Analysis of visual processing capabilities and neural coding strategies of a detailed model for laminar cortical microcircuits in mouse V1

Guozhang Chen, Franz Scherr, and Wolfgang Maass*

Institute of Theoretical Computer Science, Graz University of Technology, Graz, Austria

April 2022

Abstract

Neural networks of the brain that process visual information have structural properties that differ significantly from those of neural networks which are commonly used for visual processing in AI, such as Convolutional Neural Networks (CNNs). But it has remained unknown how these structural differences are related to network function. We analyze visual processing capabilities of a large-scale model for area V1 that arguably provides the most comprehensive accumulation of anatomical and neurophysiological data that is currently available. Its network structure turns out to induce a number of characteristic visual processing capabilities of the brain, in particular the capability to multiplex different visual processing tasks, also on temporally dispersed visual information, with remarkable robustness to noise. This V1 model also exhibits a number of characteristic neural coding properties of the brain, which provide explanations for its superior noise robustness. Since visual processing in the brain is substantially more energy-efficient than implementations of CNNs in common computer hardware, such brain-like neural network models are likely to have also an impact on technology: As blueprints for visual processing in more energy-efficient neuromorphic hardware.

1 Introduction

The comprehensive model [Billeh et al., 2020] for a patch of cortical area V1 in mouse provides an unprecedented window into the dynamics of this brain area. We show that it also provides a unique tool for studying brain-style visual processing and neural coding. The architecture of V1 exhibits an interesting combination of feedforward and recurrent connectivity: Neurons are distributed over several parallel 2D sheets, commonly referred to as layers or laminae in neuroscience. The neurons are recurrently connected, but not randomly or in an all-to-all manner. Rather, synaptic connections exist primarily between nearby neurons, both within a layer and between layers. Connectivity between layers supports a strong feedforward stream of visual information from L4 to L2/3 to L5/6, that is complemented by a host of recurrent loops. The dominance of short connections makes it possible to combine in V1 extensive recurrent connectivity with a really small total wire length, which is essential for its physical realization.

The model of [Billeh et al., 2020] integrates besides these anatomical details also a host of neurophysiological data about area V1. The point neuron version of this model that we are considering employs

*Corresponding author: maass@igi.tugraz.at

generalized leaky integrate-and-fire neurons, more precisely $GLIF_3$ neurons. These have in addition to the membrane potential two further hidden variables that model slower processes in biological neurons. The large diversity of neurons in the brain is reflected in the model of Billeh et al. through the use of 111 different types of $GLIF_3$ neuron models that have each been fitted to experimental data in the Allen Brain Atlas [Allen Institute, 2018].

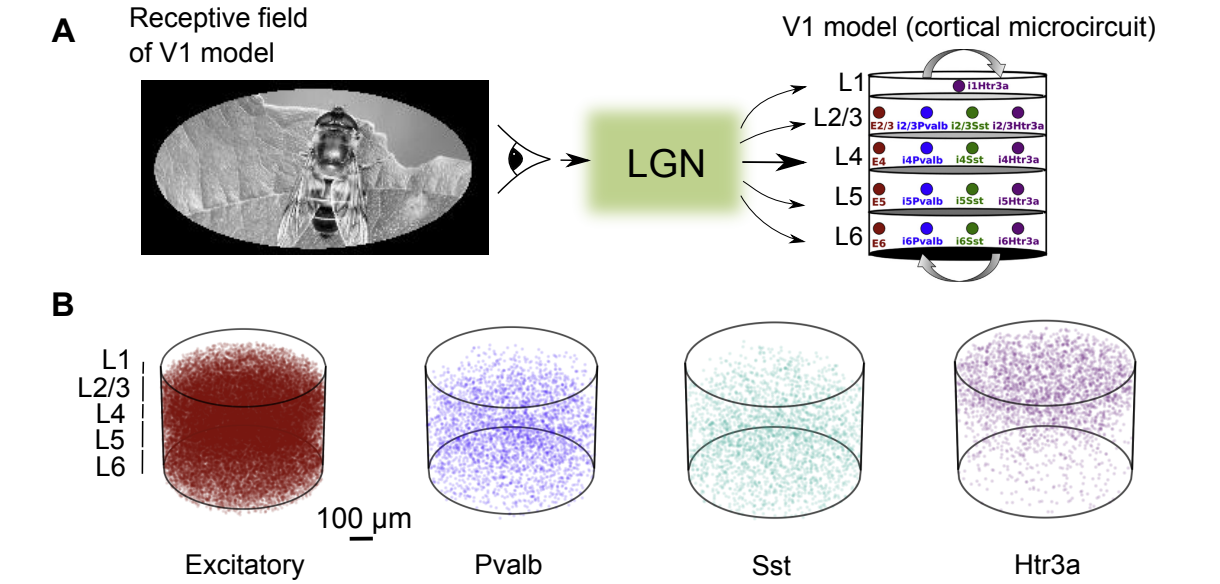
The original model of [Billeh et al., 2020] is not able to solve nontrivial computing tasks, since its synaptic weights were chosen on the basis of sparse experimental data about the mean and variance of synaptic weights. In contrast, synaptic weights in the living brain are individually tuned through a host of synaptic plasticity processes, and these processes induce higher-order correlations between weights that are crucial for computing capabilities of the network. At present we do not have enough data about these plasticity processes to reproduce them in a model. But we can address the question of what visual processing capabilities are supported by the model if synaptic weights are aligned for visual processing tasks through stochastic gradient descent. We applied this strategy to 5 different visual processing tasks that have commonly been considered in biological experiments [Rumyantsev et al., 2020, Stringer et al., 2021, Garrett et al., 2020a, Siegle et al., 2021, Morcos and Harvey, 2016, Engelhard et al., 2019]. Afterward, our model achieved high accuracy simultaneously for all 5 tasks, while working in a biologically realistic sparse firing regime close to criticality [Barth and Poulet, 2012, Wilting and Priesemann, 2019]. Surprisingly, its performance level remained in the same high-performance regime as the brain, even when we subjected the V1 model to noise in the images and in the network that it had not encountered during training. We demonstrate that this out-of-distribution (OOD) generalization capability of the V1 model with regard to new perturbations is far superior to that of CNNs. We provide an explanation for that through an analysis of neural coding properties of these two types of models: Both use high-dimensional neural codes for images. But the neural representation in the model of Billeh et al. is more robust because it employs, like the brain [Stringer et al., 2019], a power law for the explained variance in higher PCA components that is close to a theoretically optimal compromise between the opposing goals to create noise-robust neural codes and to capture many details of visual inputs. In contrast, neural codes in CNNs have a different power law which reveals a focus on the latter [Kong et al., 2022]. In addition, we demonstrate that the model of Billeh et al. uses preferentially those dimensions of population activity for coding that are orthogonal to the largest noise dimensions, like the brain does [Rumyantsev et al., 2020].

Altogether our results show that the currently available anatomical and neurophysiological data, as compiled in [Billeh et al., 2020], provide the basis for a new generation of neural network models for visual processing that can multiplex diverse visual processing capabilities in a highly robust manner. Furthermore, these neural network models provide new paradigms for neuromorphic computing since they combine multiplexing capability and robustness to noise with small total wire length and highly energy-efficient sparse activity.

2 Results

2.1 Integration of anatomical and neurophysiological data, as well as data on noise in the brain, into a neural network model of area V1

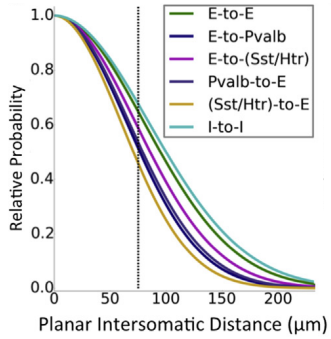
Several decades of intense research efforts have accumulated a large body of knowledge about the anatomy and neurophysiology of the visual cortex, especially for primary visual cortex, i.e., for area V1. But it has remained unknown to what extent this insight into the structure of V1 can be related to its function. We have examined this question for the case of the large-scale model for a patch of V1 in mouse from [Billeh et al., 2020], which is arguably the most comprehensive integration of anatomical and neurophysiological

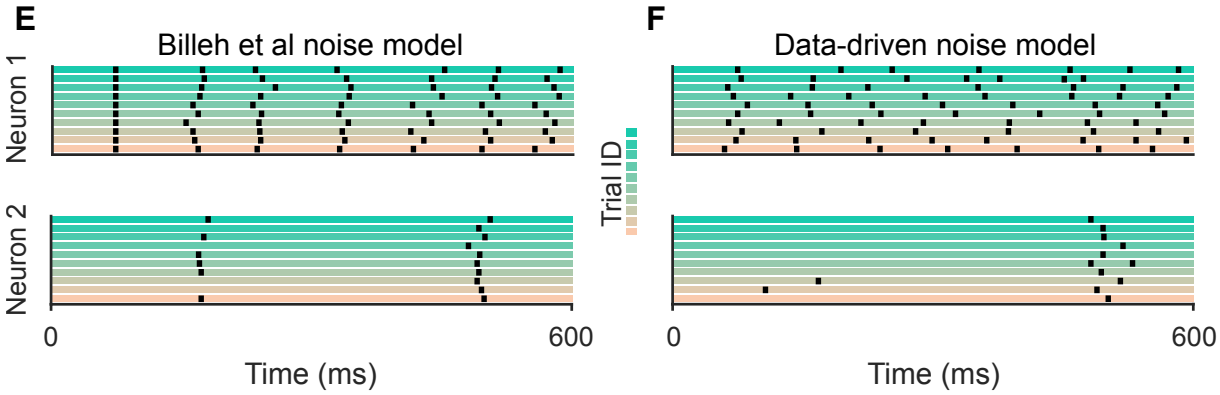


C

Connection probabilities at 75µm

Trg / Src	i1Htr3a	E2/3	i2/3 Pvalb	i2/3 Sst	i2/3 Htr3a	E4	i4 Pvalb	i4 Sst	i4 Htr3a	E5	i5 Pvalb	i5 Ssst	i5 Htr3a	E6	i6 Pvalb	i6 Sst	i6 Htr3a
i1Htr3a	0.656	0.356	0.093	0.068	0.464	0.148				0.148	0	0		0.148			
E2/3	0	0.16	0.395	0.182	0.105	0.016	0.083	0.083	0.083	0.083	0.081	0.102		0			
i2/3Pva	0.024	0.411	0.451	0.03	0.22	0.05	0.05	0.05	0.05	0.07	0.073						
i2/3Sst	0.279	0.424	0.857	0.082	0.77	0.05	0.05	0.05	0.05	0.021		0					
i2/3Htr	0	0.087	0.02	0.625	0.028	0.05	0.05	0.05	0.05	0							
E4		0.14	0.1	0.1	0.1	0.243	0.43	0.571	0.571	0.104	0.101	0.128	0.05	0.032			
i4Pvalb		0.25	0.05	0.05	0.05	0.437	0.451	0.03	0.22	0.088	0.091	0.03	0.03				
i4Sst	0.241	0.25	0.05	0.05	0.05	0.351	0.857	0.082	0.77	0.026	0.03	0	0.03				
i4Htr3a		0.25	0.05	0.05	0.05	0.351	0.02	0.625	0.028	0	0.03	0.03	0.03				
E5	0.017	0.021	0.05	0.05	0.05	0.007	0.05	0.05	0.05	0.116	0.083	0.063	0.105	0.047	0.03	0.03	0.03
i5Pvalb	0	0	0.102			0	0.034	0.03	0.03	0.455	0.361	0.03	0.22	0.03	0.01	0.01	0.01
i5Sst	0.203	0.169		0.017		0.056	0.03	0.006	0.03	0.317	0.857	0.04	0.77	0.03	0.01	0.01	0.01
i5Htr3a						0.03	0.03	0.03	0.03	0.125	0.02	0.625	0.02	0.03	0.01	0.01	0.01
E6		0				0				0.012	0.01	0.01	0.01	0.026	0.145	0.1	0.1
i6Pvalb		0.1				0.1				0.1	0.03	0.03	0.03	0.1	0.08	0.1	0.08
i6Sst										0.03	0.03	0.03	0.03	0.1	0.05	0.05	0.05
i6Htr3a										0.03	0.03	0.03	0.03	0.1	0.05	0.05	0.03





(caption next page)

Figure 1 ***(previous page)*****: V1 model of [Billeh et al., 2020]. (A)** The model consists of 4 classes of neurons on 5 layers. It comes together with a model for LGN, that transforms visual inputs into input currents to neurons in the V1 model. The LGN model receives visual input from an oval in the central part of an image [Billeh et al., 2020]. **(B)** The model contains one excitatory and three inhibitory neuron classes. Each dot denotes the position of a neuron. **(C)** The data-based base connection probabilities of [Billeh et al., 2020] depend on the cell class to which the presynaptic (row labels) and postsynaptic neuron (column labels) belongs. White grid cells denote unknown values. **(D)** The base connection probability from **(C)** is multiplied according to [Billeh et al., 2020]) for any given pair of neurons by an exponentially decaying factor that depends on the lateral distance between them. Panels A, C, D are reproduced from [Billeh et al., 2020]. **(E)** Spike outputs of 2 randomly selected neurons from the Billeh et al. model for 10 trials with the same input (a trial of visual change detection task for natural images), using the noise model of Billeh et al. **(F)** Same as in **(E)** but for the version of the data-based noise model with $s = q = 2$ that we used as default-noise model during testing. It causes substantially larger trial-to-trial variability.

data on area V1 that is currently available. We will refer to the point neuron version of this model, also in combination with our data-based noise model that we discuss below, as the Billeh et al. model.

The Billeh et al. model is a spatially structured model for a patch of V1 that consists of 51,978 neurons from four main classes: One class of excitatory neurons and three classes of inhibitory neurons (Fig. 1B) that are distributed over 5 horizontal layers of neurons, labeled as L1, L2/3, L4, L5, and L6. Synaptic connections between these neurons are generated from data-based connection probabilities. These are defined in terms of base connection probabilities (Fig. 1C) that depend on the class and layer of the pre- and postsynaptic neuron. These base connection probabilities are scaled for each concrete pair of neurons by an exponentially decaying function of the lateral distance between their somata (Fig. 1D). This distance-dependent scaling entails that the vast majority of synaptic connections are between nearby neurons, and hence that the total wire length is small. But it also impacts the specific style of computational processing in the V1 model: Information is not continuously spread out all over the network as in randomly connected recurrent neural networks, that are frequently used as models for neural networks of the brain. To transform images and movies into input currents to neurons in this V1 model, we employed the preprocessing module (LGN model) of [Billeh et al., 2020], see Fig. 1A. It consists of 17,400 filters that model in a qualitative manner the responses of four classes of experimentally observed LGN neurons (sustained ON, sustained OFF, transient ON/OFF, and transient OFF/ON), that are further subdivided according to preferred temporal frequencies.

Individual neurons are modeled as point neurons. But in contrast to the customary leaky integrate-and-fire (LIF) neuron models, the Billeh model employs 111 different variations of the LIF model, which are referred to as generalized leaky integrate-and-fire ($GLIF_3$) neuron models because they have in addition to the membrane potential two other internal variables that model after-spike currents in the neuron on slower time scales. These 111 different neuron types have been fitted to experimental data for 111 selected neurons from the neocortex according to the cell database of the Allen Brain Atlas [Allen Institute, 2018].

The neurons in the model of [Billeh et al., 2020] received besides inputs from the LGN model and inputs from other neurons also a small noise current. This noise was generated by a single Poisson source for all neurons. Hence this noise is highly correlated, but its amplitude is so small that it has only little impact on neural firing (Fig. 1E). We used this noise model during training, but instead used a data-based noise model during testing. This noise model is based on experimental data from area V1 of the awake mouse [Stringer et al., 2019]. More precisely, we employed the heavy-tailed distribution of noise amplitudes that arises from their experimental data (Fig. S1). Furthermore, we superimposed two forms of noise: A quick form noise with scaling factor q where a new value is drawn every ms from this distribution, mimicking for example noise that arises from stochastic synaptic release, and a slow form of noise with scaling factor s, where a new value is drawn from this heavy-tailed distribution once at the beginning of

