The landscape of functional neuron embeddings depends on regularization

Polina Turishcheva (turishcheva@cs.uni-goettingen.de)
Institute of Computer Science and Campus Institute Data Science, University of Göttingen
Goldschmidtstr. 1, 37077 Göttingen, Germany

Viktor Dobrev (viktor.dobrev@stud.uni-goettingen.de)
Institute of Computer Science and Campus Institute Data Science, University of Göttingen
Goldschmidtstr. 1, 37077 Göttingen, Germany

Max Burg (max.burg@uni-goettingen.de)
Institute of Computer Science and Campus Institute Data Science, University of Göttingen
Goldschmidtstr. 1, 37077 Göttingen, Germany;
International Max Planck Research School for Intelligent Systems
Max-Planck-Ring 4, 72076 Tübingen, Germany;
Tübingen AI Center, University of Tübingen
Maria-von-Linden-Straße 6, 72076 Tübingen, Germany

Michaela Vystrčilová (michaela.vystrcilova@uni-goettingen.de)
Institute of Computer Science and Campus Institute Data Science, University of Göttingen
Goldschmidtstr. 1, 37077 Göttingen, Germany

Alexander Ecker (ecker@cs.uni-goettingen.de)
Institute of Computer Science and Campus Institute Data Science, University of Göttingen
Goldschmidtstr. 1, 37077 Göttingen, Germany;
Max Planck Institute for Dynamics and Self-Organization
Am Faßberg 17, 37077 Göttingen, Germany.

Abstract

Understanding the functional landscape of neurons is crucial for developing a taxonomy of neuronal cell types. Recent work proposed an approach to identify functional cell types by learning a predictive model that approximates the input-output function of a population of neurons and represents each neuron's function by an embedding. These neurons' embeddings have been used to investigate the landscape of cortical computation in the early visual system, but it remains unclear how the structure of the embedding space depends on the design choices of the predictive model. There were two major differences in architectures: (1) a change of spatial sampling strategy for neurons receptive field; and (2) using dynamic video stimuli instead of static images. Here we investigate the impact of such design choices on the functional landscape in the mouse primary visual cortex. We find that strong L1 regularization of the final linear layer, essential for earlier models, is vital for structured embeddings, even with more recent architectures that do not require regularization to achieve strong predictive performance. Varying the backbone architecture did not significantly impact the embeddings structure. Overall, our work is an important step towards interpretable brain modeling and taxonomy of cell types in the visual system.

Keywords: neural response modeling; visual cortex; representational learning; cell types

Introduction

Recent work proposed a framework for functional cell type identification (Ustyuzhaninov et al., 2022; Tong et al., 2023). The idea is to train a deep neural network with a two-stage architecture (core + readout; Fig. 1) to predict how a large population of neurons responds to arbitrary visual stimuli (Klindt, Ecker, Euler, & Bethge, 2017). The *core* is shared among all neurons and maps the visual input to a shared feature space (Fig. 1, left). From this feature space, the *readout* predicts a neuron's response by taking a (neuron-specific) linear combination (Fig. 1, right). These linear readout weights can then be thought of as an embedding of the neuron's input-output function and have been used to map the landscape of neuronal function in the visual cortex (Ustyuzhaninov et al., 2022; Wang et al., 2023). However, the organization of the resulting embedding space appears to depends on architectural choices: earlier work (Ustyuzhaninov et al., 2022) showed a clustered embedding space where neurons in high-density modes share functional properties, whereas more recent, high-performing models result in much less structured embedding spaces (Wang et al., 2023). Here we investigate how architecture differences could cause these differences.

