Connectome-Constrained Unsupervised Learning Reveals Emergent Visual Representations in the Drosophila Optic Lobe

Keisuke Toyoda (toyoda-keisuke527@g.ecc.u-tokyo.ac.jp)
Department of Systems Innovation, School of Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

Naoya Nishiura (naoya-nishiura@g.ecc.u-tokyo.ac.jp)
Department of Systems Innovation, School of Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

Rintaro Kai (kai-rintaro944@g.ecc.u-tokyo.ac.jp)
Department of Systems Innovation, School of Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

Masataka Watanabe (watanabe@sys.t.u-tokyo.ac.jp)
Department of Systems Innovation, School of Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

Abstract

Understanding how brain structure enables visual processing is crucial. While *Drosophila* offers a complete connectome, computational models often use biologically implausible supervised signals. We address this by building a large-scale autoencoder constrained by the complete *Drosophila* right optic lobe connectome ($\sim$45k neurons, FlyWire dataset). Using photoreceptors (R1-R6) as both input and output, the model incorporates anatomical feedforward and feedback loops and was trained unsupervised on naturalistic video stimuli to minimize reconstruction error. Temporal offsets were included to probe predictive capacity. The autoencoder accurately reconstructed photoreceptor inputs with high fidelity. Deeper layer neurons (medulla, lobula) showed moderate, stable activity under sustained input, consistent with efficient engagement and functional recurrent loops. Temporal offsets improved short-term prediction, indicating learned input dynamics. We demonstrate that a connectome-based autoencoder can learn meaningful visual representations via biologically plausible unsupervised learning. This highlights how anatomical structure shapes emergent function and provides a digital twin framework for studying visual processing beyond task-specific supervised approaches, suggesting complex representations can arise from self-organization on detailed neural circuits.

Keywords: connectomics; unsupervised learning; autoencoder; Drosophila; optic lobe; visual processing; computational modeling; FlyWire

Introduction

The optic lobe functions as the primary visual processing center in *Drosophila melanogaster*, executing critical functions such as motion detection (Borst & Groschner, 2023; Borst, Haag, & Mauss, 2020), which relies on precisely wired neural circuits. Recent advances in connectomics have yielded nearly complete synaptic maps of the fly brain, including the visual system (Scheffer et al., 2020; Lappalainen et al., 2024; Dorkenwald et al., 2024), opening unprecedented opportunities for detailed circuit analysis and computational modeling. A notable recent study utilized a connectome-constrained neural network to replicate the motion detection properties characteristic of T4 and T5 neurons (Lappalainen et al., 2024). However, this model relied on supervised learning employing vector-based teaching signals (e.g., optical flow), which represent information not explicitly available to the biological system.

In contrast, we leverage the comprehensive synaptic connectivity data from the entire right optic lobe, obtained via the FlyWire project (Dorkenwald et al., 2022, 2024), to construct a large-scale autoencoder model. Crucially, the model uses the visual input itself as the teaching signal, mirroring potential self-supervised learning mechanisms in biology and incorporating known feedback pathways (Hu, Hillion, Gu, Hardie, & Juusola, 2015). Our primary objective is to develop a "digital twin" of the *Drosophila* optic lobe that operates under biologically plausible training conditions, thereby investigating how visual information is processed and represented within the constraints of its anatomical structure.

Methods

We constructed the neural network model based on a synaptic adjacency matrix derived from the complete right optic lobe connectome provided by the FlyWire dataset (Dorkenwald et al., 2022, 2024). This resulted in a network comprising approximately 45,000 neuronal nodes interconnected by over 4.5 million synaptic edges, representing the known synaptic partners and strengths. The model was configured as an autoencoder, where the input layer consisted of photoreceptor neurons (R1–R6), and the output layer aimed to reconstruct the activity of these same photoreceptors. The network architecture between the input and output layers strictly adhered to

