Deep learning models reveal internal structure and diverse computations in the retina under natural scenes

Niru Maheswaranathan[1,5]*, Lane McIntosh[1]*, David B. Kastner[1,6], Josh Melander[1], Luke Brezovec[1], Aran Nayebi[1], Julia Wang[2], Surya Ganguli[3] and Stephen A. Baccus[4]
Stanford University. [1]Currently at: Google Brain. [2]Currently at: UCSF
[1]*Neuroscience Program, Stanford University School of Medicine, Stanford, CA*
[2] *Stanford University, Stanford, CA*
[3] *Department of Applied Physics, Stanford University, Stanford, CA*
[4] *Department of Neurobiology, Stanford University, Stanford, CA*
[5]*Present address: Google Brain, Mountain View, CA*
[6]*Present address: Department of Psychiatry, University of California, San Francisco, CA*
* These authors contributed equally

Abstract

The normal function of the retina is to convey information about natural visual images. It is this visual environment that has driven evolution, and that is clinically relevant. Yet nearly all of our understanding of the neural computations, biological function, and circuit mechanisms of the retina comes in the context of artificially structured stimuli such as flashing spots, moving bars and white noise. It is fundamentally unclear how these artificial stimuli are related to circuit processes engaged under natural stimuli. A key barrier is the lack of methods for analyzing retinal responses to natural images. We addressed both these issues by applying convolutional neural network models (CNNs) to capture retinal responses to natural scenes. We find that CNN models predict natural scene responses with high accuracy, achieving performance close to the fundamental limits of predictability set by intrinsic cellular variability. Furthermore, individual internal units of the model are highly correlated with actual retinal interneuron responses that were recorded separately and never presented to the model during training. Finally, we find that models fit only to natural scenes, but not white noise, reproduce a range of phenomena previously described using distinct artificial stimuli, including frequency doubling, latency encoding, motion anticipation, fast contrast adaptation, synchronized responses to motion reversal and object motion sensitivity. Further examination of the model revealed extremely rapid context dependence of retinal feature sensitivity under natural scenes using an analysis not feasible from direct examination of retinal responses. Overall, these results show that nonlinear retinal processes engaged by artificial stimuli are also engaged in and relevant to natural visual processing, and that CNN models form a powerful and unifying tool to study how sensory circuitry produces computations in a natural context.

While evolution sculpted the retina to process natural visual stimuli, our inability to model such processing has led to a focus on understanding retinal responses to white noise or simple artificial stimuli. This raises a fundamentally unanswered question: are the earliest visual computations elicited by these artificial stimuli at all relevant to understanding computations elicited by natural stimuli?

In recent years deep learning has led to dramatic advances in our ability to discriminate and classify natural images using feedforward convolutional networks.[1] However, when deep or recurrent neural networks are used to model neurobiological systems, the comparison between model activity and brain activity is often only verified at a coarse resolution, at the level of entire population dynamics[2,3], or linear combinations of neurons [4-6], and in contexts that are not very different from the contexts that the networks were originally trained in. Thus, the advent of deep learning as a modeling approach in neuroscience raises two more fundamental unanswered questions. First, can artificial deep neural network computations capture neurobiological circuit computations, at the level of individual neurons? And second, can such network models generalize to contexts that are vastly different from those in which they were trained, providing support that they actually capture ground truth information about neurobiological circuit computation?

In this work, we address the above questions in the context of the first steps of vision by developing highly accurate convolutional neural network models of the retinal response to natural scenes. The internal functional architecture of our models matches that of the retina at the level of individual neurons, and moreover our models generalize from natural scenes, but not white noise, to a wide range of artificially structured stimuli with vastly different statistics. Thus this work provides quantitative validation for the deep learning approach to neuroscience in an experimentally accessible sensory circuit, places decades of work[7-15] on retinal responses to artificially structured stimuli on much firmer foundations of ethological relevance, and highlights the fundamental importance of studying sensory circuit responses to natural stimuli.

CNNs learn accurate models of the retinal response to natural scenes

Given the vertebrate retina has three layers of cell bodies, we tested whether three layer CNN models (Figure 1) could predict the responses of populations of salamander retinal ganglion cells responding to a 50 minute sequence of either natural images or spatiotemporal white noise. Natural scene images changed every second, and were jittered with the statistics of fixational eye movements[16,17], creating a spatiotemporal stimulus. The model had up to eight different model cell types in each of the first and second layers, with each cell type having a distinct receptive field, and a final fully connected layer that represented the responses of individual ganglion cells. We found that CNN models could predict the responses of ganglion cells to either natural scenes or white noise nearly up to a fundamental limit of precision set by intrinsic neural variability, and were substantially more accurate than linear-nonlinear (LN) models [18] or generalized linear models (GLMs) [19] (Figure 1B, C).