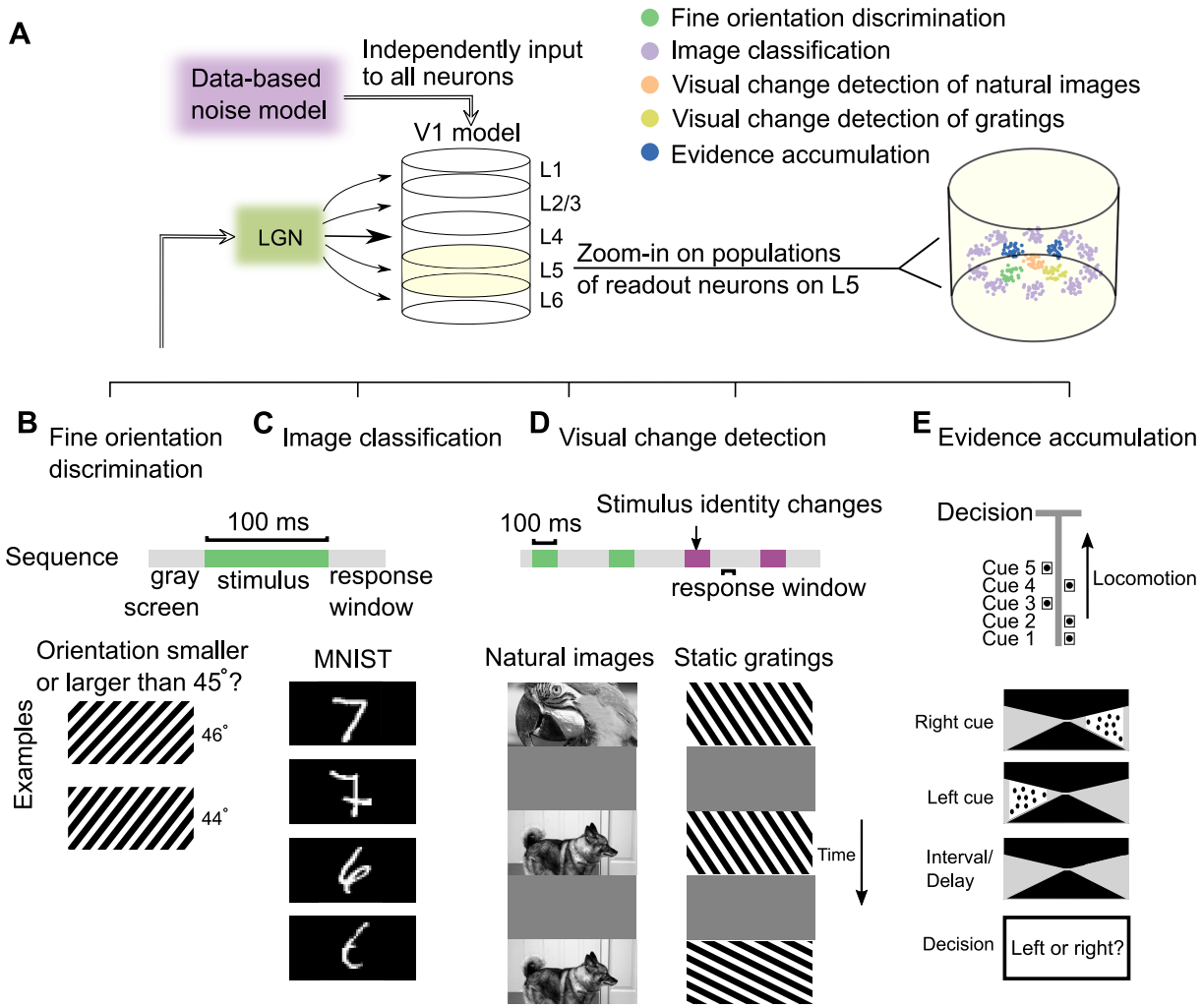
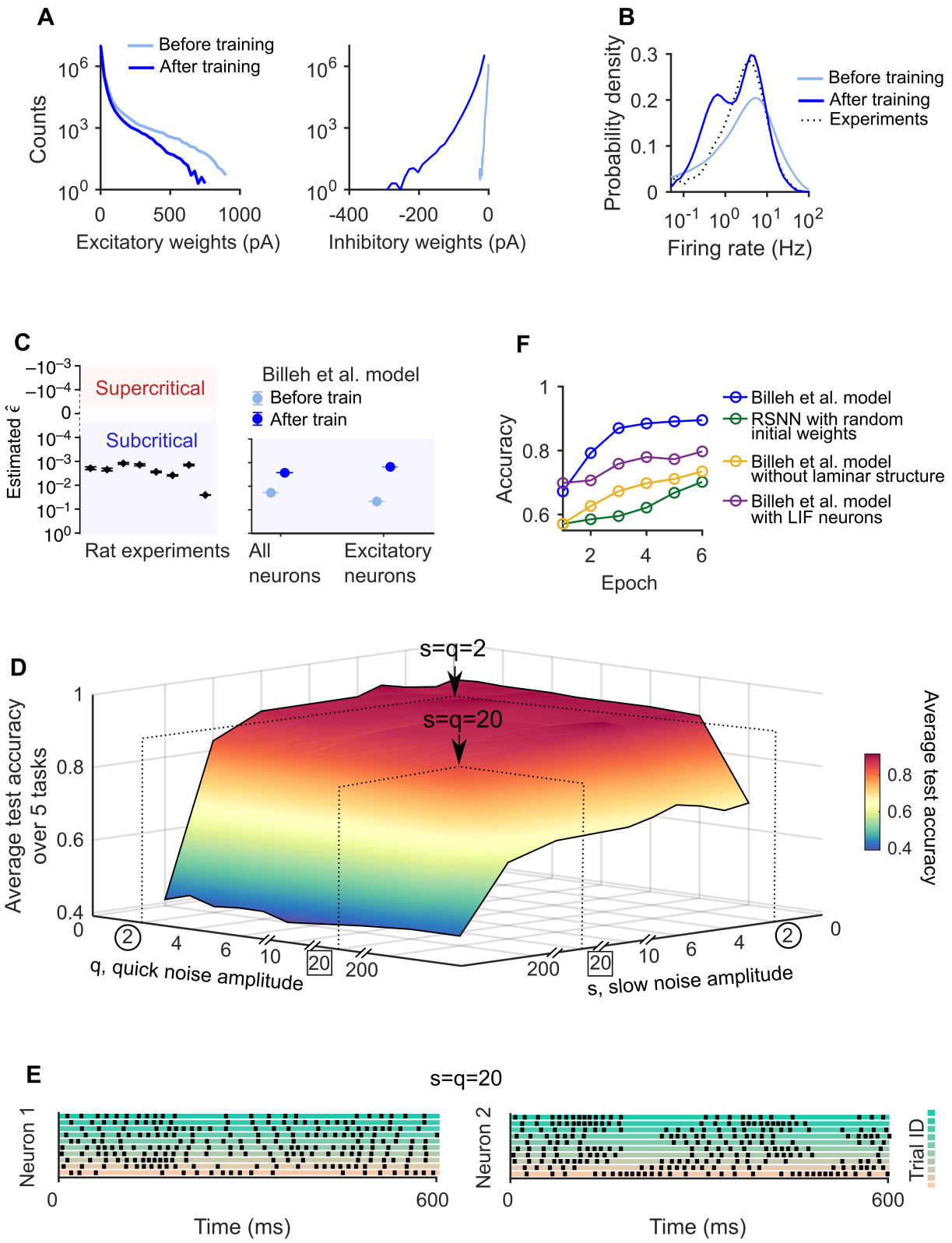


Figure 2: Illustration of the readout convention and the 5 visual processing tasks for which the V1 model was trained. (A) The visual stimuli for all 5 tasks were presented to the LGN model. Separate competing pools of pyramidal cells on L5 were chosen as readout neurons for each task. **(B-E)** Schematic diagrams and timings of 5 visual tasks (Methods). **(B)** In the fine orientation discrimination task, the network received a drifting grating with an orientation very close to 45°, and neurons in the corresponding readout pool had to fire if the orientation was larger than 45°. **(C)** For the image classification task, the network received a handwritten sample of a digit from 0 to 9 from the MNIST data set, and the corresponding one of the 10 readout pools for this task had to fire stronger than the others (two samples for digits 7 and 6 are shown). **(D)** For the visual change detection task, a long sequence of images was presented, with gray screens in between. A corresponding readout pool had to become active during the response window if the most recent image differed from the preceding one. Both natural images and static gratings were used. **(E)** In the evidence accumulation task, 7 cues were presented sequentially, and after a delay, a corresponding readout pool had to indicate through stronger firing whether the majority of the cues had been presented on the left or the right.



(caption next page)

Figure 3 ***(previous page)*****: Analysis of the V1 model after training. (A)** Distributions of excitatory weights (left) and inhibitory weights (right) before and after training. **(B)** Average distributions of firing rates for the 5 task before and after training. The distribution is moved through training closer to the one recorded in V1 [Siegle et al., 2021]. **(C)** Criticality of the Billeh model is analyzed and compared with experimental data. The y-axis shows estimates of $\hat{\epsilon} = 1 - \hat{m}$, where $\hat{m}$ is the estimated branching ratio. Estimates of this value in the rat brain from [Wilting and Priesemann, 2018a] are reproduced on the left. The Billeh et al. model produces almost the same branching ratio as the brain, especially after training. Error bars on the left represent 16 to 84% confidence intervals. Error bars on the right represent SEM over 10 trials. **(D)** The average accuracy of the V1 model on test data for the 5 tasks, as function of the amplitude q of quick and the amplitude s of slow noise. The model had been trained just with the default noise of the Billeh model, which has much less impact on neural activity according to Fig. 1 E, F. The arrow in the back points to the accuracy for the default values $s = q = 2$ for the data-based noise model. One sees that the average accuracy for the 5 tasks is also robust to much larger noise amplitudes, e.g. for $s = q = 20$, see arrow in front, it still has an average accuracy of 83.07%. The resulting trial-to-trial variability of neural firing is substantial for this noise level, see **(E)** for samples of spiking activity of the same 2 neurons as in Fig. 1E, F for 10 trials with the same network input (image). **(F)** Average test accuracy across 5 tasks of the V1 model and of control models that lacked salient structural features of the V1 model; plotted as a function of the number of training epochs (see Methods). Deleting salient structural features of the V1 model caused substantially slower progress of stochastic gradient descent training.

each trial. The latter mimics the well-known dependence of neural responses to the state of the network at the beginning of a trial, see e.g. [Arieli et al., 1996]. We use the default values $s = q = 2$ for scaling these two forms of noise. The resulting noise model causes a qualitatively similar trial-to-trial variability of network responses in the V1 model as in the brain (compare Fig. S2B with Extended Data Fig. 5 of [Stringer et al., 2019]). This trial-to-trial variability is shown for 2 sample neurons from the V1 model in Fig. 1F. The resulting Fano factor of spike counts in 10 ms windows has then a value of 1.46 in the V1 model, which is close to the measured value of 1.39 in mouse V1 [Montijn et al., 2014]. To get a clearer picture of the noise robustness of the V1 model, we tested its computational performance also for substantially larger values of the scaling factors q and s (Fig. 3D).

2.2 The V1 model can multiplex diverse computations on visual input streams

Classification of static images is a very popular test for neural networks. But brains have visual processing capabilities that go far beyond that, since visual information arrives in natural environments, especially in the presence of active vision, in a piecemeal manner. Hence brains need to be able to integrate temporally dispersed information, which can in general not be carried out by a feedforward neural network. Therefore we tested our model not only on a standard image classification task (handwritten digits from the MNIST dataset), but also on 4 tasks that require temporal integration of visual information. The latter ones have all been used in mouse experiments, and data on their behavioral performance are available. The 5 selected tasks are illustrated in Fig. 2: Discrimination of subtle differences in the orientation of drifting gratings (Fig. 2B), as in the experiments of [Rumyantsev et al., 2020, Stringer et al., 2021], a generic image classification task (Fig. 2C), visual change detection tasks for natural images and static gratings (Fig. 2D), as considered in [Garrett et al., 2020a, Siegle et al., 2021], and accumulation of temporally dispersed cues on the left and right (Fig. 2E), as considered in [Morcos and Harvey, 2016, Engelhard et al., 2019], some of them with slightly longer delay periods.

In order to test the performance of the V1 model on these tasks, one has to specify a convention for extracting the network decision. One frequently used convention, see e.g. [Haeusler et al., 2009], is to let an external "readout neuron" that receives synaptic input from all neurons in the network produce the network decision (Fig. S4C). Obviously, there are no such readout neurons in the brain that receive

synaptic inputs from all neurons in a patch of the neocortex. Furthermore, this convention is not suitable for probing the computational capability of such a network model. Theoretical results ([Maass et al., 2002]) imply that if the network model is sufficiently large and has diverse units, such readout neurons tend to become computationally quite powerful when its weights have been properly trained, and is likely to mask the computational contribution of the neural network model itself. Therefore we demanded that in V1 model, like in the brain [Harris and Shepherd, 2015], projection neurons within the network extract computational results from the model, and project them to downstream networks. In particular, a large fraction of pyramidal cells on L5 projects to subcortical areas, and can therefore use the computational result of the network to trigger a behavioral response. Therefore we selected for each computational task and each possible outcome of a network decision a population of 30 excitatory neurons on L5 (Fig. 2A). If this population produced more spikes during the response window than competing populations that voted for other outcomes, then the outcome for which it "voted" was viewed as the network decision. One important difference to the convention of using global readout neurons is that the set of neurons in the network that provides synaptic inputs to a neuron within the network is substantially smaller and spatially constrained.

With the values of synaptic weights provided by [Billeh et al., 2020], the V1 model is incapable of performing any of the 5 tasks; the accuracy is close to chance level. We then applied stochastic gradient descent, like in [Scherr and Maass, 2021], to the synaptic weights of connections within the V1 model of Billeh et al., and to connections from the LGN model to the V1 model. No synaptic connections were added during this process. We also made sure that the signs of synaptic weights could not change, i.e., we maintained the validity of Dale's law. We used a loss function for gradient descent that penalized inaccurate decisions by the chosen populations of readout neurons, see Methods. The loss function also penalized biologically unrealistic high firing rates. Stochastic gradient descent was implemented through a variation of BPTT, with the help of a suitable pseudo derivative for handling the discontinuous dynamics of spiking neurons, as suggested by [Bellec et al., 2018]. To avoid artifacts arising from gradient descent caused by the hard reset of $GLIF_3$ neurons after a spike, we subtracted instead a fixed value from the membrane potential after each spike. Control experiments show that this modification causes no significant difference in the spike output of a neuron (Fig. S5). We trained the model of Billeh et al. with its original noise model (Fig. 1E), and tested its performance with the biologically more realistic noise model of Fig. 1F, and also in the presence of even more noise.

After training, the Billeh et al. model achieved on all 5 tasks a performance that was in the same range as reported behavioral data (Table 1), with an average accuracy of 89.10%. This performance did not depend on our particular choice of readout neurons in L5: Choosing randomly distributed instead of co-located pyramidal cells yielded an average accuracy of 91.56%. As expected, choosing instead global linear readout neurons for each task led to a substantially higher accuracy of 97.73%. The differences of readout scenarios can explain why behavioral performance lags behind neural coding fidelity in area V1 [Stringer et al., 2021] (Supplementary Note 1). Sample computations of the V1 model for each of the 5 tasks are shown in Fig. S6-S10.

The median strength of inhibitory synapses increased from 0.03 to 1.65 pA during training; the median weight of excitatory synapses decreased from 2.96 to 1.43 pA, see Fig. 3A, and Fig. S3 for more detailed analyses in terms of the neuron types involved. Also, the distribution of neural firing activity was after training still close to the measured distribution in the brain, see Fig. 3B, with an average firing rate of 4 Hz.

Experimental data suggest that neural networks of the brain typically operate in a critical regime [Mora and Bialek, 2011, Shew and Plenz, 2012, Wilting and Priesemann, 2019]. We evaluated the criticality of the V1 model by measuring its branching ratio of neural activity, as suggested by [Wilting and Priesemann, 2019]. We found that both the untrained and the trained V1 model operates in a slightly subcritical regime. Training moved the model somewhat closer to the critical regime, reaching values of the branching

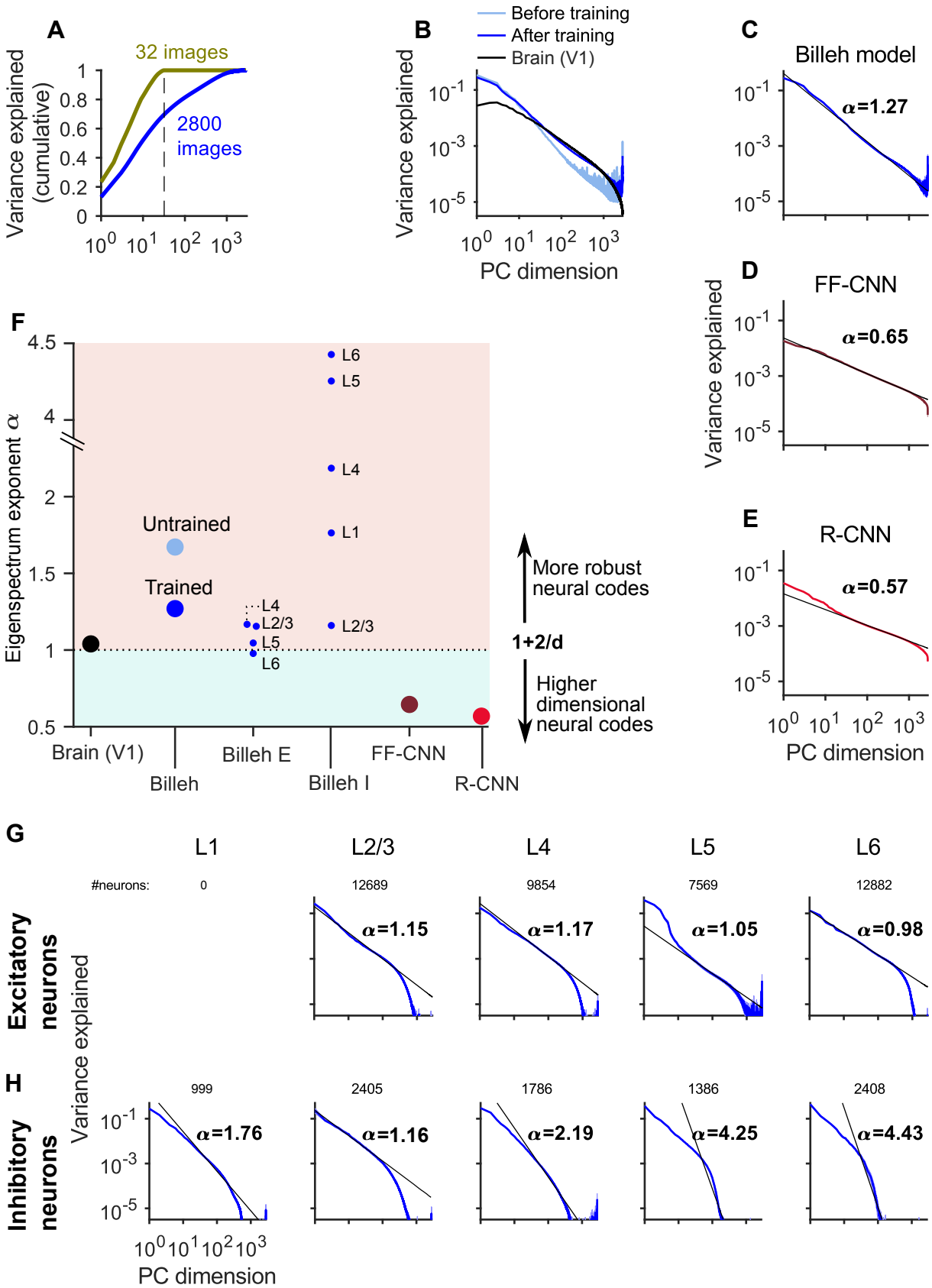
ratio that almost perfectly matched recorded data from the brain (Fig. 3C). Hence the V1 model operated in a dynamic regime that closely matches experimental data.

