

Zero-Shot Learning

Noor Bafageeh

Outline

- What is Zero-shot learning?
- Motivation
- How Does Zero-Shot Learning work?
- Zero-Shot Learning Models with Localization
- Zero-shot Localization with Attributes

What is Zero-shot learning?

- Recognize the **Wampimuk**:
 - Impossible?
- Solution: Side Information
 - Wampimuk := small, horns, furry, cute.
 - Wikipedia Page.
- Zero-shot:
 - The model is doing a task with **Zero training examples**
 - Solved with having external knowledge source (side information)



Supervised Vs. Zero-Shot Learning

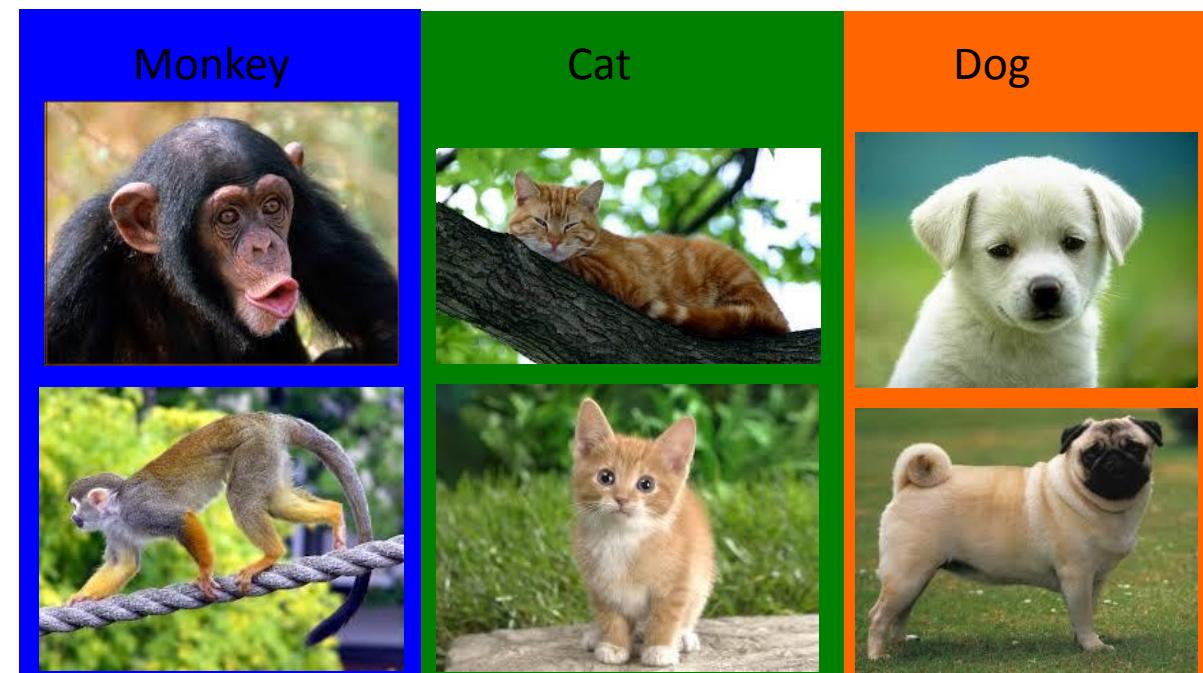
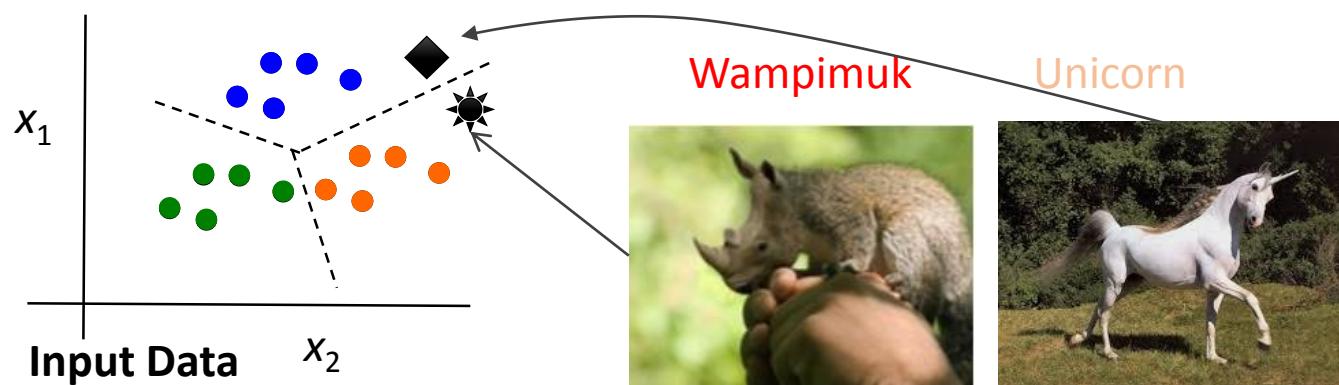
Output Labels:

Dog

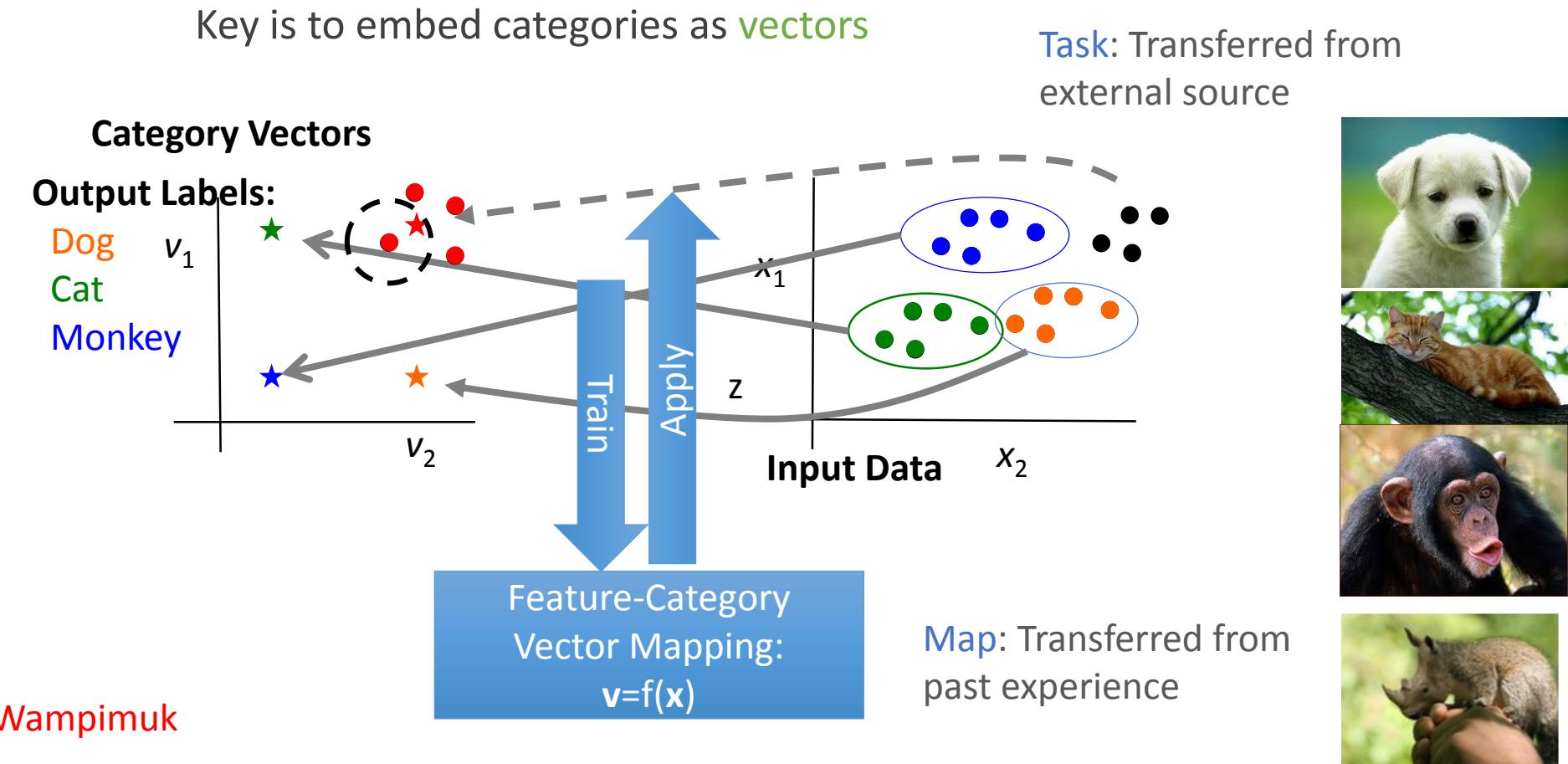
Cat

Monkey

???



Supervised Vs. Zero-Shot Learning



Outline

- What is Zero-shot learning?
- Motivation
- How Does Zero-Shot Learning work?
- Zero-Shot Learning Models with Localization
- Zero-shot Localization with Attributes

We have labeled data, why bother?



An image



Classification

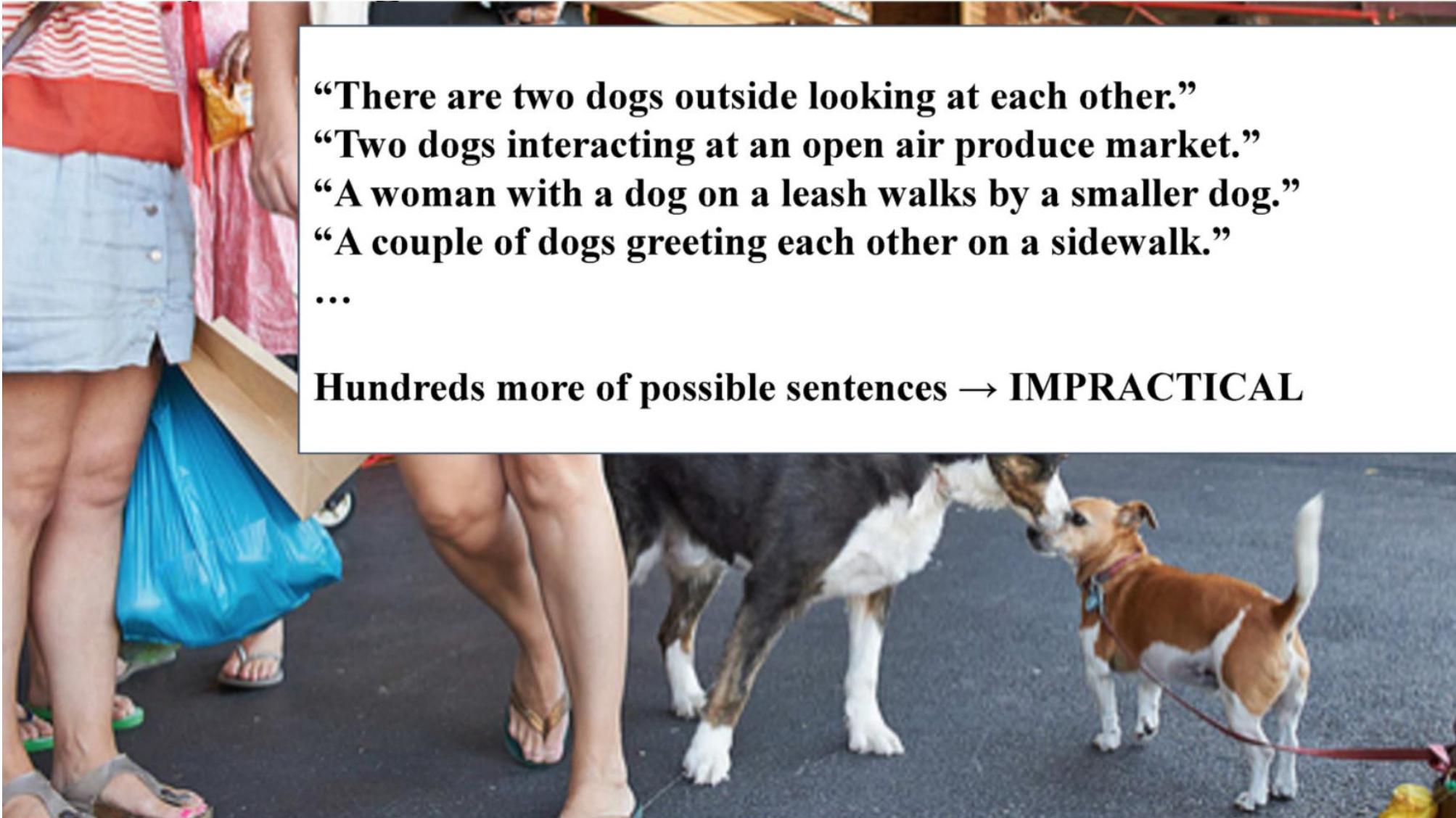


Person, dog, bicycle, bag, apples

Segmentation

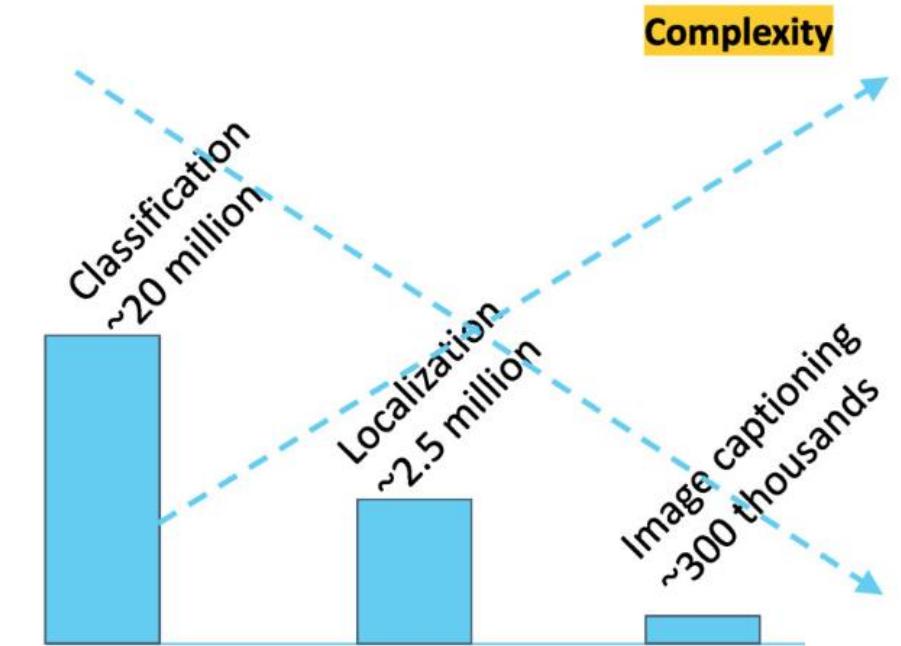


Captioning



Why Zero-shot learning?