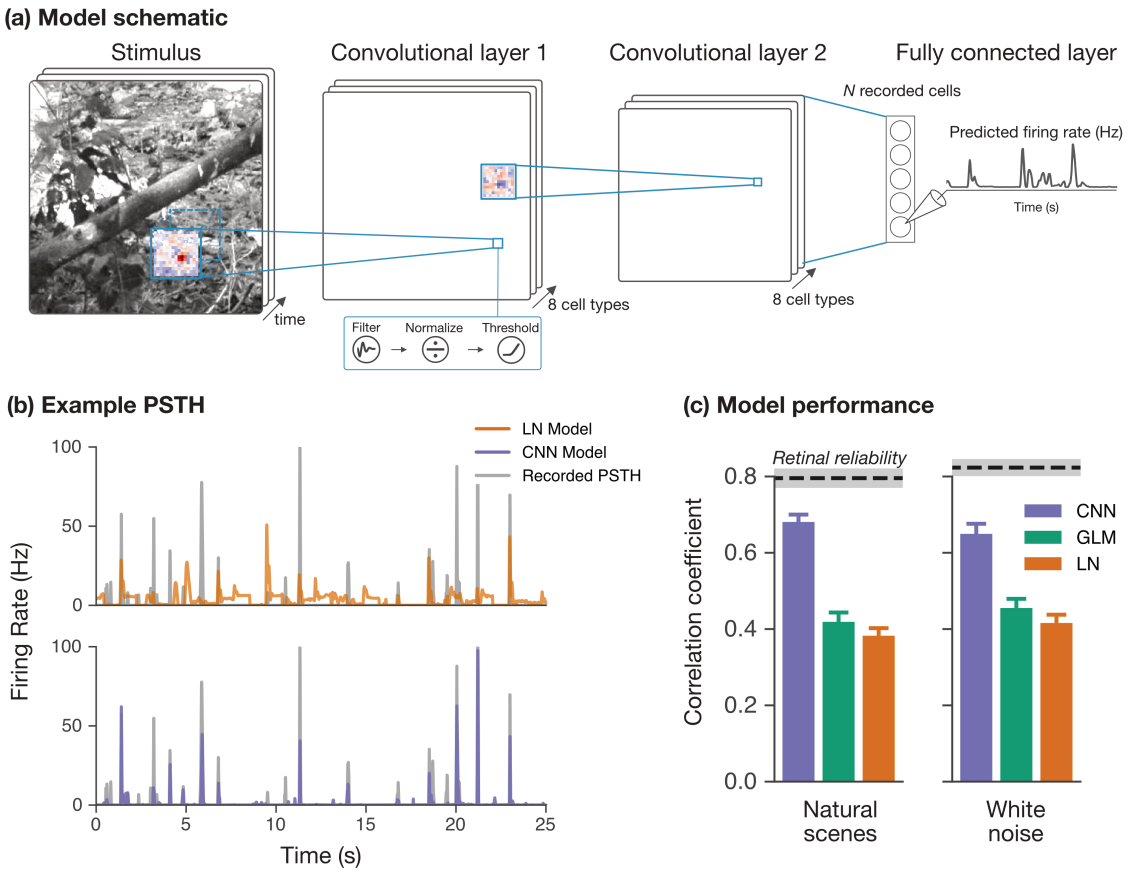


Figure 1. Convolutional neural networks provide accurate models of the retinal response to natural scenes. (A) Convolutional neural network model trained to predict the firing rate of simultaneously recorded retinal ganglion cells from the spatiotemporal movie of natural scenes. The first layer is a spatiotemporal convolution, the second is a spatial convolution, and the third is a final dense layer, with rectifying nonlinearities in between each layer. Each location within the model also has a single parameter that scales the amplitude of the response. (B) PSTHs comparing recorded data and Linear-Nonlinear (LN) or CNN models for the test data set. (C) Comparison of LN, Generalized Linear Model (GLM) and CNN model predictions for a 25 second segment of a natural scene movie. Correlation coefficients are for the test data set, as compared to the retinal reliability of ganglion cell PSTHs correlated between different sets of trials (dotted line is mean, grey bar is 1 s.e.m.)

CNN internal units have receptive field structure matching retinal interneurons

To examine whether the internal computations of CNN models were similar to those expected in the retina, we computed receptive fields for first and second layer units (model cells) in CNNs trained on responses to natural scenes. Receptive fields were computed by the standard method of reverse correlating the activity of model units with a white noise input. We found that the receptive fields of CNN units had the well-known structure of retinal interneurons[20,21], with a spatially localized center-surround structure (Fig. 2A-B, Extended Data, Fig. 1).

Comparing the responses to full-field flashes of CNNs trained on natural scenes and a previously recorded dataset of amacrine and bipolar cells, we observed that both first and second CNN layer responses exhibited sustained and transient flash responses [22,23], depending on the unit type, that qualitatively matched the flash responses of real

bipolar and amacrine cells (Figure 2C). Bipolar cells do not have strongly rectified responses, which matched the first layer units of the model by construction, as at this level signals have only passed through a linear filter. The second layer of the model contained both nearly-linear and more rectified responses as found in the amacrine cell population [24] (Figure 2C-D). Thus, CNNs reproduce the progression of diversity in the retina, with the first layer having stereotyped, bipolar-like units and the second layer having a diverse set of both linear and nonlinear units.

