Caltech-UCSD Birds 200

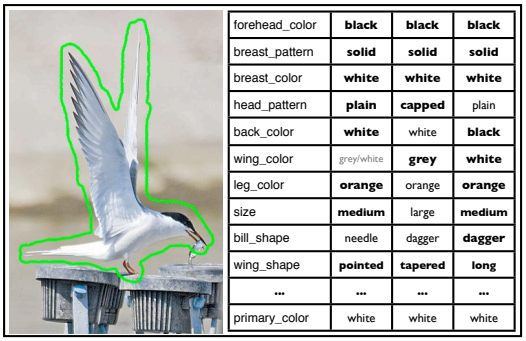
Peter Welinder[†] **Steve Branson**[⋆] **Takeshi Mita**[†] **Catherine Wah**[⋆] **Florian Schroff**[⋆]

Serge Belongie[⋆]
[⋆] University of California, San Diego

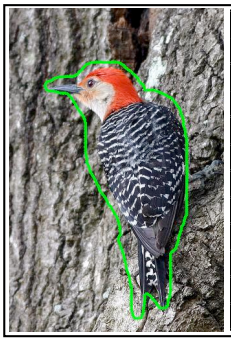
Pietro Perona[†]
[†] California Institute of Technology

Abstract

Caltech-UCSD Birds 200 (CUB-200) is a challenging image dataset annotated with 200 bird species. It was created to enable the study of *subordinate categorization*, which is not possible with other popular datasets that focus on basic level categories (such as PASCAL VOC, Caltech-101, etc). The images were downloaded from the website Flickr and filtered by workers on Amazon Mechanical Turk. Each image is annotated with a bounding box, a rough bird segmentation, and a set of attribute labels.



forehead_color	**black**	**black**	**black**
breast_pattern	**solid**	**solid**	**solid**
breast_color	**white**	**white**	**white**
head_pattern	**plain**	**capped**	plain
back_color	**white**	white	**black**
wing_color	grey/white	**grey**	**white**
leg_color	**orange**	orange	**orange**
size	**medium**	large	**medium**
bill_shape	needle	dagger	**dagger**
wing_shape	**pointed**	**tapered**	**long**
...	...	...	...
primary_color	white	white	white



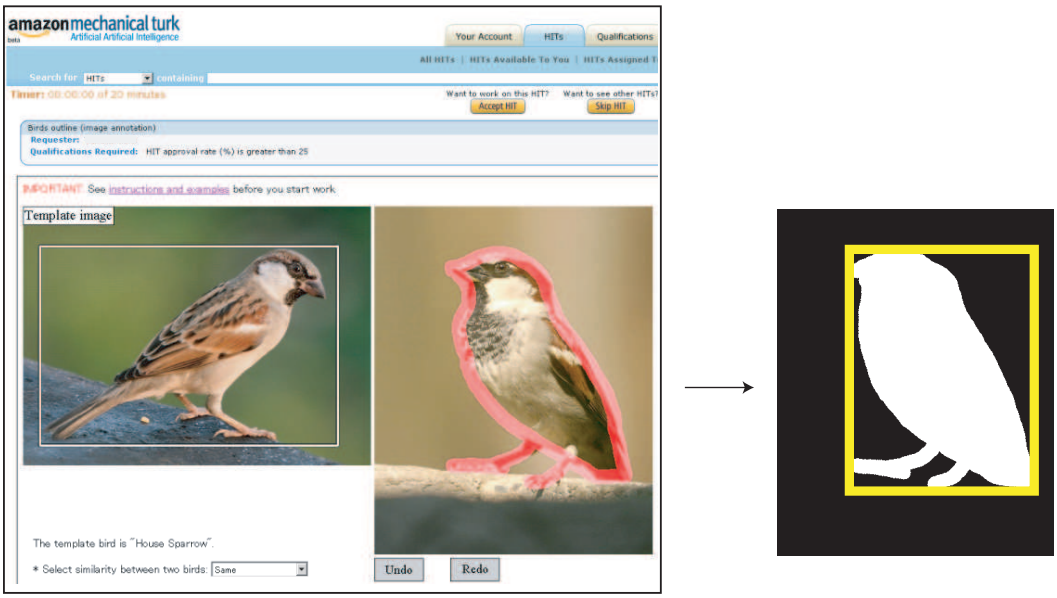
forehead_color	**red**	red	**red**
breast_pattern	multi-colored	solid	solid
breast_color	white	white/red	white
head_pattern	**capped**	capped	**capped**
back_color	**white/ black**	**white/ black**	**white/ black**
wing_color	**white/ black**	white/black	white/black
leg_color	buff	black	black
size	**small**	medium	medium
bill_shape	**all-purpose**	**dagger**	**all-purpose**
wing_shape	**pointed**	tapered	pointed
...	...	...	...
primary_color	black, red	white, black	**white, black**

Figure 1: Images and annotations from CUB-200. Each example image is shown with a rough outline (segmentation) in green. To the right of each image is a table of attributes (one per row, 11 out of a total of 25 attributes shown), and attribute-values provided by Amazon Mechanical Turk workers looking at the image. The attribute-values in the three right-most columns in the tables are provided by different workers (across both columns and rows). The font of the attribute-value indicates the confidence of the worker: bold font means the worker was 'definitely' sure of the label, thin means 'probably', and grey means 'guessing'.