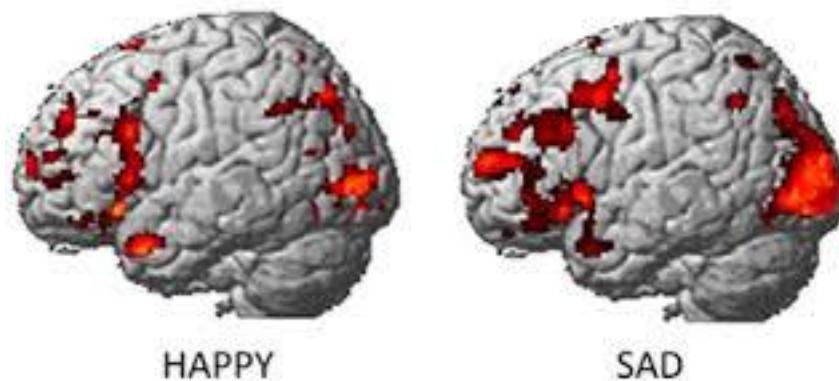
- Annotation Vs. Complexity
 - The more complex tasks we target, the fewer annotation we have, the more relevant zero-shot learning is



Imagenet+Open Images+MS COCO

Why Zero-shot learning?

- Examples:
 - Object Recognition
 - fMRI mind reading
 - Every annotation is a brain-scan
 - Every word is a category

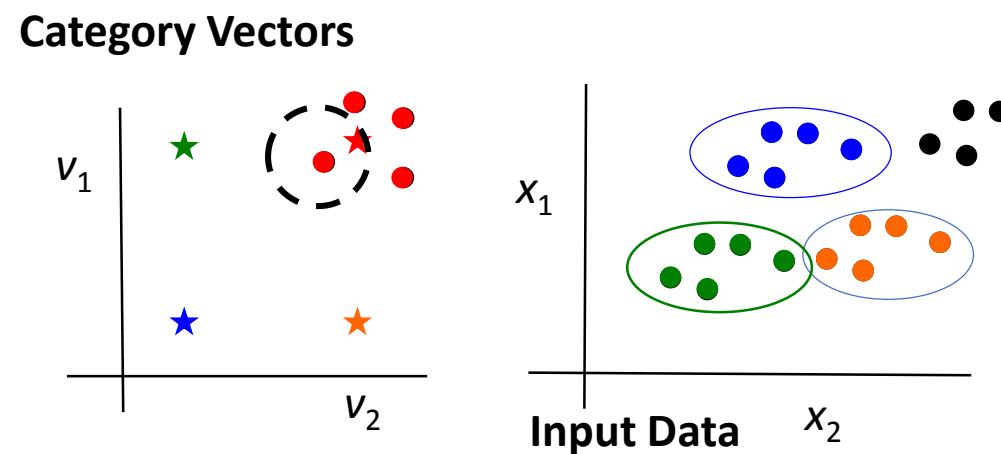


Outline

- What is Zero-shot learning?
- Motivation
- How Does Zero-Shot Learning work?
- Zero-Shot Learning Models for Image Classification
- Zero-Shot Learning Models with Localization
- Zero-shot Localization with Attributes

Regression/Classification Approach

- A simple category vector: class-level attribute description.
 - Wampimuk := small, cute, furry, horns.
- Train:
 - Given some known class-category vectors v and data x :
 - Learn image-attribute classifiers/regressors $v=f(x)$.
 - E.g., SVM/SVR. Deep Neural Nets.
- Test:
 - Specify vec v^* for new class
 - Map new data $f(x^*)$
 - NN matching of v^* vs $f(x^*)$
- Pros:
 - Easy and fast!
- Cons
 - Category separability

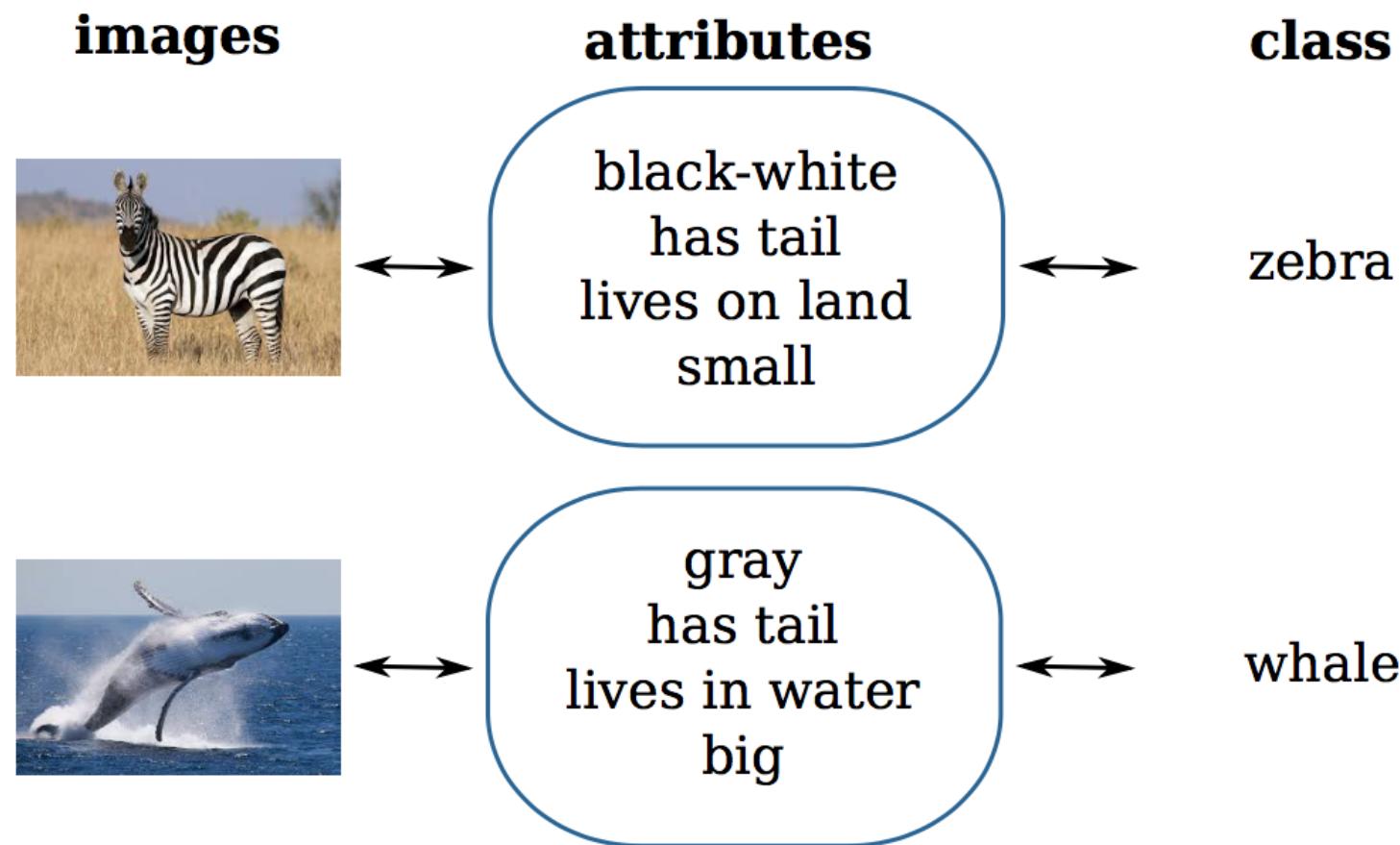


Energy Function Approach

- Given: \mathbf{x} : data, \mathbf{v} category vectors.
- Train an energy function $F(\mathbf{x}, \mathbf{v})$.
 - E.g. Bilinear: $F(\mathbf{x}, \mathbf{v}) = \mathbf{x}' \mathbf{W} \mathbf{v}$.
 - \mathbf{W} such that $F(\mathbf{x}, \mathbf{v})$ is large when data and task match. $\mathbf{x} = \mathbf{v}$.
 - \mathbf{W} such that $F(\mathbf{x}, \mathbf{v})$ is small when data and semantics mismatch. $\mathbf{x} \neq \mathbf{v}$.
 - => Max margin ranking objective.
- Test time: Classify novel class examples \mathbf{x}^* .
 - Specify \mathbf{v}^* vectors for some novel classes.
 - Evaluate $F(\mathbf{x}^*, \mathbf{v}^*)$ for each \mathbf{v}^* .
 - Max response gives classification.
- Max margin separability => accuracy. But: more complex and slower.

