Data-driven deep neural network models of visual processing in *Drosophila*

Jonathan S. Skaza
Cold Spring Harbor Laboratory
Cold Spring Harbor, NY, USA

Erin Wong
Great Neck South High School
Lake Success, NY, USA

Arie Matsliah
Princeton Neuroscience Institute
Princeton University
Princeton, NJ, USA

Benjamin R. Cowley
Cold Spring Harbor Laboratory
Cold Spring Harbor, NY, USA

Abstract

Visual projection neurons (VPNs) in *Drosophila melanogaster*'s visual system integrate and project optic lobe information to the central brain. However, the specific visual features integrated by various VPNs are not well understood. Understanding this neural code is crucial for uncovering the inner workings of visuomotor transformations during behaviors like courtship and flight. We utilized VPN recordings from multiple studies to train deep neural network (DNN) models, including classic convolutional and connectome-inspired DNNs, to predict neural responses of VPNs within the optic glomeruli. Our models revealed the stimulus preferences and temporal properties for each optic glomerulus (OG). We found that despite large differences in architecture, the DNN models had similar accuracy in predicting OG responses. Thus, the artificial stimuli traditionally used to probe visual function—moving spots and bars—are too impoverished to distinguish competing models. We propose a new class of stimuli, optimized by our models, that maximize the differences in predicted responses between models. Presenting these "controversial" stimuli in future experiments will better refine our DNN models and unlock further insights into fruit fly visual processing.

Keywords: visual system; Drosophila; optic lobe; optic glomeruli; lobula complex; deep learning; data-driven modeling

Introduction

Many insights about early visual processing have arisen by carefully investigating the visual system of the fruit fly (*Drosophila melanogaster*) (Currier, Pang, & Clandinin, 2023), especially about direction selectivity (Strother et al., 2017). Still, we are far from characterizing all neurons in this visual system, especially the population of neurons that read out from the optic lobe and project to the central brain. These neurons, called visual projection neurons or LC neurons (as they project from the lobula's columns), are highly organized. Each LC neuron type projects to a single optic glomerulus (OG) with ∼57 glomeruli in total (Fig. 1a). The stimulus tuning of each OG has typically been characterized by recording OG responses to a battery of visual stimuli (e.g., looming and moving spots and bars). Although general tuning properties can be inferred from OG responses to these stimuli, the precise tuning of each OG and its underlying computations remain open questions.

To precisely characterize OG function and tuning, we sought a computational model that accurately predicts OG responses given the same visual stimulus presented to the fly. To date, proposed models of LC function are either task-driven (Cowley et al., 2023; Lappalainen et al., 2023) or normative (Hindmarsh Sten, Li, Otopalik, & Ruta, 2021). Here, we directly train deep neural network (DNN) models on recorded OG responses. Given enough recordings and stimuli, our data-driven DNN model should be highly accurate in predicting OG responses. We test different DNN architectures, including "black-box" networks whose sole purpose is prediction as well as "connectome-inspired" networks whose connections reflect those of the recently-released FlyWire connectome (Flywire Consortium, 2024). We then use these DNN models to characterize the tuning properties of the OGs, including each OG's preferred stimulus that maximizes its response. Our key finding is that the current artificial stimuli used to probe OG function—moving dots and bars—are not diverse enough to lead to different prediction performances between models. To overcome this, we synthesize a new, diverse set of stimuli by pitting model against model and generating "controversial" stimuli (Golan, Raju, & Kriegeskorte, 2020) for use in future experiments.

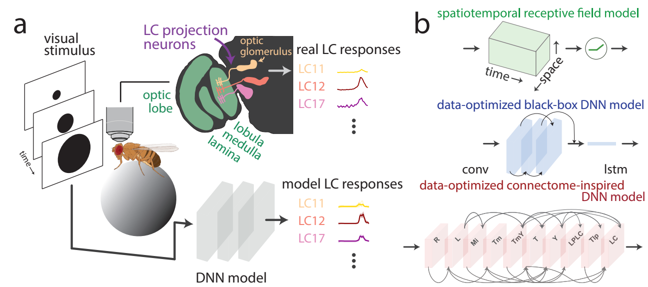

Figure 1: Data-optimized DNN models of fruit fly visual neurons a. We train DNN models with previous recordings of LC neurons that comprise the optic glomeruli to predict OG responses to any stimulus. **b**. We consider three different model architectures.

Model architectures to predict LC responses

We sought a computational model that could predict OG responses from a sequence of images. We considered three classes of models (Fig. 1b): 1) a simple spatiotemporal model, 2) a "black-box" DNN model treated as a function approximator, and 3) a DNN model with a connectome-inspired architecture. Within each class of DNN models, we evaluated a population of models that varied in hyperparameter values.

