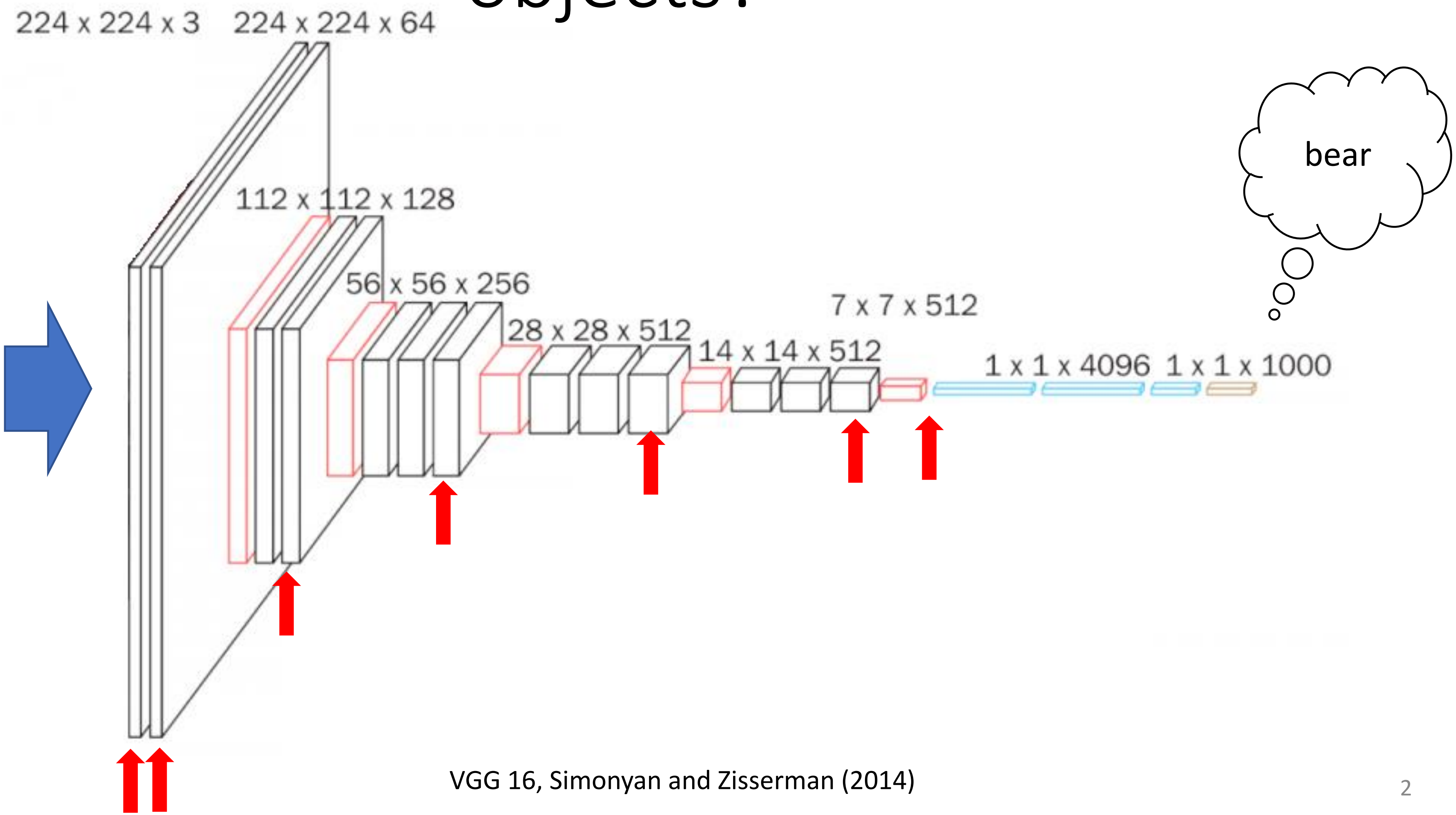
Representation of non-local shape information in deep neural networks

Shaiyan Keshvari, Ingo Fründ, & James H. Elder

ECVP 2019

What information do DCNNs use to classify objects?

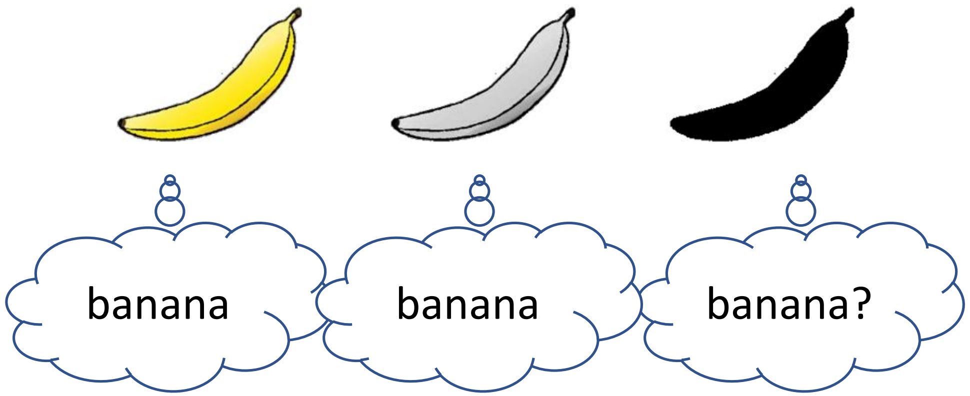


VGG 16, Simonyan and Zisserman (2014)

