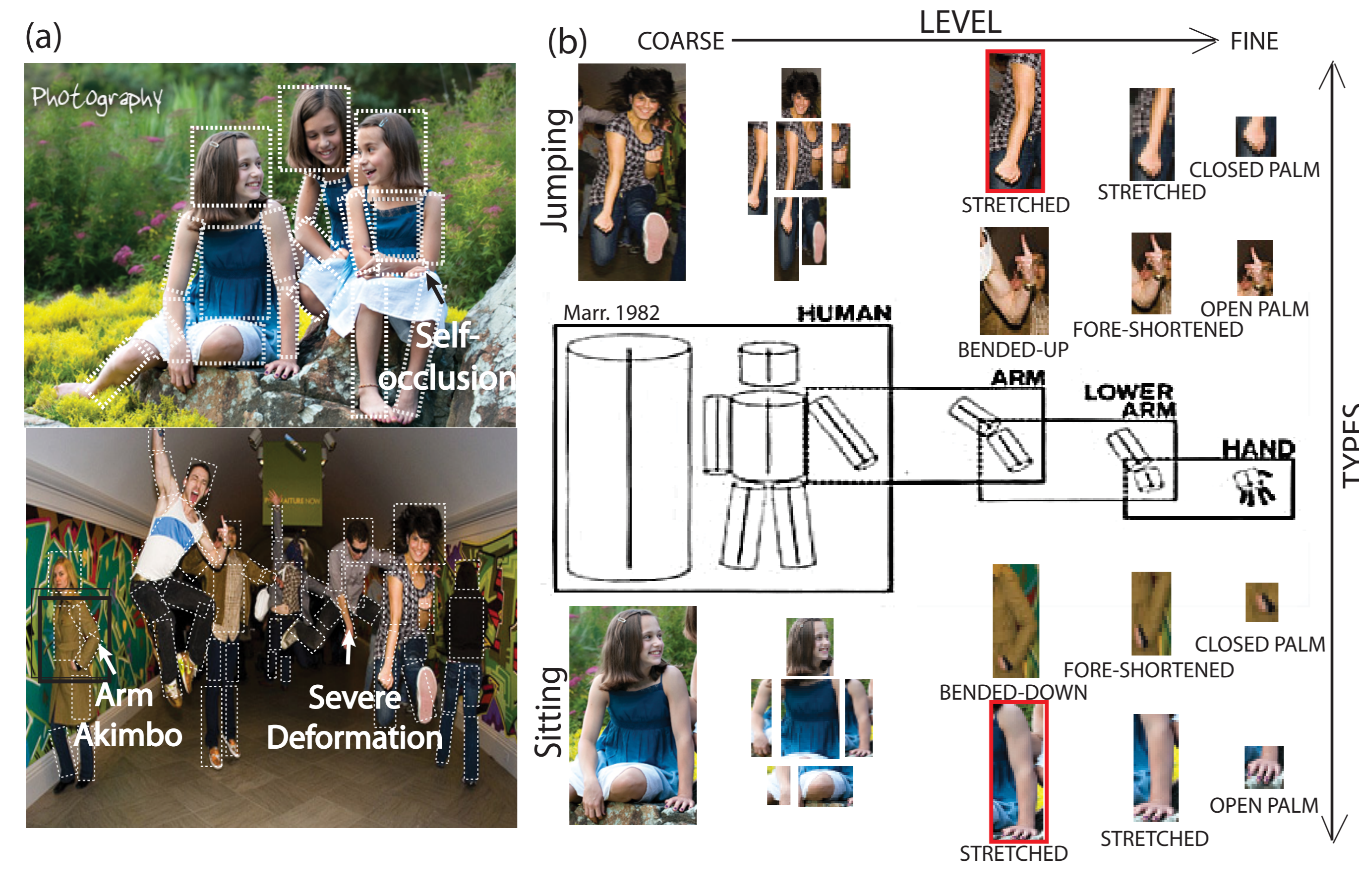


1. Overview

-Goal: Joint Object Detection and Pose Estimation

-Articulated Part-based Model:

- recursive coarse-to-fine representation
- multiple part-types
- parents-child relationship



