A generative diffusion model reveals V2's representation of natural images

Neel Agarwal[1], Gabriel Yancy[1], Zahra Kadkhodaie[2], Justin D Lieber[1], J Anthony Movshon[1], Eero P Simoncelli[1,2]

[1]Center for Neural Science, New York University, [2]Center for Computational Neuroscience, Flatiron Institute

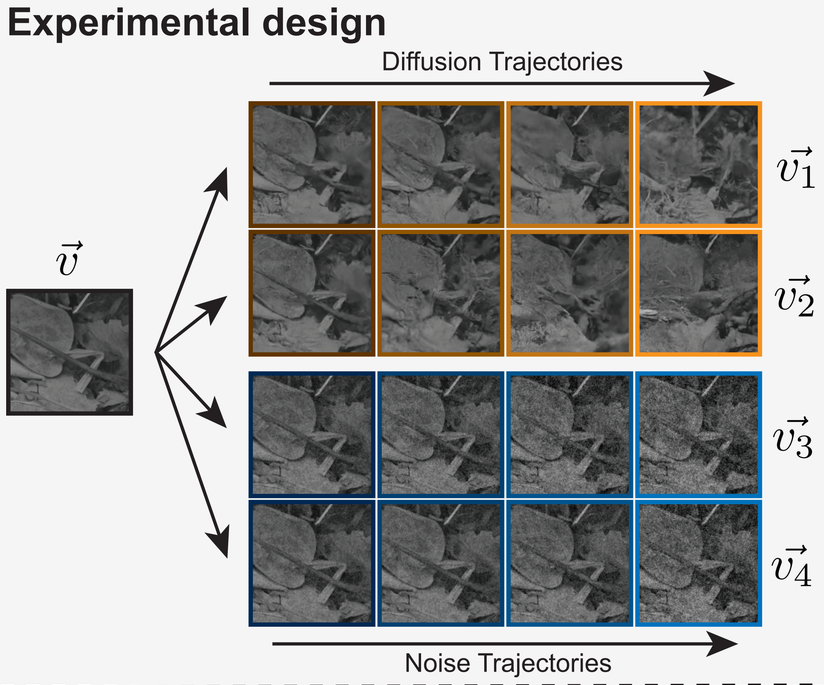
Motivation

Vision science relies on stimuli to probe and understand neural responses. Parametric stimuli such as gratings, noise, and synthetic textures are highly controlled but capture only limited aspects of natural vision. In contrast, photographic images contain the full structure of real scenes but are difficult to manipulate systematically because they are highly complex and too numerous for tractability. As a result, there is a gap in stimulus generation between simple parametric stimuli and natural images, limiting our ability to study mid-level visual processing, which is thought to encode intermediate visual structure. Recently, generative diffusion models have emerged as a powerful framework for synthesizing realistic images by iteratively denoising, effectively capturing the statistical structure of natural images while allowing controlled sampling. Here, we use a diffusion model to generate visual stimuli that we use to examine population response properties of macaque V2.

