Biophysical neural adaptation mechanisms enable deep learning models to capture dynamic retinal computation

Saad Idrees [1,2,✉], **Fred Rieke** [3], **Greg D. Field** [4], **and Joel Zylberberg** [1,2,5✉]

[1]Department of Physics and Astronomy, York University, Toronto, ON M3J 1P3, Canada; [2]Centre for Vision Research, York University, Toronto, ON M3J 1P3, Canada; [3]Department of Physiology and Biophysics, University of Washington, Seattle, WA 98195, U.S.A.; [4]Stein Eye Institute, University of California, Los Angeles, CA 90095, U.S.A.; [5]Learning in Machines and Brains Program, Canadian Institute for Advanced Research, Toronto, ON M3J 1P3, Canada

Neural adaptation is a universal feature of neural systems that modulates the output based on input conditions, enabling efficient encoding of sensory inputs without saturation or loss. In contrast, conventional artificial neural networks (ANNs) lack these adaptational mechanisms, resulting in inaccurate neural predictions under changing input conditions. Can embedding neural adaptive mechanisms in ANNs improve their performance and generate biologically-plausible models? To address this question, we introduce a new type of convolutional neural network (CNN) layer incorporating photoreceptor biophysics and adaptation mechanisms to model light-dependent photoreceptor sensitivity and kinetics. Under changing input light conditions, CNNs that include the new photoreceptor layer integrated as a front-end perform better than conventional CNN models at predicting: (1) monkey and rat retinal ganglion cell responses; and (2) context-dependent changes in neural sensitivity. This study demonstrates the potential of embedding neural adaptive mechanisms in deep learning models, enabling them to adapt dynamically to evolving inputs.

retina ganglion cells | photoreceptor biophysics | adaptation | convolutional neural networks | deep learning | adaptive convolution

Correspondence: ***saidrees@yorku.ca joelzy@yorku.ca***

Introduction

Neural adaptation (Fain et al., 2001; Benda, 2021) is a nearly universal feature of neural systems, underlying a stimulus- or activity-driven change in a system's input-output relation. This causes, for example, the fading of perceived intensity of odors (Zufall and Leinders-Zufall, 2000) and accommodation to sounds (Willmore and King, 2023). Within the visual system, neural adaptation mechanisms such as gain control (Shapley and Victor, 1978; Chander and Chichilnisky, 2001; Beaudoin et al., 2007; Schneeweis and Schnapf, 1995; Baylor et al., 1984; Demb, 2008; Dunn et al., 2006) and spike frequency adaptation, govern the diverse response properties of retinal ganglion cells (RGCs) (Kim and Rieke, 2001, 2003; O'Brien et al., 2002; Wong et al., 2012). These adaptation mechanisms are themselves highly dynamic processes (Shapley and Victor, 1978; Chander and Chichilnisky, 2001; Beaudoin et al., 2007; Kim and Rieke, 2001, 2003; O'Brien et al., 2002; Wong et al., 2012; Angueyra et al., 2021), modulated by the neuron's state and prevailing input statistics. Neural adaptation therefore enables neurons to constantly adjust their responses based on the prevailing input conditions that change at multiple time scales. In contrast, artificial neural networks (ANNs) primarily employ homogeneous, static nonlinear functions, limiting their ability to dynamically adapt to changing input conditions. This raises the question: Can incorporating neural adaptation mechanisms in ANNs enhance the development of brain models capable of accurately predicting neural responses under varied input conditions? In this study, we answer this question by creating a new type of ANN layer that incorporates photoreceptor biophysics and adaptation mechanisms to model light-dependent photoreceptor sensitivity and kinetics. This new layer, used as a front-end to deep convolutional networks (CNNs), enables these models to predict retina responses at untrained light levels. The development of such high-fidelity models that can faithfully emulate the human visual system and those of other mammals holds immense significance in advancing our fundamental understanding of how we perceive the visual world (Maheswaranathan et al., 2018; Cadena et al., 2017; Lindsey et al., 2019; Goldin et al., 2022; Yamins and Dicarlo, 2016; Tanaka et al., 2019; Drinnenberg et al., 2018; McIntosh et al., 2016; Gollisch and Meister, 2010; Yamins et al., 2014; Keat et al., 2001; Olshausen and Field, 2004). Such biologically tractable models could also provide insights into underlying neural processing and mechanisms.

Given their strong performance in computer vision applications like object recognition (Chollet, 2017; Simonyan and Zisserman, 2015; Krizhevsky et al., 2017), ANNs employing deep learning algorithms have quickly become the leading approach to model the biological visual system. For example, when trained and tested under the same lighting conditions, CNNs can successfully predict visually-evoked responses of neurons in visual cortex (Kindel et al., 2019; Cadena et al., 2017; Yamins

et al., 2014; Yamins and Dicarlo, 2016) and retina (McIntosh et al., 2016; Tanaka et al., 2019; Yan et al., 2022; Goldin et al., 2022; Maheswaranathan et al., 2018). Despite these myriad successes, their ability to handle highly dynamic lighting conditions inherent in natural vision is uncertain.

During natural vision, the amount of light falling on the retina can change by several orders of magnitude at multiple timescales. For example, light input can change locally on a region of the retina by an order of magnitude in less than a second following rapid eye movements such as saccades. Global light levels may fluctuate by an order of magnitude every few seconds on a cloudy day, as clouds pass between the observer and the sun. The retina already compensates for some dynamic changes in the input statistics, thereby enabling stable signaling to higher visual areas (Shapley and Enroth-Cugell, 1984; Demb, 2008; Rieke and Rudd, 2009). At the same time, moderate alteration of ambient luminance or sudden changes in luminance lead to altered retinal output (Tikidji-Hamburyan et al., 2015; Ruda et al., 2020; Idrees et al., 2020, 2022; Farrow et al., 2013; Schwartz and Rieke, 2013).

These dynamic retinal computations already start in photoreceptors (Fain et al., 2001), the very first cells in the visual processing cascade. Photoreceptors adapt strongly and at multiple timescales, controlling both the gain and kinetics with which light inputs are converted into electrical signals in the retina. The adaptation itself is highly dynamic (Angueyra et al., 2021; Yu et al., 2022; Clark et al., 2013): the gain and kinetics are modulated in a history-dependent manner. This adaptation is also highly local: each photoreceptor can adapt independently.

State-of-the-art retina predictors, such as Deep Retina (McIntosh et al., 2016), can predict responses to visual stimuli with similar image statistics to the training set, but may not generalize to the same stimuli with different image statistics, such as different ambient luminance levels. This is because conventional CNN models lack the ability to adjust their output's gain or kinetics based on prevailing input conditions. This can result in inaccurate predictions when evaluated with stimuli at different light levels. While normalizing the mean light level and dynamic range (e.g., with the commonly-used batch normalization and/or layer normalization methods) can compensate for some changes in input statistics, the prevailing methods do so by keeping the CNN output statistics unaltered. Capturing changes in RGC response kinetics and feature selectivity at different light levels (Tikidji-Hamburyan et al., 2015; Ruda et al., 2020) would require models that can mimic the retinal adaptive mechanisms.

In this study, we first tested the ability of conventional CNNs (similar to Deep Retina (McIntosh et al., 2016)) to predict stimulus-evoked retinal ganglion cell (RGC) firing patterns under lighting conditions that differed from those under which they were trained. Observing a profound failure of the CNNs to generalize to new lighting conditions, we then created a new class of CNN layer that builds upon a biophysical model (Angueyra et al., 2021) of the phototransduction cascade within the photoreceptors. This layer therefore mimics the transformation of light into electrical signals at the retina's photoreceptors. The photoreceptor layer can be used as an input to conventional CNNs, and can be trained end-to-end along with the other CNN layers. It thus equips deep learning CNN models of the retina with biorealistic adaptation mechanisms, enabling those models to capture dynamic retinal computations arising from changes in ambient light conditions.

The photoreceptor-CNN model therefore outperforms conventional CNN models in accurately capturing dynamic retinal computations across varying light levels. We validated this by (1) predicting monkey and rat RGC responses under test lighting conditions different from the training conditions, (2) predicting context-dependent changes in RGC sensitivity at different light levels, and (3) demonstrating that the model can capture light-dependent changes in RGC response properties such as higher latencies at darker light levels. This more accurate and biologically plausible deep learning model has several potential applications, including investigating context-dependent changes in RGC receptive fields for different cell types and for studying the role of adaptation in the retina. Incorporating it as front-end to cortical models could enable investigating higher visual processing under dynamic conditions. Moreover, this model could advance the field of visual prosthetics by providing better neural predictors.

Results

Existing deep learning models of the retina do not generalize well across light levels

We first tested whether existing deep learning models of the retina could generalize across light levels. To achieve this goal, we recorded the spiking activity of RGCs in isolated *ex vivo* macaque monkey retina using a high density multielectrode array (Methods). The retina was exposed to the same binary white noise checkerboard movie (Methods) at three different light levels. Neutral density filters in the light path allowed for the mean intensity of the stimulus to be altered without changing its pixel values. As a result, the contrast of the stimuli remained constant during the experiment, independent of the light level. All stimulus intensity values are described in units of effective photoisomerizations per receptor and per second (R*receptor^{-1}s^{-1}). Recordings were made at mean intensities of 30 R*receptor^{-1}s^{-1}, 3 R*receptor^{-1}s^{-1}and 0.3 R*receptor^{-1}s^{-1}; rod pho-

toreceptors dominate retinal input at these light levels. We recorded the RGC activity for 60 minutes at each light level. Our analysis focused on 37 RGCs that included midget and parasol RGCs of ON and OFF types.

We then built a model using existing state-of-the-art Deep Retina architecture (McIntosh et al., 2016) (i.e., the architecture in Fig. 1 with the photoreceptor layer and batch normalization layers removed, Methods), and optimized the numbers of convolutional layers, filters in each layer, and the filter sizes with a grid search. The resulting model had three CNN layers followed by a dense layer with 8, 16 and 18 channels in each of the CNN layers, respectively, and filter sizes of 9x9, 7x7 and 5x5 respectively.

