Connectome-constrained deep mechanistic networks predict neural responses across the fly visual system at single-neuron resolution

Janne K. Lappalainen,[1,2] Fabian D. Tschopp,[2] Sridhama Prakhya,[2] Mason McGill,[2,3]
Aljoscha Nern,[2] Kazunori Shinomiya,[2] Shin-ya Takemura,[2] Eyal Gruntman,[2]
Jakob H. Macke,[1,4] Srinivas C. Turaga[2*]
[1]Machine Learning in Science, Tübingen University and Tübingen AI Center, Germany
[2]HHMI Janelia Research Campus, Ashburn, VA, USA
[3]Computation and Neural Systems, California Institute of Technology, Pasadena, CA, USA
[4]Max Planck Institute for Intelligent Systems, Tübingen, Germany
*turagas@janelia.hhmi.org.

Abstract

We can now measure the connectivity of every neuron in a neural circuit, but we are still blind to other biological details, including the dynamical characteristics of each neuron. The degree to which connectivity measurements alone can inform understanding of neural computation is an open question. Here we show that with only measurements of the connectivity of a biological neural network, we can predict the neural activity underlying neural computation. We constructed a model neural network with the experimentally determined connectivity for 64 cell types in the motion pathways of the fruit fly optic lobe but with unknown parameters for the single neuron and single synapse properties. We then optimized the values of these unknown parameters using techniques from deep learning, to allow the model network to detect visual motion. Our mechanistic model makes detailed experimentally testable predictions for each neuron in the connectome. We found that model predictions agreed with experimental measurements of neural activity across 24 studies. Our work demonstrates a strategy for generating detailed hypotheses about the mechanisms of neural circuit function from connectivity measurements. We show that this strategy is more likely to be successful when neurons are sparsely connected—a universally observed feature of biological neural networks across species and brain regions.

Introduction

Electrical signals propagating through networks of neurons in the nervous system form the basis of computations such as the visual detection of movement. The propagation of neural activity is shaped by both the functional properties of individual neurons and their synaptic connectivity. In addition, multiple additional factors[1,2] including electrical synapses[3,4], neuromodulation[5], and glia[6] are known to further influence neural activity on multiple time-scales. Volume electron microscopy can now be used to comprehensively measure the connectivity of every neuron in a neural circuit, and even entire nervous systems[7–24]. However, we do not yet have the means to also comprehensively measure all other biological details, including the dynamical properties of every neuron and synapse in the same circuit[2]. For these reasons, there has been considerable debate about the utility of connectomic measurements for understanding brain function[25]. Is it possible to use measurements of only neural connectivity to generate accurate predictions about how the neural circuit functions? Even in the complete absence of direct measurements of neural activity from a living brain?

There is considerable evidence from computer science and neuroscience that there is not usually a strong link between the connectivity of a neural network and its computational function. Universal function approximation theorems for artificial neural networks[26–28] imply that the same computational task can be performed by many different networks with very different neural connectivity. Empirically, there exist many classes of general purpose artificial neural network architectures which can trained to perform the same large diversity of computational tasks[29]. Such differences in connectivity can correspond to qualitatively different computational mechanisms[30,31]. Similarly in neuroscience, there have been competing proposals by theorists for example, for the neural circuit mechanisms of the computation of visual motion[32,33], and for the integration of eye velocity commands into eye position signals[34,35], with each proposal suggesting different neural connectivity. Further, circuits with the same connectivity can function differently[5]. Thus in general, neither the connectivity of a circuit alone, nor its computational task alone, can uniquely determine the mechanism of circuit function[36–38].

Here we show that the connectivity of a neural circuit, together with knowledge of its computational task, enables accurate predictions of the role played by individual neurons in the circuit in the computational task. We constructed a differentiable[39,40] model neural network with a close correspondence to the brain, whose connectivity was given by connectomic measurements and with unknown single neuron and single synapse parameters. We optimized the unknown parameters of the model network using techniques from deep learning[41–43], to enable the model network to accomplish the computational task[44–47]. We call such models connectome-constrained and task-optimized deep mechanistic networks (DMNs; Fig. 1a).

We applied this approach to model the motion pathways in the optic lobe of the *Drosophila* visual system. We constructed a DMN with experimentally measured connectivity[48–50], and unknown parameters for the single neuron dynamics and the strength of a unitary synapse. We optimized the unknown model parameters on the computer vision task of computing visual motion from dynamic visual stimuli. Visual motion computation in the fly visual system and its mechanistic underpinnings have been extensively studied[51–56]. Thus, we were able to compare the detailed predictions of our model with experimental measurements of neural activity in response to visual stimuli, on a neuron-by-neuron basis. We found that our connectome-constrained and task-optimized DMN accurately predicts the well-known segregation of the visual system into light increment (ON) and light decrement (OFF) channels[46,57–61], as well as the direction selectivity of the well-known T4 and T5 motion detector neurons[53,62–64]. Our model further suggests that TmY3 might also detect motion, a prediction that has yet to be experimentally tested. We release our model as a resource for the community.[1]

Results

A connectome-constrained deep mechanistic network of the fruit fly visual system

The optic lobes of the fruit fly are involved in early visual processing. They comprise several layered neuropiles whose columnar arrangement has a one-to-one correspondence with the ommatidia, both possessing a remarkably crystalline organization in a hexagonal lattice. Visual input from the photoreceptors is received by the lamina, which sends projections to the medulla, lobula, and lobula plate (Fig. 1b[65]). Many components of the optic lobe are highly periodic, with columnar cell types appearing once per column, and multi-columnar neurons appearing with only small deviations from a well-defined periodicity in columnar space[65,66]. A complete reconstruction of the optic lobe is yet unavailable, but several studies have reported on the local connectivity within the lamina[50], primarily focused on the light increment (ON) and light decrement (OFF) selective components of the motion pathways of visual processing in the medulla[67,68], and in the medulla, lobula, and lobula plate[48,49,69]. We assembled these separate local reconstructions into a coherent local connectome spanning the retina, lamina, medulla, lobula, and lobula plate (Fig. 1c, SI Fig.1). We approximated the circuitry across the entire visual field as perfectly periodic[66,68], and tiled this local connectivity architecture in hexagonal lattice across retinotopic space to construct a connectome for 64 cell types across the central visual field of the right eye (Fig. 1d; Methods).

---

[1] `https://github.com/TuragaLab/flyvis`

We built a recurrent neural network modeling these first stages of visual processing in the optic lobe based on the connectome for the right eye. Each neuron in this DMN corresponds to a real neuron in the fly visual system, belonging to an identified cell type, and is connected to other neurons only if they are connected by synapses in the connectome (Fig. 1e). Our goal was to investigate whether precise synaptic connectivity and task-constraints are sufficient to account for neural tuning across the fly visual system. We therefore constructed a model with detailed connectivity, but simplified models of single neurons and chemical synapses (Fig. 1f). As many neurons in the early visual system are non-spiking, we used passive leaky linear non-spiking voltage dynamics to model the time-varying activity of single neurons. We modeled neurons as point-neurons with a single electrical compartment, as this has been previously shown to be a good approximation given the small size of many neurons in the optic lobe[54]. As the exception, we modeled the CT1 neuron with multiple compartments since it is an exceptionally large single neuron spanning the entire optic lobe and is highly electrotonically compartmentalized[70]. We effectively modeled the neuron as two columnar "cell types", with one CT1 compartment per column in the 10th layer of the medulla CT1(M10) and one per column in the 1st layer of the lobula CT1(Lo1) (SI Supplementary Note 2). We coupled neurons with chemical synapses whose connectivity was determined by the connectome. We developed a simplified model for a graded release chemical synapse between non-spiking neurons: A threshold-linear nonlinear function models the nonlinearity of the time-averaged concentration of synaptic release as a function of presynaptic voltage. The resulting network model follows well-known threshold-linear dynamics and is differentiable. Such dynamics are typically used to approximate the firing rates of a network of spiking neurons[71,72], whereas in our network, the threshold nonlinearity results from the nonlinear voltage-gated release of neuro-transmitter.

We used the cell type structure of the connectome, and assumption of perfect translation invariance across retinotopic space to reduce the number of free parameters (Fig. 1f). We assumed that neurons of the same cell type shared the same neuron time constant and resting membrane potential. We modeled a synaptic weight as proportional to the number of discrete synapses measured experimentally between those two neurons[73], with a scale factor representing the strength of a unitary synapse. The unitary synapse scaling and the sign of each synapse was the same across all pairs of neurons with the same pre- and postsynaptic cell type. Likewise, the synapse count between each pair of neurons was the same across all pairs of neurons with the same pre- and postsynaptic cell type, and their relative location in retinotopic space.

In total, the connectome-constrained model comprises 45,669 neurons and 1,513,231 connections, across 65 cell types arranged in a hexagonal lattice consisting of 721 columns, modeling the central visual field of the roughly 700-900 ommatidia typically found in the fruit fly retina[74,75]. Connectomic constraints, and our assumption of spatial homogeneity (i.e., the hexagonally convolutional structure of the network) result in a dramatic reduction to just 734 free parameters for this large network model. The only free parameters in our model are the single neuron time constants and resting membrane potentials (65 parameters each), and the unitary synapse strengths (604 parameters). In the absence of connectomic measurements, if we had to infer the type-to-type as well as spatial connectivity between all pairs of cell types, we would instead have needed to estimate well over a million parameters (Methods).

We further constrained the parameters of the model by task-optimization, i.e. by training the model to perform a computational task which is thought to approximate the computations carried out by the system[44,76]. We therefore implemented our recurrent DMN using the PyTorch library[39,40] (Methods) and used automatic differentiation to optimize the model using gradient based deep learning training methods[41–43].

As the functional task, we chose the computation of visual motion from naturalistic visual stimuli[77]. Visual motion computation in the fly visual system and its mechanistic underpinnings have been extensively studied[51–56]. This challenging computation requires the neural circuit to compare visual stimuli across space and time, and thereby critically relies on temporal integration of visual information by the dynamics of the network. We hypothesized that training our network model to perform the computer vision task of optic flow computation[77] could help us identify circuit elements involved in motion computation. Since our model contains many of the circuit elements which have been experimentally characterized and implicated in the computation of visual motion, we could then validate our model predictions.

To decode optic flow from the DMN, we used a decoding network which maps its output to the computer vision representation of optic flow. This two-layer convolutional decoding network is allowed to use only the instantaneous neural activity of the medulla and downstream areas as input. Importantly, the decoding network cannot by itself detect motion, which requires the comparison of current and past visual stimuli, but must instead rely on the temporal dynamics of the connectome-constrained network to compute motion-selective visual features. The resulting combination of our recurrent connectome-constrained DMN model and the feedforward decoding network was then trained end-to-end: We rendered video sequences from the Sintel database[77] as direct input to the photoreceptors of the connectome-constrained model (ignoring neuronal superposition[78]), and used gradient descent (backpropgation through time[79]) to minimize the task error in predicting optic flow (Fig. 1g, Methods).

Ensembles of connectome-constrained and task-optimized DMNs robustly predict known tuning properties

We used only connectome and task-constraints to construct our DMN, without any measurements of neural activity. We can therefore validate the model by computing predictions of neural selectivity for each of the 64 identified cell types in the model, and comparing them to experimental measurements. Since it is possible that connectome and task-constraints might not uniquely constrain all model parameters[80–82], we generated an ensemble of 50 models, all constrained with the same connectome, and optimized to perform the same task. Each model in the ensemble corresponds to a local optimum of task performance. Since the models achieved similar (but not identical) task performance, the ensemble reflects the diversity of possible models consistent with the connectome and task-constraints.

The ensemble of models found a variety of parameter configurations (Extended Data Fig. 2), and models with random parameter configurations—but trained decoding networks—consistently performed worse on the task than task-optimized models (Extended Data Fig. 3). We focused on the 10 models which achieved the best task performance (Fig. 2a). We simulated neural responses to multiple experimentally characterized visual stimuli, and comprehensively compared model responses for each cell type to experimentally reported responses from 24 previous studies[53–55,58–60,62,64,70,83–97] (Overview in Supplemental Data).

First, neural responses in the fly visual system are known to segregate into ON- and OFF-channels[57], a hallmark of visual computation across species[98,99]. We probed the contrast preference of each cell type using flash stimuli[59] and found that the ensemble predicts the segregation into ON- and OFF-pathways with high accuracy: The median flash response index (FRI) across the ensemble predicts the correct ON- and OFF-preferred contrast selectivity for 31 of the 31 cell types for which contrast selectivity has been experimentally established ($p = 4.7 \times 10^{-10}$, binomial test; note that in these analyses the M10 and Lo1 terminals of CT1 are treated separately). This is also the case for the best-performing model, which correctly assigns 29 out of 31 cells into the correct tuning category (Fig. 2b, $p = 2.3 \times 10^{-7}$, binomial test). Furthermore, the ensemble provides predictions for the remaining 34 cell types, and consistency across the ensemble provides a measure of confidence in the predictions (Fig. 2b).

Second, a major result in fly visual neuroscience has been the identification of the T4 and T5 neurons as directional selective neurons with cardinal direction tuning[62,64,100]. We characterised the motion selectivity of neurons by their responses to ON- or OFF-edges moving in 12 different directions. We found that the ensemble of models correctly predicts that T4 neurons are ON-motion selective, and T5 neurons are OFF-motion selective (Fig. 2c). The ensemble also correctly predicts the lack of motion tuning in the input neurons to T4 and T5 motion detector neurons (Mi1, Tm3, Mi4, Mi9, Tm1, Tm2, Tm4, Tm9, CT1; see Methods, Supplemental Data).

Our models also suggest the possibility that the transmedullary cell types, TmY3, TmY4, TmY5a, TmY13, and TmY18, might be tuned to ON-motion. We asked if our model predicted motion selectivity for all cell types with asymmetric, multicolumnar inputs, as this is a necessary connectivity motif for motion computation. Based on their local spatial connectivity profiles, we estimated that 19 cell types receive asymmetric, multi-columnar inputs (Methods), but found that only 12 are predicted to be motion selective ($p < 0.05$, binomial test; Methods) by the ensemble (SI, Supplemental Data). This suggests that our model integrates connectivity across the entire network, rather than simply focusing on local connectivity to determine which neurons are most likely to be motion selective. Next, we asked how well these tunings are predicted by models with random parameter configurations in comparison. We found that task-optimized models predict the known direction selectivity indices and flash response indices more accurately than random models ($p = 2 \times 10^{-11}$ and $p = 5.3 \times 10^{-5}$ respectively, Mann-Whitney-U test, Extended Data Fig. 5a and b).

Finally, we found that models which exhibited lower task error (Methods) also had more realistic tuning: Models with higher task performance predict the direction selectivity index of T4 and T5 cells and their inputs better ($p = 5.6 \times 10^{-6}$, Wald test, Extended Data Fig. 4b and 6).

Best task performing model accurately predicts tuning properties of T4 and T5 neurons and their inputs

Our DMN modelling approach enables a large number of model-based analyses which can illuminate the mechanistic basis of computation in a circuit, as well as suggest novel visual stimuli for experimental characterization. We illustrate these analyses using a single model from the ensemble with the best task performance (Fig. 3), focusing on the well-studied T4 and T5 neurons. A more comprehensive set of these analyses for every cell type and every model in the ensemble can be found in the Supplement.

First, we found that in the best task performing model, the four subtypes of the T4 respond strongly to dark edges, and the four subtypes of the T5 neurons to bright edges, moving in the four cardinal directions, in agreement with previous experimental findings[54,55,62,64] (Fig. 3a and Extended Data Fig. 8).

Second, we probed the mechanism of direction selectivity in T4 neurons (Fig. 3b). Examining the input currents to a single T4c neuron, we found that edges of the preferred contrast moving in the preferred direction of T4c cells elicit large responses through fast excitation and delayed inhibition, in agreement with experimental findings[54]. Conversely, edges moving to the null direction of T4c cells elicit no (or very small) responses because inhibition cancels excitation.

Third, we computed and compared the spatial and temporal receptive fields of the major columnar input neurons to T4 and T5 neurons. These input neurons have been the focus of multiple experimental studies of the motion detection pathways[53,60,70,91,93,94,101] (Fig. 3c). We characterized these receptive fields in our model by computing the spatial (Fig. 3d) and temporal (Fig. 3e) impulse responses to brief single ommatidium flashes (5 ms duration, Methods). We find that the model correctly predicts the spatial scale of the spatial receptive field for all major inputs to T4 and T5 neurons, and correctly predicts the preferred contrast for all but one of the major inputs to T4 and T5 neurons.

In agreement with experimental findings[70,93], we find that Tm3 and Tm4 have broad spatial receptive fields (two column radius, 11.6 °), while Mi1, Mi4, Mi9, Tm1, Tm2, Tm9, and CT1 compartments in both medulla and lobula have narrow spatial receptive fields (single column radius, 5.8 °). Further, the model accurately predicts the ON- vs OFF-contrast selectivity of neurons involved in motion detection, including lamina monopolar cells (L1-L5), which anatomically separate the ON- and OFF-pathways[57], and the direct input neurons to T4 and T5. These cells either depolarize (ON-selective) or hyperpolarize (OFF-selective) in response to light increment flashes. These temporal response properties are correctly predicted for all except the Tm4 cell in this model, which is incorrectly predicted to be ON-selective by its temporal receptive field.

For the motion selective T4 and T5 neurons, the spatiotemporal receptive fields are not separable in space and time. We characterized the full spatiotemporal receptive field for T4c and T5c neurons (Fig. 3f). ON-impulses on the leading side of the receptive field of the ON-contrast, upwards direction selective T4c cell lead to its fast depolarization, whereas impulses on the trailing side of the receptive field lead to a delayed hyperpolarization, again matching experimental findings[54]. ON-impulses on the leading side of the receptive field of the OFF-contrast, upwards direction selective T5c cell lead to its fast hyperpolarization, whereas impulses on the trailing side of its receptive field lead to a delayed depolarization. Because T5c is OFF-selective, its OFF-impulse responses are inverted, resembling the T4c spatiotemporal receptive field (Extended Data Fig. 7a). This suggests that in our model, T5c implements a similar motion tuning mechanism to OFF-edges as T4c to ON-edges. Again, this mechanism also agrees with experimental findings[55].

Finally, we show that the model can be used to design optimized stimuli: We used the model to screen for video sequences from the Sintel dataset that elicited the largest responses in the motion selective neurons (Fig. 3g; Methods). One might expect that pure ON- or OFF-stimuli would elicit the largest responses in T4 and T5, respectively. However, we find both ON-and OFF-elements in optimized stimuli, suggesting an interplay between ON- and OFF-pathways. We found that the stimulus that elicits the strongest response in the T4c cell is a central OFF-disc followed by an ON-edge moving upwards, matching the preferred direction of the cell. Similarly, for the T5c cell, the stimulus that elicits the strongest response is a central ON-disc followed by an OFF-edge moving upwards in the preferred direction of the cell (Extended Data Fig. 7b for corresponding full-field naturalistic stimuli, numerically optimized stimuli, and preferred moving edge stimuli). Taken together, this individual model predicts a large number of tuning properties for the T4 and T5 cells and their inputs.

Neural responses across the model ensemble cluster strongly for many cell types

The predictions in the previous section were derived from a single model. How similar or dissimilar are the predictions of different task-optimized models constrained with the same connectome? To address this question, for each cell type, we simulated the neural activity of a single neuron, in response to naturalistic video sequences from the Sintel dataset. We then used UMAP[102] to perform nonlinear dimensionality reduction on high-dimensional activity vectors of that neuron across the model ensemble, and clustered the models in the resulting 2D projections (Fig. 4a, see Methods, Supplement). For many cell types, we found that models predict strongly clustered neural responses. For T4c neurons for example, we found three clusters corresponding to qualitatively distinct responses of this cell type for naturalistic inputs: Two clusters contain models with direction selective T4c cells (Fig. 4b and c) with up- and down-selective cardinal tuning, respectively, whereas neurons in the third cluster are not direction tuned. The direction selective cluster with the (correct) upward preference has lowest average task error (circular marker, average task error 5.297) and contains the best task performing model analyzed previously (shown in Fig. 3), followed by the cluster with opposite preference (triangular marker, average task error 5.316). The non-selective cluster has the worst performance (square marker, average task error 5.357), suggesting that models with accurate tuning correlate with lower task error (see also Extended Data Fig. 6).

