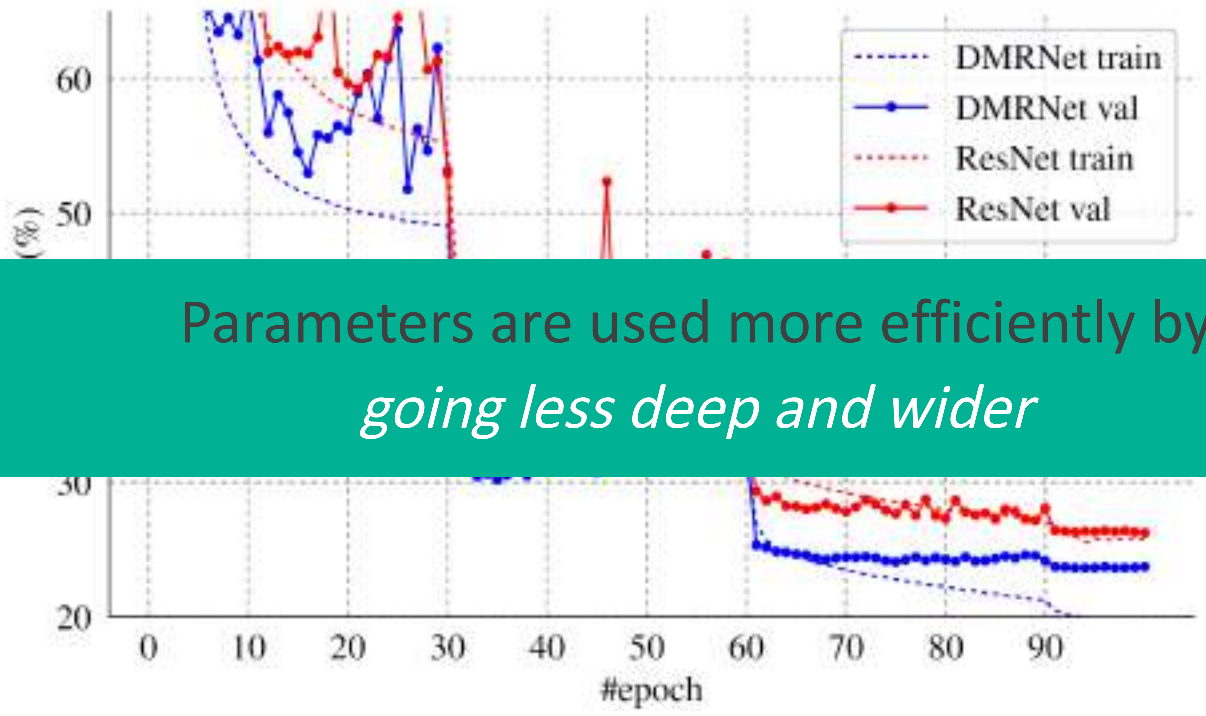
Comparison with ResNets

#parameters	L	ResNets	DMRNets
0.4M	12	1.90	2.00
0.6M	18	1.97	1.87
0.8M	24	1.93	1.86
1.0M	30	1.89	1.81
1.2M	36	1.90	1.77
1.5M	48	1.91	1.84
1.7M	54	2.00	1.68
3.1M	96	1.85	1.70

