Bird species classification

Adrian EL BAZ
ENS Paris Saclay
adrian.el_baz@ens-paris-saclay.fr

Abstract

This project is about multi-class classification of bird species from a small dataset. We used transfer learning from pre-trained ImageNet models to achieve our best score. We used Mask R-CNN to detect birds in the original images and create a new dataset from these crops. Finally, an ensemble method containing 3 models was used for the final prediction. Several models were tested but it appears that architecture with residual modules were able to perform the best.

I.Introduction

Our dataset consists of 1078 images that are well distributed over classes. Our validation has a little bit more than 100 images. At the beginning we first explored the data to spot potential mislabels. We found that several images were very hard to classify due to backlighting or shades. These images are often black birds.

II. Data augmentation

We considered several types of data augmentation for this dataset. First of all we changed the default size to 400x400 for original images and 299x299 for images produced by the bird detection algorithm. Then we also added Horizontal flip and a ColorJitter to change the brightness, saturation and contrast of the image. The previous 2 transformations were applied randomly to the images generated in our Data loader. Images generated by the detection algorithm were also randomly cropped (200x200), this choice will be justified in the next section.

III.Mask R-CNN

After some investigations with standard architecture, we noticed the difficulty for our network to predict test images classes above 75% accuracy. Therefore we decided to use Mask R-CNN detection algorithm to detect birds and then train the networks on these new cropped images. Indeed we couldn't crop the original images without risking that it wouldn't even contain the bird because of the different proportion of birds shape with respect to image. The code used to perform this task was heavily inspired by this GitHub repository[1].

IV.Neural Networks Architecture and training

We considered several architecture models in this project. Firstly, all of our models were using pre-trained weights from ImageNet classification. The maximum number of layers were kept un-freezed (depending to Colab's GPU capacity). The more we had layers un-freezed, the more validation accuracy we got from models. Details about training hyper-parameters can be found in our code.

We used 3 networks : 2 Resnet152 networks, mainly trained on the two different set of images and InceptionResnetV2 trained on both set of images. Pre-trained weights of this more recent architecture can be found in Model Zoo[2].

Inception networks are very interesting because they are testing the effect of different type of convolutions at the same time. We wanted to benefit from this general architecture that should in our eyes tackle various tasks.

From our experiments, the aggregation of these models performed better than any single model. The final prediction was *based* on the highest predicted class probability among these networks. The reason why we thought ensemble method would perform better was that the different networks had different errors among classes, and we noticed that when a network wrongly predicted a class, it did so lower confidence in average. The following figure represents the confusion matrix of our predictions from the ensemble method. Errors are now much more distributed among black birds, indicating that these classes are not well learnt.

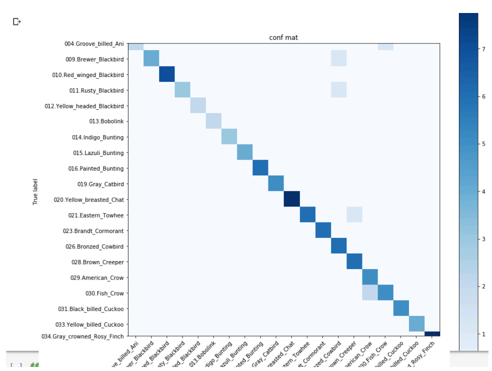

Figure1. Confusion matrix. Errors are from black birds.

V. Conclusion

We obtained 83.225% accuracy on the Kaggle public score. The results could be made better by doing a better data augmentation process. Indeed doing different data augmentation per class could have been interesting, especially to make black birds more recognizable among them. We also thought about data augmentation with GAN which could be very promising by looking at this paper results [3]. The GAN consists of creating a new image of bird using the pose of another bird, making the dataset much more diverse.

References

1. https://github.com/matterport/Mask_RCNN
2. https://modelzoo.co/model/pytorch-cnn-finetune
3. Shuang Ma, Jianlong Fu, Chang Wen Chen, DA-GAN : Instance-level Image Translation by Deep Attention Generative Adversarial Network. CVPR 2018