How Do You Tell a Blackbird from a Crow?

Thomas Berg and Peter N. Belhumeur
Columbia University
{tberg,belhumeur}@cs.columbia.edu

Abstract

How do you tell a blackbird from a crow? There has been great progress toward automatic methods for visual recognition, including fine-grained visual categorization in which the classes to be distinguished are very similar. In a task such as bird species recognition, automatic recognition systems can now exceed the performance of non-experts – most people are challenged to name a couple dozen bird species, let alone identify them. This leads us to the question, "Can a recognition system show humans what to look for when identifying classes (in this case birds)?" In the context of fine-grained visual categorization, we show that we can automatically determine which classes are most visually similar, discover what visual features distinguish very similar classes, and illustrate the key features in a way meaningful to humans. Running these methods on a dataset of bird images, we can generate a visual field guide to birds which includes a tree of similarity that displays the similarity relations between all species, pages for each species showing the most similar other species, and pages for each pair of similar species illustrating their differences.

1. Introduction

How do you tell a blackbird from a crow? To answer this question, we may consult a guidebook (*e.g.*, [22, 23]). The best of these guides, products of great expertise and effort, include multiple drawings or paintings (in different poses and plumages) of each species, text descriptions of key features, and notes on behavior, range, and voice.

From a computer vision standpoint, this is in the domain of *fine-grained visual categorization*, in which we must recognize a set of similar classes and distinguish them from each other. To contrast this with general object recognition, we must distinguish blackbirds from crows rather than birds from bicycles. There is good, recent progress on this problem, including work on bird species identification in particular (*e.g.*, [1, 29]). These methods learn classifiers which can (to some standard of accuracy) recognize bird species but do not explicitly tell us what to look for to recognize

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

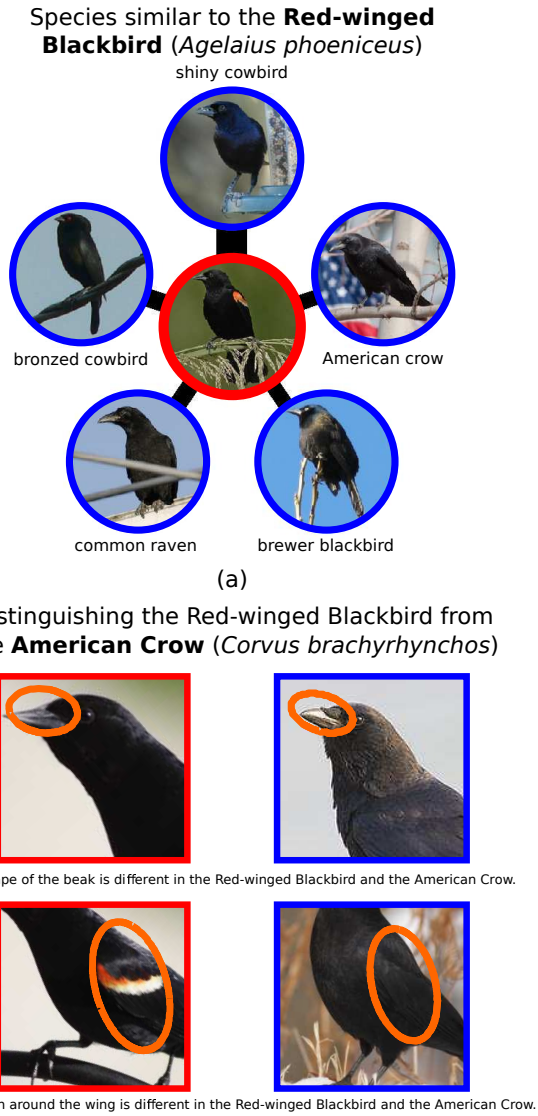


Figure 1. (a) For any bird species (here the red-winged blackbird, at center), we display the other species with most similar appearance. More similar species are shown with wider spokes. (b) For each similar species (here the American crow), we generate a "visual field guide" page highlighting differences between the species.

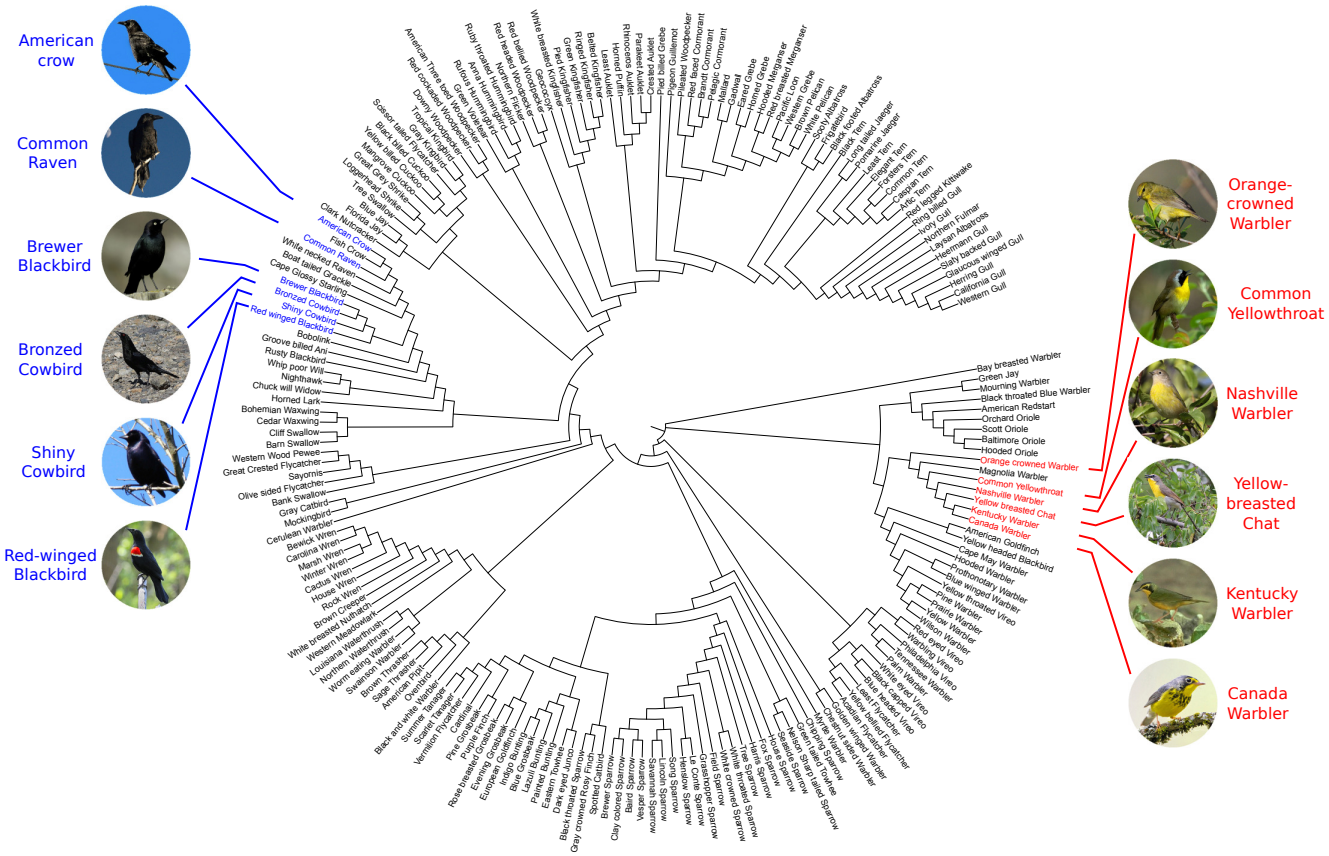


Figure 2. A similarity tree of bird species, built from our visual similarity matrix using the neighbor-joining method of [20]. Species visually similar to the red-winged blackbird (see Figure 1) are in blue, and those similar to the Kentucky warbler (see Figure 4) are in red.

bird species on our own.

In this paper, we consider the problem not of performing fine-grained categorization by computer, but of using computer vision techniques to show a human how to perform the categorization. We do this by learning which classes appear similar, discovering features that discriminate between similar classes, and illustrating these features with a series of carefully chosen sample images annotated to indicate the relevant features. We can assemble these visualizations into an automatically-generated digital field guide to birds, showing which species are similar and what a birder should look for to tell them apart. Example figures from a page we have generated for such a guide are shown in Figure 1.

In addition to the visualizations in these figures, we borrow a technique from phylogenetics, the study of the evolutionary relations between species, to generate a tree of visual similarity. Arranged in a wheel, as shown in Figure 2, this tree is suitable as a browsing interface for the field guide, allowing a user to quickly see each species and all the species similar to it. We compare our similarity-based tree with the phylogenetic "tree of life," which describes the evolutionary relations between bird species. Places where the trees are not in agreement – pairs of species that are close in the similarity tree but far in the evolutionary tree – are of special interest, as these may be examples of *convergent evolution* [11], where similar traits arise independently in species that are not closely related.

We base our similarity calculations on the "part-based one-vs-one features" (POOFs) of [1], for two reasons. First is the POOFs' strong performance on fine-grained categorization; in particular they have done well on bird species recognition. Second is their part-based nature. Fine-grained classification encourages part-based approaches because the classes under consideration have many of the same parts (for birds, every species has a beak, wings, legs, and eyes) so it is possible to distinguish classes by the appearance of corresponding parts. Experiments by Tversky and Hemenway [24] suggest that people also use properties of parts to distinguish similar classes, and bird guides often describe species in terms of their parts, as shown in Figure 3. All this suggests that part-based features may be the best way to show humans the key features. POOFs have the additional advantage of being easy to illustrate; each is associated with a learned support region that can be highlighted in our visualizations.

This paper makes the following contributions:

1. We propose and explore a new problem: using com-

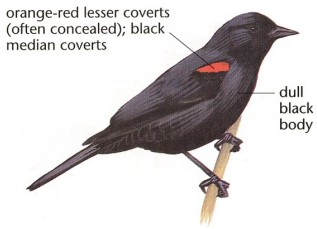


Figure 3. A picture of the red-winged blackbird from the Sibley Guide [22] shows part-based features that distinguish this species.

puter vision techniques, in particular methods of fine-grained visual categorization, to illustrate the differences between similar classes. The goal is not to perform identification, but to see if we can show a human what to look for when performing identification.

