Large-Scale Long-Tailed Recognition in an Open World

Ziwei Liu* Zhongqi Miao* Xiaohang Zhan Jiayun Wang Boqing Gong Stella X. Yu

The Chinese University of Hong Kong

UC Berkeley / ICSI
Train

Cat

Fox

Panda

Test

Cat

Fox

Panda
Train

Cat (many-shot class)

Fox (medium-shot class)

Panda (few-shot class)

Test

Cat

Fox

Panda

Cat

Fox

Panda

?

(open class)
Faces [Zhang et al. 2017]

Places [Wang et al. 2017]

Species [Van Horn et al. 2019]

Actions [Zhang et al. 2019]
Open Long-Tailed Recognition

Head Classes  Tail Classes  Open Classes

Open World
Open Long-Tailed Recognition

Head Classes → Tail Classes → Open Classes

Knowledge Transfer

Open World
Open Long-Tailed Recognition

- Head Classes
- Tail Classes
- Open Classes

Knowledge Transfer

Sensitivity to Novelty
Open Long-Tailed Recognition

Avoid Forgetting

Knowledge Transfer

Head Classes

Tail Classes

Open Classes

Open World

Sensitivity to Novelty
Open Long-Tailed Recognition

Imbalanced Classification

Few-shot Learning

Head Classes

Tail Classes

Open Set Recognition

Open World

Open Classes
Imbalanced Classification
(metric learning, re-sampling, re-weighting)

Few-Shot Learning
(meta learning, classifier dynamics)

Open Set Recognition
(distribution rectification, out-of-distribution detection)

Open Long-Tailed Recognition
(dynamic meta-embedding)

Sensitivity to Novelty
Avoid Forgetting
Knowledge Transfer
Open Long-Tailed Recognition
(dynamic meta-embedding)

- Knowledge Transfer
- Sensitivity to Novelty
- Avoid Forgetting
visual memory

FLY visual memory

top-down attention

FLY

bottom-up attention

direct embedding

enhanced embedding

familiarity

FLY
The diagram illustrates the concept of visual memory and its relationship with head classes, tail classes, and open classes.

- **Head Classes** are the most familiar and easy to recognize.
- **Tail Classes** are less familiar and harder to recognize.
- **Open Classes** are entirely unknown and require top-down attention.

**Visual Memory** is involved in the recognition process, with top-down attention focusing on open classes and bottom-up attention on head classes.
Avoid Forgetting

Knowledge Transfer

Sensitivity to Novelty

Head Classes

Tail Classes

Open Classes

bottom-up attention

visual memory

familiarity

top-down attention

visual memory

open classes
bottom-up attention

original feature map

attentive feature map

visual memory

top-down attention

familiarity

Head Classes

Tail Classes

Open Classes

Tench

Hand

Fish
**Head Classes**

**Tail Classes**

**Open Classes**

- **bottom-up attention**
- **visual memory**
- **direct embedding**
- **enhanced embedding**

**top-down attention**

- **associative memory**
- **feature selection**

- **familiarity**
Head Classes

Tail Classes

visual memory

familiarity

bottom-up attention

embedding

Open Sample

Top-down attention

Tail Class ‘Buckeye’

Tail Class ‘African Grey’

Open Classes

rescaled embedding
ImageNet-LT Benchmark
Absolute Performance Gain: ~20%

Places-LT Benchmark
Absolute Performance Gain: ~10%

MS1M-LT Benchmark
Absolute Performance Gain: ~2%
## Overall F1 Score on ImageNet-LT, Places-LT and MS1M-LT Benchmarks

<table>
<thead>
<tr>
<th>Methods</th>
<th>ImageNet-LT</th>
<th>Places-LT</th>
<th>MS1M-LT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plain Model</td>
<td>0.295</td>
<td>0.366</td>
<td>0.738</td>
</tr>
<tr>
<td>Sample Re-weighting (Focal Loss)</td>
<td>0.371</td>
<td>0.453</td>
<td>-</td>
</tr>
<tr>
<td>Metric Learning (Range Loss)</td>
<td>0.373</td>
<td>0.457</td>
<td>0.722</td>
</tr>
<tr>
<td>Open Set Recognition (OpenMax)</td>
<td>0.368</td>
<td>0.458</td>
<td>-</td>
</tr>
<tr>
<td>Few-shot Learning (FSLwF)</td>
<td>0.347</td>
<td>0.375</td>
<td>-</td>
</tr>
<tr>
<td><strong>Dynamic Meta-Embedding</strong></td>
<td><strong>0.474</strong></td>
<td><strong>0.464</strong></td>
<td><strong>0.745</strong></td>
</tr>
</tbody>
</table>
Overall F1 Score on ImageNet-LT, Places-LT and MS1M-LT Benchmarks

