A data-based large-scale model for primary visual cortex enables brain-like robust and versatile visual processing

Guozhang Chen, Franz Scherr, and Wolfgang Maass*

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

August 4, 2022

Abstract

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 to date. We find that this brain-like neural network model can reproduce a number of characteristic visual processing capabilities of the brain, in particular the capability to solve diverse visual processing tasks, also on temporally dispersed visual information, with remarkable robustness to noise. This V1 model, whose architecture and neurons markedly differs from those of deep neural networks used in current AI, such as Convolutional Neural Networks (CNNs), reproduces also a number of characteristic neural coding properties of the brain, which provides explanations for its superior noise robustness. Since visual processing is substantially more energy-efficient in the brain compared with CNNs in AI, such brain-like neural networks are likely to have an impact on future technology: As blueprints for visual processing in more energy-efficient neuromorphic hardware.

Teaser

A biologically detailed large-scale model for primary visual cortex can robustly solve multiple visual processing tasks.

1 Introduction

The comprehensive model (*1*) 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 (Fig. 1A), 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, which is complemented by a host of recurrent loops. The dominance of short

*Corresponding author: maass@igi.tugraz.at

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 (*1*) 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 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 (*1*) 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 (*2*).

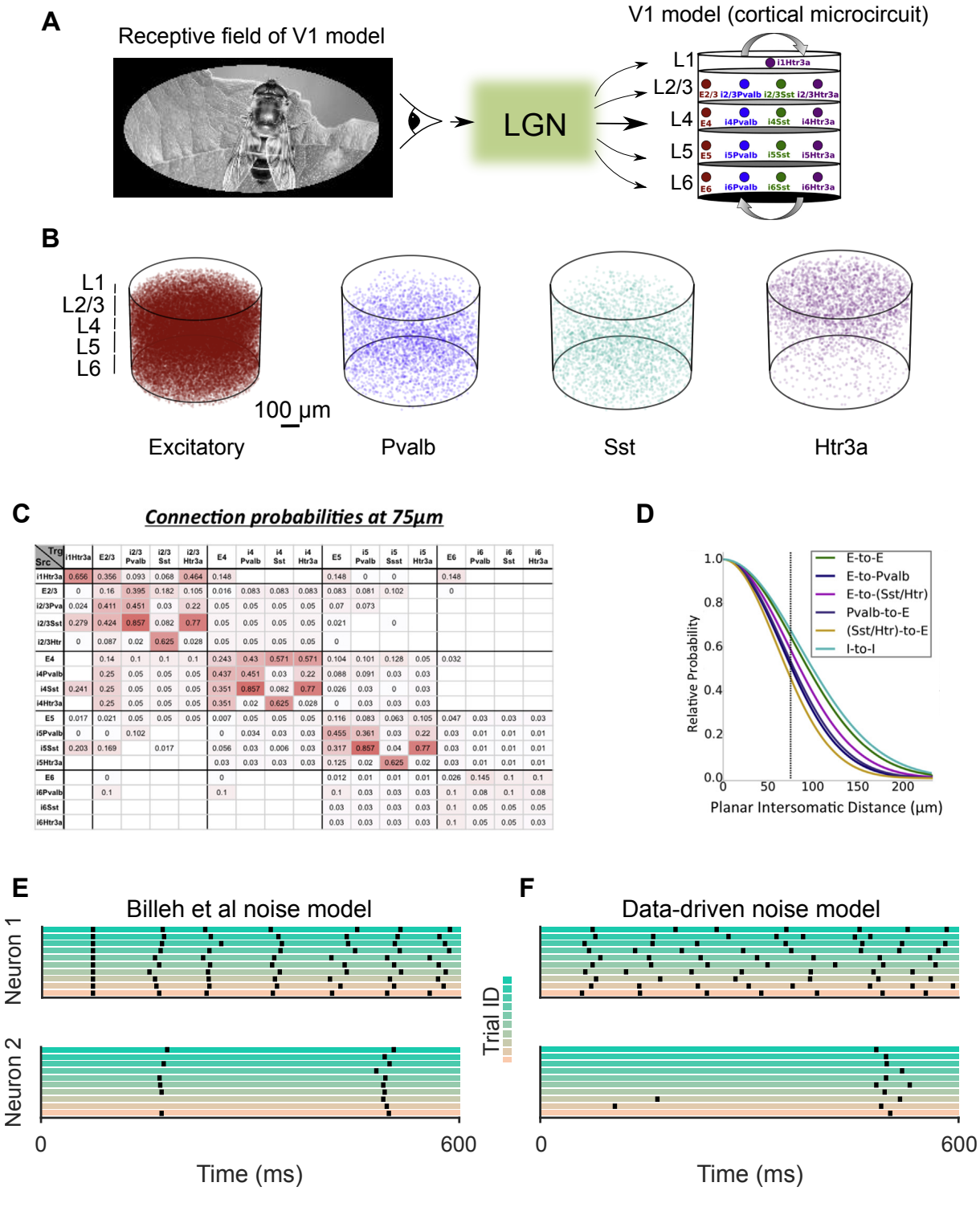
The original model of (*1*) is not able to solve nontrivial computing tasks, since its synaptic weights were chosen on the basis of sparse experimental data about the mean and variance of synaptic weights. In contrast, synaptic weights in the living brain are individually tuned through a host of synaptic 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 (*3, 4, 5, 6, 7, 8*). Afterward, our model achieved high accuracy simultaneously for all 5 tasks, while working in a biologically realistic sparse firing regime close to criticality (*9, 10*). 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 (*1*) is more robust because it employs, like the brain (*11*), a power law for the explained variance in higher PCA components that is close to a theoretically optimal compromise between the goal to be sensitive to details of visual inputs, and the goal to be robust to noise from the visual input and within the network. In contrast, neural codes in CNNs were shown to have a different power law (*12*) that favors the first (coding precision) over the second goal (noise robustness). In addition, we demonstrate that the model of (*1*) uses preferentially those dimensions of population activity for coding that are orthogonal to the largest noise dimensions, like the brain does (*3*).

Altogether our results show that the currently available anatomical and neurophysiological data, as compiled in (*1*), provide the basis for a new generation of neural network models for visual processing that can solve diverse visual processing capabilities in a highly robust manner. Furthermore, these neural network models provide new paradigms for neuromorphic computing since they combine versatility 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.



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 (*1*). (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 (*1*). **(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 (*1*) 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 (*1*)) 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 (*1*). **(E)** Spike outputs of 2 randomly selected neurons from the V1 model for 10 trials with the same input (a trial of visual change detection task for natural images), using the noise model of (*1*). **(F)** Same as in **(E)** but for the version of the data-driven noise model with $s = q = 2$ that we used as default-noise model during testing. It causes substantially larger trial-to-trial variability.

We have examined this question for the case of the large-scale model for a patch of V1 in mouse from (*1*), which is arguably the most comprehensive integration of anatomical and neurophysiological data on area V1 that is currently available.

The network 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 (*1*), 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. We will refer in the following to this model of V1 in conjunction with the LGN model of (*1*) as V1 model.

Individual neurons are modeled as point neurons. But in contrast to the customary leaky integrate-and-fire (LIF) neuron models, the V1 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 (*2*).

The neurons in the model of (*1*) 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). Hence we focused instead on a data-driven noise model. This noise model is based on experimental data from area V1 of the awake mouse (*11*). 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

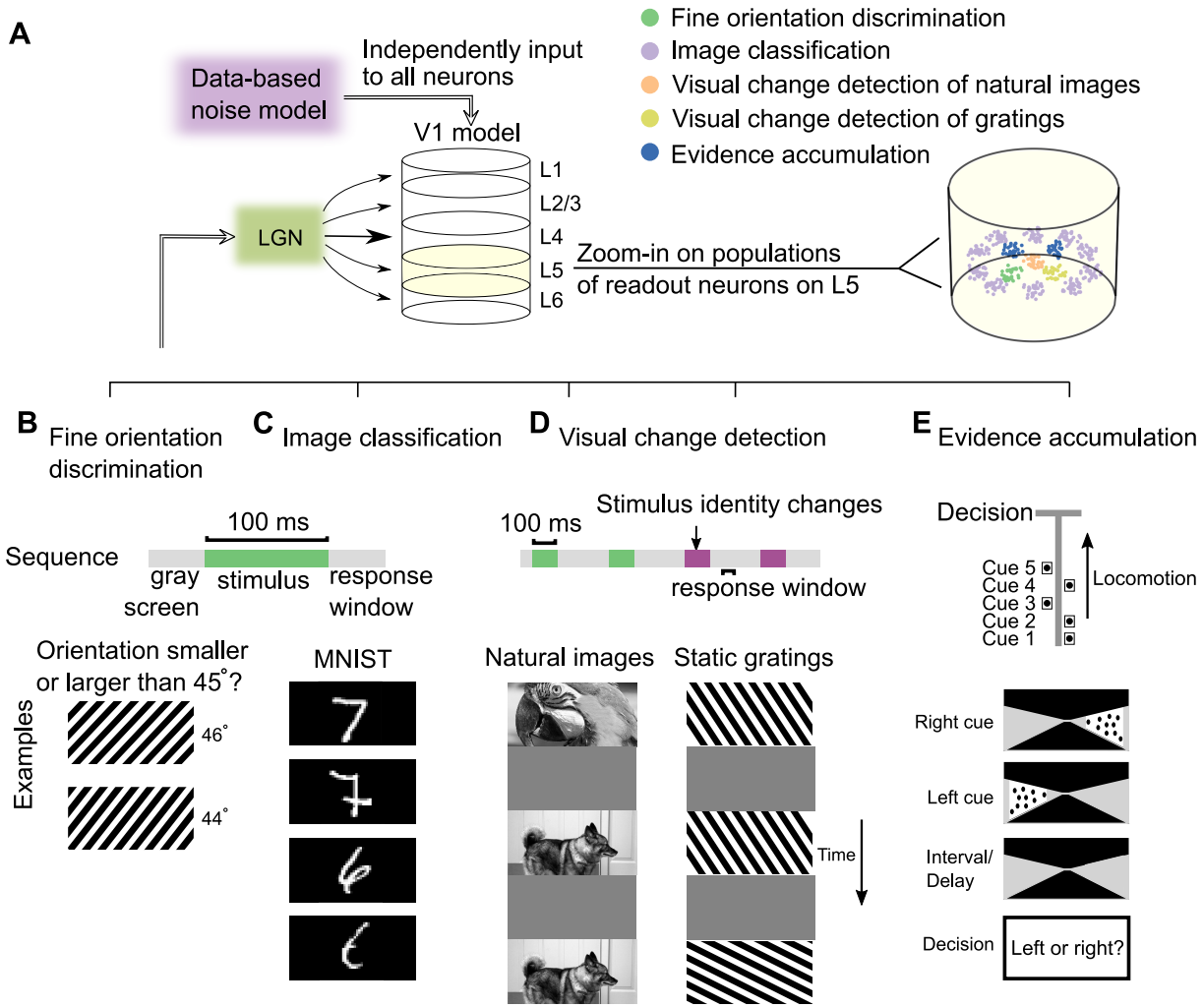
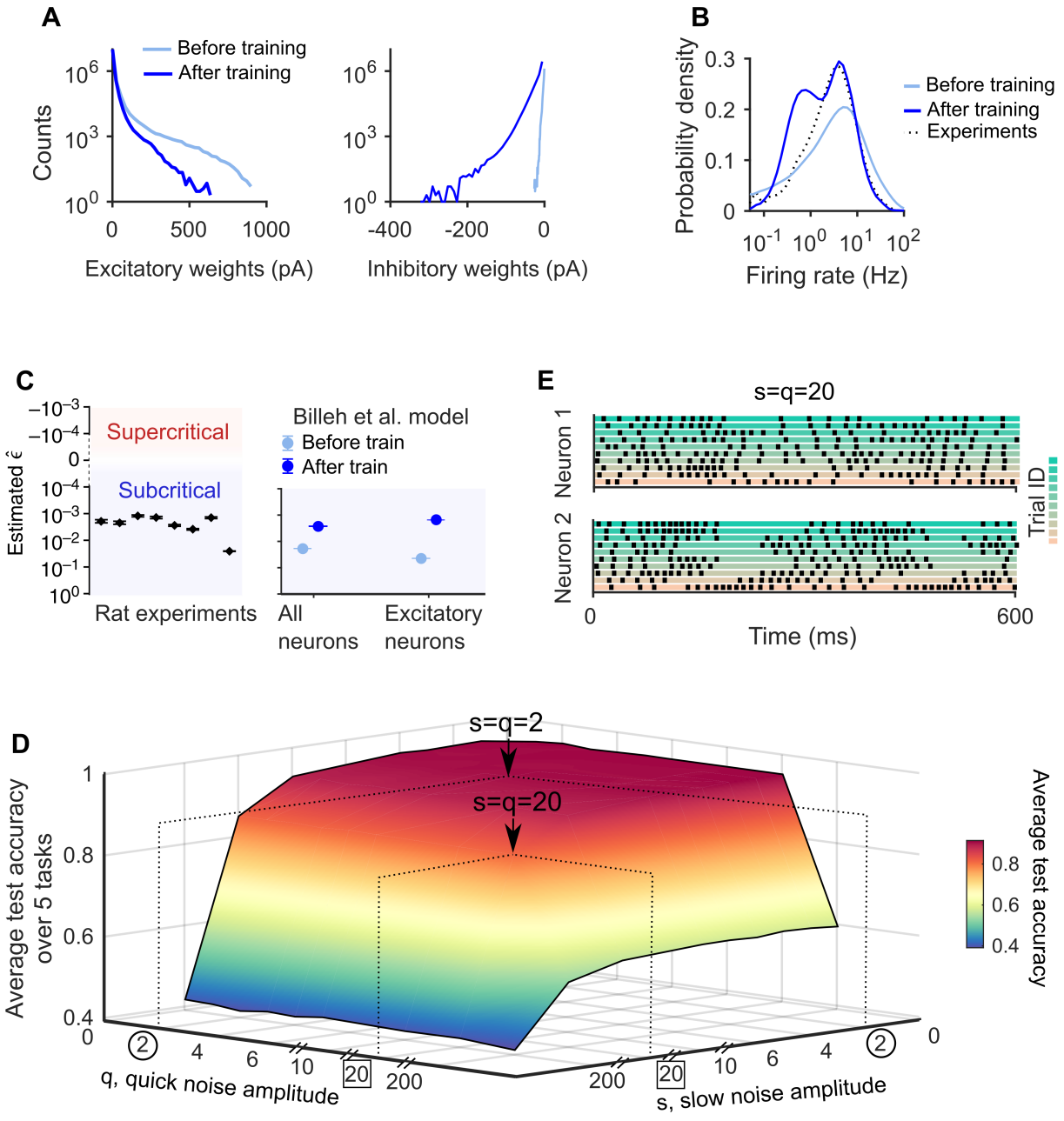


Figure 2: Illustration of the readout convention and the 5 visual processing tasks for which the V1 model was trained. (A) The visual stimuli for all 5 tasks were presented to the LGN model. Separate competing pools of pyramidal cells on L5 were chosen as readout neurons for each task. **(B-E)** Schematic diagrams and timings of 5 visual tasks (Materials and 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 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)*****: Analyses 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 (*6*). **(C)** Criticality of the V1 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 (*13*) are reproduced on the left. The V1 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 standard error of mean (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 (*1*), which has much less impact on neural activity according to Figs. 1 E, F. The arrow in the back points to the accuracy for the default values $s = q = 2$ for the data-driven 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 84.53%. 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 Figs. 1E, F for 10 trials with the same network input (image).

distribution once at the beginning of 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. (*14*). 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. S2A with Extended Data Fig. 5 of (*11*)). 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 (*15*). 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.

