Natural images are reliably represented by sparse and variable populations of neurons in visual cortex

Takashi Yoshida1,2,3 & Kenichi Ohki1,2,3,4

Natural scenes sparsely activate neurons in the primary visual cortex (V1). However, how sparsely active neurons reliably represent complex natural images and how the information is optimally decoded from these representations have not been revealed. Using two-photon calcium imaging, we recorded visual responses to natural images from several hundred V1 neurons and reconstructed the images from neural activity in anesthetized and awake mice. A single natural image is linearly decodable from a surprisingly small number of highly responsive neurons, and the remaining neurons even degrade the decoding. Furthermore, these neurons reliably represent the image across trials, regardless of trial-to-trial response variability. Based on our results, diverse, partially overlapping receptive fields ensure sparse and reliable representation. We suggest that information is reliably represented while the corresponding neuronal patterns change across trials and collecting only the activity of highly responsive neurons is an optimal decoding strategy for the downstream neurons.

1 Department of Physiology, The University of Tokyo School of Medicine, Tokyo, Japan. 2 Department of Molecular Physiology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. 3 CREST, Japan Science and Technology Agency, Tokyo, Japan. 4 International Research Center for Neurointelligence (WPI-IRCN), The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. Email: takashiy@m.u-tokyo.ac.jp, kohki@m.u-tokyo.ac.jp
Sensory inputs sparsely activate neurons in the sensory cortex\textsuperscript{1-7}, which is postulated to be an efficient information coding\textsuperscript{10}. However, how sparsely active neurons represent information and how the information is optimally decoded from the sparse representation have not been determined.

In the primary visual cortex (V1), simple cells each have a receptive field (RF) structure that is modeled by a two-dimensional (2D) Gabor function\textsuperscript{11}. Theoretically, a single natural scene is represented by a relatively small number of neurons using Gabor-like RFs, whereas information about multiple scenes is distributed among the neuronal population\textsuperscript{10,12,13}. Indeed, V1 neurons sparsely respond to natural scenes\textsuperscript{2,3,5,14}. Population activity with higher sparseness exhibits greater discriminability between natural scenes\textsuperscript{5}.

What types of information from natural scenes are represented in the sparsely active neurons in the brain? The visual contents of natural scenes are reconstructed from single-unit activity within the lateral geniculate nucleus collected from several experiments\textsuperscript{15} and functional magnetic resonance imaging data from the visual cortex\textsuperscript{16-19}. However, whether the visual contents of natural images could be represented by simultaneously recorded, sparsely active neurons has not been addressed experimentally. It also remains to be revealed which decoding strategy is the most optimal for sparse representation. Whether only the small number of strongly active neurons represent information, or whether the remaining neurons contribute additional information remains to be determined. Furthermore, do the sparsely active neurons reliably represent the natural image contents, regardless of trial-to-trial response variability? Although a computational model\textsuperscript{20} has suggested that sparse and overcomplete representation is optimal for natural image representation with unreliable neurons, this model has not been examined experimentally.

We also address how visual information is distributed in a neuronal population. Some neurons are “unresponsive” to simple visual stimuli, such as moving gratings (e.g., the response rate of the mouse V1 is 26–68\%\textsuperscript{21-27}. However, this may be partly because visual stimuli are not complex enough to completely cover the RF properties of all neurons. The proportion of neurons involved in the information processing is a matter of debate\textsuperscript{28,29}.

Here, we examined how sparsely active neurons represent natural image contents and how the information was optimally decoded from this sparse representation. Using two-photon calcium (Ca\textsuperscript{2+}) imaging, we recorded visual responses to natural images from the V1 of anesthetized and passively viewing awake mice. A small percentage of neurons responded to each image, which was sparser than that predicted by a linear encoding model. On the other hand, most neurons were involved in processing of natural images. The visual contents of a natural image were linearly decodable from a small number of highly responsive neurons, and the additional use of the remaining neurons even degraded the reconstruction performance. The responsive neurons reliably represent the image, regardless of trial-to-trial response variability, which was supported by multiple neurons with partially overlapping representation. These results revealed a new representation of a natural image by a small number of neurons in which information is reliably represented, while the corresponding neuronal patterns change across trials and implies that collecting only the activity of highly responsive neurons is an optimal decoding strategy for sparse representation. The preliminary results of this study have been published in abstract form\textsuperscript{30} and on a preprint server\textsuperscript{31}.

Results
Sparse visual responses to natural images. We presented flashes of natural images as visual stimuli (200 images for main dataset, dataset 1; 1000–2000 images for another dataset, dataset 2). Each image was consecutively flashed three times in a trial (three 200 ms presentations interleaved with 200 ms of a gray screen) and presented at least 12 times for dataset 1 and three to eight times for dataset 2 in a recording session (Fig. 1a, Methods section). Using two-photon Ca\textsuperscript{2+} imaging, we recorded single neurons’ activity from anesthetized mouse V1 (560 (284–712) cells/plane, median (25–75th percentiles); n = 24 planes from 14 mice for dataset 1, n = 4 planes from three mice for dataset 2, 260–450 microns in depth; see Fig. 1b for representative response traces). Results described below were obtained from dataset 1 unless otherwise stated.

Most neurons in a local population were visually responsive. Responsive neurons were determined by one-way analysis of variance (ANOVA, p < 0.01; data with responses to 200 images and one baseline in each trial, see Methods section). Across planes, 85\% (77–90\%) of neurons were responsive (Fig. 1d). Using label-shuffled data, we found a 1.0\% (0.9–1.0\%) false positive rate. Thus, most neurons were involved in visual processing of natural images.

We next examined the percentage of neurons that responded to each image. For each responsive neuron determined by ANOVA, a significant response to each image was defined by a t-test and amplitude threshold (p < 0.01, using a t-test and >10\% trial-averaged visual response; see Methods section and Supplementary Fig. 1). Figure 1c presents plots of significant visual response events for all images (x-axis) across all neurons (y-axis) in an example plane (n = 726 cells, depth: 360 microns from the brain surface). Across planes, 2.5\% (1.8–3.0\%) of neurons were responsive for each image (Fig. 1g, Supplementary Fig. 1; Fig. 1c bottom for the example case). This low response rate was not due to poor recording conditions. The same neurons responded well to moving gratings (49\% (45–56\%) of cells were responsive for at least one grating direction, and 12.5\% (11–14.8\%) responded to each direction of grating, Fig. 1h, 1i).

The responsive neurons only slightly overlapped between images. In examples for three natural images (Fig. 1d), each image activated different subsets of neurons that exhibited small overlaps between images (Fig. 1d, right column). Of the responsive cells, 5.4\% (4.8–6.0\%) exhibited overlap between two images (Fig. 1j). Furthermore, only a small number of neurons exhibited visual responses with large amplitudes, which is a property of sparse representations (Fig. 1e). The population sparseness\textsuperscript{2,3} was comparable to that from a previous report on mouse V1 (ref. 5; 0.36 (0.30–0.42), Fig. 1k, see Methods section). Together, each natural image activated a small number of neurons, whereas most neurons were visually responsive. The latter result represents the first report of the visual responsiveness of most neurons in mouse V1 to natural image stimuli\textsuperscript{28,29}.

Partially overlapping representations of visual features. We constructed an encoding model for the visual responses of an individual neuron to examine the visual features represented by each neuron. To extract the visual features from the natural images, we used a set of Gabor wavelet filters with self-inverting properties (1248 filters, Fig. 2a, b, see Methods section). Natural images (I) were subjected to Gabor filters (G\textsubscript{Gmvd}) and transformed into sets of Gabor feature values (F, see Methods section).

$$F = G_{Gmvd} I$$

(1)

For each neuron, we first selected the Gabor features that exhibited strong correlations with the visual response. The correlation threshold for the selected feature was adjusted to maximize the visual response prediction in each neuron (Fig. 2c, Supplementary Fig. 2a, b). Then, the visual responses of a single
cell \( R_k \), the \( k \)th cell's responses) were represented by a linear regression of the selected feature values \( F_{\text{select}} \) followed by nonlinear scaling (NL), Fig. 2c, see Methods section.

\[
R_k = \text{NL}(W_k F_{\text{select}} + b_k)
\] (2)

Parameters of the regression model (weights, \( W_k \) and a bias, \( b_k \)) were estimated with 90% of the dataset, and the prediction performance was estimated with the remaining 10% of the dataset (ten-fold cross-validation (CV)), where \( W_k \) and \( b_k \) were estimated in each CV. A model was obtained independently for each cell (i.e., for each \( k \)).
In example neurons (Fig. 2d, e), the response prediction performances (the correlation coefficients between the observed and the predicted responses) were 0.82 and 0.88. These neurons were represented by 12 (out of 1248) Gabor features (Fig. 2d, e, right panels). We defined a forward filter as a weighted sum of the Gabor filters; the forward filters of the example neurons were spatially localized (Fig. 2d, e, insets in the right panels).