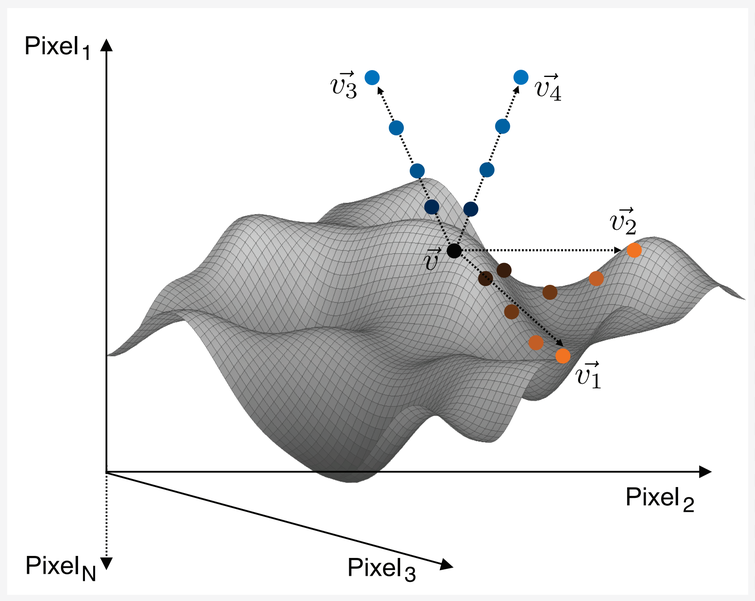
Neural data

We collected the responses of single-units and multi-units using Neuropixels probes in anaesthetised macaque V2. Stimuli were presented for 100 ms separated by 100 ms periods of gray in a 4 deg circulate vignette that covered the shared population receptive field.

Experimental design



$$\|\vec{v}-\vec{v_1}\| \approx \|\vec{v}-\vec{v_2}\| \approx \|\vec{v}-\vec{v_3}\| \approx \|\vec{v}-\vec{v_4}\|$$

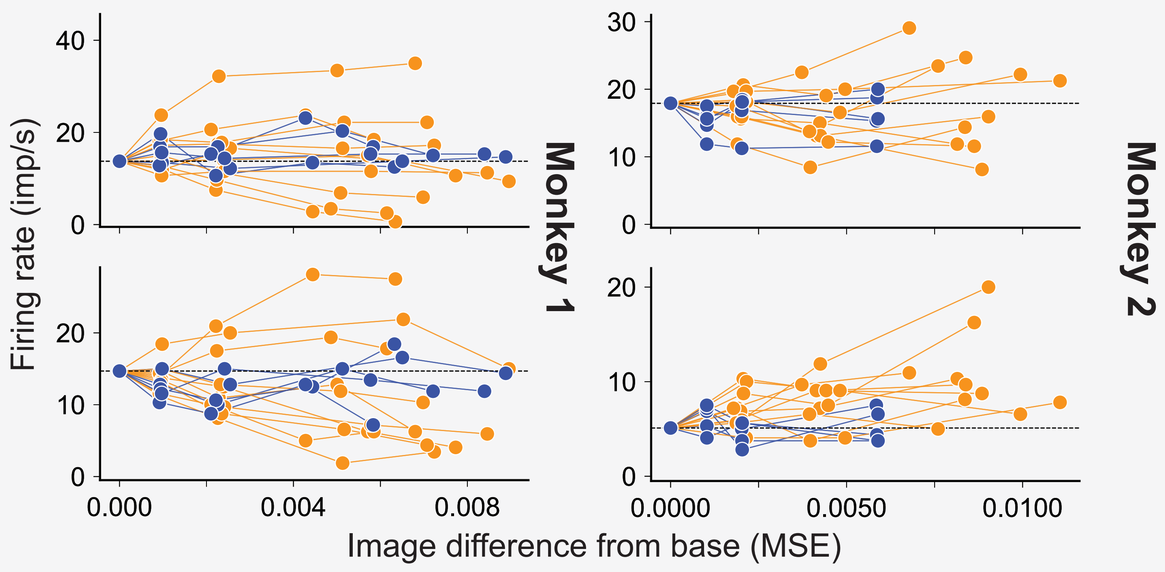


Right: Four single unit responses to off- and on-manifold image trajectories. Each plot shows responses to stimuli derived from a single base image v. Each line corresponds to a particular diffusion (orange) or white noise (blue) trajectory, starting from the base image v.

Base
Diffusion
White Noise
Baseline

Below: Alignment of population response vectors for diffusion images (orange) and noise images (blue) with those of the corresponding base image.