2. We propose an approach to this problem. We demonstrate a fully automatic method for choosing, from a large set of part-based features, those that best show the difference between two similar classes, choosing sample images that exemplify the difference, and annotating the images to show the distinguishing feature.
3. We explore the relation between visual similarity and phylogeny in bird species. Species which are visually similar but not close in the evolutionary "tree of life" may be examples of convergent evolution.

2. Related Work

There is a good deal of recent work on fine-grained categorization, much of it dealing with species or breed recognition of *e.g.*, trees [13], flowers [16], butterflies [8, 27], dogs [15, 18, 19], and birds [1, 6, 8, 9, 29, 30]. All of this work uses part-based features in one way or another, although it is mostly concerned with performing recognition rather than explaining *how* to perform recognition. We use the POOFs of [1] for reasons discussed in Section 1.

Work on fine-grained visual categorization with "humans in the loop" [3, 25], proposes a model in which a person answers questions to assist a recognition system that makes a final identification. Our proposal is conceptually opposite: the recognition system shows what features to look for, which will help a person to perform identification.

Our goal is not classification itself, but an understanding of what features are most relevant and understandable. A similar task is set by Doersch *et al*. [7], who discover the architectural features best suited to recognizing the city in which a street scene was photographed. With a much smaller dataset and a much larger number of classes, we take a careful approach based on labeled parts rather than their random image patches. Shrivastava *et al*. [21] weight regions in an image by their distinctiveness for purposes of cross-domain similar image search. This is similar to our method for finding regions to annotate in our illustrative images, but they work with a single image to find its *distinctive* regions, while we work with two classes of image to find the most *discriminative* regions. Both [7] and [21] deal with image rather than object classification, so use unaligned image patches rather than our part-based features. Deng *et al*. [6] found discriminative regions in bird by explicit human labeling in the guise of a game.

Although we take a part-based approach to allow us to annotate our images, there is also non-part-based work that attempts to describe the features of a class. Parikh and Grauman [17] discover discriminative image attributes and ask users to name them. Yanai and Barnard [28] consider the opposite problem, starting with a named concept and learning whether or not it is a visual attribute, while Berg *et al*. [2] discover attribute names and classifiers from web data. This could be used to provide supplementary, non-part-based text descriptions of species differences in our visual field guide.

3. Visual Similarity

Our goal is, in a set of visually similar classes, to determine which classes are most similar to each other, and among those most similar classes, to understand and visualize what it is that still distinguishes them.

To make this concrete, we consider the problem of bird species recognition, using the Caltech-UCSD Birds 200 dataset (CUBS-200) [26]. We use the 2011 version of the dataset, which includes 11,788 images of 200 bird species, annotated with the locations of 15 parts (beak, forehead, crown, throat, breast, belly, nape, back, tail, and left and right eyes, wings, and legs). The dataset is divided into training and test sets of roughly equal size. With many examples of species with similar appearance, and also many species with widely varying appearance, the dataset presents a difficult recognition task.

3.1. A Vocabulary of Part-based One-vs-One Features (POOFs)

The first step toward our goal is to construct a vocabulary of features suitable for differentiating among classes in our domain. For this we use a collection of POOFs [1], which we describe briefly here.

Each POOF is defined by the choice of two classes, two parts (a "feature part" and an "alignment part"), and a base feature, in our case either a color histogram or a histogram of oriented gradients (HOG) [5]. For example, a POOF may discriminate between the *red-winged blackbird* and the *rusty blackbird*, based on *color histograms* at the *wing* after alignment by the wing and the *eye*.

To build a POOF, all the training images of the two classes are rotated and scaled to put the two chosen parts in fixed locations. Each image is then cropped to a fixed-size rectangle enclosing the two parts, and the cropped image partitioned into a grid of square tiles. We extract the base feature from each tile, concatenate these features to get a feature vector for the image, and train a linear SVM

to distinguish the two classes. Tiles with low SVM weights are discarded, and a connected component of the remaining tiles about the feature part is taken as the support region for the POOF. The SVM is retrained using the base feature on just this region to get the final classifier. We use the parameter settings from [1] unaltered: images are aligned to put the two parts on a horizontal line with 64 pixels between them, the crop is 128 pixels wide and 64 pixels tall, and we use two grids, of 8 x 8 and 16 x 16 pixel tiles. The output of the POOF is the signed distance from the decision boundary of the classifier, and while each POOF is trained on just two classes, [1] show that a collection of these POOF outputs is an effective feature vector for distinguishing other subclasses of the same basic class (*e.g.*, other species of birds). With 200 classes, fifteen parts, and two base features, we can train millions of POOFs, although in practice we will always use a subset.

POOFs are suited to our task for two reasons. First of all, they have been shown to be effective at fine-grained categorization. Second, and of special importance to us, POOFs are relatively easy to interpret. If we discover that two bird species are well-separated by a color histogram-based POOF aligned by the beak and the back, and the SVM weights are large at the grid cells around the beak, we can interpret this as "These two species are differentiated by the color of the beak." This kind of understanding is our goal.