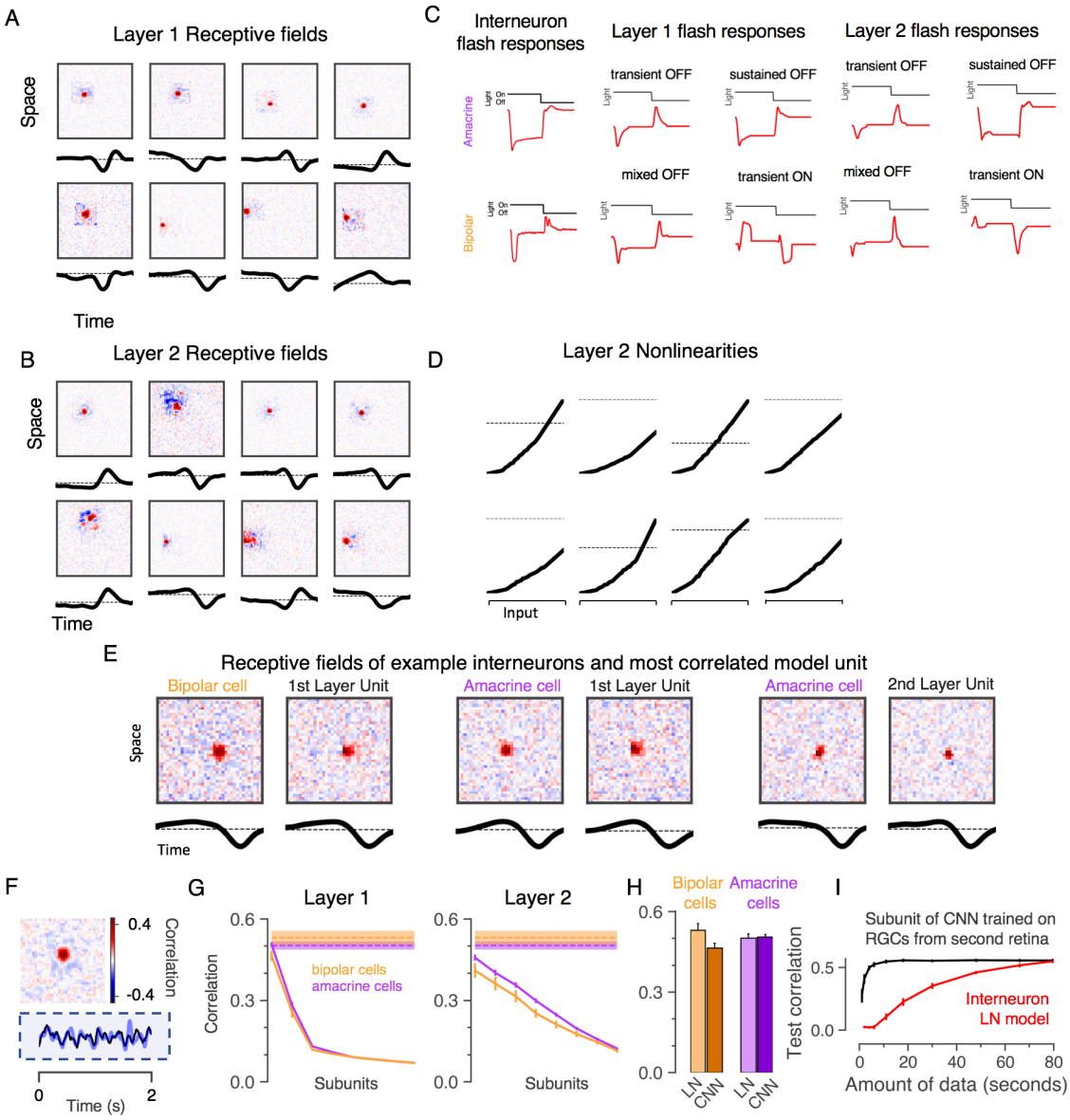


Figure 2. Model internal units are correlated with interneuron responses. (A) Receptive fields of model units in Layer 1 shown as separable spatial and temporal components. (B) Same for Layer 2. (C) Flash responses of (Left) an example OFF sustained, narrow field amacrine cell and an OFF bipolar cell, (Middle) example first layer units, and (Right) example second layer units. (D) Nonlinearities of an LN model computed for Layer 2 units, corresponding to units shown in (B). (E). Spatiotemporal receptive fields of example interneurons recorded from a separate retina, and the model unit that was most correlated with that interneuron. The model was never fit to the interneuron's response. (F) Top. Correlation map of a model cell type with the response of

an interneuron recorded from a different retina to a white noise stimulus. Each pixel is the correlation between the interneuron and a different spatial location within a single model cell type. Bottom. Responses compared to the most correlated model unit and the interneuron. (G) For Layers 1 and 2 the average correlation between different interneuron types (7 bipolar, 26 amacrine) and model cell types ranked from most correlated model unit (left) to least (right). Dotted lines indicate the maximum correlation in our dataset observed between neurons, computed by fitting an LN model to Cell 1, and then spatially shifting that model to the location of Cell 2 so that the LN model was presented the stimulus experienced by Cell 2. Thus, the correlation between model units and interneurons approaches the variability between interneurons themselves. (H) Average correlations between an interneuron's response and an LN model fit to the same interneuron, or the most correlated unit from a CNN model fit to a different retina. (I). Correlation between interneuron recordings and the most correlated CNN unit from a different retina or an LN model fit to the same interneuron a function of the amount of data.

Because retinal cell types are not perfectly homogeneous [25,26], the model contained for each location a parameter that scaled the receptive field amplitude. These parameters created a modest improvement in performance (0.69 $\pm$ 0.02 vs 0.66 $\pm$ 0.02 correlation coefficient without scaling parameters), and created receptive fields in layer 2 that had small differences within a cell type (Extended Data, Fig. 2) as seen in retinal neurons. Although layer 2 did contain receptive fields that varied to a greater extent, these arose from retinal locations that did not contribute to the recorded neurons, and were therefore unconstrained in the model (Extended Data, Fig. 3).

CNNs internal units are highly correlated with interneuron responses

In inferotemporal cortex, units of CNNs have been shown to be correlated with a linear combination of the activity of individual neurons [4] making it difficult to draw conclusions about individual neurons by an examination of CNN units. We compared the activity of CNN units to interneuron recordings performed on separate retinae that the model was never fit to. The stimulus presented to the retina and separately to the model was a spatiotemporal white noise checkerboard, a stimulus that has no spatiotemporal correlations except for the 50 µm size and 10 ms duration of square stimulus regions. We compared each interneuron recording with 8 units of the first layer and 8 units of the second layer at each location to find the most correlated unit in the model at the location of the cell. We found that each recorded interneuron was highly correlated with a particular unit type, and only at a single location (Figure 2E-H). Spatiotemporal receptive fields were highly similar between recorded interneurons, and their most correlated model cell type (Fig. 2E). The magnitude of this correlation approached that found between recorded interneurons in our dataset, correcting for different the spatial locations of recorded interneurons (Figure 2F,G). This correlation was specific for individual unit types, as could be observed by ranking the unit types from most to least correlated, and finding that the second and lower most correlated unit types were substantially less correlated with the interneuron than the most correlated unit (Fig. 2G).