The median prediction performance of the encoding model was 0.32 (0.15–0.50) in the example plane presented in Fig. 1 (median (25–75th percentiles); \( n = 726 \) cells) and 0.21 (0.06–0.42) in all cells across planes (\( n = 12,755 \) cells, Supplementary Fig. 2d). The nonlinear scaling suppressed weak predicted responses and enhanced strong predicted responses (Fig. 2f, g), suggesting that this nonlinear step enhanced the sparseness of the predicted response obtained from the linear step. The nonlinear step also slightly enhanced the performance (Supplementary Fig. 2c).

Individual neuron encoded a small number of Gabor features. On average, each neuron encoded 21 (8–51) features in all recorded neurons (\( n = 12,755 \) cells, Fig. 2h, i, Supplementary Fig. 2f).
The Gabor features recorded by each neuron were spatially localized and had similar orientations (Supplementary Fig. 3a–d). The regression weights of the Gabor features in the encoding model were positively correlated with the similarity between the corresponding Gabor filter and the RF structure estimated using the regularized inverse method (see Methods section, Supplementary Fig. 3e–h).32–34

The Gabor features that overlapped between neurons (Fig. 2j). In the example neurons (Fig. 2d, e), none of the features overlapped. For all neuron pairs, the median overlap was 1.0% (0.0–5.6%) (relative to the features represented by each cell; Fig. 2j, Supplementary Fig. 2f). The feature overlap between neurons was positively correlated with the similarity of the forward filter (pixel-to-pixel correlation between forward filters; Supplementary Fig. 3i, j). Furthermore, the features encoded by neuronal populations were diverse, which was reflected in the distribution of forward filter similarity between cells (correlation coefficient between forward filters: 0.0 (−0.03–0.04); Supplementary Fig. 3k). Thus, the Gabor features encoded by individual neurons were highly diverse and only slightly overlapped.

This analysis also revealed how the individual Gabor features were encoded across neurons (Fig. 2h). As the spatial frequency (SF) of the Gabor filter increased (i.e., the scale decreased), the corresponding feature contributed to the visual responses of fewer neurons (Fig. 2h bottom), reflecting that Gabor filters with a low SF covered more of the neuron’s RF, whereas Gabor filters with a high SF affected the responses of fewer neurons. Furthermore, almost all features contributed to the responses of at least one cell (Supplementary Fig. 2g).

Image reconstruction from the population activity. We next examined whether the features encoded in a neural population represent the visual contents of the natural images. To examine this, we reconstructed stimulus images from neuronal activity.13–19

In the image reconstruction model, each Gabor feature value ($F^i$, the $j$th Gabor feature value) was subjected to a linear regression of the activity of multiple neurons ($\mathbf{R}$) with model parameters of weights ($\mathbf{H}^j$) and a bias term ($c^j$; Fig. 3a, b):

$$F^i = \mathbf{H}^j \mathbf{R} + c^j \quad (3)$$

Each Gabor feature value was independently reconstructed (i.e., $\mathbf{H}^j$ and $c^j$ were estimated for each feature, $j$). Then, the sets of reconstructed feature values ($F = [F^1, \ldots, F^{128}]$; $F^i$: reconstructed $j$th feature values) were transformed into images ($\mathbf{I}$) based on almost self-inverting property of the Gabor filters (Gabor filter matrix for image reconstruction: $\mathbf{G}^\text{rev}$; Figs. 2a and 3b, see Methods section).

$$\mathbf{I} = \mathbf{G}^\text{rev} \mathbf{F} \quad (4)$$

The reconstruction performance was estimated with a different test dataset from the training dataset used in the regression analysis (ten-fold CV with the same data split as in the encoding model, $\mathbf{H}^j$ and $c^j$ were estimated in each CV).

We first used a model in which each feature value was reconstructed from all neurons (all-cell model, Fig. 3a). In the example plane ($n = 726$ neurons, presented in Figs. 1 and 2), the robust structures of the stimulus images were reconstructed from the population activity (Fig. 3c). The pixel-to-pixel correlation between stimulus and reconstructed images (similarity of image patterns) was 0.43 (0.35–0.55) (median (25–75th percentiles) of 200 images (dataset 1) in the example plane (Fig. 3e upper panel) and 0.33 (0.30–0.37) across all planes ($n = 24$ planes, Fig. 3f). The coefficient of determination (CD, goodness of model prediction, see Methods section) was 0.14 (0.02–0.26) in the example plane and 0.08 (0.07–0.1) across planes (Fig. 3e, bottom, and 3g). We applied the same method for another dataset (dataset 2, 1000–2000 images that did not include original 200 images). The performances of dataset 2 were similar to those of the original dataset (R: 0.45 (0.34–0.52), CD: 0.19 (0.11–0.25), $n = 4$ planes from three mice (Fig. 3d, green lines in Fig. 3f, g). Thus, the visual contents of natural images were extracted linearly from the population activity.

We next used another reconstruction model, in which each feature was reconstructed from a subset of neurons that showed strong correlations between the feature values and responses (cell-selection model, Fig. 3a). In the cell-selection model, each neuron participated in the reconstructions of subsets of features that the neuron encoded in the response prediction model (Fig. 3b, see Methods section). The reconstruction performances of the cell-selection model were almost comparable or even slightly higher than those of the all-cell model (dataset 1, R: 0.33 (0.3–0.38), $p = 0.006$ by signed-rank test compared to all-cell model. CD: 0.10 (0.08–0.14), $p = 1.8 \times 10^{-5}$ by signed-rank test. Dataset 2, R: 0.45 (0.33–0.53), $p = 0.9$ by signed-rank test compared to all-cell model. CD: 0.19 (0.11–0.25), $p = 0.9$ by signed-rank test, Fig. 3c–g). We confirmed that the performances were not substantially affected by whether the test data were used for the cell-selection step or not (Supplementary Fig. 4, see Methods section).

In the cell-selection model, we changed the number of features for which each neuron participated in the reconstruction by manipulating the feature-selection threshold (Figs. 2c and 3a, Supplementary Fig. 5a) and examined how the reconstruction performance was affected by the number of features. The number of features of the original cell-selection model led to nearly
optimal performances (Supplementary Fig. 5a–c). Thus, the cell-selection model captures nearly optimal feature-cell assignment for the reconstruction.

V1 neurons mainly represented low SF components of natural images. Reconstructions of lower SF components (0.02 and 0.04 cycle/degree, cpd) were better than those of higher spatial frequencies (0.09 and 0.18 cpd; R: 0.65, 0.53, 0.18, and 0.04, and CD: 0.24, 0.23, 0.02, and −0.03 for 0.02, 0.04, 0.09, and 0.18 cpd, respectively; Supplementary Fig. 5d–f). This may reflect the SF preferences of mouse V1 neurons35.
A small number of neurons represents a natural image. We next examined whether a small number of responsive neurons mainly represented an image or whether the remaining neurons also contained information. For this purpose, we changed the number of neurons used in the reconstruction of each image and examined how the reconstruction performance was affected. In the image reconstruction from a subset of cells, the weights of the image reconstruction model were based on the weights of the cell-selection model. Because a different set of cells was used for each image, the weights were separately computed for each image (see Eq. (9) in Methods section).

A small number of responsive neurons primarily represented a natural image. In each example image, neurons were sorted by visual response amplitude (descending order) first among the responsive neurons (red dots in Fig. 4a–c) and then among the remaining neurons (black dots in Fig. 4a–c). The image was reconstructed by the top N neurons (N = 1–726 cells), and the reconstruction performances were plotted against the number of neurons used (Fig. 4a–e). On average, ~20 neurons reconstructed the images with a level of peak performance (Fig. 4d, e). The reconstruction performances of the responsive neurons were greater than those of all cells (left panels in Fig. 4g, h). Furthermore, among the responsive neurons for each image, the peak performances were obtained with highly responsive neurons that was fewer than the number of responsive neurons (right panels in Fig. 4g, h). Therefore, highly responsive neurons mainly represent the image, and the additional use of the remaining neurons even decreases the performance.