Methods

Model architectures. Neurons in primary visual cortex are orientation-selective. To obtain neuronal embeddings invariant to neurons' preferred orientation, we use a rotation equivariant framework (Fig. 2) (Ecker et al., 2018). We compare

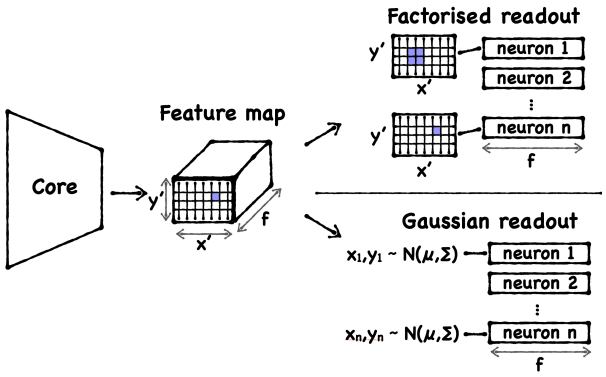

Figure 1: **Core-readout framework.** The core consists of four convolutional layers and outputs a $(y \times x \times f)$ tensor, where x is width, y is height, and f is channels. The readout selects a receptive field location for each neuron and computes a linear combination of the features to predict neuronal responses.

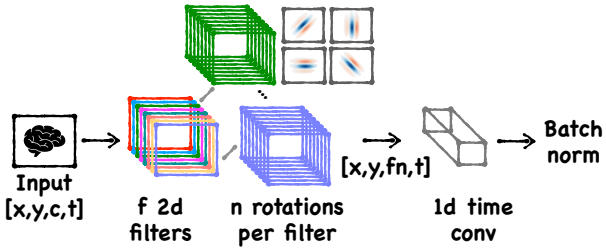

Figure 2: **Rotation-equivariant layer in the dynamic core.** We modified dynamic models from (Hoefling et al., 2022; Vystrčilová et al., 2024) by making the space convolutions rotation equivariant. Each of the f learnt features is rotated n times by $360/n$ degrees, resulting in $f \cdot n$ output channels. Batch norm is applied across f dimensions only. Scale and bias of the last layer's batch norm are not trainable as this would interfere with the readout regularisation.

two types of readout mechanism, both of which disentangle the readout into a spatial component representing the neuron's receptive field location and a vector of feature weights – the neuron's embedding, which represents its nonlinear computation. The earlier *factorised readout* (Klindt et al., 2017) learns a mask for each neuron. It requires strong $L1$ regularisation to learn a sparse mask, and regularizes the spatial mask jointly with the neuronal embedding. The more recent *Gaussian readout* (Lurz et al., 2020) represents each neuron's receptive field location as (x, y) coordinates and does not necessarily require regularization. Here we explore how the strength of regularisation γ affects both readouts. Another architecture choice is the regularization of the convolution kernels in the core. We investigate two: smoothness via a Laplace filter (γ_{inp}) and group sparsity (γ_{group}) (Ecker et al., 2018). After training, we aligned embeddings to ensure rotation invariance (Ustyuzhaninov et al., 2019) and clustered them using k-means with 30 clusters.

Training data. The static model was trained on responses to natural images of seven mice from the Sensorium Competition 2022 (Willeke et al., 2022). For the dynamic model, we used the responses to natural video of ten mice from the Sensorium Competition 2023 (Turishcheva et al., 2023). Our models account for behavioural activity (locomotion speed, pupil dilation), a known modulator for neuronal responses, by adding the behavioural variables as input channels to the core.

Evaluation. Following Turishcheva et al. (2023), we use single-trial correlation (ρ_{st}). ρ_{st} is computed independently per neuron and then averaged.

Results

First, the degree of clustering of the embeddings depends primarily on the strength γ of readout regularisation, not the type of readout (Fig. 3). However, this comes at a cost of decreasing performance (Fig. 3 E). Second, neither making the core dynamic (Fig. 4) nor varying the core regularization hyperparameters (Fig. 5) affects the presence of high-density modes.

