Fine-grained Long-tailed Food Image Classification

FRANC: Food Recognition with Attention-based ensemble Network & imbalanced Cosine sampler

- R09942074 趙昱傑 R09942171 黃繼綸 R09942066 陳昱瑋 R09921A20 蔡東霖 -

DLCV final project 授課教師:王鈺強 教授

Data Preprocessing

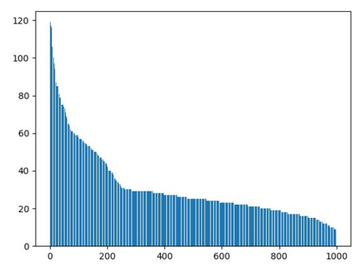
There are many factors in an imbalanced dataset that will deteriorate the performance of the model. For this reason, we adopt several techniques to alleviate the long-tailed distribution of the dataset.

Outlier Removal

In the food dataset, some frequent food categories contain 1,000+ image samples. While for some rare categories, they may contain only 1 or 2 samples. Further, according to our observation, some image samples in frequent food categories are noisy (including tableware, glasses, advertising words, etc.). Therefore, we decide to remove those noisy image samples from the dataset for model training.
For each frequent category, we first use food dataset-pretrained ResNeSt-50 to generate a vector representation for each image in the category. Next, we calculate the mean representation of these vectors, and remove those vectors that are far from the center. More specifically, we remove images with vectors that are two standard deviation from the mean vector.

Semi-supervised Learning for pseudo-labelling

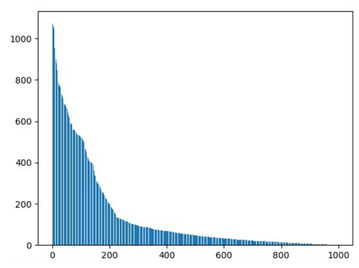
Since the lack of the rare data in the training dataset, the most direct idea is to apply psuedo label to expand the training dataset. Therefore, we first pretrain ResNest-50 on the original training dataset, and treat it as a teacher network after pretraining. Secondly, we apply our teacher network on the validation set, and calculate the probabilities of each category. We select the most significant probability as our new training data if the largest probability is greater than 0.9.



Compute Dataset Statistics

Since the food dataset is large, it is appropriate to calculate the mean and standard deviation of the dataset for data normalization.
We randomly sample 70 image samples (if ≤ 70, select all) from each category, then calculate the mean and standard deviation of each channel of the sampled images.

Finally, the mean and the standard deviation in our work are as follow:
Mean = [0.637, 0.545, 0.426]
Std = [0.280, 0.287, 0.312]

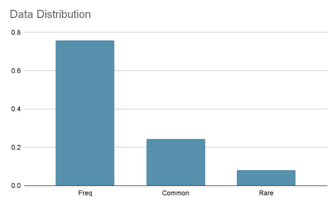


Data Sampler

Random Sampler

$$P(i) = 1$$

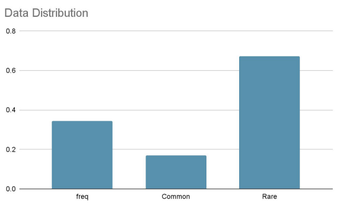
$where\ P(i)\ is\ weighting\ of\ instance\ i$



Class-balanced Sampler

$$P(i) = \frac{1}{C_i}$$

$where\ P(i)\ is\ weighting\ of\ class\ for\ instance\ i,$
$C_i\ is\ number\ of\ class\ for\ instance\ i$



