Discovering Localized Attributes for Fine-grained Recognition

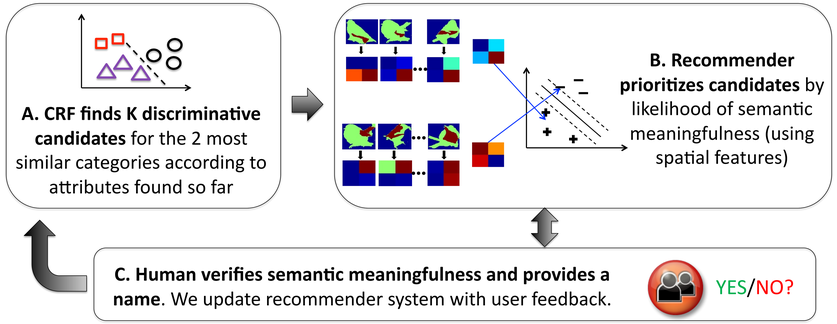
Kun Duan[1], Devi Parikh[2], David Crandall[1], and Kristen Grauman[3]

[1]Indiana University, [2]Toyota Technological Institute at Chicago, [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**

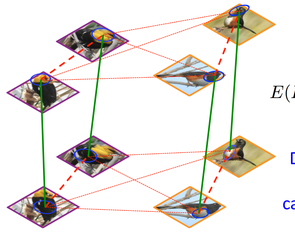
- **We employ an iterative and interactive approach:**
  A. Find discriminative attribute candidates for 2 similar classes with a **latent CRF**
  B. Use a **recommender system** to identify candidates likely to be meaningful
  C. Present them to a human for naming and verification. Repeat.



2. Discovering Localized Attributes with Latent CRFs

- **Goal:** Find K 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

Nodes: images, Labels: segments



For candidate k, find regions $L_k = (l_1^k, \ldots, l_M^k)$ that minimize,

$$E(L_k|\mathcal{I}) = \sum_{i=1}^{M} \phi_k(l_i^k|\mathcal{I}_i) + \sum_{i=1}^{M}\sum_{j=1}^{M} \psi_k(l_i^k, l_j^k|\mathcal{I}_i, \mathcal{I}_j)$$

Discriminative score of regions on category classification

Pairwise similarity or distance between regions

- Encourage diversity across multiple candidates:

$$E(\mathcal{L}|\mathcal{I}) = \sum_{k=1}^{K} E(L_k|\mathcal{I}) + \sum_{i=1}^{M}\sum_{k,k'} \delta(l_i^k, l_i^{k'}|\mathcal{I}_i)$$

Spatial overlap between regions

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

3. Attribute Discovery Results

- **Leeds Butterfly** (10 categories, 832 images)

- Three subsets of **Caltech-UCSD Birds 200** (60 images per category):

Warbler (5 categories)

Random (10 categories)

Hard (10 categories)

Local attributes discovered by our approach with names provided by users (MTurk)

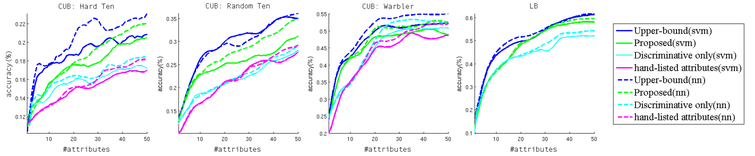
4. Image-to-text Generation Results

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

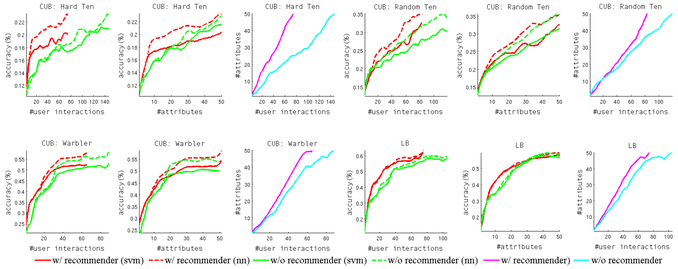
5. Image Classification Results

- 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
  - **Hand-listed** (focus on semantics): expert-generated attributes
  - **Discriminative only** (focus on discrimination): non-semantic candidates removed in post-process
  - **Upper bound** (price paid for semantics): *all* discriminative candidates including non-semantic ones

Our proposed method performs significantly better than existing approaches



Recommender gathers more attributes and achieves higher accuracy for same amount of user effort



6. Summary and Conclusions

- Finds local attributes that are both discriminative and human understandable.
- Recommender system prioritizes candidates likely to be meaningful, saving user time.
- Compares favorably to existing attribute discovery approaches.
- More information at **http://vision.soic.indiana.edu/attributediscovery**