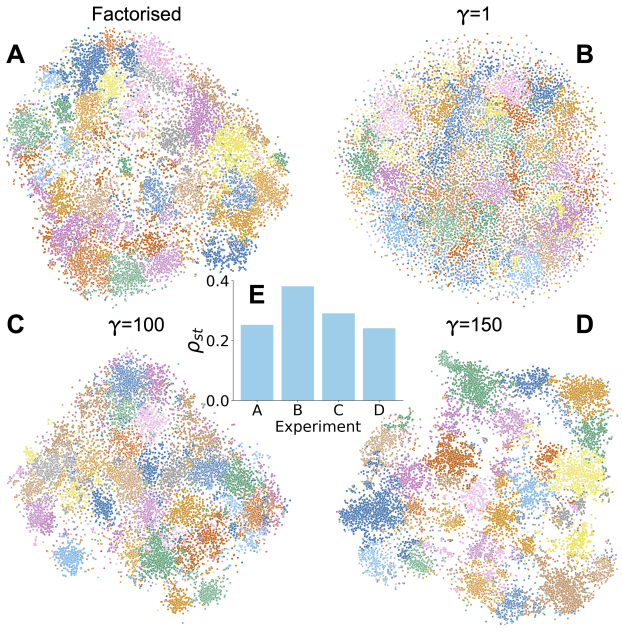

Figure 3: **t-SNE in static case**. **A–D:** Models with 8 channels, 8 rotations. t-SNE following Kobak and Berens (2019). 14000 neurons, 2000 per animal, same neurons are used across pictures. Each color corresponds to a cluster from k-means. **A:** Factorised readouts with best performing regularisation. **B–D:** Gaussian readouts. **E:** ρ_{st} score.

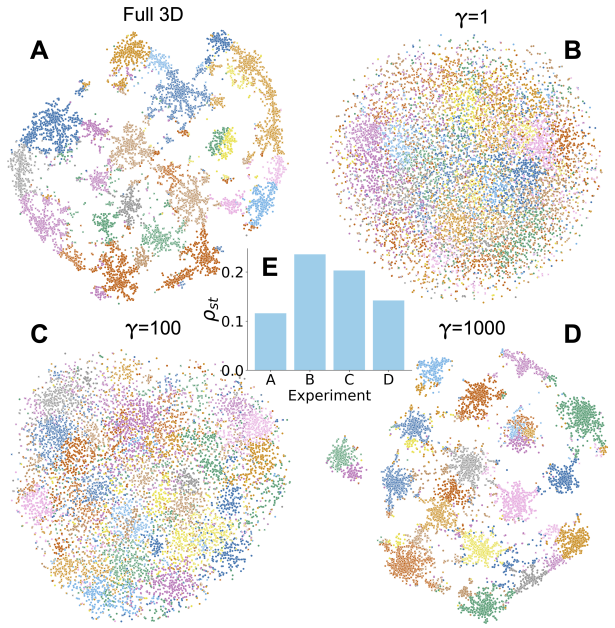

Figure 4: **t-SNE in dynamic case**. **A–D:** 10 animals, 1000 neurons per animal. 16 channels, 8 rotations and gaussian readouts. **A:** full 3d convolutions (γ 500), **B–D:** factorised convolutions. **E:** ρ_{St} score. As in the static case, readout regularisation induces structure at the cost of performance.

