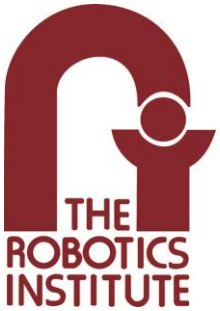
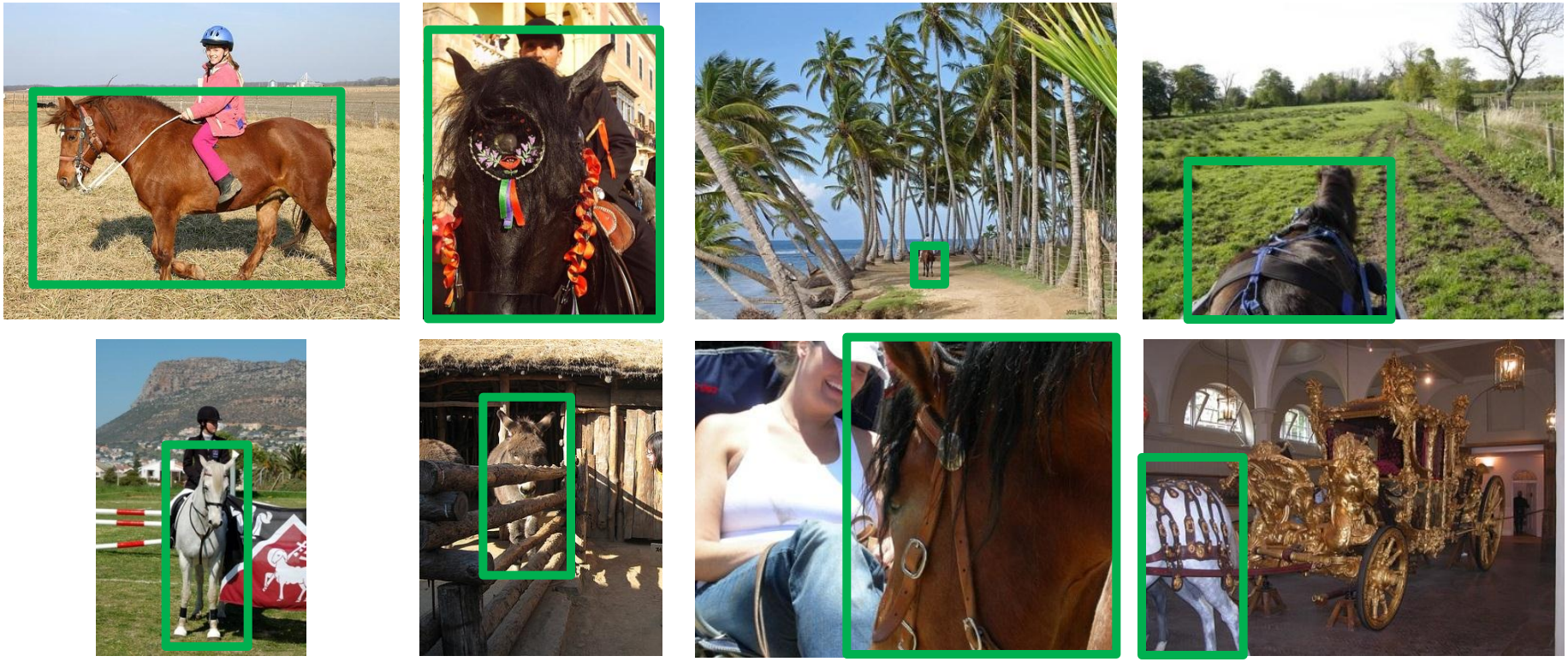
How important are “Deformable Parts” in the Deformable Parts Model?

Santosh K. Divvala
Alexei A. Efros
Martial Hebert



Intra Category Diversity



Example images for “Horse” from PASCAL VOC

Variation due to change in camera viewpoint, object pose, and occlusion

State-of-the-art Detector: Deformable Parts Model (DPM)

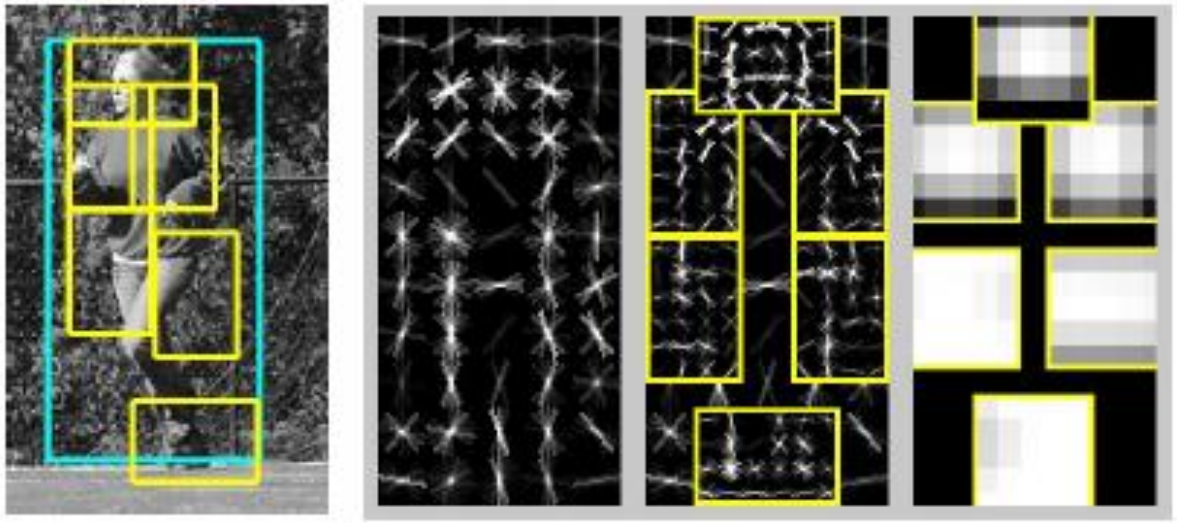
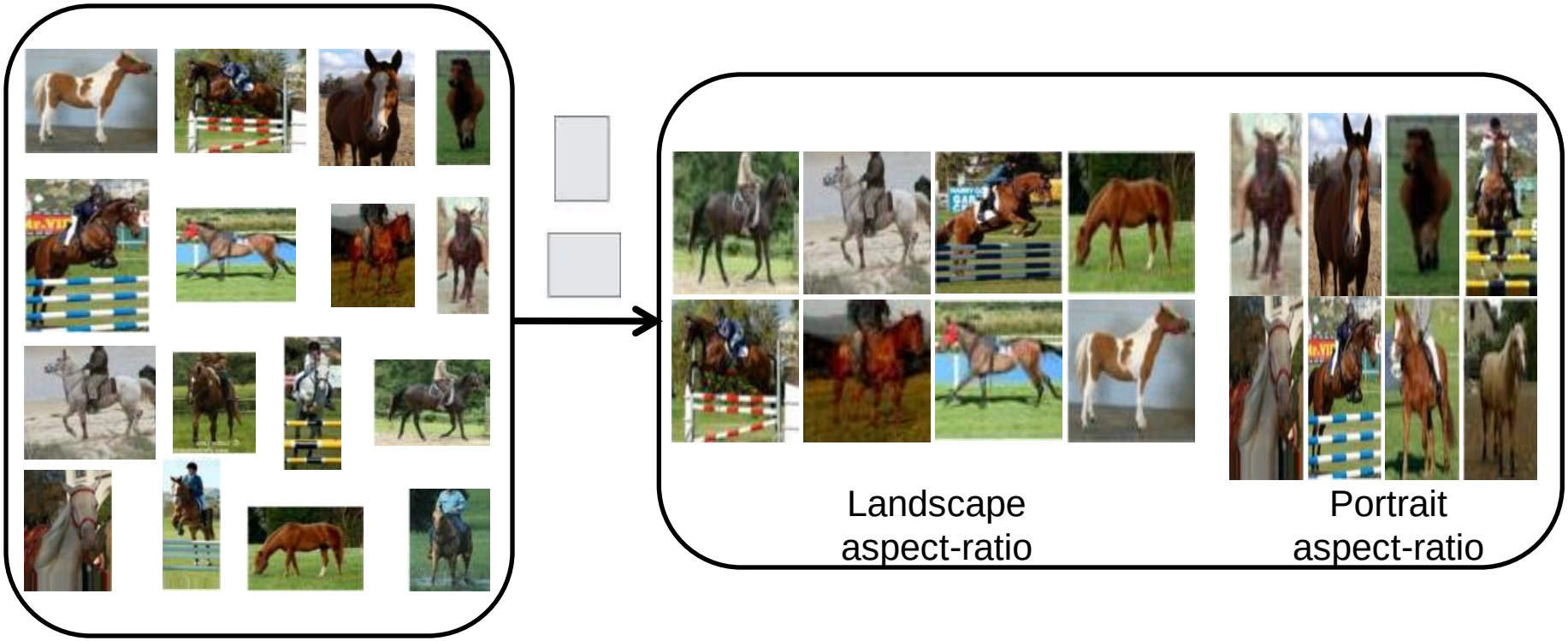


Image | Root filter (Coarse resolution) | Part filters (Fine resolution) | Deformation Models

Felzenszwalb et al., 2008, 2010, 2011

Mixture Model ("Subcategories")

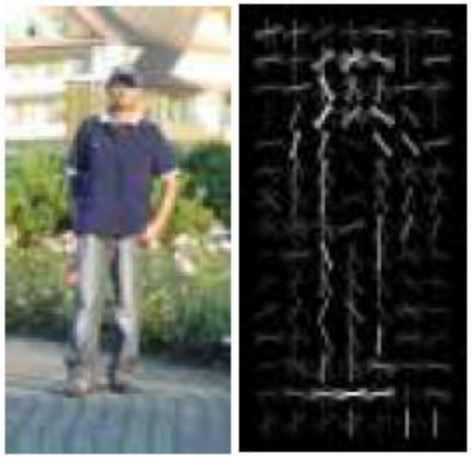


Felzenszwalb et al., 2010

The Evolution of DPM

The Evolution of DPM

K=1
no Parts
[DT'05]



HOG (Histogram of Oriented Gradients)

The Evolution of DPM

K=1 no Parts [DT'05]
mAP = 0.17

Parts vs. Subcategories: Performance

K=1
6 Parts
[PFF'08]

K=1
no Parts
[DT'05]

mAP = 0.17

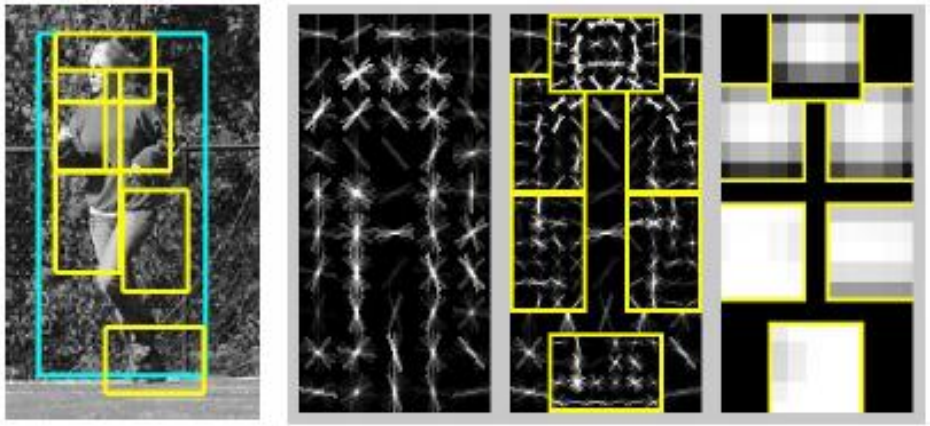


Image | Root filter (Coarse resolution) | Part filters (Fine resolution) | Deformation Models

Parts vs. Subcategories: Performance

K=1 6 Parts [PFF'08]
mAP = 0.21

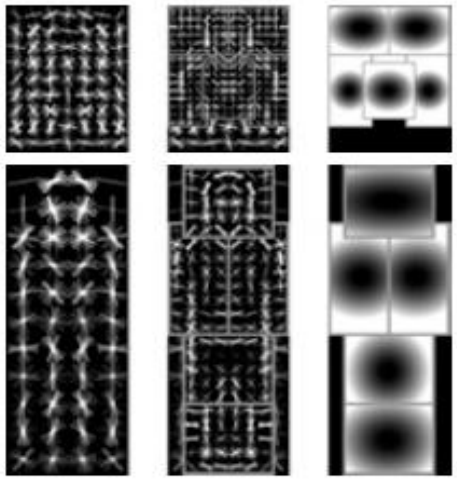
K=1 no Parts [DT'05]
mAP = 0.17

Parts vs. Subcategories: Performance

K=1 no Parts [DT'05]
mAP = 0.17

K=1 6 Parts [PFF'08]
mAP = 0.21

K=2 (ar) 6 Parts [PFF'10]



