Image classification challenge report

Quang Anh Nguyen
MVA, ENS Paris Saclay
quanganhnguyen95@gmail.com

Abstract

The objective of this challenge is to solve the problem of image classification for 20 species of birds. This report summarizes some attempted approaches and their results.

1. Introduction

To resolve this problem, we adopt the transfer learning approach by adapting pretrained models from ImageNet dataset and perform fine-tuning on these models. Furthermore, to improve the results, we apply these models on cropped images of the bird obtained from a detection model.

1.1. Dataset

Here we are provide with the Caltech-UCSD Birds-200-2011 bird dataset, containing images of 20 specied of birds.

The images in the training set and validation set are clean, the birds are nicely framed and visible. However, in the test set, some birds are small and there are sometimes occlusion or other objects. That is why we try cropping the bird from images before training.

1.2. Pretrained models

For the classification task, we try using pretrained models: ResNet[2], EfficientNet [4], ResNext [5], ViT[1], all pretrained on the ImageNet dataset.

We replace only the last linear layers of these models to adapt to the number of classes and freeze other layers. We also tried to unfreezed the second to last layer but this does not yield significant improvements on the validation set.

1.3. Data preprocessing

To crop the birds in images, we use Faster R-CNN [3] for instance detection. For each image, the bounding box with highest probabilities among those containing in the bird class is chosen for cropping. This procedure is performed on all three datasets.

1.4. Training and regularization

The optimized used was SGD with learning rate 0.01, momentum 0.9 and scheduled to decrease by 0.1 if no improvement on the validation set after 10 epochs.

We also apply L^2 regularizer with coefficient 0.1 for the trained layer to avoid overfitting. Data augmentation includes: random cropping and resizing, horizontal flipping, small random rotation.

2. Result

2.1. Experimental result

We have experimented 4 methods, on the original and cropped images. The results are summarized in table 1.

Data	Method	Validation	Public test
Original	ResNet	87%	74%
	EfficientNet	79%	70%
	ResNext	89%	74%
	ViT	91%	80%
Cropped	ResNet	92%	77%
	EfficientNet	83%	71%
	ResNext	92%	76%
	ViT	93%	85%

Table 1. Accuracy on validation set and public test set

2.2. Conclusion

We observe that employing the data cropped from Faster R-CNN helps to improve the overall performance of classification models. In particularly, vision transformer (ViT) outperformed other models on both datasets. Contrarily, EfficientNet gives the worst results among the four methods.

References

[1] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale, 2021.

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

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

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

[5] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1492–1500, 2017.