Title: Foreground Enhancement and Background Suppression in Human Early Visual System During Passive Perception of Natural Images

Abbreviated title: Scene Segmentation of Natural Images in Humans

Author names and affiliations: Paolo Papale[1], Andrea Leo[1,2], Luca Cecchetti[1], Giacomo Handjaras[1], Kendrick Kay[3], Pietro Pietrini[1*] and Emiliano Ricciardi[1]

1. Molecular Mind Lab, IMT School for Advanced Studies Lucca, Lucca, 55100 Italy

2. Research Center "E.Piaggio", University of Pisa, Pisa, 56122, Italy

3. Center for Magnetic Resonance Research, Department of Radiology, University of Minnesota, Twin Cities, Minneapolis, MN, 55455, USA

***Corresponding author**: Pietro Pietrini, MD, PhD. IMT School for Advanced Studies Lucca, Piazza San Francesco, 19, 55100, Lucca, Italy. *pietro.pietrini@imtlucca.it*

Number of pages: 23

Number of words: abstract(242); introduction(640); discussion(1221):

Conflict of Interest: The authors declare no competing financial interests.

Number of figures: 6

Abstract

One of the major challenges in visual neuroscience is represented by foreground-background segmentation, a process that is supposed to rely on computations in cortical modules, as information progresses from V1 to V4. Data from nonhuman primates (Poort et al., 2016) showed that segmentation leads to two distinct, but associated processes: the enhancement of cortical activity associated to figure processing (i.e., foreground enhancement) and the suppression of ground-related cortical activity (i.e., background suppression). To characterize foreground-background segmentation of natural stimuli in humans, we parametrically modulated low-level properties of 334 images and their behaviorally segmented counterparts. A model based on simple visual features was then adopted to describe the filtered and intact images, and to evaluate their resemblance with fMRI activity in different visual cortices (V1, V2, V3, V3A, V3B, V4, LOC). Results from representational similarity analysis (Kriegeskorte et al., 2008) showed that the correspondence between behaviorally segmented natural images and brain activity increases throughout the visual processing stream. We found evidence of foreground enhancement for all the tested visual regions, while background suppression occurs in V3B, V4 and LOC. Our results suggest that foreground-background segmentation is an automatic process that occurs during natural viewing, and cannot be merely ascribed to differences in objects size or location. Finally, "neural images" reconstructed from V4 and LOC fMRI activity revealed a preserved spatial resolution of foreground textures, indicating a richer representation of the salient part of natural images, rather than a simplistic model of objects shape.

Significance Statement

In the path from continuous sensory percepts to discrete categorical representations, foreground-background segmentation has been considered a pivotal step, in order to make sense

of the surrounding visual environment. Our findings provide novel support to the hypothesis that foreground-background segmentation of natural scenes during passive perception is an automatic process sustained by the distributed activity of multiple areas across the visual processing stream. Specifically, V3B, V4 and LOC show a background suppression effect, while retaining texture information from the foreground. These observations challenge the idea that these regions of the visual system may primarily encode simple object representations based on silhouette or shape features only.

Introduction

In the scientific journey toward a satisfying understanding of the human visual system, scene segmentation represents a central problem "for which no theoretical solution exists" (Wu et al., 2006). Indeed, segmentation into foreground and background is crucial to make sense of the surrounding visual environment, and its pivotal role as an initial step of visual content identification has long been theorized (Biederman, 1987). However, although humans naturally segment during active visual processing, no computational model is currently able to achieve comparable performances in scene segmentation (Arbelaez et al., 2011). Furthermore, several appearance-based computational models could successfully perform, albeit with sub-optimal accuracy, visual content recognition of natural images without the aid of foreground-background segmentation, thus challenging its role in visual identification (Oliva and Torralba, 2001; Lazebnik et al., 2006).

To date, numerous neurophysiological studies found evidence of texture segmentation and figure-ground organization in the early visual cortex of nonhuman primates (Lamme, 1995; Lee et al., 1998; Poort et al., 2012; Self et al., 2013; Kok and de Lange, 2014) and humans (Kastner et al., 2000; Scholte et al., 2008). In particular, a recent study on nonhuman primates attending artificial

stimuli revealed an early enhancement of V1 and V4 neurons when their receptive fields covered the foreground, and a later response suppression when their receptive fields were located in the stimulus background (Poort et al., 2016). This demonstrates that foreground enhancement and background suppression are distinct but associated processes involved in segmentation.

Other authors questioned the classical view of figure-ground segmentation as a compulsory bottom-up process in visual content recognition and proposed that identification precedes segmentation in a top-down manner (Peterson, 1994; Peterson and Gibson, 1994). In addition, from an experimental viewpoint, the role of visual segmentation has been demonstrated only by means of non-ecological stimuli (e.g., binary figures, random dots, oriented line segments and textures). Although two recent studies investigated border-ownership in monkeys with both artificial and natural stimuli (Hesse and Tsao, 2016; Williford and von der Heydt, 2016), a proof of the occurrence of scene segmentation in the human brain during visual processing of naturalistic stimuli (e.g., natural images and movies) is still lacking.

In light of this, we specifically investigated foreground enhancement and background suppression, as specific processes involved in segmentation, during passive viewing of natural images in humans. We used fMRI data, previously published by Kay and colleagues (Kay et al., 2008), to study brain activity from seven visual regions of interest (ROIs): V1, V2, V3, V3A, V3B, V4 and lateral occipital cortex (LOC) during the passive perception of 334 natural images, whose "ground-truth" segmented counterparts have been included in the Berkeley Segmentation Dataset (BSD) (Arbelaez et al., 2011).

Notwithstanding, as a reliable computational model of scene segmentation has not been achieved yet, we developed a novel pre-filtering modeling approach to study the response to complex, natural images without relying on explicit models. Our method is similar to other approaches where explicit computations are performed on representational features rather than

on the original stimuli (Naselaris et al., 2011). For instance, these methods have been recently adopted to investigate semantic representation (e.g. Huth et al., 2012; Handjaras et al., 2016) or scene segmentation (Lescroart et al., 2016).

However, as opposed to the standard modeling framework – according to which alternative models are computed from the stimuli to predict brain responses – here, low-level features of the stimuli are parametrically modulated and simple descriptors of each filtered image (e.g., edges position, size and orientation) are aggregated in a fixed biologically plausible model (Figure 1). The correspondence between the fixed model and fMRI patterns evoked by the intact naturalistic images, was then assessed using representational similarity analysis (RSA) (Kriegeskorte et al., 2008). Notably, this approach can also be exploited to obtain highly informative "neural images" representing the computations of different brain regions and may be generalized to investigate different phenomena in visual neuroscience.

