Third report - Bird classifier

Ivan Dávila Puente, Giovana Guadalupe Rodríguez Donjuan

UASLP, Machine Learning.
Dr. Juan Carlos Cuevas Tello

May 30, 2024

Abstract

The following work describes the model used for the task of image classification of different bird species with the help of machine learning, as well as the data that was used as input for training the model and its subsequent evaluation. To clarify these points, we must first understand the information we are working with. Online communities have formed to promote the development of code based on machine learning, allowing people from different parts of the world to upload their own datasets on any subject. In this work, after comparing several datasets found on the internet, as well as models that performed the task, an implementation was recreated to see what results we would obtain, in addition to the dataset being updated over the years, adding more images and classes. Once the results were obtained, it is also important to evaluate the model and how accurately it performed its task.

1 WHAT IS A DATASET?

According to the Oxford Dictionary [Oxford(n.d.)], a dataset definition in machine learning is "a collection of data that is treated as a single unit by a computer". This means that a dataset contains a lot of separate pieces of data, but can be used to teach the machine learning algorithm to find predictable patterns inside the whole dataset.

Choosing a good dataset and collecting a substantial amount of data is very important in machine learning because the more images our model has to train on, the better it will identify the patterns that occur among images belonging to the same class, resulting in more accurate classification.

A dataset is not only used for training but can also be grouped into three phases:

1. Training images, which, as mentioned earlier, are used to feed the model.
2. Validation images, which is an optional phase considered a sub-phase of training. Sydorenko [Sydorenko(2023)] states that Validation is useful to avoid training the algorithm with the same type of data and making biased predictions.
3. Finally, the testing phase, where several images are given to the model and its accuracy is measured based on how it labels the images.

Is important to have enough data to train the algorithm properly, but there is also a posibility of overtraining an algorithm (known as overfitting). In conclusion, there is no recipe for how much data you need, sometimes you will have to resort to trial and error or ask for help to a data scientist.

2 BIRDS 525 SPECIES- IMAGE CLASSIFICATION

Birds 525 species is a dataset with a lot of images of birds for classification purposes. This dataset has been updated over time, so the number of classes and images has varied over the years. As of today, the dataset consists of three folders: one with training images, another with validation images, and another with test images. Each folder contains 525 sub-folders, each representing a different bird species. The number of images within the training sub-folders varies, but all have at least 130 images. The validation and test sub-folders each contain only 5 images. The images are 224 x 224 x 3 in size and are in jpg format. There are 84635 training images, 2625 test images and 2625 validation

images, making a total of 89,885 images. Piosenka says [Piosenka(2023)] that Images were gathered from internet searches by species name. Once the image files for a species was downloaded they were checked for the presence of duplicate or near duplicate images. All duplicate or near duplicate images detected were deleted in order to prevent their being images common between the training, test and validation sets. After that the images were cropped so that the bird in most cases occupies at least 50% of the pixel in the image. Then the images were resized to 224 X 224 X3 in jpg format. The cropping ensures that when processed by a CNN their is adequate information in the images to create a highly accurate classifier. Even a moderately robust model should achieve training, validation and test accuracies in the high 90% range.

3 CONVOLUTIONAL NEURONAL NETWORK

After doing an extensive analysis and comparison of the different classifiers using the dataset, we decided to use "Transfer Learning:500 Bird Species Classification", since its accuracy level is almost 97% and its training time is much less compared to the other classifiers. A convolutional neural network (CNN) is a specialized type of deep learning algorithm designed primarily for tasks that require object recognition, such as image classification, detection and segmentation [IBM(2024)]. This classifier uses the CNN "ResNet50(Residual Network)",this architecture learns residual functions with reference to layer inputs, instead of learning functions without reference [with code(2021)]. Instead of expecting each of the few stacked layers to directly match a desired underlying map representation, residual networks allow these layers to match a residual map representation. They stack residual blocks on top of each other to form a network,ResNet-50 has fifty layers that use these blocks.