1 Introduction

Large-scale annotated image datasets have been instrumental for driving progress in object recognition over the last decade. Most datasets contain a wide variety of basic level classes, such as different kinds of animals and inanimate objects. Examples of popular such datasets include Caltech-101 and Caltech-256 [4, 5], LabelMe [8], PASCAL VOC [3], and ImageNet [2]. One property shared by all these datasets is that an average human being would have little difficulty in achieving near-perfect classification accuracy. Computer vision systems, on the other hand, still do quite poorly.

We introduce Caltech-UCSD Birds 200 (CUB-200), a dataset aimed at subordinate category classification. CUB-200 includes 6,033 annotated images of birds, belonging to 200, mostly North American, bird species. Each image is annotated with a rough segmentation, a bounding box, and



(a) (b)

Figure 2: Annotations obtained from MTurk workers. (a) Screenshot of the web-based annotation tool used by workers. The image to be annotated is on the right (superimposed in red is the rough outline provided by a worker), a template image is on the left. The worker has to assess whether the right image contains a bird, and if it does, whether the species of the two birds is the 'same', 'similar', 'different' or 'difficult to compare'. The worker is then asked to provide either a bounding box or trace the outline of the bird on the right (details in the Section 3). (b) The resulting annotations (the similarity label was 'same' and is not shown).

binary attribute annotations. There is only one other dataset known to us with a similar scope, the Flowers dataset [6] with 102 different types of flowers common in the United Kingdom. In contrast to the datasets mentioned above, accurately classifying more than a handful of birds is something only a small proportion of people can do without access to a field guide. Moreover, since few people do well on subordinate categorization tasks, it is arguably an area where a visual recognition system would be useful even if it was not perfect.

With CUB-200 we hope to facilitate research on applications where computer vision helps people classify objects that are unknown to them. For example, if an accurate bird classifier were developed, a user could submit a photo of a recently spotted bird to query a knowledge database, such as Wikipedia [7]. Such classifiers could also help to automate other areas of science[1].

2 Image Collection

A list of 278 bird species was compiled from an online field guide[2]. Next, we downloaded all images on the corresponding Wikipedia[3] page for each species. Species with no Wikipedia article, or no images on their article page, were eliminated from the list. The remaining names were fed to Flickr[4] as query terms, and up to 40 images were downloaded for each species. If a name returned less than 20 images from the Flickr search, it was removed from the list, which left 223 species with 20 or

[1] An example of something that could be automated is the Great Backyard Bird Count that crowdsources the counting of bird species in North America, http://www.birdsource.org/gbbc/.

[2] http://www.birdfieldguide.com/

[3] http://www.wikipedia.org

[4] http://www.flickr.com/

more Flickr images. We manually ensured that the example images downloaded from Wikipedia actually contained a bird.

All the Flickr images were annotated with a rough segmentation by workers on Amazon Mechanical Turk[5] (MTurk), as described in the next section. Each image was annotated by two workers per image and annotation type. The workers were shown a representative image exemplar from the Wikipedia page of the species that was used to query Flickr to find the image to be annotated. In addition to providing the annotation, they were asked to rank the similarity between the image and the exemplar using the following system:

- **Same:** the bird in the image looks like it is of the same species as the exemplar,
- **Similar:** the bird in the image and the exemplar look similar, maybe of the same species,
- **Different:** the bird in image differs from the one in the exemplar,
- **Difficult:** chosen if occlusion or scale differences make the comparison difficult.

From the annotated Flickr images, we kept only images that were labeled as 'same' or 'similar' by both workers, and where there was an overlap of the bounding boxes enclosing the rough outline annotations; the rest of the images were eliminated. The remaining images were checked by us, so that each image was reviewed by a total of three different people. After excluding all species that had less than 20 Flickr images remaining, 200 species were left with a total of 6,033 images. See the Appendix for example images from all species.

3 Annotations

We collected two kinds of annotations from MTurk: rough outlines and attribute annotations, see Figure 1. Bounding boxes were deduced from the rough outlines.