SVHN classification error, average over 5 runs

ImageNet classification

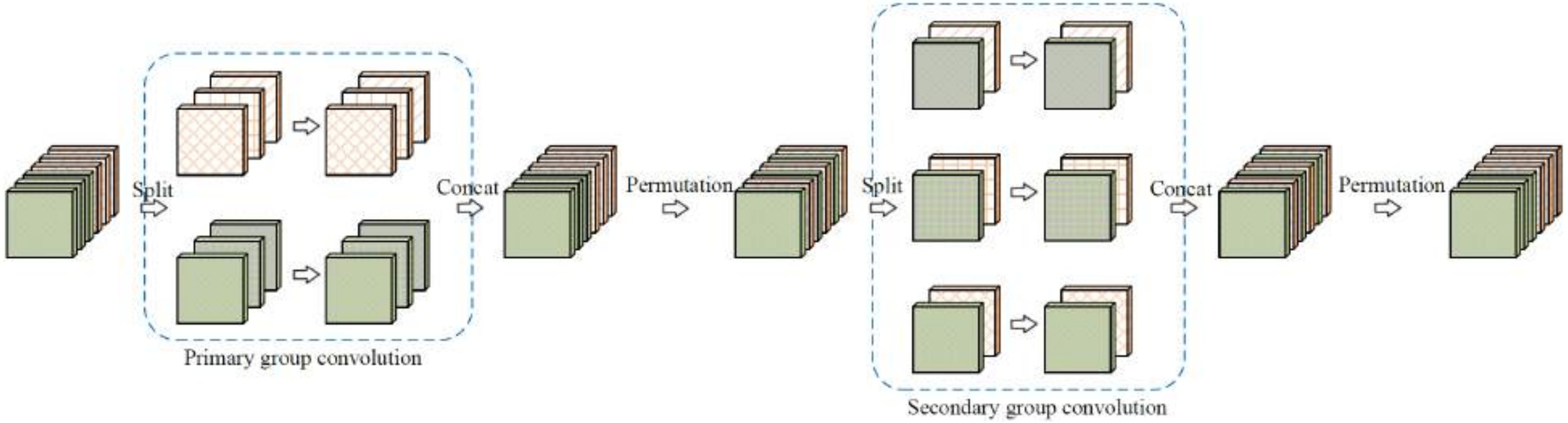


Parameters are used more efficiently by

going less deep and wider

	ResNet-101	DMRNet
	44.5M	43.3M
Top1 validation error	26.41	23.66
Top5 validation error	8.50	**6.81**
Top-1 training error	25.75	19.72
Top-5 training error	8.12	6.59

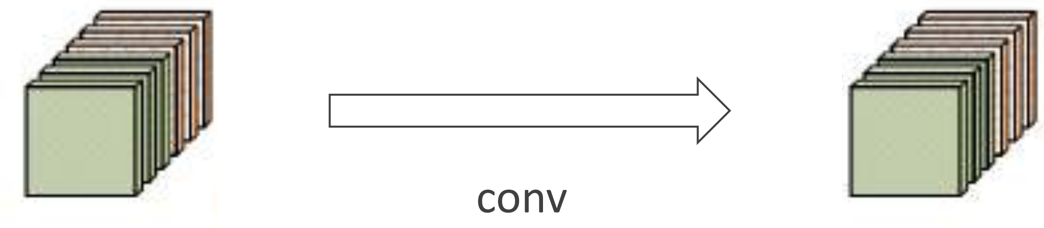
Going Wider with Interleaved Group Convolutions



Ting Zhang, Guo-Jun Qi, Bin Xiao, Jingdong Wang: Interleaved Group Convolutions, ICCV 2017.

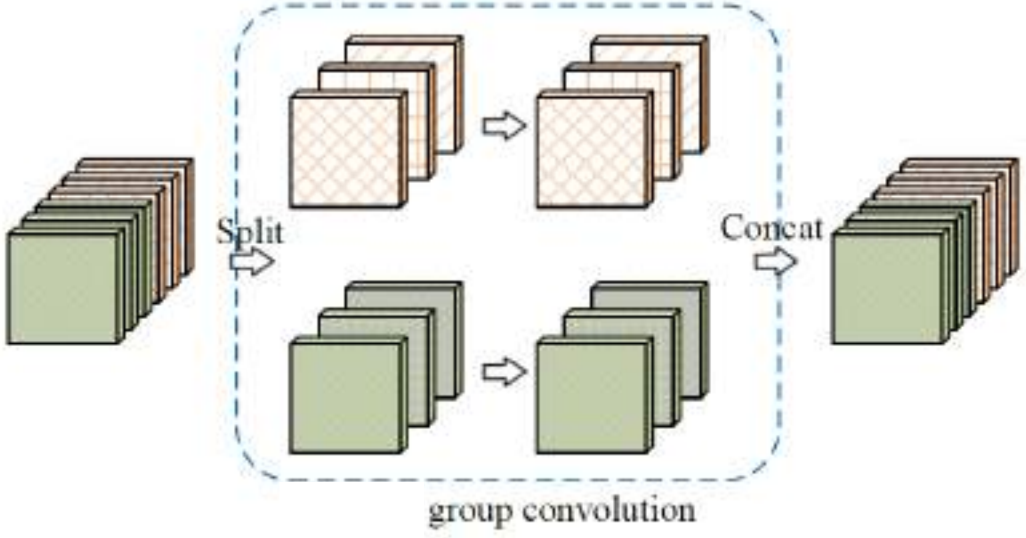
Blog: https://mp.weixin.qq.com/s/PiQB2AvhtDceMJxYN8O8jA