2.2 The V1 model can solve diverse computational tasks 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. Furthermore, brains can solve many different visual processing tasks with the same neural network, without changing their synaptic weights. We wondered whether the V1 model also has this capability. Therefore we tested it after training not only on a standard image classification task, classification of handwritten digits from the MNIST (modified national institute of standards and technology) database, but also on 4 tasks that require temporal integration of visual information (Fig. 2A). 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 (*3, 4*), a generic image classification task (Fig. 2C), visual change detection tasks for natural images and static gratings (Fig. 2D), as considered in (*5, 6*), and accumulation of temporally dispersed cues on the left and right (Fig. 2E), as considered in (*7, 8*), 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. (*16*), is to let an external

"readout neuron" that receives synaptic input from all neurons in the network produce the network decision (Fig. S3A). 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 (*17, 18*) imply that if the network model is sufficiently large and has diverse units so that it approximately satisfies the point-wise separation property for relevant input streams, such global readout neurons becomes computationally quite powerful even if only the weights of these readouts are trained for a specific task. Hence global linear readouts tend to mask the computational contribution of the neural network model itself. This theoretical prediction turns out to be valid also for the V1 model: If one takes the V1 model as defined in (*1*), without changing any of its synaptic weights or other parameters, and just trains linear readouts from all of its neurons for the 5 tasks, one gets already a very high average accuracy of all 5 tasks: 86.99%. Therefore, we used for the V1 model a more brain-like readout convention, where projection neurons within the network (*19*) extract and transmit computational results of the network. In particular, a large fraction of pyramidal cells on L5 projects computational results of the network to subcortical areas where behavioral responses are triggered. Consequently, 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. The size of this readout population turned out to have no significant impact, see Materials and Methods. It also caused no significant difference whether the neurons of each readout pool were co-located or are randomly distributed on L5 (Fig. S3). Note that a co-located readout pool can possibly benefit from synergy, whereas a distributed pool can integrate information from more and more widely distributed presynaptic neurons. Hence it is not a-priori clear which works better. In our test, both versions perform about equally well. 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 (*1*), the V1 model is incapable of performing any of the 5 tasks with this biologically more realistic readout convention: The accuracy is close to chance level (= 42%). Even if one trains linear readouts from these small populations of neurons on L5 for the 5 tasks, one achieves in the case of the untrained V1 model an accuracy of just 53.74%. We therefore applied stochastic gradient descent, like in (*20*), to the synaptic weights of synapses within the V1 model of (*1*), and to connections from the LGN model to the V1 model. No synaptic connections were added or deleted 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 Materials and Methods. The loss function also penalized biologically unrealistic high firing rates.

The principle of stochastic gradient descent is to let the network compute on a given input for one of the computational tasks, or in our case for a batch of such inputs for all 5 tasks. Then the loss function is computed for these computations, and gradient descent is applied to the network computations to determine in which direction and by how much each synaptic weight should be changed in order to decrease the loss function. Then the whole process is iterated. Stochasticity arises in this gradient descent from random selection of network inputs, and is in general useful for avoiding to get stuck in local minima of the loss function. We applied stochastic gradient descent to the V1 model through a variation of BPTT (backpropagation through time), with the help of a suitable pseudo derivative for handling the discontinuous dynamics of spiking neurons, as suggested by (*21*). To avoid artifacts arising from an application of gradient descent to 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 of the neuron model causes no significant difference in the spike output of a neuron (Fig. S4). The pseudo derivative is just an approximation to ideal stochastic gradient descent, especially

when gradients have to be propagated backward through several spikes. But it appears to be the most powerful tool for training networks of spiking neurons that is currently available. It turns out to work especially well for the GLIF_3 neuron models of (*1*), especially in a biologically realistic sparse firing regime, since gradients can be propagated with high precision through the slowly changing continuous variables of GLIF_3 neuron models.

We found that the type of noise that is applied during training has little influence on the accuracy that is reached during testing: Virtually the same accuracy is achieved when during training the noise model from (*1*), our data-driven noise model, or no noise at all is applied, while testing with the data-driven noise model, see Fig. S5. We think that this effect is due to the implicit noise that is introduced through the natural in-class variance of network inputs during training, since this variance is likely to dominate from the perspective of network function the impact of any milder form of noise within the network. We always report accuracy on test data for the case of data-driven noise within the network. Training was carried out for each network model over 5 different runs, 2 with the noise model of (*1*), 2 with the data-driven noise model (see Fig. 1F), and one run without any noise applied during training. Our motivation for also considering the noise model of (*1*) was to maintain comparability with their original model, and for enabling an unbiased evaluation of the impact of various types of noise.

After training, the V1 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 94.28%. Sample computations of the V1 model for each of the 5 tasks are shown in Figs. S6-S10. The performance after training 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 94.67%. As expected, choosing instead global linear readout neurons for each task and training both the readout weights and the weights in the V1 model led to substantially higher accuracy of 98.13%. Hence we find that the neural codes in the network could in principle support even better task performance, but biologically realistic readout from V1 reduces the amount of information which can be used by downstream networks for decision making. This effect suggests an interesting new explanation for why behavioral performance of mice lags behind neural coding fidelity in area V1 (*4*) (Supplementary Note 1).

The median strength of inhibitory synapses increased from 0.03 to 4.80 pA during training, while the median weights of excitatory synapses decreased from 2.96 to 2.08 pA, see Fig. 3A, and Fig. S11 for more detailed analyses in terms of the neuron types involved. Importantly, 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. Furthermore, training preserves spatial clustering of orientation tuning in L2/3 (Fig. S12), a neural coding feature that had been demonstrated experimentally for mouse V1 in (*22*).

Experimental data suggest that neural networks of the brain typically operate in a critical regime (*23, 24, 10*). We evaluated the criticality of the V1 model by measuring its branching ratio of neural activity, as suggested by (*10*). 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.

We tested the robustness of the resulting versatile visual processing capability of the V1 model after training by exposing its neurons to substantially larger amplitudes q and s of the data-driven noise model, although it had never been exposed to such noise during training (where we only applied the really small noise considered in (*1*)). Surprisingly, it is almost impossible to destroy its versatile 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 Impact of anatomical and neurophysiological details of the V1 model on learning speed and resulting computational performance

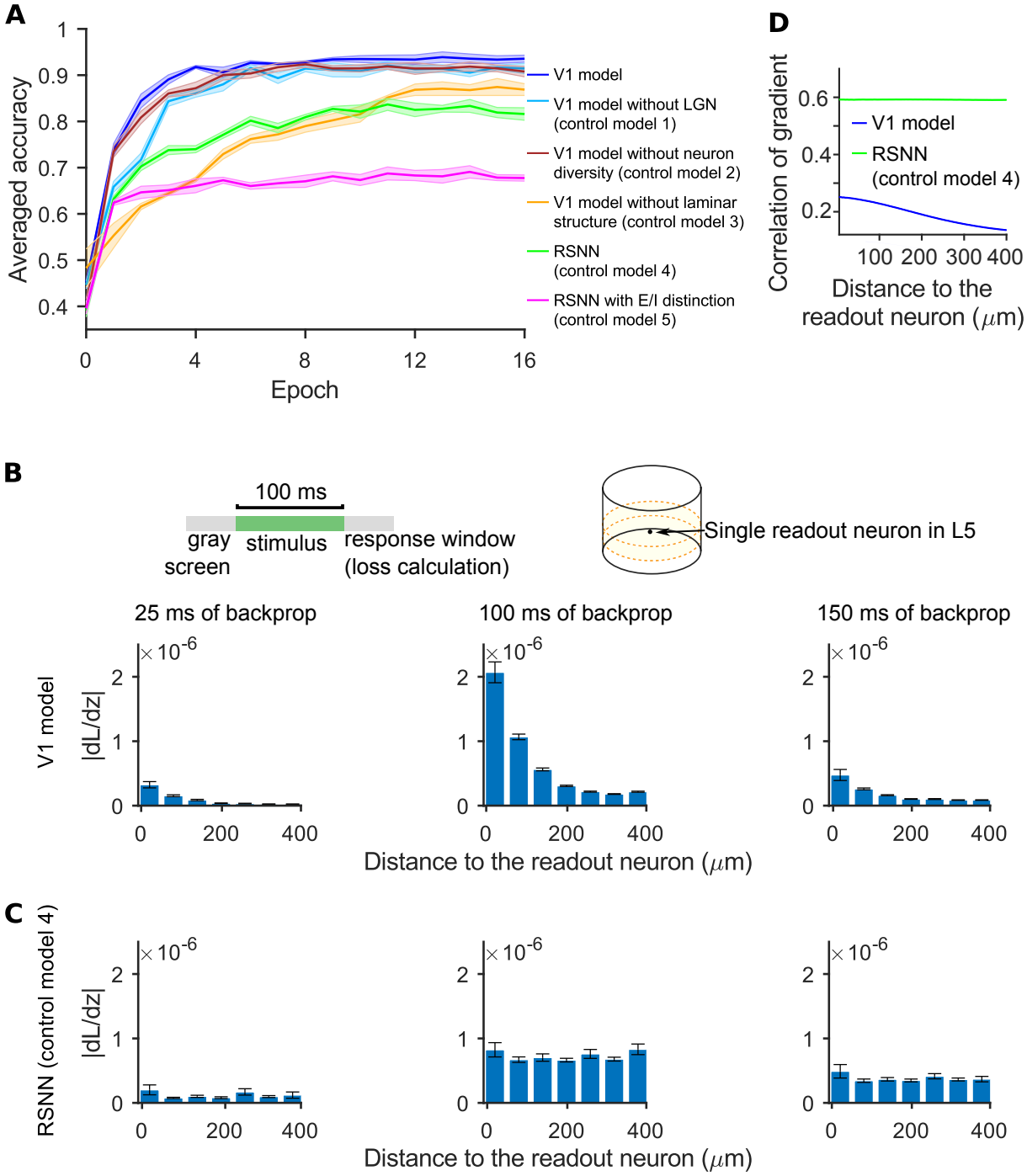
We show in Fig. 4A the performance of the V1 model and of 5 control models on test data, in dependence of the training duration. In control model 1, we removed the LGN model of (*1*), and injected pixel values directly into the V1 model, targeting the same neurons as the LGN model. This turns out to have just a minor effect on the properties that we have analyzed. In control model 2, we removed the diversity of the 111 data-based neuron types in the V1 model, replacing them with one generic model for excitatory neurons and one for inhibitory neurons. In control model 3, we removed instead the laminar spatial structure with distance-dependent connection probabilities of the V1 model, replacing them with an equal number of randomly chosen connections (without dependence on spatial proximity). Control model 4 is a randomly connected recurrent network of standard LIF neurons (randomly connected networks of spiking neurons, RSNN) with the same number of neurons and connections. This is arguably the most commonly considered type of spiking neural network model. Control model 5 is a variation of it, where the neurons are split like in the V1 model into excitatory and inhibitory neurons, and Dale's law is preserved during training. This type of model is sometimes seen as an intermediate step from generic RSNNs toward more biologically oriented network models. Two remarkable phenomena are demonstrated in Fig. 4A:

i) The V1 model learns much faster than control models, reaching close to optimal performance after just 4 training epochs. A closer look at training progress for individual tasks (Fig. S13) shows that the V1 model makes progress on each task in a continuous and smooth manner, whereas control models tend to reach early plateaus from which they can only escape after very long further training.

Interestingly, the laminar structure with primarily local synaptic connections of the V1 model also affects the dynamics of stochastic gradient descent: Gradients that result from errors at a readout neuron are not immediately spread out all over the network, which happens in generic RSNNs, but are rather propagated in a wave-like manner over larger and larger parts of the network (Fig. 4B, C). Consequently, correlations between learning gradients have for two neurons in the V1 model a more pronounced dependence on their distance than in a randomly connected control model (Fig. 4D). Hence the V1 connectivity structure better supports localized computation and learning within a large network.

With regard to the weight initialization before training, we show in Fig. S14 that if one replaces the weight values of the V1 model from (*1*), which were used as weight initialization in all of the previously described experiments, with a random initialization that is commonly used in neural network studies, this drastically reduces performance after the first 5 training epochs, but later provides the same accuracy.

ii) The V1 model reaches higher accuracy than the control models, even when each control model is trained much longer. Removal of the laminar structure (control model 3) reduces the accuracy more than the removal of neuron diversity (control model 2). But both of these control models still reach higher accuracy than the standard RSNN (control model 4). Finally, partitioning the neurons of the standard RSNN into excitatory and inhibitory neurons (control model 5), which could be seen as an interesting interpolation between the most abstract RSNN model and the V1 model, reduces performance in comparison to the generic RSNN.



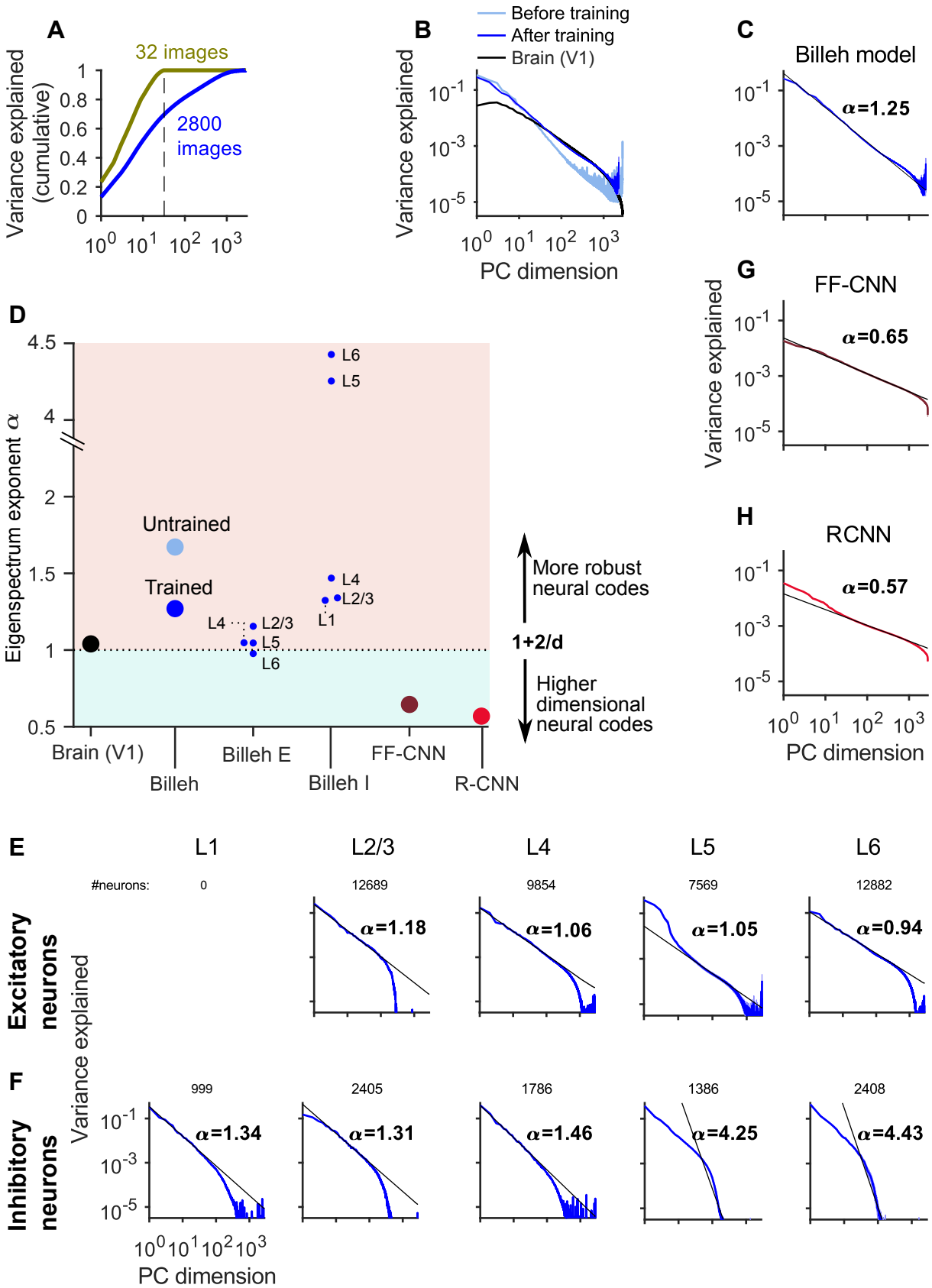
(caption next page)

Figure 4 ***(previous page)*****: Impact of details of the V1 model on learning speed and resulting computational performance. (A)** Accuracy on test examples of the V1 model and of 5 control models after varying numbers of training epochs. Whereas neuron diversity and the inclusion of the LGN model have only moderate impact on the learning speed and resulting performance, deletion of the laminar structure drastically reduced learning speed and significantly reduces the performance level that can be reached. RSNN control models have similar deficiencies, with further reduction in achievable performance. Shaded areas represent the SEM over 5 runs. **(B, C)** Spread of gradient information from a single readout neuron in L5 to other neurons depends on their distance, shown after 25, 100, 150 ms of backpropagation from the readout neuron. Whereas these learning signals primarily reach nearby neurons in the V1 model, they are uniformly spread out all over the network in a randomly connected control model 4 (RSNN). The loss function is the one used in the image classification task, but applied here only to a single readout neuron in the center of L5. **(D)** Resulting correlation of total gradient information between neuron pairs in the V1 model and the RSNN, collected during a full backprop pass (600 ms) of the loss function used in the visual change detection task of natural images. Nearby neurons in the V1 model receive correlated gradient information, which supports coherent learning in local network modules.