The features represented by individual neurons should be diverse in order to represent a natural image using a small number of neurons. Figure 4f illustrates how individual responsive neurons contribute to image reconstruction in the case presented in Fig. 4a. Each neuron had a specific pattern of contributions (reverse filter: sum of Gabor filters x weights), and the reverse filters varied across neurons (Fig. 4f, top panels) while partially overlapping in the visual field. For neuron pairs that were responsive to the same image, the number of overlapping Gabor features slightly increased compared to all pairs, but the percentage of overlapping features was still <5% (responsive cell pairs: 3.2% (0–13%) for all pairs and 4.3% (3.4–5.5%) across planes; all-cell pairs: 1.0% (0–5.6%) for all pairs and 1.0% (0.9–1.2%) across planes; Fig. 4i–k, cf. Fig. 2j, Supplementary Fig. 2f). These slightly overlapping and diverse features among neurons should be useful for image representation by a small number of neurons.

Robust image representation by overlapping representation. We next examined whether a single image was robustly represented by a small number of responsive neurons. We computed the reconstruction performance after dropping individual responsive neurons (Fig. 5a, b; the cell # on the x-axis is the same as in Fig. 4d). Dropping a single cell reduced the reconstruction performance by ~5% for the best-responding neurons and did not change the performance for most neurons (middle panels in Fig. 5a, b). Thus, an image was robustly represented by responsive neurons against dropping a single cell.

This robustness against cell drop was due to the spatial overlap of reverse filters among responsive neurons (Fig. 5c). To analyze this, we selected a set of overlapping cells for each responsive neuron; the overlapping cells consisted of one responsive neuron as a reference (cell 1) and the responsive neurons whose representation areas partially overlapped with that of the reference cell (top panels in Fig. 5c and Supplementary Fig. 6a–d, see Methods section). The maximal reverse filter similarity between a responsive cell and other overlapping cells was 0.57 (0.41–0.70) (Supplementary Fig. 6d), indicating that each responsive neuron often had at least one neuron with a similar reverse filter. Then, we examined the effects of cell drop on the reconstruction of a local part of the image from the overlapping cells. Although dropping a single cell had almost no effect on the reconstructed local image (bottom panels in Fig. 5c), a sequential dropping of these cells gradually degraded the part of the image (Fig. 5d). The pixel-to-pixel correlation between the stimulus and the reconstructed image in the overlapping area gradually decreased as the number of dropped cells increased (Fig. 5e, f). Thus, the robust image representation was due to neurons with spatially overlapping representations.

Reliable image representation across trials. We further analyzed whether the overlapping representation was useful in reducing the trial-to-trial variability in the image representation. Cortical neurons often show trial-to-trial response variability. However, the integration of activity among neurons with spatially overlapping or similar representations should reduce the variability in image representations across trials36–38.

The pattern of the reconstructed image was reliably represented across trials regardless of the trial-to-trial response variability. In the example case (shown in Fig. 5), single-trial reconstructed images from all responsive neurons (55 cells) were generally stable across trials and blurred in only a few trials (e.g., trial 10, Fig. 6a, top). In contrast, the activity patterns of responsive neurons seemed variable across trials, although a few cells were reliably active (Fig. 6a, bottom, b). To evaluate the trial-to-trial reliability of the reconstructed images (or population responses), we used two measures, across-trial similarity and across-trial variability. Across-trial similarity was defined as the Pearson’s correlation between single-trial reconstructed images (or response patterns) and their trial average (Fig. 6c). Across-
trial variability was the normalized squared error between single-trial reconstructed images (or response patterns) and their trial average (Fig. 6d, see Methods section). For all planes, the across-trial similarity of the reconstructed images was high (0.85 (0.81–0.89), Fig. 6c, left), whereas the across-trial similarity of the response was relatively low (0.40 (0.36–0.44), Fig. 6c, right). The across-trial variability of the image was relatively low (Fig.6d, left), and that of the response pattern was relatively high (Fig. 6d, right).

We next examined how the images were reliably represented across trials. Neurons with partially overlapping or similar reverse filters will provide reliable representations across trials (for at least part of the image), if these neurons are active on different trials. In the example case, among the nine overlapping cells, the
Supplementary Fig. 7d) and with signal correlations were positively correlated with reverse for the calculation, Supplementary Fig. 7a. The reconstructed image gradually decreased as the number representation. We also analyzed whether reverse filters encoded by the responsive neurons can represent any natural image as basis functions. If so, a set of features of the natural image will be accurately represented by a linear regression of the weights (i.e., feature sets) of responsive neurons in the cell-selection model independent of actual responses (see Methods section, Fig. 7e). We also separately analyzed the reconstruction performance from excitatory and inhibitory cells. In some experiments, we used the mice that expressed tdTomato in inhibitory cells (gad2-cre x Ai14)42,43 and identified excitatory and inhibitory cells based on the tdTomato (n = 5 planes from two mice). Excitatory cells were more visually responsive than inhibitory cells (Supplementary Fig. 10a–c). Furthermore, reconstruction

three example neurons (cell 1–3 in Fig. 6e) represented a partially similar pattern of the image (the correlations of reverse filters among the three neurons were 0.31–0.68), while they were active on different trials. Combining the nine overlapping cells reliably represent a local part of the image in most trials (Fig. 6e, bottom). In the example case, the across-trial variability of the local part of the reconstructed image gradually decreased as the number of cells used for the reconstruction increased (Fig. 6f). This was also observed in data from all planes (Fig. 6g, black lines), suggesting that multiple overlapping cells support reliable image representation.

We further examined the relationship between reverse filter overlap or similarity and noise correlations (see Methods section for the calculation, Supplementary Fig. 7a–c). Noise correlations were positively correlated with reverse filter correlations (correlation: 0.25, Supplementary Fig. 7d) and with signal correlations (correlation: 0.44, Supplementary Fig. 7e)37,39. However, the noise correlations were mainly distributed near 0 even in cell pairs with high overlap or reverse filter similarity (i.e., in pairs where the correlation of the reverse filter >0.5, the median noise correlation was 0.03; Supplementary Fig. 7b, c). Together, neurons with overlapping or similar reverse filters were almost independently active across trials, which was useful for reliable representation.

We also analyzed the effect of noise correlations on the reliable image representation. Removing the noise correlations (see Methods section) increased the reconstruction performances and decreased the across-trial variability (Supplementary Fig. 7g–i), suggesting that the noise correlations are rather detrimental to both reconstruction performance and the reliable image representation. In the relationship between the across-trial variability and the number of cells used for the reconstruction, the reduction in variability by removing noise correlations was much smaller than that by increasing the number of cells used for the reconstruction (Fig. 6g), suggesting that noise correlations have a relatively small impact on the reliable representation.

Representation of multiple images in a population. We next examined how multiple natural images were represented in a population of responsive neurons (Fig. 7a–d). In the example plane (n = 726 cells), natural images were sorted by reconstruction performance (y-axis in Fig. 7a, b), and the cells responding to each image are plotted in each row. First, as the number of images increased, new responsive cells were added, and the total number of responsive cells used for the reconstructions quickly increased (right end of the plot on each row, Fig. 7b). At ~50 images, the number of newly added responsive cells quickly decreased, and the increase in the total number of responsive cells slowed, indicating that the newly added image was represented by a combination of the already plotted responsive cells (i.e., neurons that responded to other images), due to the small overlap in responsive cells between images (Fig. 1j). In summary, the number of newly added cells quickly decreased to zero as the number of images increased (red lines in Fig. 7c, d). Therefore, although only ~5% of responsive neurons overlapped between images (Fig. 1j), this small overlap is useful for the representation of many natural images by a limited number of responsive neurons.