What are the circuit mechanisms in the ON-motion detection pathway (Fig. 4d) underlying tuning computations in the different clusters? We found that direction selectivity in the two tuned clusters is associated with opposite preferred contrast tuning of Mi4 and Mi9 neurons which provide direct flanking inhibitory input to T4 neurons (Fig. 4e). Models with the correct direction selectivity for T4 neurons also predict the correct contrast selectivity for Mi4 and M9 neurons, and vice versa (Fig. 4f).

This shows how the space of task-optimized and connectome-constrained models can be used to provide hypotheses about different circuit mechanisms which might underlie the tuning properties of individual cells. Conversely, it shows that experimentally measuring the tuning of one neuron automatically translates to constraints on other neurons in the circuit. Here, ‘clamping’ the T4c neurons to their measured tuning properties (by only selecting models from the correct cluster) is sufficient to correctly constrain the tuning of both Mi4 and Mi9 neurons.

Models predict motion tuning for TmY3

Amongst models with the best task performance, TmY3, TmY4, and TmY18 are often ON-motion selective (Fig. 2c). As these neurons have yet to be experimentally characterized, we analyzed these prediction in our models. Since TmY3 neurons do not receive inputs from other known motion selective neurons, we were intrigued by the possibility that it might directly compute a motion signal and possibly constitute a parallel pathway to the well-known T4 and T5 neurons. In contrast, TmY4 and TmY18 cells receive inputs from T4 cells, potentially inheriting their motion tuning.

In the model ensemble, we found four distinct clusters for TmY3 (Fig. 5a). In the best-performing cluster (circular marker) TmY3 responds to ON-edges from front to back or downwards (Fig. 5b). In contrast, in the second cluster (triangular marker), TmY3 is not direction selective. In the third cluster (square marker) TmY3 is direction selective to ON-edges moving from the back to the front. In the fourth cluster (star marker), TmY3 is, again, not direction selective. Together, the ensemble suggests ON-motion sensitivity for TmY3, but different clusters disagree in their predictions for direction and contrast selectivity.

In our connectome data, the strongest input elements of TmY3 by number of synapses are L4, L5, Tm2, Tm3, Mi1, Mi9, and Mi4 (Fig. 5c and d). While none of these input neurons are motion-selective, the asymmetries in their connectivity to TmY3 might allow it to detect motion. We asked if we could better constrain our predictions by asking which clusters also predicted the correct preferred contrast for these input neurons. We found that the first model cluster (Fig. 5a, circular marker), in which TmY3 is tuned to front-back or downards motion, most accurately captures the known contrast selectivity of all TmY3 input cells (Fig. 5e). In contrast, all three other clusters fail to consistently capture the OFF-selectivity of Mi9. Thus our model proposes TmY3 as a novel candidate motion detector independent of the well-known T4 and T5 motion pathways.

Sparse synaptic connectivity enables accurate prediction of neural responses from synthetic connectomes

We sought to understand the conditions under which connectomic measurements might strongly constrain models of neural computation. Universal function approximation theorems for artificial neural networks[26, 103] suggest that a single general-purpose connectivity can underlie many possible computations. Empirically, different deep neural network architectures trained to solve the same task have very different selectivity at the level of single units[38, 82, 104–106], even when they share population-level representations[30, 107, 108]. We hypothesized that while general-purpose neural architectures used in machine learning have a many-to-many relationship between the computational task and neural connectivity, biological neural circuits might have a tighter relationship with their computational task due to their sparse structured connectivity, and therefore connectomic measurements and task-constraints in such circuits would strongly constrain mechanistic computational models.

We investigated this counter-factual using synthetic feedforward networks designed to perform a classic handwritten digit recognition network task. We constructed a variety of synthetic networks which perform the same task, but with different degrees of sparsity in their connectivity (Fig. 6). While the synthetic networks all perform the same task with near perfect accuracy, they use dramatically different connectivity, and so use different mechanisms to perform the same task. We then used connectomic measurements from each synthetic network to construct a corresponding connectome-constrained model optimized to perform the same task (Fig. 6a). We then compared the neural responses of hidden neurons in each synthetic network to corresponding neurons in the corresponding task and connectome-constrained model. We considered two settings, one where the connectomic measurements only indicated whether a pair of neurons were connected but not their strength, and one where the measurements indicated both the connectivity and a noisy estimate of the strength of the connection. In our model of the fly visual system, we have connectomic measurements of the synapse count for each connection, which can inform the relative strength of a connection, but not the absolute strength. This constitutes an intermediate regime between known strength and unknown strength.

We found that densely connected synthetic networks, where each neuron in a given layer connects to every neuron in the next layer, are essentially uncorrelated with their task- and connectome-constrained models constructed with known connectivity but unknown strength (Fig. 6b, Extended Data Fig. 9a). In contrast, sparsely connected synthetic networks, where each neuron only connects to a small percentage of neurons, are well correlated with their task and connectome-constrained models (median Pearson's correlation of 0.856 for networks with 10% average connectivity). The correlation between synthetic networks and their simulations drops smoothly as the connection density increases. For connectome-constrained models constructed with known connectivity and a noisy estimate of connection strength, we found that connectivity sparseness is not required for models to correlate with their corresponding synthetic network (Fig. 6b, Extended Data Fig. 9b).

Discussion

Here we constructed a neural network with neural connectivity measured at the microscopic scale. We also required that at the macroscopic scale, the collective neural activity dynamics across the entire network result in an ethologically relevant computation. The combination of microscopic and macroscopic constraints enabled us to construct a large-scale computational model spanning many tens of cell types and tens of thousands of neurons. We showed that such large-scale mechanistic models could accurately make detailed predictions of the neural responses of individual neurons to dynamic visual stimuli, revealing the mechanisms by which computations are performed. Knowledge of the connectome played a critical role in this success, in part by leading to a massive reduction in the number of free model parameters.

We have taken a reductionist modeling approach to emphasize the important role played by the connectivity of a neural network. We found that for the motion pathways of the fruit fly visual system, this model correctly predicts many aspects of visual selectivity. However our reductionist model cannot, for example, account for the role played in this circuit by electrical synapses[109], complex chemical synapse dynamics[96], and neuromodulation[110]. Richer models of neurons, synapses, and extra-synaptic modulation will be essential to correctly predict these effects. Further, we only considered the role of this circuit in detecting motion, which is but one of many computations performed by the visual system[14, 111–114].

Our study provides a direct link between artificial neural networks and biological circuit models. In our task-optimized deep network, every model neuron and synapse corresponds to a real neuron and synapse in the brain. This correspondence enables detailed experimentally testable predictions for each neuron. In contrast, most previous studies using task-optimized deep neural networks have predominantly focused on modeling computations without such detailed connectomic measurements: in fly motion vision pathways[115, 116] and olfactory system[36], in fish oculomotor system[117] in mammalian visual pathways[44, 76, 106, 118] and prefrontal cortex[45]. More recently, there has been tantalizing early evidence, in small scale models[46] and larger models[119] of the fly motion pathways, that our approach to combining connectome and task-constraints in a single model might prove successful.

Our modeling approach provides a discovery tool, aimed at using connectomic measurements to generate detailed, experimentally testable hypotheses for the computational role of individual neurons. Measurements of neural activity are necessarily sparse and involve difficult trade-offs. Activity can frequently only be measured in a limited number of contexts, and for either a limited number of neurons or for a larger number of neurons with poorer temporal resolution. Connectome-constrained DMN models generate meaningful predictions even in the complete absence of neural activity measurements, but can be further constrained by sparse measurements of neural activity as we showed (Fig. 4), or even be directly optimized to match measured neural activity[120].

Whole nervous system connectome projects are nearing completion for the adult fruit fly[2, 13], and whole mouse brain connectome projects are now being discussed[121]. Large-scale whole nervous system models[120] will be of critical importance for integrating connectomic, transcriptomic, neural activity and animal behavior measurements across labs, scales, and the nervous system[2]. Further, with the recent development of detailed biomechanical body models for the fruit fly[122] and rodent[123], we can now contemplate constructing whole animal models spanning brain and body.

Author Contributions

Conceptualization, Methodology: JKL, FDT, JHM, SCT. Data curation: JKL, FDT, AN, KS, S-yT. Software and Investigation: JKL, MM, SP, FDT. Analysis: JKL, EG, AN, SCT. Writing: JKL, JHM, SCT. Writing (Review & Editing): EG, AN, KS, MM, SP, FDT. Supervision and funding: SCT, JHM.

Funding

This project was supported by the Howard Hughes Medical Institute. JKL and JHM were supported by the German Research Foundation (DFG) through Germany's Excellence Strategy (EXC-Number 2064/1, Project number 390727645) and the German Federal Ministry of Education and Research (BMBF; Tübingen AI Center, FKZ: 01IS18039A). JKL is a member of the International Max Planck Research School for Intelligent Systems (IMPRS-IS).

Conflict of Interest Statement

The authors declare that there is no conflict of interest.

Acknowledgements

We are grateful to Lou Scheffer and Lowell Umayam for assistance with accessing connectomic reconstructions. We thank Axel Borst, James Fitzgerald, Nathan Klapoetke, Gerry Rubin, Michael Reiser, and Karel Svoboda for valuable discussions. We thank James Fitzgerald, David Stern, Nathan Klapoetke, Albert Lee, Richard Gao, Jakob Voigts, Brett Mensh for valuable feedback on the manuscript. We thank Tory Herman for sharing the colorization of the optic lobe figure[65] (Fig. 1b).

This article is subject to HHMI's Open Access to Publications policy. HHMI lab heads have previously granted a nonexclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. Pursuant to those licenses, the author-accepted manuscript of this article can be made freely available under a CC BY 4.0 license immediately upon publication.

Figures

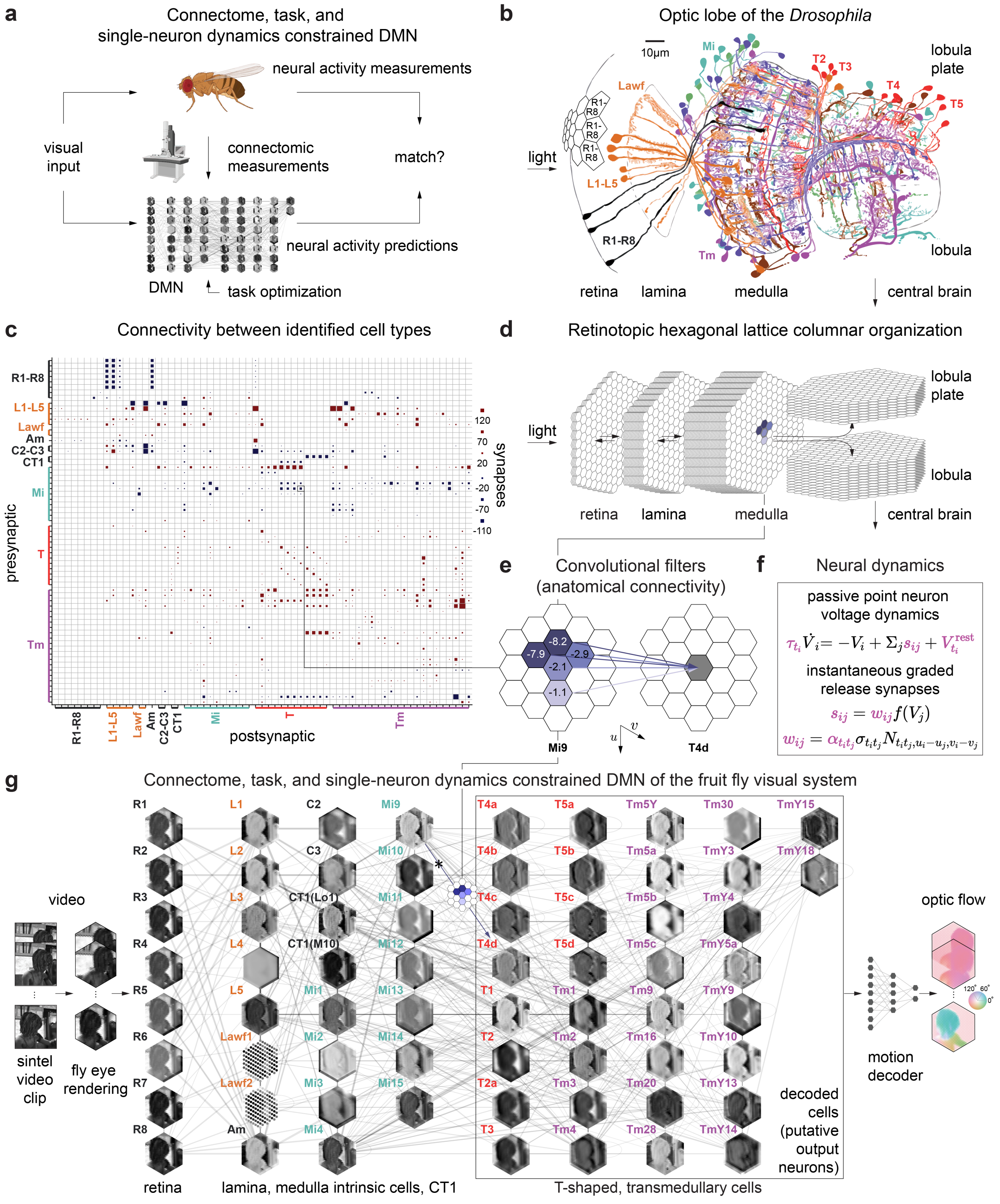

a
Connectome, task, and single-neuron dynamics constrained DMN
neural activity measurements
visual input
connectomic measurements
match?
neural activity predictions
DMN
task optimization
b
Optic lobe of the *Drosophila*
10μm
Mi
Lawf
T2
T3
T4
T5
lobula plate
light
R1-R8
L1-L5
R1-R8
Tm
lobula
retina
lamina
medulla
central brain
c
Connectivity between identified cell types
R1-R8
L1-L5
Lawf
Am
C2-C3
CT1
Mi
T
Tm
presynaptic
postsynaptic
synapses
120
70
20
-20
-70
-110
d
Retinotopic hexagonal lattice columnar organization
light
lobula plate
lobula
retina
lamina
medulla
central brain
e
Convolutional filters (anatomical connectivity)
-8.2
-7.9
-2.9
-2.1
-1.1
Mi9
T4d
u
v
f
Neural dynamics
passive point neuron voltage dynamics
$\tau_{t_i}\dot{V}_i = -V_i + \Sigma_j s_{ij} + V_{t_i}^{\text{rest}}$
instantaneous graded release synapses
$s_{ij} = w_{ij} f(V_j)$
$w_{ij} = \alpha_{t_i t_j} \sigma_{t_i t_j} N_{t_i t_j, u_i - u_j, v_i - v_j}$
g
Connectome, task, and single-neuron dynamics constrained DMN of the fruit fly visual system
video
sintel video clip
fly eye rendering
R1
R2
R3
R4
R5
R6
R7
R8
retina
L1
L2
L3
L4
L5
Lawf1
Lawf2
Am
C2
C3
CT1(Lo1)
CT1(M10)
Mi1
Mi2
Mi3
Mi4
Mi9
Mi10
Mi11
Mi12
Mi13
Mi14
Mi15
lamina, medulla intrinsic cells, CT1
T4a
T4b
T4c
T4d
T1
T2
T2a
T3
T5a
T5b
T5c
T5d
Tm1
Tm2
Tm3
Tm4
Tm5Y
Tm5a
Tm5b
Tm5c
Tm9
Tm16
Tm20
Tm28
Tm30
TmY3
TmY4
TmY5a
TmY9
TmY10
TmY13
TmY14
TmY15
TmY18
decoded cells (putative output neurons)
T-shaped, transmedullary cells
motion decoder
optic flow
120°
60°
0°

Figure 1: **Task-optimizing connectome-constrained models of the fly visual system.**
(a) The 'deep mechanistic network model' (DMN) aims to satisfy three constraints simultaneously: Its archiecture is based on connectomic measurements (see b-e), single-neuron and synaptic dynamics are given by simple mechanistic models (see f), and free parameters of this network model (see f, magenta) are optimized by training the model to perform an optic flow estimation task (see g).
(b) Schematic of optic lobe of *Drosophila* melanogaster comprising several processing stages (neuropils) and cell types (adapted from[65]). Our model includes retinal cells, lamina monopolar cells, medulla intrinsic cells, transmedullary cells, and T-shaped cells, i.e., all columnar cell types.
(c) Identified connectivity between 64 cell types, represented by total number of input synapses from all neurons of a given presynaptic cell type to a single postsynaptic of a given cell type. Blue color indicates putative hyperpolarizing inputs, red putative depolarizing inputs as inferred from neurotransmitter and receptor profiling. Size of squares indicates number of input synapses.
(d) Retinotopic hexagonal lattice columnar organization of the visual system model. Each lattice represents a cell type, each hexagon an individual cell. Positions of photoreceptor columns are aligned with positions of downstream columns. The model comprises synapses from all neuropils (SI Fig. 1) and downstream and upstream projecting connections from the retina, lamina, and medulla. Convolutional filter between Mi9 cells and a T4d cell (see Panel d) is highlighted in the final lattice of the medulla.
(e) Example of convolutional filter, representing Mi9 inputs onto T4d cells. Values represent the average number of synapses projecting from presynaptic Mi9 cells in columns with indicated offset onto the postsynaptic dendrite of T4d cells. Values indicate connection strength derived from electron microscopy data.
(f) Single-neuron and synaptic dynamics are given by simple mechanistic models. Free parameters of this network model (magenta) are optimized by training the model to perform optic flow estimation task.
(g) Visualization of DMN performing optic flow estimation on a video clip rendered from the Sintel dataset. Given the input to the photoreceptors (R1-R8), we simulated the response of each neuron in the system. Each hexagonal lattice depicts a snapshot of the voltage levels of all cells from the corresponding cell type. Edges illustrate connectivity between identified cell types. A decoding network receives the simulated neural activity of all output neurons (T-shaped and transmedullary cells) to compute optic flow. Parameters of the recurrent network model and the decoding network are optimized using optimization methods from deep learning.

