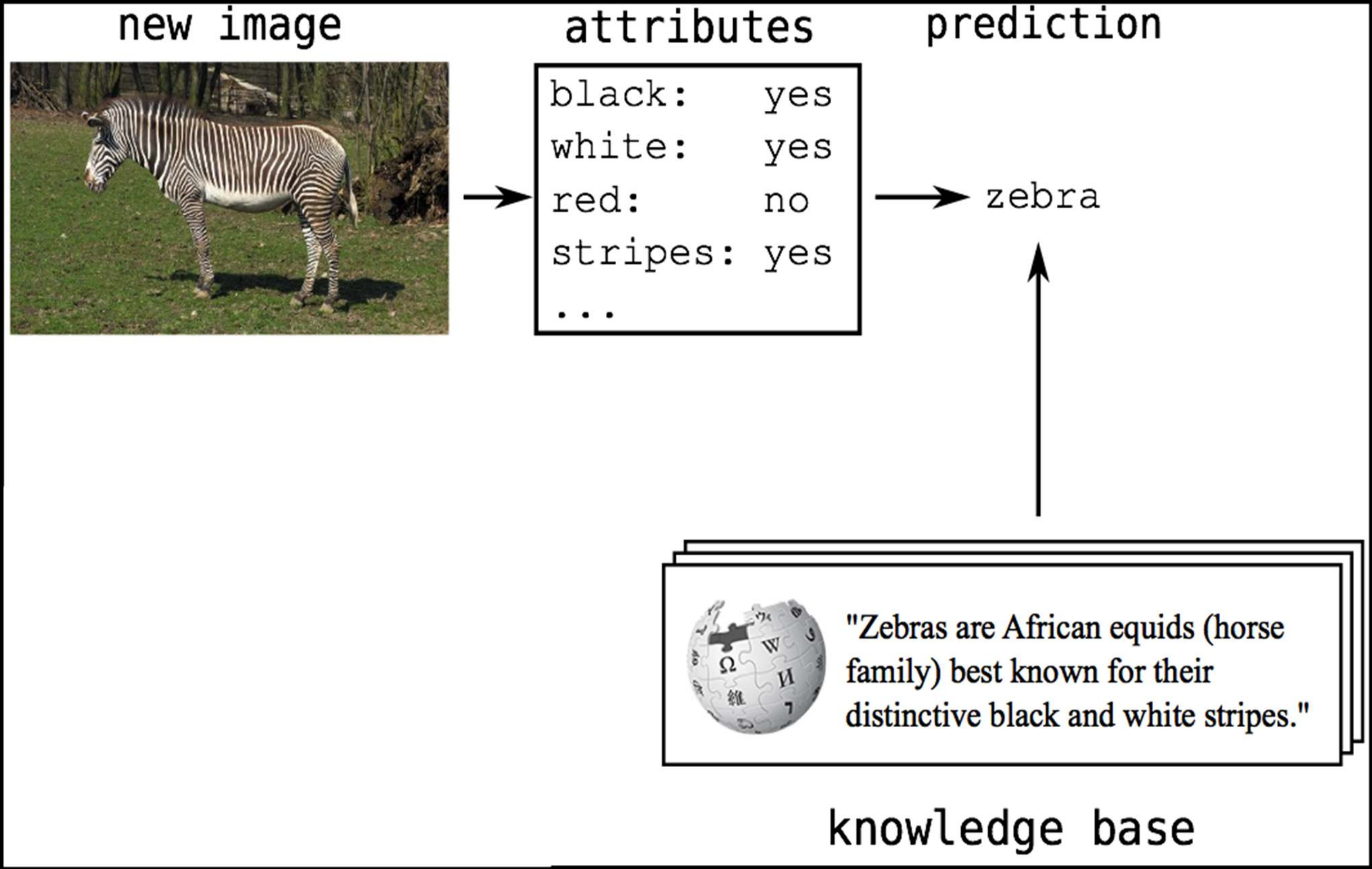
COSTA: Co-occurrence statistics for zero-shot classification

Thomas Mensink – University of Amsterdam

Parts & Attributes Workshop – ECCV 2014
September 12th

Parts & Attributes



Parts & Attributes

- Semantic representation of images
  - Properties of class / context of class
  - Each attribute relevant for a few classes

