Discovering Localized Attributes for Fine-grained Recognition

Kun Duan\textsuperscript{1}, Devi Parikh\textsuperscript{2}, David Crandall\textsuperscript{3}, and Kristen Grauman\textsuperscript{3}
\textsuperscript{1}Indiana University, \textsuperscript{2}Toyota Technological Institute at Chicago, \textsuperscript{3}University of Texas at Austin

1. Overview

- **Objective**: Given images with fine-grained category-labels, discover a vocabulary of localized attributes that are both semantically meaningful and discriminative.

- **Nodes**: images, **Labels**: segments

- **Implementaion details**:
  - **Attribute detection**: add test image to CRF and run inference
  - **Implementation details**:
    - **Generating regions**: hierarchical segmentation
    - **Region features**: color, pGB contour, size, shape, and spatial location
    - **Distances**: L2 for spatial location, chi-squared for other features

- **We employ an iterative and interactive approach**:
  - A. Find discriminative attribute candidates for 2 similar classes with a latent CRF
  - B. Use a recommender system to identify candidates likely to be meaningful
  - C. Present them to a human for naming and verification. Repeat.

2. Discovering Localized Attributes with Latent CRFs

- **Goal**: Find discriminative local attribute candidates given M images from two categories
- **Define a latent CRF to find regions in positive images that are similar to one another but dissimilar from negative image regions**

3. Attribute Discovery Results

- **Leeds Butterfly (10 categories, 232 images)**
- **Three subsets of Caltech-UCSD Birds 200 (60 images per category)**:
  - Warbler (5 categories)
  - Random (10 categories)
  - Hard (10 categories)

4. Image-to-text Generation Results

- **Our approach can annotate unseen images with region labels using our discovered local attributes**:

5. Image Classification Results

- **We run attribute detection to produce a binary feature vector for each image, then use these vectors for fine-grained recognition using Nearest Neighbor and SVM classifiers**.
- **We compare four approaches to generate localized attributes**:
  - **Proposed**: our proposed method that focuses on discriminative power and semantics
  - **Discriminative only**: focus on discrimination, non-semantic attributes removed in post process
  - **Upper bound**: price paid for semantics, all discriminative candidates including non-semantic ones

6. Summary and Conclusions

- **Recommender gathers more attributes and achieves higher accuracy for same amount of user effort**
- **Our proposed method performs significantly better than existing approaches**