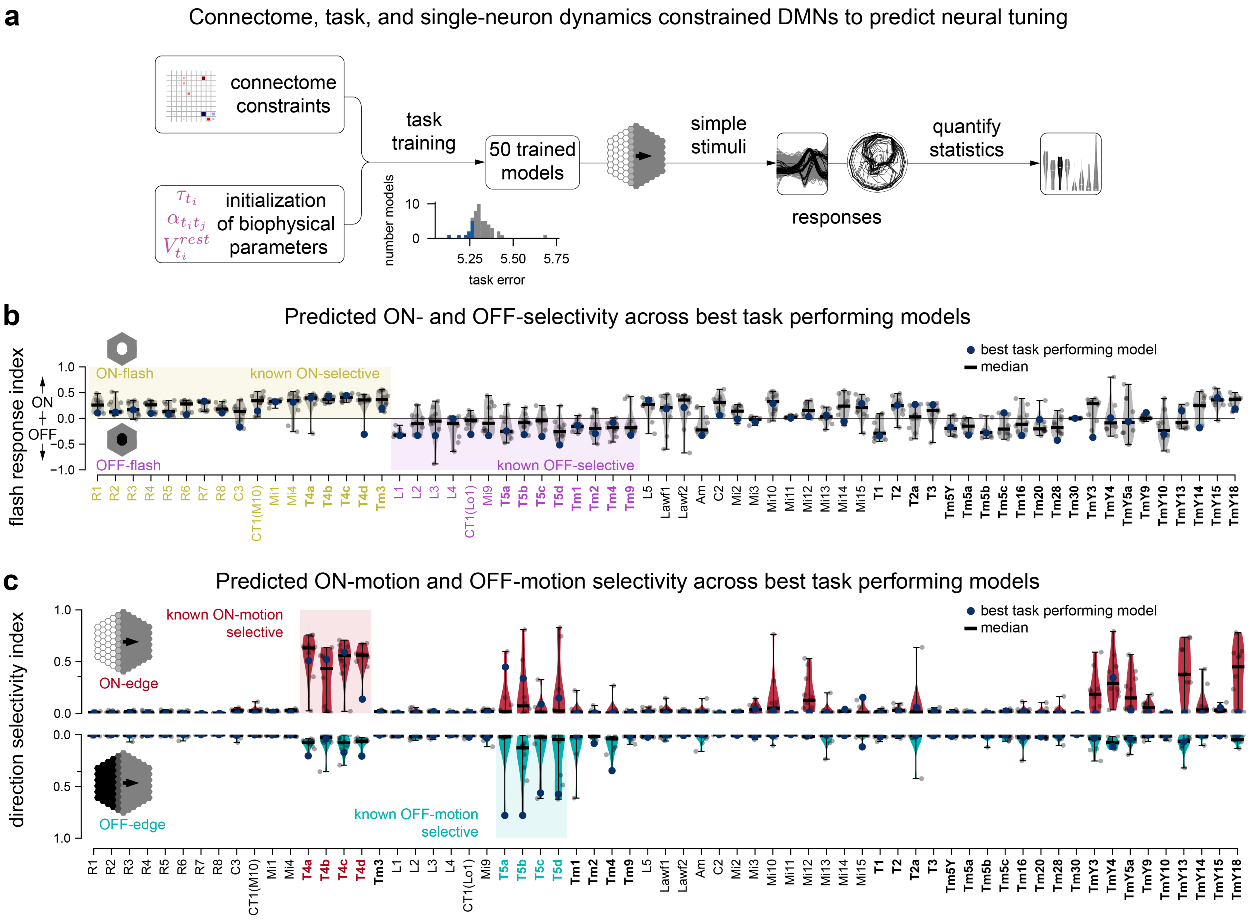


Figure 2: **Ensembles of connectome-constrained task-optimized DMNs predict tuning properties.**
(a) We task-optimized 50 connectome-constrained DMNs, yielding different solutions for the biophysical parameters. Inset: Distribution of task errors across the ensemble, 10 best models in blue. We characterized the responses and tuning properties of model neurons from each cell type to experimentally characterized visual stimuli and compare them to known tuning properties.
(b) ON- and OFF-contrast selectivity indices for each cell type based on peak transient responses for flash stimuli (Methods, Equation 3) for 10 models with best task performance (10 worst-performing models in Extended Data Fig. 4). Cell types with experimentally determined ON-/OFF-selectivity cell types are colored in yellow and violet respectively. Bold labels indicate cell types which provided input to optic flow decoder during training. Cell type names in black have not yet been experimentally characterized.
(c) Direction selectivity index (DSI) from neural responses to moving edges (Methods, Equation 4) for the same 10 best models as in (b).

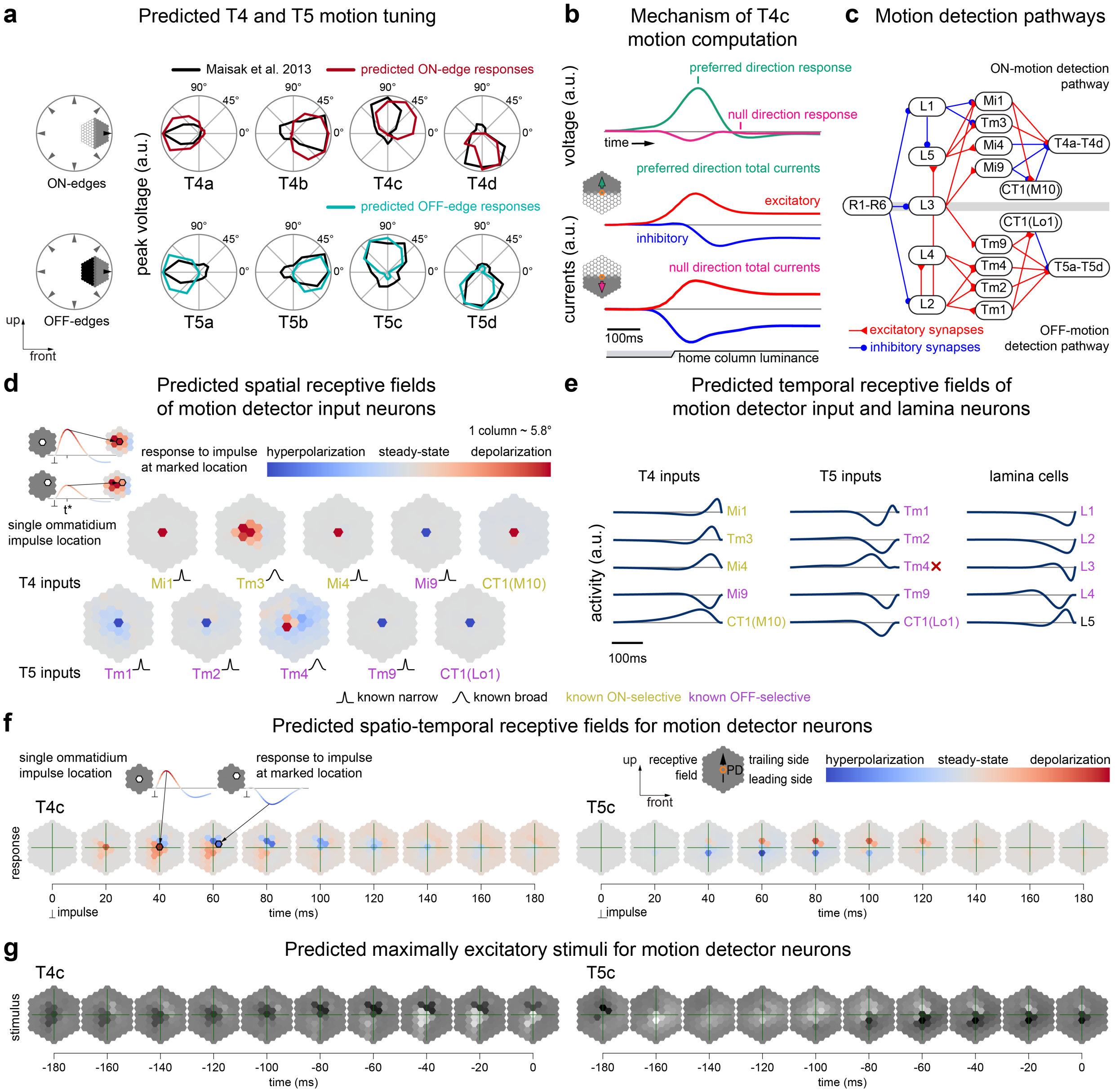

a
Predicted T4 and T5 motion tuning
Maisak et al. 2013
predicted ON-edge responses
predicted OFF-edge responses
ON-edges
OFF-edges
up
front
peak voltage (a.u.)
90°
45°
0°
T4a
T4b
T4c
T4d
T5a
T5b
T5c
T5d
b
Mechanism of T4c motion computation
preferred direction response
null direction response
time
voltage (a.u.)
preferred direction total currents
excitatory
inhibitory
null direction total currents
currents (a.u.)
100ms
home column luminance
c
Motion detection pathways
ON-motion detection pathway
OFF-motion detection pathway
R1-R6
L1
L2
L3
L4
L5
Mi1
Tm3
Mi4
Mi9
CT1(M10)
T4a-T4d
CT1(Lo1)
Tm9
Tm4
Tm2
Tm1
T5a-T5d
excitatory synapses
inhibitory synapses
d
Predicted spatial receptive fields of motion detector input neurons
1 column ~ 5.8°
response to impulse at marked location
hyperpolarization
steady-state
depolarization
t*
single ommatidium impulse location
T4 inputs
Mi1
Tm3
Mi4
Mi9
CT1(M10)
T5 inputs
Tm1
Tm2
Tm4
Tm9
CT1(Lo1)
known narrow
known broad
known ON-selective
known OFF-selective
e
Predicted temporal receptive fields of motion detector input and lamina neurons
T4 inputs
T5 inputs
lamina cells
activity (a.u.)
Mi1
Tm3
Mi4
Mi9
CT1(M10)
Tm1
Tm2
Tm4
Tm9
CT1(Lo1)
L1
L2
L3
L4
L5
100ms
f
Predicted spatio-temporal receptive fields for motion detector neurons
single ommatidium impulse location
response to impulse at marked location
up
front
receptive field
PD
trailing side
leading side
hyperpolarization
steady-state
depolarization
T4c
T5c
response
0
20
40
60
80
100
120
140
160
180
impulse
time (ms)
g
Predicted maximally excitatory stimuli for motion detector neurons
T4c
T5c
stimulus
-180
-160
-140
-120
-100
-80
-60
-40
-20
0
time (ms)

Figure 3: **DMN with the best task error largely recapitulates known mechanism of motion computation.**
(a) Responses to moving edges in different directions for T4 and T5 subtypes from the DMN with best task performance. Predictions of both direction and polarity of tuning agree with experimental measurements[62,64,100] (Null-contrasts in Extended Data Fig. 8).
(b) Fast excitation and delayed, offset inhibition enable T4c to detect motion, in agreement with experimental measurements[54]. An edge moving in the preferred direction elicits fast excitatory input currents (red) and delayed inhibitory input currents (blue) to T4c, leading to large depolarization (green). In contrast, an edge moving in the null direction elicits simultaneous arrival of excitatory and inhibitory inputs to T4c, leading to a null response (magenta).
(c) Major cell types and connectivity in the ON- (T4) and OFF- (T5) motion detection pathways (simplified).
(d) Spatial receptive fields of major motion detector input neurons revealed by single-ommatidium flashes are in agreement with experimental measurements[70,93]. For major T4 inputs, the best-performing DMN correctly predicts narrow spatial receptive fields for Mi1, Mi4, Mi9, and CT1(M10), and a wide receptive field for Tm3. Mi1, Tm3, Mi4, and CT1(M10) respond with depolarization to the ON-impulses, Mi9 responds with hyperpolarization. For major T5 inputs, the DMN correctly predicts narrow spatial receptive fields for Tm1, Tm2, Tm9, and CT1(Lo1), and a wide receptive field for Tm4. Tm1, Tm2, Tm4, Tm9, and CT1(Lo1) respond with hyperpolarization. Depending on the input ommatidium, Tm4 also responds with depolarization.
(e) Temporal receptive fields for inputs also are in agreement with experimental measurements[60,93], with the exception of Tm4 (red cross). For major T4 inputs, in this DMN, Mi1, Tm3, and Mi4 respond with transient depolarization. In contrast, CT1(M10) responds with a longer sustained depolarization to a central ON-impulse. Mi9 hyperpolarizes. For major T5 inputs, in the model, Tm2, Tm9, and CT1(Lo1) respond with transient hyperpolarization. Tm1 depolarizes fast followed by strong hyperpolarization and Tm4 is incorrectly predicted to depolarize (red cross). For lamina cell types, this DMN predicts hyperpolarization in L1, L2, L3, and L4 and depolarization in L5 in response to a central ON-impulse.
(f) Spatiotemporal receptive fields for motion detector neurons agree with experimental measurements[54]. Receptive field orientation of the motion detectors T4c and T5c align with their preferred motion axis. ON-impulses on the leading side of the receptive field quickly cause the T4 cell to depolarize, while ON-impulses on the trailing side of the receptive field cause slower hyperpolarization in the T4 cell. Conversely, for T5, ON-impulses on the leading side of the receptive field quickly cause the T5 cell to hyperpolarize, while ON-impulses on the trailing side of the receptive field cause slower depolarization in the T5 cell (Receptive fields for OFF-impulses in Extended Data Fig. 7a). **(g)** Video sequence predicted to elicit the strongest responses in T4c and T5c cells. A central OFF-disc followed by an ON-edge moving upwards elicits the strongest response in a T4c cell. A central ON-disc followed by an OFF-edge moving upwards elicits the strongest response in the central column T5c cell. We regularized the full field naturalistic stimuli to show only the content of the input that the cell is responsive to (Full field naturalistic and artificial maximally excitatory stimuli in Extended Data Fig. 7b).

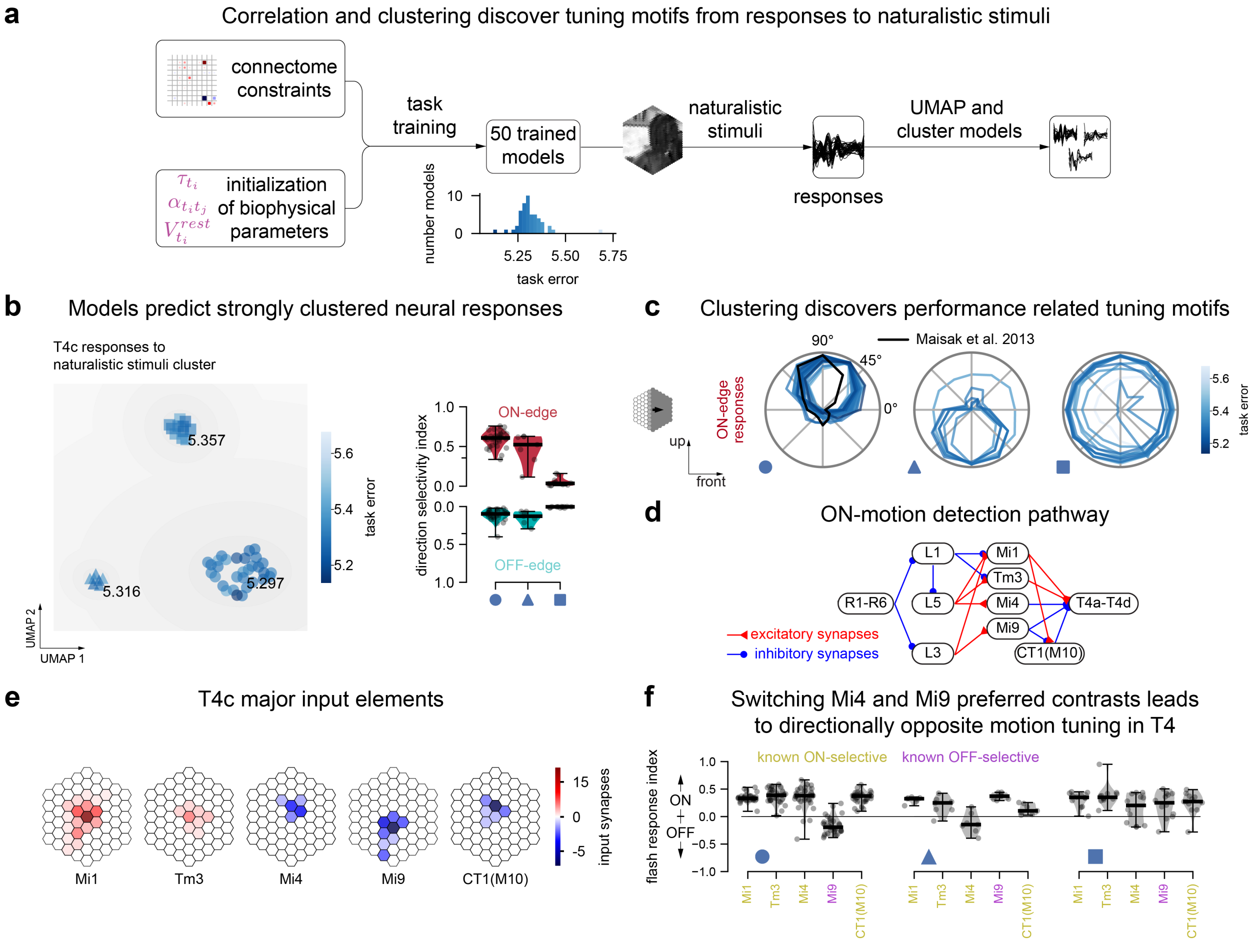

a
Correlation and clustering discover tuning motifs from responses to naturalistic stimuli
connectome constraints
τ_{t_i}
$\alpha_{t_i t_j}$
$V_{t_i}^{rest}$
initialization of biophysical parameters
task training
50 trained models
number models
10
0
5.25
5.50
5.75
task error
naturalistic stimuli
responses
UMAP and cluster models
b
Models predict strongly clustered neural responses
T4c responses to naturalistic stimuli cluster
5.357
5.316
5.297
5.6
5.4
5.2
task error
UMAP 2
UMAP 1
direction selectivity index
1.0
0.5
0.0
0.0
0.5
1.0
ON-edge
OFF-edge
c
Clustering discovers performance related tuning motifs
90°
45°
0°
Maisak et al. 2013
ON-edge responses
up
front
5.6
5.4
5.2
task error
d
ON-motion detection pathway
L1
L5
L3
R1-R6
Mi1
Tm3
Mi4
Mi9
CT1(M10)
T4a-T4d
excitatory synapses
inhibitory synapses
e
T4c major input elements
Mi1
Tm3
Mi4
Mi9
CT1(M10)
15
0
-5
input synapses
f
Switching Mi4 and Mi9 preferred contrasts leads to directionally opposite motion tuning in T4
known ON-selective
known OFF-selective
flash response index
ON
OFF
1.0
0.5
0.0
−0.5
−1.0
Mi1
Tm3
Mi4
Mi9
CT1(M10)

Figure 4: **Cluster analysis of DMN ensembles enables hypotheses generation and suggests experimental tests.**
(a) We clustered 50 DMNs after performing nonlinear dimensionality reduction of their predicted responses to naturalistic scenes for each cell type (Inset: Distribution of task performance across models). We compared their tuning to simple stimuli to identify whether clusters correspond to qualitatively different tuning mechanisms.
(b) Responses of T4c cells to naturalistic scenes reveal three distinct clusters. T4c cells in DMNs in the first and second clusters (circular and triangular marker) are ON-motion direction selective, whereas those in the third cluster (square marker) are not.
(c) The DMNs reveal three distinct solutions for the T4c cells (which are known to be tuned to upwards ON-motion): (1) upwards tuning (cluster with lowest average task error of 5.297, circular marker in panel b), (2) downwards tuning (5.316 average error, triangular marker), or (3) no motion tuning to ON-edges (5.357 average error, square marker).
(d) ON-motion detection pathway from Fig 3c.
(e) Connectivity of major input elements to T4c. Blue and red represent putative hyper- and depolarizing inputs. Saturation represents average number of input synapses for each offset location in the T4c dendrite (see Fig. 1e).
(f) Tuning properties within each cluster reveal dependencies between T4 tuning and that of Mi4 and Mi9 cells in the ensemble: Switching Mi4 (known ON-contrast selective) and Mi9 (known OFF-contrast selective) contrast preferences results in directionally opposite motion tuning solutions in T4. Cluster 1 (T4c in DMN upwards tuned, circle) indicates ON-selectivity for Mi1, Tm3, Mi4, and CT1(M10), and OFF-selectivity for Mi9. For ON-motion stimuli, in these DMNs T4c receives central depolarizing input from Mi1 and Tm3 and dorsal hyperpolarizing input from Mi4 and CT1(M10). For cluster 2 (T4c in DMN downwards tuned, triangle), Mi1, Tm3, Mi9, and CT1(M10) are ON-selective and Mi4 is OFF-selective. For ON-motion stimuli, in these DMNs, T4c receives central depolarizing input from Mi1 and Tm3, ventral hyperpolarizing input from Mi9 and dorsal hyperpolarizing input from CT1(M10). For cluster 3 (T4c in DMN not tuned, square), all major input elements are ON-selective. For ON-motion stimuli, in these DMNs, T4c receives central depolarizing input from Mi1 and Tm3, dorsal hyperpolarizing input from Mi4 and CT1(M10), and ventral hyperpolarizing input from Mi9.