2 DPM (PAMI10)

Parts vs. Subcategories: Performance

K=1 6 Parts [PFF'08]
mAP = 0.21

K=2 (ar) 6 Parts [PFF'10]
mAP = 0.26

K=1 no Parts [DT'05]
mAP = 0.17

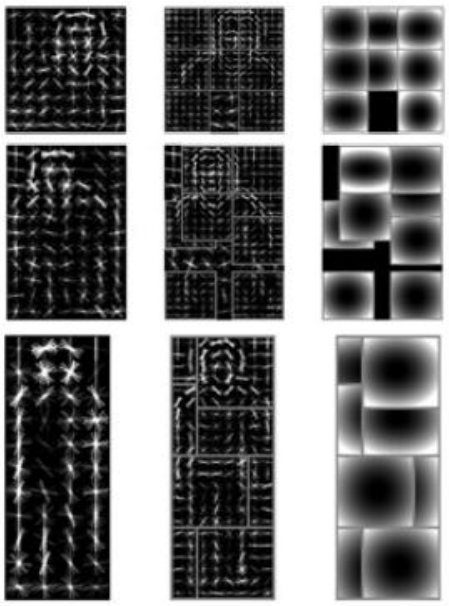
Parts vs. Subcategories: Performance

K=1 no Parts [DT'05]
mAP = 0.17

K=1 6 Parts [PFF'08]
mAP = 0.21

K=2 (ar) 6 Parts [PFF'10]
mAP = 0.26

K=6 (ar) 8 Parts [PFF'11]



6 DPM (voc-release4)

Parts vs. Subcategories: Performance

K=1 6 Parts [PFF'08]	K=2 (ar) 6 Parts [PFF'10]	K=6 (ar) 8 Parts [PFF'11]
mAP = 0.21	mAP = 0.26	mAP = 0.32

K=1 no Parts [DT'05]
mAP = 0.17

Parts vs. Subcategories: Performance

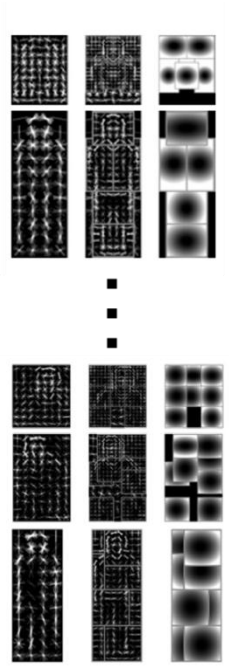
K=1 no Parts [DT'05]
mAP = 0.17

K=1 6 Parts [PFF'08]
mAP = 0.21

K=2 (ar) 6 Parts [PFF'10]
mAP = 0.26

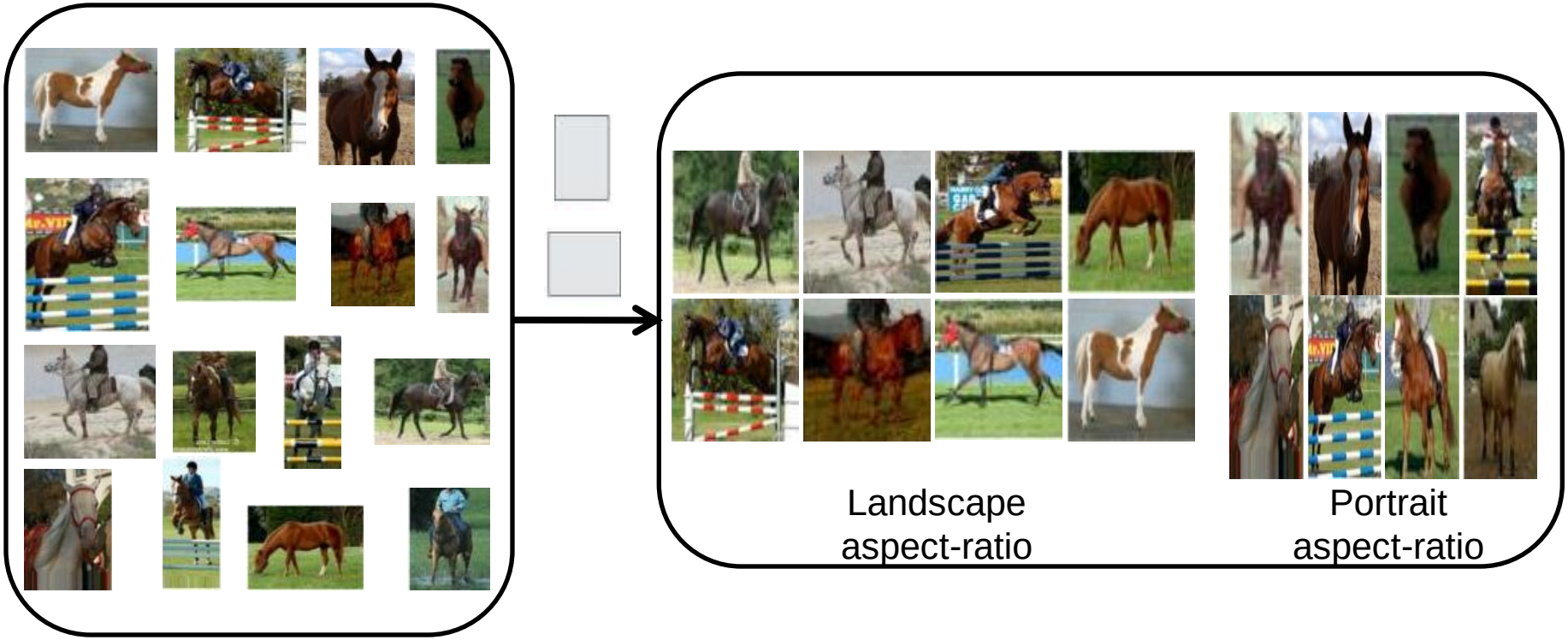
K=6 (ar) 8 Parts [PFF'11]
mAP = 0.32

K=15 (ar) 8 Parts [This work]
mAP = ?