Talk outline

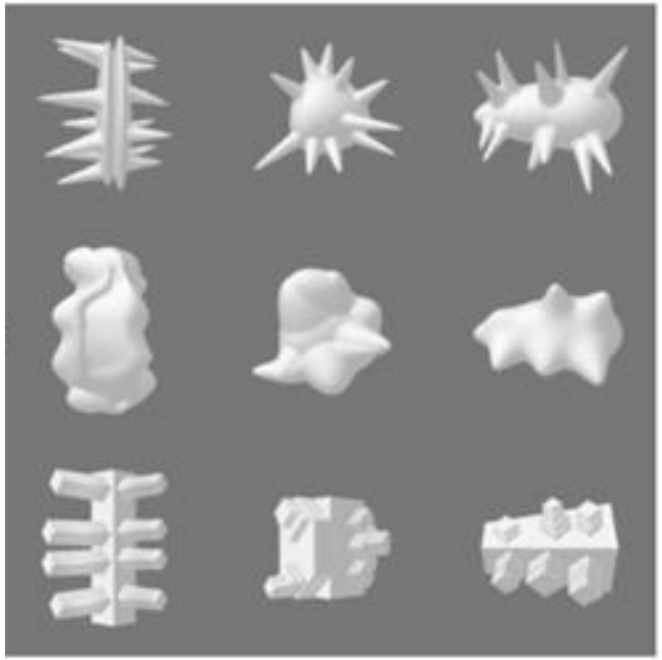
- Prior work
- Experiment 1: Network sensitivity to local shape information
- Experiment 2: A sparse components model of shape
- Conclusions

Kubilius et al. (2016): Shape representations **similar** between humans and DCNNs



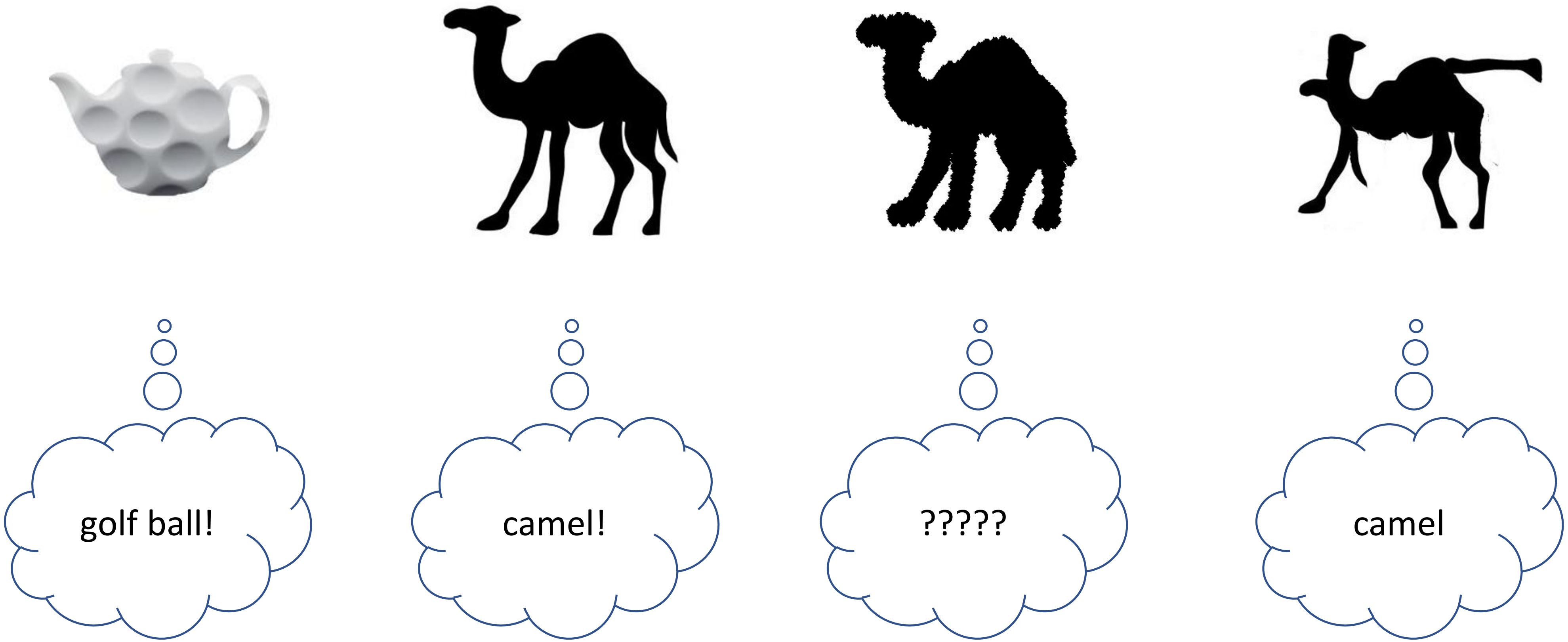
Similar perceived shape

Similar pixels



Figures from Kubilius et al. (2016)

Baker et al. (2018): DCNNs **no not** classify using global shape



Figures from Baker et al. (2018)

Summary: There is a disagreement as to what degree DCNNs are coding local vs global information

- Part of the disagreement between findings may be due to inconsistent stimuli

- Directly controlling for global vs local shape cues in the stimuli can give a clearer picture

- 1000-way object classification -> bad probe of shape sensitivity?
  - Task inherently difficult, may underestimate shape information
  - Fully connected layers may be over-fitted to ImageNet, unlike conv. layers

