Learning Deep Convolutional Neural Networks for Places2 Scene Recognition

WM Team

Li Shen

li.shen@vipl.ict.ac.cn

University of Chinese Academy of Sciences

Zhouchen Lin

zlin@pku.edu.cn

Peking University

Summary of Our Submissions

- 1st place in Places2 Scene Classification Challenge with provided training data

Team name	Entry description	Classification error
WM	Fusion with product strategy	0.168715
WM	Fusion with learnt weights	0.168747
WM	Fusion with average strategy	0.168909
WM	A single model (model B)	0.172876
WM	A single model (model A)	0.173527

Key Components

- Optimization: **Relay Back-Propagation**
- Network Architectures
- Class-aware Sampling

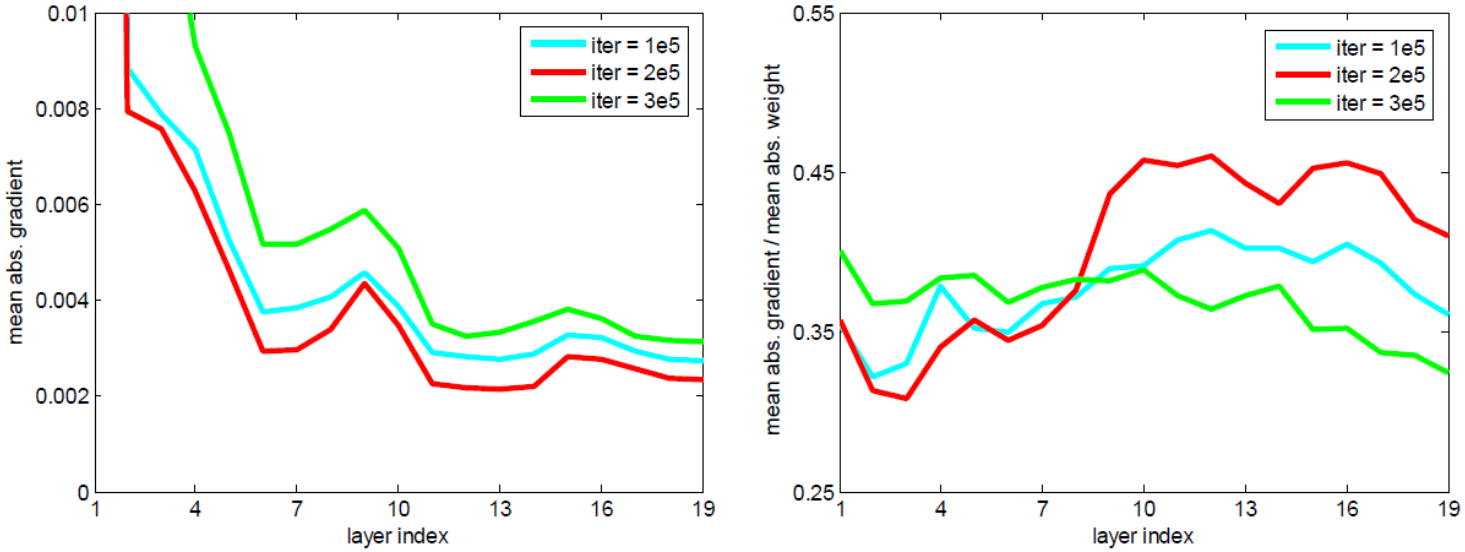
Motivation

- “Going deeper” is promising to improve the accuracy
- Difficulty: The improvement on accuracy cannot be trivially achieved by simply increasing the depth of network.

Depth	19	22	25
top-5 err. (%)	18.93	19.00	19.21

Why this phenomenon happens?

- Gradient vanishing / exploding?
  - ➢Using refined initialization [1], Batch Normalization [2] etc. has greatly reduced the risk of this issue.

