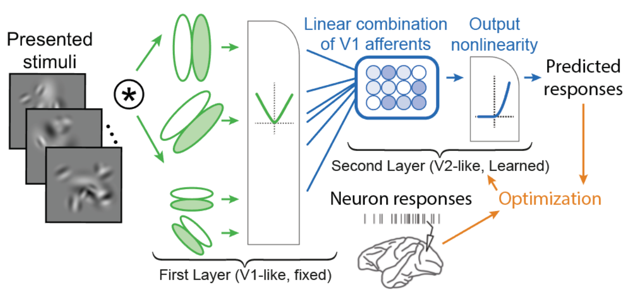
A Two-Layer Model Explains Higher-Order Feature Selectivity of V2 Neurons

Timothy D. Oleskiw, Justin D. Lieber, J. Anthony Movshon & Eero P. Simoncelli

Neurons in cortical area V2 respond selectively to higher-order visual features, such as the quasi-periodic structure of natural texture. However, a functional account of how V2 neurons build selectivity for complex natural image features from their inputs – V1 neurons locally tuned for orientation and spatial frequency – remains elusive.

We made single-unit recordings in area V2 in two fixating rhesus macaques. We presented stimuli composed of multiple superimposed grating patches that localize contrast energy in space, orientation, and scale. V2 activity is modeled via a two-layer linear-nonlinear network, optimized to use a sparse combination of V1-like outputs to account for observed activity.

Analysis of model fits reveals V2 neurons to be well-matched to natural images, with units combining V1 afferent tuning dimensions to effectively capture natural scene variation. Remarkably, although the models are trained on responses to synthetic stimuli, they can predict responses to novel image classes, *i.e.* naturalistic texture, reproducing single-unit selectivity for higher-order image statistics. Thus, we demonstrate state-of-the art performance of modeling V2 selectivity, and provide a mechanistic account of single-unit tuning for higher-order natural features.



$$\arg\min_{\vec{w},\theta} \|\vec{r} - f(\mathbf{S}\vec{w};\theta)\|_2 + \lambda \frac{\|\vec{w}\|_1}{\|\vec{w}\|_2}$$

Figure 1: Model schematic. Synthetic stimuli are generated by superimposing sparsely sampled grating patches from a set of six orientations and three scales spatially distributed over hexagonal lattices. Observed neuronal activity $\vec{r}$ is modeled by a linear combination over nonlinear V1-like afferent activity ($\mathbf{S}$, the matrix of rectified stimulus responses of a steerable pyramid; Simoncelli & Freeman, 1995), followed by an output nonlinearity f. The connection weights $\vec{w}$ and nonlinearity parameters θ are jointly optimized via stochastic gradient descent. The objective (see equation) is comprised of a residual squared error term, and an ℓ_1/ℓ_2 regularization term that enforces sparse connections while constraining the total activity energy. The regularization constant λ is chosen via cross-validation to predict data withheld during training.

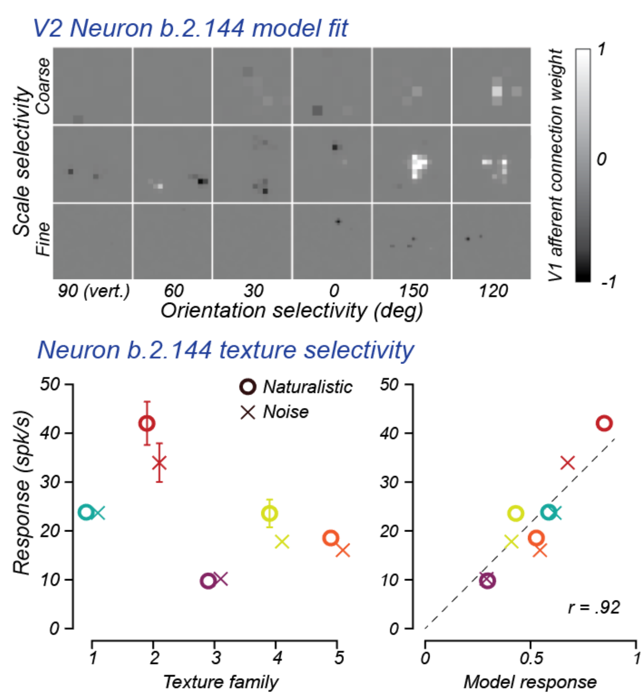


Figure 2: Explaining V2 texture selectivity. Model fits for an example neuron are visualized as the excitatory (white) and inhibitory (black) weights associated with V1 afferents; each subimage depicts weights of a spatial array of V1 afferents, tuned for a particular orientation (columns) and scale (rows). For this neuron, we also record activity to patches of visual texture. As previously reported (Freeman *et al.*, 2013), V2 neurons are sensitive to the statistics of naturalistic textures, exhibiting significant reductions in response to "noise" textures with locally-matched spectral content (*e.g.* family #2). The model, trained on sparse synthetic stimuli, accounts for responses to texture images, predicting selectivity for different texture families, and sensitivity to their higher-order statistics.