Regular convolution



Complexity: $6\times 5\times 5\times 6$

Group convolution

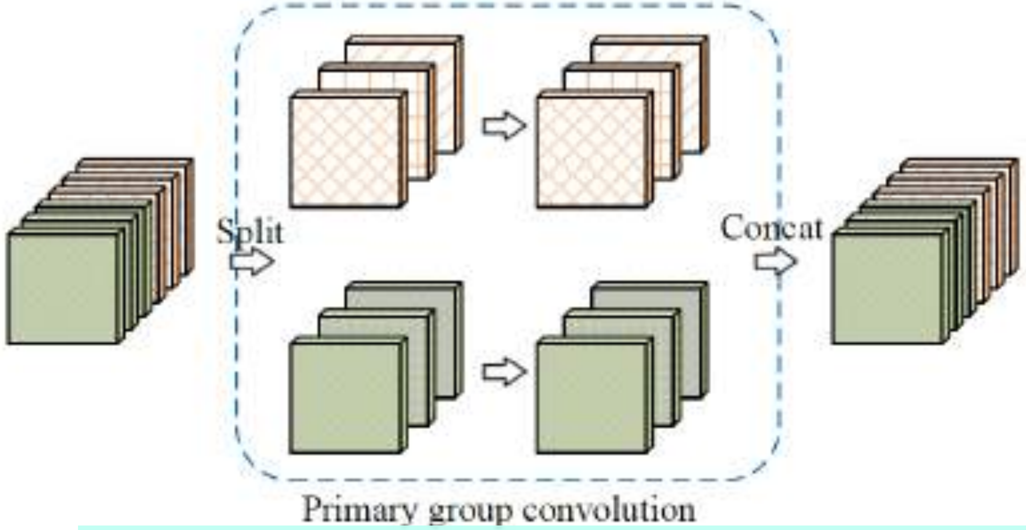


Complexity: $2\times(3\times5\times5\times3)$

Conduct convolutions *separately* over the partitions

Computation cost is lower than regular convolutions

Interleaved group convolutions



Each output channel is connected to each input channel fed into the block

L=2 primary partitions

3 channels in each partition

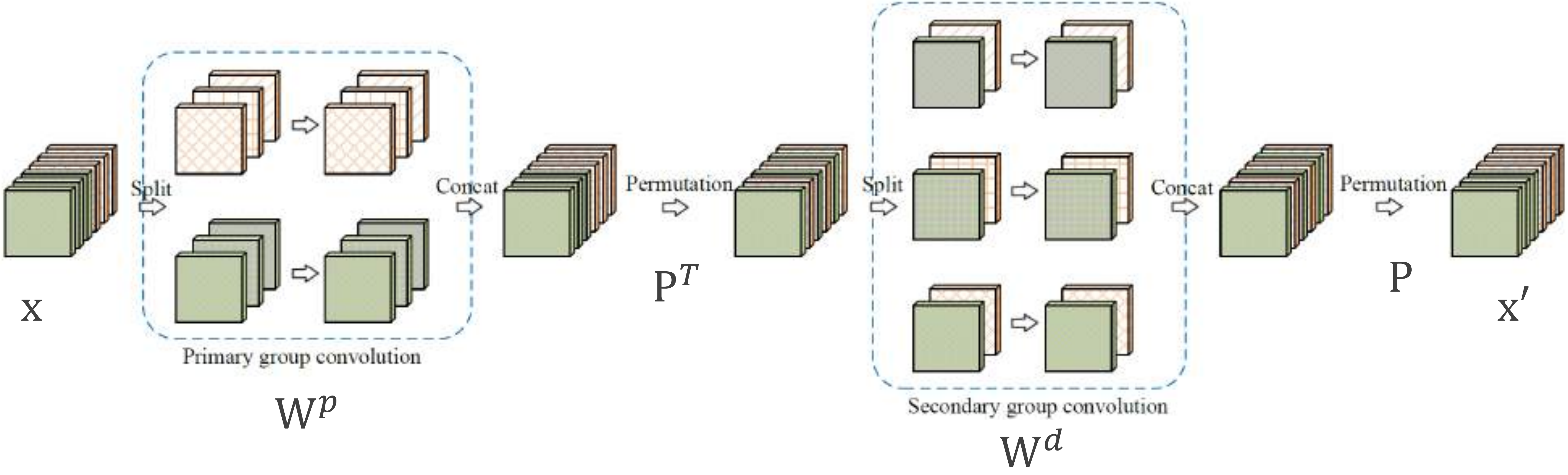
M=3 secondary partitions

2 channels in each partition

An output channel in *primary group convolution* is connected to only a subset of input channels

Orthogonality: The channels in *the same secondary partition* come from *different primary partitions*

Equivalent to a single convolution



$$x' = PW^d P^T W^p x$$

Wider than regular convolutions

- Condition:

$$\frac{L}{L-1} < MS$$

#(Primary partitions) #(Secondary partitions) Primary kernel size

- In the widely-used kernels, $S > 1$
  - Our IGC is wider except $L = 1$
  - Under the same #parameters

