Learning Deep Convolutional Neural Networks for Places2 Scene Recognition

Li Shen
li.shen@vipl.ict.ac.cn
University of Chinese Academy of Sciences

Zhouchen Lin
zlin@pku.edu.cn
Peking University

WM Team
Summary of Our Submissions

• 1st place in Places2 Scene Classification Challenge with provided training data
Key Components

• Optimization: Relay Back-Propagation

• Network Architectures

• Class-aware Sampling
Motivation

• “Going deeper” is promising to improve the accuracy

• Difficulty: The improvement on accuracy cannot be trivially achieved by simply increasing the depth of network.

<table>
<thead>
<tr>
<th>Depth</th>
<th>19</th>
<th>22</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>top-5 err. (%)</td>
<td>18.93</td>
<td>19.00</td>
<td>19.21</td>
</tr>
</tbody>
</table>
Why this phenomenon happens?

• Gradient vanishing / exploding?

➢ Using refined initialization [1], Batch Normalization [2] etc. has greatly reduced the risk of this issue.


Insight

• Although the gradient does not vanish, if we view the BP as an information propagation process, then by information theory, e.g., the Data Processing Theorem, the amount of information still diminishes.
Relay Back-Propagation
## Network Architectures

<table>
<thead>
<tr>
<th>input size</th>
<th>model A</th>
<th>model B</th>
</tr>
</thead>
<tbody>
<tr>
<td>224×224</td>
<td>[ 3×3, 64 ]×2 maxpool 2×2, 2</td>
<td>[ 7×7, 128, stride 2 ]×1</td>
</tr>
<tr>
<td>112×112</td>
<td>[ 3×3, 128 ]×2 maxpool 2×2, 2</td>
<td>maxpool 2×2, 2</td>
</tr>
<tr>
<td>56×56</td>
<td>[ 3×3, 256 ]×5 maxpool 2×2, 2</td>
<td>[ 1×1, 64; 3×3, 64; dbl 3×3, 128 ]×4 maxpool 2×2, 2</td>
</tr>
<tr>
<td>28×28</td>
<td>[ 3×3, 512 ]×5 maxpool 2×2, 2</td>
<td>[ 1×1, 128; 3×3, 128; dbl 3×3, 256 ]×4 maxpool 2×2, 2</td>
</tr>
<tr>
<td>14×14</td>
<td>[ 3×3, 512 ]×5 spp, {7, 3, 2, 1}</td>
<td>[ 1×1, 128; 3×3, 128; dbl 3×3, 256 ]×4 spp, {7, 3, 2, 1}</td>
</tr>
<tr>
<td>1×1</td>
<td>fc, 4096</td>
<td>fc, 4096</td>
</tr>
<tr>
<td>1×1</td>
<td>fc, 401, softmax</td>
<td></td>
</tr>
</tbody>
</table>
Class-aware Sampling

• Training data in Places2 dataset
  ➢ **large scale**: 8 million in total
  ➢ **non-uniform class distribution**: between 4,000 and 30,000 per class
Class-aware Sampling

Class list & 401 class-specific image lists

Training batch

Class A

Class B

Class C

~0.6% improvement
Class-aware Sampling

Class list & 401 class-specific image lists

~0.6% improvement
Error Rates (%) on Validation Set

Our model ensemble achieves 47.21% top-1 error and 15.74% top-5 error. In the brackets are the improvements over the baseline.

<table>
<thead>
<tr>
<th>Method</th>
<th>Testing Method</th>
<th>model A</th>
<th>model B</th>
</tr>
</thead>
<tbody>
<tr>
<td>loss 1 + BP (baseline)</td>
<td>center crop</td>
<td>50.91</td>
<td>50.62</td>
</tr>
<tr>
<td>loss 1&amp;2 + BP [3]</td>
<td></td>
<td>50.72(0.19)</td>
<td>50.59(0.03)</td>
</tr>
<tr>
<td>loss 1&amp;2 + Relay BP</td>
<td></td>
<td>49.75(1.16)</td>
<td>49.77(0.85)</td>
</tr>
<tr>
<td>loss 1 + BP (baseline)</td>
<td>single model</td>
<td>48.67</td>
<td>48.29</td>
</tr>
<tr>
<td>loss 1&amp;2 + BP [3]</td>
<td></td>
<td>48.55(0.12)</td>
<td>48.27(0.02)</td>
</tr>
<tr>
<td>loss 1&amp;2 + Relay BP</td>
<td></td>
<td>47.86(0.81)</td>
<td>47.72(0.57)</td>
</tr>
</tbody>
</table>

Input image size: 256×N        Crop size: 224×224        Single model: multi-view, multi-scale (256×N, 320×N, etc.)

Our team “WM” won the 1st place in the Places2 Scene Classification Challenge, and our five submissions won the top five places.

<table>
<thead>
<tr>
<th>Team name</th>
<th>top-5 err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>WM (model ensemble)</td>
<td>16.87</td>
</tr>
<tr>
<td>WM (model B)</td>
<td>17.28</td>
</tr>
<tr>
<td>WM (model A)</td>
<td>17.35</td>
</tr>
<tr>
<td>SIAT_MMLAB</td>
<td>17.36</td>
</tr>
<tr>
<td>Qualcomm Research</td>
<td>17.59</td>
</tr>
<tr>
<td>Trimps-Soußen</td>
<td>17.98</td>
</tr>
<tr>
<td>Ntu_rose</td>
<td>19.33</td>
</tr>
</tbody>
</table>
Successfully Classified Examples

1. art studio
2. art gallery
3. artists loft
4. art school
5. museum

1. sushi bar
2. restaurant kitchen
3. delicatessen
4. bakery shop
5. pantry

1. amusement park
2. carrousel
3. amusement arcade
4. water park
5. temple

1. oilrig
2. islet
3. ocean
4. coast
5. beach
Incorrectly Classified Examples

1. hotel room
2. bedroom
3. bedchamber
4. television room
5. balcony interior

GT: pub indoor

1. lift bridge
2. tower
3. bridge
4. viaduct
5. river

GT: skyscraper

1. aqueduct
2. viaduct
3. bridge
4. arch
5. hot spring

GT: waterfall block

1. corridor
2. hallway
3. elevator lobby
4. lobby
5. reception

GT: entrance hall
Future Work

• Theoretical support for Relay BP
• Exploration of Relay BP with other techniques (e.g., skip connections)
• Multiple Relay BP

Details and more experimental evaluation will be described in our arXiv paper (stay tuned).

Acknowledgement

Thank Yadang Chen and Chunyan Hao from University of Macau for helping us download dataset.
Welcome to our poster for further discussion!

Thank you!