We also analyzed whether reverse filters encoded by the responsive neurons can represent any natural image as basis functions. If so, a set of features of the natural image will be accurately represented by a linear regression of the weights (i.e., feature sets) of responsive neurons in the cell-selection model independent of actual responses (see Methods section, Fig. 7e). We also separately analyzed the reconstruction performance from excitatory and inhibitory cells. In some experiments, we used the mice that expressed tdTomato in inhibitory cells (gad2-cre x Ai14)42,43 and identified excitatory and inhibitory cells based on the tdTomato (n = 5 planes from two mice). Excitatory cells were more visually responsive than inhibitory cells (Supplementary Fig. 10a–c). Furthermore, reconstruction

Image representation in awake mice. We finally examined whether the results obtained from anesthetized mice could be generalized to awake mice. We performed an additional experiment with awake, passively viewing mice that expressed GCaMP6s in the cortical cells (see Methods section)40,41. We also separately analyzed the reconstruction performance from excitatory and inhibitory cells. In some experiments, we used the mice that expressed tdTomato in inhibitory cells (gad2-cre x Ai14)42,43 and identified excitatory and inhibitory cells based on the tdTomato (n = 5 planes from two mice). Excitatory cells were more visually responsive than inhibitory cells (Supplementary Fig. 10a–c). Furthermore, reconstruction

This schematic of the analysis. j Percentages of overlapping features for all-cell pairs responding to the same image. k The median percentages of overlapping features for all planes. Each dot indicates the median in each plane. d, e, g, h, k N = 24 planes. The stimulus images in a–c are adapted from the databases in ref. 56 and with permission. Source data are provided as a Source Data file.
performances by inhibitory cells were low compared to those only by excitatory cells and by all cells, whereas performances by excitatory cells were comparable to those by all cells (Supplementary Fig. 10d, e). Therefore, the images are mainly represented by excitatory cells at least under our experimental conditions.

We further examined how image reconstruction was affected by the locomotion state, which modulates V1 activity\(^{44,45}\). Although visual responses during running were greater than during resting (Supplementary Fig. 11a, b), the response patterns of a population of responsive cells were not largely different during running and resting (Supplementary Fig. 11c). Consequently, the reconstructed image patterns were not substantially affected by the locomotion state (Supplementary Fig. 11d, e).

**Discussion**
Visual responses to natural images are sparse in V1 (refs. \(^{2,3,5–9,14,46}\)). We confirmed that a single natural image activated only a small percentage of neurons. According to the encoding model analysis, the visual responses of individual neurons were sparser than predicted from a linear model (Fig. 2f, g). Information has been proposed to be easily decoded from sparse representations\(^4\). Indeed, sparse activity increases the discriminability of two natural scenes by rendering the representations of the two scenes separable\(^5\). Our results extend this finding by showing that information about visual contents represented by sparsely active neurons are linearly accessible, suggesting that downstream areas easily and optimally decode images by collecting the activity of highly responsive neurons in the sparse representation.
The visual features encoded by individual neurons should be diverse, in order to ensure that a small number of active neurons represent the complex visual features of the image. Although the RF structures of mouse V1 have already been reported21,22,33,34, their diversity has not been analyzed with respect to natural image representation. In the present study, the visual features represented by sparsely active neurons were sufficiently diverse to represent the visual contents of natural images (Fig. 7c–g). Computational models for natural image representation suggest that sparse activity and the number of available neurons affect the diversity of RF structure20,47–49.

We also revealed how multiple natural images were represented in a local population. A single natural image activated specific subsets of neurons, whereas most neurons responded to at least one of the images, supporting the sparse, distributed code proposed in a previous study50. Thus, most V1 neurons are involved in visual processing. Due to the small overlap of responsive neurons between images (Fig. 1i), many natural images were represented by a limited number of responsive neurons (Fig. 7a–d). Furthermore, the RFs of all responsive neurons were sufficient to represent all the natural images used in the present study (Fig. 7c–g). Together, any natural image could be represented by a combination of responsive neurons. These findings also suggest that the representation of multiple natural images is high dimensional, consistent with a report about high-dimensional representation in mouse V150. Thus, a single natural image can be low dimensionally represented in a high-dimensional representation space for a large number of natural scenes.

Sparsely active neurons reliably represented an image across trials, regardless of trial-to-trial response variability. Although a computational model proposed sparse and overcomplete representation as the optimal representation of natural images by unreliable neurons52, this model has never been investigated experimentally. The reliable representation was mainly achieved by diverse, partially overlapping representations, consistent with overcomplete representation. The RF subregions of some V1 neurons partially overlap31, which may be useful for reliable image representation. The reliable representation was also likely to be helped by almost independent activity across trials among neurons with similar RFs. Our results suggest a new representation scheme in which information is reliably represented, while the corresponding neuronal patterns change across trials. This model seems to be similar to “dropout” in deep learning51 and may be useful for avoiding overfitting and local minimum problems in learning.

Methods
Preparation for two-photon imaging in anesthetized mice. All experimental procedures were approved by the local Animal Use and Care Committee of Kyushu University and the University of Tokyo. C57BL/6 mice (male and female) were used (Japan SLC Inc., Shizuoka, Japan). Mice were anesthetized with isoflurane (5% for induction, 1.5% for maintenance during surgery, –0.5% during imaging with a sedation of ~0.5 mg/kg chloroprocaine, Sigma-Aldrich, St. Louis, MO, USA). The skin was removed from the head, and the skull over the cortex was exposed. A custom-made metal plate for head fixation was attached with dental cement (Super Bond, Sun Medical, Shiga, Japan), and a craniotomy (~3 mm in diameter) was performed over left V1 (center position: 0–1 mm anterior to lambda, ±2.5–3 mm laterally to the Rolandic ridge, and 0.8 mm lateral to the midline). A mixture of 0.8 mm Oregon Green BAPTA-AM (OGB1, Life Technologies, Grand Island, NY, USA) dissolved in 10% Pluronic (Life Technologies) and 0.025 mM sulforhodamine 101 (ref. 52; SR101, Sigma-Aldrich) was pressure-injected with a Picospritzer III (Parker Hannifin, Cleveland, OH, USA) at a depth of 300–500 μm from the brain surface. The cranial window was sealed with a coverslip and dental cement. The imaging experiment began at least 1 h after the OGB1 injection.

Preparation for two-photon imaging in awake mice. In awake mouse experiments, we used two lines of transgenic mice: Thy1-GCaMP6s (GPL4.3) transgenic mouse41 (JAX 0424275, n = 1 mouse, two imaging planes) and the mice obtained by crossing gad2-ires-cre mice43 (JAX #010802) with Ai14 mice42 (JAX #007914; Gad2-Ai14, n = 2 mice, five imaging planes). Thy1-GCaMP6s mice express GCaMP6s40 in cortical neurons. Gad2-Ai14 mice express tdTomato in almost all inhibitory neurons53.

GCaMP6s was introduced into Gad2-Ai14 mice via an adeno-associated virus (AAV). The Gad2-Ai14 mice were anesthetized with isoflurane as described above. A small incision was made on a scalp, and a small hole (~0.3 mm diameter) was made in the skull over left V1. AAV2/1-syn-GCaMP6s54 (vector core; University of Pennsylvania, Philadelphia, PA, USA) was injected into V1 through the hole (titer: 3.0–5.0 × 10^{12} genomes/mL, volume: 500 nl, depth: 250-300 micron from a brain surface). After suturing the incision, the Gad2-Ai14 mice were recovered at least 3 days after the injection. The mice were anesthetized with isoflurane to attach a metal plate for head fixation and to make a cranial window as described above. The mice were recovered after the surgery.

The mice were daily habituated with a head fixation on a disc-type treadmill. Duration of the head fixation started with a few minutes and gradually prolonged up to ~2 h over several days. If mice were calmly head-fixed for 2 h without any stressful sign, imaging experiments started on the next day. The imaging started at least 1 week after the cranial window surgery, and 3 weeks after the virus injection for the gad2-Ai14 mice.

Two-photon Ca^{2+} imaging. Imaging was performed with a two-photon microscope (AI15 MP, Nikon, Tokyo, Japan) equipped with an 830 nm × 100 mm PlanApo, Nikon) and Tisapphire mode-locked laser (MatTai Deep See, Spectra Physics, Santa Clara, CA, USA)53,54. OGB1, SR101, GCaMP6s, and tdTomato were excited at a wavelength of 920 nm. Emission filters with a passband of 325/50 nm were used for the OGB1 and GCaMP6s signals, and filters with a passband of 629/ 56 nm for the SR101 and tdTomato signals. The fields of view (FOVs) were 338 × 338 μm (10 planes from seven anesthetized mice) and 507 × 507 μm (14 planes from seven anesthetized mice and seven planes from 3 awake mice) at 512 × 512 pixels. The sampling frame rate was 30 Hz using a resonant scanner.