2.4 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 (*11*). 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 (*1*), and found that the model exhibited the same coding property (Fig. 5A). It was also shown in (*11*) 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, 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 (*25*) also predicts that a value of α close to 1 enhances under mild conditions downstream generalization performance. In-vivo recordings of (*11*) found that the value of α for primary visual cortex of mouse is 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 (*12*). 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 (*11*) to the V1 model, for a set of 2,800 randomly drawn natural images. Figure 5B shows the eigenspectrum of PCA component for the model of (*1*) before and after training, and also the measured eigenspectrum of V1 responses from (*11*). 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 power law (Fig. 5C) is for the model somewhat higher than in the brain. Figures 5D-F 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 (*12*) that feedforward CNN has a substantially smaller value of α (Fig. 5G) and found that the recurrent CNN model of (*26*), which also achieves very high accuracy for



(caption next page)

Figure 5 ***(previous page)*****: Comparing the eigenspectrum of neural codes in the V1 model of (*1*) 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 (*1*) with an ensemble of 2,800 randomly chosen natural images, and for mouse V1 (*11*). **(C)** Exponents of the power law for the V1 model of (*1*); all for the same ensemble of 2,800 randomly drawn natural images. **(D)** Summary result for exponents of the power law for the brain the V1 model of (*1*), and for CNNs. Billeh E and Billeh I mark results where the eigenspectrum is computed for specific populations of excitatory and inhibitory neurons of the model of (*1*).

(E) Eigenspectra of excitatory neurons on different layers of the model by (*1*) exhibiting values close to the measured value from a large sample of neurons in V1. **(F)** The same for inhibitory neurons. These exhibit on all layers except L2/3 substantially smaller coding fidelity. **(G, H)** Same as in **(C)** but for feedforward CNNs and recurrent CNNs, respectively.

image classification, has an even smaller value of α (Fig. 5H). 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 (*11*) to its remarkable noise robustness.

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

Our preceding analyses of neural coding in the V1 model of (*1*) 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. 6A for samples. Fig. 6B 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. To check whether CNNs become more noise robust when they are, like the V1 model, subjected to internal noise during training, we applied dropout (*27*) to them: We randomly replaced tensor outputs of ReLu layers with probability p with 0. It turns out that the robustness of CNNs is not improved substantially by that, see the results for various values of p in Fig. S15.

Since neither of these networks had been trained with the noisy images for which they were tested, we analyzed here a particular OOD (out-of-distribution) generalization capability of the V1 model and of CNNs. Figure 6C demonstrates that even if CNNs are trained with 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), even if no noise had not been present during training (Fig. S5B), can be viewed as an OOD generalization capability.

2.6 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 (*3*). They reported that dimensions in which differences between visual stimuli were encoded in area V1 of

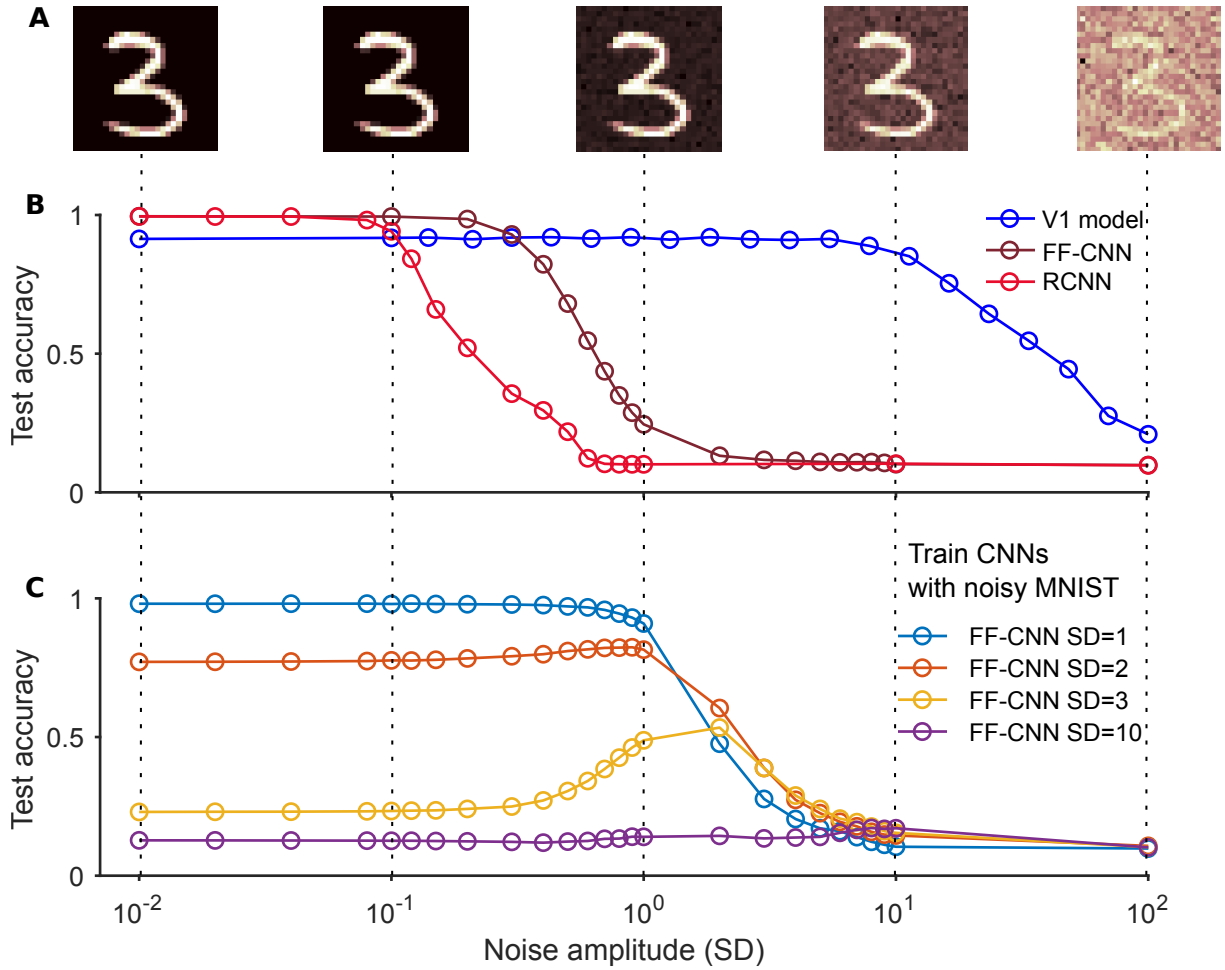


Figure 6: Robustness of the V1 model and of CNNs to noise in images. (A) Samples of handwritten MNIST digits with Gaussian noise drawn independently from $\mathcal{N}(0, \mathrm{SD})$ for each pixel, for different values of the SD. **(B)** While the V1 model never quite reaches the peak performance of 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 particular noise statistic (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.

mouse, i.e., neural coding dimensions, were nearly orthogonal to the largest noise modes. This result provided evidence that neural coding in V1 was even robust to correlated noise. In other words, cortical design principles appear to enhance coding robustness by restricting most noise fluctuations in area V1 to dimensions of the population activity that do not impede neural coding of visual inputs. We wondered whether the V1 model would inherit this important property. Therefore we carried out the analysis of (*3*) also for the V1 model, using the same visual stimuli (moving gratings with orientation differences that were close to the perception threshold). The discriminability index d' from (*3*) is a measure that they proposed as a measure for the fidelity of neural population coding, see the illustration in Fig. 7A. 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. 7B 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. 7C). Furthermore, compared with the untrained V1 model, training of this model moved the signaling dimensions to become more orthogonal to the largest noise dimension, see Fig. 7D. 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 (Materials and Methods).

It had been argued in (*3*) 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 (*11*). A possible explanation of this discrepancy was offered in (*4*): 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 (*3*). They suggested that the apparent "ceiling" would rise if one records from more neurons. We tested this hypothesis in the V1 model of (*1*), 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. 7E.

2.7 Concrete anatomical and neurophysiologal features of the V1 model that are responsible for its superior noise robustness

We had shown in Fig. 3D that the V1 model is robust to rather high levels of internal noise. Figure 8A demonstrates that the laminar spatial organization of the V1 model and its diversity of neuron types are both important factors for that. This is corroborated by the analysis of the discriminability index d' for control models, where also the projection of neural codes onto the noise dimensions with the largest eigenvectors assumes significantly larger values for the corresponding control models 2 and 3 than for the V1 model (Fig. 8B). Figure 8C shows that the laminar spatial organization and diversity of neuron types of the V1 model are also both major factors for the robustness of the V1 model to external noise in visual inputs (compare with Fig. 6). Furthermore, in comparison with CNNs, neurons in the V1 model are affected much less by this external noise, see Fig. 8D. We hypothesize that a major factor for that is the substantially larger in-degree of neurons in the connectivity graph of the V1 model (on average, 278) compared with a CNN (9), which enables them to integrate over a much larger number of image pixels, thereby making their output less dependent on noise in individual pixels. This is insofar interesting, as the small indegree of stereotypical CNN units is a cornerstone of the CNN architecture. The control models also exhibited a power spectrum of neural codes, like the V1 model (Fig. 5). The exponent α of the power spectrum was further away from the ideal value $1 + 2/d$ than that of the V1 model (Fig. S16),

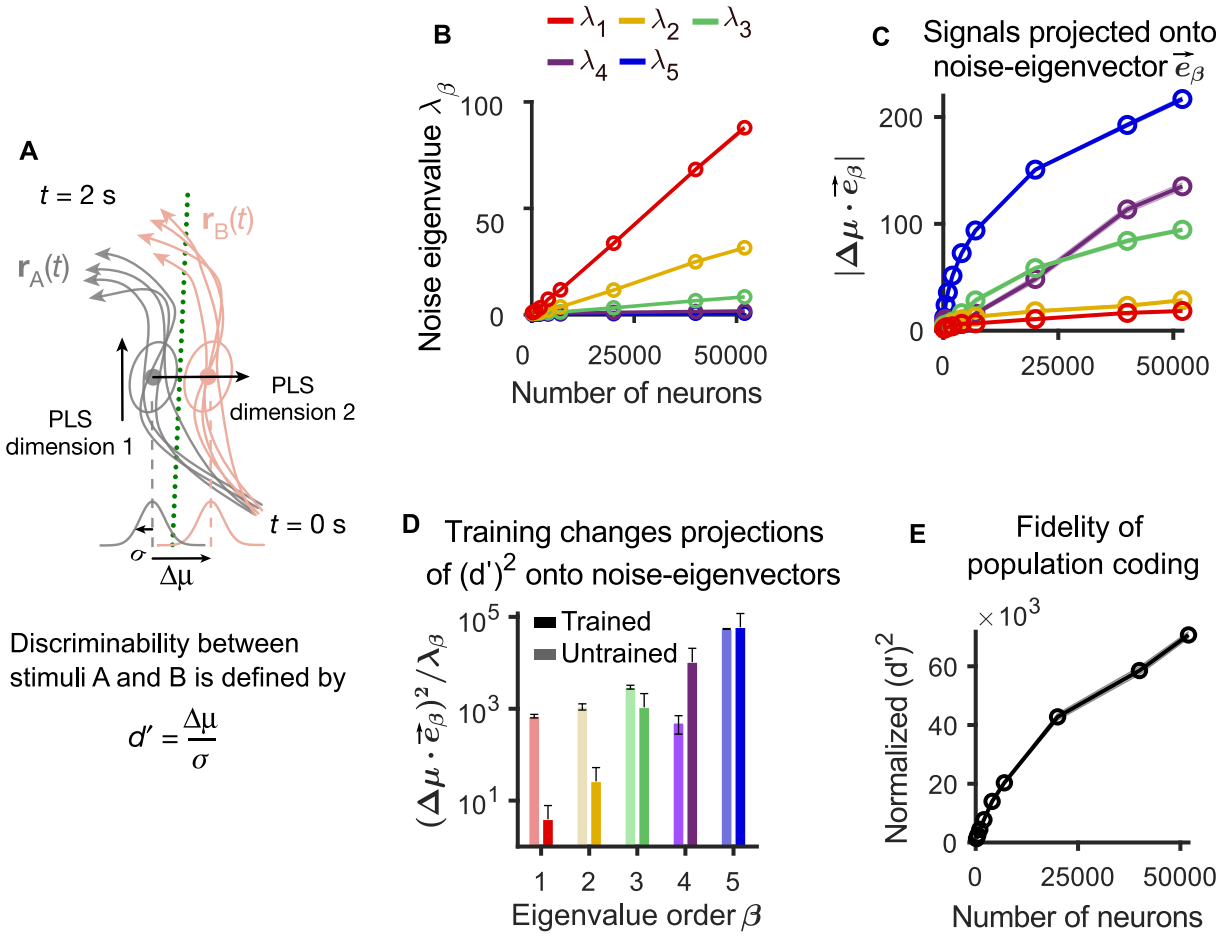
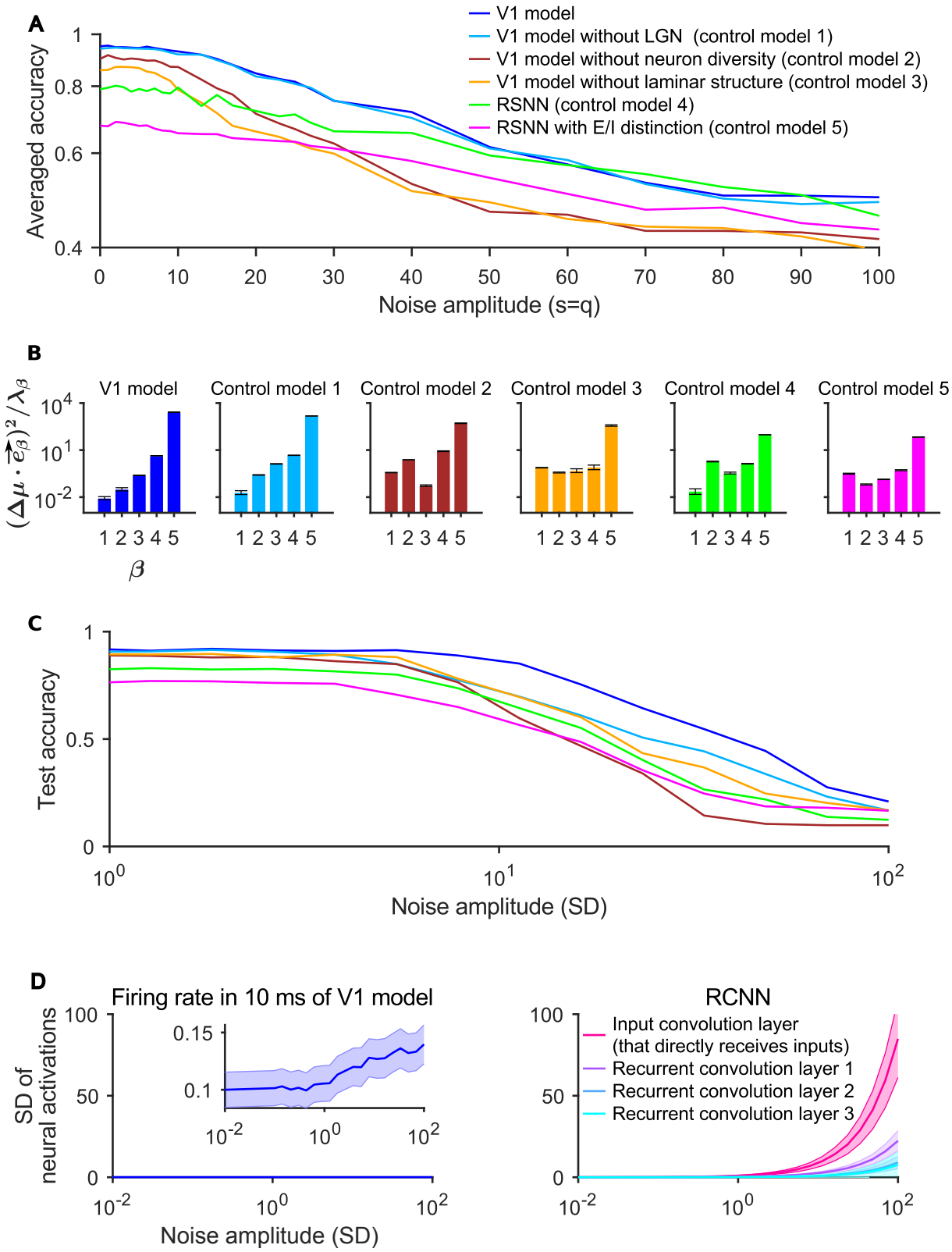


Figure 7: Analyses of relations between signaling and noise dimensions in the V1 model, and of the impact of noise correlations when one considers larger numbers of neurons. (A) Schematic of the calculation of the discriminability index d' according to (*3*). **(B)** The 5 largest eigenvalues λ_β ($\beta = 1, 2, \cdots, 5$) 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 SEM over 100 trials.

suggesting that these control models are operating in a less desirable regime. But their α values were larger than that of the V1 model. This would predict better noise robustness according to the underlying theory, but this theoretical prediction is inconsistent with our previously discussed empirical analysis.

