JSAl Cup 2018

Yk
Aizawa Laboratory
University of Tokyo
Self-introduction

• Name: 郁 (Yu) 青 (Qing)
• Hometown: Shanghai, China
• Affiliation: Aizawa Laboratory, University of Tokyo
• Grade: M1
Task

• Image Classification (55 Classes)
Dataset

• 55 classes of food ingredient images provided by cookpad
Dataset

• 55 classes of food ingredient images provided by cookpad

• Train: 11,995 images (labeled)
• Test: 3,937 images (unlabeled)
Rule

• Ensemble ✗
• K-fold ensemble ✓
• Pseudo labeling ✓
• Test time augmentation ✓
• Pretrained model ✗

• Model evaluation: Accuracy
Problem

• Pretrained model 🚫
• Small dataset

→ Overfitting
Method: Semi-supervised Learning

• Train: 11,195 images (labeled) + 3,937 images (unlabeled)
• Test: 3,937 images (unlabeled)

• Semi-supervised Learning: 
  Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results [Tarvainen, NIPS 2017]

• Implemented in PyTorch
Method

random augmentation -> labeled

CNN

back prop

softmax cross entropy loss

Groundtruth Label
Method

random augmentation

unlabeled

CNN

0.4 0.6 0.2 0.2
Method

random augmentation 1 → unlabeled → random augmentation 2

CNN

0.4 0.6 0.2 0.2

Teacher CNN

0.2 0.9 0.1 0.1
Method

random augmentation 1

unlabeled

random augmentation 2

CNN

0.4 0.6 0.2 0.2

Teacher CNN

0.2 0.9 0.1 0.1

More accurate result
Method

random augmentation1 → unlabeled → random augmentation2

CNN

Teacher CNN

MSE loss

back prop

More accurate result
Method

random augmentation1

unlabeled

random augmentation2

CNN

0.4 0.6 0.2 0.2

back prop

MSE loss

Teacher CNN

0.2 0.9 0.1 0.1

How to generate teacher CNN?

More accurate result
Method

random augmentation 1

unlabeled

random augmentation 2

CNN

0.4 0.6 0.2 0.2

same structure

Teacher CNN

0.2 0.9 0.1 0.1

More accurate result
Method

random augmentation1 → unlabeled → random augmentation2

CNN

θ → θ'

exponential moving average of past iterations

Teacher CNN

Parameters of CNN

More accurate result
Method

Train:

random augmentation1  
random augmentation2

labeled batch

back prop

CNN

0.1 0.3 0.6 0.3

back prop

MSE loss

softmax cross entropy loss

Teacher CNN

0.1 0.2 0.8 0.3

Groundtruth Label

0 1 0 0

More accurate result
Method

Train:

random augmentation 1

unlabeled batch

random augmentation 2

CNN

MSE loss

back prop

Teacher CNN

More accurate result
Method

Train:

\[ \theta_t' = \theta_{t-1}' \times \alpha + \theta_t \times (1 - \alpha) \]
\[ \alpha = 0.999 \]
Method

Test:

unlabeled

CNN

Teacher CNN

result
Data Augmentation

- Train: RandomRotation, RandomCrop, RandomHorizontalFlip, ColorJitter, RandomErasing, Mixup
  (Size 224 x 224)
Data Augmentation

- Train: RandomRotation, RandomCrop, RandomHorizontalFlip, ColorJitter, RandomErasing, Mixup
  (Size 224 x 224)

[Zhong, arXiv:1708.04896]
Data Augmentation

• Train: RandomRotation, RandomCrop, RandomHorizontalFlip, ColorJitter, RandomErasing, Mixup (Size 224 x 224)
Data Augmentation

• Train: RandomRotation, RandomCrop, RandomHorizontalFlip, ColorJitter, RandomErasing, **Mixup**
(Size 224 x 224)

0.7 * dog + 0.3 * cat

[Zhang, ICLR 2018]
Data Augmentation

• Train: RandomRotation, RandomCrop, RandomHorizontalFlip, ColorJitter, RandomErasing, Mixup
  (Size 224 x 224)

• Test: 144 Crop
  (Size: 320 x 320)
144 Crop

Sum = 4 x 3 x 6 x 2 = 144 [Szegedy, CVPR 2015]
CNN

- SE-ResNet-50 [Hu, CVPR 2018]

Squeeze-and-Excitation block
Hyperparameter

- Optimizer: SGD with Nesterov Momentum
- Learning rate: cosine annealing from 0.1
- Momentum: 0.9
- Weight decay: 1e-3
- Epoch: 300
- Batch size: 300
- Labeled batch size: 280
- Training needs 3 days on 4 Titan Xp
- Spent one month on hyperparameter tuning
Result

Score vs Date graph showing an increase in score over time from 2/7/18 to 3/28/18, peaking around 3/18 and stabilizing afterwards.
## Result

<table>
<thead>
<tr>
<th>Date</th>
<th>Score</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018/2/7</td>
<td>0.91071</td>
<td>ResNet-34 + Random Erasing</td>
</tr>
<tr>
<td>2018/2/8</td>
<td>0.92593</td>
<td>+ Mixup</td>
</tr>
<tr>
<td>2018/2/12</td>
<td>0.96093</td>
<td>+ Mean Teacher</td>
</tr>
<tr>
<td>2018/2/14</td>
<td>0.97210</td>
<td>+ Test image size 320</td>
</tr>
<tr>
<td>2018/2/15</td>
<td>0.97869</td>
<td>ResNet-34 -&gt; SE-ResNet-50</td>
</tr>
<tr>
<td>2018/2/22</td>
<td>0.97971</td>
<td>+ 144 crop</td>
</tr>
<tr>
<td>2018/3/29</td>
<td>0.98326</td>
<td>+ Hyperparameter tuning</td>
</tr>
</tbody>
</table>

Score: 0.98326  Final score: 0.98118
Summary

• Semi-supervised learning is awesome.

• Data augmentation is important.

• Hyperparameter tuning is arduous but indispensable.