We trained this CNN model to predict RGC responses (N = 30 ON; N = 7 OFF RGCs) (Methods) to a total of 40 minutes of white noise movies at mean light levels of 30 R*receptor^{-1}s^{-1}and 3 R*receptor^{-1}s^{-1}, the two higher light levels in our dataset. The model takes as input a movie segment (120 consecutive frames, where each frame corresponds to 8 ms) and outputs an instantaneous spike rate for each RGC at the end of that movie segment. By shifting the input forward by 1 frame along the input movie, the model can thus predict the full time series of RGC responses evoked by the stimulus movie. A Layer Normalization layer at the input normalized each pixel in the input movie segment by the mean and variance of all pixels in that movie segment. This had an effect similar to z-scoring the input, effectively removing the mean light level from each movie segment.

To evaluate the model's performance, we used a test data set that included held-out segments of white noise movies at all light levels (Fig. 2a), including the testing light level (0.3 R*receptor^{-1}s^{-1}) not used during the training. The corresponding responses of an example RGC to these white noise movies is shown in Fig. 2b. The model accurately predicted responses to movies at the two training light levels, 30 R*receptor^{-1}s^{-1}and 3 R*receptor^{-1}s^{-1}(Fig. 2c, columns 1-2). We quantified the model performance with the fraction of explainable variance in each RGC's response that was explained by the model (FEV; see Methods). For ease in interpretation, FEV is represented as a percentage throughout our results. By definition, a perfect model will have FEV value of 100%.

This CNN model with an architecture similar to Deep Retina performed with FEV ($median \pm 95\% c.i.$) of $84\% \pm 11\%$ and $78\% \pm 3\%$ for the two training light levels (Fig. 3a). However, this model performed poorly at the testing light level of 0.3 R*receptor^{-1}s^{-1}(Fig. 2c, column 3) explaining only $16\% \pm 15\%$ of the variance in the population data (Fig. 3a).

To overcome internal covariate shift and stabilize the neural network, we next tried including batch normalization layers (Ioffe and Szegedy, 2015) prior to each activation function in the CNN model. This marginally increased the overall performance of the model at all light levels (Fig. 3b). The generalization performance at the test light level was still quite low with FEV of $24\% \pm 15\%$ (Fig. 3b).

We hypothesized that the failure of the CNN model at the test light level of 0.3 R*receptor^{-1}s^{-1}(Fig. 3b) occurred because it predicted the same output for a given movie segment, regardless of the light level. However, RGC responses vary in both gain and kinetics in a light-dependent manner: at the lower light level, they exhibit slower response kinetics (see inset in Fig. 2b) and different feature selectivity indicated by different peak responses in Fig. 2b. Conventional CNNs lack the ability to adaptively regulate gain and kinetics of input conversion in a light-dependent manner. While Layer Normalization layer at the input compensates for mean changes in light level, without which CNNs fail to mitigate response gain adjustment, changes in output statistics like response kinetics would require inherent knowledge of how the retina alters its output with changing light levels.

Photoreceptor layer enables prediction of RGC responses across light levels

To overcome the shortcoming of conventional CNNs, which cannot adaptively adjust their output gain and kinetics in a light dependent manner, we developed a new type of convolutional neural network layer that builds upon a biophysical model of the phototransduction cascade by Angueyra et al. (2021). The biophysical model has been demonstrated to faithfully capture photoreceptor adaptation (Angueyra et al., 2021). The fully-trainable neural network layer has twelve parameters (Methods) that could be trained together with the downstream network through backpropogation. In this architecture, the new photoreceptor layer was followed by a Layer Normalization layer Ba et al. (2016) at the input stage of the deep CNN (Fig. 1) described in the previous section. The subsequent input to the CNN was thus a normalized and transformed version of the original input (e.g., as in Fig. 6f). Our hypothesis is that the trained photoreceptor layer will provide deep CNNs with the ability to alter both the gain and kinetics of their output in a light-dependent manner.

To test whether the new photoreceptor-CNN (PR-CNN) model was better at generalizing across lighting conditions, we repeated the same task as above, i.e., trained the model end-to-end including the photoreceptor layer to predict monkey RGC responses to the binary checkerboard movie at the two higher light levels, 30 R*receptor^{-1}s^{-1}and 3 R*receptor^{-1}s^{-1}. Similar to the conventional CNN model without the photoreceptor layer, the PR-CNN model reliably predicted responses to test stimuli at the two training light levels (Fig. 2d columns 1-2 and Fig. 3c). However, because now the

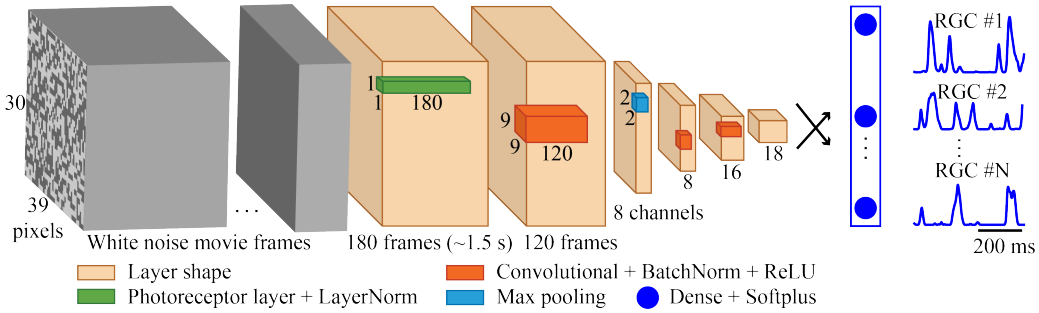

Figure 1. PR-CNN model architecture. The photoreceptor layer at the input takes 180 frames of a movie as a single sample and converts each pixel (green color) into photocurrents. The photocurrents are then passed to the first convolution layer that has 8 channels where each kernel (orange color) operates over spatial dimensions of 9 pixel x 9 pixel and temporal depth of 120 frames. The result is normalized through a Batch Normalization layer and then passes through a Rectified Linear Unit (ReLU). The output is then max pooled to reduce the spatial dimension and then passes through another two convolution layers. The model output layer is a dense layer that has N units based on the number of RGCs in the dataset. It converts previous layer outputs into RGC spiking output (blue traces). By traversing through the input movie 180 frames at a time, an entire time series of RGC responses is obtained.

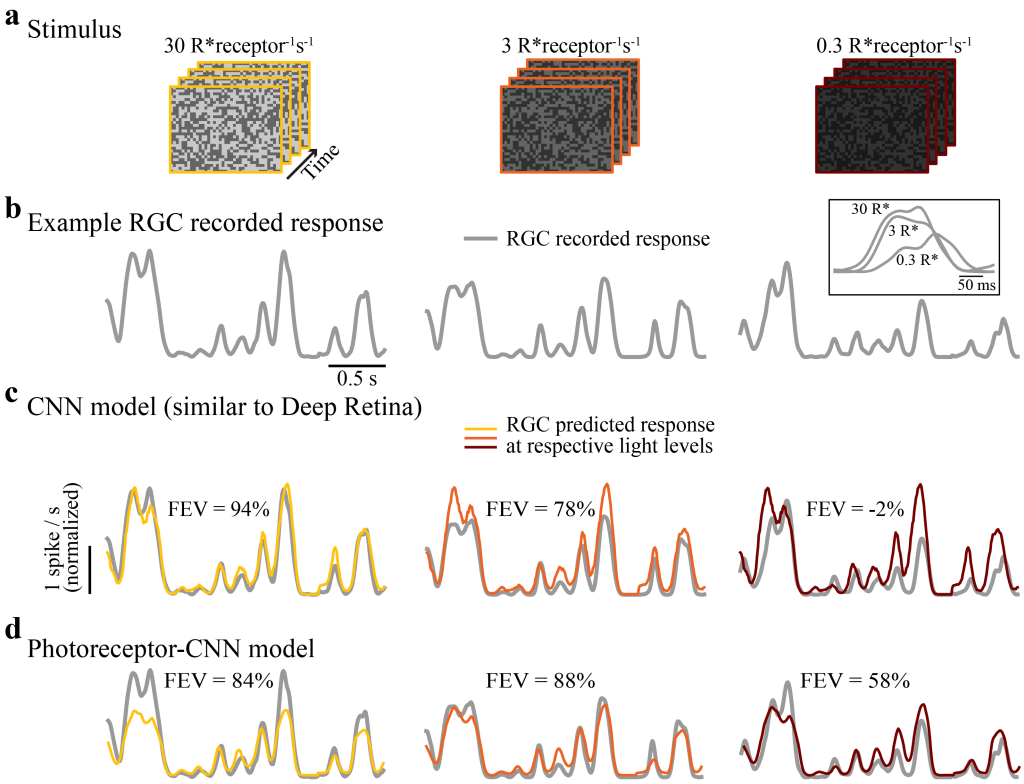

Figure 2. Predicted and actual responses of an example monkey retinal ganglion cell (RGC) at 3 different light levels. **a.** White noise movie at three different light levels (columns). **b.** Recorded response of an example RGC (gray lines) to white noise movie at the three different light levels in **a** (columns). Inset above the last column shows an overlay of a segment of the responses at the three light levels. **c.** Responses predicted by a CNN model similar to Deep Retina (colored) at each light level in **a** (columns). FEV values above each trace quantifies the performance of the model for this RGC at the corresponding light levels. **d.** Same as in **c** but for the proposed photoreceptor-CNN model. Models were trained on data at 30 R*receptor^{-1}s^{-1} (column 1) and 3 R*receptor^{-1}s^{-1} (column 2).

model output properties can depend on mean light level, the PR-CNN model could also explain $54\% \pm 11\%$ of the variance in responses at a light level lower than those at which it was trained (Fig. 2d column 3 and Fig. 3c). This was a notable improvement over the conventional CNN model without the photoreceptor layer, which only explained $24\% \pm 15\%$ of the variance (Fig. 3b).