Improvement over regular convolutions

CIFAR-10 classification accuracy

depth	RegConv-18	IGC
20	92.55 ± 0.14	92.84 ± 0.26
38	91.57 ± 0.09	92.24 ± 0.62
62	88.60 ± 0.49	90.03 ± 0.85

$+1.43$

Model size: #params ($\times 10^6$)

depth	RegConv-18	IGC
20	0.34	0.15
38	0.71	0.31
62	1.20	0.52

Computation complexity: FLOPS ($\times 10^8$)

depth	RegConv-18	IGC
20	0.51	0.29
38	1.1	0.57
62	1.7	0.95

Improvement over regular convolutions

CIFAR-100 classification accuracy

depth	RegConv-18	IGC
20	68.71 ± 0.32	70.54 ± 0.26
38	65.00 ± 0.57	69.56 ± 0.76
62	58.52 ± 2.31	65.84 ± 0.75

$+7.32$

Model size: #params ($\times 10^6$)

depth	RegConv-18	IGC
20	0.34	0.15
38	0.71	0.31
62	1.20	0.52

Computation complexity: FLOPS ($\times 10^8$)

depth	RegConv-18	IGC
20	0.51	0.29
38	1.1	0.57
62	1.7	0.95

ImageNet classification

	#params $(\times 10^7)$	FLOPS $(\times 10^9)$	Training error		Validation error	
			Top-1	Top-5	Top-1	Top-5
ResNet (Reg. Conv.)	1.333	2.1	21.43	5.96	30.58	10.77
Our approach	0.861	1.3	13.93	2.75	26.95	8.92
					$+3.63$	$+1.85$