3. Learning

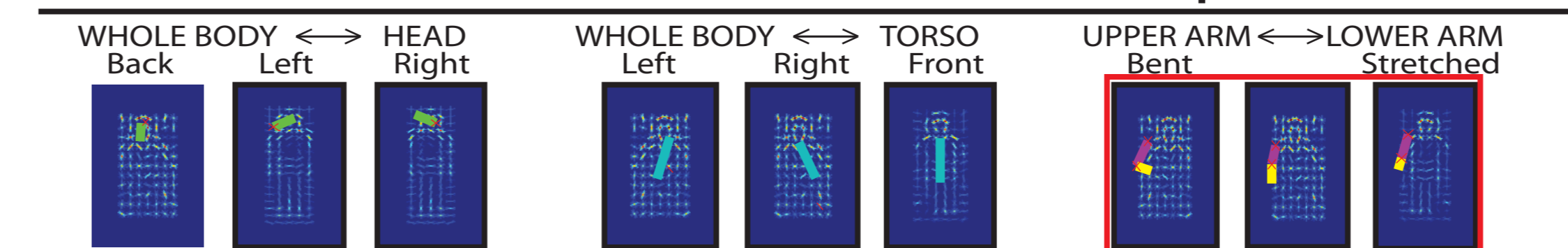
-Linear Weights

$$w^T \Psi(H; I) = \sum_i A_{(i_s)}^T \Psi_a(h_i; I) + \sum_{ij} (b_{ij}^{s_i, s_j} - d_{(i_s, j_s)}^T \Psi_d(h_j, T(h_i, t_{ij}^{s_i, s_j})))$$