Spatiotemporal receptive field model

Similar to previous modeling in the primate retina (Pillow et al., 2008), we first identified the spatiotemporal receptive field (STRF) model—implemented as a generalized linear model (GLM)—for each OG. To account for direction selectivity, we used two linear STRFs whose ReLU outputs were linearly combined (i.e., a cascade model).

Data-optimized black-box DNNs

We designed black-box networks purely for accurate prediction. Each consisted of 3 convolutional layers with residual connections and we used either 3-D kernels or a small LSTM network to model temporal dynamics; in the case of the LSTM, the last hidden state was mapped to each OG. We varied the

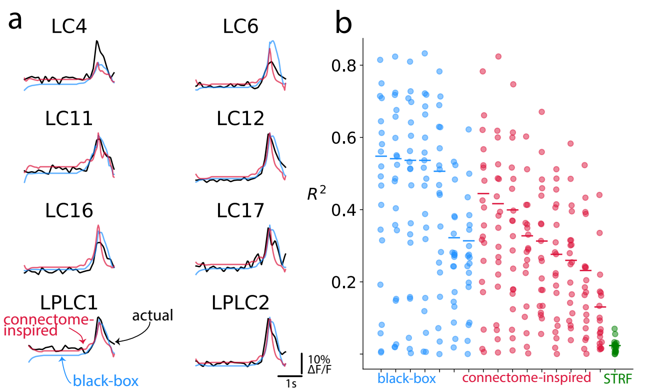

Figure 2: Model predictions a. Predicted responses (color) and real, held-out responses (black) to a looming stimulus. **b**. Held-out R^2 values for each DNN model. Each dot denotes a single OG, and lines denote means.

number of convolutional filters, kernel sizes, and number of hidden layers in the LSTM network to obtain a population of models.

Data-optimized connectome-inspired DNNs

We designed connectome-inspired models whose architecture reflected the anatomy and neuronal connectivity of the highly-organized optic lobe. These models contained convolutional layers to reflect the three main optic lobe regions–lamina, medulla, and lobula–as well as more granular subregions. The number of filters and kernel sizes were chosen to match the number of neuron types at each region based on the FlyWire connectome (Flywire Consortium, 2024); temporal dynamics were captured with 3-D kernels or small LSTMs in putative regions that integrate temporally. We considered a population of models by varying kernel sizes and pathways.

Results

To train and evaluate these proposed models, we curated a dataset of visual stimuli and OG responses comprising data from four different studies (Turner, Krieger, Pang, & Clandinin, 2022; Cowley et al., 2023; Städele, Keleş, Mongeau, & Frye, 2020; Klapoetke et al., 2022). Each model took as input a sequence of 30×30 pixel images (representing the $\sim$900 ommatidia of the retina) from the past 600 ms of visual history of the right optic lobe. OG responses were calcium imaging traces of head-fixed fruit flies viewing stimuli presented on a projection screen. We computed a held-out test R^2 (taken across all held-out stimuli) via cross-validation (Fig. 2).

We found that the best black-box model (mean $R^2 = 0.55$ across all OGs) outperformed the best connectome-inspired model (mean $R^2 = 0.44$). The STRF model, with its two filters, performed poorly (mean $R^2 = 0.02$). Responses of LC25, whose neurons are thought to detect complex bar motion (Klapoetke et al., 2022), were the best predicted for both the best black-box model ($R^2 = 0.81$) and best connectome-inspired model ($R^2 = 0.78$), whereas LC24 (black-box) and LC22 (connectome-inspired) responses were the worst predicted ($R^2 \approx 0$), likely because these LC types had less data than other LCs. We suspect that LC25 was best predicted because the stimulus set was largely biased for moving bars (ideal for identifying tuning of bar detectors) but inadequate for identifying other feature preferences present in natural vision.

The preceding model classes are all *data-driven*–i.e., optimized to best fit recorded neuron activity. In the future, we plan to investigate the predictive performance of task-optimized models of both the black-box (Cowley et al., 2023) and connectome-constrained (Lappalainen et al., 2023) varieties.