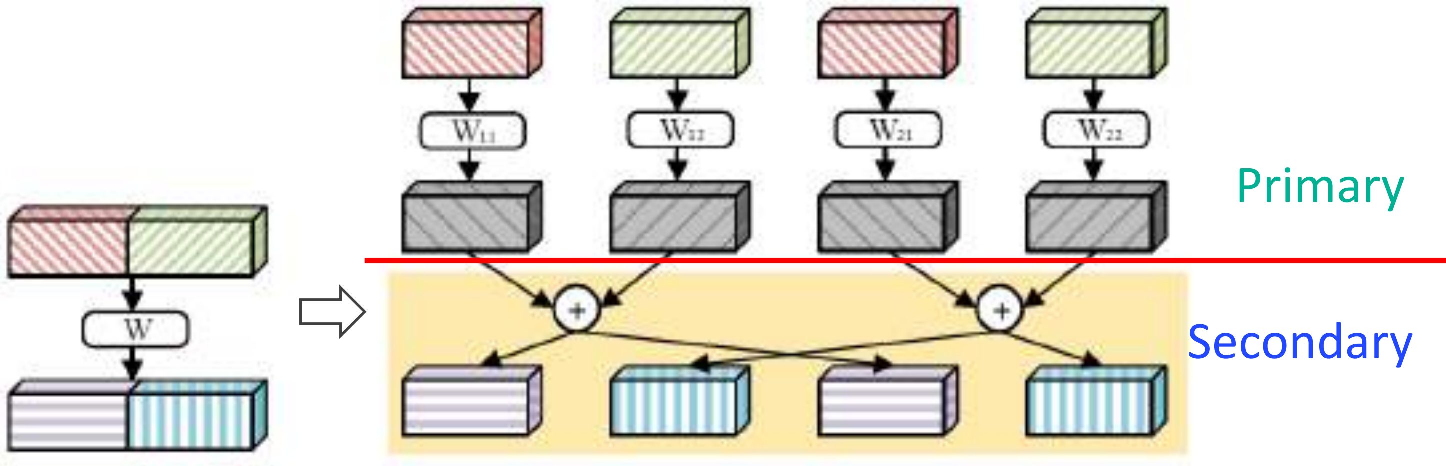
Our approach: replace regular convolutions with our interleaved group convolutions

Regular convolutions are interleaved group convolutions

- Four-branch representation
- Primary group convolution

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{bmatrix}$$

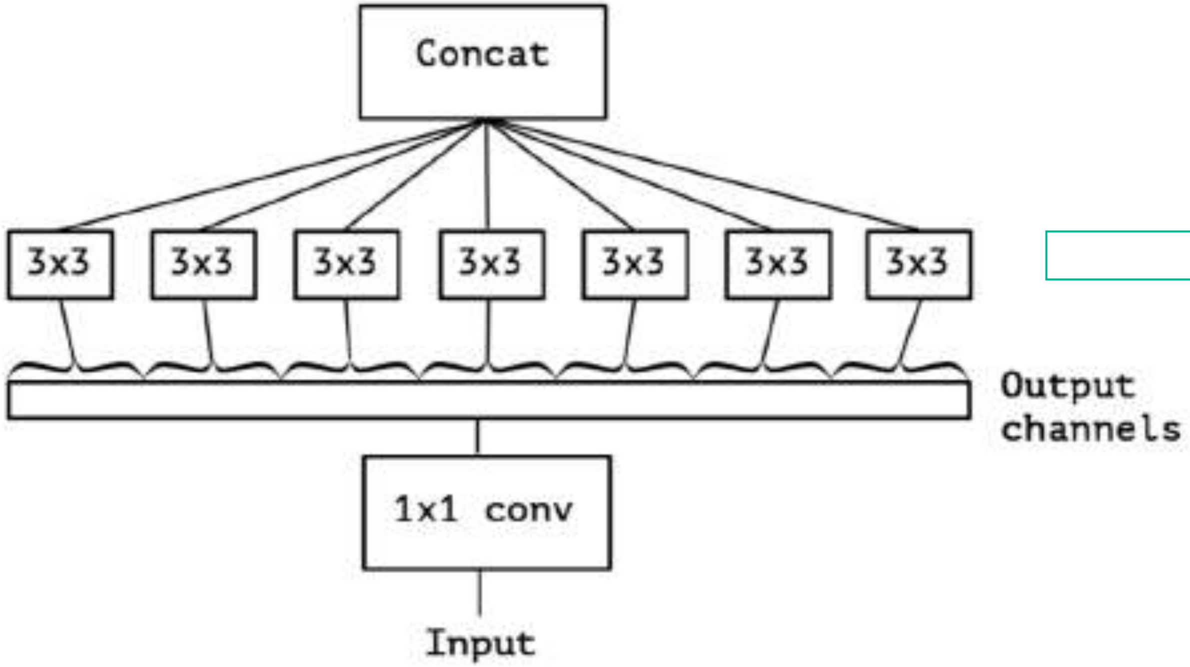
$$\mathbf{W}^p = \mathrm{diag}(\mathbf{W}_{11}, \mathbf{W}_{12}, \mathbf{W}_{21}, \mathbf{W}_{22})$$

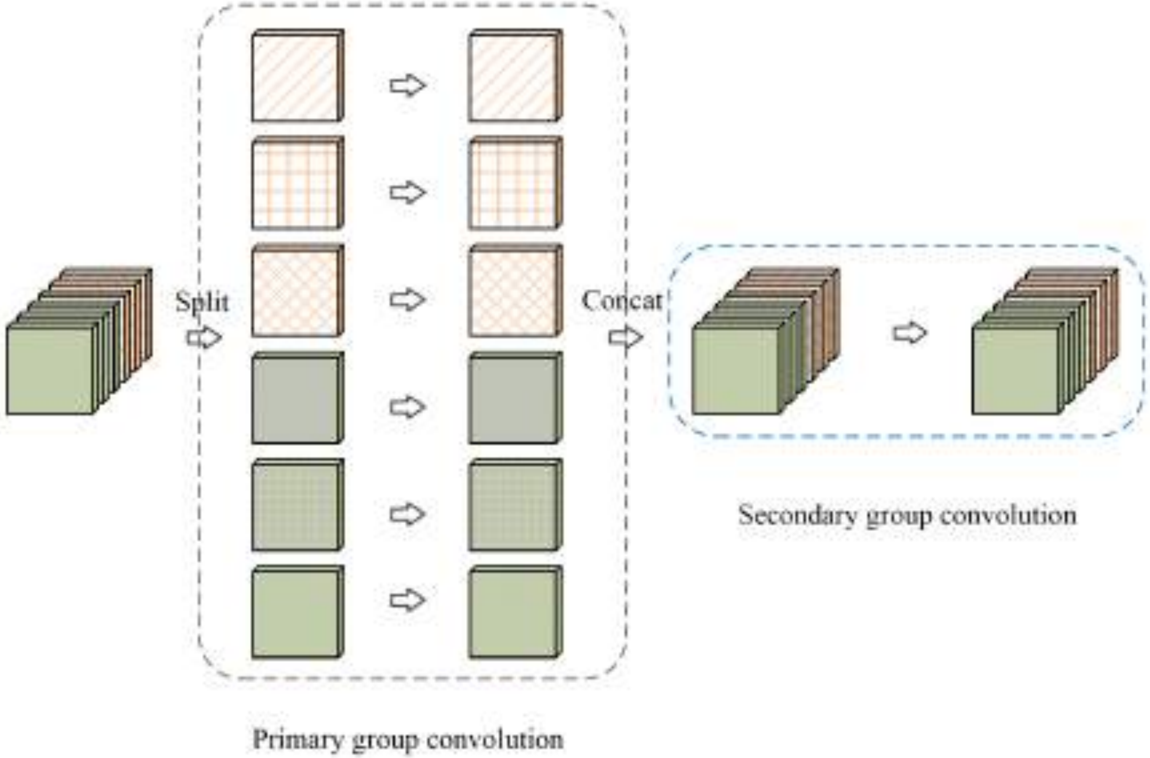


- Secondary group convolution

$$\mathbf{W}^d_{11} = \mathbf{W}^d_{22} = \cdots = \mathbf{W}^d_{MM} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

Google's Xception: an improved version of Inception and applied to mobile apps. A special case of our approach



