# A Connectome Based Hexagonal Lattice Convolutional Network Model of the Drosophila Visual System

**Fabian David Tschopp**
Institute of Neuroinformatics
University of Zurich and ETH Zurich
tschopfa@student.ethz.ch

**Michael B. Reiser**
Janelia Research Campus
Howard Hughes Medical Instutute
reiserm@janelia.hhmi.org

**Srinivas C. Turaga**
Janelia Research Campus
Howard Hughes Medical Instutute
turagas@janelia.hhmi.org

## Abstract

What can we learn from a connectome? We constructed a simplified model of the first two stages of the fly visual system, the lamina and medulla. The resulting hexagonal lattice convolutional network was trained using backpropagation through time to perform object tracking in natural scene videos. Networks initialized with weights from connectome reconstructions automatically discovered well-known orientation and direction selectivity properties in T4 neurons and their inputs, while networks initialized at random did not. Our work is the first demonstration, that knowledge of the connectome can enable in silico predictions of the functional properties of individual neurons in a circuit, leading to an understanding of circuit function from structure alone.

## 1  Introduction

Universal function approximation results for artificial neural networks [1, 2] imply that many possible neural network architectures, with different connectivity matrices and different activation functions, can compute the same function. This suggests that it is difficult to predict the precise neural circuitry underlying a given neural computation. For this reason, we have many different proposals for instance, for how visual motion is computed through elementary motion detectors (EMD) such as Barlow-Levick [3] and Hassenstein-Reichardt [4], and how neural signals might be integrated [5, 6]. In contrast, since neural connectivity is generally sparse, knowledge of the connectivity of a neural circuit might constrain its possible computational function [7, 8]. Taken together, this suggests that it might be easier to predict function from structure, rather than the other way around.

Recent reconstructions of the first two stages of the drosophila visual system [9–14] enable us to test this hypothesis that neural circuit function can be predicted from neural circuit structure. We constructed a connectome-based artificial network model of the drosophila visual system and trained this model to track objects in videos of natural scenes. We found that a network model derived from the connectome is able to recapitulate well-known orientation and direction selectivity tuning properties of cell types in the medulla of the circuit.

---

Work in progress.
Supplemental Material can be found at https://github.com/naibaf7/dvsc_2017_supplemental

## 1.1 Prior work

Existing work has focused on roughly correlating single cell responses in different brain brain regions to layers in a deep convolutional neural network [15, 16], but not identify individual cell types and their connectivity. A different hexagonal lattice based model has also been proposed [17, 18], however they do not model any circuitry beyond the photoreceptors, focusing only on the simulation of ommatidia. Specific to the *Drosophila*, mathematical models to fit the physiological motion responses have been suggested [19], however these simpler models were explicitly trained to learn the neural response. Other recent work tries to infer structure from function [20], without a connectome, and shows that a deep network can learn the neural responses of retinal ganglion cells when a convolutional network based encoding model is trained on natural images and white noise. They further show that recurrent lateral connections and feedforward inhibition help to reconstruct temporally accurate retinal responses. Physiological experiments [13] suggest that this is also the case for directional motion detection in the *Drosophila*.

## 2 Hexagonal lattice convolutional network model

Based on publications of lamina [12–14] and medulla [9–11] connectomes, we modeled a connectome consisting of 43 neuron types. The neurons in these layers have a repeating columnar architecture with one column per ommatidium, each spanning $5°$ of visual angle, and they form a hexagonal lattice [21, 22]. Since the published connectomes only correspond to reconstructions of a few columns and not the entire hexagonal lattice, we repeat the locally described connectome in a spatially invariant manner, leading to a hexagonal lattice convolutional network.

The network begins with the bundled rhabdomeres R1-R8. We simplified the neural superposition mapping [22, 23] by directly associating the inputs with their target column in the lamina and medulla, where they form inhibitory synapses [24]. For the lamina, we simulate 10 neuron classes (L1-L5, C2-C3, T1, Lawf2, Am) that are known to contribute to motion processing. L1-L5, C2-C3 and T1 are synperiodic [9] and therefore present in each column. Amacrine cells (Am) are multicolumnar. Lawf2 cells are likely presynaptic to most cells in the lamina, have wide receptive fields in the medulla and feedback to the lamina. As there are $\sim 140$ Lawf2 neurons per optic lobe, and each column is innervated by $\sim 5$ Lawf2 cells [14], we can distribute Lawf2 sparsely in our model (Figure 2, blacked out neurons are set to be constantly disabled). We had to omit Lawf1, Lai, Lat and glia cells due to a lack of dense reconstructions of their connectivity.

For the medulla, we modeled the modular types identified in a dense 7 column (Figure 1 (left)) reconstruction by Takemura *et al*. [10]. These include Mi1, Mi4, Mi9, T2, T2a, T3, Tm20, Tm1, Tm2, Tm4, Tm9, TmY5a, Dm8, Dm2 which can be mapped once per column [10]. Additionally, we simulate T4a-T4d and T5a-T5d, which are ultraperiodic and connect with 4 distinct layers in the lobula plate [9]. A recent update by Takemura *et al*. [11] looked at T4 connections beyond 7 columns and found a better connectome model (Figure 6a versus 6b previously), which also significantly improved our simulation results (Section 4). For T5 however, we know that Tm1, Tm2 and Tm9 synapse onto T5, but the spatial configuration and delays are not yet known, since T5 resides in the lobula. This also had consequences for recovering T5 functionality (Section 4.2). Tm6 and Mi15 appear to be tiled since they were found only 4 times in 7 columns [9] (Figure 2), and we also simplified the ultraperiodic occurence of Tm3 to synperiodic. Many cells are transmedullar and project to the lobula or lobula plate [25]. Transmedullar cells are the outputs of our simulated medulla and connect to a 3-stage fully connected decoder (Figure 2), with 128, 32 and 4 neurons, which integrates visual information to solve our proxy-task (Section 4.1).

Additionally, we used unpublished RNA sequence data to determine which neurons and synapses are inhibitory (such as Mi9) and excitatory, based on the neurotransmitters and receptors expressed. These constraints were also imposed during fine-tuning of the network weights (Section 3.2).

The electron microscopic reconstructions contain detailed information about the morphologies of the neurons being modeled. However, we ignore these details, instead choosing to build a simplified network of linear-nonlinear point neurons with continuous non-negative activations and instantaneous synapses. Since neurons at this stage of the fly visual system are non-spiking and produce graded potentials, this is a reasonable approximation.
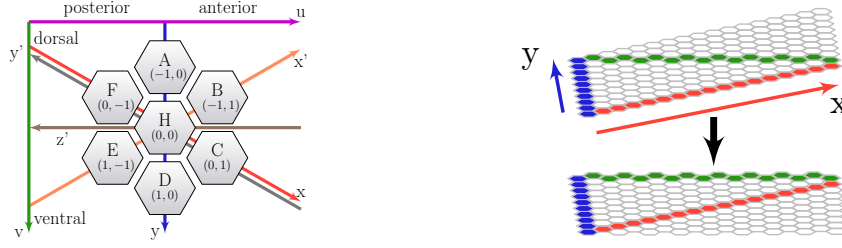
Figure 1: (left) 7 column structure with 6-neighborhood offsets along the principal $y$ and $x$ axis. The axis $y'$, $x'$ and $z'$ are equivalent to Braitenberg's hexagonal arrangement of ommatidia [21]. (right) Semantic arrangement of the hexagonal lattice of size $h \times w$ on the GPU as a 2D tensor. Additional border conditions at $y + \frac{1}{2} \cdot x = h$ (green) and $y = 0$ (red) allow us to get a rectangular lattice.