Materials and Methods

To assess differences between cortical processes involved in foreground-background segmentation, we employed a low-level description of images, defined by averaging the representational dissimilarity matrices (RDMs) of four well-known computational models (Figure 2D). The average model is based on simple features - such as edge position, size and orientation - whose physiological counterparts are well known (Marr, 1982). This model was kept constant while the images were parametrically filtered and iteratively correlated with brain activity through RSA. For each cortical module, this pre-filtering modeling approach led to a visual representation of the optimal features (contrast and spatial frequencies) of foreground and background of natural images. The analytical pipeline is schematized in Figure 2.

Stimuli.

From the 1870 images used by (Kay et al., 2008) a sub-sample of 334 pictorial stimuli, which are represented also in the Berkeley Segmentation Dataset (BSD), was selected (Arbelaez et al., 2011). For every BSD image, five subjects manually performed an individual "ground-truth" segmentation, which is provided by the authors of the dataset (http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/resources.html). Thus, for each of the 334 images, we manually selected the largest foreground patch from one of the five behavioral segmentations, in order to build the foreground binary mask. This mask was down-sampled and applied to the original stimulus (Kay et al., 2008). Stimuli are publicly available and can be downloaded at: http://crcns.org/data-sets/vc/vim-1.

fMRI Data.

The fMRI data used in this study are also publicly available at http://crcns.org/data-sets/vc/vim-1. Two subjects were acquired using the following MRI parameters: 4T INOVA MR, matrix size 64x64, TR 1s, TE 28ms, flip angle 20°, spatial resolution 2 x 2 x 2.5 mm^3. For additional details on pre-processing, acquisition parameters, retinotopic mapping and ROI localizations, please refer to (Kay et al., 2008).

Computer Vision Models.

In accordance with a previous fMRI study, we selected four well-assessed untrained computational models which showed significant correlations with brain activity patterns in early visual areas as well as LOC (Khaligh-Razavi and Kriegeskorte, 2014). The four models comprise: GIST (Oliva and Torralba, 2001), Dense SIFT (Lazebnik et al., 2006), Pyramid Histograms of Gradients (PHOG) (Bosch et al., 2007) and Local Binary Patterns (LBP) (Ojala et al., 2001). For an

exhaustive description of the four models – and links to Matlab codes – see the work by Khaligh-Razavi (2014) and Khaligh-Razavi and Kriegeskorte (2014).

Permuted segmentations.

A permutation test was performed to assess the statistical significance of the foreground selection obtained from the behavioral segmentations, and to rule out a possible "fovea-to-periphery" bias (see Results). In each iteration of this procedure, the 334 foreground masks were shuffled and a random foreground segmentation was associated to each stimulus. Of note, this set of randomly-segmented images had the same distribution of masked portions of the visual field as the one from the behavioral segmentation. This procedure was repeated 1,000 times, to build a null distribution of alternative segmentations: four examples of random segmentation are shown in Figure 2B. For each permutation, features were extracted from every image obtained by applying a random foreground mask to a stimulus, and RSA was performed using the procedure described below.

Parametric filtering procedures

In order to investigate differential processing of foreground and background in the early visual system, we employed three different filtering procedures (alpha channel modulation, low- and high-pass filtering of spatial frequencies) applied parametrically (100 steps each) to the foreground or the background of each image. For each procedure, three examples of filtered images are represented in Figure 2C. For low- and high-pass filtering, we employed a Butterworth filter, linearly sampling from a log-transformed distribution of frequencies ranging from 0.05 to 25 cyc/°, while keeping the root mean squared (RMS) contrast fixed. In addition, for each step and each ROI, we computed the differences between the Spearman's correlation coefficients of the

fMRI representational dissimilarity matrix (RDM) and the background and foreground feature-based RDMs, respectively. For each filtering procedure (i.e. alpha channel, low- and high-pass), these differences were then averaged, to represent their impact on both foreground and background (Figure 4A).

In order to assess whether low-level properties of the foreground borders might explain the similarity between the isolated foreground mask and brain activity, a control filtering procedure has been computed. The BSD behavioral masks were processed using a parametric Gaussian filter, whose radius increased by 2 pixels at each step while keeping the segmented area constant. The resulting mask was then applied to the original stimuli. For each of these steps the correlation with fMRI activity patterns was computed and compared with the BSD behavioral segmentation. Three examples of this procedure are represented in Figure 5G and the results are displayed in Figure 5H.

Representational Similarity Analysis (RSA).

For each filtered image, we collected feature vectors from the four computational models (GIST, PHOG, LBP and Dense SIFT), and RDMs were then obtained (1 minus the Pearson correlation metric). These four RDMs were normalized in a range between 0 and 1, and averaged to obtain the fixed biologically plausible model of the stimuli (for a graphical representation of the process, see Figure 2D). Single subject RDMs were similarly computed using fMRI activity patterns for each of the seven ROIs, and then averaged across the two subjects. We used Spearman's rho (ρ) to assess the correlation between the RDMs from each step of the filtering procedures and the RDMs from the brain ROIs. In addition, as different ROIs may show different levels of signal-to-noise ratio, we computed a noise estimation by correlating the RDMs from each ROI between subjects. These ROI-specific noise estimations were used to normalize the correlation coefficients, reported as

188 normalized Spearman's rho (Nρ) in the figures. The same normalization procedure has been
189 employed also for voxel-wise encoding by (Huth et al., 2016).
190
191 *Neural images*
192 For each ROI, the effects of the three filtering procedures were then combined, to build the
193 *post-hoc* "neural image". To this aim we used the filtering step with the highest correlation
194 between the fixed model and brain activity, for foreground and background. In detail, we
195 averaged the best images for the low- and high-pass filters, and multiplied each pixel for the
196 preferred alpha-channel value. Lastly, the foreground mask employed for the neural images was
197 chosen as the best step in Gaussian filtering procedure described above.
198 All analyses have been performed with Matlab (The Mathworks Inc.).
199
200 **Results**
201 *Comparison of intact and behaviorally segmented images*
202 To compare whether the RDMs of the intact stimuli and RDMs of the isolated foreground
203 differentially correlate with brain activity, two fixed descriptions of the stimuli were created
204 (Figure 2). RSA results (Kriegeskorte et al., 2008) showed that the intact and segmented version of
205 the stimuli have different correlation patterns (Figure 3A): the correlation between the segmented
206 RDM and fMRI activity increases as progressing along the hierarchy of the visual cortices, from V1
207 to LOC, with maximum correlation values in V4 and LOC (V1: Nρ = 0.07; V2: Nρ = 0.11; V3: Nρ =
208 0.14; V3A: Nρ = 0.32; V3B: Nρ = 0.27; V4: Nρ = 0.35; LOC: Nρ = 0.40). On the contrary, the intact
209 description reveals a decrease in correlation beyond V1 and V2, with the exception of V3A only
210 (V1: Nρ = 0.43; V2: Nρ = 0.49; V3: Nρ = 0.32; V3A: Nρ = 0.45; V3B: Nρ = 0.27; V4: Nρ = 0.26; LOC:

211 Nρ = 0.35). These different trends related to the intact and segmented descriptions fostered
212 further analyses.
213
214 *Foreground Enhancement*
215 A higher correlation for the foreground description as compared to the intact images was
216 found only in V4 and LOC, however a recent electrophysiological study on monkeys found
217 evidence for foreground enhancement also in earliest visual cortices (Poort et al., 2016). Thus, to
218 test whether the behavioral foreground segmentation from BSD was more tied to brain activity as
219 compared to alternate configurations obtained by shuffling the segmentation patterns across
220 stimuli (Figure 2B), we performed a specific analysis based on a permutation test.
221 As depicted in Figure 3A, for all the ROIs, the correct foreground configuration yielded a
222 significantly higher correlation as compared to the examples from the shuffled dataset, thus
223 suggesting that foreground enhancement is actually involved in scene segmentation of natural
224 images during passive perception (V1: $p = 0.002$; V2: $p < 0.001$; V3: $p < 0.001$; V3A: $p < 0.001$; V3B:
225 $p < 0.001$; V4: $p < 0.001$; LOC: $p < 0.001$).
226 This analysis also accounted for a potential confounding effect related to a "fovea-to-
227 periphery bias" in our image set - represented in Figure 3B. In fact, as already observed in
228 literature, natural images are typically characterized by objects located at the center of the scene
229 (see for instance the object location bias represented in Figure 3B in (Alexe et al., 2010)). However,
230 we replicated the same "fovea-to-periphery bias" in the null distribution, to rule out that
231 foreground enhancement could be driven by differences between the representation of fovea and
232 periphery across the set of images.
233
234 *Background Suppression*

The different correlation trends showed by RDMs of intact and segmented descriptions also suggested that the background-related information was suppressed in higher visual cortices, thus explaining the lowest performance of the intact description in V4 and LOC as compared to the description of the isolated foreground. Notably, Poort and colleagues (2016) described background suppression as a different, but associated, phenomenon with respect to foreground enhancement. Thus, in order to better characterize where and how background suppression occurs in humans attending to natural images, a further analysis was performed by parametrically filtering out the foreground, or the background, of each image, varying their contrast or spatial frequencies (low- and high-pass filtering; Figure 2C). RSA results for the parametric filtering approach are summarized in Figure 4, while results relative to each single procedure are shown in Figure 5A-F. Independently from the filtering procedure employed, background and foreground filtering showed different correlation trends: while filtering out the foreground (i.e., isolating the background) results in a correlation drop in all the ROIs, filtering out the background (i.e., isolating the foreground) leads to an increased correlation in higher regions such as V3B, V4 and LOC (Figure 4A). This effect is accounted neither by differences in the extent of visual field occupied by foreground or background nor by the "fovea-to-periphery bias". In fact, we replicated the same filtering procedures using a foveal mask whose area was kept constant and equal to the mean area of the actual foreground masks. As depicted in Figure 6, the difference between background and foreground was not accounted by differential processing of periphery and fovea.

Moreover, an additional control analysis was performed to assess the impact of low-level properties of foreground borders. A Gaussian filter was parametrically applied to the foreground masks and the resulting correlation pattern in each ROI was measured (Figure 5 G-H). The unfiltered behavioral mask showed high correlations in all ROIs (V1: max step = 6 out of 100; 12px radius; V2: max step = 1 out of 100; 0px radius; V3: max step = 1 out of 100; 0px radius; V3A: max

step = 1 out of 100; 0px radius; V3B: max step = 1 out of 100; 0px radius; V4: max step = 4 out of 100; 8px radius; LOC: max step = 3 out of 100; 6px radius).

Discussion

In the present study, we illustrated how the manipulation of low-level properties of natural images, and the following correlation with brain responses during passive viewing of the intact stimuli, could disclose the behavior of different brain regions along the visual pathway.

Employing this pre-filtering modeling approach, we were able to collect three different evidence indicating that scene segmentation is an automatic process that occurs during passive perception in naturalistic conditions, even when individuals are not required to perform any particular tasks, or to focus on any specific aspect of the images.

First, we demonstrated that the correlation of fMRI patterns with foreground-related information increases along the visual hierarchy, culminating in V4 and LOC. In addition, foreground-related information in these two regions is more linked to brain activity than intact stimuli.

Second, our analyses specifically found that foreground enhancement is present in all the selected visual ROIs, and that this effect is driven neither by the foreground inked area, nor by its location in the visual field. Thus, indirect evidence of scene segmentation of natural images could be retrieved in the activity of multiple early areas of the visual processing stream. This is consistent with a recent study, which reported that border-ownership of natural images cannot be resolved by single cells, but requires a population of cells in monkey V2 and V3 (Hesse and Tsao, 2016).

Finally, an additional proof of segmentation can be represented by the suppression of background-related information in V3B, V4 and LOC. On the contrary, earlier regions across the

visual stream - from V1 to V3 – have a uniform representation of the whole image, as evident at first glance in Figure 4B. Overall, these results further support the idea that foreground enhancement and background suppression are distinct, but associated, processes involved in scene segmentation of natural images.

Foreground segmentation as a proxy for shape processing

The success of the segmented description over the intact counterpart in explaining the functioning of V4 and LOC is consistent with several investigations on shape features selectivity in these regions, and in their homologues in monkey (Hung et al., 2012; Lescroart and Biederman, 2013; Vernon et al., 2016). In fact, the extraction of shape properties requires a previous segmentation (Lee et al., 1998), and presumably occurs in brain regions where background is already suppressed. Notably, the "neural images" reconstructed from V3B, V4 and LOC are characterized by a strong background suppression, while the foreground is preserved. This is consistent with a previous neuropsychological observation: a bilateral lesion within area V4 led to longer response times in idefintying overlapping figures (Leek et al., 2012). Hence, this region resulted to be crucial for accessing foreground-related computations, performed in earlier stages of visual processing, and presumably plays a role in matching the segmented image with stored semantic content in figure recognition. In accordance with this, a recent hypothesis suggests a role of V4 in higher-level functions, such as features integration or contour completion (Roe et al., 2012).

The preserved spatial resolution of foreground descriptive features (i.e., texture) in V4 and LOC represent an additional noteworthy aspect that arises from our data. The progression from V1 towards higher-level regions of the cortical visual pathway is associated with a relative increase in receptive fields size (Dumoulin and Wandell, 2008; Freeman and Simoncelli, 2011; Kay et al., 2015).

