Designing salient, naturalistic "super-stimuli" with deep generative models

Shankhadeep Mukherjee (shankhadeepm@iisc.ac.in)
IMI Ph.D. program and Centre for Neuroscience,
Indian Institute of Science, Bangalore-560012, Karnataka, India

Achin Parashar (achinp@iisc.ac.in)
Centre for Neuroscience,
Indian Institute of Science, Bangalore-560012, Karnataka, India

Chandra R Murthy (cmurthy@iisc.ac.in)
Department of Electrical Communication Engineering,
Indian Institute of Science, Bangalore-560012, Karnataka, India

Devarajan Sridharan (sridhar@iisc.ac.in)
Centre for Neuroscience and Computer Science and Automation,
Indian Institute of Science, Bangalore-560012, Karnataka, India

Abstract

Attention can be deployed voluntarily, "top-down", based on task goals, or captured automatically, "bottom-up", by salient stimuli. Most previous studies have controlled stimulus salience by altering low-level target features, for example, by increasing luminance, or inducing popouts, or by creating motion dynamics. Can we design static, naturalistic images that capture bottom-up attention robustly? We address this question in three stages. First, we hypothesize that such tailor-made "super-stimuli" would evoke strong responses in visual cortical areas. We advance a deep generative framework with a heuristic optimization algorithm (XDream) to design category-specific, naturalistic images that can combinatorially activate and suppress multiple regions across human visual cortex. Second, with human functional MRI recordings, we show that such super-stimuli differentially activate visual cortical regions targeted by the optimization algorithm, thereby validating the approach. Third, we show that, in a working memory task, super-stimuli optimized for specific regions are more accurately recalled than control stimuli, thereby demonstrating their behavioral salience. Our behaviorally validated super-stimuli open up new avenues of research for investigating neural mechanisms of exogenous attention control with salient, naturalistic images.

Keywords: fMRI, attention, visual, deep networks, super stimuli, saliency, working memory

Introduction

Attention can be engaged in multiple forms – goal-directed (top-down or endogenous) and stimulus-driven (bottom-up or exogenous). While task protocols for engaging endogenous attention have been extensively studied (Corbetta & Shulman, 2002), comparatively little is known about the nature of salient stimuli that engage exogenous attention (Itti & Koch, 2000). Recent work suggests that exogenous attention is directly mediated by activity in visual cortex (Fernández & Carrasco, 2020). Here, we hypothesize that stimuli that evoke strong responses in visual cortical regions would also naturally engage exogenous attention.

The human visual system processes information hierarchically, with higher brain areas progressively encoding more complex features (Hubel & Wiesel, 1959; Kanwisher, McDermott, & Chun, 1997; Pasupathy & Connor, 1999). Yet, which combinations of natural image features drive the strongest response in each visual area remains unknown. The high dimensionality of natural images renders this optimization prohibitively challenging at the pixel level. Furthermore, the large number of potential feature combinations in natural images renders trial-and-error approaches infeasible.

To address this challenge, recent studies have employed deep learning models, particularly deep generative networks (DGNs), to generate images that activate specific brain regions (Bashivan, Kar, & DiCarlo, 2019; Gu et al., 2022; Ponce et al., 2019; Walker et al., 2019). Specifically, we leverage XDream (Ponce et al., 2019; Xiao & Kreiman, 2020) – a recent framework that employs a deep generative network (DGN) (Brock, Donahue, & Simonyan, 2018) in combination with a heuristic optimization (genetic) algorithm. We advance the XDream algorithm to design salient "super-stimuli" — high-resolution, naturalistic images tailor-made to evoke the strongest responses in specific brain areas – and evaluate these as candidate images for engaging exogenous attention.

Our study makes the following key contributions: (a) We advance XDream on two fronts: (i) by generating super-stimuli constrained by specific object categories and (ii) by designing "chimeric" super-stimuli that can combinatorially activate or suppress multiple brain regions at once. (b) We show that synthesized super-stimuli activate the specific region targeted by the optimization algorithm, by directly recording participants' visual cortical responses with functional MRI. (c) We validate the behavioral salience of these super-stimuli by showing that they are more accurately recalled than control stimuli in a 2-back working memory (WM) task.