Interestingly, when the same training procedure was applied to control models that lacked salient structural features of the Billeh et al. model, their task performance advanced substantially slower, see Fig. 3F and Supplementary Note 2. This result suggests that the laminar cortical circuitry of V1 supports the efficiency of stochastic gradient descent. In particular, slow internal processes within GLIF_3 neuron models contribute to it because gradients move more effectively through the corresponding slowly changing internal variables than through spikes. We propose that the dominance of short connections within and between laminae of laminar cortical microcircuits in the brain also supports gradient descent learning, since local errors in computation are not immediately and continuously spread out all over the network like in randomly connected networks without this characteristic architecture of cortical microcircuits.

We tested the robustness of the resulting multiplexed visual processing capability of the V1 model after training by exposing its neurons to substantially larger amplitudes q and s of the data-based noise model, although it had never been exposed to such noise during training (where we only applied the really small noise considered in [Billeh et al., 2020]). Surprisingly, it is almost impossible to destroy its multiplexed visual processing capability (Fig. 3D): It remained stable even when the amplitudes q and s of quick and slow noise were increased by several orders of amplitude.

2.3 The power spectrum of neural codes provides an explanation for the astounding robustness of visual processing by the V1 model

An explanation for the robustness of visual processing in area V1 has been provided by [Stringer et al., 2019]. They verified through large-scale recordings from V1 in mouse a theoretically predicted link between noise robustness of visual processing and neural codes for images. They found that V1 employs high-dimensional neural codes for images, but the power of higher PCA components decays sufficiently fast so that its neural codes remain noise-robust. More precisely, they introduced a cross-validated principal component analysis that provides unbiased estimates of the stimulus-related variance. They found that the amount of explained variance continues to increase as further PCA dimensions were included without saturating below the dimensionality (= size) d of the image ensemble. We applied exactly the same analysis to the trained model of Billeh et al., and found that the model exhibited the same coding property (Fig. 4A). It was also shown in [Stringer et al., 2019] that the explained variance of the nth principal component of network representations of images follows a power-law $n^{-\alpha}$. The exponent α characterizes how fast the variance that is explained by higher PCA dimensions decays. Their theoretical analysis predicts that $\alpha = 1 + 2/d$ is the optimal value (Fig. 4F), since this value provides a theoretically optimal compromise between encoding too many details (leading to smaller values of α), and keeping the neural code robust to perturbations (leading to larger values of α). Other theoretical work [Ghosh et al., 2022] also predicts that a value of α close to 1 enhances under mild conditions downstream generalization performance. In-vivo recordings of [Stringer et al., 2019] found that the value of α for primary visual cortex of mouse is actually close to this optimal value $\alpha = 1 + 2/d$. This neural coding property of area V1 in the brain has gained additional interest through the contrasting result of [Kong et al., 2022]. They found that feedforward CNNs, which are viewed to be substantially less noise robust than the brain, have in fact a smaller α, as predicted by the theory. Hence we wondered whether our more brain-like neural network model for visual processing would exhibit a value that is closer to the theoretical optimum. We applied for that purpose the same measurement procedure as [Stringer et al., 2019] to the V1 model, for a set of 2,800 randomly drawn natural images. Fig. 4B shows the eigenspectrum of PCA component for the model of Billeh et al. before and after training, and also the measured eigenspectrum of V1 responses from [Stringer et al., 2019]. One sees that the eigenspectrum of the model is already before training quite close to that of the brain, and is moved by training even closer. The resulting exponent α of the



(caption next page)

Figure 4 ***(previous page)*****: Comparing the eigenspectrum of neural codes in the V1 model of Billeh et al. with data from the brain, and with CNNs. (A)** As in the brain, the cumulative fraction of explained variance saturates only at the dimension of the input ensemble, here shown for $d = 32$ and 2,800 natural images. **(B)** Eigenspectra of the untrained/trained V1 model of Billeh et al. with an ensemble of 2,800 randomly chosen natural images, and for mouse V1 [Stringer et al., 2019]. **(C, D, E)** Exponents of the power law for the V1 model of Billeh et al., for feedforward CNNs, and for recurrent CNNs; all for the same ensemble of 2,800 randomly drawn natural images. **(F)** Summary result for exponents of the power law for the brain the V1 model of Billeh et al., and for CNNs. **(G)** Eigenspectra of excitatory neurons on different layers of the model by Billeh et al. exhibiting values close to the measured value from a large sample of neurons in V1. **(H)** The same for inhibitory neurons. These exhibit on all layers except L2/3 substantially smaller coding fidelity.

power law (Fig. 4C) is for the model somewhat higher than in the brain. Fig. 4F, G, H suggest that this is largely due to the contributions of inhibitory neurons. They are generally found to have less precise neural codes for sensory stimuli, and consistent with that, their eigenspectra decayed substantially faster in the V1 model.

We have also reproduced the result of [Kong et al., 2022] that feedforward CNN has a substantially smaller value of α (Fig. 4D) and found that the recurrent CNN model of [Wang and Hu, 2021], which also achieves very high accuracy for image classification, has an even smaller value of α (Fig. 4E). Altogether our results imply that the V1 model has, unlike CNNs, similar neural coding properties as area V1 in the brain. Furthermore, these can be linked according to the theory of [Stringer et al., 2019] to its remarkable noise robustness.

2.4 Comparing noise robustness and OOD generalization of the V1 model and CNNs

Our preceding analyses of neural coding in the V1 model of Billeh et al. and CNNs suggest that the former is more noise-robust. Since it is hard to compare their robustness to noise within the networks with that of CNNs, because their computational units are so different, we compared instead their robustness to noise in the visual input, concretely to Gaussian pixel noise that was added to handwritten digits from the MNIST dataset. We used Gaussian noise with mean 0 and different SD. We first trained each type of neural network on the original dataset without noise, and then tested their classification performance on images with noise, see Fig. 5A for samples. Fig. 5B shows that the classification performance of the V1 model is substantially more robust to the added noise during testing, as predicted by the preceding analysis of the different neural coding strategies of the V1 model and CNNs.

Since neither of these networks had been trained with the noisy images for which they were tested, we are analyzing here a particular OOD (out-of-distribution) generalization capability of the V1 model and of CNNs. Fig. 5C demonstrates that even if CNNs are trained for a particular noise statistics, they do not perform well if they are tested on images with a different SD of Gaussian noise. In contrast, the V1 model exhibited perfect OOD generalization in this respect. Also, the remarkable robustness of the V1 model to noise within the network (Fig. 3D), which had not been present during training, can be viewed as an OOD generalization capability.

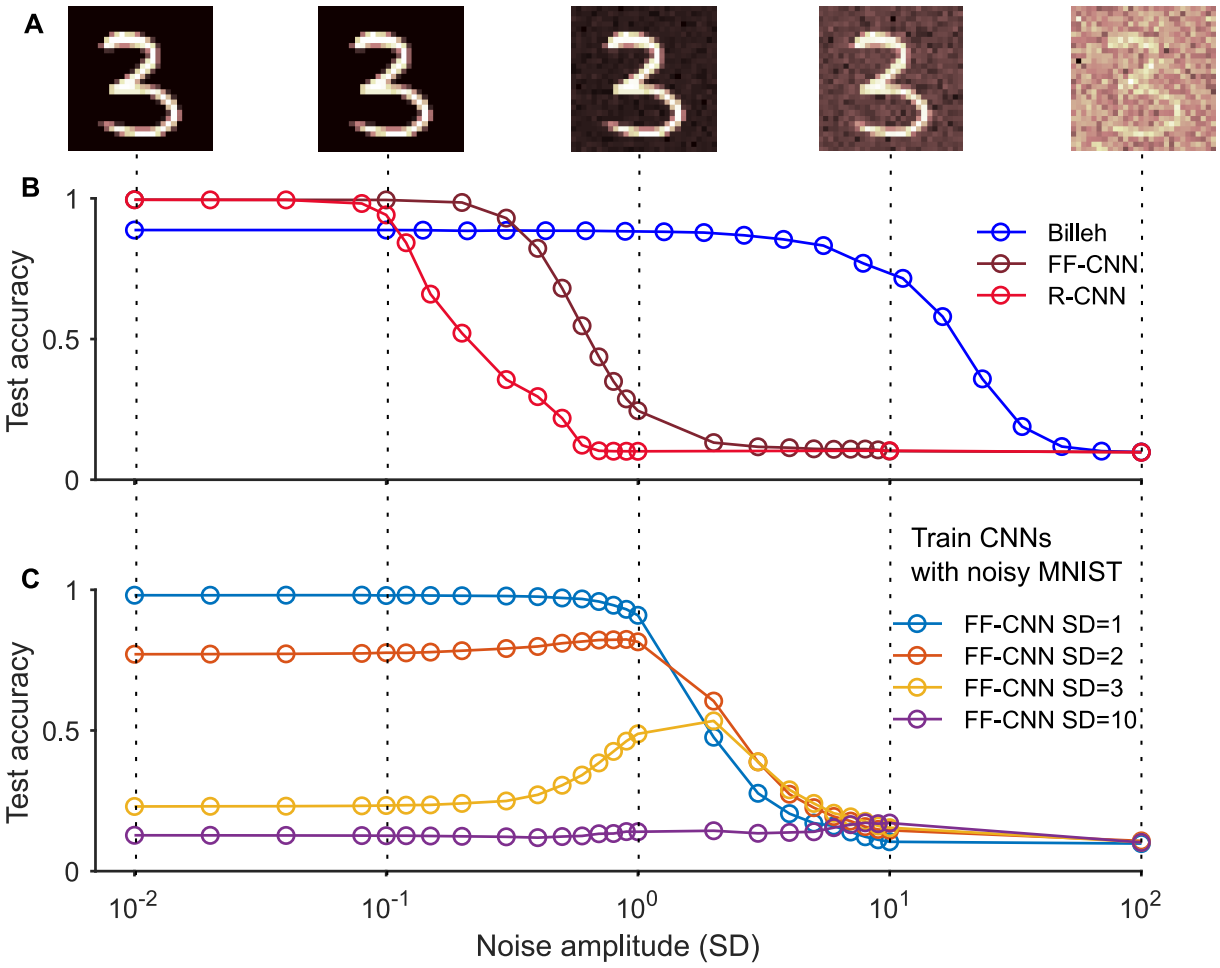


Figure 5: Robustness of the V1 model and of CNNs to noise in images. (A) Samples of handwritten MNIST digits with Gaussian noise drawn independently from $\mathcal{N}(0, \mathrm{SD})$ for each pixel, for different values of the SD. **(B)** While the V1 model never quite reaches the peak performance of the CNNs, it tolerates noise with fairly high SD, whereas the performance of feedforward (FF)-CNNs and recurrent (R)-CNNs is substantially degraded even by noise with small SD. **(C)** Even when CNNs are trained on images with a particular noise statistics (SD), they do not generalize well to test images with a different value of SD. Furthermore, they do not achieve for SD between 1 and 10 the same noise robustness as the V1 model even when they were trained on images with that type of noise.

2.5 Neural coding dimensions for visual inputs are in the trained V1 model largely orthogonal to noise, like in the brain

A further explanation for robust coding capabilities of the brain was provided by experimental data of [Rumyantsev et al., 2020]. They reported that dimensions in which differences between visual stimuli were encoded in the brain were nearly orthogonal to the largest noise mode, which therefore had little effect on coding fidelity. We carried out the same analysis for the V1 model, using the same visual stimuli (moving gratings with orientation differences that were close to the perception threshold). The definition of the discriminability index d', according to [Rumyantsev et al., 2020] a proposed measure for the fidelity of neural population coding, is illustrated in Fig. 6A. Neural population responses $\mathbf{r}_{\mathrm{A}}(t)$ and $\mathbf{r}_{\mathrm{B}}(t)$ to two stimuli A and B form two distributions (ellipses). PLS (partial least square) analysis projects them onto a subspace where they become most distinct. The discriminability d' is defined as the separation, $\Delta\boldsymbol{\mu}$, of the two distributions along the dimension orthogonal to the optimal boundary (green line) for classifying stimuli in this subspace, divided by the SD of each distribution along this dimension.

The eigenvalues of the noise covariance matrix are plotted in Fig. 6B as function of the number of neurons that are sampled in the V1 model. We found that also in the V1 model the projection of the signal difference $\Delta\boldsymbol{\mu}$ onto the eigenvectors for the largest noise eigenvalues is relatively small (Fig. 6C). Furthermore, compared with the untrained Billeh model, training of this model moved the signaling dimensions to become more orthogonal to the largest noise dimension, see Fig. 6D. The projection of $(d')^2$ on an eigenvector can be interpreted as the signal-to-noise ratio because it is the ratio of the projected signal difference and noise eigenvalue (Methods).

It had been argued in [Rumyantsev et al., 2020] that the amount of visual information that is encoded in area V1 reaches a ceiling due to noise correlations. This result appeared to contradict the results of [Stringer et al., 2019]. A possible explanation of this discrepancy was offered in [Stringer et al., 2021]: They conjectured that the seemingly limited coding capability of V1 could be explained by the relatively small number of up to 1,300 neurons from which simultaneous recordings had been carried out in [Rumyantsev et al., 2020]. They suggested that the apparent "ceiling" would rise if one records from more neurons. We tested this hypothesis in the V1 model of Billeh et al., and found that indeed the corresponding measurement d' for the total amount of encoded information keeps increasing, although at a somewhat slower rate, when the number of neurons from which one records in the model rises from 1,300 to 51,978, see Fig. 6E.

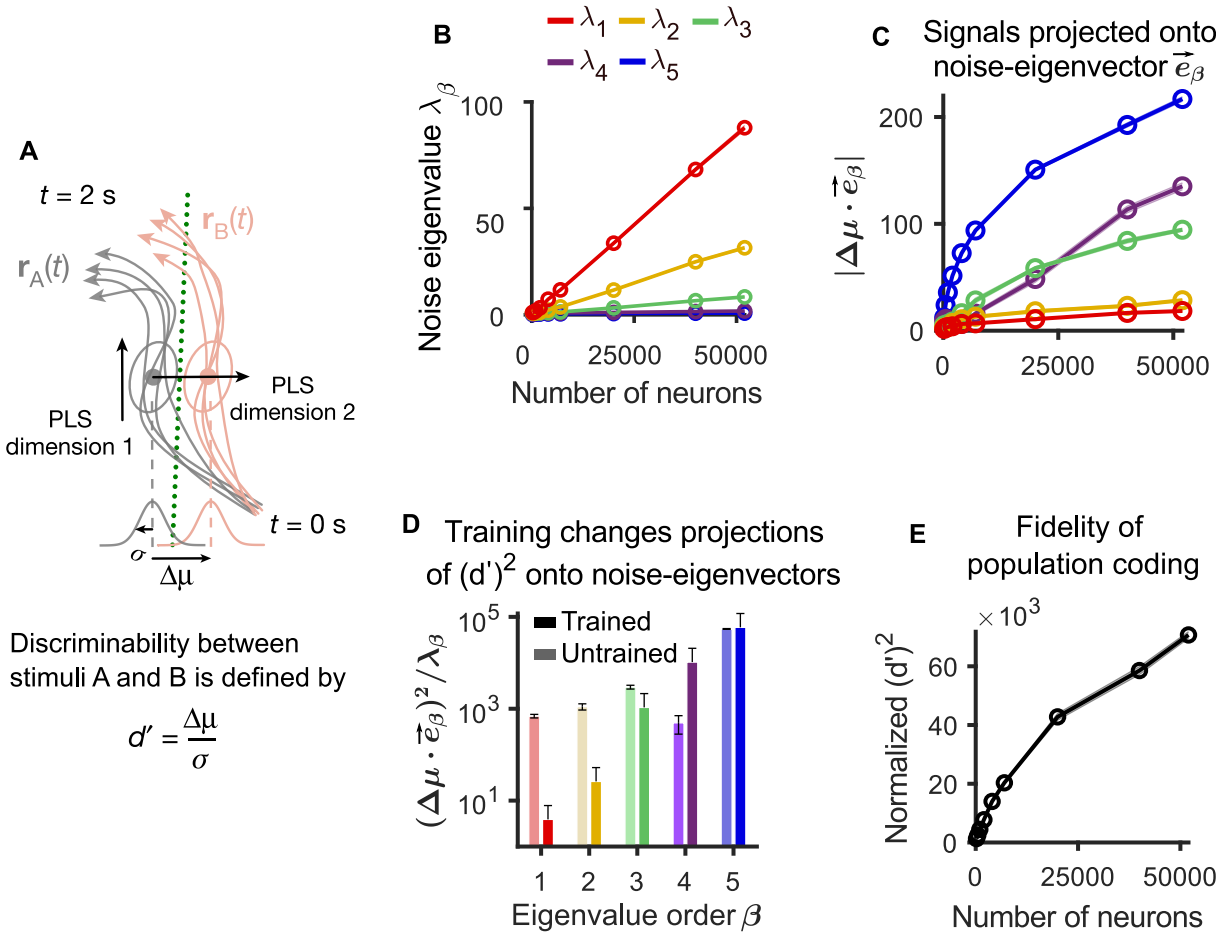


Figure 6: Analyses of relations between signaling and noise dimensions in the V1 model, and of the impact of noise correlations when one considers larger numbers of neurons. (A) Schematic of the calculation of the discriminability index d' according to [Rumyantsev et al., 2020]. **(B)** The 5 largest eigenvalues λ_β ($\beta = 1, 2, \cdots, 5$) of the noise covariance matrix in the trained V1 model increase linearly with the number of sampled neurons. **(C)** Neural coding dimensions ($\Delta\boldsymbol{\mu}$) in the trained V1 model were nearly orthogonal to the dominant eigenvectors of the noise covariance matrix, $\vec{e}_\alpha$, invariant to neuron numbers. **(D)** Training moved the neural coding dimensions so that they became more orthogonal to the dominant noise dimensions. **(E)** The squared discriminability index $(d')^2$ kept increasing when applied to more neurons om the V1 model. d' values were normalized by those obtained for trial-shuffled data (averaged across 1 s). Shaded areas in **(B, D, E)** and error bars in **(C)** represent the standard error of mean (SEM) over 100 trials.