For the rough segmentations, the workers were asked to draw with a thick brush to touch all the boundary pixels of the foreground object, see Figure 2. The rough segmentation was chosen over a more detailed segmentation, such as the segmentations in [8], since the former takes shorter time for a worker to complete, thus increasing the overall throughput.

In addition to location information, in another task we instructed MTurk workers to provide attribute annotations. We used 25 visual attributes from an online bird field guide[6], listed in Table 1. We created a user interface for MTurk workers to provide attribute annotations, see Figure 3, where the user was shown the query image to the left and a set of attribute values (and explanations) to the right. They were also asked to provide the confidence of their label in three grades: 'definitely' sure, 'probably' sure, and 'guessing'. We obtained five annotations per image and attribute from a total of 1,577 workers. Figure 4 shows how the work was distributed among the workers and Figure 5 the sizes of the images downloaded and the obtained annotations.

4 Baseline Experiments

In order to establish a baseline performance on the dataset, we used a nearest neighbor (NN) classifier to classify images from a test set using different features. We chose two simple features as the baseline: image sizes and color histograms. In the case of the image sizes, we represented each image by its width and height in pixels. For the color histograms, we used 10 bins per channel (making 10^3 bins in total) and then applied Principal Component Analysis (PCA) and kept only the top 128 principal components. Figure 6 shows how the performance of the NN classifier degrades as the number of classes in the dataset is increased. The performance of the image size features are close to chance at 0.6% for the 200 classes, while the color histogram features increase the performance to 1.7%. We also compare the NN classifier to the baseline method in [1], which is the first paper to use the dataset and achieves 19% classification performance.

One disadvantage of searching for images on Flickr is that images returned by a query are often distributed over only a few photographers. This poses a problem because it is quite common that a

[5] http://www.mturk.com
[6] http://www.whatbird.com

Attribute	Values
Crown color	blue, black, orange, buff, brown, grey, white, red, pink, rufous, iridescent, yellow, olive, purple, green
Nape color	white, black, brown, buff, grey, yellow, red, orange, iridescent, olive, green, blue, rufous, pink, purple
Bill shape	cone, all-purpose, dagger, hooked seabird, hooked, curved (up or down), spatulate, needle, specialized
Head Pattern	malar, eyebrow, capped, eyering, unique pattern, striped, spotted, crested, masked, plain, eyeline
Belly Pattern	solid, striped, spotted, multi-colored
Belly color	grey, white, black, buff, yellow, brown, green, blue, iridescent, olive, orange, red, rufous, pink, purple
Wing shape	pointed-wings, tapered-wings, long-wings, rounded-wings, broad-wings
Shape	perching-like, tree-clinging-like, gull-like, duck-like, swallow-like, upright-perching water-like, sandpiper-like, upland-ground-like, chicken-like-marsh, pigeon-like, long-legged-like, hummingbird-like, hawk-like, owl-like
Primary Color	brown, grey, white, black, rufous, yellow, buff, red, blue, olive, iridescent, green, orange, pink, purple
Size	small (5 - 9 in), very small (3 - 5 in), medium (9 - 16 in), very large (32 - 72 in), large (16 - 32 in)
Forehead Color	grey, buff, red, black, orange, brown, white, blue, iridescent, rufous, green, yellow, pink, olive, purple
Throat Color	brown, buff, black, white, orange, grey, yellow, blue, iridescent, olive, rufous, green, pink, purple, red
Eye color	yellow, black, red, rufous, orange, white, brown, grey, olive, buff, blue, green, purple, pink
Underparts Color	grey, yellow, brown, white, black, buff, orange, iridescent, olive, blue, red, green, rufous, pink, purple
Breast Pattern	striped, solid, spotted, multi-colored
Breast Color	white, grey, orange, yellow, buff, black, brown, rufous, green, iridescent, blue, red, pink, olive, purple
Upperparts Color	buff, brown, grey, black, white, yellow, red, purple, olive, orange, iridescent, green, blue, rufous, pink
Back pattern	spotted, solid, multi-colored, striped
Back color	buff, white, black, grey, brown, purple, pink, blue, iridescent, olive, rufous, yellow, green, red, orange
Leg color	white, blue, grey, black, orange, buff, brown, pink, yellow, red, purple, olive, rufous, iridescent, green
Tail pattern	striped, solid, spotted, multi-colored
Under tail color	grey, buff, orange, yellow, black, brown, white, rufous, olive, iridescent, blue, green, red, purple, pink
Upper tail color	brown, black, grey, buff, white, yellow, rufous, olive, blue, iridescent, orange, green, red, pink, purple
Wing Pattern	striped, spotted, solid, multi-colored
Wing Color	black, buff, grey, white, brown, yellow, purple, iridescent, blue, olive, rufous, orange, red, green, pink