Methods

Dataset and Encoder We utilized the Algonauts Challenge subset (Gifford et al., 2023) of the Natural Scenes Dataset (NSD) (Allen et al., 2022), which includes beta maps of multiple visual cortex region activations measured with fMRI. Each of the 8 participants was shown 10,000 images in a 7T MRI scanner. We selected data from two participants, #2 and #5 out of the four who completed all the trials. A key component of our image generation algorithm is an fMRI-"Encoder": a neural network trained with natural images, and their corresponding fMRI activations to predict fMRI activations to novel images. Here, we trained a separate encoder (Gaziv et al., 2022) for each region and participant using the dataset.

Generating super-stimuli Our model comprises two major components (Fig 1A). The first component, an unconstrained optimizer based on the XDream framework (purple shading), itself comprised of 3 modules: i) a deep generative network (BigGAN-deep (Brock et al., 2018)) that synthesizes images based on vector "image codes", ii) an Encoder (Gaziv et al., 2022) that predicts regional fMRI brain responses to the synthesized images, and iii) a "genetic" optimization algorithm (GA) that iteratively refines image codes to enhance fMRI activity over generations stochastically. The second component comprises a "constrained optimizer" (yellow shading) that seeks to enhance category-relevant features in the generated images.

The constrained optimizer utilizes the output probabilities from the classifier (Tu et al., 2022). A minimum threshold of 0.8 is applied to the classifier scores of the images belonging to their respective classes. If an image's class score falls below this threshold, the corresponding codes are updated using the classifier loss backpropagated through the DGN. This process modifies the images to increase region-specific activity and maintain their class membership simultaneously. Control

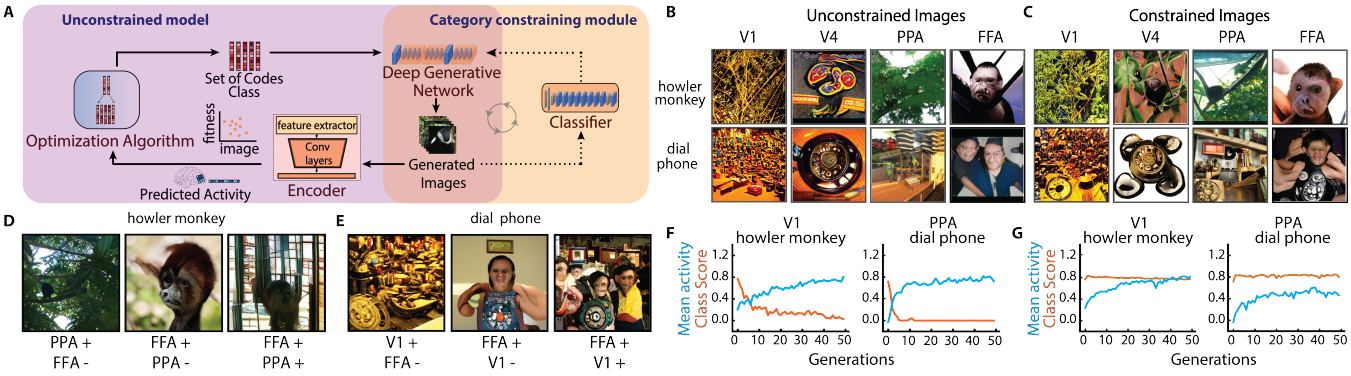

Figure 1: Generating naturalistic "super-stimuli". A) The proposed model (see text). (B) and (C) Images generated with the unconstrained model and constrained optimizer respectively. Rows: Categories; Columns: Visual regions. V1, V4: primary and object-selective visual areas; FFA- fusiform face area; PPA- parahippocampal place area. (F-G) Predicted fMRI activity (cyan) and class score (orange) as a function of generations for V1 (left) PPA (right) in the unconstrained (F) & constrained (G) models. (D-E) Chimeric images generated by jointly optimizing two regions. Conventions: +: enhanced, -: suppressed.

stimuli were generated using the same initial seeds but by randomly shuffling the image scores obtained from the encoder in each generation of the genetic algorithm. To modify images jointly for two regions, we combined image scores from two separate encoders, one for each region of interest, with a weighted softmax function.

Behavioral and fMRI evaluation To assess the behavioral salience of these super-stimuli, we tested participants (n=10) on a 2-back working memory task (Kirchner, 1958) (Fig. 2B). Participants pressed a button when the current image matched the image presented two frames earlier and withheld responses otherwise. We selected images (both super-stimuli and controls) optimized for V1v and hV4 regions based on previous research, which suggests the involvement of these regions in exogenous attention (Burrows & Moore, 2009; Chen, Zhang, Wang, Zhou, & Fang, 2016). These same images were presented to participants in a functional MRI scanner (Siemens Prisma) to identify visual cortical regions they activate.