3 Discussion

We have demonstrated that the neural network model of [Billeh et al., 2020], which arguably provides the largest currently available accumulation of anatomical and neurophysiological data on area V1 in the mouse brain, provides not only a window into brain dynamics, but also into visual processing capabilities and neural coding properties that are entailed by these data. CNNs are currently in AI one of the most commonly considered types of neural networks for visual processing. They were also inspired by some aspects of visual processing in the brain, especially the existence of simple and complex cells in area V1. But a closer look shows that they differ from V1 in the brain in almost all other respects: With regard to their computational units (artificial neurons in CNNs versus spiking neurons in the brain), the diversity of their units (very few simple units versus a large diversity of neurons with different temporal dynamics), their large-scale architecture (usually feedforward, versus recurrent with laminar structure), their small-scale architecture (very simple network motifs versus a complex combination of feedforward and recurrent processing in cortical microcircuits), and total wire length (almost quadratic versus just linear growth with the number of processing neurons). But V1 in the brain also differs from CNNs with regard to two important visual processing capabilities: The brain can multiplex strategies for solving diverse visual tasks within the same network, in particular also tasks that require integration of sequentially arriving visual information. In addition, visual processing in the brain is very noise-robust, also to new types of noise (OOD generalization). We have shown here that the previously listed fundamental differences between the structure of V1 in the brain and CNNs are causally related to these two superior visual processing capabilities of the brain: The V1 model of [Billeh et al., 2020], that integrates these structural features of V1 in the brain, is able to multiplex robust visual processing.

Since we can reproduce now these two important functional capabilities of area V1 in a model, we have a new research platform at our disposal for studying how neural coding properties of the brain emerge from its anatomical and neurophysiological features, and how they are related to its visual processing capabilities. We have demonstrated here the feasibility of this new research strategy by applying to the V1 model, an analysis of its neural coding that had already been used successfully for elucidating neural codes for images in area V1 of the brain: We analyzed the eigenspectrum of the explained variance of principal components of its neural codes for images. We found that the listed structural features of V1 in the brain do in fact induce a salient feature of its neural coding strategy: The V1 model exhibits a similar power law for neural codes of images as the brain. In contrast, CNNs have a power law with a substantially slower decay of the eigenspectrum. According to the theoretical analysis of [Stringer et al., 2019], this implies that neural codes in CNNs are less noise-robust. Our V1 model also demonstrates that a further salient aspect of neural coding in area V1 of the brain results from its anatomical and neurophysiological details: According to [Rumyantsev et al., 2020] about 90% of the noise fluctuations in area V1 of the brain are constrained to dimensions of the population activity that are orthogonal to noise dimensions. We found that this is in fact an emergent property of the V1 model of [Billeh et al., 2020].

On a more general level, we have shown that the V1 model of [Billeh et al., 2020] can be seen as the first prototype of a new generation of neural network models for visual processing that capture substantially more features of brain processing than CNNs. Further work is needed to tease apart the functional implications of each of its structural features, and to port similar advanced brain-like visual processing capabilities into simpler neural network models. Often one uses instead of data-based models for neural networks of the brain randomly connected recurrent networks of strongly simplified neuron models. In our experience, neural coding and computational properties of recurrent neural networks vary substantially in dependence of their connectivity structure and neuron models. This highlights the need to test brain-like features not only in abstract models, but also in neural network models that integrate our available knowledge about the actual structure of these neural networks in the brain.

We propose that our method can also be applied to investigate how anatomical and neurophysiological details of interconnected higher and lower brain areas carry out distributed computations, in particular how higher cortical areas enhance visual processing capabilities of V1. A substantial body of anatomical and neurophysiological data on higher cortical areas and their connectivity to V1 is currently available for that, see e.g. [Siegle et al., 2021, Campagnola et al., 2021, Iavarone et al., 2022]. We expect that deficits in visual processing capabilities of the V1 model, such as limited spatial integration of image features and a relatively short working memory time span, will disappear when the V1 model is combined with models of higher brain areas.

The analysis of neural coding in the V1 model has produced a number of predictions for future biological experiments. In particular, we have shown in Fig. 6E that correlated noise reduces the coding fidelity of the network, but does not produce an a-priori bound for its sensory discrimination capability (this had already been hypothesized by [Stringer et al., 2021]). A further prediction of our model is that the PCA eigenspectrum of neural codes for inhibitory neurons does not obey a power law for higher dimensions, see Fig. 4H. Finally, the values of excitatory synaptic weights in V1 are predicted to generally shrink through training (Fig. 3A left), while inhibitory weights are predicted to become stronger (Fig. 3A right).

Visual processing in the brain exhibits also with regard to two aspects of physical implementations substantial advantages over CNNs: Most synaptic connections in V1 are between nearby units, which is essential for an efficient physical realization of synaptic connections in neuromorphic hardware. This architectural feature is also likely to support faster learning (Fig. 3F). In addition, computations are carried out in the V1 model through event-based processing with very sparse firing activity. This computing regime is not only very energy efficient, but it also supports computations on tasks where temporal aspects play an important role, because it allows to let time represent itself in network computations. Hence, our analysis of a neural network model for area V1 of the brain paves the way for substantially more energy-efficient neuromorphic implementations of visual processing [Plank et al., 2022].

4 Methods

4.1 Neuron models

We based our study on the "core" part of the point-neuron version of the realistic V1 model introduced by [Billeh et al., 2020]. To make it gradient-friendly, we replaced the hard reset of membrane potential after a spike emerges with the reduction of membrane potential $z_j(t)v_{th}$, where $z_j(t) = 1$ when neuron j fires at time t and $z_j(t) = 0$ otherwise. v_{th} is the firing threshold of membrane potential. This causes no significant change in the neural response (Fig. S5). We simulated each trial for 600 ms. The dynamics of the modified GLIF_3 model was defined as

$$\begin{aligned} v_j(t+\delta t) &= \alpha v_j(t) + \frac{1-\alpha\tau}{C}\left(I_j^e(t+1) + \sum_m I_j^m(t+1) + gE_{\mathrm{L}} + I_j^{\mathrm{syn}}(t)\right) - z_j(t)v_{\mathrm{th}} \\ z_j(t) &= H\left(v_j(t) - v_{\mathrm{th}}\right) \\ I_j^e(t) &= \sum_i W_{ji}^{\mathrm{in}} x_i(t) + qK_j^{\mathrm{quick}}(t) + sK_j^{\mathrm{slow}}, \end{aligned} \tag{1}$$

where C represents the neuron capacitance, E_L the resting membrane potential, I^e the external current, I^{syn} the synaptic current, g the membrane conductance, and v_{th} the spiking threshold. W_{ji}^{in} is the synaptic weight from LGN neuron i to V1 neuron j. The scales of the quick noise $K_j^{\mathrm{quick}}(t)$ and the slow noise K_j^{slow} to neuron j are $q = 2$ and $s = 2$, respectively, unless otherwise stated. K_j was randomly

drawn from the empirical noise distribution which will be elaborated on later. The decay factor α is given by $e^{-\delta t/\tau}$, where τ is the membrane time constant. δt denotes the discrete-time step size, which is set to 1 ms in our simulations. H denotes the Heaviside step function. To introduce a simple model of neuronal refractoriness, we further assumed that $z_j(t)$ is fixed to 0 after each spike of neuron j for a short refractory period depending on the neuron type. The after-spike current $I^m(t)$ was modeled as

$$I^m(t+\delta t) = f^m I^m(t) + z(t)\delta I^m;\ m = 1, \ldots, N_{\text{asc}}, \tag{2}$$

where the multiplicative constant $f^m = \exp(-k^m \delta t)$ and an additive constant, δI^m. In our study, $m = 1$ or 2. Neuron parameters have been fitted to experimental data from 111 selected neurons according to the cell database of the Allen Brain Atlas [Allen Institute, 2018], see [Teeter et al., 2018, Billeh et al., 2020], including neuron capacity C, conductance g, resting potential E_{L}, the length of the refractory period, as well as amplitudes δI^m and decay time constants k^m of two types of after-spike currents, $m = 1, 2$.

4.2 Synaptic inputs

The Billeh et al. model specifies the connection probability between neurons, based on experimental data. The base connection probability for any pair of neurons from the 17 cell classes is provided in [Billeh et al., 2020] by a table (reproduced in Fig. 1C); white grid cells denote unknown values. The entries in this table are based on measured frequencies of synaptic connections for neurons at maximal 75 µm horizontal inter-somatic distance. This base connection probability was scaled by an exponentially decaying factor in terms of the horizontal distance of the somata of the two neurons (Fig. 1D). This distance-dependent scaling is also based on statistical data from experiments (leaving aside finer details of connection probabilities). The synaptic delay was spread in $[1, 4]$ ms, which was extracted from the Fig. 4E of [Billeh et al., 2020] and converted to integers as the integration step is 1 ms.

The postsynaptic current of neuron j was defined by the following dynamics [Billeh et al., 2020]:

$$I_j^{\text{syn}}(t+\delta t) = e^{-\frac{\delta t}{\tau_{\text{syn}}}} I_j^{\text{syn}}(t) + \delta t e^{-\frac{\delta t}{\tau_{\text{syn}}}} C_j^{\text{rise}}(t) \tag{3}$$

$$C_j^{\text{rise}}(t+\delta t) = e^{-\frac{\delta t}{\tau_{\text{syn}}}} C_j^{\text{rise}}(t) + \sum_i W_{ji}^{\text{rec}} z_i(t) \frac{e}{\tau_{\text{syn}}}, \tag{4}$$

where τ_{syn} is the synaptic time constant, W_{ji}^{rec} is the recurrent input connection weight from neuron i to j, and z_i is the spike of presynaptic neuron i. The τ_{syn} constants depend on neuron types of pre- and postsynaptic neurons [Billeh et al., 2020].

4.3 Initial conditions

The initial conditions of spikes and membrane potentials were zero unless stated otherwise. The initial conditions of $\mathbf{W}^{\text{in}}$ and $\mathbf{W}^{\text{rec}}$ were given by the values in [Billeh et al., 2020] unless stated otherwise.

4.4 Data-driven noise model

The noise currents $K_j^{\text{quick}}(t)$ and K_j^{slow} in Eq. 1 were randomly drawn from an empirical noise distribution. The quick noise $K_j^{\text{quick}}(t)$ was drawn independently for all neurons in every 1 ms; the slow noise K_j^{slow} was drawn independently for all neurons once 600 ms. The empirical noise distribution (Fig. S1) was from the additive noise decoded from experimental data of mice response to 2,800 nature images [Stringer

et al., 2019] (https://figshare.com/articles/Recordings_of_ten_thousand_neurons_in_visual_cortex_in_response_to_2_800_natural_images/6845348). The decoding method was cross-validation principal component analysis (cvPCA) [Stringer et al., 2019] which will be elaborated later. It measures the reliable variance of stimulus-related dimensions, excluding trial-to-trial variability from unrelated cognitive and/or behavioral variables or noise. We collected the variability (additive noise) to form the empirical noise distribution. We refer to the methods and supplementary materials of [Stringer et al., 2019] for a detailed mathematical analysis of this method.

4.5 Readout populations

By default we employed 15 readout populations in the V1 model, whose firing activity during the response window encoded the network decisions for the 5 visual processing tasks. Each population consisted of 30 randomly selected excitatory neurons in layer 5, located within a sphere of a radius of 55 µm, with some distance between these spheres for different readout populations (Fig. S4A). The results were not sensitive to the number of neurons in each population [Scherr and Maass, 2021]. We also considered the case where the neurons in these readout populations were randomly distributed in L5 (Fig. S4B), and the case where each population was replaced by global linear readout neurons which received synaptic inputs from all neurons with activity ($\mathbf{Z}$) in the V1 model, i.e. $\mathbf{Y}_{\text{global}} = \mathbf{W}_{\text{readout}}\mathbf{Z} + \mathbf{B}$, $\mathbf{B}$ is the bias (Fig. S4C).

4.6 Visual processing tasks

We designed details of these 5 tasks to be as close as possible to corresponding biological experiments while keeping them as simple as possible. Only for the image classification task (MNIST) there exist no corresponding mouse experiments.

LGN model. The visual stimuli were preprocessed by the LGN model (Fig. 2A) according to [Billeh et al., 2020] (it is actually meant to model preprocessing by the retina and LGN in a qualitative manner). This LGN model consists of 17,400 spatiotemporal filters that model responses of LGN neurons in mouse to visual stimuli [Durand et al., 2016]. Each filter produces a positive output that is interpreted as firing rates of a corresponding LGN neurons.

According to the requirements of this LGN model, each visual input pixel was first converted to gray-scale and scaled into an interval $[-Int, Int]$, $Int > 0$. The output of the LGN model was injected into the V1 model as external currents, i.e.,

$$I_{\text{sti}} = \boldsymbol{W}^{\text{in}} \cdot \text{LGN}(G_{Int}), \tag{5}$$

where G_{Int} represents images scaled into $[-Int, Int]$ for $Int = 2$.

Fine orientation discrimination task. In mouse experiments, mice were trained to distinguish orientation of drifting grating stimuli [Rumyantsev et al., 2020, Stringer et al., 2021]. The stimuli were presented for 750 ms or longer. To reproduce this task under the limitations of GPU memory, we input drifting grating to the V1 model through the LGN model for 100 ms (Fig. 2B). As in [Stringer et al., 2021], stimuli were sinusoidal drifting gratings (spatial frequency, 0.05 cycles per degree, drifting rate, 2 Hz). Both in the training and testing processes, the orientation was uniformly drawn from $[43, 47]^\circ$ (i.e., 45 ± 2) with the precision of 0.1°. The orientation difference was the same as in [Stringer et al., 2021]. The initial phase was randomly sampled. The simulation sequence included 50-ms delay, 100-ms drifting gratings, and 50-ms response window in order.

In the response window, we defined the mean firing rate of readout population as

$$r_{\text{readout}} = \frac{1}{T_{\text{resp}} \cdot N_{\text{readout}}} \sum_{t=1}^{T_{\text{resp}}} \sum_{j=1}^{N_{\text{readout}}} z_j(t), \tag{6}$$

where the sum over j is over the $N_{\text{readout}} = 30$ readout neurons and the sum over t is over the time length of response window $T_{\text{resp}} = 50$ ms. If $r > r_0 = 0.01$, then this reported a network decision that the orientation was larger than 45°. Otherwise, it reported that the orientation was smaller than 45°.

Image classification task. To demonstrate that the V1 model is also able to classify images, we included the task to classify handwritten digits from 0 to 9 from the MNIST dataset (Fig. 2C). The timing of input images and response windows was the same as in the preceding task. The task was to decide which digit was denoted by the handwritten image (two samples for 7 and 6 are shown in Fig. 2C). Each of the 10 readout populations for this task was assigned to one of the 10 digits. The network decision was taken to be that digit for which the readout population fired most strongly during the response window.

Visual change detection task with natural images. In mouse experiments [Garrett et al., 2020b, Siegle et al., 2021], mice were trained to perform the visual change detection task with natural images. A sequence of static natural images (250 ms), interleaved by short phases (500 ms) of gray screens, was presented as visual input; mice had to report whether the most recently presented image was the same as the previously presented one. To reproduce this task under the limitations of GPU memory, we presented natural images for 100 ms each, with the gray delays between them lasting for 200 ms (Fig. 2D). Note that the first image was presented after 50 ms. All images were selected from a set of 40 randomly chosen images from the ImageNet dataset [Deng et al., 2009]. The probability that the next image differed from the preceding one was set to 50%. In case of a changed image identity, the model had to report within a time window of 50 ms length that started 150 ms after image onset (response window). If the mean firing rate of the readout population in the response window $r_{\text{readout}} > r_0$, it reported a network decision that the image had changed. The computation of the V1 model on this task has been further analyzed in [Scherr and Maass, 2021].

Visual change detection task with drifting gratings. We also replaced the natural images above with static gratings which have different orientations and kept the input sequence the same (Fig. 2D). The setting of the static grating is the same as in the fine orientation discrimination task except it is static. The changing probability of orientation is 50%; the orientation of static gratings was uniformly drawn in $[120, 150]$ (i.e., 135 ± 15) with the precision of 0.1°.

Evidence accumulation task. A hallmark of cognitive computations in the brain is the capability to go beyond a purely reactive mode: to integrate diverse sensory cues over time, and to wait until the right moment arrives for an action. A large number of experiments in neuroscience analyze neural coding after learning such tasks (see e.g., [Morcos and Harvey, 2016, Engelhard et al., 2019]). We considered the same task that was studied in the experiments of [Morcos and Harvey, 2016, Engelhard et al., 2019]. There a rodent moved along a linear track in a virtual environment, where it encountered several visual cues on the left and right (Fig. 2E). Later, when it arrived at a T-junction, it had to decide whether to turn left or right. The network should report the direction from which it had previously received the majority of visual cues. To reproduce this task under the limitations of a GPU implementation, we used a shorter duration of 600 ms for each trial. The right (left) cue was represented by 50 ms of cue image in which the black dots on the right (left) side of the maze. Visual cues were separated by 10 ms, represented by the gray wall of the maze. After a delay of 250 ms, the network had to decide whether more cues had been presented on the left or right, using two readout populations for left and right. The decision was indicated by the more vigorously firing readout pool (left or right) within the response window of 50 ms.

4.7 Loss function

The loss function was defined as

$$L = L_{\text{cross-entropy}} + \lambda_f L_{\text{rate reg.}} + \lambda_v L_{\text{v reg.}}, \tag{7}$$

where $L_{\text{cross-entropy}}$ represents the cross-entropy loss, λ_f and λ_v represent the weights of firing-rate regularization $L_{\text{rate reg.}}$ and voltage regularization $L_{\text{v reg.}}$, respectively. As an example, the cross-entropy loss of visual change detection tasks was given by

$$L_{\text{cross-entropy}} = -\sum_m \left[T^{(m)} \log \sigma \left(\theta \left(r^{(m)}_{\text{readout}} - r_0 \right) \right) + \left(1 - T^{(m)} \right) \log \sigma \left(\theta \left(r_0 - r^{(m)}_{\text{readout}} \right) \right) \right], \tag{8}$$

where the sum over m is organized into chunks of 50 ms and $r^{(m)}_{\text{readout}}$ denotes the mean readout population firing rate defined in Eq. 6. Similarly, $T^{(m)}$ denotes the target output in time window m, being 1 if a change in image identity should be reported and otherwise 0. The baseline firing rate r_0 was 0.01. σ represents the sigmoid function. θ is a trainable scale ($\theta > 0$) of firing rate.