4 RESULTS

This model uses pre-trained weights from the ImageNet dataset, to pre-train the model we did not include the original ResNet50 classification layer because this classifier has its own classification layers. A fully connected classification layer is added with the number of neurons equal to the number of classes, it uses the 'softmax' activation function to convert the outputs into probabilities for each class. It also uses a global averaging pooling layer after the convolutional layers. This converts the output of the last convolutional layer into a feature vector. And finally it uses the 49 predefined convolutional layers of the network.

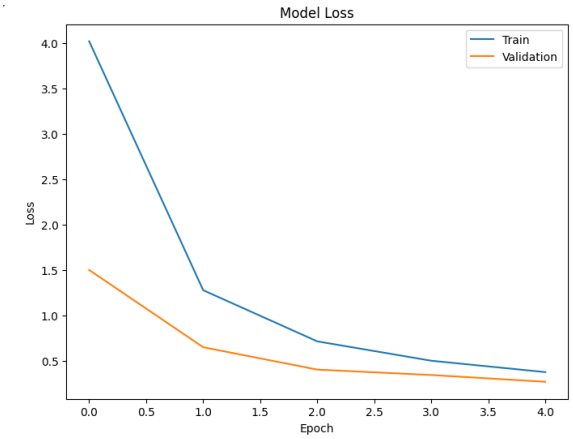

Figure 1: Model Loss

Figure 1 shows the training and validation accuracy over the epochs. It can be observed that the model achieves high accuracy on both the training and validation sets, indicating good generalization performance.

Figure 2: Results: Classification and comparison

Figure 2 illustrates some examples of the model's predictions on the test set, showing both correctly and incorrectly classified images.Each row represents a different class, and the images display the actual class and the predicted class. This provides insight into the model's performance and highlights areas where the model performs well and where it struggles.

The results indicate that the model performs well on the classification task, achieving an accuracy of 96.89% on the test set. The confusion matrix reveals that most classes are correctly classified. Future work could involve further tuning of hyperparameters and exploring advanced data augmentation techniques to improve performance.

The training and validation loss curves in Figure 1 show a steady decrease, indicating that the model is learning effectively and not overfitting. The accuracy curves (not shown) likely exhibit a similar trend, supporting the conclusion that the model generalizes well to unseen data.

5 CONCLUSION

The results obtained were quite good, largely because the images in the dataset were distributed in such a way that the model did not struggle to recognize the patterns within the same species. As for the model, we opted for the most common approach in image classification, which is CNNs. We had planned to use EfficientNetB0, but when training the model in Colab, the interface would stop because we exceeded the available time for users who do not pay for more computing units. For this reason, we decided to look for an alternative that used transfer learning so that the model training would not be so heavy, and indeed it was, as within minutes the model was ready to be put to the test.

We were satisfied with the work done; the original work [SHAH(2023)] used 515 classes, whereas we tested it with the current version of the dataset that added another 10 bird species. With this update, a slight decrease in the

model's performance can be observed, which dropped from 96.89% accuracy to 96.23%

References

[IBM(2024)] IBM. ¿qué son las redes neuronales convolucionales?, May 2024. URL https://www.ibm.com/es-es/topics/convolutional-neural-networks.

[Oxford(n.d.)] Oxford. Definition of data set noun from the oxford advanced learner's dictionary, n.d. URL https://www.oxfordlearnersdictionaries.com/definition/english/data-set?q=data+set.

[Piosenka(2023)] Gerry Piosenka. Birds 525 species- image classification, 2023. URL https://www.kaggle.com/datasets/gpiosenka/100-bird-species.

[SHAH(2023)] AASHNA SHAH. Transfer learning: 500 bird species classification, April 2023. URL https://www.kaggle.com/code/aashnaashahh1504/transfer-learning-500-bird-species-classification.

[Sydorenko(2023)] Iryna Sydorenko. What is a dataset in machine learning: Sources, features, analysis, december 2023. URL https://labelyourdata.com/articles/what-is-dataset-in-machine-learning#:~:text=A%20machine%20learning%20dataset%20is,same%20way%20as%20humans%20do.

[with code(2021)] Papers with code. Resnet, February 2021. URL https://paperswithcode.com/model/resnet.