Results

Stimuli Combinatorially Optimized for Visual Cortex Images optimized for specific brain regions contained features consistent with prior knowledge about these regions. V1v and hV4 optimized images contained textures or primitive objects, respectively (Fig.1B, 1st, 2nd columns). By contrast, FFA and PPA optimized images contained face-like or scene-based features, respectively (Fig.1B, 3rd, 4th columns). Yet, with the unconstrained model, category-specific information (class score) reduced steeply over generations (Fig. 1F, orange).

Addition of the constrained optimizer yielded images that clearly carried more category-specific features (Fig. 1C). For example, conditional generation with the "telephone" class optimized for FFA yielded the image of a person's face superimposed over a rotary phone dial (Fig. 1C). Moreover, class scores remained high over generations (Fig 1G, orange).

Optimizing for two brain regions jointly yielded "chimeric" stimuli (Fig. 1D & E). Notably, suppressing the activity of one region while enhancing the other diminished the features associated with former. For example, enhancing FFA while suppressing PPA produced "face" features but blurred out the background, rendering "place" features challenging to identify (Fig. 1D, FFA+/PPA-).

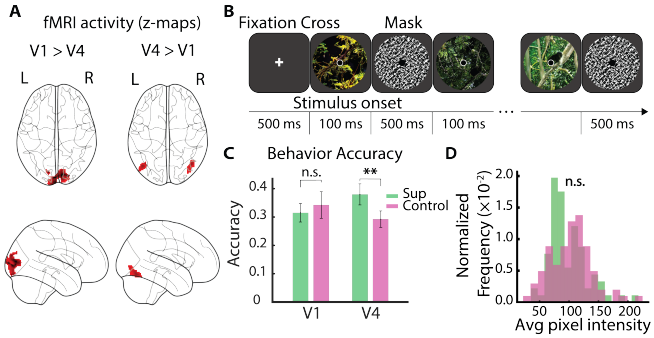

Figure 2: (A) Thresholded z-map of fMRI activity (n=5 participants, $z>2.5$) showing regions activated more by V1v than hV4 optimized images (left) and vice versa (right). (B) A 2-back task design (see text). (C) Accuracy in the 2-back task for V1v-optimized (left) and hV4-optimized (right) super-stimuli (green) and control stimuli (purple). (D) Average pixel intensities of the super-stimuli and control stimuli.

Super-stimuli Activate Visual Cortex Selectively Five participants underwent fMRI scans while viewing images optimized for the V1v and hV4 regions. Thresholded brain activity maps (Fig. 2A, $p<0.01$, FPR) revealed greater activation near the primary visual cortex (medial cuneus and lingual gyrus) for V1v-optimized stimuli. Conversely, hV4-optimized stimuli elicited stronger activations in ventrolateral visual cortex consistent with the locus of the LOC (lateral occipital cortex).

Super-stimuli are Behaviorally Salient 2-back recall accuracies were higher for the hV4-optimized super-stimuli than control stimuli ($p=0.003$; permutation test, Fig. 2C, right) indicating higher salience; a similar difference was not evident for V1v-optimized stimuli ($p=0.791$, Fig. 2C, left). Average pixel intensities did not differ between the super-stimuli (Fig. 2D, green) and control stimuli (Fig. 2D, purple) ($p = 0.11$, Wilcoxon test), indicating that the results could not be attributed to overall intensity differences between the image groups.

Acknowledgments

This work was supported by the following funding sources: a Wellcome Trust-Department of Biotechnology India Alliance Intermediate fellowship, DST Swarna Jayanti fellowship, a Pratiksha Trust Intramural grant, an India-Trento Programme for Advanced Research (ITPAR) grant and a Department of Biotechnology-Indian Institute of Science Partnership Program grant (all to D.S.).