Next, we repeated this analysis with all different combinations of training and testing light levels: each time, the model was trained on data from two light levels, and then evaluated on test data from a different light level (Fig. 3d). For all combinations of training and testing light levels, the PR-CNN model performed better at generalizing to new light levels than the conventional CNN model without the photoreceptor layer. The difference was smallest for the case where the testing light level (3 R*receptor^{-1}s^{-1}) was intermediate between the training light levels (30 R*receptor^{-1}s^{-1}and 0.3 R*receptor^{-1}s^{-1}) (Fig. 3d, column 3). This suggests that while conventional CNN based models can interpolate between different sets of training conditions, they are not well equipped to extrapolate to conditions that are outside the range under which they were trained, more so if the change in input statistics are as-

sociated with change in output statistics, as is the case here. The new PR-CNN is much better at this challenging task as the photoreceptor layer enables the CNNs to adjust output statistics in a light-dependent manner.

Photoreceptor-CNN model more accurately predicts context-dependent changes in RGC sensitivity

Frequent and significant changes in light intensity on the retina, such as those caused by saccades, alter RGC encoding (Idrees et al., 2020, 2022; Tikidji-Hamburyan et al., 2015; Ruda et al., 2020) and feature selectivity (Grimes et al., 2014). The use of CNNs to investigate such context-dependent changes in RGC feature selectivity is becoming increasingly useful (Maheswaranathan et al., 2018; Goldin et al., 2022; Walker et al., 2019). This is because instantaneous receptive fields of CNN-modeled RGCs can be calculated from gradients of the model-predicted response with respect to the input pixel values (Fig. 4). These derivatives can be evaluated for individual segments of input movie, without the need to average over multiple segments as would be needed in the typical reverse correlation analysis. As a result, the CNN-based methods provide a new tool for researchers to probe rapid stimulus-dependent changes in neuronal receptive fields.

We hypothesized that PR-CNN model could capture context-dependent changes in receptive fields better than conventional CNNs, due to the fact that it can better capture how the retina adapts to dynamically changing stimuli. Our focus was specifically on quantifying changes in RGC sensitivity, as defined by the gain (amplitude) of its receptive field.

To test the hypothesis, we computed the spatio-temporal receptive fields (STRFs) of the model RGCs from gradients (Fig. 5a) of the model trained on the same monkey dataset used in the previous section. We decomposed each STRF into its spatial and temporal components and normalized the spatial component to have unit mean. This enabled us to represent the STRF's amplitude solely in the temporal component. We observed substantial variation in gain, i.e. the peak amplitude of these receptive fields varied across the movie segments for a given model RGC (Fig. 5a). This variation constitutes a model-derived prediction that the real RGCs (whose activities the model was trained to predict) would exhibit the same pattern of stimulus-dependent gain variation. To test these model predictions, we sorted the movie segments by their model-predicted gains, and then divided the segments into 10 equally-sized bins (i.e., lowest-gain bin 1 up to highest-gain bin 10; Fig. 5b). We then estimated the RGC STRF separately based on the data in each bin, using the standard reverse correlation (Dayan and Abbott, 2001) approach. This procedure was performed separately for each RGC, because the stimulus segments that yield higher or lower gain can differ between RGCs. From these experimentally-derived STRFs, we then extracted the spatial and temporal components. The gains of these temporal receptive fields within each bin (Fig. 5c) were consistent with model-derived predictions obtained from averaging the temporal receptive fields from the model's gradients (Fig. 5b) over the set of stimuli within each bin. Thus, the retrospective analysis of RGC receptive fields in these stimulus bins confirmed that the PR-CNN model could veridically predict context-dependent RGC sensitivity.

To quantify the accuracy with which the model could predict RGC sensitivity changes, we calculated the squared error between the predicted (derived from the model gradients) and the actual (derived from reverse correlation) gain within each stimulus bin and averaged the resulting errors across bins. Figure 5d shows a scatter plot of the resulting mean squared error (MSE) values for the PR-CNN and the conventional CNN model for different RGCs (N = 22) and different light levels (columns). These RGCs constitute a subset of the total RGCs that displayed FEV values greater than 50% at the training light levels of 30 R*receptor^{-1}s^{-1}and 3 R*receptor^{-1}s^{-1}for both models: thus, these were the RGCs for which both models' predictions were most reliable. The PR-CNN model consistently exhibited lower MSE values than the conventional CNN model at all light levels (Fig. 5d-f) with the difference being statistically significant for 3 R*receptor^{-1}s^{-1}(Fig. 5e; p = 0.01, one-tail Wilcoxon signed-rank test, N = 22RGCs) and the held-out test light level of 0.3 R*receptor^{-1}s^{-1}(Fig. 5f; p = 0.007, one-tail Wilcoxon signed-rank test, N = 22 RGCs). This result indicates that the PR-CNN model is superior in predicting context-dependent changes in RGC sensitivity.

Photoreceptor-CNN model can capture light-dependent changes in RGC response kinetics

It is well-established that RGC responses exhibit slower kinetics at darker light levels (Ruda et al., 2020). This effect was also apparent in the dataset used in this study: RGCs response to white noise movie presented at 3 R*receptor^{-1}s^{-1}was slower than the response to the same movie at 30 R*receptor^{-1}s^{-1}(Fig. 6a shows an example RGC). These kinetics are typically measured by calculating response latency, given as the time-to-peak of the temporal receptive fields obtained from reverse correlation (Fig. 6b). Consistent with the slower RGC responses at darker light level, the response latencies obtained from receptive fields were higher at 3 R*receptor^{-1}s^{-1}in comparison to 30 R*receptor^{-1}s^{-1}(Fig. 6b). This difference was statistically significant across the population of RGCs (Fig. 6e, left column;p = 2×10^{-4}, two-tailed Wilcoxon rank-sum test, N = 22 RGCs).

We therefore wondered whether the neural adaptation

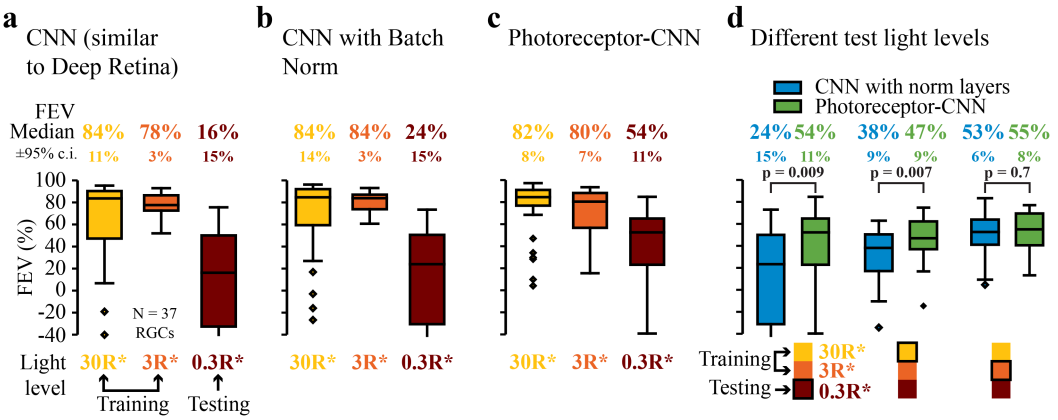

Figure 3. Model performance comparisons. **a-c.** Performance of **(a)** a CNN model that had a similar architecture as Deep Retina, **(b)** CNN model of **a** but with added normalization layers included, and **(c)** the photoreceptor-CNN model. Each model was evaluated at three light levels (labelled below each box plot): 30 R*receptor^{-1}s^{-1}and 3 R*receptor^{-1}s^{-1}, at which the models were trained, and 0.3R*receptor^{-1}s^{-1}(testing condition) which the models did not see during the training. The box plot at each light level shows the distribution of FEVs across 37 monkey RGCs. Numbers at the top of each box plot are the median FEVs $\pm 95\%c.i.$. **d.** Performance of the CNN model with normalization layers (blue color; same model as in **b**), and the photoreceptor-CNN model (green color; same model as in **c**) at all combinations of training and test light levels. For each column, the legend below the box plot panel shows the two light levels the models were trained at (black outline) and the third light level at which it was tested. The box plots show the distribution of FEVs at this testing light level. Testing light levels were 0.3 R*receptor^{-1}s^{-1}(column 1), 3 R*receptor^{-1}s^{-1}(column 2), and 30 R*receptor^{-1}s^{-1}(column 3). p-values were calculated by performing two-sample t-test on the FEV distributions from the CNN and PR-CNN model at each testing light level.

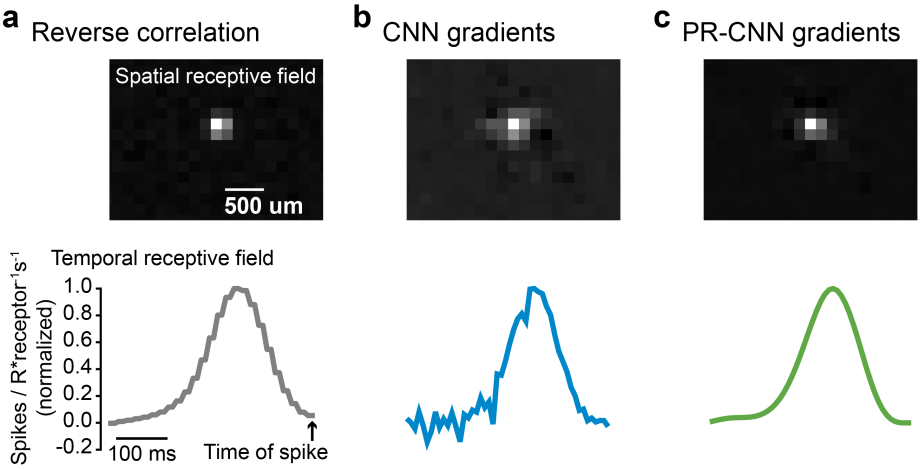

Figure 4. a. The receptive field of an example RGC calculated from the reverse correlation of white noise movie (~ 55 mins), decomposed into its spatial (top row) and temporal (bottom row) components. **b-c.** Spatial (top) and temporal (below) components of the receptive field of the same RGC calculated from the gradient outputs of the conventional CNN model, **b**, and the photoreceptor-CNN model, **c**, evaluated for a single movie segment (~ 1.5 s).