3.2. Finding Similar Classes

Few would confuse a cardinal and a pelican. It would be difficult and not useful to describe the particular features that distinguish them; any feature you care to look at will suffice. The interesting problem is to find what details distinguish classes of similar appearance. To do this we must first determine which classes are similar to each other.

Our starting point is our vocabulary of POOFs. For efficiency we take a subset of 5000 POOFs, so each image is describe by the 5000-dimensional vector of their outputs. An L1 or L2 distance-based similarity in this space is appealing for its simplicity, but considers all features to be equally important, which is unlikely to be a good idea. The POOFs are based on random classes and parts. Some of these choices will be good, looking at two species that differ in a clear way at the parts being considered. Others will look at parts that are not informative about those two classes. We wish to downweight features that are not discriminative, and emphasize those that are. A standard tool for this is linear discriminant analysis (LDA) [10], which, from a labeled set of samples with n classes, learns a projection to an $n-1$ dimensional space that minimizes the ratio of within-class variance to between-class variance. We apply LDA, and use the negative L1 distance in the resulting 199-dimensional space as a similarity measure.

By applying this image similarity measure to mean feature vectors over all the images in a class, we obtain a similarity measure between classes, with which we can determine the most similar class to any given class. The red-winged blackbird and its five most similar species are shown at the top of Figure 1.

3.3. Choosing Discriminative Features

Given a pair of very similar classes, we are now interested in discovering what features can be used to tell them apart. We consider as candidates all the features from our vocabulary that are based on this pair of classes. With the birds dataset, with twelve parts and two low-level features, there are 264 candidate features. We rank the features by their *discriminativeness*, defining the discriminativeness of feature f as

$$d_f = \frac{(\mu_2 - \mu_1)^2}{\sigma_1 \sigma_2}, \tag{1}$$

where μ_1 and μ_2 are the mean feature values for the two classes, and σ_1 and σ_2 are the corresponding standard deviations. Maximizing discriminativeness is similar in spirit to the optimization performed by LDA, which maximizes interclass variation and minimizes intraclass variation. Here we seek a individual score for each feature rather than a projection of the feature space, as it allows us to report particular features as "most discriminative."

3.4. Visualizing the Features

Once we have determined which features are most useful to distinguish between a pair of classes, we would like to present this information in a format that will help a viewer understand what he should look for. We present each feature as a pair of illustrative images, one from each species, with the region of interest indicated in the two images.

The first step is to choose the illustrative images. In doing this, we have several goals:

1. The images should exemplify the difference they are intended to illustrate. If the feature is beak color, where one class has a yellow beak and the other gray, then the images must have the beak clearly visible, with the beak distinctly yellow in one and gray in the other.
2. The images should minimize differences other than the one they are intended to illustrate. If the yellow and gray-beaked species above can both be either brown or black, it is misleading to show one brown and one black, as this difference does not distinguish the classes.
3. To facilitate direct comparison of the feature, the two samples should have their parts in similar configurations, *i.e.*, the birds should be in the same pose.

We translate these three goals directly into three objective expressions to be minimized. For the first, we take the view that the images should be somewhat farther from the decision boundary than average for their class, but not too

far. This corresponds to the feature being somewhat exaggerated, but avoids extreme values from the POOF which may be outliers or particularly unusual in some way. Taking c_1 and c_2 as the classes associated with positive and negative feature values respectively, let b_1 be the 75$^{\text{th}}$ percentile of feature values on c_1, and let b_2 be the 25$^{\text{th}}$ percentile of feature values on c_2. We take these exaggerated, but not extreme feature values as "best," and attempt to minimize

$$F(I_1, I_2) = (1 + |f(I_1) - b_1|)\,(1 + |f(I_2) - b_2|)\,, \quad (2)$$

where I_1 and I_2 are the candidate illustrative images from classes c_1 and c_2, and $f()$ is the feature to be illustrated.

To achieve the second goal, we consider an additional set of features, based on POOFs trained on classes other than c_1 and c_2. We use the 5000 POOFs used to determine interclass similarity in Section 3.2, less those with the same feature part as the POOF to be illustrated, and attempt to minimize the L1 distance between the resulting "other feature" vectors $\mathbf{g}(I_1)$ and $\mathbf{g}(I_2)$.

$$G(I_1, I_2) = ||\mathbf{g}(I_1) - \mathbf{g}(I_2)||_1 \quad (3)$$

To achieve the third goal, we consider the part locations in the two images. We resize the images so that in each, the mean squared distance between parts is 1, then find the best fit similarity transformation from the scaled locations $\mathbf{x}_1$ in image I_1 to the scaled locations $\mathbf{x}_2$ in image I_2. We minimize the squared error of the transformation, which we denote $H(I_1, I_2)$. Overall, we choose the image pair that minimizes

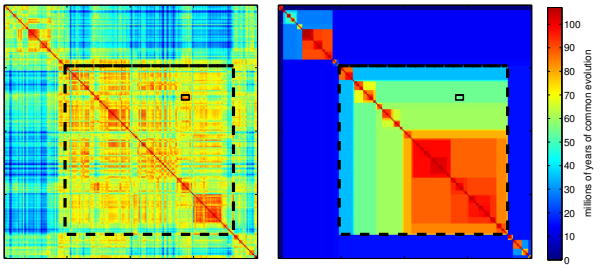
$$k_F F(I_1, I_2) + k_G G(I_1, I_2) + k_H H(I_1, I_2), \quad (4)$$

where coefficients k_F, k_G, and k_H determine the importance of each objective. To make them equally important, we set each to the multiplicative inverse of the standard deviation of its term, *i.e.*, $k_F = \frac{1}{\sigma_F}$, $k_G = \frac{1}{\sigma_G}$, and $k_H = \frac{1}{\sigma_H}$.