We also used regularization terms to penalize unrealistic firing rates as well as unrealistic membrane voltages. Their weights, $\lambda_{\text{f}} = 0.1$ and $\lambda_{\text{v}} = 10^{-5}$. The rate regularization is given by the Huber loss [Huber, 1992] between the target firing rates, y, calculated from the model in [Billeh et al., 2020], and the firing rates, r, sampled the same number of neurons from the network model:

$$\begin{aligned} L_{\text{rate reg.}} &= \sum_j^N \left| \tau_j - \mathbb{I}\left\{ \delta_j < 0 \right\} \right| \frac{\mathcal{L}_\kappa\left(\delta_j\right)}{\kappa}, \quad \text{with} \\ \mathcal{L}_\kappa\left(\delta_j\right) &= \begin{cases} \frac{1}{2}\delta_j^2, & \text{if } |\delta_j| \le \kappa \\ \kappa\left(|\delta_j| - \frac{1}{2}\kappa\right), & \text{otherwise} \end{cases} \end{aligned} \tag{9}$$

where j represents neuron j, N the number of neurons, $\tau_j = j/N$, $\delta = 0.002$, $\delta_j = r_j - r_j^{\text{target}}$. $\mathbb{I}(x) = 1$ when x is true; $\mathbb{I}(x) = 0$ when x is false.

The voltage regularization was given by

$$L_{\text{v reg.}} = \frac{1}{N} \sum_{j=0}^{j=N} \left(\left[\frac{v_j - E_{\text{L}}}{E_{\text{L}}} - 1 \right]^+ \right)^2 + \left(\left[-\frac{v_j - E_{\text{L}}}{E_{\text{L}}} + 1 \right]^+ \right)^2, \tag{10}$$

where N represents the total number of neurons, v_j, the membrane potential of neuron j, E_{L}, the resting membrane potential, $[\cdots]^+$, rectifier function.

4.8 Training and testing

We trained the model for all 5 tasks together. Pairs of visual inputs and target outputs were collected in separate 64 batches for each task and these batches were interlaced during training. Apart from the change detection tasks, the spikes and membrane potentials were reset to 0 after each trial that consisted of 600 ms.

We applied back-propagation through time (BPTT) [Scherr and Maass, 2021] to minimize the loss function. The non-existing derivative $\frac{\partial z_j}{\partial v_j}$ was replaced in simulations by a simple nonlinear function of the

Table 1: The Billeh et al. model achieves high accuracy in all 5 tasks, consistent with the behavior performance of mice in similar tasks, after 6 training epochs.

	Test accuracy	Behavior accuracy	Mean firing rate (Hz)	Spike raster
Fine orientation discrimination	93.15%	$\sim 83\%^{a}$	3.97	Fig. S6
Image classification	88.92%	N/A	4.11	Fig. S7
Visual change detection of natural images	84.13%/83.25%[b]	$\sim 73\%/77\%^{c}$	3.97	Fig. S8
Visual change detection of gratings	89.25%	$\sim 60\%^{d}$	3.90	Fig. S9
Evidence accumulation	90.92%	$\sim 85\%^{e}$	3.96	Fig. S10

(a) Estimated from Fig. 4C of [Lee et al., 2012] when the orientation difference was 90°.
(b) In the visual change detection task for natural images, the two values refer to testing with familiar and novel images, respectively.
(c) Estimated from Fig. 1I of [Garrett et al., 2020a] when familiar and novel images were presented to mice, respectively.
(d) Estimated from Fig. 3A of [Glickfeld et al., 2013] when the orientation difference was 5°.
(e) Estimated from Fig. 1C of [Morcos and Harvey, 2016] when the number of cues was 6.
Note that the behavioral experiments had longer time delays that made the tasks somewhat more difficult.

membrane potential that is called the pseudo-derivative. Outside of the refractory period, we chose a pseudo-derivative of the form

$$\begin{aligned} \psi^t &= \frac{\gamma_{\rm pd}}{v_{\rm th} - E_{\rm L}} \exp\left(\frac{-\left(v^t_{sc}\right)^2}{\sigma_p^2}\right), \\ v^t_{sc} &= \frac{v^t - v_{\rm th}}{v_{\rm th} - E_{\rm L}}, \end{aligned} \tag{11}$$

where the dampening factor $\gamma_{\rm pd} = 0.5$, the Gaussian kernel width $\sigma_p = 0.28$. During the refractory period, the pseudo derivative was set to 0. During the training, we added the sign constraint on the weights of the neural network to keep Dale's law. Specifically, if an excitatory weight was updated to a negative value, it would be set to 0; vice versa.

We would like to emphasize that we tested the trained model -whenever possible- for new visual stimuli that had not been shown during training (this was not possible for the gratings because there were not sufficiently many different visual stimuli for them). In that sense, we evaluated the generalization capability of the trained Billeh et al. model, rather than its capability to handle a fixed set of stimuli correctly (which often suffices to solve behavioral tasks in experiments). The model achieved on all 5 tasks a performance that is in the same range as reported behavioral data from corresponding mouse experiments (Table 1).

4.9 Other simulation details

The BPTT training algorithm was coded, as the simulation of the model, in TensorFlow, which runs very efficiently on GPUs, and also on multiple GPUs for training in parallel. We employed independent simulations in parallel by distributing trials for all 5 tasks over batches. Every batch consisted of 320 trials, 64 for each of the 5 tasks. In every trial, the model of Billeh et al. was simulated for 600 ms of biological time, which took, together with the calculation of gradients around 5 s on a NVIDIA A100 GPU. Once all batches had finished (one step), gradients were calculated and averaged to update the weights by BPTT. We define an epoch as 781 iterations/steps, because this represents one cycle through

the full training dataset of MNIST. This computation had to be iterated for 6 epochs until the average performance on the 5 tasks was saturated. This took 20 h of wall clock time on 160 GPUs.

4.10 Control models

We used three control models in Fig. 3F. The first one was a randomly connected recurrent network of spiking neurons with the same numbers of neurons and connections, referred to as RSNN. In the RSNN, all data-based features of Billeh et al. model except for the number of neurons and synapses were removed: The GLIF_3 neuron model was replaced by the standard LIF model; neural connectivity representing laminar structure with primarily local connection was replaced by random connectivity; diverse neuron types were replaced by a single neuron type: the excitatory neuron on L2/3 (node type id in Allen brain atlas: 487661754); Dale's law was removed; initial weights were replaced by random values drawn from the Gaussian distribution with the same mean and variance as the initial values of the Billeh et al. model. In the second control model (Billeh et al. model without laminar structure), we kept the diverse GLIF_3 neuron models, but the laminar connectivity structure was replaced by random connections, and we kept the same number of synaptic connections. In the third control model (Billeh et al. model with LIF neurons), the diverse GLIF_3 neuron models were replaced by the standard LIF neuron model, all other features were kept.

4.11 Branching ratio as a measure for criticality

Based on the work of [Wilting and Priesemann, 2018b], where the branching ratio was recommended as a rather reliable measure for criticality of a network, we examined this branching ratio for the V1 model. In particular, this measure was shown there to be more robust to subsampling. The branching ratio is defined as the ratio of the number of neurons spiking at time $t+1$ to the number of spiking neurons at time t. Critical regimes, by their nature, are balanced and avoid runaway gain (positive or negative) and have a branching ratio of 1.0. We stimulated the V1 model as in the visual change detection task of nature images for 15 s.

In a network with A active neurons at time t, if the branching ratio has a fixed value m then $\langle A_{t+1} \mid A_t \rangle = mA_t + h$ where $<|>$ denotes the conditional expectation, m is the branching ratio and h is a mean rate of an external drive/stimulus. Considering subsampling, a_t is proportional to A_t on average $< a_t \,|A_t\rangle = \eta A_t + \xi$, where η and ξ are constants. This subsampling leads to a bias: $m\left(\eta^2 \operatorname{Var}[At]/\operatorname{Var}[at] - 1\right)$. Instead of using time t and $t+1$, this method focuses on times t and $t+k$ with different time lags $k = 1, \ldots, k_{\text{maximum}}$. With this, the branching ratio m_k is $< a_{t+k} \mid a_t >= m_k = \eta^2 \operatorname{Var}[A_t] / \operatorname{Var}[a_t]\, m^k = bm^k$, where b is a constant. To compute m_k with different k, we obtained an exponential curve and extracted m from this curve. $m < 1$ indicates a subcritical regime; $m > 1$ indicates a supercritical regime; $m = 1$ indicates a critical regime.

4.12 Convolution neural networks

Feedforward CNN. We used ResNet-18 [He et al., 2016] as FF-CNN. To calculate its eigenspectra, we used the pre-trained version on ImageNet provided by PyTorch. To evaluate its robustness against pixel noise, we trained ResNet-18 on MNIST with the Adadelta optimizer. The batch size was 64; learning rate was 1; weight decay was 0.0001; the coefficient used for computing a running average of squared gradients was 0.9; the term added to the denominator to improve numerical stability was 1×10^{-6}, the number of training epochs was 10.

Recurrent CNN. We used the gated recurrent convolution neural network [Wang and Hu, 2021] as R-CNN, inspired by abundant recurrent connections in the visual systems of animals. The gates control the amount of context information inputted to the neurons. We used the code, GRCNN-55 (weight sharing), in https://github.com/Jianf-Wang/GRCNN. To calculate its eigenspectra, we used the pre-trained version on ImageNet provided by [Wang and Hu, 2021]. To evaluate its robustness against pixel noise, we trained R-CNN on MNIST with the stochastic gradient descent (SGD) optimizer. The batch size was 64; the learning rate was 0.1; the momentum was 0.9; the weight decay was 0.0001; the number of training epochs was 10.

4.13 Eigenspectrum analysis

cvPCA. Eigenspectra of Billeh et al. model were estimated by the explained variance of the neural response along with the nth principal component (computed from the first presentation). It is achieved by cvPCA that computes the covariance of the projections of neural responses for the two repeats onto this component. cvPCA measures the reliable variance of stimulus-related dimensions, excluding trial-to-trial variability from unrelated cognitive and/or behavioral variables or noise. It accomplishes this by computing the covariance of responses between two presentations of an identical stimulus ensemble (Fig. S2A). Because only stimulus-related activity will be correlated across presentations, cvPCA provides an unbiased estimate of the stimulus-related variance. Briefly, the algorithm operates as follows:

$$
\begin{aligned}
\boldsymbol{X}^{(1)} &= \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\top} && \text{(singular value decomposition)} \\
\widetilde{\boldsymbol{X}}^{(1)} &= \boldsymbol{X}^{(1)}\boldsymbol{V} && \text{(project data onto eigenvectors)} \\
\widetilde{\boldsymbol{X}}^{(2)} &= \boldsymbol{X}^{(2)}\boldsymbol{V} && \\
\boldsymbol{\lambda}_j &= \sum_{i=1}^{s} \widetilde{\boldsymbol{X}}^{(1)}_{ij}\widetilde{\boldsymbol{X}}^{(2)}_{ij}, \quad \text{for } j \in \{1,\ldots,C\} && \text{(compute eigenvalue)}
\end{aligned}
$$

where $\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)} \in \mathbb{R}^{S\times N}$ (S is the number of stimuli and N is the number of neurons) are the neural responses for the first and second half of the trials (and averaged across trials), $\boldsymbol{V} \in \mathbb{R}^{N\times C}$ are the C eigenvectors of the covariance matrix of $\boldsymbol{X}^{(1)}$ and $\boldsymbol{\lambda} \in \mathbb{R}^{C}$ are the cross-validated eigenvalues associated with each of the eigenvectors ($\boldsymbol{\lambda_j}$ is the $\boldsymbol{j}$ th eigenvalue).

The first step of the cvPCA algorithm computes the eigenvectors of the neural response covariance from one set of the trials. The second and third steps project the neural responses from each half of the trials onto each eigenvector. The final step computes the (scaled) variance of the neural responses when projected onto an eigenvector (that was computed using one half of the trials). Thus, each cross-validated eigenvalue is related to the amount of stimulus-related variance of the neural responses along the eigenvalue's corresponding eigenvector.

To be consistent with [Stringer et al., 2019], we summed up spikes over 500 ms in response to visual stimuli. We ran cvPCA ten times on the response of the neural network fed with the same images that are used in [Stringer et al., 2019]. On each iteration randomly sampling the population responses of each stimulus from the two repeats without replacement. We ran ten different runs and found they were very similar to each other, i.e, the SD was close to 0. For the trained Billeh et al. model, we calculated the eigenspectra in three models trained with different noise and randomly generated data, and found the SD is 5.95×10^{-5}. The displayed eigenspectra of the trained Billeh et al. model were averaged over these three models. The code is available in https://github.com/MouseLand/stringer-pachitariu-et-al-2018b.

Billeh et al. model. We analyzed the neural representation in the trained Billeh et al. model in the same way as responses of V1 neurons were analyzed in [Stringer et al., 2019]: Without loss of generality, we used 2,800 generic images which were randomly drawn from ImageNet validation dataset in all panels Fig. 4. We also tried the 2,800 nature images used in [Stringer et al., 2019] and found they gave rise to a slower decaying speed of eigenspectrum (1.15); there were only used in Fig. S2B to compare with the noise level in mouse V1 experiment. We also used a smaller set of 32 images, repeated 90 times. All stimuli were input 50 ms after the simulation onset and sustained for 500 ms in each trial to be the same as experimental procedures. They were presented twice to allow cross-validated analysis. The initial condition of membrane potentials and spikes was set to zeros, unless otherwise stated. We input the 2,800-nature-image stimuli 5 times with different random seeds that were used to draw the noise and initial conditions of membrane potential and after-spike current from uniform distributions. We found that the results were not sensitive to the initial condition and noise.

CNNs. Generic images were resized so that their shorter dimension was 256 pixels and then center-cropped to 224×224 pixels. Padding the image and resizing it to 224×224 pixels achieved similar results. Images were additionally preprocessed by normalizing each image channel (RGB channels) using the mean and standard deviation that was used during model training (mean: 0.485, 0.456, 0.406 and SD: 0.229, 0.224, 0.225). For gray images, we repeated it to 3 channels. Using these preprocessed images, we extracted activation from every layer of each CNN and computed their eigenspectra using principal components analysis (PCA), because artificial neural responses are deterministic.

Power-law fitting of eigenspectra. Using the least-squares method, we fit power laws to the eigenspectra, $f(n)$, against PC dimension, n. The fitting function is $f(n) = n^{-\alpha}(n \in [n_{\min}, n_{\max}])$, where $n_{\min}$ and $n_{\max}$ are lower and higher bounds, respectively. For most cases, we chose $n_{\min} \in [1, 20]$ and $n_{\max} \in [301, 2800]$. For the 32-grating recordings, owing to noise and the length of the spectrum, we chose $n_{\min} \in [1, 10]$ and $n_{\max} \in [14, 35]$. For each possible pair of $n_{\min}$ and $n_{\max}$, we estimated the exponent α and its goodness-of-fit by the coefficient of determination (R^2). We then selected as our estimate of $n_{\min}$, $n_{\max}$, and α that gave the maximum R^2 (> 0.99) over all possibilities.

4.14 Discriminability index d' for neural responses to visual stimuli

To estimate how much information the neural activity conveyed about the stimulus identity, following [Rumyantsev et al., 2020], we used the metric d', which characterizes how readily the distributions of the neural responses to the two different sensory stimuli can be distinguished [Bishop, 2007]. The quantity $(d')^2$ is the discrete analog of Fisher information [Averbeck and Lee, 2006].

Fig. 6A and E. To be consistent with the experimental study [Rumyantsev et al., 2020], we calculated the neural response as the spike counts in each bin of 200 ms and evaluated two different approaches to compute d' values for the discrimination of the two different visual stimuli (gratings in the fine orientation discrimination task); the difference between two gratings is 2 °; each stimulus was presented in 500 trials. We analyzed the neural responses in a specific time bin relative to the onset of visual stimulation, which was called as instantaneous decoding approach used in [Rumyantsev et al., 2020]. The alternative way, cumulative decoding, i.e., analyzing neural responses that were concatenated over time from the start of the trial up to a chosen time, demonstrated similar results. To determine d' accurately despite having about fewer trials than neuron number in the Billeh et al. model, we reduced dimensional by using partial least squares (PLS) analysis [Geladi and Kowalski, 1986] to identify and retain only 5 population vector dimensions in which the stimuli were highly distinguishable as in [Rumyantsev et al., 2020]. In this 5-dimensional representation, the neural dynamics evoked by the two stimuli become distinguishable over the first 200 ms of stimulus presentation. In the reduced space, we calculated the $(d')^2$ value of the

optimal linear discrimination strategy as:

$$(d')^2 = \Delta\boldsymbol{\mu}^{\mathrm{T}}\Sigma^{-1}\Delta\boldsymbol{\mu} = \Delta\boldsymbol{\mu}^{\mathrm{T}}\mathbf{w}_{\mathrm{opt}} \tag{12}$$

where $\Sigma = \frac{1}{2}(\Sigma_{\mathrm{A}} + \Sigma_{\mathrm{B}})$ the noise covariance matrix averaged across two stimulation conditions, $\Delta\boldsymbol{\mu} = \boldsymbol{\mu}_{\mathrm{A}} - \boldsymbol{\mu}_{\mathrm{B}}$ is the vector difference between the mean ensemble neural responses to the two stimuli and $\mathbf{w}_{\mathrm{opt}} = \Sigma^{-1}\Delta\boldsymbol{\mu}$, which is normal to the optimal linear discrimination hyperplane in the response space [Averbeck and Lee, 2006]. Each entry of a covariance matrix is the covariance of spike counts of two neurons a_i and b_i ($i \in \{1, 2, \cdots, N\}$, N is the number of trials): $\frac{1}{N-1}\sum_{i=1}^{N}(a_i - \overline{a})(b_i - \overline{b})$ where $\overline{x}$ is the mean of $\{x_i\}$.

