Functional organization and natural scene responses across mouse visual cortical areas revealed with encoding manifolds

Luciano Dyballa[1,2], Greg D. Field[3], Michael P. Stryker[4,5], and Steven W. Zucker[2,6]

[1]School of Science and Technology, IE University, Madrid, Spain
[2]Department of Computer Science, Yale University, New Haven, USA
[3]Jules Stein Eye Institute, Department of Ophthalmology, David Geffen School of Medicine, University of California, Los Angeles, CA, USA
[4]Department of Physiology, University of California, San Francisco, CA, USA
[5]Kavli Institute for Fundamental Neuroscience, University of California, San Francisco, CA, USA
[6]Department of Biomedical Engineering, Yale University, New Haven, CT, USA

Abstract

A challenge in sensory neuroscience is understanding how populations of neurons operate in concert to represent diverse stimuli. To meet this challenge, we have created "encoding manifolds" that reveal the overall responses of brain areas to diverse stimuli with the resolution of individual neurons and their response dynamics. Here we use encoding manifold to compare the population-level encoding of primary visual cortex (VISp) with five higher visual areas (VISam, VISal, VISpm, VISlm, and VISrl). We used data from the Allen Institute Visual Coding–Neuropixels dataset from the mouse. We show that the encoding manifold topology computed only from responses to grating stimuli is continuous, for V1 and for higher visual areas, with smooth coordinates spanning it that include orientation selectivity and firing-rate magnitude. Surprisingly, the manifolds for each visual area revealed novel relationships between how natural scenes are encoded relative to static gratings—a relationship that was conserved across visual areas. Namely, neurons preferring natural scenes preferred either low or high spatial frequency gratings, but not intermediate ones. Analyzing responses by cortical layer reveals a preference for gratings concentrated in layer 6, whereas preferences for natural scenes tended to be higher in layers 2/3 and 4. The results reveal how machine learning approaches can be used to organize and visualize the structure of sensory coding, thereby revealing novel relationships within and across brain areas and sensory stimuli.

Significance Statement

Manifolds have become a commonplace for analyzing and visualizing neural responses. However, prior work has focused on building manifolds that organize diverse stimuli in neural response coordinates. Here, we demonstrate the utility of an alternative approach: building manifolds to represent neurons in stimulus/response coordinates, which we term 'encoding manifolds'. This approach has several advantages, such as being able to directly visualize and compare how different brain areas encode diverse stimulus ensembles. We use the approach to reveal novel relationships between layer-specific responses and the encoding of natural versus artificial stimuli.

Introduction

Understanding how populations of neurons are organized to represent external stimuli is a core goal of sensory neuroscience. This is made challenging by the fact that both artificial and sensory stimuli are typically high dimensional as are the patterns of spiking activity that these stimuli produce across populations of neurons. This presents a challenge to visualizing, let alone understanding, how complex stimuli are represented by neural population activity. To overcome this challenge, we have recently developed *encoding manifolds* to organize neurons according to their responses to an ensemble of visual stimuli [23]; these manifolds allow visualizing how neural populations encode ensembles of stimuli. In the encoding manifold, each point is a neuron, and nearby neurons respond similarly in time to members of the stimulus ensemble. The encoding manifold differs from more standard applications of manifolds in neuroscience, in which each point on the manifold is a stimulus (or a point along a trajectory) embedded in neural response coordinates (see, e.g., [14] for a review). This distinction is relevant as the latter approach emphasizes reading out the stimulus (or movement trajectory) from the population response (i.e., decoding), while the former organizes neurons according to their functional properties (i.e., encoding), which can drive hypotheses about cell types and their anatomical and functional connectivity in a circuit.

One feature of encoding manifolds exploited in prior work [23] is they allow for comparing the topology of sensory encoding at different stages of sensory processing. When comparing the encoding manifolds between retina and V1 in mouse to an ensemble of artificial stimuli (drifting gratings) and more naturalistic stimuli (flows), they were strikingly different: the encoding manifold inferred from retinal responses consisted of largely separate clusters of neurons that corresponded to known retinal ganglion cell types. This result served to lend credence to the encoding manifold as a useful approach for revealing underlying biological structure because the stimuli used to generate the manifold were distinct from those used to classify the retinal ganglion cell types. The V1 manifold derived from responses to the same stimuli was much more continuous than the retinal manifold, with neurons effectively carpeting stimulus/response space. Like the retinal manifold, the structure of the V1 manifold was quite informative about the physiological organization of V1. While no information about laminar position or putative excitatory versus putative inhibitory neurons (inferred from spike waveforms) was used in its construction, these properties were not randomly located across the manifold; rather, functional cell types were distributed in a complementary fashion across different layers and in different regions.

This prior work raised several important questions: to what extent does the structure of the V1 encoding manifold depend on the stimulus set? To what extent are the encoding manifolds of higher visual areas similar to that observed in V1? Finally, can the encoding manifold reveal aspects of how natural scenes are encoded by visual cortex? Fortunately, these three questions could be answered by utilizing the Allen Institute's large dataset of visual cortical responses of V1 and higher cortical visual areas to a variety of stimuli [57, 18, 19]. The stimulus set includes stationary gratings of a wide range of orientations, spatial frequencies, and spatial phases, a set of a similar number of stationary natural images, and an ensemble of moving gratings across a range of spatial frequencies and orientations (Fig. 1a,b). Using these data, we have constructed encoding manifolds for each of the higher visual areas using only the responses to static and drifting gratings. The resulting V1 encoding manifold was similar to that in our earlier work despite substantive differences in the stimulus sets. Namely, both the present and previous manifolds were continuous, with foci for different cortical layers and excitatory versus inhibitory cell types. The higher visual areas also exhibited similar manifold structures. One major similarity among the manifolds generated from each visual area was a preference for low spatial frequency gratings in layer 6, a layer that projects back to sub-cortical structures and elsewhere [50]; neurons in all other layers, across all visual areas, tended to prefer natural scenes.

Further analysis of the natural scene responses using the encoding manifold revealed a particularly interesting geometry. The resulting manifolds had a coordinate that ran from neurons with strong to weak

orientation and stimulus selectivity. However, those neurons that responded better to natural images relative to gratings formed a band in the manifold orthogonal to the orientation axis in VISp, which suggests that there exist natural-scene preferring neurons over the full range of selectivity for stimulus orientation. Furthermore, these neurons preferred either high or low rather than the intermediate spatial frequencies, which were preferred by the neurons driven best by gratings. This structure was also present in each of the higher visual areas, although the axes differed from VISp. These results further validate the use of encoding manifolds for assaying the structure and topology of sensory coding across distinct stimulus ensembles and stages of sensory processing.

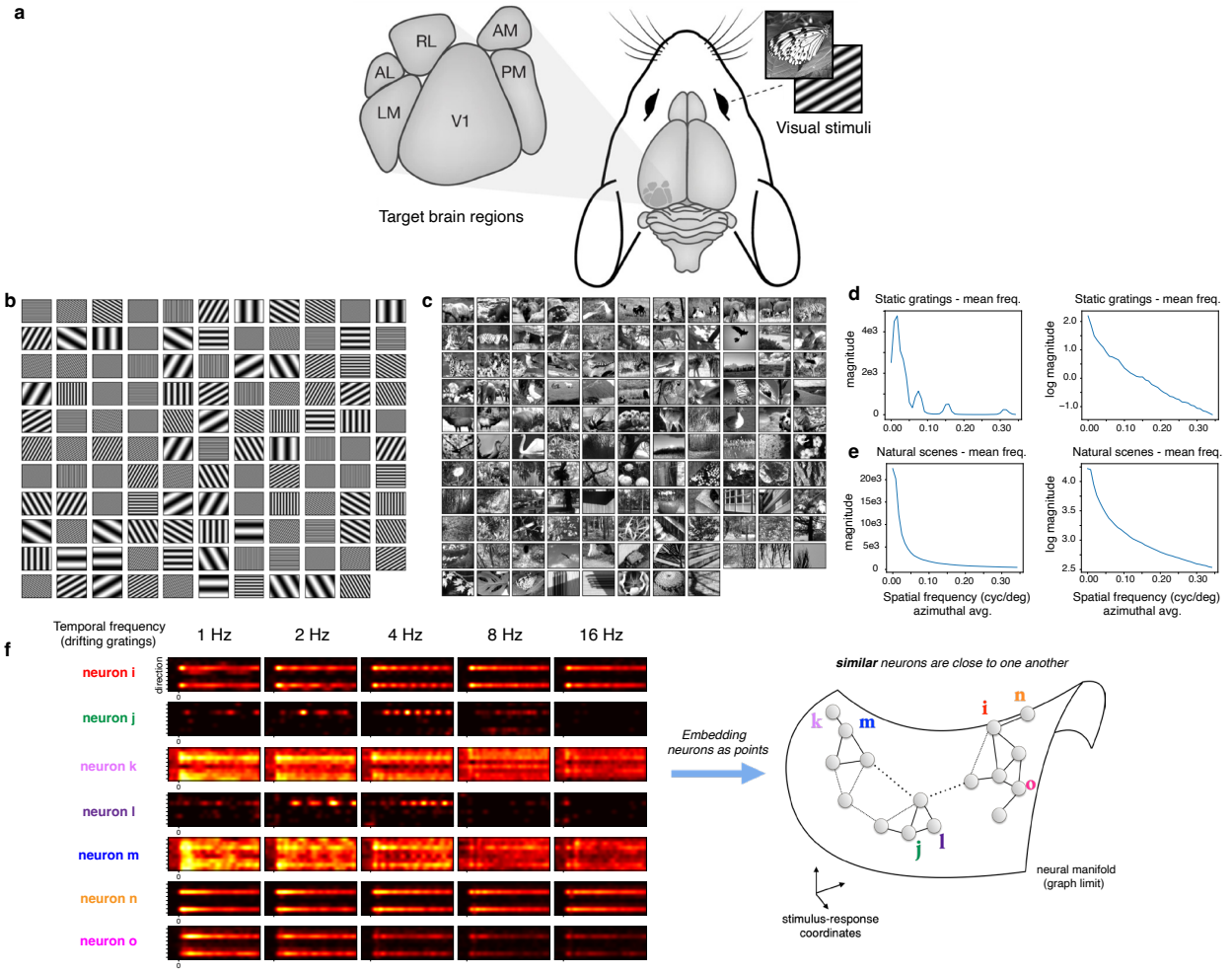

Fig. 1: **Building encoding manifolds from visual cortical responses to artificial and naturalist stimuli.** **a**, The locations of the visual cortical areas recorded at the Allen Institute while presenting drifting gratings, flashed static gratings, and flashed natural scenes. **b–c**, The ensemble of static gratings (**b**) and natural scenes (**c**) used in the Allen dataset. **d–e**, The spatial frequency contents of the flashed gratings (**d**) and natural scenes (**e**); they are similar. **f**, PSTH of responses to an ensemble of drifting gratings that vary in spatial frequency (columns) for several neurons (rows). The encoding manifold organizes neurons so that those nearby on the manifold respond similarly in time to similar stimuli. Neurons i and k exhibit similar activity patterns, and are close, while neurons i and j are different. This is illustrated by the underlying data graph (see Methods) superimposed on the manifold. Part **a** modified from [19]; part **f** after [23].

Results

We began by determining the extent to which the fundamental structure of the encoding manifold of mouse primary visual cortex (VISp) generalizes from our prior visual stimulus set to that utilized by the Allen Institute [3]. In previous work, we analyzed responses to drifting gratings and naturalistic 'flow' stimuli, composed of points and line segments moving semi-coherently [22, 23]. However, the previous data were limited to VISp and did not thoroughly examine how the structure of the encoding manifold depended on the temporal frequency composition of the stimulus. The Allen Institute dataset has several features that allow a useful comparison (Fig. 1). First, it consists of measurements not only from VISp but also from five higher cortical visual areas in mouse: VISam, VISal, VISpm, VISlm, VISrl (Fig. 1a). Second, the stimulus set contains drifting gratings presented at five temporal frequencies. Third, it contains static gratings presented at multiple spatial frequencies (five), orientations (six), and phases (four) (Fig. 1b). Finally, the Allen Institute dataset contains static images of natural scenes (Fig. 1c). The spatial frequency content of the flashed gratings and natural scenes were similar (Fig. 1d,e). Thus, we used the Allen Institute dataset to generate, analyze, and compare encoding manifolds along the mouse cortical visual hierarchy.

The encoding manifold reveals novel relationships between stimulus selectivity and neuronal classes

Encoding manifolds were calculated from the Allen Institute data using the responses to both drifting and static gratings. In brief, the encoding manifold arises from a machine learning algorithm that learns to place individual neurons in a relatively low-dimensional space according to their similarity in stimulus selectivity and response dynamics: neurons nearby on the manifold exhibit similar responses to similar stimuli (Fig. 1f) [23]. Responses to the natural scene stimuli were not included in the calculations of the encoding manifolds.

Consistent with prior results [23], the encoding manifold of VISp was relatively smooth and continuous, with neurons uniformly carpeting stimulus/response space (Fig 2a). Note that this topology is quite distinct from the manifold produced by this analysis when applied to retinal responses [23]. Representative peri-stimulus histograms (PSTHs) for preferred stimuli of 5 temporal frequencies are shown for several sets of neurons; differences in selectivity for stimulus orientation and in the temporal responses to drifting gratings across the manifold are hinted at by these PSTHs.

To more clearly visualize how response properties are organized across the manifold, neurons (points) are colored according to their preferred stimuli and/or response features. Clear gradients or geometries emerged (Fig. 2b,c) in particular coordinates (see Methods). For example, a gradient in the orientation selectivity index (OSI) of each cell was revealed by the manifold (Fig. 2b). However, not all stimulus features were so well organized. Entropy, a measure of the range of stimuli for which a neuron gives robust responses, exhibited generally lower values on one side of the manifold, but was distributed more variably than OSI (Fig. 2c). Meanwhile, neurons were not particularly well organized with respect to their spatial frequency preference (Fig. 2d), a point to which we shall return in the last subsection of the Results.

Broadly, the organization of stimulus features across the manifold had a similar topology to that from our previous studies that used distinct stimulus ensembles [22][23]. This suggests that the topology of the VISp encoding manifold is not strongly dependent on the chosen stimulus set, provided that the set is sufficiently diverse. Of course, we cannot rule out that other topological features (i.e., distinct clusters) might emerge with yet-to-be tested stimuli.

