Rapid dynamics in the retina

Baptiste Lorenzi

CentraleSupélec, Unversité Paris Saclay, Institut de la Vision

Abstract

We trained a convolutional neural network (CNN) model on large-scale functional recordings of RGC responses to natural mouse movies, and then used this model to investigate the role of past images in the activity of RGC.

Our work showcases how a combination of experiments with natural stimuli and computational modelling allows discovering ...

1 Introduction

During my internship in Olivier Marre's team at l'Institut de la Vision, I am focusing on computational modeling of the retina. Olivier Marre's team is an interdisciplinary laboratory, hosting four professors and a dozen of interns, Ph.D. students and post-docs, working hand in hand to advance research on the retina. They all have various backgrounds mainly from biology, theoretical physics and engineering. In the context of this project, I've been working closely with Samuele Virgili, a third-year Ph.D. student, whose previous and current projects all focus on the modeling of retinal ganglion cells.

The ability of the visual system to process complex stimuli on different temporal and spatial scales is remarkable. Natural environments are such complex stimuli, and extracting the relevant features at all times is crucial for many species.

Both the accessibility and apparent complexity of the retina make it a perfect candidate for the study of the front-end of visual processing [Gollisch and Meister, 2010]. In

the mouse, the retina is composed of more than 30 parallel feature channels, embodied by ganglion cell types. Through their axons, the optic nerve, they provide information to numerous visual areas in the brain. A few channels are active in the encoding of basic features including luminance changes and motion, that are only combined in more downstream areas. Other channels however are known to play a role in the extraction of specific features of natural scenes.

An important feature of natural scenes is the diversity of contrast levels in one scene. This diversity can be both spatial (plain and detailed object side to side) and temporal (sudden apparition of an object). It is known that the visual system adapts to the diversity of fluctuations in the scene. This adaptation begins in the retina. By probing the retina with flickers of different intensities, Baccus and colleagues revealed two distinct phases of sensitivity change in the retina: one fast (<0.1s) and one slow (over 10s) [Baccus and Meister, 2002]. In another study, Garvert et al. revealed that contrast adaptation also happens at different scales, either at the whole scene scale (global contrast adaptation) or within one ganglion cell receptive field, the part of the visual field that the cell receives inputs from (local contrast adaptation) [Garvert and Gollisch, 2013].

Still, we currently lack an explanation of the features extracted by many channels and how they adapt to diverse scenes. One of the historical reasons for this is that synthetic stimuli used to study retinal responses are not complex enough to activate these channels. Hence, they cannot uncover critical response properties encountered in natural environments.

In practice, Karamanlis and colleagues [Kim et al., 2020] have probed a larger complexity in retinal spatial non-linearities thanks to stimuli capturing the statistics of natural environments. As those non-linearities cannot be captured by Linear-Nonlinear (LN) models, convolutional neural networks (CNNs) have become the state-of-the-art approach

for predictive modeling of visual processing, not only in the retina but also in higher visual areas [McIntosh et al., 2017].

Insight on methods Here, we combined the power of CNN-based modeling with large-scale mutli-electrode array (MEA) recordings from RGCs to investigate the mechanisms of fast adaptation in the retina under natural stimulus conditions. To this end, we recorded RGC responses to flashed images paired together. Each pair is composed of a synthetic adaptation image followed by a natural image. We were able to identify different trends in the responses of RGCs to natural images, depending on the adaptation image.

To investigate the diversity of this adaptation process and its implementation, we paired deep convolutional models with more traditional modeling. We trained a CNN model on RGC responses to a movie of flashed images. After training, we study how this model generalized to images after being adapted with patterns it wasn't trained to. By tweaking part of the model at the inference level, We hope to show that temporal mechanisms such as gain-control play a major role in the fast adaptation of RGCs in a natural context.

2 Background

The retina is part of the central nervous system in vertebrates. It is made of only a handful of layers of neurons. Its first layer is composed of photo-sensitive neurons called photoreceptors, that act as light sensors for the network. They give their excitatory output to bipolar cells, which can be divided into 14 different types and each type responds differently to the same stimulus, allowing for a vast functional diversity. Bipolar cells excite in turn ganglion cells, which finally send the pre-processed visual information to the rest of the brain through the optic nerve. Ganglion cells can also be divided into different

functional types (at least 32) and each type is believed to extract a different feature from the visual scene. The retina also has two classes of inhibitory neurons, horizontal and amacrine cells, that further modulate the processing of excitatory cells. Compared to the rest of the brain, its relative simplicity and its relatively easy experimental accessibility make the retina an ideal neural tissue to study using computational models.

Adaptation in the retina To operate optimally in a wide range of stimulation conditions, the retina adapts its responses to the statistics of the visual scene. In particular, it was observed to adapt both to the average luminance (stimulus average) and the average contrast (stimulus distance from the average or variance).