Table 1: Multi-valued bird attributes. For each image, we asked workers to select the values that were most appropriate for the attribute in question.

photographer has taken many images of the same individual bird in a very short time period, resulting in near-identical images in the Flickr search results. Thus, if a large proportion of the images in a class come from one photographer, a simple nearest neighbor based method will perform artificially well on the classification task. To overcome this problem, for each species we chose a date that split the images into roughly equal-sized sets: the images before the date to be used as training set and the images after the date to be used as test set. We strongly suggest that our dataset is always used this way. Different choices of the training-testing sets will likely produce vastly different classification performance figures. We have released the training/test set splits on the CUB-200 project website.

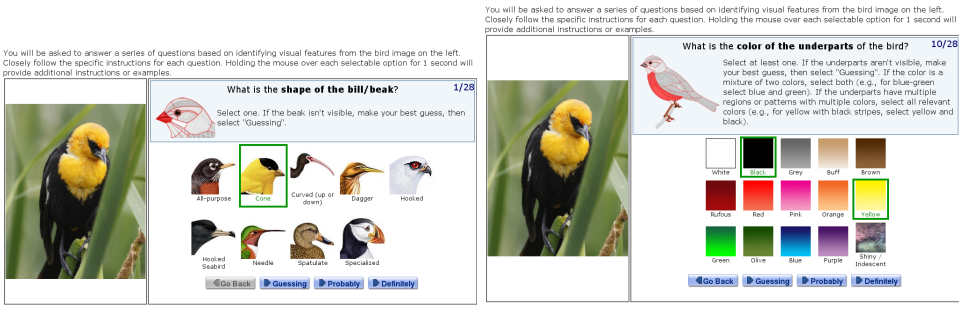


Figure 3: The interface used by MTurk workers to provide attribute labels. The query image is shown to the left and the choice of attribute values on the right in each diagram.

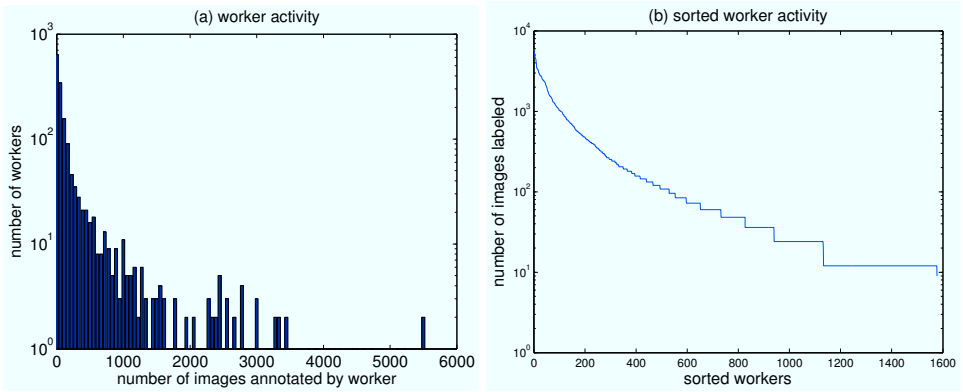


Figure 4: The distribution of activity of the MTurk workers. (a) A histogram of the number of images annotated per worker. (b) All workers sorted by the number of labels they provided.

5 Conclusion

CUB-200[7] has a total of $6,033$ images allocated over 200 (mostly North American) bird species, see Figure 5. The large number of categories should make it an interesting dataset for subordinate categorization. Moreover, since it is annotated with bounding boxes, rough segmentations and attribute labels, it is also ideally suited for benchmarking systems where the users take an active part in the recognition process, as demonstrated in [1].

References

[1] Steve Branson, Catherine Wah, Florian Schroff, Boris Babenko, Peter Welinder, Pietro Perona, and Serge Belongie. Visual Recognition with Humans in the Loop. In *ECCV*, 2010. 3, 5, 6

[2] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR*, 2009. 1

[3] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, 88(2):303–338, June 2010. 1

[4] L. Fei-Fei, R. Fergus, and P. Perona. One-shot learning of object categories. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 28(4):594–611, 2006. 1

[5] G. Griffin, A. Holub, and P. Perona. Caltech-256 object category dataset. Technical Report 7694, California Institute of Technology, 2007. 1

[7] Download at http://www.vision.caltech.edu/visipedia.