- Interesting for
  - Zero-shot prediction
  - Few-shot prediction
  - Recounting of visual content

Parts & Attributes: Disadvantages

- Unnatural distinction between
  - Attributes to be detected
  - Classes of interest
- Binary map from classes to attributes
- Inherently multi-class zero-shot prediction

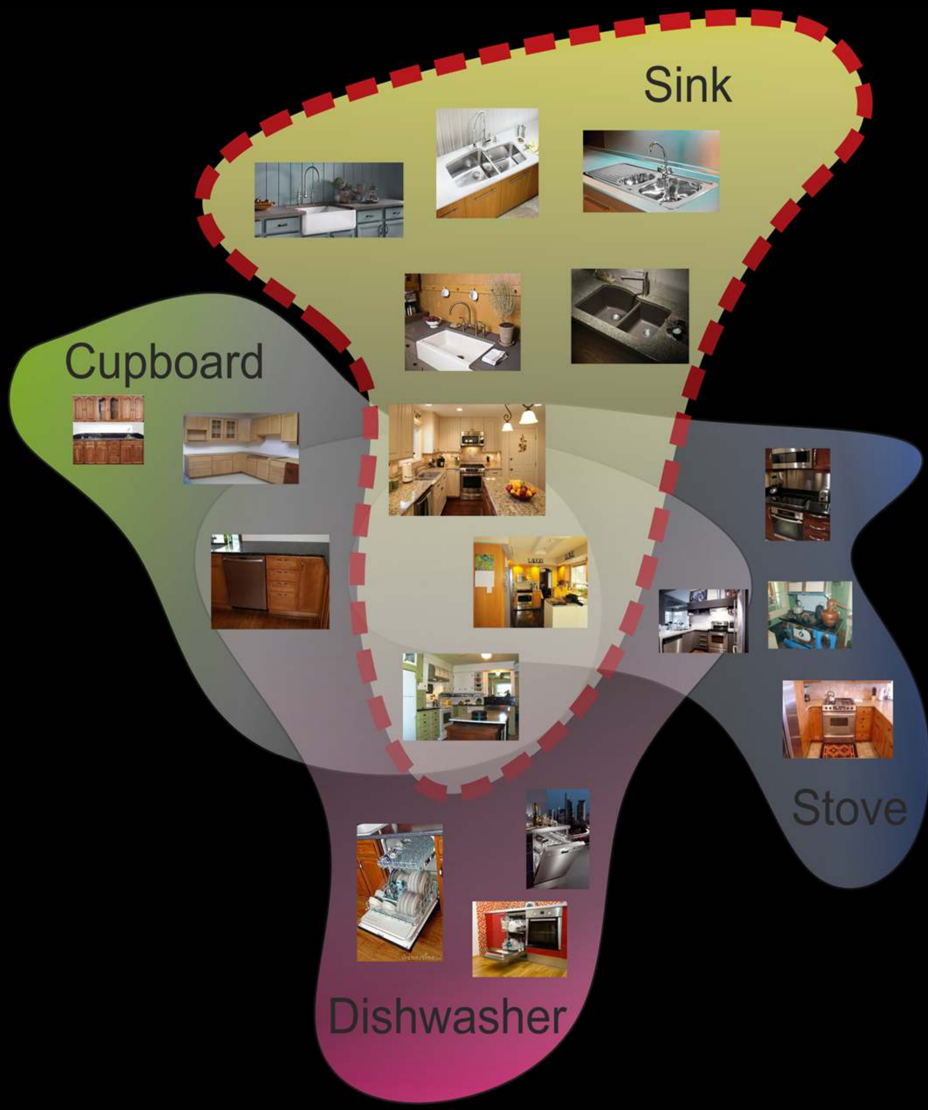
CAN'T WE DO ZERO-SHOT PREDICTION IN MULTI-LABELED DATASETS?

