Ensemble Max-Pooling:
Is Only the Maximum Activation Useful When Pooling?

Hao Zhang and **Jianxin Wu**

National Key Laboratory for Novel Software Technology, Nanjing University

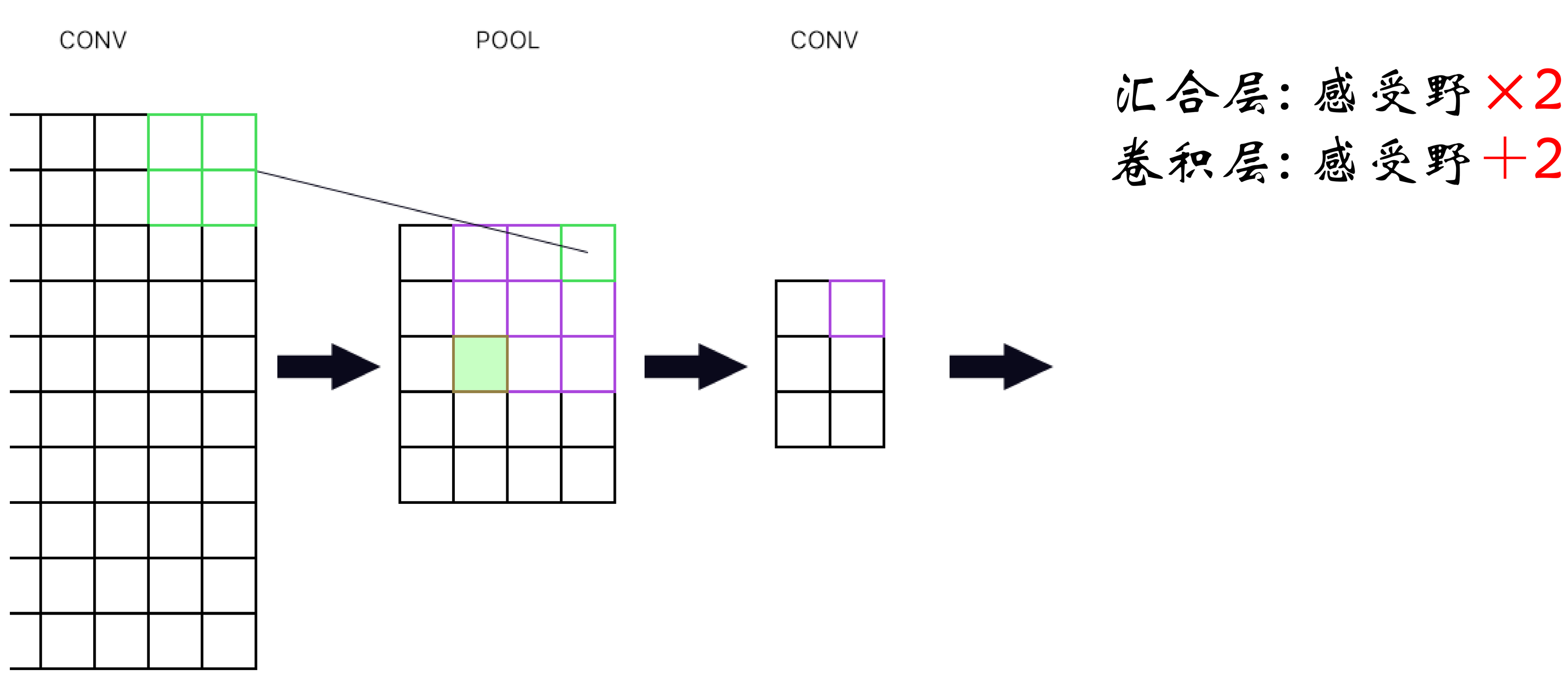
