An Exploration of Computer Vision Techniques for Bird Species Classification

Anne L. Alter (annealtr@stanford.edu), Karen M. Wang (kmwang14@stanford.edu)

Introduction

Objective

Classifying organisms, such as specific bird species, is a challenging task and active machine learning research field. Current state-of-the-art computer vision algorithms can achieve a maximum of 85% accuracy with most around 60% [1]. Our objective was to try to implement our own machine learning and computer vision algorithms with Softmax regression, an SVM, and a CNN with transfer learning to see what accuracy we could achieve in classifying bird species and whether we could identify specific features on the birds.

Cardinal

Difficulties in Bird Classification

- Complex foreground / background settings (trees, water, etc…)
- Similarities in subspecies of birds
- Photo lighting conditions
- Only a subset of bird features appearing in photos (ex. right eye and right wing visible, cannot see bird belly and left side)

Western Meadowlarks in different background settings

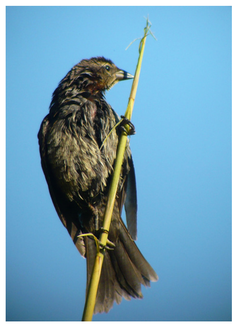
Dataset

Caltech-UCSD-Birds-200-2011

- 200 categories of bird species
- 11,788 total number of images
- Other information: labeled visible bird parts (in pixels), binary attributes, bounding box of bird

Sooty Albatross

Blue Grosbeak



Shiny Cowbird

Features and Models: Summary

Softmax Regression on Binary Attributes

The Softmax regression model was based on binary data of 312 bird attributes (wing color, beak shape, tail pattern, etc…) that was manually collected in the Caltech-UCSB study and thus did not use computer vision. Rather, this model was intended to serve as a baseline for our classification performance. The data was run for specifically classified (ex. Prairie and Pine Warblers in separate classes with 200 classes total) and also on 'broadly classified data' (ex. all Warblers fall into one class with 71 classes total).

Multiclass SVM on HOG+RGB features

The feature vectors used in this model are the histogram of oriented gradients (HOG) concatenated with RGB histogram values. HOGs are feature descriptors which had widespread success in image detection (Dalal and Triggs 2005) due to its ability to detect silhouettes and invariance to geometric and photometric transforms (except object orientation). Since bird color is essential in identifying its class, RGB histograms of the images are used to aid in identification.

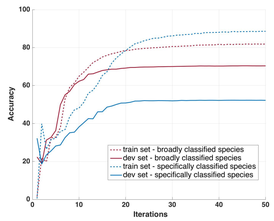
CNNs and Transfer Learning

This method uses the pretrained AlexNet model for transfer learning. We input RGB images, import the first 22 layers of the pretrained AlexNet model, and train the weights of the last 3 layers using images in the CUB-200-2011 and a softmax layer to classify into 200 categories.

Results: Softmax Regression on Binary Attributes

- Baseline model to determine classification accuracy *without* computer vision
- Trained data with regularized Softmax regression on 312 manually collected, binary attributes from Caltech / UCSB study such as:
  - bill shape
  - wing color
  - tail pattern
  - general size
  - eye color
  - belly pattern
  - bill length
  - general shape
- Ran model on 'specifically classified data' (ex. Prairie Warbler and Pine Warbler in separate classes with 200 classes total) and also on 'broadly classified data' (ex. all Warblers fall into one class with 71 classes total)
- Broadly Classified Acc.: **~70% ;** Specifically Classified Acc.: **~53%**

Accuracy



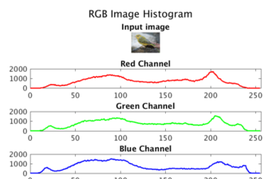
Ruby-Throated Hummingbird

Results: Multiclass SVM on HOG and RGB features

- Images used are bounding boxes of the original images which are then resized to 160x160 pixels
- Some of test included using image preprocessing to mask the background of the birds inside the bounding box
- Features are fed into a linear SVM; using only HOG (Histogram of Oriented Gradients) features gave an accuracy of **3.4%**; using HOG+RGB features boosted the accuracy to **9%**
- Preliminary studies concatenating localized bird head and beak parts and extracting HOG+RGB features to existing features were made with promising results

Example of image preprocessing to mask background and cropping and resizing of birds using given bounding box data

HOG cell sizes were varied, and an optimum cell size of 16x16 was chosen for HOG feature extraction to capture the most informative spatial features of the bird



RGB histogram features were concatenated with HOG feature information and fed into a linear SVM

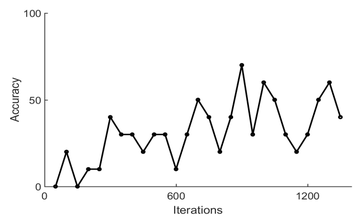
Least Tern

Results: CNNs and Transfer Learning

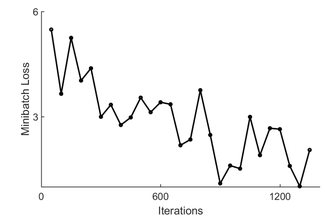
- **Initial prototyping**: a simplified neural network is used (two convolutional layers and 2 fully connected layers) with inputs of 32x32 RGB images from the dataset for fast training. Initial results are promising, so we move on to next step with confidence
- **Transfer learning**: used pre-trained AlexNet model (5 convolutional layers and 3 fully connected layers) which has been trained on over a million images from ImageNet (Krizhevsky et. Al. 2012)
  - Inputs are 227x227 RGB images from the CUB dataset
  - Last three layers are replaced by a fully connected layer, a Softmax layer, and a classification output layer
  - Learning rate in new layers are tuned much larger than learning rate in previous layers to speed up learning for our particular bird dataset

Loggerhead Shrike

Accuracy



Loss



Initial accuracy and loss for two epochs of training. More iterations are needed (and are still being run) until the validation loss converges.

Conclusion

Discussion

- Classification of specific species with manually identified attributes is only ~50% accurate, demonstrating the difficult nature of bird classification; accuracy improved to ~70% with more broadly defined species; perhaps more attributes would need to be identified to differentiate closely related species
- Using HOG+RGB features of the overall bird gives better results than HOG alone
- Adding bird head and beak parts as features gives only slight gains in accuracy, possibly due differences in frontal versus side views of the bird
- Transfer learning of the pre-trained AlexNet CNN has the most promising results but is slow due to many layers of backpropagation and training of weights

Future Work

- See how SIFT+RGB features can improve accuracy or if it would give comparable results to HOG+RGB features
- Use frozen pre-trained AlexNet weights (leveraging the features extracted) with just a multi-class SVM in the outermost layer is currently being tested for improved speed. Accuracy will be compared.

Cape Glossy Starling

References

[1] Branson, Steve, et al. "Improved Bird Species Recognition Using Pose Normalized Deep Convolutional Nets." *Proceedings of the British Machine Vision Conference 2014*, 2014, doi:10.5244/c.28.87.
[2] photos: Welinder P., Branson S., Mita T., Wah C., Schroff F., Belongie S., Perona, P. "Caltech-UCSD Birds 200". California Institute of Technology. CNS-TR-2010-001. 2010.

Stanford University