Decoupling Representation and Classifier for Long-Tailed Recognition

Bingyi Kang, Saining Xie, Marcus Rohrbach, Zhicheng Yan, Albert Gordo, Jiashi Feng, Yannis Kalantidis
Long-tailed classification

Problem statement
- Training set: long-tailed distribution
  - Head v.s. Tail
- Testing set: balanced distribution
- Evaluation: three splits based on cardinality

Existing methods
- Rebalancing the data
  Up/Down sampling tail/head classes.
- Rebalancing the loss
  Assign larger/smaller weight to tail/head classes.
  e.g., CB-Focal[1], LDAM[2]

The problem behind long-tail classification performance:

- Classification performance
  - Representation Quality
  - Classifier Quality
The problem behind long-tail classification performance is influenced by representation quality and classifier quality.
The problem behind long-tail

Classification performance \equiv \text{Representation Quality} \oplus \text{Classifier Quality}

NOTE: Such observations are drawn empirically!
Notations

- Feature representation: \( f(x; \theta) = z \)
- Linear classifiers: \( g_i(z) = W_i^T z + b \)
- Final prediction: \( \hat{y} = \text{argmax} \ g_i(z) \)
What is the problem with the classifier?

- After joint training with instance-balanced sampling, the norms of the weights \( ||w_j|| \) are correlated with the size of the classes \( n_j \).
How to improve the classifier? -- Three ways

KEY: break the norm v.s. class size correlation.

I. Classifier Retraining (cRT)

- Freeze the representation.
- Retrain the linear classifier with class-balanced sampling.
How to improve the classifier? -- Three ways

**KEY:** break the norm v.s. #data correlation.

I. **Classifier Retraining (cRT)**
   - Freeze the representation.
   - Retrain the linear classifier with class-balanced sampling.

II. **Tau-Normalization (τ-Norm)**
   - Adjust the classifier weight norms directly
     \[ \tilde{w}_i = \frac{w_i}{\|w_i\|^\tau} \]
   - Tau is “temperature” of the normalization.
How to improve the classifier? -- Three ways

KEY: break the norm v.s. #data correlation.

I. Classifier Retraining (cRT)
   - Freeze the representation.
   - Retrain the linear classifier with class-balanced sampling

II. Tau-Normalization (τ-Norm)
   - Adjust the classifier weight norms directly.
     \[ \tilde{w}_i = \frac{w_i}{\|w_i\|^\tau} \]
   - Tau is “temperature” of the normalization.

III. Learnable Weight Scaling (LWS)
   - Tune the scale of each weight vector
     \[ \tilde{w}_i = f_i \times w_i, \text{ where } f_i = \frac{1}{\|w_i\|^\tau} \]
- Without classifier rebalancing (i.e. Joint training), progressively-balanced sampling works best.
- When instance-balanced sampling is used and classifiers are re-balanced, medium-shot, and few-shot performance increases significantly, and achieve best results.
How Does Classifier Rebalancing Work?

- Larger weights ==> Wider classification cone
- Un-normalized weights ==> Unbalanced decision boundaries
- Classifier rebalancing ==> More balanced decision boundaries

\[ \tilde{w}_i = \frac{w_i}{\|w_i\| \tau} \]

\( \tau \to 0 \) vs \( \tau \to 1 \)
Can we finetune both trunk and classifier?

The best performance is achieved when only classifier is retrained, and backbone model is fixed.

<table>
<thead>
<tr>
<th>Re-train</th>
<th>Many</th>
<th>Medium</th>
<th>Few</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>B+C</td>
<td>55.4</td>
<td>45.3</td>
<td>24.5</td>
<td>46.3</td>
</tr>
<tr>
<td>B+C(0.1×lr)</td>
<td><strong>61.9</strong></td>
<td>45.6</td>
<td>22.8</td>
<td>48.8</td>
</tr>
<tr>
<td>LB+C</td>
<td>61.4</td>
<td>45.8</td>
<td>24.5</td>
<td>48.9</td>
</tr>
<tr>
<td>C</td>
<td>61.5</td>
<td><strong>46.2</strong></td>
<td><strong>27.0</strong></td>
<td><strong>49.5</strong></td>
</tr>
</tbody>
</table>
Experiments

Datasets

I. ImageNet_LT
   - Constructed from ImageNet 2012
   - 1000 categories, 115.8k images

II. iNaturalist 2018
   - Contains only species.
   - 8142 categories, 437.5k images

III. Places_LT
   - Constructed fromPlaces365
   - 365 classes
## Experiments

### Datasets

I. ImageNet_LT

- Constructed from ImageNet 2012
- 1000 categories, 115.8k images

- From joint to LWS/cRT/tau-norm, with little sacrifice on many shot
- New SOTA can be achieved
- Improvement on Medium: ~10, few: 20+

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Many</th>
<th>Medium</th>
<th>Few</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLTR</td>
<td>43.2</td>
<td>35.1</td>
<td>18.5</td>
<td>35.6</td>
</tr>
<tr>
<td>OLTR(rerun)</td>
<td>40.7</td>
<td>33.3</td>
<td>18.1</td>
<td>34.1</td>
</tr>
<tr>
<td>Joint</td>
<td>65.9</td>
<td>37.5</td>
<td>7.7</td>
<td>44.4</td>
</tr>
<tr>
<td>NCM</td>
<td>56.6</td>
<td>45.3</td>
<td>28.1</td>
<td>47.3</td>
</tr>
<tr>
<td>cRT</td>
<td>61.8</td>
<td>46.2</td>
<td>27.4</td>
<td>49.6</td>
</tr>
<tr>
<td>$\tau$-normalized</td>
<td>59.1</td>
<td>46.9</td>
<td>30.7</td>
<td>49.4</td>
</tr>
<tr>
<td>LWS</td>
<td>60.2</td>
<td><strong>47.2</strong></td>
<td>30.3</td>
<td><strong>49.9</strong></td>
</tr>
</tbody>
</table>
Experiments

Datasets

II. iNaturalist 2018

- Contains only species.
- 8142 categories, 437.5k images

➢ From joint to cRT/tau-norm, little sacrifice on head classes, Large gain on tail classes.
➢ Once representation is sufficiently trained, New SOTA can be easily obtained.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Many</th>
<th>Medium</th>
<th>Few</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB-Focal</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>61.1</td>
</tr>
<tr>
<td>LDAM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>64.6</td>
</tr>
<tr>
<td>LDAM+DRAW</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>68.0</td>
</tr>
<tr>
<td>Joint</td>
<td>72.2</td>
<td>63.0/66.9</td>
<td>57.2/61.7</td>
<td>61.7/65.8</td>
</tr>
<tr>
<td>NCM</td>
<td>55.5</td>
<td>57.9/63.5</td>
<td>59.3/63.6</td>
<td>58.2/63.1</td>
</tr>
<tr>
<td>cRT</td>
<td>69.0</td>
<td>66.0/68.8</td>
<td>63.2/66.1</td>
<td>65.2/68.2</td>
</tr>
<tr>
<td>$\tau$-normalized</td>
<td>65.6</td>
<td>65.3/68.9</td>
<td>65.9/69.3</td>
<td>65.6/69.3</td>
</tr>
</tbody>
</table>

* Notation: 90 epochs/200 epochs
Take home messages

- For solving long-tailed recognition problem, representation and classifiers should be considered separately.

- Our methods achieve performance gain by finding a better tradeoff (currently the best one) between head and tail classes.

- Future research might be focusing more on improving representation quality.

https://github.com/facebookresearch/classifier-balancing