Winning Model Documentation

Name: Tao Shen

Location: China

Email: taoshen.seu@gmail.com

Competition: Humpback Whale Identification

Private Leaderboard Score: 0.97208

Private Leaderboard Place: 2nd

Background:

1. What is your academic/professional background?

- Tao Shen is a computer vision engineer in ReadSense.

2. Did you have any prior experience that helped you succeed in this competition?

- I have participated in some competitions in Kaggle and some other platforms in China, and these experiences do help me.

Model Summary:

1. Dataset:

- Training set: all images including new whales.
- Validation set: randomly sampled 400 identities that has over 2 images + 200 new whales.
- Horizontal flip to create more ids -> 5004*2
- Indentity sampler: controls the id sampling in every batch

2. Model Design

- Input: 256x512 or 512*512 cropped images;
- Backbone: resnet101, seresnet101, seresnext101;
- Loss function: arcface loss + triplet loss + focal loss;
- optimizer: adam with warm up lr strategy；
- Augmentation: blur,grayscale,noise,shear,rotate,perspective transform；

3. Training

- optimizer: adam optimizer
- LR schedule: warm up lr strategy

3. Single Model performance

Single model performace

single model	privare LB
resnet101_fold0_256x512	0.9696
seresnet101_fold0_256x512	0.9691
seresnext101_fold0_256x512	0.9692
resnet101_fold0_512x512	0.9682
seresnet101_fold0_512x512	0.9664
seresnext101_fold0_512x512	-

4. Pseudo Labeling

I generate a pseudo label list containing 1.5k samples when I reached 0.940 in public LB, and I kept using this list till the competition ended. I used the bottleneck feature of the arcface model (my baseline model) to calculate cosine distance of train test images. For those few shot classes (less than 2 samples), I choose 0.65 as the threshold to filter high confidence samples. I think it will be better result using 0.970 LB model to find pseudo label.

Single model performace with pseudo labeling

single model	privare LB
resnet101_fold0_256x512	0.9705
seresnet101_fold0_256x512	0.9704
seresnext101_fold0_256x512	-

5. Model Ensemble

Weighted average is used as our ensemble strategy. The final submission is the weight average result of 10 ckpts.

Model ensemble performace

single model	privare LB
resnet101_seresnet101_seresnext101_fold0_256x512	0.97113
resnet101_seresnet101_seresnext101_fold0_512x512_pseudo	0.97072
10 models(final submisson)	0.97209

Appendix

A1. Model Execution Time

- What software did you use for training and prediction?

Pytorch 1.0.1

- What hardware?

cpu：Intel(R) Xeon(R) CPU E5-2620；

gpu： 4 RTX Titan

- How long does it take to train your model?

resnet101 with 256x512 input size : 12 hours; To get the final ensemble submission, it took about 4 days training；

- How long does it take to generate predictions using your model?

resnet101 with 256x512 input size: a few mins;

A2. References

- https://arxiv.org/abs/1503.03832 Triplet Loss
- https://arxiv.org/abs/1801.07698 Arcface Loss
- https://github.com/layumi/Person_reID_baseline_pytorch