BIRD DETECTIVE

Udacity Machine Learning Engineer Nanodegree Capstone Proposal

WILLIAM SKERTIC – 05 APRIL 2020

DOMAIN BACKGROUND

As human beings continue to encroach on wildlife's habitat and introduce new technological perils it is imperative that wildlife resource management agencies can determine the impact of proposed changes and take appropriate steps to mitigate negative impacts. The first step in these efforts is likely the identification of wildlife that may be impacted, and being able to fly, birds present an interesting challenge in identification. The most accessible means of identification and tracking is visual. Currently (last update published in 2016) the North American Bird Conservation Initiative lists 432 species most in need of conservation action. A recent falconry experience at a bird of prey conservancy inspired this project.

PROBLEM STATEMENT

The problem is effective visual identification of bird species. The challenge will be to create a model for visual identification of birds using a specialized dataset containing bird species images, ideally producing a generalized bird classifier that performs well across the large number of bird species represented in the dataset.

DATASETS AND INPUTS

Data provided by the Cornell Lab of Ornithology, with thanks to photographers and contributors of crowdsourced data at AllAboutBirds.org/Labs. This material is based upon work supported by the National Science Foundation under Grant No. 1010818.

The NABirds V1 dataset available from The Cornell Lab of Ornithology (http://birds.cornell.edu/nabirds/) has been selected due to the specialized nature of the dataset.

> *NABirds V1 is a collection of 48,000 annotated photographs of the 400 species of birds that are commonly observed in North America. More than 100 photographs are available for each species, including separate annotations for males, females and juveniles that comprise 700 visual categories. This dataset is to be used for fine-grained visual categorization experiments.*
>
> *More than 550 visual categories, organized taxonomically*
>
> *Photos curated in collaboration with domain experts*

The NABirds V1 dataset comprises 48,562 images provided by members of the birding community. Experts from the Cornell Lab of Ornithology curated the images and created the taxonomy of 555 categories specifically to allow for the creation of bird identification tools for novice birders. These images were subsequently annotated with 11 bird parts and bounding boxes by a combination of Citizen Scientist and Amazon Mechanical Turk workers.

Solution Statement

The solution will be an application of transfer learning to the specialized NABirds V1 dataset. Using pre-trained networks with the final layer replaced by a new fully connected layer, the entire network will be retrained using the predefined weights for initialization.

Benchmark Model

The benchmark model is described in [1] and defined in [2]. Accuracy for image classification without pose normalization is 35.7%.

Evaluation Metrics

The evaluation metrics will be Top-1 and Top-5 Accuracy. These are the percentage of images for which the correct class label was in the Top-1 and Top-5 ranked label predictions respectively. The papers leading to the creation of the specialized NABirds V1 dataset (being used for model training in this project) utilized Top-1 accuracy. Top-5 accuracy is another metric commonly used in the evaluation of image classification models [4] and will therefore also be considered.

Project Design

Keras with a TensorFlow backend will be used to provide both the framework for implementing the methods and models, but it will also provide the pre-trained model itself – Xception.

The data will be moved into an S3 bucket where it can be manipulated. The data will be preprocessed (images resized – 299x299 for Xception, classes balanced, etc.) to provide consistent model inputs and will then be split to provide appropriate holdout examples.

With the data prepared the model objects will be created, their last layer replaced by a new fully connected layer. One of these models will have its pre-trained weights frozen, training only the output layer on the training data. This model serves as a baseline for understanding the added benefit of fine-tuning weights. The second model will have all weights initialized to the pre-trained values then the entire model will be retrained on the training data. These models will be compared with each other as well as with the benchmark model to show the value of retraining the entire network versus only the final layers.

Due to the model design, output for an input is a vector of class predictions. These predictions can be ranked as they are normalized probabilities of class membership. Once trained, the models will be compared using Top-1 and Top-5 Accuracy of these ranked class membership probabilities (the top-5 labels when ranked by class membership probability).

References

[1] G. Van Horn, S. Branson, R. Farrell, S. Haber, J. Barry, P. Ipeirotis, P. Perona, and S. Belongie, 2015. Building a bird recognition app and large scale dataset with citizen scientists: The fine print in fine-grained dataset collection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 595-604)*. https://ieeexplore.ieee.org/abstract/document/7298658

https://vision.cornell.edu/se3/wp-content/uploads/2015/05/Horn_Building_a_Bird_2015_CVPR_paper.pdf

[2] S. Branson, G. Van Horn, S. Belongie, and P. Perona., 2014. Bird Species Categorization Using Pose Normalized Deep Convolutional Nets. *arXiv preprint arXiv:1406.2952v1 [cs.CV]*.
https://arxiv.org/abs/1406.2952v1

[3] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, 2013. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. *arXiv preprint arXiv:1310.1531v1 [cs.CV].*
https://arxiv.org/abs/1310.1531v1

[4] F. Chollet, 2017. Xception: Deep Learning with Depthwise Separable Convolutions. *arXiv preprint arXiv:1610.02357v3 [cs.CV].*
https://arxiv.org/abs/1610.02357v3