Monitoring of eye and treadmill motion in awake mice. In awake mouse experiments, right eye images and rotation of the disc-type treadmill were recorded during the imaging. The right eye was monitored with a USB camera (NET New Electronic Technology GmbH, Germany). The treadmill rotation was monitored with a rotary encoder (OMRON, Japan). The eye images, the encoder signals and time stamps of frame acquisition of the two-photon imaging were simultaneously recorded using a custom-written program in LabView (National Instruments, Austin, TX, USA).
**Visual stimulation.** Before beginning the recording session, the retinotopic position of the recorded FOV was determined using moving grating patches (lateral or upper directions, 99.9% contrast, 0.04 cycle/degrees, 2 Hz temporal frequency, 20 and 50 degrees in diameter), while monitoring the changes in signals over the entire FOV. The lateral or upper motion directions of the grating were used to activate many cells because the preferred directions of mouse V1 neurons are slightly biased toward the cardinal directions54,55. First, the grating patch of 50 degrees in diameter was presented in 1 of 15 (5 × 3) positions that covered the entire monitor to roughly determine the retinotopic position. Then, the patch of 20 degrees in diameter was presented on the 16 (4 × 4) positions covering an 80 × 80-degree space to finely identify the retinotopic position. The stimulus position that induced the maximum visual response of the entire FOV was set as the center of the retinotopic position of the FOV.

A set of circular patches of greyscale, contrast-enhanced natural images (200 image types in main datasets, dataset 1) was used as the visual stimuli for predicting the response and reconstructing the natural image (256 intensity level, 60 degrees in diameter, 512 × 512 pixels, with a circular edge (5 degrees) that was gradually mixed to a gray background). Contrasts of natural images were enhanced (>90%). Original natural images were obtained from the van Hateren Natural Image Dataset (http://pirsquared.org/research/#van-hateren-database)56 and the McGill Calibrated Color Image Database (http://tabby.vision.mcgill.ca/html/welcome.html)57. Square image patches (512 × 512 pixels) were obtained from...
**Fig. 6 Reliable image representation across trials.** a Examples of single-trial reconstructed images (top) and response patterns in a FOV (bottom). The first panel is a stimulus image, and the last panel is a trial-averaged image. FOV size: 507 micron each side. The color code for each dot indicates the response amplitude of each cell. b Single-trial evoked responses to the image in a. c Across-trial similarity of the reconstructed images (left) and the response patterns of responsive cells (right). The across-trial similarity was the Pearson’s correlation coefficient between a single-trial reconstructed image (or response pattern) and their trial average. d Across-trial variability of the reconstructed images (left) and the response patterns of responsive cells (right). The normalized squared error between a single-trial image (or response pattern) and their trial average was computed for the across-trial variability. e Reconstructed image from a set of overlapping cells (cell #1-9 in b). Upper left: stimulus image and trial-averaged reconstructed images from the nine overlapping cells. Cyan dotted line: the overlapping area. Upper right: representation (reverse filters) of the overlapping cells. Red line: the representation area. Cell #1 was the reference cell. Lower: single-trial representation patterns of three example cells (cell #1, 2, and 3) selected from the overlapping cells. Bottom: single-trial reconstructed images (upper) and single-trial response patterns in an FOV (lower) obtained from the nine overlapping cells. The brightness of each color dot in the lower panels indicates response amplitude of each cell. f Across-trial variability of a local part of reconstructed image (cyan dotted line in e) against the number of overlapping cells used for the reconstruction in the example case in e. g Across-trial variability against the percentage of overlapping cells used for the reconstruction. Black lines: raw data. Orange lines: trial-shuffled data. Thick and thin lines are median and 25th or 75th percentile, respectively (n = 200 random sequences of cell adding). h Across-trial variability against the number of responsive cells (black line) and of newly added responsive cells (red line) are plotted against the number of overlapping cells used for the reconstruction in the example case in e. i Across-trial variability against the number of responsive cells (black line) and of newly added responsive cells (red line) are plotted against the number of overlapping cells used for the reconstruction. j Across-trial variability against the number of responsive cells (black line) and of newly added responsive cells (red line) are plotted against the number of overlapping cells used for the reconstruction. k Across-trial variability against the number of responsive cells (black line) and of newly added responsive cells (red line) are plotted against the number of overlapping cells used for the reconstruction. l Across-trial variability against the number of responsive cells (black line) and of newly added responsive cells (red line) are plotted against the number of overlapping cells used for the reconstruction. m Across-trial variability against the number of responsive cells (black line) and of newly added responsive cells (red line) are plotted against the number of overlapping cells used for the reconstruction. n Across-trial variability against the number of responsive cells (black line) and of newly added responsive cells (red line) are plotted against the number of overlapping cells used for the reconstruction. o Across-trial variability against the number of responsive cells (black line) and of newly added responsive cells (red line) are plotted against the number of overlapping cells used for the reconstruction. p Across-trial variability against the number of responsive cells (black line) and of newly added responsive cells (red line) are plotted against the number of overlapping cells used for the reconstruction. q Across-trial variability against the number of responsive cells (black line) and of newly added responsive cells (red line) are plotted against the number of overlapping cells used for the reconstruction. r Across-trial variability against the number of responsive cells (black line) and of newly added responsive cells (red line) are plotted against the number of overlapping cells used for the reconstruction. s Across-trial variability against the number of responsive cells (black line) and of newly added responsive cells (red line) are plotted against the number of overlapping cells used for the reconstruction. t Across-trial variability against the number of responsive cells (black line) and of newly added responsive cells (red line) are plotted against the number of overlapping cells used for the reconstruction. u Across-trial variability against the number of responsive cells (black line) and of newly added responsive cells (red line) are plotted against the number of overlapping cells used for the reconstruction. v Across-trial variability against the number of responsive cells (black line) and of newly added responsive cells (red line) are plotted against the number of overlapping cells used for the reconstruction. w Across-trial variability against the number of responsive cells (black line) and of newly added responsive cells (red line) are plotted against the number of overlapping cells used for the reconstruction. x Across-trial variability against the number of responsive cells (black line) and of newly added responsive cells (red line) are plotted against the number of overlapping cells used for the reconstruction. y Across-trial variability against the number of responsive cells (black line) and of newly added responsive cells (red line) are plotted against the number of overlapping cells used for the reconstruction.
image flashes to reduce the total recording time and maximize the number of repetitions. In this design, the tail of the Ca\textsuperscript{2+} response to one image invaded the time window of the next image presentation (Fig. 1b). Although this overlap may have affected the visual responses to two adjacent images, the use of many trial repetitions with the pseudorandom image sequences and the sparse responses to natural images (Fig. 1) minimized the effects of response contamination between two consecutive images.

In another set of experiments with anesthetized mice, we used different image sets (1000–2000 images that did not contain the 200 images described above, \( n = 4 \) planes from three mice). The original images for these sets were derived from the image datasets described above, the Caltech 101 dataset (http://www.vision.caltech.edu/Image_Datasets/Caltech101/\cite{58}), a free image website (https://www.pakutaso.com/), and images that we photographed. In experiments using these image sets, each image was presented three to eight times.

Moving square gratings (eight directions, 0.04 cycles/degree, 2 Hz temporal frequency, 60-degree patch diameter) were presented at the same position as the natural image on the screen. Each direction was presented for 4 s interleaved by 4 s of the gray screen. The sequence of directions was pseudo-randomized, and each direction was presented ten times for anesthetized mice.

All stimuli were presented with PsychoPy\cite{59} on a 32-inch, gamma-corrected liquid crystal display monitor (Samsung, Hwaseong, South Korea) with a 60-Hz refresh rate, and the timing of the stimulus presentation was synchronized with the
timing of imaging acquisition using a transistor–transistor logic pulse counter (USB-6501, National Instruments). The data recording portion for one frame was divided into several sub-sessions (4–6 trials/sub-session and 15–25 min for each sub-session). Each sub-session was interleaved by ~5–10 min of rest time, during which the slight drift of the FOV was manually corrected. In anesthetized mice, the retinotopic position of the FOV was confirmed with the grating patch stimuli during the rest time every two or three sub-sessions, and the recording was terminated if the retinotopic position had shifted (probably due to eye movement). The recordings were performed in one to three planes of different depths and/or positions in each anesthetized mouse (1.7 ± 0.8 planes, mean ± standard deviation). In awake mice, the recording continued independently of the eye position and terminated if the mouse showed any stressful sign. The recording was performed in one plane per day, and one or two planes were obtained from each awake mouse.