We also calculated $(d'_{\mathrm{shuffled}})^2$, the optimal linear discrimination performance using trial-shuffled datasets, which we created by shuffling the responses of each cell across stimulation trials of the same stimulus. Owing to this shuffling procedure, the off-diagonal elements of Σ_{A} and Σ_{B} became near zero. $(d'_{\mathrm{shuffled}})^2$ increased much faster with the increase of sampled neurons than $(d')^2$.

Fig. 6B. Noise eigenvalue λ_β and its eigenvector $\vec{e}_\beta$ was calculated by eigen-decomposition of noise covariance matrix Σ.

Fig. 6C. To quantify the signals projected onto the eigenvector $\vec{e}_\beta$, we projected $\Delta\boldsymbol{\mu}$ onto $\vec{e}_\beta$ and calculated its norm $|\Delta \cdot \boldsymbol{\mu}\vec{e}_\beta|$

Fig. 6D. To demonstrate that training makes signaling dimensions more orthogonal to the largest noise dimension, we decompose $(d')^2$ into a sum of projections:

$$(d')^2 = \Delta\boldsymbol{\mu}^{\mathrm{T}}\Sigma^{-1}\Delta\boldsymbol{\mu} = \sum_{\beta}\left(\frac{|\Delta\boldsymbol{\mu}\cdot\vec{e}_\beta|^2}{\lambda_\beta}\right). \tag{13}$$

Because $\Delta\boldsymbol{\mu}\cdot\vec{e}_\beta$ corresponds to the signal projected on noise eigenvector $\vec{e}_\beta$ and noise eigenvalue λ_β corresponds to the noise scale on $\vec{e}_\beta$, $\frac{|\Delta\boldsymbol{\mu}\cdot\vec{e}_\beta|^2}{\lambda_\beta}$ can be interpreted as singal-to-noise ratio on eigenvector $\vec{e}_\beta$. Clearly, the eigenvectors well aligned with $\Delta\boldsymbol{\mu}$ are the most important for discriminating between the two distributions of neural responses.

Acknowledgements

We would like to thank Eben Kadile, Du Kai, Oleg Kolner, Christoph Stöckl, and Yujie Wu for helpful comments on an earlier version of the manuscript. Also, we would like to thank Sandra Diaz for advice and help regarding large-scale computations. Most of computations were carried out on the Human Brain Project PCP Pilot Systems at the Jülich Supercomputing Centre, which received co-funding from the European Union (Grant Agreement number 604102). We also would like to thank the Beijing Academy of Artificial Intelligence for providing computational resources. This research was partially supported by the Human Brain Project (Grant Agreement number 785907) of the European Union and a grant from Intel.

References

[Allen Institute, 2018] Allen Institute (2018). © 2018 Allen Institute for Brain Science. Allen Cell Types Database, cell feature search. Available from: celltypes.brain-map.org/data.

[Arieli et al., 1996] Arieli, A., Sterkin, A., Grinvald, A., and Aertsen, A. (1996). Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science*, 273(5283):1868–1871.

[Averbeck and Lee, 2006] Averbeck, B. B. and Lee, D. (2006). Effects of noise correlations on information encoding and decoding. *Journal of neurophysiology*, 95(6):3633–3644.

[Barth and Poulet, 2012] Barth, A. L. and Poulet, J. F. (2012). Experimental evidence for sparse firing in the neocortex. *Trends in neurosciences*, 35(6):345–355.

[Bellec et al., 2018] Bellec, G., Salaj, D., Subramoney, A., Legenstein, R., and Maass, W. J. a. p. a. (2018). Long short-term memory and learning-to-learn in networks of spiking neurons.

[Bellec et al., 2020] Bellec, G., Scherr, F., Subramoney, A., Hajek, E., Salaj, D., Legenstein, R., and Maass, W. (2020). A solution to the learning dilemma for recurrent networks of spiking neurons. *Nature Communications*, 11(1):3625.

[Billeh et al., 2020] Billeh, Y. N., Cai, B., Gratiy, S. L., Dai, K., Iyer, R., Gouwens, N. W., Abbasi-Asl, R., Jia, X., Siegle, J. H., Olsen, S. R., et al. (2020). Systematic integration of structural and functional data into multi-scale models of mouse primary visual cortex. *Neuron*, 106(3):388–403.

[Bishop, 2007] Bishop, C. M. (2007). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 1 edition.

[Campagnola et al., 2021] Campagnola, L., Seeman, S. C., Chartrand, T., Kim, L., Hoggarth, A., Gamlin, C., Ito, S., Trinh, J., Davoudian, P., Radaelli, C., et al. (2021). Local connectivity and synaptic dynamics in mouse and human neocortex. *bioRxiv*.

[Deng et al., 2009] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee.

[Durand et al., 2016] Durand, S., Iyer, R., Mizuseki, K., de Vries, S., Mihalas, S., and Reid, R. C. (2016). A comparison of visual response properties in the lateral geniculate nucleus and primary visual cortex of awake and anesthetized mice. *Journal of Neuroscience*, 36(48):12144–12156.

[Engelhard et al., 2019] Engelhard, B., Finkelstein, J., Cox, J., Fleming, W., Jang, H. J., Ornelas, S., Koay, S. A., Thiberge, S. Y., Daw, N. D., Tank, D. W., et al. (2019). Specialized coding of sensory, motor and cognitive variables in vta dopamine neurons. *Nature*, 570(7762):509–513.

[Garrett et al., 2020a] Garrett, M., Manavi, S., Roll, K., Ollerenshaw, D. R., Groblewski, P. A., Ponvert, N. D., Kiggins, J. T., Casal, L., Mace, K., and Williford, A. (2020a). Experience shapes activity dynamics and stimulus coding of VIP inhibitory cells. *eLife*, 9:e50340.

[Garrett et al., 2020b] Garrett, M., Manavi, S., Roll, K., Ollerenshaw, D. R., Groblewski, P. A., Ponvert, N. D., Kiggins, J. T., Casal, L., Mace, K., Williford, A., Leon, A., Jia, X., Ledochowitsch, P., Buice, M. A., Wakeman, W., Mihalas, S., and Olsen, S. R. (2020b). Experience shapes activity dynamics and stimulus coding of vip inhibitory cells. *eLife*, 9:e50340.

[Geladi and Kowalski, 1986] Geladi, P. and Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. *Analytica chimica acta*, 185:1–17.

[Ghosh et al., 2022] Ghosh, A., Mondal, A., Agrawal, K., and Richards, B. (2022). Investigating power laws in deep representation learning. *arXiv preprint arXiv:2202.05808*.

[Glickfeld et al., 2013] Glickfeld, L. L., Histed, M. H., and Maunsell, J. H. (2013). Mouse primary visual cortex is used to detect both orientation and contrast changes. *Journal of Neuroscience*, 33(50):19416–19422.

[Haeusler et al., 2009] Haeusler, S., Schuch, K., and Maass, W. (2009). Motif distribution, dynamical properties, and computational performance of two data-based cortical microcircuit templates. *Journal of Physiology-Paris*, 103(1-2):73–87.

[Harris and Shepherd, 2015] Harris, K. D. and Shepherd, G. M. G. (2015). The neocortical circuit: themes and variations. *Nature neuroscience*, 18(2):170.

[He et al., 2016] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

[Huber, 1992] Huber, P. J. (1992). Robust estimation of a location parameter. In *Breakthroughs in statistics*, pages 492–518. Springer.

[Iavarone et al., 2022] Iavarone, E., Simko, J., Shi, Y., Bertschy, M., García-Amado, M., Litvak, P., Kaufmann, A.-K., O'Reilly, C., Amsalem, O., Abdellah, M., et al. (2022). Reconstruction and simulation of thalamoreticular microcircuitry. *bioRxiv*.

[Kong et al., 2022] Kong, N. C. L., Margalit, E., Gardner, J. L., and Norcia, A. M. (2022). Increasing neural network robustness improves match to macaque v1 eigenspectrum, spatial frequency preference and predictivity. *PLOS Computational Biology*, 18(1):1–25.

[Lee et al., 2012] Lee, S.-H., Kwan, A. C., Zhang, S., Phoumthipphavong, V., Flannery, J. G., Masmanidis, S. C., Taniguchi, H., Huang, Z. J., Zhang, F., Boyden, E. S., et al. (2012). Activation of specific interneurons improves v1 feature selectivity and visual perception. *Nature*, 488(7411):379–383.

[Maass et al., 2002] Maass, W., Natschläger, T., and Markram, H. (2002). Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural computation*, 14(11):2531–2560.

[Montijn et al., 2014] Montijn, J. S., Vinck, M., and Pennartz, C. (2014). Population coding in mouse visual cortex: response reliability and dissociability of stimulus tuning and noise correlation. *Frontiers in computational neuroscience*, 8:58.

[Mora and Bialek, 2011] Mora, T. and Bialek, W. (2011). Are biological systems poised at criticality? *Journal of Statistical Physics*, 144(2):268–302.

[Morcos and Harvey, 2016] Morcos, A. S. and Harvey, C. D. (2016). History-dependent variability in population dynamics during evidence accumulation in cortex. *Nature neuroscience*, 19(12):1672–1681.

[Plank et al., 2022] Plank, P., Rao, A., Wild, A., and Maass, W. (2022). A long short-term memory for ai applications in spike-based neuromorphic hardware. *Nature Machine Intelligence, in press. arXiv preprint arXiv:2107.03992*.

[Rumyantsev et al., 2020] Rumyantsev, O. I., Lecoq, J. A., Hernandez, O., Zhang, Y., Savall, J., Chrapkiewicz, R., Li, J., Zeng, H., Ganguli, S., and Schnitzer, M. J. (2020). Fundamental bounds on the fidelity of sensory cortical coding. *Nature*, 580(7801):100–105.

[Scherr and Maass, 2021] Scherr, F. and Maass, W. (2021). Analysis of the computational strategy of a detailed laminar cortical microcircuit model for solving the image-change-detection task. *bioRxiv*.

[Shew and Plenz, 2012] Shew, W. L. and Plenz, D. (2012). The Functional Benefits of Criticality in the Cortex. *The Neuroscientist*, 19(1):88–100.

[Siegle et al., 2021] Siegle, J. H., Jia, X., Durand, S., Gale, S., Bennett, C., Graddis, N., Heller, G., Ramirez, T. K., Choi, H., Luviano, J. A., Groblewski, P. A., Ahmed, R., Arkhipov, A., Bernard, A., Billeh, Y. N., Brown, D., Buice, M. A., Cain, N., Caldejon, S., Casal, L., Cho, A., Chvilicek, M., Cox, T. C., Dai, K., Denman, D. J., de Vries, S. E. J., Dietzman, R., Esposito, L., Farrell, C., Feng, D., Galbraith, J., Garrett, M., Gelfand, E. C., Hancock, N., Harris, J. A., Howard, R., Hu, B., Hytnen, R., Iyer, R., Jessett, E., Johnson, K., Kato, I., Kiggins, J., Lambert, S., Lecoq, J., Ledochowitsch, P., Lee, J. H., Leon, A., Li, Y., Liang, E., Long, F., Mace, K., Melchior, J., Millman, D., Mollenkopf, T., Nayan, C., Ng, L., Ngo, K., Nguyen, T., Nicovich, P. R., North, K., Ocker, G. K., Ollerenshaw, D., Oliver, M., Pachitariu, M., Perkins, J., Reding, M., Reid, D., Robertson, M., Ronellenfitch, K., Seid, S., Slaughterbeck, C., Stoecklin, M., Sullivan, D., Sutton, B., Swapp, J., Thompson, C., Turner, K., Wakeman, W., Whitesell, J. D., Williams, D., Williford, A., Young, R., Zeng, H., Naylor, S., Phillips, J. W., Reid, R. C., Mihalas, S., Olsen, S. R., and Koch, C. (2021). Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature*.

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

[Stringer et al., 2019] Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M., and Harris, K. D. (2019). High-dimensional geometry of population responses in visual cortex. *Nature*, 571:361–365.

[Teeter et al., 2018] Teeter, C., Iyer, R., Menon, V., Gouwens, N., Feng, D., Berg, J., Szafer, A., Cain, N., Zeng, H., and Hawrylycz, M. (2018). Generalized leaky integrate-and-fire models classify multiple neuron types. *Nature communications*, 9(1):1–15.

[Wang and Hu, 2021] Wang, J. and Hu, X. (2021). Convolutional neural networks with gated recurrent connections. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[Wilting and Priesemann, 2018a] Wilting, J. and Priesemann, V. (2018a). Inferring collective dynamical states from widely unobserved systems. *Nature communications*, 9(1):2325.

[Wilting and Priesemann, 2018b] Wilting, J. and Priesemann, V. (2018b). On the ground state of spiking network activity in mammalian cortex. *arXiv preprint arXiv:1804.07864*.

[Wilting and Priesemann, 2019] Wilting, J. and Priesemann, V. (2019). 25 years of criticality in neuroscience—established results, open controversies, novel concepts. *Current opinion in neurobiology*, 58:105–111.

Supplementary materials

Table S1: Performances on 5 tasks are not sensitive to the noise level. The light cyan highlights the noise level used in this study unless otherwise stated. VCDN is tested on novel images. FOD, fine orientation discrimination; IC, image classification; VCDN, visual change detection of nature images; VCDG, visual change detection of gratings; EA, evidence accumulation.

	FOD acc.	IC acc.	VCDN acc.	VCDG acc.	EA acc.
$q = 1,\ s = 0$	93.67%	89.14%	83.61%	92.37%	93.75%
$q = 1,\ s = 2$	93.09%	88.49%	83.42%	89.22%	92.92%
$q = 1,\ s = 3$	93.47%	88.97%	82.52%	88.73%	90.17%
$q = 1,\ s = 3.5$	93.73%	88.49%	81.11%	88.58%	91.00%
$q = 1,\ s = 4$	93.06%	88.81%	82.30%	87.86%	91.50%
$q = 2,\ s = 2$	93.15%	88.92%	83.25%	89.25%	90.92%
$q = 2,\ s = 3.5$	93.93%	89.19%	82.20%	87.51%	90.58%
$q = 3,\ s = 3$	94.10%	88.81%	81.87%	89.52%	92.25%
$q = 3,\ s = 4$	93.53%	89.02%	83.06%	88.14%	89.33%
$q = 4,\ s = 3$	93.91%	88.91%	82.53%	88.16%	91.33%
$q = 4,\ s = 4$	93.47%	88.16%	82.16%	88.17%	89.08%
$q = 5,\ s = 4$	93.32%	88.49%	81.41%	89.12%	90.58%
$q = 10,\ s = 10$	94.03%	88.54%	82.97%	89.56%	84.33%
$q = 20,\ s = 20$	91.01%	86.63%	77.41%	85.95%	77.75%
$q = 200,\ s = 200$	54.58%	39.56%	53.09%	53.96%	52.67%

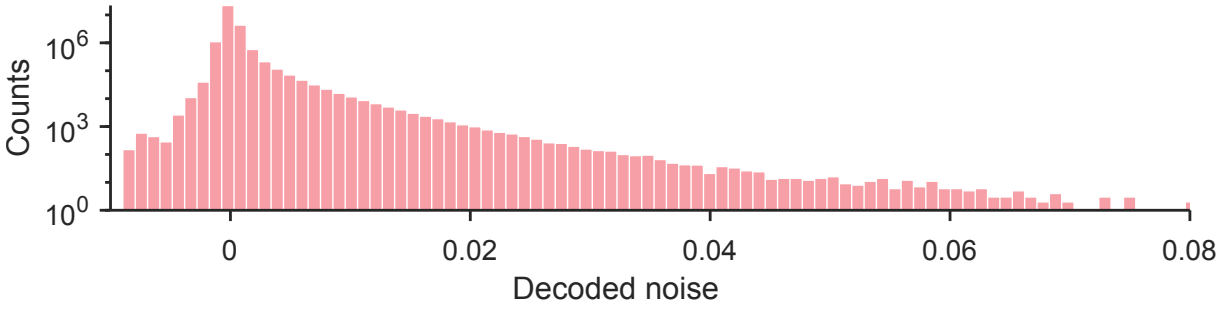


Figure S1: Empirical noise distribution where we drew the noise values. They are based on experimental data from [Stringer et al., 2019].

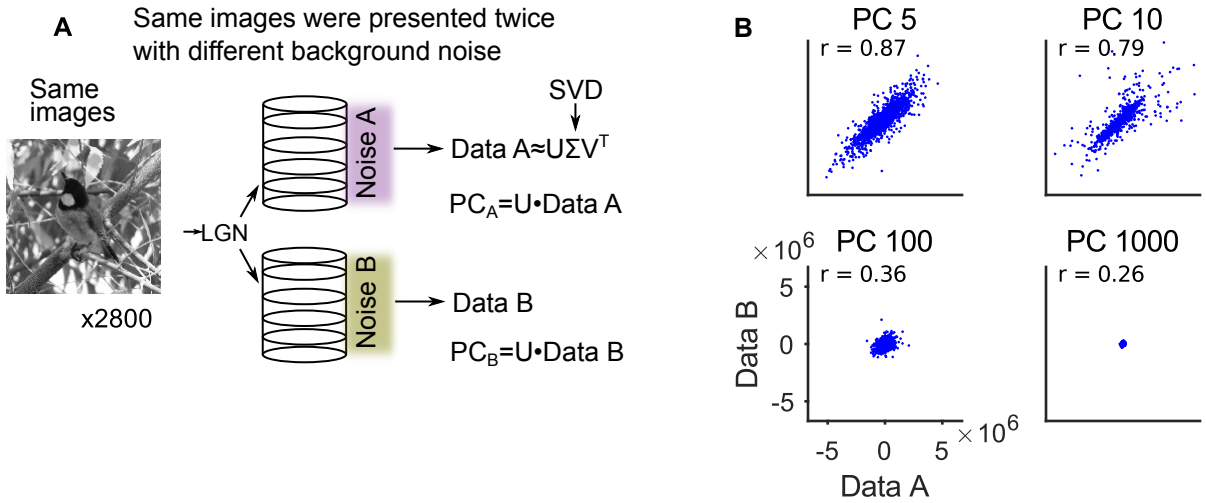


Figure S2: Schematic diagram of cvPCA (cross-validated PCA) to extract signal and noise. (A) Schematic diagram of cvPCA (cross-validated PCA). A set of generic images was presented twice with independently drawn noise for each image (Noise A and Noise B). Neural responses to the first presentation (Data A) were factorized by singular value decomposition (SVD) to estimate eigenspectrum of neural responses. **(B)** Correlations (r) of neural responses for two presentations of the same natural image, using our data-based noise model with $s = q = 2$ and projected onto selected principal components (PC).