{zhangh, wujx}@lamda.nju.edu.cn

Learning And Mining from DatA
http://lamda.nju.edu.cn

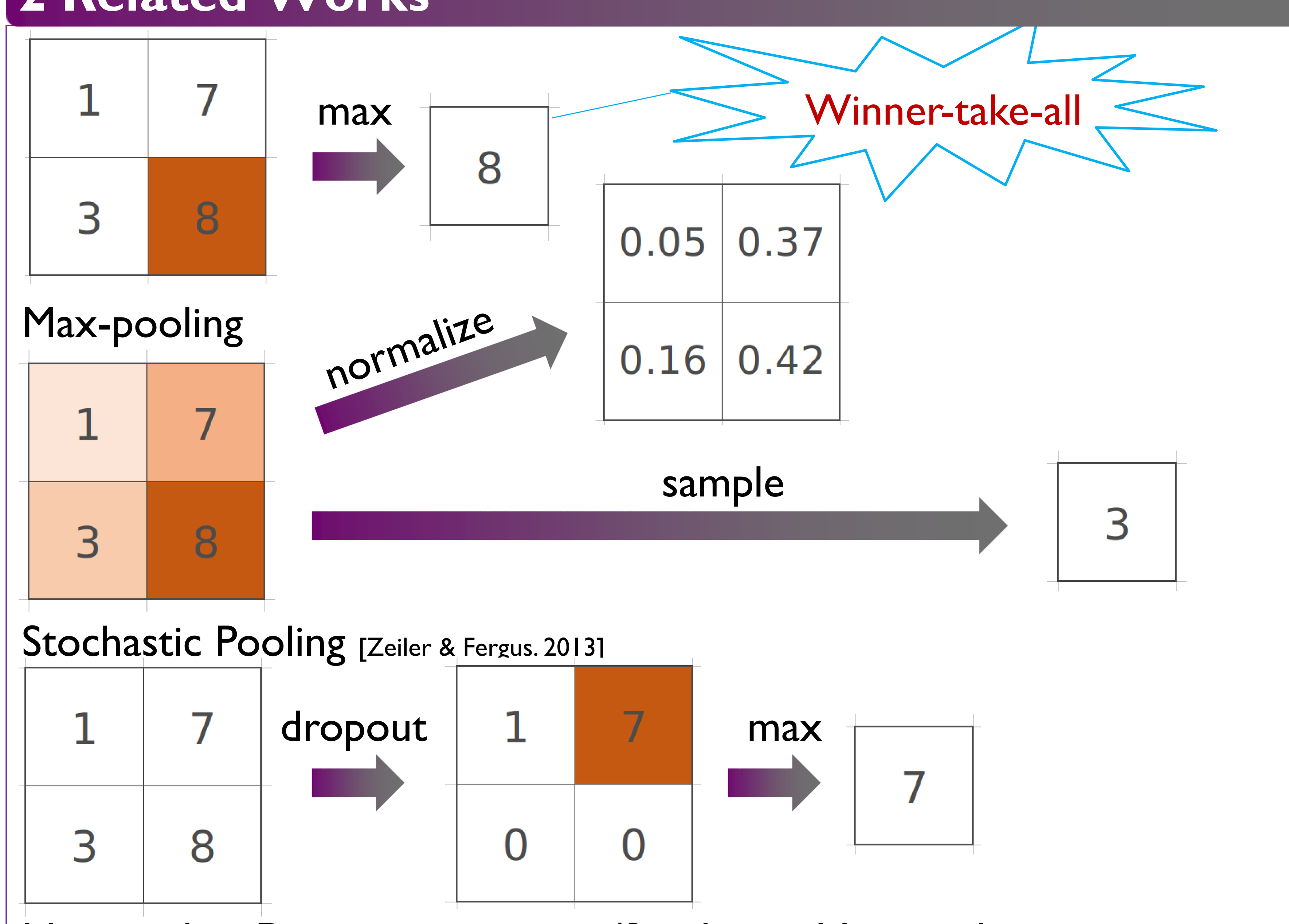
1 Introduction

汇合层(pooling layer)的作用

- 减少特征映射(feature map)的空间大小，降低对局部敏感程度
- 有效提高感受野(receptive field)