mechanisms embedded in the PR-CNN model could also allow for dynamic adjustment of response kinetics, in addition to sensitivity changes reported above. To investigate this, we used the PR-CNN and conventional CNN models trained at mean light levels of 30 R*receptor^{-1}s^{-1}and 3 R*receptor^{-1}s^{-1}and asked whether the temporal receptive fields of model RGCs show similar differences as observed in the experimental data (Fig. 6b, Fig. 6e, left column). The temporal receptive fields of model RGCs (see Fig. 6c,d for an example RGC) were calculated by averaging the instantaneous temporal receptive fields of Fig. 5a across all movie segments. Consistent with our hypothesis, the temporal receptive fields of PR-CNN model RGCs exhibited light-dependent changes in response kinetics, with higher latencies at the darker light level of 3 R*receptor^{-1}s^{-1}(Fig. 6c). This difference in latencies was statistically significant at the population level (Fig. 6e; $p = 1 \times 10^{-4}$, two-tailed Wilcoxon rank-sum test, N = 22 RGCs). This change in kinetics could already be observed at the output of the photoreceptor layer (Fig. 6f), simply as a function of input light intensity. In contrast, conventional CNN model showed no changes in latencies across the two light levels the models were trained at (Fig. 6d; right column in Fig. 6e). This suggests that the photoreceptor layer in the models is necessary for modulating the response kinetics, highlighting the value of equipping ANNs with mechanisms that allow them to adjust their outputs dynamically.

Generalization across extremely different light levels

Having observed that the PR-CNN model performs well across light levels that differ by 1-2 orders of magnitude, we wondered whether it could also generalize well

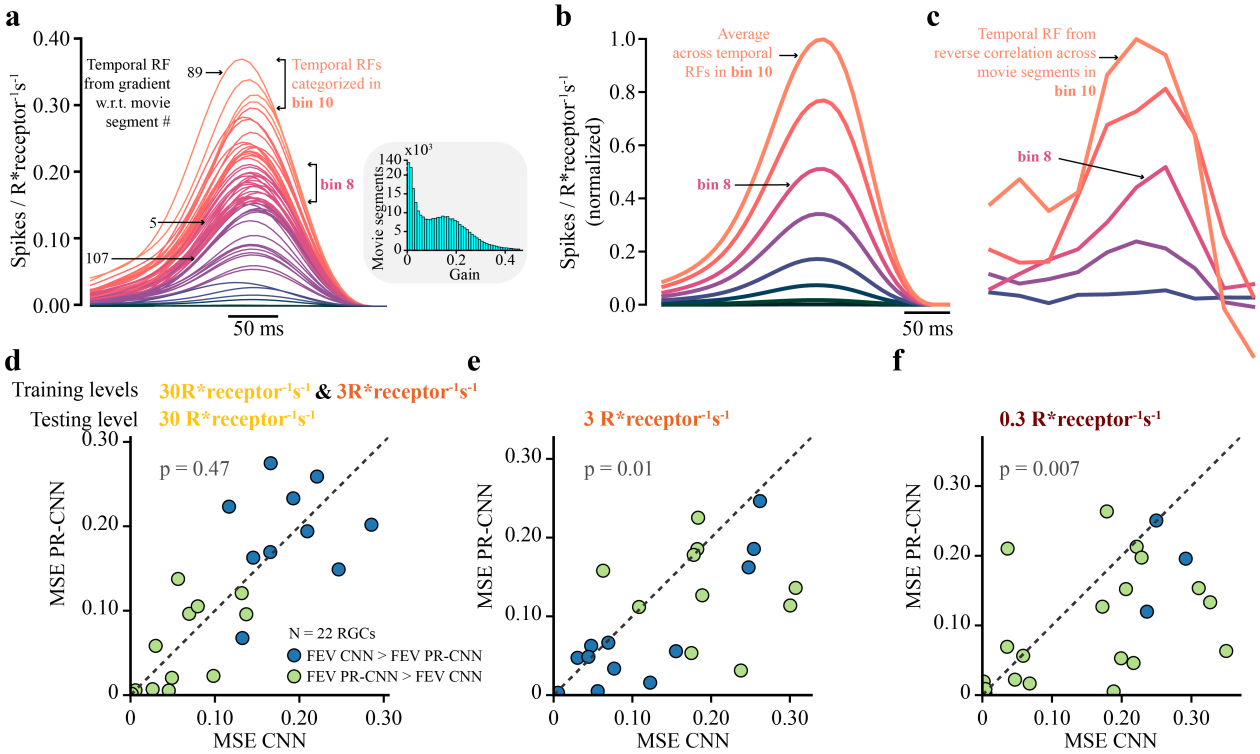

Figure 5. **a.** The temporal receptive fields of a model RGC obtained from gradients of the output of a PR-CNN model, evaluated for different movie segments (only a subset is shown here for clarity). Inset shows the distribution of receptive field gains obtained as the peak of the temporal filters across 400,000 individual movie segments. **b.** The temporal receptive fields in **a** were ranked by their gain, divided into 10 bins (different colors in **a**), and averaged within each bin. The averaged temporal receptive fields were normalized by the peak amplitude of the highest-amplitude curve. **c.** Temporal receptive fields calculated directly from reverse correlation on experimental data, using movie segments that were grouped into the same bins as in **b**. **d-f.** Mean squared error between the predicted gains, **b**, and actual gains, **c**, of the temporal receptive fields plotted for different RGCs (N=22) for conventional CNN model (x-axis) and PR-CNN model (y-axis). The models were trained at 30 R*receptor^{-1}s^{-1}and 3 R*receptor^{-1}s^{-1}, and predicted gains were obtained from gradients of model output w.r.t. movie segments at 30 R*receptor^{-1}s^{-1}**(d)**, 3 R*receptor^{-1}s^{-1}**(e)**, and 0.3 R*receptor^{-1}s^{-1}**(f)**. p-values were obtained by conducting a one-tailed Wilcoxon signed-rank test to analyze whether the PR-CNN MSE values for the population of RGCs were lower than those of CNN.

across more extreme variations in lighting. To answer this question, we trained the model to predict rat RGC responses at a relatively bright (photopic) light level (10,000 R*receptor^{-1}s^{-1}) where cone photoreceptors predominantly contribute to vision, and evaluated the model's generalization performance at a much dimmer (scotopic) light level (1 R*receptor^{-1}s^{-1}) where mostly rod photoreceptors are active. For this analysis, we used rat RGC recordings that we previously published (Ruda et al., 2020).

For comparison, we first performed this experiment with the conventional CNN model including normalization layers (as in Fig. 3b). That CNN model failed (FEV of $-52\% \pm 9\%$) to predict rat RGC (N = 55 RGCs) responses at the dim 1 R*receptor^{-1}s^{-1}light level after having been trained at the bright 10,000 R*receptor^{-1}s^{-1}level (Fig. 7a,c). The proposed PR-CNN model, however, did surprisingly well (Fig. 7b), achieving FEV of $54\% \pm 8\%$ on this task (Fig. 7c). For this experiment, we first trained the PR-CNN model at 10,000 R*receptor^{-1}s^{-1}and then replaced that model's photoreceptor parameters (which correspond to cone cells at this light level) with those corresponding to rod cells, as explained below and in Supplementary Fig. 1. The remaining CNN parameters were unchanged by this procedure. Thus, our finding demonstrates that the changes in photoreceptor layer parameters alone can account for much of the difference in how the PR-CNN model predicts the RGC responses at these two light levels.

In more detail, the PR-CNN experiment was performed as follows. First, we trained the model end-to-end (including photoreceptor layer parameters) to predict rat RGC responses from Retina A (N = 61 RGCs) to stimuli at the bright (photopic) light level (10,000 R*receptor^{-1}s^{-1}) (Step 1 in Supplementary Fig. 1). This led to an estimate of photoreceptor and CNN parameters for the cone pathway. We then estimated parameters reflecting rod photoreceptors by retraining the model at the dim (scotopic) light level (1 R*receptor^{-1}s^{-1}), while keeping the CNN weights and biases fixed to their already-learnt values (Step 2 in Supplementary Fig. 1). Thus, the retraining altered only the photoreceptor parameters based on the data from Retina A.

We then trained the PR-CNN model to predict RGC responses of another rat's retina (Retina B) to stimuli at the bright light level (Step 3 in Supplementary Fig. 1). Here, we fixed the photoreceptor parameters to the values already learned from Retina A at the bright light