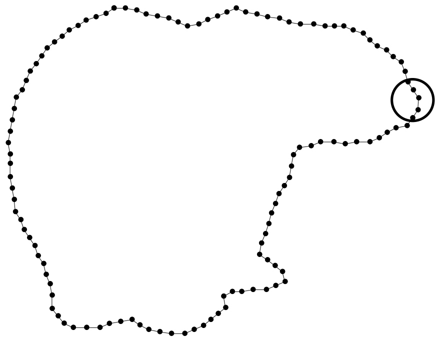
Carefully control shape information available to DCNNs

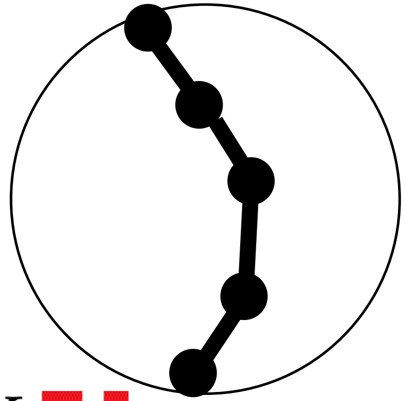
- Control the presence of local shape information using a generative model of shape

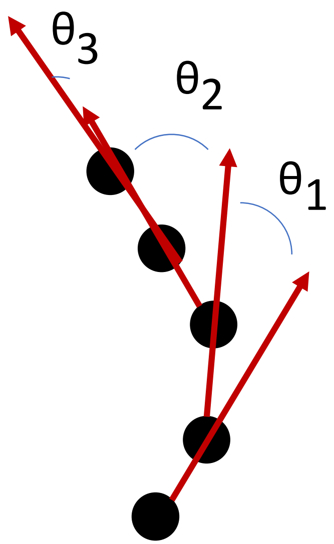
- Compare different “amounts” of local shape information

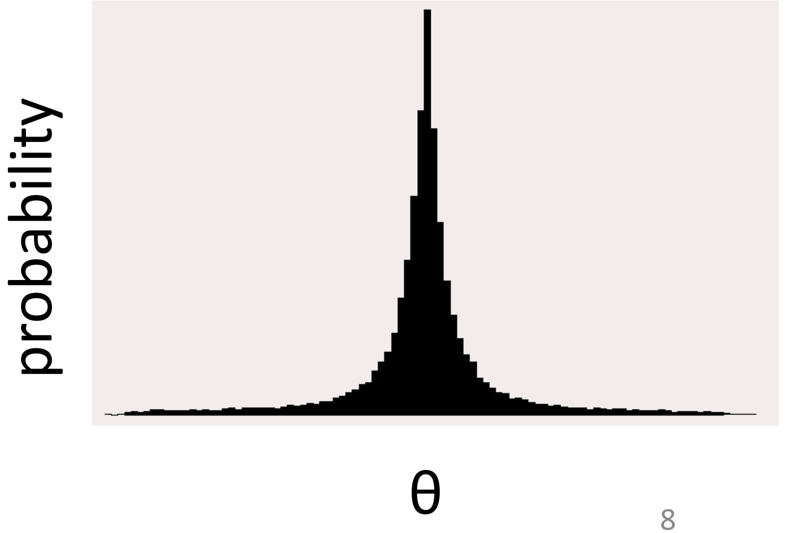
- Measure information as decodability throughout the network

Defining shape in terms of local curvature







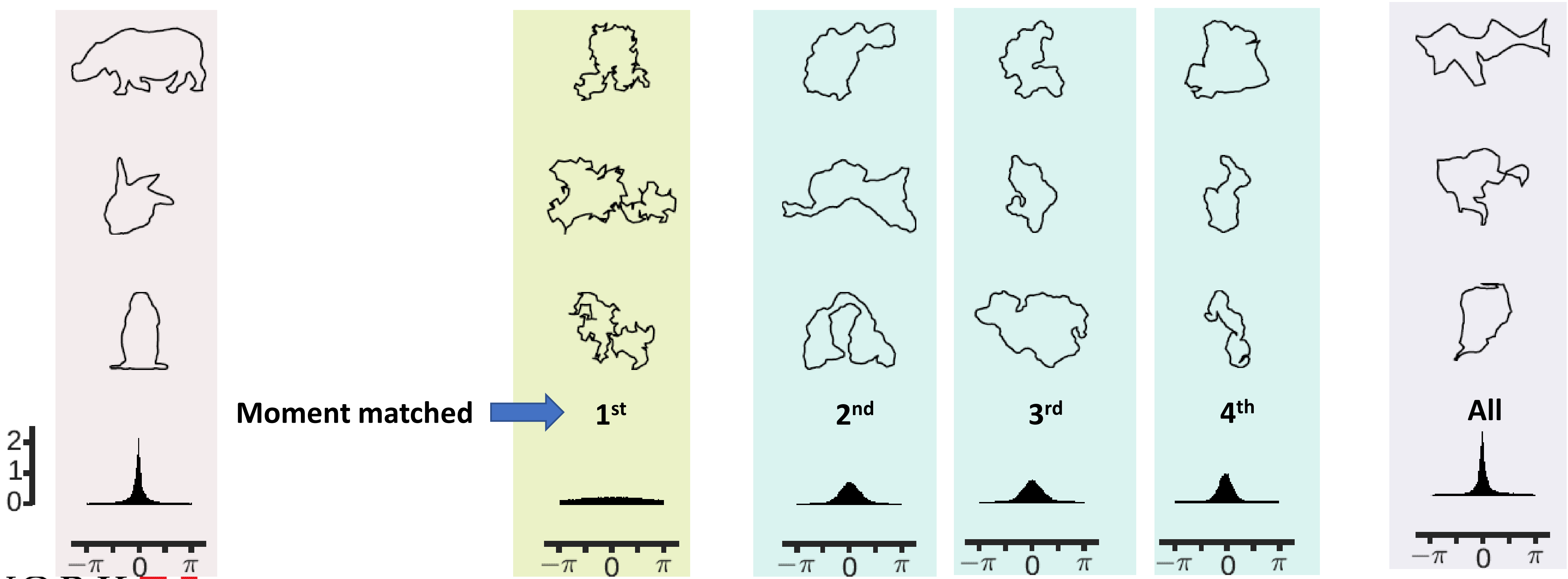


Talk outline

- Prior work
- Experiment 1: Network sensitivity to local shape information
- Experiment 2: A sparse components model of shape
- Conclusions

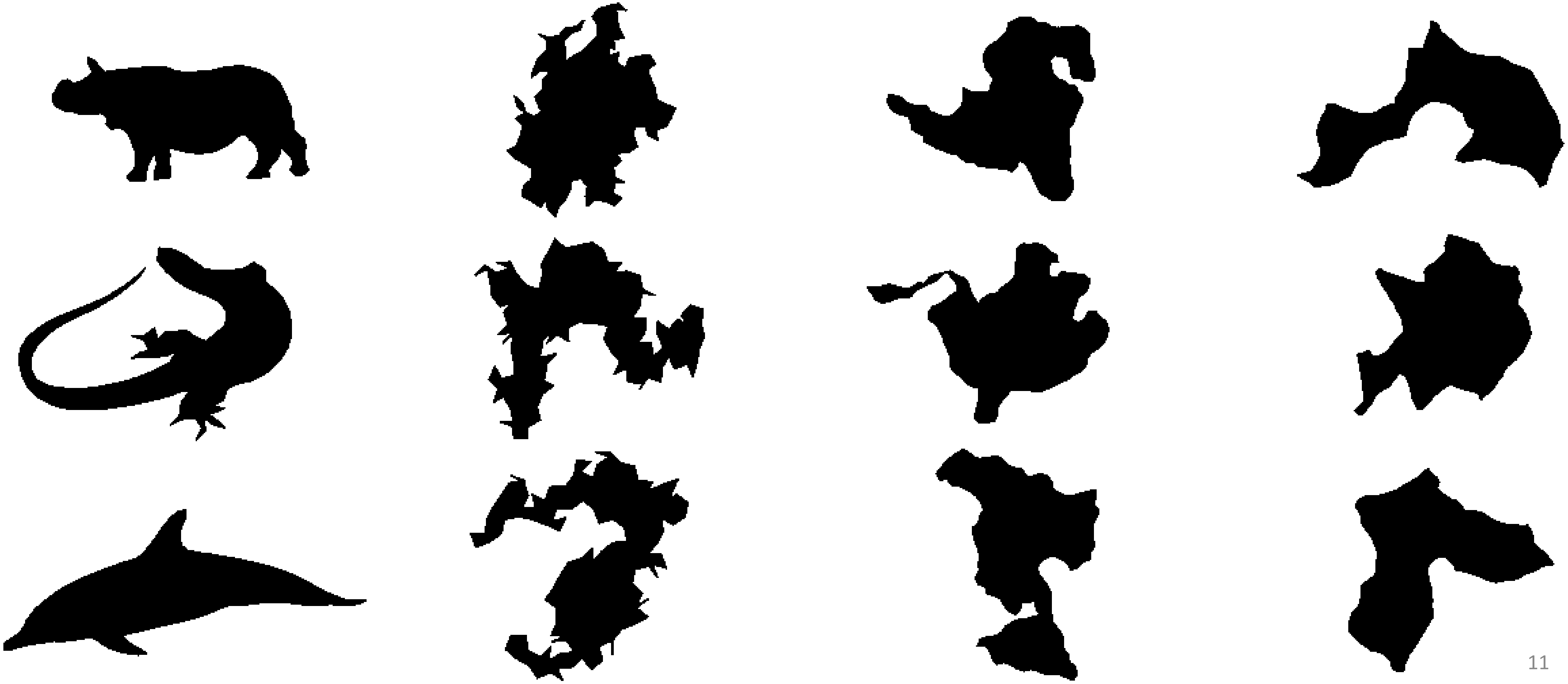
Experiment 1: local shape metamers

Successively added constraints



Experiment 1: Example local-matched shapes

Animals	Mean-matched	Variance-matched	Full-distribution



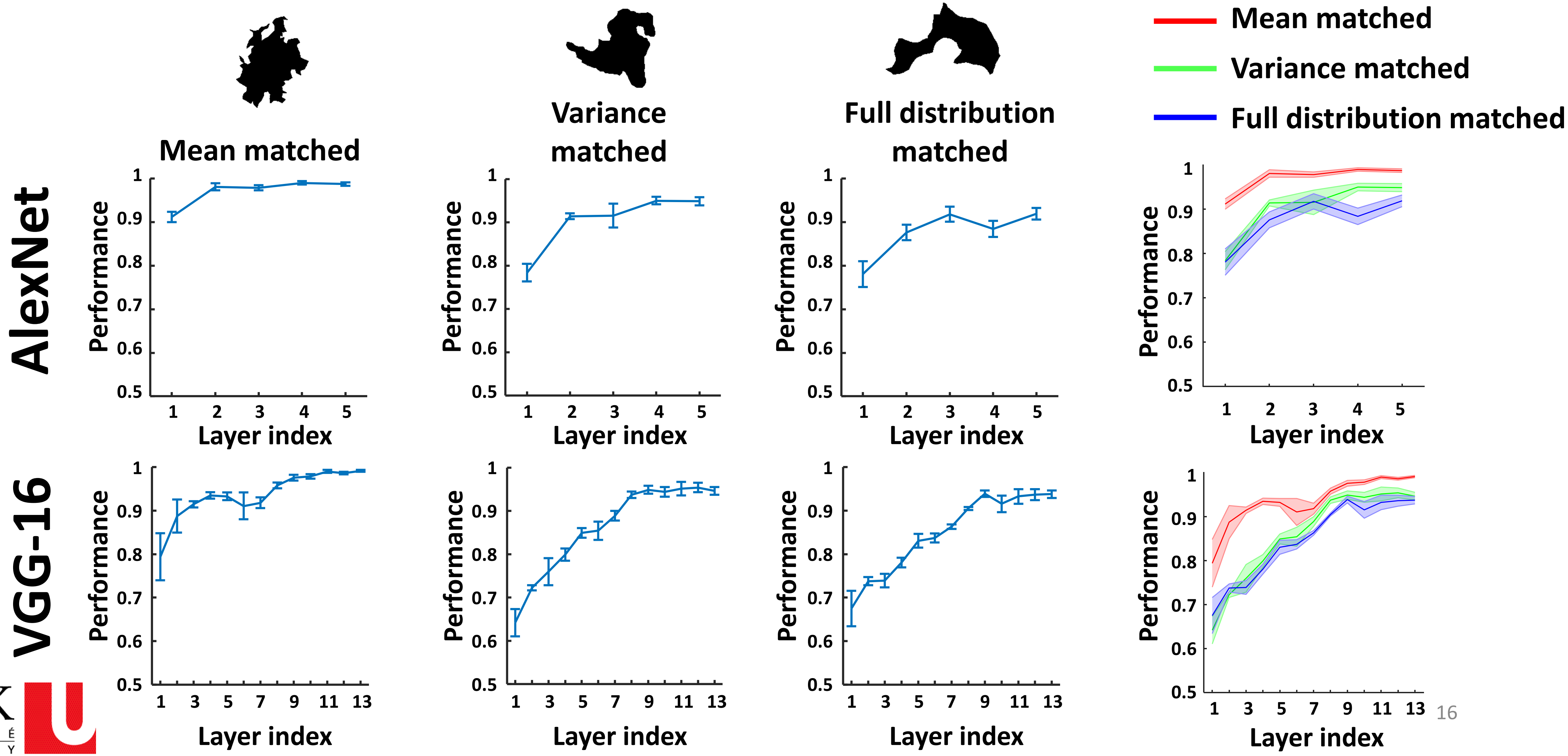
Approach to testing DCNN shape sensitivity

- 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

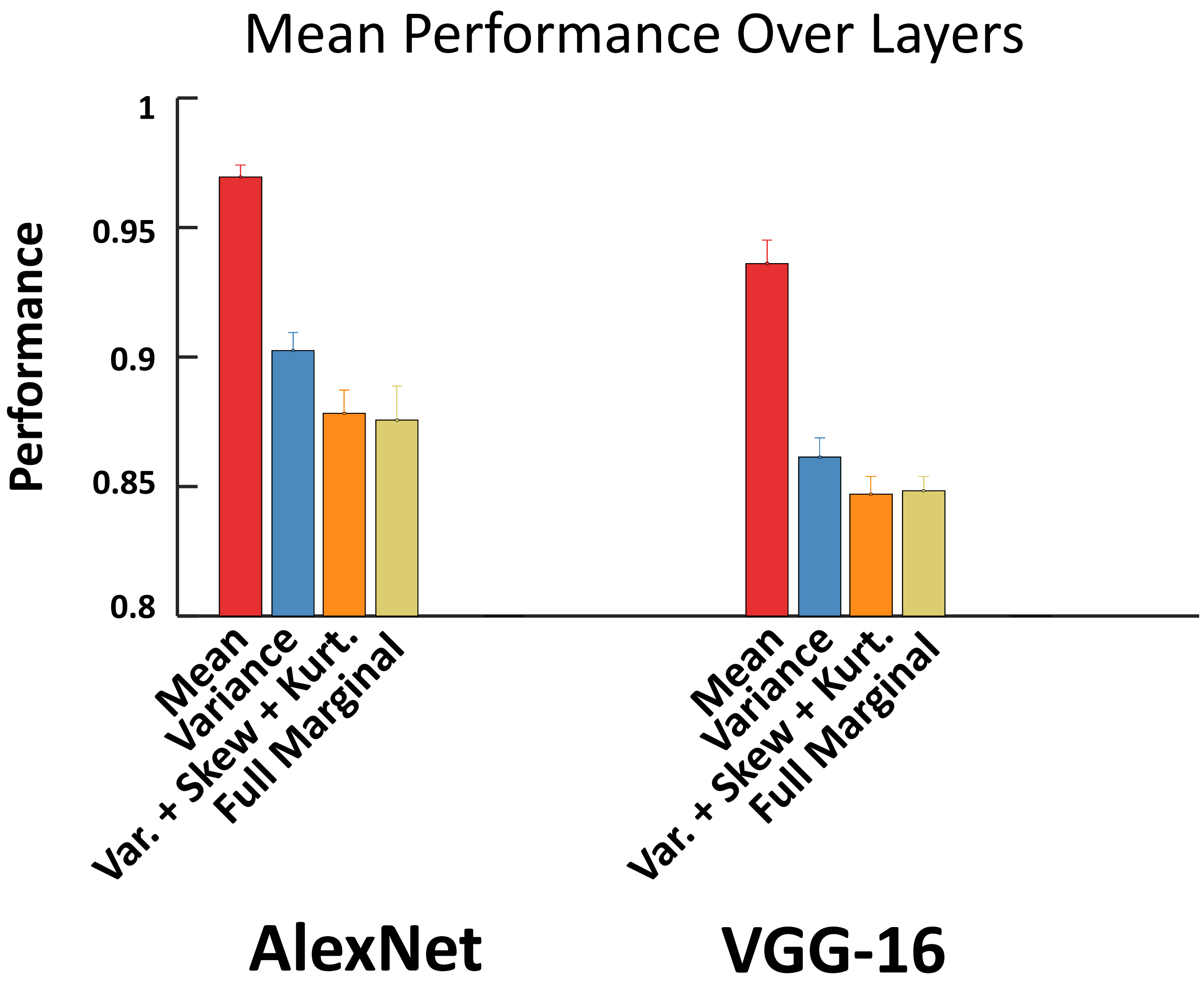
Experimental methods

- Natural shapes: 391 animal shapes from Hemera dataset
- 391 synthetic shapes per dataset
- Silhouettes are area-matched, resized to 224 by 224 pixels.
- Train **linear classifier** to distinguish animals from synthetic shapes based on activations at each convolutional layer.
  - 8 epochs of training per category
  - 5-fold cross validation
  - Small rotations and flips during training

Experiment 1 results: networks compute nonlocal shape information, and it increases with depth



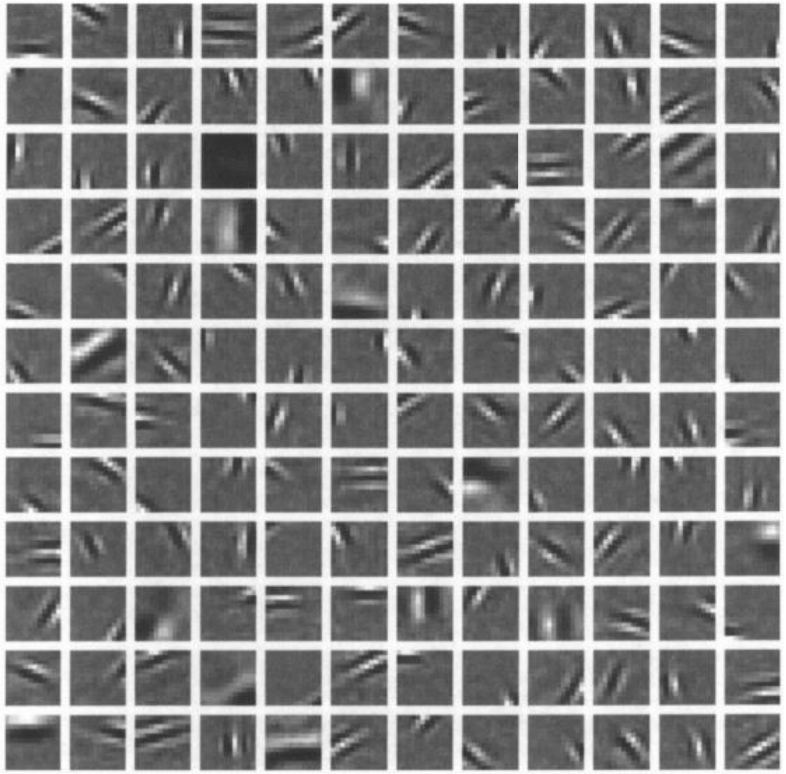
Average performance: sparse components most metameric



Talk outline

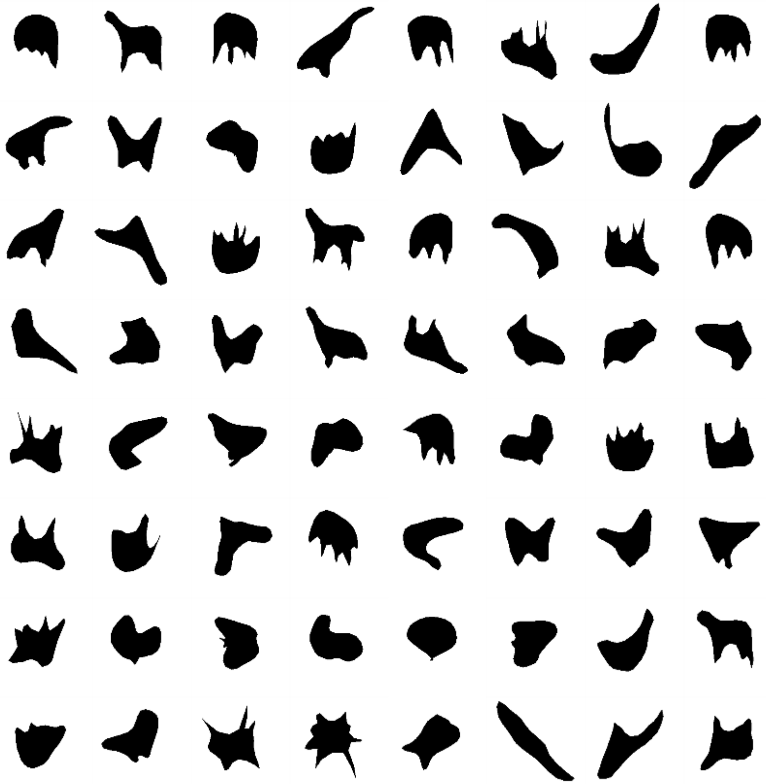
- Prior work
- Experiment 1: Network sensitivity to local shape information
- Experiment 2: A sparse components model of shape
- Conclusions

Experiment 2: sparse components of global shape



Sparse components of images

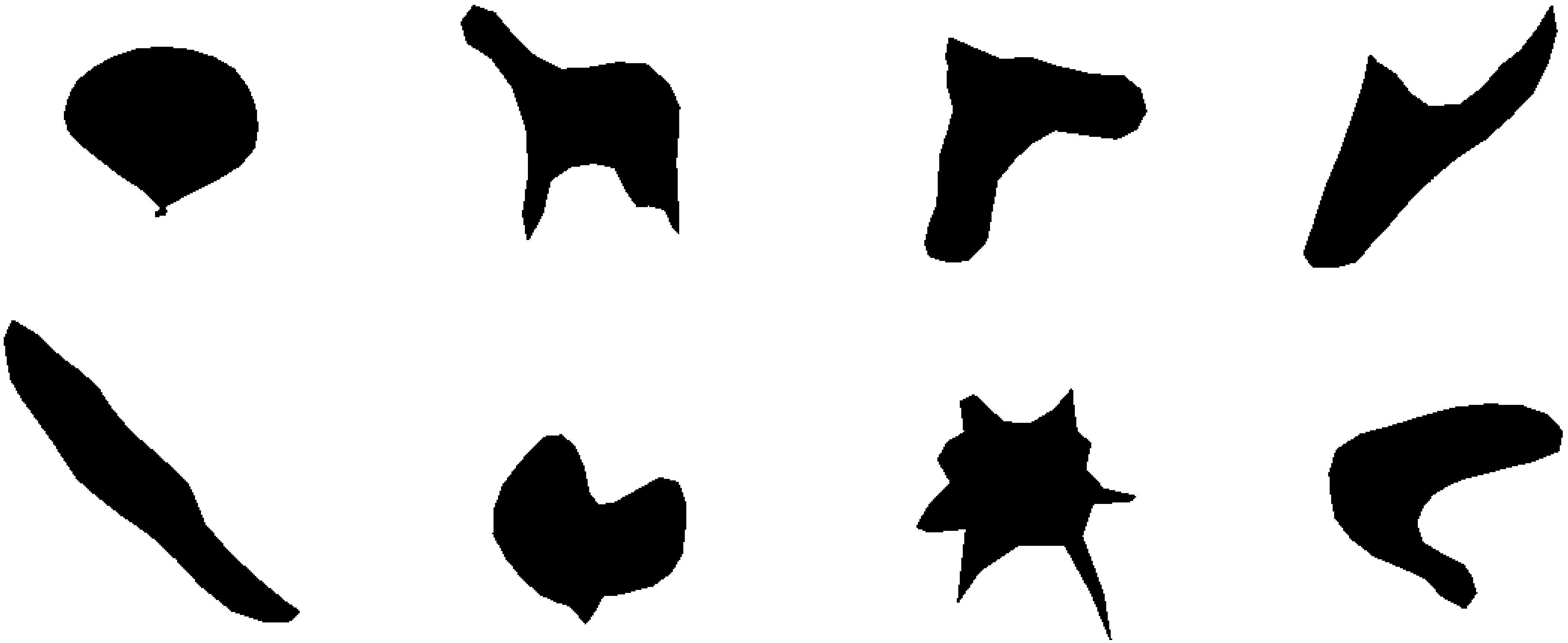
Olshausen & Field (1997)



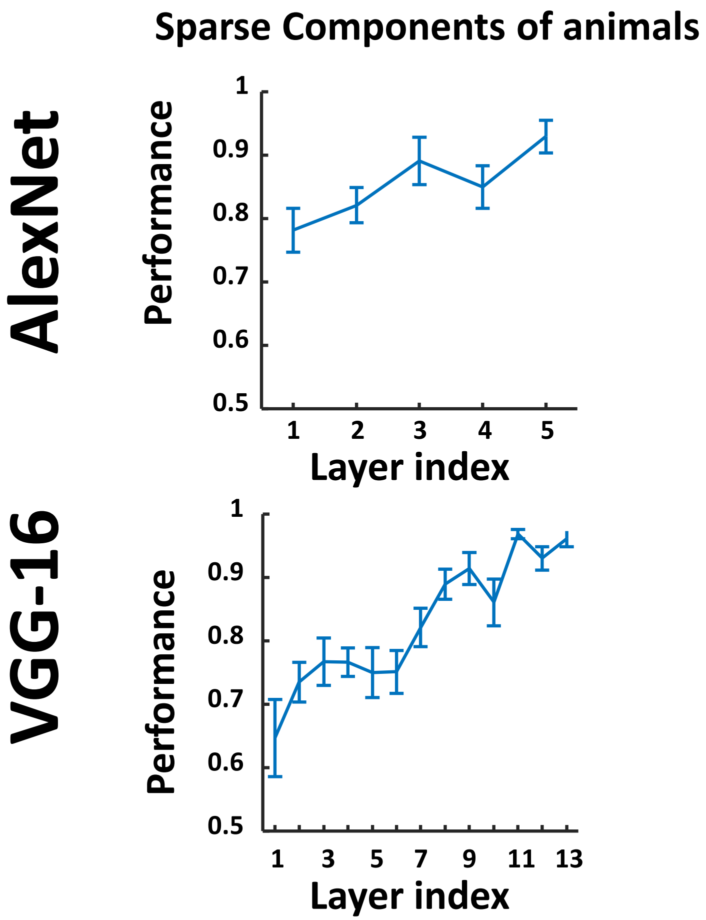
Sparse components of animal shape

Clément & Elder (2018)