Moreover, we find that these model units fit to a different retina predict the response of an interneuron as well as LN models directly fit to the interneuron membrane potential (Figure 2H), and can do so with much less data (Figure 2I). Using a CNN model trained on the retinal ganglion cell responses to natural scenes from a different retina, it took less than 10 seconds of data to find the model unit most correlated with the recorded

interneuron, producing a model on par with an LN model directly fit to 80 seconds of interneuron membrane potential (Figure 2H). Thus, fitting a CNN model to the natural scene responses of retinal ganglion cells alone generates a model of an entire population of interneurons, many of which have high correlation with measured interneuron responses created with a different stimulus and a different retina.

CNNs replicate wide range of retinal phenomena

Numerous nonlinear computations have been identified by presenting artificial stimuli to the retina, including flashing spots, moving bars and white noise. However, we neither understand to what degree natural vision engages these diverse retinal computations elicited by artificial stimuli, nor understand the relationship between these computations under natural scenes and underlying retinal circuitry. We tested models fit either only to natural scenes or white noise by exposing them to a battery of structured stimuli previously used in the literature to identify and describe retinal phenomena. We focused on effects shorter than 400 ms, which was the longest timescale our model could reproduce as limited by the first layer spatiotemporal filter. Remarkably, the CNN model exhibited fast contrast [7-9] adaptation (Fig. 3A), latency encoding [10] (Fig. 3B), synchronized responses to motion reversal [11](Fig. 3C), motion anticipation [12] (Fig. 3D), the omitted stimulus response [13] (Fig. 3E), frequency doubling in response to reversing gratings [14] (Fig. 3F) and polarity reversal [15](Fig. 3G). All of these response properties arose in a single CNN model simply as a by-product of optimizing the models to capture ganglion cell responses to natural scenes. CNN models trained on white noise did not exhibit all of these phenomena, indicating that natural scene statistics trigger nonlinear computations that white noise does not. Even though these natural scenes consisted only of a sequence of images jittered with the statistics of fixational eye movements (the stimulus contained no explicit object motion or periodic patterns), the CNNs still exhibited motion anticipation and reversal, and the omitted stimulus response.

The only retinal phenomenon we tested that was not captured by the model was the object motion sensitive (OMS) response[17]. We hypothesized that the absence of an OMS response in the model was due to the lack of differential motion in the training stimulus, and trained additional models on the retinal response to movies of swimming fish that include differential motion. We found that these models did indeed exhibit an OMS response (Fig. 3H). This observation supports the conclusion that the presence of a computation in the model indicates that the computation is engaged under the stimulus used to fit the model. We thus conclude that the nonlinear circuit properties that produce these phenomena are indeed engaged during natural scenes, and that all of these identified computations, previously identified only with artificial stimuli, are nonetheless highly relevant to natural vision.