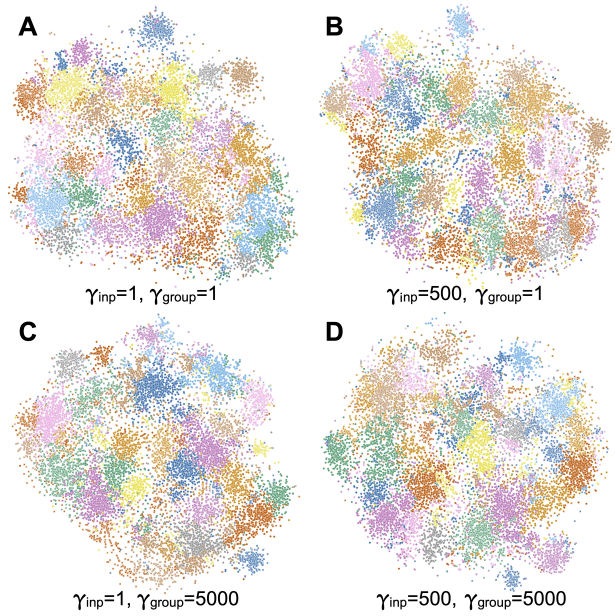

Figure 5: **Regularization of the core**. Two different regularizers have been employed for the convolution kernels of the core: smoothness via a Laplace filter (γ_{inp}) and group sparsity (γ_{group}) (Ecker et al., 2018). **A–B:** Changing γ_{inp}. **C–D:** Changing γ_{group}. Both don't influence the presence of density modes or performance. All models with Gaussian readout and readout regularisation $\gamma = 100$.

References

Ecker, A. S., Sinz, F. H., Froudarakis, E., Fahey, P. G., Cadena, S. A., Walker, E. Y., ... Bethge, M. (2018). A rotation-equivariant convolutional neural network model of primary visual cortex. *arXiv preprint arXiv:1809.10504*.

Hoefling, L., Szatko, K., Behrens, C., Qiu, Y., Klindt, D., Jessen, Z., ... others (2022). A chromatic feature detector in the retina signals visual context changes. biorxiv (2022). *Google Scholar*.

Klindt, D., Ecker, A. S., Euler, T., & Bethge, M. (2017). Neural system identification for large populations separating "what" and "where". *Advances in neural information processing systems*, *30*.

Kobak, D., & Berens, P. (2019). The art of using t-sne for single-cell transcriptomics. *Nature communications*, *10*(1), 5416.

Lurz, K.-K., Bashiri, M., Willeke, K., Jagadish, A. K., Wang, E., Walker, E. Y., ... others (2020). Generalization in data-driven models of primary visual cortex. *BioRxiv*, 2020–10.

Tong, R., da Silva, R., Lin, D., Ghosh, A., Wilsenach, J., Cianfarano, E., ... Trenholm, S. (2023). The feature landscape of visual cortex. *bioRxiv*. doi: 10.1101/2023.11.03.565500

Turishcheva, P., Fahey, P. G., Hansel, L., Froebe, R., Ponder, K., Vystrčilová, M., ... others (2023). The dynamic sensorium competition for predicting large-scale mouse visual cortex activity from videos. *arXiv preprint arXiv:2305.19654*.

Ustyuzhaninov, I., Burg, M. F., Cadena, S. A., Fu, J., Muhammad, T., Ponder, K., ... others (2022). Digital twin reveals combinatorial code of non-linear computations in the mouse primary visual cortex. *bioRxiv*, 2022–02.

Ustyuzhaninov, I., Cadena, S. A., Froudarakis, E., Fahey, P. G., Walker, E. Y., Cobos, E., ... others (2019). Rotation-invariant clustering of neuronal responses in primary visual cortex. In *International conference on learning representations*.

Vystrčilová, M., Sridhar, S., Burg, M. F., Gollisch, T., & Ecker, A. S. (2024). Convolutional neural network models of the primate retina reveal adaptation to natural stimulus statistics. *bioRxiv*. doi: 10.1101/2024.03.06.583740

Wang, E. Y., Fahey, P. G., Ponder, K., Ding, Z., Chang, A., Muhammad, T., ... Tolias, A. S. (2023). Towards a foundation model of the mouse visual cortex. *bioRxiv*. doi: 10.1101/2023.03.21.533548

Willeke, K. F., Fahey, P. G., Bashiri, M., Pede, L., Burg, M. F., Blessing, C., ... others (2022). The sensorium competition on predicting large-scale mouse primary visual cortex activity. *arXiv preprint arXiv:2206.08666*.