For the analyses described below, the natural images were scaled such that the maximum (255) and minimum (0) intensities were 1 and −1, respectively, and the gray intensity (127) was 0. A square (43 × 43 degrees) positioned in the center of the natural image was extracted and downscaled to a 32 × 32-pixel image. The downsampled image was used to analyze the Gabor features, response predictions and image reconstructions.

### Analysis of eye positions and treadmill rotations

Each eye image for the awake mice was binarized based on pixel intensities. The contour of the binarized pupil area was fitted with an ellipse whose center was used as the eye position (Fig. 8a, b). The eye positions on the image were transformed to angular positions. In this transformation, a previously reported reported value was used for the radius of the mouse eye. In the distribution of the eye position during the entire recording session, a peak position was manually selected. Only the time points at which the eye was within 3.5 degrees (or ~70 microns on the image) of the peak position for each eye was selected as described below (except for the time course extraction of the Ca2+ signal from the two-photon imaging data).

In the analysis of the treadmill rotation, position signals from the rotary encoder on the treadmill were transformed to velocity and smoothed with a Savitzky–Golay filter. Periods were defined as periods during which the velocity was >2 cm/sec.

### Analysis of two-photon imaging data

All data analysis procedures were performed using MATLAB (Mathworks, Natick, MA, USA). The recorded images were processed and aligned between frames. The average image across frames was used to determine the region of interests (ROIs) of individual cells. After removing the slow SF component (obtained with a Gaussian filter with a sigma of approximately five times the soma diameter), the frame-averaged image was subjected to a template matching method in which the 2D, difference of sigma of approximately 0.02, 0.04, 0.09, and 0.18 cycle/degrees (cpd).

The time course of the calcium signal in each cell was computed as an average of all pixels within an ROI. Signal contamination from an out-of-focus plane was removed using a previously reported method. Brieﬂy, the contamination ratio was set to 0.7 for all cells in the FOV. In awake mice, the contamination ratio was set to 0.7 for all cells following a previous study, because it was difﬁcult to identify the blood vessels in the GCaMP imaging. After removing the out-of-focus signal, slow temporal frequency components (>60 sec/cycle) were removed from the time course of each cell (a Gaussian low-cut filter applied on the frequency domain for anesthetized data or a median low-cut filter for awake data), followed by smoothing with the Savitzky–Golay ﬁlter (forth order, 15 frame points length (~500 ms)). Then, the ﬁltered time course (Frameaved) was transformed to ratio change (dF/F) by using the 20th percentile value across frames (F) (dF/F = Frameaved – F)/F. Frame-averaged activity (dF/F) during 200 baseline (six frames immediately before the stimulus) and during stimulus (average of the last 200 ms for each stimulus period) were used for subsequent analyses. The evoked response was obtained by subtracting the activity during baseline from that during stimulus. In anesthetized mice, data for moving grating from one plane was discarded because of a large drift of FOV during recording. In awake mice, only data for images that contained at least six trials were used for subsequent analyses.

### Analysis of visual responses

Visually responsive neurons were determined by one-way ANOVA (p < 0.01) with a dataset of N stimuli and one baseline (mean across stimuli) activity in each trial (size: N = 1 activity × no. of trials, N = number of stimuli used for the analysis). To validate this criterion, ANOVA was applied to a randomized dataset in which data labels were shuffled in each trial. The false positive rate was only a small fraction of the percentage of the responsive cells (anesthetized: 85% responsive cells and 1.0% false positive rate. Awake: 82% responsive cells and 1.8% false positive rate).

For each responsive neuron identiﬁed by ANOVA, responsiveness for each image was determined by using a t-test (p < 0.01, comparison of activity between stimulus and baseline) and a cross-correlation image. The averaged evoked response (≥10). The evoked response threshold was used to reduce the false positive rate (Supplementary Fig. 1). The false positive rate was determined with the label-shuffled data. Without the evoked response threshold, the false positive rate was relatively high compared to the responsive cell rate of responsiveness (anesthetized: 3.5% for observed and 0.4% for shuffled data. Awake: 1.8% for observed and 0.3% for shuffled data). With the 10 amplitude threshold, the false positive rate decreased (anesthetized: 2.5% for observed and 0.1% for shuffled data. Awake: 1.5% for observed and 0.1% for shuffled data, Supplementary Fig. 1). Thus, we used the 10% evoked response threshold. For the dataset with 1000–2000 stimulus images (dataset 2), responsive cells were not determined because of fewer trials.

The population sparseness (s) was computed using the equation described in previous studies. The signal from the two-photon imaging data (Gabor filter with an almost self-inverting feature was prepared to extract the visual features of the natural images). The downsampled images were first subject to the set of Gabor filters to obtain Gabor feature values. A single feature value corresponds to a single wavelet filter.

Gabor filters have four orientations (0, 45, 90, and 135 degrees), two phases, and four sizes (8 × 8, 16 × 16, 32 × 32, and 64 × 64 pixels) located on 11 × 11, 5 × 5, 3 × 3, and 1 × 1 grids (Fig. 2a, b). Therefore, the three smaller scale filters spatially overlapped with each other. The spatial frequencies of the four size scales of the Gabor wavelets were 0.02, 0.04, 0.09, and 0.18 cycle/degrees.

The Gabor filter set was almost self-inverting, i.e., the feature values obtained by applying an image to the wavelet set could be transformed to the image by summing the filters after multiplying by the feature values.

\[ F = G_{\text{Gabor}} I \] (5)

\[ I' = G_{\text{Gabor}} F \] (6)

In Eq. (5) (corresponding to Eq. (1) in the main text), F is the feature value matrix (matrix size: f × s; the number of features, 1248, and s: the number of images). G_{\text{Gabor}} is the Gabor filter matrix that transforms images to feature values, in which each row contains the reshaped 2D-Gabor filters (f × p; p: the number of the image pixels, 1024), and I is the downsampled stimulus image matrix (p × s). In Eq. (6) (corresponding to Eq. (4)), I' is the reconstructed image matrix (p × s) from F, and G_{\text{Gabor}} is the Gabor filter matrix that transforms the features to images (p × f).

In an ideal situation, G_{\text{Gabor}}^T G_{\text{Gabor}} equals G_{\text{ideal}} (i.e., G_{\text{Gabor}} = G_{\text{ideal}} × A^{-1} and A are inverse and transposed matrices of A, respectively), resulting in I' equaling I (ref. 50). However, our Gabor transformation was not perfect; the pixel-to-pixel correlation between I and I' was 0.93 ± 0.026 (mean ± standard deviation, n = 200 images). To minimize the effect of this information loss on the evaluations of image reconstruction performance (see below), we used I' instead of I as the target images for the evaluation of image reconstruction. The G_{\text{ideal}} in Eq. (6) was different from G_{\text{Gabor}} in Eq. (5) in terms of scaling: G_{\text{ideal}} = G_{\text{ideal}} κ (a: scaling factor). a was computed to minimize the sum of the mean squared error between I' and I. The Gabor filters and the transformations were based on an open source program (originally written by Dr. Daisuke Kato and Dr. Izumi Ohzawa, Osaka University, Japan, https://visioneuroinf.jp/modules/soonips/detail.php?item_id=6894).

### Encoding model (response prediction model)

In the encoding model, single-cell responses (R_{ik} = [R_{i1}, . . . , R_{ik}, . . . , R_{in}]) and cell number (n) trial number across stimuli and trials. The number of trials across stimuli and trials) were predicted using linear regression analysis of the selected Gabor feature values (F_{select} = [F_{select}], select Gabor feature number. F_{select} = [F_{select} × r_{select}]. Number of selected features, 2a–c, Supplementary Fig. 2a, b).

\[ R_{ik} = NL(W_{ik} F_{select} + b) \] (7)

where W_{ik} ([W_{ik}] size: 1 × fselect) is the weight vector, F_{select} ([fselect × rselect] is the matrix of selected feature values, b_{ik} (size: 1 × 1) is bias, and NL() is the nonlinear logistic function (Eq. (7) corresponds to Eq. (2)). The encoding model was created independently for each cell. The features used in the regression were determined as follows. First, Pearson’s correlation coefficients between the response and feature values were computed for each feature. Then, using one of the preset values for the correlation coefficient significant as a threshold (13 points ranging from 0.05 to 0.35, Supplementary Fig. 2a, b), only the more strongly correlated features were selected (feature selection) and used in the regression analysis. W_{ik} and b_{ik} were estimated to...
minimize the loss function: \( \sum [R_i - (W_iF + B_i)^2] + \lambda \sum W^2 \) (\( \lambda \) is a regularization parameter). This was solved using Bayesian linear regression with an expectation-maximization algorithm that is approximately equivalent to linear regression with regularization. After the regression analysis, the nonlinearity of the predicted response was adjusted with a rectification step using the following function: \( \text{NL}(s) = A[1 + \exp(\alpha s + C)] + D \), where \( A, B, C, \) and \( D \) are parameters estimated using a built-in Matlab function (\text{loglin}()). This step merely scaled the regression output without changing the regression parameters (\( W^* \) and \( B^* \)).