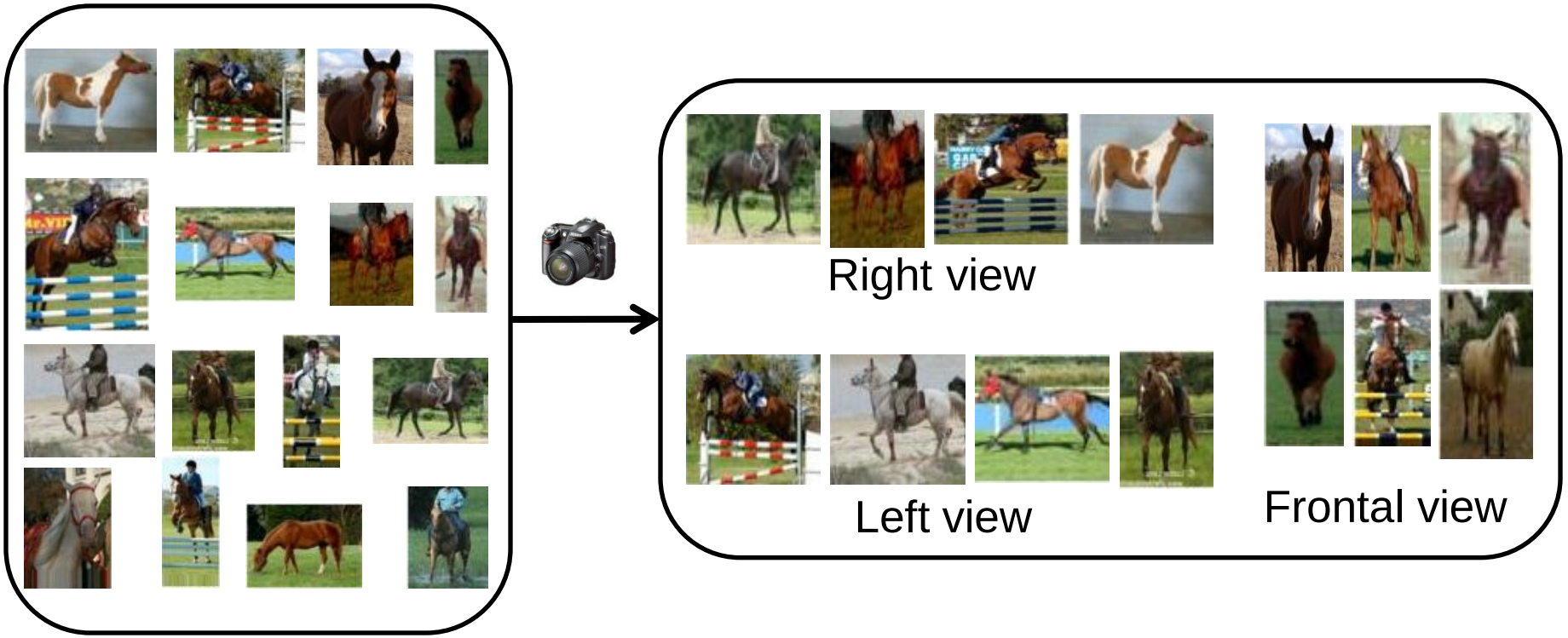
15 DPM

Aspect Ratio Subcategories



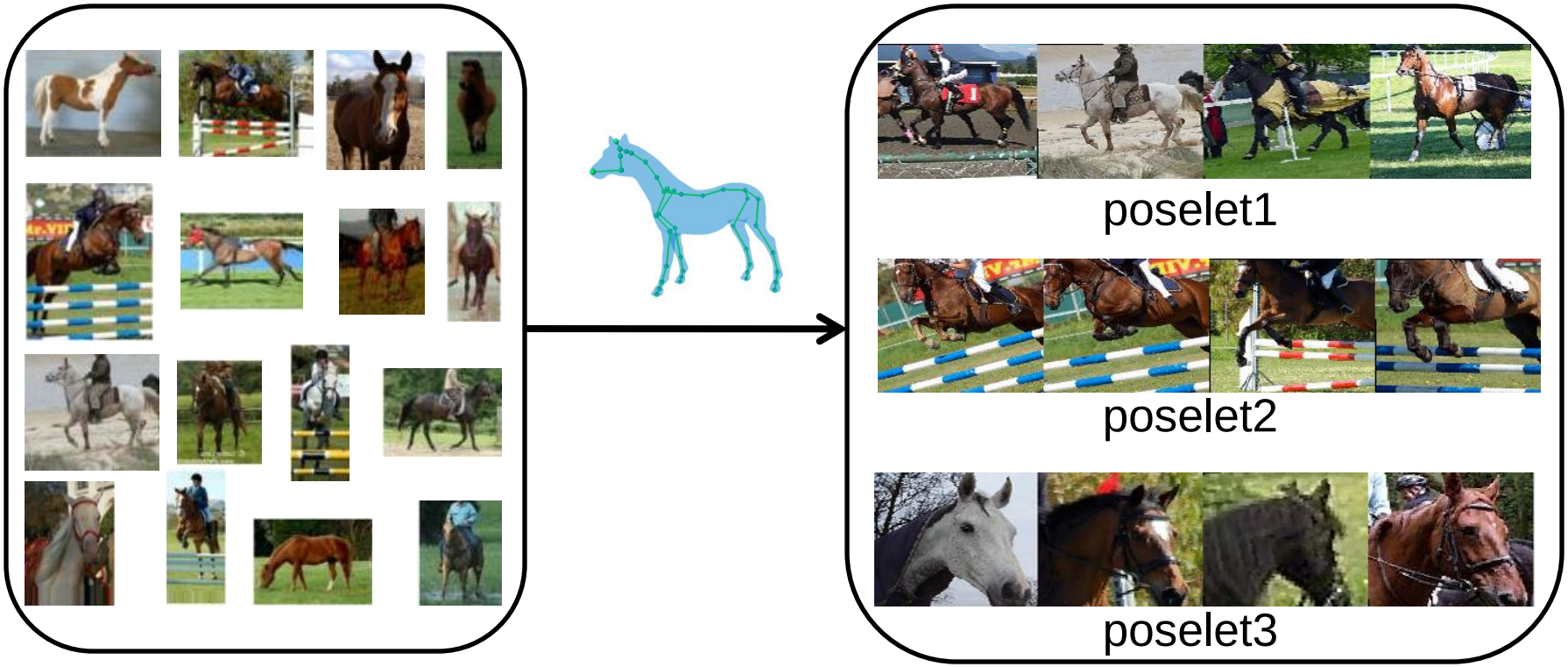
Felzenszwalb et al., 2010

Viewpoint Subcategories



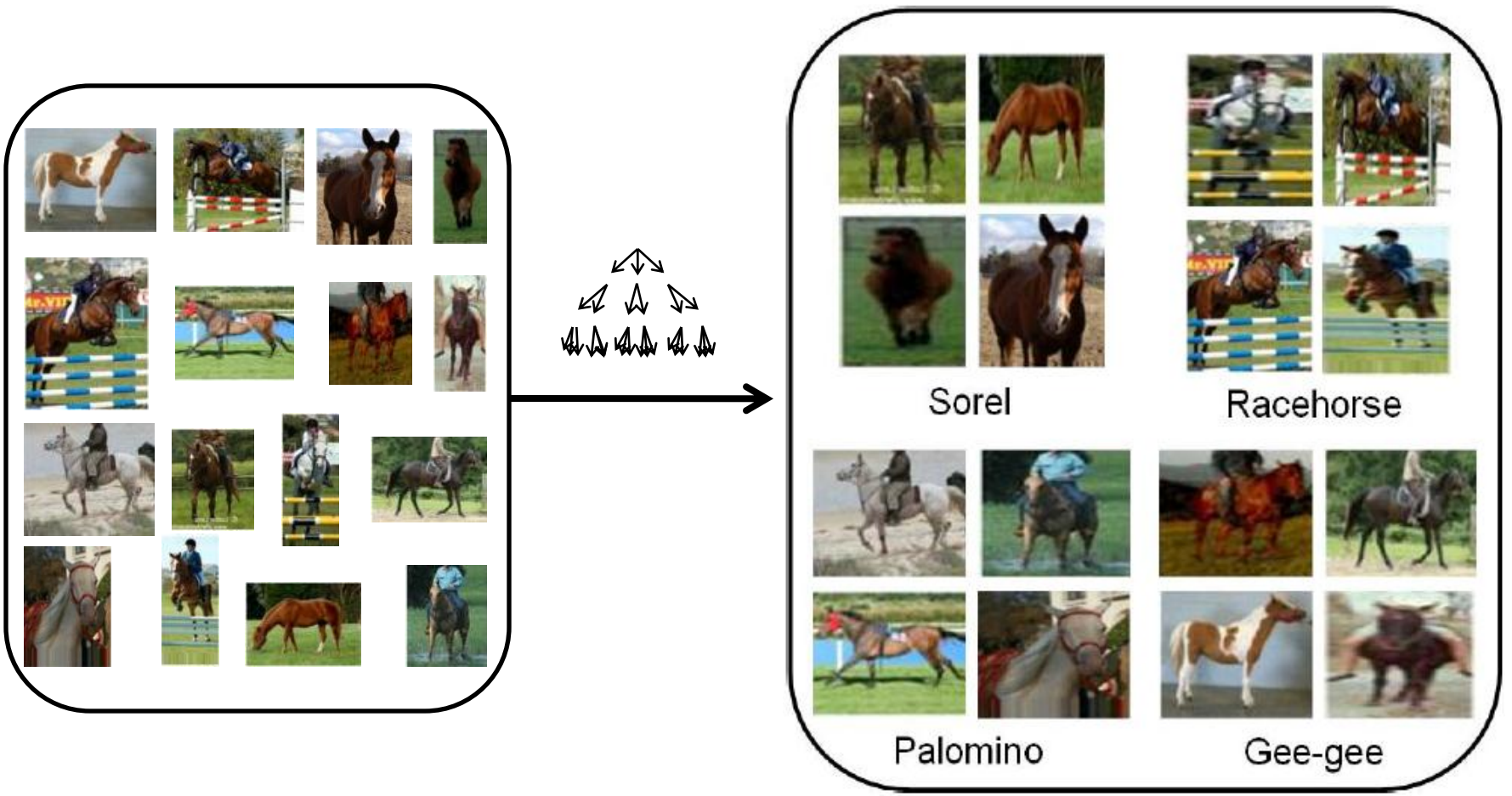
Chum & Zisserman 2007, Harzallah and Schmid 2008

3D Configuration Subcategories



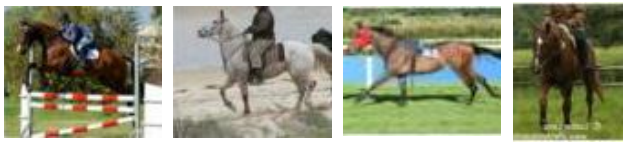
Bourdev & Malik, 2009

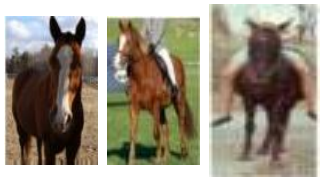
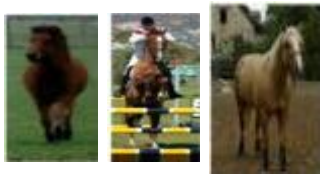
Taxonomy Subcategories



"ImageNet", Deng et al., 2009


Right

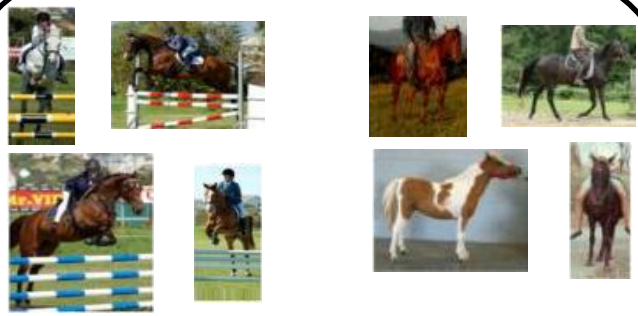

Left



Front

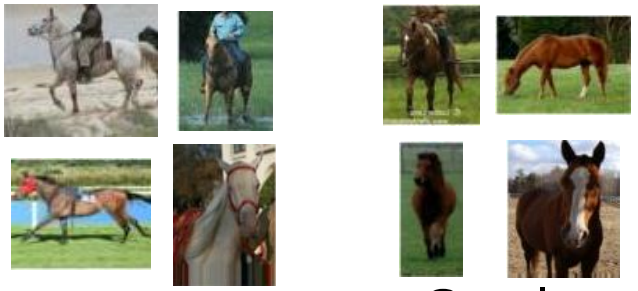

poselet1


poselet2


poselet3


Racehorse

Gee-gee


Palomino

Sorel


Landscape


Portrait

Similarity in Appearance!