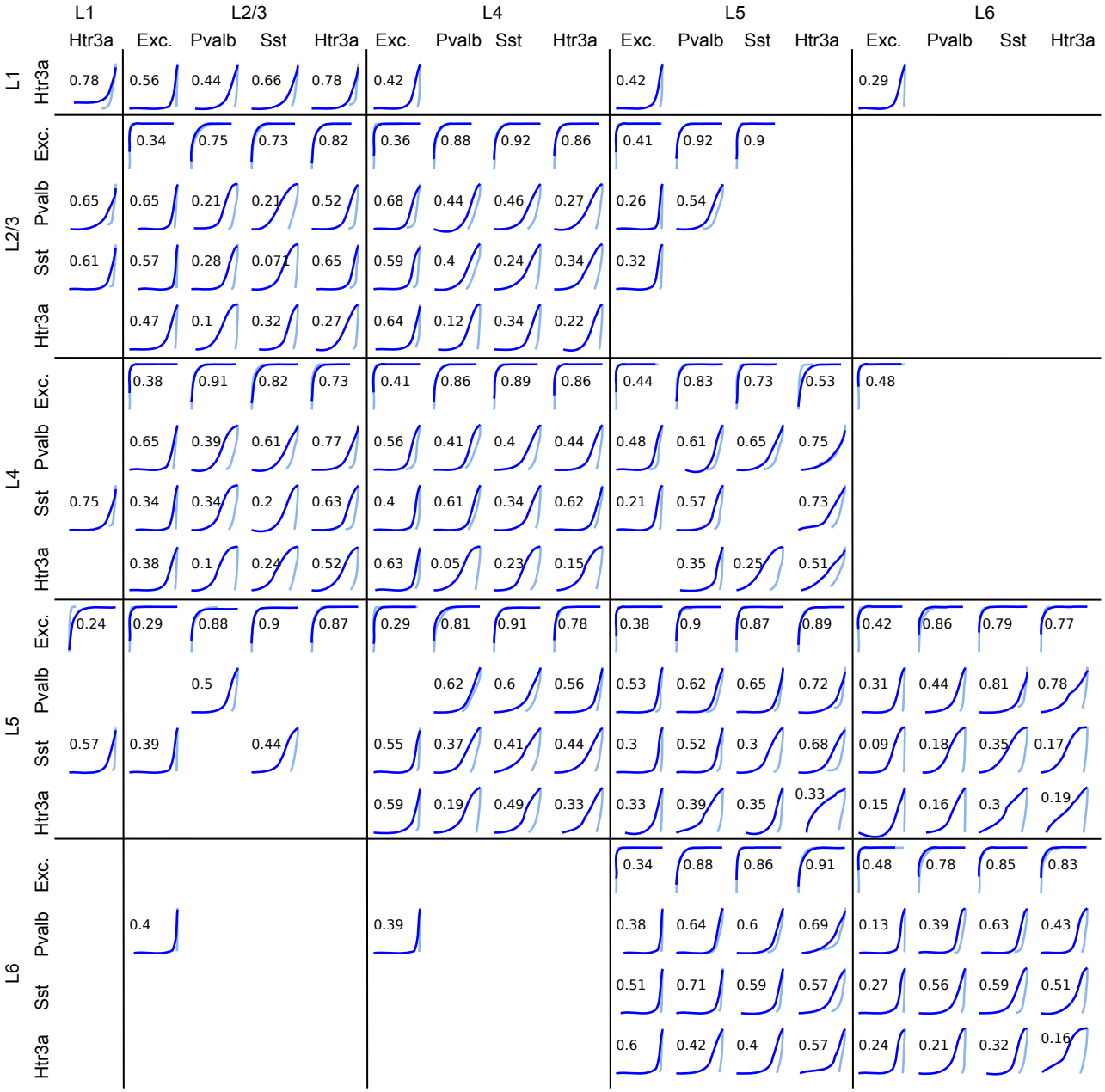


Figure S3: Distribution of recurrent weights between each population before (light blue) and after learning (dark blue) the task. Each row represents a pre-synaptic neuron population, and each column represents a post-synaptic neuron population. The histogram represents the distribution of synaptic weights of all synaptic connections that share the same pre-synaptic and post-synaptic neuron population. Vertical axis in each panel is log-scale. Horizontal axis is linear scale and horizontal range is from the smallest value to the largest value of each population. The number is $1 - D$ where D is from the Kolmogorov–Smirnov test, quantifying the similarity between distributions [Billeh et al., 2020]. Exc., excitatory neurons.

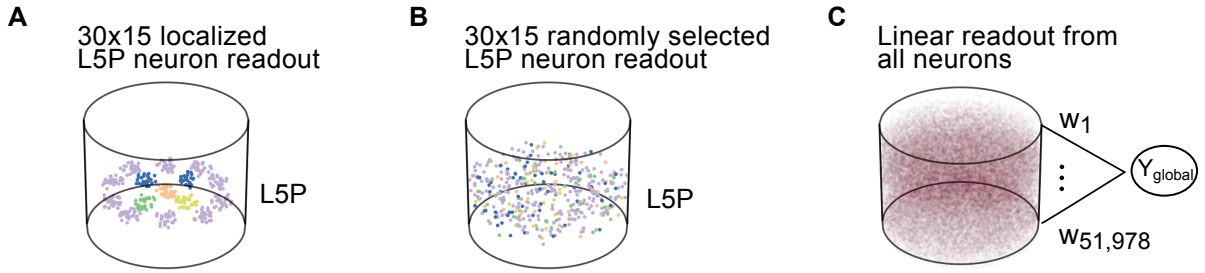


Figure S4: Schematic diagram of difference readout scenarios. (A) 15 spatially separated groups of 30 pyramidal neurons in L5 were selected to signal specific network outputs for 5 different tasks. Color encodes for 5 chosen tasks, same as in Fig. 2A. **(B)** Alternative selection of these 15 populations in L5 without spatial clustering leads to very similar performance. **(C)** Schematic diagram of a linear readout receives synaptic input from all 51,978 neurons in the microcircuit model, using a corresponding number of weights that can all be optimized for one particular task.

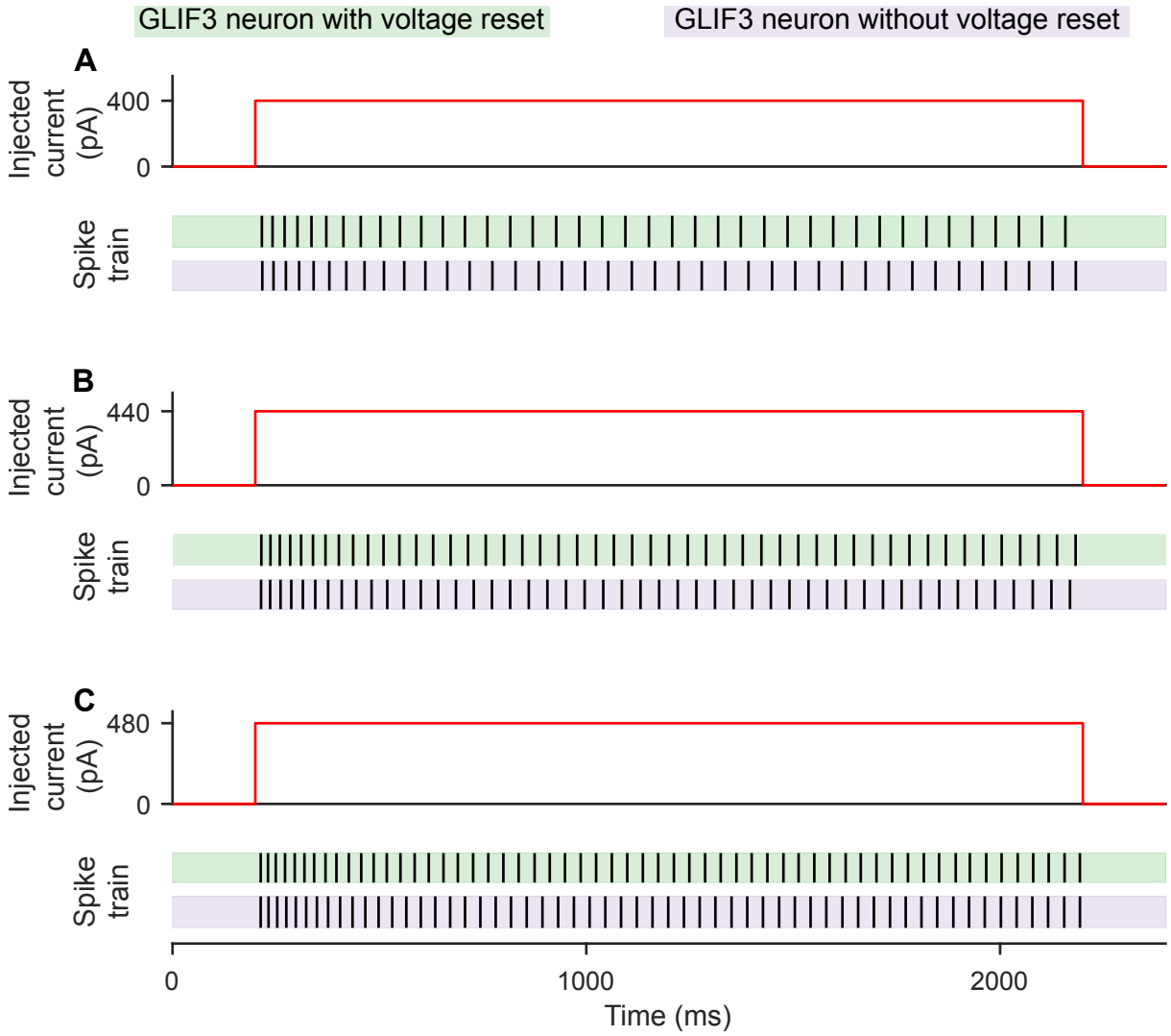


Figure S5: Adding voltage rest does not significantly affect the GLIF$_3$ model. (A) Stimulus in form of a step function (top) was inputted to the GLIF$_3$ model with voltage reset used in this study (middle) and the GLIF$_3$ model without voltage reset in [Teeter et al., 2018]. **(B-C)** Same as in **(A)** but for two different intensities of step stimuli. The difference between two models is so small that it can be ignored. Henceforth, we call our modified GLIF$_3$ model with voltage reset also as GLIF$_3$ model for convenience.

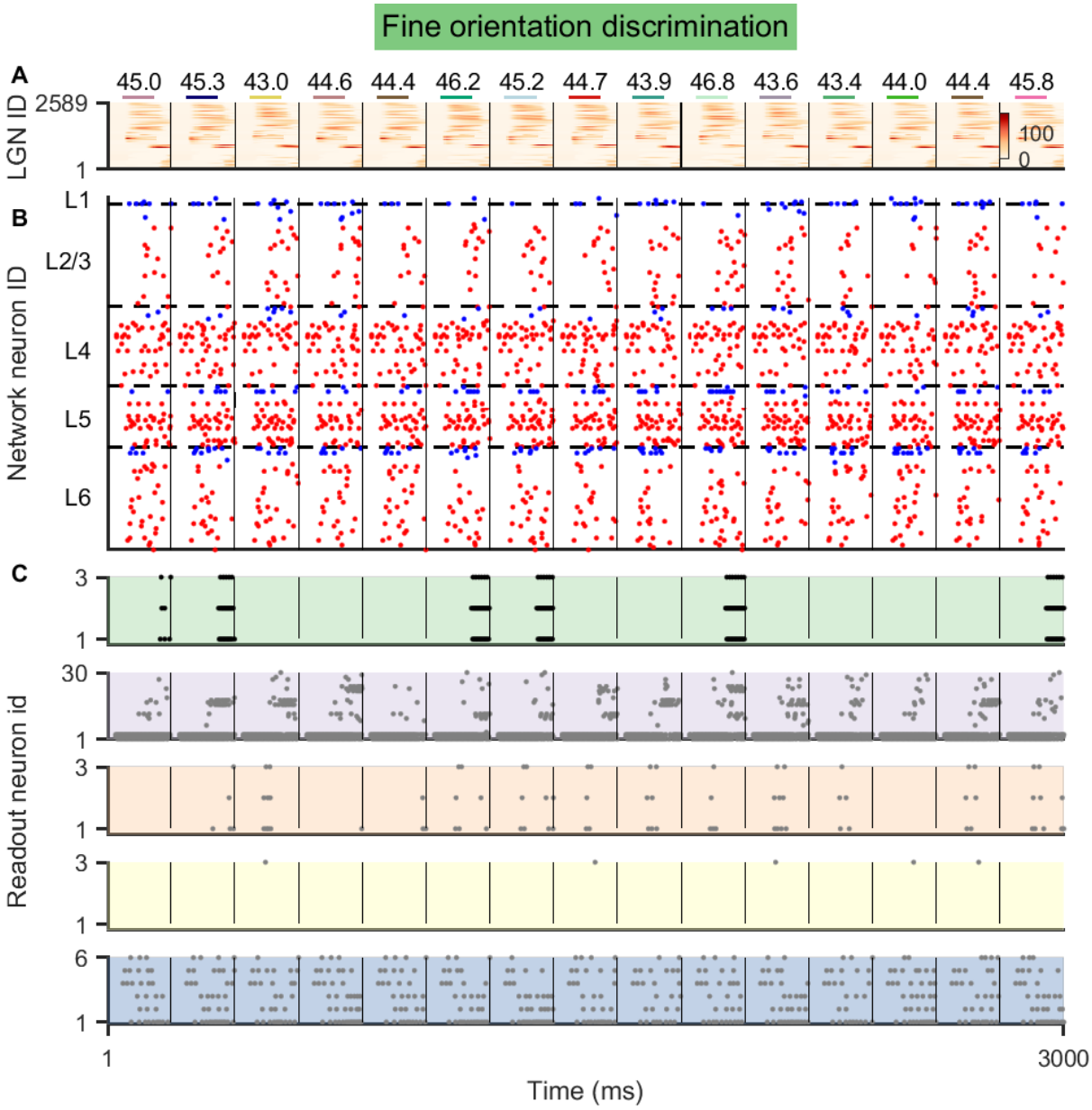


Figure S6: After learning, the model reliably distinguishes the orientation of gratings in the fine orientation discrimination task. **(A)** Colorful lines represent the timing of input images. Numbers on them represent the orientations of input gratings. The bottom colormap demonstrates the activity of LGN neuron activity. **(B)** Spike raster of the laminar V1 model. 200 neurons are sampled. Red and blue dots represent the spikes of excitatory and inhibitory neurons, respectively. Note that the spike and membrane potential of the model was reset to 0 after one classification was done (separated by the think black line). **(C)** Spike raster of readout neurons. 10% of neurons are sampled in every readout population. Color codes of panels are the same as in Fig. 2A. From the top to bottom, there are readout populations of the fine orientation discrimination, the image classification, the visual change detection of nature images and gratings, and the evidence accumulation tasks.

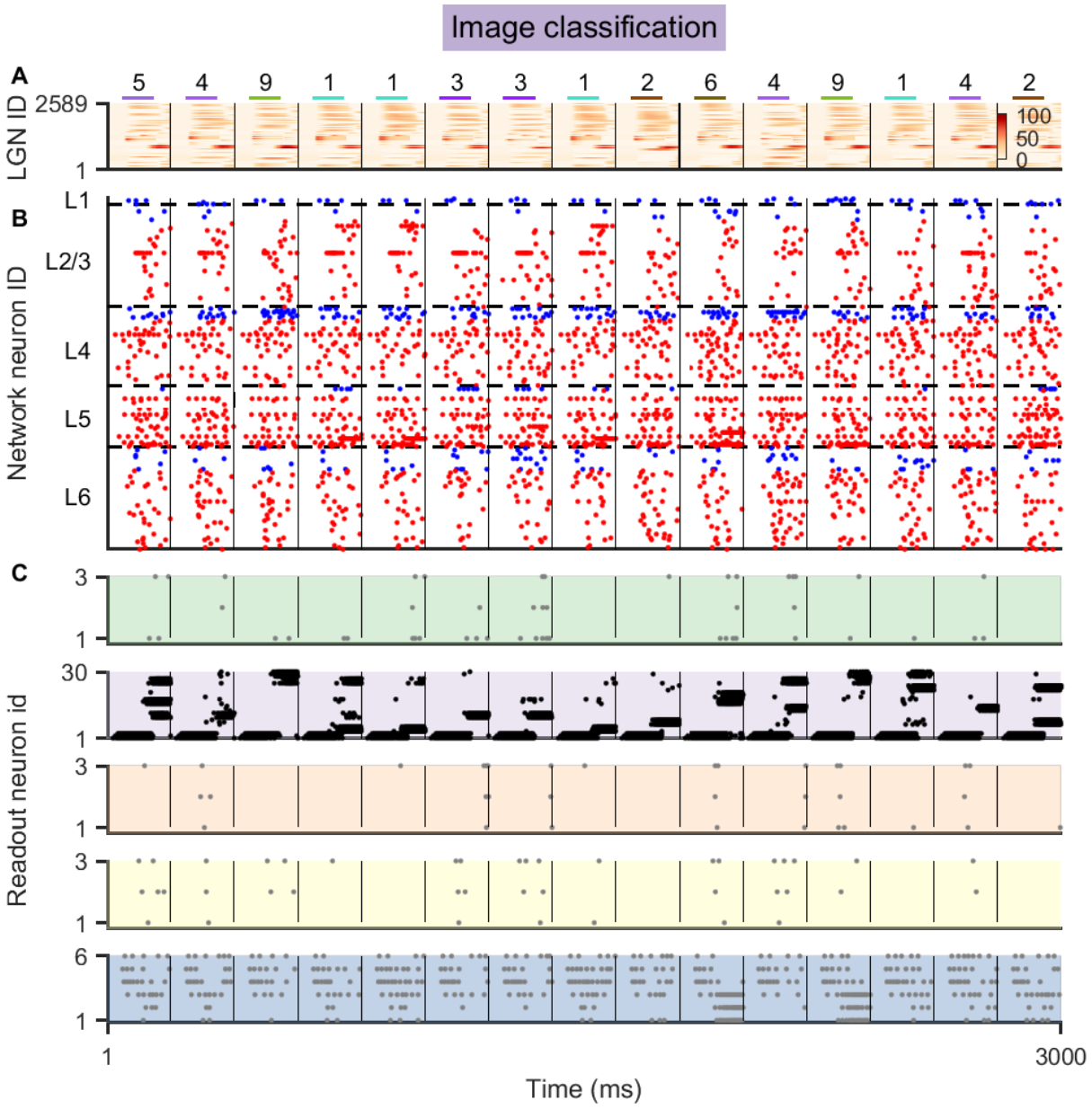


Figure S7: After learning, the model reliably classified the MNIST images. **(A)** Colorful lines represent the timing of input images. Numbers on them represent the digits in the input images. The bottom colormap demonstrates the activity of LGN neuron activity. **(B)** Spike raster of the laminar V1 model. 200 neurons are sampled. Red and blue dots represent the spikes of excitatory and inhibitory neurons, respectively. Red and blue dots represent the spikes of excitatory and inhibitory neurons, respectively. Note that the spike and membrane potential of the model was reset to 0 after one classification was done (separated by the think black line). **(C)** Spike raster of readout neurons. 10% of neurons are sampled in every readout population. Color codes of panels are the same as in Fig. 2A. From the top to bottom, there are readout populations of the fine orientation discrimination, the image classification, the visual change detection of nature images and gratings, and the evidence accumulation tasks.