In addition, it should be kept in mind that regions such as V4 demonstrate a complete representation of the contralateral visual hemifield, rather than selective responses to stimuli locate above or below the horizontal meridian (Wandell and Winawer, 2011). The evidence that the foreground portion of "neural images" maintains fine-grained details in V4 and LOC seems to contrast the traditional view according to which these regions are more tuned to object shape (i.e., silhouettes), instead of being selective for the internal configuration of images (e.g. Malach et al., 1995; Grill-Spector et al., 1998; Moore and Engel, 2001; Stanley and Rubin, 2003). However, it has been shown that foveal and peri-foveal receptive fields of V4 do accomodate fine details of the visual field (Freeman and Simoncelli, 2011) and that the topographic representation of the central portion of this area is based on a direct sampling of the primary visual cortex retinotopic map (Motter, 2009). Therefore, given the "fovea-to-periphery" bias found in our stimuli and in natural images, it is reasonable that an intact configuration of the foreground may be more tied to the activity of these brain regions, and that a richer representation of the salient part may overcome simplistic models of objects shape (i.e., silhouettes).

Lastly, it is well known that selective attention represents one of the "active" cognitive mechanisms supporting figure segmentation (Qiu et al., 2007; Poort et al., 2012), as suggested, for instance, by bistable perception phenomena (Sterzer et al., 2009) or by various neuropsychological tests (e.g. De Renzi et al., 1969; Bisiach et al., 1976). In the present experiment, participants were asked to simply gaze a central fixation point without performing any overt or covert tasks related to the presented image. Nonetheless, we found evidence of a clear background suppression and foreground enhancement in several regions of the visual stream, suggesting that scene segmentation is mediated by an automatic process tha may be driven either by bottom-up (e.g., low-level properties of the foreground configuration), or top-down (e.g., semantic knowledge) attentional mechanisms.

331

332 *Facing the challenge of explicit modeling in visual neuroscience*

333 As predicting brain responses in ecological conditions is one the major goals of visual
334 neuroscience, our study showed that the sensitivity of fMRI pattern analysis can represent an
335 adequate tool to investigate complex phenomena through the richness of natural stimuli.

336 The standard approach in investigating visual processing in ecological conditions implies
337 testing the correlation of brain responses from a wide range of natural stimuli with features
338 extracted by different alternative computational models. This approach facilitates the comparison
339 between the performances of competing models and could ultimately lead to the definition of a
340 more plausible model of brain activity. However, the development of explicit computational
341 models for many visual phenomena in ecological conditions is difficult, as testified by the
342 extensive use of artificial stimuli in visual neuroscience (e.g. Carandini et al., 2005; Wu et al., 2006).

343 Actually, even if computer vision is a major source of computational models and feature
344 extractors, often its objectives hardly overlap with those of visual neuroscience. Computer
345 scientists are mainly interested in solving single, distinct tasks (e.g., segmentation, recognition,
346 etc.), while, from the neuroscientific side, the visual system is considered as a general-purpose
347 system that could adapt itself to perform different behaviors (Medathati et al., 2016).
348 Consequently, while computer science typically employs solutions that rely only seldom on
349 previous neuroscientific knowledge, visual neuroscience frequently lacks of solid computational
350 models, ending up with several arbitrary assumptions in modeling, especially for mid-level vision
351 processing, such as scene segmentation or shape features extraction (for a definition see: Kubilius
352 et al., 2014).

353 In light of this, we believe that the manipulation of a wide set of natural images, and the
354 computation of a fixed model based on low-level features, can offer a simple and biologically

plausible tool to investigate brain activity related to higher-order computations. In fact, the results of this procedure can be depicted and are more intuitive as compared to the description obtained through formal modeling (Figure 4B), thus highlighting interpretable differences rather than data predictions.

Figure legends:

Figure 1. Comparing the Standard Modeling Approach and the Pre-filtering Modeling Approach. A) In the standard modeling pipeline, different models are compared. After extracting features from the stimuli, competing feature vectors can be used in order to predict brain activity in an encoding procedure, or stimuli dissimilarities can be used in a representational similarity analysis. Finally, the model that better predicts brain responses is discussed. B) In our pre-filtering modeling approach, different filtered versions of the original stimuli are compared. Various biologically plausible filtering procedures are applied to the stimuli prior to compute a unique feature space according to a given fixed and easily interpretable model. In our approach a single model is employed and the best step of each filtering procedure is used to build a *post-hoc* "neural image", to visually interpret the results. While the standard modeling approach is theoretically more advantageous, as its output is a fully computable model of brain activity, it can not be applied when reliable explicit models of the perceptual process do not exist yet, as in the case of scene segmentation. Alternative attempts to reconstruct visual stimuli from brain activity have been previously reported using decoding techniques (e.g. Stanley et al., 1999; Thirion et al., 2006; Miyawaki et al., 2008; Nishimoto et al., 2011).

Figure 2. Analytical Pipeline.

A) An example of intact image and its behaviorally segmented counterpart B) The set of segmented stimuli is tested against a null distribution of 1,000 permutations. Each permutation is built by randomly shuffling the 334 behavioral foreground masks C) Three steps (20, 50 and 80 out of 100) for the contrast or spatial frequencies filtering of foreground and background. D) In clockwise order: features for each model were extracted from the stimuli; the dissimilarity (1 - Pearson's r) between each stimulus pair was computed and aggregated in four representational dissimilarity matrices (RDMs); the obtained RDMs were normalized in a 0-1 range; finally, the four RDMs were averaged in the unique appearance-based RDM, which was correlated to brain activity patterns in the subsequent analyses.

Figure 3. Foreground Enhancement in the Human Early Visual System.

A) Results for RSA: the correlation between the segmented version of the images and brain activity increased across the ROIs in a way respectful of the hierarchical organization of visual cortices; conversely the intact version does not show a similar trend. In addition, to test foreground enhancement and rule out a "fovea-to-periphery" bias, the behavioral segmentation was tested against a null distribution of shuffled masks made of 1000 permutations, and yielded a significant correlation for all the tested ROIs. B) The biased distribution of foreground masks in the 20° of visual field covered by the stimuli from Kay and colleagues (Kay et al., 2008). The color-bar represents the number of times each pixel is comprised in a foreground mask.

Figure 4. Background Suppression in the Human Early Visual System.

A) Mean correlation difference between background and foreground filtering. For each ROI and each iteration, the mean difference between the correlation of brain activity with background and foreground filtering is represented. Positive values indicate higher correlation due to filtering-out

the background (i.e., isolating the foreground), while negative values indicate higher correlation due to filtering-out the foreground (i.e., isolating the background). B) Neural images have been obtained as the combination of the steps of the filtering procedures (contrast, Gaussian, low- and high-pass filtering) which show the higher correlation with brain activity in each ROI (see Methods).

