Bird Part Localization Using Exemplar-Based Models with Enforced Pose and Subcategory Consistency

Jiongxin Liu and Peter N. Belhumeur
{luijx09, belhumeur}@cs.columbia.edu
Columbia University

Problem

The goal of our work is to localize the parts automatically and accurately for fine-grained categories. We evaluate our method on bird images in the CUB-200-2011 [1] dataset.

Approach

Does $X_{k,t}$ match the image $I$? $\iff P(X_{k,t}|I) =$?

$P(X_{k,t}|I) = P(X_{k,t}|D_p)^{\alpha} P(X_{k,t}|D_s)^{1-\alpha}$ (1)

$P(X_{k,t}|D_p) = G_{avg}\{P(x_{k,t}^i|d_{i,p}^c,s_{i,k,t})\}$ (2)

$P(X_{k,t}|D_s) = G_{avg}\{P(x_{k,t}^i|d_{i,s}^l,s_{i,k,t})\}$ (3)

$P(X_{k,t}|l,D_s) = \max_l P(X_{k,t}|l,D_s)$ (4)

We use the most likely models $M$ to predict the part locations of the testing sample:

$\hat{x}^i = \arg\max_{x^i} \sum_{k,t \in M} P(\Delta x_{k,t}^i) P(x^i|d_{i,p}^c,s_{i,k,t})$ (5)

Pose Detectors

For each pose cluster $c$ of part $i$, we build a detector. The detector scans the image over scales, and the response map of this detector at a particular scale $s^i$ is denoted as $d_{i,p}^c(s^i,s_{i,k,t})$.

Subcategory Detectors

For each species $l$ of part $i$, we build a detector after aligning the samples. Assuming the detector scans the image over scales and orientations, then the response map of this detector at a particular scale $s^i$ and orientation $\theta^r$ is denoted as $d_{i,s}^l(s^i,\theta_{i,k,t})$.

Enforcing Consistency

$P(x_{k,t}^i|d_{i,p}^c(s^i,s_{i,k,t})) \cdot P(x_{k,t}^i|d_{i,s}^l(s^i,\theta_{i,k,t}))$

Localization Examples

Comparisons

References