The organization of features that were not used in the calculation of the manifold is more interesting. Neurons were classified on the basis of their electrical waveforms as putative excitatory (broad-spiking) or inhibitory (narrow-spiking). The putative excitatory neurons were concentrated on the right in this view of the manifold, with a secondary concentration salient at the lower left, and were present at lower concentration throughout the middle of the manifold (Fig. 2e). In contrast, the putative inhibitory neurons were

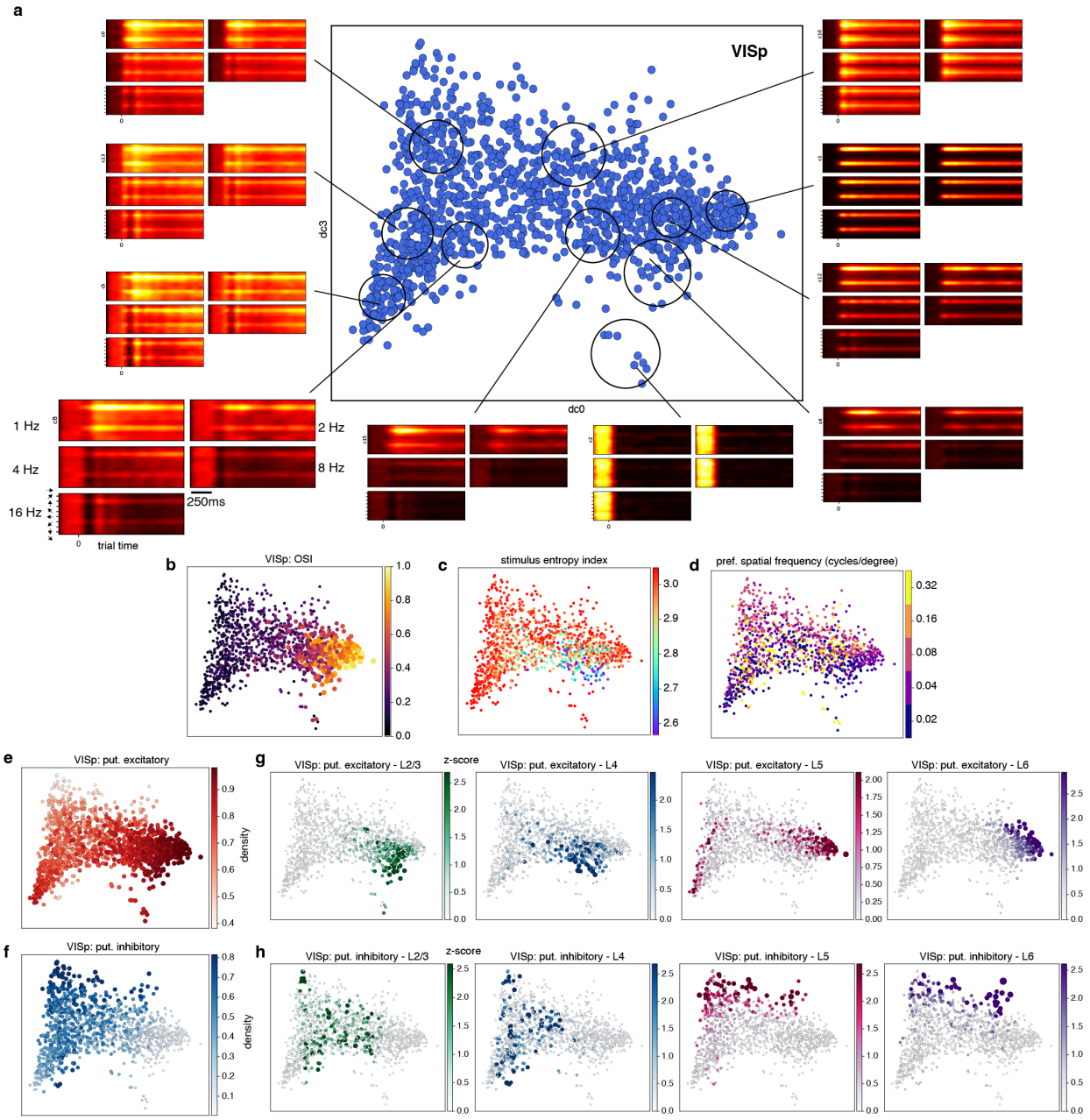

Fig. 2: **The encoding manifold for V1 (VISp) computed from responses to drifting gratings.** **a**, A projection onto the first (dc0) and fourth (dc3) diffusion coordinates of the encoding manifold (see Methods) shows a smooth transition in ensemble PSTHs, from neurons sharply tuned for orientation (right side) to those broadly tuned (left side), indicating a continuous topology. The mean PSTHs for 10 different groups of neurons indicated on the manifold are arranged by temporal frequency from left to right then top to bottom at 1, 2, 4, 8, and 15 Hz. The PSTHs have been aligned by the preferred orientation to de-emphasize the specific orientation to which each neuron is tuned. **b–d**, Encoding manifold in **a** reproduced, with points colored according to orientation selectivity index (OSI) (**b**), stimulus entropy (**c**), and preferred spatial frequency (**d**). Note the OSI values varies smoothly and continuously across the manifold, but are progressively less organized for stimulus entropy and preferred spatial frequency. **e–f**, Encoding manifold from **a** with points colored by the density of putative excitatory neurons (**e**) or putative inhibitory neurons (**f**): note the distributions are complementary. **g–h**, Encoding manifolds reproduced showing the locations of putative excitatory and inhibitory neurons by cortical layers 2/3, 4, 5, and 6.

concentrated in the upper left corner and were also present at a lower concentration throughout the middle of the manifold; they were less well represented in the regions of the manifold in which the excitatory neurons were concentrated (Fig. 2f). The encoding manifold recapitulates the well known observation that inhibitory neurons tend to be less selective for stimulus orientation and to fire at higher rates than excitatory neurons in mouse V1 [51, 34].

The encoding manifold for V1 also organized neurons according to cortical layers even though layer information was not used to generate the manifold (Fig. 2g,h). Putative excitatory neurons in layer 2/3 were concentrated in the lower right, as are those in layer 4 though less tightly, while excitatory neurons in layer 6 were much more tightly concentrated in the upper right. It is known that layer 6 excitatory neurons tend to have low firing rates [66], which is confirmed by the manifold. Excitatory neurons in layer 5 have two foci of density: one together with those of layer 6, but more tightly focused, and another in the salient concentration at the lower left. This bimodal distribution reveals that layer 5 contains (at least) two populations of excitatory neurons with quite different response properties [35]. The distributions of putative inhibitory neurons in layers 5 and 6 along the top of the manifold toward the left are similar to each other, whereas those in layers 2 and 4 are concentrated lower down except for a cluster at the very top left corner. While the layers differ from one another in the distributions of excitatory and inhibitory neurons, within each layer these two major classes of neurons are largely complementary. Again, we emphasize that these features of the data were not utilized in generating the encoding manifold, and thus they demonstrate the ability of the manifold approach to reveal novel relationships between stimulus selectivity, response dynamics and neuronal classes.

These results recapitulate several observations from our previous use of encoding manifolds to understand mouse retina and VISp [23]. However, they add to that work by demonstrating the approach is robust when the experiments use (somewhat) different stimuli, different neuronal recording technology, and are performed by a different laboratory.

Encoding manifolds organize natural scene responses relative to gratings responses

A novel feature of the Allen Institute dataset relative to previous applications of the encoding manifold is the inclusion of responses to natural scenes. This feature provided an opportunity to examine the topology of neural responses to natural scenes relative to gratings. While we did not utilize the responses to natural scenes in generating the encoding manifolds, we could display how neurons were organized on the grating-derived manifold according to their responses to natural scenes [32]. Given the (perhaps) unbounded variability of natural scenes [62, 42], it is difficult to know how to appropriately arrange the responses, so we began by simply plotting the mean firing rate to natural scenes across the manifold (Fig. 3a). Surprisingly, the geometry of the natural scene responses, like that of the grating responses, varied smoothly across the manifold from right to left (Fig. 3b). Comparing the kinetics of the natural scene to grating responses in individual neurons reveals similar dynamics (Fig. 3c,d), but some neurons exhibited larger responses to natural scenes (Fig. 3c), while others responded more strongly to gratings (Fig. 3d).

It should be noted that the static grating stimuli can be said to tile the (static) stimulus space for gratings, encompassing the relevant range of orientations and spatial frequencies so as to stimulate all grating-responsive neurons well, regardless of the location of their receptive fields on the stimulus screen. Therefore, the best response measured with these stimuli for each neuron is likely to be the near-optimal grating response. However, the natural scenes are a small selection from all possible natural scenes and, since they are flashed (not drifted), the responses to them presumably depend strongly on the location of each neuron's receptive field in the visual field.

Nevertheless, the comparable co-variation in the geometries of natural scenes to gratings in the manifolds (Fig. 3a,b) inspired us to consider the ratio of each neuron's responses to the two types of stimuli. This comparison produced a striking result (Fig. 3e): the relative responses to natural scenes were organized

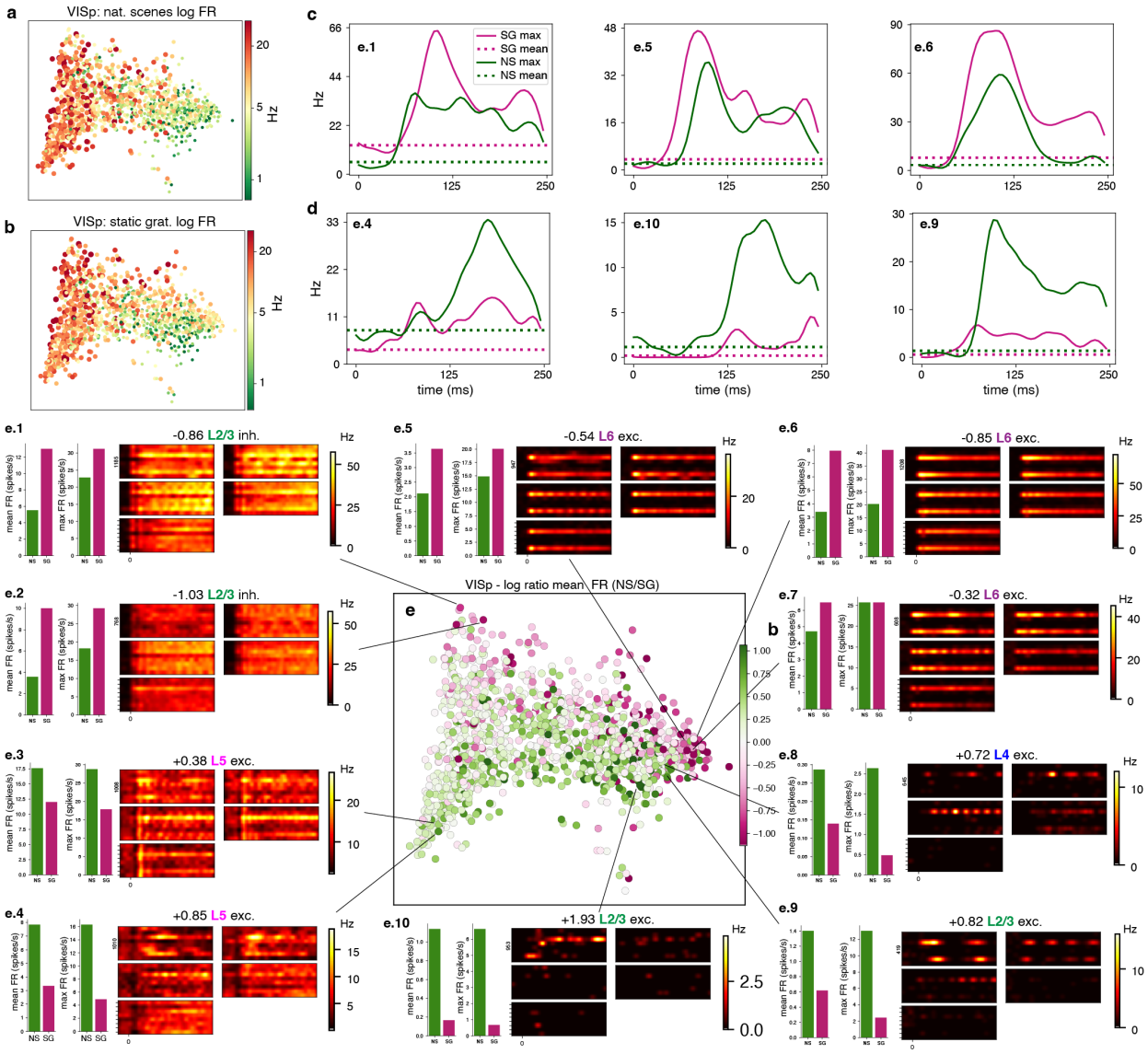

Fig. 3: **The encoding manifold inferred from gratings organizes the ratio of natural scene to static grating responses. a–b**, Neurons on the encoding manifold from Fig. 2 colored by log mean firing rate to all natural scenes (**a**) and log mean firing rate to all static gratings (**b**). **c–d**, Firing rates (kernel-smoothed, see Methods) over at least 39 trials as a function of time for three representative neurons that respond better to static gratings than to natural scenes (**c**), or respond better to natural scenes than to static gratings (**d**). **e**, Encoding manifold with points (neurons) coloring by the ratio of their mean response to gratings versus natural scenes: note a gradient from top to bottom of neurons perfering gratings to those natural scenes. **e.1–e.10**) PSTHs as in Fig. 2 and bar plots of firing rates to gratings and natural scenes are shown for representative neurons whose positions are indicated on the manifold. Bar plots show the result is robust when using either 'mean' (using all stimulus variations) or 'max' (using the preferred stimulus variation) firing rates. Notice the top-to-bottom organization of natural scene preference is approximately orthogonal to that of firing rate and orientation selectivity (cf. Fig. 2).

along an axis *orthogonal* to the the gradient of orientation selectivity on the manifold (Fig. 2b). Specifically, the average response of each neuron to the 118 natural scenes (Fig. 1c) was compared to its average response to all the static grating stimuli (Fig. 1b). Note that the two types of stimuli were presented identically, in separate blocks that were interleaved between the repetitions of each block. The neurons preferring natural scenes (green) were distributed along one side of the manifold while those preferring static gratings were more prevalent on the other side (magenta). An alternative analysis compared the maximum response to any individual stimulus from among the sets of 120 static gratings and 118 natural scenes; this is potentially a noisier measure of selectivity than using the average responses to the entire stimulus set, but it reveals the same result: when the manifold is colored by the ratio of each neuron's response to the best natural scene versus its response to the best static grating, the picture is very similar (Fig. S1). This surprising geometry across the manifold for natural scenes (relative to gratings) is not a trivial consequence of differences in firing rates; the distributions of average firing rates to static gratings and to natural scenes are similar (Fig. 3a,b).