Multi-label zero-shot classification

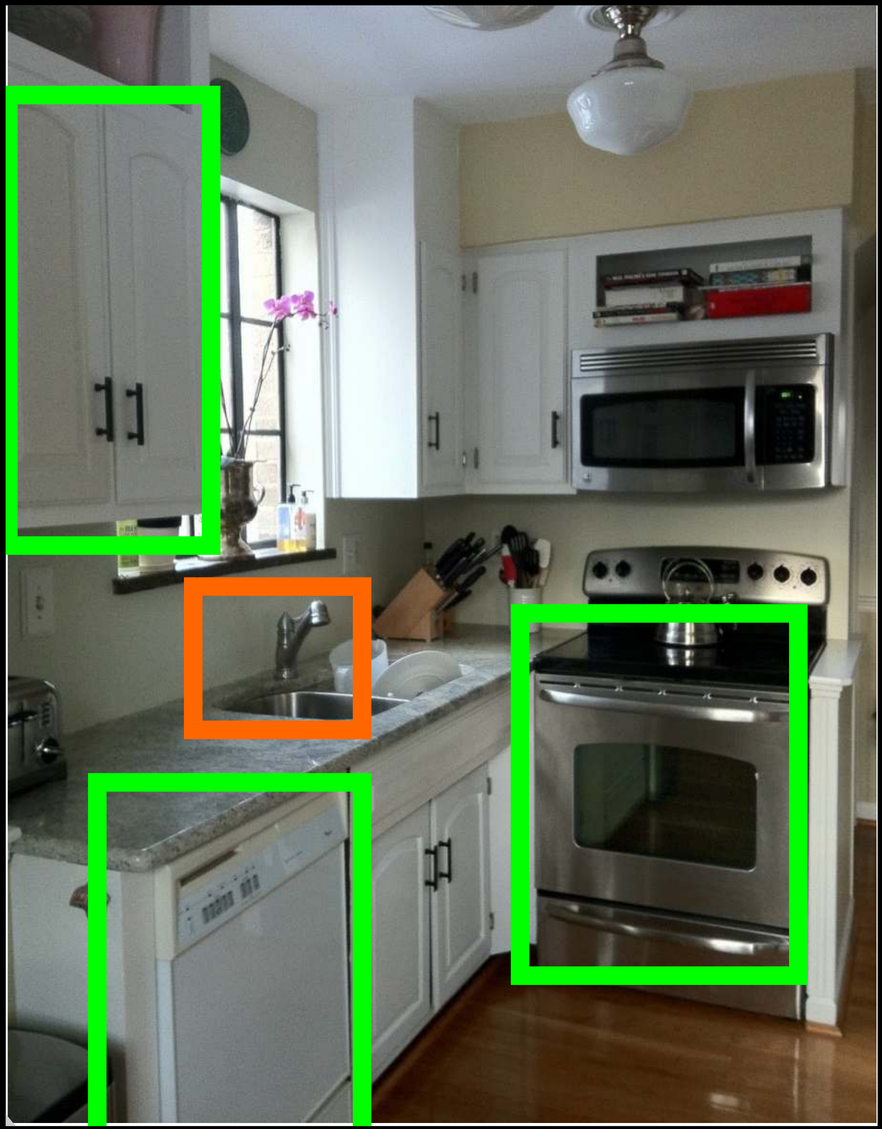
- I’m looking for a label, which I have not seen before. However, this picture contains also:
  - Indoor
  - Living room
  - Table
  - …

- We can classify based on context

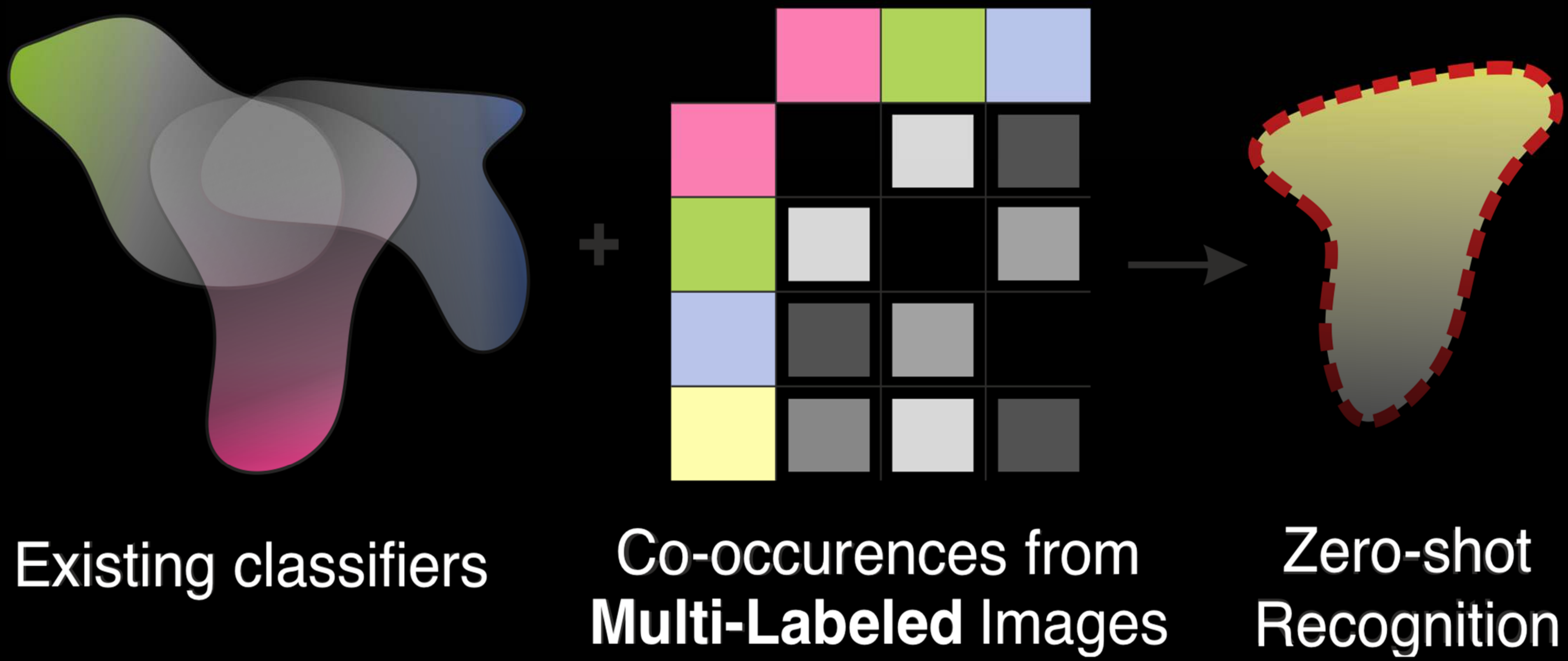
COSTA: Intuition



COSTA: Intuition (2)



COSTA: Design



COSTA: Classifier

- Goal: Estimate classifier $\hat{\boldsymbol{w}}_l$ for unseen label
- Assumption: *k* trained classifiers $\boldsymbol{w}_k \in \mathbb{R}^{d \times 1}$
- Zero-shot classifier:

$$\hat{\boldsymbol{w}}_l = \sum_k \boldsymbol{w}_k \, s_{lk}$$

- Where s_{lk} is based on co-occurrence stats

Co-Occurrence Statistics

How to set a weight s, based on counts c

- Normalized $$s^{\mathrm{n}}_{ij} = \frac{c_{ij}}{c_i}$$
- Binarized $$s^{\mathrm{b}}_{ij} = [\![c_{ij} \geq t]\!]$$
- Burstiness corrected $$s^{\mathrm{s}}_{ij} = \sqrt{c_{ij}}$$
- Dice coefficient $$s^{\mathrm{d}}_{ij} = \frac{c_{ij}}{c_i + c_j}$$

Co-Occurrence Statistics (2)

