Representation of non-local shape information in deep neural networks

Shaiyan Keshvari, Ingo Fründ, and James H. Elder

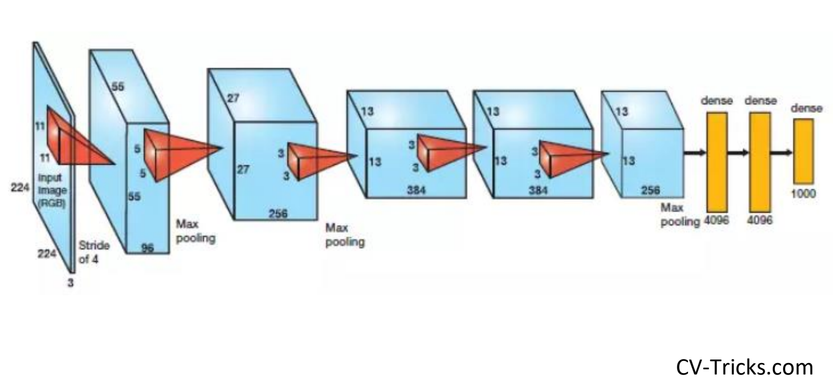
INTRODUCTION

- Neurons in primate visual areas such as V4 and IT are known to be selective for global shape.
- How explicitly do deep convolutional neural networks (DCNNs) represent global object shape?
  - Baker et al. (2018) suggest very little[1]
  - others find evidence for global shape coding[2,3]
- To resolve this debate, we need generative shape models that allow local and global cues to be precisely controlled.
- Model 1: Local shape metamers[4,5]
  - Match local curvature statistics of natural shapes
  - Global shape is random (maximum entropy)
- Model 2: Global sparse components[6]
  - Capture global shape without semantics
- These synthetic shapes can be used to systematically probe representation of local vs global shape in DNNs.

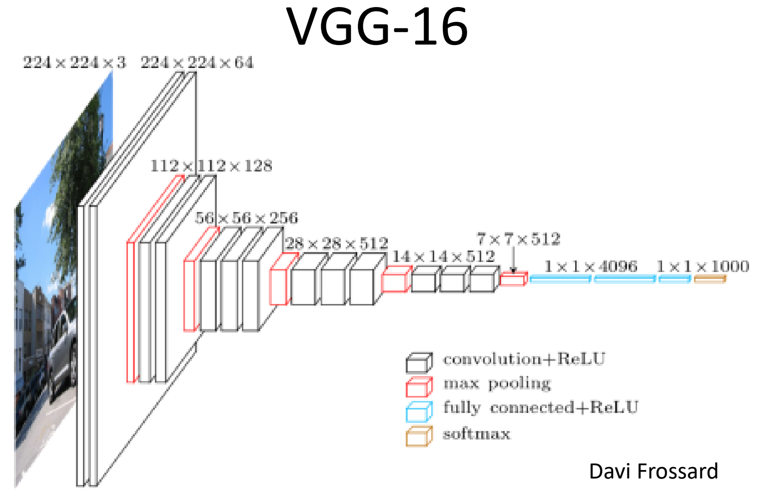
APPROACH

1. Input natural and synthetic shape silhouettes to **AlexNet and VGG-16** networks **pre-trained on ImageNet**
2. Measure linear separability of natural vs synthetic shapes at each layer of the networks

AlexNet



VGG-16



METHODS

- Natural shapes: 391 120-point polygons of animal shapes drawn from Hemera Photo-Object dataset
- Model 1: Match moments of the local curvature distribution
- Model 2: Sparse components of global shape
- Silhouettes are area-matched, resized to 224 by 224 pixels.
- Train **linear classifier** to distinguish natural from synthetic shapes based on activations at each convolutional layer.

