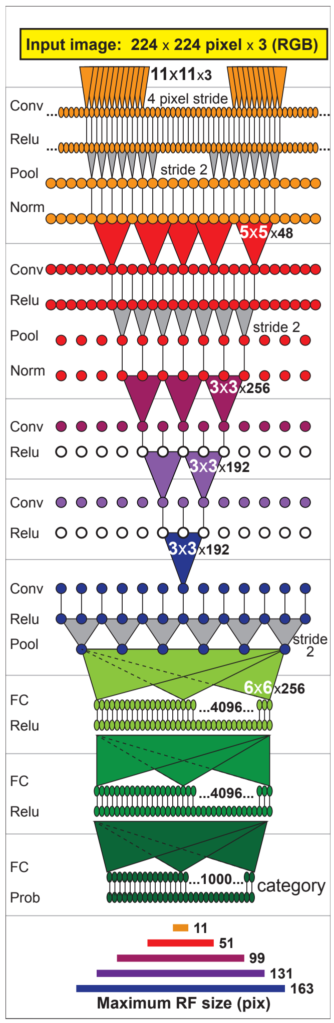
Title: Evaluating and Interpreting a Convolutional Neural Net as a Model of V4

Authors: Dean A. Pospisil, Anitha Pasupathy, and Wyeth Bair, Dept. of Biological Structure, Univ. of Washington

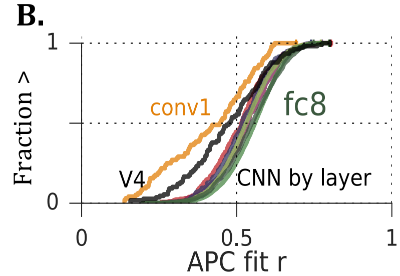
Convolutional neural nets (CNNs) are currently the highest performing image recognition computer algorithms. Of interest is whether these CNNs, following extensive supervised training, perform computations similar to those in the ventral visual stream. We investigated whether CNN units' tuning for shape boundaries was similar to V4's as described in the angular position and curvature (APC) model of Pasupathy and Connor 2001. From units in all layers of AlexNet (see Figure A), an object recognition CNN, we recorded responses to the original study's set of shape stimuli (51 simple closed shapes at up to 8 rotations) presented at 51 spatial translations (2 pixel increments). We found many units in all layers with V4-like APC shape tuning, but only the later layers had the translation invariance to deem them truly V4-like (Figure B). We then asked whether the CNN could directly predict responses of V4 neurons better than the simpler APC model (Figure D). We found that even model units in the second layer could serve as good V4 models so we have started to probe quantitatively the representation of the early layers in terms of form and chromatic representation. We have found the first layer (Figure C.) can be described with a handful of parameters (orientation, peak frequency, bandwidth) and the pattern of weights in the second layer approximate classical properties of V1 including cross-orientation suppression. We will discuss the implications of these results for mid-level visual encoding and the development of state-of-the-art image-computable models for mid-level visual representation.

A.

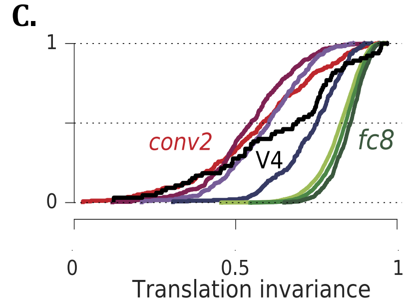
Major layer index	Sub-layer name	Number of unique units	Spatial locations (x x y)	Kernel size (x x y x depth)
1	**Conv1**	96	55 x 55	11 x 11 x 3
	Relu1	96	55 x 55	
	Pool1	96	27 x 27	3 x 3
	Norm1	96	27 x 27	
2	**Conv2**	256	27 x 27	5 x 5 x 48
	Relu2	256	27 x 27	
	Pool2	256	13 x 13	3 x 3
	Norm2	256	13 x 13	
3	**Conv3**	384	13 x 13	3 x 3 x 256
	Relu3	384	13 x 13	
4	**Conv4**	384	13 x 13	3 x 3 x 192
	Relu4	384	13 x 13	
5	**Conv5**	256	13 x 13	3 x 3 x 192
	Relu5	256	13 x 13	
	Pool5	256	6 x 6	3 x 3
6	**FC6**	4096	-	6 x 6 x 256
	Relu6	4096	-	
7	**FC7**	4096	-	4096
	Relu7	4096	-	
8	**FC8**	1000	-	4096
	Prob	1000	-	
Total:	**21** sub-layers	**22,096** unique units	**1,553,986** total units	**60,954,656** unique weights; **724,406,816** total weights



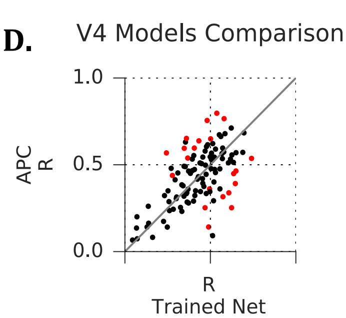
B.



C.



D.



E.

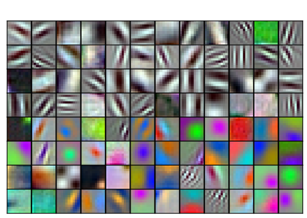


Figure. **(A)** Layer names, number of unique units, spatial locations and kernel sizes in Caffe's AlexNet. In addition to the weights (right column) associated with each kernel, there is also one bias value per kernel, which adds 10,568 free parameters to the ~60.9 million unique weights. Pooling has a 3 x 3 spatial region (faded values in right column), but no free parameters. **(B)** Cumulative distribution by layer of CNN (coloring matches table) and V4 (black) fit to APC model showing a mild increase over layers. **(C)** Translation invariance shows a marked difference between layers. **(D)** First layer filters of AlexNet show clear orientation, frequency, and chromatic preferences. **(E)** APC model and trained net perform similarly in predict many units (black) but for some (red) significantly diverge.

<u>References:</u>

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks." NIPS.

Pasupathy, A., and C. E. Connor. "Shape Representation in Area V4: Position-Specific Tuning for Boundary Conformation." Journal of Neurophysiology

<u>Funding</u>

National Science Foundation (NSF) Graduate Research Fellowship DGE-1256082 (D.A.P.).

NSF Collaborative Research in Computational Neuroscience Grant IIS-1309725 (W.B. and A.P.).

Google Faculty Research Award (W.B. and A.P.)