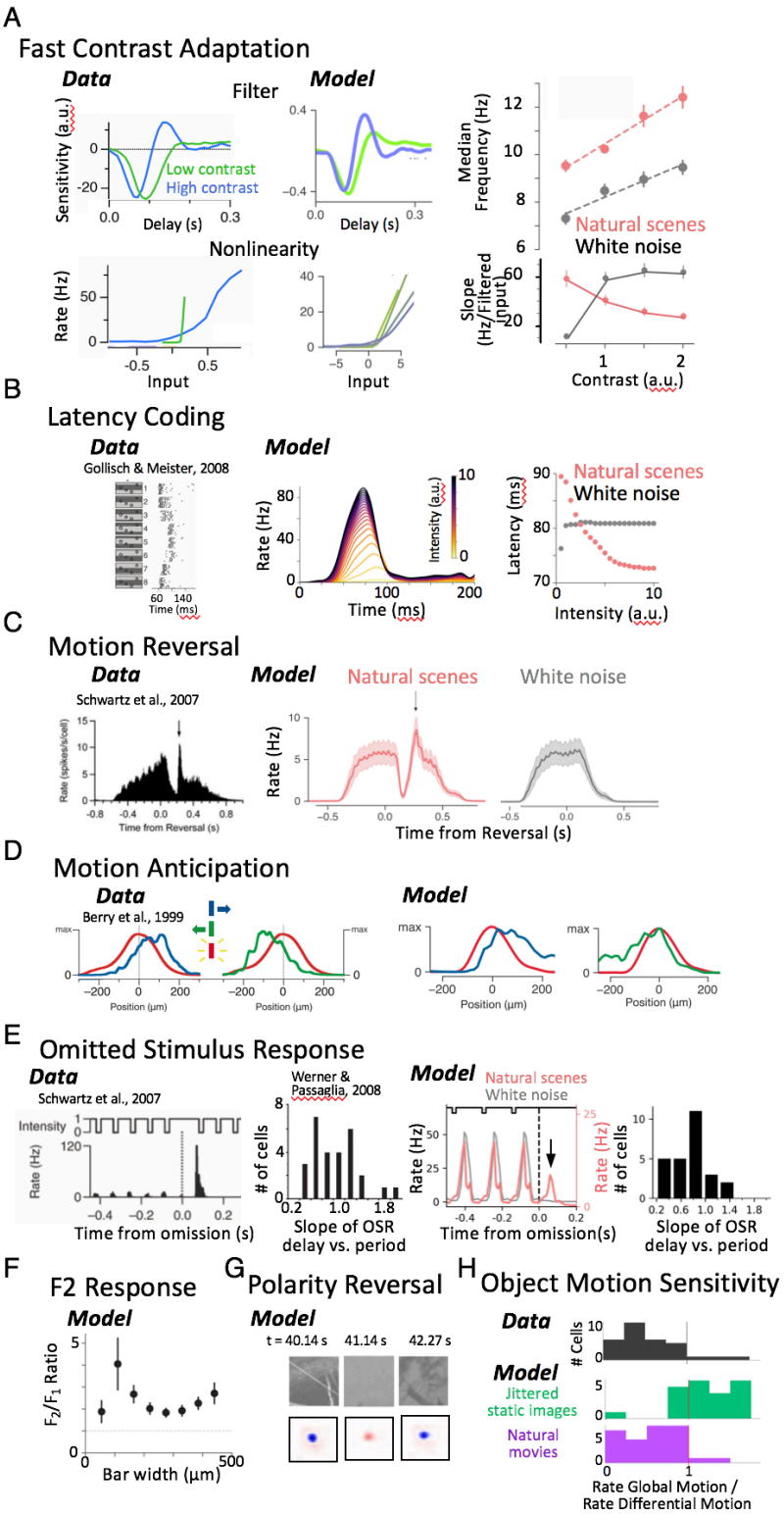


Figure 3. CNN models reveal that many nonlinear retinal computations are engaged in natural scenes. (A) Contrast adaptation. Left: LN model during high (35 %) and low (5 %) contrast, showing changing temporal filters and gain as shown by the slope of the nonlinearity. Middle: Temporal filters of a CNN model ganglion cell at high and low contrast, as well as nonlinearities at several contrasts from low (green) to high (blue). Right, Top: Median temporal

frequency taken from the Fourier transform of the temporal filter, averaged over a population of ganglion cells as a function of contrast. Results shown for models fit to natural scenes and white noise. Right, Bottom: Averaged gain measured as the slope of the nonlinearity as a function of contrast, showing that CNN models decrease their decrease their gain with contrast when fit to natural scenes, but not when fit to white noise. (B) Latency encoding. Published results showing latency encoding as a function of the strength of a flashed stimulus that varies in position [10]. Middle: Flash response with intensities ranging from weak (yellow) to strong (purple). Right: Latency of the peak response vs. stimulus intensity for models trained on natural scenes or white noise. (C) Motion reversal. Stimulus consists of a moving bar that abruptly reverses direction at different positions. Left. Published results of a population of ganglion cells showing a synchronous response (arrow) to the reversal. Also shown is the population response of CNN model cells, trained for natural scenes (middle) or white noise (right). (D) Motion anticipation. Population ganglion cell responses to a flashed bar (red) vs motion to the right (blue) or left (green), from published results[12] (left) or the CNN model (right). (E) Omitted stimulus response (OSR). Left. Published results[13] showing the response to a missing stimulus following a train of flashes and a histogram of the slope of the OSR delay vs. the stimulus period. A slope of one indicates the OSR tracks the stimulus period exactly. Right. CNN model response to a sequence of three flashes. The OSR (arrow) appears for models trained on natural scenes but not white noise. Also shown is a histogram of the slope of the OSR delay vs. the stimulus period as it was varied from 8 – 20 Hz (F) Frequency doubling in response to reversing gratings of different width, computed as the ratio of the response at twice the stimulus frequency (F2) and the response at the stimulus frequency (F1). (G) Polarity reversal. Example reversal of polarity during a natural image sequence. Each panel shows the current image (top) and corresponding instantaneous receptive field (bottom) for an example cell at a fixed delay (~100 ms) relative to the stimulus at different times during the sequence, showing fast kernel reversal from an OFF- feature (blue) to ON- (red) and back. Rapid receptive field changes are further analyzed in Fig. 4. (H). Object Motion Sensitivity. CNN models were fit to either jittered static images or natural movies consisting of swimming fish in the presence of image jitter and saccade-like transitions. Stimuli were then shown to the model consisting of a jittering central grating surrounded by a jittering background grating. Gratings moved either synchronously (Global motion) representing eye movements, or asynchronously (Differential Motion) representing object motion. Shown is the ratio of firing rates in Global Motion to Differential Motion. A ratio much less than one indicates Object Motion Sensitivity. Results for (A-F) are from a population of 26 ganglion cells. Figures reproduced with permission from authors.

CNNs reveal extreme context dependence of the retinal code

Receptive fields in sensory neuroscience are typically thought of as representing a static sensory feature, although it is known that this feature can change due to adaptation to the statistics of the stimulus [27-29]. An attractive feature of CNN models is that the instantaneous receptive field can be easily computed as the gradient of the model output with respect to the current stimulus. Computing the gradient of the model revealed that in addition to the previously described property of polarity reversal (Fig. 3G) the instantaneous receptive field showed an extreme context sensitivity (Fig. 4), thereby revealing the full complexity of how visual feature sensitivities change during natural images. This shows that multiple parallel pathways, each encoding different features, are dynamically selected to generate the ganglion cell response to natural scenes even on a timescale as short as 10 ms.