We simulate a hexagonal lattice of $y = 20$, $x = 35$ point neurons with continuous activations, resulting in the typical $\sim 700$ ommatidia and retinotopic columns found in a *Drosophila* fruit fly. Every neuron is defined as node with intrinsic properties: Name, sparsity along the $y$ and $x$-axis, activation function (ReLU), bias value and operator (addition). We selected the network orientation and reference frame (Figure 1 (left)) to be aligned with the dorsal half of the right *Drosophila* eye and built the model so that we can disregard the optic chiasma between lamina and medulla.

## 2.1 Connectome-based network weights

Connections between neurons in our model are defined as edge between source and target neuron. All weights are replicated spatially for all neurons of the same type, like in regular convolution filters. Depending on how two neuron types are connected laterally, arbitrary sparse filters within hexagonal neighborhoods can arise (Figure 6a). In our model, we replaced all synapses counted between two connected distinct neurons in the connectome by a single synapse. The resulting synapse was then initialized with a weight equal to the number of counted synapses divided by an arbitrary normalization factor of 75. This factor was chosen as small as possible, but large enough so that gradient-based optimization (Section 3.2) was stable. Since we used leaky rectified linear units and a graded potential model (non-spiking) for all neurons, the choice of normalization factor is arbitrary and can be rescaled exactly by the activities of the neurons. For the neuron bias values, there is no information that can be gained from the connectome and we therefore initialized them to a small positive constant value of $3.5/75.0$.

The complete model with all intercolumnar connection pattern and sparse filters can be found in the *Supplemental Material*. In Figure 2 we present our complete network graph. Each edge starts at the receptive field, which is marked in the same color as the edge, and projects to its target neuron. Note that the hexagonal lattice is shown as $15 \times 20$, whereas we used a $20 \times 35$ lattice for our simulation.

## 2.2 Physiology-based synaptic delays

As we simulate our model as a convolutional neural network with recurrent components, the activity of every downstream neuron can only be computed when all inputs to a neuron are determined for step $t$. This means we needed to linearize the connections so that the network graph becomes a directed acyclic graph through time. For symmetric connections between neurons, autapses to the same neuron and feedback connections going back from the medulla to the lamina and ommatidia, we added a synaptic delay of $1$ a.u., meaning the neurons read their input from state $t-1$, which is well-defined for the whole network from the previous forward pass.

Additional synaptic delays were introduced where physiological data suggested necessary delays: For T4, because of the temporal frequency optimum [26] and the spatial offsets of Mi1 to Tm3 [10] and only small temporal delay between them [27], we expect a delay at the negative input [11] (Figure 6a), which would form a more complex type of BL [3] detector with additional properties not described by such simple models (Section 4.2). Our previous model based on older data [9] suggested a configuration more akin to a HR [4] model. For the T5 cells, due to the lack of a complete connectome, we assumed the same configuration of temporal and spatial offsets.
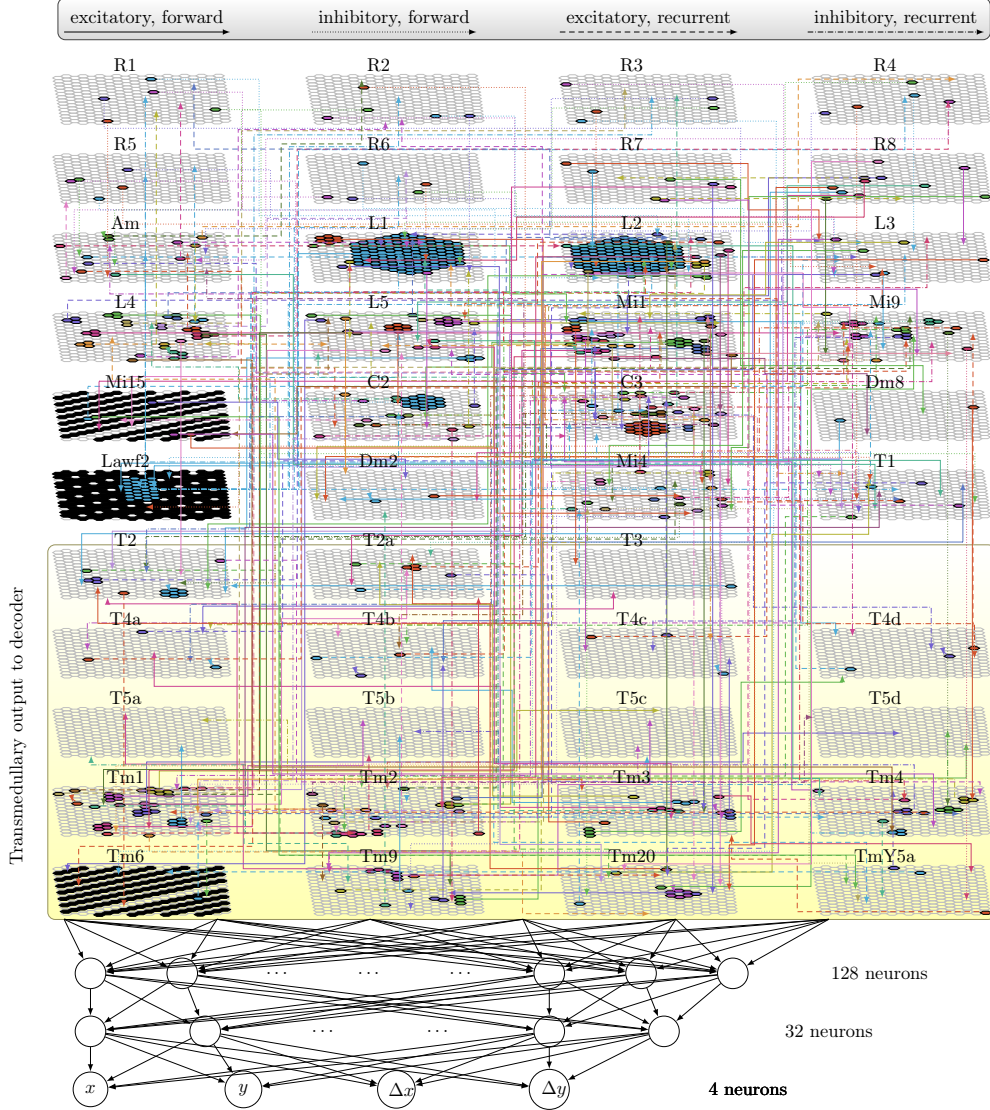
3

Figure 2: *Drosophila* visual system connectome-based model with 3 layer decoder.