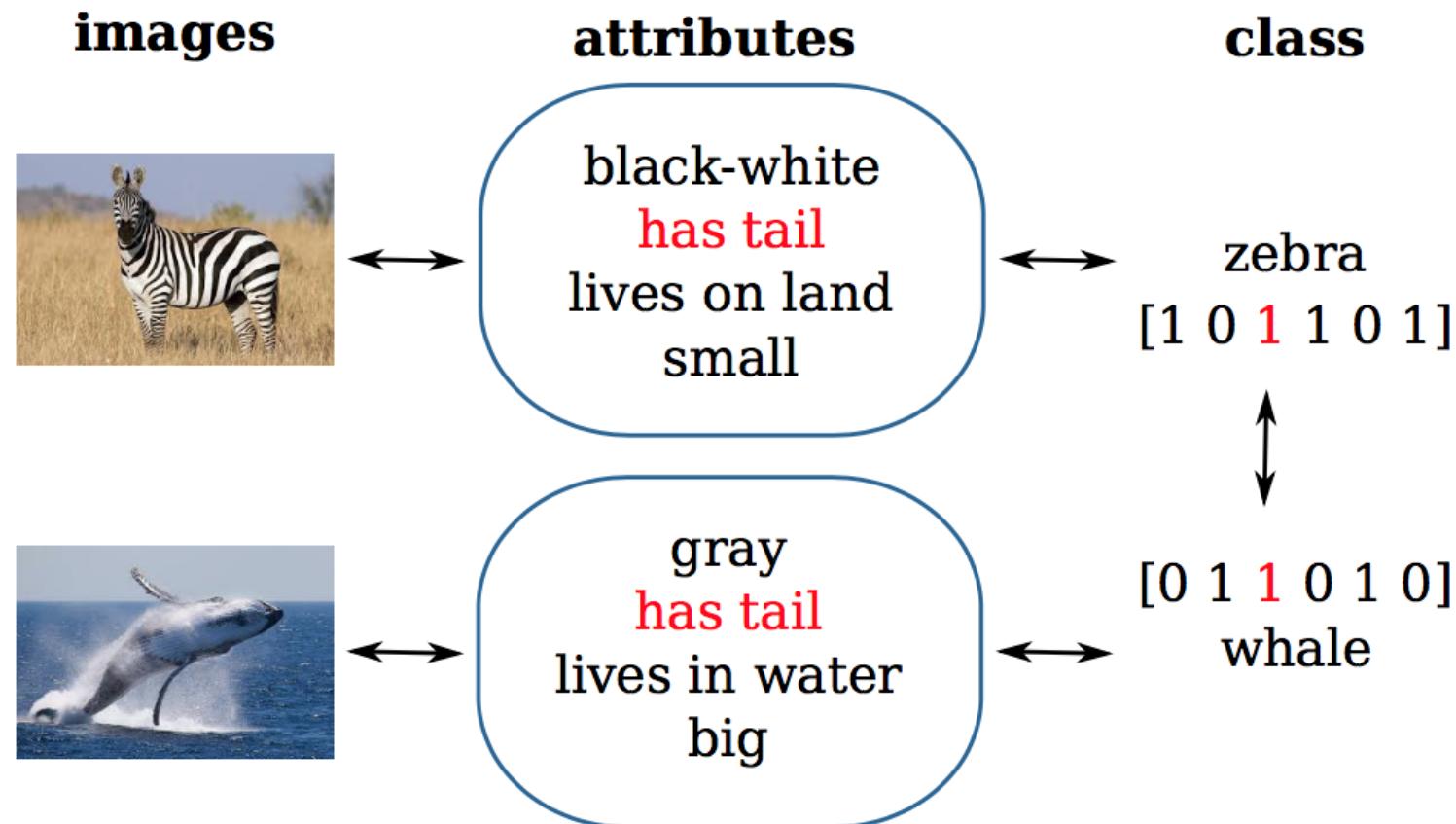
Where to get Category Vectors?

- Attributes as side information



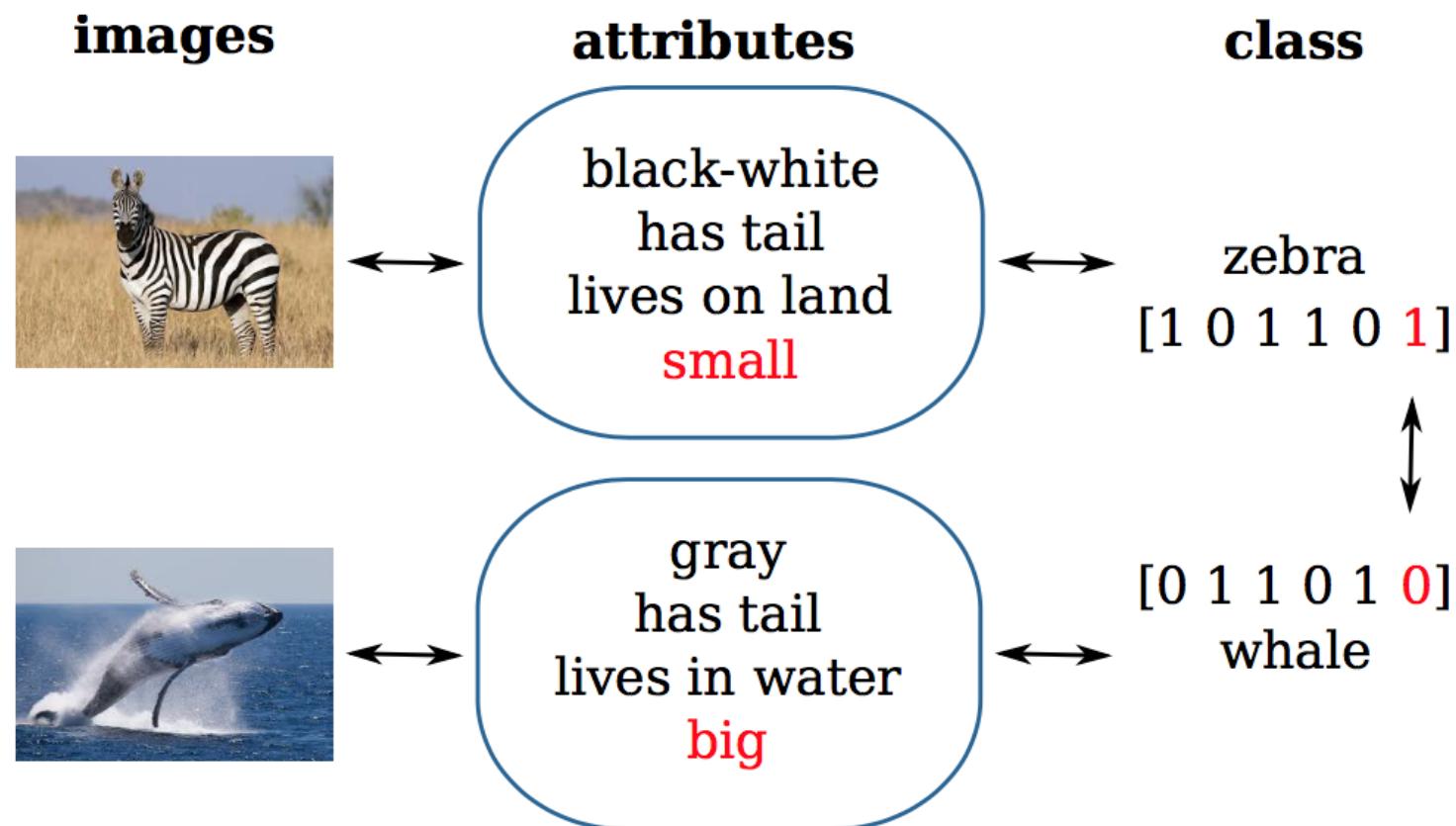
Where to get Category Vectors?