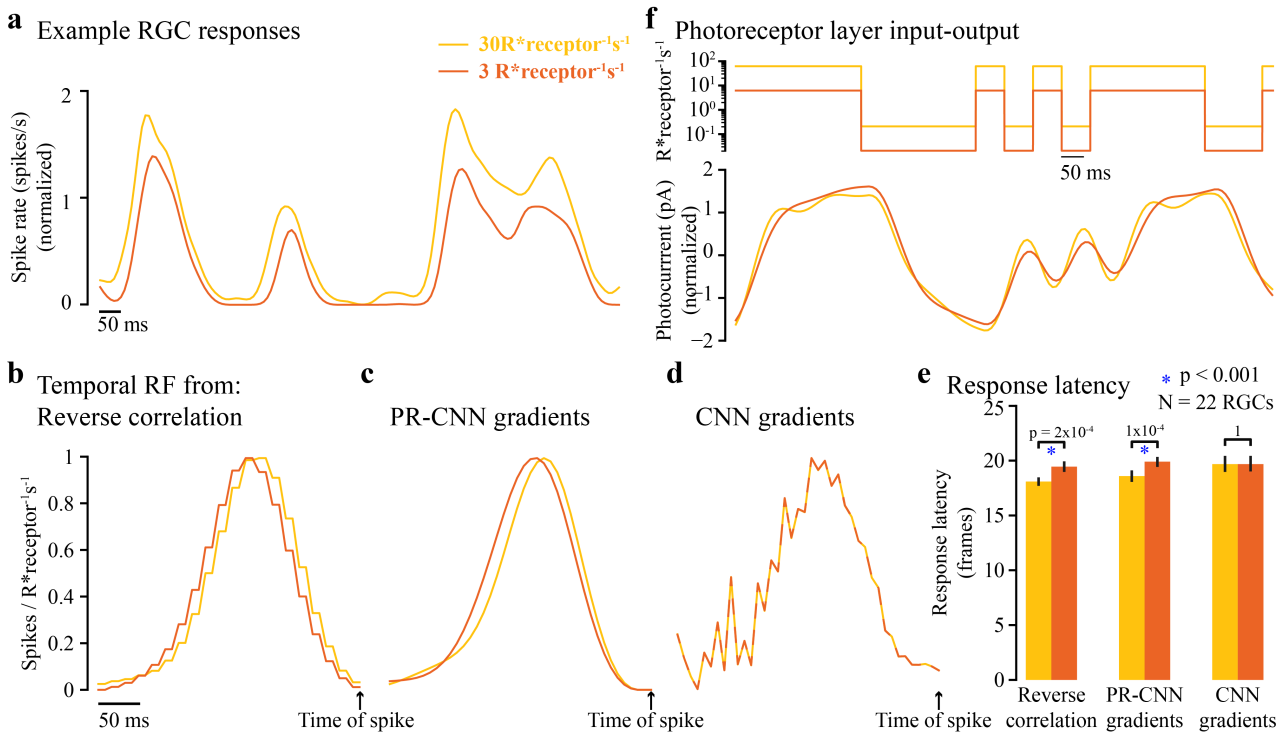

Figure 6. **a**. Normalized recorded spiking activity of an example RGC in response to a white noise stimulus at two light levels used in model training: 30 R*receptor^{-1}s^{-1}(yellow), and 3 R*receptor^{-1}s^{-1}(orange). **b-d.** Temporal receptive field of the same RGC calculated using reverse correlation of a white noise movie (55 minutes) at the two different light levels. **c.** Temporal receptive fields of the same RGC obtained by averaging across multiple instantaneous temporal receptive fields from PR-CNN output gradients with respect to multiple input movie segments. **d.** Same as in **c** but for the conventional CNN model without the photoreceptor layer. **e.** Mean response latency (N = 22 RGCs), calculated as the number of frames (1 frame = 8 ms) between the time of spike and peak of the temporal receptive field. Error bars indicate 95% confidence interval of the mean. Colors (legend in **a**) represent the light level at which temporal receptive fields were calculated from reverse correlation of experimental data (left column), PR-CNN model gradients (middle column) and conventional CNN model gradients (right column). An asterisk indicates a significant difference in response latencies between two light levels ($p < 0.001$, N = 22 RGCs, two-tailed Wilcoxon rank-sum test). **f. Top.** Intensity changes over time for a single pixel in the binary checkerboard white noise movie at the two different mean light levels. **Bottom.** The output of the photoreceptor layer from the PR-CNN model after the Layer Normalization layer that immediately follows the photoreceptor layer. This output is fed into subsequent CNNs. Models in **c-e** were trained at mean light levels of 30 R*receptor^{-1}s^{-1}and 3 R*receptor^{-1}s^{-1}. Legend in **a** is valid for all panels (line style varies for clarity).

level. We then exchanged those photoreceptor parameters with the ones learned from Retina A at the low light level, while keeping the remaining CNN parameters fixed (Step 4 in Supplementary Fig. 1). In this manner, we took the model where the CNN weights were learned on model of Retina B at the bright light level, and simply exchanged the cone photoreceptor parameters for rod parameters learned from retina A. Notably, in this procedure, none of the data from Retina B at the dim light level was used in training the model: these data were used only in evaluating the generalization ability of the PR-CNN model.

Discussion

This paper introduces a new class of convolutional neural network for vision models that builds upon a biophysical model of the phototransduction cascade (Angueyra et al., 2021) within the photoreceptors. When used as a front-end to CNNs, this new photoreceptor layer equips the model with adaptive mechanisms that adjust the model output in a light-dependent manner. As a result, the PR-CNN model is significantly better than conventional CNN models at predicting RGC responses at light levels different from those to which the model was exposed during training.

Existing artificial neural networks (ANNs) struggle to accurately predict neural responses in dynamic input conditions, like changing ambient light levels. This is important because during natural visual behavior, the input statistics, specifically light levels, can vary drastically, making it challenging for models trained on limited data to perform well on out-of-distribution inputs. Unlike in computer vision tasks where output labels remain unchanged in dynamic conditions, the statistics of neural responses, such as their kinetics, also change.

While ANNs are universal function approximators (Hornik et al., 1989) and therefore capable of implementing any transformation with simple nonlinear units, their operating range is limited by the input and output statistics of the training set. A sufficiently large ANN can therefore predict neural responses to stimuli with same statistics as the training set, without the need for intricate adaptive mechanisms. However, ANNs can benefit from having the right inductive biases that represent prior knowledge about the underlying data, as demonstrated by the success of CNNs in computer vision over non-convolutional

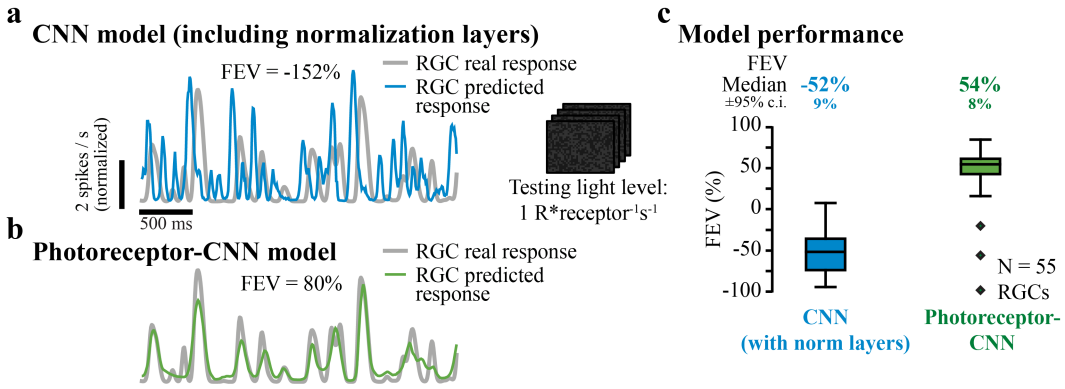

Figure 7. Generalization across extreme light levels. **a-b.** Rat Retina B example RGC responses at scotopic light level (1 R*receptor^{-1}s^{-1}). Actual responses shown in gray. Predicted responses generated by the CNN model (trained at photopic level, 10,000 R*receptor^{-1}s^{-1}) without the photoreceptor layer (**a**, blue) and with a photoreceptor layer(**b**, green). **c.** Box plots show the distribution of FEVs across RGCs (N = 55 RGCs, Retina B) for the CNN model with normalization layers (blue) and for the CNN model with the photoreceptor layer added, when the models were trained at photopic light level and tested at scotopic light level.

neural networks (LeCun et al., 1989, 1998; Alzubaidi et al., 2021). In the same way, our paper demonstrates that equipping CNN based deep learning models with photoreceptor adaptive mechanisms greatly improves their ability to generalize to out-of-distribution tasks, such as extrapolating light conditions (Fig. 3d, 7c). The new PR-CNN is much better at this challenging task as the photoreceptor layer enables the CNNs to learn response properties such as sensitivity (Fig. 5) and kinetics (Fig. 6) in a light-dependent manner. In addition the photoreceptor layer dynamically adjusts its adaptive mechanisms based on the prevailing light conditions. This unique capability is demonstrated by the smoother photoreceptor layer output (darker line in (Fig. 6f), bottom) at the darker light level of 3 R*receptor^{-1}s^{-1}, reflecting a linear transformation of the input signal (darker line in Fig. 6f, top). In contrast, the output at light level of 30 R*receptor^{-1}s^{-1}is less linear. This behavior is consistent with biological photoreceptors that exhibit adaptation, that includes changes in kinetics (Dunn et al., 2006; Angueyra et al., 2021; Yu et al., 2022; Clark et al., 2013) across light levels.

Although the PR-CNN model outperformed the conventional CNN model at held-out testing light levels, it falls short of achieving the same level of performance at those test light levels as it did at the light levels used for training (Fig. 3c,d). This could be because the network lacks other adaptive mechanisms found downstream of the photoreceptors in the retina, such as spike frequency adaptation at the level of RGCs (Chang et al., 2022; Brien et al., 2002). Introducing adaptive recurrent units (ARU; developed by (Geadah et al., 2022)) to the model's output layer, which implement spike frequency adaption through dynamic control of nonlinearity, could be a potential solution. ARUs at the output layer would also enable the network to have output units with diverse properties, similar to the diversity at the level of RGCs (Wong et al., 2012). Additionally, adaptation in the retina's intervening circuitry between photoreceptors and RGCs (i.e., in bipolar and amacrine cells) can include unexpected changes in feature selectivity (i.e., subunit rectification) (Grimes et al., 2014). To capture this adaptation, an adaptive-convolution layer based on a model for divisive gain control (Clark et al., 2013; Cui et al., 2016) could be used. This trainable layer would incorporate two pathways with distinct kinetics, with the output of one pathway controlling the gain of the other, allowing for greater adaptability to changes in stimuli. We leave these investigations for a future study.

