Assignment 3: Fine-Grained Image Classification on a Birds Dataset

Pierre Fernandez
pierre.fernandez@polytechnique.edu

Abstract

In this project, we deal with a subset of the Caltech-UCSD Birds-200-2011 bird dataset. The objective is to produce a model that gives the highest possible accuracy on a test dataset containing the same categories. All the resources needed for the project can be found here [1].

1. Introduction

Fine-grained image recognition is a challenging computer vision problem, due to the small inter-class variations caused by highly similar subordinate categories. In this project, we first used a network to detect birds on the images and created an auxiliary cropped dataset, to make sure the further models would only use the most valuable information. We then tried different methods that are used in FGIC, namely finetuning a pretrained neural-network, feature extraction and classification based on [1] and attentive pairwise interaction network based on [2].

2. Bird Detection

The first step was to create a dataset made of the cropped images containing only valuable information for the further steps. To do so, a pretrained RetinaNet model was used to detect birds, with a relatively low threshold of 0.1. The cropped image was then created from the bounding box of highest probability. Doing so, the model did not find a bird on only 5 images (the remaining images were manually cropped).

3. Bird Classification

Different classification methods have been tried out. Then a custom majority vote based on the predictions of the best model in each of the following categories was used on the test dataset.

3.1. Finetuning a Pre-trained Model

The more straight-forward approach is to use a pytorch model pre-trained on ImageNet. Several of them have been trained and validated, for instance ResNet-50, ResNet-101, ResNet-152, ResNeXt-101. We finetuned the network's last module with a classic training process. The best results were obtained with an SGD optimizer with a learning rate 0.01, a momentum 0.9 and a weight decay $3.0e-4$, with a cross-entropy loss and a cosine-annealing.

3.2. Feature Extraction and Classification

This method transfers the knowledge learned from a large scale dataset: ImageNet, via feature extraction. It processes the image and extracts the information of the "Mixed_7c" layer of an InceptionV3 trained on ImageNet (with an input size of 299). The extracted features can then be used in usual classifiers (Random Forest, Logistic Regression, Gradient Boosting, etc.).

3.3. API-Net

Here, the goal is to learn a mutual vector first to capture semantic differences in input pair of images. It then compares this mutual vector with individual vectors to generate gates for each input image. For more details, see [2].

3.4. Results

Here is the accuracy of the best model found with the different methods in the validation dataset classification task:

Method	Accuracy
Finetuning	0.93
API-Net	0.89
LR Classification	0.93

4. Conclusion

To conclude with, this project points out the importance of transfer learning and showcases several SOTA methods in FGIC. The Kaggle submission gave 0.8129 accuracy on the test dataset. This score should be improved by a more rigourous hyperparameter search and by pre-training on a source domain that is more similar to the target domain (namely not ImageNet, but INaturalist for instance, as mentioned by [1]).

[1] https://github.com/willowsierra/recvis20_a3

References

[1] Yin Cui, Yang Song, Chen Sun, Andrew Howard, and Serge Belongie. Large scale fine-grained categorization and domain-specific transfer learning. 2018. 1

[2] Peiqin Zhuang, Yali Wang, and Yu Qiao. Learning attentive pairwise interaction for fine-grained classification. 2020. 1