update_june4

June 4, 2020

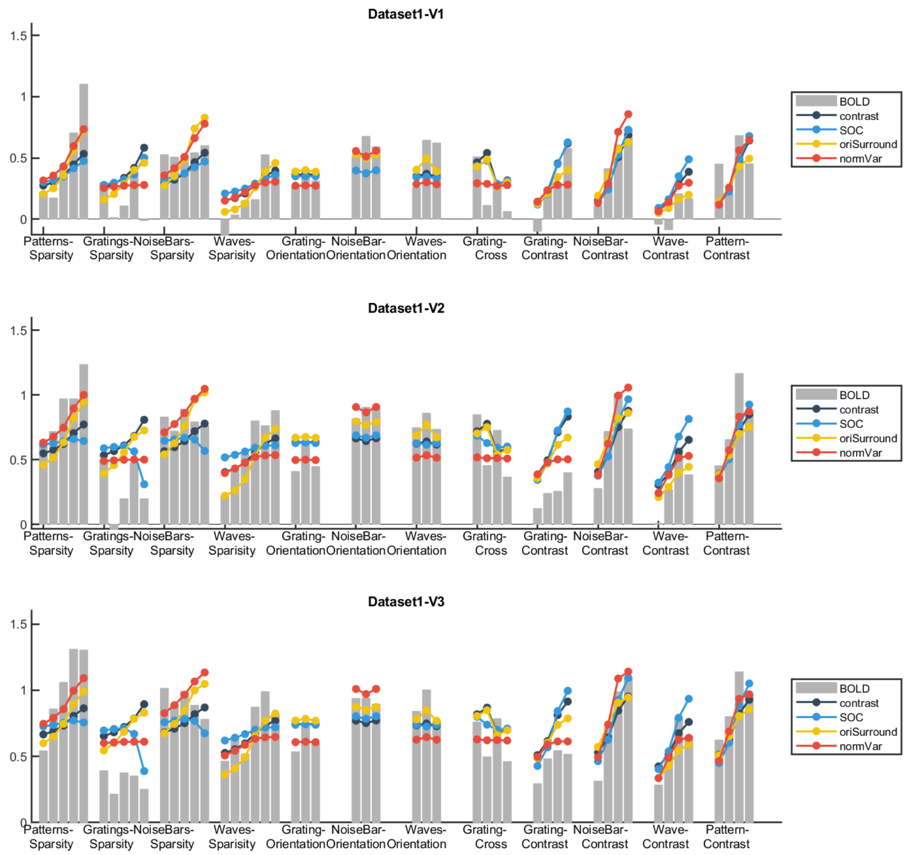
1 What's left

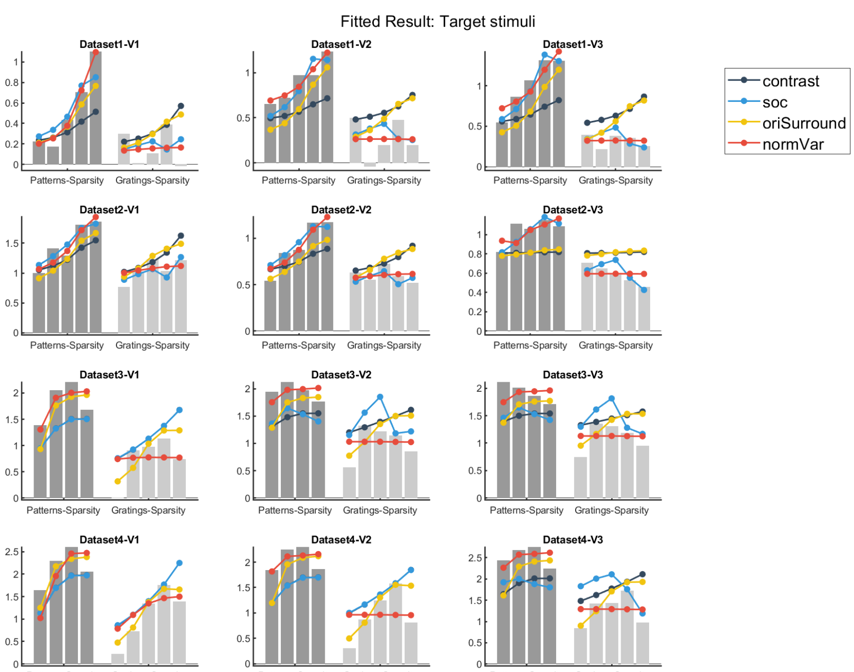
- Fix some bugs in the soc fit.
- Switch the sequence of the plot & table
- Cross_validation + boostrapping
- Any explanation of the variance