The performance of the model was estimated for each neuron on ten-fold CVs, in which the response data for 90% images were used to estimate the parameters, and the remaining data for 10% images were used to evaluate the prediction, (thus, \( W^* \) and \( B^* \) were estimated and fixed in each CV). In the ten-fold CVs, all images were used once as test data. The prediction performances were estimated using Pearson’s correlation coefficient between the observed (trial average) and predicted responses. Encoding models were created for all preset threshold values for feature selection, and the model that exhibited the best prediction performance was selected as the final model.

In the analysis of overlapping weights (i.e., feature) between two cells, the percentage of overlapping weights relative to the number of non-zero weights was computed for each cell and averaged between the two cells in the pair.

We independently obtained the weights of the image reconstruction model (\( H = [H_1, \ldots, H_N] \), size: \( f \times N_{cell} \)), the weights of encoding model (\( W = [W_1, \ldots, W_{N_{cell}}] \), size: \( f \times N_{cell} \) and RF coefficient, \( b = [b_1, \ldots, b_{N_{cell}}] \), size: \( f \times N_{cell} \)) (Fig. 3) of the image reconstruction model to optimally reconstruct the image by a population of neurons. In the image reconstruction model, \( H \) was estimated to directly optimize the image reconstruction considering the responses of multiple cells. By contrast, in the encoding model (or RF estimation), weights were estimated independently in each cell without considering the other cells’ responses. Because \( H \) was likely to be more optimized to represent images with multiple cells than \( W \) or RF, we chose the model scheme for the image reconstruction.

Reconstruction performance against the number of features. In the analysis in Supplementary Fig. 5a-c, the cell-selection model was used for the image reconstruction. In the cell-selection model, each neuron participated in the reconstructions of a small number of features that were strongly correlated with the neuron’s responses. In the cell-selection model shown in Fig. 3 (the original model), the threshold for the correlation coefficient was selected based on the encoding model for each neuron (Fig. 3a, right panel). For each neuron, the threshold of the correlation coefficient was adjusted to increase (or decrease) the number of features for which each neuron participated in their reconstruction (0.1–20-fold change in the number of features per neuron relative to the original model, Supplementary Fig. 5a–c). For each fold change, the reconstruction model was trained with training data (i.e., weights and bias parameters were estimated in each fold change of the number of features), and the performance was estimated with test data using ten-fold CV as described above.

Image reconstruction from a small number of cells. In the analyses shown in Figs. 4a–h and 5a, b, cells in each image were separated into responsive and remaining cells and sorted by their response amplitude in descending order (i.e., from highest to lowest response amplitude). Then, the cells were selected first from the responsive cells and then from the remaining cells (i.e., Figs. 4a–h) or dropping (Fig. 5a, b). The analyses only used data for images including at least ten responsive cells in Fig. 4a–h and at least five responsive cells in Fig. 5a–d. In the image reconstruction from a subset of cells for each image (Figs. 4–6), the weights of the cell-selection model (\( H = [H_1, \ldots, H_{N_{cell}}], f \times N_{cell} \)) were scaled because \( H \) was estimated by more cells than those used during the cell addition or removal.

Robustness of image reconstruction against cell drop. In the robustness (Fig. 5c–f), a representation area for each cell was determined using the z-scored reverse filter (sum of weights \( \times \) Gabor filters). The representation area was defined as a cluster of pixels whose absolute z-scores were > 1.5 and whose contours were filled (e.g., red contours in Fig. 5c and Supplementary Fig. 6a). If multiple areas were obtained, the largest was used. Then, using the representation area, the overlap index was computed between responsive cells for each stimulus; overlap index = \((A \cap B)/(A \cup B)\), where \((A \cap B)\) is the overlapping representation area between cell A and cell B, and \((A \cup B)\) is a combined representation area between cells A and cell B (Supplementary Fig. 6a). Using the overlap index, a set of overlapping cells was selected for each responsive cell; the overlapping cells consisted of one responsive cell (the reference cell) and the responsive cells that overlapped with the reference cell (overlap index > 0.2). This analysis did not care whether other overlapping cells overlapped with each other or with other non-selected cells. The overlap index was greater than 0.2 once among the overlapping cells, and the reconstructed image was computed after each cell was dropped. The reference cell was initially removed, and then other remaining overlapping cells were removed in each cell drop sequence. Changes in the reconstructed images were estimated by quantifying the pixel-to-pixel correlation (\( R \)) of a local part of the image. The local part of the image was determined as the representation area of the reference cell that was overlapped by the area of at least one of the other overlapping cells (overlapping area in Fig. 5c). This random dropping of overlapping cells was repeated 120 times, and the results were averaged across the random orders for each reference cell. All responsive cells were used to reconstruct the reference cell for each stimulus image. This analysis only used data for images that included at least five responsive cells and sets of overlapping cells that included at least five overlapping cells.
Across-trial similarity and variability. To estimate the reliability of the reconstructed images (or response patterns) across trials, two measures were used: across-trial similarity and across-trial variability. For the across-trial similarity of the reconstructed images (or response patterns; see Fig. 6c), Pearson’s correlations between single-trial reconstructed images (or response patterns) and their trial average were computed and averaged across trials.

For the across-trial variability (Fig. 6d-g), the normalized squared error between single-trial images (or response patterns) and trial-averaged images (or response patterns) were computed using the following equation:

\[
\text{Across-trial variability} = \frac{\sum \sum (A_{n} - \bar{A})^2 / (N_{\text{trials}} - 1)}{\sum \sum (A_{n} - \bar{A}) / (N_{\text{trials}} - 1)}
\]

where \(A_{n}\) is a single-trial reconstructed image or response pattern, \(\bar{A}\) is the trial-averaged reconstructed image or response pattern, \(N_{\text{trials}}\) is the number of trials, \(X_{t}\) is the number of pixels or cells, and \(N_{\text{cell}}\) is the number of cells. The squared error between single-trial images and their trial average was computed and averaged across trials. Two measures were used: (1) squared error, which reflects deviations of single-trial images or response patterns from that of trial average, and (2) averaged across trials. However, this squared error could be larger when the pixel intensities are larger in the reconstructed image or when the response amplitudes are larger in the cell population. To control this problem, this squared error was normalized by another squared error between \(A_{n}\) and \(\bar{A}\) (i.e., \(\sum \sum (A_{n} - \bar{A})^2/\sum \sum (A_{n} - \bar{A})^2\)), which reflects deviations of pixels’ intensity or cells’ responses from the mean value across pixels or cells in the trial-averaged data). To correct differences between the number of pixels and cells (i.e., the number of degrees of freedom), each squared error was divided by the degree of freedom \(N_{\text{cell}} - 1\) for numerator and \(N_{\text{trials}} - 1\) for denominator), following the correction of squared error by degree of freedom in ANOVA wherein pixels or cells correspond to groups and trials correspond to samples. This across-trial variability equals the inverse of the F-value in the ANOVA.

In the analysis depicted in Fig. 6f-g, the overlapping cells were selected as described above for each responsive cell (i.e., the reference cell; see the section titled “Robustness of image reconstruction against cell drop”). The reference cell was initially selected, and then other overlapping cells were randomly selected for a set of cells that were used for image reconstruction (the sequences of random cell selection were repeated 200 times). The image was reconstructed from the subset of overlapping cells, and across-trial variability of a local part (i.e., overlapping area) of the reconstructed image was computed for each subset of cells. Only data for images that contained at least five responsive cells were used in the analyses in Fig. 6.

