Identifying Bird Species with Few-Shot Learning and Prototypical Networks

Michael Albert
College of Science & Engineering
Western Washington University
Bellingham, WA
albertmichael746@gmail.com

Jonah Douglas
College of Science & Engineering
Western Washington University
Bellingham, WA
jonah.douglas23@gmail.com

Ethan Lindell
College of Science & Engineering
Western Washington University
Bellingham, WA
eclindell@gmail.com

Archan Rupela
College of Science & Engineering
Western Washington University
Bellingham, WA
archanrupela@outlook.com

River Yearian
College of Science & Engineering
Western Washington University
Bellingham, WA
yeariar@wwu.edu

***Abstract*—This project involved creating a model to identify bird species based on images. The identification is a deep learning model using few-shot learning and prototypical networks to train the model. We achieved an accuracy of 80% identification with 200 possible classes. This accuracy is better than the 20% baseline accuracy determined if a class was to be selected at random since the implementation was 5-way 5-shot learning.**

***Index Terms*—deep learning, training, image classification, few-shot, prototypical networks**

I. Overview

Implementing a deep learning model using few-shot learning with prototypical networks, the goal of this project is to accurately identify different bird species given an input image while only having a few training images. With over 10,000 bird species, some rarely ever photographed, this project idea was inspired by mitigating the great time, resources, and expertise required to identify a bird in the wild. A model that is easily accessible and able to accurately distinguish birds based on images would be useful for a bird spotter or even casual hiker. Finding only a small dataset from kaggle.com containing a set of 200 bird species, our model is only trained to identify those.

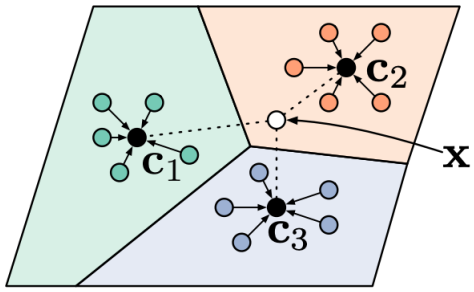
II. Background

Deep learning often uses large data sets in order to train models. Few-shot learning is intended to train a model using a small training set with few representations of each class. Prototypical networks are used to be able to generalize so that classes not seen during training can be identified during testing given only a few number of representations of each class.

A. Few-Shot Learning

Few-shot learning, as stated in the background, feeds a limited amount of data into a learning model which would be strongly overfitted when using a traditional deep learning model that necessitates large datasets to train. Since our Kaggle dataset is relatively small for deep learning, approximately 30,000 photos, a few-shot learning model was decided to be the most reasonable deep learning technique to use. Few-shot learning can be classified for multiple styles as n-shot, k-way with the most popular being 5-shot, 1-shot and 0-shot. Since the bird identification training set utilizes 5 images of a given bird, the model is classified as 5-shot. K-way specifies how many bird species the input sample will be compared against. With this model, it was found that 5-way yielded the most accurate results, so in summation our few-shot learning model is 5-shot, 5-way.

B. Prototypical Networks



(a) Few-shot

Fig. 1. Few-Shot Learning Illustration (Snell et al., 2015)

The few-shot learning model utilizes Prototypical Networks to achieve its picture classification. The networks gradually learn metric spaces where classifications can be determined by calculating the distances to a prototype representation for every class. The prototypes are created by averaging all points confirmed to be a part of the given class. For example, consider figure 1. If Blue jays might be found in the blue region,

Dr. Brian Hutchinson

Cardinals in the red, and Parrots in the green, the new sample image X is closest to the Cardinal prototype so that bird image would be classified as a Cardinal. Whenever a new image is added to a given class, the associated prototype is then updated, achieving a more accurate representation of the class through each pass. The network is semi-supervised, which provides the framework for establishing the prototype for every given class. Since the bird dataset has an uneven distribution of pictures per bird species, some prototypes will be built off 100 images, while others may draw from several hundred or thousand. The prototypes are then adjusted with each new episode through back-propagation.

III. Prior Work

"Learning to Compare: Relation Networks for Few-Shot Learning" goes into the use of using deep learning from a limited dataset by using a technique called few-shot learning. Datasets can be too small and risk learning a model that will overfit for the specific case rather than generalize, and few-shot learning is a strategy to fix overfitting. There were multiple papers that used convolutional neural networks (CNNs) for image classification including the papers by Huang and H. Basanta, and Meena and Agilandeeswari. CNNs accomplished image classification via feature extraction, which could include examples of bird size, color, wing type, etc. "Optimal Approach for Image Recognition using Deep Convolutional Architecture" addresses 5 different deep convolutional architectures for various image recognition tasks and concluded that on average Inception-ResNet had the highest accuracy while also keeping the computational requirements in-check.

IV. Methodology

We utilized the vgg11 model with the final classification layer removed. The layers of this model can be seen in column A of figure 2. This modified deep convolutional neural network allowed us to create tensors of size 4096 which represent various features extracted from our given images. Five classes are randomly selected for each episode, with ten samples from each class then being chosen. Prototypes for each class are created using five of the samples, while the remaining 25 samples are used as the queries. Standard euclidean distance calculations were done between the tensors created from the prototypes and the tensors created from the queries. Cross-entropy loss was then used to determine the probabilities of a given query being a given class.

