

A Convolutional Network Architecture Driven by Mouse Neuroanatomical Data

Jianghong Shi¹, Michael A. Buice^{1,2,✉}, Eric Shea-Brown^{1,2}, Stefan Mihalas^{1,2}, and Bryan Tripp³

¹Applied Mathematics and Computational Neuroscience Center, University of Washington, Seattle, WA, United States

²Allen Institute, Seattle, WA, United States

³Centre for Theoretical Neuroscience, University of Waterloo, Waterloo, Ontario, Canada

Convolutional neural networks trained on object recognition derive some inspiration from the neuroscience of the visual system in primates, and have been used as models of the feedforward computation performed in the primate ventral stream. In contrast to the hierarchical organization of primates, the visual system of the mouse has flatter hierarchy. Since mice are capable of visually guided behavior, this raises questions about the role of architecture in neural computation. In this work, we introduce a framework for building a biologically constrained convolutional neural network model of lateral areas of the mouse visual cortex. The structural parameters of the network are derived from experimental measurements, specifically estimates of numbers of neurons in each area and cortical layer, the interareal connectome, and the statistics of connections between cortical layers. This network is constructed to support detailed task-optimized models of mouse visual cortex, with neural populations that can be compared to specific corresponding populations in the mouse brain. The code is freely available to support such research.

mouse | convolutional network | neuroanatomy

Correspondence: michaelbu@alleninstitute.org

Introduction

Convolutional neural networks (CNNs) trained on object recognition derive some inspiration from the neuroscience of the visual system in primates, and have been used as models of the feedforward computation performed in the primate ventral stream (1–3). Indeed, these networks show functional similarity to areas of the primate visual system (4, 5) and the task-training paradigm for artificial neural networks is known to correlate with more predictive models and more similar representations to neural activity in both visual and auditory areas (6–9). Despite these successes and the clear power of CNNs to solve machine learning problems in the visual domain (10, 11), they are not precise analogues for the biological circuit. Recent endeavors (12, 13) show that imposing brain like structure such as shallowness and recurrence can improve the functional similarity to the primates’ brain. The interplay of architecture and computation remains an open problem in both machine learning and neuroscience.

In contrast to the hierarchical organization of primates, the visual system of the mouse has a shallower organization in which multiple higher visual areas are at similar levels of the hierarchy (14). Since mice and other rodents are capable of visually guided behavior including invariant object recognition (15, 16), this raises questions about the role of architecture in neural computation.

In the mean time, large scale tract tracing data sets have revealed neuro-anatomical structure in unprecedented detail (17–19). Using such data sets, we construct detailed estimates for circuit properties of the mouse visual system, including the interareal connectome and the numbers of neurons per region and layer.

Convolutional networks share weights across the visual field, and thus form an approximation of the likely functional properties of the real neural system, an approximation likely partially imposed by a degree of translation invariance of stimuli in the natural world leading to equivariant representations in neural systems (1–3). This weight sharing makes them much easier to train, which is an important practical consideration for establishing representations without a lifetime of experience.

Python code for this project is available at https://github.com/mabuice/Mouse_CNN.

Methods

Construction of the MouseNet CNN

In this section, we introduce our framework for constructing the MouseNet CNN. Fig 1 shows an overview of this framework. The basic idea is to use available sources of anatomical data (e.g. tract tracing data, cell counts, and statistics of short-range connections) to constrain the CNN network structure and architectural hyperparameters. We discuss the details of each step below.

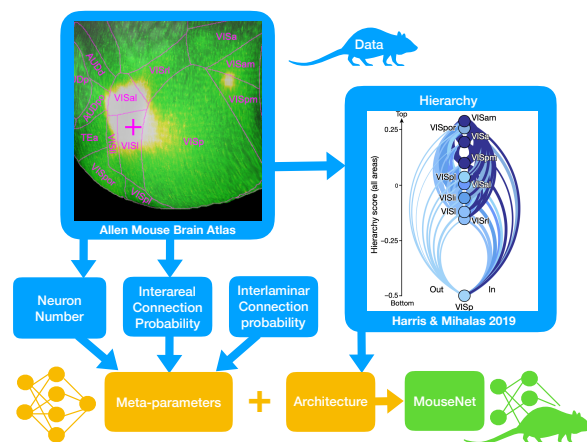


Fig. 1. Modeling framework. Framework for constructing biologically constrained MouseNet.

Network architecture. MouseNet spans the ventral lateral geniculate nucleus (LGNv) and six visual areas (Fig 2A). Input to the network passes first through LGNv, and then to the primary visual area VISp. After VISp, the architecture branches into five parallel pathways, representing five lateral visual areas: VISl (lateral visual area), VISal (anterolateral), VISpl (posterolateral), VISli (laterointermediate), and VISrl (rostromedial). Finally, the output of VISp together with all five lateral visual areas provide input to VISpor (postrhinal). We include only the lateral areas because they are more associated to object recognition while the medial areas are more involved in multimodal integration (20). The three-level architecture among the VIS areas was derived from an analysis of the hierarchy of mouse cortical and thalamic areas (Figure 6e in (14)), which considered feedforward and feedback connections from each area. In this analysis, VISp was clearly low in the hierarchy, and VISpor was clearly high, but the other lateral visual areas had similar intermediate positions. In the MouseNet model, each VIS area is represented by three separate cortical layers: layer 4 (L4), layer 2/3 (L2/3) and layer 5 (L5). We call a specific cortical layer within a specific area a “region”. Here we only consider the feedforward pathway, thought in primates to drive responses within ≈ 100 ms of stimulus presentation (2, 6). Following depictions of the canonical microcircuit (e.g. as summarized in Fig 5 in (21)), we consider the feedforward pathway to consist of laminar connections from L4 to L2/3, and from L2/3 to L5. Input from other areas feed into L4 and all of L4, L2/3 and L5 output to downstream areas, as shown in Fig 2B. This is consistent with broad connectivity among visual areas from each of these layers (Figure 2f of (14)). Fig 2C shows the MouseNet architecture in full detail, including all 22 regions and associated connections.