Despite the marked difference in the two stimulus sets, a relationship between grating preferences and natural scene preferences was revealed by the encoding manifold. We note that it has been notoriously difficult to find any consistent characterization of natural scene responses in visual cortex [24, 55, 57, 9, 72], especially when considering average responses [48]. Yet the encoding manifold has found an organization whereby at any given range of grating preferences, there appears to be an ensemble of neurons that display a range of preferences for gratings versus natural scenes.

Higher cortical areas exhibit a relationship between natural scenes and gratings similar to VISp

After analyzing VISp, we next computed encoding manifolds from responses to the same gratings for the five higher cortical visual areas in mouse with the most data in the Allen Institute Neuropixels database: VISam, VISal, VISpm, VISlm, and VISrl (Fig. 4). The manifolds from all areas were continuous in a manner similar to VISp and distinct from the retina [23]. Furthermore, when the orientation selectivity index (OSI) and firing rate are plotted as functions on the manifolds, they vary smoothly and similarly, as they did for VISp (Fig. 4 left and center-left columns). Once again, high firing rates are organized with low OSI, and conversely. In the diffusion coordinates chosen, spatial frequency preference does not seem well-organized for any of the higher visual areas (Fig. 4 center-right column). However, the ratio of the firing rates elicited by flashed natural scenes (NS) versus flashed static gratings (SG) does vary smoothly for all areas (Fig. 4 right column and Fig. S2), a point to which we shall return in the last section of the Results.

We now focus our analysis on two of the areas, VISpm and VISal, thought to be representative of the two major higher cortical visual streams, ventral and dorsal, respectively ([25, 44], but see [41]).

VISpm

The predominant responses of VISpm neurons to high spatial and low temporal frequencies suggest that it may initiate the mouse homologue of the primate "ventral stream", which runs through the inferior temporal lobe and is thought to participate in object recognition[25, 44, 57]. The bias in VISpm toward lower temporal frequencies is indicated by the example PSTHs shown (Fig. 5a.1–a.9), especially toward to lower portion of the manifold). As mentioned above, firing rate and OSI are well-organized coordinates across this projection of the manifold (Fig. 5b,c), as is the complementary distribution of excitatory versus inhibitory densities (Fig. 5d,e). Note how neurons with high OSI are on the left with a gradient to lower OSIs on the right (Fig. 5c). As in VISp, there is a complementary distribution of excitatory versus inhibitory neuronal densities for each layer (Fig. 5f,g), with inhibitory neurons tending to align with higher firing rates.

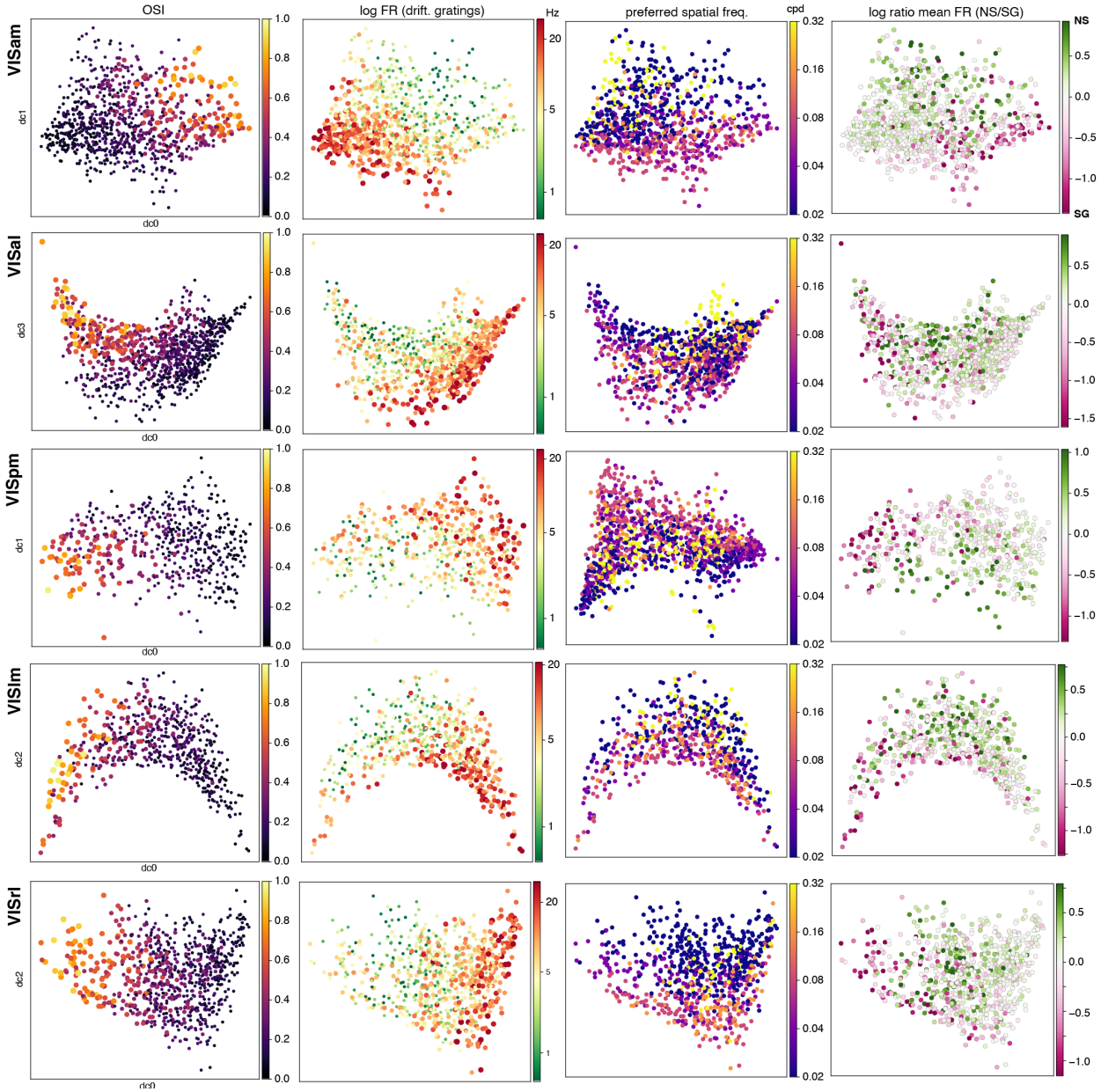

Fig. 4: **Encoding manifolds for five higher visual areas (HVAs) computed from grating responses.** Rows indicate the different HVAs and columns indicate features shown in color scale bars. OSI and firing rates to drifting gratings (first two columns) are similarly well organized (see Methods), exhibiting clear gradients; note similarity of organization of these properties across areas and with VISp (Figs 2 & 3). Preferred spatial frequency (third column) is less well organized across the HVA manifolds, also like in VISp (Fig 2). The ratio of Natural scene firing rates versus grating firing rates (fourth column) is well organized. Unlike in VISp, neurons in the HVAs that prefer gratings tend to have high orientation selectivity (OSI) and lower firing rates (FR).

These observations recapitulate at a population level much of what was known previously for VISpm at the individual cell level [25, 44, 57].

While there were many similarities to VISp, there were interesting differences in the layer concentrations. In VISpm, layer 2/3 had significantly different concentrations of excitatory neurons from layer 4, and layer 5 showed two populations of excitatory neurons. Layer 6, on the other hand, was closer to layer 2 in distribution. We note that previous analyses of layer responses did not distinguish putative excitatory from inhibitory cell types [68].

We next compared the relative responses of natural scenes to static gratings for VISpm (Fig. 5a). As with VISp, a distributed organization was observed for the ratio of natural scene responses to grating responses, but it is no longer orthogonal to orientation selectivity: unlike VISp, the main concentration of static grating preferring neurons overlaps with the most orientation-selective neurons on the left side (cf. Fig. 5a–c). Examining the layer distribution, these grating-preferring neurons tended to be excitatory neurons in layer 2/3 and layer 6 (Fig. 5f). The inhibitory neurons exhibited little preference for static gratings or natural scenes (cf. Fig 5a–g). These relationships held both when comparing the mean response to all natural scenes compared to the mean response for static gratings, or when comparing the maximum responses for both stimuli. The most curious feature of the layer distributions was the bi-lobed distributions of layer 5 putative excitatory neurons, suggesting that this layer contains two populations of neurons with distinct response dynamics and/or stimulus selectivity.

VISal

We repeated these analyses for VISal (Fig. 6), an area that prefers higher temporal frequencies and lower spatial frequencies, suggesting it may initiate the mouse homologue of the primate "dorsal stream" (cf. the PSTHs between Fig 6a.1–a.10 and Fig. 5a.1–a.9) [44, 57]. As expected, cells were more broadly tuned than in VISpm, however, there remains a clear gradient of OSIs across the manifold (Fig. 6c). Furthermore, similar to VISpm, a gradient was present for neurons preferring natural scenes versus gratings (Fig. 6a). About half the neurons, particularly those responding more strongly to natural scenes than to gratings, show little evidence of orientation selectivity. In general, then, there is substantial similarity in functional organization between VISpm and VISal, as has been reported previously, even though the projections from VISp are largely segregated [36].

The distributions of putative excitatory and inhibitory neurons in the different cortical layers of VISal are strikingly diverse (Fig. 6d,e). In general, within any one layer, the concentrations of excitatory and inhibitory neurons are largely but not fully complementary (Fig. 6f,g). Excitatory neurons in layer 6 are distributed over a region consisting mostly of neurons that prefer gratings over natural scenes, whereas the layer 6 inhibitory neurons are concentrated in the right-most portion of the manifold. Layer 2/3 is similar to layer 6. Layer 4's excitatory neurons extend into the center of the manifold, differently from any of the other layers, and its distribution of inhibitory neurons largely intersects that of layer 2/3. Layer 5 in VISal has the most unusual distribution: excitatory neurons are most concentrated in the upper right, where we find a similar overall concentration of both excitatory and inhibitory neurons (compare with panels **d** and **e**). Layer 5 inhibitory neurons occupy the lowest portion of the manifold, different than any other layer, indicating that they exhibit unique response properties among neurons in VISal. These relations between laminar distributions on the manifold are quite distinct from those found for VISpm, which can provide insights to the key functional differences among higher visual areas.

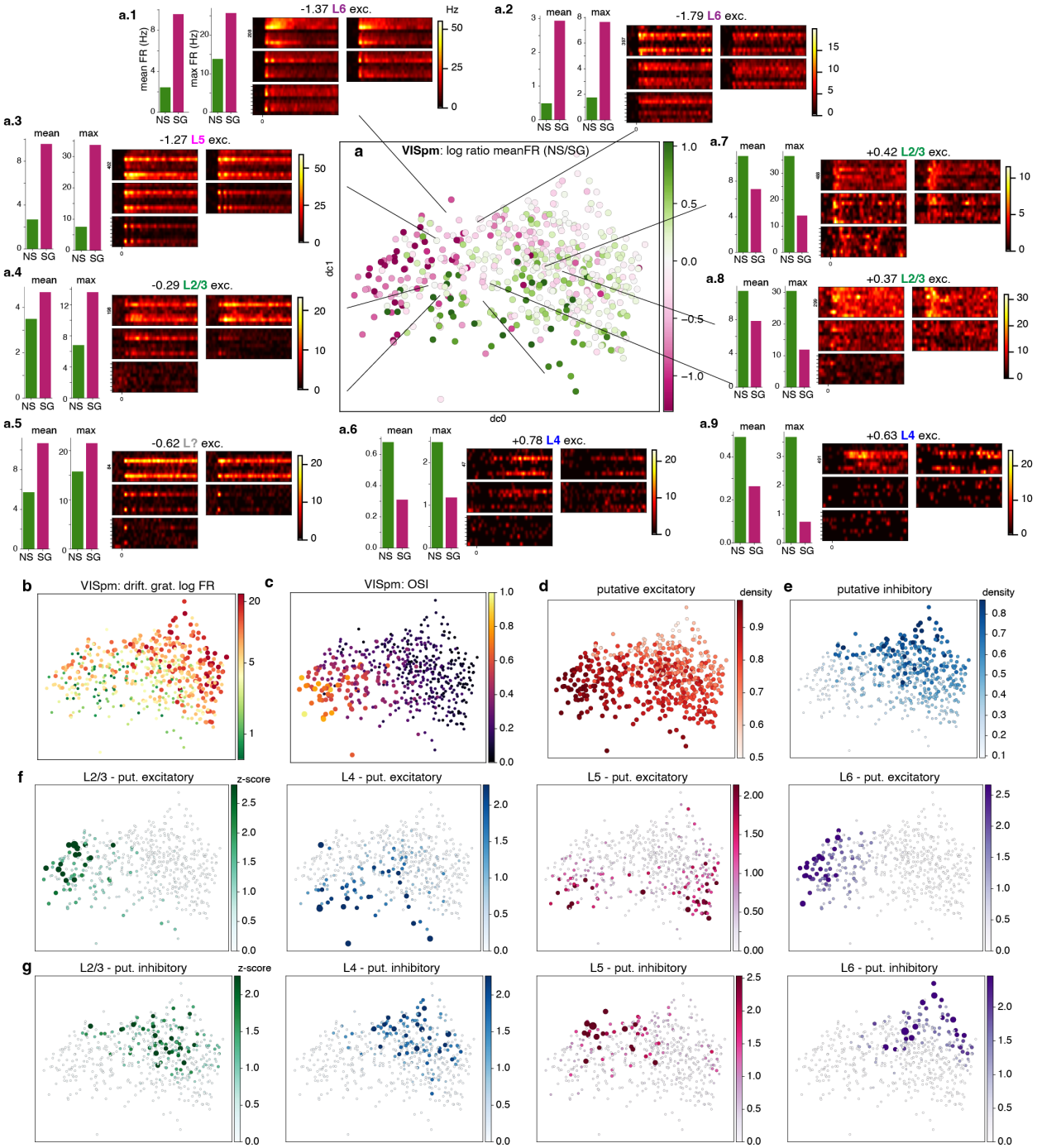

Fig. 5: **The VISpm encoding manifold computed from grating responses.** **a**, The encoding manifold for VISpm colored to indicate ratio of responses to natural scenes versus gratings is again continuous, varying from right (preference for gratings) to left (preference for natural scenes). Details of responses of nine representative neurons (**a.1–a.9**) shown as in Fig. 3. **b–c**, Encoding manifold with points colored according to log mean firing rates to all drifting gratings (**b**) or OSI for responses to drifting gratings (**c**). **d–e**, Encoding manifolds highlighting the relative densities of putative excitatory (broad-spiking) neurons (**d**) or putative inhibitory (narrow-spiking) neurons (**e**). **f-g**, Encoding manifolds indicating relative densities by cortical layer of putative excitatory neurons (**f**) or putative inhibitory neurons (**g**).

