Deformable Part Descriptors for Fine-grained Recognition and Attribute Prediction

Ning Zhang[1], Ryan Farrell[1,2], Forrest Iandola[1], Trevor Darrell[1]

Introduction

- Fine-grained Recognition

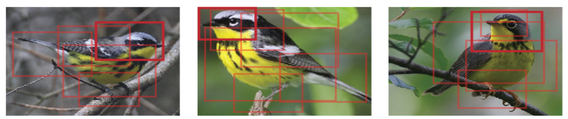
anna's hummingbird

ruby-throated hummingbird

- Human Attribute Prediction

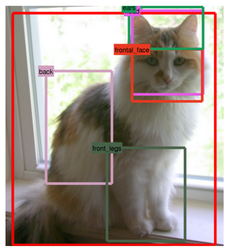
male or female?
long hair?
wear dress?
is baby?
wear shorts?
wear glasses?
wear hat?
wear jeans?
age?
…

- Pose-normalized representations [1]



Deformable Part Model (DPM)

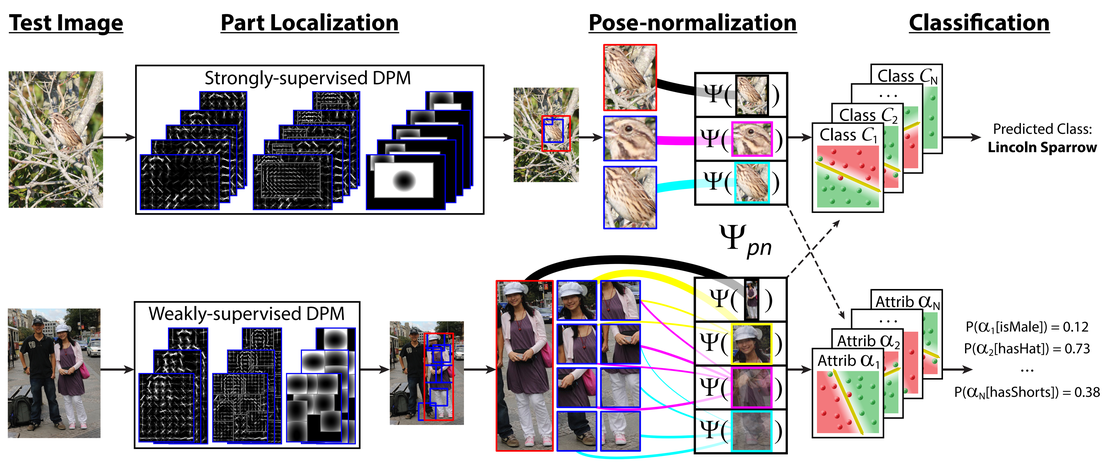
- Weakly supervised DPM
  - Fix-sized part filters initialized by heuristics.
  - Components initialized by clustering aspect ratio.
- Strongly supervised DPM [2]
  - Semantic part filters initialized by part annotations.
  - Clusters pose information to initialize the components.



- Computational efficient DPM detections [3].
- Strong DPM provides semantic part localizations for pose-normalized representations.
- What about simpler weak DPM without pose annotations?

Method

Deformable part descriptors (DPD)



- The first descriptor (top row) applies a strong DPM for part localization then pool features from these inherently semantic parts.
- The second descriptor employs a weakly supervised DPM for part localization and then used a learned semantic correspondence weights to pool features from the latent parts into semantic regions.

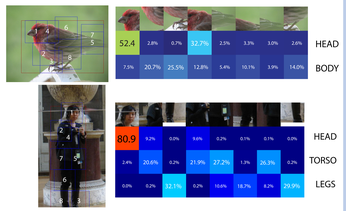
How Weights Get Computed

- $w_{il}^{(j)} \in \mathcal{W}$ of size $|\mathcal{P}| \times |\mathcal{R}| \times |\mathcal{C}|$.

$$w_{il}^{(j)} = \sum_{k=1}^{A} \rho_{kl} \cdot overlap\left(a_k, p_i^{(j)}\right)$$

- $p_i^{(j)}$: i-th part of component $c^{(j)}$. r_l: semantic region.
- $a_k \in \mathcal{A}$: keypoints or other semantic labels.
- $\rho_{kl} \in [0, 1]$: relevance of a_k to region r_l.
- $\mathcal{I}_{jk}$: training images with a_k and component $c^{(j)}$.

$$overlap\left(a_k, p_i^{(j)}\right) = \frac{|\{I \in \mathcal{I}_{jk} | a_k(I) \cap p_i^{(j)} \neq \emptyset\}|}{|\mathcal{I}_{jk}|}.$$



Pooling/Classification

- Pose-normalized representation

$$\Psi_{pn}(I) = [\Psi(I, r_0), \ldots, \Psi(I, r_R)].$$

- Pooled image feature for semantic region $\Psi(I, r_l)$.

$$\Psi(I, r_l) = \frac{1}{N} \sum_{i=1}^{P} w_{il}^{(j)} \cdot \Psi\left(I, p_i^{(j)}\right)$$

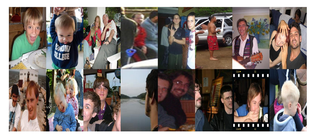
- 1 vs all linear SVM using Ψ_{pn} for final classification.

Example Results and Failure Cases

Top scored people with long hair.

Top scored people wearing long sleeves.

Most confused failure case of males.



Experimental Results

Fine-grained Recognition

Method	Mean Accuracy(%)
MKL	19.0
Random Forest	19.2
KDES	26.4
TriCos	26.7
Template matching	28.2
Segmentation	30.2
Bubblebank	32.5
DPD-strong-2	**34.5**

Results on CUB200-2010 dataset .

Method	Mean Accuracy(%)
PPK	28.18
KDES	42.53
Template matching	43.67
Oracle	64.53
DPD-weak-8	50.98
DPD-strong-2	50.05
DPD-weak-8-DeCAF [4]	**64.96**

Results on CUB200-2011 dataset.

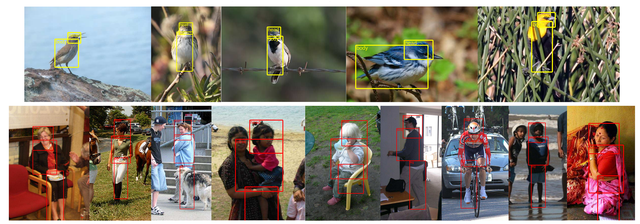
Human Attribute Prediction

Attribute	Freq	SPM	Poselets	Per component	DPD-weak-8	DPD-strong-2
is male	59.3	68.1	82.4	80.5	82.9	**83.7**
has long hair	30.0	40.0	**72.5**	60.8	67.8	70.0
has glasses	22.0	25.9	**55.6**	33.6	40.7	38.1
has hat	16.6	35.3	60.1	61.3	70.3	**73.4**
has t-shirt	23.5	30.6	**51.2**	43.7	46.1	49.8
has long sleeves	49.0	58.0	74.2	74.3	76.5	**78.1**
has shorts	17.9	31.4	45.5	50.3	59.4	**64.1**
has jeans	33.8	39.5	54.7	72.3	77.1	**78.1**
has long pants	74.7	84.3	90.3	90.6	93.0	**93.5**
Mean AP	36.31	45.91	65.18	63.03	68.20	**69.88**

Results on the Human Attributes dataset.

Localization Results of strong DPM

Samples of correct part localizations.



Failure cases of part localizations.

References

[1] Ning Zhang, Ryan Farrell and Trevor Darrell. Pose Pooling Kernels for Sub-Category Recognition. In CVPR 2012.
[2] Hossein Azizpour and Ivan Laptev. Object Detection Using Strongly-Supervised Deformable Part Models. In ECCV 2012.
[3] Charles Dubout and François Fleuret. Exact Acceleration of Linear Object Detectors. In ECCV 2012.
[4] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng and Trevor Darrell. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. On Arxiv.

[1]ICSI / EECS, University of California, Berkeley [2]Brigham Young University
Mail: {nzhang, forresti, trevor}@eecs.berkeley.edu, farrell@cs.byu.edu

code available at: http://dpd.berkeleyvision.org