The LGN model also contributes to the robustness of internal noise. As shown in Fig. 8A, control model 1 (where the LGN model is removed) is less robust to external noise than the V1 model with the LGN. Consistent with that, adding this LGN model (without further training of its parameters) as preprocessor to the RCNN drastically improves its noise robustness (Fig. S17). Hence the spatiotemporal filters of this LGN model can filter out some of the noise, which is consistent with the preceding results of (*28*).



(caption next page)

Figure 8 ***(previous page)*****: Concrete factors of the V1 model that are responsible for its superior noise robustness.** **(A)** Decay of accuracy of the V1 model and 5 control models for increasing amplitudes of internal noise (see Fig. 3D for a preceding analysis of this for the V1 model). One sees that both neuron diversity and laminar structure are important factors for the robustness of the V1 model to internal noise. **(B)** Orthogonality of neural coding and noise dimensions in the V1 model (see Fig. 7D) and 5 control models after training. Only control model 1 (V1 model without LGN) has similarly small projections onto the noise eigenvectors with the largest eigenvalues, i.e., for small values of β. Neuron diversity and laminar structure turn out to be important factors for that. **(C)** Decay of accuracy of the V1 model and 5 control models for image classification (MNIST task) with increasing amount of noise in the visual input. Neuron diversity and laminar structure of the V1 model turn out to be also essential for the superior robustness of the V1 model to noise in the input. Color scheme is the same as in **(A)**. **(D)** Analysis of the impact of the same noise as in **(C)** on neurons in the V1 model and in the RCNN of (*26*). When the SD of the pixel noise becomes larger than 1, which is the range where the performance of the RCNN becomes substantially worse than that of the V1 model according to Fig. 6B, the output of RCNN units becomes substantially stronger affected by this input noise than the firing activity of neurons in the V1 model. Shaded areas in left and right represent 0.1-fold SD and SD across neurons, respectively. Inset of left panel shows the same data but for a different order of magnitude.

3 Discussion

We have demonstrated that the V1 model of (*1*), 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. In particular, we found that this V1 model exhibits interesting advantages with regard to learning speed and visual processing performance over models that are closer to common artificial neural network models, such as randomly connected networks of spiking neurons (RSNNs) and CNNs. Furthermore, we have isolated concrete anatomical and neurophysiological features of the V1 which are responsible for that.

CNNs are currently most commonly used for visual processing in AI. 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 is more versatile since it can solve a number of diverse visual tasks with the same synaptic weights, 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 (*1*), which integrates a large body of experimental data on area V1 in the brain, is able to perform similarly versatile and robust visual processing. Furthermore, we could identify concrete anatomical and neurophysiological features of the V1 model which are responsible for this.

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 it 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, applying the same training process to CNNs, CNNs exhibit a power law with a substantially slower decay of the eigenspectrum. According to the theoretical analysis of (*11*), this implies that neural codes in CNNs are less noise-robust. It should, however, be noted that by regularizing CNNs (*29*) or training them with different methods (*25*) one can move the exponent of their eigenspectra closer to $1 + 2/d$, which can improve the robustness of CNNs. We have also shown that the anatomical and neurophysiological data on V1 that have been integrated into the V1 model suffice to reproduce a further salient aspect of neural coding in area V1 of the brain, where according to (*3*) about 90% of the noise fluctuations in area V1 are constrained to dimensions of the population activity that are orthogonal to noise dimensions. We found that this is an emergent property of the V1 model of (*1*).

On a more general level, we have shown that the V1 model of (*1*) 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 on 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.

Our method can also be applied to investigate more detailed models of V1, e.g., models that include data on the lattice of microcolumns of neurons in L5 (*30*), on short-term plasticity (*31*), and functionally salient aspects of dendritic spikes (*32*). It also provides a paradigm for elucidating how anatomical and neurophysiological details of interconnected higher and lower brain areas carry out distributed computations, in particular how interaction of neurons in superficial layers of V1 with higher cortical areas enhances visual processing capabilities of V1. Although we have focused here on readout neurons in L5, the sparsely firing pyramidal cells in L2/3 of the trained V1 model contain already most of the information that is needed to solve the computational tasks of the network, and hence they could potentially transmit this to higher areas: Readouts from these neurons (which might be seen as proxies for neurons in higher cortical areas that receive input from V1) can be trained to solve all 5 tasks with high accuracy (Fig. S18). 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. (*6, 33, 34*). 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. Also, neurons on L2/3 of the V1 model will then be placed into a biologically more realistic context where they send computational results to higher areas, and receive inputs from them.

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. 7E 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 (*4*)). 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. 5F. 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 implementation two attractive features: 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. 4A). In addition, computations are carried out in V1 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. Since we have shown that both of these features can be reproduced in a corresponding neural network model for visual processing, this V1 model suggests that it will also be possible to recruit them for the design of substantially more energy-efficient neuromorphic implementations of visual processing (*35*).

4 Materials and 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 (*1*). To make it gradient-friendly, we replaced the hard reset of membrane potential after a spike emerges with the reduction of membrane potential $z_j(t)\,(v_{th} - E_L)$, 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. E_L the resting membrane potential. This causes no significant change in the neural response (Fig. S4). We simulated each trial for 600 ms. The dynamics of the modified GLIF_3 model was defined as

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

where C represents the neuron capacitance, 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

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

where the multiplicative constant $f^m = \exp\left(-k^m \delta t\right)$ 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 (*2*), see (*36, 1*), 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 V1 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 (*1*) 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 (*1*) and converted to integers as the integration step is 1 ms.

The postsynaptic current of neuron j was defined by the following dynamics (*1*):

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

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

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 (*1*).

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 (*1*) 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 (*11*) (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) (*11*) 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 (*11*) 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. S3B). The results were not sensitive to the number of neurons in each population. We tested populations consisting of 15 and 60 readout neurons, and found that the performance difference is less than 0.2%. We had also shown in (*20*) that picking just 2 neurons from the trained readout pool to produce network outputs provided almost the same accuracy for the visual change detection task.

We also considered the case where the neurons in these readout populations were randomly distributed in L5 (Fig. S3C), 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. S3A).

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 (*1*) (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 (*37*). 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 (*3, 4*). 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 (*4*), 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 (*4*). 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 (*38, 6*), 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 (*39*). 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 (*20*).

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., (*7, 8*)). We considered the same task that was studied in the experiments of (*7, 8*). 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, $\lambda_{\mathrm{f}} = 0.1$ and $\lambda_{\mathrm{v}} = 10^{-5}$. The rate regularization is given by the Huber loss (*40*) between the target firing rates, y, calculated from the model in (*1*), and the firing rates, r, sampled the same number of neurons from the network model:

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

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

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

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

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

where the dampening factor $\gamma_{\mathrm{pd}} = 0.5$, the Gaussian kernel width $\sigma_p = 0.28$. During the refractory period, the pseudo derivative was set to 0.

We drew a batch of visual stimuli (in our case batch size is 320, consisting of 64 network inputs for each of the 5 tasks) and calculated the gradient after every trial for each synaptic weight whether an increase or decrease of it (but without changing its sign) would reduce the loss function. Weights were then updated by the average gradient across the batch. This method had originally only been applied to

Table 1: The V1 model achieves high accuracy in all 5 tasks, consistent with the behavior performance of mice in similar tasks, after 16 training epochs. The chance level of performance in image classification task is 10%, and those in other tasks are 50%.

	Test accuracy	Behavior accuracy	Mean firing rate (Hz)	Spike raster
Fine orientation discrimination	93.17%	$\sim 83\%^{a}$	3.97	Fig. S6
Image classification	92.11%	N/A	4.11	Fig. S7
Visual change detection of natural images	92.13%/92.14%[b]	$\sim 73\%/77\%^{c}$	3.97	Fig. S8
Visual change detection of gratings	94.64%	$\sim 60\%^{d}$	3.90	Fig. S9
Evidence accumulation	99.32%	$\sim 85\%^{e}$	3.96	Fig. S10

(a) Estimated from Fig. 4C of (*41*) 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 (*5*) when familiar and novel images were presented to mice, respectively.
(d) Estimated from Fig. 3A of (*42*) when the orientation difference was 5°.
(e) Estimated from Fig. 1C of (*7*) when the number of cues was 6.
Note that the behavioral experiments had longer time delays that made the tasks somewhat more difficult.

neuron networks with differentiable neuron models and was normally referred to as stochastic gradient descent.

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.

In every training run, we used a different random seed in order to draw fresh noise samples from the empirical distribution, and for randomly generating/selecting training samples.

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 V1 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 (*1*) 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 16 epochs to make sure the average performance on the 5 tasks was saturated. This took 60 h of wall clock time on 160 GPUs.

Table 2: The control model settings.

Model name	Neuron model	Connectivity[b]	Neuron type[c]	Weight sign constraint	Input via LGN model
V1 model	$GLIF_3$	V1	V1	Yes	Yes
V1 model without LGN (control model 1)	$GLIF_3$	V1	V1	Yes	No
V1 model without neuron diversity (control model 2)	$GLIF_3$	V1	1E 1I	Yes	Yes
V1 model without laminar structure (control model 3)	$GLIF_3$	random	V1	Yes	Yes
RSNN (control model 4)	LIF[a]	random	single	No	Yes
RSNN with E/I distinction (control model 5)	LIF	random	single	Yes	Yes

(a) LIF represents standard LIF neuron model.
(b) V1: the connectivity in the V1 model; random: randomly selected connectivity but the number of connectivity is fixed as in the V1 model.
(c) V1: the diverse neuron types as in the V1 model (in total 111 types); 1E 1I: all excitatory neurons are the same (the excitatory neuron on L2/3, node type id in Allen brain atlas: 487661754) and all inhibitory neurons are the same (PV neuron on L2/3, node type id:484635029); Single: all neurons are the same.

4.10 Control models

We used 5 control models in Fig. 4. In control model 1, we removed the LGN model, and directly injected pixel values of the image into the V1 model. The image was resized to a number of pixels that roughly matched the number of LGN channels (2544). The pixel values were scaled by a factor 0.04 to bring the resulting firing activity into a reasonable regime. In control model 2, we removed the diversity of the 111 data-based neuron types in the V1 model, replacing them with one generic model for excitatory neurons (the excitatory neuron on L2/3, node type id in Allen brain atlas: 487661754) and one for inhibitory neurons (PV neuron on L2/3, node type id:484635029). In control model 3, we removed instead the laminar spatial structure with distance-dependent connection probabilities of the V1 model, replacing them with an equal number of randomly chosen connections. Control model 4 is a randomly connected recurrent network of standard LIF neurons (RSNN) with the same number of neurons and connections as the V1 model. The membrane constant of these LIF neurons is 10 ms, and their refractory period is 5 ms long. These are common values for spiking neural network models. Control model 5 is a variation of this RSNN where excitatory and inhibitory neurons are distinguished like in the V1 model, and Dale's law is observed during training. This type of model is sometimes seen as an intermediate step from generic RSNNs toward more biologically oriented network models. The detailed settings are shown in Table 2. All control models were trained in the same way as the V1 model, i.e., for the same tasks and loss function, including the same sparsity regularization term, and using the same hyperparameters for training.

4.11 Branching ratio as a measure for criticality

Based on the work of (*43*), 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 Convolutional neural networks

Feedforward CNN. We used ResNet-18 (*44*) as FF-CNN. To calculate its eigenspectra, we used the pre-trained version on ImageNet provided by PyTorch. To evaluate its robustness to 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 (*26*) as RCNN, 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 (*26*). To evaluate its robustness to pixel noise, we trained RCNN on MNIST with the stochastic gradient descent 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.

Adding dropout to CNN training. To introduce internal noise to the FF-CNN and the RCNN during training, we independently set outputs of ReLu units with probability p to 0, using samples from a Bernoulli distribution. This procedure is known as dropout training (*27*).

4.13 Eigenspectrum analysis

cvPCA. Eigenspectra of V1 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. S2B). 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 (*11*), 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 (*11*). 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 V1 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 V1 model were averaged over these three models. The code is available in https://github.com/MouseLand/stringer-pachitariu-et-al-2018b.

V1 model. We analyzed the neural representation in the trained V1 model in the same way as responses of V1 neurons were analyzed in (*11*): Without loss of generality, we used 2,800 generic images which were randomly drawn from ImageNet validation dataset in all panels of Fig. 5. We also tried the 2,800 nature images used in (*11*) and found they gave rise to a slower decaying speed of eigenspectrum (1.15); there were only used in Fig. S2A 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 (*3*), we used the metric d', which characterizes how readily the distributions of the neural responses to the two different sensory stimuli can be distinguished (*45*). The quantity $(d')^2$ is the discrete analog of Fisher information (*46*).

Figs. 7A and E. To be consistent with the experimental study (*3*), 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 (*3*). 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 V1 model, we reduced dimensional by using partial least squares (PLS) analysis (*47*) to identify and retain only 5 population vector dimensions in which the stimuli were highly distinguishable as in (*3*). 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 (*46*). 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. 7B. Noise eigenvalue λ_β and its eigenvector $\vec{e}_\beta$ was calculated by eigen-decomposition of noise covariance matrix Σ.

Fig. 7C. 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. 7D. 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.

Acknowledgments: 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. **Funding:** 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). This research was partially supported by the Human Brain Project (Grant Agreement number 785907) of the European Union and a grant from Intel. **Author contributions:** G.C. and W.M. designed the study. G.C. and F.S. built the software. G.C. performed the experiments and analyses. G.C. and W.M. wrote the paper. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

References

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

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

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

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

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

6. J. H. Siegle, *et al.*, Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature* (2021).

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

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

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

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

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

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

13. J. Wilting, V. Priesemann, Inferring collective dynamical states from widely unobserved systems. *Nature communications* **9**, 2325 (2018).

