1. Introduction

This document contains the supplemental materials for [1]. Specifically, it contains the following figures:

- Fig. 1: Time courses with which clutter, contrast, and luminance representations emerged in the brain.
- Fig. 2: Results of cross-classification analysis plotted with 95% confidence intervals as determined by bootstrapping participants (colored shaded areas).
- Fig. 3: Scene clutter representations emerged with a distinct time course and were robust to changes in viewing conditions and other scene properties.
- Fig. 4: Persistent components in scene representations. The details of this figure are given in Section 1.1 below.
- Fig. 5: Layer-resolved prediction of emerging neural representations of single scene images by computational models.
- Fig. 6: Comparison of the deep scene network against other computational models in predicting neural representations of single scene images.
- Fig. 7: Complete layer-wise comparisons of scene representations in computational models of object and scene categorization.
- Fig. 8: Comparison of size effect in the deep scene network without and with partialling out the effect of other experimental factors (clutter, contrast and luminance).
- Fig. 9: Neural representations of scene clutter, contrast, and luminance explained by computational models.
- Movie: MEG decoding accuracy matrix and multidimensional scaling solution in two dimensions. The movie shows the MEG decoding accuracy matrices over time in 1 millisecond steps, accompanied with a two-dimensional scaling solution (criterion: metric stress). Experimental conditions (images) are plotted in the two dimensional space, color- and shape-coded to indicate scene size and clutter level.

Link: http://brainmodels.csail.mit.edu/scene-size

1.1. Explanation of Fig. 4

(A) While early visual processes are highly dynamic, changing rapidly one transient representation to another, emerging scene representations critical for behavioral guidance may be persistent, lasting hundreds of milliseconds, to ensure subsequent access [4, 2, 3]. We reasoned that if visual representations persist over time, MEG signals should be similar across time as well. To evaluate, we trained a SVM classifier to distinguish scene images from MEG brain responses at time $t_x$ and tested on data at other times $t_y$. Repeating this procedure for all pairwise combinations of conditions and all time points produced a 4D condition $\times$ condition $\times$ time $\times$ time matrix.

(B-F) Averaging over or comparing by subtraction different parts of the first two dimensions (left column, same as in Figure 1C (scene identity), Figure 2A (scene size) and Supplementary Figure 1) yielded 2D time-time decoding accuracy matrices (subject-averaged results in middle column; significant values in right column). Most cases exhibited strong transient components, as evidenced by higher decoding accuracies along the diagonal ($t_x \approx t_y$) than surrounding points in the time-time decoding matrices, with only exemption luminance. Persistent components had significant points outside the diagonal ($t_x \neq t_y$), and were evidenced for single scene image classification, scene size and scene clutter (B-D), but not for contrast or luminance (E, F). ($n = 15$, sign-permutation tests, cluster-definition threshold $P < 0.0005$ for B and $P < 0.05$ for C-F, cluster threshold $P < 0.05$).

References


Figure 1: Time courses with which clutter, contrast, and luminance representations emerged in the brain. The MEG decoding matrix was partitioned by (A) clutter, (B) contrast and (C) luminance level (left column), and the mean within-subdivision decoding accuracies (dark gray, −) were subtracted from between-subdivision decoding accuracies (light gray, +) to estimate time-resolved representations (right column). Horizontal lines indicate significant time points ($n = 15$, cluster-definition threshold $P < 0.05$, corrected significance level $P < 0.05$); gray shaded area indicates 95% confidence intervals determined by bootstrapping participants; gray vertical line indicates image onset.


Figure 2: **Results of cross-classification analysis plotted with 95% confidence intervals as determined by bootstrapping participants (colored shaded areas).** Horizontal lines indicate significant time points ($n = 15$, cluster-definition threshold $P < 0.05$, corrected significance level $P < 0.05$); gray vertical line indicates image onset.

Figure 3: **Scene clutter representations emerged with a distinct time course and were robust to changes in viewing conditions and other scene properties.** (A) Cross-classification analysis, exemplified for cross-classification of scene clutter across scene size. (B) Results of cross-classification analysis indicated tolerance of visual representations of scene clutter to changes in other scene and image properties (scene size, contrast and luminance). Horizontal lines indicate significant time points ($n = 15$, cluster-definition threshold $P < 0.05$, corrected significance level $P < 0.05$); colored shaded areas indicate 95% confidence intervals determined by bootstrapping participants; gray vertical line indicates image onset.
Figure 4: Persistent components in scene representations. Please see Section 1.1 for details.
Figure 5: **Layer-resolved prediction of emerging neural representations of single scene images by computational models.** (A) Deep scene network, (B) Deep object network, (C) GIST, (D) HMAX. Horizontal lines indicate significant time points ($n = 15$, cluster-definition threshold $P < 0.05$, corrected significance level $P < 0.05$ divided by model-specific layer number); gray vertical line indicates image onset.

Figure 6: **Comparison of the deep scene network against other computational models in predicting neural representations of single scene images.** To directly compare the time series in Figure 3D, we subtracted the time series of the HMAX, GIST, and deep object network from the time series of the deep scene network. While the deep scene and object networks did not differ significantly, the deep scene network predicted MEG better than HMAX and GIST. For details see Table 2b. Horizontal lines indicate significant time points ($n = 15$, cluster-definition threshold $P < 0.05$, corrected significance level $P < 0.05$); gray vertical line indicates image onset.
Figure 7: Complete layer-wise comparisons of scene representations in computational models of object and scene categorization. Comparisons for (A) scene size, (B) clutter level, (C) contrast level and (D) luminance level. Bars indicate correlation between the computational model RDMs and explicit models of size, clutter, contrast, or luminance (e.g. the explicit size model in Figure 4E inset: an RDM with entries 0 for images of similar size, 1 for images of dissimilar size). While abstract scene properties emerged with increasing layer size in the deep scene and object networks progressively with increasing layer number, the low-level image properties contrast and luminance were progressively abstracted away. Lines above bars indicate significant differences across layers. Comparisons across equivalent layers of the deep scene and object networks are colored red ($n = 48$, label permutation tests for statistical inference, $P < 0.05$, FDR-corrected for multiple comparisons for all layer-wise comparisons).
Figure 8: Comparison of size effect in the deep scene network without and with partialling out the effect of other experimental factors (clutter, contrast and luminance). To control whether the size effect observed in the deep scene network was explained by correlation with other experimental factors, we calculated the correlation between layer-specific RDMs and the size model before and after partialling out the model RDMs for clutter, contrast and luminance. We found that the size effect persisted in the model RDM in the partial correlation analysis, indicating that the size effect is not explained by clutter. Stars above bars indicate statistical significance ($n = 48$; label permutation tests for statistical inference, $P < 0.05$, FDR-corrected for multiple comparisons).
Figure 9: Neural representations of scene clutter, contrast, and luminance explained by computational models. (A) MEG representations of scene clutter, (termed MEG clutter signal) before (black) and after (color-coded by model) partialling out the effect of different computational models. (B-C) same as (A) for contrast and luminance. Only partialling out the deep scene network abolished experimental effects for all factors ($n = 15$; cluster-definition threshold $P < 0.05$, significance threshold $P < 0.05$ corrected for multiple comparisons by 5 for all panels).