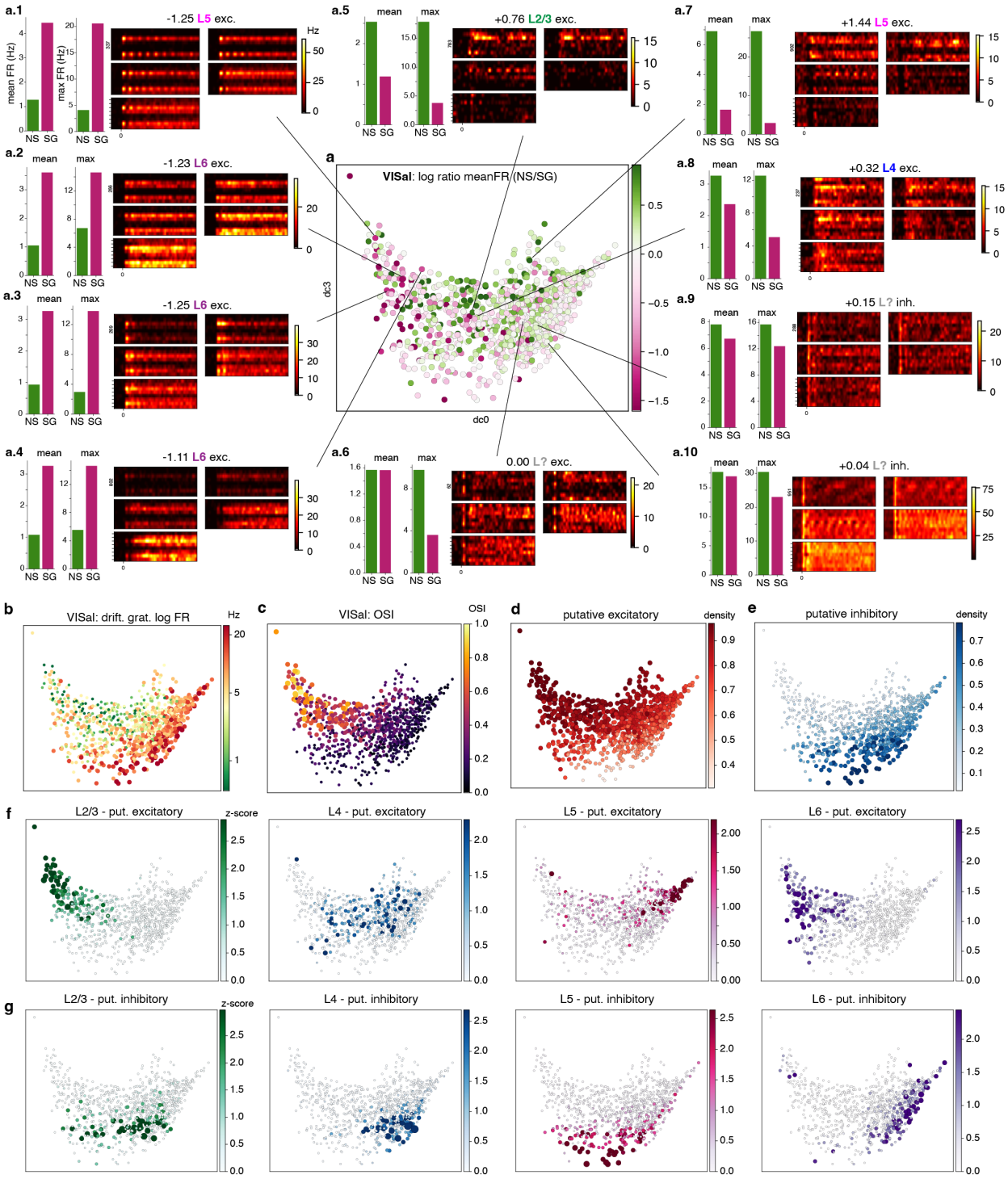

Fig. 6: **The VISal manifold computed from responses to drifting gratings.** Panel organization and analyses are the same as in Fig. 5, but results shown for VISal.

Natural scene preference differs by layer

The previous results hint at some layer-specific preferences for natural scenes vs. static gratings across visual cortical areas in mouse. To investigate this further, we compared the selectivity for natural scenes vs. gratings for each cortical layer and each cortical area (Fig. 7). Specifically, we generated histograms of the log of the ratio of the mean firing rates for neurons responding to natural scenes vs. gratings for each layer and visual area (layers 2 and 6 shown: 7a). Histograms shifted toward positive values indicate a preference for natural scenes, while a shift toward negative values indicates a preference for static gratings. As with our previous comparisons between natural scene and static grating responses, the data were analyzed in two ways. First, the mean response across all displayed natural scenes was compared to the mean response across all static gratings ('mean analysis'). Second, the mean response to the single natural scene that generated the largest response was compared to the mean response to the single static grating that generated the largest response ('max analysis').

The 'mean analysis' revealed a preference among layer 6 neurons for static gratings over natural scenes in each cortical area (Fig. 7b, left columns: magenta (green) indicates grating (natural scene) preference). This preference was less pronounced under the 'max analysis' (Fig. 7b, right columns). In VISp, the 'mean analysis' revealed that layers 2/3 and 4 tend to have stronger responses to natural scenes. This was not the case in the higher cortical areas for layers 2/3 and 4. However, the 'max analysis' did indicate a rather consistent preference for natural scenes over static gratings for layers 2/3, 4. We emphasize these preferences are represented in the average performance across the population of neurons, but that within each cortical layer and cortical area, there was substantial heterogeneity (as revealed by the histograms in Fig. 7a).

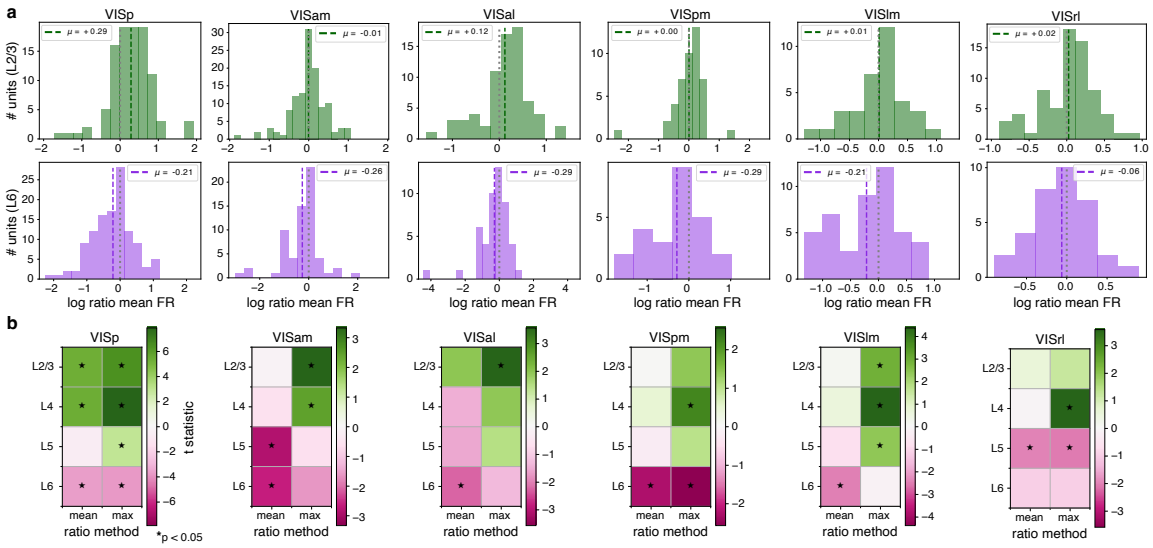

Fig. 7: **Preference for natural scenes versus gratings across cortical areas and layers.** **a**, Histograms of the distribution of log ratios (using mean firing rate across all variations of natural scenes vs. static gratings) for layer 2/3 (**top row**) and layer 6 (**bottom row**); each column shows a different visual cortical area; μ indicates the mean. In layer 2/3, the VISp histogram is shifted toward positive values, indicating a preference for natural scenes; all other areas have histograms centered approximately around zero, indicating no layer-specific preference for natural scenes compared with gratings. By contrast, neurons in layer 6 show a significant preference for gratings in all areas (negative shift) except for VISrl. Two-tailed t-tests were performed to determine if each distribution differed significantly from a normal distribution (same variance) centered at 0. Significance ($p < 0.05$) is indicated by $\star$. **b**, Summary of t-test statistics comparing the log ratio between natural scenes (positive, green) and static gratings (negative, magenta), per cortical layer and across visual areas. Values for ratios considering both the mean (left column) and maximum firing rates (right) are shown. A preference for gratings dominates in layer 6 across all visual cortical areas, especially in the average responses, while most neurons in other layers respond more strongly to at least one natural scene.

Global relationship between spatial frequency tuning and natural scene preferences

The encoding manifolds uncovered a topological relationship between natural scene responses relative to grating responses in VISp and the higher visual areas (Figs. 2a, 6a, 5a). This is surprising because natural scene responses were not used in building these manifolds. This indicates that some aspect of the responses to gratings must signify a preference for natural scenes. To uncover this signifier, a plausible place to start is with the selectivity to spatial frequencies. We first binned neurons in each visual area by their spatial frequency preference (Fig. 8a). Consistent with previous work, there was a preponderance of neurons preferring lower spatial frequencies [25, 44]. We then calculated the log ratio of the mean response across natural scenes to the mean response across static gratings for each neuron as a function of that neuron's preferred spatial frequency (to drifting gratings) (Fig. 8b). The result was a new invariant across cortical areas: a "U"-shaped distribution, showing that neurons in the group preferring either low or high spatial frequencies tended to have the largest responses to natural scenes, while those preferring intermediate frequencies tended to responded more vigorously to gratings. A two-tailed t-test was used to determine if the distribution differed significantly ($p < 0.05$) from a normal distribution (with the same variance) centered at zero. Importantly, this distribution held not only in VISp but also in all five of the higher visual areas studied. To check whether it held at the population level and considering only the grating responses, we combined the low- and high-frequency preferring neurons together, versus those preferring intermediate frequencies, and plotted this new statistic on the encoding manifold. We found complementary densities of neurons preferring low- + high frequencies versus neurons preferring intermediate frequencies, and their distributions over the manifold vary almost identically to the distributions of neurons preferring natural scenes versus gratings (Fig. 9e–g, and Fig. S6). Thus, the population of natural scene preferring neurons distribute on the manifold in a similar manner to neurons preferring high or low spatial frequencies. In the Discussion, we comment on the potential significance of this observation.

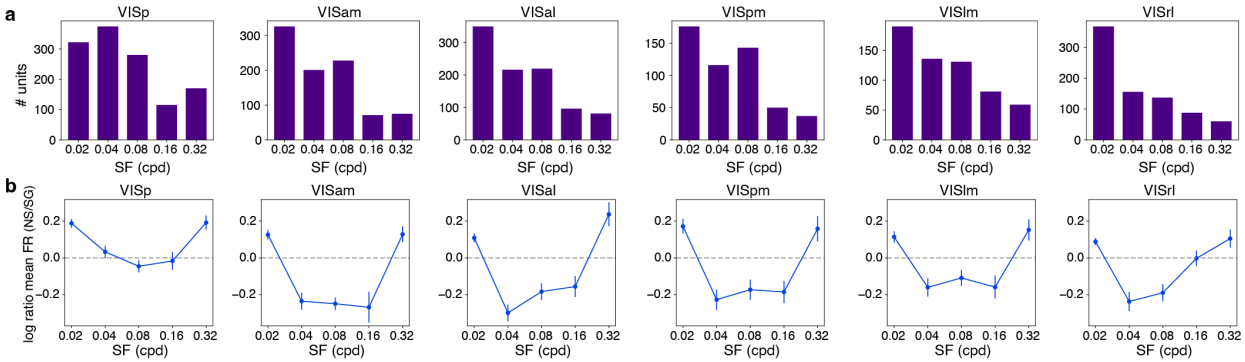

Fig. 8: **Comparing preference for spatial frequency and natural scenes versus gratings.** **a**, Distribution of numbers of units preferring a given spatial frequency (SF) in all visual areas. As expected, it drops off with increasing spatial frequency (cf. Fig. 1d, although many neurons prefer higher SFs). **b**, Mean $\pm$ s.e.m. log ratios between mean firing rates to natural scenes vs. static gratings of the neuronal populations in each area as a function of their preferred SF. Neurons that prefer low or high SFs tend to respond more strongly to natural scenes, while those that prefer intermediate SFs respond more strongly to gratings. The "U"-shaped curve holds across all areas..

Discussion

The encoding manifold organizes large populations of neurons based on their responses to an ensemble of stimuli, thereby incorporating both circuit and stimulus properties. Responses to gratings collected by the Allen Institute confirmed the continuous topology of the encoding manifold for VISp that we found earlier

using a different stimulus ensemble [23]. This lends validation to the approach as different visual stimuli and different neural recording technologies were used and the experiments were performed in a different laboratory. We further showed that higher visual areas in mouse shared a similar continuous topology to VISp, while displaying some similarities and differences with respect to the organization of neurons on their respective manifolds from different cortical layers and different neuronal classes (excitatory versus inhibitory). We discuss below two curious aspects of the Allen data revealed by the encoding manifold.