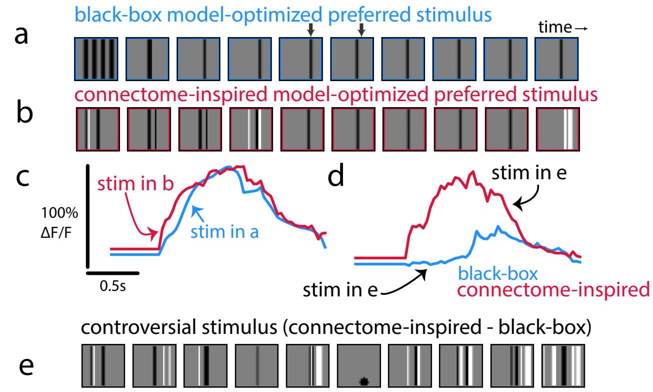

Figure 3: Novel stimuli a. Black-box model-optimized preferred stimulus for LC15. **b**. Connectome-inspired model-optimized preferred stimulus for LC15. **c**. Predicted responses to preferred stimuli. **d**. Predicted responses to the controversial stimulus. **e**. Controversial stimulus that maximizes the disagreement between models for LC15 (connectome-inspired − black-box).

Novel stimulus generation

We used our DNN models to identify the preferred stimulus (i.e., the one that maximizes an OG's response) for each LC neuron type. To find these maximizing stimuli, we took a greedy approach by sequentially selecting the stimulus image for a given time frame that maximized the model's output and then continued to the next frame. For LC15, the most predictive black-box DNN identified a stimulus which prominently features black, laterally moving vertical bars (Fig. 3a), consistent with LC15's detection of bar motion (Klapoetke et al., 2022). Interestingly, the preferred stimulus identified by the most predictive connectome-inspired DNN also had vertical lines but these were both light and dark (Fig. 3b). This motivated us to find a "controversial" stimulus in which the black-box and connectome-inspired models differed the most in their predictions (Fig. 3e). These controversial stimuli are highly informative: by comparing model predictions and real responses to these stimuli (collected in future experiments), we can determine the more accurate model.

References

Cowley, B. R., Calhoun, A. J., Rangarajan, N., Turner, M. H., Pillow, J. W., & Murthy, M. (2023). One-to-one mapping between deep network units and real neurons uncovers a visual population code for social behavior. *bioRxiv*. doi: 10.1101/2022.07.18.500505

Currier, T. A., Pang, M. M., & Clandinin, T. R. (2023, June). Visual processing in the fly, from photoreceptors to behavior. *Genetics*, *224*(2), iyad064. doi: 10.1093/genetics/iyad064

Flywire Consortium. (2024). *FlyWire Connectome.* Retrieved from `https://codex.flywire.ai/`

Golan, T., Raju, P. C., & Kriegeskorte, N. (2020, October). Controversial stimuli: adjudicating between deep neural network models of biological vision with synthetic images. *Journal of Vision*, *20*(11), 947. doi: 10.1167/jov.20.11.947

Hindmarsh Sten, T., Li, R., Otopalik, A., & Ruta, V. (2021, July). Sexual arousal gates visual processing during Drosophila courtship. *Nature*, *595*(7868), 549–553. (Publisher: Nature Publishing Group) doi: 10.1038/s41586-021-03714-w

Klapoetke, N. C., Nern, A., Rogers, E. M., Rubin, G. M., Reiser, M. B., & Card, G. M. (2022, May). A functionally ordered visual feature map in the *Drosophila* brain. *Neuron*, *110*(10), 1700–1711.e6. doi: 10.1016/j.neuron.2022.02.013

Lappalainen, J. K., Tschopp, F. D., Prakhya, S., McGill, M., Nern, A., Shinomiya, K., . . . Turaga, S. C. (2023, March). Connectome-constrained deep mechanistic networks predict neural responses across the fly visual system at single-neuron resolution.
(Pages: 2023.03.11.532232 Section: New Results) doi: 10.1101/2023.03.11.532232

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., & Simoncelli, E. P. (2008, August). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, *454*(7207), 995–999. (Publisher: Nature Publishing Group) doi: 10.1038/nature07140

Strother, J. A., Wu, S.-T., Wong, A. M., Nern, A., Rogers, E. M., Le, J. Q., . . . Reiser, M. B. (2017, April). The Emergence of Directional Selectivity in the Visual Motion Pathway of *Drosophila*. *Neuron*, *94*(1), 168–182.e10. doi: 10.1016/j.neuron.2017.03.010

Städele, C., Keleş, M. F., Mongeau, J.-M., & Frye, M. A. (2020, July). Non-canonical Receptive Field Properties and Neuromodulation of Feature-Detecting Neurons in Flies. *Current Biology*, *30*(13), 2508–2519.e6. doi: 10.1016/j.cub.2020.04.069

Turner, M. H., Krieger, A., Pang, M. M., & Clandinin, T. R. (2022, October). Visual and motor signatures of locomotion dynamically shape a population code for feature detection in Drosophila. *eLife*, *11*, e82587. doi: 10.7554/eLife.82587