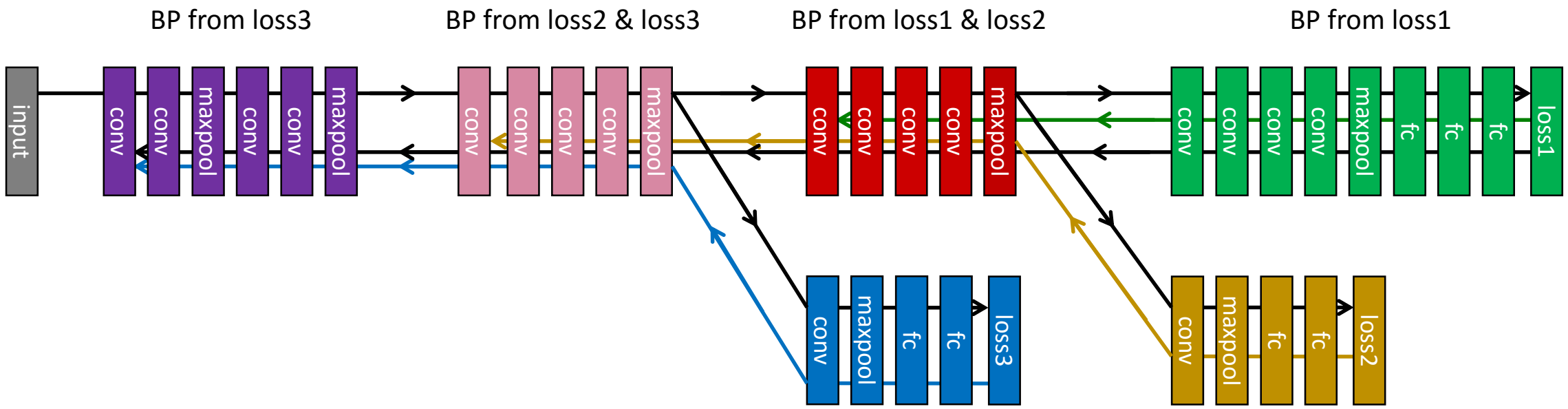


[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun. *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification*. In ICCV 2015.
[2] Sergey Ioffe and Christian Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. In ICML 2015.

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

input size	model A	model B
224×224	[3×3, 64]×2 maxpool 2×2, 2	[7×7, 128, stride 2]×1
112×112	[3×3, 128]×2 maxpool 2×2, 2	maxpool 2×2, 2
56×56	[3×3, 256]×5 maxpool 2×2, 2	[1×1, 64; 3×3, 64; dbl 3×3, 128]×4 maxpool 2×2, 2
28×28	[3×3, 512]×5 maxpool 2×2, 2	[1×1, 128; 3×3, 128; dbl 3×3, 256]×4 maxpool 2×2, 2
	branch	branch
14×14	[3×3, 512]×5 spp, {7, 3, 2, 1}	[1×1, 128; 3×3, 128; dbl 3×3, 256]×4 spp, {7, 3, 2, 1}
-	fc, 4096	
-	fc, 4096	
-	fc, 401, softmax	

Interim loss2

Propagation path of loss2

Propagation path of loss1

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

~0.6% improvement

Class A

Class B

Class C

Class-aware Sampling

Class list & 401 class-specific image lists

Training batch



~0.6% improvement

Class B



Class C








Class A





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.

Method	Testing Method	model A		model B	
		top-1 err.	top-5 err.	top-1 err.	top-5 err.
loss1 + BP (baseline)	center crop	50.91	19.00	50.62	18.69
loss1&2 + BP [3]		$50.72_{(0.19)}$	$18.84_{(0.18)}$	$50.59_{(0.03)}$	$18.68_{(0.01)}$
loss1&2 + Relay BP		$49.75_{(1.16)}$	$17.83_{(1.17)}$	$49.77_{(0.85)}$	$17.86_{(0.83)}$
loss1 + BP (baseline)	single model	48.67	17.19	48.29	16.89
loss1&2 + BP [3]		$48.55_{(0.12)}$	$17.05_{(0.14)}$	$48.27_{(0.02)}$	$16.89_{(0.00)}$
loss1&2 + Relay BP		$47.86_{(0.81)}$	$16.33_{(0.86)}$	$47.72_{(0.57)}$	$16.36_{(0.53)}$

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

[3] Chen-Yu Lee, Saining Xie, Patrick Gallagher, Zhengyou Zhang and Zhuowen Tu. *Deeply-Supervised Nets.* In Proceedings of AISTATS 2015.

Error Rates (%) on Test Set

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

Team name	top-5 err.
WM (model ensemble)	16.87
WM (model B)	17.28
WM (model A)	17.35
SIAT_MMLAB	17.36
Qualcomm Research	17.59
Trimps-Soushen	17.98
Ntu_rose	19.33

Successfully Classified Examples

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

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

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

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

GT: waterfall block

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

GT: skyscraper

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 technique (e.g., skip connections)

Details and more experimental evaluation will be described in our arXiv paper.

Thank you !