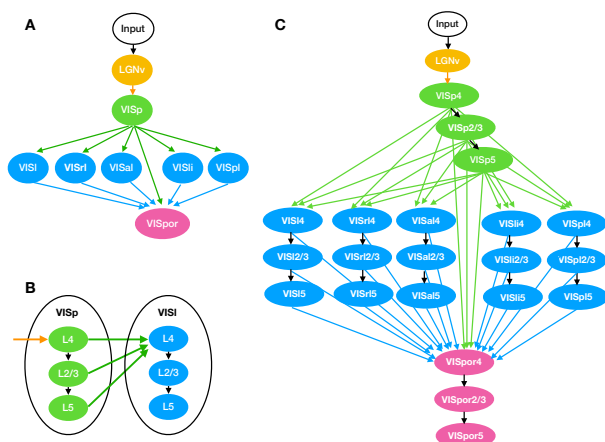


Fig. 2. Illustration of MouseNet architecture. Only feedforward connections are included. (A) High-level organization of MouseNet, based on analysis of the hierarchy of lateral visual areas ((14)). (B) Connection patterns at the level of cortical layers. (C) Full MouseNet architecture.

From architecture to convolutional neural net. Similar to the CALC model for the primate visual cortex by one of the authors (22), the general idea is to use convolution (Conv) operations to model the projections between different regions

in the mouse visual cortex. Conv operations are linear combinations of many inputs, so they impose the assumption of linear synaptic integration. They are widely used in machine learning, because they run efficiently on graphical processing units, and they share parameters across the visual field, leading to reduced memory requirements and faster learning, relative to general linear maps. Each connection from source brain region i to target brain region j is modelled with a Conv operation, Conv^{ij} . The input to Conv^{ij} corresponds to the neural activities in source region i , and the output of Conv^{ij} drives neural activities in the target region j . For example, as shown in Fig 3A, the projection from Region 1 to Region 2 (Proj 1 \rightarrow 2) is modeled by Conv^{12} . The neural activities in Region 1 correspond to the input to Conv^{12} , while the neural activities in Region 2 are a nonlinear function (ReLU, as shown in Fig 3C) of the output of Conv^{12} . In MouseNet, L4 of all areas except VISp receive multiple converging inputs, similar to Region 4 in the figure. In these cases, each projection is modeled by a separate Conv layer, and a nonlinear function (ReLU) is applied to the sum of the output from all these Conv layers, to produce the neural activities in Region 4.

Finding meta-parameters consistent with mouse data.

After fixing the architecture, we need to determine the meta-parameters for constructing the kernels for each Conv operation (Fig 3). The standard Conv operation is defined in terms of a four-dimensional kernel. The output of the kernel is a three-dimensional tensor of activations, A^j , which pass through element-wise ReLU nonlinearities to produce non-negative spike rates. Element $A_{\alpha\beta\gamma}^j$ is the activation of the neuron at the α^{th} vertical and β^{th} horizontal position in the visual field, in the γ^{th} channel (or feature map). The γ^{th} channel of the activation tensor for region j , A_{γ}^j , depends on inbound connections as,

$$A_{\gamma}^j = \sum_{i \in I^j} \sum_{\delta} C_{\gamma\delta}^{ij} * A_{\delta}^i, \quad (1)$$

where I^j is the set of regions that provide input to region j . Note that both $C_{\gamma\delta}^{ij}$ and A_{δ}^i are two-dimensional, and they undergo standard two-dimensional convolution. The meta-parameters of kernel C^{ij} are: number of input channels c_{in}^{ij} , number of output channels c_{out}^{ij} , stride s^{ij} , padding p^{ij} , and finally kernel size k^{ij} , i.e. the height and width (which are set equal) of $C_{\gamma\delta}^{ij}$. To make the connections realistically sparse, we add a binary Gaussian mask on the Conv filters, whose parameters are also estimated from data. See Fig 3B for an illustration of Conv operation with Gaussian mask. We constrain these meta-parameters with quantitative data whenever possible, and reasonable assumptions indicated by experimental observations otherwise.

Cortical population constraints. We set the horizontal and vertical resolution of the input (in pixels) based on mouse visual acuity. According to (23), the upper bound for visual acuity in mice is 0.5 cycles/degree, corresponding to a Nyquist sampling rate of 2 pixels/cycle x 0.5 cycles/degree =

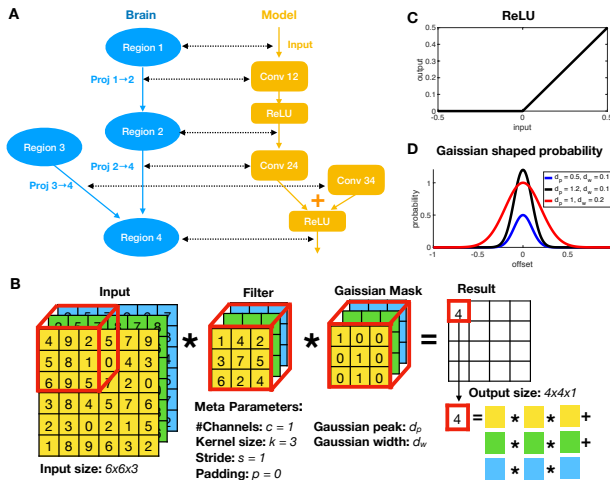


Fig. 3. From mouse brain to CNN model. (A) From mouse brain hierarchy to CNN architecture. (B) An example of Conv operation with Gaussian mask. (C) ReLU operation in the CNN architecture. (D) Gaussian shaped probability for Gaussian mask.