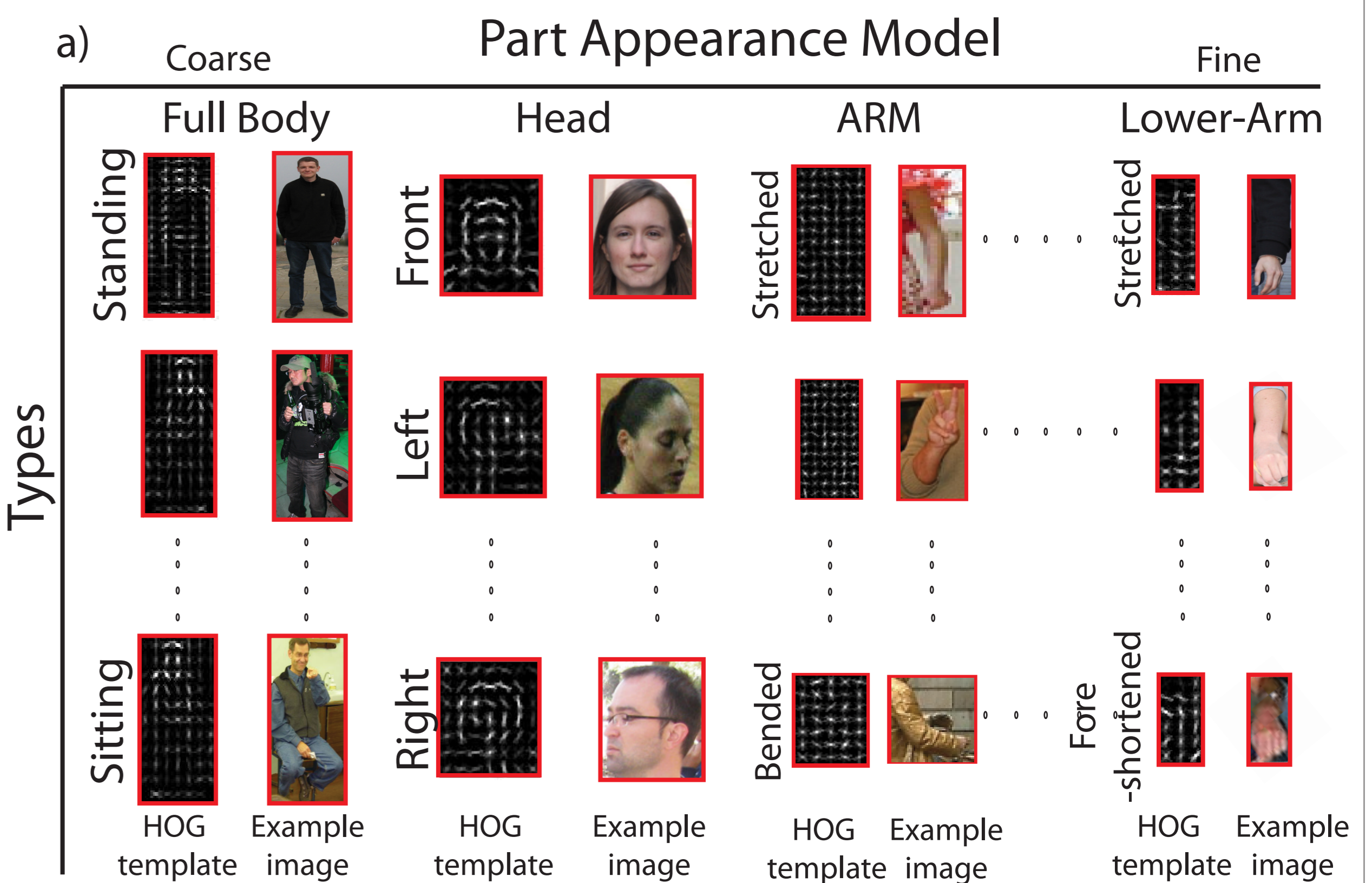
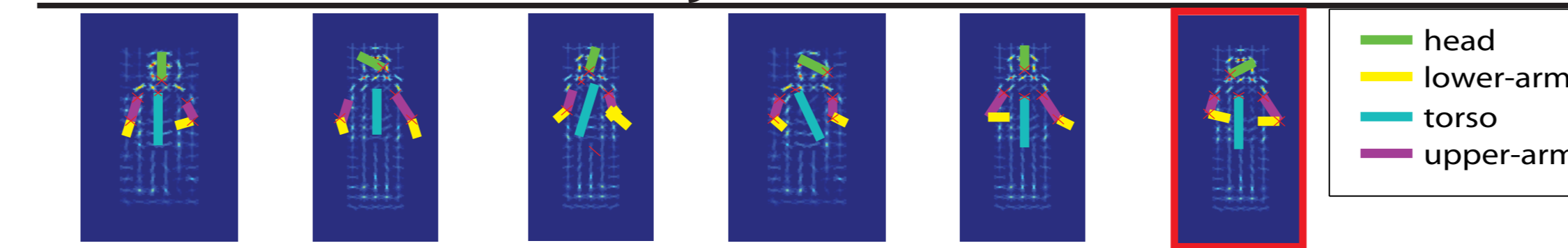
-Struct SVM

$$\min_{w, \xi^n > 0} w^T w + C \sum_n \xi^n(H) \quad \text{s. t.} \quad \xi^n(H) = \max_H (D(H; H^n) + w^T \Psi(H; I^n) - w^T \Psi(H^n; I^n))$$