## 2.3 Implementation

In order to simulate our model efficiently, we used the tensor data structures and gradient based optimizers from the Caffe deep learning library [28] combined with runtime code generating modules from LibDNN [29]. We introduce a new layer, *sparse repeated pattern recurrent neural network*, which takes care of translating our connectome-based network defined in Python to runtime generated CUDA or OpenCL code. Caffe's Python interface allows to conduct experiments rapidly, as it can be combined with Python tools to generate artificial stimuli and analyze the neuronal responses.

The newly introduced layer supports backpropagation through time (BPTT, [30][31]) during training, and a variable history length can be chosen. In our setup, we used a batchsize of 5 steps and a history of 10 steps. Images have to be resampled and rearranged for the hexagonal lattice (Figure 1 (right)). To this end, we implemented nearest neighbor and bilinear interpolations as well as an ommatidia-mode which resamples an RGB image to grayscale for the rhabdomeres R1-R6, which all receive the same input in our model, R7 receives red (simulates UV), 30% of R8 sample blue and 70% of R8 sample green [23]. Note that we normalize the images and do not simulate any spherical projections that occur in real fly vision [18], since our focus is on the lamina and medulla circuitry.
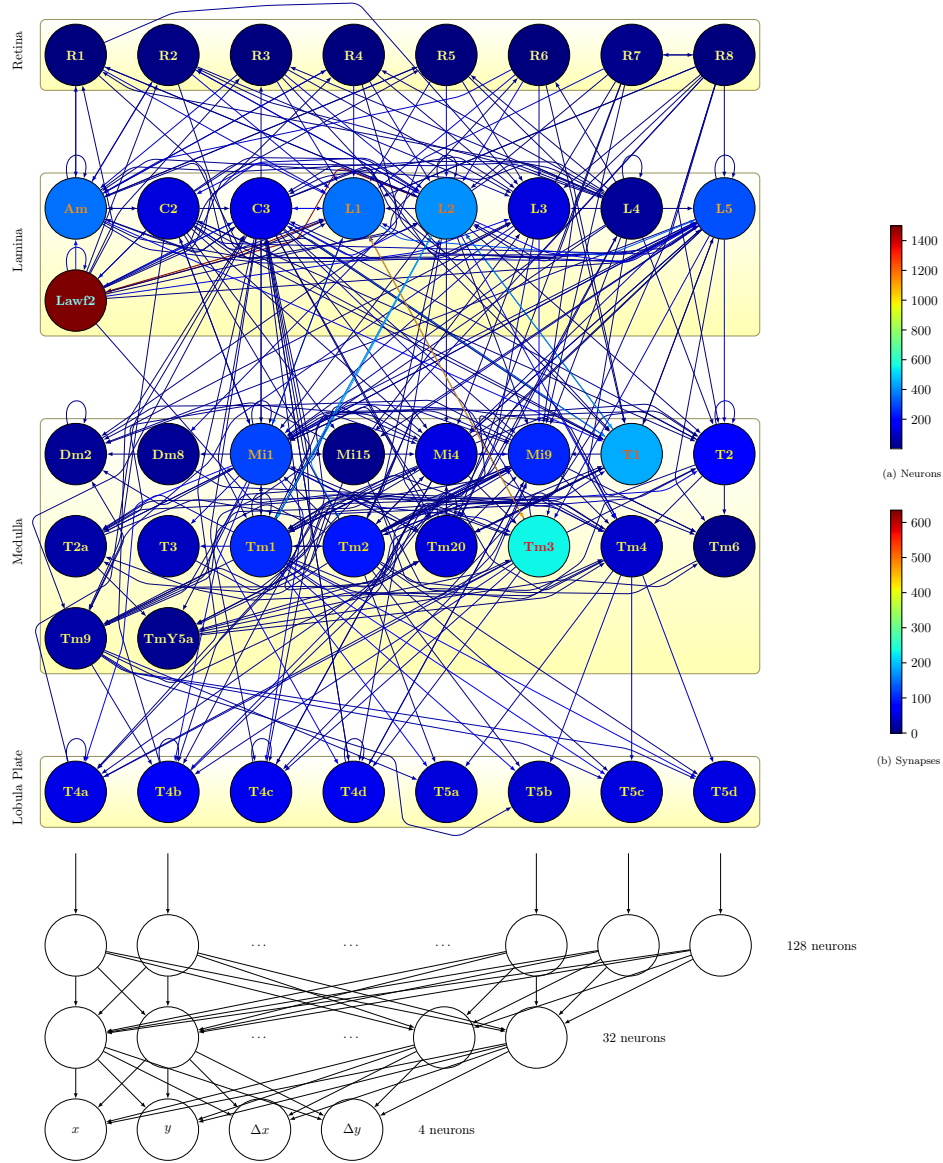
4