1.0 pixel/degree. The visual stimulus covers about 112 degrees of visual angle in the horizontal direction and 64 degrees in the vertical direction, but we simplified this to square input of size 64 by 64 pixels.

The resolution of the other regions depends on both the resolution of the input, and strides of the connections. The stride of a connection is the sampling period with respect to its input. For example, a Conv with a stride of one samples every element of its input, whereas a Conv with a stride of two samples every other element (both horizontally and vertically), leading to output of half the size in each dimension. Because cortical neurons are not organized into discrete channels in the same way as convolutional network layers, there is no strong anatomical constraint on the stride. However, the mean stride has to be somewhat less than two; there are ten steps in the longest path through MouseNet, but if only six of them had a stride of two, the 64x64 input would be reduced to 1x1 in VISpor, with no remaining topography. Lacking strong constraints, for simplicity, we first attempted to set all the strides to one, but this left very few channels in some of the smaller regions (due to an interaction between channels and strides that we describe below). We therefore set the strides of the connections outbound from VISp to two, and others to one. The feature maps of LGNv and VISp were therefore 64x64 (the same as the input), and all others were 32x32.

Given the resolutions of the channels in each region, the numbers of channels are constrained by the number of neurons. Specifically, Let n^i be the number of neurons in region i and (l_x^i, l_y^i) be the size of the output in area i , then the number of channels in area i is determined by $c^i = n^i / (l_x^i * l_y^i)$.

Estimating number of neurons in each area from data

We only model the excitatory neural population in our model, consistent with the fact that all neurons in the model project to other visual areas. In fact, neurons in convolutional networks are neither excitatory or inhibitory, but have both positive and negative output weights. However, past modelling work (24, 25) has shown that such mixed-weight projections

can be transformed so that the original neurons are all excitatory, and an additional population of inhibitory neurons recovers the functional effects of the original weights. We used estimates of the numbers of excitatory neurons in LGNv, VISp, VISal, VISam, VISl, VISpl, and VISpm from (26) (summarized in Table 1).

Table 1. Excitatory population number given by (26).

	L4	L2/3	L5	Total
LGNv				21313
VISp	93791	187111	166616	
VISal	23062	53964	54587	
VISam	18952	50749	43659	
VISl	9822	23396	19121	
VISpm	15557	54756	43216	
VISpl	8699	19385	18859	

For VISrl, VISli, and VISpor, we use the estimation of surface area from the voxel model (see section, *Estimating d_w^{ij} and d_p^{ij} for interareal connections*, for details) and estimate their numbers with respect to VISl, assuming these areas have similar neural density. The estimated surface areas with population numbers are given in Table 2.

Table 2. Excitatory population number estimated by voxel model with respect to VISl in (26).

	surface area	L4	L2/3	L5
VISl	0.92791	9822	23396	19121
VISrl	0.69805	7389	17600	14384
VISli	0.43561	4611	10983	8976
VISpor	1.39366	14752	35139	28719

Cortical connection constraints. Neurons tend to receive relatively dense inputs from other neurons that are above or below them, in other cortical layers, and the connection density decreases with increasing horizontal distance. Similarly, inputs from other cortical areas tend to have a point of highest density, with smoothly decreasing density around that point. We approximate such connection-density profiles with two-dimensional Gaussian functions. Specifically, we model the fan-in connection probability from source region i to target region j at position (x, y) (position offset from center in μm) as,

$$P^{ij}(x, y) = d_p^{ij} \exp\left(-\frac{x^2}{2(d_x^{ij})^2} - \frac{y^2}{2(d_y^{ij})^2}\right). \quad (2)$$

For simplicity, we assume $d_x^{ij} = d_y^{ij} \triangleq d_w^{ij}$ and let $r = \sqrt{x^2 + y^2}$ denote the offset from the center of the source layer, the above equation then simplifies to,

$$P^{ij}(r) = d_p^{ij} \exp\left(-\frac{r^2}{2(d_w^{ij})^2}\right), \quad (3)$$

where d_p^{ij} is the peak probability at the center and d_w^{ij} (μm) is the Gaussian width.

Both d_p^{ij} and d_w^{ij} are estimated from mouse data. The parameters for interlaminar connections are estimated from hit-rate

probabilities in functional connection studies, and the parameters for interareal connections are estimated from the mouse connectome (details below).

Conv layer with Gaussian mask. To translate our Gaussian models of connection density into network meta-parameters, we apply a binary mask to the weights of the Conv layers (Fig 3B). To do that, we first change the unit of d_w^{ij} in Eq.4 from micrometers to source area-dependent “pixels” (unit of output size of source area i) by multiplying it with $\sigma_i = \sqrt{(l_x^i * l_y^i)/a_i}$ (pixel/ μm). Here a_i denotes the surface size of area i , estimated from the voxel model (See section, *Estimating d_w^{ij} and d_p^{ij} for interareal connections*). We then have,

$$P^{ij}(\tilde{r}) = d_p^{ij} \exp\left(-\frac{\tilde{r}^2}{2(\tilde{d}_w^{ij})^2}\right), \quad \tilde{d}_w^{ij} = \sigma_i d_w^{ij}, \quad (4)$$

where both \tilde{r} and \tilde{d}_w^{ij} are in the “pixel” unit. The kernel size of the Conv layer is set to be $k^{ij} = 2\tilde{d}_w^{ij} + 1$, with padding calculated by $p^{ij} = (k^{ij} - s^{ij})/2$. During initialization, a mask containing zeros and ones is generated for each Conv layer, with size $(c_{out}^{ij}, c_{in}^{ij}, k^{ij}, k^{ij})$. The probability of each element being one is $P^{ij}(\tilde{r})$, where \tilde{r} (pixel) is the offset from the center of mask. The weights of the Conv layer are then multiplied by the mask. This gives the connections realistic densities (or sparsities), with realistic spatial profiles.