Figure 5. Results of the Filtering Procedures.

Correlation pattern between brain activity and the contrast, high- and low-pass filtering applied to the foreground (A, C, E) and to the background (B, D, F). G) Three examples of the Gaussian filtering procedure (at step 20, 50 and 80 out of 100). H) Correlation pattern of the Gaussian filter.

Figure 6. Segmentation is driven by differential processing of foveal and peripheral information.

Mean difference between periphery and fovea (see Results). In order to test whether background suppression could be explained by the fovea-to-periphery bias or by the different area of foreground and background, we repeated the filtering analysis using a fixed foveal mask equal to the mean area of the foreground masks. As depicted, the differences between background and foreground (in black) are not driven by the differences between periphery and fovea (red to blue).

References

Alexe B, Deselaers T, Ferrari V (2010) ClassCut for Unsupervised Class Segmentation. Lect Notes Comput Sc 6315:380-393.

Arbelaez P, Maire M, Fowlkes C, Malik J (2011) Contour detection and hierarchical image segmentation. IEEE Trans Pattern Anal Mach Intell 33:898-916.

Biederman I (1987) Recognition-by-Components - a Theory of Human Image Understanding. Psychological Review 94:115-147.

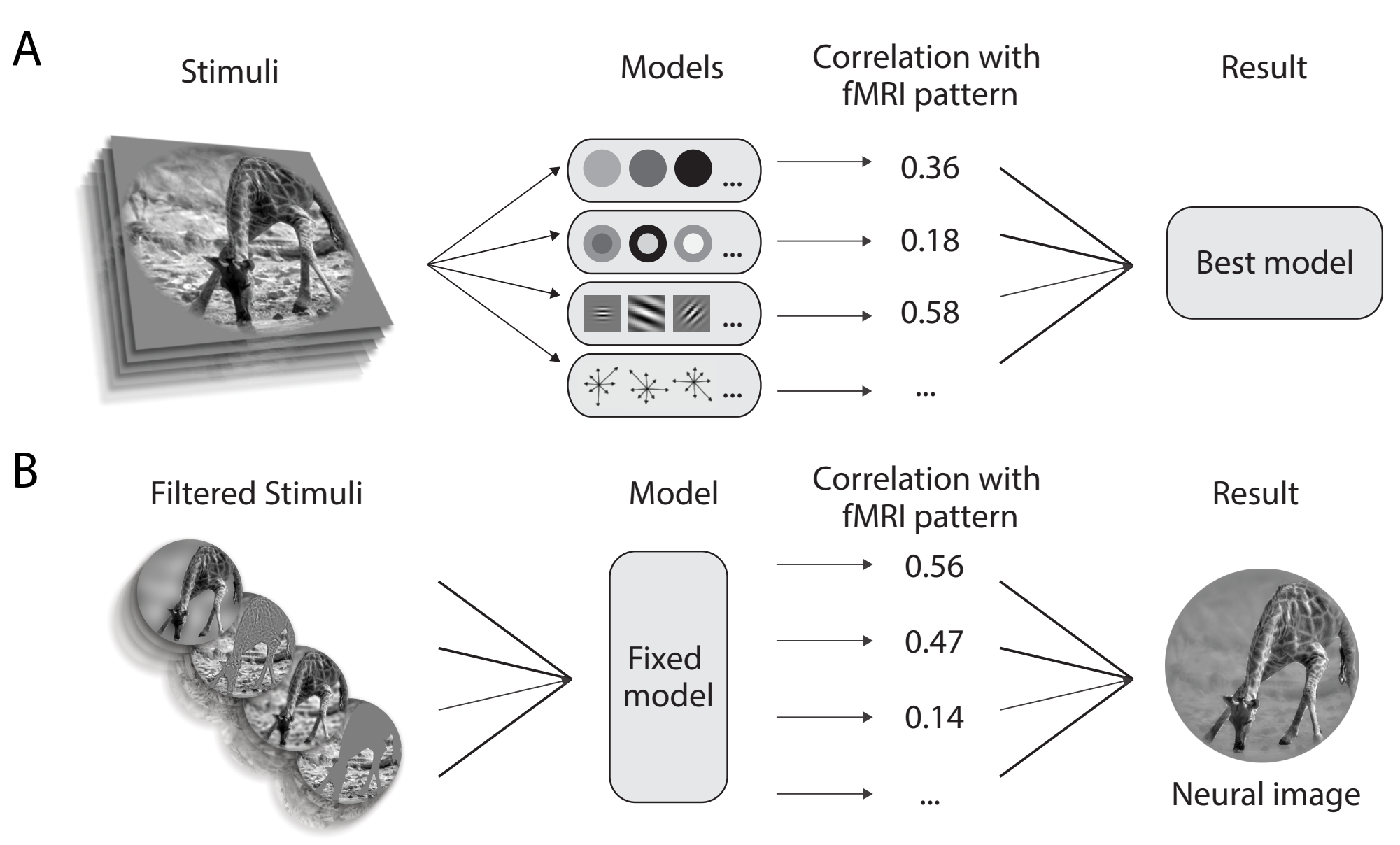
427 Bisiach E, Capitani E, Nichelli P, Spinnler H (1976) Recognition of overlapping patterns and focal
428 hemisphere damage. Neuropsychologia 14:375-379.
429 Bosch A, Zisserman A, Munoz X (2007) Representing shape with a spatial pyramid kernel. In, pp
430 401-408: ACM.
431 Carandini M, Demb JB, Mante V, Tolhurst DJ, Dan Y, Olshausen BA, Gallant JL, Rust NC (2005) Do
432 we know what the early visual system does? J Neurosci 25:10577-10597.
433 De Renzi E, Scotti G, Spinnler H (1969) Perceptual and associative disorders of visual recognition.
434 Relationship to the side of the cerebral lesion. Neurology 19:634-642.
435 Dumoulin SO, Wandell BA (2008) Population receptive field estimates in human visual cortex.
436 Neuroimage 39:647-660.
437 Freeman J, Simoncelli EP (2011) Metamers of the ventral stream. Nat Neurosci 14:1195-1201.
438 Grill-Spector K, Kushnir T, Edelman S, Itzchak Y, Malach R (1998) Cue-invariant activation in object-
439 related areas of the human occipital lobe. Neuron 21:191-202.
440 Handjaras G, Ricciardi E, Leo A, Lenci A, Cecchetti L, Cosottini M, Marotta G, Pietrini P (2016) How
441 concepts are encoded in the human brain: A modality independent, category-based
442 cortical organization of semantic knowledge. NeuroImage 135:232-242.
443 Hesse JK, Tsao DY (2016) Consistency of Border-Ownership Cells across Artificial Stimuli, Natural
444 Stimuli, and Stimuli with Ambiguous Contours. J Neurosci 36:11338-11349.
445 Hung CC, Carlson ET, Connor CE (2012) Medial axis shape coding in macaque inferotemporal
446 cortex. Neuron 74:1099-1113.
447 Huth AG, Nishimoto S, Vu AT, Gallant JL (2012) A continuous semantic space describes the
448 representation of thousands of object and action categories across the human brain.
449 Neuron 76:1210-1224.