Figure 3: *Drosophila* visual system model based on [11] with a 3 layer decoder, displaying the total number of input synapses to a neuron (a) and the synapses per neuron type-to-type edge (b).
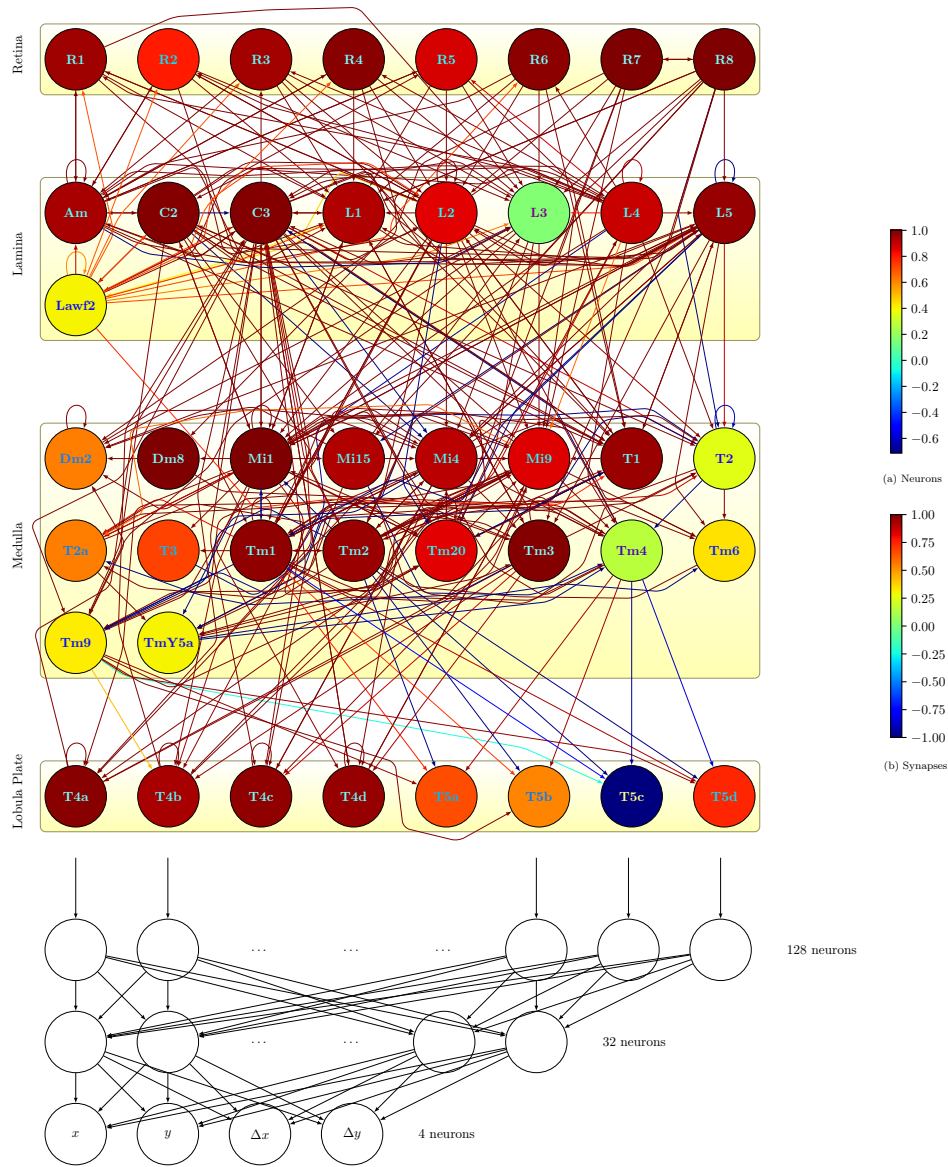
Figure 4: Correlation coefficient computed between trained weights and the connectome initialized weights based on [11] show that the on pathway weights are largely unchanged. Significant weight changes occur largely in the off pathway for which high quality connectome data was not available.
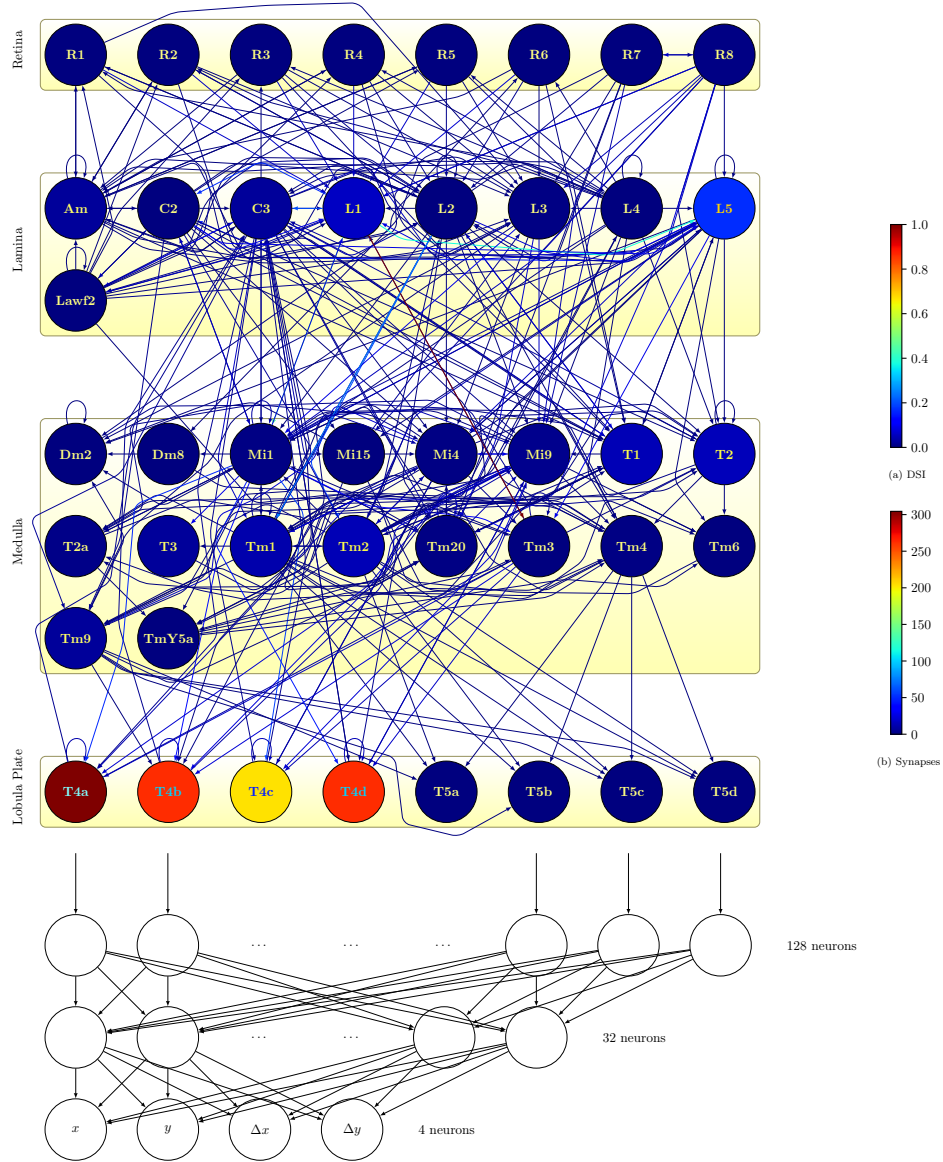
Figure 5: Takemura 2017 initialized and trained network shows strong direction selectivity arising in the T4 neurons but not in their inputs, consistent with known physiology. Cell type nodes are color-coded by the direction selectivity index (DSI) of neurons (a) and the synapses per neuron type-to-type edge (b).

(a) T4a input filters according to the Takemura 2017 model [11].



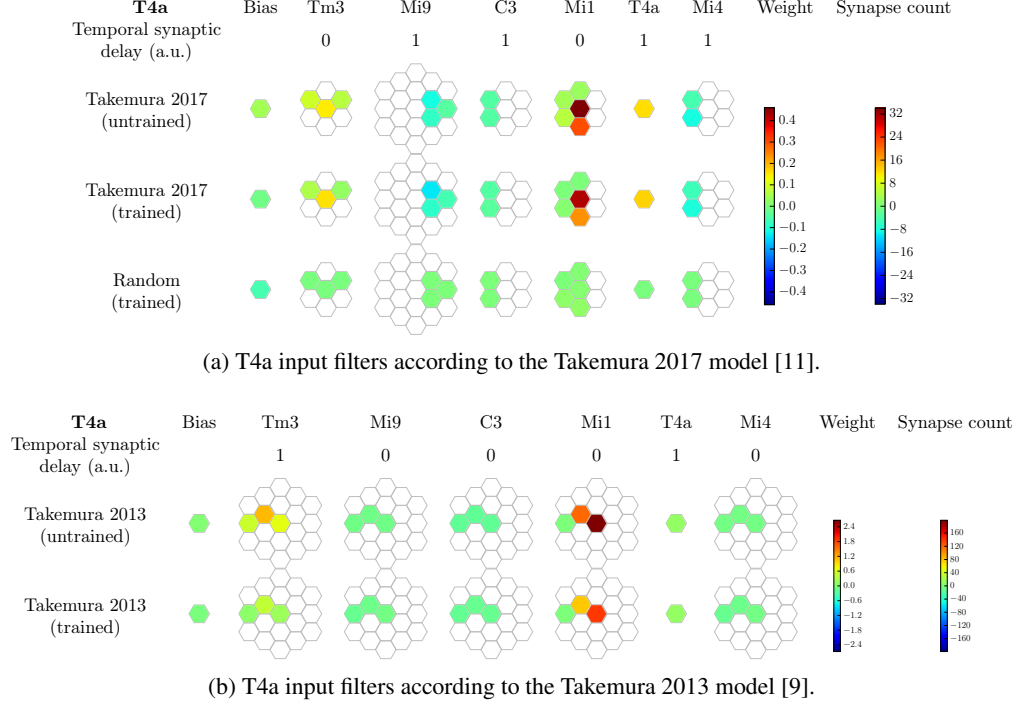(b) T4a input filters according to the Takemura 2013 model [9].