The second step in visualizing the features is annotating the chosen images to indicate the feature in question. Recall that the feature is the output of a POOF, which at its core is a vector of weights to be applied to a base feature extracted over a spatial grid. By taking the norm of the sub-vector of weights corresponding to each grid cell, we obtain a measure of the importance of each cell. An ellipse fit to the grid cells with weight above a small threshold then illustrates the feature.

4. A Visual Field Guide to Birds

As a direct application of the techniques in Section 3, we can construct a visual field guide to birds. While this guide will not have the notes on habitat and behavior of a traditional guide, it will have a couple advantages. First, it is automatically generated, and so could easily be built for another domain where guides may not be available. Second, it can be in some sense more comprehensive. While a traditional, hand-assembled guide will have an entry for each species, it is not combinatorially feasible to produce an entry on the differences between every *pair* of species. For an automatically-generated, digital guide, this is not an issue.



(a) (b)

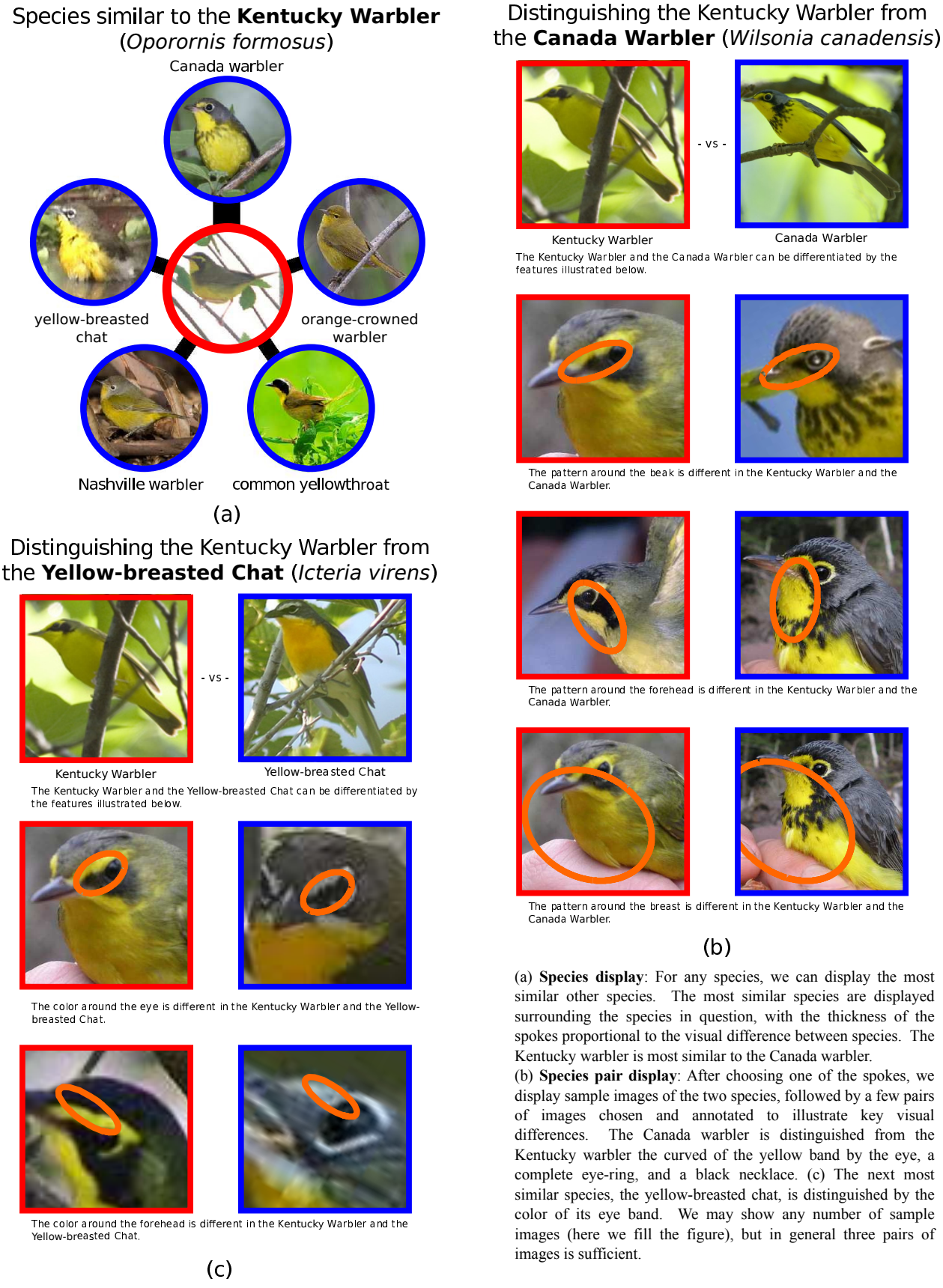
Figure 6. Similarity matrices. (a) Visual similarity. (b) Phylogenetic similarity. In both, rows/columns are in order of a depth-first traversal of the evolutionary tree, ensuring a clear structure in (b). The large dashed black box corresponds to the passerine birds ("perching birds," mostly songbirds), while the small solid black box holds similarities between crows and ravens on the y-axis and blackbirds and cowbirds on the x-axis.

