Texture Synthesis Using Convolutional Neural Networks

Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge

Introduction

Goal of Texture Synthesis:

Infer a generating process which allows to produce arbitrary many new samples.

Related Works

Resampling pixels/whole patches

- Pros: efficiency | Cons: no actual model

Parametric texture model

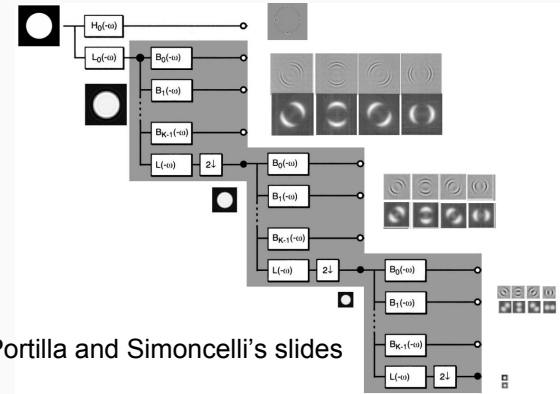
- Nth-order joint histograms of pixels
- Statistical measurements taken on the filter responses

Related Works

Portilla and Simoncelli's framework:

- Based on a set of carefully handcrafted summary statistics computed on the responses of a linear filter bank called Steerable Pyramid.

1. Pick an initial set of constraints/statistics
2. Run synthesis-by-analysis using these constraints on large dataset and find "failures"
3. Choose a new set of constraints/statistics to add that captures the most notable missing feature in failures
4. Verify this new constraint helps improve results
5. Verify the original constraints are still necessary
6. Loop back to step 2



cited from Portilla and Simoncelli's slides

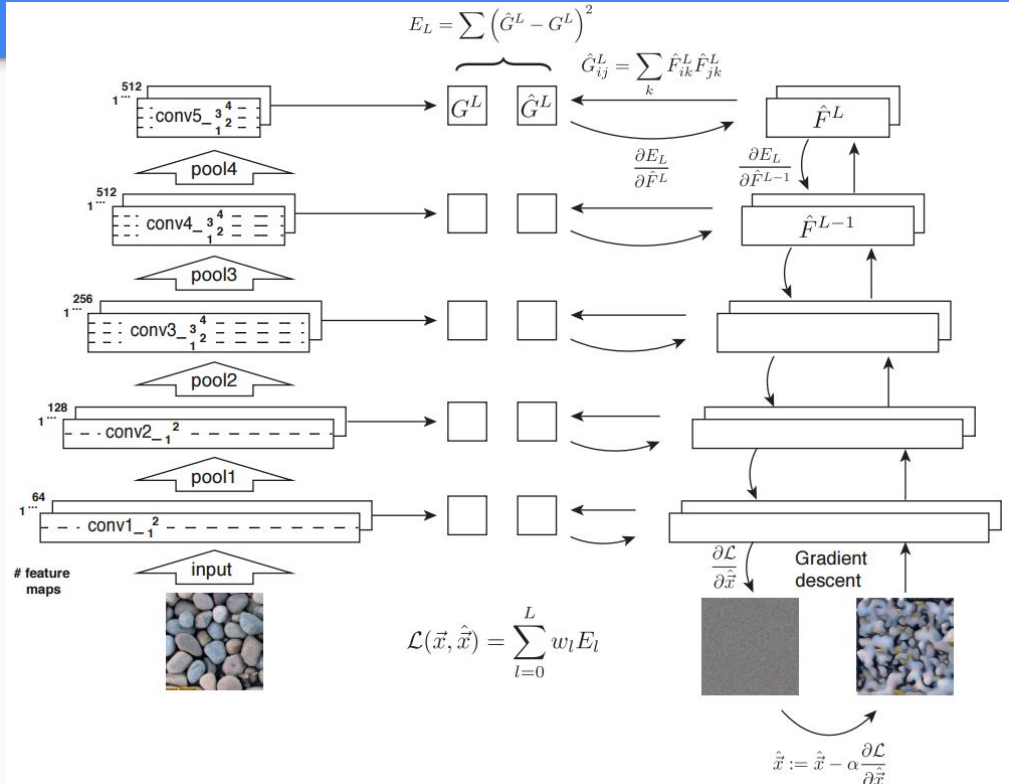
Convolutional Neural Networks

VGG-19

interleaved conv and pooling layer

conv layer: 3*3*k filters (k:# of input feature maps)