Figure 6: Initial and emerging convolutional receptive field filters of T4a. Mi1 and Tm3 are the main excitatory contributions, while Mi9, C3 and Mi4 are inhibitory. Weights initialized from the Takemura 2017 connectome are similar to the initialization, however weights initialized randomly do not recover the connectome structure.

# 3 Gradient-based circuit optimization for object tracking

In order to fine-tune the weights of our networks, we used a proxy-task that depends on the circuit's ability to compute a moving object's position and velocity. This approach is in contrast to explicitly training the network to learn an encoding model based on physiologically measured neural responses [20], and thus requires only anatomical information, and no recordings of neural activity.

## 3.1 DAVIS dataset

We used the DAVIS dataset [32, 33] in its 2016 version with 480p image resolution, which is used as a video object segmentation challenge. Due to its dense single-object annotation, it is easy to compute ground truth for the object centroid $(x, y)$ and velocity $(\Delta x, \Delta y)$ between consecutive frames. The network's task is to predict these four parameters. The whole dataset consists of 50 independent sequences of 40 to 100 frames at $24\,\text{FPS}$. For our purposes, we selected 8 sequences (*bmx-bumps*, *drift-chicane*, *motorbike*, *paragliding*, *rollerblade*, *soccerball*, *surf* and *kite-walk*) that contain primarily object motion and maintain a static camera position. Other sequences were not suitable, as the camera follows the object too closely, leading to near-zero object motion relative to the camera. The selected sequences were split into blocks of 10 consecutive frames. Per set, 2 blocks of 10 frames were used for testing, and the remaining 2 to 8 blocks were used as training set. Additionally, we augmented the dataset by mirroring the training data along the vertical and horizontal axis.

## 3.2 Gradient based optimization of network parameters

During circuit optimization, we used the Adam optimizer [34] with weight decay to avoid large weights. The 164 decoder neurons were always initialized with Xavier [35] weights without constraints. For our connectome-based network, six different training configurations were tested for functionality and robustness.
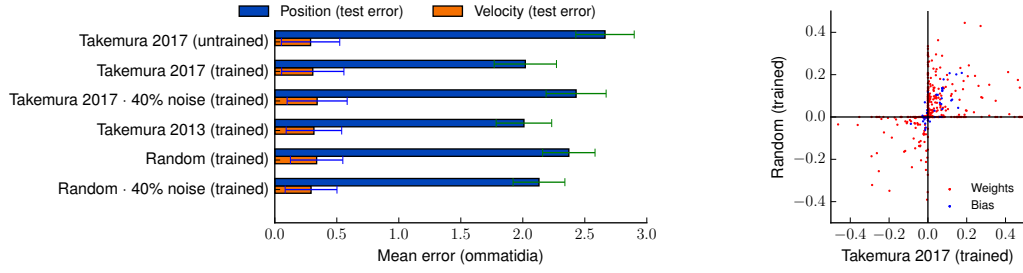
Figure 7: (left) Test scores (mean $L_2$ error) on the DAVIS dataset based on 20 frames for each of the selected 8 sequences in the DAVIS dataset (Section 3.1). (right) Different weight and bias solutions are discovered by training from connectome vs random initialization.

- **Takemura 2017 (trained)** Sign constrained weights initialized from the connectome.
- **Takemura 2017 · 40% noise (trained)** Initial weights perturbed with $40\%$ uniformly sampled multiplicative noise.
- **Takemura 2017 (untrained)** Weights fixed to connectome, with only biases and decoder weights trained.
- **Random (trained)** Sign constrained weights initialized at random with MSRA [36].
- **Random · 40% noise (trained)** Sign constrained weights using the previous MSRA initialization with $40\%$ uniformly sampled multiplicative noise.
- **Takemura 2013 (trained)** Sign constrained weights using our older model [9] for T4 inputs (Figure 6a).

For the random (MSRA [36]) initialized weights, the sign constraints were taken from the connectome initialized weights, since we only want to consider models which keep known inhibitory and excitatory properties. Bias values are not limited since their initial values are arbitrary (Section 2.1). The weight constraints were enforced with a projected gradient descent update, $\theta_t' = \text{sign}(\theta_0) \cdot |\theta_t|$, that mirrors the weights on the 0-axis rather than setting them to 0, where $\theta_t$ is the weight obtained by Adam [34] optimization in step $t$. This keeps as many weights from being disabled as possible. Each configuration was trained for 30,000 iterations to minimize the squared error in predicting object location and velocity.

## 4   Results

### 4.1   DAVIS object tracking

Our networks are able to track objects with a mean $L_2$ error as low as 2.0 ommatidia in distance on the test set. The network which was constrained to only train the decoder and the bias values of the network was found to perform worst, with 2.6 ommatidia error (Figure 7 (left)). Predicted velocities had a mean $L_2$ error within a range of 0.28 to 0.33 ommatidia/step.

### 4.2   Predicted neural tuning properties

We recorded all cell types both qualitatively as video clip (see *Supplemental Material*) and quantitatively by checking for linear and nonlinear behavior. Most notably, we found that direction and orientation selective properties were recovered during training (Section 3.2). To compute the direction selectivity index (DSI) and orientation selectivity of all 43 neuron types in our simulation (see Figure 9), we moved light and dark bars of 180 directions across the visual field of our simulated fly eye and measured the peak responses $R_k^{peak}$. For direction selectivity, the bars had a width of $\sim 1$ ommatidia $(\overset{\triangle}{=} 5°)$ and a speed of $\sim 0.7$ ommatidia/timestep. Orientation selectivity was determined with a bar of 2 ommatidia width and a speed of $\sim 0.1$ ommatidia/step, which is slightly below the threshold for direction selectivity.

A robust method based on vector addition [37] was used to determine a noise-free selectivity index:

$$DSI = \sqrt{(\sum_{k=0}^{2\pi} \cos(k) \cdot R_k^{peak})^2 + (\sum_{k=0}^{2\pi} \sin(k) \cdot R_k^{peak})^2}$$

9

(a) T4a NULL response ($0°$).