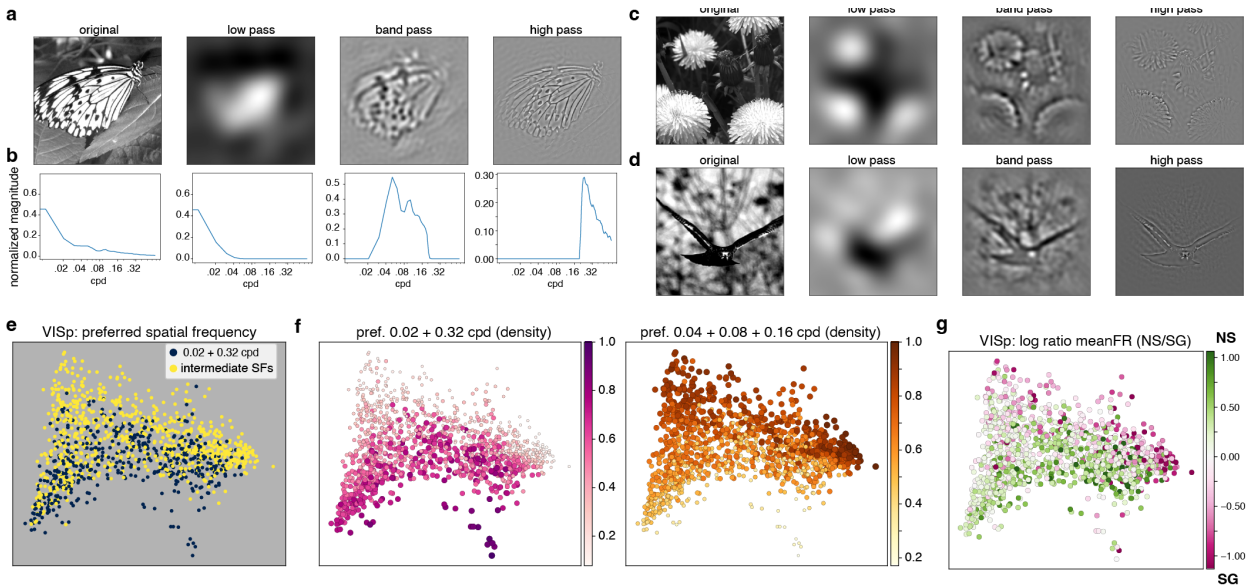

Fig. 9: **Natural scene statistics and neuronal preference on the encoding manifold. a**, An example natural scene image and its low-pass, intermediate-, and high-pass filtered versions, and **b** associated spectra. **c–d**, Two more natural scenes and their filtered versions. **e**, Encoding manifold for VISp highlighting spatial frequency preferences of neurons. Neurons are labeled according to their spatial frequency preference for static grating stimuli, grouping 0.02+0.32 (extreme frequencies, in yellow) vs. 0.04+0.08+0.16 (intermediate frequencies, in blue), agreeing with the natural scene ratios plotted on the encoding manifold in Fig. 3e. **f**, Spatial frequency preferences summarized by the density over the data graph (see Methods) for extreme frequencies (left) and for intermediate frequencies (right). **g**, The log ratios from Fig. 3e, shown here for comparison with the distributions in **f**. The different spatial frequency bands thus provide insight into the U-shaped curves from Fig. 8b, suggesting a grouping of the low and high frequency content signals separately from the intermediate ones. For comparable plots in the higher visual areas, see Fig. S6.

The functional organization of laminar preferences differ across visual areas

Differences in grating selectivity between visual areas have been observed previously [25, 44], and have been interpreted as a hint toward dorsal/ventral stream models [69, 61, 41]. The encoding manifold approach allowed us to refine the population analysis to individual layers, where further, novel differences emerged. For example, in VISp, the greatest concentrations of high orientation selectivity (putative) excitatory neurons were in layers 5 and 6 (Fig. 2b,g). The same was true for excitatory neurons in layers 4 and 2/3, even though they occupied regions practically disjoint from those of layers 5 and 6. In contrast, although in VISal the (putative) excitatory neurons of layer 2/3 lie in the regions of the manifold with high orientation selectivity, those in layer 4 do not (Fig. 6c,f). The same was true to a lesser extent in VISpm (Fig. 5c,f). This is perhaps surprising considering that the dominant feedforward projection from superficial layers of VISp goes to layer 4 of higher cortical visual areas [57]; one would expect the recipient layer 4 populations not to lose selectivity.

The deeper layers are also interesting, since they involve both cortical and sub-cortical projections [66]. Projections to the dLGN may sharpen receptive fields in the mouse [10] and cat [49, 58, 6]), while the cortico-cortical projection could be involved in generating endstopping and other surround effects [20]. Layer 5 also projects to the pulvinar [38], a projection that could be relevant to coordinating natural scene responses [74, 8].

Layers 5 and 6 show different organization for excitatory and inhibitory neurons, depending on the area. In VISp (Fig. 2g), layer 5 exhibits two distinct concentrations, one of which overlaps almost completely with layer 6 in the region of highest orientation selectivity. Similarly, layer 5 neurons in the VISpm manifold are organized in two clusters, although the overlap of the one cluster with the highly selective layer 6 neurons is less complete (Fig. 5c,f). On the other hand, in VISal, layer 5 has practically no overlap with layer 6, even though the latter remains highly selective (Fig. 6c,f). This difference between VISpm and VISal is striking, and consistent with anatomical observations (cf. [36]). The distributions of laminar density for VISam are shown in Fig. S3, for VISlm in Fig. S4, and for VISrl in Fig. S5. Layer 5 in all three cases indicates a bimodal distribution on the respective manifolds, one of which remains consistent with the density of (putative) excitatory neurons in layer 6. Fig. S8) summarizes all of the above comparisons between layers 5 and 6.

In summary, building encoding manifolds from responses to diverse stimulus sets and then painting the locations of excitatory and inhibitory neurons from different layers can provide insights and generate novel hypotheses about the common themes and distinct specialization in the functional organization of visual cortical areas. Expanding these approaches to population recordings from primate visual cortical areas may also provide important insights to the conserved versus divergent features of visual processing across species.

A novel relationship between spatial frequency tuning and natural scene preferences

Natural scenes have been among the most difficult stimuli to analyze, and the encoding manifold revealed a unique invariant at the population level. Although the manifolds computed in this study were based solely on gratings, a surprising and consistent relationship emerged between the natural scene preferences of high- and low-spatial frequency preferring neurons compared to neurons preferring intermediate frequencies. Specifically, neurons tuned to low or high spatial frequencies tend to generate larger responses to natural scenes than gratings, while neurons tuned to intermediate spatial frequencies tend to generate larger responses to gratings (Fig. 8b). How might this curious relationship be explained (cf. [72, 9])? One possibility is that it has to do with object content, or at least a proxy for it. Examining the statistics of image content is instructive (Fig. 9a–d. Low frequencies signal "where" an object might be, while high frequencies signal "what" it might be made of. Such an identification of spatial frequency content with scene layout has long been proposed [63, 53], and blobs do indeed matter at low frequencies [56]. But it is unlikely that simply extracting filtered features is the complete story (e.g., see [70]).

Classically, natural scene analysis has been thought to proceed by a coarse-to-fine visual processing [29]. Object perception, figure/ground discrimination and hunting behavior [31, 60] are complex tasks, implicating computations across multiple visual areas [33], including feedback within and between them [26, 28] and recurrence [37, 73]. A recent study [59] found direct low- and high-frequency dynamics in natural scene viewing, extending classical results on dynamical frequency tuning [2, 12, 46, 67]. Since the encoding manifold considers the entire time course of responses [23], it is able to capture these dynamics—a possible signature of coarse-to-fine processing [43, 30]—in the manifold inference process. Perhaps while building biologically-motivated perceptual systems, following this study, one should consider coarse-*and*-fine processing.

Encoding Manifolds

Encoding manifolds capture similarity in the responses over time of neurons exposed to a wide range of stimuli. They display neurons in relation to as many different stimulus features as one is able to illustrate. No other procedure of which we are aware lets one simultaneously visualize how the neurons that respond particularly well to one stimulus respond to many others. They disclose visually the relationships between, for example in the present data: firing rate, selectivity for stimulus orientation, spatial and temporal frequencies, cortical laminae, electrical waveform (related to excitatory versus inhibitory classes), overall selectivity (stimulus entropy), and preference for gratings versus natural scenes. This property should make encoding manifolds widely useful across biology when many relationships may be present but could go unnoticed.

Methods

Dataset

The "Neuropixels" single neuron recording dataset shared by the Allen Institute includes recordings from many cortical visual areas in alert wild type (C57BL6/J) mice Harris et al. [28]. Recording sites were localized to specific cortical areas by reference to intrinsic signal imaging maps and to cortical layer by transformation into the Allen Common Coordinate Framework. Details of all procedures are described at links from [3].

To summarize, visual stimuli were were displayed on an LCD monitor at a resolution of 1920×1200 pixels at 60 Hz refresh rate. Stimuli were presented monocularly, and the monitor spanned $120^\circ \times 95^\circ$ of visual space prior to stimulus warping. Each monitor was gamma corrected and had a mean luminance of 50 cd/m^2. To account for the close viewing angle of the mouse, a spherical warping was applied to all stimuli to ensure that the apparent size, speed, and spatial frequency were constant across the monitor as seen from the mouse's perspective.

In our experiments, we analyzed responses to a subset of their "Brain Observatory 1.1" stimulus set, including 40 different drifting gratings, 120 stationary gratings, and 118 natural scenes [5]. The drifting gratings were shown at a spatial frequency of 0.04 cycle/deg, 80% contrast, 8 directions of motion (0, 45, 90, 135, 180, 225, 270, and 315 degrees) and 5 temporal frequencies (1, 2, 4, 8, and 15 Hz), with 15 repeats per condition. The stationary grating set, also presented at 80% contrast, consisted of 120 gratings of 6 different orientations (0, 30, 60, 90, 120, and 150 degrees, clockwise from 0 = vertical), 5 spatial frequencies (0.02, 0.04, 0.08, 0.16, 0.32 cycle/degree), and 4 phases (0, 0.25, 0.5, 0.75), each presented for 0.25 seconds with no intervening gray period. The natural scene stimuli consisted of 118 natural images taken from the Berkeley Segmentation Dataset [45], the van Hateren Natural Image Dataset [65], and the McGill Calibrated Colour Image Database [54]. The images were presented in grayscale and were contrast normalized and resized to 1174×918 pixels. The images were presented in a random order for 0.25 seconds each, with no intervening gray period.

Data preprocessing

As in Dyballa et al. [23], responses of a neuron to each stimulus (PSTHs) were kernel-smoothed using the Improved Sheather-Jones algorithm [11], using a bandwidth of 25 ms (typical value obtained for all visual areas when using automatic bandwidth selection) for drifting gratings and 10 ms for static gratings and natural scenes (between 39 and 50 trials). The `KDEpy` Python implementation [52] was used.

Responses to the drifting gratings were tested for significance by performing a two-tailed Mann-Whitney test comparing the mean activity during the second half of the ISI immediately preceding the stimulus versus mean activity over any interval of the same length within the stimulus presentation. This strategy allowed transient responses to be able to reach the same level of significance as sustained responses. Responses were deemed significant if they had a p-value of at most 0.001. Only neurons that had a significant response to at least one variation of drifting grating were included in the data tensor (see below).

Data analysis

Neural encoding manifolds. Neural encoding manifolds were constructed from the spikes elicited by gratings in VISp and each of 5 higher cortical visual areas using the procedure described in Dyballa et al. [23]. Briefly, it involves a dimensionality reduction pipeline consisting of two main steps: applying a permuted non-negative tensor factorization (pNTF) to a tensor built from the temporal responses of each neuron to each variation of grating stimuli, followed by manifold inference using iterated adaptive neighborhoods (IAN) [21] in combination with diffusion maps [17] to produce an embedding in which

each point is a neuron. We briefly overview this material here for completeness; please refer to Dyballa et al. [23] for details and further discussion. Code is available at `https://github.com/dyballa/NeuralEncodingManifolds`.

Tensor decomposition. We organize the data as a *3-way tensor*, NEURONS$^{(1)}\times$ STIMULI$^{(2)}\times$ RESPONSE IN TIME$^{(3)}$. (For an introduction to tensors, see [71].) In standard notation [39]:

$$\tilde{\boldsymbol{T}} = [\![X^{(1)}, X^{(2)}, X^{(3)}]\!] \equiv \sum_{r=1}^{R} v_r^{(1)} \circ v_r^{(2)} \circ v_r^{(3)}, \tag{1}$$

where $\circ$ denotes vector outer product and the *factors* are collected into the factor matrices $X^{(k)}$, with individual factors $v_r^{(k)}$ as columns. A *component* is an associated set of factors, one from each tensor mode.

Factors are normalized by computing $\lambda_r = \prod_{k=1}^{3} \|v_r^{(k)}\|$, where r indexes the component and k the factor mode. Collecting these scalars into a vector $\lambda \in \mathbb{R}^R$, we have:

$$\tilde{\boldsymbol{T}} = [\![\lambda\,;\, X^{(1)}, X^{(2)}, X^{(3)}]\!] \equiv \sum_{r=1}^{R} \lambda_r v_r^{(1)} \circ v_r^{(2)} \circ v_r^{(3)}. \tag{2}$$

Since neuronal firing rates are non-negative, we adopt a non-negative tensor factorization (NTF) algorithm [13] based on CP decomposition. This minimizes the squared reconstruction error [15]:

$$\min_{X^{(1)},X^{(2)},X^{(3)}} \frac{1}{2}\|\boldsymbol{T} - \tilde{\boldsymbol{T}}\|^2 \tag{3}$$

subject to the non-negativity constraint $X^{(k)} \geq 0, k \in [1, 2, 3]$ [40, 15]. Non-negative tensor factorization is further advantageous in the sense that it imposes a part-like decomposition [40, 15].

For implementation, we use a gradient-based direct optimization approach (OPT) [1] in the Tensor Toolbox [7], modified to allow permutation of the response factors. This allowed us to exploit data limitations, since two narrowly-tuned simple cells (for example) are similar except in their preferred orientations, effectively collapsing the orientation-preference coordinate on the manifolds.

Because drifting and static gratings have different trial lengths, two tensors were constructed for each cortical area: one containing the drifting grating responses (in which the stimulus mode consisted of the 5 temporal frequencies) and another for the static grating responses (in which the stimulus mode consisted of the 5 spatial frequencies). The temporal mode was a concatenation of the various directions/orientations (8 directions of motion for the drifting gratings and 5 orientations at the optimal phase for the static gratings).

Neural encoding space. The first mode of each decomposition is a neural factor; together they can be composed into a neural matrix (factors as columns). We interpret this as a linear space spanned by the outer product of the other modes (stimulus and response) after vectorization and normalization. Thus, each neuron can now be viewed as a function of its response to (weighted combinations of) the stimuli. The number of tensor components—technically the dimension of the neural encoding space—is determined by the "explained variance" condition discussed in Dyballa et al. [23], *Methods* section.