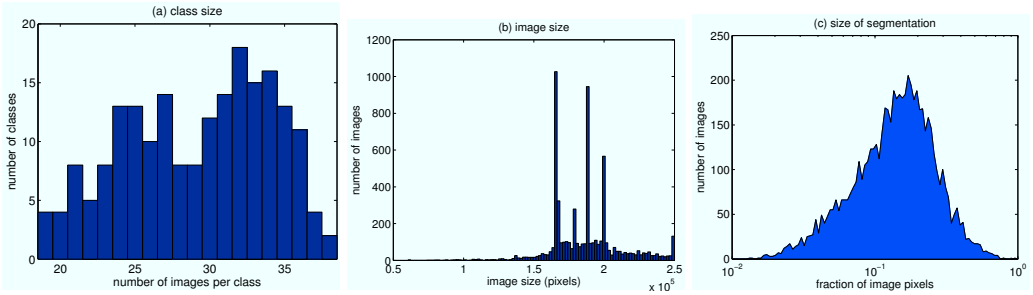
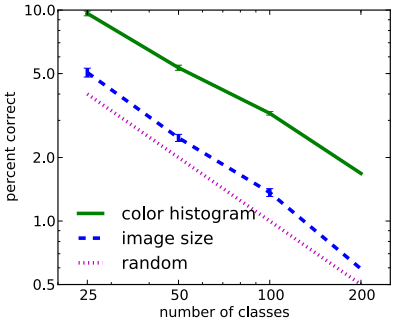


Figure 5: Distribution of images. (a) A histogram of the sizes (in number of images) per bird species. (b) Distribution of image sizes (in pixels) in the dataset out of 6,033 images. (c) Distribution of the fractions of pixels that the segmented bird occupies with respect to the total size in the image.



Method	Performance
NN (image size)	0.6%
NN (color histogram)	1.7%
SVM (SIFT, spatial pyramid)	19%

Figure 6: Baseline performance on CUB-200. Left: Performance of a nearest neighbor classifier using image size and color histogram features as the number of classes is increased. The error bars show the standard error from 10 trials where a subset of the 200 classes was randomly sampled without replacement. Also shown (labeled 'random') is the probability of making a correct classification by chance. Right: Performance on the full dataset with 200 classes. We also compare against the baseline method used in [1] which is based on a 1-vs-all SVM classifier using SIFT features and a spatial pyramid.

[6] M-E. Nilsback and A. Zisserman. Automated flower classification over a large number of classes. In *Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing*, Dec 2008. 2

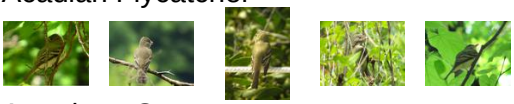
[7] P. Perona. Vision of a Visipedia. *Proceedings of the IEEE*, 98(8):1526 –1534, 2010. 2

[8] B.C. Russell, A. Torralba, K.P. Murphy, and W.T. Freeman. LabelMe: A Database and Web-Based Tool for Image Annotation. *Int. J. Comput. Vis.*, 77(1–3):157–173, 2008. 1, 3

Appendix: Example Images

Here we show five random example images from each of the 200 bird categories.