Artificial neural networks offer the potential to simulate entire networks of biological neurons with shared circuits, making them highly attractive in visual neuroscience. These models are capable of automatically learning meaningful representations through multiple layers of abstraction. Establishing correspondence between ANN layers and neural layers (McIntosh et al., 2016; Maheswaranathan et al., 2018; Cowley et al., 2022; Yamins and Dicarlo, 2016; Mano et al., 2021) is therefore being increasingly used to gain insight into biological circuits. One challenge in using current ANNs to elucidate biological circuits is that ANNs are primarily designed to optimize performance on specific tasks rather than mimic biological circuits. As a result, the structure and function of ANNs may not accurately reflect the complexity and organization of biological circuits which often include feedback loops and dynamic interactions between neurons. Moreover, biological neurons are highly complex whereas ANNs typically consist of simple, homogeneous units repeated throughout the network, which may oversimplify neural circuitry and fail to capture the full complexity of biological neural circuits. In contrast, the biophysical photoreceptor model we use in the photoreceptor layer while complex, has parameters that map directly onto the biology, providing an opportunity for investigating the role of photoreceptor adaptation in the retina. For example, slower rod-mediated RGC responses at dim conditions compared to cone-mediated responses at brighter light levels (Ruda et al., 2020; Baylor and Fettiplace, 1977) may explain the temporal lag between predicted and actual response at dim light level (Fig. 7a): the model trained at bright light levels (10,000 R*receptor^{-1}s^{-1})

learnt faster kinetics of the cone pathway. While some of the differences in RGC responses may arise due to faster cone response kinetics (Cao et al., 2007; Baylor and Hodgkin, 1973; Ingram et al., 2016; Schneeweis and Schnapf, 1995), the relative contribution of photoreceptors and downstream retinal adaptation are not well understood. Fixing the photoreceptor layer parameters to empirically measured values would help dissect photoreceptor and circuit mechanisms, which may have be confounded by setting the parameters to trainable in the current study. Similarly, such biologically plausible models could provide insights into mechanisms underlying neural adaptation in other areas of the brain.

The ability of the PR-CNN model to capture context-dependent sensitivity changes (Fig. 5d-f) in RGCs also makes it a promising tool for investigating how feature-selectivity changes with the dynamics during natural vision. Moreover, the paper presents a retrospective analysis pipeline that validates the sensitivities estimated from receptive fields using model gradients (Maheswaranathan et al., 2018; Goldin et al., 2022; Walker et al., 2019). This approach therefore provides an opportunity to investigate and validate context-dependent changes in receptive fields without requiring closed-loop experiments to validate those findings. For instance, the spatial receptive field of an RGC may depend on stimulus context simply because different subunits within the RGC may be sensitive to distinct segments of a stimulus movie. Consequently, the model's predicted sensitivities will exhibit highly localized spatial components for different segments of the stimulus movie. By analyzing the properties of these specific segments, it becomes possible to investigate the context-dependency of RGC spatial receptive fields. These findings can then be retrospectively validated by reverse correlating the movie segments for which the model predicted high sensitivities in a specific region of the receptive field. This analysis involving spatial component of receptive fields was not possible in the current study because of the coarse checker size of the white noise stimuli used: we leave that analysis for future work.

In general, models of retina that can leverage the power of deep learning to model multiple ganglion cells simultaneously, together with biologically interpretable components, could be used to dissect the relative contributions by different cell types in the retina. They could also serve as an input stage to visual-cortical models to investigate higher visual processing under dynamic conditions. From a wider neuroscience perspective, our approach demonstrates the power of embedding neural dynamics in deep learning models of the brain where biophysical layers match the sensitivity to changing input conditions while the downstream layers extract relevant features from dynamically adapting input stages. Beyond neuroscience, these models could also pave the way for medical interventions, such as prosthetic devices, that restore sight to the blind (Bosking et al., 2017).

Future work will address the limitations of this study, such as the manual switch from cone to rod photoreceptors required for the model trained at cone light levels to generalize to rod light levels (Fig. 7). Additionally, the learned photoreceptor parameters may reflect adaptive mechanisms present in downstream retinal circuits, which could constrain certain biological inferences. Initially, this issue was mitigated by fixing the photoreceptor layer parameters to values that recapitulate direct photoreceptor recordings. However, this approach resulted in reduced model performance, likely because the model parameters no longer captured some of the downstream adaptive mechanisms. To overcome this challenge in the future, an alternative strategy would be to fix the photoreceptor layer parameters to their experimental values and introduce additional layers downstream of the photoreceptor layer that incorporate other known sources of adaptation in the retina. Future studies will also evaluate the PR-CNN model at finer spatial scale white noise and natural movies to investigate how retinal code is altered during rapid image shifts across the retina resulting from eye movements such as saccades. Here, instantaneous receptive fields could be used to determine the stimulus features that cause active subunits to change, the time scales of such changes, and how much of that change can be traced back to adaptation at different levels in the retina.

Methods

Retina electrophysiology

We performed electrophysiological experiments on one ex vivo macaque retina (N = 1) and two ex vivo rat retina. The rat data was previously published by Ruda et al. (2020). All electrophysiology experiments were performed in accordance with Duke University's Institutional Animal Care and Use Committee.

The basic experimental approach was the same in both sets of experiments and described in detail in Ruda et al. (2020). Briefly, the retina of an euthanized animal was extracted and placed with the RGC-side down on a multi-electrode array (MEA). The MEA consisted of 512 electrodes with 30 μm spacing. The voltage on each electrode was sampled at 20 kHz and filtered between 80 and 2000 Hz. Spikes on each electrode were identified by thresholding the voltage traces at 4 s.d. of a robust-estimate of the voltage s.d. Spike sorting was performed by an automated PCA algorithm and verified by hand with a custom software (Shlens et al., 2006; Field et al., 2007). Spike waveform clusters were identified as neurons only if they exhibited a refractory period (1.5 ms) with $<$10% estimated contamination. To track identified RGCs across light conditions, cell clusters were sorted in the same PCA subspace at each

light level. Spike times were converted to estimates of firing rate by convolving spike times with a Gaussian of $\sigma = 10ms$ standard deviation and amplitude of $0.25\sigma^{-1}e^{1/2}$.

Visual stimulation

Visual stimuli were created with custom Matlab code. Stimuli were presented with a gamma-corrected OLED display. The image from the display was focused onto the photoreceptors using an inverted microscope (Ti-E, Nikon Instruments) with a x4 objective. The ambient light level (i.e. mean stimulus intensity) was set using neutral density filters in the light path. In each recording, stimuli were first presented at the darker light level. For every subsequent higher light level, the retina tissue was first adapted to that light level before continuing the recordings. Stimuli consisted of non-repeated, binary checkerboard white noise interleaved with repeated, binary white noise segments (5 or 10 s) to estimate noise.

We recorded monkey RGC activity to the same 60 minute white noise sequence at three different mean light levels, each differing by 1 log unit: 0.3 R*receptor^{-1}s^{-1}, 3 R*receptor^{-1}s^{-1}and 30 R*receptor^{-1}s^{-1}. These light levels fall under the scotopic regime, where mostly rod photoreceptors contribute to vision. The movies across the three light levels only differed in their mean pixel values which were 0.3 R*receptor^{-1}s^{-1}, 3 R*receptor^{-1}s^{-1}and 30 R*receptor^{-1}s^{-1}. The checkerboard stimuli in this experiment had 39 pixels x 30 pixels, where each pixel edge corresponded to approximately 140 μm on the retina surface. The refresh rate of the stimulus was set to 15 Hz (~ 67 ms per frame). In this work, we used a subset of 37 recorded RGCs that were classified as high quality units after spike sorting. This subset contained both the midget RGCs (that process fine details given their small receptive fields) and parasol RGCs (sensitive to motion given their large receptive fields) of ON and OFF types.

The rat experiments of Ruda et al. (2020) were performed at two light levels differing by 4 log units: 1 R*receptor^{-1}s^{-1}(scotopic light level where mostly rod photoreceptors contribute to vision) and 10,000 R*receptor^{-1}s^{-1}(photopic light level where cone photoreceptors predominantly contribute to vision). The white noise checkerboard movie in these experiments had 10 pixels x 11 pixels, with each pixel edge corresponding to approximately 252 μm on the retina. The refresh rate of the stimulus was 60 Hz and 30 Hz at the photopic and scotopic light levels, respectively. In this work, we used data from two rat experiments: a subset of 61 RGCs from Retina A and a subset of 58 RGCs from Retina B that were classified as high quality units after spike sorting. This subset contained OFF brisk sustained and OFF brisk transient RGC subtypes.

Photoreceptor layer

The proposed photoreceptor CNN layer builds upon a biophysical model of the phototransduction cascade by Angueyra et al. (2021) that incorporates the various feedforward and feedback molecular processes that convert photons into electrical signals, and therefore faithfully captures the photoreceptors' adaptation mechanisms. In the model, adaptation emerges through activity-dependent changes in the cyclic guanosine monophosphate (cGMP) turnover, produced by a light-induced increase in phosphodiesterase (PDE) activity and the calcium dependence of the rate of cGMP turnover; the time scale of adaptation depends largely on the kinetics of this process.

We implemented this biophysical model as a fully-trainable neural network layer in Keras (Chollet et al., 2015) in Python. This layer converts time-varying light intensity signal in units of receptor activations per photoreceptor per second (R*receptor^{-1}s^{-1}) into photocurrents (pA) by solving the differential equations that underlie the phototransduction cascade (biophysical model from Angueyra et al. (2021) reproduced in Supplementary Note 1). The photoreceptor layer has twelve parameters corresponding to the twelve interaction parameters of the biophysical model, all of which can (if needed) be trained through backpropagation using the Keras and TensorFlow package in Python. For the experiments presented herein, we fixed 8 of the parameters to their values from fits to photoreceptor data obtained as part of other experiments. Some of these parameters like the concentration of cyclic guanosine monophosphate (cGMP) in darkness vary across rod and cone photoreceptor types (rods and cones). Other parameters governing cGMP conversion into current, calcium concentration in the dark, affinity for Ca^{2+}, and hills coefficient are comparable across photoreceptor types. The remaining 4 parameters, set to be trainable, consisted of the photopigment decay rate σ, the phosphodiesterase (PDE) activation rate η, the PDE decay rate ϕ and the rate of Ca^{2+} extrusion β. These trainable parameters also differ across photoreceptor types. In the current version of the model, the photoreceptor parameters are shared by all the input pixels, and each pixel acts as an independent photoreceptor: I.e., the conversion of each pixel into photocurrents only depends on that pixel's previous values and not on the values of the other pixels.