Single unit activity



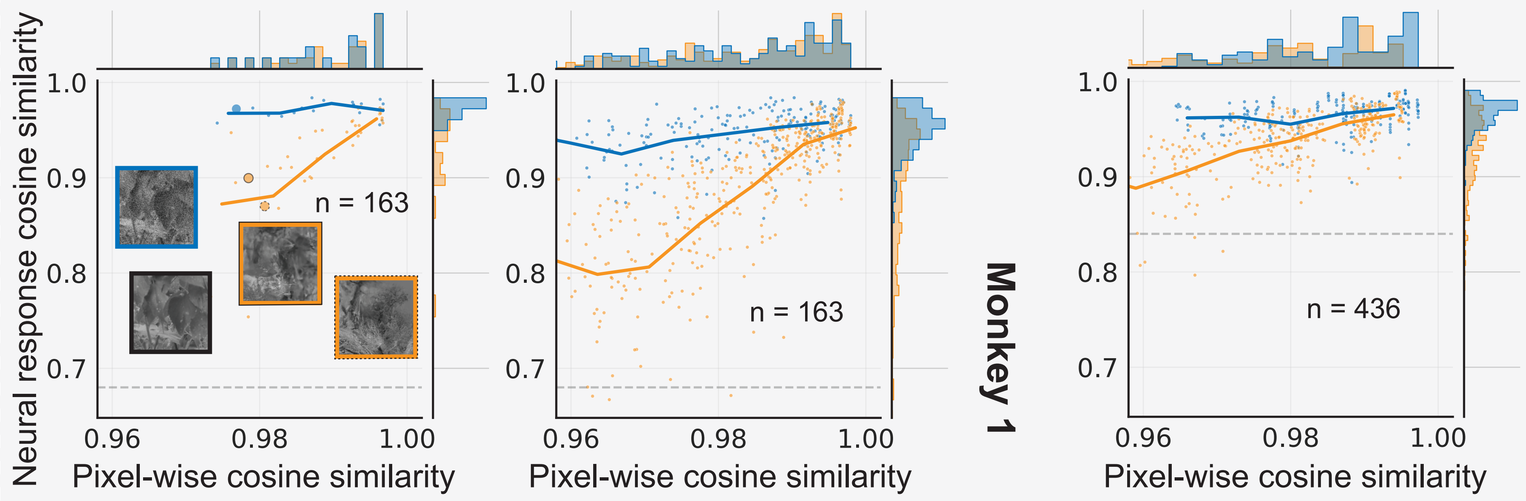
Population activity

Response pattern similarity

Single base image family

All base image families

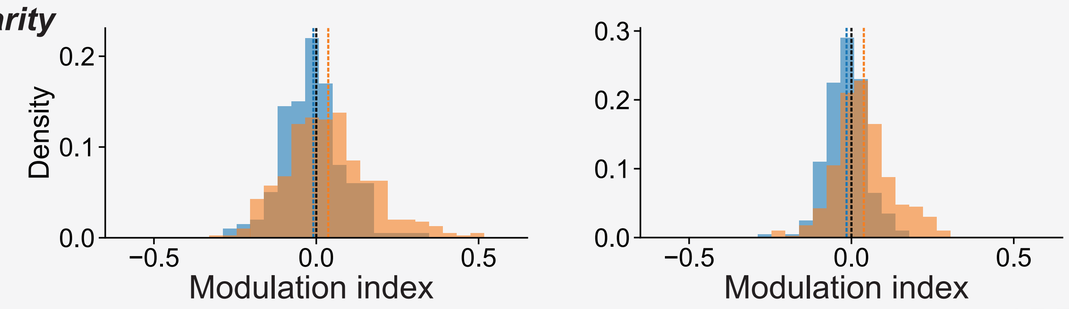
All base image families



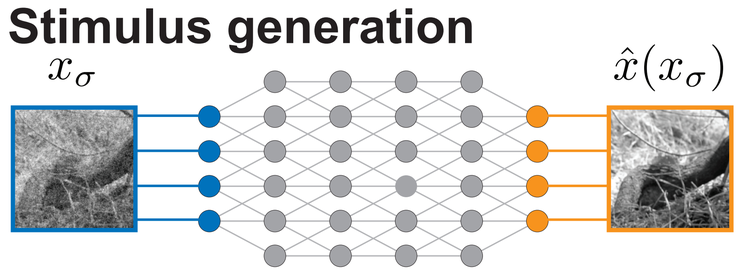
Response magnitude similarity

Right: Modulation index is calculated as:

$$\frac{\|\vec{v_i}\| - \|\vec{v_{base}}\|}{\|\vec{v_i}\| + \|\vec{v_{base}}\|}$$