Acadian Flycatcher



American Crow

American Goldfinch

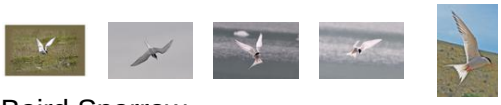
American Pipit

American Redstart

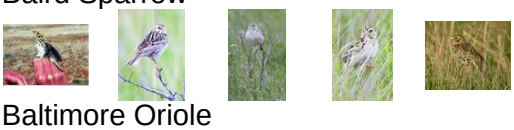
American Three toed Woodpecker

Anna Hummingbird

Artic Tern



Baird Sparrow



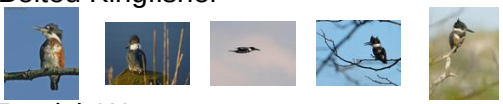
Baltimore Oriole

Bank Swallow

Barn Swallow

Bay breasted Warbler

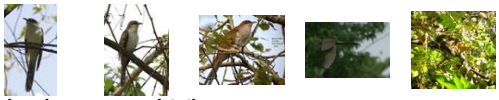
Belted Kingfisher



Bewick Wren

Black and white Warbler

Black billed Cuckoo

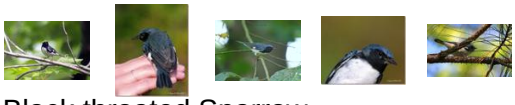


Black capped Vireo

Black footed Albatross

Black Tern

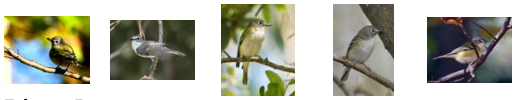
Black throated Blue Warbler



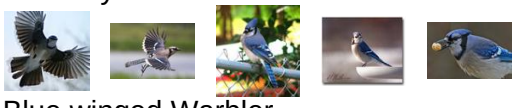
Black throated Sparrow

Blue Grosbeak

Blue headed Vireo



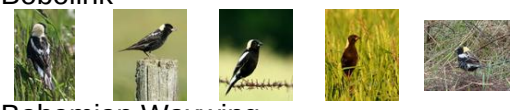
Blue Jay



Blue winged Warbler

Boat tailed Grackle

Bobolink

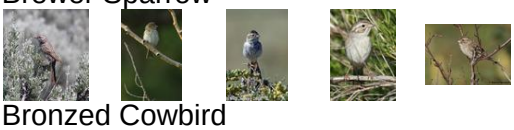


Bohemian Waxwing

Brandt Cormorant

Brewer Blackbird

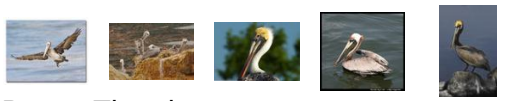
Brewer Sparrow



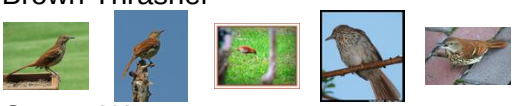
Bronzed Cowbird

Brown Creeper

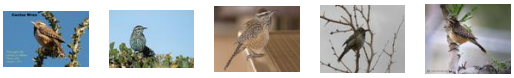
Brown Pelican



Brown Thrasher



Cactus Wren



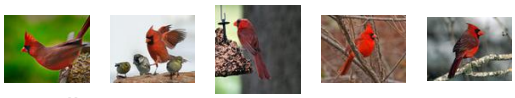
California Gull

Canada Warbler

Cape Glossy Starling

Cape May Warbler

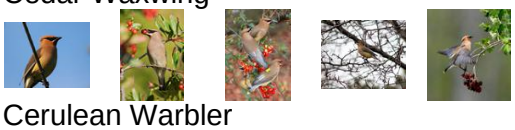
Cardinal



Carolina Wren

Caspian Tern

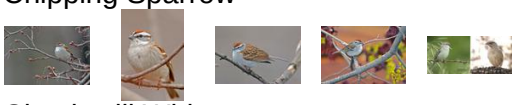
Cedar Waxwing



Cerulean Warbler

Chestnut sided Warbler

Chipping Sparrow



Chuck will Widow

Clark Nutcracker

Clay colored Sparrow

Cliff Swallow

Common Raven

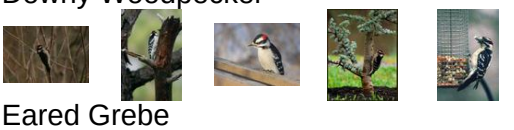
Common Tern

Common Yellowthroat

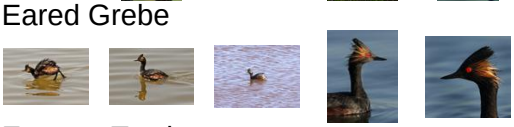
Crested Auklet

Dark eyed Junco

Downy Woodpecker



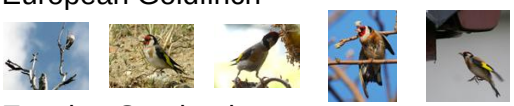
Eared Grebe



Eastern Towhee

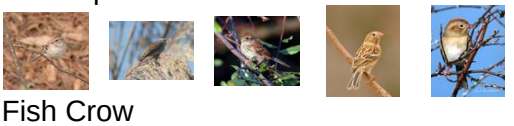
Elegant Tern

European Goldfinch



Evening Grosbeak

Field Sparrow



Fish Crow

Florida Jay

