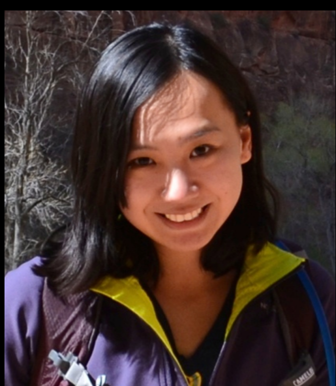
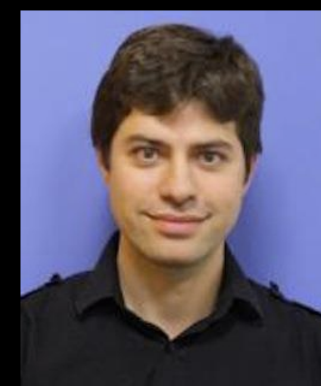
PANDA: Pose Aligned Networks for Deep Attribute Modeling







Ning Zhang[1,2] Manohar Paluri[1] Marc'Aurelio Ranzato[1] Trevor Darrell[2] Lubomir Bourdev[1]

[1] Facebook AI Research [2] EECS, UC Berkeley

men who wear helmet and sunglasses

Why is attribute classification challenging?

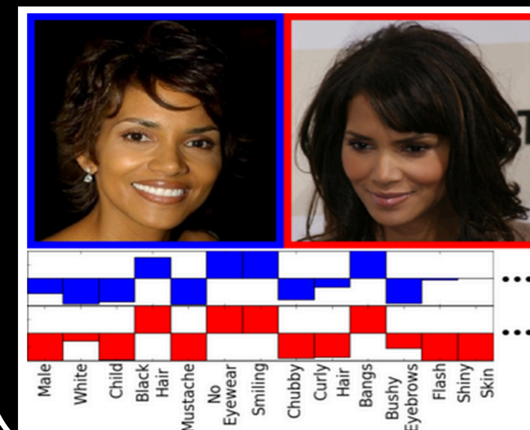

Low resolution
Occlusion
Pose variations

Toward attribute classification

Transfer knowledge

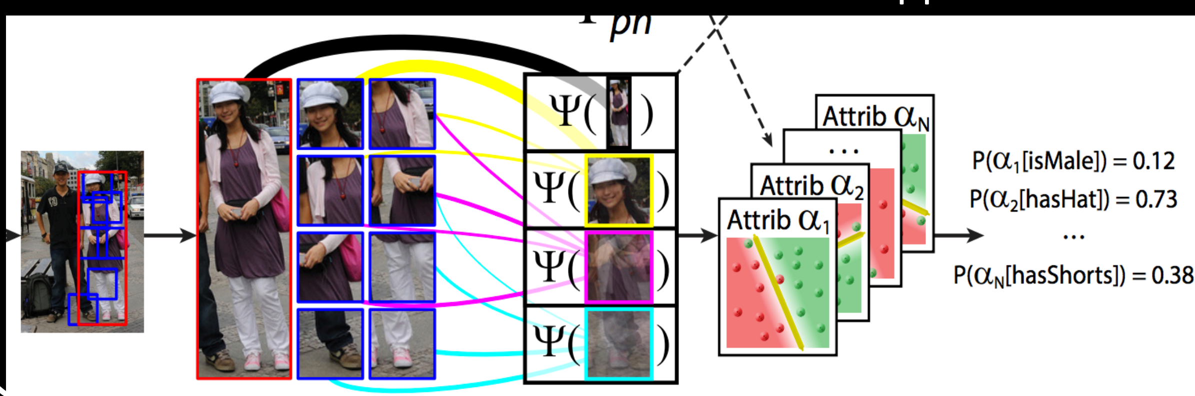
[Lampert et al. (CVPR 09), Farhadi et al. (CVPR 09)]

Facial Attribute



[Kumar et al. (ICCV 09)]

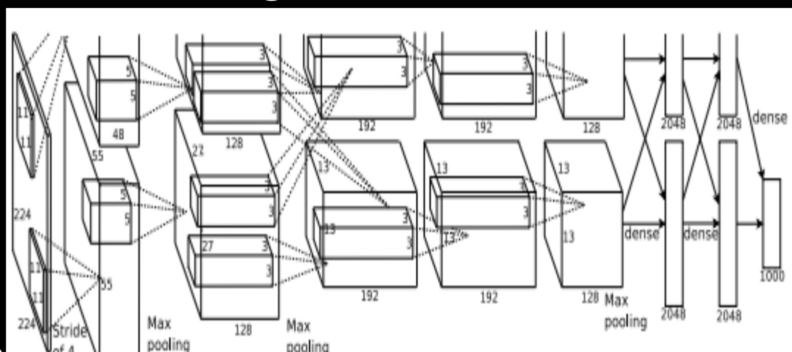
Part-based approach



[Bourdev et al. (ICCV11), Zhang et al. (ICCV 13) Joo et al. (ICCV 13)]

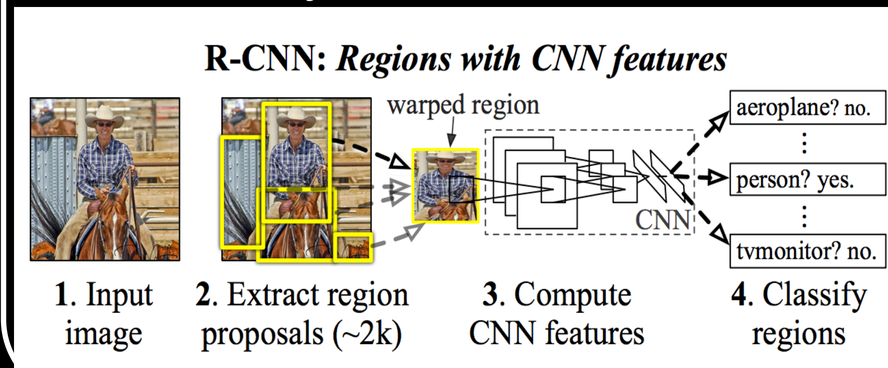
Progress in deep learning

image classification



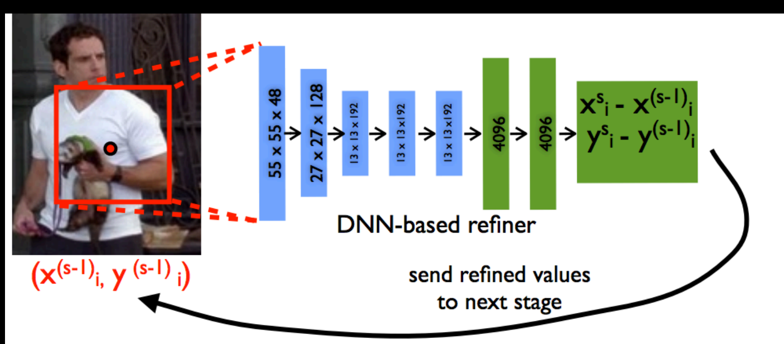
[Krizeshsky et al. NIPS 12, Zeiler et al. ICLR 14]

object detection



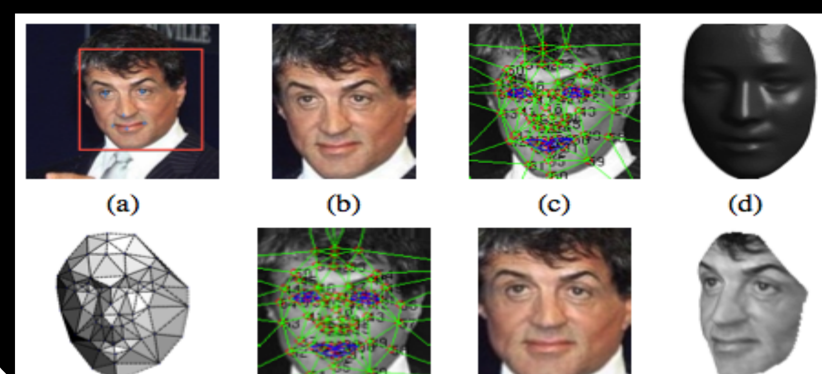
[Girshick et al. CVPR 14]

human pose estimation



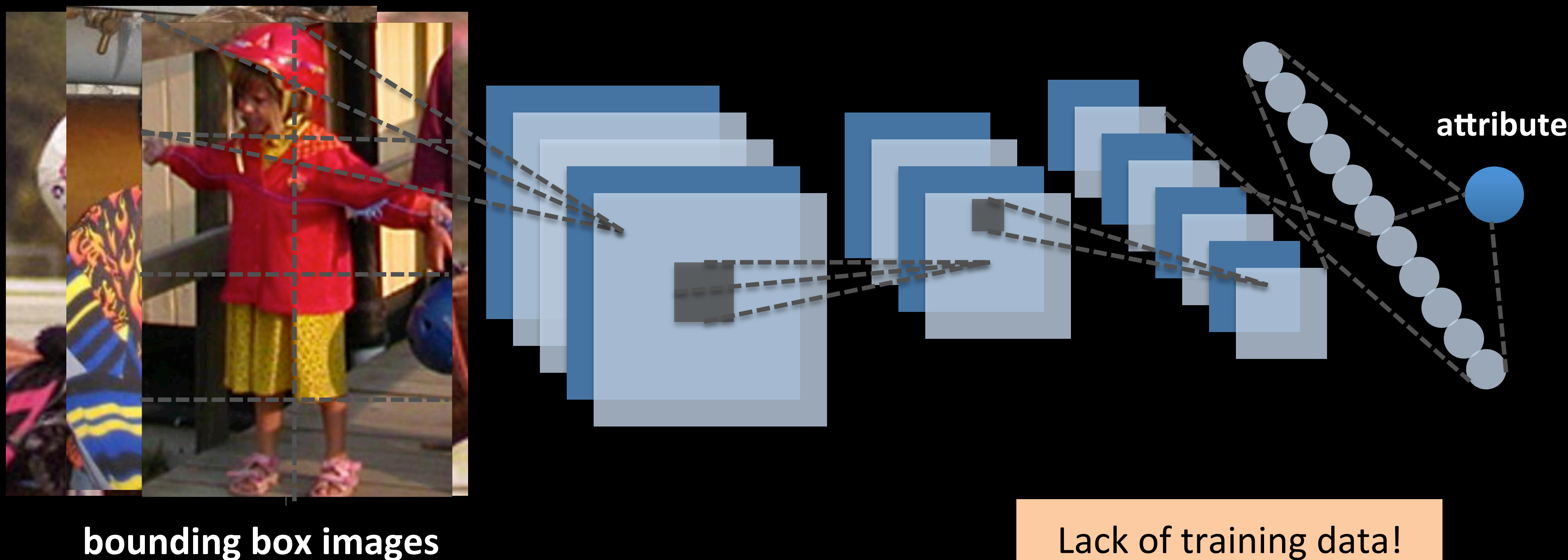
[Toshev et al. CVPR 14]

face verification



[Taigman et al. CVPR 14]

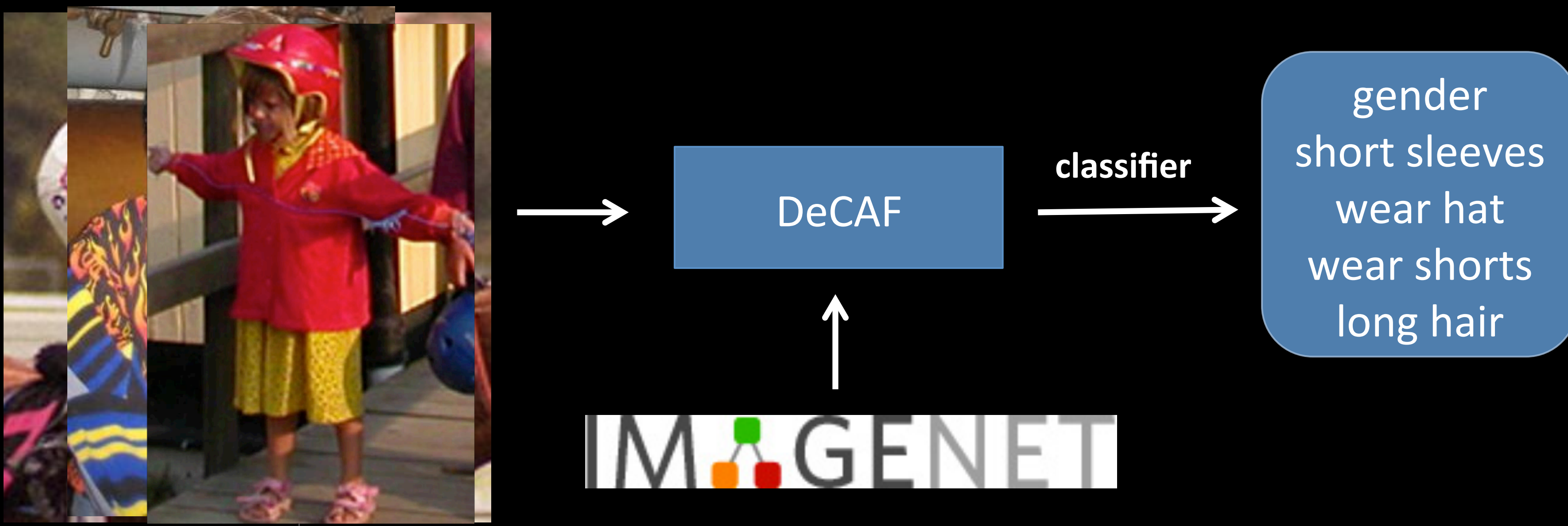
Can we train CNN from scratch?



Lack of training data!

method	Joo et al. ICCV 2013	CNN from scratch
mean AP	70.7	**58.11**

What if we finetune from ImageNet?



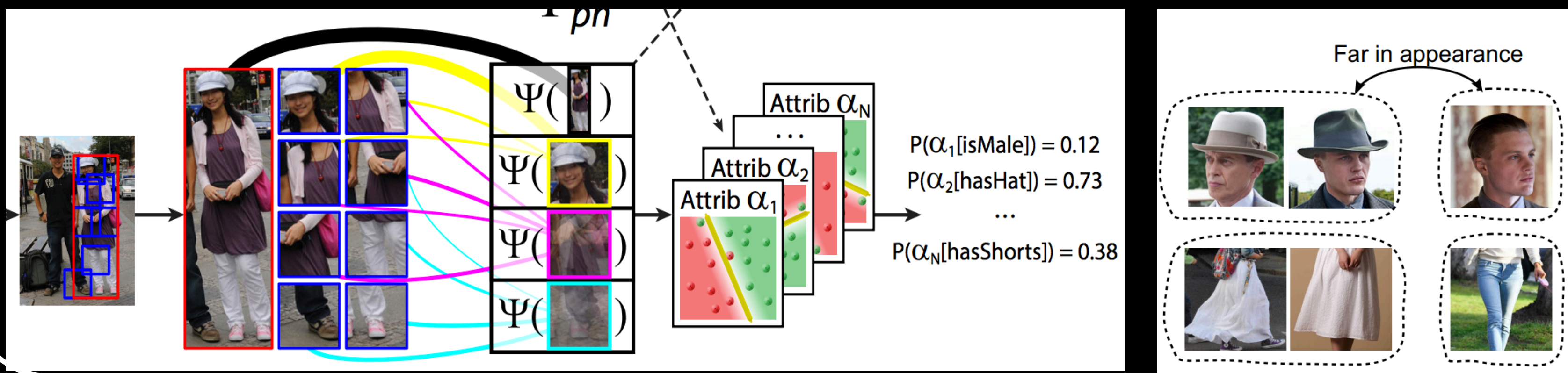
method	Joo et al	from scratch	from ImageNet
mean AP	70.7	58.11	**67.49**

How can we simplify the task?

[Donahue et al. ICML 2014]

Decompose the image into parts

Part-based approach



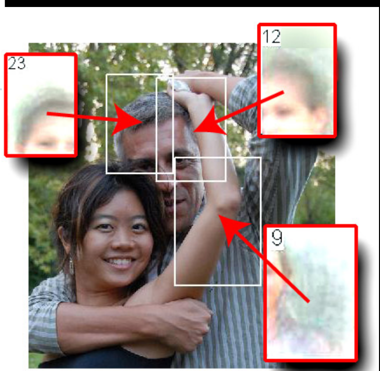
[Bourdev et al. (ICCV11), Zhang et al. (ICCV 13) Joo et al. (ICCV 13)]

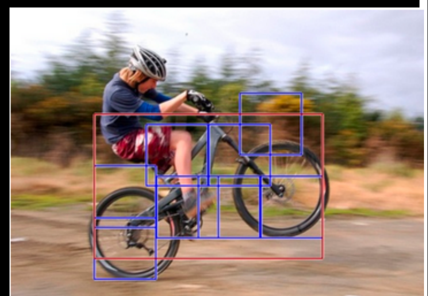
Decompose the image into parts

gender
long hair
wear long pants
wear jeans
wear hat
Wear glasses
is sitting
is playing tennis
is dancing
…

Our approach

Part-based models

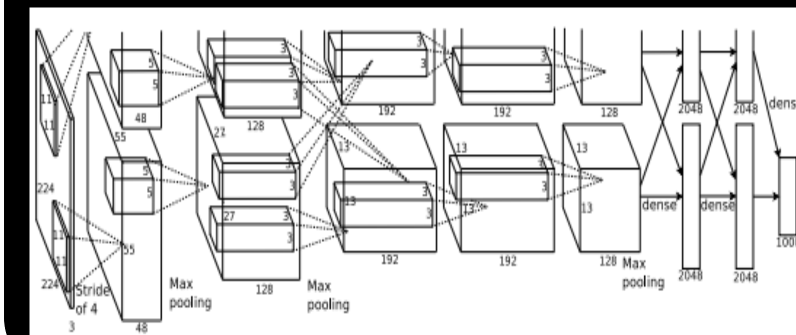




Pose normalization

+

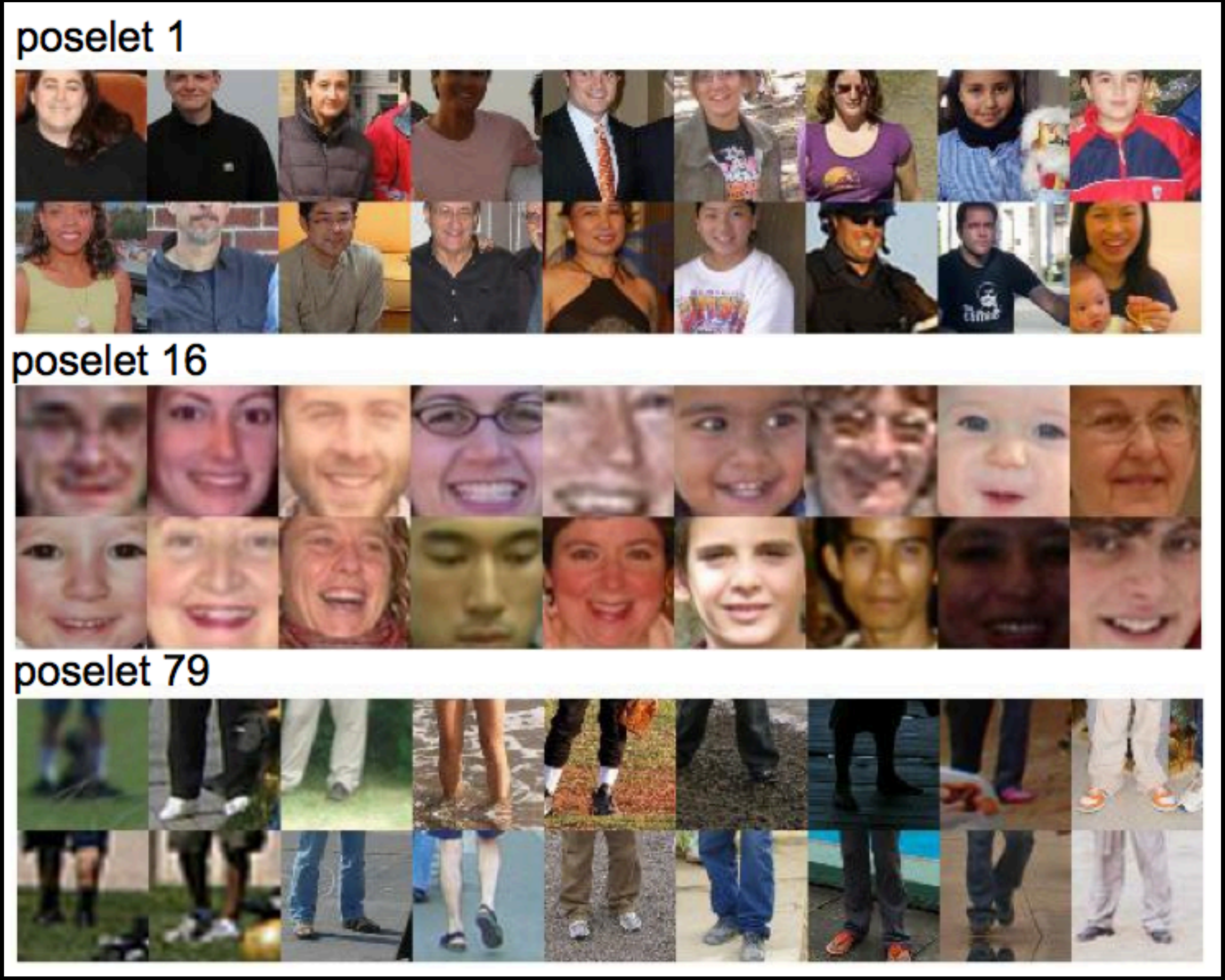
Deep convolutional networks



Discriminative feature representation

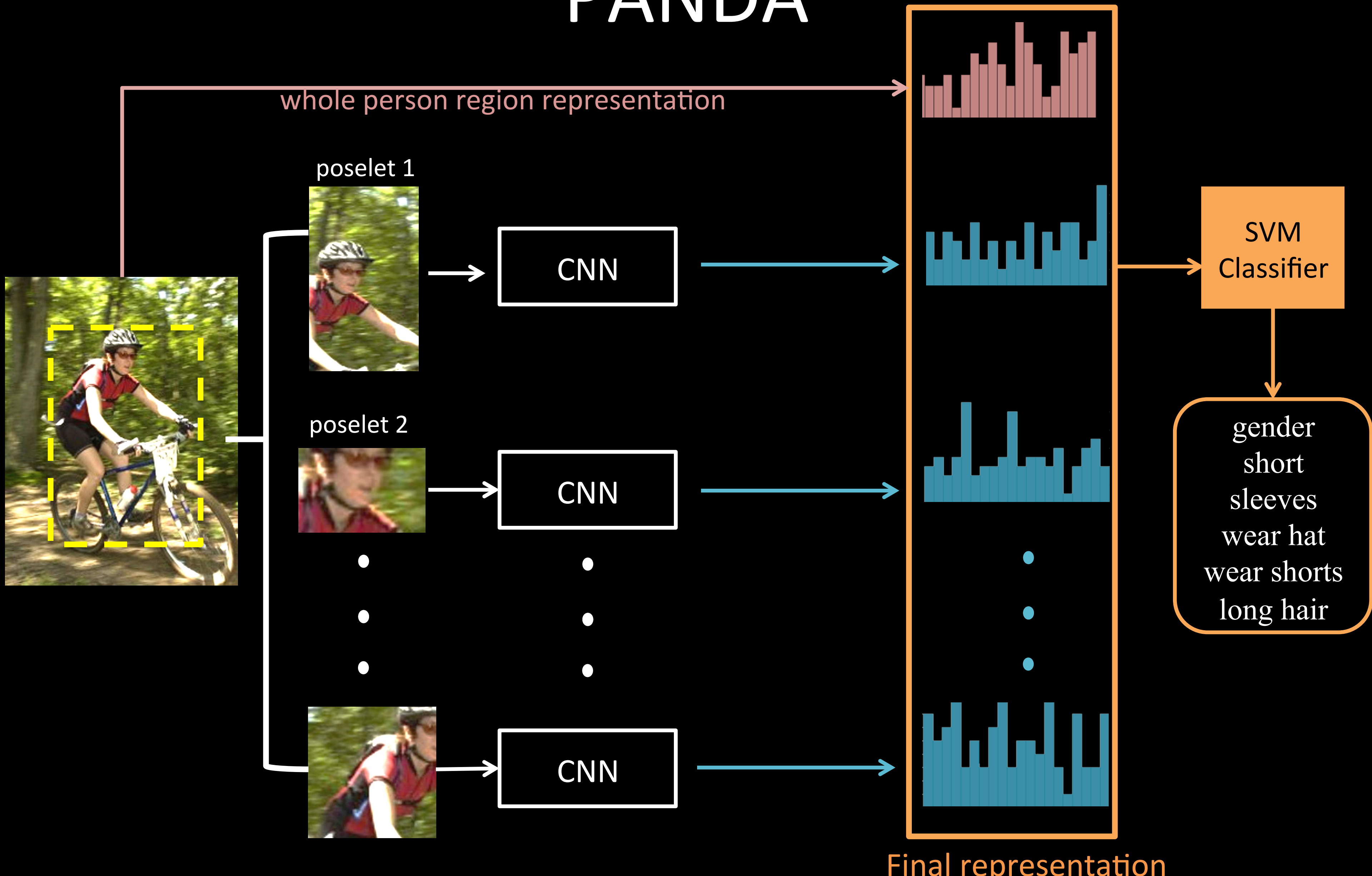
Pose Aligned Networks for Deep Attribute modeling (PANDA)

Poselets capture part of the pose from a given viewpoint

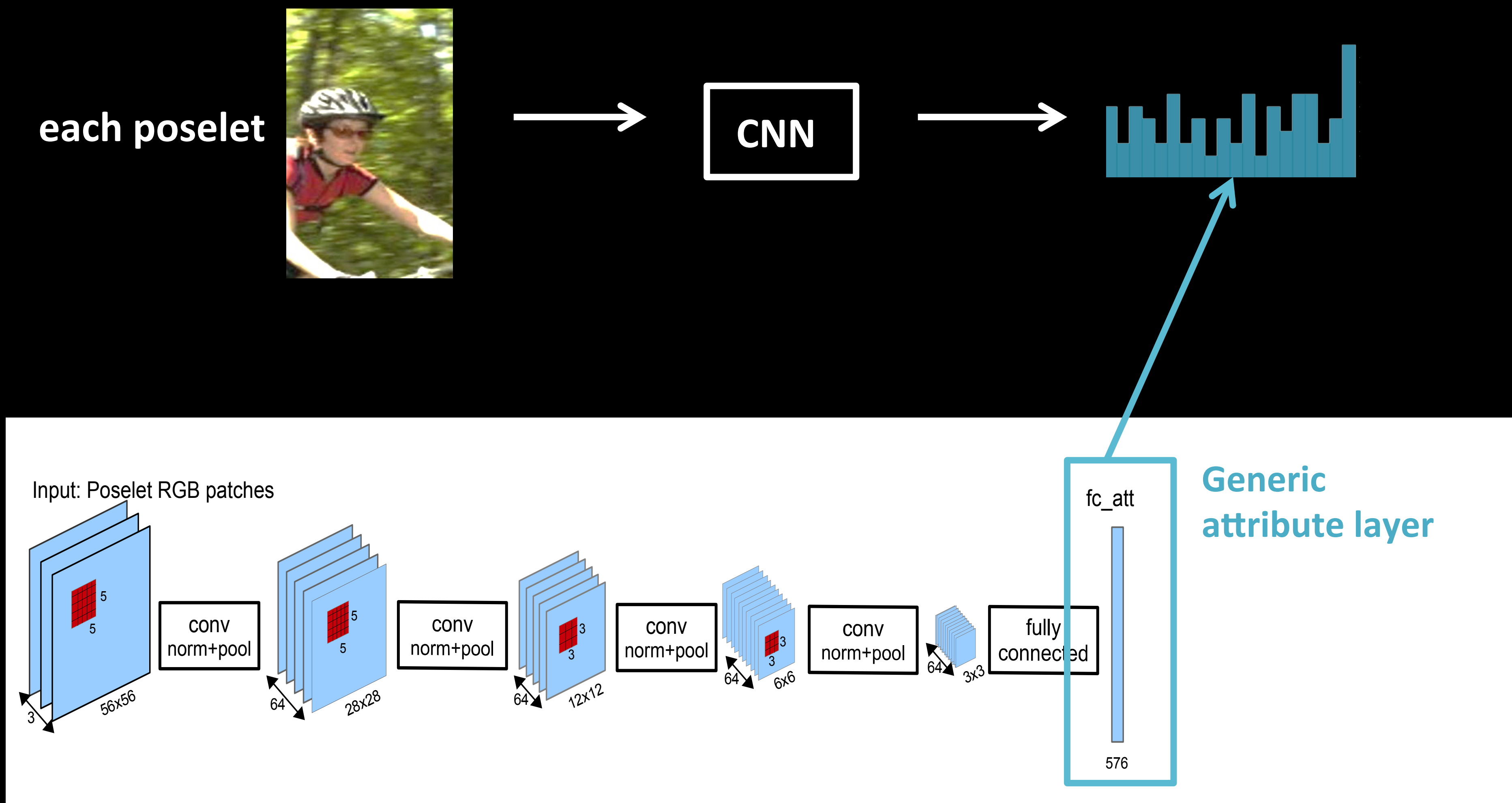


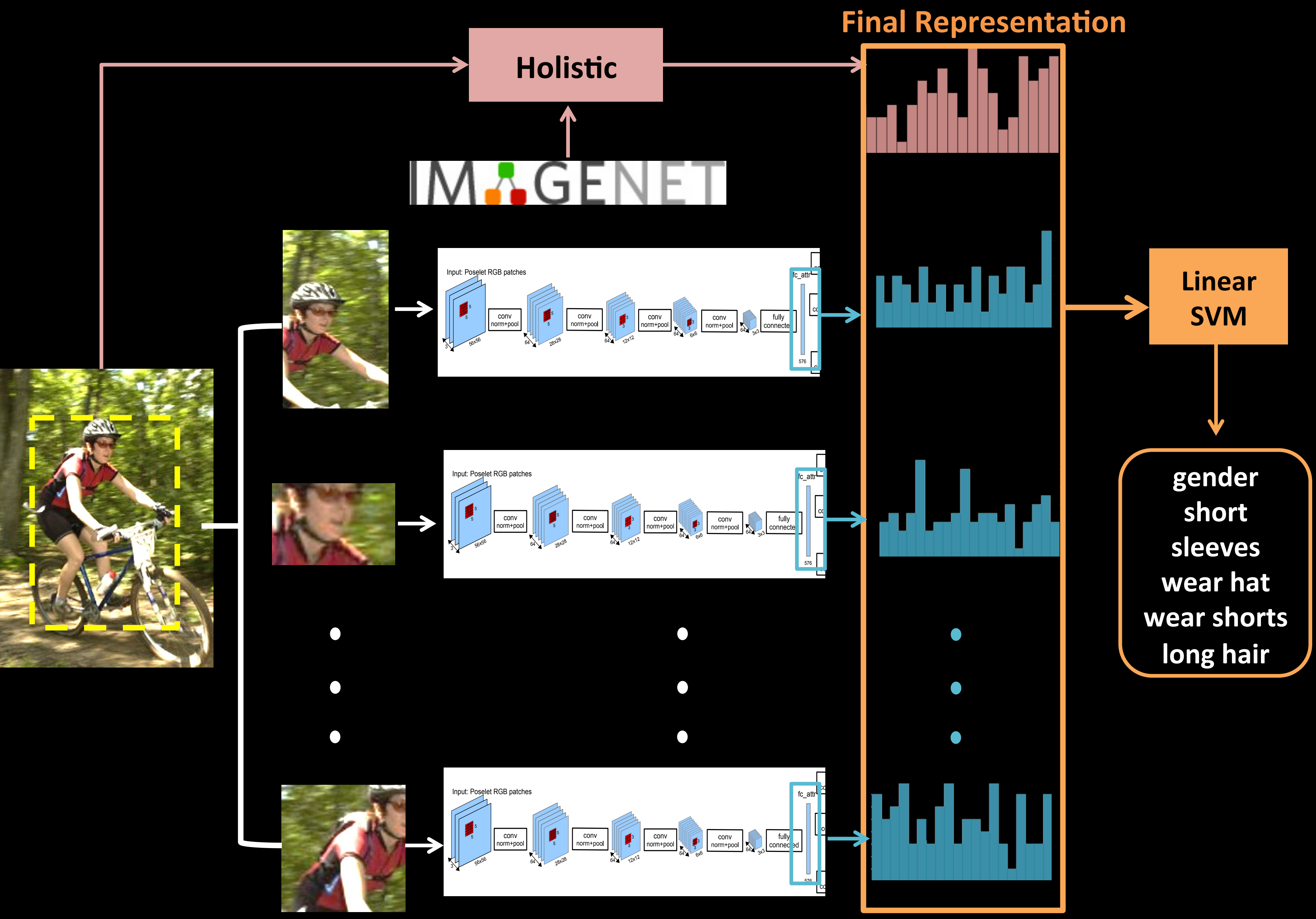
[Bourdev & Malik, ICCV 2009]

PANDA

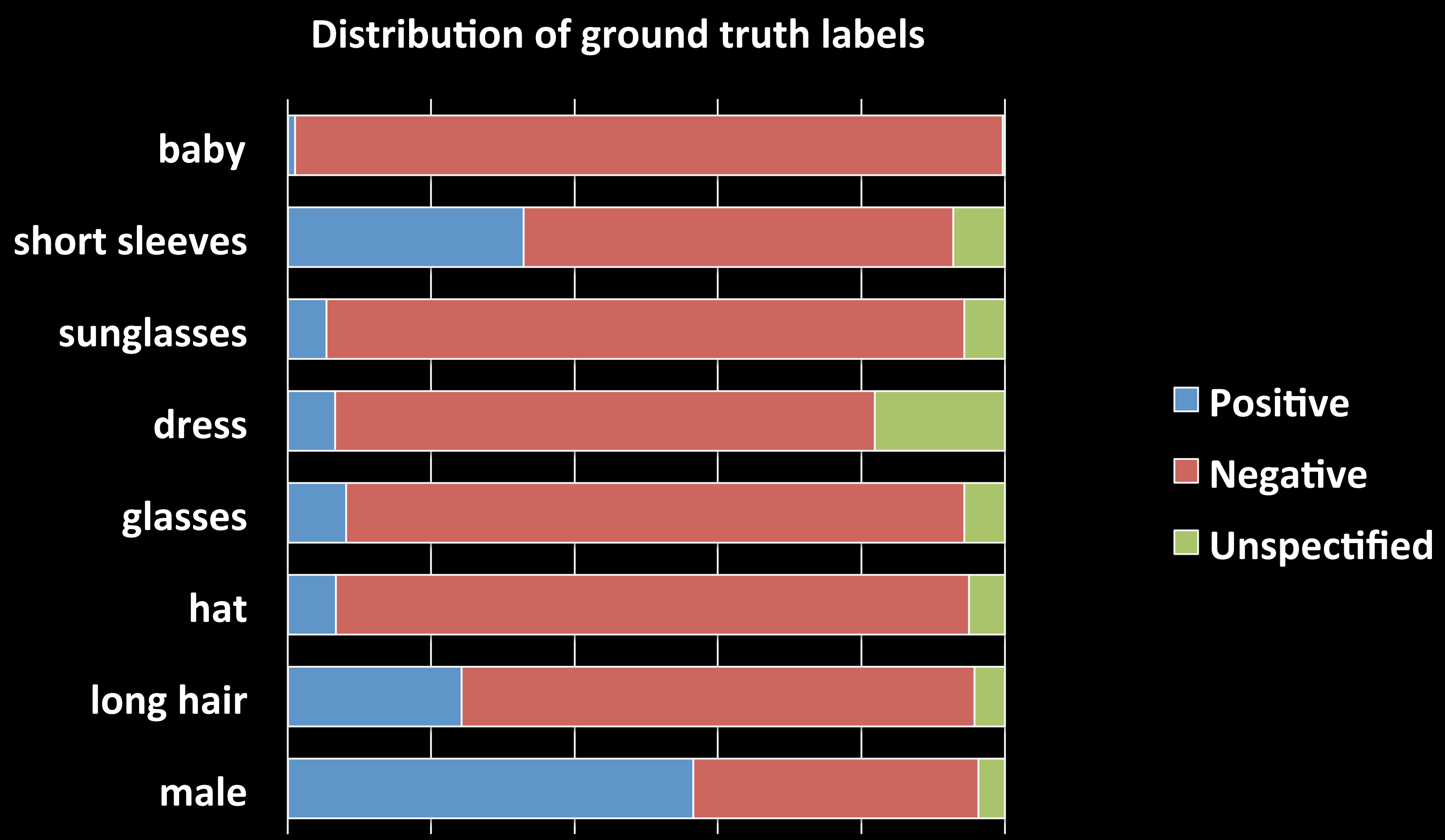

whole person region representation
poselet 1
poselet 2
CNN
CNN
CNN
SVM Classifier
gender
short
sleeves
wear hat
wear shorts
long hair
Final representation

Part-Level CNN




Final Representation
Holistic
IMAGENET
Input: Poselet RGB patches
conv
norm+pool
fully
connected
fc_attr
Linear
SVM
gender
short
sleeves
wear hat
wear shorts
long hair

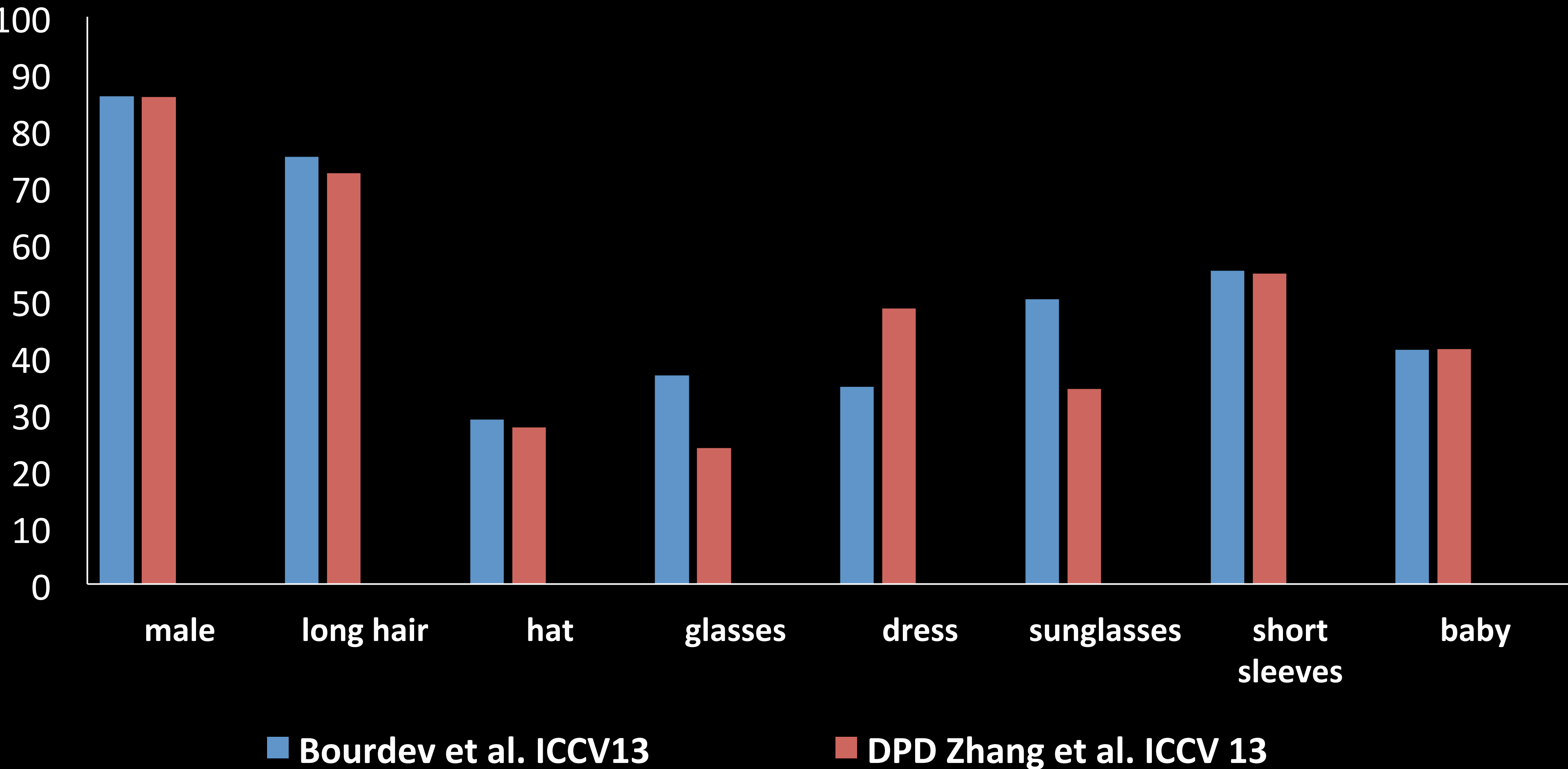
Dataset: Attribute 25k



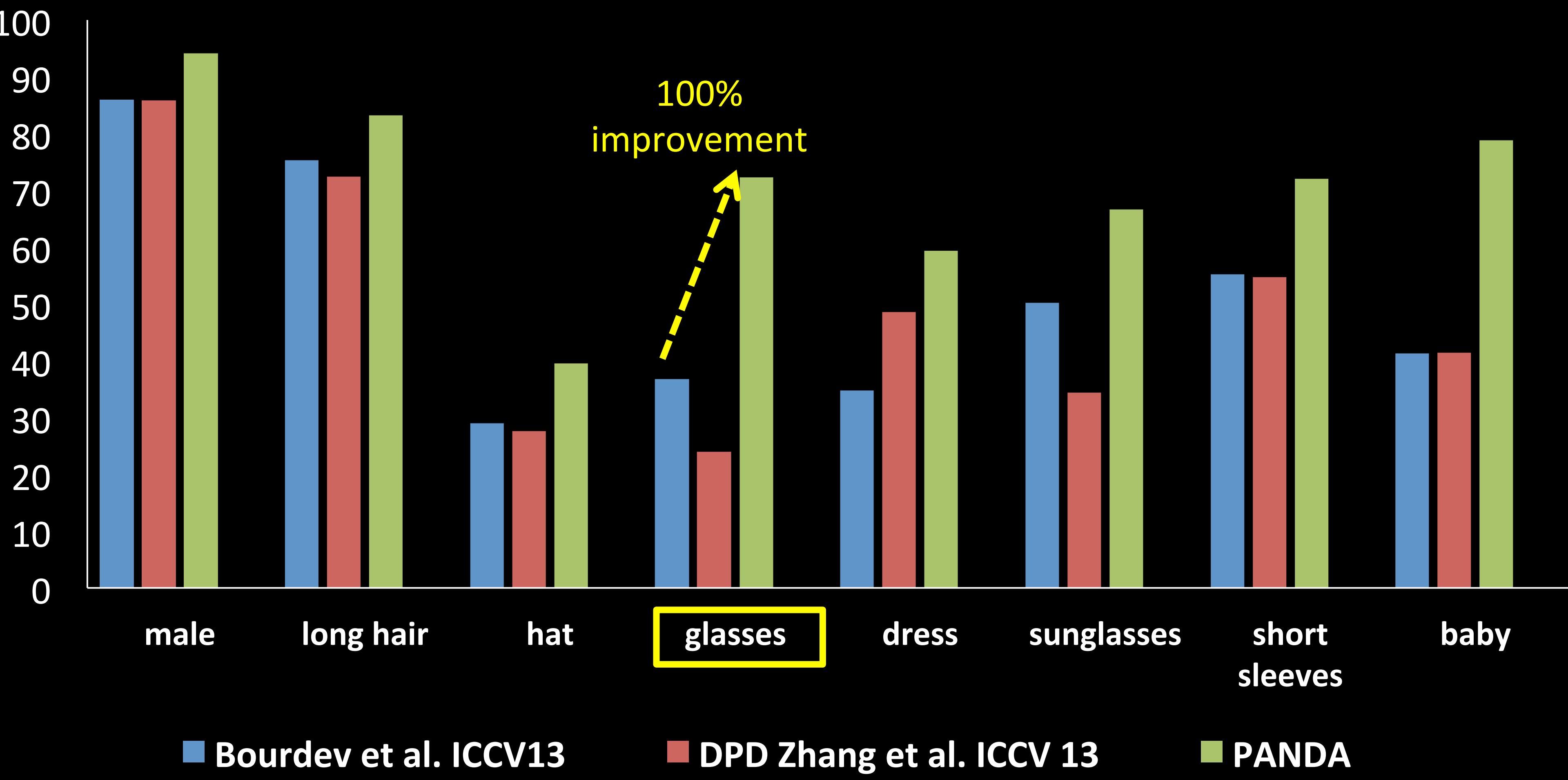
2061 training examples per poselet on average

RESULTS

Average Precision (AP) on Attribute 25k



Average Precision (AP) on Attribute 25k



Component Evaluation

method	mean AP
PANDA (Holistic + Poselets)	70.74

Component Evaluation

method	mean AP
PANDA (Holistic + Poselets)	70.74
Holistic only	44.97
Poselets only	64.72

Component Evaluation

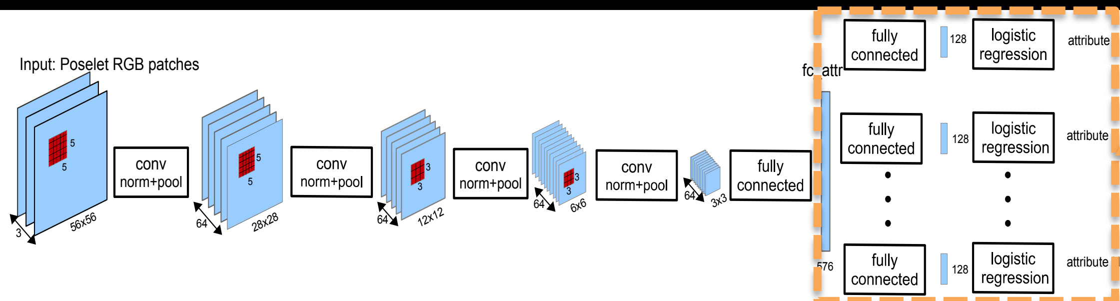
method	mean AP
PANDA (Holistic + Poselets)	70.74
Holistic only	44.97
Poselets only	64.72
Holistic + DPM	61.20

Poselets vs DPM

Frontal face poselet

Head DPM

Transfer learning



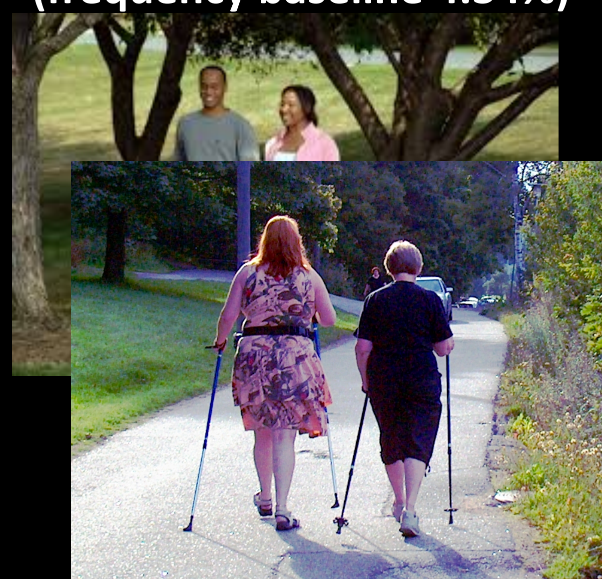
~~Adding n attributes and r a CNNs~~

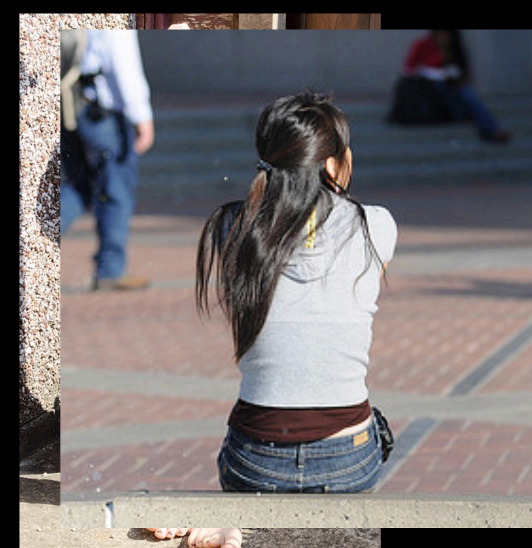
Use the same CNNs only retrain SVM classifier

smiling: AP 84.7%
(frequency baseline 40.67%)

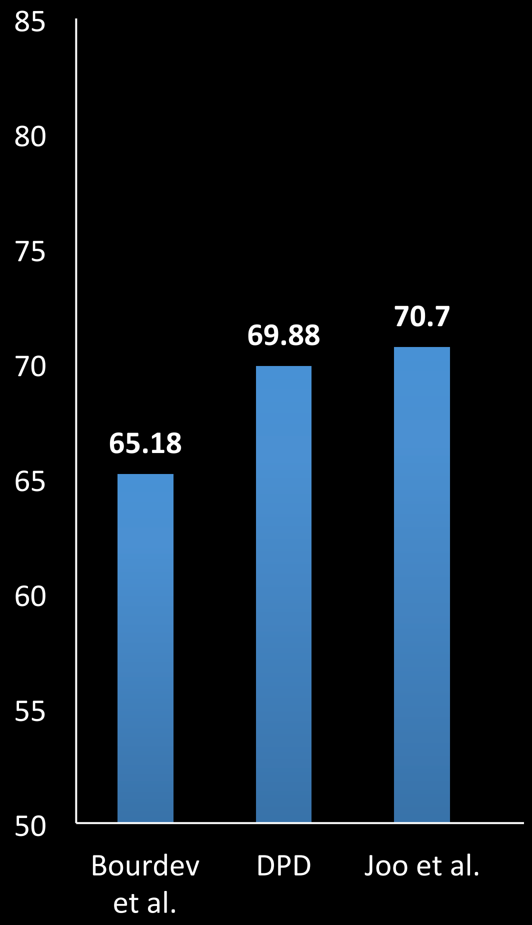
walking: AP 26.0%
(frequency baseline 4.34%)

sitting: AP 25.70%
(frequency baseline 7.65%)

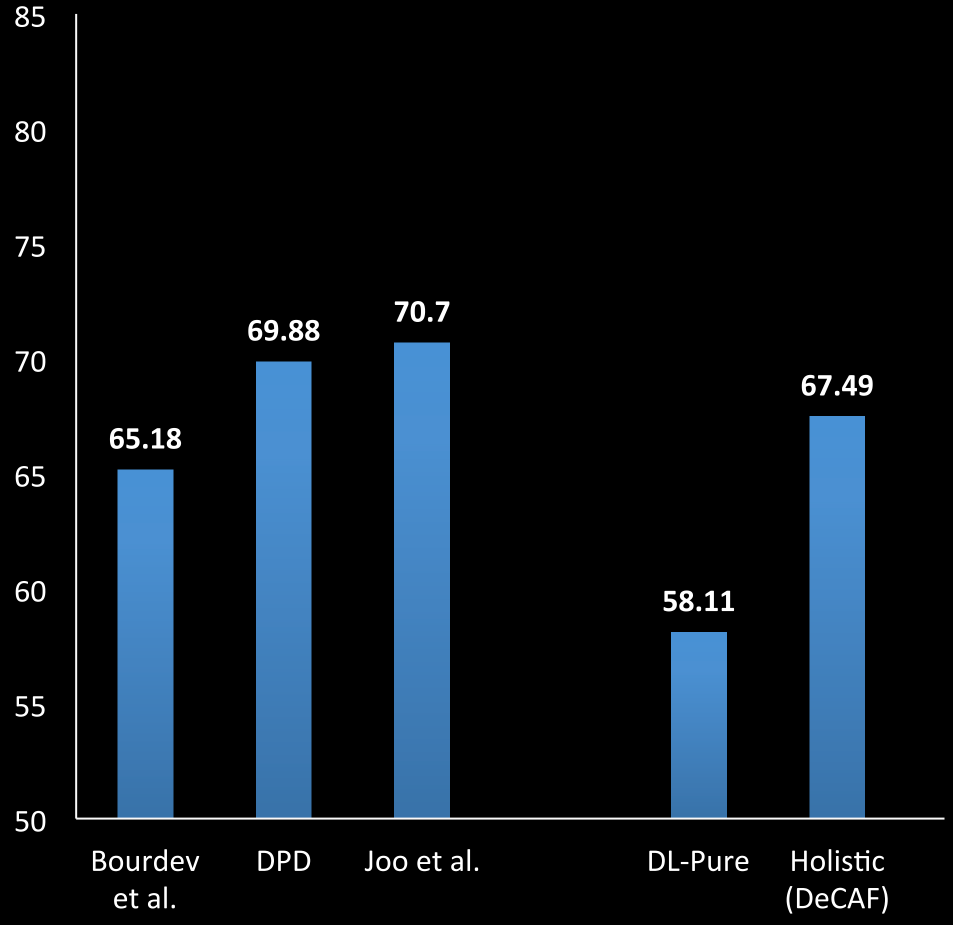




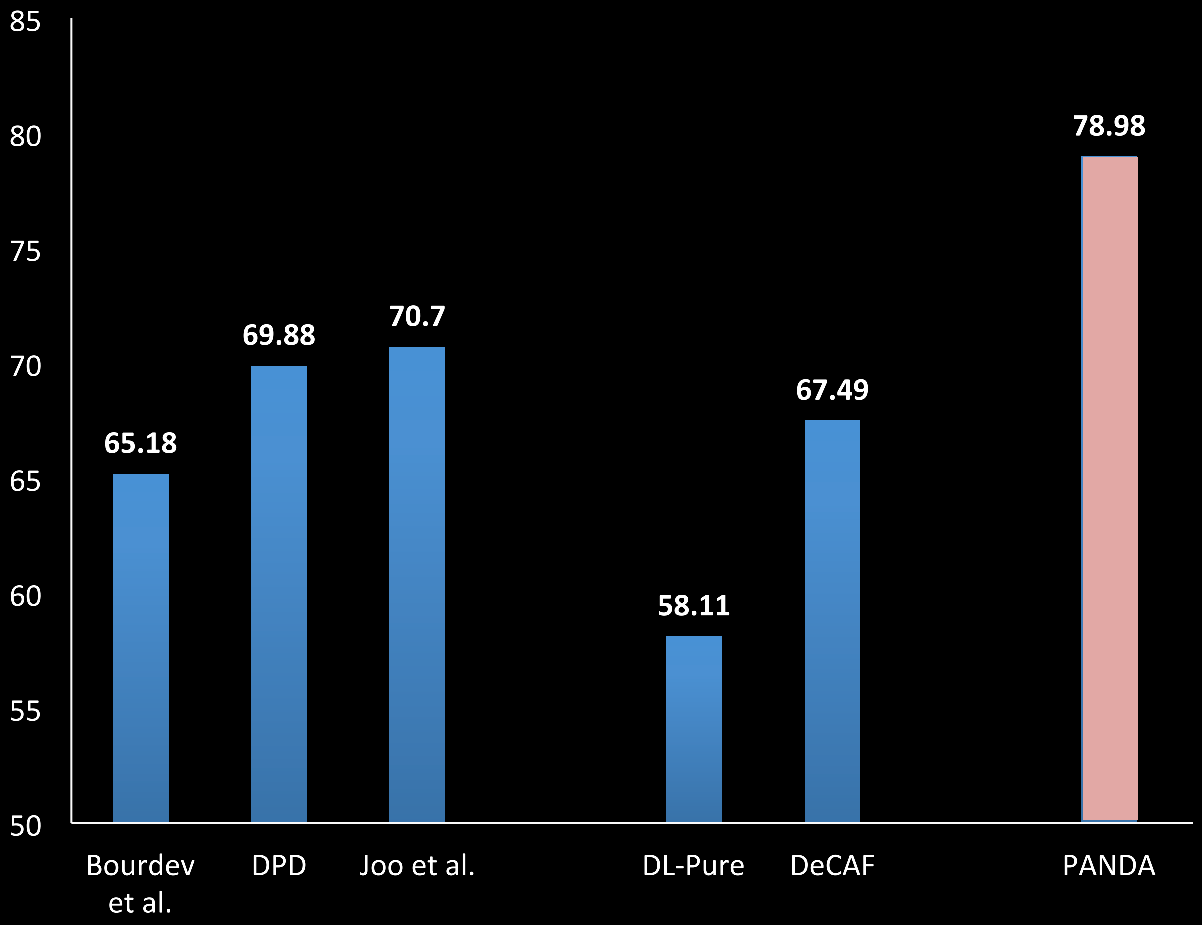
AP on Berkeley Attributes of People Dataset



AP on Berkeley Attributes of People Dataset



AP on Berkeley Attributes of People Dataset



The part-level CNNs are trained using Attribute 25k data.