2 Tables

see google doc

2.1 Plots


Dataset1-V1
Dataset1-V2
Dataset1-V3
BOLD
contrast
SOC
oriSurround
normVar
Patterns-Sparsity
Gratings-Sparsity
NoiseBars-Sparsity
Waves-Sparsity
Grating-Orientation
NoiseBar-Orientation
Waves-Orientation
Grating-Cross
Grating-Contrast
NoiseBar-Contrast
Wave-Contrast
Pattern-Contrast


Fitted Result: Target stimuli
Dataset1-V1
Dataset1-V2
Dataset1-V3
Dataset2-V1
Dataset2-V2
Dataset2-V3
Dataset3-V1
Dataset3-V2
Dataset3-V3
Dataset4-V1
Dataset4-V2
Dataset4-V3
contrast
soc
oriSurround
normVar
Patterns-Sparsity
Gratings-Sparsity

3 One possible explanation of the normalization by variance over orientation channel

3.1 Alexnet use the channel normalization

3.3 Local Response Normalization

ReLUs have the desirable property that they do not require input normalization to prevent them from saturating. If at least some training examples produce a positive input to a ReLU, learning will happen in that neuron. However, we still find that the following local normalization scheme aids generalization. Denoting by $a^i_{x,y}$ the activity of a neuron computed by applying kernel i at position (x, y) and then applying the ReLU nonlinearity, the response-normalized activity $b^i_{x,y}$ is given by the expression

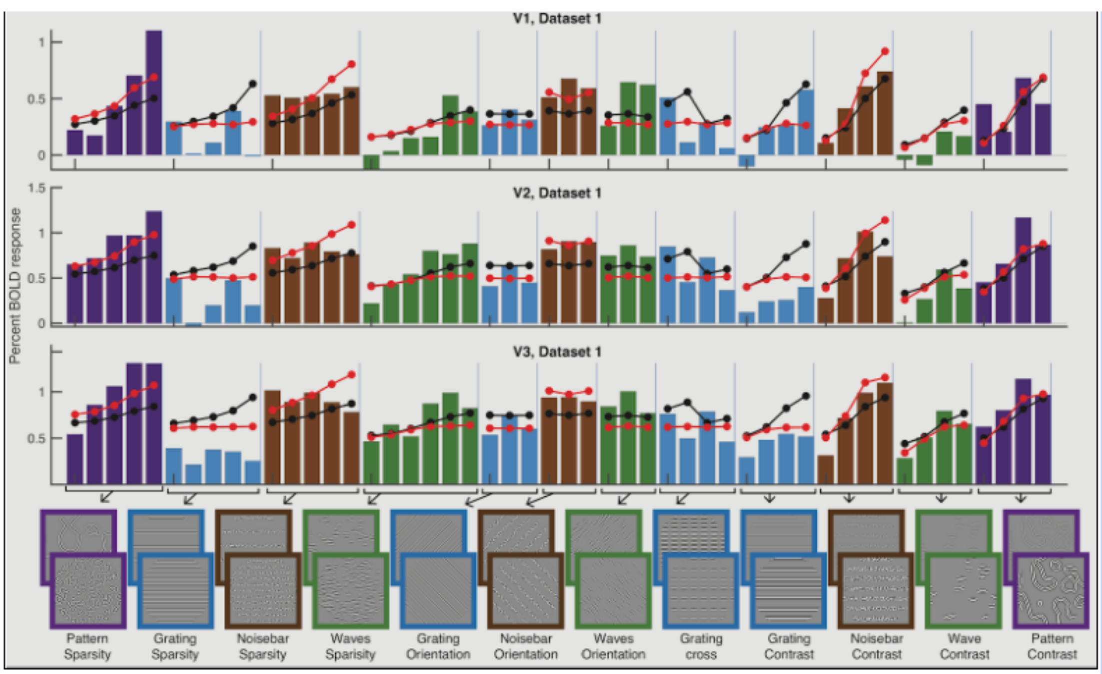
$$b^i_{x,y} = a^i_{x,y} / \left(k + \alpha \sum_{j=\max(0,i-n/2)}^{\min(N-1,i+n/2)} (a^j_{x,y})^2 \right)^\beta$$

where the sum runs over n "adjacent" kernel maps at the same spatial position, and N is the total number of kernels in the layer. The ordering of the kernel maps is of course arbitrary and determined before training begins. This sort of response normalization implements a form of lateral inhibition inspired by the type found in real neurons, creating competition for big activities amongst neuron outputs computed using different kernels. The constants k, n, α, and β are hyper-parameters whose values are determined using a validation set; we used $k = 2$, $n = 5$, $\alpha = 10^{-4}$, and $\beta = 0.75$. We applied this normalization after applying the ReLU nonlinearity in certain layers (see Section 3.5).

3.2 Linked to the shortest description length

shortest description length ≈ entropy

Less neuron participanting in grating processing (and also wave) than the other images.



[ ]: