

Selected Topics in Visual Recognition using Deep Learning Homework 1

Github link of the codes

Project Repository: [VRDL_21au](#)

Reference

Github repository [PMG-Progressive-Multi-Granularity-Training](#) is used and modified in this project.

Introduction

The project is an experiment of implementing deep learning model to deal with bird image classification task. The dataset used here is Caltech-UCSD Birds 200 2010, which is an image dataset with photos of 200 bird species. It is composed of 6,033 images and the size of training/testing sets are 3,000 / 3,033 images respectively. The main feature of dataset CUB 200-2010 is that the size of it is relatively small compared to other open-datasets such as ImageNet 1k or ImageNet 21k, and the scope of the labels is narrow too. Hence, the major difficulty of this problem is that we need to identify the slight difference between bird species. We compare the performance of basic CNN models with fine-grained-based models, and also apply some techniques such as transfer learning, learning rate scheduling and data augmentation to improve the accuracy of model.

Data Preprocessing

For each model mentioned below, we use specific preprocessing methodology. However, some common operations are shared. Firstly, since the size between images are different, random-cropping / center resizing is implemented. Secondly, 0-1 scaling and normalization are used to transfer RGB color bit to floating numbers, fitting the general input layer of convolutional neural network.

We do some file organization on dataset too. The images are rearranged into pytorch Imagefolder style to fit the commonly-used pytorch interface. Also, since we lack the information about the labels of testing set, 10 percent of the original training images are used as testing sets to evaluate the model performance. All images are used to retrain the model eventually after all hyperparameters are determined.

Transfer Learning

To utilize the domain knowledge learned from other image classification tasks, the weights of

ResNet model are initialized with the ones pretrained on ImageNet, and only the last layer (fully-connected part) is replaced to finetune the whole model. For fine-grained model mentioned later, we also use pretrained model in its counterpart.

Data Augmentation

Data augmentation in data analysis are techniques used to increase the amount of data by adding slightly modified copies of already existing data. It acts as a regularizer and helps reduce overfitting when training a machine learning model. Here we apply AutoAugment as part of our preprocessing workflow to generate more training samples. It's an image transformer pipeline composed of several different type of transforming polociies, such as rotation, sharpness, colorjitter etc. More details about its implementation could be found [here](#).

Fine-Grained Visual Classification via Progressive Multi-

Granularity Training of Jigsaw Patches

We also test the performance of some fine-grained-based methodology. Progressive Multi-Granularity Training (PMG) was proposed by *Ruoyi Du, Dongliang Chang, Ayan Kumar Bhunia, Jiyang Xie, Zhanyu Ma, Yi-Zhe Song, Jun Guo* in paper "Fine-Grained Visual Classification via Progressive Multi-Granularity Training of Jigsaw Patches" in 2020. The core idea of PMG is that selecting appropriate granularity to locate the discriminative part of images is an important factor to affect the performance of image classifier.

To be more concrete, the author proposes (extracted from the paper) a novel progressive training strategy that adds new layers in each training step to exploit information based on the smaller granularity information found at the last, and a simple jigsaw puzzle generator to form images contain information of different granularity levels.

Experiment Result

Model	Feature	Accuracy
ResNet34		62.64%
ResNet50		68.38%
ResNet50	Auto Augment	65.28%
ResNet50	PMG	84.57%
ResNet101	PMG	85.30%
Ensemble	PMG50 + PMG101	86.12%

Summary

As we could see above, the basic implementation of ResNet family models could achieve an accuracy of around 65%, and Res50 performs better than Res34. However, applying advanced

data augmentation on current model does not improve the performance, and some further analysis should be done here. Also, we find that fine-grained methodology PMG-Training could significantly increase the accuracy of the model by a ratio of 20%.