# Object recognition and computer vision
# Assignment 3: object classification

Clémence Grislain
clemence.grislain@eleves.enpc.fr

## Abstract

*We are interested in classifying 20 species of birds thanks to a dataset of 1702 labelled images - divided into training, validation, and testing sets. To address this problem, we used pre-trained ResNet models on Imagenet from the torchvision library to perform transfer learning, data augmentation. We also used a pre-trained detection model to crop the data around the birds.*

## 1. Data

The labelled dataset as a whole is made up of 1082 pieces of training data (approximately 55 images for each species), 1033 pieces of validation data, and 517 data of testing data. The data in the training and validation sets are mainly of good quality and centered on the bird. However, some of the testing data are noisy, and the birds are not centered but flying or at the top of trees far away. We increased the size of the training image to (256, 256) to ensure the model had enough information to learn. To fit these higher resolution data in memory, we reduced the batch size from 64 to $16 or 32$ depending on the model size. In addition, to artificially increase the training data base and avoid overfitting, we implemented data augmentation inside the data loader with horizontal flips (with probability 0.5) and rotations (with probability 1 less than 20 degrees). Finally, we used a pre-trained detectron2 model trained on COCO to extract a bounding box around the bird [2] and crop all the dataset (train, validation, and test) around this bounding box.

## 2. Architecture

In the literature, Resnet models [1] have proven to be really effective architectures for image classification on relatively small datasets. We used pre-trained ResNet on Imagenet of sizes 50, 101, and 152 from the library torchvision and trained them on the bird dataset with different frozen layers. These architectures are composed of four blocks and a final dense layer. We fine-tuned the pre-trained models on the bird dataset with the three first blocks' weights frozen. We also changed the output size of the dense layer to match the number of classes in our problem (20). Finally, because we observed overfitting as a result of only optimizing a few parameters, we replaced the last dense layer with the block $head = linear, Relu, Dropout(0.25)$, $linear, Relu, Dropout(0.5), linear$. Yet, this last head provided worse results than the classic dense layer.

## 3. Parameters

We tried to optimize the model using SGD with momentum of $0.9$ and a learning rate of $0.1$ and Adam with a learning rate of $0.01$ (default parameters: $\beta_1 = 0.9$, $\beta_2 = 0.999$). I also used a learning rate scheduler, which multiplies the starting learning rate by $gamma = 0.1$ every $10$ epochs. The best results were obtained by combining the Adam optimizer with learning rate decay (this result was observed on the validation set, and all testing results were obtained by using the Adam optimizer). To avoid overfitting, we trained our fine-tuned models from 10 to 30 epochs.

## 4. Results

Pre-trained ResNet-50 and ResNet-101 with unfrozen $block_4$ and the last layer replaced by $head$ achieve accuracy of $56.128\%$ and $58.709\%$ on the provided bird dataset. Using the biggest ResNet model, ResNet-152, with the same fine-tuning configuration as above increased the accuracy to $71.612\%$. Fine-tuning the classic ResNet-152 architecture with unfrozen $block_4$ and the last dense layer reaches an accuracy of $75.483$. Finally, the best result, $84.516\%$ of accuracy, was obtained by fine-tuning the ResNet-152 as before but training and evaluating it on the dataset cropped around the birds' bounding boxes.

## References

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015. 1

[2] The Imperfect. Object Detection Tutorial using Detectron2, 2022. 1