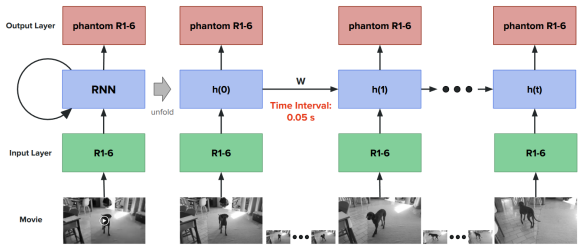


Figure 1: Schematic diagram of the Recurrent Neural Network (RNN) architecture for the autoencoder model. The left side shows the general loop structure of the RNN, and the right side shows its unfolding over time. The input layer (R1-6) receives movie frames, which are processed through the RNN (state $h(t)$, weights W) to generate reconstructed signals at the output layer (phantom R1-6). The time interval shown is 0.05 s.

the connectome data, preserving both feedforward processing pathways and recurrent/feedback connections, including those known to modulate photoreceptor sensitivity (Hu et al., 2015).

The model was trained using sequences of naturalistic visual stimuli (e.g., video clips). During training, synaptic weights were iteratively adjusted using an optimization algorithm (e.g., gradient descent) to minimize the mean squared reconstruction error between the original photoreceptor signals and the signals reconstructed by the network's output layer. To assess the model's predictive capabilities regarding temporal dynamics, each training iteration incorporated slight temporal offsets between the input stimulus presented and the target reconstruction signal.

Following training, the activity patterns of neurons throughout the network were analyzed. Neurons were categorized based on their topological distance (shortest synaptic path length) from the input photoreceptors, allowing for the examination of signal propagation and transformation through successive processing layers, analogous to biological analyses (Borst & Groschner, 2023).

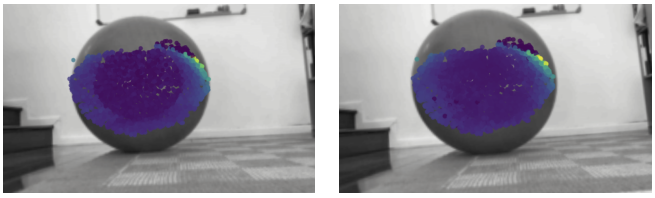
Results

Upon completion of the training phase, the connectome-based autoencoder demonstrated high fidelity in reconstructing the initial photoreceptor input signals. This visual reconstruction quality is illustrated in Figure 2, which compares the input photoreceptor activation (Fig. 2a) with the corresponding reconstructed output activation (Fig. 2b) for an example stimulus frame. Quantitatively, the model achieved a low mean squared error across a diverse range of visual contexts presented during testing.

Analysis of neuronal activity revealed that neurons situated beyond the superficial lamina layers (i.e., in deeper regions like the medulla and lobula) exhibited moderate levels of activation. This indicates that while these deeper circuits were engaged in the reconstruction task, they were not driven to maximal activity levels, potentially reflecting efficient coding strategies.

When the network was presented with prolonged, stable visual stimulation, the activation patterns across many neuronal populations tended to stabilize after an initial transient response. This stabilization suggests the functional engagement of recurrent loops inherent in the optic lobe's connectome, which may serve to dampen excessive fluctuations or adapt neuronal responses, consistent with biological reports on feedback mechanisms modulating photoreceptor gain and sensitivity (Hu et al., 2015).

Furthermore, performance analyses incorporating minor temporal offsets during training showed improved predictive accuracy. This finding hints that the network successfully captured not only static spatial features but also short-term temporal correlations present within the natural visual input streams.



(a) Input frame (photoreceptor activation)

(b) Reconstructed output (phantom photoreceptor activation)

Figure 2: Visual comparison of model input and reconstructed output for an example stimulus frame. (a) Input photoreceptor activation visualized on the stimulus. (b) Corresponding reconstructed activation ('phantom R1-6') generated by the autoencoder. The similarity highlights the model's high reconstruction fidelity.