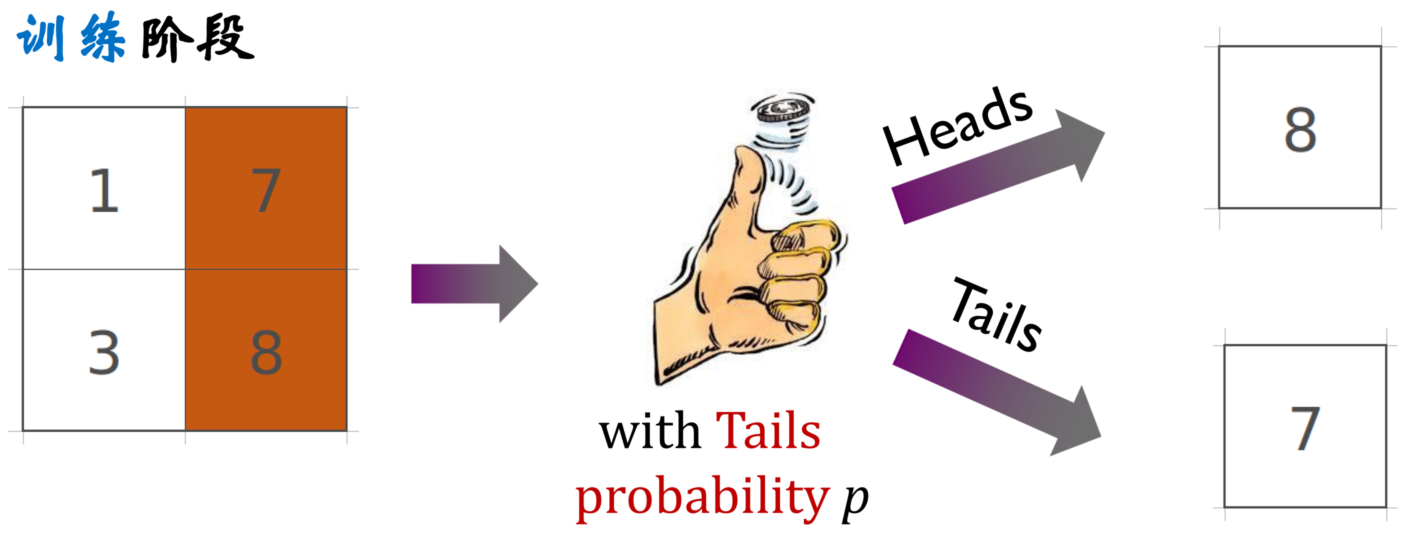
2 Related Works



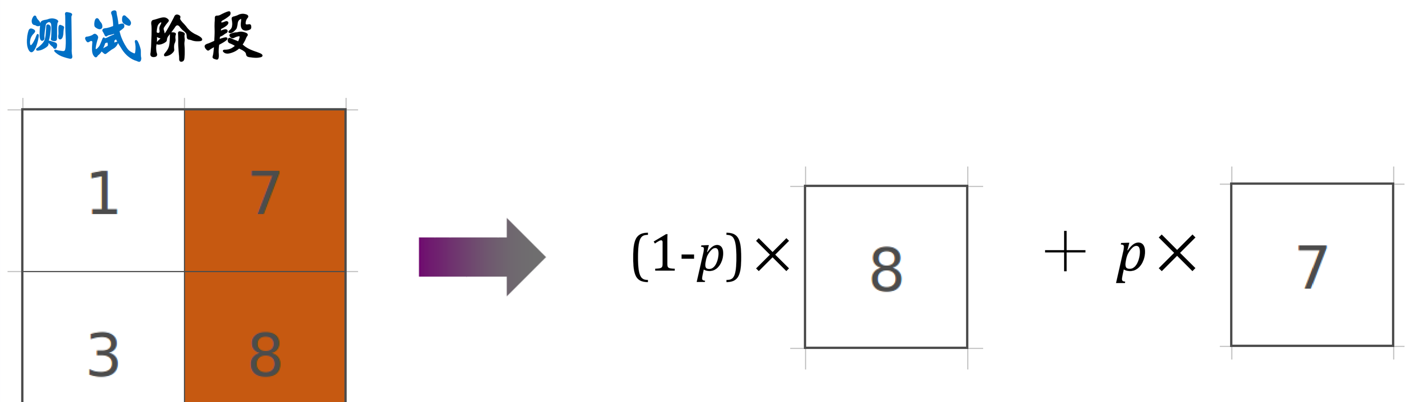
Max-pooling Dropout [Wu & Gu. 2015]/Stochastic Max-pooling [Huang et al. 2015]