V. Experimental Results

The dataset utilized from kaggle.com consisted of 200 bird species. Although the dataset is not perfectly balanced, there are a minimum of 100 images for each species, each of size 3x224x224. There is a significant bias in the male to female ratio of 4:1, this is because female birds tend to be more bland, while male birds are often diverse in color. To train and test the model the dataset was divided into three disjoint sets. The sets were a train, test, and dev set consisting of 140 species, 30 species, and 30 species respectively. Initially,

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Fig. 2. Very Deep Convolutional Networks (Simonyan & Zisserman, 2014)

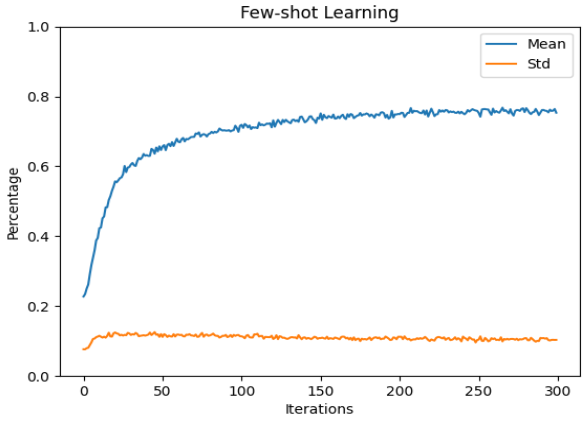


Fig. 3. Results of 300 Iterations at 500 Episodes with no pre-training

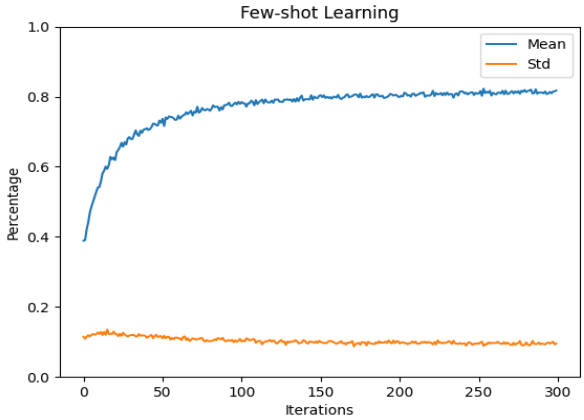


Fig. 4. Results of 300 Iterations at 500 Episodes with pre-training

the bird identification model went through 100 iterations with 100 episodes in each iteration, which achieved 40% accuracy. Increasing the number of episodes per iteration to 500 and the number of iterations to 300 increased the model's accuracy to 75%. Our highest accuracy was obtained by pre-training the model on imagenet and then doing 300 iterations of 500 episodes with 5-way, 5-shot learning, which attained 80% accuracy. Each trained model achieved similar standard deviations, ranging from 0.09 to 0.11. Image augmentation through affine transformations as well as slight variations in the color, contrast, saturation, and hue was also done on the testing dataset, though the accuracy and convergence time of the model remained similar to when it was pre-trained on imagenet.

VI. Conclusion and Future Work

The bird identification model operating with Few-Shot Learning with Prototypical Networks is able to converge to a sufficiently high accuracy of 80% after being pre-trained on imagenet. The major strength of this system is its capability to accurately classify classes that it has very little or no training with. A major weakness of a system is that it likely is not very accurate at classifying female birds, as there is a distinct amount of overlap in their appearances as compared to that of male birds. A way to possibly fix this issue would be to separate the data into male and female datasets and train two models on each one. In regards to prototypical networks, future work could involve further testing of image augmentation in order to better train the model. This effectively creates a larger training dataset by altering the current images and rotating them to create unique additional images that would appear to be completely different to the model. In addition, had time permitted, it may have been worth modifying the vgg11 model further in order to achieve faster or more accurate convergence.

References

[1] Gerry. (2020, May 5). 200 Bird Species. Retrieved from https://www.kaggle.com/gpiosenka/100-bird-species

[2] Snell, J., Swersky, K., Zemel, R. (2015, March 15). Prototypical Networks for Few-shot Learning. Retrieved from https://arxiv.org/abs/1703.05175

[3] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. S. Torr and T. M. Hospedales, "Learning to Compare: Relation Network for Few-Shot Learning," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018, pp. 1199-1208.

[4] Y. Huang and H. Basanta, "Bird Image Retrieval and Recognition Using a Deep Learning Platform," in IEEE Access, vol. 7, pp. 66980-66989, 2019.

[5] S. Divya, Meena and L. Agilandeeswari, "An Efficient Framework for Animal Breeds Classification Using Semi-Supervised Learning and Multi-Part Convolutional Neural Network (MP-CNN)," in IEEE Access, vol. 7, pp. 151783-151802, 2019.

[6] Simonyan, K., Zisserman, A. (2014, September 4). Very Deep Convolutional Networks for Large-Scale Image Recognition. Retrieved from https://arxiv.org/abs/1409.1556

[7] P. Shah, V. Bakrola, S. Pati, "Optimal Approach for Image Recognition using Deep Convolutional Architecture,".

[8] Branson, Steve & Horn, Grant & Belongie, Serge & Perona, Pietro. (2014). Bird Species Categorization Using Pose Normalized Deep Convolutional Nets. BMVC 2014 - Proceedings of the British Machine Vision Conference 2014.