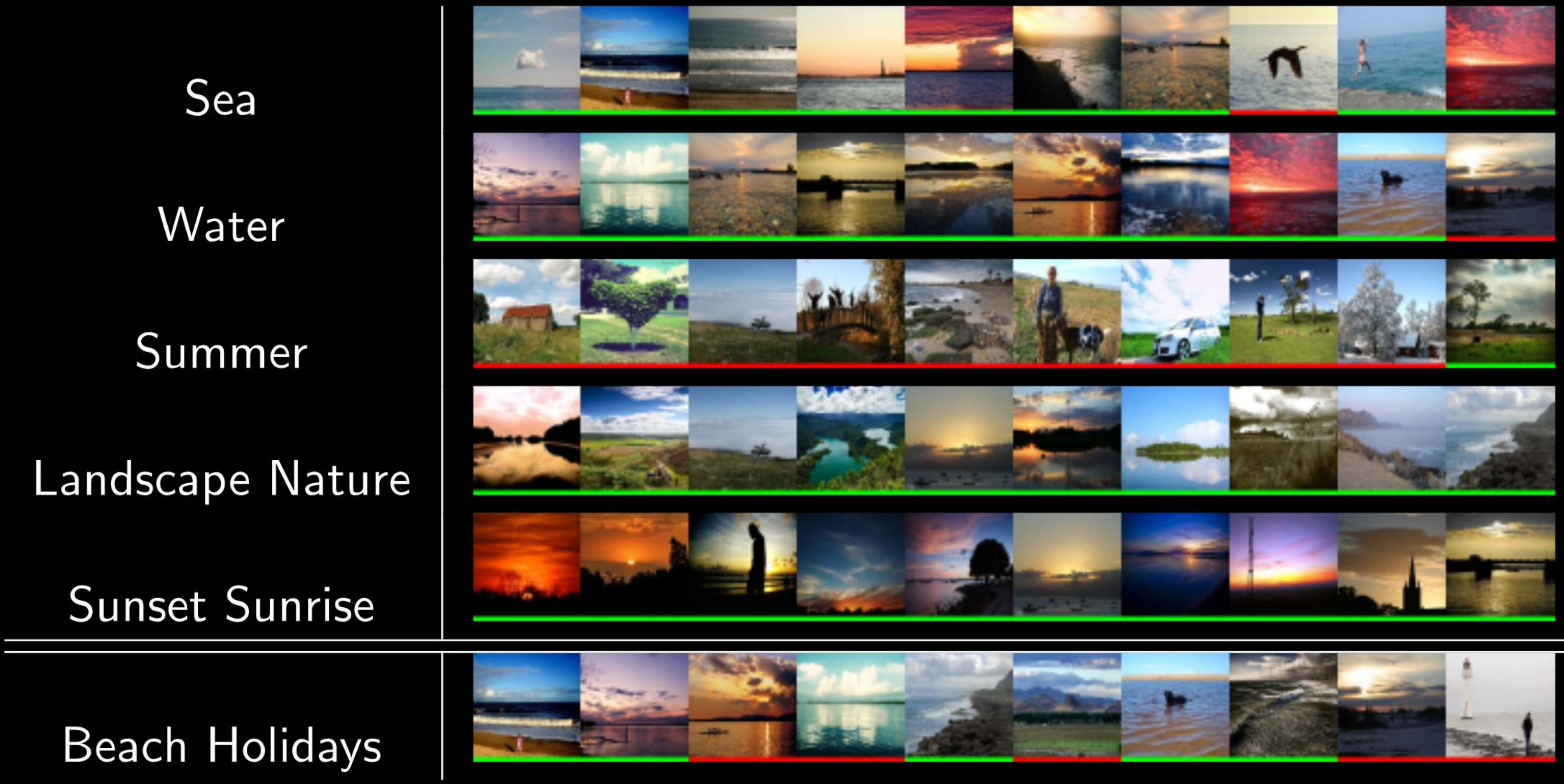
Co-occurrences can be obtained from:

- Ground-truth data (proof-of-concept)
- Web search engines
- Flickr Tags
- Microsoft COCO

Example: Beach Holiday

Concept	Normalized Co-Oc Weight
Sea	0.1810
Water	0.0992
Summer	0.0548
LandscapeNature	0.0435
SunsetSunrise	0.0383
Sports	0.0367
Travel	0.0347
Ship	0.0346
Sunny	0.0319
Big Group	0.0282