(b) T4a preferred direction ($192°$) response.



(c) DS, Takemura 2017 (untrained).



(d) OS, Takemura 2017 (untrained).



(e) DS, Takemura 2017 (trained). Each T4 cell is direction selective in one of four principal directions (T4a: $192°$, T4b: $359°$, T4c: $51°$ and T4d: $275°$).



(f) OS, Takemura 2017 (trained). Two-lobed orientation selectivity with strong center-surround inhibition for bars orthogonal to the preferred direction as shown by [38].



(g) DS, Random (trained).



(h) OS, Random (trained).



(i) DS, Takemura 2013 (trained).



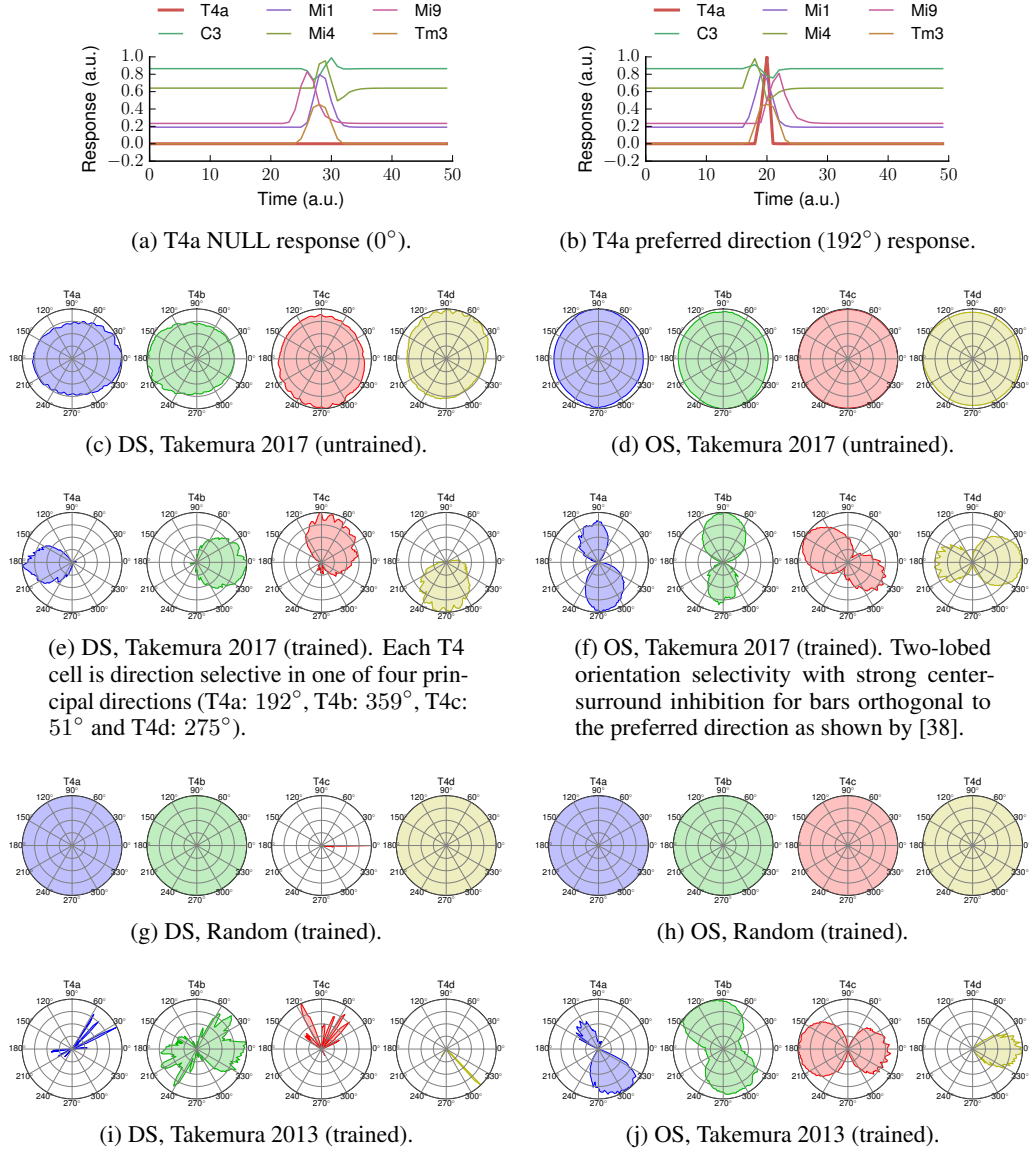(j) OS, Takemura 2013 (trained).

Figure 8: Orientation (OS) and direction (DS) selectivity of T4 cells. Responses were tested with light and dark bars of 180 directions.

We found that training the Takemura 2017 initialized weight model with polarity constraints obtained from connectome data recovers and tunes T4 cells to well-known physiological responses [19, 26, 38]. The neurons became highly direction selective in their preferred orientation (T4a: $192°$, T4b: $359°$, T4c: $51°$ and T4d: $275°$, Figure 8e, angles are oriented as in Fisher's measurements [38]). They also show a two-lobed orientation selectivity with strong center-surround inhibition for bars orthogonal to the preferred direction (Figure 8f). Previous models [9] were also able to tune the cells, although less stable (Figures 8i-8j). Networks initialized with random weights were unable to recover these properties (Figures 8g-8h). The characteristic tuning properties were also absent with initial weights, however slight DS tuning is visible (Figures 8c-8d).

## 4.3 Robustness of connectome-initialized networks

Although networks initialized with random weights were also able to track objects (Section 4.1), the circuitry recovered to achieve this does not correspond to physiology, unlike networks initialized
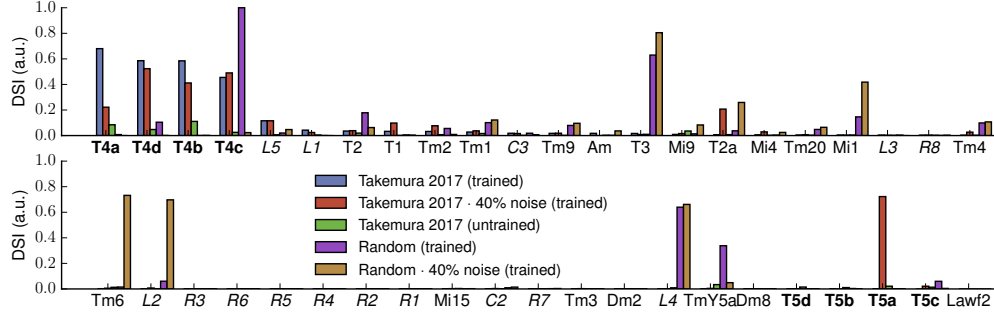
Figure 9: Direction selectivity index (DSI) scores. **Bold** neurons are known to be direction selective, while *italic* neurons are known to be non-selective. The remaining neurons have unknown tuning properties. The Takemura 2017 model successfully recovers **T4a-T4d** direction selectivity, while random initialization does not.

with synapse counts. The high disparity between connectome-based and random weights after training (Figure 7 (right)) are also an indicator that the network has many local minima to solve the task, as predicted by the universal function approximation [1, 2]. We found that random weights trained networks also recovered DS responses for some neurons (Figure 9) at sufficiently dissimilar angles (T3: $10°$ and L4: $238°$). Solutions initialized at random are also very unstable, and adding $40\%$ multiplicative noise to the initialization leads to different neurons becoming direction selective (such as L2 and Tm6 instead of T4c and TmY5a). This also shows that many neurons have the spatiotemporal wiring required to potentially become direction selective.