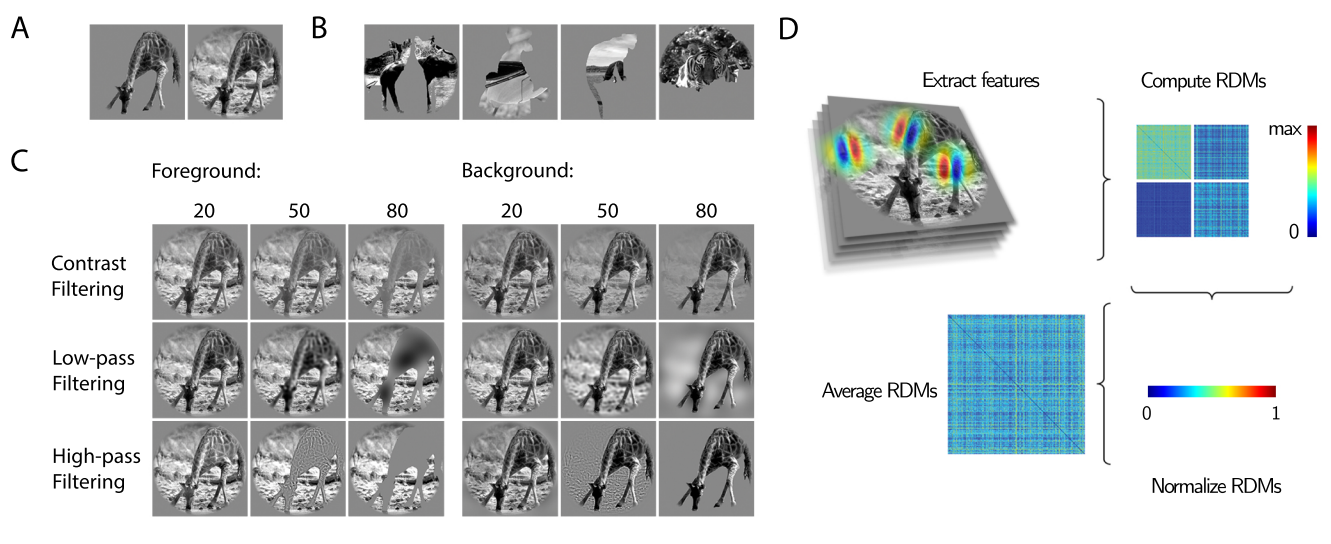
450 Huth AG, de Heer WA, Griffiths TL, Theunissen FE, Gallant JL (2016) Natural speech reveals the
451 semantic maps that tile human cerebral cortex. Nature 532:453-458.
452 Kastner S, De Weerd P, Ungerleider LG (2000) Texture segregation in the human visual cortex: A
453 functional MRI study. J Neurophysiol 83:2453-2457.
454 Kay KN, Weiner KS, Grill-Spector K (2015) Attention reduces spatial uncertainty in human ventral
455 temporal cortex. Curr Biol 25:595-600.
456 Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain
457 activity. Nature 452:352-355.
458 Khaligh-Razavi S-M (2014) What you need to know about the state-of-the-art computational
459 models of object-vision: A tour through the models. arXiv preprint arXiv:14072776.
460 Khaligh-Razavi SM, Kriegeskorte N (2014) Deep supervised, but not unsupervised, models may
461 explain IT cortical representation. PLoS computational biology 10:e1003915.
462 Kok P, de Lange FP (2014) Shape perception simultaneously up- and downregulates neural activity
463 in the primary visual cortex. Curr Biol 24:1531-1535.
464 Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA (2008)
465 Matching categorical object representations in inferior temporal cortex of man and
466 monkey. Neuron 60:1126-1141.
467 Kubilius J, Wagemans J, Op de Beeck HP (2014) A conceptual framework of computations in mid-
468 level vision. Front Comput Neurosci 8:158.
469 Lamme VA (1995) The neurophysiology of figure-ground segregation in primary visual cortex. J
470 Neurosci 15:1605-1615.
471 Lazebnik S, Schmid C, Ponce J (2006) Beyond bags of features: Spatial pyramid matching for
472 recognizing natural scene categories. In, pp 2169-2178: IEEE.

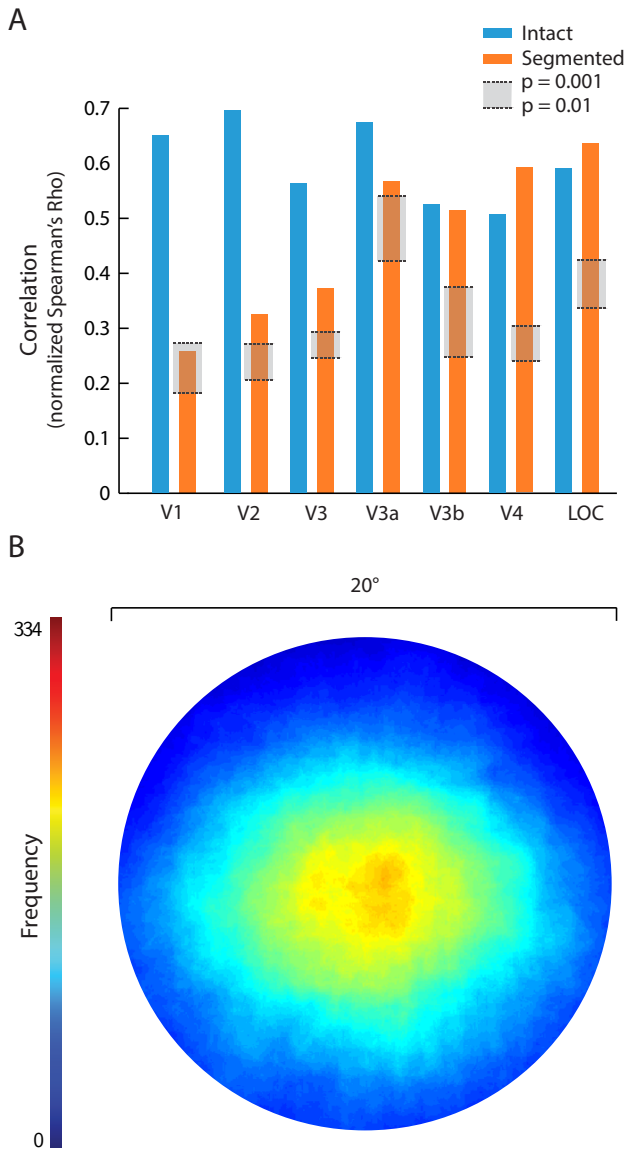
473 Lee TS, Mumford D, Romero R, Lamme VA (1998) The role of the primary visual cortex in higher
474 level vision. Vision Res 38:2429-2454.
475 Leek EC, d'Avossa G, Tainturier MJ, Roberts DJ, Yuen SL, Hu M, Rafal R (2012) Impaired integration
476 of object knowledge and visual input in a case of ventral simultanagnosia with bilateral
477 damage to area V4. Cogn Neuropsychol 29:569-583.
478 Lescroart M, Agrawal P, Gallant J (2016) Both convolutional neural networks and voxel-wise
479 encoding models of brain activity derived from ConvNets represent boundary-and surface-
480 related features. J Vis 16:756-756.
481 Lescroart MD, Biederman I (2013) Cortical representation of medial axis structure. Cereb Cortex
482 23:629-637.
483 Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR,
484 Tootell RB (1995) Object-related activity revealed by functional magnetic resonance
485 imaging in human occipital cortex. Proc Natl Acad Sci U S A 92:8135-8139.
486 Marr D (1982) Vision : a computational investigation into the human representation and
487 processing of visual information. San Francisco: W.H. Freeman.
488 Medathati NVK, Neumann H, Masson GS, Kornprobst P (2016) Bio-inspired computer vision:
489 Towards a synergistic approach of artificial and biological vision. Computer Vision and
490 Image Understanding 150:1-30.
491 Miyawaki Y, Uchida H, Yamashita O, Sato M-a, Morito Y, Tanabe HC, Sadato N, Kamitani Y (2008)
492 Visual image reconstruction from human brain activity using a combination of multiscale
493 local image decoders. Neuron 60:915-929.
494 Moore C, Engel SA (2001) Neural response to perception of volume in the lateral occipital complex.
495 Neuron 29:277-286.