- Attributes as side information



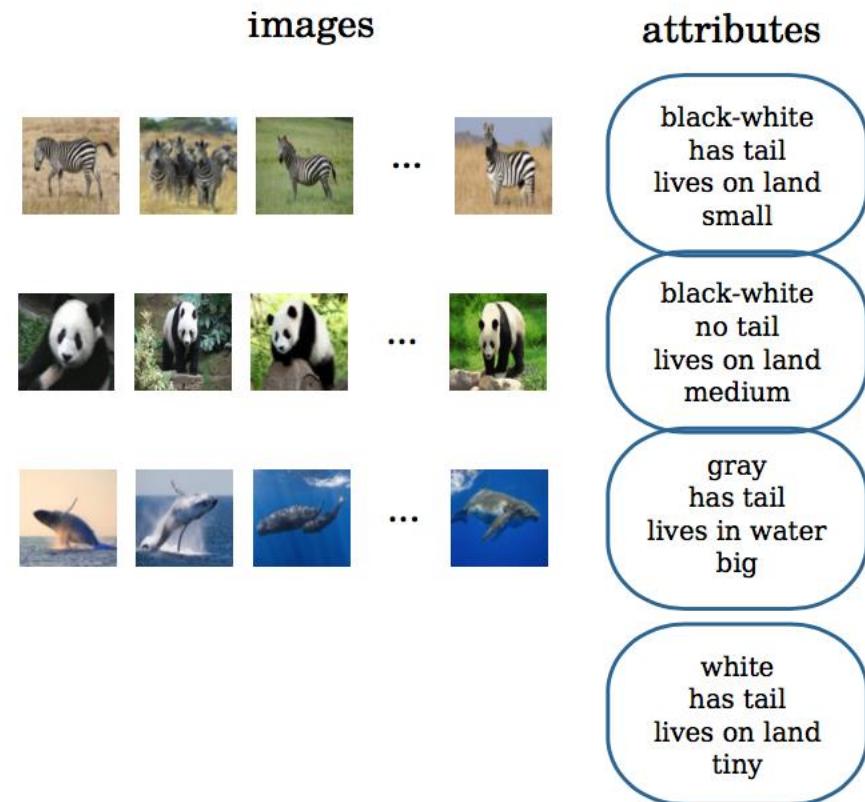
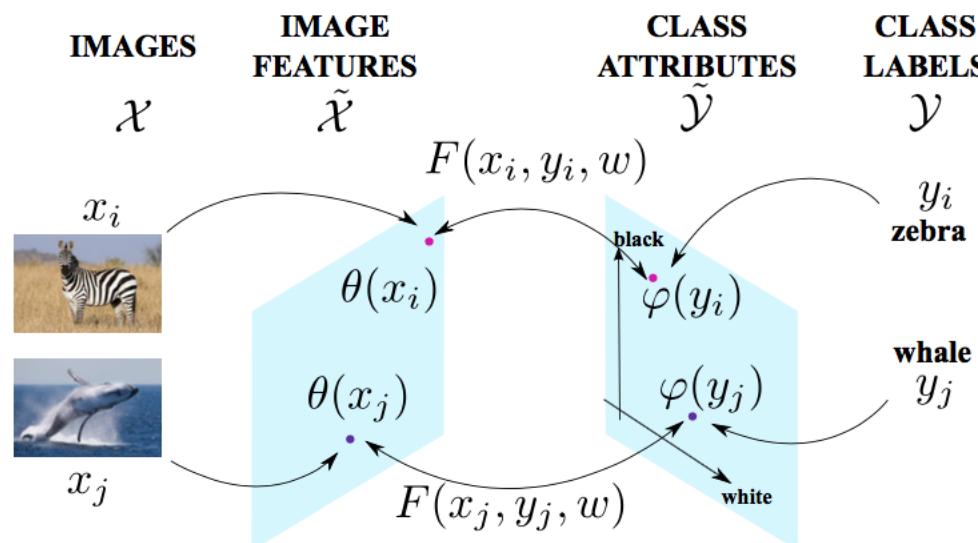
Where to get Category Vectors?

- Attributes as side information



Where to get Category Vectors?

- Attributes as side information

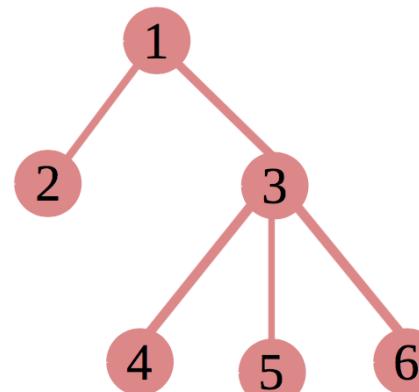


Where to get Category Vectors?

- Wikipedia and WordNet as Side Information
 - object descriptions or hierarchies