pooling: non-overlapping 2*2



Texture Model

1) Extract features of different sizes homogeneously from the image
2) Compute a spatial summary statistic on the feature responses
3) Find a new image with the same stationary description

Each layer = Non-linear filter bank

Activations in response to an image = A set of filtered images (feature maps)

Texture Model

Feature maps can be stored in a matrix: $F^l \in \Re^{N_l \times M_l}$

F^l_{jk} is the activation of the jth filter at position k in layer l.

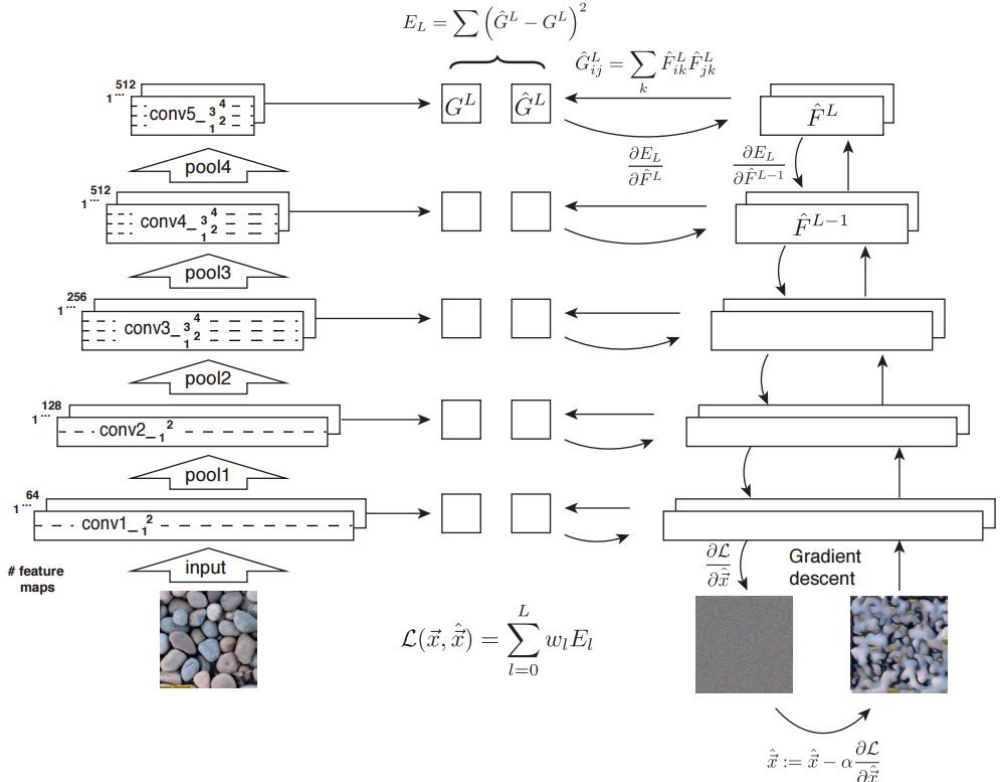
Feature correlations are presented in Gram Matrix: $G^l \in \Re^{N_l \times N_l}$