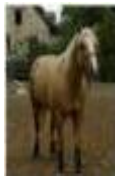

Right




Left





Front

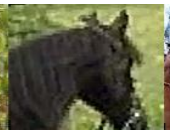



poselet1




poselet2






poselet3



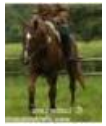

Racehorse




Gee-gee




Palomino

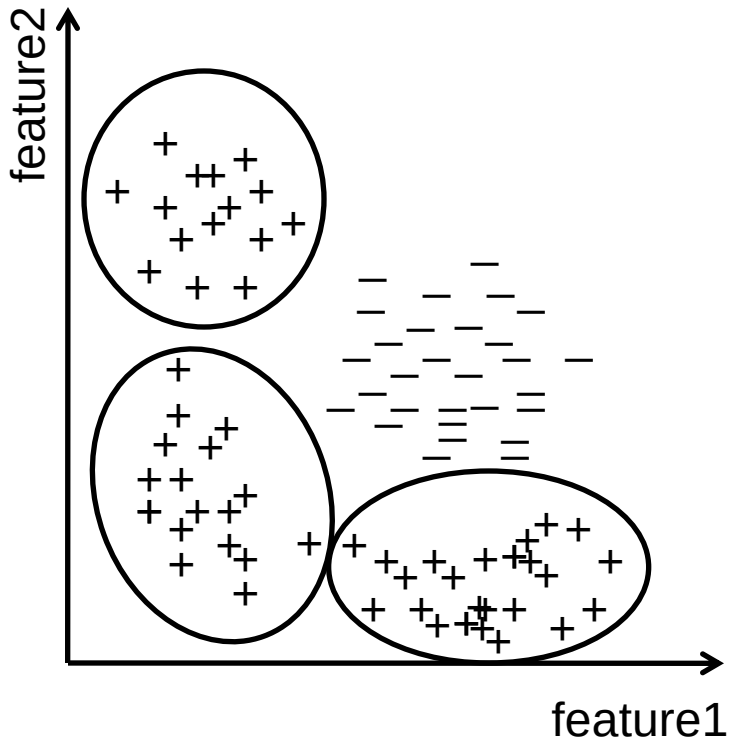



Sorel

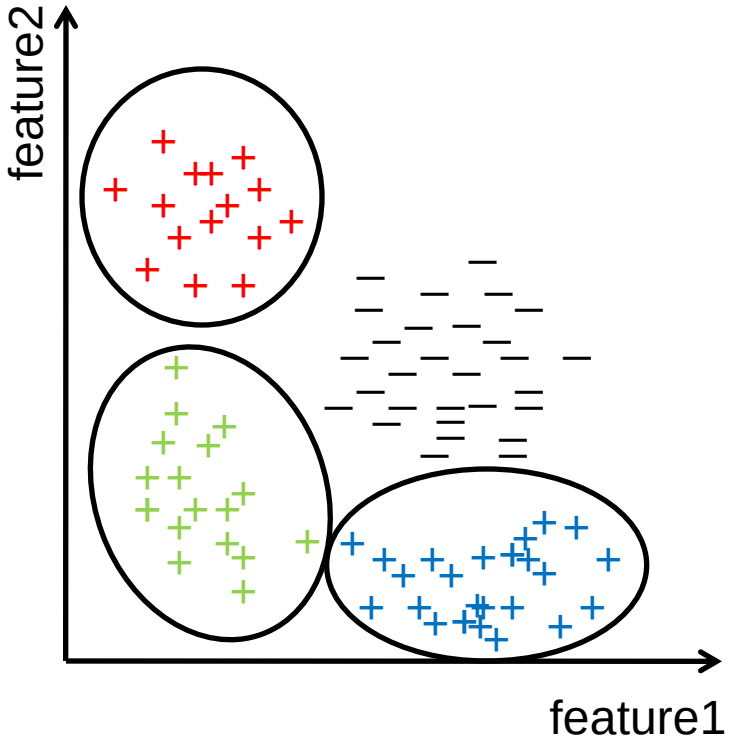
Aspect 1

Aspect 2

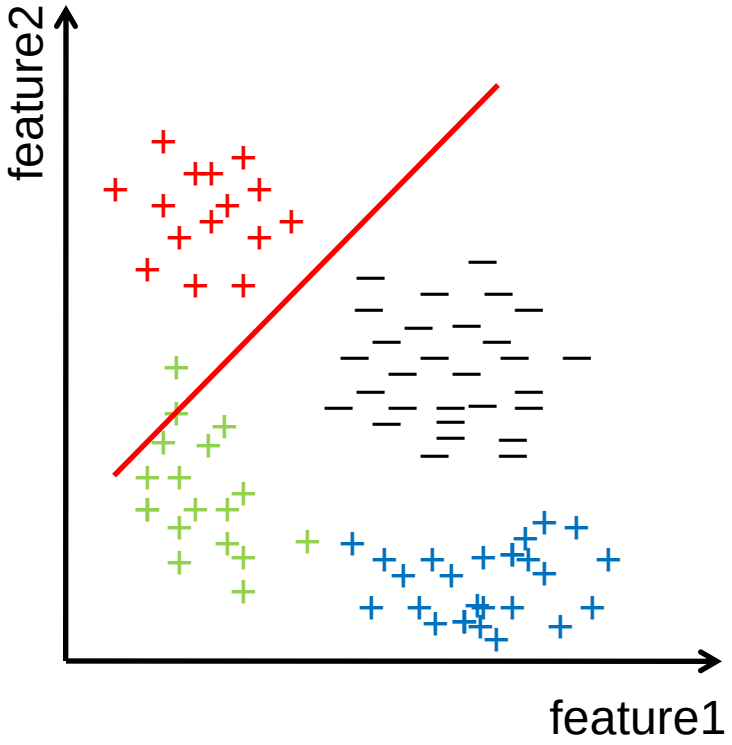
Subcategories



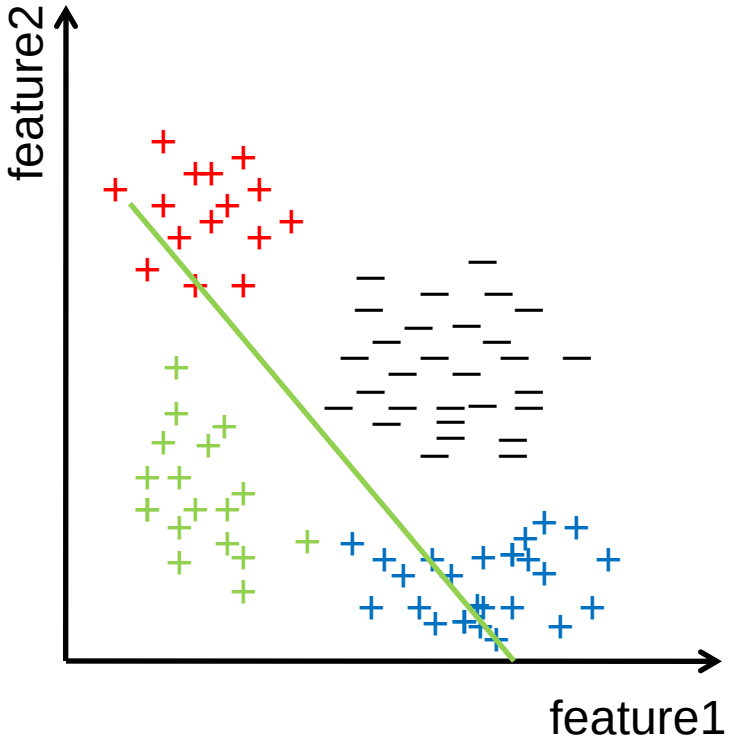
Subcategories



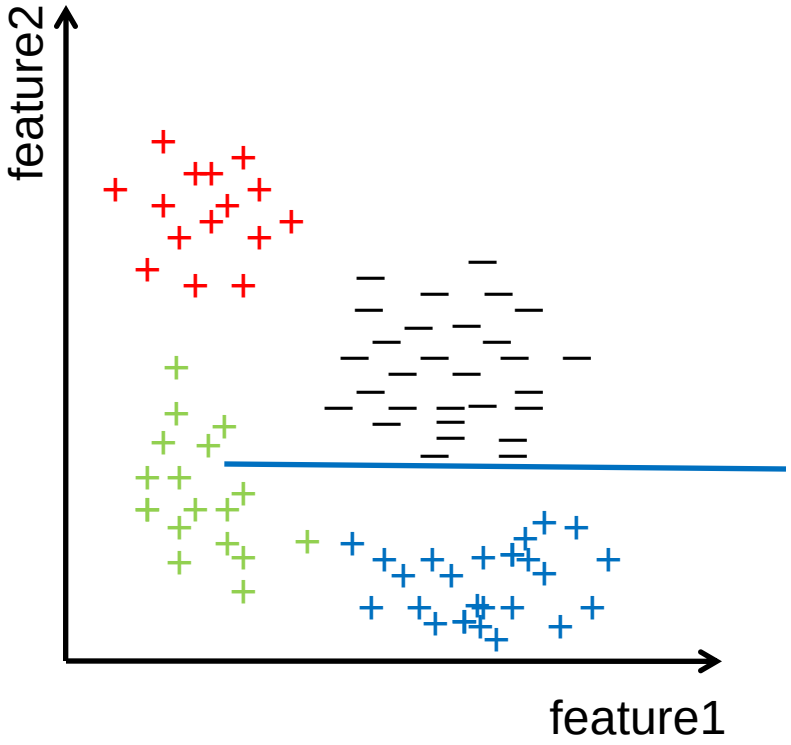
Subcategories



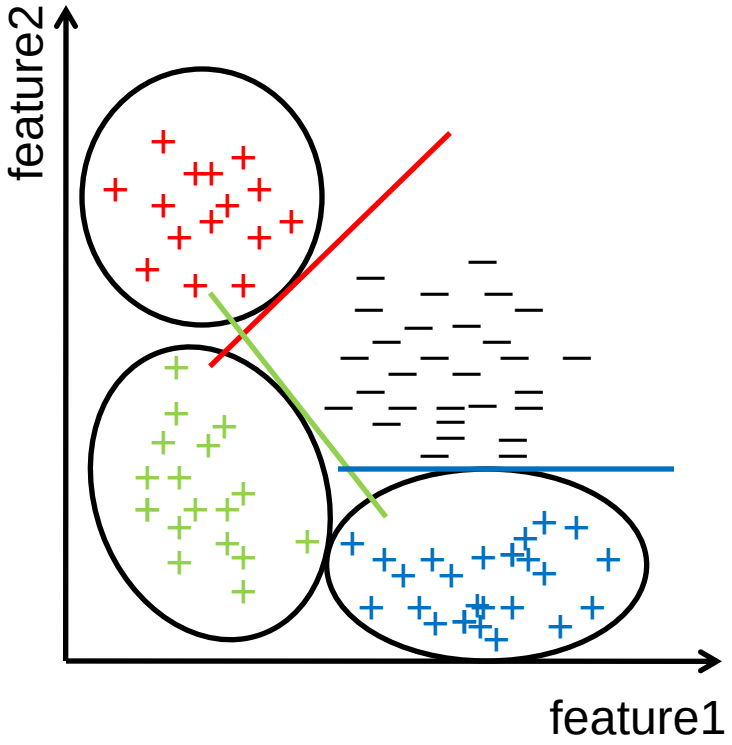
Subcategories



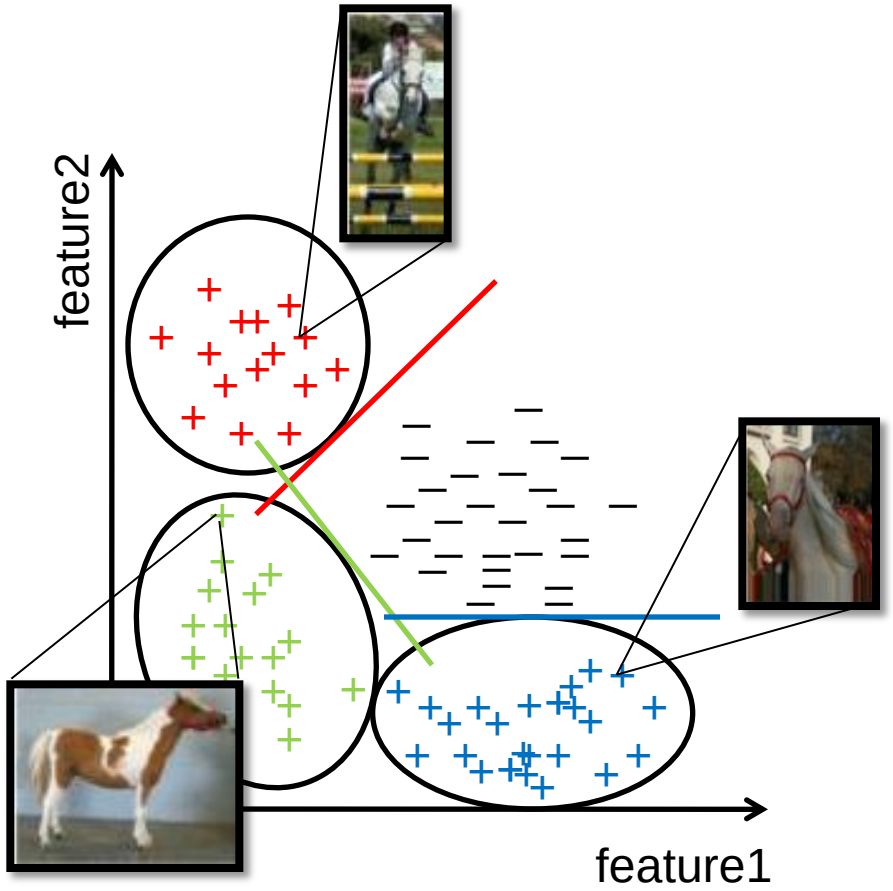
Subcategories



Subcategories



Subcategories



Parts vs. Subcategories: Performance

K=1 no Parts [DT'05]
mAP = 0.17

K=1 6 Parts [PFF'08]
mAP = 0.21

K=2 (ar) 6 Parts [PFF'10]
mAP = 0.26

K=6 (ar) 8 Parts [PFF'11]
mAP = 0.32

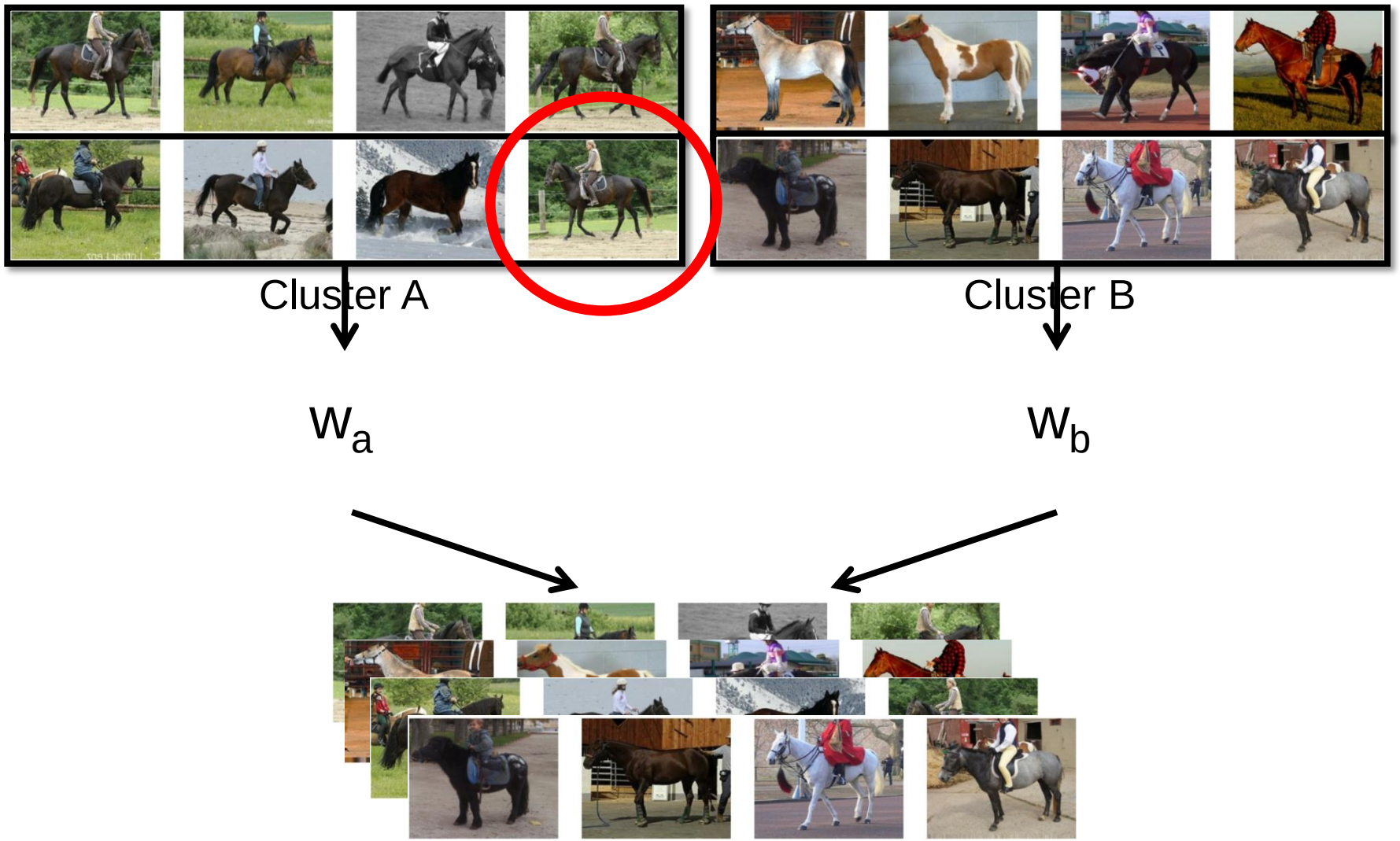
K=15 (app) 8 Parts [This work]