Word2vec or GloVe

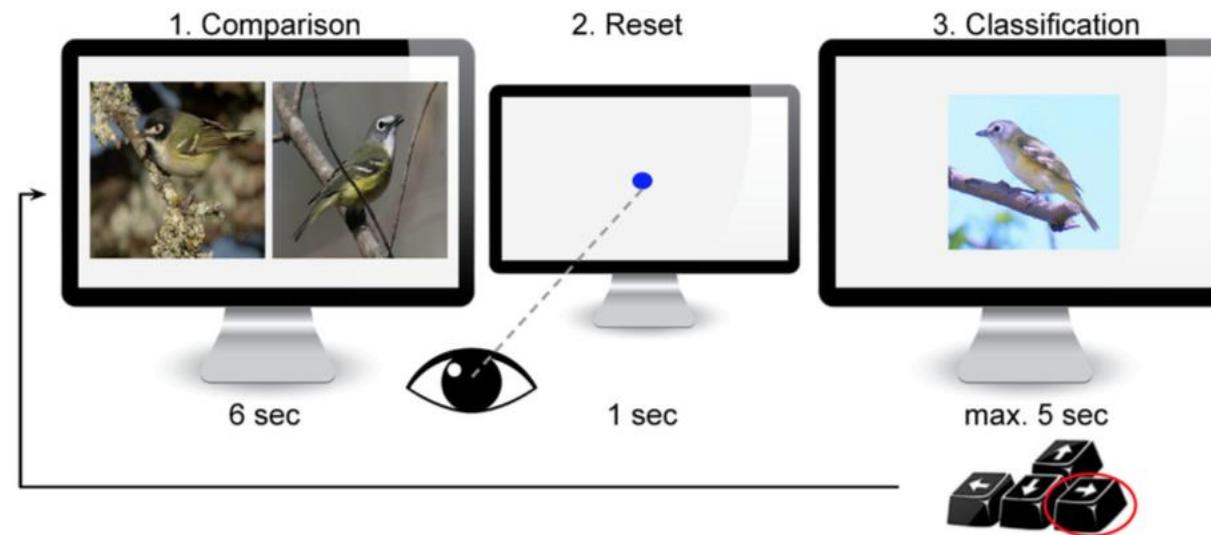


$$2 = [1 \ 0 \ 2 \ 3 \ 3 \ 3]$$

Hierarchical similarity measures

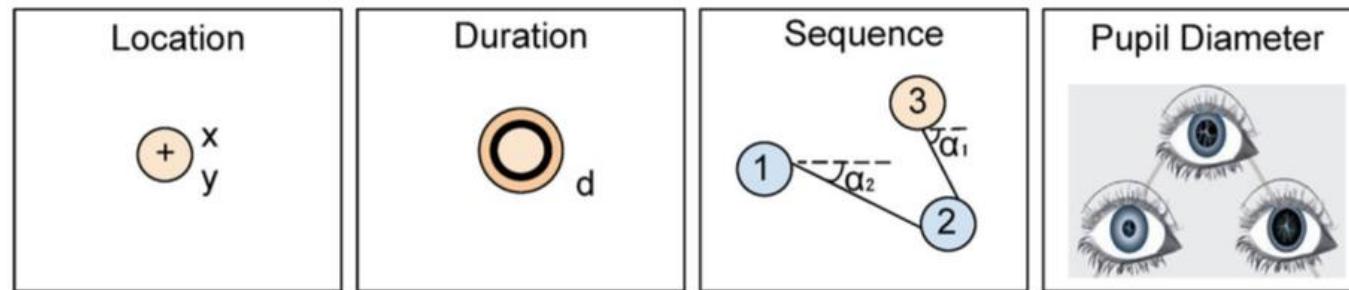
Where to get Category Vectors?

- Human Gaze as Side information



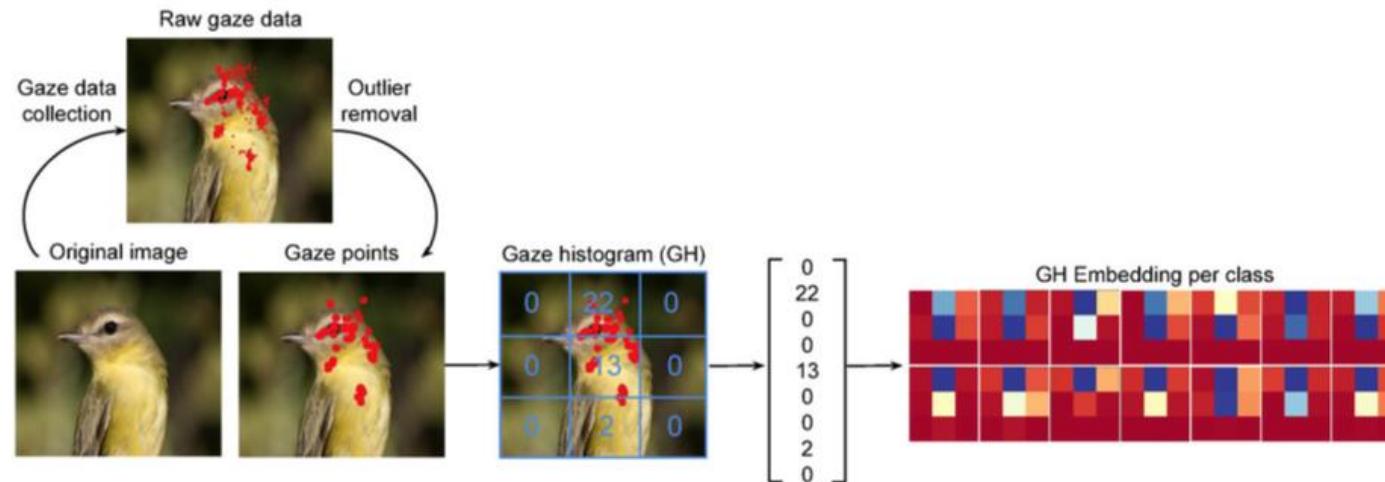
Where to get Category Vectors?

- Human Gaze as Side information



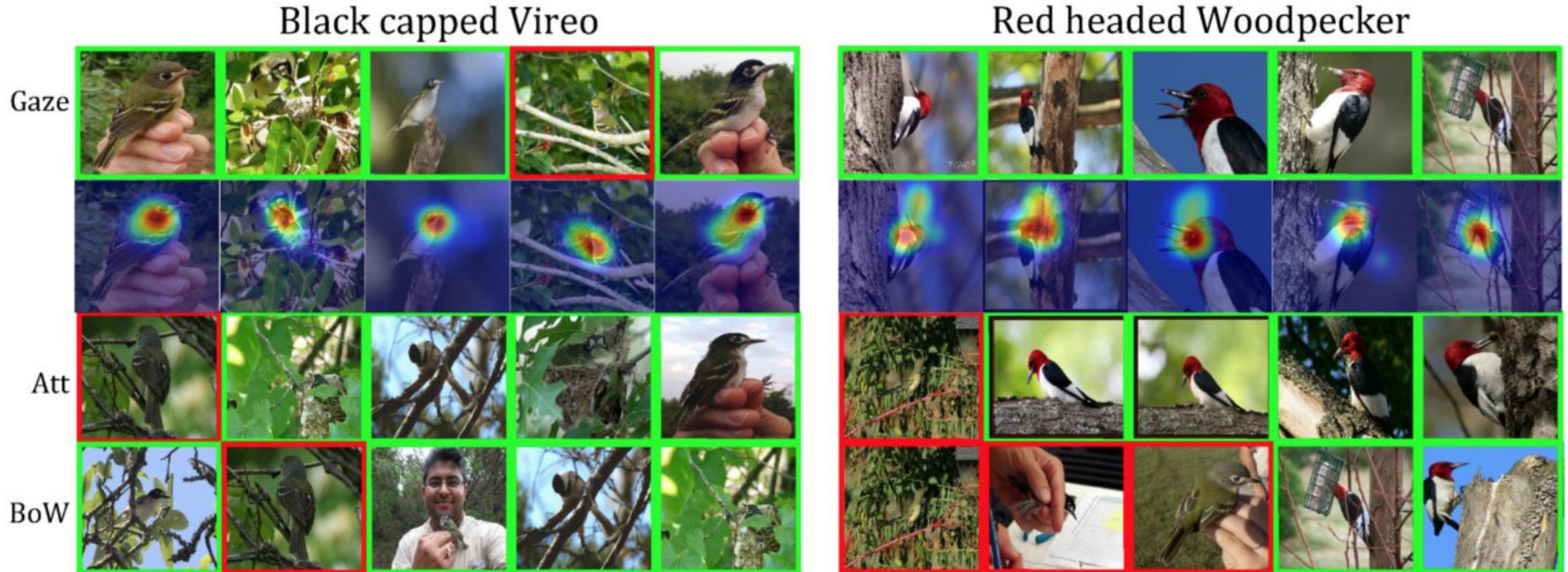
Where to get Category Vectors?