Noise correlation. In the analysis depicted in Supplementary Fig. 7, the noise correlation was computed using the responses across stimuli. Evoked responses to each stimulus were shuffled across stimuli and across trials. Using the shuffled data, an image reconstruction model was obtained as described above for the analyses in Fig. 6g. Supplementary Fig. 7g-i,p-r, and Supplementary Fig. 9f.

Capacity for image representation of encoded features. The analysis in Fig. 7e-g illustrates whether the features encoded by responsive cells could represent images as a basis function independent of actual neural responses. If the features encoded by responsive neurons can represent any image, a set of features of a given image (F) will be linearly regressed by the weights of the responsive cells in the cell-selection model, \(H = H^{t} \ldots H_{1}^{t} f \times N_{\text{cell}} H^{t}\) in Eq. (8):

\[
F = HB + d + e
\]

where \(B\) and \(d\) are free parameters that are calculated to minimize the sum of the squared error, \(\sum (F - (HB + d))^{2}\), and \(e\) is an error term. \(F\) was selected from a test dataset, and \(H\) was obtained from a training dataset in the ten-fold CV. The fitting was evaluated by calculating the fitting error (Error) on the image space as follows:

\[
\text{Error} = \sum (F - \bar{F})^{2} / \sum (F_{\text{mean}} - \bar{F})^{2} \times 100
\]

where \(G_{w}\) is the Gabor filter matrix used for reconstruction (Eq. (6)), and \(F_{\text{mean}}\) is the mean pixel intensity of F. Thus, this analysis estimates how well the features of individual neurons in a local population could represent the image features independent of actual neuronal activity. In other words, this analysis estimates the upper-bound capacity of a local population to represent any image with an ideal combination of cell features (with parameters \(B\) and \(d\)).
15. Stanley, G. B., Li, F. F. & Dan, Y. Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus. *J. Neurosci.* 19, 8036–8042 (1999) (2008).

16. Miyawaki, Y. et al. Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron* 60, 915–929 (2008).

17. Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M. & Gallant, J. L. Bayesian reconstruction of natural images from human brain activity. *Neuron* 63, 902–915 (2009).

18. Nishimoto, S. et al. Reconstructing visual experiences from brain activity evoked by natural movies. *Curr. Biol.* 21, 1641–1646 (2011).

19. Horikawa, T., Tamaki, M., Miyawaki, Y. & Kamitani, Y. Neural decoding of visual imagery during sleep. *Science* 340, 639–642 (2013).

20. Doi, E. & Lewicki, M. S. Sparse coding of natural images using an overcomplete set of limited capacity units. *Adv. Neural Inf. Process. Syst.* 17, 377–384 (2005).

21. Smith, S. L. & Hauser, M. Parallel processing of visual space by neighboring neurons in mouse visual cortex. *Nat. Neurosci.* 13, 1144–1149 (2010).

22. Bonin, V., Histed, M. H., Yurgenson, S. & Reid, R. C. Local diversity and fine-scale organization of receptive fields in mouse visual cortex. *J. Neurosci.* 31, 18506–18512 (2011).

23. Kampa, B. M., Roth, M. M., Gobel, W. & Helmchen, F. Representation of visual scenes by local neuronal populations in layer 2/3 of mouse visual cortex. *Front. Neural Circuits* 5, 18 (2011).

24. Ko, H. et al. Functional specificity of local synaptic connections in neocortical networks. *Nature* 473, 87–91 (2011).

25. Marschel, J. H., Garrett, M. E., Nauhaus, I. & Callaway, E. M. Functional specialization of seven mouse visual cortical areas. *Neuron* 72, 1040–1054 (2011).

26. Miller, J. E., Ayyezhatat, I., Carrillo-Reid, L. & Yuste, R. Visual stimuli recruit intrinsically generated cortical ensembles. *Proc. Natl Acad. Sci. USA* 111, E4053–E4061 (2014).

27. Rikhye, R. V. & Sur, M. Spatial correlations in natural scenes modulate response reliability in mouse visual cortex. *J. Neurosci.* 35, 14661–14680 (2015).

28. Olshausen, B. A. & Field, D. J. How close are we to understanding V1? *Neural Comput.* 17, 1665–1699 (2005).

29. Shoham, S., O’Connor, D. H. & Segev, R. How silent is the brain: is there a “dark matter” problem in neuroscience? *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* 192, 777–784 (2006).

30. Yoshida, T. & Ohki, K. Visual image reconstruction from neuronal activities in the mouse primary visual cortex. Program No. 415.17. 2015 Neuroscience Meeting Planner. (Chicago, IL: Society for Neuroscience, 2015).

31. Yoshida, T. & Ohki, K. Robust representation of natural images by sparse and variable population of active neurons in visual cortex. Preprint at https://www.biorxiv.org/content/10.1101/300863v2 (2018).

32. Smyth, D., Willmore, B., Baker, G. E., Thompson, I. D. & Tolhurst, D. J. The response field organization of simple cells in primary cortical areas of ferrets under natural scene stimulation. *J. Neurosci.* 23, 4746–4759 (2003).

33. Ko, H. et al. The emergence of functional microcircuits in visual cortex. *Nature* 496, 96–100 (2013).

34. Corseil, L. et al. Functional organization of excitatory synaptic strength in primary visual cortex. *Neuron* 67, 858–871 (2010).

35. Shadlen, M. N. & Newsome, W. T. Noise, neural codes and cortical computation with cellular resolution reveals precise micro-architecture in visual cortex. *Nature* 433, 597–603 (2005).

36. Hagihara, K. M., Murakami, T., Yoshida, T., Tagawa, Y. & Ohki, K. Neuronal activity is not required for the initial formation and maturation of visual selectivity. *Nat. Neurosci.* 18, 1780–1788 (2015).

37. Mank, M. et al. A genetically encoded calcium indicator for chronic in vivo2-photon imaging. *Nat. Methods* 5, 805–811 (2008).

38. van Hateren, J. H. & van der Schaaf, A. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. Biol. Sci. 265, 359–366 (1998).

39. Olmos, A. & Kingdom, F. A. A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception* 33, 1463–1473 (2004).

40. Fei-Fei, L., Fergus, R. & Perona, P. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. In Proc.2004 Conference on Computer Vision and Pattern Recognition Workshop, 178–178 (2004).

41. Peirce, J. W. Generating stimuli for neuroscience using PsychoPy. *Front. Neuroinform.* 2, 10 (2008).

42. Stahl, J. S., van Alphen, M. A. & De Zeeuw, C. I. A comparison of video and magnetic search coil recordings of mouse eye movements. *J. Neurosci.* Methods 99, 101–110 (2000).

43. Kerlin, A. M., Andermann, M. L., Berezovski, V. K. & Reid, R. C. Broadly tuned response properties of diverse inhibitory neuron subtypes in mouse visual cortex. *Neuron* 57, 109–120 (2008).

44. Trees, A. & Rolls, E. T. What determines the capacity of autoassociative memories in the brain? *Netw. Comput. Neural Syst.* 2, 371–397 (1991).

45. Lee, T. S. Image representation using 2D gabor wavelets. *IEEE Trans. Pattern Anal. Mach. Intell.* 18, 959–971 (1996).

46. Kay, K. N., Naselaris, T., Prenger, R. J. & Gallant, J. L. Identifying natural images from human brain activity. *Nature* 452, 352–355 (2008).

47. Bishop, C.M. Pattern Recognition and Machine Learning (Information Science and Statistics) (Springer-Verlag, New York, Inc. 2006).

Acknowledgements
We thank Ms. Y. Sono, A. Hayashi, T. Inoue, A. Ohmori, A. Honda, and M. Nakamichi for animal care, and all members of Ohki laboratory for support and discussions. This work was supported by grants from Core Research for Evolutionary Science and Technology (CREST)—Japan Agency for Medical Research and Development (AMED; to K.O.), Japan Society for the Promotion of Science (JSPS) KAKENHI (grant number 19H05642, 25221001, and 25171004 to K.O. and 15K16573, 17K13276, and 19K07793 to T.Y.), Brain Mapping by AMED (to K.O.), Strategic International Research Cooperative Program (SICP)—AMED (to K.O.), grants from the Ichiro Kanehara Foundation for the Promotion of Medical Sciences and Medical Care, and the Uchara Memorial Foundation (to T.Y.).

Author contributions
T.Y. and K.O. designed the research. T.Y. performed the experiments. T.Y. and K.O. analyzed the data and wrote the manuscript. K.O. supervised the research.

Competing interests
The authors declare no competing interests.