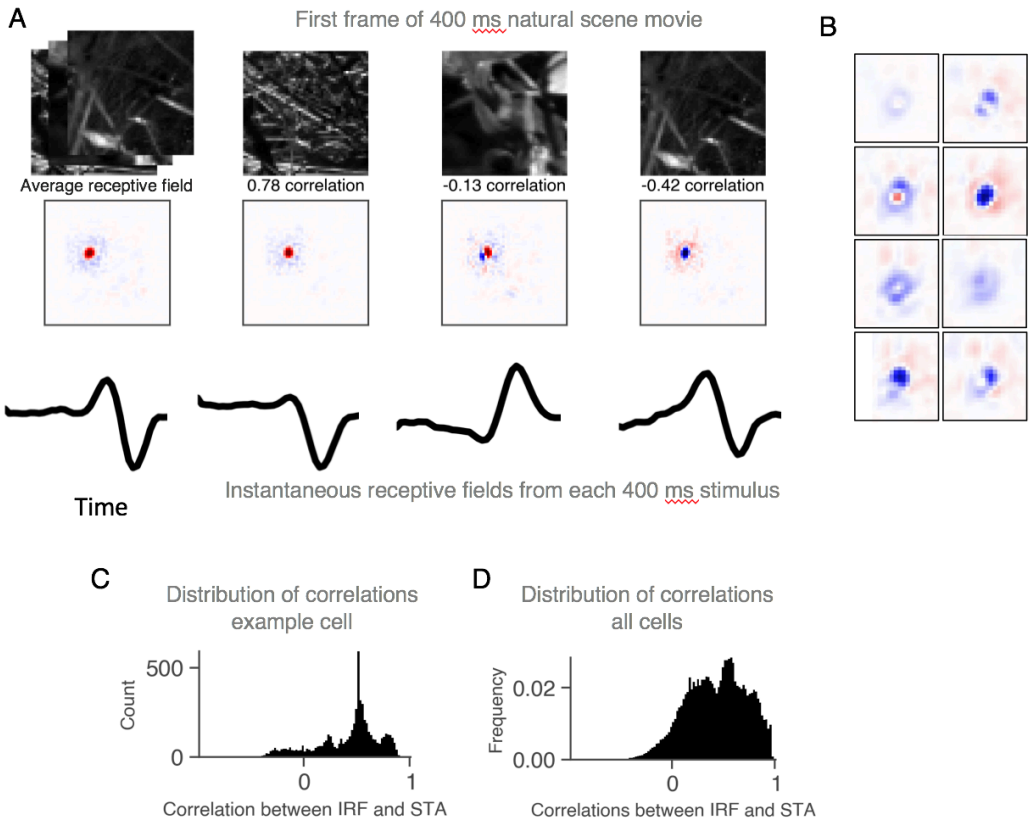


Figure 4. Extreme context dependence of retinal receptive fields. (A) Left. The average receptive field computed as the average gradient of the model output with respect to the stimulus. Right. The instantaneous spatiotemporal receptive field (IRF) computed as the instantaneous gradient. At top are shown different images at a fixed latency relative to the time that each IRF was computed. (B) Expanded view of different IRFs for a single cell at different times during the stimulus. (C). For a single model ganglion cell, the distribution of correlation coefficients of each IRF with the average receptive field, indicating the range of variation of the IRF. (D) Same as (C) averaged over a population of 37 ganglion cells.

Overall, our results indicate that CNN models accurately capture the responses of the retina to natural scenes and generalize to reproduce much of the phenomenology previously described using an internal representation that is highly correlated with actual interneuron responses. Although we cannot state at this point whether there is a one-to-one correspondence of model units to interneurons, the current results are sufficient to indicate the feasibility of a program of successive refinement of the model. Such a program would use recorded interneuron responses, directly measured effects of current injection into those interneurons on retinal ganglion cell responses [24,30] and connectomics data[31] to constrain and re-optimize these models, with the promise of a complete computational and circuit level description of the retina under natural scenes.

Methods

Visual Stimuli. A video monitor projected the visual stimuli at 30 Hz controlled by Matlab (Mathworks), using Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Stimuli had a constant mean intensity of $10mW/m^2$. Images were presented in a 50 x 50 grid

with a square size of 25 μm at a frame rate of 100 Hz. Static natural jittered scenes consisted of images drawn from a natural image database [32] and drifted in two dimensions with the approximate statistics of fixational eye movements [17]. The image also changed to a different location every one second, representing a saccade-like transition. Natural movies consisted of fish swimming in an aquarium, and contained both drift and saccade-like transitions that matched static jittered natural scenes. For analysis of model responses to artificial stimuli (Fig. 3), unless otherwise stated stimuli were chosen to match published values for each phenomenon.

Electrophysiology. Retinal ganglion cells of larval tiger salamanders of either sex were recorded using an array of 60 electrodes (Multichannel Systems) as previously described [33]. Intracellular recordings were performed using sharp as previously described [24].

Model training. We trained convolutional neural network models to predict retinal ganglion cell responses to either a white noise or natural scenes stimulus, simultaneously for all cells in the recorded population of a given retina [34]. Model parameters were optimized to minimize a loss function corresponding to the negative log-likelihood under Poisson spike generation,

$$L(y_t, \hat{y}_t) = \frac{1}{T}\sum_{t=0}^{T} \hat{y}_t - y_t \log \hat{y}_t,$$

where y_t and $\hat{y}_t$ are the actual and predicted firing rates of the retinal ganglion cells at time *t*, respectively with a batch size of *T* , chosen to be 50 s. To help with model fitting, we smoothed retinal ganglion responses during training with a 10 ms standard deviation Gaussian, the size of a single time bin in our model.