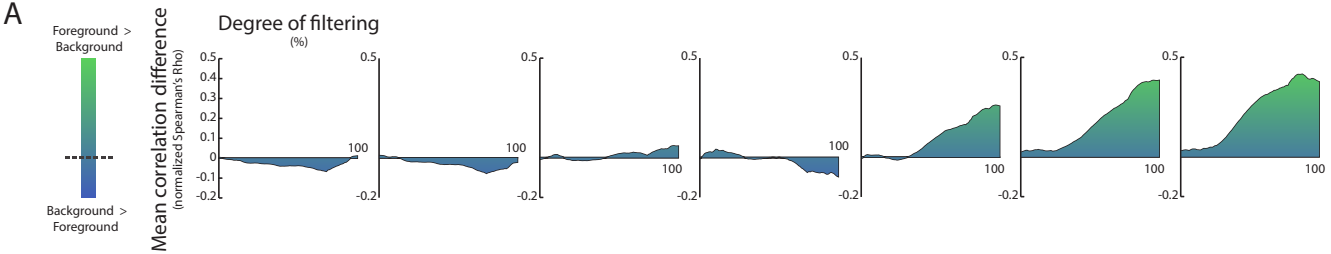
496 Motter BC (2009) Central V4 receptive fields are scaled by the V1 cortical magnification and
497 correspond to a constant-sized sampling of the V1 surface. J Neurosci 29:5749-5757.
498 Naselaris T, Kay KN, Nishimoto S, Gallant JL (2011) Encoding and decoding in fMRI. Neuroimage
499 56:400-410.
500 Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu B, Gallant JL (2011) Reconstructing visual
501 experiences from brain activity evoked by natural movies. Current Biology 21:1641-1646.
502 Ojala T, Pietikäinen M, Mäenpää T (2001) A generalized local binary pattern operator for
503 multiresolution gray scale and rotation invariant texture classification. In, pp 399-408:
504 Springer.
505 Oliva A, Torralba A (2001) Modeling the shape of the scene: A holistic representation of the spatial
506 envelope. International journal of computer vision 42:145-175.
507 Peterson MA (1994) Object Recognition Processes Can and Do Operate before Figure Ground
508 Organization. Current Directions in Psychological Science 3:105-111.
509 Peterson MA, Gibson BS (1994) Must Figure-Ground Organization Precede Object Recognition - an
510 Assumption in Peril. Psychological Science 5:253-259.
511 Poort J, Self MW, van Vugt B, Malkki H, Roelfsema PR (2016) Texture Segregation Causes Early
512 Figure Enhancement and Later Ground Suppression in Areas V1 and V4 of Visual Cortex.
513 Cereb Cortex 26:3964-3976.
514 Poort J, Raudies F, Wannig A, Lamme VA, Neumann H, Roelfsema PR (2012) The role of attention
515 in figure-ground segregation in areas V1 and V4 of the visual cortex. Neuron 75:143-156.
516 Qiu FT, Sugihara T, von der Heydt R (2007) Figure-ground mechanisms provide structure for
517 selective attention. Nat Neurosci 10:1492-1499.
518 Roe AW, Chelazzi L, Connor CE, Conway BR, Fujita I, Gallant JL, Lu H, Vanduffel W (2012) Toward a
519 unified theory of visual area V4. Neuron 74:12-29.

