Large-scale Scene Understanding
Yinda Zhang, Fisher Yu, Shuran Song, Pingmei Xu, Ari Seff, Jianxiong Xiao
Duplicate Removal

- Training Set (Million)
- Testing Set (Thousand)

- Within Training
- Within Testing
- Between Training & Testing
Duplicate Removal

Within Training

Reflect internet image distribution. Should be maintained.

Within Testing

Redundant images must be removed.

Between Training & Testing

Must be removed, otherwise causes cheating.
Types of Duplication

Identical

Flip
Types of Duplication

- Slightly different color tone
- Slight misalignment
- Tiny different text change
Major Challenge

1. Pixel-wise difference: too sensitive.
2. Gist feature: lose details.
3. Deep learning feature: class relevant
Pipeline

1. Fc7 layer of PlaceNet (AlexNet on Places), compute correlation distance
2. Resize image to 100x100, gray-scale, compute intensity difference.
Computational Cost

1. Within testing: $1500 \times 1500$
   
   Exhaustive Search

2. Between training&testing: $1500 \times 1,500,000$
   
   Dimension reduction + Approximate NN
## Challenge Result

<table>
<thead>
<tr>
<th>Team Name</th>
<th>Top 1 Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google</td>
<td>0.9120</td>
</tr>
<tr>
<td>SIAT_MMLAB</td>
<td>0.9030</td>
</tr>
<tr>
<td>Contextual CNN</td>
<td>0.8821</td>
</tr>
<tr>
<td>isia_ICT</td>
<td>0.8799</td>
</tr>
<tr>
<td>Baseline*</td>
<td>0.8204</td>
</tr>
</tbody>
</table>

*Baseline: AlexNet trained with 500K per category.