Participants. A total of 100 participants were recruited from Amazon Mechanical Turk to participate. Ten participants (seven from the VLG condition; three from the Control condition) were excluded from analysis due to either providing incomplete responses or failing the attention check questions (discussed below). All participants gave informed consent and had normal or corrected-to-normal vision.

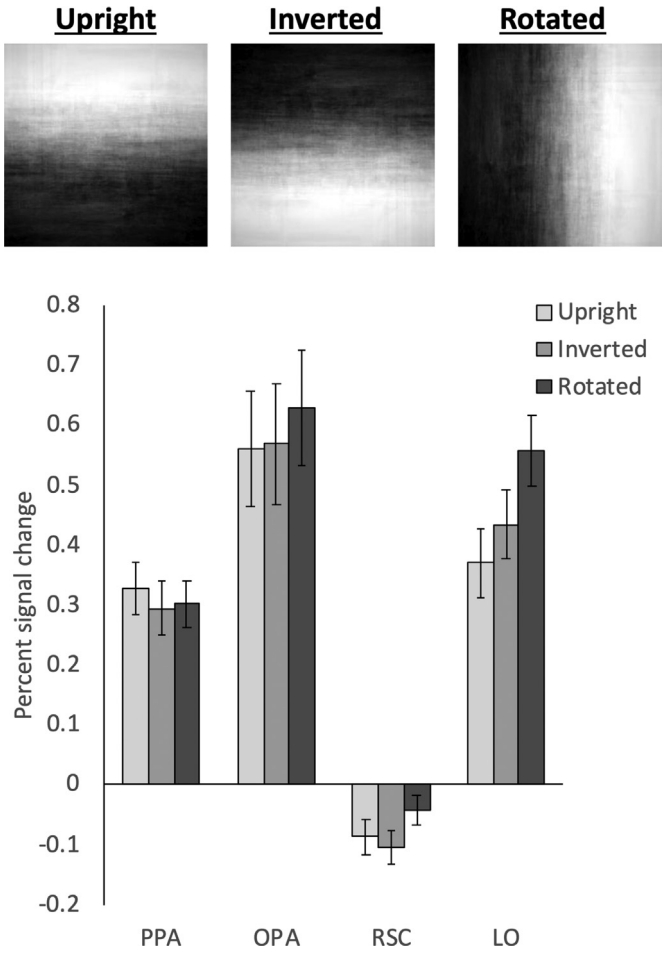


Fig. 6. Neural response of the ROIs to the VLG stimuli across different orientations. LO shows a greater response to Rotated over Upright and Inverted VLG, indicating the response to the Rotated VLG Condition is not specific to the scene-selective regions. Error bars represent ± 1 standard error of the mean.