$$G^l_{ij} = \sum_k F^l_{ik} F^l_{jk}$$

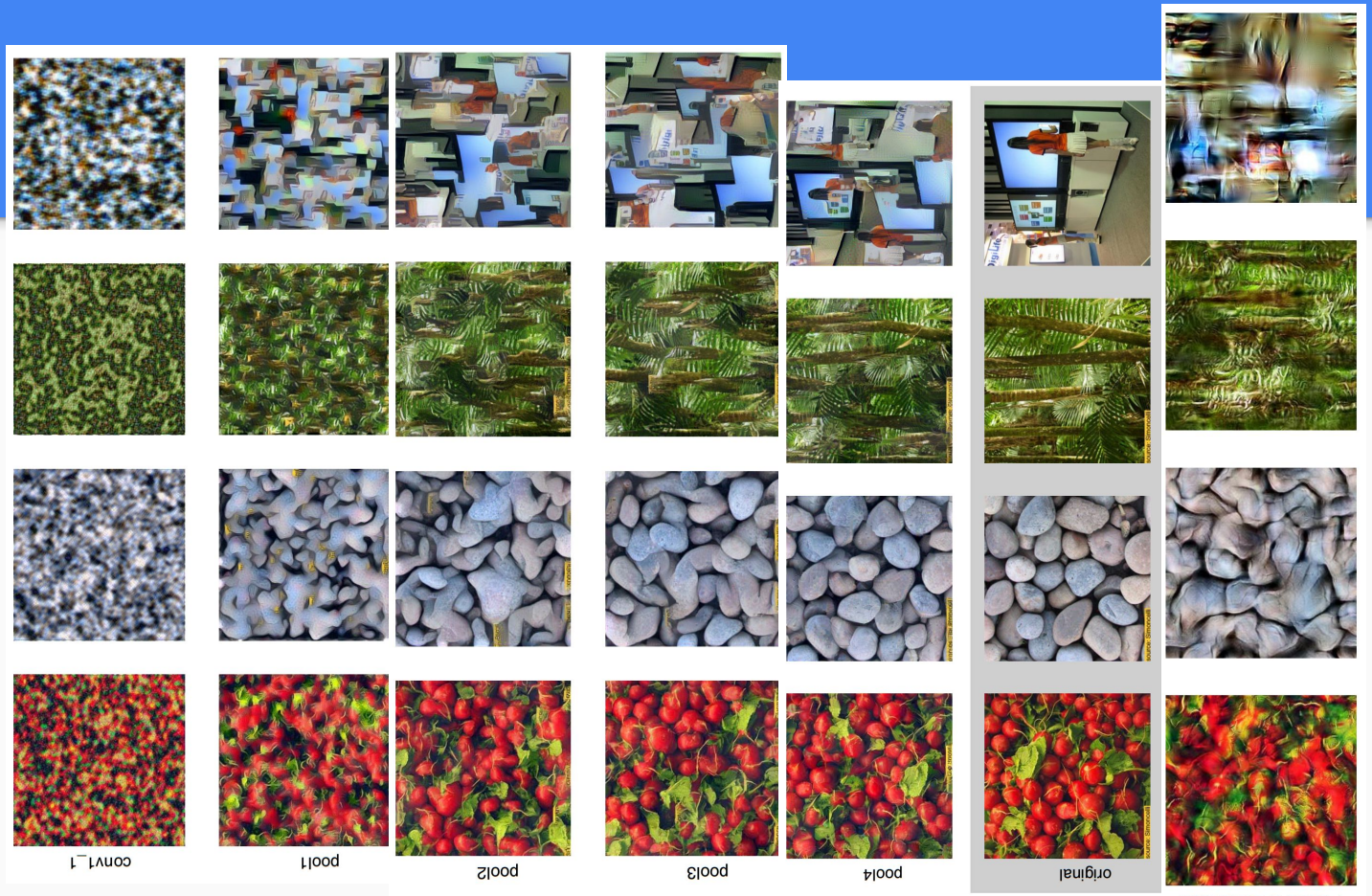
Texture Generation

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G_{ij}^l - \hat{G}_{ij}^l)^2$$