We envision our field guide with a main entry for each species. Examples are shown in Figures 1 (a) and 4 (a). The main entry shows the species in question, and the top k most similar other species (we use $k = 5$) as determined by the method of Section 3.2. Selecting one of the similar species will lead to a pair entry illustrating the differences between the two species as described in Sections 3.3 and 3.4. Figures 1 (b) and 4 (b) and (c) show examples of pair entries. We find that many of the highlighted features, including the dark auriculars (feather below and behind the eye) of the Kentucky warbler, the black "necklace" of the Canada warbler, and the white "spectacles" of the yellow-breasted chat (all shown in Figure 4), correspond to features mentioned in bird guides (all included in the Sibley Guide [22]).

4.1. A Tree of Visual Similarity

Visual similarity as estimated from the POOFs is the basis for our visual field guide. In similarity estimation, unlike straight classification, there is no obvious ground truth. If we say a blackbird is more like a crow than like a raven, who can say we are wrong? One way to get a ground truth for similarity is to consider the evolutionary "tree of life," the tree with a root representing the origin of life, a leaf for every extant species or evolutionary dead end, and a branch for every speciation event, with edge lengths representing time between speciations. Species close to each other in the tree of life are in a sense "more similar" than species that are not close, although this will not necessarily correspond to visual similarity.

The scientific community has not reached consensus on the complete structure of the tree of life, or even the subtree containing just the birds in CUBS-200. However there is progress in that direction. Recently Jetz *et al.* [12] proposed the first complete tree of life for all 9993 extant bird



(a) **Species display**: For any species, we can display the most similar other species. The most similar species are displayed surrounding the species in question, with the thickness of the spokes proportional to the visual difference between species. The Kentucky warbler is most similar to the Canada warbler.
(b) **Species pair display**: After choosing one of the spokes, we display sample images of the two species, followed by a few pairs of images chosen and annotated to illustrate key visual differences. The Canada warbler is distinguished from the Kentucky warbler the curved of the yellow band by the eye, a complete eye-ring, and a black necklace. (c) The next most similar species, the yellow-breasted chat, is distinguished by the color of its eye band. We may show any number of sample images (here we fill the figure), but in general three pairs of images is sufficient.

Figure 4. Visual field guide pages for the Kentucky warbler.

