Enemble Max-Pooling

Is Only the Maximum Activation Useful When Pooling?

Hao Zhang and **Jianxin Wu**

{zhangh, wujx}@lamda.nju.edu.cn

LAMDA Group
National Key Laboratory for Novel Software Technology
Department of Computer Science and Technology

Nanjing University

November 24, 2017

Table of Contents

Table of Contents

Hao Zhang and **Jianxin Wu**
{zhangh, wujx}@lamda.nju

NANJING UNIVERSITY
LAMDA Group
Nanjing University

Image Classification

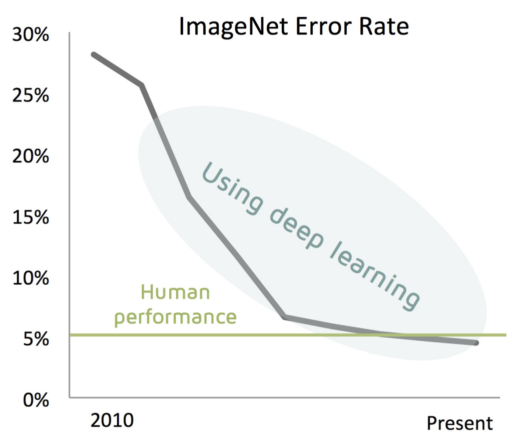

Figure: ImageNet error rate.

[N Rao. Deep learning and the need for unified tools.]

Models are Getting Larger

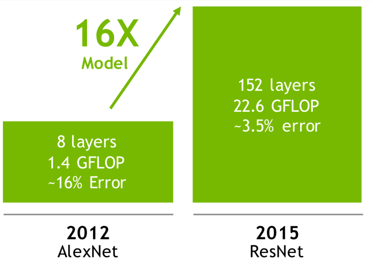

Figure: Models are getting larger.

[W Dally. Efficient methods for deep neural networks. NIPS'16 Workshop.]

Table: Model error and its training time.

Model	Error	Time
ResNet-18	10.76%	2.5 d
ResNet-50	7.02%	5 d
ResNet-101	6.21%	1 w
ResNet-152	6.16%	1.5 w

[S Han. Benchmark with fb.resnet.torch using four M40 GPUs.]

Hao Zhang and **Jianxin Wu**
{zhangh, wujx}@lamda.nju

Other Challenges with Large CNN Model

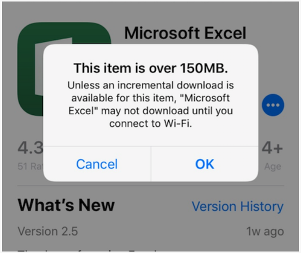

Figure: Mobile solution.

[Jane. How to Download Over 150MB Apps on iPhone via Cellular.]

Figure: AlphaGo costs $3000 electric bill per game.

[S Shead. Here's how much computing power Google DeepMind needed to beat Lee Sedol at Go.]

Motivation

Large model

- Pros.: The deeper, the better.
- Cons.: Higher resource costs, such as computing time, storage space, money, etc.

Motivation

- Similar model complexity, better performance.

Idea

- Revealing treasures from convolutional activations.

Hao Zhang and **Jianxin Wu**
{zhangh, wujx}@lamda.nju

LAMDA Group
Nanjing University

Table of Contents

Max-pooling

Figure: Max-pooling.

[D Scherer et al. Evaluation of pooling operations in convolutional architectures for object recognition. ICANN'10.]

Open problem:

Is only the maximum activation useful when max-pooling?

Related Works

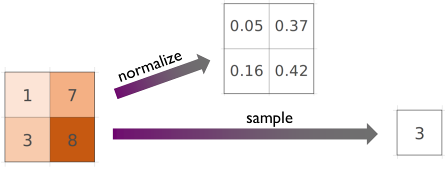

Figure: Stochastic pooling.

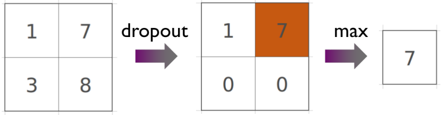

Figure: Max-pooling dropout/Stochastic max-pooling.

[M Zeiler & R Fergus. Stochastic pooling for regularization of deep convolutional neural networks. ICLR'13.]
[H Wu & X Gu. Max-pooling dropout for regularization of convolutional neural networks. NIPS'15.]
[Y Huang et al. Channel-max, channel-drop and stochastic max-pooling. CVPR'15 Workshop.]