1. Initializing Subcategories

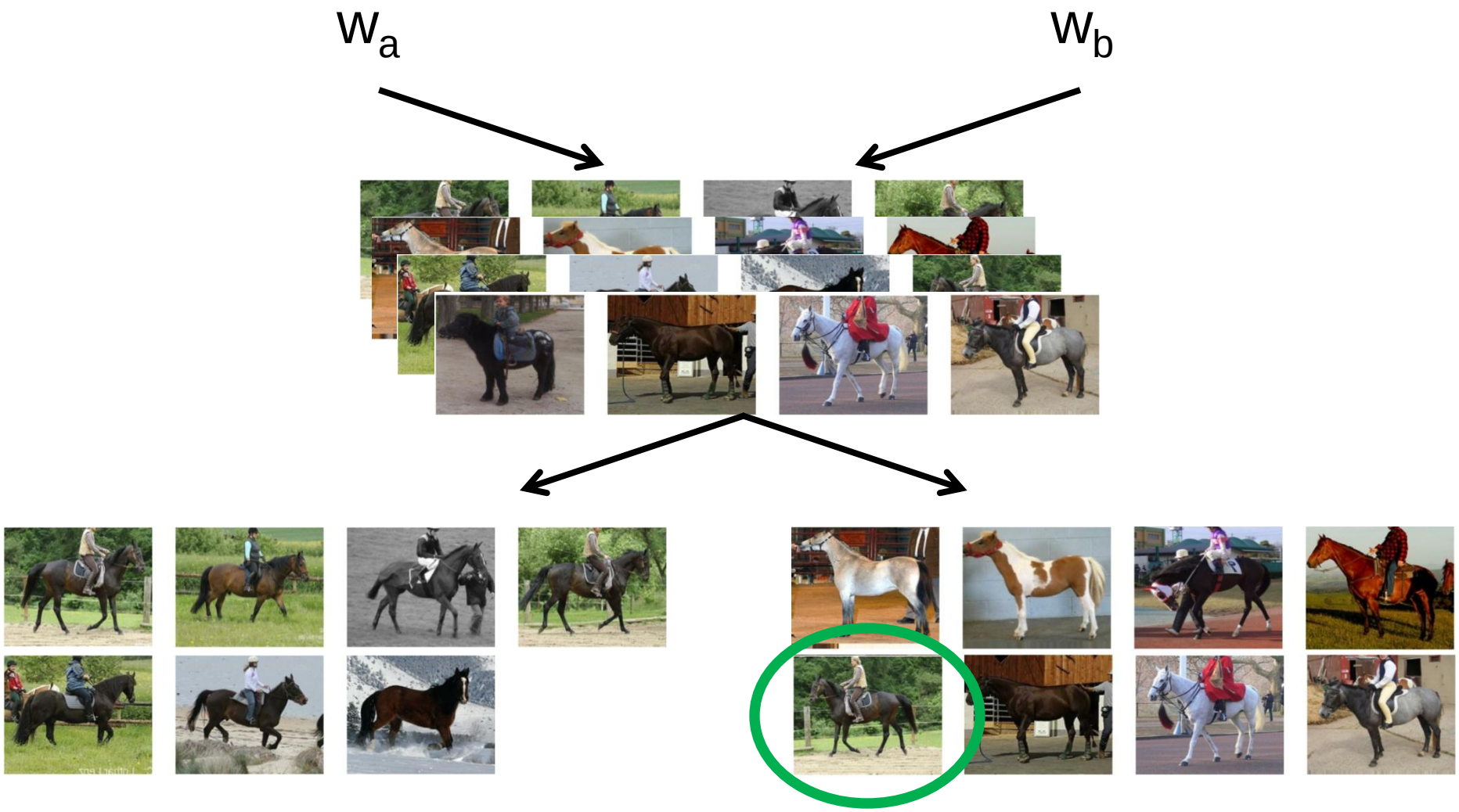
All ground-truth ‘horse’ instances

K-means
feature-space
clustering

Latent Re-clustering



Latent Re-clustering



2. Number of Subcategories

- Use a large K to accommodate intra category diversity

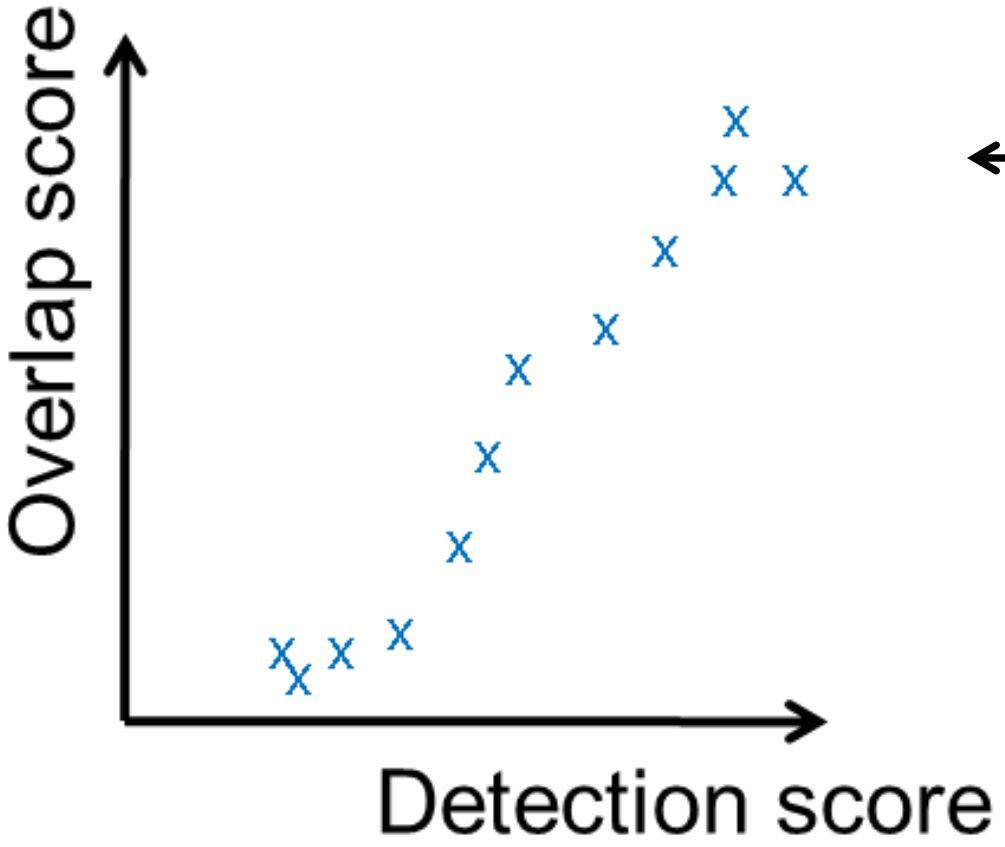
3. Calibration

- K varies with the object category

- Suppress the influence of “noisy” clusters
  - Defined based on the performance on the validation set

Calibration: Suppressing noisy clusters

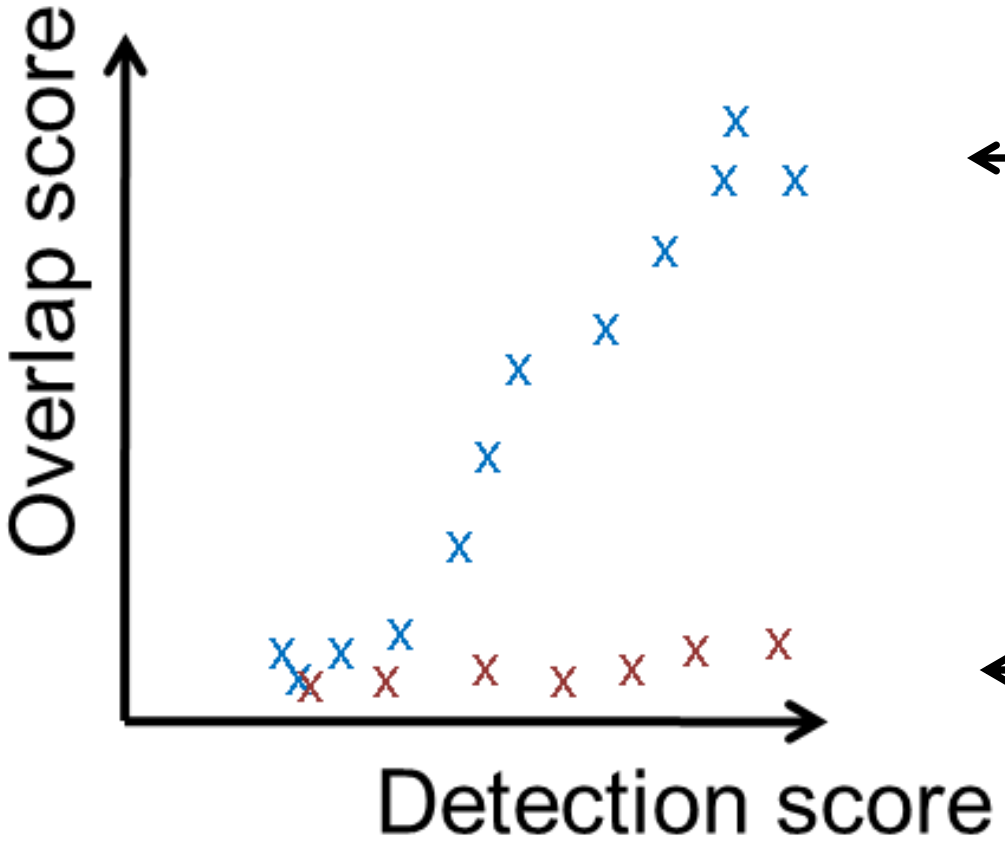
- Relate detection score to overlap score



"Good" Subcategory

Calibration: Suppressing noisy clusters

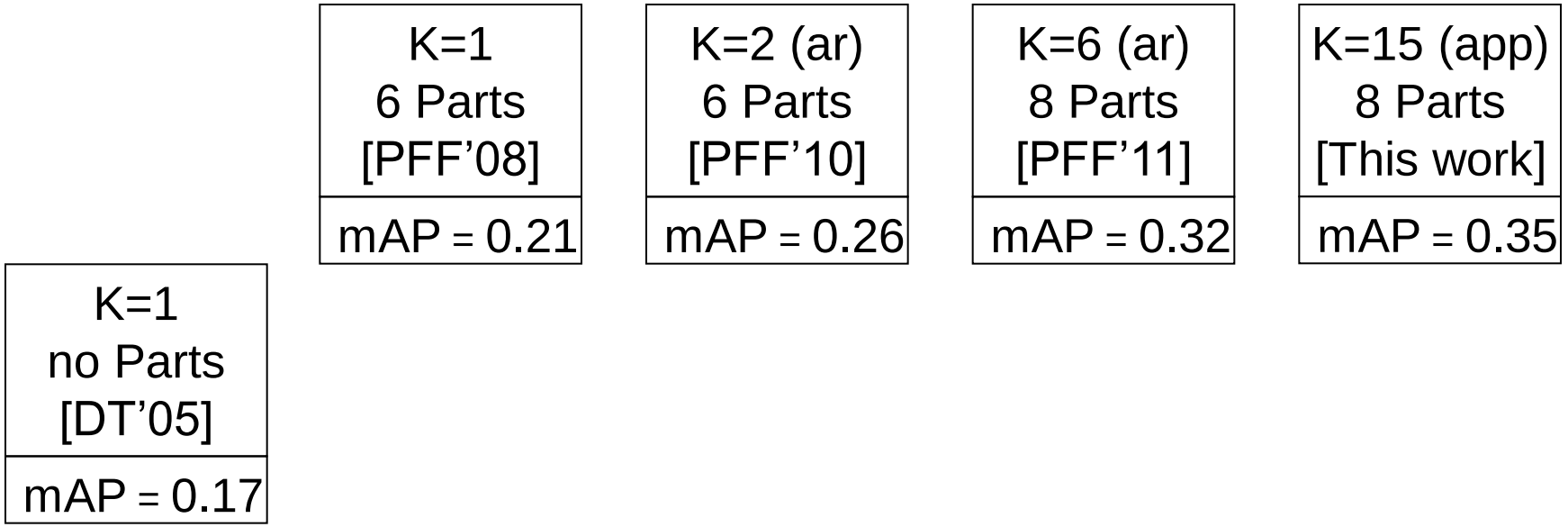
- Relate detection score to overlap score



