POOF: **P**art-Based **O**ne-vs-**O**ne **F**eatures for Fine-Grained Categorization, Face Verification, and Attribute Estimation

Thomas Berg and Peter N. Belhumeur, Columbia University

POOFs

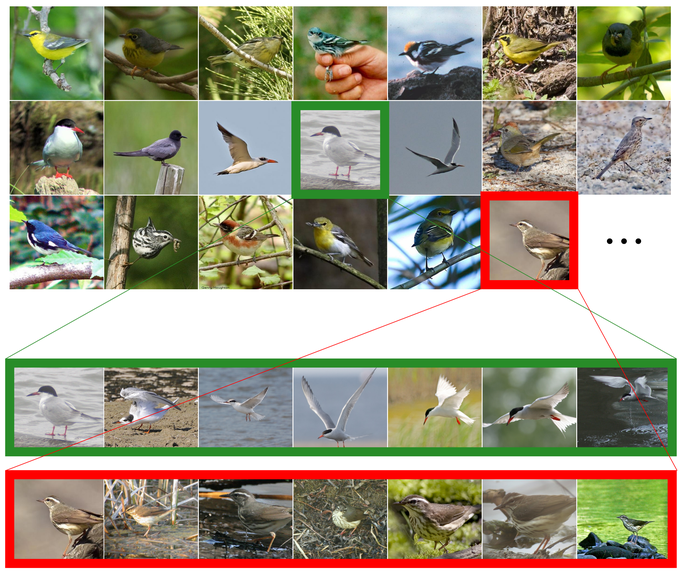
Part-based One-vs-One Features (POOFs) are high-level features well-suited to fine-grained categorization. They are:

- Learned in a ***fully automatic*** way, given a set of images with category and part labels.
- ***Discriminative***, based on SVMs.
- ***Diverse***, each based on a pair of categories *(i, j)*, pair of parts *(f, a)*, and base feature *(b)*.

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

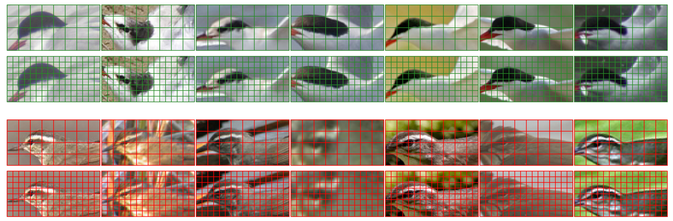
Construction

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

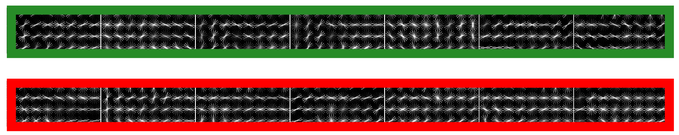


2. Choose a feature part and an alignment part (e.g. f = eye and a = back), align and crop

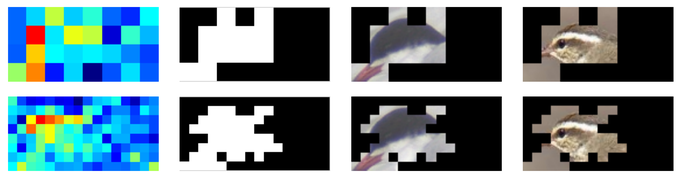
3. Divide cropped images into grids at multiple scales



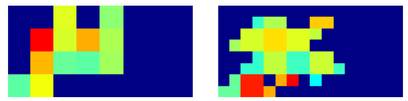
4. Extract base features (e.g. b = gradient direction histograms (shown) or color histograms)



5. Train a linear SVM to separate the classes, then threshold the weights to get the discriminative region



6. Retrain linear SVM on discriminative region to get a POOF



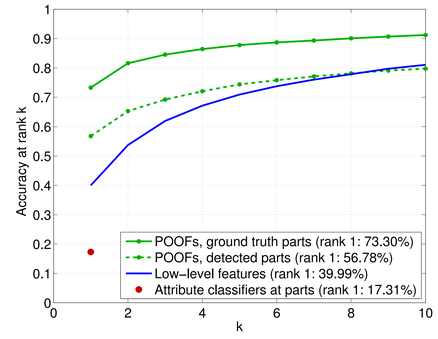
7. Repeat (e.g. with CUB-200, can build millions of POOFs, but thousands suffice)

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

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

Experiment: Bird Identification

Classification Accuracy



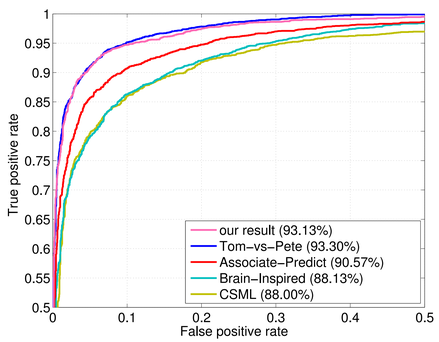
Mean Average Precision

	14 cat.	200 cat.
Zhang et al., GT parts	57.44	28.18
POOFs, det. parts	70.16	56.89
POOFs, GT parts	85.57	74.32

The CUB-200-2011 dataset has images of 200 bird species, with about 30 training images per species. Birds are labeled with a bounding box and 15 part locations (beak, eye, wing, etc.).

We learn 5000 POOFs, which are fed into 200 one-vs-all species classifiers (linear SVMs).

Experiment: Face Verification



LFW has 6000 pairs of faces, half "same person" pairs and half "different person," in ten 600-pair cross-validation folds. Our separate reference dataset contains images of 120 people not in LFW.

We learn 10,000 POOFs from the reference set, getting a 10,000-dimensional POOF vector *f(I)* for each face image *I* in LFW. Each face pair *(I, J)* is represented by vector

$$(|f(I) - f(J)|, f(I) \odot f(J)).$$

These are fed into a same-vs-different classifier (linear SVM).

Experiment: Attribute Estimation

Attribute	Method	Number of training samples 6	20	60	200	600	Kumar et al.
Male	low-level feat.	50.7	61.0	66.9	81.4	87.8	90.5
	POOFs	**86.2**	**89.9**	**89.7**	**91.3**	**91.7**	
Asian	low-level feat.	53.9	53.9	68.4	78.2	83.2	86.5
	POOFs	**75.2**	**75.8**	**84.3**	**87.6**	**89.8**	
...	...	...	...	...	...	...	...
Average improvement		12.3	13.4	8.0	4.3	2.7	2.8

Using the same POOFs as above, we train attribute classifiers (linear SVMs) using the 10,000-dimensional POOF vectors.