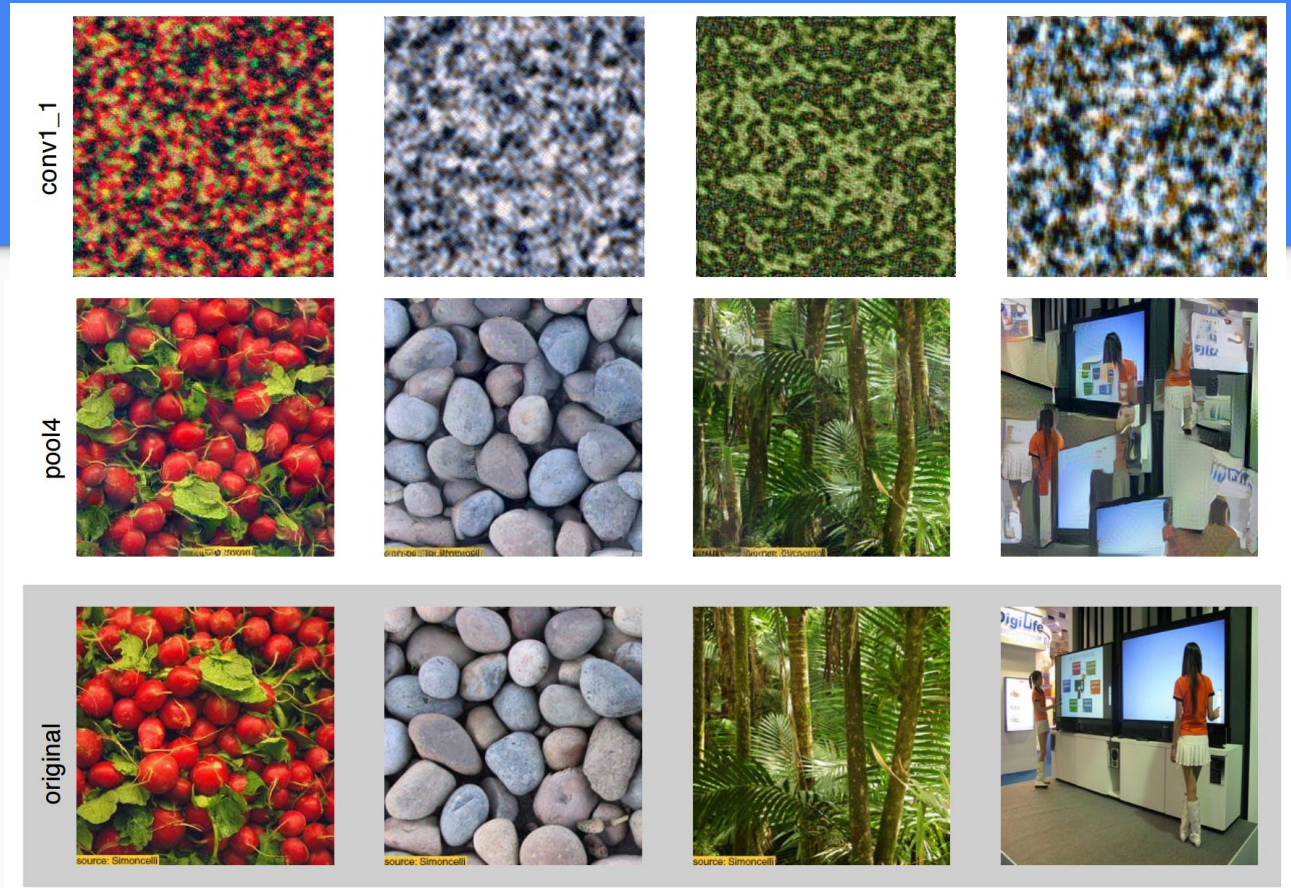
$$\frac{\partial E_l}{\partial \hat{F}_{ij}^l} = \begin{cases} \frac{1}{N_l^2 M_l^2}((\hat{F}^l)^T (G^l - \hat{G}^l)) & \text{if } \hat{F}_{ij}^l > 0 \\ 0 & \text{if } \hat{F}_{ij}^l < 0 \end{cases}$$