Top scoring examples

wear glasses

short hair

female

Top scoring examples

wear hat

wear shorts

wear jeans

Failure Cases

Hard to see skin

Unusual pose

Predicted: Long sleeves, Ground truth: short sleeves

Predicted: short pants, ground truth: Long pants

Annotation errors

ambiguous

Gender Recognition on Labeled Faces in the Wild

Much easier dataset – no occlusion, high resolution, centered frontal faces

Method	Gender AP
Kumar et al	95.52
Frontal face poselet	96.43

[Kumar et al, ICCV 2009]

Gender Recognition on Labeled Faces in the Wild

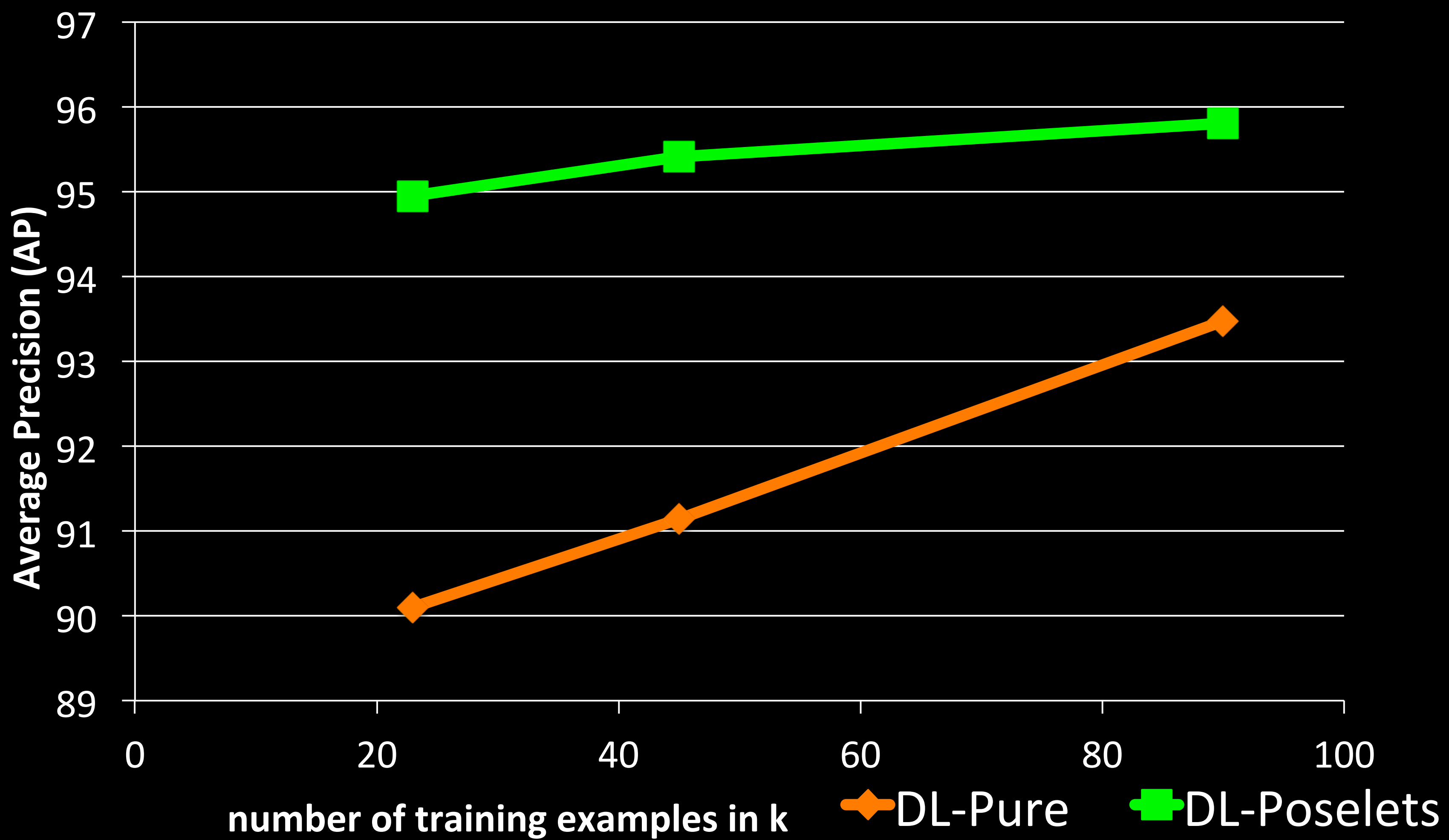
Much easier dataset – no occlusion, high resolution, centered frontal faces

Method	Gender AP
Kumar et al	95.52
Frontal face poselet	96.43
PANDA	99.54

Male of female?

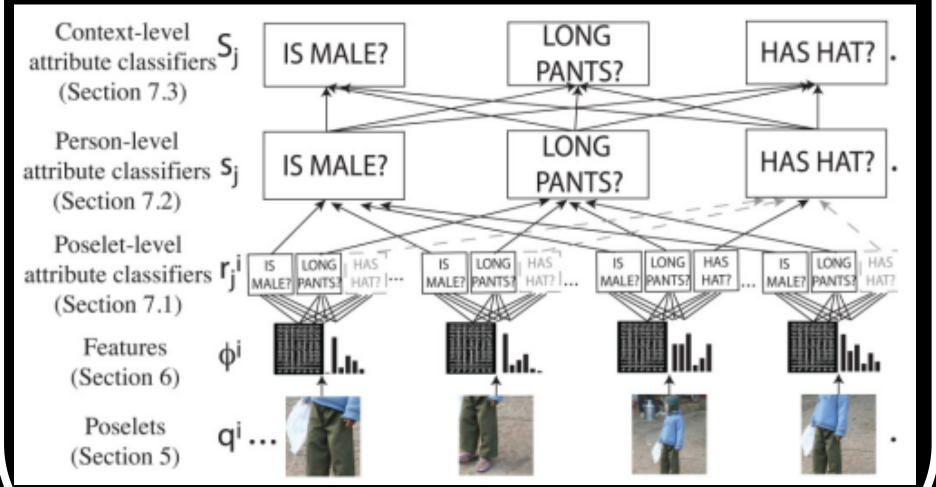
[Kumar et al, ICCV 2009]

Does more data help?

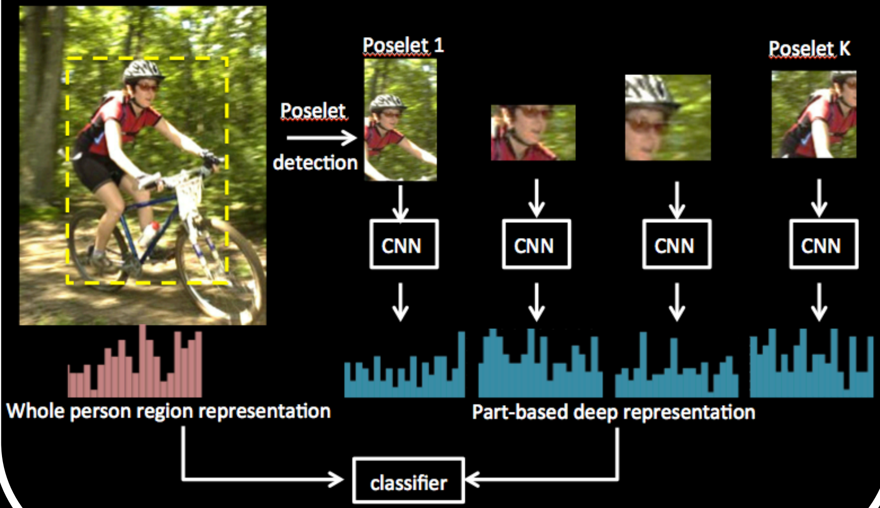


Comparison

Bourdev et al. ICCV 11



PANDA

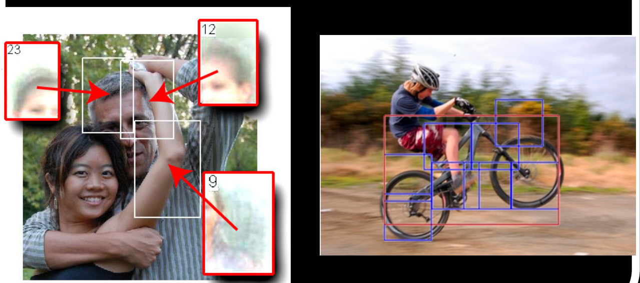


- Use poselet as part-based model
- Has context-level attribute classifier
- Use HOG+color+skin+part masks

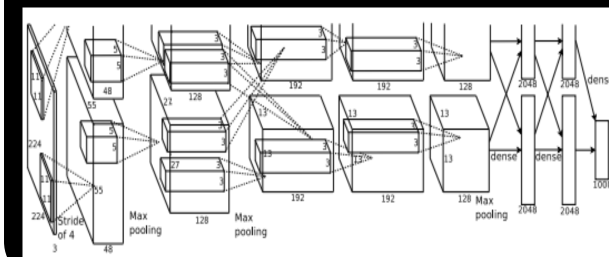
- Use poselets as part-based model
- Attributes are jointly trained
- Training part-level CNN for powerful discriminative feature
- Generalized much better to new attributes

Conclusion

Part-based models



Deep convolutional networks



Pose normalization

Discriminative feature representation

- Pose-normalization significantly helps deep convolutional networks in the task of attribute classification.
- Mid-level parts remain important in the context of CNNs.

Thanks!

- Code and pre-trained models will be released soon.

*None of the images in this slides are taken from Facebook.

Running time

- Single CPU
- 13s (poselet detection) +2s(feature extraction)