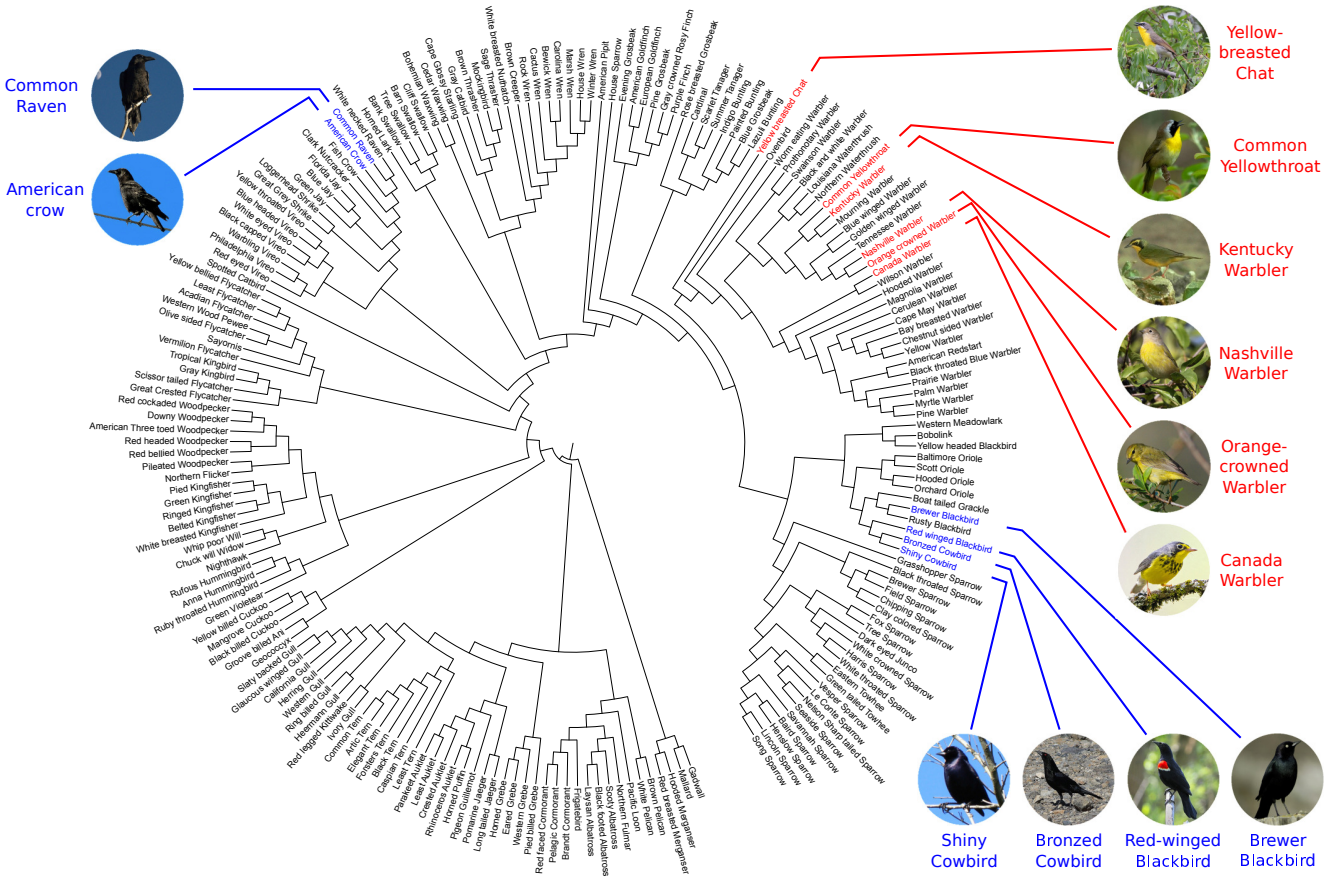


Figure 5. The phylogenetic "tree of life" representing evolutionary history. As in Figure 2, species visually similar to the red-winged blackbird are in blue, and those similar to the Kentucky warbler are in red. Although the American crow and common raven are visually similar to blackbirds, they are not close in terms of evolution.

Rank	Species Pair
1	Gadwall vs Pacific Loon
2	Hooded Merganser vs Pigeon Guillemot
6	Red-breasted Merganser vs Eared Grebe
11	Least Auklet vs Whip poor Will
16	Black billed Cuckoo vs Mockingbird

Table 1. Species pairs with high visual and low phylogenetic similarity.

Figure 7. The gadwall (left) and the pacific loon (right) have similar overall appearance but are not closely related.

species, complete with estimated dates for all splits, based on a combination of the fossil evidence, morphology, and genetic data. Pruning this tree to include only the species in CUBS-200 yields the tree shown in Figure 5 (produced in part with code from [14]). This tree shows the overall phylogenetic similarity relations between bird species.

As a browsing interface to our digital field guide, we propose a similar tree, in the same circular format. This tree, however, is based on visual similarity rather than phylogenetic similarity. Producing a tree from a similarity matrix is a basic operation in the study of phylogeny, for which standard methods exist (note the tree of life in Figure 5 is based on more advanced techniques that use additional data beyond a similarity matrix). We calculate the full similarity matrix of the bird species using the POOFs, then apply one of these standard methods, Saitou and Nei's "neighbor-joining" [20], to get a tree based not on evolutionary history but on visual similarity. This tree is shown in Figure 2. In an interactive form, it will allow a user to scroll through the birds in an order that respects similarity and shows a hierarchy of groups of similar birds.

We can compare the similarity-based tree in Figure 2 with the evolutionary tree in Figure 5. They generally agree as to which species are similar, but there are exceptions. For example, crows are close to blackbirds in the similarity tree, but the evolutionary tree shows that they are not closely related. Such cases may be examples of convergent evolution, in which two species independently develop similar traits.

We can find such species pairs, with high visual similarity and low phylogenetic similarity, in a systematic way. The phylogenetic similarity between two species can be quantified as the length of shared evolutionary history, *i.e.*, the path length, in years, from the root of the evolutionary tree to the species' most recent common ancestor (techniques such as the neighbor-joining algorithm [20] also use this as a similarity measure). Figure 6 (a) shows a similarity matrix calculated in this way for the 200 bird species, with the corresponding matrix based on visual similarity as Figure 6 (b). Potential examples of convergent evolution correspond to high values in (a) and relatively low values in (b). The blackbirds-crows region is marked as an example.

We rank all $\binom{200}{2}$ species pairs by visual similarity (most similar first) and by phylogenetic difference (least similar first). We then list all species pairs in order of the sum of these ranks.Table 1 shows the top five pairs, excluding pairs where one of the species has already appeared on the list to avoid excessive repetition (as the pacific loon scores highly when paired with the gadwall, it will also score highly with all near relatives of the gadwall). The top ranked pair is a duck and a loon, two species the author had mistakenly assumed were closely related based on their visual similarity. Figure 7 shows samples of these two species. Space precludes including images of the other pairs in Table 1, but images can be viewed on Cornell's All About Birds site [4].