Example animal shapes from Hemera dataset

Model 1: Example locally-matched shapes

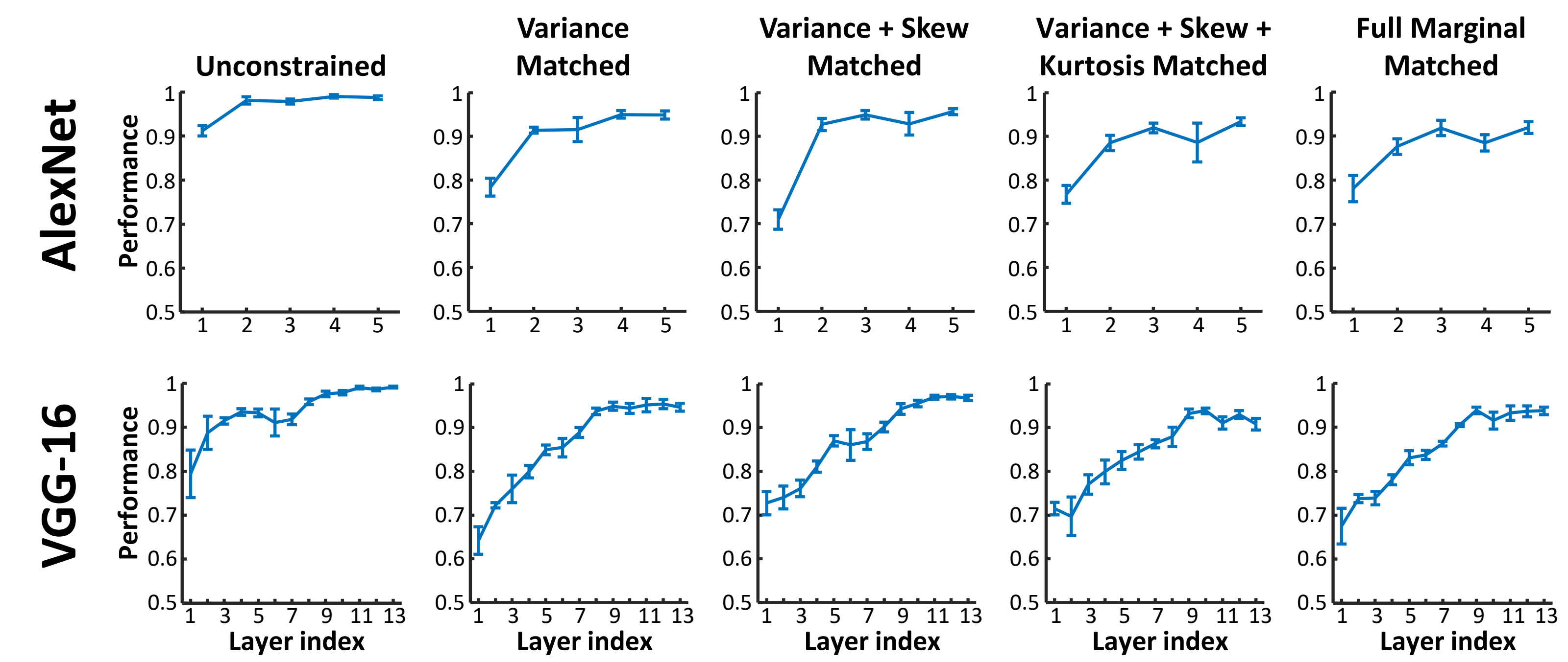
Model 2: Example sparse components of global shape

