How do you tell a blackbird from a crow?

Thomas Berg and Peter N. Belhumeur, Columbia University

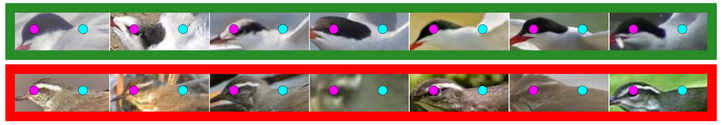
Goal: To Understand a Visual Domain (*e.g.* Birds)

In a domain of fine-grained visual categorization:
- Which classes (species) are most visually similar to each other?
- What distinguishes similar classes from each other?
- Is there a natural hierarchy that relates the classes to each other?

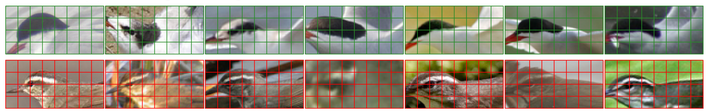
Part-based One-vs-One Features (POOFs) [1]

1. Choose two classes (e.g. *i* = common tern and *j* = Louisiana waterthrush)

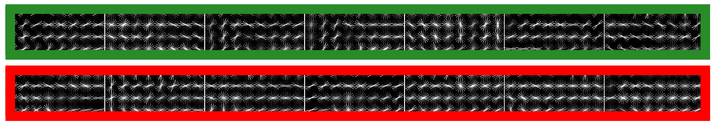
2. Choose a feature part and an alignment part (e.g. *f* = eye and *a* = back), align and crop



3. Divide cropped images into grids



4. Extract base features (e.g. *b* = gradient direction histograms (shown) or color histograms)



5. Train a linear SVM to separate the classes, then threshold the weights and retrain on just the discriminative region to get a POOF.

$$\Rightarrow T^{i,j}_{f,a,b}$$

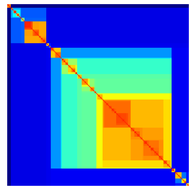
7. Repeat (e.g. with CUB-200, can build millions of POOFs; we use 5000).

$$\binom{200}{2}\text{class pairs} \cdot (12 \cdot 11)\text{part pairs} \cdot 2\,\text{base features}$$
$$= 5,253,600\,\text{POOFs}$$

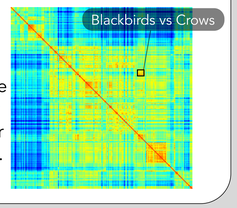
[1] T. Berg and P. N. Belhumeur, "Part-based One-vs-One Features for Fine-grained Categorization, Face Verification, and Attribute Estimation," CVPR 2013
[2] N. Saitou and M. Nei, "The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees," Molecular Biology and Evolution, 4(4), 1987

Visual Similarity

Phylogenetic Similarity
Based on time since split from most recent common ancestor.

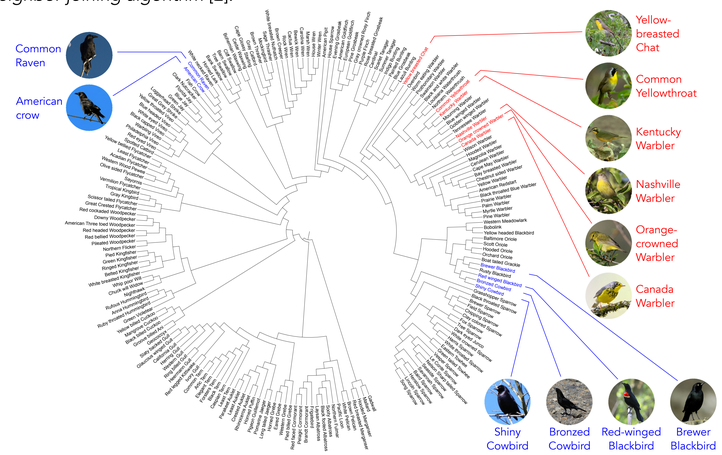


Visual Similarity
Based on L1-distance in POOF space after projection with linear discriminant analysis.

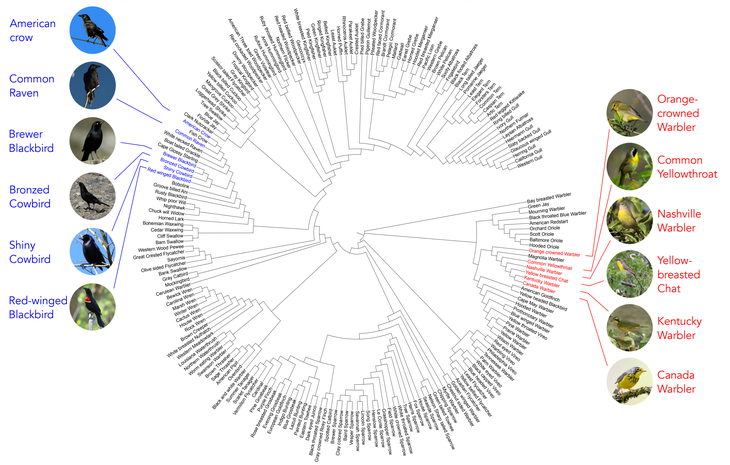


A Tree of Similarity

The phylogenetic *tree of life* can be estimated from the genetic similarity matrix using the *neighbor-joining algorithm* [2].



Use the same algorithm with our visual similarity matrix to get a tree of visual similarity.



This work was supported by ONR award N00014-08-1-0638, Moore Foundation grant 2987, and NSF grant 1116631 .

Illustrating Visual Differences

To find differences between classes *i* and *j*, find the most discriminative POOFs.

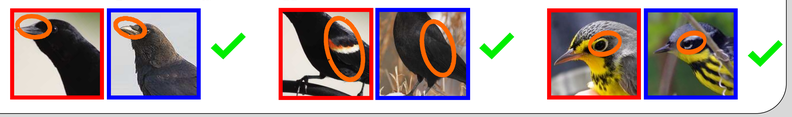
$$T^{i,j}_{\hat{f},\hat{a},\hat{b}} = \text{argmax}\frac{(\mu_2 - \mu_1)^2}{\sigma_1\sigma_2}$$

Show the support region of a POOF as a level set of a Gaussian fit to the SVM weights.

To illustrate a POOF, T, choose image pairs to minimize three objectives:

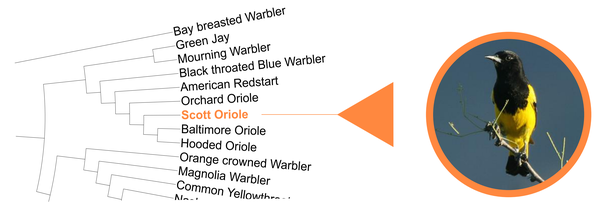
1. Prefer strong, but not extreme, values of T.
   $F(I_1, I_2) = (1 + |T(I_1) - b_1|)(1 + |T(I_2) - b_2|)$
   where b_1, b_2 are 75th and 25 percentile of T scores.
2. Prefer similar scores for non-(*i*,*j*) POOFs.
   $G(I_1, I_2) = ||\mathbf{g}(I_1) - \mathbf{g}(I_2)||$
   where $\mathbf{g}(I)$ is the vector of non-(*i*,*j*) POOF scores on image I.
3. Prefer similar poses.
   $H(I, J)$ = sum of squared part location differences

Minimize $k_F F + k_G G + k_H H$.



Generating a Visual Field Guide

Users scroll through the tree of similarity to find something that looks like what they saw.



Connect each species to its *k* most similar species. Users navigate through the graph.

For any pair of similar species, show the POOFs that distinguish them from each other.