Towards end-to-end cell-typing in large-scale recordings

[**Summary**] Reliably identifying cell types is crucial for understanding neural systems (*Zeng & Sanes, 2017*; *Weis et al., 2023*). However, classifying the neurons recorded in large-scale population recordings into well-defined types remains a challenge. While some methods rely on genetic targeting or anatomy for classification, another approach employed in sensory systems is to analyze the responses of neurons to stimuli that elicit discriminable responses in different types (*Vlasits et al, 2019*; *Baden et al., 2016*). Previous approaches of this nature have been limited by a restricted set of stimuli and scalability issues. In this study, we present a novel, *end-to-end* approach for the functional classification of retinal ganglion cells (RGCs), the retinal output, that combines recording responses to standardized stimuli with modeling using Convolutional Neural Networks (CNNs). Initially, we recorded the responses of a large set of ganglion cells to a novel noise stimulus that incorporated multiple spatial frequencies (**MSF**). These data were then used as a training set to learn and predict how each recorded neuron would respond to a variety of standardized stimuli, which can be used to determine cell types. Notably, previous research has demonstrated that surround properties can distinguish many cell types (*Farrow & Roska, 2013*; *Goetz et al., 2022*). Our method successfully predicted surround suppression in several cell types and significantly outperformed previous methods used for this purpose. Therefore, our approach serves as a powerful tool for determining RGC types in large-scale recordings. While classifications based on transcriptomic analysis (*Tran et al., 2019*), anatomical features (*Bae et al., 2020*), and functional characteristics (*Baden et al., 2016*) have been proposed in the retina, our method suggests a general approach for performing classification in population recordings by comparing the models learned for each cell.

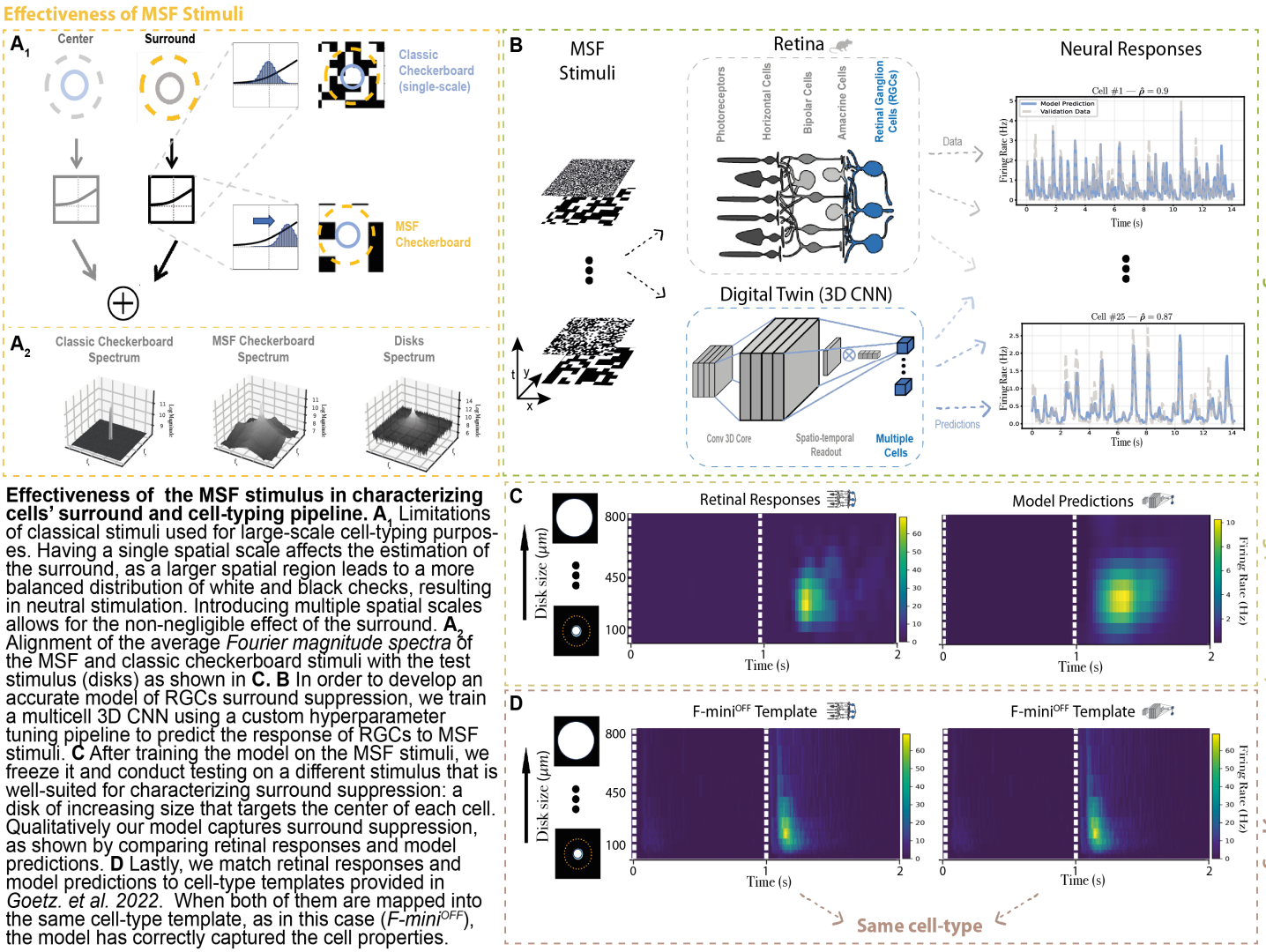
[**Additional details**] Mouse RGCs exhibit a response to variations in light primarily within their receptive field (RF) center, a specific region of the visual field. Additionally, many RGCs possess a suppressive surround, which refers to the area surrounding the RF center. When a light stimulus covers both the center and surrounding areas, the cell's response is weaker compared to when the stimulus is limited to the RF center. This characteristic is prevalent among ganglion cells and has been utilized to differentiate various types of ganglion cells in the mouse retina (*Goetz et al., 2022*). While estimating the suppressive surround for single-cell recordings is relatively straightforward, it becomes more challenging when dealing with a population of ganglion cells recorded using electrode arrays or calcium imaging. This is because each cell's receptive field is located in a different position. Traditional methods of quantifying receptive fields in these large-scale recordings, such as reverse correlation from white noise, tend to underestimate the amount of surround suppression.