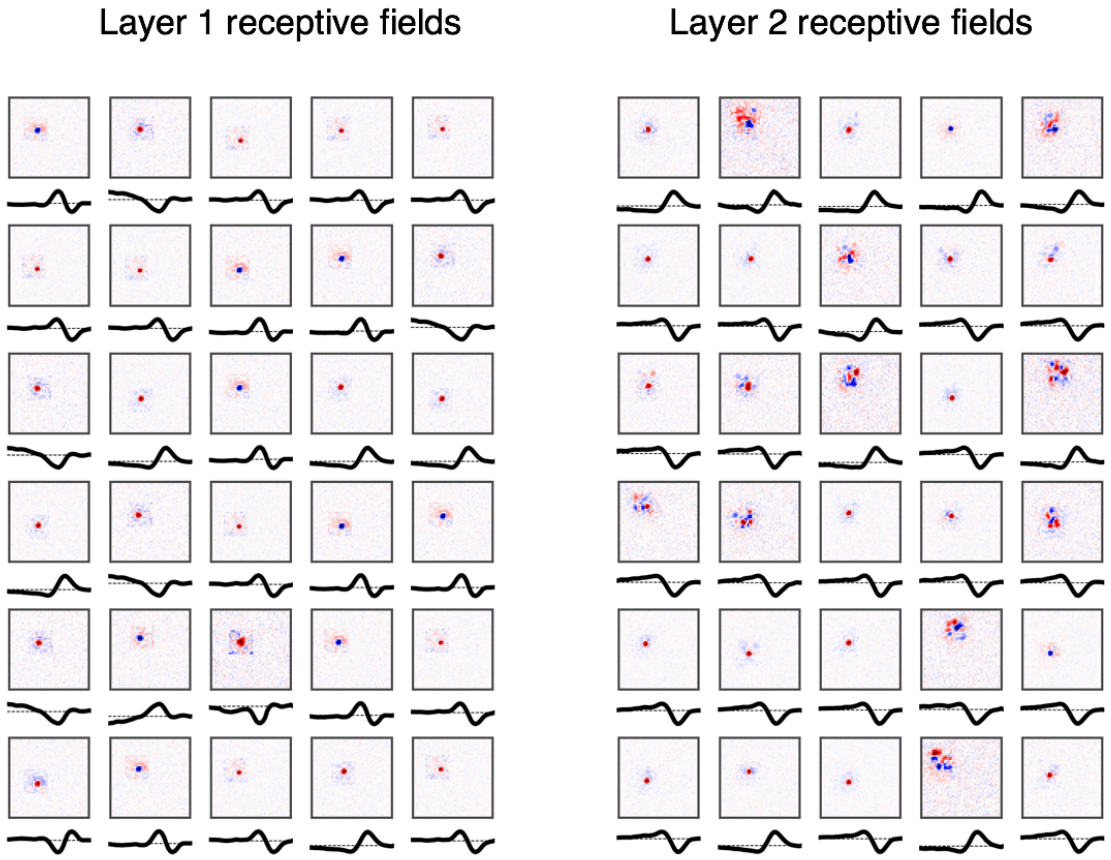
The architecture of the convolutional neural network model consisted of three layers, with 8 cell types (or channels, in the language of neural networks) per layer. Each layer consisted of a linear spatiotemporal filter, followed by a rectification using a rectified linear unit (ReLU). For each unit, an additional parameter scaled the activation of the model unit prior to the rectified nonlinearity. This scaling parameter could vary independently with location.

Optimization was performed using Adam [35], a variant of stochastic gradient descent. Models were trained using TensorFlow[36] on NVIDIA Titan X GPUs. Training an individual model to convergence required ~8 hours on a single GPU. The networks were regularized with an L2 weight penalty at each layer and an L1 activity penalty at the final layer, which helped maintain a baseline firing rate near 0 Hz.

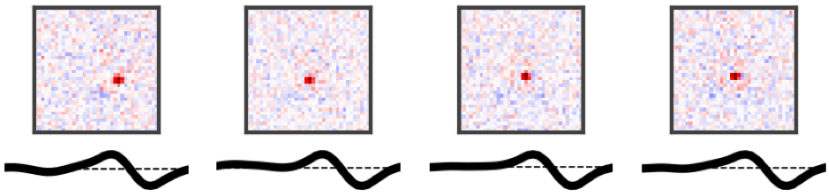
We split our dataset into training, validation, and test sets, and chose the number of layers, number of filters per layer, the type of layer (convolutional or fully connected), size of filters, regularization hyperparameters, and learning rate based on performance on the validation set. We found that increasing the number of layers beyond three did not improve performance, and we settled on eight filter types in both the first and second layers, with filters that were much larger (Layer 1,15 x 15 and Layer 2, 11 x 11) compared to traditional deep learning networks used for image classification (usually 5 x 5 or smaller). Values quoted are mean $\pm$ s.e.m. unless otherwise stated.

Linear-Nonlinear Models. Linear-nonlinear models were fit by the standard method of reverse correlation to a white noise stimulus [18]. We found that these were highly susceptible to overfitting the training dataset, and imposed an additional regularization procedure of zeroing out the stimulus outside of a 500 µm window centered on the cell's receptive field.

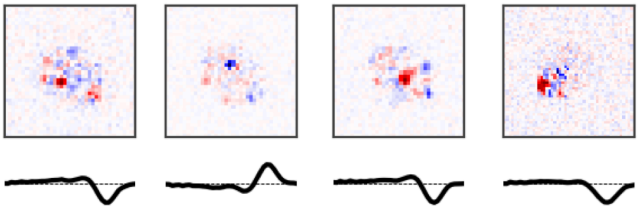
Generalized Linear Models. Generalized linear models (GLMs) were fit by minimizing the same objective as used for the CNN, the Poisson log-likelihood of data under the model. We performed the same cutout regularization procedure of only keeping the stimulus within a 500 µm region around the receptive field (this was critical for performance). The GLMs differed from the linear-nonlinear models in that they have an additional spike history feedback term used to predict the cell's response (Pillow et. al. 2008). Instead of the standard exponential nonlinearity, we found that using soft rectified functions log(1+exp(x)) gave better performance.



Extended Data Figure 1. Example spatiotemporal receptive fields of Model Units. Shown are example receptive fields of model units that contributed most strongly in each layer to the responses of the recorded ganglion cell population. The strength of the contribution for each unit in each location was computed as the gradient of the model output with respect to each cell type in each location.



Extended Data Figure 2. Variation of model units within a single cell type. Shown are receptive fields in layer 2 in different locations for the same cell type in a single model, all of which were in the thirty units that contributed most strongly to the recorded ganglion cells.



Extended Data Figure 3. Unconstrained model units that did not contribute strongly to model output. Example receptive fields in layer 2 that generated a contribution that was among the weakest to model output, as assessed by computing the gradient of the model's output with respect to the unit.

Acknowledgments. The authors wish to acknowledge William Newsome, Jennifer Raymond, Tom Clandinin, Kwabena Boahen and Leonidas Guibas for helpful discussions. This work was supported by grants from the NEI, Pew Charitable Trusts, McKnight Endowment Fund for Neuroscience and the Ziegler Foundation, (SAB); Burroughs Wellcome, McKnight, James S. McDonnnell, Simons Foundations, and the Office of Naval Research (SG), an NSF fellowship (NM) an NRSA (LTM), by the Stanford Medical Scientist Training Program (DBK) and an NSF IGERT graduate fellowship (DBK).

1. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521,** 436–444 (2015).
2. Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503,** 78–84 (2013).
3. Sussillo, D., Churchland, M. M., Kaufman, M. T. & Shenoy, K. V. A neural network that finds a naturalistic solution for the production of muscle activity. *Nat. Neurosci.* **18,** 1025–1033 (2015).
4. Yamins, D. L. K. *et al.* Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* **111,** 8619–8624 (2014).
5. Yamins, D. L. K. & DiCarlo, J. J. Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* **19,** 356–365 (2016).
6. Kell, A. J. E., Yamins, D. L. K., Shook, E. N., Norman-Haignere, S. V. & McDermott, J. H. A Task-Optimized Neural Network Replicates Human Auditory Behavior, Predicts Brain Responses, and Reveals a Cortical Processing Hierarchy. *Neuron* **98,** 630–644.e16 (2018).
7. Smirnakis, S. M., Berry, M. J., Warland, D. K., Bialek, W. & Meister, M. Adaptation of retinal processing to image contrast and spatial scale. *Nature* **386,** 69–73 (1997).

8. Baccus, S. A. & Meister, M. Fast and slow contrast adaptation in retinal circuitry. *Neuron* **36,** 909–919 (2002).
9. Kim, K. J. & Rieke, F. Temporal contrast adaptation in the input and output signals of salamander retinal ganglion cells. *J. Neurosci.* **21,** 287–299 (2001).
10. Gollisch, T. & Meister, M. Rapid neural coding in the retina with relative spike latencies. *Science* **319,** 1108–1111 (2008).
11. Schwartz, G., Taylor, S., Fisher, C., Harris, R. & Berry, M. J. Synchronized firing among retinal ganglion cells signals motion reversal. *Neuron* **55,** 958–969 (2007).
12. Berry, M. J., Brivanlou, I. H., Jordan, T. A. & Meister, M. Anticipation of moving stimuli by the retina. *Nature* **398,** 334–338 (1999).
13. Schwartz, G., Harris, R., Shrom, D. & Berry, M. J. Detection and prediction of periodic patterns by the retina. *Nat. Neurosci.* **10,** 552–554 (2007).
14. Hochstein, S. & Shapley, R. M. Linear and nonlinear spatial subunits in Y cat retinal ganglion cells. *J. Physiol. (Lond.)* **262,** 265–284 (1976).
15. Geffen, M. N., de Vries, S. E. J. & Meister, M. Retinal ganglion cells can rapidly change polarity from Off to On. *PLoS Biol.* **5,** e65 (2007).
16. Martinez-Conde, S. & Macknik, S. L. Fixational eye movements across vertebrates: comparative dynamics, physiology, and perception. *J Vis* **8,** 28.1–16 (2008).
17. Olveczky, B. P., Baccus, S. A. & Meister, M. Segregation of object and background motion in the retina. *Nature* **423,** 401–408 (2003).
18. Chichilnisky, E. J. A simple white noise analysis of neuronal light responses. *Network* **12,** 199–213 (2001).
19. Pillow, J. W. *et al.* Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* **454,** 995–999 (2008).
20. Diamond, J. S. Inhibitory Interneurons in the Retina: Types, Circuitry, and Function. *Annu Rev Vis Sci* **3,** 1–24 (2017).
21. Kaneko, A. Receptive field organization of bipolar and amacrine cells in the goldfish retina. *J. Physiol. (Lond.)* **235,** 133–153 (1973).
22. Zhang, A.-J. & Wu, S. M. Responses and receptive fields of amacrine cells and ganglion cells in the salamander retina. *Vision Res.* **50,** 614–622 (2010).
23. Pang, J.-J., Gao, F. & Wu, S. M. Relative contributions of bipolar cell and amacrine cell inputs to light responses of ON, OFF and ON-OFF retinal ganglion cells. *Vision Res.* **42,** 19–27 (2002).
24. Manu, M. & Baccus, S. A. Disinhibitory gating of retinal output by transmission from an amacrine cell. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 18447–18452 (2011).
25. Gauthier, J. L. *et al.* Receptive fields in primate retina are coordinated to sample visual space more uniformly. *PLoS Biol.* **7,** e1000063 (2009).
26. Liu, Y. S., Stevens, C. F. & Sharpee, T. O. Predictable irregularities in retinal receptive fields. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 16499–16504 (2009).
27. Barlow, H. B., FITZHUGH, R. & Kuffler, S. W. Change of organization in the receptive fields of the cat's retina during dark adaptation. *J. Physiol. (Lond.)* **137,** 338–354 (1957).
28. Hosoya, T., Baccus, S. A. & Meister, M. Dynamic predictive coding by the retina. *Nature* **436,** 71–77 (2005).
29. Nagel, K. I. & Doupe, A. J. Temporal processing and adaptation in the songbird auditory forebrain. *Neuron* **51,** 845–859 (2006).
30. Asari, H. & Meister, M. The projective field of retinal bipolar cells and its modulation by visual context. *Neuron* **81,** 641–652 (2014).

31. Helmstaedter, M. *et al.* Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature* **500,** 168–174 (2013).
32. Tkačik, G. *et al.* Natural Images from the Birthplace of the Human Eye. *PLoS ONE* **6,** e20409 (2011).
33. Kastner, D. B. & Baccus, S. A. Coordinated dynamic encoding in the retina using opposing forms of plasticity. *Nat. Neurosci.* **14,** 1317–1322 (2011).
34. McIntosh, L. T., Maheswaranathan, N., Nayebi, A., Ganguli, S. & Baccus, S. A. Deep Learning Models of the Retinal Response to Natural Scenes. *Adv Neural Inf Process Syst* **29,** 1369–1377 (2016).
35. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *arXiv.org* **cs.LG,** (2014).
36. Abadi, M. TensorFlow: A System for Large-Scale Machine Learning. *OSDI* **16,** (2016).