Estimating d_w^{ij}, d_p^{ij} for interlaminar connections. For the interlaminar connections, we estimate the Gaussian width d_w^{ij} from multiple experimental resources. Firstly, from Table 3 in (27), we get the estimation of d_w^{ij} to be 114 micrometers for functional connections between pairs of L4 pyramidal cells in mouse auditory cortex. Secondly, manually extracted from (28) Fig 8B, we obtain the variation of the Gaussian width depending on source and target layer from cat V1. Finally, we use this variation to scale the L4 to L4 width 114 μm to other layers in the mouse cortex. We summarize the Gaussian widths from cat cortex, along with corresponding scaled estimates for mouse cortex, in Table 3. Note that in the current model, we only use the values for connections from L4 to L2/3 and from L2/3 to L5 (Fig 2B).

Table 3. Estimated Gaussian width d_w^{ij} for interlaminar excitatory connections. The values outside of the parenthesis are extracted from (28); the values inside the parenthesis are scaled to mouse cortex, using the width 114 μm for L4-to-L4 connections in mouse auditory cortex (27). Units are micrometers (μm).

		Target		
		L2/3 (scaled)	L4 (scaled)	L5 (scaled)
Source	L2/3	225 (142.5)	50 (31.67)	100 (63.33)
	L4	220 (139.33)	180 (114)	140 (88.67)
	L5	150 (95)	100 (63.33)	210 (133)

To estimate the Gaussian peak probability d_p^{ij} , we first collect the connection probability between excitatory populations offset at 75 micrometer d_{75}^{ij} (Figure 4A in (29)). We then get

the peak probability d_p^{ij} by the relation

$$d_p^{ij} = d_{75}^{ij} / \exp\left(-\frac{75^2}{2(d_w^{ij})^2}\right) \quad (5)$$

We summarize the probability at 75 micrometers d_{75}^{ij} along with the peak probability d_p^{ij} in Table 4.

Table 4. The connection probability between excitatory populations offset at 75 micrometer d_{75}^{ij} . The numbers are from Fig 4A in (29). The calculated Gaussian peak probability d_p^{ij} are given in parenthesis.

		Target		
		L2/3 (peak)	L4 (peak)	L5 (peak)
Source	L2/3	0.160 (0.184)	0.016 (0.264)	0.083 (0.167)
	L4	0.140 (0.162)	0.243 (0.302)	0.104 (0.149)
	L5	0.021 (0.029)	0.007 (0.014)	0.116 (0.136)

Estimating d_w^{ij} and d_p^{ij} for interareal connections. To estimate interareal connection strengths and spatial profiles, we use the mesoscale model of the mouse connectome (19, 30). This model estimates connection strengths between 100 micrometer resolution voxels, based on 428 individual anterograde tracing experiments mapping fluorescent labeled neuronal projections in wild type C57BL/6J mice.

Flat map The voxel model is in 3 dimensional space. For the purpose of our analysis, we need to map the 3 dimensional structure into 2 dimensions. First, we fit the visual area positions by a sphere with center $c \in \mathcal{R}^3$ and radius r . Each position $x \in \mathcal{R}^3$ is then mapped to $\bar{x} \in \mathcal{R}^2$ with relation

$$\bar{x}_1 = v \cdot r \cdot \arctan\left(\frac{x_1 - c_1}{x_2 - c_2}\right), \quad (6)$$

$$\bar{x}_2 = v \cdot r \cdot \arctan\left(\frac{x_3 - c_3}{\sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2}}\right) \quad (7)$$

where $v = 100\mu m$ is the size of the voxel.

Area size Estimates of the size of each area/layer are needed to convert the widths of connection profiles from voxels in the mesoscale model to convolutional-layer pixels in MouseNet. For this purpose, each area’s size is approximated by the area of a convex hull of its mapped two-dimensional positions. These estimates are summarized in Table 5.

Estimating d_w^{ij} For each connection from source region i to target region j , we estimate d_w^{ij} from the mesoscale model. The first step is to estimate the widths of connections to individual voxels in j . The incoming width d_k^{ij} for target voxel k in j is estimated by the standard deviation of the connection strength about its center of mass. Specifically, $d_k^{ij} = (\sum_l w_{lk} d_l^2 / \sum_l w_{lk})^{1/2}$, where l indexes the voxels in source region j , w_{lk} is the connection weight between source and target voxels l and k in the mesoscale model, and d_l is the distance from voxel l to the center of mass of these connection weights. We then estimate d_w^{ij} as the average of these widths over the voxels in j . We omit from this average any target voxels that have multi-modal input profiles. This procedure provides an upper bound for d_w^{ij} , because a target voxel

Table 5. Area size (mm^2) estimated from the voxel model.