[**Originality**] To address these challenges, our work introduces several novel contributions. **1)** Firstly, we present a new approach to quantify the suppressive surround of mouse RGCs using large-scale multi-electrode array recordings and a modified white noise stimulus, the MSF (see **A** for intuitions on its effectiveness). **2)** Secondly, we introduce an optimized Bayesian hyperparameter tuning and neural architecture search pipeline that enables efficient selection of a high-performance CNN model (see **B**). **3)** Lastly, we provide a way to map the model's predictions on a testing task to cell-typing templates, as outlined in (*Goetz et al., 2022*), effectively establishing an end-to-end connection from the input (the stimulus) to the output (predicted cell type). Template matching is performed by means of a time-shift invariant (biologically relevant delays of *100 ms*) algorithm which is also robust to firing rate scalings in intensity (see results in **C**). Finally, this framework allows us to integrate transcriptomic (*Tran et al., 2019*), anatomical (*Bae et al., 2020*), and functional (*Baden et al., 2016*) methods without the need for human expertise in labeling cell types.

[**Significance**] Our study aims to tackle the challenges associated with quantifying the suppressive surround in mouse retinal ganglion cells (RGCs) and provide an integrated approach for determining cell types. To achieve this, we leverage large-scale recordings and employ advanced modeling techniques, specifically

employing spatio-temporal 3D CNNs. The classification of cell types is not only crucial in systems neuroscience but also facilitates the development of more accurate computational models of sensory systems. This is because there exists a profound connection between RGC families and the features, to which different cells are selective or sensitive in a stimulus *(Kerschensteiner 2022)*. Moreover, beyond its potential applications in retinal research, our method holds promise for investigating the suppressive surrounds of other types of neurons, such as those found in the visual cortex.

[**Relevance**] This work lies at the interface of systems neuroscience, in this case understanding the function of RGCs, i.e. determining cell-types; and computational neuroscience, by employing CNN models learned on cell responses. More than that, it establishes connections between the two fields of interest for Cosyne. In our approach, the conception of the MSF stimuli takes into account how on the one hand, *biologically*, RGCs surround response detection can benefit from introducing multiple spatial scales and how, on the other hand, the retina extracts information from second order statistics (Fourier magnitude spectrum). From the *modeling* perspective, we investigate how to learn meaningful features that can be transferred across tasks involving multiple spatial scales. This is exemplified by how the magnitude spectra of the MSF stimuli and disks are aligned (see **A**), i.e. it is easier from a machine learning viewpoint to generalize in a task transfer setting if in the training set the model has seen enough statistics which is shared between the two tasks (*Tripuranemi et al. 2020*): predicting responses to MSF and disks. In conclusion, this work presents a compelling conjunction between experiments and modeling, offering valuable insights for the Cosyne audience.



Effectiveness of the MSF stimulus in characterizing cells' surround and cell-typing pipeline. A$_1$ Limitations of classical stimuli used for large-scale cell-typing purposes. Having a single spatial scale affects the estimation of the surround, as a larger spatial region leads to a more balanced distribution of white and black checks, resulting in neutral stimulation. Introducing multiple spatial scales allows for the non-negligible effect of the surround. **A**$_2$ Alignment of the average *Fourier magnitude spectra* of the MSF and classic checkerboard stimuli with the test stimulus (disks) as shown in **C. B** In order to develop an accurate model of RGCs surround suppression, we train a multicell 3D CNN using a custom hyperparameter tuning pipeline to predict the response of RGCs to MSF stimuli. **C** After training the model on the MSF stimuli, we freeze it and conduct testing on a different stimulus that is well-suited for characterizing surround suppression: a disk of increasing size that targets the center of each cell. Qualitatively our model captures surround suppression, as shown by comparing retinal responses and model predictions. **D** Lastly, we match retinal responses and model predictions to cell-type templates provided in *Goetz. et al. 2022*. When both of them are mapped into the same cell-type template, as in this case (*F-mini*^{OFF}), the model has correctly captured the cell properties.