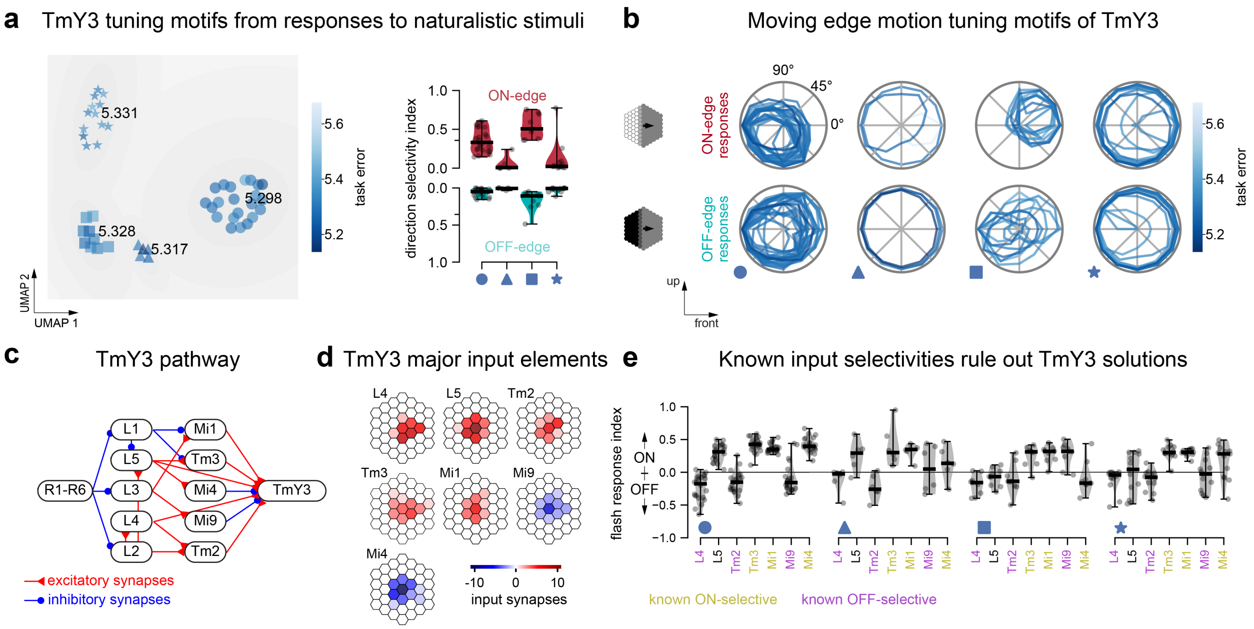


Figure 5: **DMNs suggest that TmY3 neurons compute motion independently of T4 and T5 neurons.**
(a) Dimensionality reduction on TmY3 responses to naturalistic stimuli reveals 4 clusters of DMNs with average task errors 5.298 (circle), 5.317 (triangle), 5.328 (square) and 5.331 (star). Across clusters, TmY3 shows different strengths of direction selectivity (evaluated with moving edge stimuli). ON-edge direction selectivity is strong in the first and the third cluster.
(b) Normalized peak responses of TmY3 to moving edge stimuli in the DMNs of each cluster.
(c) Major cell types and synaptic connections in the pathway that projects onto TmY3 (simplified).
(d) The input elements of TmY3 with the highest amount of synapses are L4, L5, Tm2, Tm3, Mi1, Mi9, and Mi4. The asymmetries of their projective fields could allow TmY3 to become motion selective.
(e) Dependencies between TmY3 tuning and the contrast preference of its input cells. For clusters in which TmY3 is motion selective, cluster 1 (TmY3 tuning to downwards/front-to-back motion, circular marker) indicates ON-selectivity for Tm3, Mi1, and Mi4 cells, and OFF-selectivity for L4, Tm2, and Mi9 cells, in agreement to the known selectivities. In contrast, cluster 3 (TmY3 tuning to upwards/back-to-front motion, square marker) indicates ON-selectivity for Mi9 in contradiction to the known selectivities and hence ruling out the third TmY3 tuning solution.

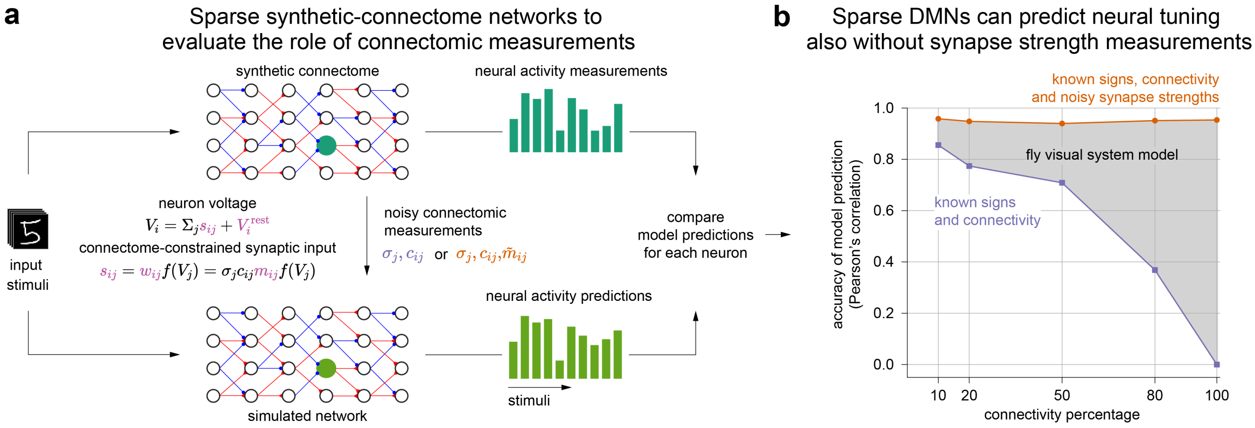


Figure 6: **Connectomic measurements can strongly constrain neural networks in circuits with sparse connectivity.**

(a) We constructed synthetic-connectome networks for classifying hand-written digits with varying degrees of sparse connectivity. For each synthetic-connectome network, we simulated connectomic measurements and constructed a connectome-constrained and task-optimized simulated DMN (Methods). We measured the correlation of the neural responses across stimuli of the same neuron in the synthetic-connectome network (dark green) and the fitted neural network (light green).

(b) Median neural response correlation coefficient from 100 randomly-sampled neuron pairs from each layer and across 25 network pairs. Simulations constrained only with connectivity measurements only correlate well for low connectivities (dark blue), while simulations constrained by measurements of both connectivity and connection strength correlate well across all average connectivity percentages (orange). The fly visual system likely falls in the region between the two curves, since measured synapse counts inform relative connection strengths between pairs of neurons for the same pair of cell types, but not absolute connection strength.

Methods

Construction of spatially invariant connectome from local reconstructions

We built a computational model of the fly visual system which is consistent with available connectomic data[48–50,68,124,125], which has biophysically plausible neural dynamics, and which can be computationally trained to solve an ethiologically relevant behavioural task, namely estimation of optic flow. To achieve this, we developed algorithms to blend annotations from two separate data-sets by transforming, sanitizing, combining and pruning the raw data sets into a coherent connectome spanning all neuropils of the optic lobe (Supplementary Note 1).

The original data stems from focused ion beam scanning EM datasets (FIBSEM) from the FlyEM project at Janelia Research Campus. The FIB-25 dataset volume comprises seven medulla columns and the FIB-19 dataset volume comprises the entire optic lobe and, in particular, detailed connectivity information for inputs to both the T4 and T5 pathways[48,49,68]. The data available to us consisted of 1801 neurons, 702 neurons from FIB-25 and 1099 neurons from FIB-19. For about 830 neurons the visual column was known from hand annotation. These served as reference positions. Of the 830 reference positions, 722 belong to neuron types selected for simulation. None of the T5 cells, whose directional selectivity we aimed to elucidate, were annotated. We therefore built an automated, probabilistic expectation maximization algorithm that takes synaptic connection statistics, projected synapse center-of-mass clusters and existing column annotations into account. We verified the quality of our reconstruction as described in Supplementary Note 1 Only the neurons consistently annotated with both 100% and 90% of reference positions used were counted to estimate the number of synapses between cell types and columns, in order to prune neuron offsets with low confidences.

Synaptic signs for most cell types were predicted based on known expression of neurotransmitter markers (primarily the cell type specific transcriptomics data from Davis et al 2020). For a minority of cell types included in the model, no experimental data on transmitter phenotypes were available. For these neurons, we used guesses of plausible transmitter phenotypes. To derive predicted synaptic signs from transmitter phenotypes, we assigned the output of histaminergic, GABAergic and glutamatergic neurons as hyperpolarizing and the output of cholinergic neurons as depolarizing. In a few cases, we further modified these predictions based on distinct known patterns of neurotransmitter receptor expression (see Davis et al. for details). For example, output from R8 photoreceptor neurons, predicted to release both acetylcholine and histamine, was treated as hyperpolarizing or depolarizing, respectively, depending on whether a target cell type is known to express the histamine receptor ort (a histamine-gated chloride channel).

Representing the model as a hexagonal convolutional neural network Our end-to-end differentiable[126] DMN model of the fly visual system can be interpreted as a continuous-time neural ordinary differential equation (neural ODE)[127] with a deep convolutional recurrent neural network (convRNN)[128] architecture that is trained to perform a computer vision task using backpropagation through time (BPTT)[41,79]. Our goal was to optimize a simulation of the fly visual system to perform a complex visual information processing task using optimization methods from deep learning. One hallmark of visual systems that has been widely exploited in such tasks are their convolutional nature[129–132], i.e. the fact that the same computations are applied to each pixel of the visual input. To model the hexagonal arrangement of photoreceptors in the fly retina, we developed a hexagonal convolutational neural network in the widely used deep learning framework Pytorch[40], which we used for simulation and optimization of the model.

Neuronal Dynamics

In detail, we simulated point neurons with voltages V_i of a postsynaptic neuron i, belonging to cell type t_i using threshold-linear dynamics, mathematically equivalent to commonly used formulations of firing-rate models[133]

$$\tau_{t_i} \dot{V}_i = -V_i + \sum_j s_{ij} + V_{t_i}^{\text{rest}} + e_i \tag{1}$$

Neurons of the same cell type share time constants, τ_{t_i}, and resting potentials, $V_{t_i}^{\text{rest}}$. Dynamic visual stimuli were delivered as external input currents e_i to the photoreceptor (R1-R8), for all other cell types, $e_i = 0$. In our model, instantaneous graded synaptic release from presynaptic neuron j to postsynaptic neuron i is described by

$$s_{ij} = w_{ij} f(V_j) = \alpha_{t_i t_j} \sigma_{t_i t_j} N_{t_i t_j \Delta u \Delta v} f(V_j), \tag{2}$$

comprising the anatomical filters in terms of the synapse count from EM-reconstruction, $N_{t_i t_j \Delta u \Delta v}$, at the offset location $\Delta u = u_i - u_j$ and $\Delta v = v_i - v_j$ in the hexagonal lattice between two types of cells, t_i and t_j, and further characterised by a sign, $\sigma_{t_i t_j} \in \{-1, +1\}$, and a non-negative scaling factor, $\alpha_{t_i t_j}$.

The synapse model in Equation 2 entails a trainable non-negative scaling factor per filter that is initialized as

$$\alpha_{t_i, t_j} = \frac{0.01}{\langle N_{t_i, t_j} \rangle_{\Delta u, \Delta v}},$$

with the denominator describing the average synapse count of the filter. Synapse counts, $N_{t_i t_j \Delta u \Delta v}$, and signs, $\sigma_{t_i t_j}$, from reconstruction and neurotransmitter and receptor profiling were kept fixed. The scaling factor was clamped during training to remain non-negative.

Moreover, at initialization, the resting potentials were sampled from a Gaussian distribution

$$V_{t_i}^{\text{rest}} \sim \mathcal{N}(\mu_{V^{\text{rest}}}, \sigma^2_{V^{\text{rest}}})$$

with mean $\mu_{V^{\text{rest}}} = 0.5$ (a.u.) and variance $\sigma^2_{V^{\text{rest}}} = 0.05$ (a.u.). The time constants were initialized at $\tau_{t_i} = 50$ms. The 50 task-optimized DMNs were initialized with the same parameter values. During training, in Euler integration of the dynamics, we clamped the time constants as $\tau_i = \max(\tau_i, \Delta t)$, so that they remain above the integration time step Δt at all times.

In total, the model comprises 45669 neurons and 1513231 synapses, across two-dimensional hexagonal arrays 31 columns across. Independently of the extent of the two-dimensional hexagonal arrays are the numbers of free parameters: 65 resting potentials, 65 membrane time constants, 604 scaling factors; and connectome determined parameters: 604 signs, and 2355 synapse counts. Thus, the number of free parameters in the visual system model is 734.

In the absence of connectomic measurements, the number of parameters to be estimated is much larger. With $T = 65$ cell types (counting CT1 twice for the compartments in the medulla and lobula) and $C = 721$ cells per type for simplicity, the number of cells in our model would be $TC = 46,865$. Assuming an RNN with completely unconstrained connectivity and simple dynamics $\tau_i V_i = -V_i + \sum_j w_{ij} f(V_j) + V_i^{\text{rest}}$ we would have to find $(TC)^2 + 2(TC) = 2,196,421,955$ free parameters. Assuming a convolutional RNN with shared filters between cells of the same postsynaptic type, shared time constants, and resting potential, the amount of parameters reduces drastically to $T^2C + 2T = 3,046,355$. Further assuming the same convolutional RNN but additionally convolutional filters are constrained to $F = 5$ visual columns, i.e. the number of presynaptic input columns in hexagonal lattice is $P = 3F(F+1)+1$, the amount of parameters reduces to $T^2P + 2T = 384,605$. Assuming as in our connectome only $Q = 604$ connections between cell types exist, this reduces the number of parameters further to $QP + 2T = 55,185$. Instead of parametrizing each individual synapse strength, we assume that synapse strength is proportional to synapse count from the connectome times a scalar for each filter, reducing the number of parameters to $Q + 2T = 734$ while providing enough capacity for the DMNs to yield realistic tuning to solve the task.

Convolutions using scatter and gather operations For training the network, we compiled the convolutional architecture specified by the connectome and the sign constraints to a graph representation containing (1) a collection of parameter buffers shared across neurons and/or connections, (2) a collection of corresponding index buffers indicating where the parameters relevant to a given neuron or connection can be found in the parameter buffers, and (3) a list of pairs (presynaptic neuron index, postsynaptic neuron index) denoting connectivity. This allowed us to efficiently simulate the network dynamics via Euler integration using a small number of element-wise, scatter, and gather operations at each time step. We found that this is more efficient than using a single convolution operation, or performing a separate convolution for each cell type, since each cell type has its own receptive field - some much larger than others - and the number of cells per type is relatively small.

Optic flow task

Model training An optic flow field for a video sequence consists of a 2D vector field for each frame. The 2D vector at each pixel represents the magnitude and direction of the apparent local movement of the brightness pattern in an image.

We frame the training objective as a regression task

$$\hat{\mathbf{Y}}[n] = \mathrm{Decoder}(\mathrm{DMN}(\mathbf{X}[0], ..., \mathbf{X}[n-1])),$$

with $\hat{\mathbf{Y}}$ the optic flow prediction, and $\mathbf{X}$ the visual stimulus sequence from the Sintel dataset, both sampled to a regular hexagonal lattice of 721 columns. With the objective to minimize the square error loss between predicted optic flow and target optic flow fields, we jointly optimized the parameters of both the decoder and the visual system network model described above.

In detail, for training the network, we added randomly augmented grey-scaled video sequences from the Sintel dataset sampled to a regular hexagonal lattice of 721 columns to the voltage of the eight photoreceptor cell types (Fig. 1f, Equation 1). We denote a sample from a minibatch of video sequences as $\mathbf{X} \in \mathbb{R}^{N,C}$ with N the number of time steps, and C the number of photoreceptor columns. The dynamic range of the input lies between 0 and 1. Input sequences during training entailed 19 consecutive frames drawn randomly from the dataset and resampled to match the integration rate. The original framerate of 24 Hz and 19 frames lead to a simulation of 792ms. We did not find that an integration time step smaller than 20 ms, i.e. a framerate of 50 Hz after resampling, yielded qualitatively superior task performance nor more realistic tuning predictions. We interpolated the target optic flow in time to 50 Hz temporal resolution, instead of resampling it. To increase the amount of training data for better generalization, we augmented input and target sequences as described further below. At the start of each epoch, we computed an initial state of the network's voltages after 500ms of grey stimuli presentation to initialize the network at a steady state for each minibatch during that epoch. The network integration given input $\mathbf{X}$ results in simulated sequences of voltages $\mathbf{V} \in \mathbb{R}^{N,T_C}$ with T_C the total number of cells. The subset of voltages, $\mathbf{V}_{\text{out}} \in \mathbb{R}^{N,D,C}$, of the D cell types in the black box in Fig. 1f was passed to a decoding network. For decoding, the voltage was rectified to avoid that the network finds biologically implausible solutions by encoding in negative dynamic ranges. Further, it was mapped to cartesian coordinates to apply Pytorch's standard spatial convolution layers for decoding and on each timestep independently. In the decoding network, one layer implementing spatial convolution, batch normalization, softplus activation, and dropout, followed by one layer of spatial convolution, transformed the D feature maps into the two-dimensional representation of the estimated optic flow, $\hat{\mathbf{Y}} \in \mathbb{R}^{N,2,C}$.

Using stochastic gradient descent with adaptive moment estimation ($\beta_1 = 0.9$, $\beta_2 = 0.999$, learning rate decreased from 5×10^{-5} to 5×10^{-6} in ten steps over iterations, batch size of four) and the automatic gradient calculation of the fully differentiable pipeline, we optimized the biophysical parameters through backpropagation through time such that they minimize the L2-norm between the predicted optic flow, $\hat{\mathbf{Y}}$, and the groundtruth optic flow, $\mathbf{Y}$:

$$L(\mathbf{Y}, \hat{\mathbf{Y}}) = ||\mathbf{Y} - \hat{\mathbf{Y}}||$$

We additionally regularized the shared resting potentials for 150,000 iterations, using stochastic gradient descent without momentum, based on time-averaged responses to naturalistic stimuli of the central column cell of each cell type, t_{central}, to encourage configurations of resting potentials that lead to nonzero and nonexploding activity in all neurons in the network. We weighted these terms independently with $\gamma = 1$, encouraging activity greater than a, and $\delta = 0.1$, encouraging activity less than a. We chose $\lambda_V = 0.1$ and $a = 5$ in arbitrary units. With B being the batch size and T the number of all cell types, the regularizer is

$$R(V) = \frac{\lambda_V}{BT} \sum_b \sum_{t_{\text{central}}} \begin{cases} \gamma(\bar{V} - a)^2, & \text{if } \bar{V} = \frac{1}{N} \sum_n V_{bt_{\text{central}}}[n] \leq a \\ \delta(\bar{V} - a)^2, & \text{if } \bar{V} > a. \end{cases}$$

We regularly checkpointed the above error measure $L(\mathbf{Y}, \hat{\mathbf{Y}})$ averaged across a held out validation set of Sintel video clips. Models generalized on optic flow computation after round about 250,000 iterations, yielding functional candidates for our fruit fly visual system models that we analyzed with respect to their tuning properties.

Task-optimization dataset We optimized the network on 23 sequences from the publicly available computer-animated movie Sintel[134]. The sequences have 20-50 frames, at a frame rate of 24 frames per second and a pixel resolution of 1024x436. The dataset provides optical flow in pixel space for each frame after the first of each sequence. Since the integration time steps we use are faster than the actual sampling rate of the sequences, we resample input frames accordingly over time and interpolate the optic flow.

Fly-eye rendering We first transformed the RGB pixel values of the visual stimulus to normalized greyscale between 0 and 1. We translated cartesian frames into receptor activations by placing simulated photoreceptors in a two-dimensional hexagonal array in pixel space, 31 columns across resulting in 721 columns in total, spaced 13 pixels apart. The transduced luminance at each photoreceptor is the greyscale mean value in the 13×13-pixel region surrounding it.

Augmentation We used (1) random flips of input and target across one of the three principal axes of the hexagonal lattice, (2) random rotation of input and target around its six-fold rotation axis, (3) adding element-wise Gaussian noise with mean zero and variance $\sigma_n = 0.08$ to the input X (then clamped at 0) (4) random adjustments of contrasts, $\log c \sim \mathcal{N}(0, \sigma_c^2 = 0.04)$, and brightness, $b \sim \mathcal{N}(0, \sigma_b^2 = 0.01)$, of the input with $X' = c(X - 0.5) + 0.5 + cb$.

In addition, we strided fly-eye rendering across the rectangular raw frames in width, subsampling multiple scenes from one. We ensured that such subsamples from the same scene do not distribute across training and validation sets. Input sequences in chunks of 19 consecutive frames were drawn randomly in time from the full sequences.

