Winner Solution for AliProducts Challenge: Large-scale Product Recognition

Yuanzhi Liang, and Wei Zhang
JD AI Research
{liangyuanzhi3,zhangwei96}@jd.com

This report presents the final 1st solution of the Aliproduct competition in CVPR2020. Our solution contains two base methods, which are the DCL and LIO. The basic backbone of networks are Efficientnet-b3, Efficientnet-b4, resnet50, seresnext50, seresnet101. We train all the backbone and finetune the backbone with DCL and LIO methods. We also design an accuracy loss to optimize the top-1 accuracy directly. Finally, we ensemble all the models and achieve the top-1 error rate 6.27% on the test set.

1. Backbone Training

In our solution, we train four backbones in original dataset first, which are resnet50, seresnext50, efficientnet-b3, efficientnet-b4. All backbones used the pretrained ImageNet models. The models have 76-82 top-1 accuracy rate in validation set after the basic training process. In the performances in test set from the leadboard, the models have 20-25 top-1 error rate.

Since training with unbalanced and overall dataset leads the models achieve basic performance in classification, and not perform well in the balanced validation set, we finetune all the backbone in a balanced training set, in which all the categories contains 30 images (using all images if the number of images in the corresponding category is less than 30). All the backbones have dramatic improvement in both validation set and test set. With the balanced finetune, all models come to 9-12 top-1 error rate in leadboard. The best single model in our submission is efficientnet-b3 and has 9.78% top-1 error rate in leadboard.

Some settings are below:

Image size: Resize all image to 256*256 and random corp 224*224 for training. Resize to 256*256 and center crop 224*224 for test.

Augmentation: RandomResizedCrop, RandomHorizontalFlip, AutoAugmentation,Cutout, Nomalization.

Optimizer: SGD

Scheduler: Manual learning rate decay

Loss function: CrossEntropy

2. DCL finetune

Destruction and Construction Learning (DCL) [1] aims to enhance the feature representation ability for backbone network. The technical details about DCL can be found in paper "Destruction and Construction Learning for Fine-grained Image Recognition" (http://ylbai.me/DCL_cvpr19.pdf). Code link: https://github.com/JDAI-CV/DCL.

In this stage, we reloaded the backbone models with balanced finetuned and use DCL to finetune them on higher resolution training images (448*448). All the backbone improve 1-2 points in top-1 error rate in leadboard.

3. LIO

In our solution, we also train a resnet50 method with LIO [2]. The "look into object" (explicitly yet intrinsically model the object structure) through incorporating self-supervisions into the traditional framework. This method show the recognition backbone can be substantially enhanced for more robust representation learning, without any cost of extra annotation and inference speed.

4. Accuracy Loss

We also design an accuracy loss to optimize accuracy in each batch. The loss can be defined as following:

$$loss = 1 - mean(\frac{\hat{y} * y}{(\hat{y} * y) + (1 - y) * \hat{y}}) \qquad (1)$$

y and $\hat{y}$ indicate the prediction and ground truth in one-hot representation.

We use this loss the finetune the model in 2-3 epoch in balanced train set, and the improvements in models are 0.2-0.5 in leadboard.

4.1. Model Ensemble

We add all the probabilities of all the models, which contains balanced finetuned resnet50, seresnext50, seresnet101, efficientnet-b3, efficientnet-b4, resnet50 with DCL finetuned, seresnext50 with DCL finetuned, efficientnet-b3 with accuracy loss, resnet50 with accuracy loss, seresnext50 with accuracy loss and resnet50 with LIO.

5. TTA

All models are evaluted with ten croped.

6. Summary of tricks

Some tricks did not work

1. Focal loss
2. Asoftmax
3. Hierarchy loss
4. BBN
5. Bitempered loss
6. Filtering duplicated labels in training set

Some tricks work, but not used due to the limited time and resources.

7. Accuracy loss
8. Pseudo label
9. Model distillation
10. DCL in efficienetnet

References

[1] Y. Chen, Y. Bai, W. Zhang, and T. Mei. Destruction and construction learning for fine-grained image recognition. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2019.

[2] M. Zhou, Y. Bai, W. Zhang, T. Zhao, and T. Mei. Look-into-object: Self-supervised structure modeling for object recognition. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2020.