Model architecture

We used the new photoreceptor layer at the input stage of a deep CNN (Fig. 1). The general architecture of the deep CNN we used was similar to Deep Retina (McIntosh et al., 2016), which we describe briefly here. In Deep Retina, the model input was a movie which passed through a convolutional layer operating in both space and time, followed by another convolutional layer

that only operated in the spatial dimension. The output of the final dense layer was the predicted spike rate of RGCs. The PR-CNN model differs from Deep Retina in two key ways: 1) the addition of the photoreceptor layer at the input stage; and 2) the use of batch normalization layers. We performed some experiments on models in which the photoreceptor layer was removed, and some in which the normalization layers were removed. These latter experiments provide the closest comparisons to Deep Retina.

Similar to Deep Retina, the PR-CNN model takes as input a movie (180 frames per training example where each frame corresponds to 8 ms) and outputs an instantaneous spike rate for each RGC at the end of that movie segment. To obtain the time series of RGC responses to longer movie stimuli, we feed into the model many 180-frame video samples taken from that longer movie, that correspond to 1-frame shifts. I.e., the model receives as inputs frames 1–180, 2–181, 3–182, etc., and outputs RGC responses at the times of movie frames 180, 181, 182, etc.

The photoreceptor layer of PR-CNN model converts each pixel of the input movie (where each pixel is in R*receptor^{-1}s^{-1}) into photocurrents (pA), based on the history of the pixel's value. The first 60 time points are then discarded to account for edge effects. The output of this layer is a movie that is 120 frames long, and the same spatial dimensions as the input visual stimuli. This photocurrents movie is then normalized by its mean and variance using a Layer Normalization layer (LayerNorm). The resulting movie is then passed through a conventional 3-layer CNN, where the first layer is a 3D convolutional layer operating in both the spatial and temporal dimensions. The output of the 3D convolutional layer is a 2D image which is normalized using Batch Normalization (BatchNorm) and then passed through a rectifying (ReLU) nonlinearity. All the temporal information from the movie is extracted by this layer as the temporal dimension of the convolutional filter is the same as the temporal dimension of the movie (120 in this case). To down sample the spatial dimensions, we applied a 2D max pool operation that took the maximum value over 2x2 patches of the previous layer's output. The subsequent 2D CNN layers are followed by a final, fully connected layer that outputs the predicted spike rate for each RGC in the dataset.

Model parameters were optimized using Adam (Kingma and Ba, 2015), where the loss function was given by the negative log-likelihood under Poisson spike generation. Photoreceptor parameters were initialized to their known values from fits to experimental data. This initialization helped stabilize the joint training of the photoreceptor model with the downstream layers. The network layers were regularized with a L_2 weight penalty at each layer, including the photoreceptor layer.In addition, a L_1 penalty was applied to the output of the dense layer because the neural activity itself is relatively sparse and L_1 penalties are known to induce sparsity. The number of channels in each CNN layer and the filter sizes were optimized by a grid search for each model type. These details therefore differed for the models with and without the photoreceptor layer or normalization layers. Learning rates were set to 0.001.

RGC sensitivity estimates from model gradients

For a given model RGC, we computed the gradient of its output spiking rate with respect to the pixel values in the input movie segments, similar to Maheswaranathan et al. (2018); Goldin et al. (2022). These gradients were evaluated for different binary white noise movie segments from the experiment. In total we had 400,000 input movie segments, generated by incrementing the white noise movie that was shown to the retina forward by one frame at a time (where 1 frame corresponds to 8 ms). These input movie segments spanned a total duration of 54 minutes. Since all the models were implemented with TensorFlow (Abadi et al., 2016), we calculated the gradients using automatic differentiation.

The resulting gradients were decomposed into spatial and temporal components (Fig. 4), similar to the way the spatial and temporal receptive fields are computed from the spike triggered average (STA) analysis applied directly to experimental data.

We normalized the spatial component to have unit mean. By doing so, our process of decomposing the spatio-temporal RF into spatial and temporal components assigned any variations in the RF's amplitude only to the temporal component. The peak of these temporal receptive fields, i.e. the sensitivity, showed substantial variation across the white noise movie segments (Fig. 5a). We grouped the temporal components into 10 equal bins based on their peak values, allowing us to identify the movie segments that composed each bin. The peak of the average temporal receptive field within each bin (Fig. 5b-c) gave us the sensitivity within that bin.

Retrospective analysis of experimental data

To validate the model's predictions about the specific set of movie segments that drove the RGC with a particular sensitivity, we re-analyzed the experimental data using reverse correlation (Dayan and Abbott, 2001) separately on the movie segments that corresponded to each sensitivity bin. We then decomposed the estimated receptive field into its spatial and temporal components. For each bin, we estimated the RGC sensitivity as the peak value of the resulting temporal receptive field. The sensitivity per bin was normalized by the sensitivity of the strongest bin.

Statistics and reproducibility

We applied statistical tests to compare the performance between models (Fig. 3d). We performed a two-tailed t-test of the null hypothesis that there was no difference between the fraction of explainable variance explained (FEV) by the CNN model and the PR-CNN model across a sample of 37 RGCs. In Fig. 5d-f, we performed a paired one-tailed Wilcoxon signed-rank test of the null hypothesis that there is no difference between the median of the distribution (N = 22 RGCs) of mean squared error (MSE) values between the CNN and PR-CNN model. The alternative hypothesis was that the median MSE of the PR-CNN model would be lower than that of the conventional CNN model. In Fig. 6e, we compared the response latencies between two different light levels obtained by different methods. For this, we performed a two-tailed Wilcoxon rank sum test of the null hypothesis that there was no difference between the distributions (N = 22 RGCs) of latencies at the two different light levels.

Data and code availability

The photoreceptor-CNN model can be downloaded from https://github.com/saadidrees/dynRet. All retina electrophysiology data used in this study can be made available upon reasonable request.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., and Zheng, X. Tensorflow: A system for large-scale machine learning. 2016.

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., and Farhan, L. *Review of deep learning: concepts, CNN architectures, challenges, applications, future directions*, volume 8. Springer International Publishing, 2021. ISBN 4053702100444. doi: 10.1186/s40537-021-00444-8. URL https://doi.org/10.1186/s40537-021-00444-8.

Angueyra, J. M., Baudin, J., Schwartz, G. W., and Rieke, F. Multiple time scales of adaptation allow cones to encode the inputs created by visual exploration of natural scenes. *bioRxiv*, page 2021.02.13.431101, 2021.

Ba, J. L., Kiros, J. R., and Hinton, G. E. Layer Normalization. 2016.

Baylor, D. A. and Fettiplace, R. Transmission from photoreceptors to ganglion cells in turtle retina. *The Journal of Physiology*, 271(2):391–424, oct 1977.

Baylor, D. A. and Hodgkin, A. L. Detection and resolution of visual stimuli by turtle photoreceptors. *The Journal of Physiology*, 234(1):163–198, oct 1973.

Baylor, D., Nunn, B., and Schnapf, J. The photocurrent, noise and spectral sensitivity of rods of. *Journal of Physiology*, 357:575–607, 1984.

Beaudoin, D., Borghuis, B., and Demb, J. Cellular basis for contrast gain control over the receptive field center of mammalian retinal ganglion cells. *The Journal of neuroscience*, 27(10):2636, 2007. doi: 10.1523/JNEUROSCI.4610-06.2007.

Benda, J. Neural adaptation. *Current Biology*, 31(3):R110–R116, 2021. doi: 10.1016/j.cub.2020.11.054.

Bosking, W. H., Beauchamp, M. S., and Yoshor, D. Electrical stimulation of visual cortex: relevance for the development of visual cortical prosthetics. *Annual review of vision science*, 3:141–166, 2017.

Brien, B. J. O., Isayama, T., Richardson, R., and Berson, D. M. Intrinsic physiological properties of cat retinal ganglion cells. *The Journal of physiology*, pages 787–802, 2002. doi: 10.1013/jphysiol.2001.013009.

Cadena, S. A., Denfield, G. H., Walker, E. Y., Gatys, L. A., Tolias, A. S., Bethge, M., and Ecker, A. S. Deep convolutional models improve predictions of macaque V1 responses to natural images Author summary. *Plos Computational Biology*, pages 1–28, 2017.

Cao, D., Zele, A. J., and Pokorny, J. Linking impulse response functions to reaction time: Rod and cone reaction time data and a computational model. *Vision Research*, 47(8): 1060–1074, apr 2007.

Chander, D. and Chichilnisky, E. J. Adaptation to temporal contrast in primate and salamander retina. *J Neurosci*, 21(24):9904–9916, 2001. doi: 21/24/9904[pii].

Chang, L., Ran, Y., Auferkorte, O., Butz, E., Hueser, L., Haverkamp, S., Euler, T., and Schubert, T. Spike desensitisation as a mechanism for high-contrast selectivity in retinal ganglion cells. *bioRxiv*, pages 1–39, 2022. doi: 10.1101/2022.08.11.503581.

Chollet, F. et al. Keras, 2015. URL https://github.com/fchollet/keras.

Chollet, F. Xception: Deep learning with depthwise separable convolutions. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-January:1800–1807, 2017.

Clark, D. A., Benichou, R., Meister, M., and Azeredo da Silveira, R. Dynamical Adaptation in Photoreceptors. *PLoS Computational Biology*, 9(11), 2013.

Cowley, B. R., Calhoun, A. J., Rangarajan, N., Pillow, J. W., and Murthy, M. One-to-one mapping between deep network units and real neurons uncovers a visual population code for social behavior. *bioRxiv*, page 2022.07.18.500505, 2022.

Cui, Y., Wang, Y. V., Park, S. J. H., Demb, J. B., and Butts, D. A. Divisive suppression explains high-precision firing and contrast adaptation in retinal ganglion cells. *eLife*, 5:1–25, 2016. doi: 10.7554/elife.19460.

Dayan, P. and Abbott, L. F. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. The MIT Press, 2001.

Demb, J. B. Functional circuitry of visual adaptation in the retina. *The Journal of physiology*, 586(Pt 18):4377–84, oct 2008. doi: 10.1113/jphysiol.2008.156638.

