Articulated Part-based Model for Joint Object Detection and Pose Estimation

Min Sun  Silvio Savarese
Dept. of Electrical and Computer Engineering, University of Michigan at Ann Arbor, USA

1. Overview
- Goal: Joint Object Detection and Pose Estimation
- Articulated Part-based Model:
  - recursive coarse-to-fine representation
  - multiple part-types
  - parents-child relationship

2. Model (a) Model Structure (b) Img Evidence
- Matching Score
  - Appearance Score: $f^A(h; i) = A^T \psi_i(h, i)$
  - Deformation Score: $f^D(h; \hat{h}) = -d^T \psi_i(h, \hat{h})$
- Score Aggregation
  - Child location selection: $f_{c_{si}}(h, \hat{h}, l) = \max_{h'} f_{c_{si}}(h'_l, \hat{h})$
  - Child alignment: $f_{c_{si}}(T(h, T^i_{f_{c_{si}}})$, $l)$, $T(h, \hat{h}) = (x - t_x, y - t_y, L - t_L, \theta - t_\theta)$
  - Child type selection: $f_{c_{si}}(h, \hat{h}; l) = \max_{h'} f_{c_{si}}(T(h, T^i_{f_{c_{si}}} l) + \theta_{c_{si}}^i$
  - Aggregation: $f_{c_{si}}(h, \hat{h}, l) = f_{c_{si}}(h, \hat{h}, l) + \sum_{f(h, l)}$

3. Learning
- Linear Weights
  $w^T \psi_{i}(H, i) + \sum_{s} w_s \psi_{i}(H, i) + \psi_{i}(H, i))$
- Struct SVM
  $\min_{w, c} \sum_{s} \epsilon^T(h)$
  $s.t. \epsilon^T(h) = \max (|D(h, H^i) - w^T \psi(H, i)|)$

4. Results on Human Dataset
- Poselet Dataset
- Iterative Image Parsing Dataset

5. Results on Cats & Dogs Dataset
- LSVM[4] head cat foreleg tail dog foreleg tail

6. Conclusion
- Improvement in both object detection and pose estimation:
  recursive coarse-to-fine and multiple part-type representation
- Novel performance measure: the part recall vs. FPPI curve

Acknowledgments
ONR grant N000141110389