Black-box decoding network The decoding network is feedforward, convolutional and has no temporal structure. Aspects of the architecture are explained in the paragraph Model training. The spatial convolutions have a filter size of 5×5. The first layer transforms the $D = 34$ feature maps to an eight-channel intermediate representation, that is further translated by an additional convolutional layer to a three-channel intermediate representation of optic flow. The third channel is used as shared normalization of each coordinate of the remaining two-dimensional flow prediction. The decoder uses Pytorch-native implementations for two-dimensional convolutions, batch normalization, softplus activation, and dropout. We initialized its filter weights homogeneously at 0.001.

Model characterization

Task error To rank models based on their task performance, we computed the standard optic flow metric of average end-to-end point error (EPE)[135] which calculates the average over all timesteps and pixels (i.e. here columns) of the error

$$\mathrm{EPE}(\mathbf{Y}, \hat{\mathbf{Y}}) = \frac{1}{NC} \sum_n \sum_c \sqrt{(y_{1c}[n] - \hat{y}_{1c}[n])^2 + (y_{2c}[n] - \hat{y}_{2c}[n])^2}.$$

between predicted optic flow and groundtruth optic flow, and averaged across the held out validation set of Sintel sequences.

Unconstrained CNN We trained unconstrained, fully convolutional neural networks on the same dataset and task yielding a lower bound for the task error of the DMN. Optic flow was predicted by the CNN from two consecutive frames

$$\hat{Y}[n] = \mathrm{CNN}(X[n], X[n-1]).$$

with the original frame rate of the Sintel movie. We chose 5 layers for the CNN with 32, 92, 136, 8, 2 channels respectively and kernel size 5 for all but the first layer which kernel size is 1. Each layer performs a convolution, batch normalization, and ELU activation except the last layer which only performs a convolution. We optimized an ensemble of 5 unconstrained CNNs with 414,666 free parameters each using the same loss function, $L(Y, \hat{Y})$, as for the DMN. We used the same dataset, i.e. hexagonal sequences and augmentations from Sintel, for training and validating the CNN as for training and validating the DMN, allowed by two custom modules mapping from hexagonal lattice to cartesian map and back.

Random DMNs We created 50 DMNs with random parameters (all sampled from Gaussians of different means and standard deviations), and task-optimized only their decoding network, yielding an upper bound for the task error of the task-optimized DMN and a lower bound for the accuracy of tuning predictions without task-optimization.

Circular flash stimuli To evaluate the contrast selectivity of cell types in task-constraint model candidates, we simulated responses of each DMN to circular flashes. The networks were initialized at an approximate steady state after 1s of grey-screen stimulation. Afterwards the flashes were presented for 1s. The flashes with a radius of 6 columns were ON (intensity $I = 1$) or OFF ($I = 0$) on grey ($I = 0.5$) background. We integrated the network dynamics with an integration time step of 5 ms. We recorded the responses of the modeled cells in the central columns to compute the flash response index.

Flash response index To derive the contrast selectivity of a cell type, t_i, we computed the flash response index as

$$\mathrm{FRI}_{\mathrm{t_i}} = \frac{r^{\text{peak}}_{\mathrm{t_{central}}}(I=1) - r^{\text{peak}}_{\mathrm{t_{central}}}(I=0)}{r^{\text{peak}}_{\mathrm{t_{central}}}(I=1) + r^{\text{peak}}_{\mathrm{t_{central}}}(I=0)} \tag{3}$$

from the non-negative activity

$$r^{\text{peak}}_{\mathrm{t_{central}}}(I) = \max_n V_{t_{\text{central}}}[n](I) + |\min_{n,I} V_{t_{\text{central}}}[n](I)|,$$

from voltage responses $V_{t_{\text{central}}}[n](I)$ to circular flash stimuli of intensities $I \in \{0, 1\}$ lasting for 1s after 1s of grey stimulus. We note that our index quantifies whether the cell depolarizes to ON- or to OFF-stimuli. However, cells like R1-R8, L1, and L2 can be unrectified, i.e., sensitive to both light increments and light decrements, which is not captured by our index.

For the p-values reported in the results, we performed a binomial test with probability of correct prediction 0.5 (H0) or greater (H1) to both test whether the median FRI from the DMN-ensemble and the best-performing model predict the contrast preferences significantly. Additionally, we found for each individual cell type across 50 DMS that predictions for 29 out of 31 cell types are significant ($P < 0.05$, binomial).

Moving edge stimuli To predict the motion sensitivity of each cell type in task-constrained DMNs, we simulated the response of each network, initialized at an approximate steady state after 1s of grey-screen stimulation, to custom generated edges moving to 12 different directions, $\theta \in [0°, 30°, 60°, 90°, 120°, 150°, 180°, 210°, 240°, 270°, 300°, 330°]$. We integrated the network dynamics with an integration time step of 5ms. ON-edges ($I = 1$) or OFF-edges ($I = 0$) moved on grey ($I = 0.5$) background. Their movement ranged from -13.5° to 13.5° visual angle and we moved them at six different speeds, ranging from $13.92°/s$ to $145°/s$ ($S \in [13.92°/s, 27.84°/s, 56.26°/s, 75.4°/s, 110.2°/s, 145.0°/s]$).

Direction selectivity index We computed a direction selectivity index of a particular type t_i as

$$\mathrm{DSI}_{\mathrm{t_i}}(I) = \frac{1}{|\mathbb{S}|} \sum_{S \in \mathbb{S}} \frac{|\sum_{\theta \in \Theta} r^{\text{peak}}_{\mathrm{t_{central}}}(\mathrm{I}, \mathrm{S}, \theta) \exp i\theta|}{\max_{I \in \mathbb{I}} |\sum_{\theta} r^{\text{peak}}_{\mathrm{t_{central}}}(\mathrm{I}, \mathrm{S}, \theta)|} \tag{4}$$

from peak voltages

$$r^{\text{peak}}_{\mathrm{t_{central}}}(I, S, \theta) = \max_n V_{t_{\text{central}}}[n](I, S, \theta), \tag{5}$$

elicited from moving edge stimuli. We parametrized movement angle $\theta \in \Theta$, intensities $I \in \mathbb{I}$, and speeds $S \in \mathbb{S}$ of moving edges. To take the response magnitudes into account for comparing DSI for on- and for off-edges, we normalized by the maximum over both intensities in the denominator. To take different speeds into account, we averaged over $\mathbb{S}$.

Determining whether a cell type with asymmetric inputs counts as direction selective We counted a cell type as direction selective if the DSIs from its synthetic measurements were larger than 99% of DSIs from non-motion selective cell types (i.e. those with symmetric filters). We note, however, that estimates of the spatial asymmetry of connectivity from existing connectomic reconstructions can be noisy.

For deriving the 99%-threshold, we first defined a distribution $p(d^*|\mathrm{t_{sym}})$ over the direction selectivity index for non-direction selective cells, from peak responses to moving edges of cell types with symmetric inputs, $\mathrm{t_{sym}}$. We computed that distribution numerically by sampling

$$d^* = \frac{|\sum_{\theta^*} r^{\text{peak}}_{\mathrm{t_{central}}}(\mathrm{I}, \mathrm{S}, \theta^*) \exp i\theta|}{|\sum_{\theta} r^{\text{peak}}_{\mathrm{t_{central}}}(\mathrm{I}, \mathrm{S}, \theta)|}$$

for 100 independent permutations of the angle θ^*. We independently computed d^* for all stimulus conditions, models, and cell types with symmetric inputs. From $p(d^*|\mathrm{t_{sym}})$, we derived the threshold $d_{\text{thresh}} = 0.357$ as the 99% quantile of the random variable d^*, meaning that the probability that a realization of $d^* > d_{\text{thresh}}$ is less than 1% for cell types with symmetric inputs. To determine whether an asymmetric cell type counts as direction selective, we tested whether synthetically measuring direction selectivity larger than d_{thresh} in that cell type is binomial with probability 0.1 (H0) or greater (H1). We identified 12 cell types with asymmetric inputs (T4a, T4b, T4c, T4d, T5a, T5b, T5c, T5d, TmY3, TmY4, TmY5a, TmY18) as direction selective ($P < 0.05$) from our models, and seven cell types with asymmetric inputs to not count as direction selective (T2, T2a, T3, Tm3, Tm4, TmY14, TmY15). See Supplemental Data for reference of cell types with symmetric and asymmetric inputs in our model.

UMAP and clustering We first simulated central column responses to naturalistic scenes (24Hz Sintel video clips from the full augmented dataset) with an integration time step of 10 ms. We clustered models in feature space of concatenated central column responses and sample dimension. Next, we computed a nonlinear dimensionality reduction to 2d (UMAP), and finally fitted Gaussian mixtures of 2 to 5 components to the embedding to label the clusters based on the Guassian mixture model with the number of components that minimize the Bayesian information criterion, using the python libraries umap-learn and scikit-learn[102, 136].

Single ommatidium impulse stimuli To derive spatio-temporal receptive fields, we simulated the response of each network to single ommatidium impulses. Impulses were ON ($I = 1$) on grey ($I = 0.5$) background and presented for 5 ms after 2 s of grey-screen stimulation and followed by 5 s of grey-screen stimulation.

Spatio-temporal, spatial and temporal receptive fields We derived the spatio-temporal receptive field (STRF) of a cell type t_i as the baseline subtracted responses of the central column cell to single ommatidium impulses $J(u, v)$ at ommatidium locations (u, v):

$$\mathrm{STRF}_{t_{\text{central}}}[n](u, v) = V_{t_{\text{central}}}[n](J(u, v)) - V_{t_{\text{central}}}[n = 0](J(u, v)).$$

We derived spatial receptive fields, SRFs, from the responses to impulses $J(u, v)$ at the point in time at which the response to the central ommatidium impulse is at its extremum:

$$\mathrm{SRF}(u, v) = \mathrm{STRF}(n = \arg\max_n |\mathrm{STRF}[n](0, 0)|, u, v).$$

We derive temporal receptive fields, TRFs, from the response to an impulse $J(0, 0)$ at the central ommatidium: $\mathrm{TRF}[n] = \mathrm{STRF}[n](0, 0)$.

Maximally excitatory naturalistic and artificial stimuli First, we found the naturalistic maximally excitatory stimulus, $\mathbf{X}^*$, by identifying the Sintel video clip, $\mathbf{X}$, from the full dataset with geometric augmentations that elicited the highest possible response in the central column cell of a particular cell type in our models.

$$\mathbf{X}^* = \underset{\mathbf{X}\in\text{Sintel}}{\arg\max} V_{t_\text{central}}(\mathbf{X}).$$

Next, we regularized the naturalistic maximally excitatory stimulus, to yield $\mathbf{X}'$, capturing only the stimulus information within the receptive field of the cell, with the objective to minimize

$$L(\mathbf{X}') = \sum_n \|V_{t_\text{central}}(\mathbf{X}^*)[n] - V_{t_\text{central}}(\mathbf{X}')[n]\|^2 + \frac{1}{C}\sum_c \|\mathbf{X}'[n,c] - 0.5\|^2.$$

The first summand maintains the exact central response to $\mathbf{X}^*$, while the second sets the redundant stimulus information outside of the receptive field to grey ($I = 0.5$).

In addition, we computed artificial maximally excitatory stimuli[137].

Training synthetic connectomes

Training feedforward synthetic-connectome networks Sparsified feedforward neural networks with 6 hidden layers (linear transformations sandwiched between rectifications) with equal number of neurons in each hidden layer functioned as synthetic-connectome networks (SCN). The main results describe networks with 128 neurons per hidden layer. We interpret the individual units in the SCN's as neurons with voltage

$$V_i = \sum_j s_{ij} + V_i^\text{rest} = \sum_j \sigma_j c_{ij} m_{ij} f(V_j) + V_i^\text{rest},$$

with presynaptic inputs s_{ij} and resting potentials V_i^rest. The connectome-constrained synapse strength, w_{ij}, is characterized by the adjacency matrix c_{ij}, the signs, σ_j, and the non-negative weight magnitudes m_{ij}. $c_{ij} = 1$ if the connection exists, else $c_{ij} = 0$. To respect Dale's law, the signs were tied to the presynaptic identity, j.

We identified the SCN's parameters σ_j, m_{ij}, and V_i^rest by task-optimization on handwritten digit classification (MNIST)[138]. We determined adjacency matrices, c_{ij}, for a given connectivity percentage using an iterative local pruning technique, the Lottery Ticket Hypothesis algorithm[139]. The algorithm decreases the connectivity percentage of the SCNs while maintaining high task accuracy.

We optimized SCNs and all simulated networks described below in Pytorch with stochastic gradient descent with adaptive moment estimation (ADAM with AMSGrad), learning rate 0.001, batch size 500, and an exponentially decaying learning rate decay factor of 0.5 per epoch. To constrain the weight magnitudes to stay non-negative, we clamped the values at zero after each optimization step (projected gradient descent). The parameters after convergence minimize the cross-entropy loss between the predicted and the groundtruth classes of the handwritten digits.

Simulated networks with known connectivity and unknown strength Simulated networks inherited connectivity, c_{ij}, and synapse signs, σ_j, from their respective SCN. In simulated networks, signs and connectivity were held fixed. Weight magnitudes, m_{ij}, and resting potentials, V_i^rest, were initialized randomly and task-optimized. Just like SCNs, simulated networks were trained on the MNIST handwritten digit classification task until convergence.

Simulated networks with known connectivity and known strength Alternatively, we imitate measurements of synaptic counts from the SCN's weight magnitudes:

$$\tilde{m}_{ij} = m_{ij}\epsilon_{ij} \text{ with } \epsilon_{ij} \sim \mathcal{U}(1-\sigma, 1+\sigma),$$

with multiplicative noise to imitate spurious measurements. We used $\sigma = 0.5$ for the main results. Weight magnitudes were initialized at the measurement, $\tilde{m}_{ij}$, and task-optimized on MNIST with the additional objective to minimize the squared distance between optimized and measured weight magnitudes, $\tilde{m}_{ij}$ (L2 constraint, Gaussian weight magnitude prior centered around the simulated network's initialization). We weighted the L2 constraint ten times higher than the cross-entropy objective to keep weight magnitudes of the simulated networks close to the noisy connectomic measurements. Resting potentials, V_i^{rest}, were again initialized randomly and task-optimized.

Measuring SCN-simulated network similarity SCN-simulated network similarity was measured by calculating the median Pearson's correlation of tuning responses (rectified voltages) of corresponding neurons in the SCN-simulated network pair. In each of the 6 hidden layers, $N = 100$ randomly-sampled neurons were used for comparison. Response tuning was measured over input stimuli from the MNIST test-set ($N = 10,000$ images). Results are medians over all hidden layers and over 25 SCN-simulated network pairs.

References

[1] Bargmann, C. I. & Marder, E. From the connectome to brain function. *Nature methods* **10**, 483–490 (2013).

[2] Scheffer, L. K. & Meinertzhagen, I. A. A connectome is not enough–what is still needed to understand the brain of drosophila? *Journal of Experimental Biology* **224**, jeb242740 (2021).

[3] Ammer, G., Vieira, R. M., Fendl, S. & Borst, A. Anatomical distribution and functional roles of electrical synapses in drosophila. *Current Biology* **32** (2022).

[4] Marder, E., Gutierrez, G. J. & Nusbaum, M. P. Complicating connectomes: electrical coupling creates parallel pathways and degenerate circuit mechanisms. *Developmental Neurobiology* **77**, 597–609 (2017).

[5] Marder, E. Neuromodulation of neuronal circuits: back to the future. *Neuron* **76**, 1–11 (2012).

[6] Mu, Y. *et al.* Glia accumulate evidence that actions are futile and suppress unsuccessful behavior. *Cell* **178**, 27–43 (2019).

[7] White, J. G., Southgate, E., Thomson, J. N., Brenner, S. *et al.* The structure of the nervous system of the nematode caenorhabditis elegans. *Philos Trans R Soc Lond B Biol Sci* **314**, 1–340 (1986).

[8] Jarrell, T. A. *et al.* The connectome of a decision-making neural network. *science* **337**, 437–444 (2012).

[9] Cook, S. J. *et al.* Whole-animal connectomes of both caenorhabditis elegans sexes. *Nature* **571**, 63–71 (2019).

[10] Witvliet, D. *et al.* Connectomes across development reveal principles of brain maturation. *Nature* **596**, 257–261 (2021).

[11] Zheng, Z. *et al.* A complete electron microscopy volume of the brain of adult drosophila melanogaster. *Cell* **174**, 730–743 (2018).

[12] Scheffer, L. K. *et al.* A connectome and analysis of the adult drosophila central brain. *Elife* **9**, e57443 (2020).

[13] Dorkenwald, S. *et al.* FlyWire: online community for whole-brain connectomics. *Nature Methods* **19**, 119–128 (2022).

[14] Hulse, B. K. *et al.* A connectome of the drosophila central complex reveals network motifs suitable for flexible navigation and context-dependent action selection. *eLife* **10**, e66039 (2021).

[15] Wanner, A. A., Genoud, C., Masudi, T., Siksou, L. & Friedrich, R. W. Dense em-based reconstruction of the interglomerular projectome in the zebrafish olfactory bulb. *Nature neuroscience* **19**, 816–825 (2016).

[16] Hildebrand, D. G. C. *et al.* Whole-brain serial-section electron microscopy in larval zebrafish. *Nature* **545**, 345–349 (2017).

[17] Svara, F. *et al.* Automated synapse-level reconstruction of neural circuits in the larval zebrafish brain. *Nature Methods* 1–10 (2022).

[18] Bock, D. D. *et al.* Network anatomy and in vivo physiology of visual cortical neurons. *Nature* **471**, 177–182 (2011).

[19] Helmstaedter, M. *et al.* Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature* **500**, 168–174 (2013). URL `http://dx.doi.org/10.1038/nature12346`.

[20] Oh, S. W. *et al.* A mesoscale connectome of the mouse brain. *Nature* **508**, 207–214 (2014).

[21] Motta, A. *et al.* Dense connectomic reconstruction in layer 4 of the somatosensory cortex. *Science* **366**, eaay3134 (2019).

[22] Loomba, S. *et al.* Connectomic comparison of mouse and human cortex. *Science* **377**, eabo0924 (2022).

[23] Kasthuri, N. *et al.* Saturated reconstruction of a volume of neocortex. *Cell* **162**, 648–661 (2015).

[24] Shapson-Coe, A. *et al.* A connectomic study of a petascale fragment of human cerebral cortex. *BioRxiv* (2021).

[25] Jabr, F. The connectome debate: Is mapping the mind of a worm worth it? *Scientific American* (2012). URL `https://www.scientificamerican.com/article/c-elegans-connectome/`.

[26] Cybenko, G. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems* **5**, 455–455 (1992).

[27] Leshno, M., Lin, V. Y., Pinkus, A. & Schocken, S. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural networks* **6**, 861–867 (1993).

[28] Montufar, G. F., Pascanu, R., Cho, K. & Bengio, Y. On the number of linear regions of deep neural networks. *Advances in neural information processing systems* **27** (2014).

[29] Goodfellow, I., Bengio, Y. & Courville, A. *Deep learning* (MIT press, 2016).

[30] Kornblith, S., Norouzi, M., Lee, H. & Hinton, G. Similarity of neural network representations revisited. In *International Conference on Machine Learning*, 3519–3529 (PMLR, 2019).

[31] Raghu, M., Unterthiner, T., Kornblith, S., Zhang, C. & Dosovitskiy, A. Do vision transformers see like convolutional neural networks? *Advances in Neural Information Processing Systems* **34**, 12116–12128 (2021).

[32] Reichardt, W. Autocorrelation, a principle for evaluation of sensory information by the central nervous system. *Principles of sensory communications* (1961).

[33] Barlow, H. B. & Levick, W. R. The mechanism of directionally selective units in rabbit's retina. *The Journal of Physiology* **178**, 477–504 (1965).

[34] Seung, H. S. How the brain keeps the eyes still. *Proc. Natl. Acad. Sci. USA* 13339–13344 (1996).

[35] Goldman, M. S. Memory without Feedback in a Neural Network. *Neuron* **61**, 621–634 (2009). URL `http://dx.doi.org/10.1016/j.neuron.2008.12.012`.

[36] Wang, P. Y., Sun, Y., Axel, R., Abbott, L. & Yang, G. R. Evolving the olfactory system with machine learning. *Neuron* **109**, 3879–3892 (2021).

[37] Biswas, T. & Fitzgerald, J. E. Geometric framework to predict structure from function in neural networks. *Physical Review Research* **4**, 023255 (2022).