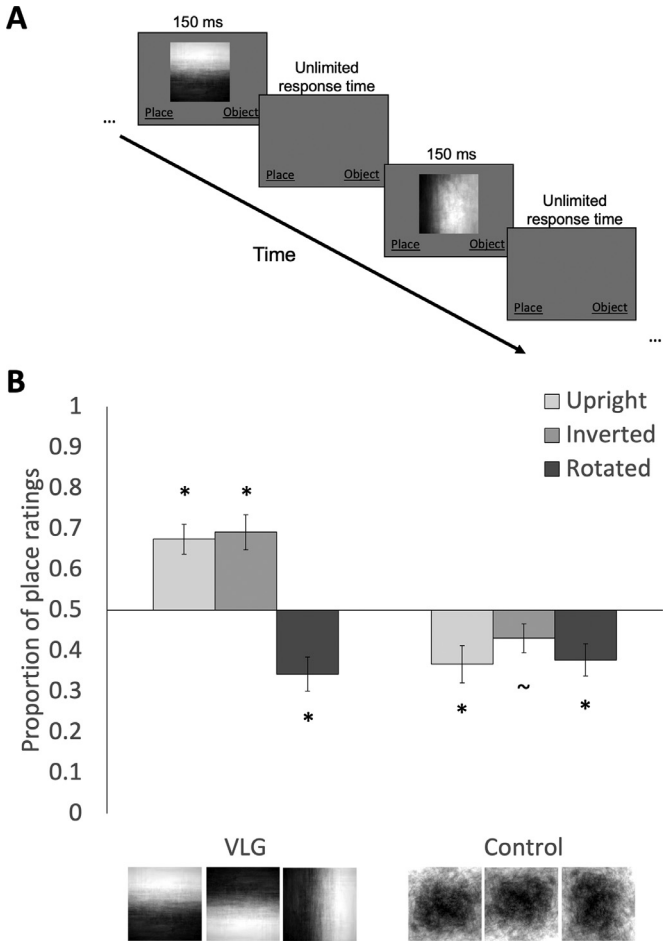


Fig. 7. A, Experimental procedure for Study 3. After an image was presented for 150 ms, participants were asked to indicate whether the image they just saw was a "place" or an "object". **B.** Participants' proportion of place ratings for the stimuli. Participants rated Upright and Inverted VLG as a "place" more often than as an "object", but not Rotated VLG. As a control comparison, participants showed qualitatively different response patterns for the Control conditions. Error bars represent ± 1 standard error of the mean. An asterisk (*) indicates that the proportion of place ratings is significantly different from chance, and a tilde (~) indicates that the difference is marginally significant at $p = 0.06$.

Visual stimuli. The same set of VLG and Control stimuli in all three Orientations (Upright, Inverted, Rotated) from Study 2 were used.

Experimental design. To test whether VLG is behaviorally relevant for visual scene recognition, we asked an independent group of participants to determine whether the stimuli used in Study 2 is a "place" or an "object" without defining either of these words. We used a between-subject design and assigned a participant to only either the VLG or Control condition to avoid participants from noticing the categorical difference between the VLG and Control stimuli and basing their judgment on that. Each participant first completed 6 practice trials to familiarize themselves with the task, and then another 12 experimental trials, which consisted of 12 unique images from either the VLG or Control condition, with 4 images per Orientation (Upright, Inverted, Rotated). Within each trial, an image was briefly presented for 150 ms, and participants were asked to indicate whether the image they just saw was a "place" or an "object" (Fig. 7A). Image order and the assignment of orientation for the images were randomized. After participants completed the experimental trials, they also completed 4 additional trials in which they performed the same task on 4 naturalistic images of real-world scenes and objects (2 per condition) to check whether they were paying attention during the experiment.

Data Analysis. We calculated the proportion of place ratings for each condition, and then used one sample t-tests to test the proportion against chance (i.e., 50%) for each condition. We also compared the place ratings across the conditions using repeated-measures ANOVAs.

4.2. Results

We hypothesized that if VLG is a visual feature that is behaviorally relevant for humans to recognize scene from non-scene stimuli, then participants will categorize even highly impoverished stimuli of VLG as a "place". To test our hypothesis, we examined whether participants' place ratings for both the Upright and Inverted VLG stimuli are significantly above chance (i.e., 50%; Fig. 7B). Consistent with our hypothesis, one-sample t-tests revealed that participants showed significantly above-chance place ratings for both the Upright and Inverted VLG stimuli (Upright: $t_{(42)}=4.74$, $p<0.001$; Inverted: $t_{(42)}=4.45$, $p<0.001$). By contrast, participants' place ratings for the Rotated VLG stimuli were significantly *below* chance ($t_{(42)}=-3.77$, $p<0.001$), indicating that participants categorized the Rotated VLG stimuli as an "object" more often than as a "place", providing converging behavioral evidence that the Upright and Inverted VLG, but not Rotated VLG, is a stimulus feature that humans use to distinguish scene from non-scene stimuli. Finally, to directly test whether participants showed a qualitatively different response to Rotated versus Upright and Inverted VLG, we conducted a three-level (Orientation: Upright, Inverted, Rotated VLG) repeated-measures ANOVA. We found a significant main effect of Orientation

($F_{(2,84)}$=30.94, p<0.001, η_p^2=0.42), with post-hoc comparisons revealing significantly greater place ratings for Upright and Inverted over Rotated VLG (both ps<0.001), and no significant difference between Upright and Inverted VLG (p = 0.70). Together, these results are consistent with the results from Studies 1 and 2 demonstrating the specific role of VLG for visual scene recognition, and the potential role of horizontal luminance gradient for object recognition.

Finally, as a control comparison, we also ran the same experiment on the Control stimuli in a separate group of participants. We found participants showed below-or-at-chance place ratings for the Control conditions (Upright: $t_{(46)}$=−2.92, p = 0.005; Inverted: $t_{(46)}$=−1.95, p = 0.06; Rotated: $t_{(46)}$=−3.10, p = 0.003). Furthermore, a 2 (Condition: VLG, Control) × 3 (Orientation: Upright, Inverted, Rotated) mixed-effects repeated-measures ANOVA revealed a significant Condition × Orientation interaction ($F_{(2,176)}$=13.51, p<0.001, η_p^2=0.13), indicating a qualitatively different response pattern for the VLG versus Control stimuli. Thus, the greater place ratings for Upright and Inverted over Rotated VLG is unlikely to be driven by any particular details in the experimental design or a general effect of inverting or rotating the stimuli.