b) Parent-Child Relationship

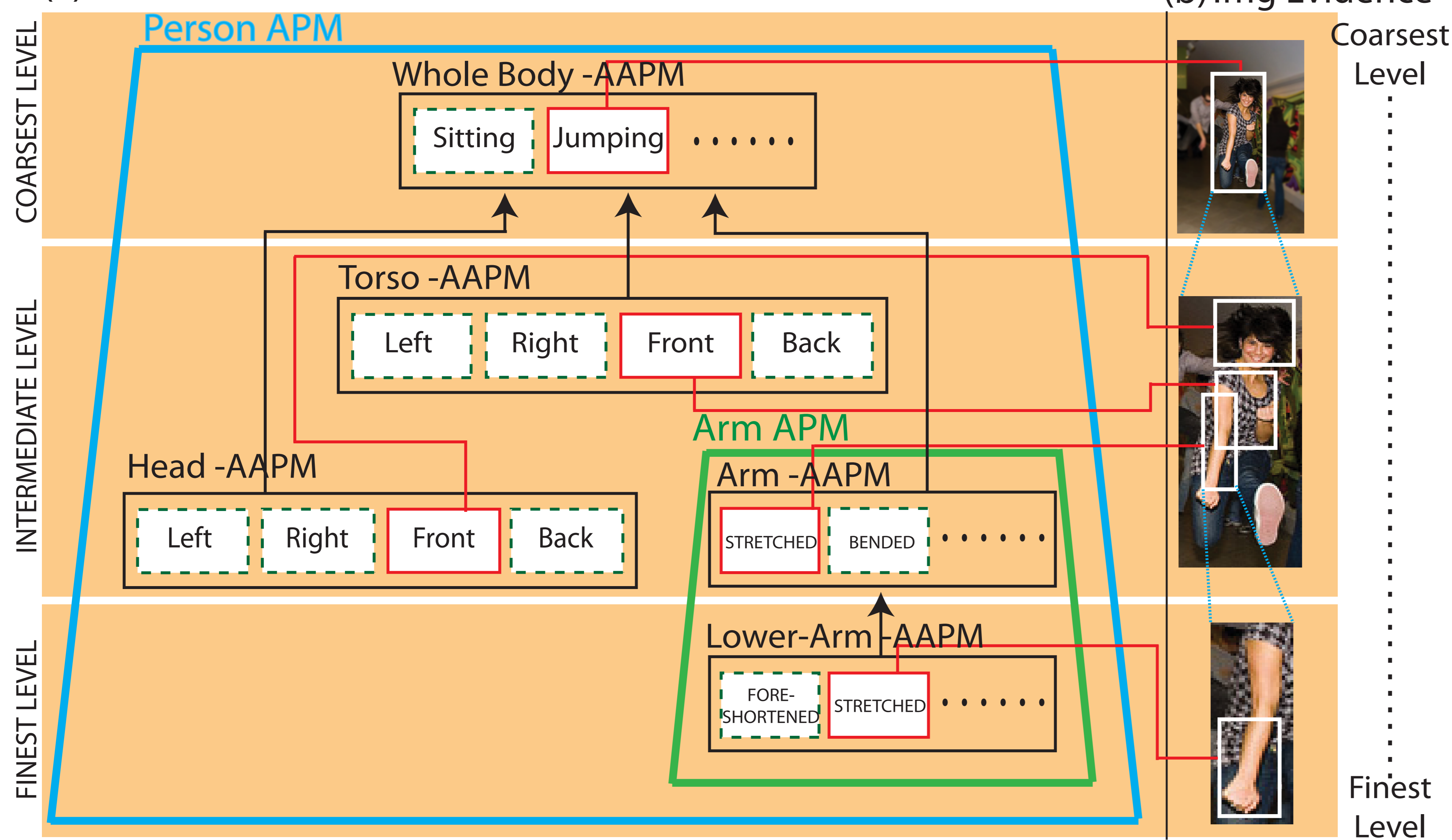


c) Object Poses



2. Model

(a) Model Structure



-Matching Score

- Appearance Score: $f^A(h; I) = A^T \psi_a(h, I)$

- Deformation Score: $f^D(h, \hat{h}) = -d^T \psi_d(h, \hat{h})$

-Score Aggregation

- Child location selection: $f_{c_s c}(\hat{h}_c, I) = \max_{h_c} f_{c_s c}(h_c, I) + f^D(h_c, \hat{h}_c)$

- Child alignment: $f_{c_s c}(T(h_i, t_{i_c}^{s_i, s_c}), I); T(h, t) = (x - t_x, y - t_y, L - t_L, \theta - t_\theta)$