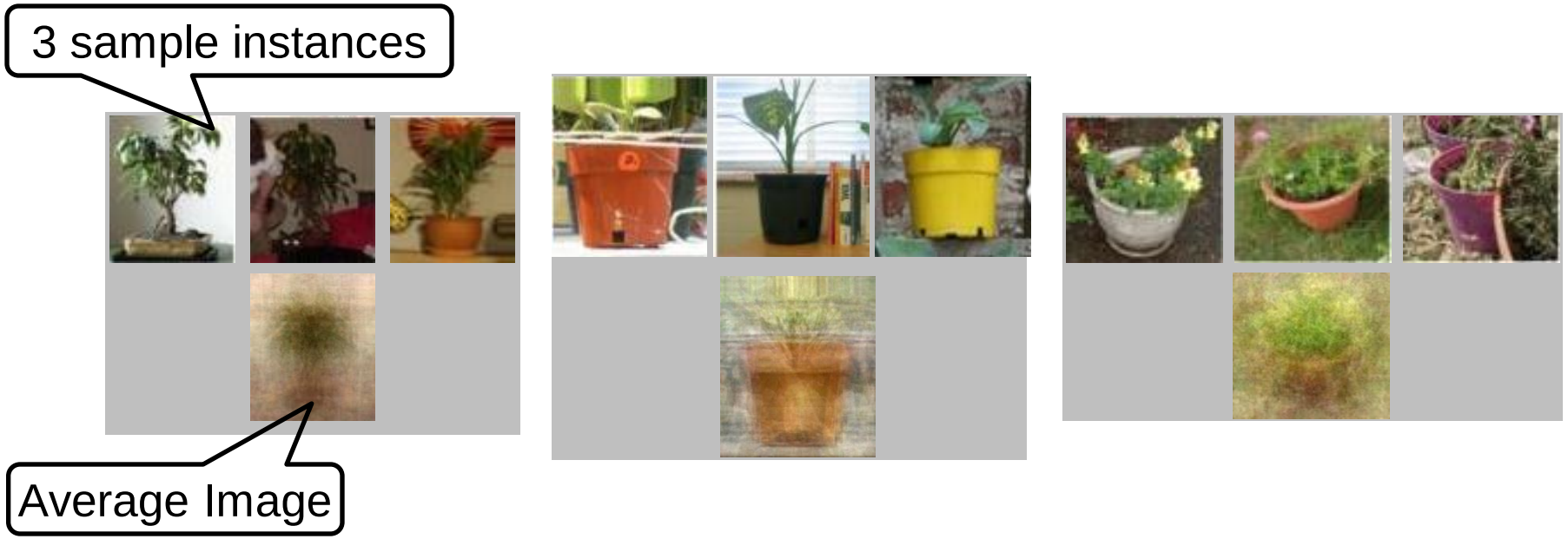
"Good" Subcategory

"Noisy" Subcategory

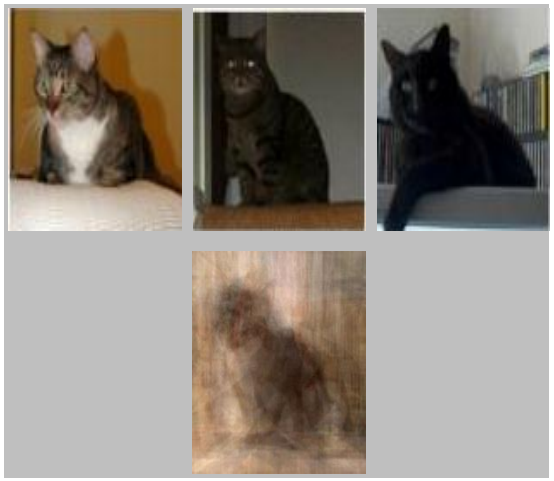
Parts vs. Subcategories: Performance

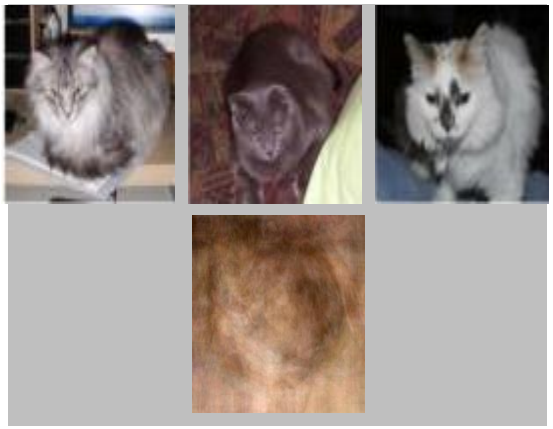


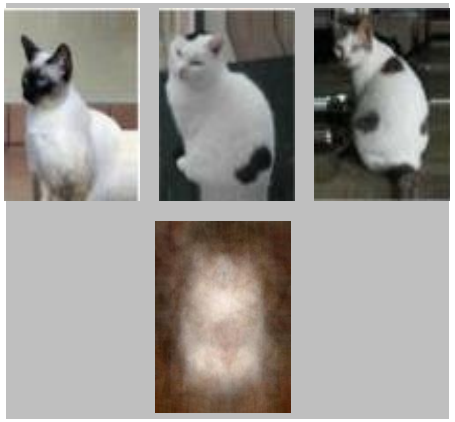
Visual Subcategories: ‘Pottedplant’

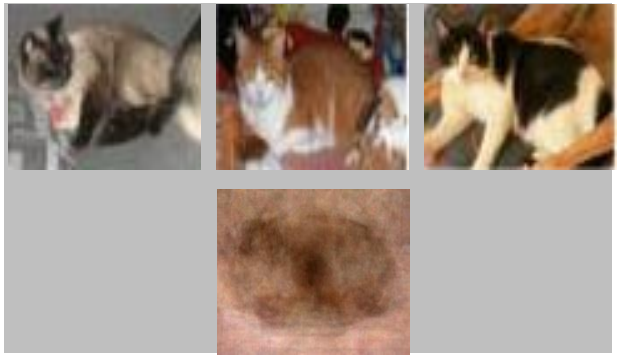


Visual Subcategories: ‘Cat’

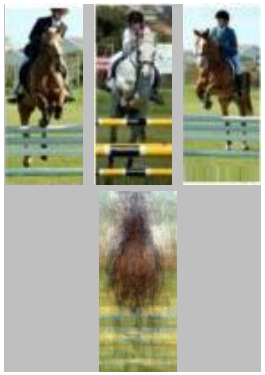


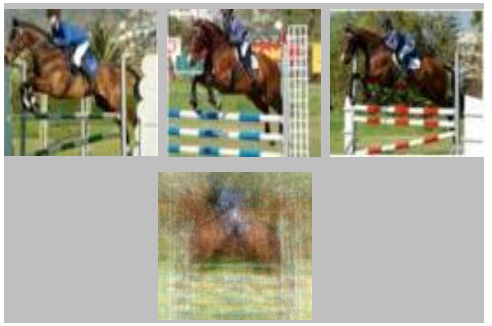


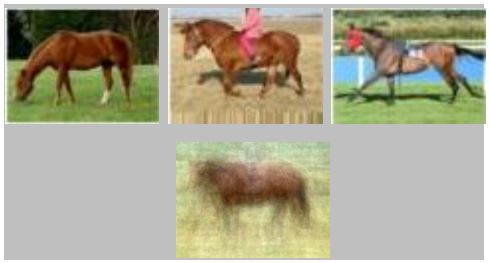


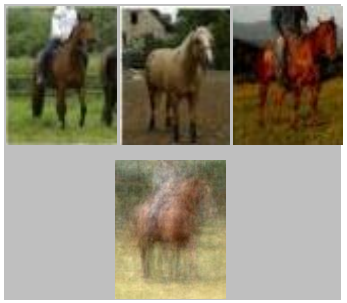


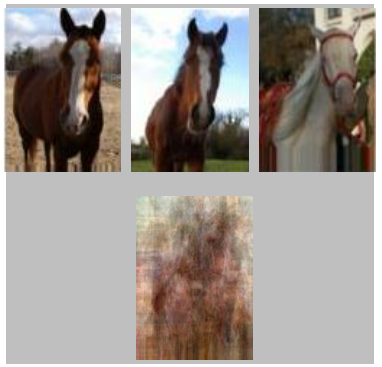
Visual Subcategories: ‘Horse’











Top Detection Results: ‘Train’