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

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

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

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

18. W. Maass, H. Markram, On the computational power of circuits of spiking neurons. *Journal of computer and system sciences* **69**, 593–616 (2004).

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

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

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

22. D. L. Ringach, P. J. Mineault, E. Tring, N. D. Olivas, P. Garcia-Junco-Clemente, J. T. Trachtenberg, Spatial clustering of tuning in mouse primary visual cortex. *Nature communications* **7**, 1–9 (2016).

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

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

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

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

27. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research* **15**, 1929–1958 (2014).

28. J. Dapello, T. Marques, M. Schrimpf, F. Geiger, D. Cox, J. J. DiCarlo, Simulating a primary visual cortex at the front of cnns improves robustness to image perturbations. *Advances in Neural Information Processing Systems* **33**, 13073–13087 (2020).

29. J. Nassar, P. Sokol, S. Chung, K. D. Harris, I. M. Park, On 1/n neural representation and robustness. *Advances in Neural Information Processing Systems* **33**, 6211–6222 (2020).

30. H. Maruoka, N. Nakagawa, S. Tsuruno, S. Sakai, T. Yoneda, T. Hosoya, Lattice system of functionally distinct cell types in the neocortex. *Science* **358**, 610–615 (2017).

31. L. Campagnola, S. C. Seeman, T. Chartrand, L. Kim, A. Hoggarth, C. Gamlin, S. Ito, J. Trinh, P. Davoudian, C. Radaelli, *et al.*, Local connectivity and synaptic dynamics in mouse and human neocortex. *Science* **375**, eabj5861 (2022).

32. M. E. Larkum, J. Wu, S. A. Duverdin, A. Gidon, The guide to dendritic spikes of the mammalian cortex in vitro and in vivo. *Neuroscience* (2022).

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

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

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

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

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

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

39. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, *2009 IEEE conference on computer vision and pattern recognition* (Ieee, 2009), pp. 248–255.

40. P. J. Huber, *Breakthroughs in statistics* (Springer, 1992), pp. 492–518.

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

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

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

44. K. He, X. Zhang, S. Ren, J. Sun, *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 770–778.

45. C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)* (Springer, 2007), first edn.

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

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

Supplementary Information for:

A data-based large-scale model for primary visual cortex enables brain-like robust and versatile visual processing

Guozhang Chen, Franz Scherr, and Wolfgang Maass

*Corresponding author: Email: maass@igi.tugraz.at

This PDF file includes:

Table S1
Figs. S1 to S18
Supplementary Note 1

Table S1: Performances on the 5 tasks is not very sensitive to the amplitude of internal noise in the network. The light cyan highlights the noise level used in this study unless otherwise stated. All tasks were tested on testing dataset.

	Fine orientation discrimination	Image classification	Visual change detection of nature images	Visual change detection of gratings	Evidence accumulation
$q = 1,\ s = 0$	94.66%	91.92%	93.30%	94.44%	99.20%
$q = 1,\ s = 2$	94.47%	92.11%	92.83%	94.64%	99.56%
$q = 1,\ s = 3$	94.18%	91.96%	92.48%	93.96%	99.48%
$q = 1,\ s = 4$	94.01%	91.77%	92.99%	94.68%	99.56%
$q = 2,\ s = 2$	93.17%	92.11%	92.13%	94.64%	99.32%
$q = 3,\ s = 3$	94.46%	91.66%	92.99%	94.56%	99.28%
$q = 3,\ s = 4$	94.61%	91.84%	92.32%	94.60%	99.28%
$q = 4,\ s = 3$	94.69%	91.50%	92.85%	93.40%	99.24%
$q = 4,\ s = 4$	94.39%	91.94%	92.43%	94.24%	99.20%
$q = 5,\ s = 4$	94.63%	92.08%	92.31%	95.12%	99.52%
$q = 10,\ s = 10$	92.27%	90.34%	88.29%	92.64%	97.20%
$q = 20,\ s = 20$	83.57%	84.39%	75.30%	90.96%	88.96%
$q = 200,\ s = 200$	45.79%	10.55%	50.52%	48.24%	49.08%

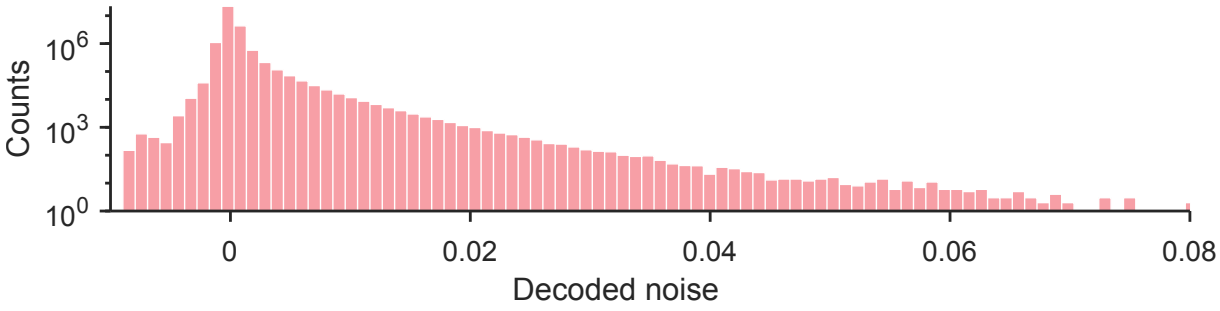


Figure S1: Empirical noise distribution from which noise values for our data-driven noise model are drawn. They are based on experimental data from (*11*) (see Data-driven noise model in Materials and Methods).

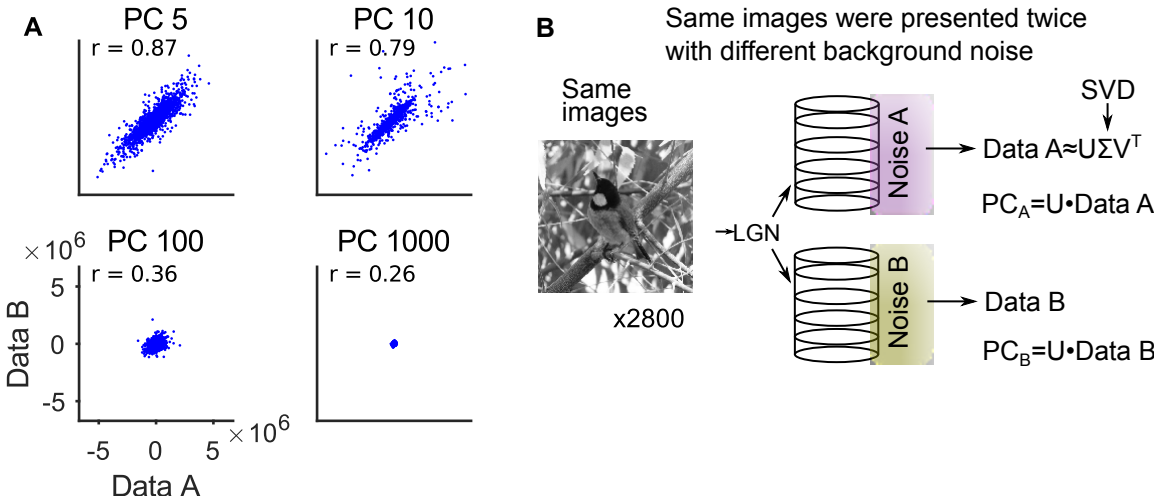


Figure S2: Schematic diagram of cvPCA (cross-validated PCA) to extract signal and noise according to (*11*). (A) Correlations (r) of neural responses for two presentations of the same natural image, using our data-driven noise model with $s = q = 2$ and projected onto selected principal components (PC). **(B)** 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.

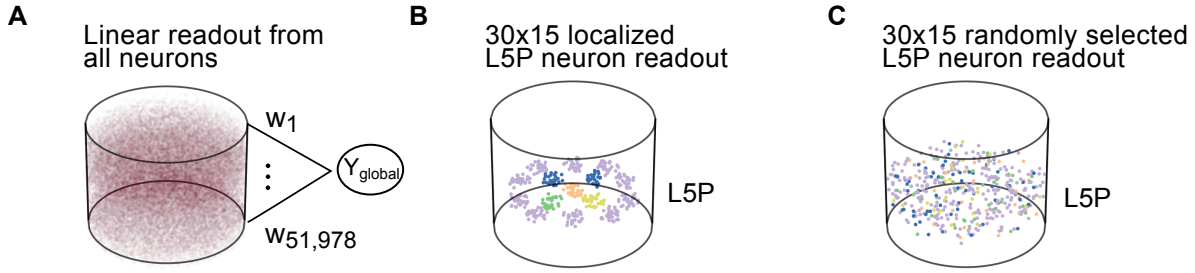


Figure S3: Schematic diagram of different readout scenarios. (A) Schematic diagram of a linear readout receives synaptic input from all 51,978 neurons in the microcircuit model, using a corresponding number of weights that can all be optimized for one particular task. **(B)** 15 spatially separated groups 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. **(C)** Alternative selection of these 15 populations in L5 without spatial clustering leads to very similar performance (an average accuracy of 94.67% instead of 94.28% for our standard readout convention from **(B)**).

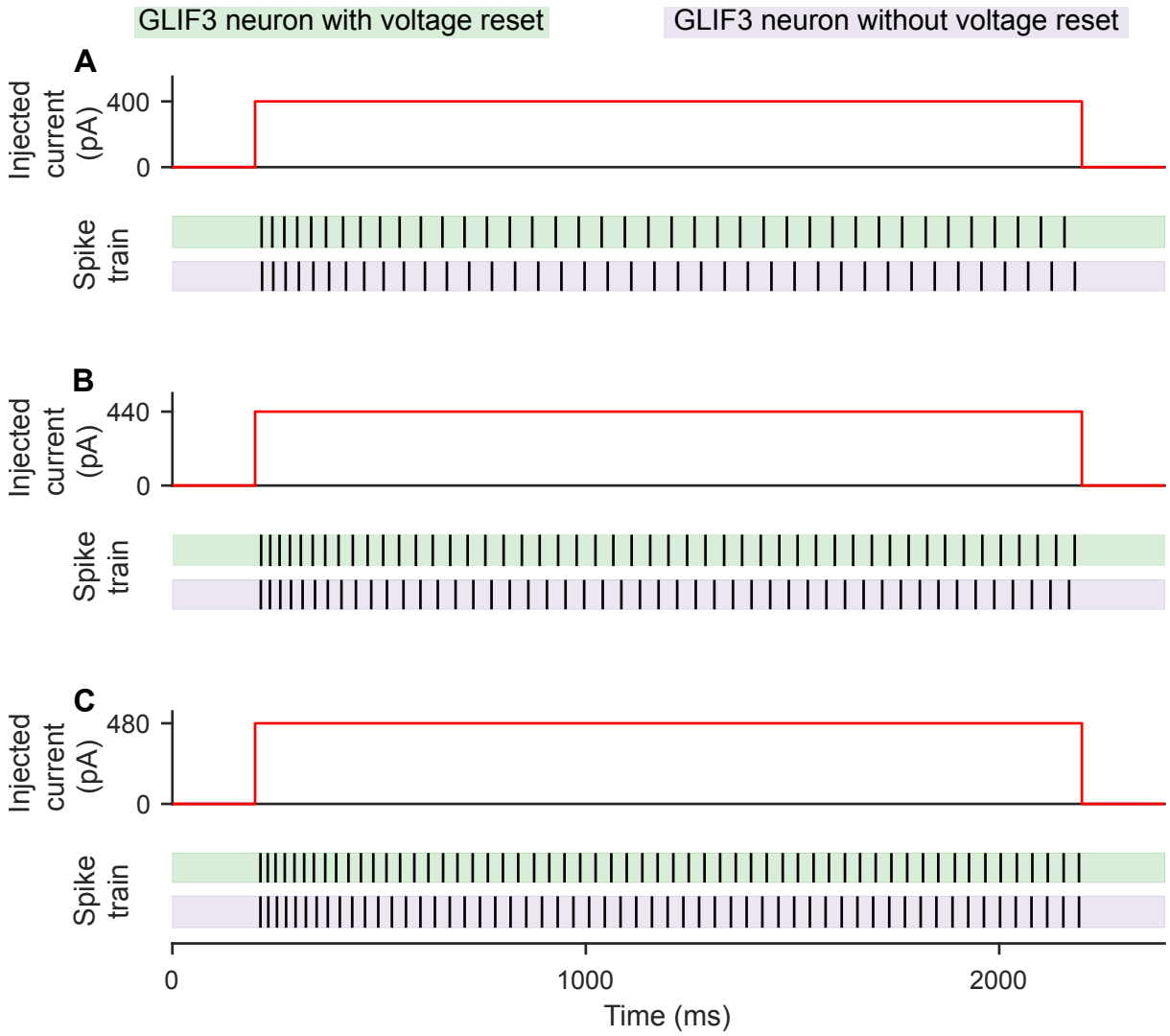


Figure S4: Adding voltage reset does not significantly affect the spike response of the GLIF$_3$ model. (A) Stimulus in form of a step function (top) was inputted to the GLIF$_3$ model with voltage reset used in this study (middle) and the GLIF$_3$ model without voltage reset in (*36*). The after-spike currents in the GLIF3 model cause a spike-frequency adaptation, i.e., they reduce the firing rate during the later part of the input current injection. **(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.

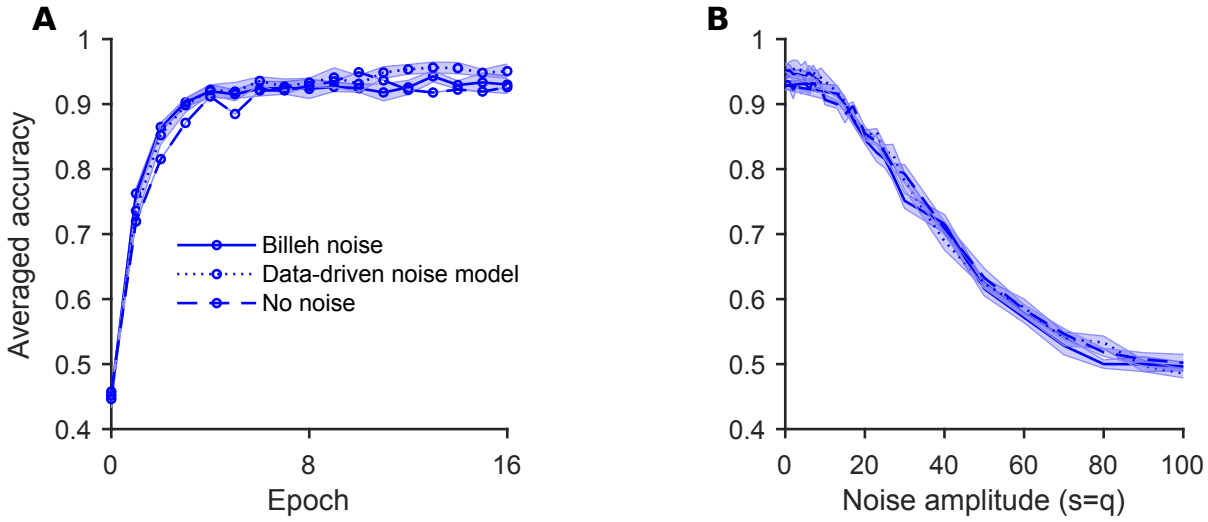


Figure S5: The effect of different noise scenarios during training on testing accuracy. (A) Accuracy on test examples of the V1 model trained with three noise scenarios: i) Single Poisson source, the noise used in (*1*) (Billeh noise), ii) data-driven noise model when $s = q = 2$, iii) no injected noise (No noise). The test accuracy is not sensitive to them. We used the data-driven noise model with $s = q = 2$ during the testing process. **(B)** Decay of accuracy of the V1 model trained with three different noise scenarios for increasing amplitudes of internal noise, see Fig. 3D for a preceding analysis for the V1 model. One sees that these noise scenarios do not dramatically affect the noise robustness of the V1 model.