Experiment 2: examples of sparse components of animal shape



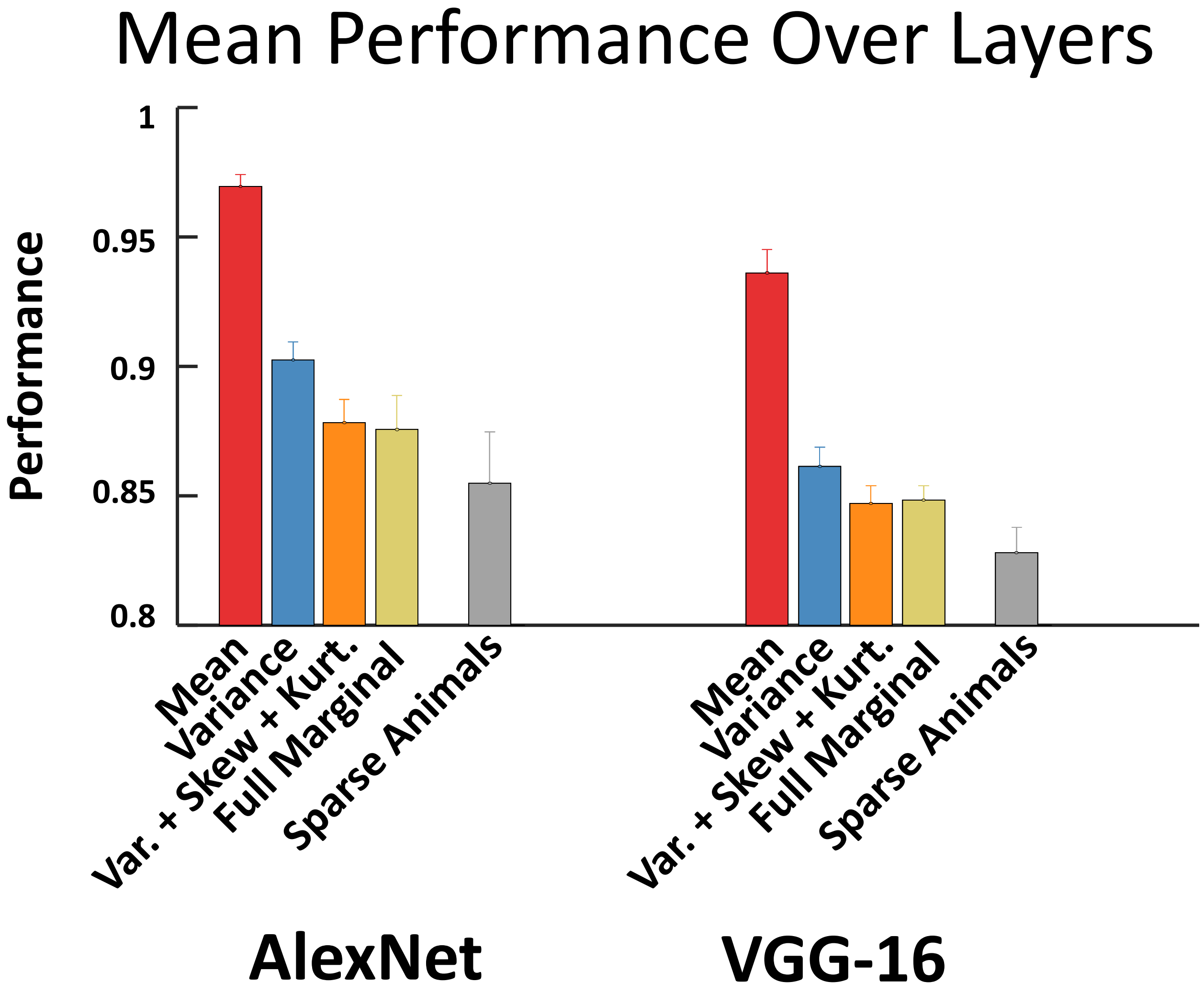
Experiment 2 results: sparse component vs animal is distinguishable in DCNNs, increasingly with layer



Talk outline

- Prior work
- Experiment 1: Network sensitivity to local shape information
- Experiment 2: A sparse components model of shape
- Conclusions

Average performance: sparse components most metameric for DCNNs



Conclusions

- Ability to discriminate animal vs synthetic shapes increases monotonically with network depth.
  - DCNNs code precise information about curvature distribution
  - Non-local shape information emerges in the hierarchy
- Sparse components of global shape (Experiment 2) more metameric than shapes matching only local curvature (Experiment 1).
  - Sparse coding captures something about non-local information computed by DCNNs

Acknowledgements



- James H. Elder
- Ingo Fründ
- Hemanth Pidaparthy & Kris Ehinger
- Michaël Clément & Yiming Qian
- Bob Hou
- VISTA Postdoctoral Grant