- Human Gaze as Side information



Where to get Category Vectors?

- Human Gaze as Side information



Outline

- What is Zero-shot learning?
- Motivation
- How Does Zero-Shot Learning work?
- **Zero-Shot Learning Models with Localization**
- Zero-shot Localization with Attributes

Traditional Localization

Training



Testing

Bicyclist



Zero-Shot Localization

Training

Known visual classes

wheels



helmet



street



Testing

Bicyclist:
“wheels”+“helmet”+“street”



Why Zero-Shot Localization is Important?

Find the object

Mammal

Brown

Hairy

Colorful

Pointy ears

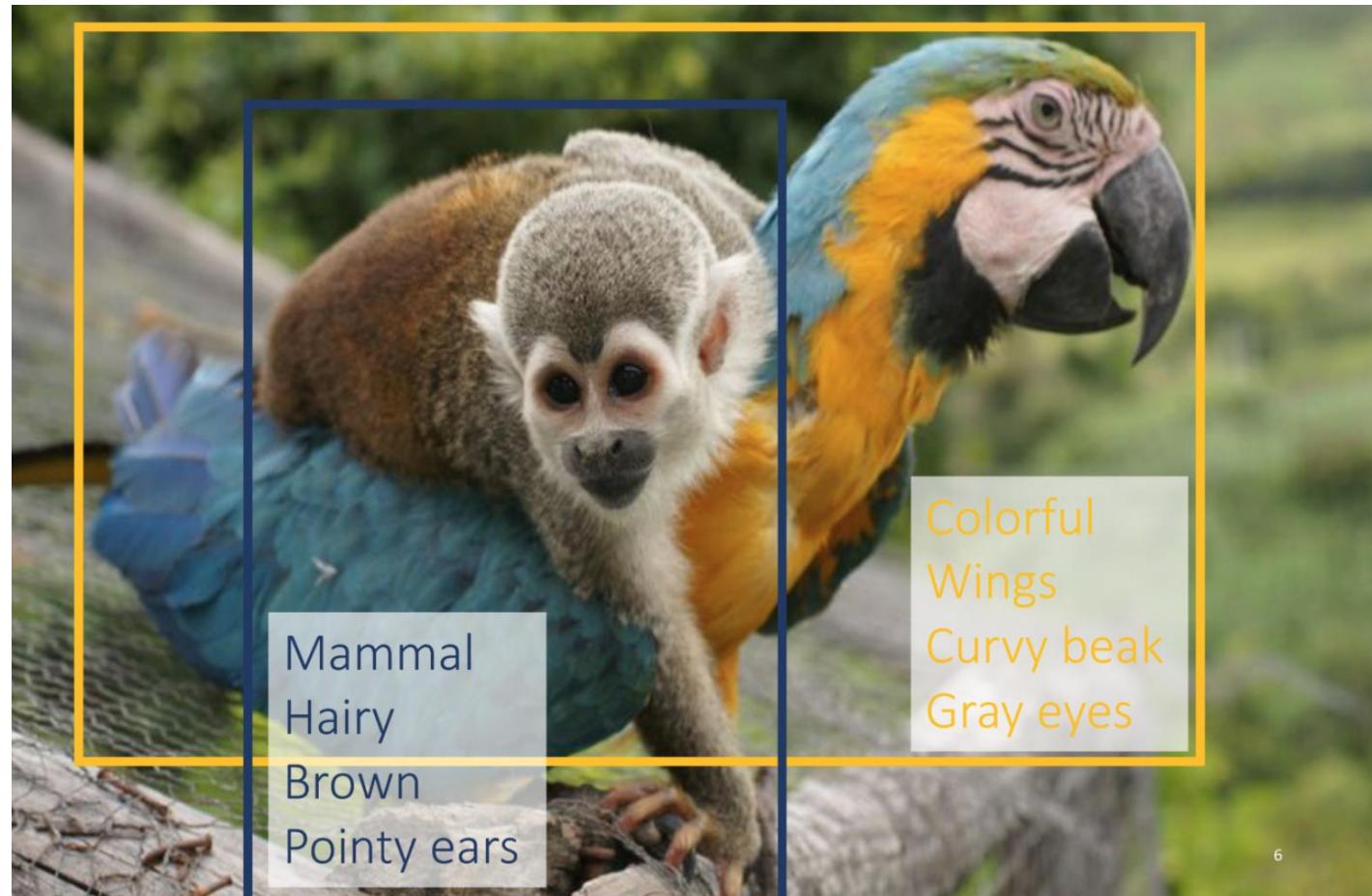
Wings

Curvy beak

Gray eyes

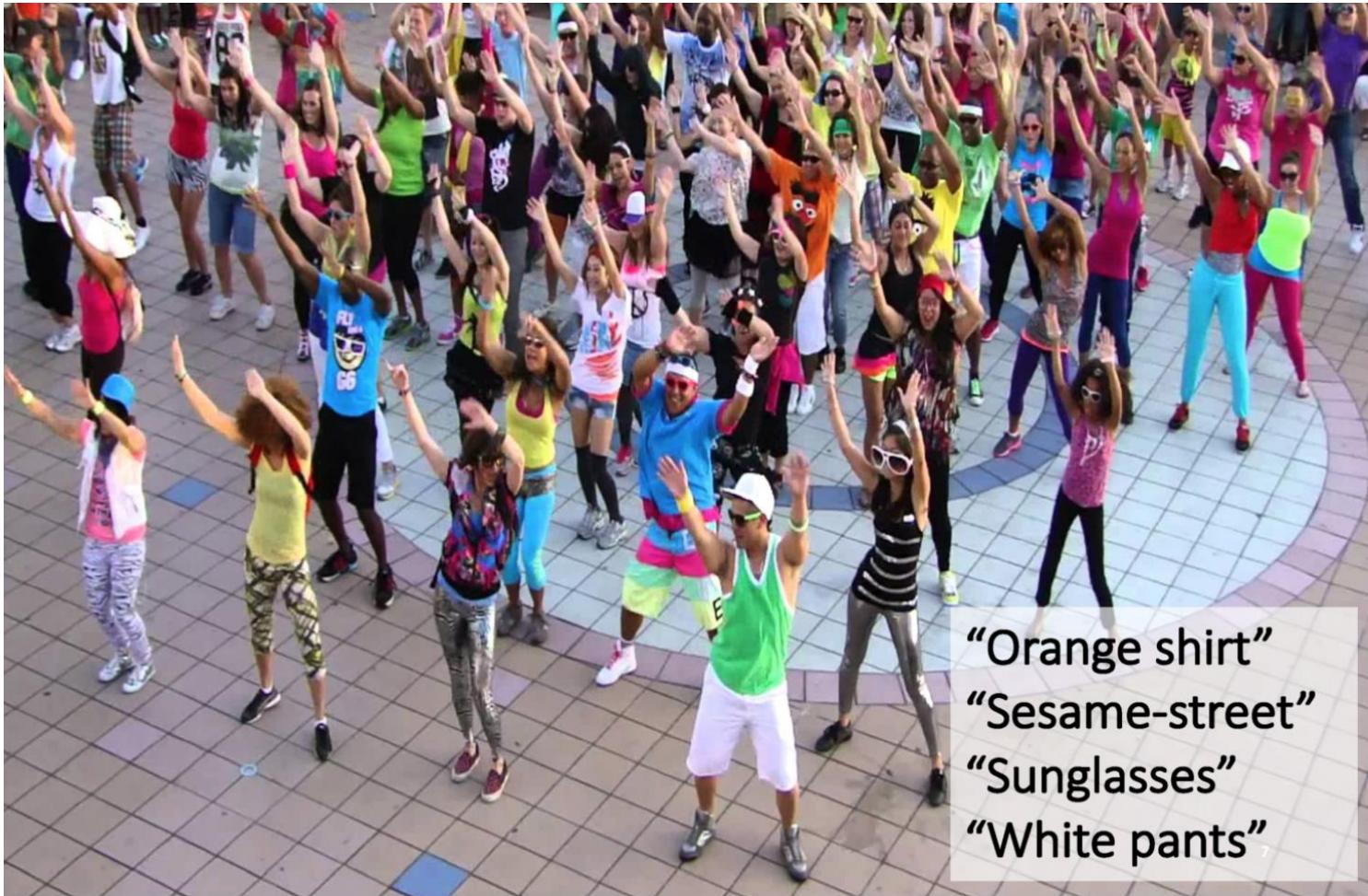
Why Zero-Shot Localization is Important?