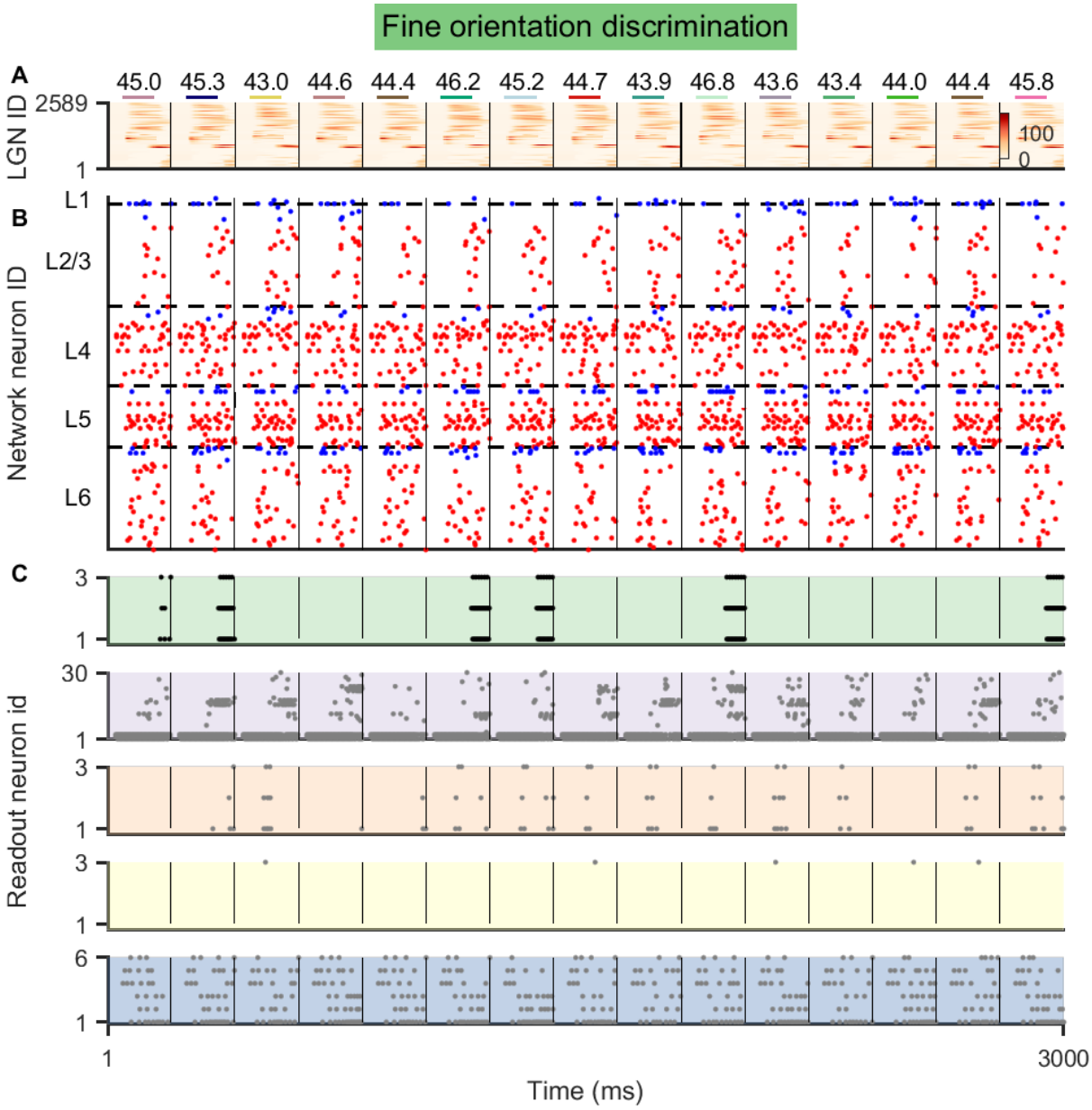


Figure S6: Sample run of the V1 model on input for the fine orientation discrimination task. (A) Colored lines represent the timing of input images. Numbers on them represent the orientations of input gratings. The bottom colormap demonstrates the activity of LGN neuron activity. **(B)** Spike raster 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 thick 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.

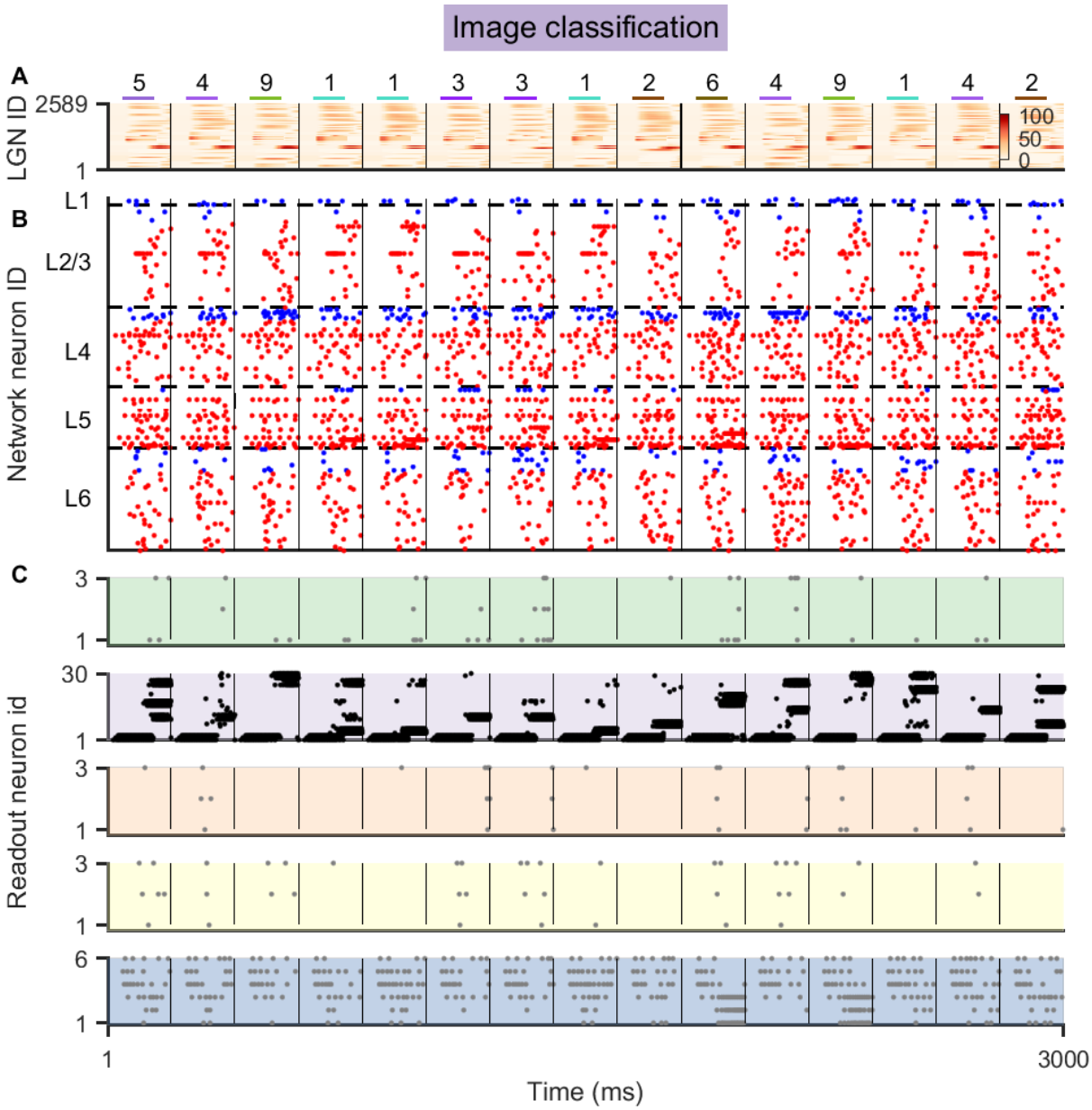


Figure S7: Sample run of the V1 model on input for the image classification task. (A) Colored lines represent the timing of input images. Numbers on them represent the digits in the input images. The bottom colormap demonstrates the activity of LGN neuron activity. **(B)** Spike raster of the laminar 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 thick 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.

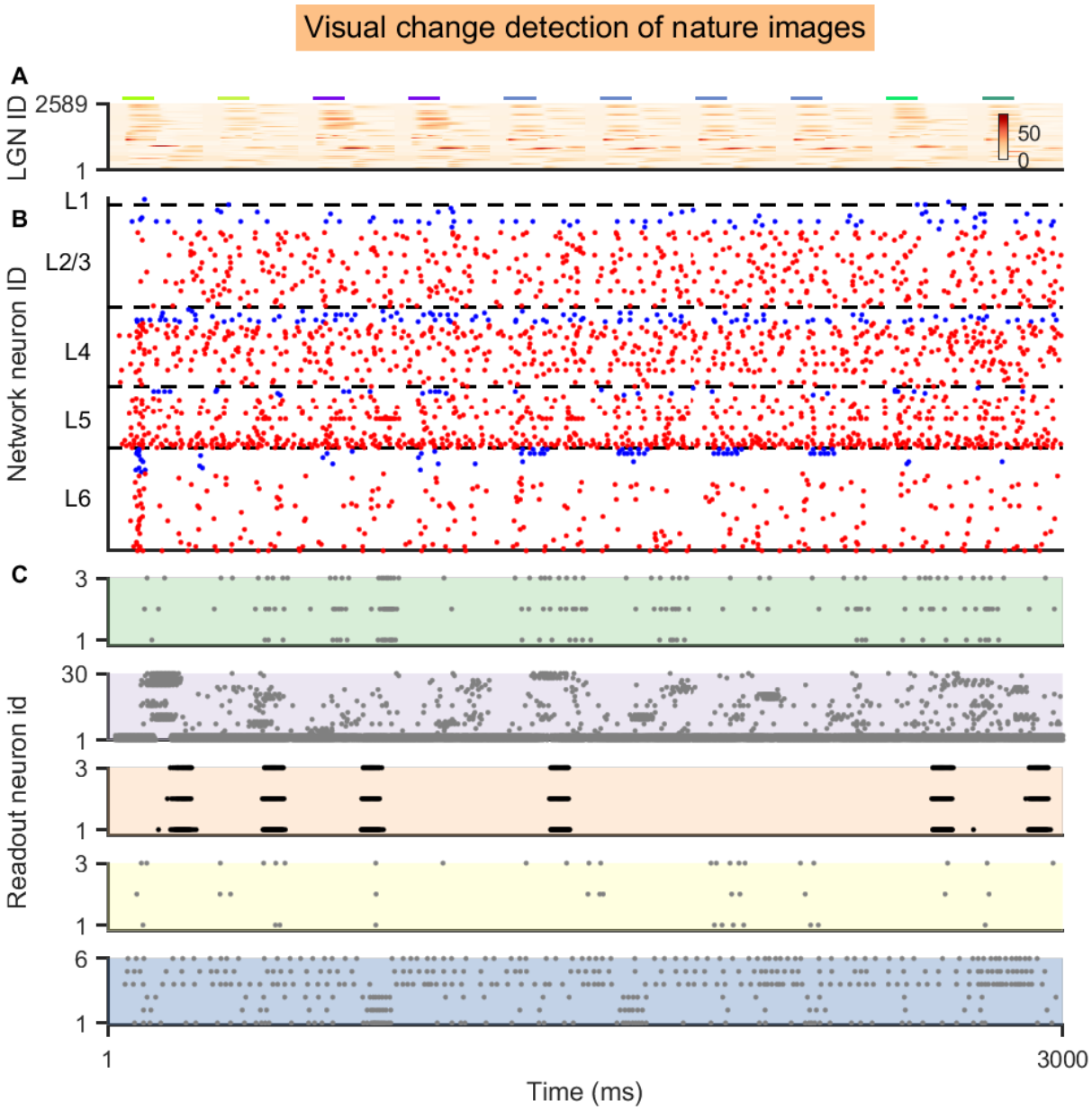


Figure S8: Sample run of the V1 model on input for the visual change detection task of natural images. (A) Colored lines represent the timing of input images and the colors code the image identity. The bottom colormap demonstrates the activity of LGN neuron activity. **(B)** Spike raster of the laminar 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.

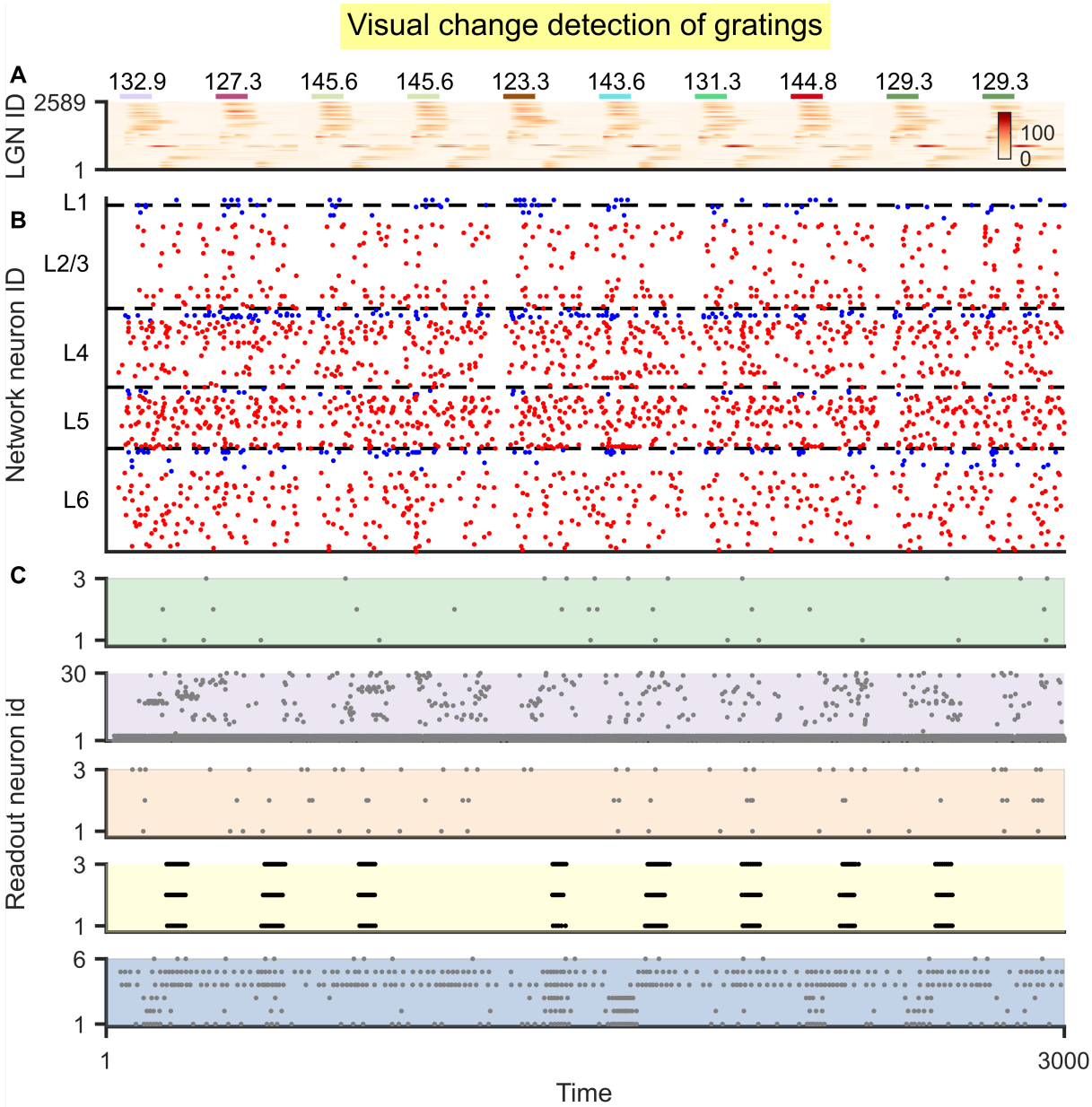


Figure S9: Sample run of the V1 model on input for the visual change detection task of gratings. **(A)** Colored lines represent the timing of input images and the colors code the image identity. Numbers on them represent the orientations of input gratings. The bottom colormap demonstrates the activity 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.