References

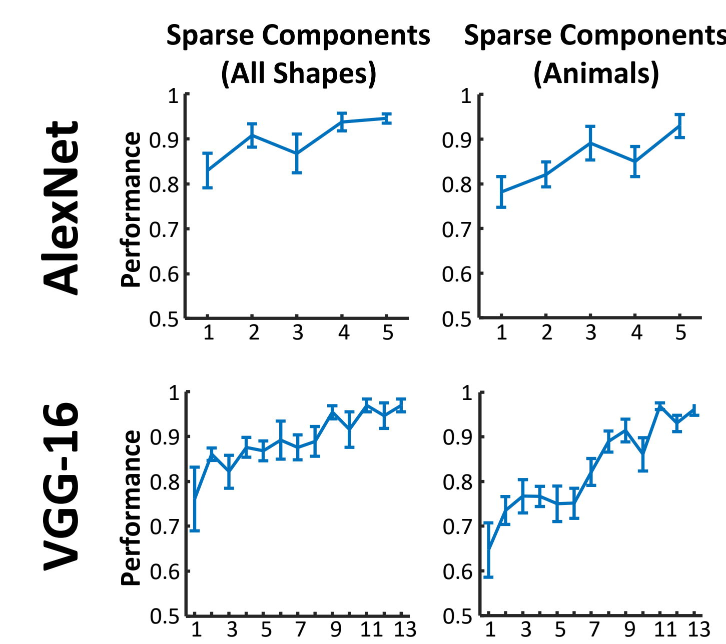
[1] Baker, N., Lu, H., Erlikhman, G., & Kellman, P.J. (2018). *PLoS computational biology 14*(12), e1006613.
[2] Kubilius, J., Bracci, S., & de Beeck, H. P. O. (2016). *PLoS computational biology 12*(4), e1004896.
[3] Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., & Brendel, W. (2019). *International Conference on Learning Representations (ICLR)* .
[4] Fründ, I. & Elder, J.H. (2015). *Computational and Systems Neuroscience (Cosyne).*
[5] Elder, J.H., Oleskiw, T.D. & Fruend, I. (2018). *Journal of Vision*, *18*(12):14.
[6] Clément & Elder, J.H. (2018). What are the sparse components of natural shapes? European Conference on Visual Perception.

RESULTS

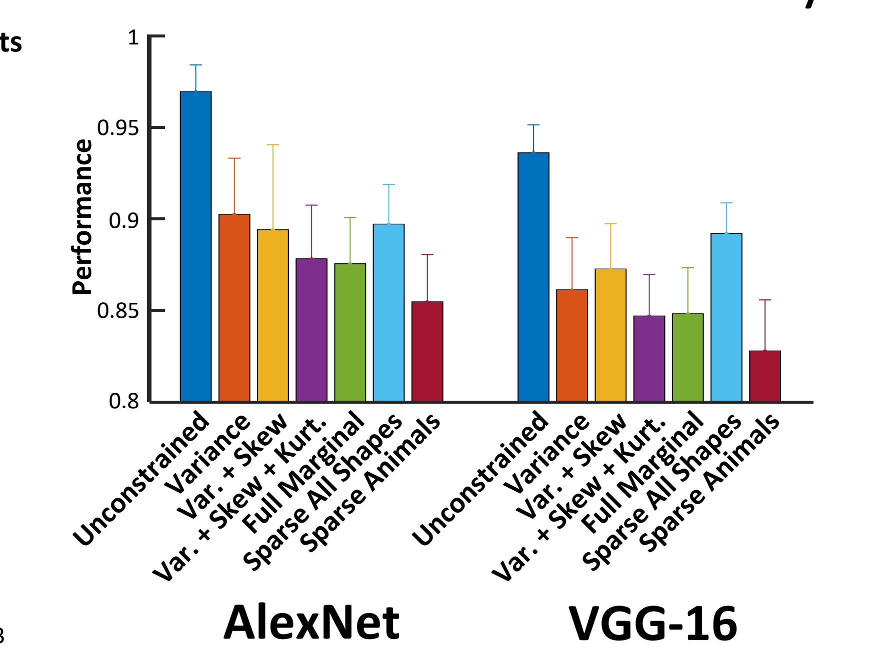
Model 1



Model 2



Mean Performance Over Layers



OBSERVATIONS

For both neural networks, we observe:

- Linear separability of shapes increasing monotonically with network depth.
- Tuning to fine details of local curvature distribution.
- Discriminative non-local shape information increasingly available in deeper layers.
- Sparse components of global shape (Model 2) more metameric than shapes matching only local curvature (Model 1).

Main Conclusion:

- Imagenet-trained DCNNs encode both local and non-local shape information.