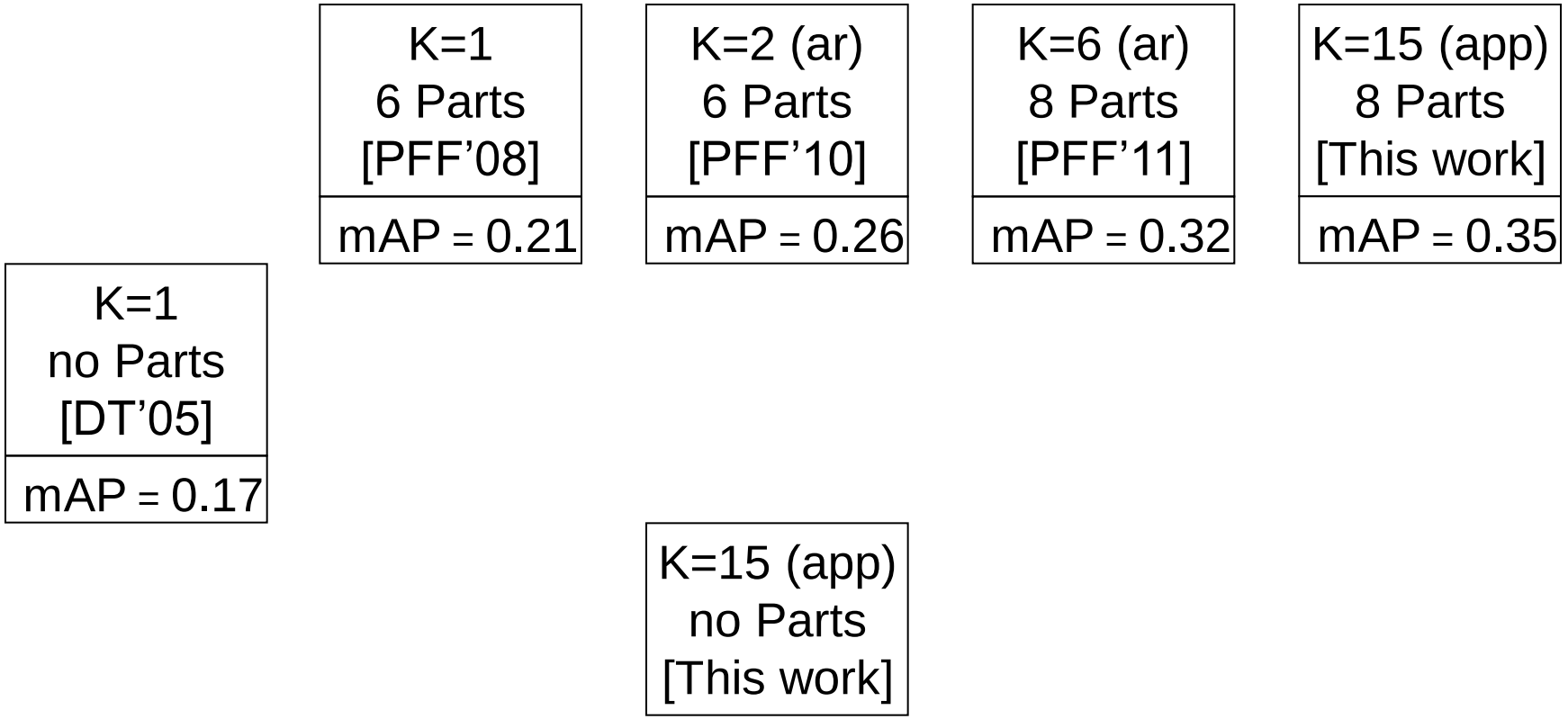
Cluster1

Cluster2

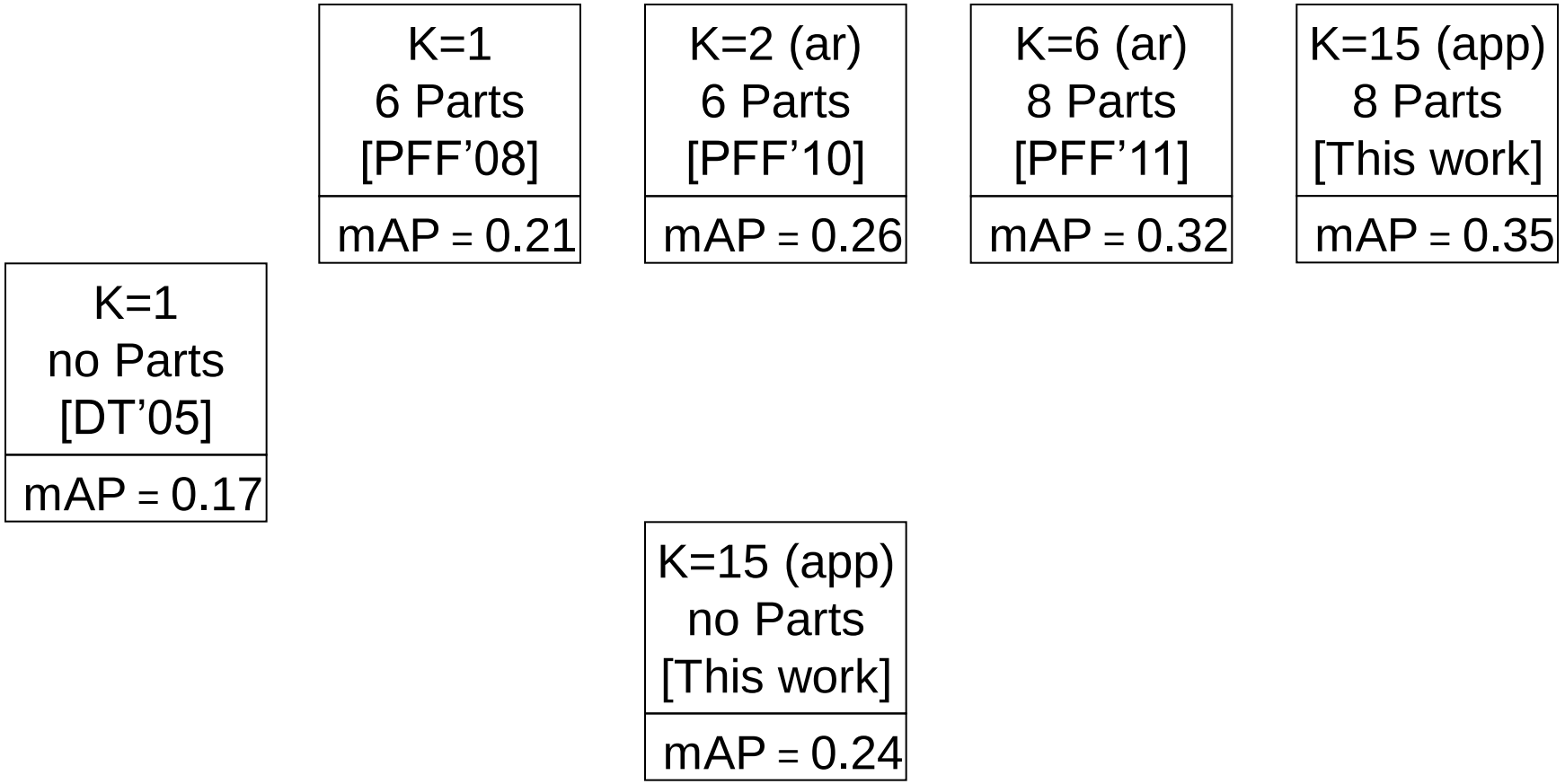
Cluster3

Cluster4

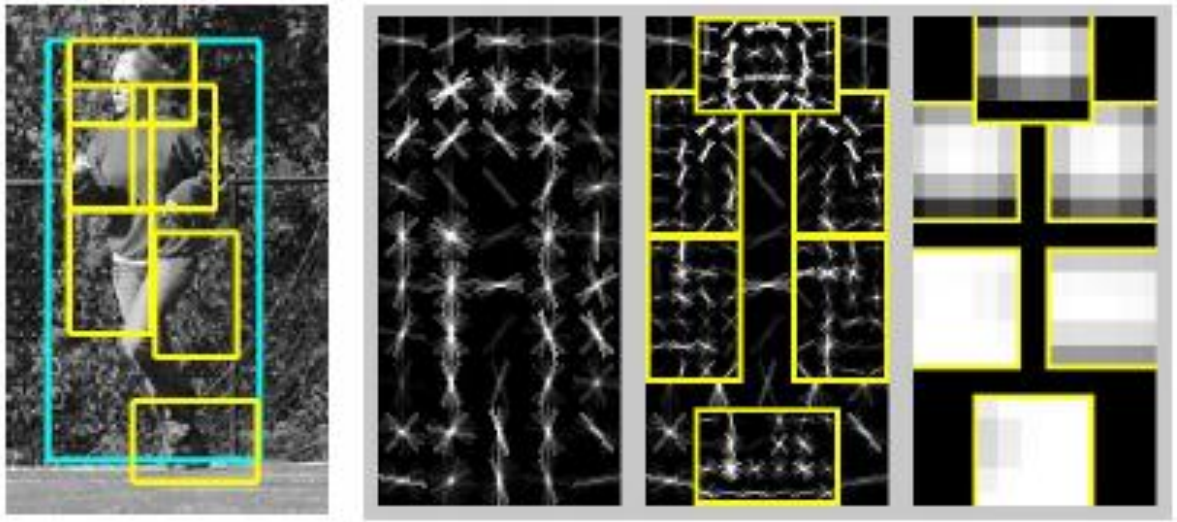
Parts vs. Subcategories: Performance



Parts vs. Subcategories: Performance



How important are “deformable parts”?



Image

Root filter (Coarse resolution)

Part filters (Fine resolution)

Deformation Models

How important are “deformable parts”?

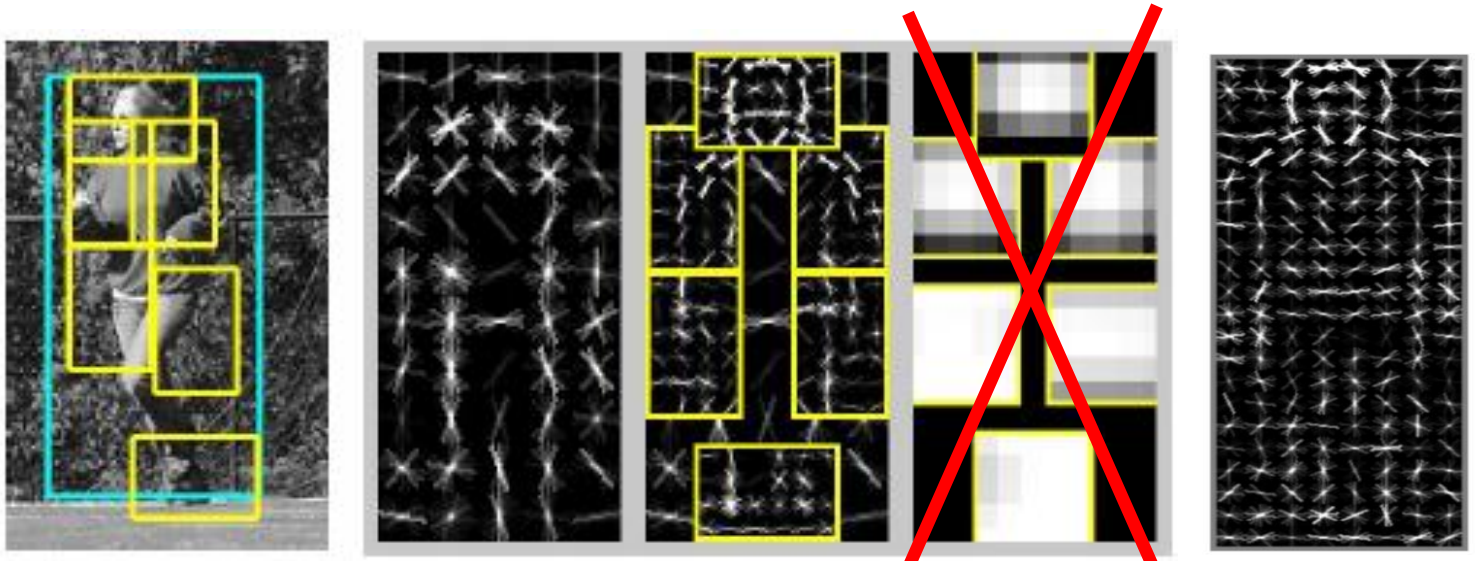
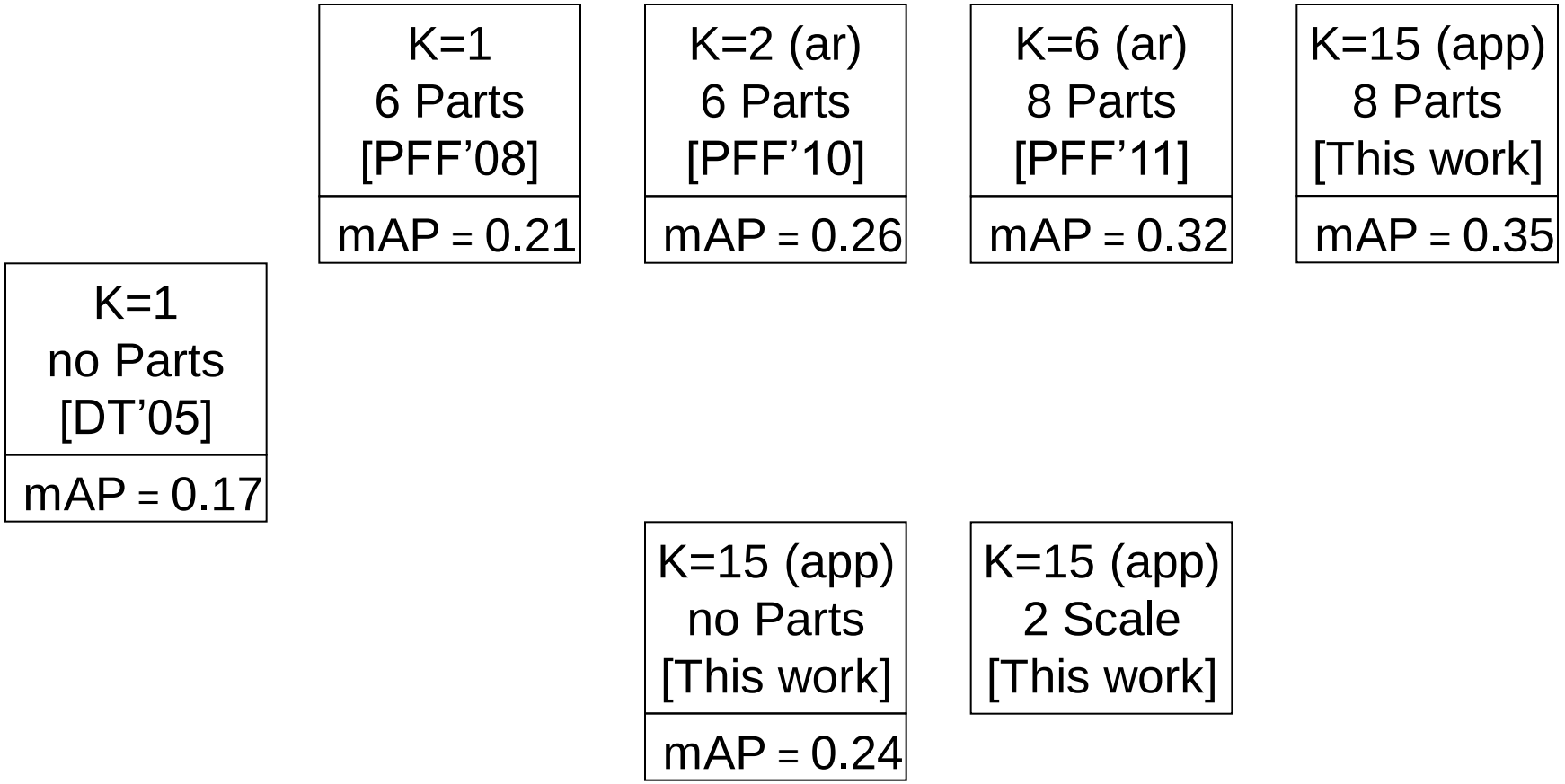
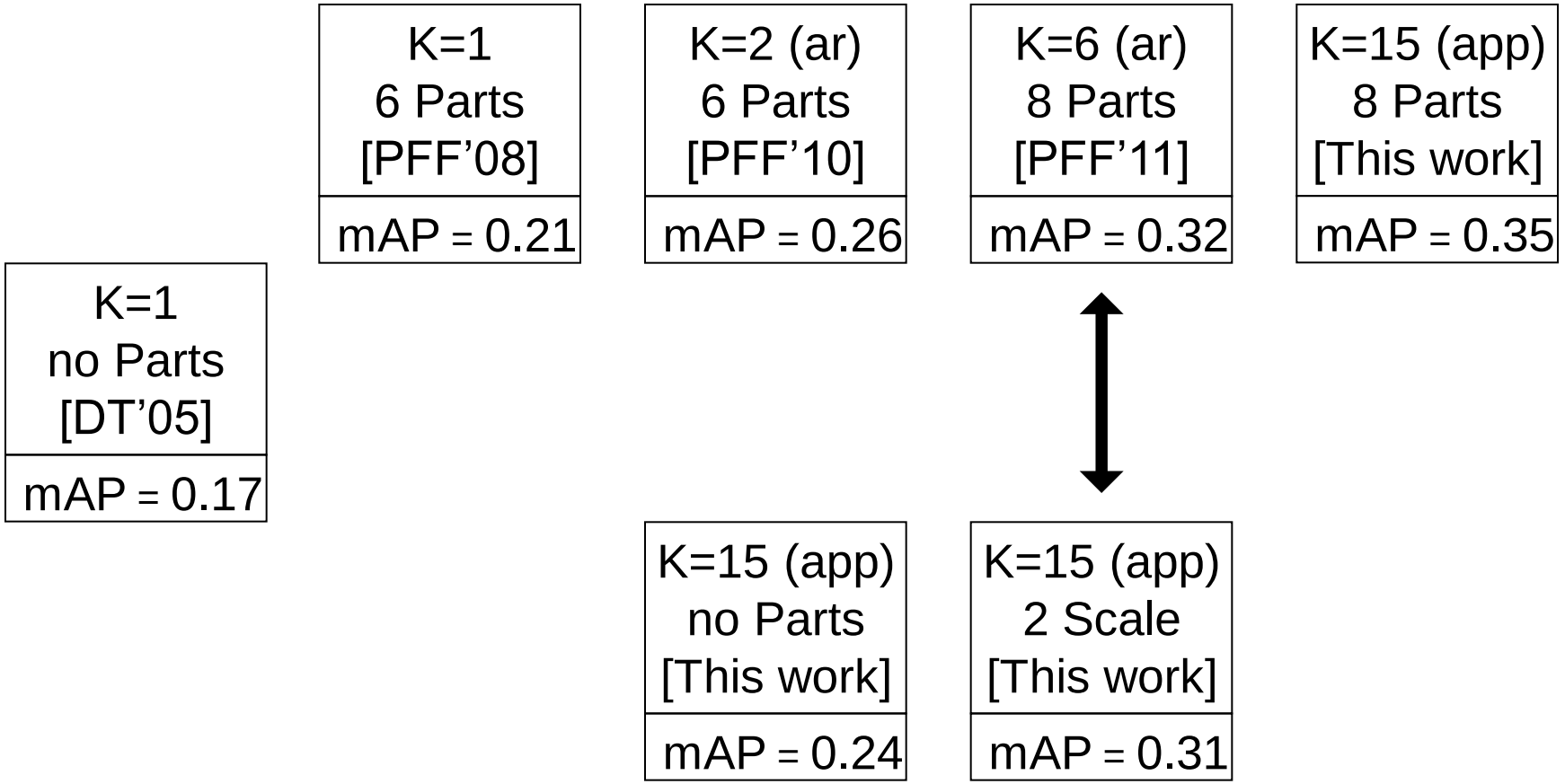