Drinnenberg, A., Franke, F., Morikawa, R. K., Jüttner, J., Hillier, D., Hantz, P., Hierlemann, A., Azeredo da Silveira, R., and Roska, B. How Diverse Retinal Functions Arise from Feedback at the First Visual Synapse. *Neuron*, pages 1–18, 2018. doi: 10.1016/j.neuron.2018.06.001.

Dunn, F. A., Doan, T., Sampath, A. P., and Rieke, F. Controlling the gain of rod-mediated signals in the mammalian retina. *Journal of Neuroscience*, 26(15):3959–3970, 2006. doi: 10.1523/JNEUROSCI.5148-05.2006.

Fain, G. L., Matthews, H. R., Cornwall, M. C., and Koutalos, Y. Adaptation in vertebrate photoreceptors. *Physiological Reviews*, 81(1):117–151, 2001. doi: 10.1152/physrev.2001.81.1.117.

Farrow, K., Teixeira, M., Szikra, T., Viney, T. J., Balint, K., Yonehara, K., and Roska, B. Ambient illumination toggles a neuronal circuit switch in the retina and visual perception at cone threshold. *Neuron*, 78(2):325–338, apr 2013.

Field, G. D., Sher, A., Gauthier, J. L., Greschner, M., Shlens, J., Litke, A. M., and Chichilnisky, E. J. Spatial properties and functional organization of small bistratified ganglion cells in primate retina. *Journal of Neuroscience*, 27(48):13261–13272, 2007.

Geadah, V., Horoi, S., Kerg, G., Wolf, G., and Lajoie, G. Goal-driven optimization of single-neuron properties in artificial networks reveals regularization role of neural diversity and adaptation. *bioRxiv*, 2022.

Goldin, M. A., Lefebvre, B., Virgili, S., Pham Van Cang, M. K., Ecker, A., Mora, T., Ferrari, U., and Marre, O. Context-dependent selectivity to natural images in the retina. *Nature Communications*, 13(1):5556, 2022.

Gollisch, T. and Meister, M. Eye smarter than scientists believed: neural computations in circuits of the retina. *Neuron*, 65(2):150–64, jan 2010. doi: 10.1016/j.neuron.2009.12.009.

Grimes, W. N., Schwartz, G. W., and Rieke, F. The synaptic and circuit mechanisms underlying a change in spatial encoding in the retina. *Neuron*, 82(2):460–73, apr 2014. doi: 10.1016/j.neuron.2014.02.037.

Hornik, K., Stinchcombe, M., and White, H. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, jan 1989.

Idrees, S., Baumann, M. P., Franke, F., Münch, T. A., and Hafed, Z. M. Perceptual saccadic suppression starts in the retina. *Nature Communications*, 11(1):1977, dec 2020.

Idrees, S., Baumann, M.-p., Korympidou, M. M., Schubert, T., Kling, A., Franke, K., Hafed, Z. M., Franke, F., and Münch, T. A. Suppression without inhibition: how retinal computation contributes to saccadic suppression. *Communications Biology*, 5(1):692, dec 2022.

Ingram, N. T., Sampath, A. P., and Fain, G. L. Why are rods more sensitive than cones? *Journal of Physiology*, 594(19):5415–5426, 2016.

Ioffe, S. and Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *32nd International Conference on Machine Learning, ICML 2015*, 1:448–456, 2015.

Keat, J., Reinagel, P., Reid, R., and Meister, M. Predicting Every Spike. *Neuron*, 30(3): 803–817, 2001. doi: 10.1016/s0896-6273(01)00322-1.

Kim, K. J. and Rieke, F. Temporal Contrast Adaptation in the Input and Output Signals of Salamander Retinal Ganglion Cells. 21(1):287–299, 2001.

Kim, K. J. and Rieke, F. Slow Na+ inactivation and variance adaptation in salamander retinal ganglion cells. *Journal of Neuroscience*, 23(4):1506–1516, 2003. doi: 10.1523/jneurosci.23-04-01506.2003.

Kindel, W. F., Christensen, E. D., and Zylberberg, J. Using deep learning to probe the neural code for images in primary visual cortex. *Journal of Vision*, 19(4):29, apr 2019.

Kingma, D. P. and Ba, J. L. Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pages 1–15, 2015.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, may 2017.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. Backpropagation applied to digit recognition, 1989.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2323, 1998. doi: 10.1109/5.726791.

Lindsey, J., Ocko, S. A., Ganguli, S., and Deny, S. A unified theory of early visual representations from retina to cortex through anatomically constrained deep cnNs. *7th International Conference on Learning Representations, ICLR 2019*, pages 1–17, 2019.

Maheswaranathan, N., McIntosh, L. T., Tanaka, H., Grant, S., Kastner, D. B., Melander, J. B., Nayebi, A., Brezovec, L., Wang, J., Ganguli, S., and Baccus, S. A. The dynamic neural code of the retina for natural scenes. *bioRxiv*, 2018. doi: 10.1101/340943.

Mano, O., Creamer, M. S., Badwan, B. A., and Clark, D. A. Predicting individual neuron responses with anatomically constrained task optimization. *Current Biology*, 31(18):4062–4075.e4, 2021. doi: 10.1016/j.cub.2021.06.090.

McIntosh, L. T., Maheswaranathan, N., Nayebi, A., Ganguli, S., and Baccus, S. A. Deep learning models of the retinal response to natural scenes. *Advances in Neural Information Processing Systems*, pages 1369–1377, 2016.

O'Brien, B. J., Isayama, T., Richardson, R., and Berson, D. M. Intrinsic physiological properties of cat retinal ganglion cells. *Journal of Physiology*, 538(3):787–802, 2002. doi: 10.1113/jphysiol.2001.013009.

Olshausen, B. A. and Field, D. J. What Is the Other 85 Percent of V1 Doing? pages 182–212, jan 2004. doi: 10.1093/acprof:oso/9780195148220.003.0010.

Rieke, F. and Rudd, M. E. The Challenges Natural Images Pose for Visual Adaptation. *Neuron*, 64(5):605–616, 2009. doi: 10.1016/j.neuron.2009.11.028.

Ruda, K., Zylberberg, J., and Field, G. D. Ignoring correlated activity causes a failure of retinal population codes. *Nature Communications*, 11(1):1–15, 2020.

Schneeweis, D. M. and Schnapf, J. L. Photovoltage of rods and cones in the macaque retina. *Science*, 268(5213):1053–1056, 1995.

Schwartz, G. W. and Rieke, F. Controlling gain one photon at a time. *eLife*, 2013(2):1–20, 2013. doi: 10.7554/eLife.00467.

Shapley, R. and Enroth-Cugell, C. Chapter 9 Visual adaptation and retinal gain controls. *Progress in Retinal Research*, 3:263–346, jan 1984. doi: 10.1016/0278-4327(84)90011-7.

Shapley, R. M. and Victor, J. D. The effect of contrast on the transfer properties of cat retinal ganglion cells. *The Journal of Physiology*, 285(1):275–298, dec 1978. doi: 10.1113/jphysiol.1978.sp012571.

Shlens, J., Field, G. D., Gauthier, J. L., Grivich, M. I., Petrusca, D., Sher, A., Litke, A. M., and Chichilnisky, E. J. The structure of multi-neuron firing patterns in primate retina. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 26(32): 8254–66, aug 2006.

Simonyan, K. and Zisserman, A. Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pages 1–14, 2015.

Tanaka, H., Nayebi, A., Maheswaranathan, N., McIntosh, L., Baccus, S. A., and Ganguli, S. From deep learning to mechanistic understanding in neuroscience: the structure of retinal prediction. (NeurIPS):1–11, 2019.

Tikidji-Hamburyan, A., Reinhard, K., Seitter, H., Hovhannisyan, A., Procyk, C. a., Allen, A. E., Schenk, M., Lucas, R. J., and Münch, T. a. Retinal output changes qualitatively with every change in ambient illuminance. *Nature Neuroscience*, 18:66–74, 2015.

Walker, E. Y., Sinz, F. H., Cobos, E., Muhammad, T., Froudarakis, E., Fahey, P. G., Ecker, A. S., Reimer, J., Pitkow, X., and Tolias, A. S. Inception loops discover what excites neurons most using deep predictive models. *Nature Neuroscience*, 22(December), 2019. doi: 10.1038/s41593-019-0517-x.

Willmore, B. D. and King, A. J. Adaptation in Auditory Processing. *Physiological Reviews*, 103(2):1025–1058, 2023. doi: 10.1152/physrev.00011.2022.

Wong, R. C. S., Cloherty, S. L., Ibbotson, M. R., and O'Brien, B. J. Intrinsic physiological properties of rat retinal ganglion cells with a comparative analysis. *Journal of Neurophysiology*, 108(7):2008–2023, oct 2012. doi: 10.1152/jn.01091.2011.

Yamins, D. L. K. and Dicarlo, J. J. Using goal-driven deep learning models to understand sensory cortex. (February), 2016. doi: 10.1038/nn.4244.

Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., and Dicarlo, J. J. Performance-optimized hierarchical models predict neural responses in higher visual cortex. 2014. doi: 10.1073/pnas.1403112111.

Yan, Q., Zheng, Y., Jia, S., Zhang, Y., Yu, Z., Chen, F., Tian, Y., Huang, T., and Liu, J. K. Revealing Fine Structures of the Retinal Receptive Field by Deep-Learning Networks. *IEEE transactions on cybernetics*, 52(1):39–50, jan 2022.

Yu, Z., Turner, M. H., Baudin, J., and Rieke, F. Adaptation in cone photoreceptors contributes to an unexpected insensitivity of primate On parasol retinal ganglion cells to spatial structure in natural images. *eLife*, 11:1–28, 2022.

Zufall, F. and Leinders-Zufall, T. The cellular and molecular basis of odor adaptation. *Chemical Senses*, 25(4):473–481, 2000. doi: 10.1093/chemse/25.4.473.

Acknowledgments

This work was supported by VISTA: Vision science to applications fellowship to SI; Canada Research Chair grant and NSERC Discovery grant (RGPIN-2019-06379) to JZ.

Competing interests

The authors declare no competing interests.