	VISp	VISal	VISl	VISli	VISpl	VISrl	VISpor
L4	4.3271	0.4909	0.8793	0.3355	0.2865	0.6182	0.5264
L2/3	4.7406	0.5477	0.9279	0.4356	0.6659	0.6980	1.3937
L5	4.2511	0.4972	0.8651	0.4039	0.6785	0.6748	1.2445

Table 6. Parameters from data or assumptions

Notation	CNN parameter	Biological source or assumptions
n^i	Number of neurons in area i	Based on (26) combined with the voxel model (19)
a_i	2 dimensional area size for area i	Estimated from voxel model data (19)
e	Total fan-in connections for all areas	Set to be 1000 based on (31)
(l_x^0, l_y^0)	input size	64x64 based on mouse visual acuity (23)
(l_x^i, l_y^i)	output size of area i	64x64 up to VISp, 32x32 after VISp (Assumption)
d_w^{ij}	Gaussian width (interlaminar)	From mouse (27) and cat (28) data
	Gaussian width (interareal)	Estimated from voxel model (19)
d_p^{ij}	Gaussian peak (interlaminar)	Based on hit rate experiments (29)
	Gaussian peak (interareal)	Estimated from voxel model (19)

may include multiple neurons with partially overlapping input areas.

Estimating d_p^{ij} The mesoscale model provides estimates of relative densities of connections between pairs of voxels. But an additional factor is needed to convert these relative densities into neuron-to-neuron connection probabilities. For this purpose, we assumed that each neuron received inputs from 1000 neurons in other areas (we call this number the extrinsic in-degree, e). This is on the order of the estimate from Fig S9 M in (31). Given this assumption, we calculated d_p^{ij} by the relation,

$$e \cdot \frac{w_{ij}}{\sum_i w_{ij}} = 2\pi(\tilde{d}_w^{ij})^2 \cdot d_p^{ij} \cdot c^i, \quad (8)$$

where w_{ij} is the connection strength from source area i to target area j , estimated from integrating the connection weights of the corresponding areas in the mesoscale model. The estimated values for d_w^{ij} and d_p^{ij} are given in Table 7.

Conv kernel size for LGNv. The above methods allowed us to set kernel sizes for intracortical connections, but not subcortical ones. We set the kernel sizes for inputs to LGNv and VISp L4 according to receptive field sizes in these regions. Receptive fields are about 9 degrees in LGNv and 11 degrees in VISp (32). As mentioned in Section, *Cortical population constraints*, mouse visual acuity is approximately 1 pixel/degree, therefore we set kernel size of the connection from input to LGNv to 9x9. We then set the kernel size of the connection from LGNv to VISp to 3x3, such that the receptive field size for VISp was 11x11 pixels.

Summary tables. In Table 8, we summarize the calculated number of channels in each area (in parenthesis) and the kernel size for each Conv layer.

The parameters used in the model based on biological sources and assumptions are summarized in Table 6 and the formulae for calculating the Conv layer meta-parameters are summarized in Table 9.

Table 7. The estimated d_w^{ij} (μm) and d_p^{ij} for interareal connections.

Source	Target	d_w^{ij} (μm)	d_p^{ij}
VISp4	VISal4	277.1	0.046
	VISl4	313.2	0.035
	VISli4	296.7	0.037
	VISpl4	290.6	0.038
	VISrl4	306.7	0.038
	VISpor4	276.8	0.016
VISp2/3	VISal4	266.1	0.058
	VISl4	325.8	0.036
	VISli4	303.3	0.042
	VISpl4	284.2	0.044
	VISrl4	339.4	0.030
	VISpor4	307.4	0.012
VISp5	VISal4	239.0	0.052
	VISl4	311.5	0.034
	VISli4	314.3	0.034
	VISpl4	278.3	0.043
	VISrl4	311.4	0.033
	VISpor4	298.3	0.013
Source	Target	d_w^{ij} (μm)	d_p^{ij}
VISal4		37.98	0.227
VISal2/3		13.81	1.258
VISal5		14.87	1.192
VISl4		204.7	0.022
VISl2/3		210.6	0.015
VISl5		215.4	0.021
VISli4		155.1	0.018
VISli2/3	VISpor4	169.3	0.020
VISli5		148.4	0.038
VISpl4		22.2	0.075
VISpl2/3		59.5	0.050
VISpl5		54.4	0.083
VISrl4		97.2	0.103
VISrl2/3		105.4	0.082
VISrl5		110.4	0.083

Table 8. The calculated meta-parameters for the Conv layers.

Source(#channel)	Target	kernel size
input(3)	LGNv	9×9
LGNv(5)	VISp4	3×3
VISp4(22)	VISp2/3	9×9
	VISal4	17×17
	VISI4	19×19
	VISli4	19×19
	VISpl4	19×19
	VISr4	19×19
VISp2/3(45)	VISp4	17×17
	VISp5	3×3
	VISal4	15×15
	VISI4	19×19
	VISli4	17×17
	VISpl4	17×17
VISp5(40)	VISr4	21×21
	VISpor4	19×19
	VISal4	15×15
	VISI4	19×19
	VISli4	19×19
VISpor4(14)	VISpl4	17×17
	VISr4	19×19
VISpor2/3(32)	VISpor4	19×19
	VISpor2/3	13×13
VISpor2/3(32)	VISpor5(28)	3×3
	VISpor2/3	13×13
VISal4(22)	VISal2/3	13×13
	VISpor4	3×3
VISal2/3(52)	VISal5	5×5
	VISpor4	1×1
VISal5(53)	VISpor4	1×1
VISI4(9)	VISI2/3	9×9
	VISpor4	15×15
VISI2/3(22)	VISI5	5×5
	VISpor4	15×15
VISI5(18)	VISpor4	15×15
VISli4(4)	VISli2/3	17×17
	VISpor4	19×19
VISli2/3(10)	VISli5	7×7
	VISpor4	17×17
VISli5(8)	VISpor4	15×15
VISpl4(8)	VISpl2/3	17×17
	VISpor4	3×3
VISp2/3(18)	VISp5	5×5
	VISpor4	5×5
VISp5(18)	VISpor4	5×5
VISr4(7)	VISr2/3	11×11
	VISpor4	5×5 (check)
VISr2/3(17)	VISr5	5×5
	VISpor4	9×9
VISr5(14)	VISpor4	9×9