Example: Beach Holidays



TWO EXTENSIONS

Defining a concept by what it is not

- Knowing what is not related to a visual concept is informative for its visual scope
- Related: used in image retrieval [Jegou&Chum ECCV 12]
- Example: a car is never* together with a table
- Solution: positive and negative co-occurrences:

$$\hat{\boldsymbol{w}}_l = \sum_k \boldsymbol{w}_k \; s_{lk}^{++} - \boldsymbol{w}_k \; s_{lk}^{+-} - \boldsymbol{w}_k \; s_{lk}^{-+} + \boldsymbol{w}_k \; s_{lk}^{--}$$

* Ok. Never say never, but it is very unlikely

Regression to improve COSTA

- Our problem is estimating a classifier:

$$\hat{\boldsymbol{w}}_l = \sum_k \boldsymbol{w}_k \, s_{lk}$$

- Objective: the estimated classifier should be as close as possible to the learned classifier if we would have visual labels.

Regression to improve COSTA (2)

- Idea: learn a weight a_k per classifier

$$\hat{\boldsymbol{w}}_l = \sum_k a_k \, \boldsymbol{w}_k \, s_{lk}$$

- Note: Weights are independent of novel class
- Solve: Regression objective

$$L_{\text{reg}} = \sum_i \|\boldsymbol{w}_i - \sum_k a_k \boldsymbol{w}_k \, s_{ik}\|_2^2$$

- Train: Using a leave-one-out setting over train classes

EXPERIMENTS

Experimental setup

- Hierarchical SUN dataset [Choi et al. CVPR'10]
  - 107 Labels
  - 4367 train 4317 test images
  - 5.34 labels per image
- Fisher Vectors (3096 dim)
- SVMs with 2 fold cross-validation

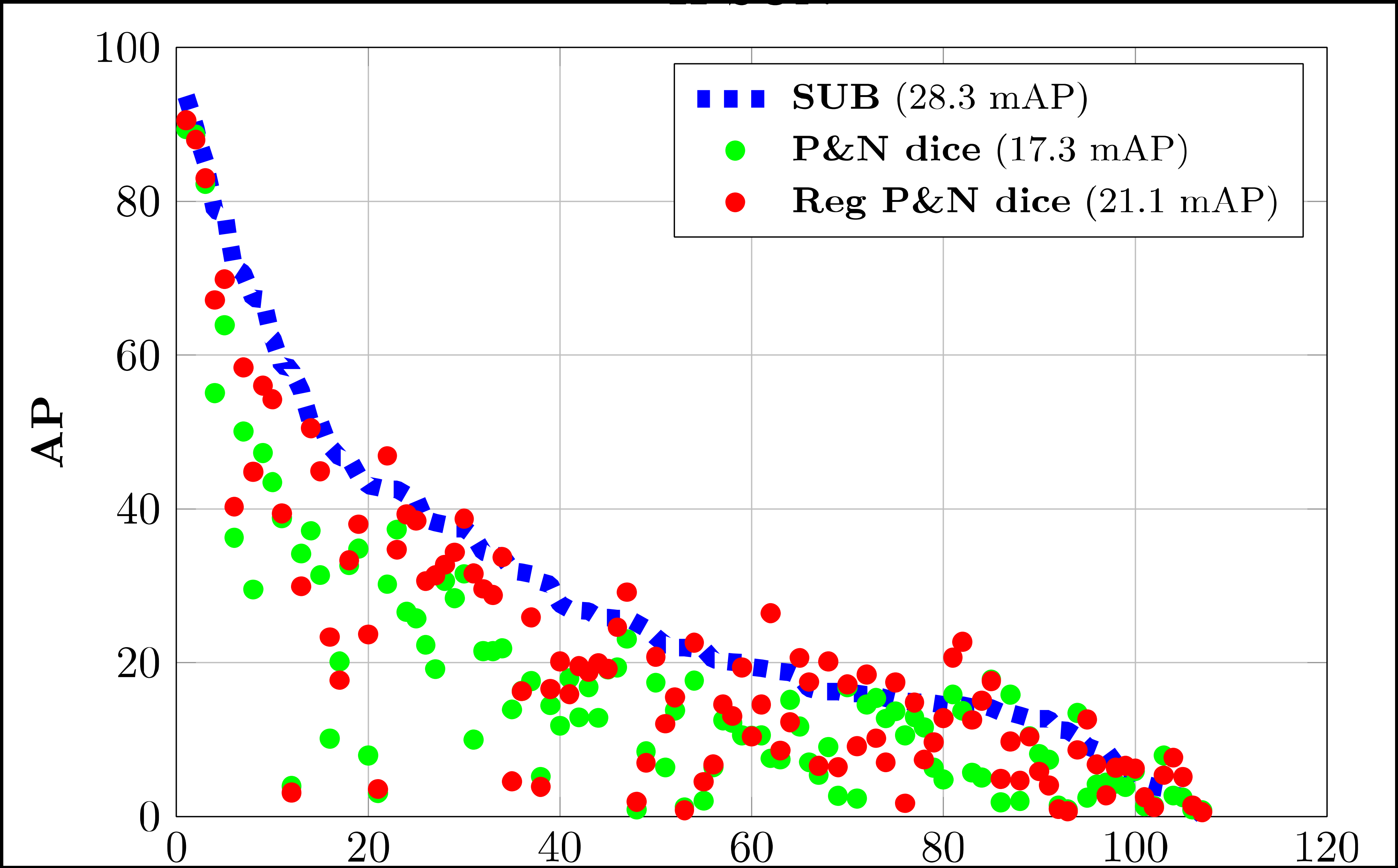
- In paper also experiments on:
  - ImageCLEF'10 and CUB-Attributes