Discussion

Our findings demonstrate that a large-scale autoencoder model, constrained by the complete anatomical connectivity of the *Drosophila* right optic lobe (Dorkenwald et al., 2022, 2024), can effectively learn to reconstruct its visual inputs using only the sensory information itself as a teaching signal. The high visual fidelity achieved in this reconstruction (exemplified in Figure 2) strongly supports this conclusion. This unsupervised learning paradigm, which incorporates biologically known feedback loops (Hu et al., 2015), supports the hypothesis that significant aspects of visual processing can emerge through self-organization guided by network structure and input statistics. By preserving the detailed anatomical wiring patterns derived from connectomics (Scheffer et al., 2020; Dorkenwald et al., 2024), the model provides a platform to investigate how structural constraints inherently shape functional computations.

Compared to previous modeling approaches that focused on specific computations like local motion detection (Ammer, Leonhardt, Bahl, Egelhaaf, & Borst, 2015) or relied on supervised learning paradigms with potentially biologically implau-

sible teaching signals (Lappalainen et al., 2024), our unsupervised autoencoder method allows for the exploration of emergent neural coding properties without imposing task-specific objectives.

Although neurons in deeper layers displayed only moderate activity levels during the reconstruction task, their consistent engagement suggests a role in hierarchical processing, potentially refining the representation or contributing to the prediction of visual inputs (Borst & Groschner, 2023).

Future investigations could involve dissecting the activity within specific, functionally characterized subnetworks (e.g., those implicated in contrast gain control or specific motion pathways (Scheffer et al., 2020; Hu et al., 2015)) within the trained autoencoder framework. This could clarify how unsupervised learning shapes representations relevant to these tasks and how feedback mechanisms contribute to refining visual perception under varying conditions.

Acknowledgments

This work has been supported by the Mohammed bin Salman Center for Future Science and Technology for Saudi-Japan Vision 2030 at The University of Tokyo (MbSC2030) and JSPS KAKENHI Grant Number 23K25257.

References

Ammer, G., Leonhardt, A., Bahl, A., Egelhaaf, M., & Borst, A. (2015). Functional specialization of neural input elements to the Drosophila on motion detector. *Current Biology*, *25*(17), 2247–2253. doi: 10.1016/j.cub.2015.07.038

Borst, A., & Groschner, L. N. (2023). How flies see motion. *Annual Review of Neuroscience*, *46*, 17–37. doi: 10.1146/annurev-neuro-111021-112408

Borst, A., Haag, J., & Mauss, A. S. (2020). How fly neurons compute the direction of visual motion. *Journal of Comparative Physiology A*, *206*(2), 109–124. doi: 10.1007/s00359-020-01401-y

Dorkenwald, S., McKellar, C. E., Azevedo, A. W., Eckstein, N., Wang, K., Wang, A., ... Scheffer, L. K. (2024). The complete adult drosophila brain connectome. *Nature*. doi: 10.1038/s41586-024-07910-3

Dorkenwald, S., Schubert, P. J., Ayala, A., Póczos, B., Kaynig, V., Lee, K., ... Seung, H. S. (2022). FlyWire: Online community for whole-brain connectomics. *Nature Methods*, *19*(1), 119–128. doi: 10.1038/s41592-021-01330-0

Hu, C., Hillion, A. D., Gu, P., Hardie, R. C., & Juusola, M. (2015). Ih channels control feedback regulation from amacrine cells to photoreceptors in the Drosophila visual system. *PLoS Biology*, *13*(4), e1002115. doi: 10.1371/journal.pbio.1002115

Lappalainen, T. J., Rule, M., Dorkenwald, S., Güneş Baydin, A., Funk, N., Turning, L., ... Turaga, S. C. (2024). Connectome-constrained networks predict neural activity across the fly visual system. *Nature*, *634*(8036), 1132–1140. doi: 10.1038/s41586-024-07990-z

Scheffer, L. K., Xu, C. S., Januszewski, M., Lu, Z., Takemura, S.-y., Hayworth, K. J., ... Rubin, G. M. (2020). A connectome and analysis of the adult Drosophila central brain. *eLife*, *9*, e57443. doi: 10.7554/eLife.57443