Our Takemura 2017 [11] model with connectome-based weights, on the other hand, is robust against up to $40\%$ multiplicative noise, at which point it still recovers identical, although less sharply tuned, responses for T4a-T4d. Additionally, it enabled T5a to become direction selective, although at the wrong angle ($335°$ instead of the expected $\sim 180°$).

## 5   Discussion

Our model predicted accurate responses for all T4 cells and demonstrates that functional information can be gained by means of fine-tuning a model initialized with weights derived from electron microscopic reconstructions with only minimal additional assumptions about synaptic delays. Remarkably, the model converged to such responses without being forced to learn them, as in encoding models [20], but instead through a visual proxy-task with 4 predicted variables.

We acknowledge that not all neurons took on their expected ON/OFF responses suggested in the literature. While for example Mi1, Tm3 and Mi4 (ON) and Tm1, Tm2, Tm4, Tm9 (OFF) are responding as expected, Mi9 shows ON characteristics instead of OFF (as visualized in the video clip in our *Supplemental Material*). Many parts of our model still contain uncertainties and incomplete filters due to missing connectomic information and the difficulty of assembling a consensus-model across multiple publications. Examples include the T5 pathway, for which we were unable robustly recover OFF direction selectivity, likely due to incomplete lateral dendritic arbor reconstructions. Indeed, the T4 neurons only recovered their known direction and orientation selectivity when their connectome was reconstructed with sufficient accuracy, as in [11]. The incomplete reconstructions from [9] were not sufficient. Note also that neurons with large tangential branches such as Mt8, Mi11 or Tm28 within the medulla were also not included in our model [25].

We also anticipate that our model could be useful to also test later stages of fly vision, such as vertical and horizontal system cells [39], object-detecting neurons such as LC11 [40] and visual projection neurons [41, 42]. These steps await further dense EM reconstructions of the lobula and lobula plate. With more complete models, we expect to be able to extract further characteristics such as object detection, color vision and rotational movement sensitivity from structure alone, which could then be verified with physiological measurements or be used to make predictions. Furthermore, we hope to automate the reconstruction of models that can be simulated directly from EM reconstructions without the need to assemble coherent networks from various publications.

# References

[1] Balázs Csanád Csáji and Huub Ten Eikelder. "Approximation with Artificial Neural Networks". In: (2001).

[2] G. Cybenko. "Approximation by superpositions of a sigmoidal function". In: *Mathematics of Control, Signals, and Systems* 5.4 (Dec. 1992), pp. 455–455. DOI: 10.1007/bf02134016. URL: https://doi.org/10.1007/bf02134016.

[3] H B Barlow and W R Levick. "The mechanism of directionally selective units in rabbit's retina." In: *The Journal of Physiology* 178.3 (June 1965), pp. 477–504. DOI: 10.1113/jphysiol.1965.sp007638. URL: https://doi.org/10.1113/jphysiol.1965.sp007638.

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

[5] H. S. Seung and H. S. Seung. "How the brain keeps the eyes still". In: *Proc. Natl. Acad. Sci. USA* (1996), pp. 13339–13344.

[6] Mark S. Goldman. "Memory without Feedback in a Neural Network". In: *Neuron* 61.4 (Feb. 2009), pp. 621–634. DOI: 10.1016/j.neuron.2008.12.012. URL: https://doi.org/10.1016/j.neuron.2008.12.012.

[7] H. Sebastian Seung. "Reading the Book of Memory: Sparse Sampling versus Dense Mapping of Connectomes". In: *Neuron* 62.1 (Apr. 2009), pp. 17–29. DOI: 10.1016/j.neuron.2009.03.020. URL: https://doi.org/10.1016/j.neuron.2009.03.020.

[8] Moritz Helmstaedter et al. "Connectomic reconstruction of the inner plexiform layer in the mouse retina". In: *Nature* 500.7461 (Aug. 2013), pp. 168–174. DOI: 10.1038/nature12346. URL: https://doi.org/10.1038/nature12346.

[9] Shin-ya Takemura et al. "A visual motion detection circuit suggested by Drosophila connectomics". In: *Nature* 500.7461 (Aug. 2013), pp. 175–181. DOI: 10.1038/nature12450. URL: http://dx.doi.org/10.1038/nature12450.

[10] Shin-ya Takemura et al. "Synaptic circuits and their variations within different columns in the visual system ofDrosophila". In: *Proceedings of the National Academy of Sciences* 112.44 (Oct. 2015), pp. 13711–13716. DOI: 10.1073/pnas.1509820112. URL: http://dx.doi.org/10.1073/pnas.1509820112.

[11] Shin-ya Takemura et al. "The comprehensive connectome of a neural substrate for 'ON' motion detection in Drosophila". In: *eLife* 6 (Apr. 2017). DOI: 10.7554/elife.24394. URL: https://doi.org/10.7554%2Felife.24394.

[12] Marta Rivera-Alba et al. "Wiring Economy and Volume Exclusion Determine Neuronal Placement in the Drosophila Brain". In: *Current Biology* 21.23 (Dec. 2011), pp. 2000–2005. DOI: 10.1016/j.cub.2011.10.022. URL: http://dx.doi.org/10.1016/j.cub.2011.10.022.

[13] John C. Tuthill et al. "Contributions of the 12 Neuron Classes in the Fly Lamina to Motion Vision". In: *Neuron* 79.1 (July 2013), pp. 128–140. DOI: 10.1016/j.neuron.2013.05.024. URL: http://dx.doi.org/10.1016/j.neuron.2013.05.024.

[14] John C. Tuthill et al. "Wide-Field Feedback Neurons Dynamically Tune Early Visual Processing". In: *Neuron* 82.4 (May 2014), pp. 887–895. DOI: 10.1016/j.neuron.2014.04.023. URL: https://doi.org/10.1016%2Fj.neuron.2014.04.023.

[15] D. L. K. Yamins et al. "Performance-optimized hierarchical models predict neural responses in higher visual cortex". In: *Proceedings of the National Academy of Sciences* 111.23 (May 2014), pp. 8619–8624. DOI: 10.1073/pnas.1403112111. URL: https://doi.org/10.1073/pnas.1403112111.

[16] Nicolas Pinto et al. "A High-Throughput Screening Approach to Discovering Good Forms of Biologically Inspired Visual Representation". In: *PLoS Computational Biology* 5.11 (Nov. 2009). Ed. by Karl J. Friston, e1000579. DOI: 10.1371/journal.pcbi.1000579. URL: https://doi.org/10.1371/journal.pcbi.1000579.