[38] Bagherian, D. *et al.* Fine-grained system identification of nonlinear neural circuits. *arXiv preprint arXiv:2106.05400* (2021).

[39] Wengert, R. E. A simple automatic derivative evaluation program. *Communications of the ACM* **7**, 463–464 (1964).

[40] Paszke, A. *et al.* Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems* **32** (2019).

[41] Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986).

[42] Kingma, D. & Ba, J. Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations (ICRL)* (2014).

[43] LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *nature* **521**, 436–444 (2015).

[44] Yamins, D. L. & DiCarlo, J. J. Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience* **19**, 356–365 (2016).

[45] Yang, G. R., Joglekar, M. R., Song, H. F., Newsome, W. T. & Wang, X.-J. Task representations in neural networks trained to perform many cognitive tasks. *Nature neuroscience* **22**, 297–306 (2019).

[46] Mano, O., Creamer, M. S., Badwan, B. A. & Clark, D. A. Predicting individual neuron responses with anatomically constrained task optimization. *Current Biology* **31**, 4062–4075 (2021).

[47] Sandbrink, K. J. *et al.* Task-driven hierarchical deep neural network models of the proprioceptive pathway. *bioRxiv* (2020).

[48] Shinomiya, K. *et al.* Comparisons between the ON- and OFF-edge motion pathways in the Drosophila brain. *Elife* **8**, 2431 (2019).

[49] Takemura, S.-y. *et al.* The comprehensive connectome of a neural substrate for 'ON' motion detection in Drosophila. *eLife* **6**, 1–16 (2017).

[50] Rivera-Alba, M. *et al.* Wiring economy and volume exclusion determine neuronal placement in the drosophila brain. *Current Biology* **21**, 2000–2005 (2011).

[51] Krapp, H. G., Hengstenberg, B. & Hengstenberg, R. Dendritic structure and receptive-field organization of optic flow processing interneurons in the fly. *Journal of neurophysiology* **79**, 1902–1917 (1998).

[52] Borst, A., Haag, J. & Reiff, D. F. Fly motion vision. *Annual Review of Neuroscience* **33**, 49–70 (2010). URL `https://doi.org/10.1146%2Fannurev-neuro-060909-153155`.

[53] Strother, J. A. *et al.* The emergence of directional selectivity in the visual motion pathway of drosophila. *Neuron* **94**, 168–182 (2017).

[54] Gruntman, E., Romani, S. & Reiser, M. B. Simple integration of fast excitation and offset, delayed inhibition computes directional selectivity in Drosophila. *Nature Neuroscience* **21**, 250–257 (2018). URL `http://dx.doi.org/10.1038/s41593-017-0046-4`.

[55] Gruntman, E., Romani, S. & Reiser, M. B. The computation of directional selectivity in the drosophila off motion pathway. *Elife* **8**, e50706 (2019).

[56] Borst, A., Haag, J. & Mauss, A. S. How fly neurons compute the direction of visual motion. *Journal of Comparative Physiology A* **206**, 109–124 (2020).

[57] Joesch, M., Schnell, B., Raghu, S. V., Reiff, D. F. & Borst, A. ON and OFF pathways in drosophila motion vision. *Nature* **468**, 300–304 (2010). URL `https://doi.org/10.1038%2Fnature09545`.

[58] Clark, D. A., Bursztyn, L., Horowitz, M. A., Schnitzer, M. J. & Clandinin, T. R. Defining the computational structure of the motion detector in drosophila. *Neuron* **70**, 1165–1177 (2011).

[59] Strother, J. A., Nern, A. & Reiser, M. B. Direct observation of on and off pathways in the drosophila visual system. *Current Biology* **24**, 976–983 (2014). URL `http://dx.doi.org/10.1016/j.cub.2014.03.017`.

[60] Behnia, R., Clark, D. A., Carter, A. G., Clandinin, T. R. & Desplan, C. Processing properties of ON and OFF pathways for drosophila motion detection. *Nature* **512**, 427–430 (2014).

[61] Borst, A. & Helmstaedter, M. Common circuit design in fly and mammalian motion vision. *Nature Neuroscience* **18**, 1067 (2015).

[62] Maisak, M. S. *et al.* A directional tuning map of Drosophila elementary motion detectors. *Nature* **500**, 212–216 (2013).

[63] Yonehara, K. & Roska, B. Motion detection: neuronal circuit meets theory. *Cell* **154**, 1188–1189 (2013).

[64] Fisher, Y. E., Silies, M. & Clandinin, T. R. Orientation Selectivity Sharpens Motion Detection in Drosophila. *Neuron* **88**, 390–402 (2015). URL `http://dx.doi.org/10.1016/j.neuron.2015.09.033`.

[65] Fischbach, K. F. & Dittrich, A. P. The optic lobe of Drosophila melanogaster. I. A Golgi analysis of wild-type structure. *Cell and Tissue Research* **258**, 441–475 (1989).

[66] Nern, A., Pfeiffer, B. D. & Rubin, G. M. Optimized tools for multicolor stochastic labeling reveal diverse stereotyped cell arrangements in the fly visual system. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E2967–76 (2015).

[67] Takemura, S.-y. *et al.* A visual motion detection circuit suggested by Drosophila connectomics. *Nature* **500**, 175–181 (2013). URL `http://dx.doi.org/10.1038/nature12450`.

[68] Takemura, S.-y. *et al.* Synaptic circuits and their variations within different columns in the visual system of Drosophila . *Proceedings of the National Academy of Sciences* **112**, 13711–13716 (2015).

[69] Shinomiya, K., Nern, A., Meinertzhagen, I. A., Plaza, S. M. & Reiser, M. B. Neuronal circuits integrating visual motion information in drosophila melanogaster. *Current Biology* **32**, 3529–3544 (2022).

[70] Meier, M. & Borst, A. Extreme Compartmentalization in a Drosophila Amacrine Cell. *Current Biology* **29**, 1545–1550.e2 (2019). URL `https://doi.org/10.1016/j.cub.2019.03.070`.

[71] Hahnloser, R. & Seung, H. S. Permitted and forbidden sets in symmetric threshold-linear networks. *Advances in neural information processing systems* **13** (2000).

[72] Curto, C., Degeratu, A. & Itskov, V. Flexible memory networks. *Bulletin of mathematical biology* **74**, 590–614 (2012).

[73] Liu, T. X., Davoudian, P. A., Lizbinski, K. M. & Jeanne, J. M. Connectomic features underlying diverse synaptic connection strengths and subcellular computation. *Current Biology* **32**, 559–569.e5 (2022). URL `https://www.sciencedirect.com/science/article/pii/S0960982221016420`.

[74] Götz, K. G. Optomotorische untersuchung des visuellen systems einiger augenmutanten der fruchtfliege drosophila. *Kybernetik* **2**, 77–92 (1964).

[75] Ready, D. F., Hanson, T. E. & Benzer, S. Development of the drosophila retina, a neurocrystalline lattice. *Developmental biology* **53**, 217–240 (1976).

[76] Yamins, D. L. K. *et al.* Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences* **111**, 8619–8624 (2014).

[77] Butler, D. J., Wulff, J., Stanley, G. B. & Black, M. J. A Naturalistic Open Source Movie for Optical Flow Evaluation (Sintel). *Eccv* 611–625 (2012).

[78] Braitenberg, V. Patterns of projection in the visual system of the fly. i. retina-lamina projections. *Experimental Brain Research* **3**, 271–298 (1967).

[79] Werbos, P. J. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE* **78**, 1550–1560 (1990).

[80] Marder, E. & Taylor, A. L. Multiple models to capture the variability in biological neurons and networks. *Nature Neuroscience* **14**, 133–138 (2011).

[81] Schaeffer, R., Khona, M. & Fiete, I. No free lunch from deep learning in neuroscience: A case study through models of the entorhinal-hippocampal circuit. *bioRxiv* (2022).

[82] Maheswaranathan, N., Williams, A., Golub, M., Ganguli, S. & Sussillo, D. Universality and individuality in neural dynamics across large populations of recurrent networks. In Wallach, H. *et al.* (eds.) *Advances in Neural Information Processing Systems*, vol. 32 (Curran Associates, Inc., 2019). URL `https://proceedings.neurips.cc/paper/2019/file/5f5d472067f77b5c88f69f1bcfda1e08-Paper.pdf`.

[83] Peretz, A. *et al.* The light response of drosophila photoreceptors is accompanied by an increase in cellular calcium: effects of specific mutations. *Neuron* **12**, 1257–1267 (1994).

[84] Reiff, D. F., Plett, J., Mank, M., Griesbeck, O. & Borst, A. Visualizing retinotopic half-wave rectified input to the motion detection circuitry of drosophila. *Nature neuroscience* **13**, 973–978 (2010).

[85] Freifeld, L., Clark, D. A., Schnitzer, M. J., Horowitz, M. A. & Clandinin, T. R. Gabaergic lateral interactions tune the early stages of visual processing in drosophila. *Neuron* **78**, 1075–1089 (2013).

[86] Silies, M. *et al.* Modular use of peripheral input channels tunes motion-detecting circuitry. *Neuron* **79**, 111–127 (2013).

[87] Meier, M. *et al.* Neural circuit components of the drosophila off motion vision pathway. *Current Biology* **24**, 385–392 (2014).

[88] Fisher, Y. E. *et al.* A Class of Visual Neurons with Wide-Field Properties Is Required for Local Motion Detection. *Current Biology* **25**, 3178–3189 (2015). URL `http://dx.doi.org/10.1016/j.cub.2015.11.018`.

[89] Hardie, R. C. & Juusola, M. Phototransduction in drosophila. *Current opinion in neurobiology* **34**, 37–45 (2015).

[90] Leonhardt, A. *et al.* Asymmetry of drosophila on and off motion detectors enhances real-world velocity estimation. *Nature neuroscience* **19**, 706–715 (2016).

[91] Yang, H. H. *et al.* Subcellular imaging of voltage and calcium signals reveals neural processing in vivo. *Cell* **166**, 245–257 (2016).

[92] Serbe, E., Meier, M., Leonhardt, A. & Borst, A. Comprehensive characterization of the major presynaptic elements to the drosophila off motion detector. *Neuron* **89**, 829–841 (2016).

[93] Arenz, A., Drews, M. S., Richter, F. G., Ammer, G. & Borst, A. The Temporal Tuning of the Drosophila Motion Detectors Is Determined by the Dynamics of Their Input Elements. *Current Biology* **27**, 929–944 (2017). URL `http://dx.doi.org/10.1016/j.cub.2017.01.051`.

[94] Ramos-Traslosheros, G. & Silies, M. The physiological basis for contrast opponency in motion computation in drosophila. *Nature communications* **12**, 1–16 (2021).

[95] Gruntman, E., Reimers, P., Romani, S. & Reiser, M. B. Non-preferred contrast responses in the drosophila motion pathways reveal a receptive field structure that explains a common visual illusion. *Current Biology* **31**, 5286–5298 (2021).

[96] Groschner, L. N., Malis, J. G., Zuidinga, B. & Borst, A. A biophysical account of multiplication by a single neuron. *Nature* **603**, 119–123 (2022).

[97] Ketkar, M. D. *et al.* First-order visual interneurons distribute distinct contrast and luminance information across on and off pathways to achieve stable behavior. *Elife* **11**, e74937 (2022).

[98] Borst, A. & Helmstaedter, M. Common circuit design in fly and mammalian motion vision. *Nature Neuroscience* **18**, 1067–1076 (2015). URL `https://doi.org/10.1038%2Fnn.4050`.

[99] Gjorgjieva, J., Sompolinsky, H. & Meister, M. Benefits of pathway splitting in sensory coding. *Journal of Neuroscience* **34**, 12127–12144 (2014).

[100] Serbe, E. Analysis of the neural circuit underlying the detection of visual motion in drosophila melanogaster (2016).

[101] Ammer, G., Leonhardt, A., Bahl, A., Dickson, B. J. & Borst, A. Functional Specialization of Neural Input Elements to the Drosophila on Motion Detector. *Current Biology* **25**, 2247–2253 (2015). URL `http://dx.doi.org/10.1016/j.cub.2015.07.014`.

[102] Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using umap. *Nature biotechnology* **37**, 38–44 (2019).

[103] Csáji, B. C. *et al.* Approximation with artificial neural networks. *Faculty of Sciences, Etvs Lornd University, Hungary* **24**, 7 (2001).

[104] Li, Y., Yosinski, J., Clune, J., Lipson, H. & Hopcroft, J. E. Convergent learning: Do different neural networks learn the same representations? In *International Conference on Learning Representations (ICLR)* (2016).

[105] Turner, E., Dabholkar, K. V. & Barak, O. Charting and navigating the space of solutions for recurrent neural networks. *Advances in Neural Information Processing Systems* **34**, 25320–25333 (2021).

[106] Tanaka, H. *et al.* From deep learning to mechanistic understanding in neuroscience: the structure of retinal prediction. *Advances in neural information processing systems* **32** (2019).

[107] Raghu, M., Gilmer, J., Yosinski, J. & Sohl-Dickstein, J. Svcca: Singular vector canonical correlation analysis for deep learning dynamics and interpretability. *Advances in neural information processing systems* **30** (2017).

[108] Barrett, D. G., Morcos, A. S. & Macke, J. H. Analyzing biological and artificial neural networks: challenges with opportunities for synergy? *Current opinion in neurobiology* **55**, 55–64 (2019).

[109] Ammer, G., Leonhardt, A., Bahl, A., Dickson, B. J. & Borst, A. Functional specialization of neural input elements to the drosophila on motion detector. *Current Biology* **25**, 2247–2253 (2015).

[110] Strother, J. A. *et al.* Behavioral state modulates the on visual motion pathway of drosophila. *Proceedings of the National Academy of Sciences* **115**, E102–E111 (2018).

[111] Card, G. & Dickinson, M. H. Visually mediated motor planning in the escape response of drosophila. *Current Biology* **18**, 1300–1307 (2008).

[112] Ofstad, T. A., Zuker, C. S. & Reiser, M. B. Visual place learning in drosophila melanogaster. *Nature* **474**, 204–207 (2011).

[113] Ribeiro, I. M. *et al.* Visual projection neurons mediating directed courtship in drosophila. *Cell* **174**, 607–621 (2018).

[114] Schretter, C. E. *et al.* Cell types and neuronal circuitry underlying female aggression in drosophila. *Elife* **9**, e58942 (2020).

[115] Fitzgerald, J. E. & Clark, D. A. Nonlinear circuits for naturalistic visual motion estimation. *Elife* **4**, e09123 (2015).

[116] Leonhardt, A., Meier, M., Serbe, E., Eichner, H. & Borst, A. Neural mechanisms underlying sensitivity to reverse-phi motion in the fly. *PloS one* **12**, e0189019 (2017).

[117] Fisher, D., Olasagasti, I., Tank, D. W., Aksay, E. R. & Goldman, M. S. A modeling framework for deriving the structural and functional architecture of a short-term memory microcircuit. *Neuron* **79**, 987–1000 (2013).

[118] Kriegeskorte, N. Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annual Review of Vision Science* **1**, 417–446 (2015).

[119] Tschopp, F. D., Reiser, M. B. & Turaga, S. C. A Connectome Based Hexagonal Lattice Convolutional Network Model of the Drosophila Visual System. *arXiv preprint arXiv:1806.04793* (2018). URL `http://arxiv.org/abs/1806.04793`. `1806.04793`.

[120] Mi, L. *et al.* Connectome-constrained latent variable model of whole-brain neural activity. In *International Conference on Learning Representations* (2022). URL `https://openreview.net/forum?id=CJzi3dRlJE-`.

[121] Abbott, L. F. *et al.* The mind of a mouse. *Cell* **182**, 1372–1376 (2020).

[122] Ríos, V. L. *et al.* Neuromechfly, a neuromechanical model of adult drosophila melanogaster. *bioRxiv* (2021).

[123] Merel, J. *et al.* Deep neuroethology of a virtual rodent. In *International Conference on Learning Representations* (2020).

[124] Tuthill, J. C., Nern, A., Rubin, G. M. & Reiser, M. B. Wide-field feedback neurons dynamically tune early visual processing. *Neuron* **82**, 887–895 (2014).

[125] Tuthill, J. C., Nern, A., Holtz, S. L., Rubin, G. M. & Reiser, M. B. Contributions of the 12 Neuron Classes in the Fly Lamina to Motion Vision. *Neuron* **79**, 128–140 (2013). URL `http://dx.doi.org/10.1016/j.neuron.2013.05.024`.

[126] AlQuraishi, M. & Sorger, P. K. Differentiable biology: using deep learning for biophysics-based and data-driven modeling of molecular mechanisms. *Nature methods* **18**, 1169–1180 (2021).

[127] Chen, R. T. Q., Rubanova, Y., Bettencourt, J. & Duvenaud, D. K. Neural ordinary differential equations. In Bengio, S. *et al.* (eds.) *Advances in Neural Information Processing Systems*, vol. 31 (Curran Associates, Inc., 2018). URL `https://proceedings.neurips.cc/paper/2018/file/69386f6bb1dfed68692a24c8686939b9-Paper.pdf`.

[128] Shi, X. *et al.* Convolutional lstm network: A machine learning approach for precipitation nowcasting. In Cortes, C., Lawrence, N., Lee, D., Sugiyama, M. & Garnett, R. (eds.) *Advances in Neural Information Processing Systems*, vol. 28 (Curran Associates, Inc., 2015). URL `https://proceedings.neurips.cc/paper/2015/file/07563a3fe3bbe7e3ba84431ad9d055af-Paper.pdf`.

[129] Fukushima, K. & Miyake, S. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and cooperation in neural nets*, 267–285 (Springer, 1982).

[130] LeCun, Y. *et al.* Backpropagation applied to handwritten zip code recognition. *Neural computation* **1**, 541–551 (1989).

[131] Riesenhuber, M. & Poggio, T. Hierarchical models of object recognition in cortex. *Nature neuroscience* **2**, 1019–1025 (1999).

[132] Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, 1097–1105 (2012).

[133] Miller, K. D. & Fumarola, F. Mathematical equivalence of two common forms of firing rate models of neural networks. *Neural computation* **24**, 25–31 (2012).

[134] Butler, D. J., Wulff, J., Stanley, G. B. & Black, M. J. A naturalistic open source movie for optical flow evaluation. In *European conference on computer vision*, 611–625 (Springer, 2012).

[135] Dosovitskiy, A. *et al.* Flownet: Learning optical flow with convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, 2758–2766 (2015).

[136] Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).

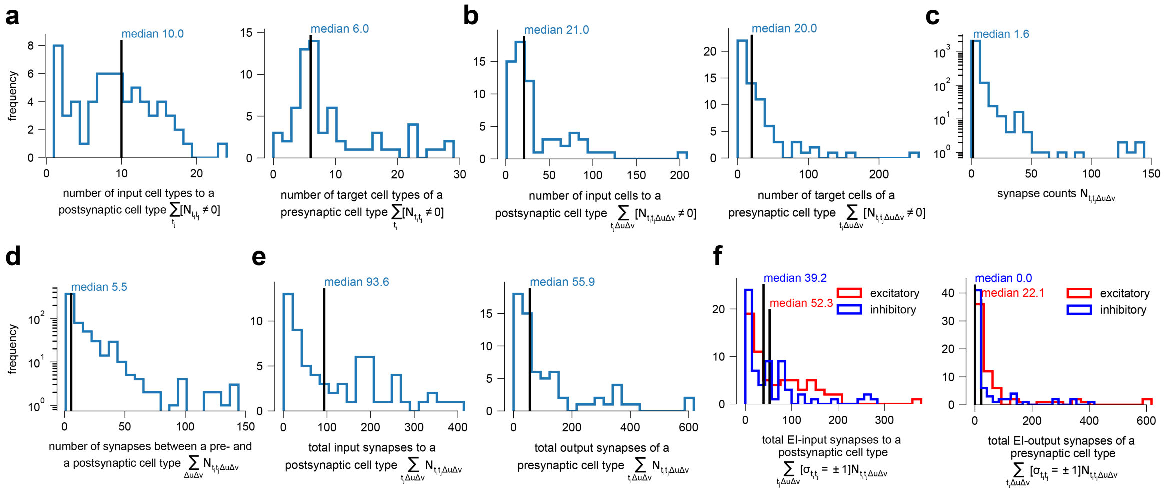
[137] Walker, E. Y. *et al.* Inception loops discover what excites neurons most using deep predictive models. *Nature Neuroscience* **22**, 2060–2065 (2019).