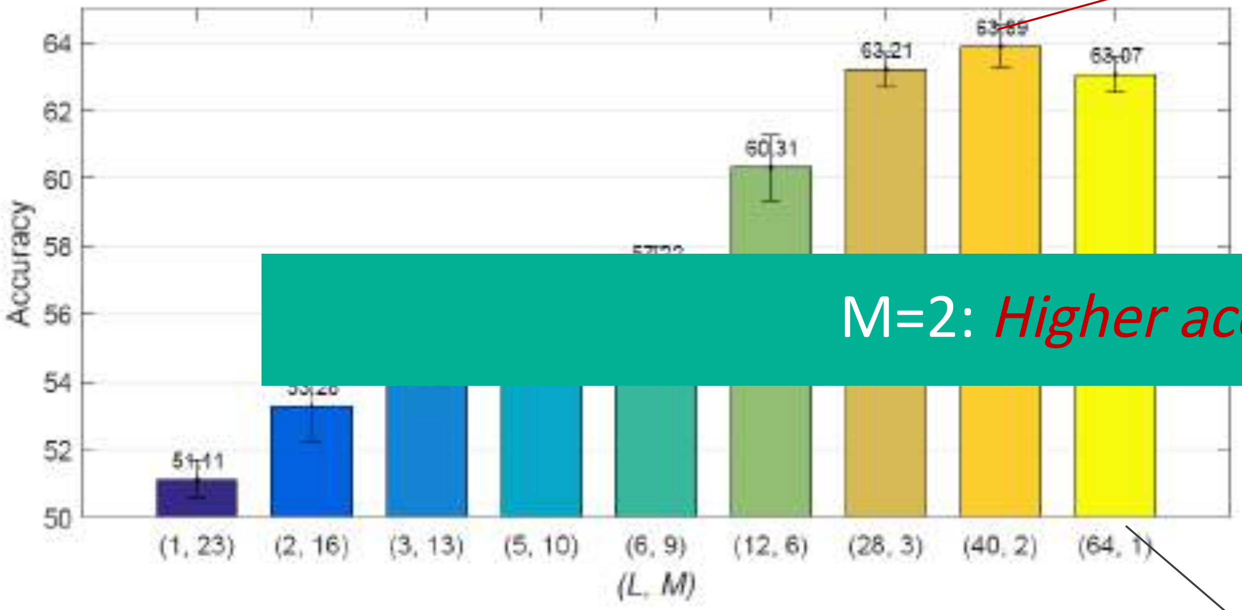


L=6

M=1

Accuracy under same model size and computation complexity

Best accuracy



width	23	32	39	50	54	72	**84**	80	64

M=2: *Higher accura*

Xception

Comparison with Google's Xception

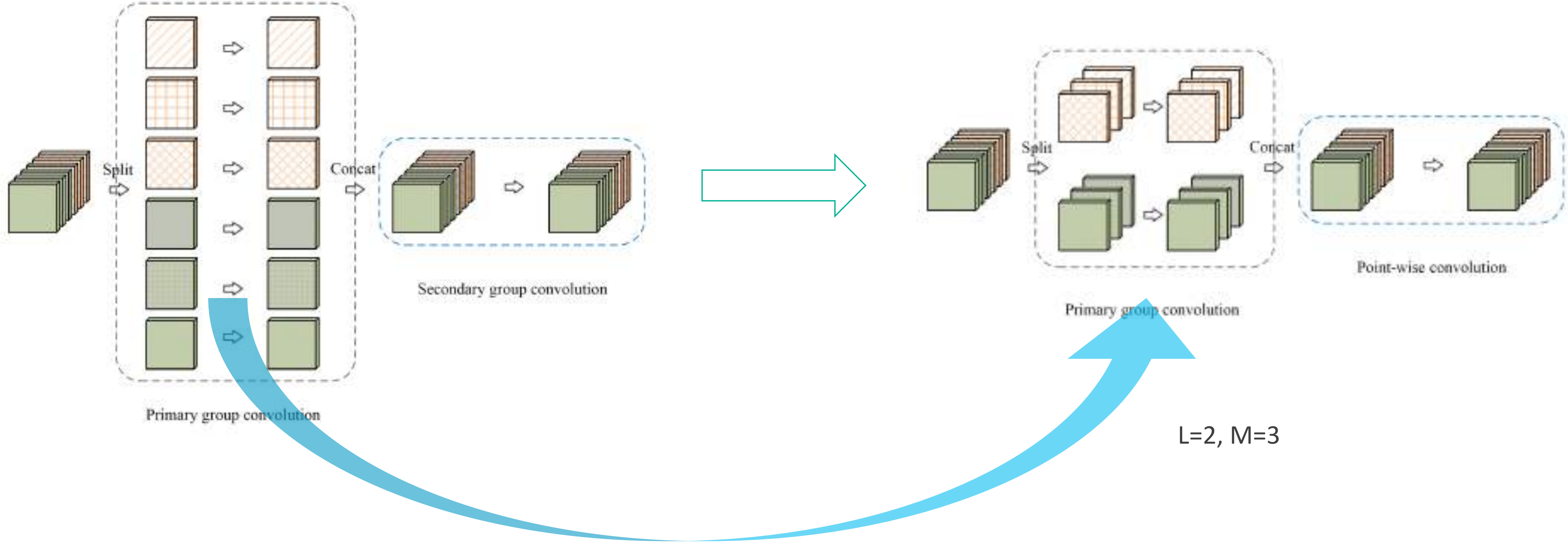
CIFAR-100, Small model

	Xception	Our approach	
testing error	36.93 ± 0.54	36.11 ± 0.62	-0.82
#params	$3.62{\times}10^4$	$3.80{\times}10^4$	
FLOPS	$3.05{\times}10^7$	$3.07{\times}10^7$	