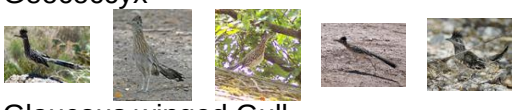
Forsters Tern

Fox Sparrow

Frigatebird

Gadwall

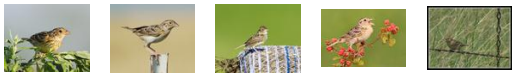
Geococcyx



Glaucous winged Gull

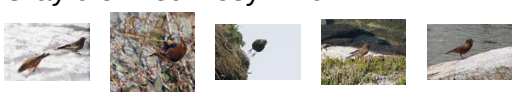
Golden winged Warbler

Grasshopper Sparrow

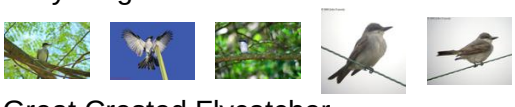


Gray Catbird

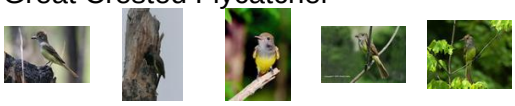
Gray crowned Rosy Finch



Gray Kingbird



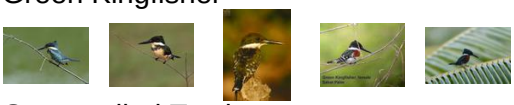
Great Crested Flycatcher



Great Grey Shrike

Green Jay

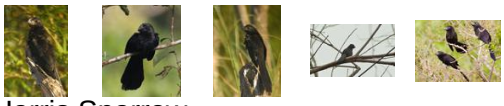
Green Kingfisher



Green tailed Towhee

Green Violetear

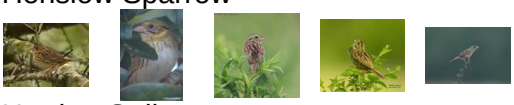
Groove billed Ani



Harris Sparrow

Heermann Gull

Henslow Sparrow



Herring Gull

Hooded Merganser

Hooded Oriole

Hooded Warbler

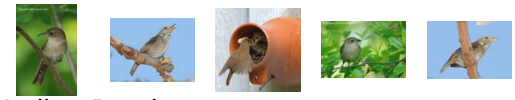
Horned Grebe

Horned Lark

Horned Puffin

House Sparrow

House Wren

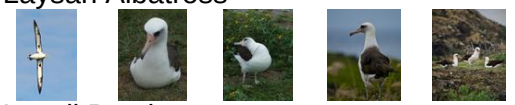


Indigo Bunting

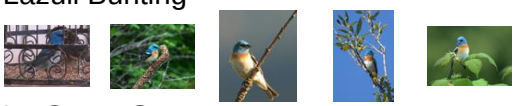
Ivory Gull

Kentucky Warbler

Laysan Albatross



Lazuli Bunting



Le Conte Sparrow

Least Auklet

Least Flycatcher

Least Tern

Lincoln Sparrow

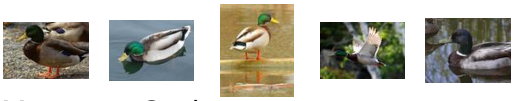
Loggerhead Shrike

Long tailed Jaeger

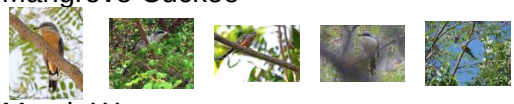
Louisiana Waterthrush

Magnolia Warbler

Mallard

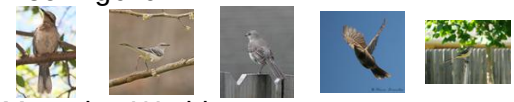


Mangrove Cuckoo

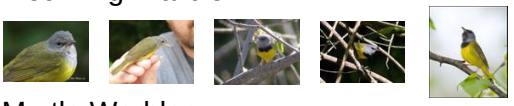


Marsh Wren

Mockingbird



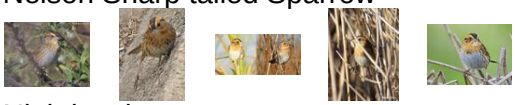
Mourning Warbler



Myrtle Warbler

Nashville Warbler

Nelson Sharp tailed Sparrow



Nighthawk

Northern Flicker

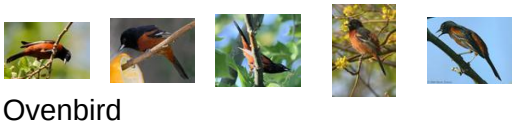
Northern Fulmar

Northern Waterthrush

Olive sided Flycatcher

Orange crowned Warbler

Orchard Oriole



Ovenbird

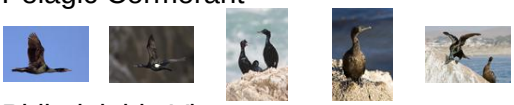
Pacific Loon

Painted Bunting

Palm Warbler

Parakeet Auklet

Pelagic Cormorant

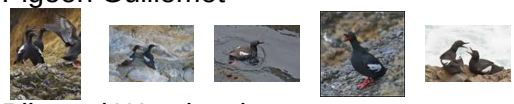


Philadelphia Vireo

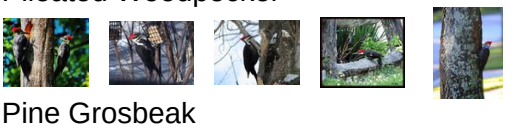
Pied billed Grebe

Pied Kingfisher