[138] LeCun, Y., Cortes, C. & Burges, C. Mnist handwritten digit database. *ATT Labs [Online]. Available: http://yann.lecun.com/exdb/mnist* **2** (2010).

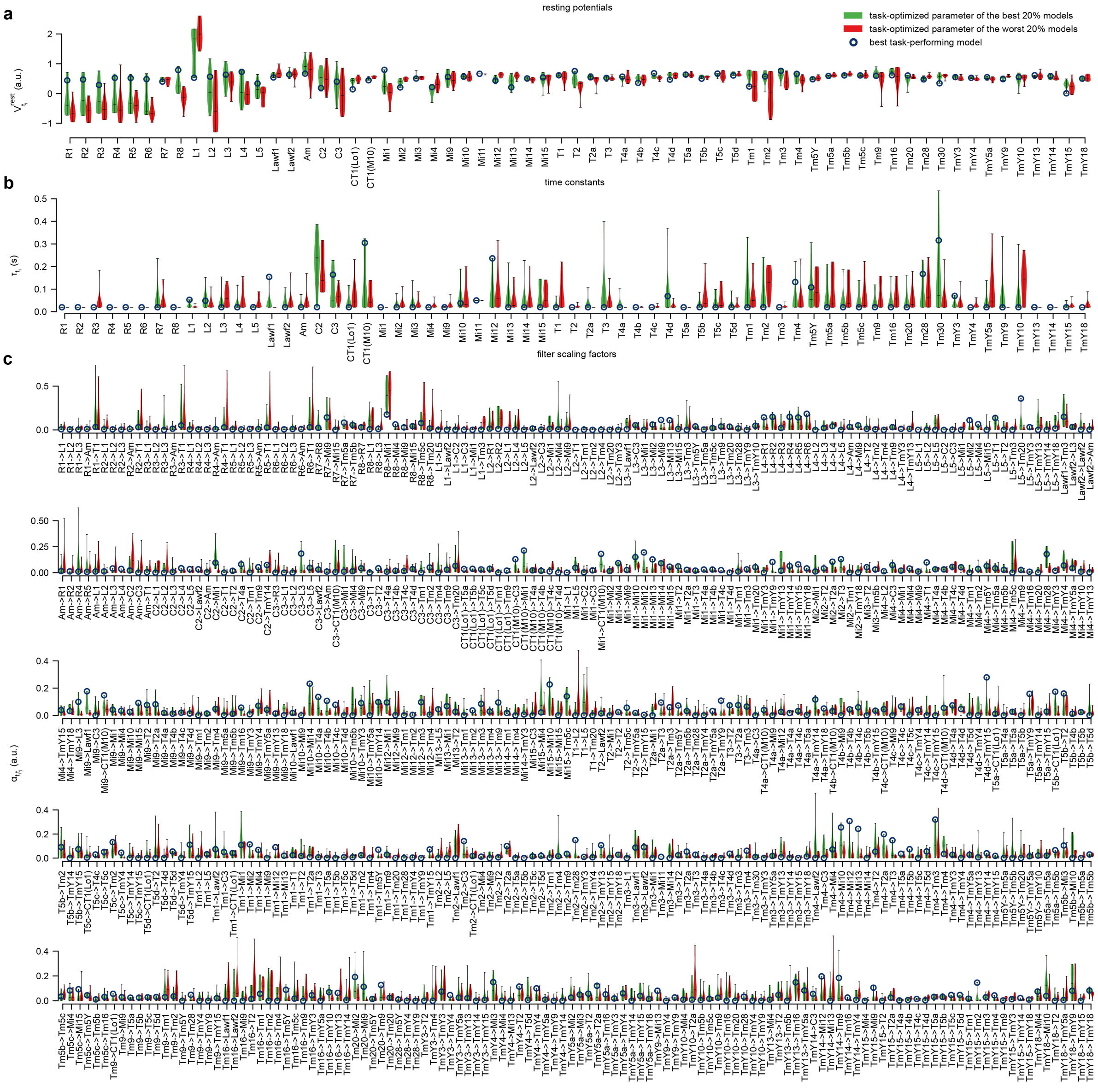
[139] Frankle, J. & Carbin, M. The lottery ticket hypothesis: Finding sparse, trainable neural networks. *arXiv preprint arXiv:1803.03635* (2018).

[140] He, K., Zhang, X., Ren, S. & Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, 1026–1034 (2015).

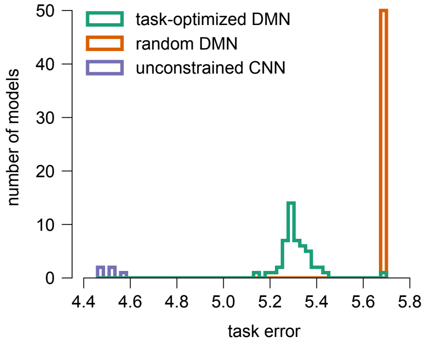
Extended Data Figures



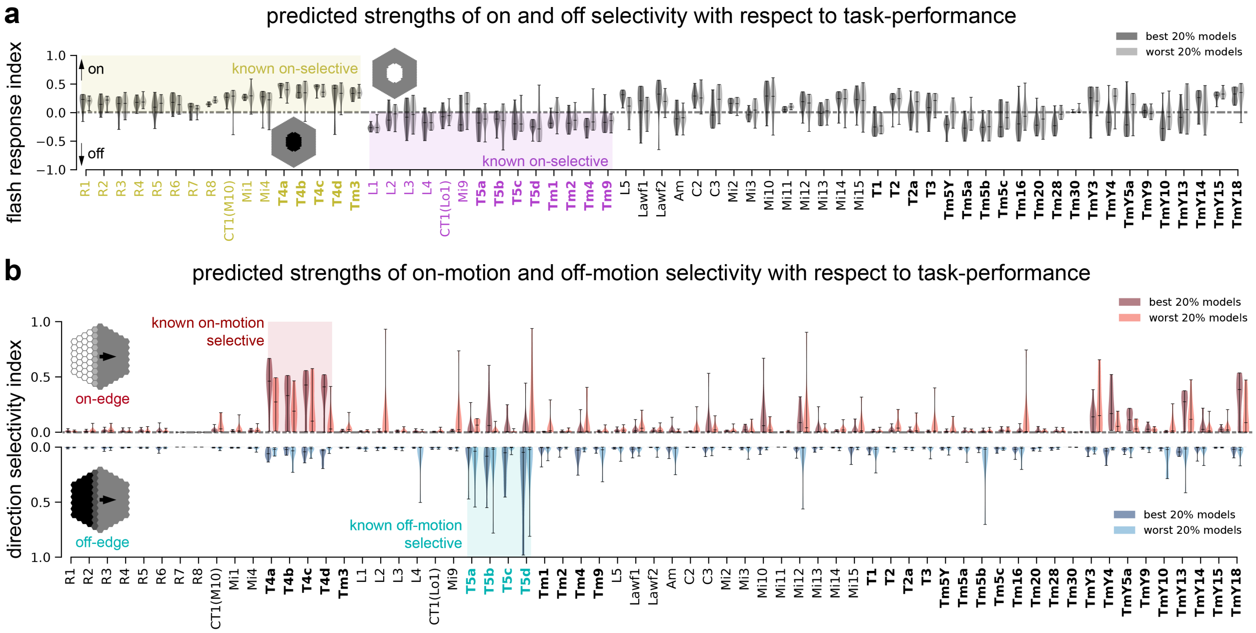
Extended Data Figure 1: **Statistics of derived connectome.** **(a)** (left) Half of the 65 cell types receive input from more than ten other cell types, while the other half receives input from less than ten. (right) Half of the 65 cell types project onto more then six other cell types, while the other half projects onto less than six. **(b)** (left) Half of the 65 cell types receive input from 21 up to 200 cells, while the other half receives input from less than 21 cells. (right) Half of the 65 cell types project output onto 20 up to 200 cells, while the other half projects output onto less than 20 cells. **(c)** Half of the connections are characterized by less than 1.6 synapses while the other half are characterized by 1.6 up to hundreds of synapses. **(d)** A pair of presynaptic and postsynaptic cell type is connected by 5.5 synapses in half of the cases and by more than 5.5 up to hundreds in the other half of the cases. **(e)** (left) Half of the 65 cell types receive input from less than 93.6 synapses and the other half between 93.6 to 400 synapses. (right) Half of the 65 cell types project less than 55.9 synapses and the other half projects between 55.9 to 600 synapses. **(f)** Separating (e) into excitatory and inihibitory synapses, (left) we see that half of the 65 cell types receive excitatory inputs from less than 52.3 synapses and the other half from 52.3 to hundreds. Half of the 65 cell types receive inhibitory inputs from less than 39.2 synapses and the other half from 39.2 to hundreds. (right) Half of the 65 cell types project less than 22.1 excitatory synapses and the other half from 22.1 to hundreds. At least half of the 65 cell types project no inhibitory synapses and the rest project between zero to hundreds.



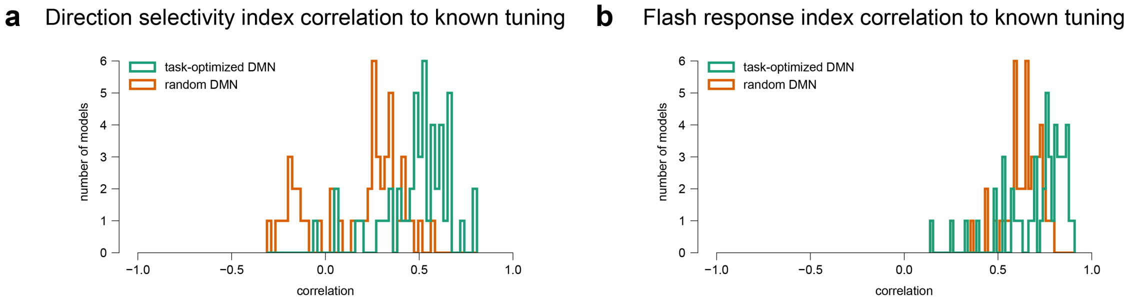
Extended Data Figure 2: **Statistics of learned parameters of best 20% models vs. worst 20% models. (a)** Task-optimized resting potentials. **(b)** Task-optimized time constants. **(c)** Task-optimized filter scaling factors.



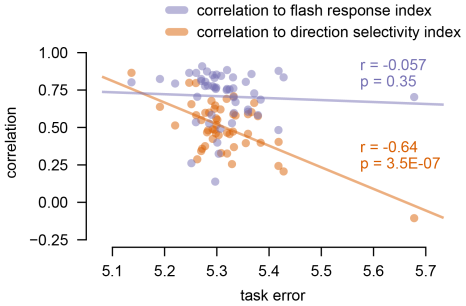
Extended Data Figure 3: **Models with random parameter configurations consistently performed worse on the task than task-optimized models.** Task error distributions over ensembles. **(green)** The task error of 50 DMNs distributes between 5.1 and 5.5 (a.u.) after optimization. **(red)** The task error of 50 random DMNs collapses at 5.7 despite optimization of the decoders. **(blue)** An unconstrained CNN with 414,666 parameters reaches a task error between 4.5 and 4.6 (here 5 models).



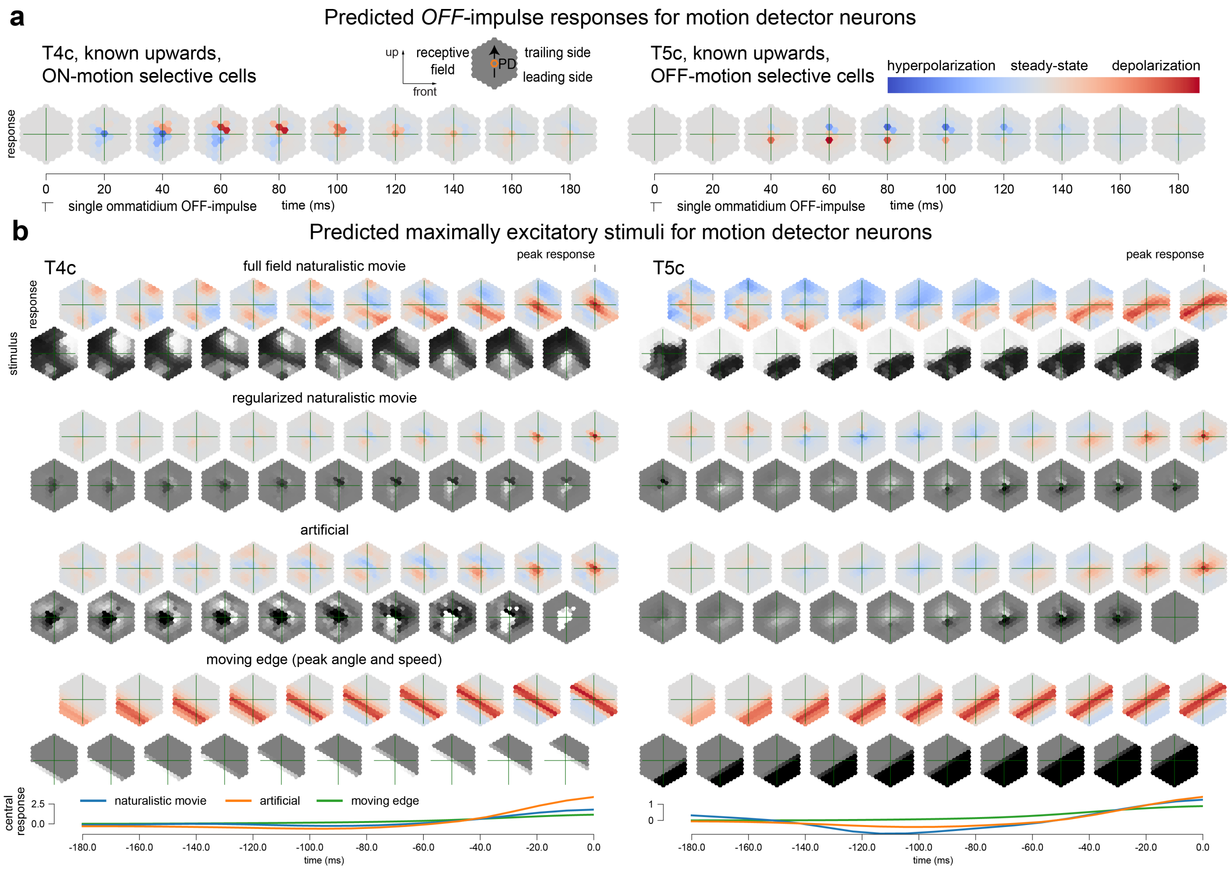
Extended Data Figure 4: **Predicted tuning with respect to task-performance. (a)** Flash response index computed as the max-abs-scaled peak response to an off flash subtracted from the max-abs-scaled peak response to an on flash – both of approximately 35° radius and presented for 1s after 2 seconds of grey input. Values above 0 indicate on-polarity, values below zero indicate off-polarity. Known on-polar and off-polar cell types are colored in yellow and magenta. **(b)** Single cell type direction selectivity of best 20% task-performing models versus worst 20% task-performing models of an ensemble of 50 models as a result of peak voltage responses in central columns to on-edges and off-edges moving towards all possible directions on grey background (Equation 9). The bolded cell types are those which optic flow is decoded from.



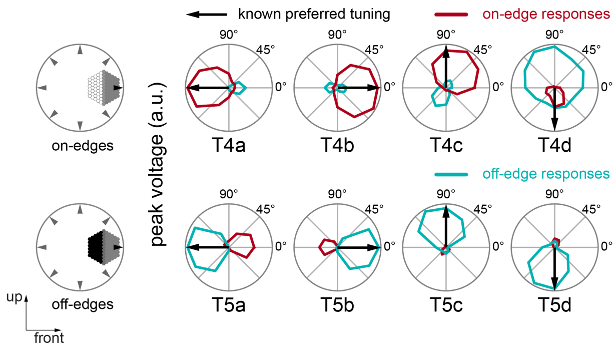
Extended Data Figure 5: **Models with random parameter configurations predict the known direction selectivity indices and flash response indices worse than task-optimized models.** We correlate predicted tuning metrics from each model to the known tuning properties to answer if models with random parameter configurations lead to similarly accurate tuning predictions as task-optimized models. **(a)** Task-optimized DMNs (green) predict more accurate direction selectivity indices than randomly parametrized models (red) ($P = 2 \times 10^{-11}$, Mann-Whitney-U). **(b)** Task-optimized DMNs (green) predict more accurate flash response indices than randomly parametrized models (red) ($P = 5.3 \times 10^{-5}$, Mann-Whitney-U).



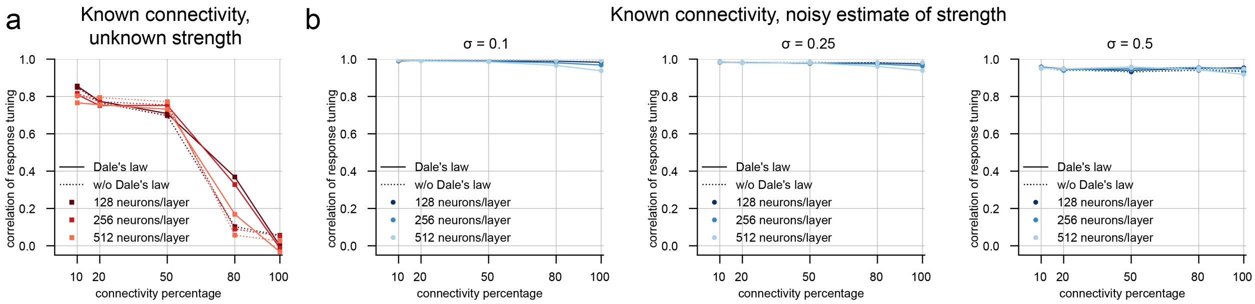
Extended Data Figure 6: **Better task performing models predict motion tuning neurons better.** We correlate predicted tuning metrics from each model to the known tuning properties to understand when better performing models give us better tuning predictions. **(orange)** When correlating the direction selectivity index of each model to the binary known properties for T4 and T5 and their input cell types, we find that this correlation is higher for better performing models. **(magenta)** While the models predicted the known contrast preferences generally well, the correlation of flash response index to the binary known contrast preferences of 31 cell types did not significantly increase with better performing models.



Extended Data Figure 7: **Spatio-temporal receptive fields mapped with OFF-impulses and maximally excitatory stimuli. (a)** Spatio-temporal receptive field mapping with single ommatidium OFF-impulses. **(b)** Maximally excitatory stimuli and baseline-subtracted responses. Including full-field naturalistic, regularized naturalistic, artificial, and moving edge stimuli and responses. Moving edge angle and speed maximize the central cell peak response. Artificial stimuli are optimized from initial noise to maximize the central cell activity using gradient ascent plus full-field regularization towards grey. The last row shows the baseline-subtracted central cell responses. Peak central cell responses at time point zero.



Extended Data Figure 8: **Motion tuning predictions for T4 and T5 subtypes to preferred and null contrast edges in the best-task-performing model.**



Extended Data Figure 9: **Investigating the role of sparse connectivity with synthetic networks for MNIST handwritten digit recognition.** **(a)** Median hidden-layer response correlation as a function of synthetic network connectivity percentage for task and connectome constrained models that had access to only connectivity information but not connection strength. **(b)** Median hidden-layer response correlation as a function of synthetic network connectivity percentage for task and connectome constrained models with access to noisy estimates of connection strength (multiplicative noise levels of $\sigma = 0.1, \sigma = 0.25$, and $\sigma = 0.5$, respectively). Connectome constrained models were task optimized with a soft (L2) constraint with the noisy connectomic measurements.