References

Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Prince, J. S., Dowdle, L. T., ... Kay, K. (2022, 1). A massive 7t fmri dataset to bridge cognitive neuroscience and artificial intelligence. *Nature Neuroscience*, *25*, 116-126. doi: 10.1038/s41593-021-00962-x

Bashivan, P., Kar, K., & DiCarlo, J. J. (2019). Neural population control via deep image synthesis. *Science (New York, N.Y.)*, *364*. doi: 10.1126/science.aav9436

Brock, A., Donahue, J., & Simonyan, K. (2018). Large scale gan training for high fidelity natural image synthesis. doi: 10.48550/ARXIV.1809.11096

Burrows, B. E., & Moore, T. (2009, 12). Influence and limitations of popout in the selection of salient visual stimuli by area v4 neurons. *The Journal of Neuroscience*, *29*, 15169-15177. doi: 10.1523/JNEUROSCI.3710-09.2009

Chen, C., Zhang, X., Wang, Y., Zhou, T., & Fang, F. (2016, 6). Neural activities in v1 create the bottom-up saliency map of natural scenes. *Experimental Brain Research*, *234*, 1769-1780. doi: 10.1007/s00221-016-4583-y

Corbetta, M., & Shulman, G. L. (2002, 3). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*, 201-215. doi: 10.1038/nrn755

Fernández, A., & Carrasco, M. (2020, 10). Extinguishing exogenous attention via transcranial magnetic stimulation. *Current Biology*, *30*, 4078-4084.e3. doi: 10.1016/j.cub.2020.07.068

Fine, M. S., & Minnery, B. S. (2009, 6). Visual salience affects performance in a working memory task. *Journal of Neuroscience*, *29*, 8016-8021. doi: 10.1523/JNEUROSCI.5503-08.2009

Gaziv, G., Beliy, R., Granot, N., Hoogi, A., Strappini, F., Golan, T., & Irani, M. (2022, 7). Self-supervised natural image reconstruction and large-scale semantic classification from brain activity. *NeuroImage*, *254*, 119121. doi: 10.1016/j.neuroimage.2022.119121

Gifford, A. T., Lahner, B., Saba-Sadiya, S., Vilas, M. G., Lascelles, A., Oliva, A., ... Cichy, R. M. (2023, 7). *The algonauts project 2023 challenge: How the human brain makes sense of natural scenes.* arXiv. Retrieved from http://arxiv.org/abs/2301.03198 (arXiv:2301.03198 [cs, q-bio]) doi: 10.48550/arXiv.2301.03198

Gu, Z., Jamison, K. W., Khosla, M., Allen, E. J., Wu, Y., St-Yves, G., ... Kuceyeski, A. (2022, 2). Neurogen: Activation optimized image synthesis for discovery neuroscience. *NeuroImage*, *247*, 118812. doi: 10.1016/j.neuroimage.2021.118812

Hubel, D. H., & Wiesel, T. N. (1959, 10). Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, *148*, 574-591. doi: 10.1113/jphysiol.1959.sp006308

Itti, L., & Koch, C. (2000, 6). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489-1506. doi: 10.1016/S0042-6989(99)00163-7

Kanwisher, N., McDermott, J., & Chun, M. M. (1997, 6). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, *17*, 4302-4311. doi: 10.1523/JNEUROSCI.17-11-04302.1997

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, *55*, 352-358. doi: 10.1037/h0043688

Pasupathy, A., & Connor, C. E. (1999, 11). Responses to contour features in macaque area v4. *Journal of Neurophysiology*, *82*, 2490-2502. doi: 10.1152/jn.1999.82.5.2490

Ponce, C. R., Xiao, W., Schade, P. F., Hartmann, T. S., Kreiman, G., & Livingstone, M. S. (2019, 5). Evolving images for visual neurons using a deep generative network reveals coding principles and neuronal preferences. *Cell*, *177*, 999–1009.e10. doi: 10.1016/j.cell.2019.04.005

Tu, Z., Talebi, H., Zhang, H., Yang, F., Milanfar, P., Bovik, A., & Li, Y. (2022). Maxvit: Multi-axis vision transformer. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, & T. Hassner (Eds.), (p. 459-479). Springer Nature Switzerland. doi: 10.1007/978-3-031-20053-3_27

Walker, E. Y., Sinz, F. H., Cobos, E., Muhammad, T., Froudarakis, E., Fahey, P. G., ... Tolias, A. S. (2019, 12). Inception loops discover what excites neurons most using deep predictive models. *Nature Neuroscience*, *22*, 2060-2065. doi: 10.1038/s41593-019-0517-x

Xiao, W., & Kreiman, G. (2020, 6). Xdream: Finding preferred stimuli for visual neurons using generative networks and gradient-free optimization. *PLOS Computational Biology*, *16*, e1007973. doi: 10.1371/journal.pcbi.1007973