Image | Root filter (Coarse resolution) | Part filters (Fine resolution) | Deformation Models | Child filter (Fine resolution)

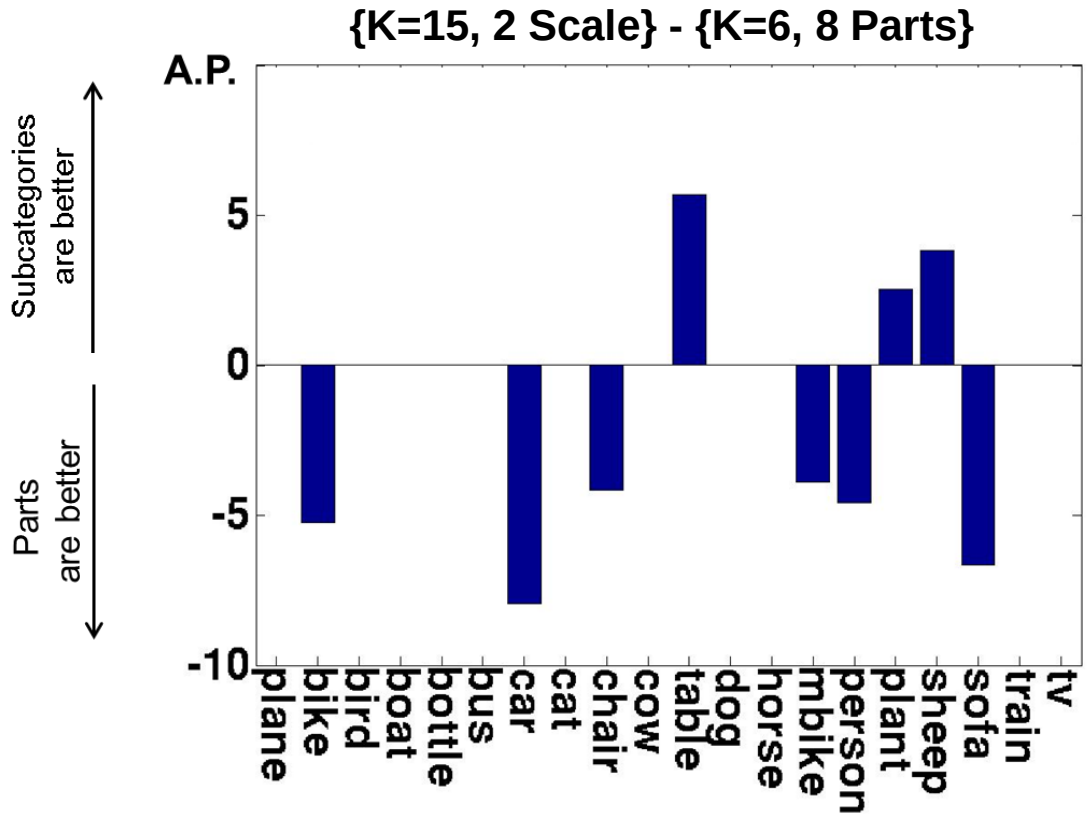
Parts vs. Subcategories: Performance



Parts vs. Subcategories: Performance



How important are “deformable parts”?



- 11 classes: Parts == Subcategories
- 6 classes: Parts > Subcategories
- 3 classes: Parts < Subcategories

Parts vs. Subcategories: Testing

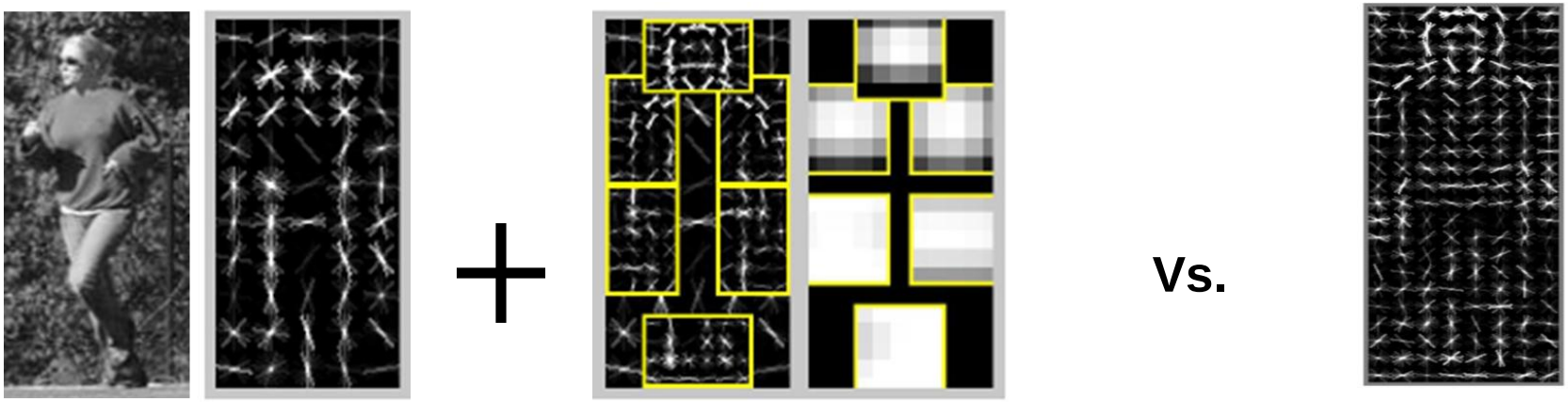


Image | Root filter (Coarse resolution) | Part filters (Fine resolution) | Deformation Models | Child filter (Fine resolution)

Vs.

K = 6, 8 parts
[Felzenszwalb et al., 2011]

K = 15, 2 Scale
[This Work]

✓ Faster Testing

8 parts x 6 subcategories +
1 root x 6 subcategories
= 54 template convolutions
(9 seconds)

1 child x 15 subcategories +
1 root x 15 subcategories
= 30 template convolutions
(7 seconds)

Parts vs. Subcategories: Training

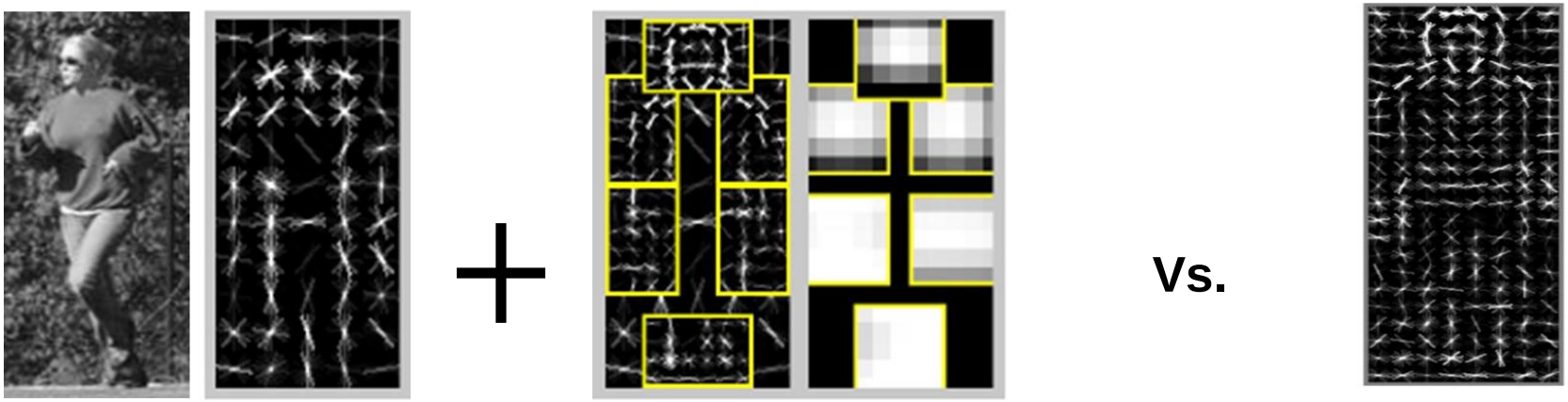


Image | Root filter (Coarse resolution) | + | Part filters (Fine resolution) | Deformation Models | Vs. | Child filter (Fine resolution)

K = 6, 8 parts
[Felzenszwalb et al., 2011]

K = 15, 2 Scale
[This Work]

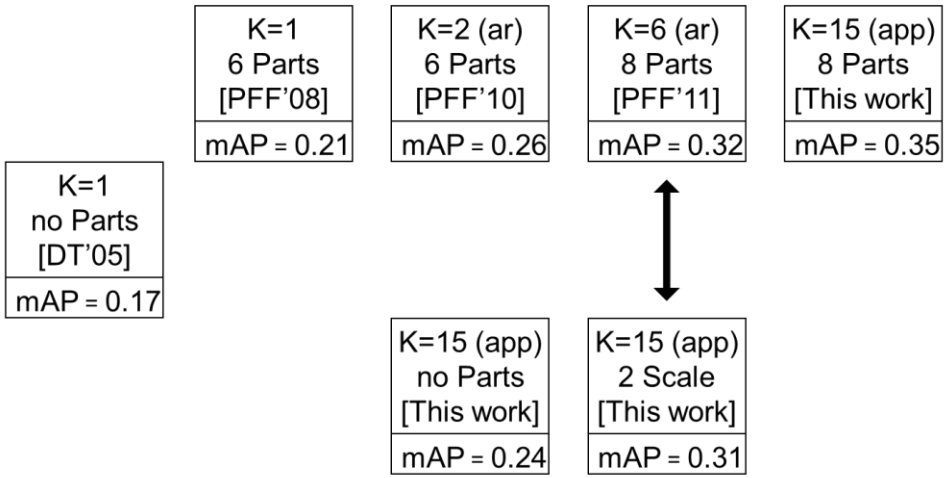
✓ Faster Testing

✓ Faster Training + less prone to local minima

(8 parts x 6 latent vars)
x 3 subcategories + 1
= 145 latent variables
(10.1 hours)

1 latent variable
(8.3 hours)

Deformable Parts *vs.* Subcategories: What have we learned



- Similar performance
  - More of one, less of the other
  - Computational tradeoffs
  - Ease of implementation
- Need for multi-scale