5. Discussion

The current study aimed to identify a stimulus feature that characterizes visual inputs as a scene, and thereby drives cortical scene processing. In Study 1, using a stimulus-driven approach, we observed a common VLG in the visual stimuli that elicited the greatest response in the scene-selective regions, including PPA, OPA and RSC. Consistently, we also found a positive and significant correlation between VLG in these complex, naturalistic stimuli and the neural response in the scene-selective regions, but not in LOC or EarlyVis. In Study 2, using a hypothesis-driven approach, we then directly tested whether VLG selectively drives cortical scene selectivity. Consistent with our hypothesis, even when we tightly controlled for visual features orthogonal to VLG, the scene-selective regions still showed a significantly greater response to the VLG images over the Control images with minimal VLG. In Study 3, we next further explored the behavioral relevance of VLG for visual scene recognition and found that participants also rated images of VLG as a "place" more often than as an "object". Taken together, these results reveal that VLG is a stimulus feature that drives cortical scene selectivity in adult human visual cortex, and that VLG is behaviorally relevant for visual scene recognition.

Our findings that the scene-selective regions respond selectively to VLG lend further support to the previous finding that cortical scene selectivity is driven by stimulus features that are common and unique to visual scenes (Cheng et al., 2021), and extend prior work by revealing VLG as another such feature that can account for cortical scene selectivity to a relatively large set of "non-Concave" scenes, such as landscape scenes (Epstein and Kanwisher 1998), convex buildings depicted in an outdoor environment (Cheng et al., 2021), and even night scenes. VLG may also explain cortical scene selectivity for objects that are large in real-world size (e.g., cars, furniture; Mullally and Maguire, 2011; Troiani et al., 2014; Kamps et al., 2016; Julian et al., 2017), as large objects also have a relatively large surface area to capture changes in luminance along the vertical dimension, albeit smaller than that of a scene.

To what extent does VLG drive cortical scene selectivity? In Study 2, since we did not directly compare cortical scene selectivity to VLG versus naturalistic scene stimuli (e.g., a real-world photograph of a forest), the precise magnitude of cortical scene selectivity driven by VLG remains unclear. However, given previous findings for cortical scene selectivity to stimuli that are not concave and do not have a prominent VLG, such as images of a building cut out from its background (Epstein and Kanwisher, 1998), it is evident that there exist other features that can drive cortical scene selectivity beyond VLG and concavity. Thus, one fruitful future direction is to identify these other features, and to pit them (together with concavity and VLG) against natural scene stimuli to elucidate their relative importance for human visual scene recognition. Furthermore, as cortical scene selectivity is likely driven by multiple common and unique scene features (i.e., concavity, VLG and others), another fruitful research direction is to investigate whether selectivity to these features is common or distinct across subpopulations of neurons within the scene-selective regions.

But how might neural sensitivity to VLG emerge in the scene-selective regions? There are two, and not necessarily mutually exclusive, plausible explanations. The first plausible explanation is that such sensitivity might be scaffolded by retinotopic biases to visual contrasts along the vertical dimension, given previous findings for retinotopic biases to the upper and lower visual field in PPA and OPA, respectively (Silson et al., 2015, 2016). The second plausible explanation is that such sensitivity is likely supported by earlier stages of visual processing. Specifically, a previous study reported that V4 neurons in non-human primates exhibited sensitivity to directions of lights and shading on objects, with a particular bias towards shading gradients that vary along vertical directions (Hanazawa and Komatsu, 2001). Coupled with findings for a distinct channel that is particularly tuned for scene-like stimuli along the ventral visual pathway of non-human primates (Vaziri et al., 2014), neurons in earlier stages of visual processing might be tuned to VLG (or at least aspects of VLG), thereby gatekeeping scene versus non-scene information into cortical scene processing. However, future research is needed to explore these plausible explanations.