Cosine-based Sampler

$$\theta = \frac{\beta}{180} \times \pi$$

$$\beta \sim U(0, 90)$$

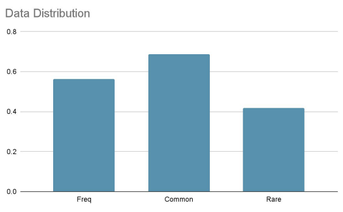
$$Cos_{comm} = 0.1 + |\cos(\theta)|$$

$$Cos_{freq} = 1.5 \times (1 + |\cos(\theta)|)$$

$$Cos_{rare} = \frac{2}{3} \times (0.1 + |\cos(\theta)|)$$

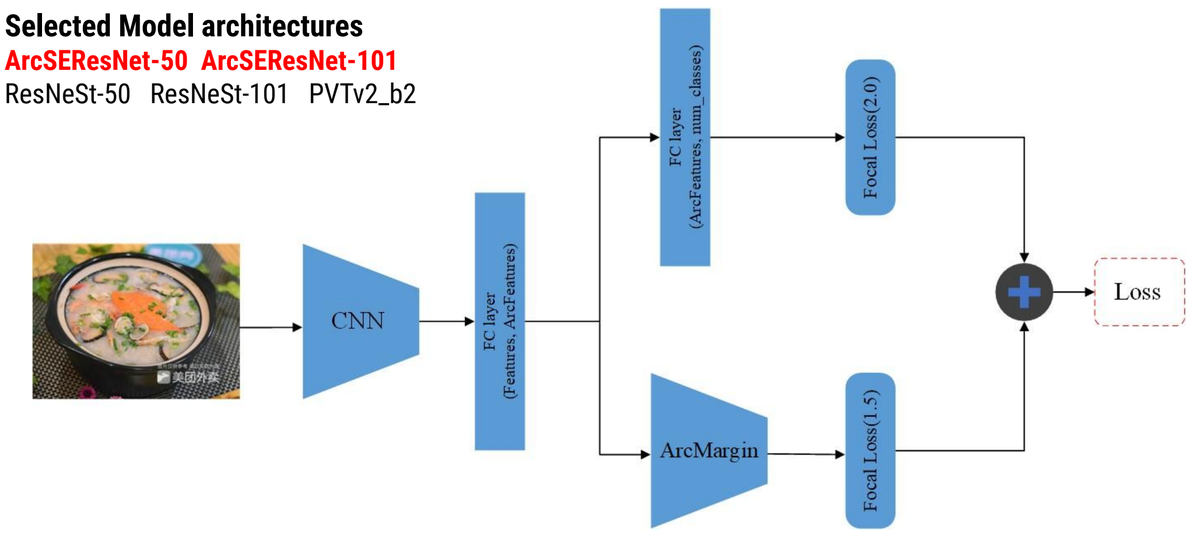
$$P(i) = \frac{1}{C_i} \times Cos_\epsilon, \epsilon = \{common, freq, rare\}$$

$where\ P(i)\ is\ weighting\ of\ instance\ i, C_i\ is\ number\ of\ class\ for\ instance\ i$
$and\ \epsilon\ is\ the\ categories\ of\ class\ for\ instance\ i$



Ensemble Framework

Selected Model architectures
ArcSEResNet-50 ArcSEResNet-101
ResNeSt-50 ResNeSt-101 PVTv2_b2



Experiments

	All	Freq.	Comm.	Rare
ArcSEResnet50 with random sampler	0.5387	0.8184	0.4625	0.0353
ArcSEResnet50 with Imbalanced sampler	0.6383	0.7565	0.6355	0.2772
ArcSEResnet50 with Imbalanced sampler and five_crop	0.6951	0.7919	0.6991	0.3489
ArcSEResnet101 with Imbalanced sampler and five_crop	0.7129	0.8083	0.7254	0.3395
Resnest50 with imbalanced sampler and five_crop	0.7243	0.7840	0.7483	**0.4205**
Resnest50 with imbalanced cosine sampler and five_crop	0.7643	0.8837	0.7663	0.3680
Ensemble model with imbalanced cosine sampler and five_crop	0.7751	0.8906	0.7853	0.3901
Ensemble model with imbalanced cosine sampler, five_crop, and test_augmentation	**0.7774**	**0.8940**	**0.7855**	0.3832

Conclusion

- ❏ For a long-tailed dataset, it is hard to train the model with the random sampler successfully. The experimental table shows the importance of a Cosine-based data sampler, which improves the classification accuracy by a considerable margin.
- ❏ From the visualization examples above, we can notice the importance of the attention mechanism, rendering the model to focus on the food instead of other confusing parts.

Future Work

- ❏ Dynamic re-weighting strategy is proved to be more effective to handle long-tail distribution, we suppose the Dynamic Curriculum sampler can be take into account in the future, which would assign lower probability after a class just sampled.
- ❏ According to the picture above, it can be found that the food is actually captured correctly, but it will still be misclassified. The main reason is that these dishes are quite similar. To our viewpoint, we think that analyzing the dishes according to the names for the classifier can further subdivide the difference between them.

Flow Chart