Supplementary Information

Supplementary Note 1 Probabilistic model for automatic construction of connectome

The EM-datasets primarily contain the lamina projections, medulla (FIB-25), lobula and lobula plate (FIB-19) cells, and the important cell types of the primary motion detection circuit (T4, T5). In total, they contain 1801 neurons (702 from FIB-25 and 1099 from FIB-19), with hand-annotated positions available for 830 of these neurons (SI Figure 1). To accurately localize the remaining neurons and synapses and to derive cell-type connectivity (Fig. 1b), we build a probabilistic expectation maximization algorithm that takes synaptic connection statistics, projected synapse center-of-mass clusters and existing column annotations into account. We verified the quality of our reconstruction, concluding that even in the absence of 90% of the hand-annotations available to us, we could accurately position the majority of the neurons in our circuit reconstruction (SI Table 1, SI Figure 2). In the absence of ground-truth annotations, we verified the quality of our reconstruction by the *recovery* and *consistency* rates (Table 1). The *recovery* rate is defined as the ratio of reference positions successfully recovered by our algorithm after removing a random proportion of reference positions from the data. For each 10% of the reference positions removed, on average, only 2.5% are not correctly recovered. The *consistency* rate is defined as the fraction of neurons obtaining the same position between evaluations of the algorithm starting with a different fraction of reference positions. For each 10% of the reference positions removed, on average, an additional 3.9% of neurons are not consistently estimated. We found that even just 10% (83 positions) of the available ground truth was sufficient to robustly position the majority of the neurons (64.3%, 534 positions) into the correct columns, and annotate 48% (865 neurons) perfectly consistent.

Probabilistic expectation maximization for unassigned neurons. Each neuron is either annotated in the dataset (in $\mathcal{K}$), assigned to a position (in $\mathcal{A}$) by our algorithm, or still under evaluation (in set $\mathcal{U}$). Iteratively, the EM-expectation step updates the normal distribution ($\mu_{s,t,(y,x)}$, $\sigma_{s,t,(y,x)}$) of expected synapse counts between neurons, while the EM-maximization step updates the positions (y, x) of all neurons not yet assigned to a column (set $\mathcal{U}$).

Synapse center-of-mass as a neuron column position proxy. For offset assignment, we take the center-of-mass of all synapses belonging (pre- or postsynaptic) to a neuron into account. These are generally a more useful hint than the physical location of the cell body, as the cell bodies are mostly positioned on the side of the neuropiles, and not near the column to which the neuron belongs to, but most cells have a majority of synapses in close proximity to their own column. Since many neurons span more than one layer in a neuropile, or even multiple neuropiles, we first group all synapses per neuron into clusters, and then assign the center-of-mass of these clusters to one of up to $N = 5$ super-clusters (approximately matching the medulla, lobula and lobula plate). Clustering is done via k-means with the ideal number of clusters determined by silhouette scores. The super-clusters allow to project 3D synapse coordinates onto a retinotopic 2D hexagonal lattice with a simple projection and affine transformation.

Hybrid cost-model for neuron-position likelihood estimates. Prior knowledge about the normal distribution ($\mu_{s,t,(y,x)}$, $\sigma_{s,t,(y,x)}$) of expected synapse counts between neurons from existing annotations is required to express the probability of any cell specimen c to be located at position (y, x). This metric correlates a neuron to all pre- and postsynaptic neurons it is connected to, of which some already have a fixed, known position. Thereby, the neighbouring neurons with known position (in $\mathcal{K}$ and $\mathcal{A}$) contribute to stabilize

Reference positions	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	average
Recovery rate	**0.870**	0.849	0.820	0.807	0.790	0.754	0.730	0.707	0.669	0.643	
Recovery rate delta		0.020	0.029	0.013	0.017	0.036	0.024	0.023	0.039	0.025	**0.025**
Consistency rate		0.790	0.728	0.701	0.682	0.638	0.609	0.552	0.546	0.480	
Consistency rate delta			0.062	0.027	0.019	0.044	0.029	0.057	0.006	0.066	**0.039**

Table 1: Recovery and consistency of columnar cell position estimation.

the probabilities of unassigned neurons (in set $\mathcal{U}$). For columnar, spatially repeated neurons, we can also assume that only one neuron per position is present. This prior rapidly discounts the number of possible positions an unassigned neuron can have, each time another neuron becomes assigned (moves from $\mathcal{U}$ to $\mathcal{A}$).

Fusion of sepratately evaluated datasets. Since FIB-19 and FIB-25 are evaluated separately, we have to combine the estimated parameters of both models to a single, coherent model. The datasets overlap partially, in terms of the neurpoiles and cell types covered, and our method therefore fuses the model by always taking the larger estimated parameter.

$$\mu_{s,t,(y,x)} = \max(\mu^{\text{FIB-19}}_{s,t,(y,x)},\ \mu^{\text{FIB-25}}_{s,t,(y,x)}) \tag{6}$$

This method of model fusion only underestimates the number of synapses between two neurons if they have connections in two different neuropiles, and if each neuropile is exclusively covered by only one dataset.

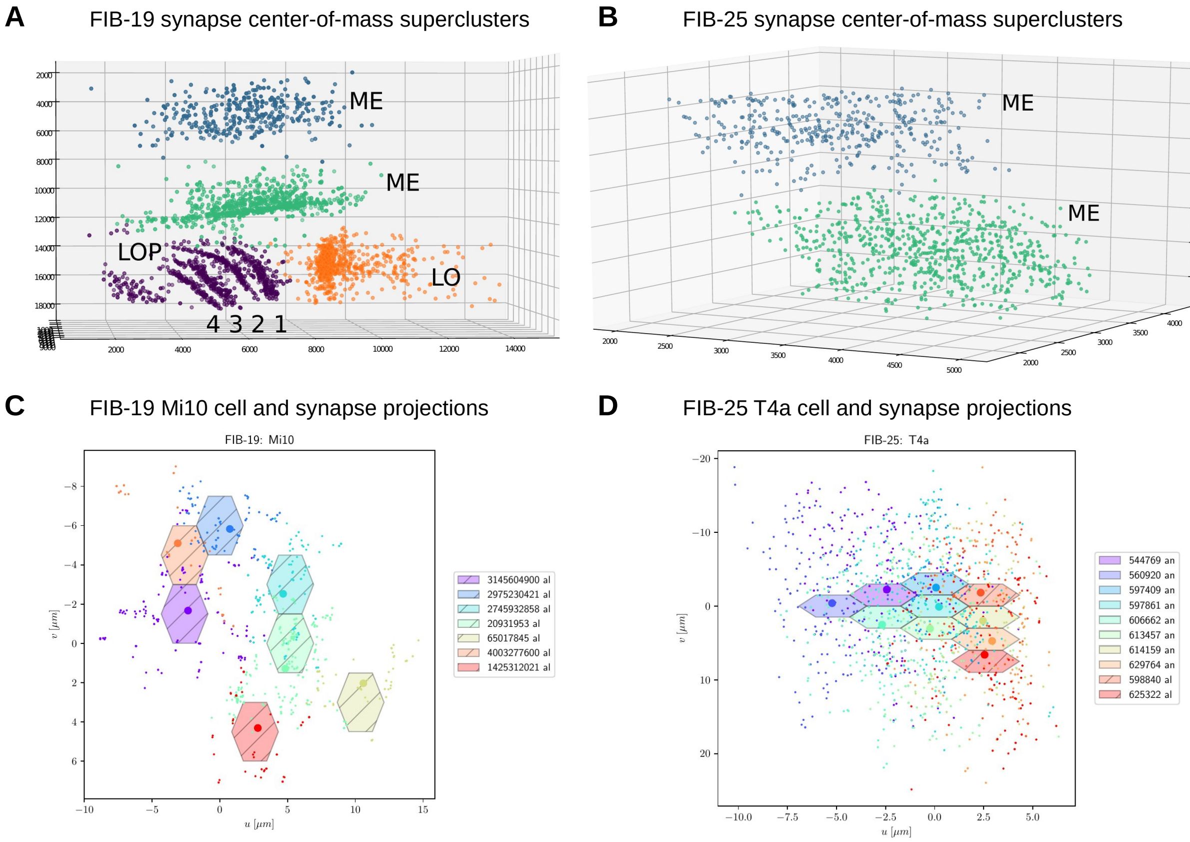
Pruning spurious synapses. Some automatic annotations, which are not proof read in the FlyEM DVID data, contain a large number of autapses per neuron on most neuron types, arising from wrongly detected synapses in the cell bodies themselves. Additionally, there are many statistically insignificant single synapses left from the assignment algorithm. We imposed the following additional filter on our estimated model parameters, to remove both autapses and spurious connections with less than one synapse on average.

$$\mu_{s,t,(y,x)} = \begin{cases} \emptyset & \text{if } s = t \land (y,x) = (0,0) \\ \emptyset & \text{if } \mu_{s,t,(y,x)} < 1 \\ \mu_{s,t,(y,x)} & \text{otherwise} \end{cases} \tag{7}$$

Finally, neuron types without connections and synapses with either missing target or source can be removed. The resulting mean synapse counts $\mu_{s,t,(y,x)}$ form the convolutional filters for our simulation.

Pos. removed [%]	0	10	20	30	40	50	60	70	80	90	100
Reference positions	830	747	664	581	498	415	332	249	166	83	0
Combined	1801	1459	1305	1346	1248	1152	1229	1081	972	944	0
FIB-25	702	592	573	566	538	509	491	464	445	399	0
FIB-19	1099	867	732	780	710	643	738	617	527	545	0
Combined (annotated)	721	708	694	684	658	643	626	576	566	499	0
FIB-25 (annotated)	393	387	379	375	360	353	339	307	304	272	0
FIB-19 (annotated)	328	321	315	309	298	290	287	269	262	227	0

Table 2: Results of the probabilistic model construction.

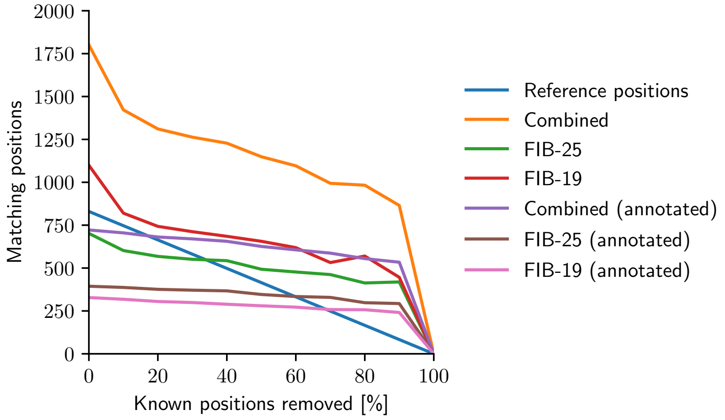

A FIB-19 synapse center-of-mass superclusters

B FIB-25 synapse center-of-mass superclusters

C FIB-19 Mi10 cell and synapse projections

D FIB-25 T4a cell and synapse projections

Supplementary Figure 1: **(A)** FIB-19 synapse center-of-mass superclusters. The clusters form two strata in the medulla (ME), and in the lobula (LO) and lobula plate layers (LOP, 1-4) additionally. Each dot corresponds to the center-of-mass of all synapses belonging to the super-cluster. Typically, each diverging arborization of a cell becomes a distinct location, which helps our probabilistic model to project 3D positions of synapses into retinotopic 2D planes, despite the lobula having a different spatial orientation (perpendicular) than the medulla and lobula (Fig. 1a and c). **(B)** FIB-25 synapse center-of-mass superclusters. The clusters form two strata (ME) in the medulla. **(C)** Mi10 cell type in FIB-19 with no pre-annotated lattice positions. The seven cell specimen (hexagons) are recovered by our probabilistic algorithm. Individual synapses and synapse center-of-mass projections are superimposed. **(D)** T4a cell type in FIB-25 with eight pre-annotated (an) and two recovered (al) lattice positions. The projected synapse positions show directional displacement consistent with the direction selectivity of T4a cells (Fig. 3a).



Supplementary Figure 2: Results of the probabilistic model construction.

Supplementary Note 2 Manually constructed connectome

Lamina and ommatidia model Since neither FIB-19 nor FIB-25 contain the connections of the ommatidia or first neuropile, the lamina[48,68], we reused and refined the existing hand-crafted model from our previous work[119], which is based on data from Rivera-Alba *et al.*[50] and Tuthill *et al.*[124,125].

Non-columnar single CT1 cell model While we did in general not model any neurons with large tangential branches, such as Mt, Mt, Pm, Dm, which span many columns and are therefore insufficiently segmented in FIB-19 and FIB-25, we did model the single CT1 cell present in the lobula CT1(Lo1) and medulla CT1(M10)[49]. Because a multi-compartment model with bidirectional electrical synapses resulted often in oscillatory dynamics in earlier modeling attempts and because CT1 terminals were found to act as functionally independent units[70] we modelled CT1 as two anatomically separate cell types CT1(Lo1) and CT1(M10).

Non-columnar periodic cells Because lamina-wide-field neurons, Lawf1 and Lawf2, do not occur in each individual column but more sparsely, we modeled them with an inferred spatial stride to occur more sparsely resulting in 123 cell of each type in our model (there are approx. 140 Lawf2 neurons per optic lobe, and each of approx. 700 columns is innervated by approx. 5 Lawf2 cells[124]).

Hexagonal lattice rendering of connectome For compilation into the hexagonal grid, the convex hull of the filters is filled with ones to remove spatial discontinuities. Although these are considered mostly false positives from the connectome reconstruction[12], this allowed for weak autapses in our hexagonal model that did not affect the tuning predictions.

Additional proofreading We manually proofread filters on the hexagonal lattice and compared them to the reported filters in the literature to ensure overall correspondence. We found that the reconstruction did not fully capture the asymmetry reported in[48] of the T5 anatomical receptive field of Tm9, which we then substituted by a Gaussian at the reported offset column scaled by the reported number of input synapses. For few T4 and T5 inputs the number of input synapses reported in the literature[48] slightly deviated from our reconstruction. To get a better initialization of our filter scale we scaled them to closely match the number of input synapses reported[48].

Supplementary Note 3 Investigating the role of sparse connectivity with synthetic networks for MNIST digit recognition

Training feedforward synthetic networks The weight matrix for each layer in Dale's-law-based synthetic networks (DLTrues) is decomposed into three components: binary adjacency matrix, non-negative weight magnitudes, and a sign vector.

$$\begin{aligned} W_{\texttt{DLTrue}} &= C_{\texttt{DLTrue}} \odot W_{\texttt{DLTrue}} \odot \left[\mathbf{1} \otimes \mathbf{s}_{\texttt{DLTrue}}\right], \\ \text{where } C_{\texttt{DLTrue}} &= \text{binary adjacency matrix}, \\ W_{\texttt{DLTrue}} &= \text{non-negative weight magnitudes of true network}, \\ \mathbf{s}_{\texttt{DLTrue}} &= \text{sign vector} \end{aligned}$$

By means of projected gradient descent, $W_{\texttt{DLTrue}}$ is enforced to be non-negative and is initialized from the absolute value of the He initialization distribution[140]. Although sign vector $\mathbf{s}_{\texttt{DLTrue}}$ is randomly initialized with equal probability to be either -1 or +1 to represent inhibitory and excitatory synapses respectively, its elements are allowed to assume values in $\mathbb{R}$ over the course of training.

Inducing sparsity Binary adjacency matrix $C_{\texttt{DLTrue}}$ is initialized to be a unit matrix and is later updated according to the desired true network connectivity level. Following the LTH algorithm, a portion of the lowest-magnitude weights, designated to be pruned, were identified from $W_{\texttt{DLTrue}} \odot [\mathbf{1} \otimes \mathbf{s}_{\texttt{DLTrue}}]$. For pruning synapses, the corresponding entries in the adjacency matrix $C_{\texttt{DLTrue}}$ were then set to zero. After each pruning iteration, weight magnitudes were reset back to their original initialization, followed by a final training run post-pruning.

Training with/without sign constraints In addition to a Dale's-law-based sign constraint, we also experimented with networks trained without any sign constraint. No restrictions were imposed on the nature of outgoing synapses i.e., a neuron can have both excitatory and inhibitory outgoing synapses. For true network variants trained without a sign constraint, $\mathbf{1} \otimes \mathbf{s}_{\texttt{DLTrue}}$ was simply replaced by a sign matrix $S_{\texttt{nonDLTrue}}$ initialized in a similar fashion; that is, all entries were initialized to be in $\{-1, +1\}$ with equal probability. We will refer to true networks trained without a sign constraint as `nonDLTrues`.

$$\begin{aligned} W_{\texttt{nonDLTrue}} &= C_{\texttt{nonDLTrue}} \odot W_{\texttt{nonDLTrue}} \odot S_{\texttt{nonDLTrue}}, \\ \text{where } C_{\texttt{nonDLTrue}} &= \text{binary adjacency matrix}, \\ W_{\texttt{nonDLTrue}} &= \text{non-negative weight magnitudes of true network}, \\ S_{\texttt{nonDLTrue}} &= \text{sign matrix} \end{aligned}$$

Training simulated networks As elements in a true network's $\mathbf{s}_{\texttt{DLTrue}}$ or $S_{\texttt{nonDLTrue}}$ are allowed to assume values in $\mathbb{R}$ while training, only the signs of these elements are inherited by the true networks's respective simulated network.

$$\begin{aligned} W_{\texttt{DLSimulated}} &= C_{\texttt{DLTrue}} \odot W'_{\texttt{DLSimulated}} \odot \left[\mathbf{1} \otimes \mathbf{s}'_{\texttt{DLTrue}}\right], \\ \text{where } C_{\texttt{DLTrue}} &= \text{binary adjacency matrix from corresponding DLTrue}, \\ W'_{\texttt{DLSimulated}} &= \text{non-negative weight magnitudes of simulated network}, \\ \mathbf{s}'_{\texttt{DLTrue}} &= \text{signs of elements in } \mathbf{s}_{\texttt{DLTrue}} \end{aligned}$$

$$\begin{aligned} W_{\texttt{nonDLSimulated}} &= C_{\texttt{nonDLTrue}} \odot W'_{\texttt{nonDLSimulated}} \odot S'_{\texttt{nonDLTrue}}, \\ \text{where } C_{\texttt{nonDLTrue}} &= \text{binary adjacency matrix from corresponding NonDLTrue}, \\ W'_{\texttt{nonDLSimulated}} &= \text{non-negative weight magnitudes of simulated network}, \\ S'_{\texttt{nonDLTrue}} &= \text{signs of elements in } S_{\texttt{nonDLTrue}} \end{aligned}$$

Extended Data Fig. 9 A shows median hidden-layer tuning correlations for networks trained with and without Dale's law sign constraint for three different architectures.

Training simulated networks that had access to weight magnitudes Three levels of multiplicative noise $\sigma = 0.1, 0.25, 0.5$ were explored, inducing low-noise, medium-noise, and high-noise weight estimates, respectively. Each noise level represents the maximum percentage by which a weight magnitude could be perturbed.

$$\begin{aligned} W_{\texttt{Simulated\_Init}} &= X \odot W_{\texttt{True}}, \\ \text{where } X_{ij} &= \text{Uniform}(1-\sigma,\ 1+\sigma) \end{aligned}$$

Simulated networks were trained with a Gaussian prior on the weights centered around the noisy initialization. In effect, this additional loss term penalizes trainable weights for deviating from their noisy initialization.

$$\mathcal{L}_{\texttt{weight\_prior}} = \mathcal{L}_{\texttt{cross\_entropy}} + \lambda \sum \left[W_{\texttt{Simulated}} - W_{\texttt{Simulated\_Init}}\right]^2$$

Fig. 9B shows median hidden-layer tuning correlations for networks with low-, medium-, and high-noise weight perturbations.

Sparse networks have larger local minima. We inspected the size of the simulated networks' local minima by analyzing how well they converged when initialized from a perturbed local optimum. Response tuning correlation in this context was used to quantify efficacy of convergence. After training the connectome simulations on the same handwritten digit classification task, we found that sparser networks were able to recover function even from highly-perturbed network initializations. By virtue of weight pruning, as the number of free parameters in the true network (and hence simulated network) are reduced, we believe that the size of the simulations' local minima increase, allowing sparser simulated networks to converge even when initialized far from optimum.

Simulated networks were initialized with a noisy version of their respective true network's weights. Three varying levels of multiplicative noise ($\sigma = 0.1, 0.25, 0.5$) were used to perturb the simulations' local minima.

Extended Data Fig **??** shows median hidden-layer tuning correlations at initialization distances of $\sigma = 0.1, 0.25, 0.5$ for networks with 128, 256, and 512 neurons per hidden layer, respectively.