Similarity kernel and diffusion maps. The data graph for manifold inference is built in the neural encoding space. Since neighborhoods may vary, we use the IAN kernel [21], a multiscale Gaussian kernel that adapts to local geometry in a (relaxed) optimal fashion. The IAN weighted graph yields the similarity matrix on which the diffusion maps algorithm [17, 16] is based. Diffusion maps are a nonlinear spectral embedding method based on the eigendecomposition of the normalized graph Laplacian that tends to better preserve

the topology of the data [21]. The resulting diffusion coordinates are then used to embed the neurons in (reduced) stimulus-response coordinates (see above). The standard parameters α = 1 (Laplace-Beltrami approximation) and diffusion time t = 1 were used.

One of the main advantages of using a spectral embedding method such as diffusion maps is the natural way in which multi-dimensional aspects of the data can be visualized by looking through the various 'diffusion coordinates' produced (denoted 'dc's in the figures). Importantly, different diffusion coordinates concentrate different aspect of the data. Other popular methods, such as t-SNE [64] and UMAP [47], approach this differently. They are based on a gradient descent optimization that focuses on preserving nearest neighbors at the expense of global structure. Both t-SNE and UMAP tend to collapse these multiple dimensions in ways that end up stochastically mixing together the global coordinates that run through the manifold geometry (see Fig. 29 in [21] for an example of this phenomenon). Diffusion maps, by contrast, yield deterministic results.

Our figures show a projection of the diffusion maps into two dimensions, while the intrinsic dimensionality was between 5 and 6 for all cortical areas; see [21] for computation details. Just as in principal component analysis, where different coordinates (principal components) emphasize different aspects of the data, so too do lower-dimensional projections of diffusion maps illustrate different features of organization. We chose the two-dimensional projections of the manifolds shown in the figures to emphasize how they organize the properties, such as orientation selectivity, noted in the text. We selected the pair of diffusion coordinates that maximized the *coefficient of determination* (R^2) for a plane fit to the function defined by the label values (e.g., OSI or putative cell type density) over the domain given by the 2-dimensional coordinate values for all data points. To illustrate for VISp, we fit a plane in diffusion coordinates dc0 and dc3 to OSI values resulting in $R^2 = 0.99$; to putative inhibitory cell densities resulting in $R^2 = 0.68$ and to putative excitatory cell densities resulting in $R^2 = 0.70$. This provides evidence for the success of the method in organizing cell properties in an unsupervised manner, since those labels were not used to compute the manifolds. Other coordinates organize different properties, such as temporal frequency preference.

Laminar and putative cell type densities. The density of categorical properties around each neuron in a manifold (such as layer, putative cell type, and preferred spatial frequency) were computed as the fraction of adjacent nodes in the non-weighted IAN graph belonging to the same putative type and layer.

Natural scene vs. static grating selectivity ratio. Selectivity ratios for individual neurons were computed as the natural logarithm of the ratio between their firing rate to natural scenes and their firing rate to static gratings. Thus a selectivity ratio of 0 means the neuron produced exactly the same magnitude of response to both stimulus classes; and a positive (respectively, negative) value indicates a stronger response to natural scenes (resp., static gratings). The magnitudes of responses to these two stimulus classes are readily comparable because both consist of static images presented during the same time interval of 250 ms.

Following the methodology from the Allen Brain Observatory SDK [4], mean firing rates (FR) for each stimulus class (namely, drifting gratings, static gratings, and natural scenes) were computed as the mean number of spikes per second of stimulus presentation across all trials of the same stimulus, kernel-averaged across all stimulus variations within the same class. For static gratings, 'variations' meant different orientation (6), spatial frequency (5), and phase (4); for natural scenes, 'variations' meant each of 118 scenes used. We also computed the maximum FR for a given stimulus class as the maximum kernel-averaged FR across all stimulus variations (in other words, as the FR to the preferred stimulus variation). Thus, two ratios were computed: one using the 'mean' responses ("mean FR ratios"), and another using the 'maximum' responses ("max FR ratios").

Additional metrics. We used precomputed stimulus metrics available from the Allen Brain Observatory SDK [4] for firing rates, global orientation selectivity index (OSI), preferred spatial frequency (drifting gratings), and preferred phase (static gratings) for each unit. In addition, a stimulus entropy index (Fig. 2c) was defined as 2^H, where H is the base-2 entropy of the vector containing the relative response magnitudes (divided by their sum) of a neuron to the 5 temporal frequencies used in the drifting gratings experiments. It therefore ranges between 0 (case in which the neuron responds to a single stimulus) and 5 (when it responds with uniform magnitude to all stimuli).

Natural scene filtering. Filtered versions of the natural scenes were created by first computing fast Fourier transforms of each image using the FFT module in the `Numpy` Python library [27]. For the low-pass filtered version, the Fourier domain was cropped with a disk of radius 0.04 cpd; for the band-pass version, an annulus with inner radius 0.04 cpd and outer radius 0.22 cpd; and for the high-pass version a disk with radius 0.22 cpd (see example in Fig. S7).

Acknowledgements

Research supported by NIH Grant EY031059, NSF CRCNS Grant 1822598, the Swartz Foundation (LD), the RPB Disney Award for Amblyopia Research (MPS), and an unrestricted fund to the The Jules Stein Eye Institute from Research to Prevent Blindness and P30 - EY000331 (GDF). We thank the Allen Institute for the use of their data and Cris Neill for discussions.

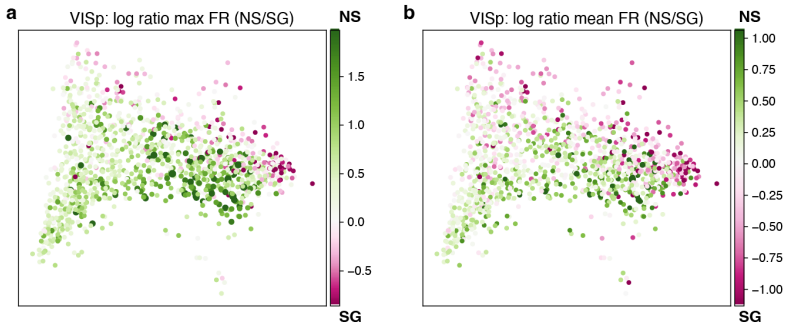

Fig. S1: **Ratio distribution is robust to maximum firing rate.** **a**, Log ratios of the maximum firing rate for all natural scenes divided by the maximum firing rate over all static gratings are plotted on the VISp encoding manifold. **b**, For comparison, an analogous plot using mean firing rates when computing log ratios (cf. Fig. 3 in main text).

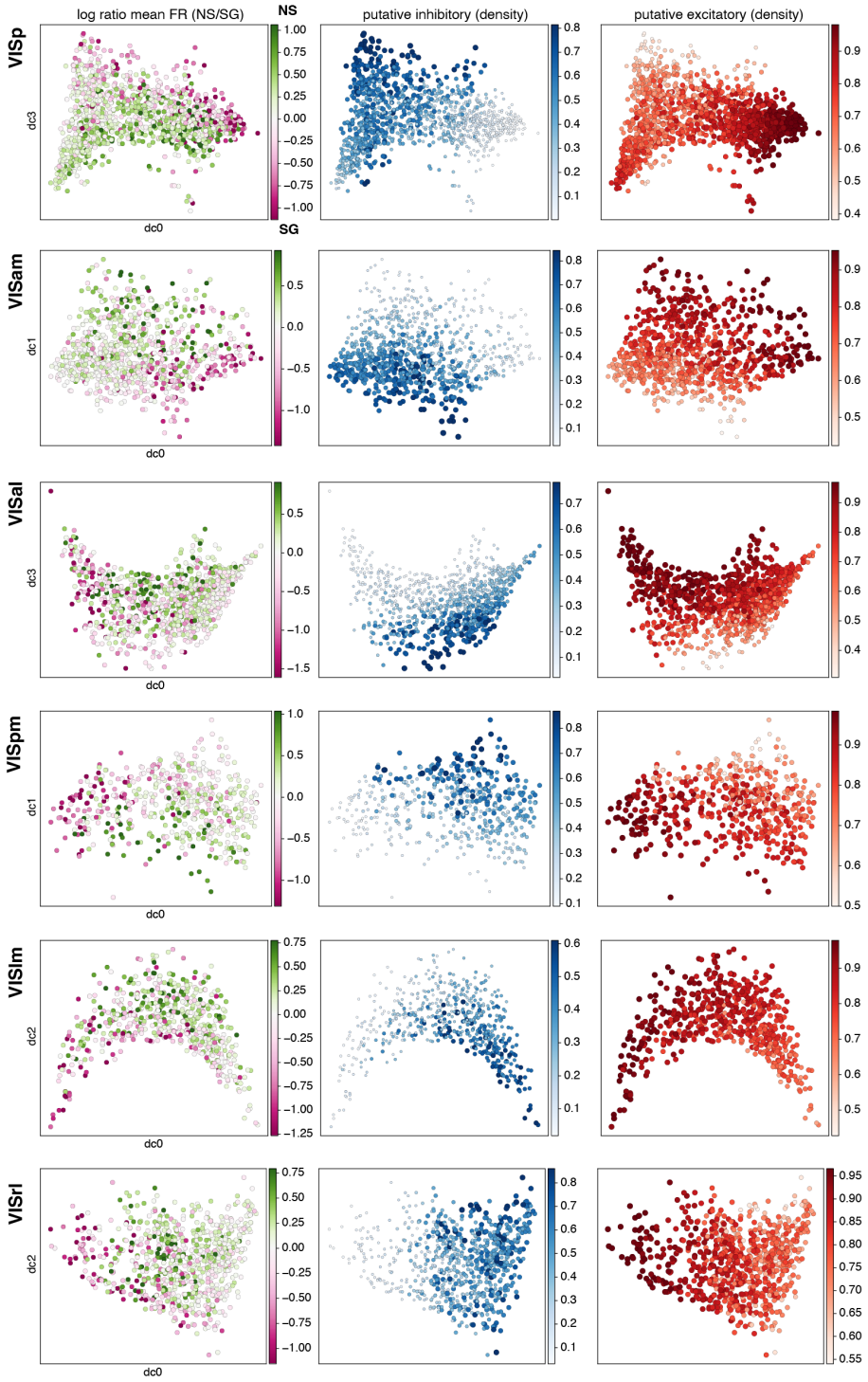

Fig. S2: **Smooth property distributions across all visual areas. left** Encoding manifolds colored by the ratio of natural scene responses to static grating responses. **middle** Encoding manifolds colored by the density of putative inhibitory and **right** putative excitatory neurons.

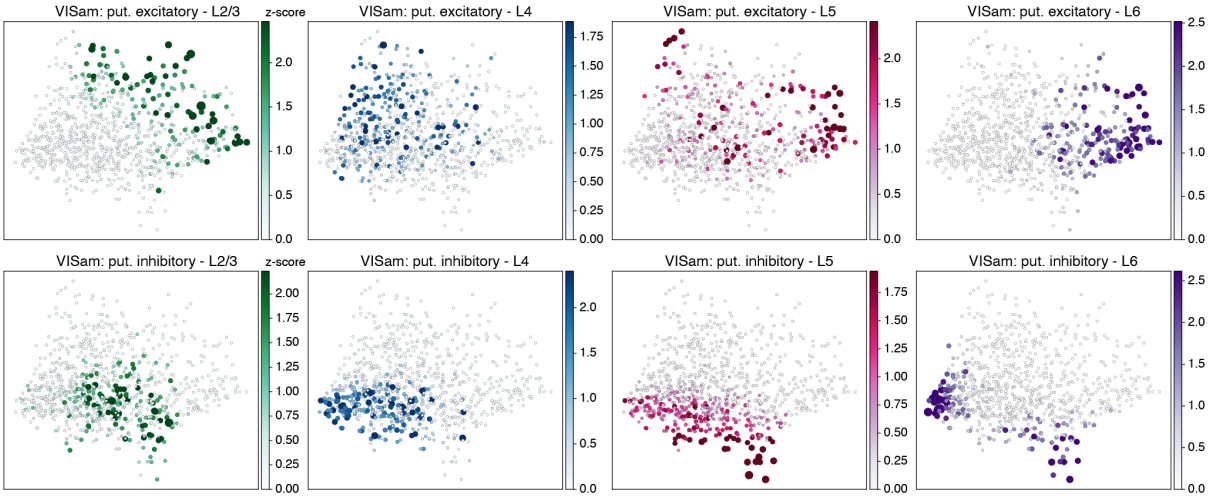

Fig. S3: **Laminar density across the VISam manifold.**. Density of putative excitatory and inhibitory neurons by layer. Note the distinctness of layer 2 vs layers 5 and 6 for excitatory neurons, and the multi-lobed density of inhibitory neurons in layer 6.

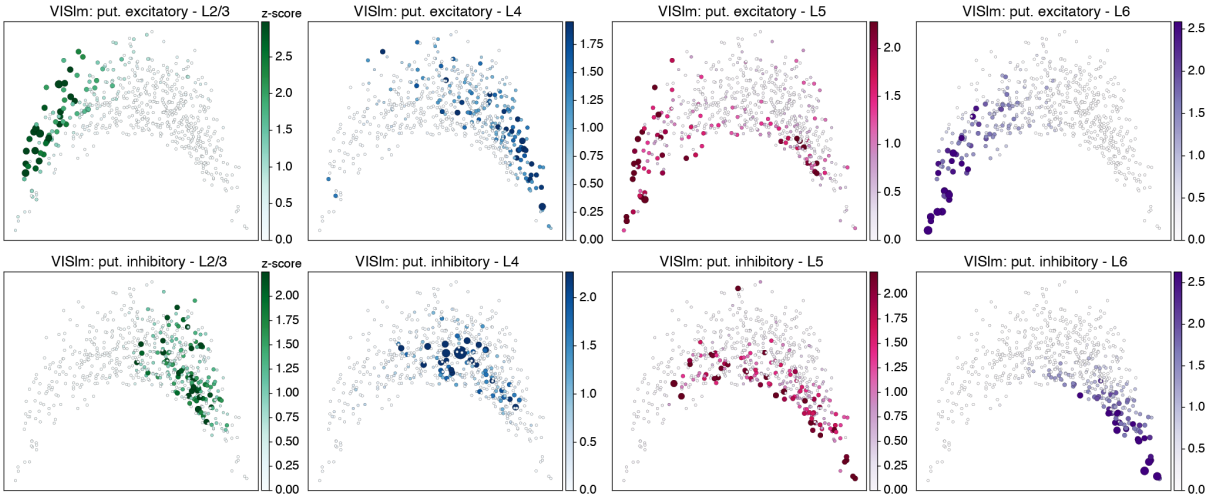

Fig. S4: **Laminar density across the VISlm manifold.** Density of putative excitatory and inhibitory neurons by layer. Note the similarity of layers 2 and 6 for putative excitatory neurons, and the distinctness of layer 4 for putative inhibitory neurons.