In light of a growing body of literature that highlights distinct roles of the scene-selective regions for visual scene processing, we would like to make clear that our findings point to VLG as a stimulus feature that differentiates scene from non-scene stimuli, and does not imply that VLG also sufficiently enables a fine-grained, in-depth understanding of a scene, which is necessary for visual scene discrimination (e.g., differentiating images of a forest versus a beach) and navigation. While differentiating scene from non-scene stimuli involves stimulus features that are *common and unique* to visual scenes (like VLG found in this paper), achieving a fine-grained, in-depth understanding of a scene (enabling scene discrimination and navigation) involves stimulus features that are *different* among visual scenes. Consistently, the existing literature has highlighted a rich representation of highly variable scene features in scene-selective cortical regions, encompassing low-level to high-level features—such as contour junctions (Choo and Walther, 2016), textures (Cant and Goodale, 2011), and the geometric properties of the local environment (Dillon et al., 2018; Epstein and Kanwisher, 1998; Kravitz et al., 2011; Lescroart and Gallant, 2019; Park et al., 2011; for a comprehensive review, see Malcolm et al., 2016; Groen et al., 2017; Dilks et al., 2021)—that could be used to achieve different behavioral goals in scene understanding. Thus, we would like to make clear that our findings point to VLG as a stimulus feature that characterizes visual scenes as a domain of inputs (distinct from faces and objects), and not that VLG is a stimulus feature that differentiates different kinds of scenes or enables the precise behaviors necessary for scene discrimination and navigation.

Next, we would like to point out two caveats regarding our findings in Studies 2 and 3. The first caveat is that the Rotated VLG stimuli tested are made by rotating the upright VLG stimuli clockwise; as such, across all Rotated VLG stimuli, the right half of a stimulus is brighter than the left half, and we did not test for cortical scene selectivity for horizontal luminance gradient in which the left half of a stimulus is brighter than the right half. However, since we did not find an effect for horizontal luminance gradient when we accounted for both left-to-right and right-to-left luminance gradients in the BOLD5000 analysis, it is unlikely that left-to-right horizontal luminance gradient drives cortical scene selectivity. Another caveat is that the scene versus object ratings in Study 3 is obtained under an alternative forced choice task. As such, one possible explanation for the above-chance object ratings for Rotated VLG is that the participants might not truly think that the Rotated VLG stimuli are indeed more like objects, but rather are less like scenes relative to the Upright and Inverted VLG stimuli. To discern whether participants

truly consider the Rotated VLG stimuli to be more like objects, or merely less like scenes relative to Upright and Inverted VLG stimuli, one future research direction is to obtain scene and object ratings for these stimuli independently. Regardless, our results nevertheless show that visual stimuli of Upright and Inverted VLG are categorized as a scene more often than non-VLG stimuli, including the Rotated VLG and the Control stimuli, providing evidence for our hypothesis that VLG is a stimulus feature that humans use for visual scene recognition.

Finally, in addition to our main findings that VLG is a visual feature that drives cortical scene selectivity, we found that horizontal luminance gradient may drive cortical object processing in LO in both naturalistic, complex stimuli in BOLD5000 and artificial, tightly-controlled stimuli in Study 2. Moreover, in Study 3, we also found tightly-controlled stimuli of horizontal luminance gradient to be behaviorally categorized as an "object" more often than a "place". Together, these findings raise the intriguing possibility that horizontal luminance gradient may be a visual feature for human visual object recognition. Why might horizontal luminance gradient be a diagnostic feature of objects? One plausible explanation is that many real-world objects are small and they tend *not* to be placed directly under the light source. Coupled with their three-dimensional and multi-faceted nature, light likely reflects unevenly on object surfaces, causing the side closer to the light source to be brighter, and thus resulting in a common horizontal luminance gradient in objects. However, this plausible explanation is pure speculation; future research is therefore needed to investigate whether horizontal luminance gradient is indeed an intrinsic property of visual object stimuli.

In sum, we asked what characterizes visual inputs as a "scene", and thereby drives human cortical scene processing, and our results indicate that VLG is one such feature. Together with findings that concavity is another such feature, our findings provide further evidence that visual scenes can be characterized by a set of diagnostic features common and unique to visual scenes, and calls for future research to identify the rest of them.

Declaration of Competing Interest

The authors declare no competing financial interests.

Credit authorship contribution statement

Annie Cheng: Conceptualization, Methodology, Software, Formal analysis, Data curation, Project administration, Visualization, Writing – original draft, Writing – review & editing. **Zirui Chen:** Conceptualization, Methodology, Software, Formal analysis, Data curation, Project administration, Visualization. **Daniel D. Dilks:** Conceptualization, Methodology, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition, Data curation.

Data and code availability

The dataset generated during this study is available at https://osf.io/y3vuj/.