Pigeon Guillemot



Pileated Woodpecker



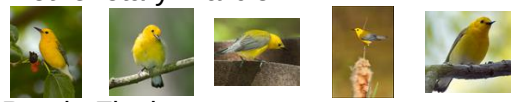
Pine Grosbeak

Pine Warbler

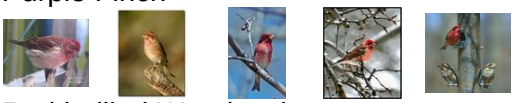
Pomarine Jaeger

Prairie Warbler

Prothonotary Warbler



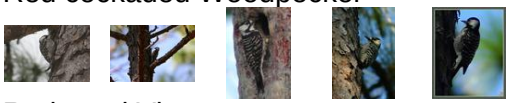
Purple Finch



Red bellied Woodpecker

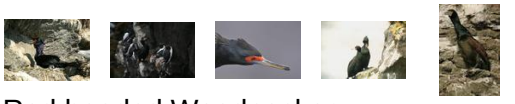
Red breasted Merganser

Red cockaded Woodpecker



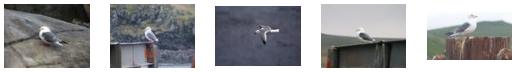
Red eyed Vireo

Red faced Cormorant

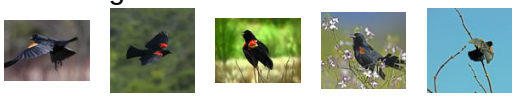


Red headed Woodpecker

Red legged Kittiwake



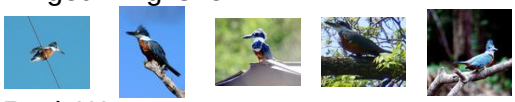
Red winged Blackbird



Rhinoceros Auklet

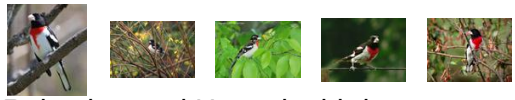
Ring billed Gull

Ringed Kingfisher

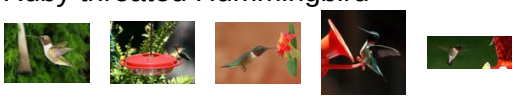


Rock Wren

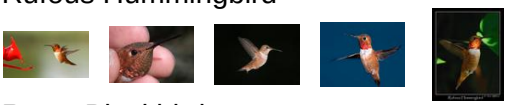
Rose breasted Grosbeak



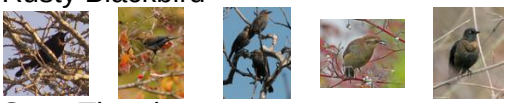
Ruby throated Hummingbird



Rufous Hummingbird



Rusty Blackbird

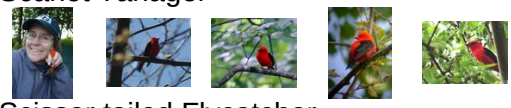


Sage Thrasher

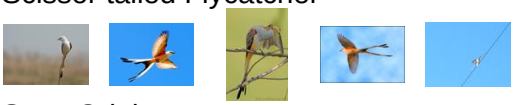
Savannah Sparrow

Sayornis

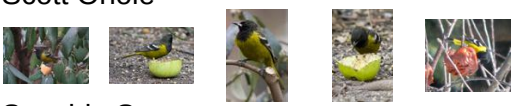
Scarlet Tanager



Scissor tailed Flycatcher

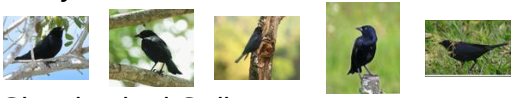


Scott Oriole



Seaside Sparrow

Shiny Cowbird



Slaty backed Gull

Song Sparrow

Sooty Albatross

Spotted Catbird

Summer Tanager

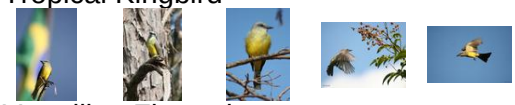
Swainson Warbler

Tennessee Warbler

Tree Sparrow

Tree Swallow

Tropical Kingbird



Vermilion Flycatcher

Vesper Sparrow

Warbling Vireo

Western Grebe

Western Gull

Western Meadowlark

Western Wood Pewee





Whip poor Will





White breasted Kingfisher





White breasted Nuthatch





White crowned Sparrow





White eyed Vireo





White necked Raven





White Pelican





White throated Sparrow

Wilson Warbler

Winter Wren

Worm eating Warbler

Yellow bellied Flycatcher

Yellow billed Cuckoo

Yellow breasted Chat

Yellow headed Blackbird

Yellow throated Vireo

Yellow Warbler