Object Recognition and Computer Vision
Assignment 3: Bird classification

Ambroise Odonnat
Master MVA
ENS Paris-Saclay
ambroise.odonnat@eleves.enpc.fr

Abstract

The purpose of the assignment is to classify images of 20 *species of birds taken from the* ***Caltech-UCSD Birds-200-2011 bird dataset****. Provided with only 1185 images, the goal is to achieve the highest possible accuracy on the test set. We rely on small models (*$\leq$ *28 millions parameters) pretrained on ImageNet and show promising results with a final accuracy on test set of 86.5* %.

1. Dataset

The training set contains 1082 images, the validation set contains 103 images while the test set contains 517 images. Images were resized to 256×256 and normalized using the mean and standard deviation of the ImageNet dataset.

2. Preprocessing

Cropped Images We cropped images using a pretrained Mask R-CNN from the Facebook Research Detectron's library ([3]). Images with a prediction confidence higher than 70% were cropped outside of the predicted bounding box. It increases the size of data and enhances the quality of images by removing some background. It significatively improved the results.

Data Augmentation To avoid overfitting, we used custom gaussian blur, random affine, random horizontal flip, color jitter, random rotation and random erasing Pytorch transformations during training. Images on the validation and test sets were only resized and normalized as explained in 1.

3. Architecture of the Model

The first model is a pretrained ResNet34 ([1]). We replaced the last two layers by a two-layer MLP with Tanh activation and added a block in the forward to fine-grained the classification. The second model is a pretrained ConvNext ([2]). We replaced the last layer by a two-layer MLP with ReLU activation.

We used a batch size of 64 for the ResNet34 and of 32 for the ConvNext. Training is done in two stages: 15 epochs training only the MLP and 15 epochs training all the layers. We used Adam optimizer with lr = 1e-4, a StepLR scheduler with $\gamma = 0.75$ and step_size $= 15$. Early stopping was used with a patience of 5 epochs.

4. Results

ConvNext performs the best with an accuracy on the test set of 86.5%. ResNet34 achieves at best an accuracy of 80% in test. Accuracy is given in % in the Table 1. The column "Crop" means that the model was trained on all images and tested only on the cropped ones. This technique decreases the validation performances but improves the test ones.

Model	Crop	Validation Accuracy	Test Accuracy
ResNet34	No	83.8	71.6
ResNet34	Yes	81.6	80.0
ConvNext	**Yes**	**91.3**	**86.5**

Table 1. Models Results

References

[1] Kaiming He, X. Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 1

[2] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s, 2022. 1

[3] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. Detectron2. https://github.com/facebookresearch/detectron2, 2019. 1