Acknowledgments

We would like to thank the Facility for Education and Research in Neuroscience (FERN) Imaging Center in the Department of Psychology, Emory University, Atlanta, GA. We would also like to thank Joshua Julian for sharing the code for calculating rectilinearity, and Dirk B. Walther for his comments and suggestions. This work was supported by a National Eye Institute (NEI) grant (R01EY029724) (DDD).

References

Aguirre, G.K., Zarahn, E., D'esposito, M, 1998. An area within human ventral cortex sensitive to "building" stimuli: evidence and implications. Neuron 21 (2), 373–383.
Bainbridge, W.A., Oliva, A., 2015. A toolbox and sample object perception data for equalization of natural images. Data Brief 5, 846–851.
Baldassano, C., Esteva, A., Fei-Fei, L., Beck, D.M., 2016. Two distinct scene-processing networks connecting vision and memory. eNeuro 3 (5).
Berman, D., Golomb, J.D., Walther, D.B., 2017. Scene content is predominantly conveyed by high spatial frequencies in scene-selective visual cortex. PLoS One 12 (12), e0189828.
Bryan, P.B., Julian, J.B., Epstein, R.A., 2016. Rectilinear edge selectivity is insufficient to explain the category selectivity of the parahippocampal place area. Front. Hum. Neurosci. 10, 137.
Cant, J.S., Goodale, M.A., 2011. Scratching beneath the surface: new insights into the functional properties of the lateral occipital area and parahippocampal place area. J. Neurosci. 31 (22), 8248–8258.
Chang, N., Pyles, J.A., Marcus, A., Gupta, A., Tarr, M.J., Aminoff, E.M., 2019. BOLD5000, a public fMRI dataset while viewing 5000 visual images. Sci. Data 6 (1), 1–18.
Cheng, A., Walther, D.B., Park, S., Dilks, D.D., 2021. Concavity as a diagnostic feature of visual scenes. Neuroimage 232, 117920.
Choo, H., Walther, D.B., 2016. Contour junctions underlie neural representations of scene categories in high-level human visual cortex. Neuroimage 135, 32–44.
Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009. Imagenet: a large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. Ieee, pp. 248–255.
Dilks, D.D., Julian, J.B., Paunov, A.M., Kanwisher, N., 2013. The occipital place area is causally and selectively involved in scene perception. J. Neurosci. 33 (4), 1331–1336.
Dilks, D.D., Kamps, F.S., Persichetti, A.S., 2021. Three cortical scene systems and their development. Trend. Cogn. Sci. (Regul. Ed.).
Dillon, M.R., Persichetti, A.S., Spelke, E.S., Dilks, D.D., 2018. Places in the brain: bridging layout and object geometry in scene-selective cortex. Cereb. Cortex 28 (7), 2365–2374.
Epstein, R., Kanwisher, N., 1998. A cortical representation of the local visual environment. Nature 392 (6676), 598.
Esteban, O., Markiewicz, C.J., Blair, R.W., Moodie, C.A., Isik, A.I., Erramuzpe, A., … Gorgolewski, K.J., 2019. fMRIPrep: a robust preprocessing pipeline for functional MRI. Nat. Method. 16 (1), 111–116.
Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzchak, Y., Malach, R., 1998. A sequence of object-processing stages revealed by fMRI in the human occipital lobe. Hum. Brain Mapp. 6 (4), 316–328.
Groen, I.I., Silson, E.H., Baker, C.I., 2017. Contributions of low-and high-level properties to neural processing of visual scenes in the human brain. Philosoph. Transact. Roy. Soc. B: Biolog. Sci. 372 (1714), 20160102.
Hanazawa, A., Komatsu, H., 2001. Influence of the direction of elemental luminance gradients on the responses of V4 cells to textured surfaces. J. Neurosci. 21 (12), 4490–4497.
Hebart, M.N., Dickter, A.H., Kidder, A., Kwok, W.Y., Corriveau, A., Van Wicklin, C., Baker, C.I., 2019. THINGS: a database of 1,854 object concepts and more than 26,000 naturalistic object images. PLoS One 14 (10), e0223792.
Julian, J.B., Ryan, J., Epstein, R.A., 2017. Coding of object size and object category in human visual cortex. Cereb. Cortex 27 (6), 3095–3109.
Kamps, F.S., Julian, J.B., Kubilius, J., Kanwisher, N., Dilks, D.D., 2016. The occipital place area represents the local elements of scenes. Neuroimage 132, 417–424.
Kauffmann, L., Ramanoël, S., Peyrin, C., 2014. The neural bases of spatial frequency processing during scene perception. Front. Integr. Neurosci. 8, 37.
Konkle, T., Brady, T.F., Alvarez, G.A., Oliva, A., 2010. Scene memory is more detailed than you think: the role of categories in visual long-term memory. Psychol. Sci. 21 (11), 1551–1556.
Kravitz, D.J., Peng, C.S., Baker, C.I., 2011. Real-world scene representations in high-level visual cortex: it's the spaces more than the places. J. Neurosci. 31 (20), 7322–7333.
Lescroart, M.D., Gallant, J.L., 2019. Human scene-selective areas represent 3D configurations of surfaces. Neuron 101 (1), 178–192.
Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ..., Zitnick, C.L., 2014. Microsoft coco: common objects in context. In: European conference on computer vision. Springer, Cham, pp. 740–755.
Maguire, E., 2001. The retrosplenial contribution to human navigation: a review of lesion and neuroimaging findings. Scand. J. Psychol. 42 (3), 225–238.
Malcolm, G.L., Groen, I.I., Baker, C.I., 2016. Making sense of real-world scenes. Trend. Cogn. Sci. (Regul. Ed.) 20 (11), 843–856.
Marchette, S.A., Vass, L.K., Ryan, J., Epstein, R.A., 2014. Anchoring the neural compass: coding of local spatial reference frames in human medial parietal lobe. Nat. Neurosci. 17 (11), 1598–1606.
Mullally, S.L., Maguire, E.A., 2011. A new role for the parahippocampal cortex in representing space. J. Neurosci. 31 (20), 7441–7449.
Nasr, S., Tootell, R.B., 2012. A cardinal orientation bias in scene-selective visual cortex. J. Neurosci. 32 (43), 14921–14926.
Nasr, S., Echavarria, C.E., Tootell, R.B., 2014. Thinking outside the box: rectilinear shapes selectively activate scene-selective cortex. J. Neurosci. 34 (20), 6721–6735.
Oliva, A., Schyns, P.G., 2000. Diagnostic colors mediate scene recognition. Cogn. Psychol. 41 (2), 176–210.
Park, S., Chun, M.M., 2009. Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. Neuroimage 47 (4), 1747–1756.
Park, S., Brady, T.F., Greene, M.R., Oliva, A., 2011. Disentangling scene content from spatial boundary: complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. J. Neurosci. 31 (4), 1333–1340.
Persichetti, A.S., Dilks, D.D., 2019. Distinct representations of spatial and categorical relationships across human scene-selective cortex. Proc. Natl Acad. Sci. 116 (42), 21312–21317.
Rajimehr, R., Devaney, K.J., Bilenko, N.Y., Young, J.C., Tootell, R.B., 2011. The "parahippocampal place area" responds preferentially to high spatial frequencies in humans and monkeys. PLoS Biol. 9 (4), e1000608.