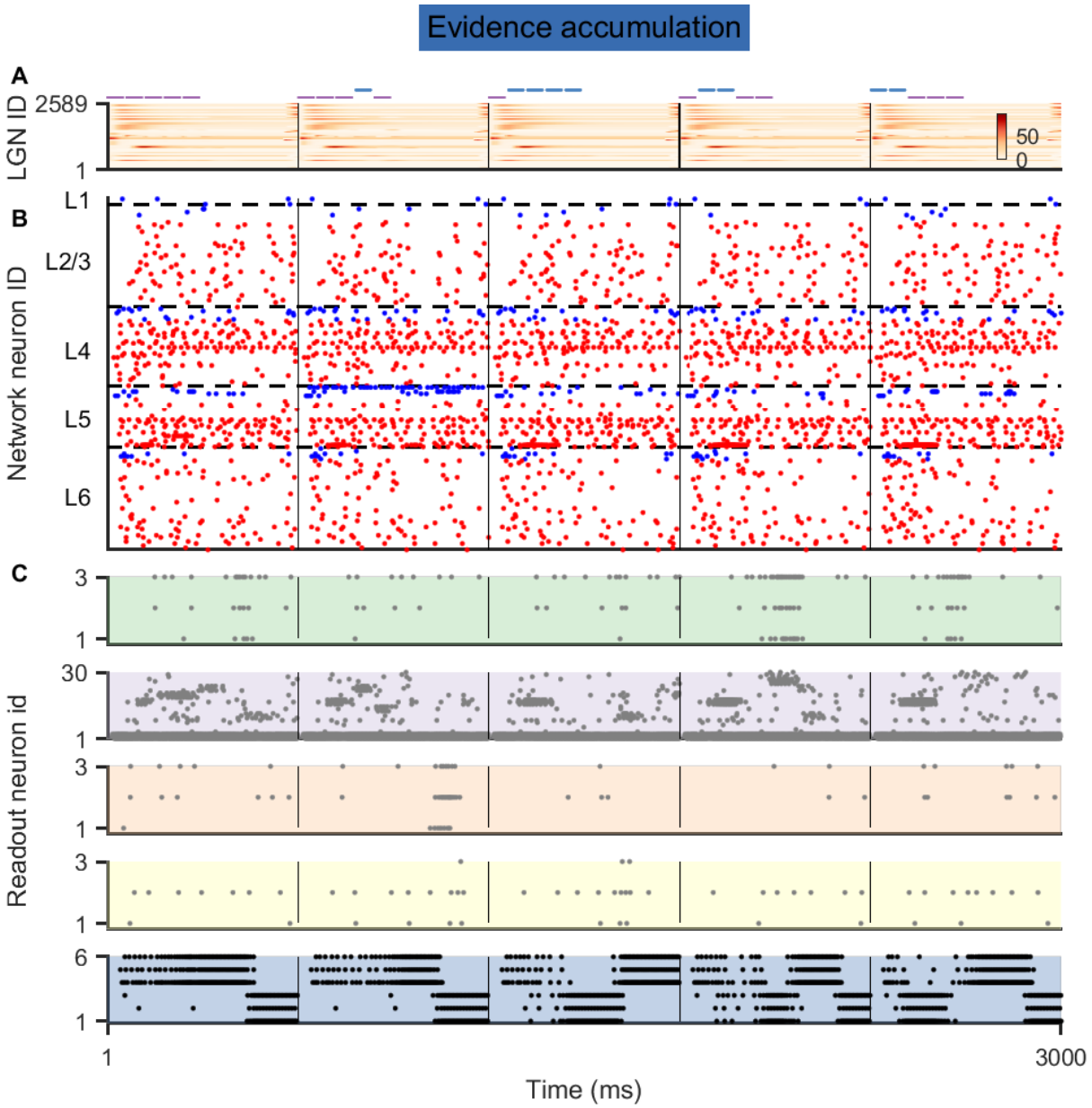


Figure S10: Sample run of the V1 model on input for the evidence accumulation task. (A) Colored lines represent the timing of input left/right cues. The bottom colormap demonstrates the activity of LGN neuron activity. **(B)** Spike raster of the laminar V1 model. Red and blue dots represent the spikes 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 thick 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.

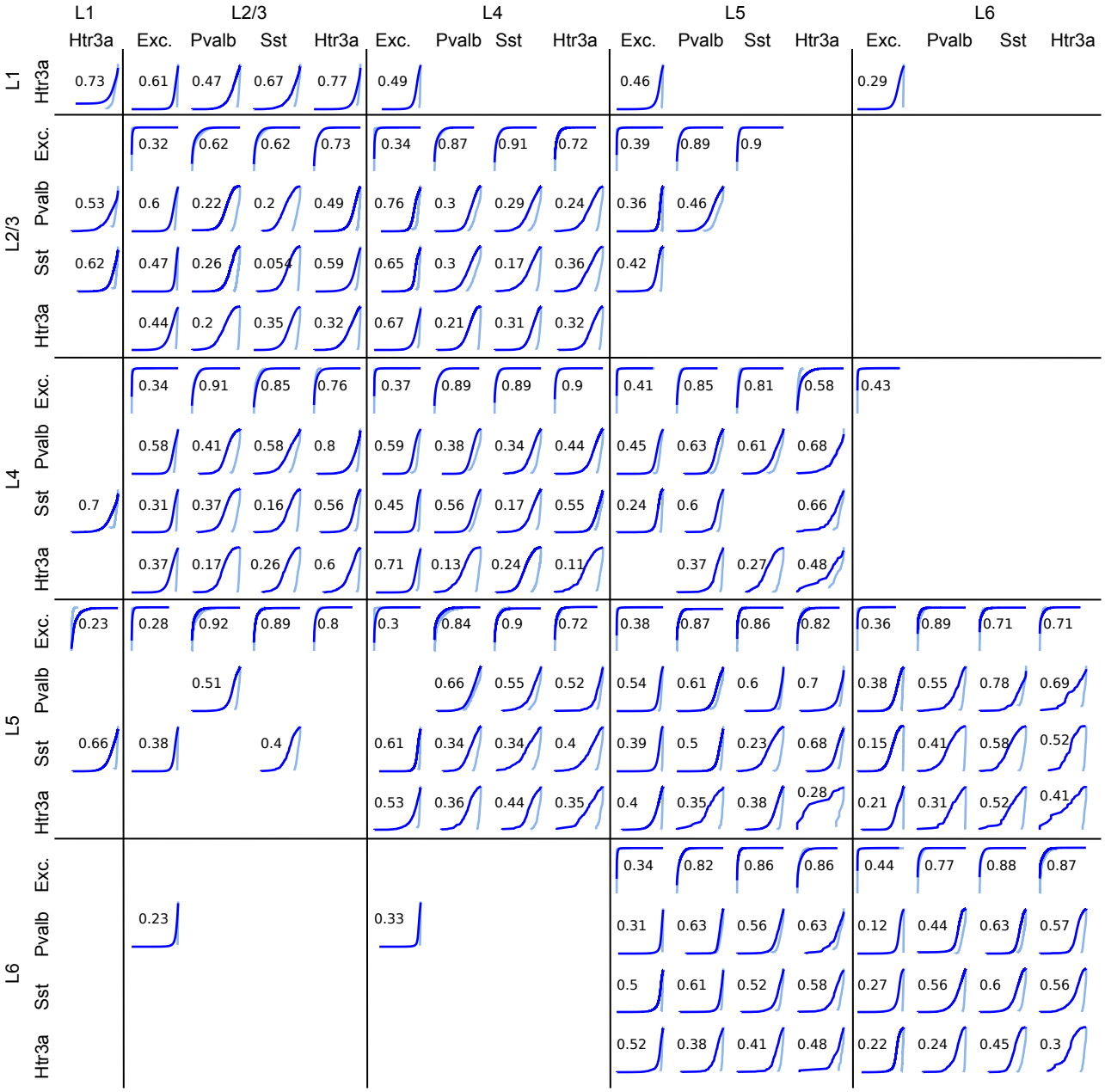


Figure S11: Distribution of recurrent weights between each pair of populations before (light blue) and after learning (dark blue) the 5 tasks. Each row represents a pre-synaptic neuron population, and each column represents a post-synaptic neuron population. The histogram represents the distribution 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 (*1*). Exc., excitatory neurons.

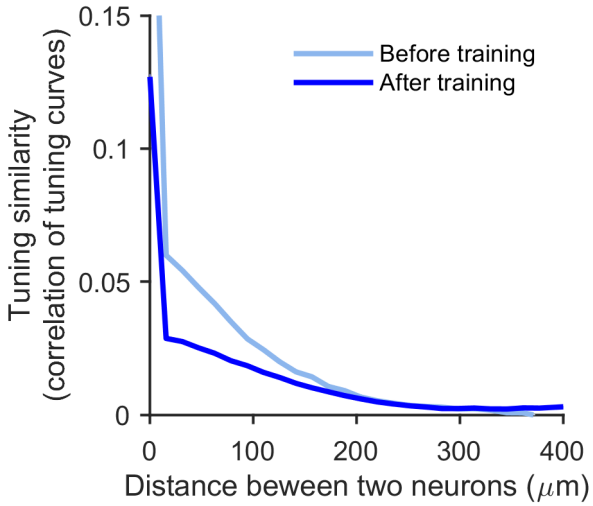


Figure S12: Spatial clustering of orientation tuning of excitatory neurons in L2/3 of the V1 model before and after training. The correlation of tuning curves of two neurons as a function of the horizontal distance between them. Firstly, we calculated the spike counts of L2/3 excitatory neurons in 100 ms when the model received gratings with different orientation (0 to 180° with the step of 5°). Then we calculated the correlation between the tuning curves of neuron pairs and plotted the correlation coefficient as the function of the horizontal distance between the neuron pairs. The spatial cluster is weaker than the experimental results in Fig. 2b of (*22*), probably because they calculated the correlation based on the joint tuning of neurons in the orientation and spatial frequency plane but we did not use the spatial frequency (we do not have a related task).

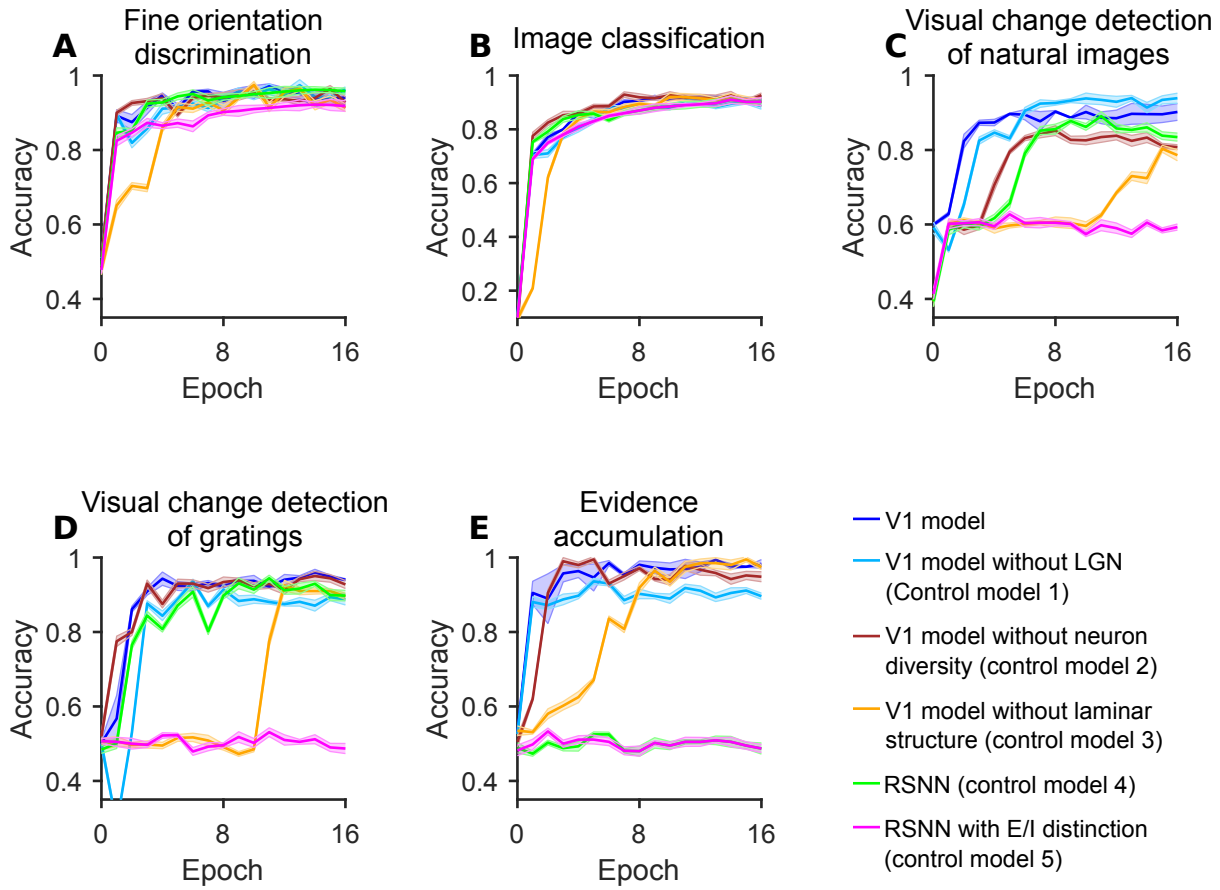


Figure S13: Task-wise accuracy on test examples for the V1 model and 5 control models, all trained for all 5 tasks. (A-E) The testing accuracy is shown for each task as function of training length (number of epochs). Shared area represent the SEM over 5 runs.

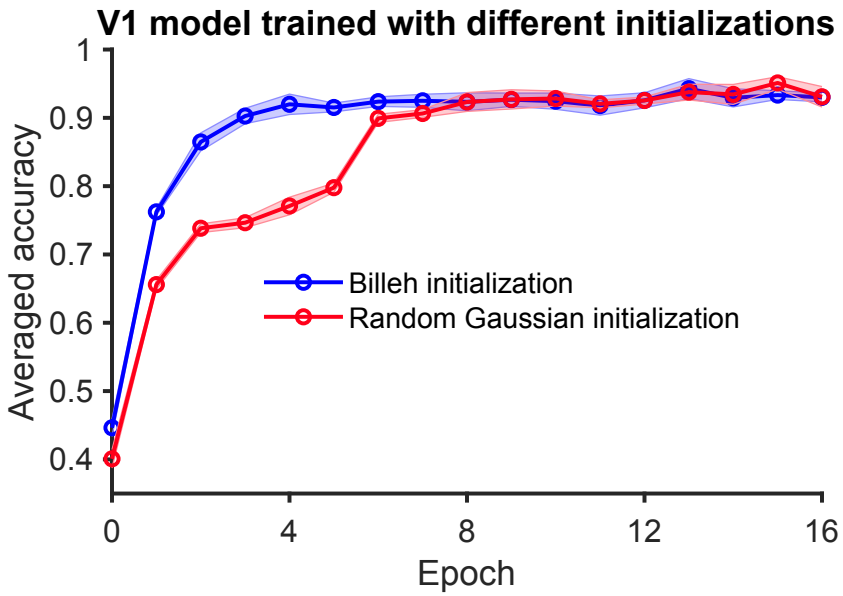


Figure S14: The effect of initialization of synaptic weights on the testing accuracy. The accuracy of V1 model trained with initialization of synaptic weights from (*1*) (Billeh initialization) and random Gaussian initialization whose mean and SD are equal to the Billeh initialization.