3 Proposed Method: Ensemble Max-pooling

训练阶段



测试阶段



作用

- 网络集成：指数级多的基础潜在网络集成
- 数据扩充(data augmentation)：作用在中间层

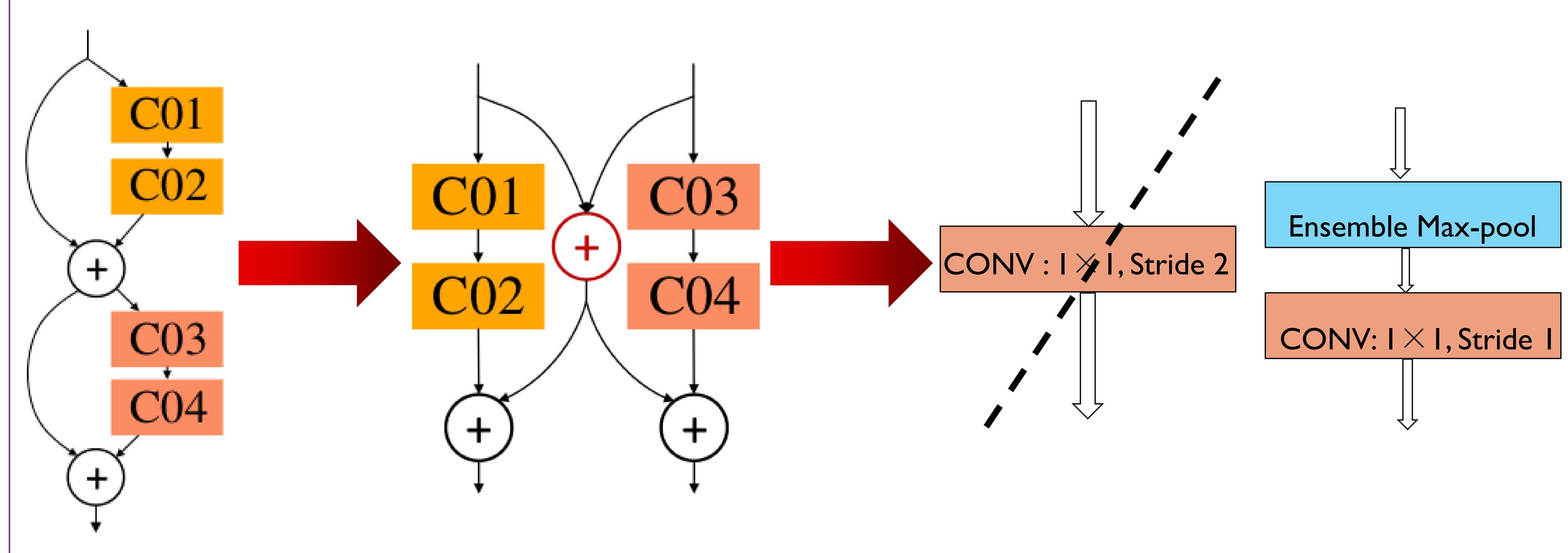
实现

- 容易实现，修改汇合源代码得到
- 可直接应用于现有网络结构

4 Comparisons with Related Works: CIFAR-10 Results

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	-

5 Towards State-of-the-art: CIFAR-10 Results



ResNet [He et al. 2016]
- 许多基础潜在网络集成

DFN-MR [Zhao et al. 2016]
- 足够多的基础潜在网络
- 避免极深网络
- 比ResNet更好的效果

DFN-MR (Ensemble Max-pooling)
- 替换DFN-MR中步长(stride)为2的卷积层

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%

6 ImageNet Results

Model	*Error Rate (Top-1/Top-5)*	***Relative** Improve Compared to Pre-trained Model (Top-1/Top-5*
CaffeNet Pre-trained Model	43.10%/19.97%	0.00%/0.00%
Ensemble Max-pooling (p=0.05)	42.69%/19.53%	0.94%/2.20%
Ensemble Max-pooling (p=0.1)	**42.37%/19.47%**	**1.68%/2.50%**
Ensemble Max-pooling (p=0.15)	42.61%/19.52%	1.13%/2.53%
Ensemble Max-pooling (p=0.3)	42.81%/19.69%	0.65%/1.41%
Ensemble Max-pooling (p=0.5)	42.71%/19.67%	0.89%/1.51%
Ensemble Max-pooling (p=0.7)	42.85%/19.81%	0.58%/0.80%
CaffeNet Pre-trained Model	43.10%/19.97%	0.00%/0.00%
Ensemble Max-pooling (p=0.05)	42.11%/19.19%	2.30%/3.93%
Ensemble Max-pooling (p=0.1)	**41.97%/19.14%**	**2.60%/4.18%**
Ensemble Max-pooling (p=0.15)	42.05%/19.23%	2.43%/3.70%
Ensemble Max-pooling (p=0.3)	42.22%/19.25%	2.03%/3.61%
Ensemble Max-pooling (p=0.5)	42.25%/19.33%	1.95%/3.21%
Ensemble Max-pooling (p=0.7)	42.28%/19.40%	1.88%/2.87%

Tain from Scratch

Fine-tuning