[17] Lev E. Givon and Aurel A. Lazar. "Neurokernel: An Open Source Platform for Emulating the Fruit Fly Brain". In: *PLOS ONE* 11.1 (Jan. 2016). Ed. by William W Lytton, e0146581. DOI: 10.1371/journal.pone.0146581. URL: https://doi.org/10.1371%2Fjournal.pone.0146581.

[18] Aurel A. Lazar et al. "A Parallel Processing Model of the Drosophila Retina". In: (2015). DOI: 10.5281/zenodo.30036.

[19] Etienne Serbe. "Analysis of the neural circuit underlying the detection of visual motion in Drosophila melanogaster". Sept. 2016. URL: http://nbn-resolving.de/urn:nbn:de:bvb:19-199173.

[20] Lane McIntosh et al. "Deep Learning Models of the Retinal Response to Natural Scenes". In: *Advances in Neural Information Processing Systems 29*. Ed. by D. D. Lee et al. Curran Associates, Inc., 2016, pp. 1369–1377. URL: http://papers.nips.cc/paper/6388-deep-learning-models-of-the-retinal-response-to-natural-scenes.pdf.

[21] V. Braitenberg. "Patterns of projection in the visual system of the fly. I. Retina-lamina projections". In: *Experimental Brain Research* 3.3 (1967), pp. 271–298. ISSN: 1432-1106. DOI: 10.1007/BF00235589. URL: http://dx.doi.org/10.1007/BF00235589.

[22] Kuno Kirschfeld. "Die projektion der optischen umwelt auf das raster der rhabdomere im komplexauge von Musca". In: *Experimental Brain Research* 3.3 (1967), pp. 248–270. ISSN: 1432-1106. DOI: 10.1007/BF00235588. URL: http://dx.doi.org/10.1007/BF00235588.

[23] Javier Morante and Claude Desplan. "Building a projection map for photoreceptor neurons in the Drosophila optic lobes". In: *Seminars in Cell & Developmental Biology* 15.1 (Feb. 2004), pp. 137–143. DOI: 10.1016/j.semcdb.2003.09.007. URL: https://doi.org/10.1016%2Fj.semcdb.2003.09.007.

[24] R. C. Hardie. "A histamine-activated chloride channel involved in neurotransmission at a photoreceptor synapse". In: *Nature* 339.6227 (June 1989), pp. 704–706. DOI: 10.1038/339704a0. URL: https://doi.org/10.1038/339704a0.

[25] K.-F. Fischbach and A.P.M. Dittrich. "The optic lobe of Drosophila melanogaster. I. A Golgi analysis of wild-type structure". In: *Cell and Tissue Research* 258.3 (Dec. 1989). DOI: 10.1007/bf00218858. URL: https://doi.org/10.1007%2Fbf00218858.

[26] Matthew S. Maisak et al. "A directional tuning map of Drosophila elementary motion detectors". In: *Nature* 500.7461 (Aug. 2013), pp. 212–216. DOI: 10.1038/nature12320. URL: https://doi.org/10.1038/nature12320.

[27] Rudy Behnia et al. "Processing properties of ON and OFF pathways for Drosophila motion detection". In: *Nature* 512.7515 (July 2014), pp. 427–430. DOI: 10.1038/nature13427. URL: https://doi.org/10.1038/nature13427.

[28] *SRPRNN Caffe*. 2017.

[29] *LibDNN*. 2017.

[30] Michael C. Mozer. "Backpropagation". In: ed. by Yves Chauvin and David E. Rumelhart. Hillsdale, NJ, USA: L. Erlbaum Associates Inc., 1995. Chap. A Focused Backpropagation Algorithm for Temporal Pattern Recognition, pp. 137–169. ISBN: 0-8058-1259-8. URL: http://dl.acm.org/citation.cfm?id=201784.201791.

[31] A. J. Robinson and F. Fallside. *The utility driven dynamic error propagation network*. Tech. rep. CUED/F-INFENG/TR.1. Cambridge: Cambridge University Engineering Department, 1987.

[32] F. Perazzi et al. "A Benchmark Dataset and Evaluation Methodology for Video Object Segmentation". In: *Computer Vision and Pattern Recognition*. 2016.

[33] Jordi Pont-Tuset et al. "The 2017 DAVIS Challenge on Video Object Segmentation". In: *arXiv:1704.00675* (2017).

[34] Diederik P. Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: *CoRR* abs/1412.6980 (2014). URL: http://arxiv.org/abs/1412.6980.

[35] Xavier Glorot and Yoshua Bengio. "Understanding the difficulty of training deep feedforward neural networks". In: *In Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS'10). Society for Artificial Intelligence and Statistics*. 2010.

[36] Kaiming He et al. "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification". In: *2015 IEEE International Conference on Computer Vision (ICCV)*. IEEE, Dec. 2015. DOI: 10.1109/iccv.2015.123. URL: https://doi.org/10.1109/iccv.2015.123.

[37] Mark Mazurek, Marisa Kager, and Stephen D. Van Hooser. "Robust quantification of orientation selectivity and direction selectivity". In: *Frontiers in Neural Circuits* 8 (Aug. 2014). DOI: 10.3389/fncir.2014.00092. URL: https://doi.org/10.3389%2Ffncir.2014.00092.

[38] Yvette E. Fisher, Marion Silies, and Thomas R. Clandinin. "Orientation Selectivity Sharpens Motion Detection in Drosophila". In: *Neuron* 88.2 (Oct. 2015), pp. 390–402. DOI: `10.1016/j.neuron.2015.09.033`. URL: `https://doi.org/10.1016%2Fj.neuron.2015.09.033`.

[39] Alexander Borst, Juergen Haag, and Dierk F. Reiff. "Fly Motion Vision". In: *Annual Review of Neuroscience* 33.1 (June 2010), pp. 49–70. DOI: `10.1146/annurev-neuro-060909-153155`. URL: `https://doi.org/10.1146%2Fannurev-neuro-060909-153155`.

[40] Mehmet F. Keleş and Mark A. Frye. "Object-Detecting Neurons in Drosophila". In: *Current Biology* 27.5 (Mar. 2017), pp. 680–687. DOI: `10.1016/j.cub.2017.01.012`. URL: `https://doi.org/10.1016%2Fj.cub.2017.01.012`.

[41] Ming Wu et al. "Visual projection neurons in the Drosophila lobula link feature detection to distinct behavioral programs". In: *eLife* 5 (Dec. 2016). DOI: `10.7554/elife.21022`. URL: `https://doi.org/10.7554%2Felife.21022`.

[42] Hideo Otsuna and Kei Ito. "Systematic analysis of the visual projection neurons of Drosophila melanogaster. I. Lobula-specific pathways". In: *The Journal of Comparative Neurology* 497.6 (2006), pp. 928–958. DOI: `10.1002/cne.21015`. URL: `https://doi.org/10.1002%2Fcne.21015`.