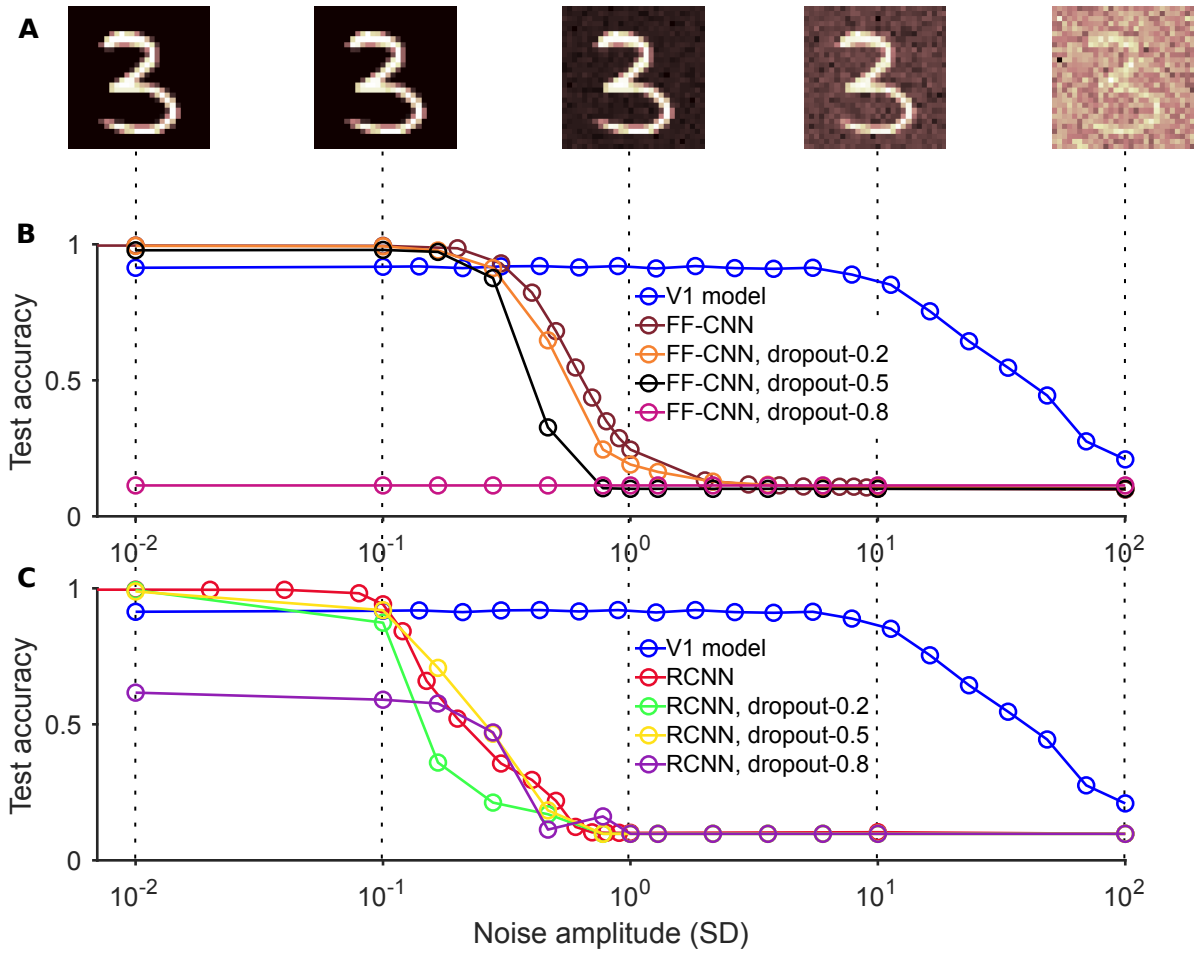


Figure S15: Robustness of CNNs trained with dropout to noise in test images. (A) Samples of handwritten MNIST digits with Gaussian noise drawn independently from $\mathcal{N}(0, \mathrm{SD})$ for each pixel, for different values of the SD. **(B)** To add internal noise during training of CNNs, we randomly set some of the elements of the input tensor with probability p using samples from a Bernoulli distribution. Each element was set as zero independently. We use dropout-p to represent this module in the legend. With dropout, the performance of feedforward (FF)-CNNs is still less robust than the V1 model when p is from 0.1 to 0.9. **(C)** Similar to **(B)** but for recurrent CNN (RCNN).

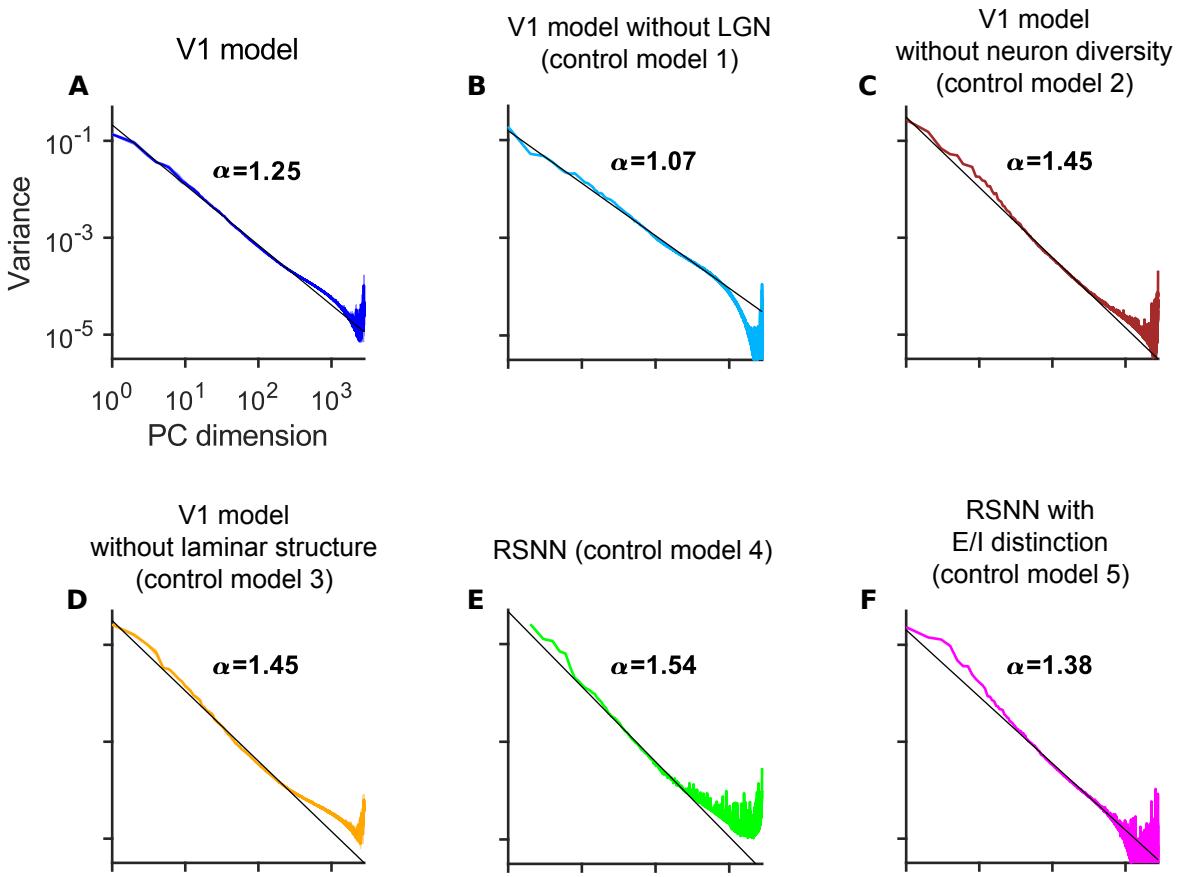


Figure S16: Comparing the eigenspectra of neural codes in the V1 model and control models. (A-F) Eigenspectra of V1 model and 5 control models. In control models except the one without LGN (control model 1), their power-law fittings are worse than that of V1 model. Either head or tail cannot be fitted, making the power-law span about one order of magnitude smaller than that of V1 model. Moreover, the fitted power-law exponent is more away from the experimental data $(1 + 2/d)$.

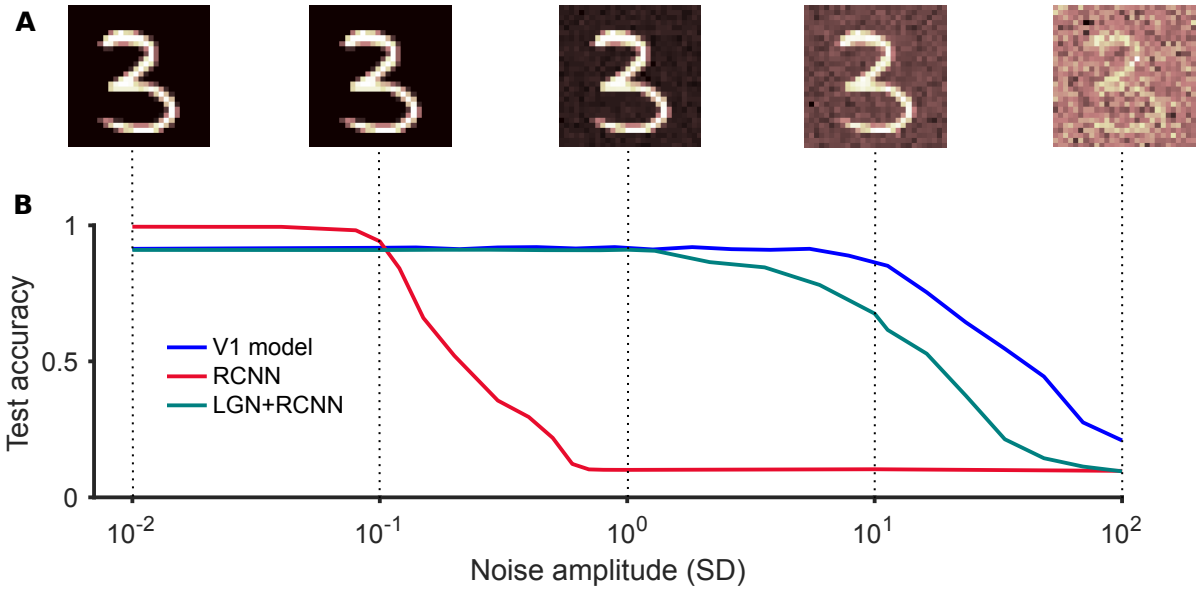


Figure S17: Robustness of CNNs with LGN to noise in images. (A) Samples of handwritten MNIST digits with Gaussian noise drawn independently from $\mathcal{N}(0, \mathrm{SD})$ for each pixel, for different values of the SD. **(B)** Adding the LGN model as preprocessor of the RCNN substantially improved their robustness, although it slightly reduces accurracy with little or no noise.

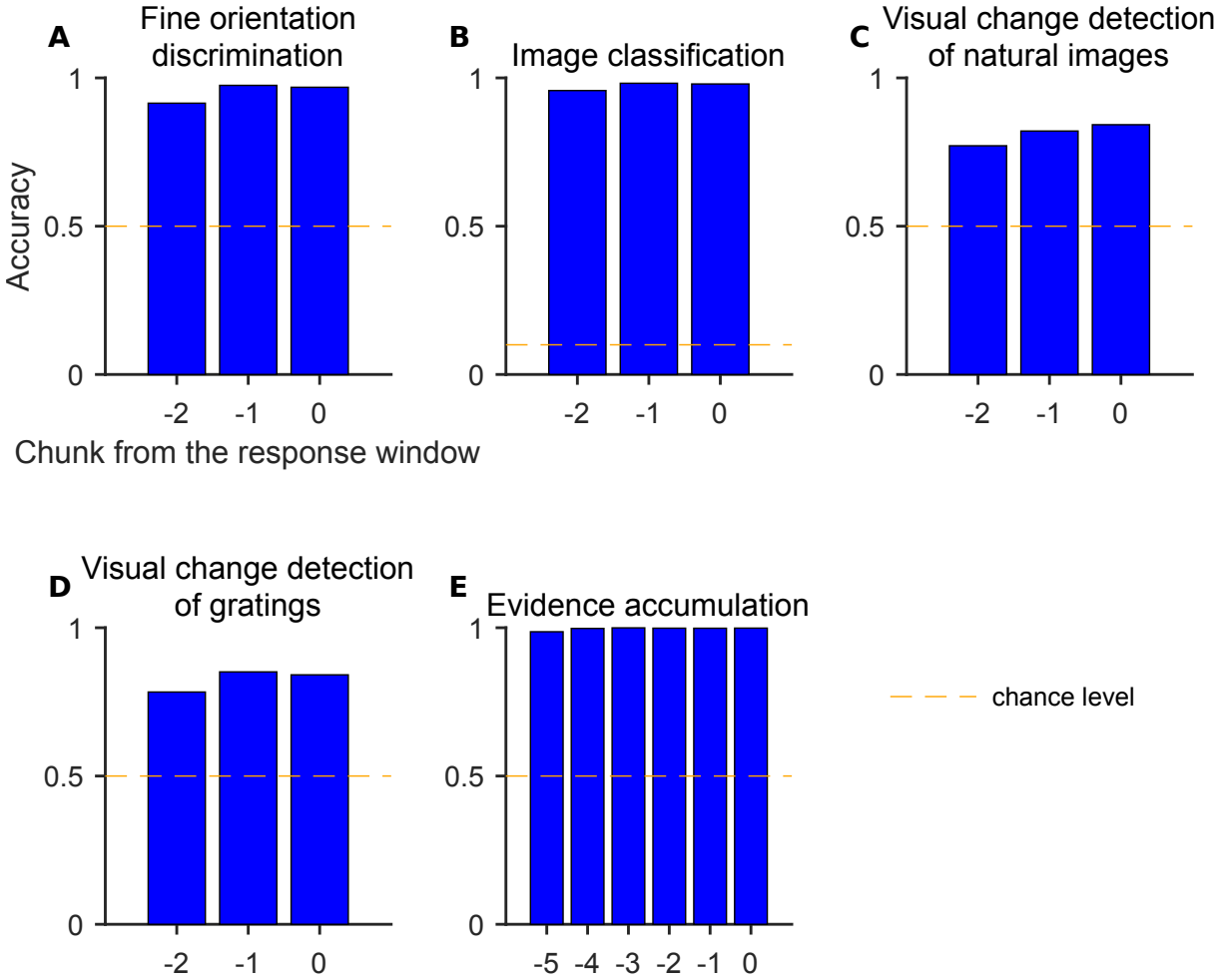


Figure S18: Excitatory neurons in L2/3 contain and transmit most of the information that is needed to solve the 5 computational tasks. To demonstrate that, we trained a multi-layer perceptron (which might be seen as proxy for higher cortical areas that are not represented in the model) that receives the spike counts during a time window of 50 ms from all excitatory neurons in L2/3 of the trained V1 model. We label the response window for each task as 0, and also show results for 50 ms time windows before that (counting backwards). The weights of the multi-layer perceptron were trained by Adam optimizer to produce the target output for each task. The multi-layer perceptron has 3 layers, with 3000 and 600 ReLu neurons on the first 2 layers, and the number of neurons on the last layer equal to the number of possible decisions for the 5 tasks. The number of training epochs is 6; the learning rate is 0.0001. **(A-E)** The readout network can transform the firing activity of L2/3E neurons into the target outputs for all 5 tasks with high accuracy. This proves that the firing activity of the L2/3E neurons contains most of the information needed for solving the 5 tasks.

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 according to (*4*) 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. They conjectured that this difference was caused by the limitations of downstream decoders. The V1 model suggests a further factor that is relevant. 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. These simultaneous recordings are conceptually related to the postulate of having a global readout that receives synaptic input from all 50,000 neurons, see Fig. S3A. However, we have demonstrated in the V1 model that a global linear readout attains for the fine orientation discrimination task an accuracy of 98.18% (the global linear readout layer is trained for 6 epochs with the frozen V1 model whose synaptic weights were trained in our default manner, simulating that the readout from the recorded neurons). On the other hand, a pool of 30 projection neurons on L5 could only achieve an accuracy of 93.17% if one assumed that they were localized closely together (Fig. S3B), and of 94.01% if they were assumed to be randomly distributed in L5 (Fig. S3C). These results suggest that how information from area V1 is extracted and projected to downstream areas is also 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.