Implementation of MouseNet. We implemented the MouseNet structure in PyTorch. Each Conv layer was followed by a batch normalization layer and a ReLU non-

linearity. For regions such as VISpor L4, which receive input from multiple Conv layers, the outputs of the Conv layers are summed before being fed into a batch normalization layer and a ReLU non-linearity.

We have trained the MouseNet model on an image classification task, by adding a simple classifier. Specifically, the L5 output from all areas was reduced to 4x4 by an average pooling layer, then flattened, concatenated, and fed to a linear fully-connected layer which reduced the dimension to the number of classes of the task. The outputs were then transformed to probabilities by the softmax function and the cross-entropy loss was used to train on the image classification task. Analysis of the results is left for future work.

Conclusions

Task-optimized deep networks show promise for brain modelling, because they are functionally sophisticated, and they often develop internal representations that overlap strongly with representations in the brain. Deep network architectures are loosely inspired by the brain, but there has been an extensive empirical exploration of the effects of architectural features in machine learning, independent of neuroscience. In parallel, a great deal has been learned about brain architectures, and mouse brain architecture has been particularly well characterized.

We have developed a deep network architecture that is consistent with a wide range of data on mouse visual cortex, including data from tract-tracing studies and studies of local connection statistics. While standard deep networks have provided useful points of comparison with neurobiological systems, more biologically realistic deep networks may enable more specific comparisons with the brain, such as comparisons between homologous groups of neurons, or modelling of specific lesions. By making our code publicly available, we hope to facilitate such work.

ACKNOWLEDGEMENTS

We thank Tianqi Chen, Blake Richards, Shahab Bakhtiari, Graham Taylor for helpful discussions and suggestions on the manuscript. We thank Saskia de Vries, Hannah Choi, Kameron Decker Harris, Timothy Lillicrap, Julie Harris, and Severine Durand for helpful discussions. We thank the Allen Institute for Brain Science founder, Paul G. Allen, for his vision, encouragement, and support. We acknowledge the NIH Graduate training grant in neural computation and engineering (R90DA033461).

Bibliography

1. Kunihiko Fukushima. Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural Networks*, 1(2):119 – 130, 1988. ISSN 0893-6080. doi: [https://doi.org/10.1016/0893-6080\(88\)90014-7](https://doi.org/10.1016/0893-6080(88)90014-7).
2. Maximilian Riesenhuber and Tomaso Poggio. Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11):1019–1025, Nov 1999. ISSN 1546-1726. doi: [10.1038/14819](https://doi.org/10.1038/14819).
3. James J. DiCarlo, Davide Zoccolan, and Nicole C. Rust. How does the brain solve visual object recognition? *Neuron*, 73(3):415 – 434, 2012. ISSN 0896-6273. doi: <https://doi.org/10.1016/j.neuron.2012.01.010>.
4. Umut Güçlü and Marcel A. J. van Gerven. Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience*, 35(27):10005–10014, 2015. ISSN 0270-6474. doi: [10.1523/JNEUROSCI.5023-14.2015](https://doi.org/10.1523/JNEUROSCI.5023-14.2015).
5. Nikolaus Kriegeskorte. Deep neural networks: A new framework for modeling biological vision and brain information processing. *Annual Review of Vision Science*, 1(1):417–446, 2015. doi: [10.1146/annurev-vision-082114-035447](https://doi.org/10.1146/annurev-vision-082114-035447). PMID: 28532370.
6. Daniel L. K. Yamins, Ha Hong, Charles F. Cadieu, Ethan A. Solomon, Darren Seibert, and James J. DiCarlo. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, 111(23):8619–8624, 2014. ISSN 0027-8424. doi: [10.1073/pnas.1403112111](https://doi.org/10.1073/pnas.1403112111).

Table 9. Meta-parameters for Conv layer connecting source area i to target area j

Notation	CNN parameter	Formula
c^i	number of channels in area i	$c^i = n^i / (l_x^i \cdot l_y^i)$
k^{ij}	kernel size	$k^{ij} = 2 * \tilde{d}_w^{ij} + 1$
s^{ij}	stride	$s^{ij} = l_x^i / l_x^j = l_y^i / l_y^j$
p^{ij}	padding	$p^{ij} = (k^{ij} - s^{ij}) / 2$
\tilde{d}_w^{ij}	Gaussian width	$\tilde{d}_w^{ij} = \sqrt{(l_x^i \cdot l_y^i) / a_i} \cdot \tilde{d}_w^i$
d_p^{ij}	Gaussian peak	d_p^{ij}