Silson, E.H., Chan, A.W.Y., Reynolds, R.C., Kravitz, D.J., Baker, C.I., 2015. A retinotopic basis for the division of high-level scene processing between lateral and ventral human occipitotemporal cortex. J. Neurosci. 35 (34), 11921–11935.
Silson, E.H., Gilmore, A.W., Kalinowski, S.E., Steel, A., Kidder, A., Martin, A., Baker, C.I., 2019. A posterior–anterior distinction between scene perception and scene construction in human medial parietal cortex. J. Neurosci. 39 (4), 705–717.
Silson, E.H., Groen, I.I., Kravitz, D.J., Baker, C.I., 2016. Evaluating the correspondence between face-, scene-, and object-selectivity and retinotopic organization within lateral occipitotemporal cortex. J. Vis. 16 (6) 14-14.
Smith, S.M., 2002. Fast robust automated brain extraction. Hum. Brain Mapp. 17 (3), 143–155.
Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., … Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23, S208–S219.
Troiani, V., Stigliani, A., Smith, M.E., Epstein, R.A., 2014. Multiple object properties drive scene-selective regions. Cereb. Cortex 24 (4), 883–897.
Vaziri, S., Carlson, E.T., Wang, Z., Connor, C.E., 2014. A channel for 3D environmental shape in anterior inferotemporal cortex. Neuron 84 (1), 55–62.
Wang, L., Mruczek, R.E., Arcaro, M.J., Kastner, S., 2015. Probabilistic maps of visual topography in human cortex. Cereb. Cortex 25 (10), 3911–3931.
Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., Torralba, A., 2010, June. Sun database: Large-scale scene recognition from abbey to zoo. In: *2010 IEEE computer society conference on computer vision and pattern recognition*. IEEE, pp. 3485–3492.