Results


conv1_1
pool1
pool2
pool3
pool4
original
Portilla & Simoncelli

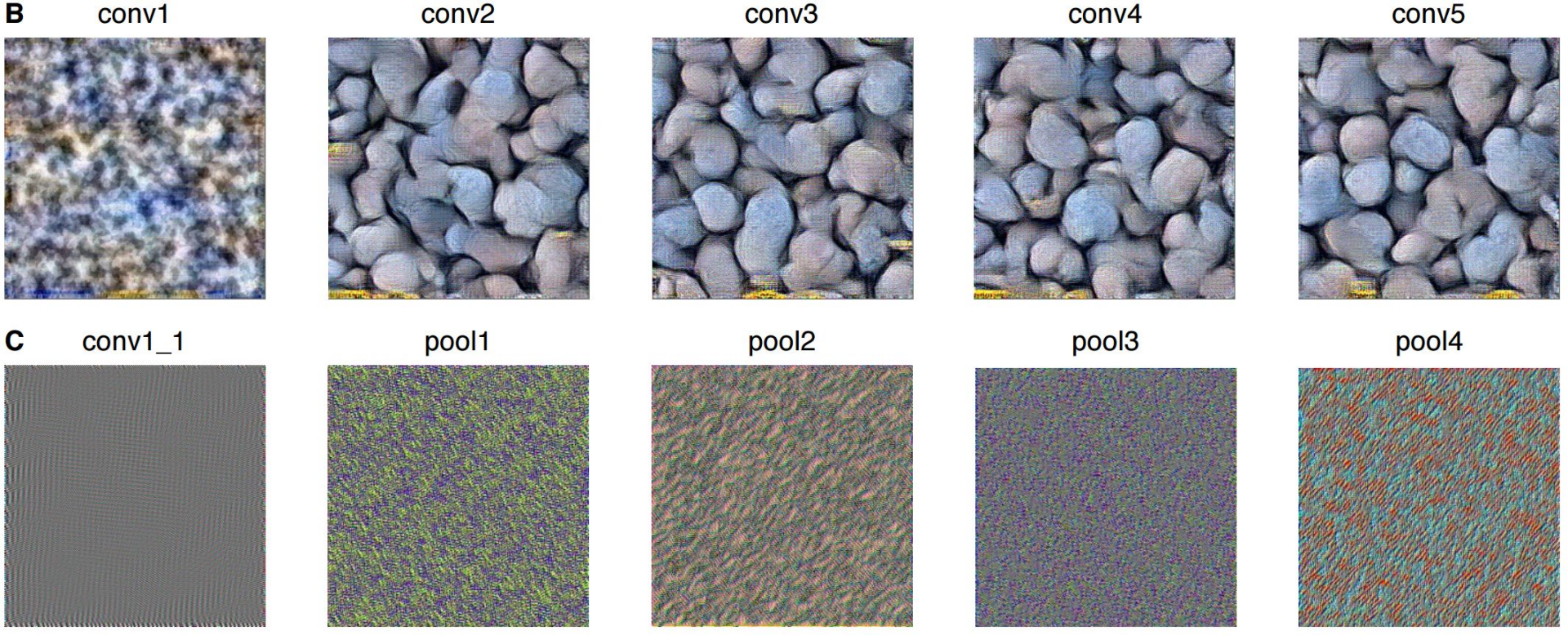
Results


conv1_1
pool4
original
source: Simoncelli
source: Simoncelli
source: Simoncelli
DigiLife