Stimulus generation



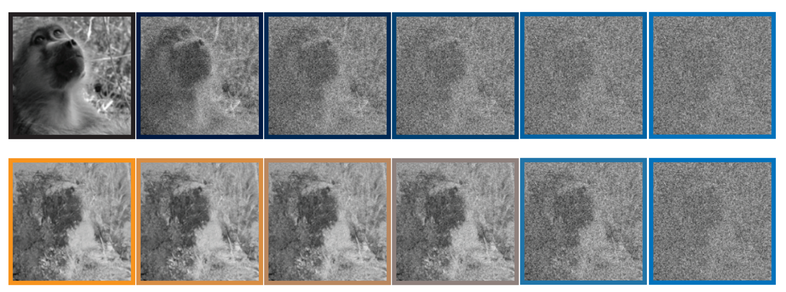
- A deep neural network trained to denoise images provides access to the gradient of the log probability ("Score") function describing the data distribution.

$$\hat{x}(x_\sigma) = \mathbb{E}[x \mid x_\sigma] = x_\sigma + \sigma^2 \nabla_{x_\sigma} \log p(x_\sigma)$$

(Miyasawa 1961)

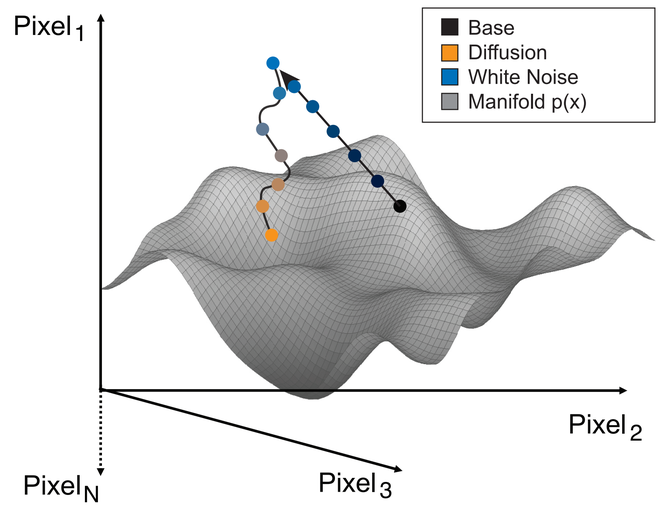
- For base image x: $x_\sigma = x + (\sigma * w)$, $w \sim \mathcal{N}(0,1)$
- Network: 4-stage U-Net (Ronneberger et al. 2015).
- Trained on: Images from the "birthplace of the human eye" (Tkačik et al 2011)
- This network can be used in an iterative ascent algorithm to draw samples from this distribution (Kadkhodaie & Simoncelli, ICLR2021).
- When initialized from a noise-corrupted version of a base image, it generates "naturalistic" images near the base image.

Forward diffusion (additive noise)



Reverse diffusion (denoising / sampling)

These models implicitly learn to represent the image manifold p(x)



Conclusions

On-manifold image variations produce more diverse changes in neural activity than distance-matched off-manifold variations, while only modestly changing the population response magnitude. As a result, as image difference increases, cosine similarity to the base image response declines more for on-manifold than off-manifold trajectories. These results suggest that V2 population responses are more sensitive to variations in the content of natural images, as defined by the diffusion model, and are therefore adapted to report the image content of natural scenes. Future work will explore how features of the naturalistic manifold are represented by the V2 population.

Contact

na3524@nyu.edu, gmy225@nyu.edu

Funding acknowledgements

Simons Foundation, NIH EY022428