Assignment 3 : Image Classification

Théophane Gregoir

1. Processing

1.1. Bird detection and localization

Before applying classification models to our dataset, it is essential to process images to ensure each of them is representing precisely an instance of a given class without being surrounded by too much noise. Therefore, I decided to use Faster R-CNN [4] pretrained on COCO dataset (containing a class 'Bird') in order to localize the bird on each data point and generate inputs for classification models (see **Figure 1**).

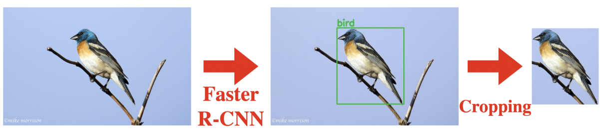

Figure 1. Example of data processing thanks to box prediction by Faster R-CNN model on a sample from Lazuli Bunting class

1.2. Balancing datasets

The provided train and validation datasets appeared to be unbalanced. Consequently, train and validation datasets were created randomly thanks to a $80\%/20\%$ split.

1.3. Data augmentation

For training and validation, classical transformations were randomly applied to images (like rotation, horizontal flip and color jitter) in order to increase robustness of models and extend datasets.

2. Multiple Transfer Learnings

Using transfer learning, 5 well-known classification models pretrained on Imagenet were finetuned on our datasets : VGG-16 [5], Resnet152 [1], Resnext50-32x4d [7], Efficientnet-B0 [6], Densenet161 [2].

Trainings were done on 20 epochs thanks to Stochastic Gradient Descent backward propagation using a learning rate between 0.001 and 0.01 (with batch size = 16).

	VGG	Res.	Resx.	Eff.	Den.
BB	$90,74$	$97,67$	$91,80$	$98,18$	$94,92$
AC	$80,88$	$90,70$	$84,06$	$94,92$	$84,29$
FC	$84,75$	$70,89$	$94,55$	$88,52$	$92,45$

Table 1. Precisions (%) of the 5 models on merged training and validation sets for 3 chosen classes : Brewer Blackbird (BB), American Crow (AC), Fish Crow (FC). Models need each other !

Each model presenting different forces and weaknesses, majority voting appeared as a relevant solution to take advantage of the diversity of models (see **Table 1**).

3. Reliable Majority Voting (RMV)

Instead of using classical majority voting, Reliable Majority Voting (RMV) was used, inspired by a recent article [3]. An ensemble model gathering the 5 models presented before was developed.

For each prediction, each model votes with a weight corresponding to their precision (on training set) linked to the class it predicted. The class predicted by the ensemble model is the class gathering the highest sum of weights.

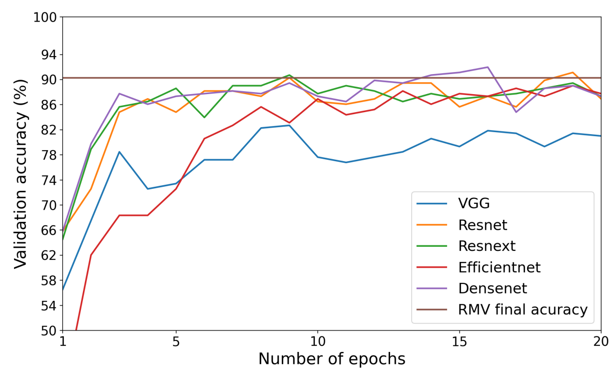

Figure 2. Validation accuracies for the 5 models trained on the same training set. Here, RMV represents the validation accuracy reached by Reliable Majority Voting gathering the 5 trained models with the best validation accuracy.

RMV ensemble model reaches a validation accuracy of $90,29\%$ which is almost better than any of the 5 models at any state (see **Figure 2**). Moreover, it is important to note that RMV model is more robust than any model taken separately as it will take advantage of the features extracted by all its components.

4. Conclusion

As a conclusion, RMV provides a robust and precise prediction method which topped the single-model results initially suggested (on the test set). To go further on this project, given a large dataset of unlabelled birds, we could train an Auto-encoder and then use its encoder part as a feature extractor for a classification model. This new classification model, once finetuned on our datasets, could be added to the RMV pool of models.

References

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

[2] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks, 2018. 1

[3] Agus Budi Raharjo, Mohamed Quafafou, and Faicel Chamroukhi. On Reliability of Majority Voting. In *Le 24th conférence de la Société Francophone de Classification (SFC 2017)*, Lyon, France, June 2017. 1

[4] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks, 2016. 1

[5] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2015. 1

[6] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks, 2020. 1

[7] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks, 2017. 1