Hao Zhang and **Jianxin Wu**
{zhangh, wujx}@lamda.nju

Table of Contents

Hao Zhang and **Jianxin Wu**
`{zhangh, wujx}@lamda.nju`

Ensemble Max-pooling (EMP)

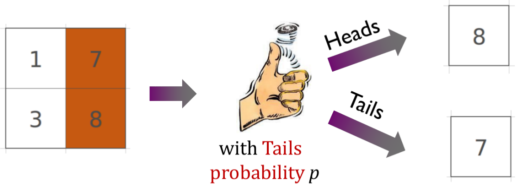

Figure: Training phase of EMP.

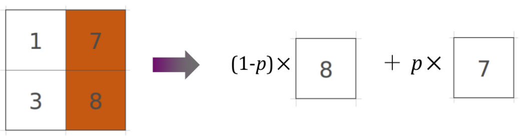

Figure: Testing phase of EMP.

Table of Contents

Comparisons with Related Works

Model	*Error Rate*	***Relative** Improve Compared to Max-pooling*
Max-pooling	15.25%	0.00%
Average Pooling	16.20%	-6.23%
Ensemble Max-pooling (p=0.05)	15.51%	-1.70%
Ensemble Max-pooling (p=0.1)	15.14%	0.72%
Ensemble Max-pooling (p=0.2)	15.07%	1.18%
Ensemble Max-pooling (p=0.3)	14.28%	6.36%
Ensemble Max-pooling (p=0.4)	**14.27%**	**6.43%**
Ensemble Max-pooling (p=0.5)	14.62%	4.13%
Ensemble Max-pooling (p=0.7)	14.89%	2.36%
Stochastic Pooling	15.81%	-3.67%
Max-pooling Dropout/Stochastic Max-pooing (p=0.05)	15.22%	0.20%
Max-pooling Dropout/Stochastic Max-pooing (p=0.1)	15.17%	0.52%
Max-pooling Dropout/Stochastic Max-pooing (p=0.2)	14.97%	1.84%
Max-pooling Dropout/Stochastic Max-pooing (p=0.3)	16.59%	-6.96%
Max-pooling Dropout/Stochastic Max-pooing (p=0.4)	25.93%	-67.18%
Max-pooling Dropout/Stochastic Max-pooing (p=0.5)	Not Converge	-
Max-pooling Dropout/Stochastic Max-pooing (p=0.7)	Not Converge	-

Figure: Comparisons with related works on CIFAR-10.

[A Krizhevsky, G Hinton. Learning multiple layers of features from tiny images. UToronto technical

Deeply Fused Networks-Merge and Run (DFN-MR)

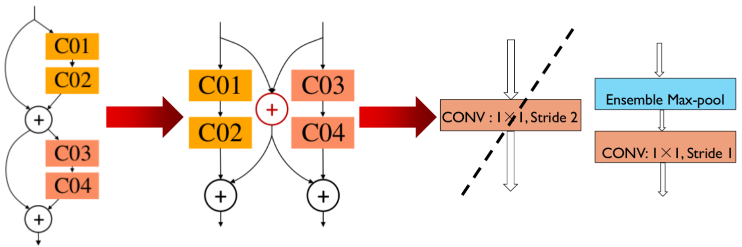

Figure: ResNet, DFN-MR, and DFN-MR (Ensemble Max-Pooling).

[K He et al. Deep residual learning for image recognition. CVPR'16.]
[L Zhao et al. On the connection of deep fusion to ensembling. arXiv'16.]

Towards State-of-the-art

Model	*Layers*	*Error Rate*	***Relative*** *Improve Compared to DFN-MR*
DFN-MR	56	5.50%	0.00%
DFN-MR (Classical Max-pooling)	56	5.44%	1.14%
DFN-MR (Ensemble Max-pooling, p=0.05)	56	5.13%	6.82%
DFN-MR (Ensemble Max-pooling, p=0.1)	56	**4.62%**	**16.00%**
DFN-MR (Ensemble Max-pooling, p=0.2)	56	5.06%	7.95%
DFN-MR (Ensemble Max-pooling, p=0.5)	56	4.69%	14.77%
ResNet[4]	110	6.61%	-20.19%
ResNet[16]	110	6.41%	-16.54%
ResNet (Pre-activation)[9]	164	5.46%	0.72%
ResNet (Pre-activation)[9]	1001	4.62%	16.00%
ResNet (Stochastic Depth)[16]	110	5.23%	4.91%
ResNet (Stochastic Depth)[16]	1202	4.91%	10.73%

Figure: Comparisons with state-of-the-art.

Experiments on CIFAR-100 and ImageNet can be referred in:

[H Zhang & J Wu. Ensemble Max-Pooling: Is Only the Maximum Activation Useful When Pooling? CCFAI'17.]

[H Zhang & J Wu. Ensemble Max-Pooling: Is Only the Maximum Activation Useful When Pooling? JUST'17.]

Thanks

Questions Please!

Hao Zhang and **Jianxin Wu**
{zhangh, wujx}@lamda.nju.edu.cn