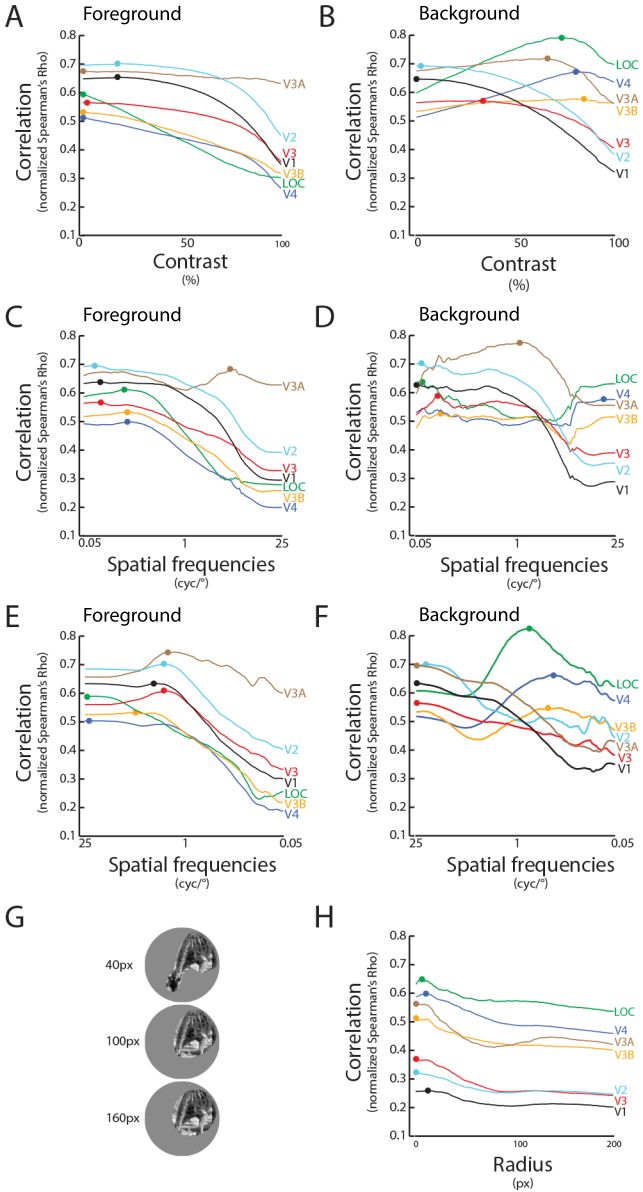
520 Scholte HS, Jolij J, Fahrenfort JJ, Lamme VA (2008) Feedforward and recurrent processing in scene
521 segmentation: electroencephalography and functional magnetic resonance imaging. J Cogn
522 Neurosci 20:2097-2109.
523 Self MW, van Kerkoerle T, Super H, Roelfsema PR (2013) Distinct roles of the cortical layers of area
524 V1 in figure-ground segregation. Curr Biol 23:2121-2129.
525 Stanley DA, Rubin N (2003) fMRI Activation in Response to Illusory Contours and Salient Regions in
526 the Human Lateral Occipital Complex. Neuron 37:323-331.
527 Stanley GB, Li FF, Dan Y (1999) Reconstruction of natural scenes from ensemble responses in the
528 lateral geniculate nucleus. J Neurosci 19:8036-8042.
529 Sterzer P, Kleinschmidt A, Rees G (2009) The neural bases of multistable perception. Trends Cogn
530 Sci 13:310-318.
531 Thirion B, Duchesnay E, Hubbard E, Dubois J, Poline JB, Lebihan D, Dehaene S (2006) Inverse
532 retinotopy: inferring the visual content of images from brain activation patterns.
533 Neuroimage 33:1104-1116.
534 Vernon RJ, Gouws AD, Lawrence SJ, Wade AR, Morland AB (2016) Multivariate Patterns in the
535 Human Object-Processing Pathway Reveal a Shift from Retinotopic to Shape Curvature
536 Representations in Lateral Occipital Areas, LO-1 and LO-2. J Neurosci 36:5763-5774.
537 Wandell BA, Winawer J (2011) Imaging retinotopic maps in the human brain. Vision Res 51:718-
538 737.
539 Williford JR, von der Heydt R (2016) Figure-Ground Organization in Visual Cortex for Natural
540 Scenes. eNeuro 3.
541 Wu MC, David SV, Gallant JL (2006) Complete functional characterization of sensory neurons by
542 system identification. Annu Rev Neurosci 29:477-505.
543

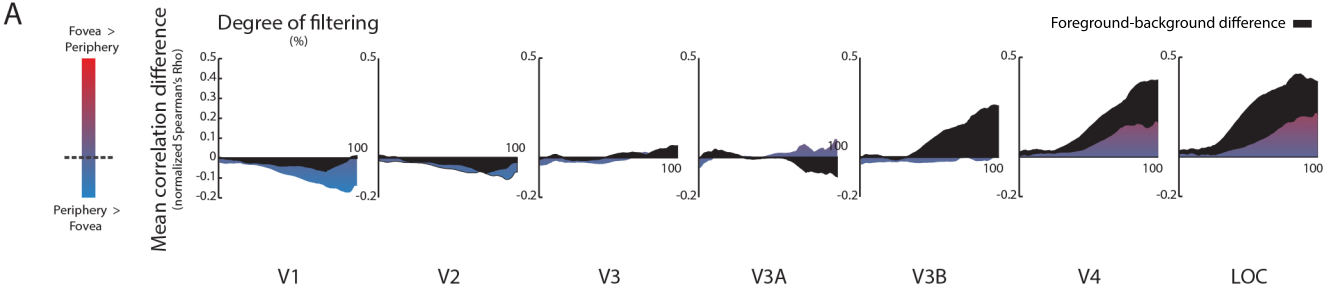

A
Stimuli
Models
Correlation with fMRI pattern
Result
0.36
0.18
0.58
...
Best model
B
Filtered Stimuli
Model
Correlation with fMRI pattern
Result
Fixed model
0.56
0.47
0.14
...
Neural image


A
B
C
Foreground:
Background:
20
50
80
20
50
80
Contrast Filtering
Low-pass Filtering
High-pass Filtering
D
Extract features
Compute RDMs
max
0
Average RDMs
0
1
Normalize RDMs


A
Intact
Segmented
p = 0.001
p = 0.01
Correlation
(normalized Spearman's Rho)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
V1
V2
V3
V3a
V3b
V4
LOC
B
20°
334
Frequency
0


A
Foreground > Background
Background > Foreground
Mean correlation difference
(normalized Spearman's Rho)
Degree of filtering
(%)
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
100
B
Stimulus
V1
V2
V3
V3A
V3B
V4
LOC


A
Foreground
Correlation
(normalized Spearman's Rho)
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
50
100
Contrast
(%)
V3A
V2
V3
V1
V3B
LOC
V4
B
Background
Correlation
(normalized Spearman's Rho)
0
50
100
Contrast
(%)
LOC
V4
V3A
V3B
V3
V2
V1
C
Foreground
Correlation
(normalized Spearman's Rho)
0.05
1
25
Spatial frequencies
(cyc/°)
V3A
V2
V3
V1
LOC
V3B
V4
D
Background
Correlation
(normalized Spearman's Rho)
0.05
1
25
Spatial frequencies
(cyc/°)
LOC
V4
V3A
V3B
V3
V2
V1
E
Foreground
Correlation
(normalized Spearman's Rho)
25
1
0.05
Spatial frequencies
(cyc/°)
V3A
V2
V3
V1
LOC
V3B
V4
F
Background
Correlation
(normalized Spearman's Rho)
25
1
0.05
Spatial frequencies
(cyc/°)
LOC
V4
V3B
V2
V3A
V3
V1
G
40px
100px
160px
H
Correlation
(normalized Spearman's Rho)
0
100
200
Radius
(px)
LOC
V4
V3A
V3B
V2
V3
V1


A
Fovea > Periphery
Periphery > Fovea
Mean correlation difference
(normalized Spearman's Rho)
Degree of filtering
(%)
Foreground-background difference
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
100
V1
V2
V3
V3A
V3B
V4
LOC