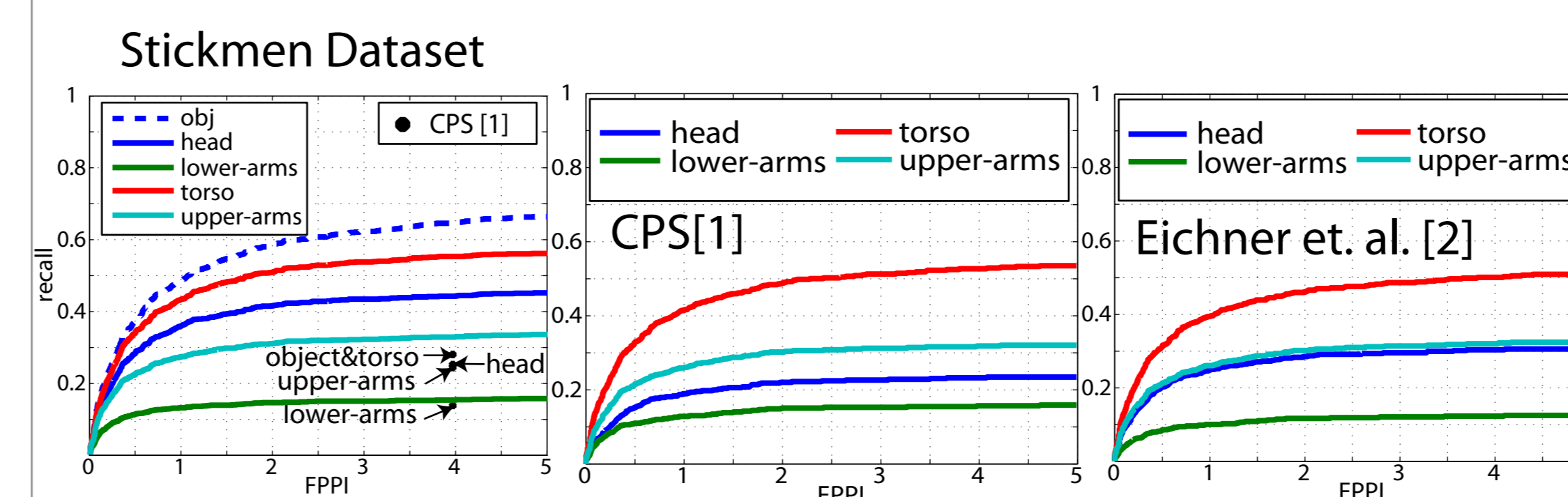
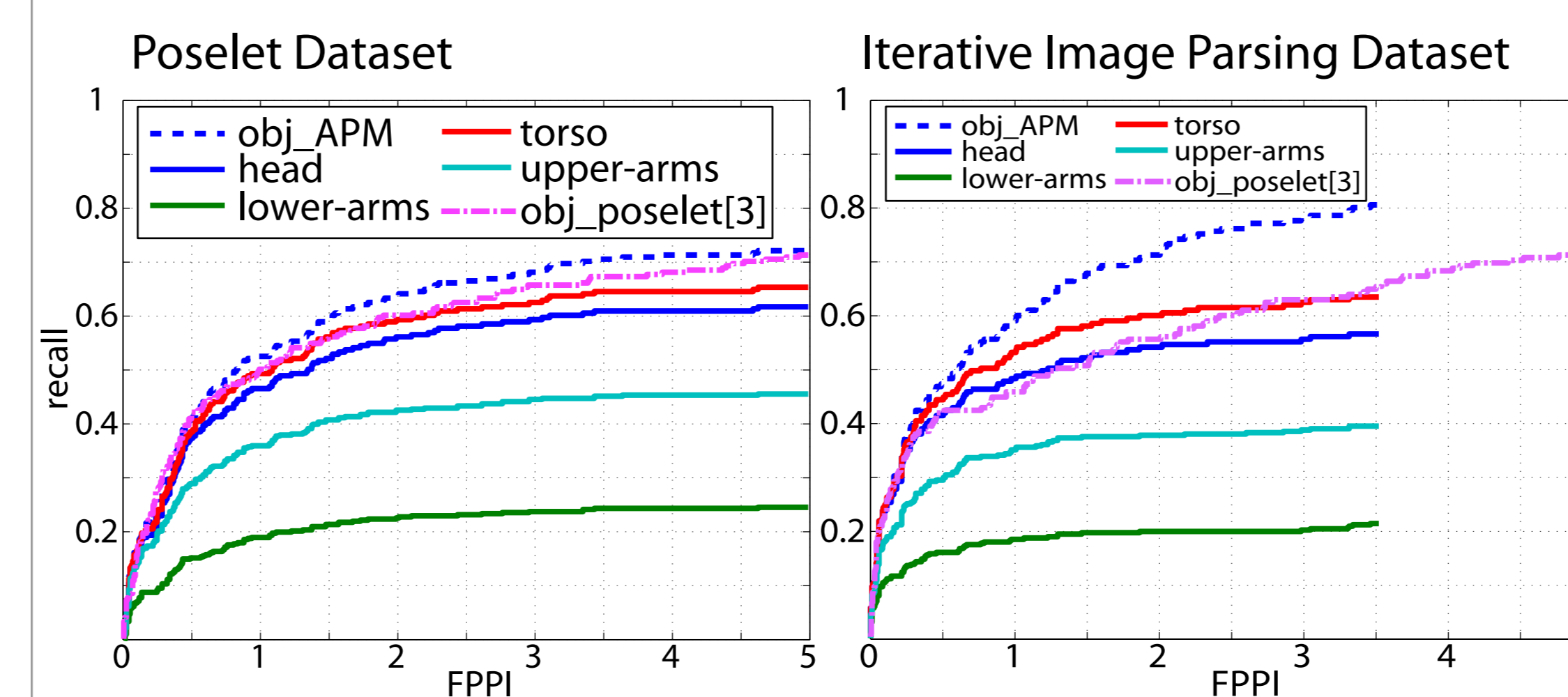
- Child type selection: $f_c(h_i, I) = \max_{s_c} f_{c_s c}(T(h_i, t_{i_c}^{s_i, s_c}), I) + b_{i_c}^{s_i, s_c}$