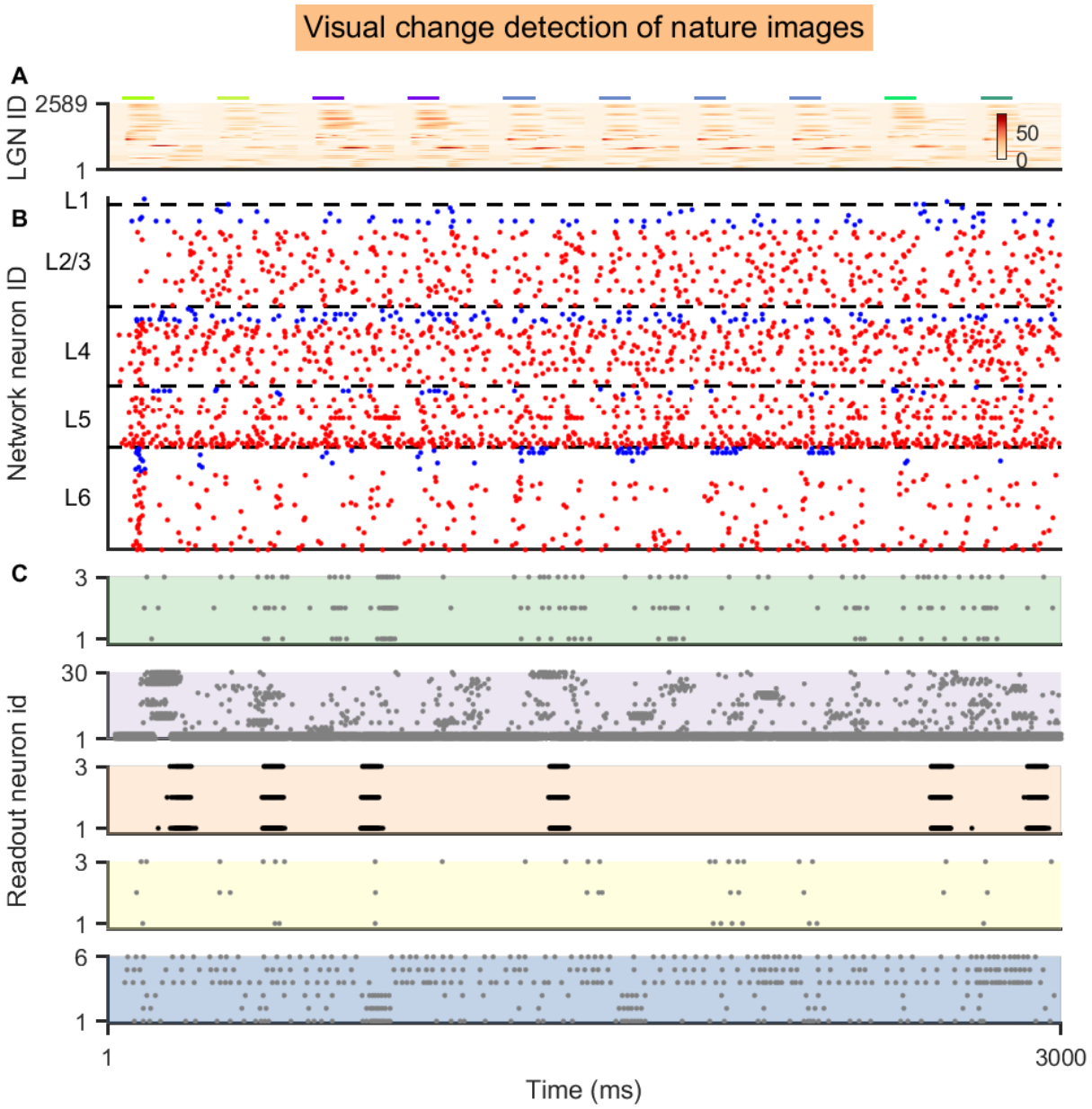


Figure S8: After learning, the model reliably reported the image identity change in the visual change detection task. **(A)** Colorful lines represent the timing of input images and the colors code the image identity. The bottom colormap demonstrates the activity of LGN neuron activity. **(B)** Spike raster of the laminar V1 model. 200 neurons are sampled. The slow noise was resampled every 600 ms. Red and blue dots represent the spikes of excitatory and inhibitory neurons, respectively. **(C)** Spike raster of readout neurons. 10% of neurons are sampled in every readout population. Color codes of panels are the same as in Fig. 2A. From the top to bottom, there are readout populations of the fine orientation discrimination, the image classification, the visual change detection of nature images and gratings, and the evidence accumulation tasks.

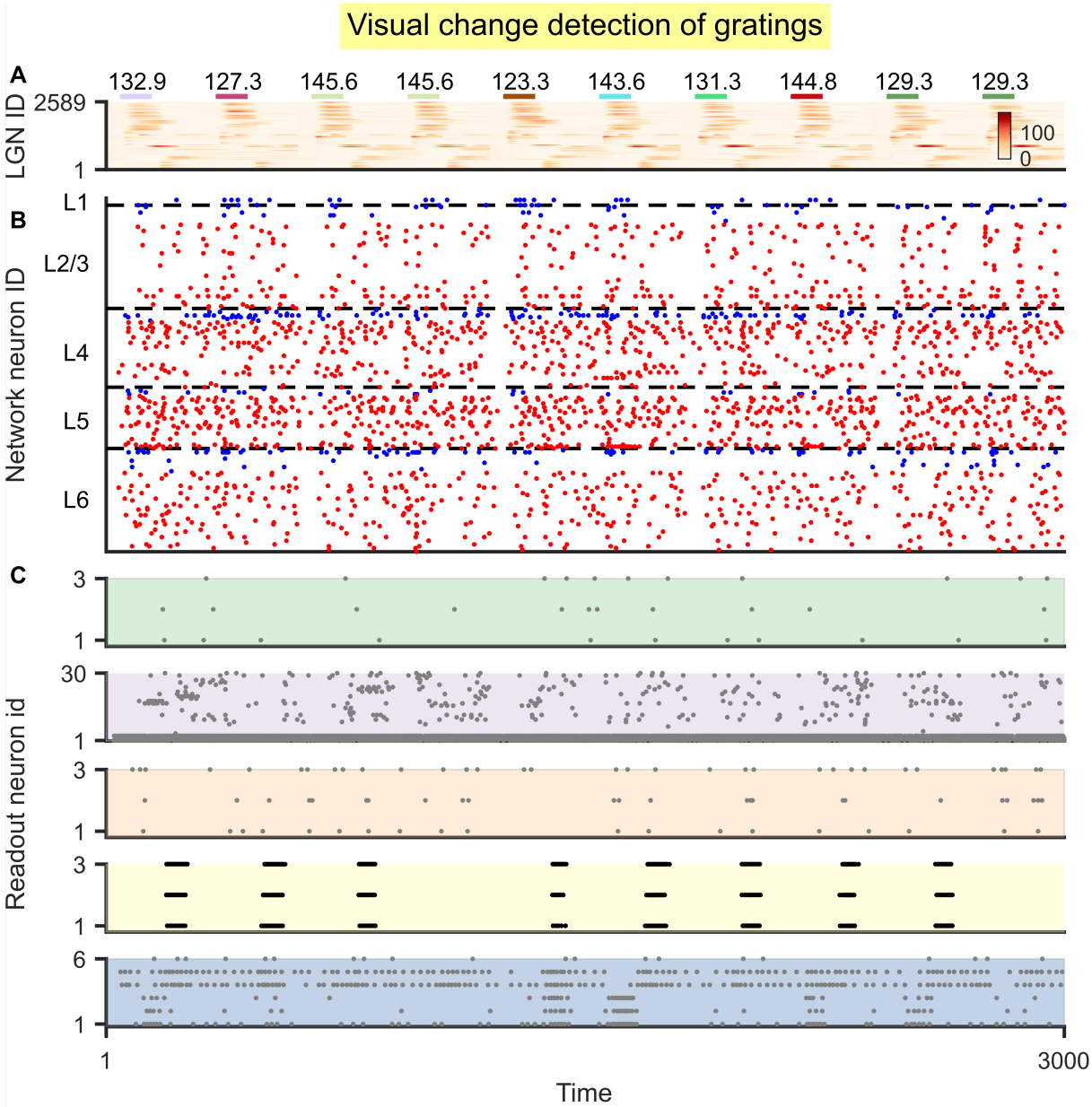


Figure S9: After learning, the model reliably reported the image identity change in the visual change detection task. **(A)** Colorful lines represent the timing of input images and the colors code the image identity. Numbers on them represent the orientations of input gratings. The bottom colormap demonstrates the activity of LGN neuron activity. **(B)** Spike raster of the laminar V1 model. 200 neurons are sampled. The slow noise was resampled every 600 ms. Red and blue dots represent the spikes of excitatory and inhibitory neurons, respectively. **(C)** Spike raster of readout neurons. 10% of neurons are sampled in every readout population. Color codes of panels are the same as in Fig. 2A. From the top to bottom, there are readout populations of the fine orientation discrimination, the image classification, the visual change detection of nature images and gratings, and the evidence accumulation tasks.

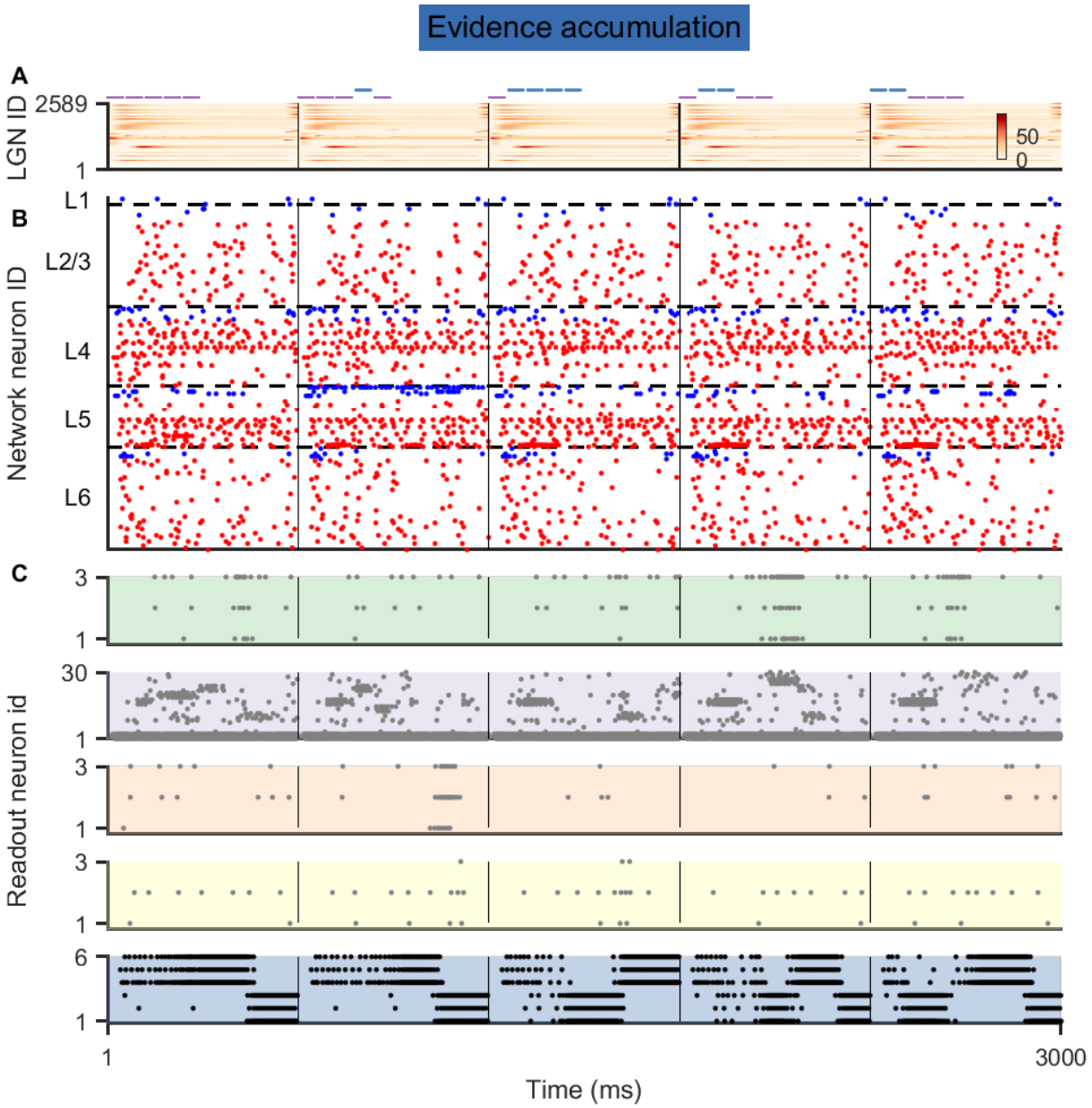


Figure S10: After learning, the spiking activity in the evidence accumulation task. **(A)** Colorful lines represent the timing of input left/right cues. The bottom colormap demonstrates the activity of LGN neuron activity. **(B)** Spike raster of the laminar V1 model. Red and blue dots represent the spikes of excitatory and inhibitory neurons, respectively. Note that the spike and membrane potential of the model was reset to 0 after one classification was done (separated by the think black line). **(C)** Spike raster of readout neurons. 10% of neurons are sampled in every readout population. Color codes of panels are the same as in Fig. 2A. From the top to bottom, there are readout populations of the fine orientation discrimination, the image classification, the visual change detection of nature images and gratings, and the evidence accumulation tasks.

4.15 Supplementary Note 1

The differences of readout scenarios can explain why behavioral performance lags behind neural coding fidelity in area V1. The behavioral discrimination threshold for orientations in the mouse V1 was almost 100 times larger than the discrimination threshold which they inferred from neural coding fidelity of populations of 50,000 neurons in area V1 of the mouse [Stringer et al., 2021]. They conjectured that this difference was caused by the limitations of downstream decoders. The Billeh et al. model suggests a slightly more refined explanation. Direct measurements of coding fidelity based on simultaneous recordings from 50,000 neurons do not account for the fact that their information content has to be extracted by neurons in V1 that project to downstream areas. They are conceptually similar to the postulate of having a global readout neuron that receives synaptic input from all 50,000 neurons, see Fig. S4C. However, one can demonstrate in the Billeh et al. model that such a global linear readout attains for the fine orientation discrimination task an accuracy of 98.81%. On the other hand, a pool of 30 projection neurons on L5 could only achieve an accuracy of 93.15% if one assumed that they were localized closely together (Fig. S4A), and of 93.61% if they were assumed to be randomly distributed in L5 (Fig. S4D). These results suggest that how information from area V1 is extracted and projected to downstream areas is a limiting factor that is likely to contribute to the gap between the performance of an ideal observer of neural activity in V1 and the behavioral performance of mice.

4.16 Supplementary Note 2

Biological features speed up gradient descent training of spiking neural networks. We applied the same training procedure also to control models that lacked salient structural features of the Billeh et al. model (Methods), and found that their task performance advances substantially slower. We found that during the training time, the Billeh et al. model achieved an approximately saturating task performance, whereas these control models (all models contain 51,978 neurons) were only able to reach a substantially lower task performance level in 6 epochs, see Fig. 3F. Substantially larger computing resources will be needed to determine the performance levels that these control models can eventually reach after sufficiently long training.

It had already been shown that neuron models with slower changing internal variables tend to enhance BPTT training, see Fig. 2D of [Plank et al., 2022], and Fig. 3C and Supplementary Movie of [Bellec et al., 2020]. Fig. 3C of the latter reference also shows that a similar training advantage holds for a biologically more plausible variant of gradient descent learning. But Fig. 3F shows that also the laminar connectivity structure of the Billeh et al. model contributes to its learning speed, even if the neuron models remain unchanged (yellow curve). One possible explanation is that the laminar structure enforces topographic maps between different layers, and hence tends to keep information more local within the network. The data-based rapid spatial decay of connection probabilities within a layer has a similar effect. This locality of information processing may facilitate learning through local learning mechanisms in such network. In contrast, in a generic randomly connected network without this connectivity structure, all information is continuously dissipated throughout the network, which is likely to impede the localization of processing errors and their correction.

It will be interesting to see whether this effect also arises for biologically more realistic learning methods, for example, *e-prop* [Bellec et al., 2020], which we did not perform because training with such methods tends to take substantially more trials, and therefore exceeded our computing resources. But since we now know that the Billeh et al. model is able to carry out the 5 visual processing tasks that we considered, one can now look for various biologically more realistic ways in the network initialization and/or learning algorithm -also for smaller instances of the Billeh et al. model- that can induce a similar computing

capability or neural coding features.