Visual system can function over a wide range of light intensities, from starlight to a bright sunny day – a luminance range of 10 10 The retinal adaptation to the luminance of the scene is quite simple by nature. For instance, it is known that the retina uses different neuronal pathways at low and high luminance. Rods and their retinal neuronal channels cover the dimmest light while cones facilitate contrast, color and motion discrimination but only in brighter light.

Contrast adaptation, by comparison, is harder to study. It was always studied through the use of simple stimuli. Contrast adaptation is known to have different timescales. While slower contrast adaptation ($\approx 10s$) is better understood, fast adaptation (<1s) is more complex to study. It is still unclear how it affects temporal processing and the sensitivity to stimulus features[Baccus and Meister, 2002]. Furthermore, contrast adaptation can also happen at different scales, either at the whole scene scale (global contrast adaptation) or within one ganglion cell receptive field, the part of the visual field that the cell receives inputs from (local contrast adaptation) [Garvert and Gollisch, 2013]. Local contrast adaptation is especially relevant in understanding how ganglion cells respond to natural images since these stimuli are full of spatial details like edges in which two contrast levels

appear simultaneously. Such images are challenging to use, as they can't be summed up to a few statistics easily.

3 Methods

Retinal recordings. Since I don't realize any experiments myself, I will try here to give as few details as necessary for the understanding of the rest of the work. Still, experimental recording of the retina is a very interesting topic and the reader can look here for more information. The laboratory has access to three experimental rooms that enable state-of-the-art experimentation on the retina. For this project, we record the activity of retinal ganglion cells using a multi-electrode array. The retina is placed on a ???

Stimuli design. The stimuli used in this project are composed of two images, one synthetic adaptation image followed by a natural image. Adaptation images are taken from a pull of three different patterns: a grey screen used as control, a checkerboard of X*X checks and the same checkerboard with inverted colors (Fig. TO ADD). Natural images are taken from ADD REFERENCE. XXXX images were used for training the CNN, 10 were used to test the CNN and among them, 3 were used to record an estimation of the LSTA of each cell. Adaptation and natural images are always paired together to form a single stimulus pair, also referenced as a clip. Each frame is XXXxXXX pixels wide and each clip is 2*400ms long.

The training set is composed of XXXX clips, each composed of the grey adaptation image followed by a natural image. The test set is composed of 30 clips, each repeated 30 times. The test clips are composed 10 different natural images preceded by each adaptation (3 different clips for each natural image). The dataset used to record LSTA is composed of 9 different clips repeated 1000 times. Each clip is composed of one of the

three selected natural images preceded by one of the adaptation patterns.

We first used 4 different natural images while computing the LSTA of each cell, each 12 different clips being repeated 12 times. We found that the estimation of the LSTA was too unstable with only 750 repetitions. In following experiments we excluded the image that yielded the least amount of stable estimations of LSTA (20% average success rate as compared to 42% average success rate for the other three images). We then used 3 different natural images while computing the LSTA of each cell, each 9 different clips being repeated 1000 times. We found that the estimation of the LSTA was stable with 1000 repetitions.

Data processing. Multi-electrode array experimental data takes the shape of a collection of temporal electrical signals tiling the recorded area. In most scenarios, including here, these signals are sorted into different cell signals using a semi-automatic algorithm. This algorithm is based on the shape of the electrical spikes as well as their spatial location. This process is quite messy due to the low signal-to-noise ratio in the data and each experiment needs to have its sorting corrected by hand. This process can take up to an entire day for a single experiment. I used spiking-circus for semi-automatic spike sorting and the UI phy for handmade corrections. [CITE] It is important to note that even though the retina is an easier organ than most to record clean spike signals from, the data is still very noisy and the sorting process is not perfect. Hence, when validating hypotheses, cells are easily rated by their reliability.

After spike sorting, we can use some standard stimuli to characterize each ganglion cell receptive field. To this end, we display a random binary checkerboard for approximately 1 h at 30 Hz. Check size is 42. A ganglion cell receptive is computed as its spike trigger average (STA), for this checkerboard stimulus. The STA of a cell can also be described as the stimulus that triggers the most spikes from that cell. It is computed as the average

of the presented checkerboard weighted by the number of spikes using a set number of samples per repetition (here 21). The spatial STA is usually shown as the 2-dimensional spatial slice at the maximum value after smoothing. Temporal STA is the one dimensional time slice at the pixel with the maximum value. For smoothing, a double Gaussian can be fitted on the resulting spatial STA.

Natural images. We used a subset of the Open Access van Hateren Natural Images Dataset [ADD TO BIB]. It consists of monochromatic and calibrated (perfect mapping from pixel value to luminance) images of diverse natural environments. These images need to be preprocessed to avoid triggering the adaptation to different ranges of light intensities in the retina, which would call unwanted dynamics. First, images with numerous saturated pixels were not included in our subset. Using a custom procedure previously developed in the laboratory, we ensured the images were normalized in the mean luminance and the root mean square (RMS) contrast.