5. Conclusions

Recognition techniques, in particular methods of estimating visual similarity, can be used for more than just identification and image search. Here we exploit a setting in which computers can do better than typical humans – fine-grained categorization in a specialized domain – to show how progress in computer vision can be turned to helping humans understand the relations between the categories.

References

[1] T. Berg and P. N. Belhumeur. POOF: Part-based One-vs-One Features for fine-grained categorization, face verification, and attribute estimation. In *Proc. CVPR*, 2013. 1, 2, 3, 4

[2] T. L. Berg, A. C. Berg, and J. Shih. Automatic attribute discovery and characterization from noisy web data. In *Proc. ECCV*, 2010. 3

[3] S. Branson, C. Wah, B. Babenko, F. Schroff, P. Welinder, P. Perona, and S. Belongie. Visual recognition with humans in the loop. In *Proc. ECCV*, 2010. 3

[4] Cornell Lab of Ornithhology. allaboutbirds.org, 2011. 8

[5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. CVPR*, 2005. 3

[6] J. Deng, J. Krause, and L. Fei-Fei. Fine-grained crowdsourcing for fine-grained recognition. In *Proc. CVPR*, 2013. 3

[7] C. Doersch, S. Singh, A. Gupta, J. Sivic, and A. A. Efros. What makes paris look like paris? *ACM Trans. Graphics*, 31(4), 2012. 3

[8] K. Duan, D. Parikh, D. Crandall, and K. Grauman. Discovering localized attributes for fine-grained recognition. In *Proc. CVPR*, 2012. 3

[9] R. Farrell, O. Oza, N. Zhang, V. I. Morariu, T. Darrell, and L. S. Davis. Birdlets: Subordinate categorization using volumetric primitives and pose-normalized appearance. In *Proc. ICCV*, 2011. 3

[10] R. A. Fisher. The use of multiple measurements in taxonomic problems. *Ann. Eugenics*, 7(2), 1936. 4

[11] D. J. Futuyma. *Evolutionary Biology*, page 763. Sinauer Associates, 1997. 2

[12] W. Jetz, G. H. Thomas, J. B. Joy, K. Hartmann, and A. O. Mooers. The global diversity of birds in space and time. *Nature*, 491(7424), 2012. 5

[13] N. Kumar, P. N. Belhumeur, A. Biswas, D. W. Jacobs, W. J. Kress, I. Lopez, and J. V. B. Soares. Leafsnap: A computer vision system for automatic plant species identification. In *Proc. ECCV*, 2012. 3

[14] I. Letunic and P. Bork. Interactive tree of life (itol): An online tool for phylogenetic tree display and annotation. *Bioinformatics*, 23(1), 2007. 7

[15] J. Liu, A. Kanazawa, D. Jacobs, and P. Belhumeur. Dog breed classification using part localization. In *ECCV*, 2012. 3

[16] M.-E. Nilsback and A. Zisserman. Automated flower classification over a large number of classes. In *Indian Conf. Computer Vision Graphics and Image Processing*, 2008. 3

[17] D. Parikh and K. Grauman. Interactively building a discriminative vocabulary of nameable attributes. In *Proc. CVPR*, 2011. 3

[18] O. M. Parkhi, A. Vedaldi, A. Zisserman, and C. V. Jawahar. Cats and dogs. In *Proc. CVPR*, 2012. 3

[19] P. Prasong and K. Chamnongthai. Face-Recognition-Based dog-Breed classification using size and position of each local part, and pca. In *Proc. Int. Conf. Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, 2012. 3

[20] N. Saitou and M. Nei. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4), 1987. 2, 7, 8

[21] A. Shrivastava, T. Malisiewicz, A. Gupta, and A. A. Efros. Data-driven visual similarity for cross-domain image matching. *ACM Trans. Graphics*, 30(6), 2011. 3

[22] D. A. Sibley. *The Sibley Guide to Birds*. Knopf, 2000. 1, 3, 5

[23] L. Svensson, K. Mullarney, and D. Zetterström. *Collins Bird Guide*. Collins, 2011. 1

[24] B. Tversky and K. Hemenway. Objects, parts, and categories. *J. Experimental Psychology: General*, 113(2), 1984. 2

[25] C. Wah, S. Branson, P. Perona, and S. Belongie. Multiclass recognition and part localization with humans in the loop. In *Proc. ICCV*, 2011. 3

[26] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The Caltech-UCSD Birds-200-2011 Dataset. Technical Report CNS-TR-2011-001, California Institute of Technology, 2011. 3

[27] J. Wang, K. Markert, and M. Everingham. Learning models for object recognition from natural language descriptions. In *Proc. British Machine Vision Conf.*, 2009. 3

[28] K. Yanai and K. Barnard. Image region entropy: A measure of "visualness" of web images associated with one concept. In *ACM Int. Conf. Multimedia*, 2005. 3

[29] B. Yao, G. Bradski, and L. Fei-Fei. A codebook-free and annotation-free approach for fine-grained image categorization. In *Proc. CVPR*, 2012. 1, 3

[30] N. Zhang, R. Farrell, and T. Darrell. Pose pooling kernels for sub-category recognition. In *Proc. CVPR*, 2012. 3