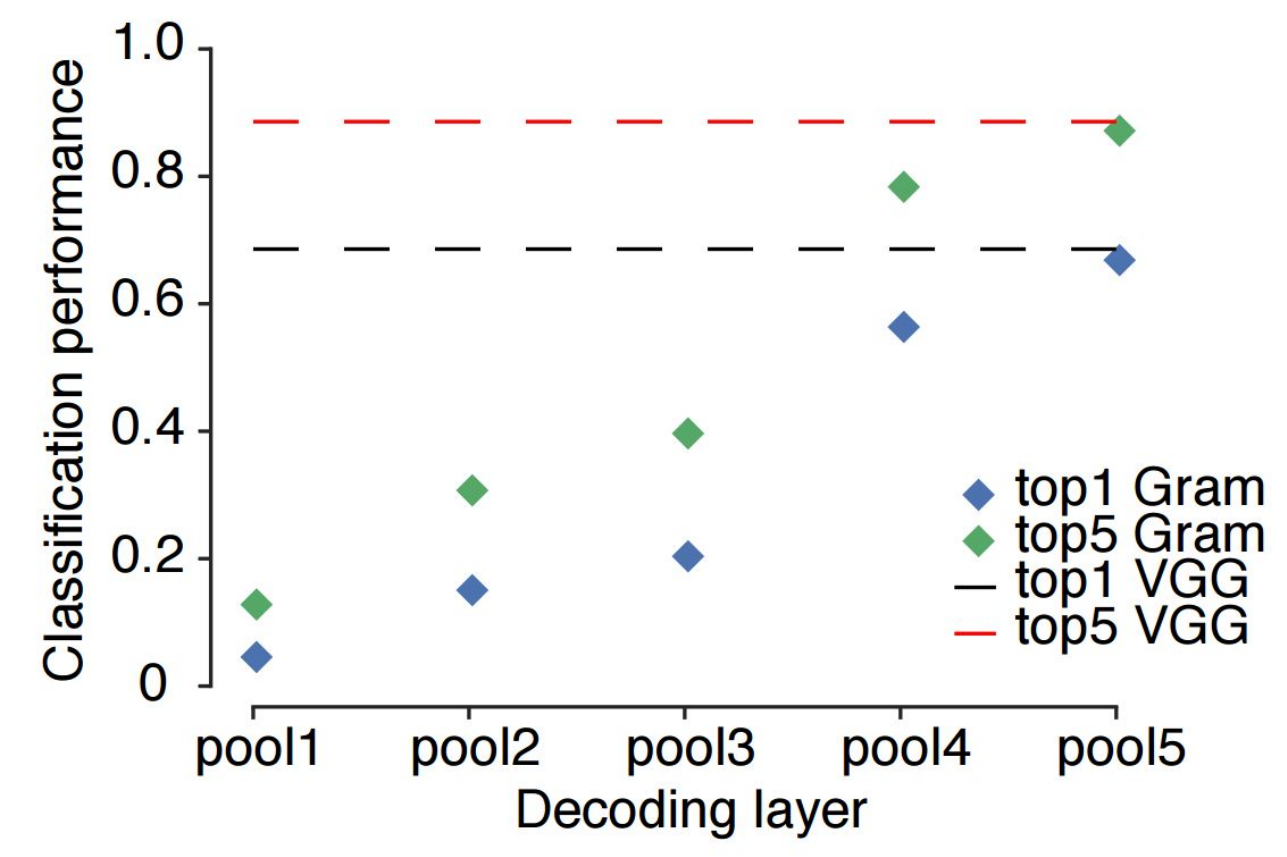
Results

observation: texture representation can be compressed greatly.
open question: how to find minimal set of parameters that reproduces the quality of the full model?

Results


B
conv1
conv2
conv3
conv4
conv5
C
conv1_1
pool1
pool2
pool3
pool4

Results


1.0
0.8
0.6
0.4
0.2
0
Classification performance
pool1
pool2
pool3
pool4
pool5
Decoding layer
top1 Gram
top5 Gram
top1 VGG
top5 VGG

Conclusion

The new parametric texture model based on CNN achieves a substantial improvement on the quality of the texture synthesised.

Can leverage on the improvements in general CNN

Compelling models for studying visual info processing in the brain