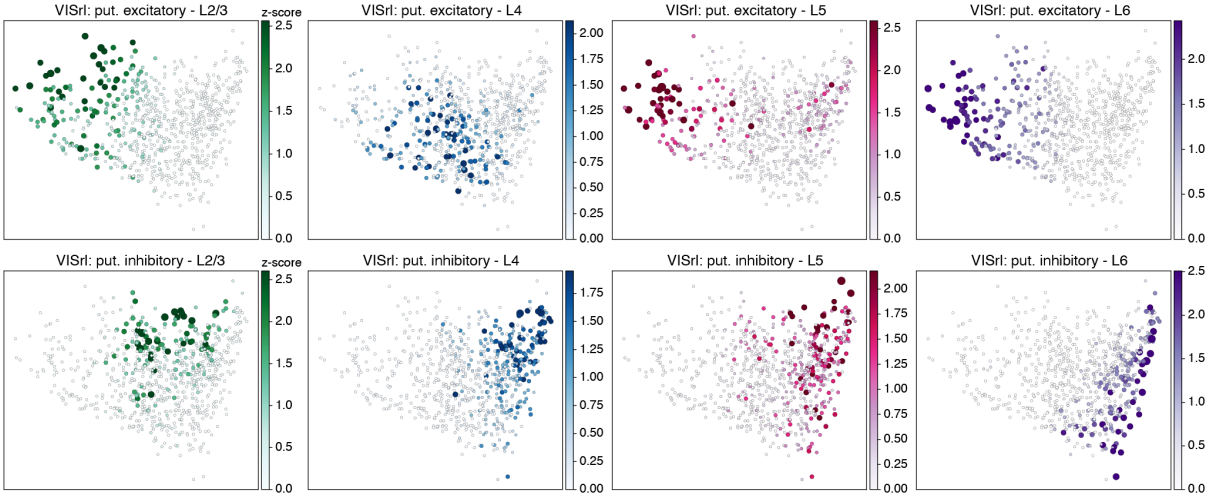

Fig. S5: **Laminar density across the VISrl manifold.** Like VISlm, there is a multi-lobed distribution of putative excitatory neurons in layer 5, and similar agreement between layers 2 and 6.

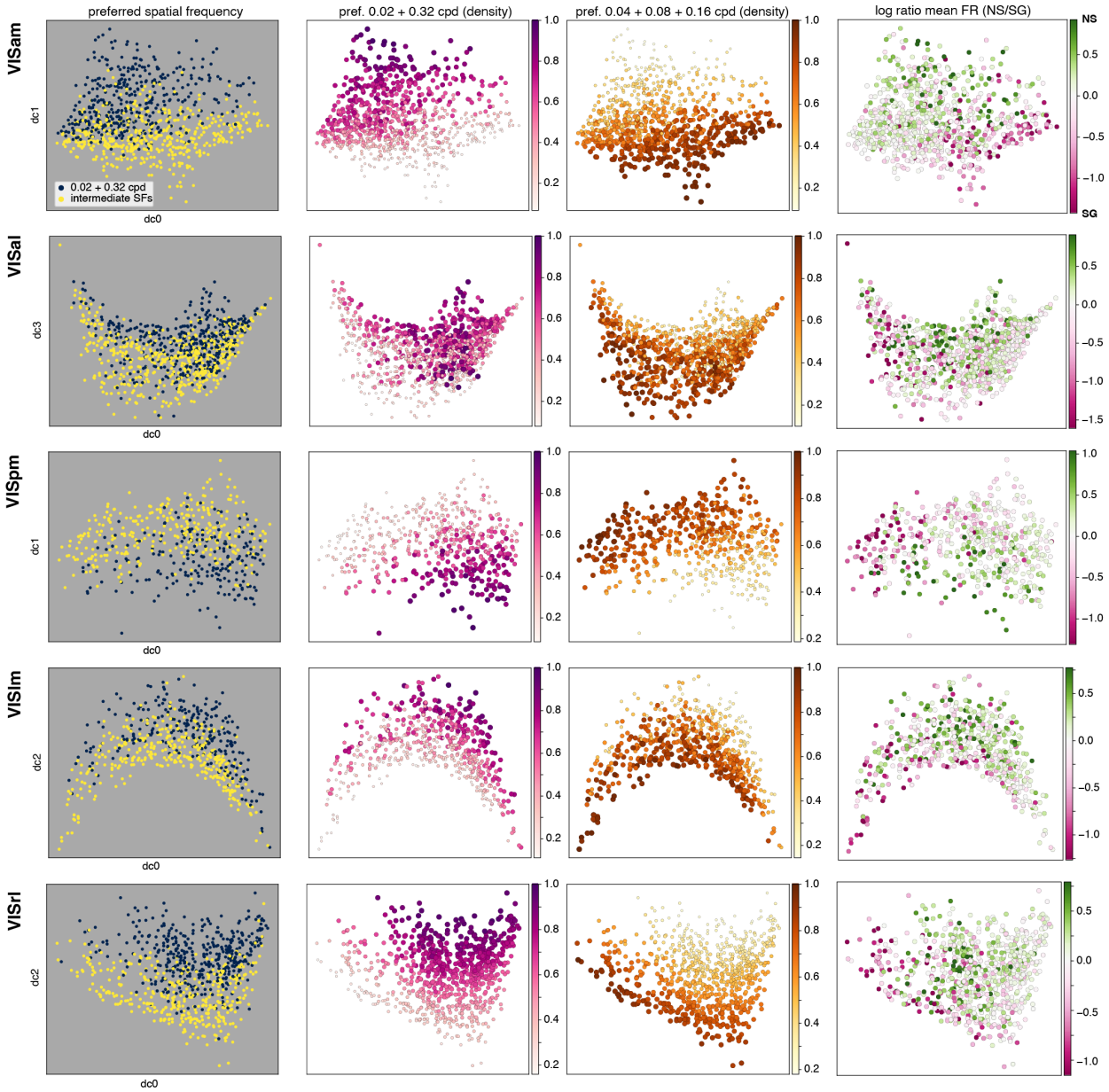

Fig. S6: **Preference for extreme vs. intermediate frequencies.** A grid of colored manifolds, in which each row is a visual area and each column is (from left to right) the ratio of high and low spatial frequencies to intermediate ones; the density of cells preferring high and low; the density of cells preferring intermediate spatial frequencies, and the log ratio plot of mean firing rate to natural scenes (NS) by static gratings (SG). Corresponding plots for VISp are shown in Fig. 8 g,h.

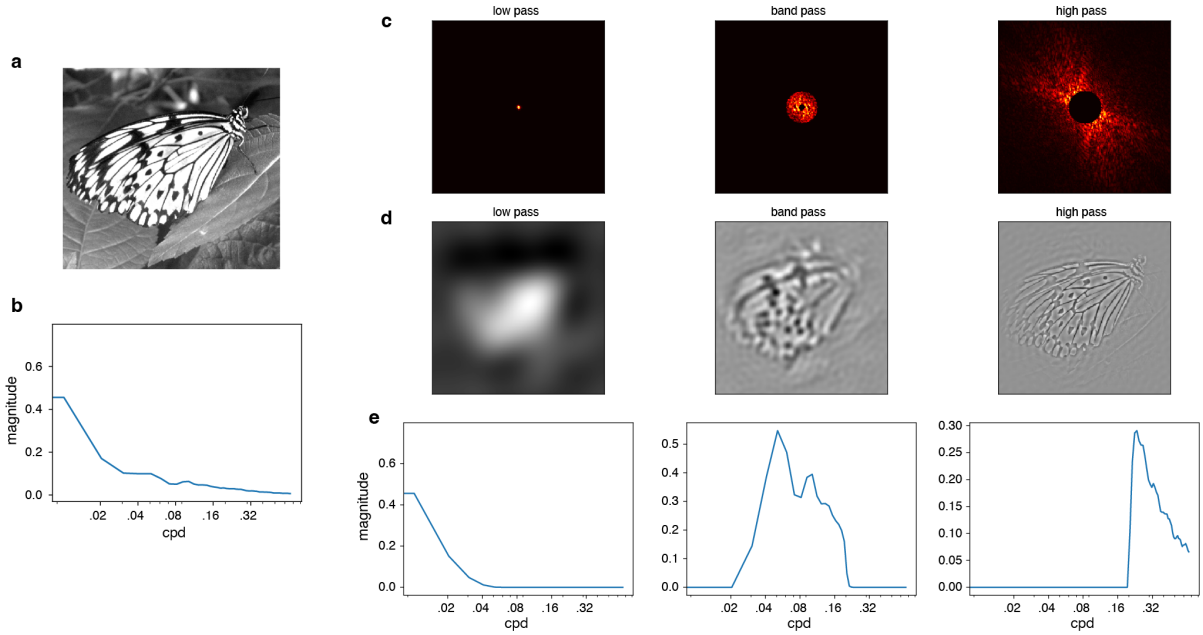

Fig. S7: **Filtering natural scenes.** **a**, An example natural scene and **b** its corresponding spatial frequency spectrum (azimuthal average), in cycles/degree (cpd) of visual angle. **c**, Images were filtered by cropping their spectra in Fourier domain (see Methods). **d** Resulting filtered images and **e** their corresponding spectra (azimuthal averages) after filtering was performed.

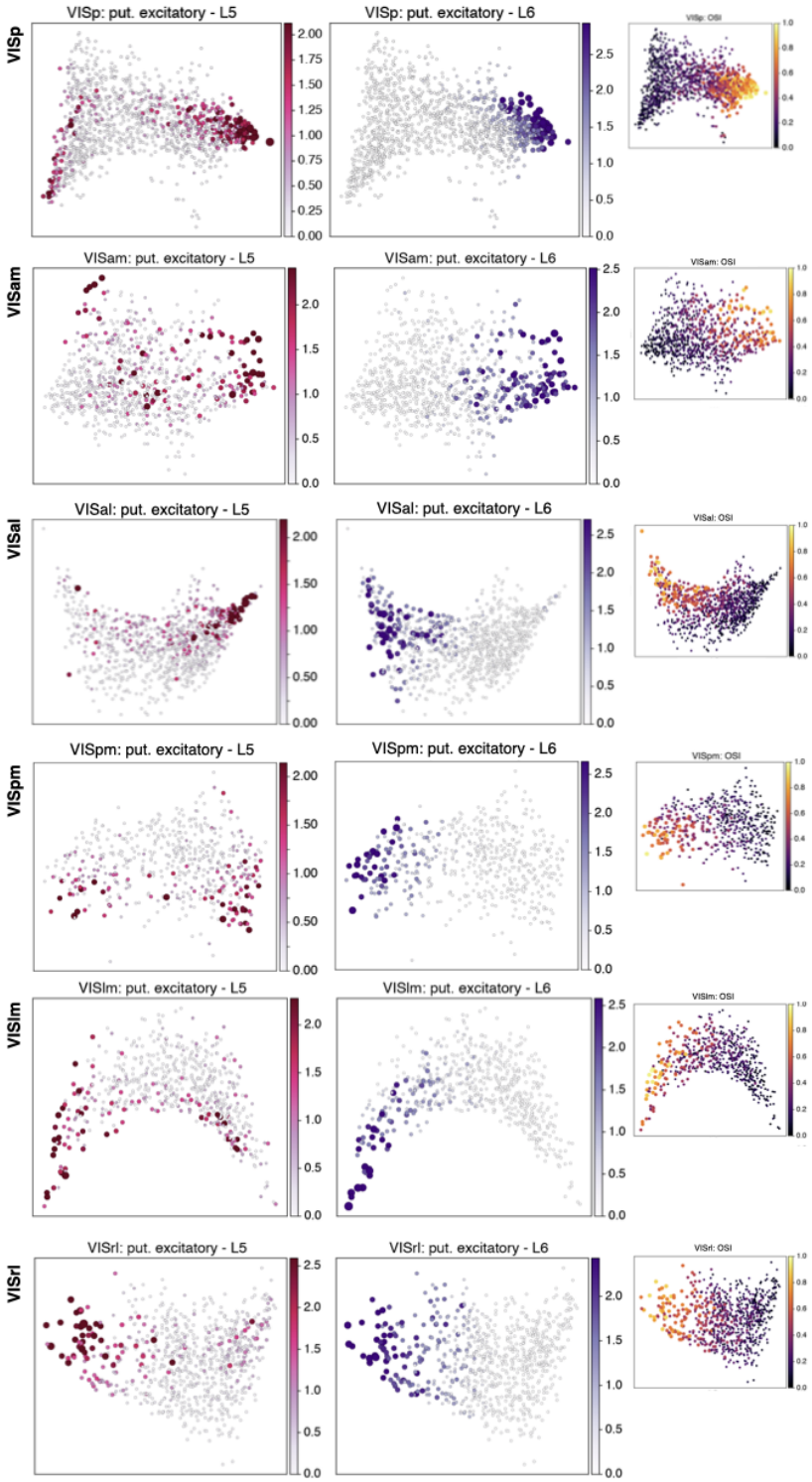

Fig. S8: **Distributions of layers 5 and 6 for putative excitatory neurons.** In VISp, layer 5 (left column) exhibits two distinct concentrations, one of which overlaps almost completely with layer 6 (middle column) in the region of highest orientation selectivity (OSI, shown in inset). Similarly, layer 5 neurons in the VISpm manifold are organized in two clusters, although the overlap of the one cluster with the highly selective layer 6 neurons is less complete. On the other hand, in VISal, layer 5 has practically no overlap with layer 6, even though the latter remains highly selective. Layer 5 in all three cases indicates a bimodal distribution on the respective manifolds, one of which remains consistent with the density of (putative) excitatory neurons in layer 6.

References

[1] E. Acar, D. M. Dunlavy, and T. G. Kolda. A scalable optimization approach for fitting canonical tensor decompositions. *J Chemom*, 25(2):67–86, 2011. doi: 10.1002/cem.1335.

[2] E. A. Allen and R. D. Freeman. Dynamic spatial processing originates in early visual pathways. *Journal of Neuroscience*, 26(45):11763–11774, 2006.

[3] Allen Institute. Visual coding – neuropixels. `https://portal.brain-map.org/circuits-behavior/visual-coding-neuropixels`, . Accessed: 04–Oct–2024.

[4] Allen Institute. Allen brain atlas software development kit (allensdk). `https://allensdk.readthedocs.io/en/latest/visual_coding_neuropixels.html`, . Accessed: 04–Oct–2024.

[5] Allen Institute. Allen Brain Observatory – Neuropixels Visual Coding. `https://brainmapportal-live-4cc80a57cd6e400d854-f7fdcae.divio-media.net/filer_public/80/75/8075a100-ca64-429a-b39a-569121b612b2/neuropixels_visual_coding_-_white_paper_v10.pdf`, 2019. Version 1.0 (October 1, 2019) [Online; accessed 04–Oct–2024].

[6] J.-M. Alonso, W. M. Usrey, and R. C. Reid. Precisely correlated firing in cells of the lateral geniculate nucleus. *Nature*, 383(6603):815–819, 1996.

[7] B. W. Bader and T. G. Kolda. Efficient MATLAB computations with sparse and factored tensors. *SIAM J Sci Comput*, 30(1):205–231, 2008.

[8] A. Blot, M. M. Roth, I. Gasler, M. Javadzadeh, F. Imhof, and S. B. Hofer. Visual intracortical and transthalamic pathways carry distinct information to cortical areas. *Neuron*, 109(12):1996–2008, 2021.

[9] F. Bolaños, J. G. Orlandi, R. Aoki, A. V. Jagadeesh, J. L. Gardner, and A. Benucci. Efficient coding of natural images in the mouse visual cortex. *Nature Communications*, 15(1):2466, 2024.

[10] G. Born, F. A. Schneider-Soupiadis, S. Erisken, A. Vaiceliunaite, C. L. Lao, M. H. Mobarhan, M. A. Spacek, G. T. Einevoll, and L. Busse. Corticothalamic feedback sculpts visual spatial integration in mouse thalamus. *Nature neuroscience*, 24(12):1711–1720, 2021.

[11] Z. I. Botev, J. F. Grotowski, and D. P. Kroese. Kernel density estimation via diffusion. *Ann Stat*, 38 (5):2916–2957, 2010.

[12] C. E. Bredfeldt and D. Ringach. Dynamics of spatial frequency tuning in macaque v1. *Journal of Neuroscience*, 22(5):1976–1984, 2002.

[13] R. Bro and S. De Jong. A fast non-negativity-constrained least squares algorithm. *J Chemom*, 11(5): 393–401, 1997.

[14] S. Chung and L. Abbott. Neural population geometry: An approach for understanding biological and artificial neural networks. *Curr Opin Neurobiol*, 70:137–144, 2021.

[15] A. Cichocki, R. Zdunek, A. H. Phan, and S.-i. Amari. *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons, 2009.

[16] R. R. Coifman and S. Lafon. Diffusion maps. *Appl Comput Harmon Anal*, 21(1):5–30, 2006.

[17] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proc Natl Acad Sci USA*, 102(21):7426–7431, 2005.

[18] S. E. de Vries, J. A. Lecoq, M. A. Buice, P. A. Groblewski, G. K. Ocker, M. Oliver, D. Feng, N. Cain, P. Ledochowitsch, D. Millman, et al. A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. *Nature neuroscience*, 23(1):138–151, 2020.

[19] S. E. de Vries, J. H. Siegle, and C. Koch. Sharing neurophysiology data from the allen brain observatory. *Elife*, 12:e85550, 2023.

[20] A. Dobbins, S. W. Zucker, and M. S. Cynader. Endstopped neurons in the visual cortex as a substrate for calculating curvature. *Nature*, 329(6138):438, 1987.

[21] L. Dyballa and S. W. Zucker. IAN: Iterated Adaptive Neighborhoods for manifold learning and dimensionality estimation. *Neural Comput*, 35, 2023.

[22] L. Dyballa, M. S. Hoseini, M. C. Dadarlat, S. W. Zucker, and M. P. Stryker. Flow stimuli reveal ecologically appropriate responses in mouse visual cortex. *Proc Natl Acad Sci USA*, 115(44):11304–11309, 2018.

[23] L. Dyballa, A. M. Rudzite, M. S. Hoseini, M. Thapa, M. P. Stryker, G. D. Field, and S. W. Zucker. Population encoding of stimulus features along the visual hierarchy. *Proceedings of the National Academy of Sciences*, 121(4):e2317773121, 2024.

[24] E. Froudarakis, P. Berens, A. S. Ecker, R. J. Cotton, F. H. Sinz, D. Yatsenko, P. Saggau, M. Bethge, and A. S. Tolias. Population code in mouse V1 facilitates readout of natural scenes through increased sparseness. *Nat Neurosci*, 17(6):851, 2014.

[25] L. L. Glickfeld and S. R. Olsen. Higher-order areas of the mouse visual cortex. *Annual review of vision science*, 3:251–273, 2017.

[26] Y. Han, J. M. Kebschull, R. A. Campbell, D. Cowan, F. Imhof, A. M. Zador, and T. D. Mrsic-Flogel. The logic of single-cell projections from visual cortex. *Nature*, 556(7699):51–56, 2018.

[27] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant. Array programming with NumPy. *Nature*, 585(7825): 357–362, Sept. 2020. doi: 10.1038/s41586-020-2649-2. URL `https://doi.org/10.1038/s41586-020-2649-2`.

[28] J. A. Harris, S. Mihalas, K. E. Hirokawa, J. D. Whitesell, H. Choi, A. Bernard, P. Bohn, S. Caldejon, L. Casal, A. Cho, et al. Hierarchical organization of cortical and thalamic connectivity. *Nature*, 575 (7781):195–202, 2019.

[29] J. Hegdé. Time course of visual perception: Coarse-to-fine processing and beyond. *Progress in Neurobiology*, 84:405–439, 2008. URL `https://api.semanticscholar.org/CorpusID:900253`.

[30] J. Hegdé. Time course of visual perception: coarse-to-fine processing and beyond. *Progress in neurobiology*, 84(4):405–439, 2008.

[31] J. L. Hoy, I. Yavorska, M. Wehr, and C. M. Niell. Vision drives accurate approach behavior during prey capture in laboratory mice. *Curr Biol*, 26(22):3046–3052, 2016.

[32] S. Katzner, G. Born, and L. Busse. V1 microcircuits underlying mouse visual behavior. *Current opinion in neurobiology*, 58:191–198, 2019.

[33] A. J. Keller, M. M. Roth, and M. Scanziani. Feedback generates a second receptive field in neurons of the visual cortex. *Nature*, 582(7813):545–549, 2020.

[34] A. M. Kerlin, M. L. Andermann, V. K. Berezovskii, and R. C. Reid. Broadly tuned response properties of diverse inhibitory neuron subtypes in mouse visual cortex. *Neuron*, 67(5):858–871, 2010.

[35] E. J. Kim, A. L. Juavinett, E. M. Kyubwa, M. W. Jacobs, and E. M. Callaway. Three types of cortical layer 5 neurons that differ in brain-wide connectivity and function. *Neuron*, 88(6):1253–1267, 2015.

[36] M.-H. Kim, P. Znamenskiy, M. F. Iacaruso, and T. D. Mrsic-Flogel. Segregated subnetworks of intracortical projection neurons in primary visual cortex. *Neuron*, 100(6):1313–1321, 2018.

[37] L. Kirchberger, S. Mukherjee, U. H. Schnabel, E. H. van Beest, A. Barsegyan, C. N. Levelt, J. A. Heimel, J. A. Lorteije, C. van der Togt, M. W. Self, et al. The essential role of recurrent processing for figure-ground perception in mice. *Science advances*, 7(27):eabe1833, 2021.

[38] M. A. Kirchgessner, A. D. Franklin, and E. M. Callaway. Distinct "driving" versus "modulatory" influences of different visual corticothalamic pathways. *Current Biology*, 31(23):5121–5137, 2021.

[39] T. G. Kolda. Multilinear operators for higher-order decompositions. Technical Report SAND2006-2081, Sandia National Laboratories, April 2006. URL `http://www.osti.gov/scitech/biblio/923081`.

[40] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

[41] F. J. Luongo, L. Liu, C. L. A. Ho, J. K. Hesse, J. B. Wekselblatt, F. F. Lanfranchi, D. Huber, and D. Y. Tsao. Mice and primates use distinct strategies for visual segmentation. *Elife*, 12:e74394, 2023.

[42] J. Manley, S. Lu, K. Barber, J. Demas, H. Kim, D. Meyer, F. M. Traub, and A. Vaziri. Simultaneous, cortex-wide dynamics of up to 1 million neurons reveal unbounded scaling of dimensionality with neuron number. *Neuron*, 2024.

[43] D. Marr and T. Poggio. A computational theory of human stereo vision. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 204(1156):301–328, 1979.

[44] J. H. Marshel, M. E. Garrett, I. Nauhaus, and E. M. Callaway. Functional specialization of seven mouse visual cortical areas. *Neuron*, 72(6):1040–1054, 2011.

[45] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings of the 8th IEEE International Conference on Computer Vision (ICCV 2001)*, volume 2, pages 416–423. IEEE, 2001.

[46] J. A. Mazer, W. E. Vinje, J. McDermott, P. H. Schiller, and J. L. Gallant. Spatial frequency and orientation tuning dynamics in area v1. *Proceedings of the National Academy of Sciences*, 99(3): 1645–1650, 2002.

[47] L. McInnes, J. Healy, N. Saul, and L. Großberger. UMAP: Uniform manifold approximation and projection. *Journal of Open Source Software*, 3(29):861, 2018. doi: 10.21105/joss.00861.

[48] J. S. Montijn, P. M. Goltstein, and C. M. Pennartz. Mouse v1 population correlates of visual detection rely on heterogeneity within neuronal response patterns. *Elife*, 4:e10163, 2015.

[49] P. Murphy and A. Sillito. Corticofugal feedback influences the generation of length tuning in the visual pathway. *Nature*, 329(6141):727–729, 1987.

[50] C. M. Niell and M. Scanziani. How cortical circuits implement cortical computations: mouse visual cortex as a model. *Annual Review of Neuroscience*, 44:517–546, 2021.

[51] C. M. Niell and M. P. Stryker. Highly selective receptive fields in mouse visual cortex. *J Neurosci*, 28 (30):7520–7536, 2008.

[52] T. Odland. KDEpy: Kernel Density Estimation in Python, Dec. 2018. v0.9.10. Available at https://doi.org/10.5281/zenodo.2392268.

[53] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision*, 42:145–175, 2001.

[54] A. Olmos and F. A. Kingdom. A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception*, 33(12):1463–1473, 2004.

[55] R. V. Rikhye and M. Sur. Spatial correlations in natural scenes modulate response reliability in mouse visual cortex. *J Neurosci*, 35(43):14661–14680, 2015.

[56] P. G. Schyns and A. Oliva. From blobs to boundary edges: Evidence for time-and spatial-scale-dependent scene recognition. *Psychological science*, 5(4):195–200, 1994.

[57] J. H. Siegle, X. Jia, S. Durand, S. Gale, C. Bennett, N. Graddis, G. Heller, T. K. Ramirez, H. Choi, J. A. Luviano, et al. Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature*, pages 1–7, 2021.

[58] A. M. Sillito, H. E. Jones, G. L. Gerstein, and D. C. West. Feature-linked synchronization of thalamic relay cell firing induced by feedback from the visual cortex. *Nature*, 369(6480):479–482, 1994.

[59] R. Skyberg, S. Tanabe, H. Chen, and J. Cang. Coarse-to-fine processing drives the efficient coding of natural scenes in mouse visual cortex. *Cell reports*, 38(13), 2022.

[60] R. J. Skyberg and C. M. Niell. Natural visual behavior and active sensing in the mouse. *Current Opinion in Neurobiology*, 86:102882, 2024.

[61] I. T. Smith, L. B. Townsend, R. Huh, H. Zhu, and S. L. Smith. Stream-dependent development of higher visual cortical areas. *Nature neuroscience*, 20(2):200–208, 2017.

[62] C. Stringer, M. Michaelos, D. Tsyboulski, S. E. Lindo, and M. Pachitariu. High-precision coding in visual cortex. *Cell*, 184(10):2767–2778, 2021.

[63] E. Switkes, M. J. Mayer, and J. A. Sloan. Spatial frequency analysis of the visual environment: Anisotropy and the carpentered environment hypothesis. *Vision research*, 18(10):1393–1399, 1978.

[64] L. Van der Maaten and G. Hinton. Visualizing data using t-sne. *J Mach Learn Res*, 9(11), 2008.

[65] J. H. Van Hateren and A. van der Schaaf. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1394):359–366, 1998.

[66] M. Vélez-Fort, C. V. Rousseau, C. J. Niedworok, I. R. Wickersham, E. A. Rancz, A. P. Brown, M. Strom, and T. W. Margrie. The stimulus selectivity and connectivity of layer six principal cells reveals cortical microcircuits underlying visual processing. *Neuron*, 83(6):1431–1443, 2014.

[67] S. Vreysen, B. Zhang, Y. M. Chino, L. Arckens, and G. Van den Bergh. Dynamics of spatial frequency tuning in mouse visual cortex. *Journal of neurophysiology*, 107(11):2937–2949, 2012.

[68] H. Wang, O. Dey, W. N. Lagos, and E. M. Callaway. Diversity in spatial frequency, temporal frequency, and speed tuning across mouse visual cortical areas and layers. *Journal of Comparative Neurology*, 530(18):3226–3247, 2022.

[69] Q. Wang, O. Sporns, and A. Burkhalter. Network analysis of corticocortical connections reveals ventral and dorsal processing streams in mouse visual cortex. *Journal of Neuroscience*, 32(13):4386–4399, 2012.

[70] F. A. Wichmann, J. Drewes, P. Rosas, and K. R. Gegenfurtner. Animal detection in natural scenes: Critical features revisited. *Journal of Vision*, 10(4):6–6, 2010.

[71] A. H. Williams, T. H. Kim, F. Wang, S. Vyas, S. I. Ryu, K. V. Shenoy, M. Schnitzer, T. G. Kolda, and S. Ganguli. Unsupervised discovery of demixed, low-dimensional neural dynamics across multiple timescales through tensor component analysis. *Neuron*, 98(6):1099–1115, 2018.

[72] Y. Yu, J. N. Stirman, C. R. Dorsett, and S. L. Smith. Selective representations of texture and motion in mouse higher visual areas. *Current Biology*, 32(13):2810–2820, 2022.

[73] Y. Yu, J. N. Stirman, C. R. Dorsett, and S. L. Smith. Visual information is broadcast among cortical areas in discrete channels. *bioRxiv*, pages 2023–12, 2023.

[74] H. Zhou, R. J. Schafer, and R. Desimone. Pulvinar-cortex interactions in vision and attention. *Neuron*, 89(1):209–220, 2016.