Multi-label Zero-Shot Classification

All methods are evaluated on a subset of 25% of the labels.

	H-SUN			
Setting	SUB	L1O	ZS75	ZS50
Nr. Train labels	107	106	81	54
Baselines				
Supervised SVM	21.5	–	–	–
Attributes, following [1]	–	12.8	13.0	12.3
COSTA				
Co-oc Dice	–	14.5	14.5	12.9
P&N Dice	–	13.7	13.8	10.8
Reg P&N Dice	–	17.0	16.4	15.0

AP per Concept



Co-occurrences from the Web

	Setting	SUB	L1O	ZS75	ZS50
H-SUN	**Label Annotations**				
	SUB	**21.5**	-	-	-
	Label Co-oc	-	**17.0**	**16.4**	**15.0**
	Internet search				
	Web hit counts	-	9.9	9.8	9.8
	Image hit counts	-	12.7	9.1	9.3
	Flickr hit counts	-	**15.1**	**13.4**	**10.1**

Ok. But?

How about DeepNets?

- Related works: DeViSe and CONSe
  - Very similar to COSTA, few differences
  - Predict 1000 ImageNet Classes
  - Measure relatedness by Word2Vec

- Preliminary result: co-occurrences capture visual semantics better than Word2Vec

Failure mode(s)?

- Fine-grained classification:
  - Co-occurrences are not sufficient to distinguish:

Italian Sparrow

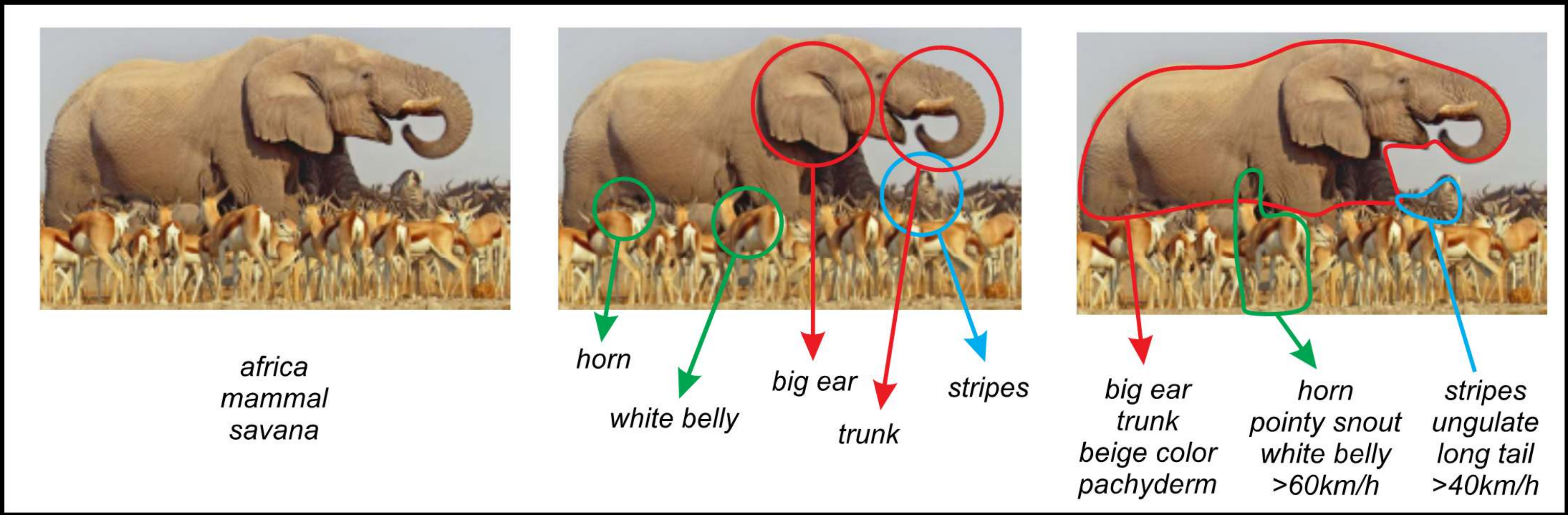
Great Sparrow

Failure mode(s)?

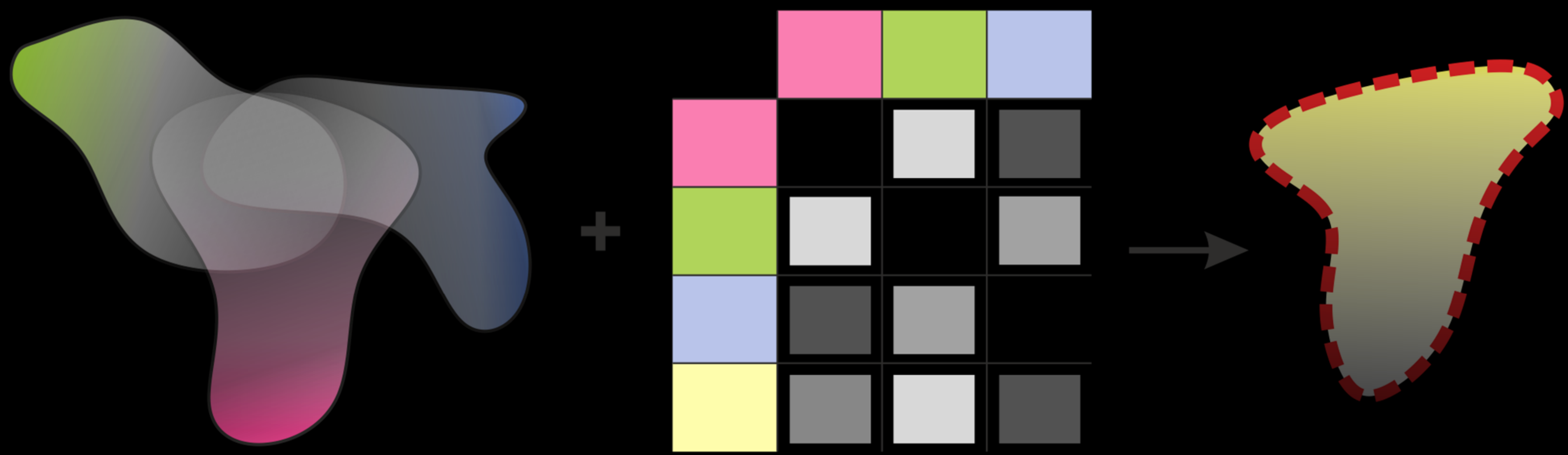
- Fine-grained classification:

  Attributes make sense on segmented objects

  Z. Li, E. Gavves, T. Mensink, and C.G.M. Snoek , ECCV 2014



Conclusion: COSTA



- First method designed for multi-label zero-shot
- Many visual concepts can be described as an open set of concept-to-concept relations
- Describe latent image semantics with co-occurrences
- Exploit natural bias in natural images

COSTA: Co-occurrence statistics for zero-shot classification

T. Mensink, E. Gavves, and C.G.M. Snoek , CVPR 2014