LSTA. To record the local specoficity of the response of a ganglion cell to a natural image, we used a method called local spike trigger average (LSTA) for its analogy with the STA. This method was previously developed in the laboratory. We first generate a set of perturbed natural images by superimposing some of the natural images (3-4 images) with various perturbation patterns in the form of random checkerboards. We once again used a checker size of 42. Following calibration guidelines measures in previous experiments, the amplitude of the perturbation was set to 12.5%, where 100% corresponds to a pixel value of 1. In the mouse retina, this amplitude was found to trigger a change in firing rate of approximately 1.5Hz in ganglion cells with high firing rates to the unperturbed images [Goldin et al., 2022].

Blablable to share my programming skills to help and improve the data pipeline of the laboratory.

Modeling. This should be the main part of my internship and also the most challenging. We are designing a dynamical model of the retinal fast adaptation. In fact, we mostly look at the evolution of the response from an image to another, meaning that the dynamic we observe only spans two points in time. This reduction makes the model more realistic to study. Most of this job can be summarized as model design, python programming, sensitivity analysis and data fitting. By comparing how different modeling strategies reproduce the observed LSTA in the data, we can gain insight on how fast adaptation to natural scene is implemented in the retina.

Our baseline model is the LNLN model of ganglion cell widely used in the literature. Each neuron is encoded as a spatial linear filter chained with a non-linearity (usually an activation linearity in the like of ReLU). A single layer of subunit neurons, representing bipolar cells, converge into a single modeled ganglion cell. We would like to add temporal dynamics to this model, either by adding a time dimension to the spatial liner filter of the cells or by considering a gain control mechanism. This last mechanism consists in scaling up or down the present output depending on past outputs (Figure TO FIND).

We will first study our models in a data agnostic manner and study its behavior for different set of parameters. We will then fit it on our own experimental data using an efficient optimization framework in python using strategies developed in the field of machine learning.

4 Results

Some idea of speech: *** responses of exemplary RGCs

Here, we investigated fast adaptation in the mouse retina under natural stimulus conditions. To this end, we trained a CNN model on RGC responses to a movie of flashed images appearing naturally in the mouse environment,

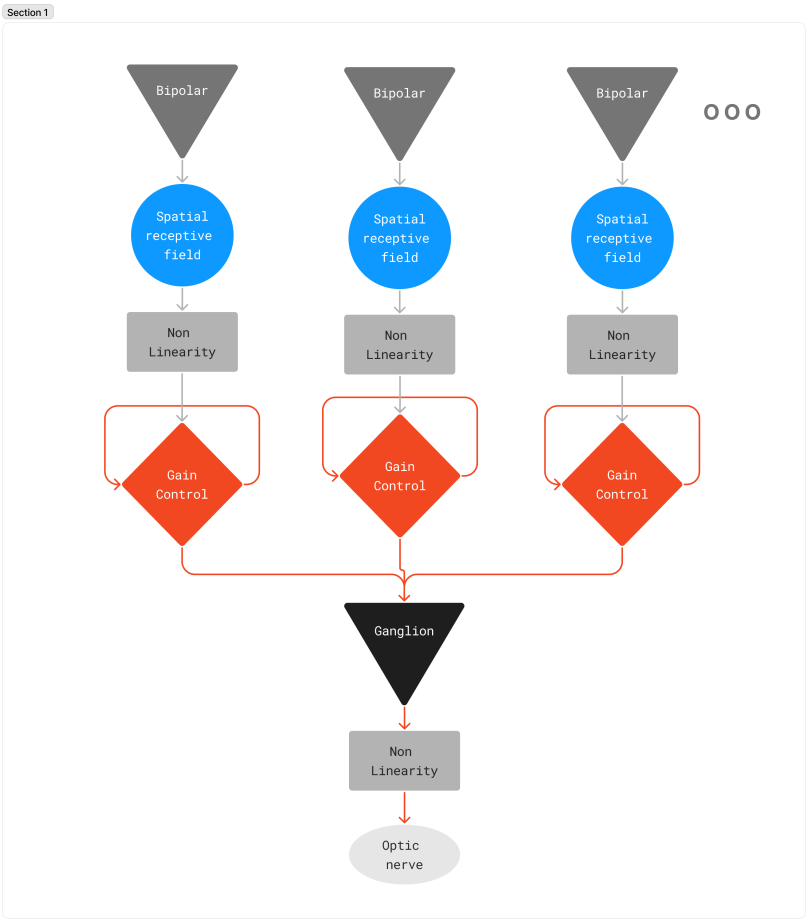

Figure 1: **Quick sketch of a gain control LNLN model.** Each bipolar cell is composed of a linear spatial filter that selectively responds to part of the scene, a non-linear activation function, and a gain control mechanism that scale its output depending on past events. They all converge into on bipolar cell (forming its receptive field) of which output is also modeled using a non-linear function.