1. Attributes belong to objects ,not images.



Why Zero-Shot Localization is Important?

2. It is more relevant in complex scenes



Outline

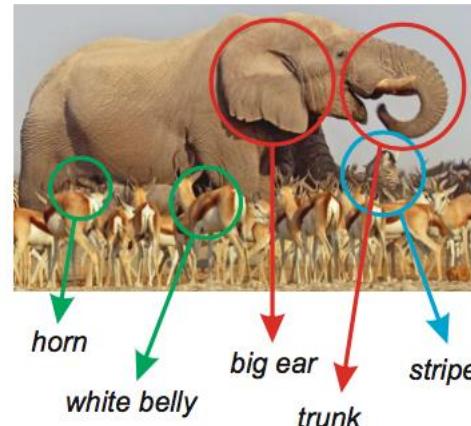
- What is Zero-shot learning?
- Motivation
- How Does Zero-Shot Learning work?
- Zero-Shot Learning Models with Localization
- Zero-shot Localization with Attributes

Zero Localization by Attributes

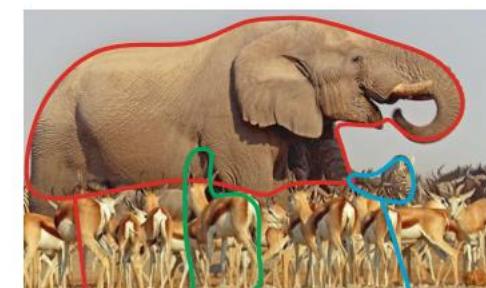
- Region-level attribute based localization



africa
mammal
savana



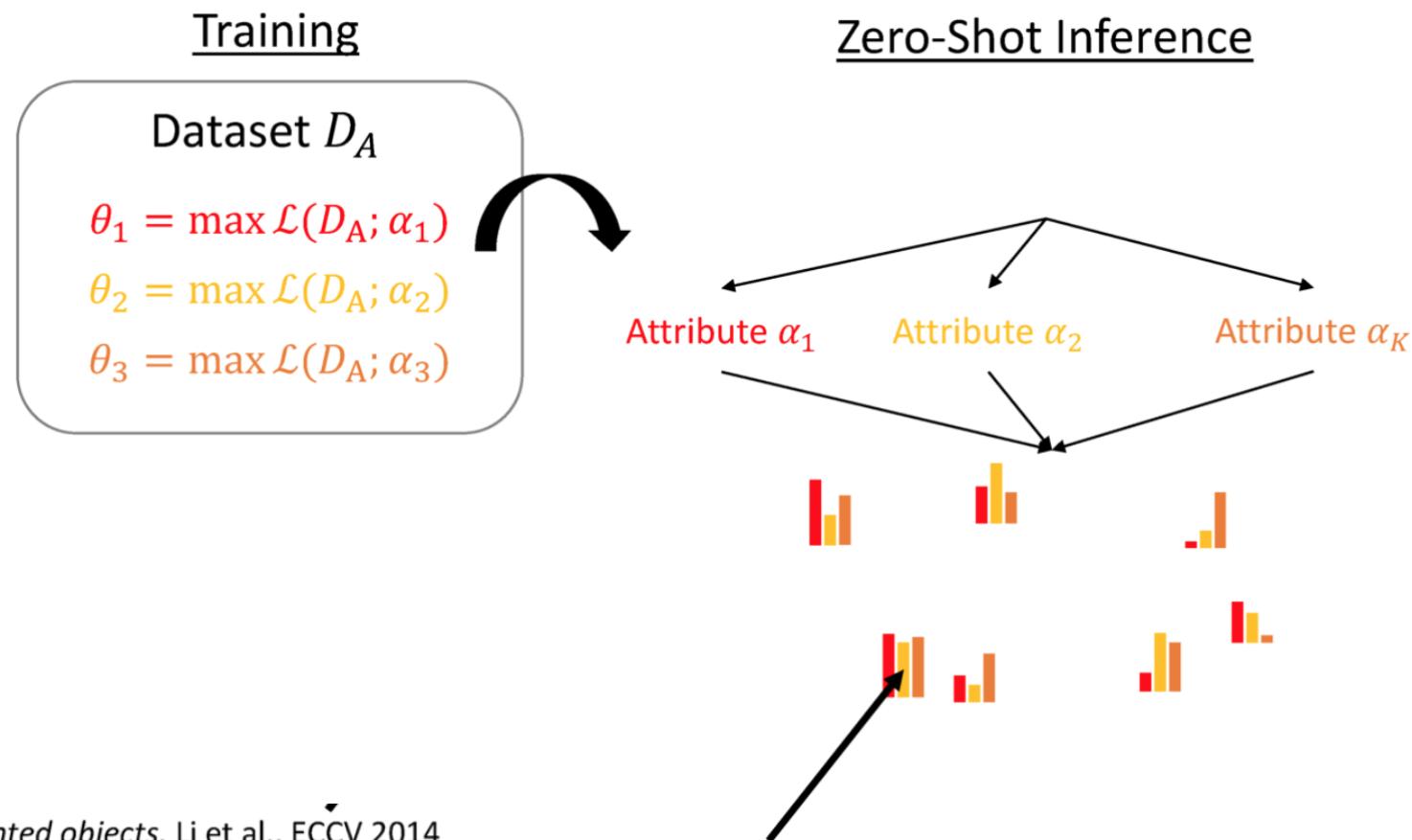
horn
white belly
big ear
trunk
stripes



big ear
trunk
beige color
pachyderm
horn
pointy snout
white belly
 $>60\text{km/h}$
stripes
ungulate
long tail
 $>40\text{km/h}$

Zero Localization by Attributes

- Region-level attribute based localization



Zero Localization by Attributes

1. Extract regions localization (CPMC, ~ 500)
2. Learn attributes with ALE

$$f(x) = \arg \max_{y \in \mathcal{Y}} \max_{z \in Z(x)} F(z, y)$$

$$F(z, y; W, \phi) = \theta(z)' W \phi(y)$$

$$\min_W \frac{\lambda}{2} \|W\|^2 + R(W, \Phi^A)$$

Per Region
Maximization

CPMC Regions



ALE Attributes

Zero Localization by Attributes

1. Extract regions localization (CPMC, ~ 500)
2. Learn attributes with ALE
3. Efficient inference by codemaps

CPMC Regions



**Per Region
Maximization**

ALE Attributes