<table>
<thead>
<tr>
<th>Methods</th>
<th>ImageNet-LT</th>
<th>Places-LT</th>
<th>MS1M-LT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plain Model</td>
<td>0.295</td>
<td>0.366</td>
<td>0.738</td>
</tr>
<tr>
<td>Sample Re-weighting (Focal Loss)</td>
<td>0.371</td>
<td>0.453</td>
<td>-</td>
</tr>
<tr>
<td>Metric Learning (Range Loss)</td>
<td>0.373</td>
<td>0.457</td>
<td>0.722</td>
</tr>
<tr>
<td>Open Set Recognition (OpenMax)</td>
<td>0.368</td>
<td>0.458</td>
<td>-</td>
</tr>
<tr>
<td>Few-shot Learning (FSLwF)</td>
<td>0.347</td>
<td>0.375</td>
<td>-</td>
</tr>
<tr>
<td><strong>Dynamic Meta-Embedding</strong></td>
<td><strong>0.474</strong></td>
<td><strong>0.464</strong></td>
<td><strong>0.745</strong></td>
</tr>
</tbody>
</table>
Overall *F1 Score* on ImageNet-LT, Places-LT and MS1M-LT Benchmarks

<table>
<thead>
<tr>
<th>Methods</th>
<th>ImageNet-LT</th>
<th>Places-LT</th>
<th>MS1M-LT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plain Model</td>
<td>0.295</td>
<td>0.366</td>
<td>0.738</td>
</tr>
<tr>
<td>Sample Re-weighting (Focal Loss)</td>
<td>0.371</td>
<td>0.453</td>
<td>-</td>
</tr>
<tr>
<td>Metric Learning (Range Loss)</td>
<td>0.373</td>
<td>0.457</td>
<td>0.722</td>
</tr>
<tr>
<td>Open Set Recognition (OpenMax)</td>
<td>0.368</td>
<td>0.458</td>
<td>-</td>
</tr>
<tr>
<td>Few-shot Learning (FSLwF)</td>
<td>0.347</td>
<td>0.375</td>
<td>-</td>
</tr>
<tr>
<td><strong>Dynamic Meta-Embedding</strong></td>
<td><strong>0.474</strong></td>
<td><strong>0.464</strong></td>
<td><strong>0.745</strong></td>
</tr>
</tbody>
</table>
# Overall F1 Score on ImageNet-LT, Places-LT and MS1M-LT Benchmarks

<table>
<thead>
<tr>
<th>Methods</th>
<th>ImageNet-LT</th>
<th>Places-LT</th>
<th>MS1M-LT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plain Model</td>
<td>0.295</td>
<td>0.366</td>
<td>0.738</td>
</tr>
<tr>
<td>Sample Re-weighting (Focal Loss)</td>
<td>0.371</td>
<td>0.453</td>
<td>-</td>
</tr>
<tr>
<td>Metric Learning (Range Loss)</td>
<td>0.373</td>
<td>0.457</td>
<td>0.722</td>
</tr>
<tr>
<td>Open Set Recognition (OpenMax)</td>
<td>0.368</td>
<td>0.458</td>
<td>-</td>
</tr>
<tr>
<td>Few-shot Learning (FSLwF)</td>
<td>0.347</td>
<td>0.375</td>
<td>-</td>
</tr>
<tr>
<td><strong>Dynamic Meta-Embedding</strong></td>
<td><strong>0.474</strong></td>
<td><strong>0.464</strong></td>
<td><strong>0.745</strong></td>
</tr>
</tbody>
</table>
### Overall F1 Score on ImageNet-LT, Places-LT and MS1M-LT Benchmarks

<table>
<thead>
<tr>
<th>Methods</th>
<th>ImageNet-LT</th>
<th>Places-LT</th>
<th>MS1M-LT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plain Model</td>
<td>0.295</td>
<td>0.366</td>
<td>0.738</td>
</tr>
<tr>
<td>Sample Re-weighting (Focal Loss)</td>
<td>0.371</td>
<td>0.453</td>
<td>-</td>
</tr>
<tr>
<td>Metric Learning (Range Loss)</td>
<td>0.373</td>
<td>0.457</td>
<td>0.722</td>
</tr>
<tr>
<td>Open Set Recognition (OpenMax)</td>
<td>0.368</td>
<td>0.458</td>
<td>-</td>
</tr>
<tr>
<td>Few-shot Learning (FSLwF)</td>
<td>0.347</td>
<td>0.375</td>
<td>-</td>
</tr>
<tr>
<td><strong>Dynamic Meta-Embedding</strong></td>
<td><strong>0.474</strong></td>
<td><strong>0.464</strong></td>
<td><strong>0.745</strong></td>
</tr>
</tbody>
</table>
Few shot
New Task
Open Long-Tailed Recognition (OLTR)

New Approach
Dynamic Meta-Embedding

New Benchmarks
ImageNet-LT  Places-LT  MS1M-LT
Thanks!

Code, models and benchmarks are available at

Project Page: https://liuziwei7.github.io/projects/LongTail.html