5 Conclusion

We observed that during natural image stimulation, many ganglion cells change their selectivity for different parts of the images depending on previous light patterns.

We combined large-scale recordings of RGC responses to natural movie stimulations with CNN-based modeling to investigate such mechanisms of fast contrast adaptation in the retina.

Under realistic regularization constraints, the CNNs learned a structure similar to retinal pathways, where a ganglion cell activation is the result of a pooling of local subunits of different types in a specific area in space, the ganglion cell receptive field.

Our result supports the idea that retinal ganglion cells encode both spatial and temporal features of natural scenes on a local scale. Previous works have described those features to be encoded as features of the retinal response (latency, firing rate). We might be able to support this theory with further comparison of measured responses with predicted responses of our model.

TO DO

Maheswaranathan and colleagues [Maheswaranathan et al., 2023] have recently been able to predict different aspect of encoding in the retina using deep comvolutional network. In comparison, our experimental approach of estimating the LSTA allow a direct comparison from the model to the data. Classical estimation of performance can't describe what is missed in the prediction, while our qualitative comparison might be able to.

The modeling of retinal responses to natural stimuli has improved our understanding of complex retinal processing. In a recent review, Karamanlis and colleagues /citep, suggested three perspectives of study on the retinal encoding of natural scenes: The

circuit perspective ('How is the retinal code implemented?'), the normative perspective ('Why is it complimented this way?) and the coding perspective ('What is the code used by the retina?'). In this work, We focus on the 'what'. By exploring the response of the retina to a portion of the spatio-temporal stimuli space we can gain insight into the code used by the retina on that subspace. To explore further the 'how' perspective, one would need to study how the different known types of cells in the retina participate in that encoding. This poses the challenge of bridging the typing of cells from functional and anatomical perspectives. The normative perspective has also been explored using deep CNNs with anatomically realistic constrained. It is likely that species with simpler cortical circuitry, as mice, have a stronger need for upstream feature extraction, in the retina. In opposition, species with computationally powerful cortexes such as primates can deal with more faithful and linear representations of the visual inputs. Some studies admirably developed approaches that allow investigation of retinal processing from all three perspectives [ADD CITE].

References

Stephen A. Baccus and Markus Meister. Fast and slow contrast adaptation in retinal circuitry. *Neuron*, 36(5):909–919, December 2002. ISSN 0896-6273. doi: 10.1016/s0896-6273(02)01050-4.

Mona M. Garvert and Tim Gollisch. Local and global contrast adaptation in retinal ganglion cells. *Neuron*, 77(5):915–928, March 2013. ISSN 1097-4199. doi: 10.1016/j.neuron.2012.12.030.

Matías A. Goldin, Baptiste Lefebvre, Samuele Virgili, Mathieu Kim Pham Van Cang, Alexander Ecker, Thierry Mora, Ulisse Ferrari, and Olivier Marre. Context-

dependent selectivity to natural images in the retina. *Nature Communications*, 13 (1):5556, September 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-33242-8. URL https://www.nature.com/articles/s41467-022-33242-8. Number: 1 Publisher: Nature Publishing Group.

Tim Gollisch and Markus Meister. Eye smarter than scientists believed: Neural computations in circuits of the retina. *Neuron*, 65(2):150–164, January 2010. ISSN 0896-6273. doi: 10.1016/j.neuron.2009.12.009. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3717333/.

Young Joon Kim, Nora Brackbill, Ella Batty, JinHyung Lee, Catalin Mitelut, William Tong, E J Chichilnisky, and Liam Paninski. Nonlinear decoding of natural images from large-scale primate retinal ganglion recordings. *bioRxiv*, 2020.

Niru Maheswaranathan, Lane T. McIntosh, Hidenori Tanaka, Satchel Grant, David B. Kastner, Joshua B. Melander, Aran Nayebi, Luke E. Brezovec, Julia H. Wang, Surya Ganguli, and Stephen A. Baccus. Interpreting the retinal neural code for natural scenes: From computations to neurons. *Neuron*, page S0896627323004671, July 2023. ISSN 08966273. doi: 10.1016/j.neuron.2023.06.007. URL https://linkinghub.elsevier.com/retrieve/pii/S0896627323004671.

Lane T McIntosh, Niru Maheswaranathan, Aran Nayebi, Surya Ganguli, and Stephen A Baccus. Deep Learning Models of the Retinal Response to Natural Scenes. 2017.

Acknowledgments

Include acknowledgments of funding, any patents pending, where raw data for the paper are deposited, etc.

Supplementary materials