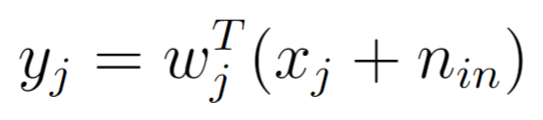
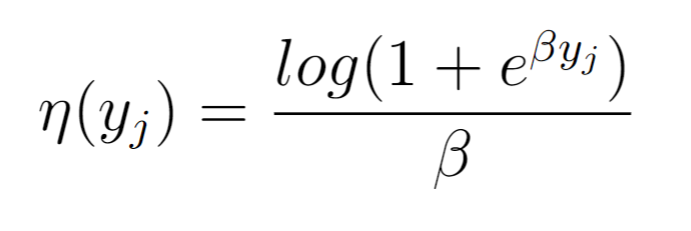
The retina plays an important role in early visual processing by separating visual information into multiple cell types that each process specific visual information. These pathways do not encode all information in visual scenes, but instead focus on encoding specific patterns of stimuli. While we know what visual information is encoded by these neurons, the physiology itself cannot tell us *why* the receptive fields of neurons are organized in this way. We want to make theories that explain retinal function, and such theories need corresponding quantitative models that can make predictions about how the retina is organized. Successful theories should be able to explain many properties of retinal ganglion cells (RGCs), such as why RGCs are separated into multiple cell types with neurons within each cell type tiling the entire retina.

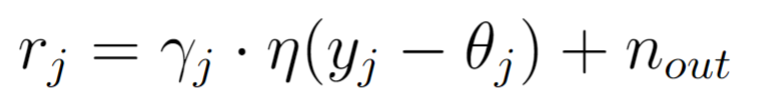
The efficient coding hypothesis is one of the most successful theories in this vein, which states that the retina should remove redundancies to maximize the amount of information transmitted with a limited number of spikes. In our model, we maximize the mutual information between the inputs X and the outputs R with a firing rate constraint. The inputs will consist of D pixels patches of Natural Images X corrupted by gaussian input noise ~ N(0, . The model RGC j takes as input a linear combination of these natural images:



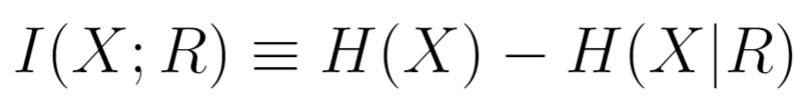
Where wj are unit-norm kernels (||wj|| = 1) and represent how much each photoreceptor contributes to the response of model RGC j. yj is then passed through a softplus nonlinearity:



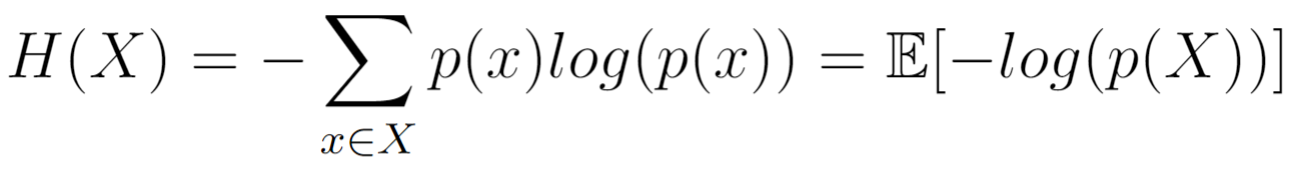
And further corrupted by output noise to produce firing rate with threshold and gain :



The above parameters are optimized via Adam to maximize the mutual information between the inputs X and the outputs R, under a firing rate constraint. Mutual information represents the amount of information, in bits, that is transmitted to RGCs from photoreceptors. Mutual information is equivalent to subtracting the entropy of the natural images X from the conditional entropy of X on R:

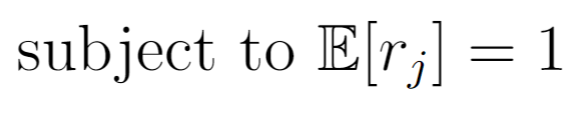


Where the entropy H(X) is defined as:



Previous work has derived a closed-form solution for the mutual information between X and R3, 14, 15. The above parameters are optimized to maximize this mutual information:





Where represents the covariance matrix of the input natural images, and are diagonal matrices that represent the covariance of the input and output noise, respectively, and W is the weight matrix. The shape of W is mxn, where m represents the number of inputs (or pixels) each model neuron receives, and n represents the number of neurons. G is a diagonal matrix that represents the local derivatives of the output responses for a specific set of input images. Since the output nonlinearity is a softmax function, the diagonal of G has approximately binary values (1 if the neuron is firing and 0 if not). This function is maximized using Adam optimization16. To represent the metabolic cost of firing spikes, each neuron is be restricted to have a fixed average firing rate.

We trained this model on randomly sampled 18x18x2 images, where 18x18 represent the spatial resolution of the image and 2 is the number of color channels. Since we had 600 model neurons, the W matrix had dimensions 648x600. The weights for each neuron converges to difference-of-gaussian functions, similarly to actual RGCs. Because difference-of-gaussians are radially symmetric, we can condense the spatial receptive fields for each color channel in each neuron into a 1D-dimensional function that represents the distance from the difference-of-gaussian center. Doing principal component analysis on these “radial receptive fields” reveals that most of the variance in the W matrix can be explained by two components (Figure 2). 90% of the variance in the W matrix lies in the first component, which adds the two colors. The second component has approximately 9% of the variance, and represents the difference between red and blue. The third component has 1% of the variance and represents the strength of the surround. These results suggest that according to efficient coding, the optimal strategy for encoding chromatic natural images is to sum the red and blue channels.