- Aggregation: $f_{i_s i}(h_i, I) = f_{i_s i}^A(h_i, I) + \sum_c f_c(h_i; I)$

-Model Properties

- Sublinearity
- Efficient Exact Inference

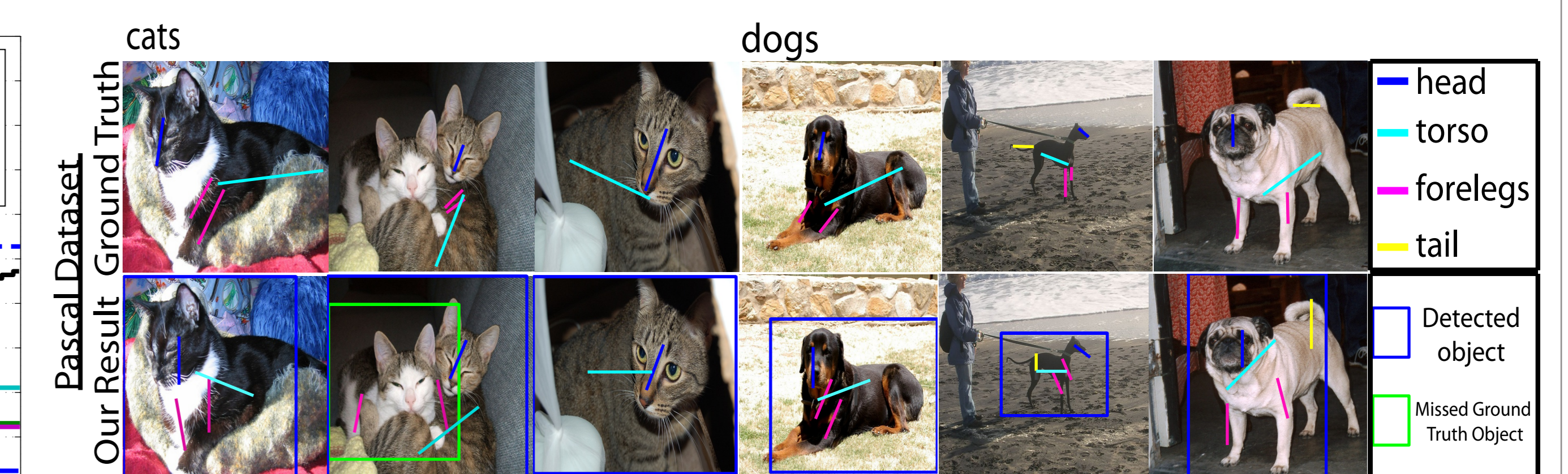
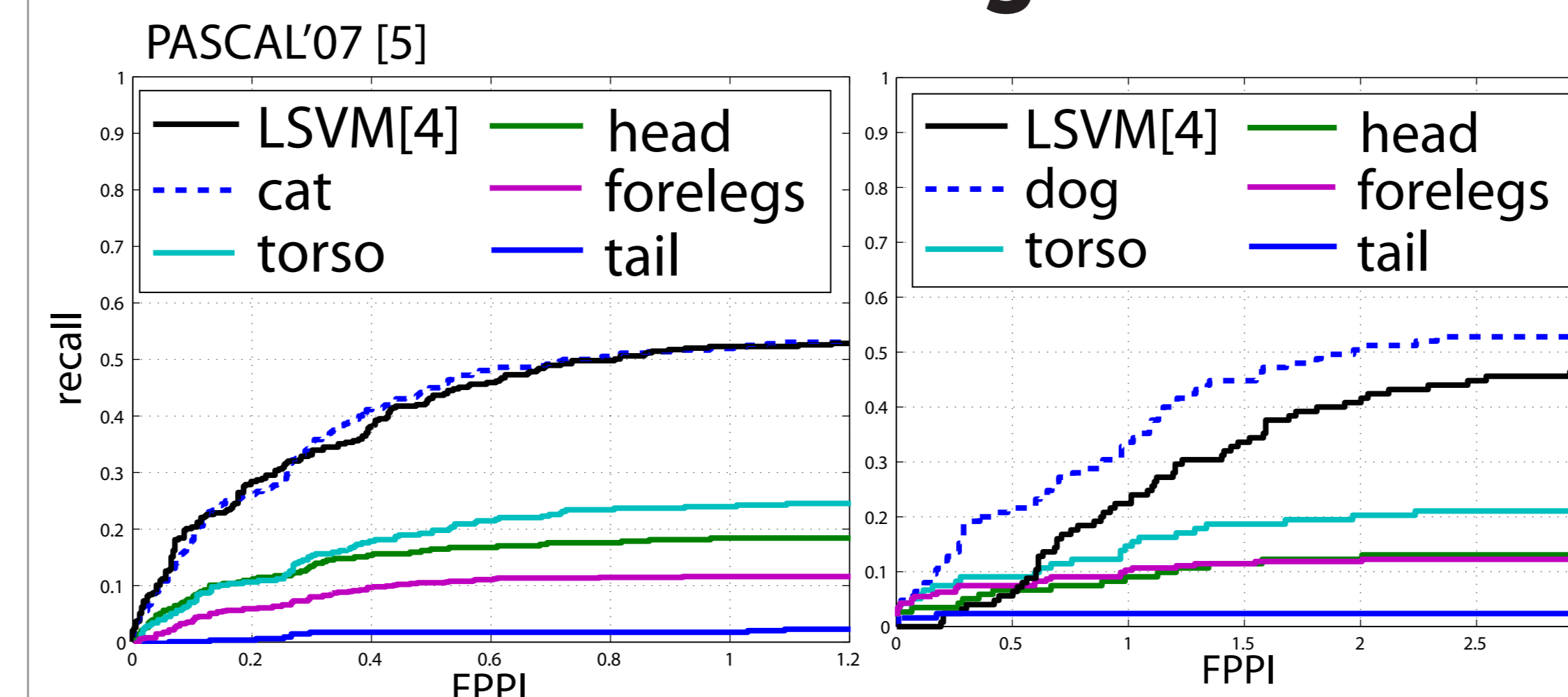
4. Results on Human Dataset



PASCAL stickmen Det.	Recall/PCP @ 4 FPPI				Recall obj
	torso	head	upper arm	lower arm	
Eichner et al[2]	0.497/77.44	0.311/48.430	0.318/49.520	0.122/19.06	
CPS[1]	0.525/81.80	0.231/35.920	0.316/49.270	0.155/24.150	0.642
APM (ours)	0.550/85.57	0.439/68.330	0.326/50.730	0.151/23.54	



5. Results on Cats & Dogs Dataset



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