7. Daniel L. K. Yamins and James J. DiCarlo. Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, 19:356, Feb 2016.
8. Alexander J.E. Kell, Daniel L.K. Yamins, Erica N. Shook, Sam V. Norman-Haignere, and Josh H. McDermott. A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy. *Neuron*, 98(3):630–644.e16, 2018. ISSN 0896-6273. doi: <https://doi.org/10.1016/j.neuron.2018.03.044>.
9. Seyed-Mahdi Khaligh-Razavi and Nikolaus Kriegeskorte. Deep supervised, but not unsupervised, models may explain it cortical representation. *PLOS Computational Biology*, 10(11):1–29, 11 2014. doi: [10.1371/journal.pcbi.1003915](https://doi.org/10.1371/journal.pcbi.1003915).
10. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In F Pereira, C J C Burges, L Bottou, and K Q Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.
11. Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.
12. Jonas Kubilius, Martin Schrimpf, Aran Nayebi, Daniel Bear, Daniel L. K. Yamins, and James J. DiCarlo. Cornet: Modeling the neural mechanisms of core object recognition. *bioRxiv*, 2018. doi: [10.1101/408385](https://doi.org/10.1101/408385).
13. Jonas Kubilius, Martin Schrimpf, Kohitij Kar, Rishi Rajalingham, Ha Hong, Najib Majaj, Elias Issa, Pouya Bashivan, Jonathan Prescott-Roy, Kailyn Schmidt, Aran Nayebi, Daniel Bear, Daniel L Yamins, and James J DiCarlo. Brain-like object recognition with high-performing shallow recurrent ans. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 12805–12816. Curran Associates, Inc., 2019.
14. Julie A Harris, Stefan Mihalas, Karla E Hirokawa, Jennifer D Whitesell, Hannah Choi, Amy Bernard, Phillip Bohn, Shiella Caldejon, Linzy Casal, Andrew Cho, Aaron Feiner, David Feng, Nathalie Gaudreault, Charles R Gerfen, Nile Graddis, Peter A Groblewski, Alex M Henry, Anh Ho, Robert Howard, Joseph E Knox, Leonard Kuan, Xiuli Kuang, Jerome Lecoq, Phil Lesnar, Yaoyao Li, Jennifer Luviano, Stephen McConoughey, Marty T Mortrud, Mitham Naemni, Lydia Ng, Seung Wook Oh, Benjamin Ouellette, Elise Shen, Staci A Sorensen, Wayne Wakeman, Quanxin Wang, Yun Wang, Ali Williford, John W Phillips, Allan R Jones, Christof Koch, and Hongkui Zeng. Hierarchical organization of cortical and thalamic connectivity. *Nature*, 575(7781):195–202, 2019. ISSN 1476-4687. doi: [10.1038/s41586-019-1716-z](https://doi.org/10.1038/s41586-019-1716-z).
15. Davide Zoccolan, Nadja Oertelt, James J. DiCarlo, and David D. Cox. A rodent model for the study of invariant visual object recognition. *Proceedings of the National Academy of Sciences*, 106(21):8748–8753, 2009. ISSN 0027-8424. doi: [10.1073/pnas.0811583106](https://doi.org/10.1073/pnas.0811583106).
16. Andrew D. Huberman and Cristopher M. Niell. What can mice tell us about how vision works? *Trends in Neurosciences*, 34(9):464 – 473, 2011. ISSN 0166-2236. doi: <https://doi.org/10.1016/j.tins.2011.07.002>.
17. Rembrandt Bakker, Thomas Wachtler, and Markus Diesmann. Cocomac 2.0 and the future of tract-tracing databases. *Frontiers in Neuroinformatics*, 6:30, 2012. ISSN 1662-5196. doi: [10.3389/fninf.2012.00030](https://doi.org/10.3389/fninf.2012.00030).
18. Seung Wook Oh, Julie A. Harris, Lydia Ng, Brent Winslow, Nicholas Cain, Stefan Mihalas, Quanxin Wang, Chris Lau, Leonard Kuan, Alex M. Henry, Marty T. Mortrud, Benjamin Ouellette, Thuc Nghi Nguyen, Staci A. Sorensen, Clifford R. Slaughterbeck, Wayne Wakeman, Yang Li, David Feng, Anh Ho, Eric Nicholas, Karla E. Hirokawa, Phillip Bohn, Kevin M. Joines, Hanchuan Peng, Michael J. Hawrylycz, John W. Phillips, John G. Hohmann, Paul Wohnoutka, Charles R. Gerfen, Christof Koch, Amy Bernard, Chinh Dang, Allan R. Jones, and Hongkui Zeng. A mesoscale connectome of the mouse brain. *Nature*, 508(7495): 207–214, Apr 2014. doi: [10.1038/nature13186](https://doi.org/10.1038/nature13186).
19. Joseph E Knox, Kameron Decker Harris, Nile Graddis, Jennifer D Whitesell, Hongkui Zeng, Julie A Harris, Eric Shea-Brown, and Stefan Mihalas. High-resolution data-driven model of the mouse connectome. *Network Neuroscience*, 3(1):217–236, 2019. doi: [10.1162/netn_a_00066](https://doi.org/10.1162/netn_a_00066).
20. Lindsey L. Glickfeld and Shawn R. Olsen. Higher-order areas of the mouse visual cortex. *Annual Review of Vision Science*, 3(1):251–273, 2017. doi: [10.1146/annurev-vision-102016-061331](https://doi.org/10.1146/annurev-vision-102016-061331). PMID: 28746815.
