Evidence for Configural Superiority Effects in Convolutional Neural Networks

Shaiyan Keshvari & Ruth Rosenholtz

INTRODUCTION

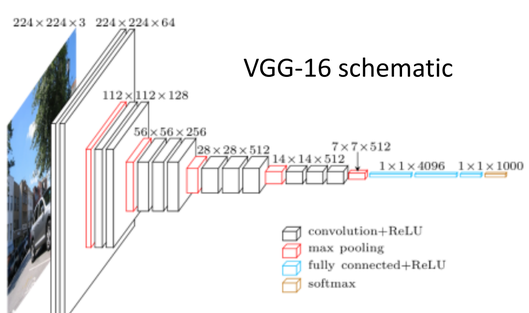
- **Configural superiority effect (CSE)** – combinations of parts are perceived more quickly and accurately than the parts alone[1,2]
- CSEs thought to be driven by **"emergent" feature (EF)** differences between target and distractors[1]
- EFs may result from the visual system learning abstract representations to support complex tasks, like object recognition, at the expense of simpler but less ecologically relevant tasks
- **Convolutional Neural Nets (CNNs)** excel at object recognition, as well as tasks for which they are not trained. Feature vectors at different layers correlate with responses of various brain areas[3]

Research question:

- Do the higher levels in a CNN show CSEs?

Approach:

- Use the VGG-16 network[4] pre-trained on ImageNet as a stand-in for the visual hierarchy
- Train a classifier to do an oddball localization task using layer activations as the input features

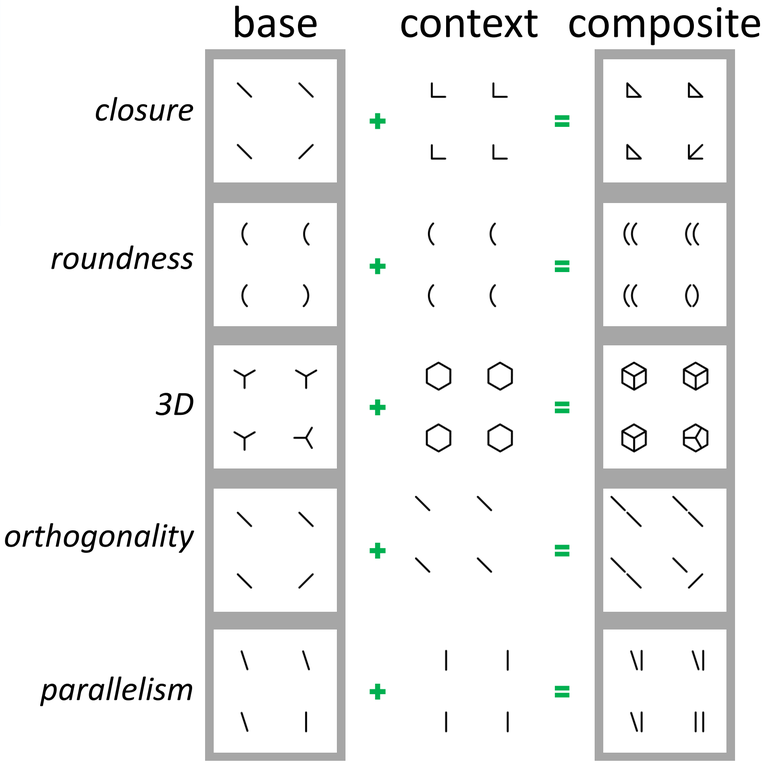


VGG-16 schematic

https://blog.heuritech.com

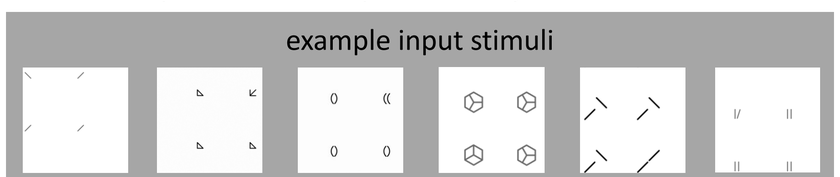
METHODS

- Base (no EF) and composite (EF) stimuli
- Noisy, translated, rotated, resized, and contrast-adjusted images to promote generalizability
- For each EF x (base, composite), trained a multi-class linear SVM on the last fully connected layer (fc7, 4096 "neurons") to locate the "odd quad"
- Compute cross-validated performance
- Also tested a network with random weights



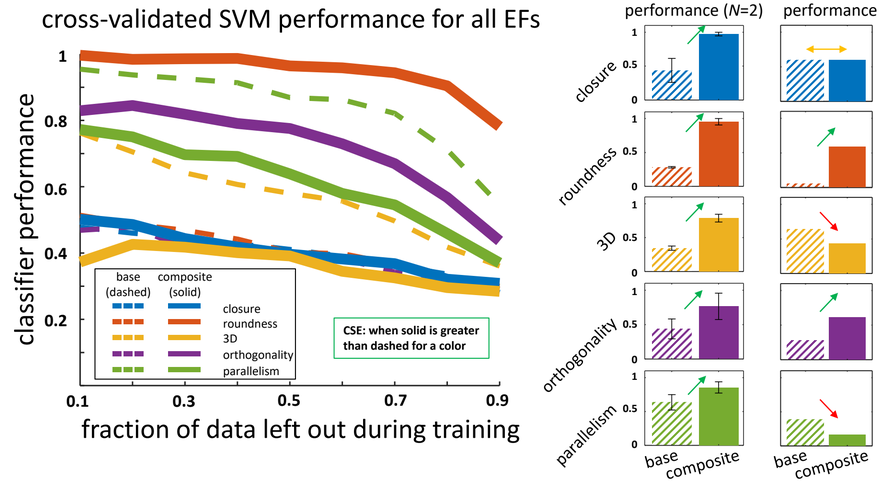
✖ noise, translations, etc. = 28,797 stimuli in total

example input stimuli



RESULTS

- Average cross-validated performance
- Composite better than base (**CSE!**):
  - *orthogonality* (Δ33 percentage points) and *roundness* (53 pp)
- No effect:
  - *closure* (< 1 pp)
- Base better than composite:
  - *parallelism* (23 pp) and *3D* (21 pp)
- Random CNN weights: no effect for any EFs
- Pilot behavioral experiment (*N*=2) confirmed CSE (44 +/- 0.06 pp)



CONCLUSIONS

- Some evidence that later CNN features compute "emergent" features, but not consistently
- Need to test other layers and networks
- Other factors like "false pop-out", may explain some effects not modeled by CNN

[1] Pomerantz, J. R., Sager, L. C., & Stoever, R. J. (1977)
[2] Pomerantz, J. R., & Cragin, A. I. (2015)
[3] Yamins D*, Hong H*, et al. (2014)
[4] Simonyan, K. A. Zisserman, K. (2014)

Special thanks to Yrvine Thelusma, Carl Vondrick, Shaoxiong Wang, and Ben Wolfe