21. Nuno Miguel Amorim Da Costa and Kevan Martin. Whose cortical column would that be? *Frontiers in Neuroanatomy*, 4:16, 2010. ISSN 1662-5129. doi: [10.3389/fnana.2010.00016](https://doi.org/10.3389/fnana.2010.00016).
22. Bryan Tripp. Approximating the architecture of visual cortex in a convolutional network. *Neural Computation*, 31(8):1551–1591, 2019. doi: [10.1162/neco_a_01211](https://doi.org/10.1162/neco_a_01211). PMID: 31260392.
23. Glen T Prusky, Paul W.R West, and Robert M Douglas. Behavioral assessment of visual acuity in mice and rats. *Vision Research*, 40(16):2201 – 2209, 2000. ISSN 0042-6989. doi: [https://doi.org/10.1016/S0042-6989\(00\)00081-X](https://doi.org/10.1016/S0042-6989(00)00081-X).
24. Christopher Parisien, Charles H Anderson, and Chris Eliasmith. Solving the problem of negative synaptic weights in cortical models. *Neural computation*, 20(6):1473–1494, 2008.
25. Bryan Tripp and Chris Eliasmith. Function approximation in inhibitory networks. *Neural Networks*, 77:95–106, 2016.
26. Csaba Erő, Marc-Oliver Gewaltig, Daniel Keller, and Henry Markram. A cell atlas for the mouse brain. *Frontiers in Neuroinformatics*, 12:84, 2018. ISSN 1662-5196. doi: [10.3389/fninf.2018.00084](https://doi.org/10.3389/fninf.2018.00084).
27. Robert B. Levy and Alex D. Reyes. Spatial profile of excitatory and inhibitory synaptic connectivity in mouse primary auditory cortex. *Journal of Neuroscience*, 32(16):5609–5619, 2012. ISSN 0270-6474. doi: [10.1523/JNEUROSCI.5158-11.2012](https://doi.org/10.1523/JNEUROSCI.5158-11.2012).
28. Armen Stepanyants, Judith A. Hirsch, Luis M. Martinez, Zoltán F. Kisvárdy, Alex S. Ferencsik, and Dmitri B. Chklovskii. Local Potential Connectivity in Cat Primary Visual Cortex. *Cerebral Cortex*, 18(1):13–28, 04 2007. ISSN 1047-3211. doi: [10.1093/cercor/bhm027](https://doi.org/10.1093/cercor/bhm027).
29. Yazan N. Billeh, Binghuang Cai, Sergey L. Graty, Kael Dai, Ramakrishnan Iyer, Nathan W. Gouwens, Reza Abbasi-Asl, Xiaoxuan Xia, Joshua H. Siegle, Shawn R. Olsen, Christof Koch, Stefan Mihalas, and Anton Arkhipov. Systematic integration of structural and functional data into multi-scale models of mouse primary visual cortex. *Neuron*, 106(3):388 – 403.e18, 2020. ISSN 0896-6273. doi: <https://doi.org/10.1016/j.neuron.2020.01.040>.
30. Kameron D Harris, Stefan Mihalas, and Eric Shea-Brown. High resolution neural connectivity from incomplete tracing data using nonnegative spline regression. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 3099–3107. Curran Associates, Inc., 2016.
31. Henry Markram, Eilif Muller, Srikanth Ramaswamy, Michael W. Reimann, Marwan Abdellah, Carlos Aguado Sanchez, Anastasia Ailamaki, Lidia Alonso-Nanclares, Nicolas Antille, Selim Arsever, Guy Antoine Atenekeng Kahou, Thomas K. Berger, Ahmet Bilgili, Nenad Buncic, Athanasia Chalimourda, Giuseppe Chindemi, Jean-Denis Courcol, Fabien Delalandre, Vincent Delattre, Shaul Druckmann, Raphael Dumusc, James Dynes, Stefan Eilemann, Eyal Gal, Michael Emiel Gevaert, Jean-Pierre Ghobril, Albert Gidon, Joe W. Graham, Anirudh Gupta, Valentin Haenel, Etay Hay, Thomas Heinis, Juan B. Hernandez, Michael Hines, Lida Kanari, Daniel Keller, John Kenyon, Georges Khazen, Yihwa Kim, James G. King, Zoltan Kisvárdy, Pramod Kumbhar, Sébastien Lasserre, Jean-Vincent Le Bé, Bruno R.C. Magalhães, Angel Merchán-Pérez, Julie Meystre, Benjamin Roy Morrice, Jeffrey Muller, Alberto Muñoz-Céspedes, Shruti Muralidhar, Keerthan Muthurasa, Daniel Nachbaur, Taylor H. Newton, Max Nolte, Aleksandr Ovcharenko, Juan Palacios, Luis Pastor, Rodrigo Perin, Rajnish Ranjan, Imad Riachi, José-Rodrigo Rodríguez, Juan Luis Riquelme, Christian Rössert, Konstantinos Stryrakis, Ying Shi, Julian C. Shillcock, Gilad Silberberg, Ricardo Silva, Farhan Tauheed, Martin Telefont, Maria Toledo-Rodriguez, Thomas Tränkle, Werner Van Geit, Jafet Villafranca Díaz, Richard Walker, Yun Wang, Stefano M. Zaninetta, Javier DeFelipe, Sean L. Hill, Idan Segev, and Felix Schürmann. Reconstruction and simulation of neocortical microcircuitry. *Cell*, 163(2):456 – 492, 2015. ISSN 0092-8674. doi: <https://doi.org/10.1016/j.cell.2015.09.029>.
32. Séverine Durand, Ramakrishnan Iyer, Kenji Mizuseki, Saskia de Vries, Stefan Mihalas, and R. Clay Reid. A comparison of visual response properties in the lateral geniculate nucleus and primary visual cortex of awake and anesthetized mice. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 36(48):12144–12156, Nov 2016. ISSN 1529-2401. doi: [10.1523/JNEUROSCI.1741-16.2016](https://doi.org/10.1523/JNEUROSCI.1741-16.2016).