CIFAR-100, Large model

	Xception	Our approach	
testing error	32.87 ± 0.67	31.87 ± 0.58	-1.00
#params	$1.21{\times}10^5$	$1.26{\times}10^5$	
FLOPS	$1.11{\times}10^8$	$1.12{\times}10^8$	

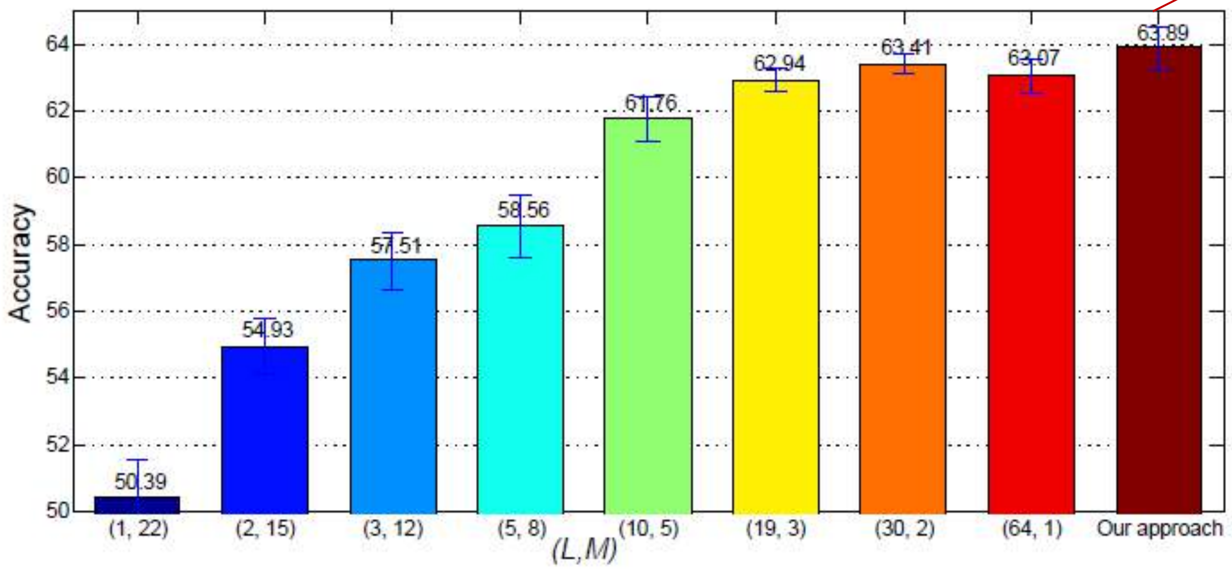
Xception's (our IGC's) variant



L=2, M=3

Comparison with group convolutions + pointwise convolution

Our approach



Comparison with state-of-the-arts

Method	Depth	#Params.	CIFAR-10	CIFAR-100	SVHN
FractalNet with DO/DP	21	38.6M	5.22	23.30	2.01
	21	38.6M	4.60	23.73	1.87
ResNet	110	1.7M	6.41	27.22	2.01
Multi ResNet	200	10.2M	4.35	20.42	-
Wide ResNet	16	11.0M	4.81	22.07	-
	28	36.5M	4.17	20.50	-
DenseNet	40	1.0M	5.24	24.42	1.79
	100	27.2M	3.74	19.25	1.59
DMRNet	56	1.7M	4.94	24.46	1.66
DMRNet-Wide	32	14.9M	3.94	19.25	**1.51**
DMRNet-Wide	50	24.8M	3.57	19.00	1.55
IGC-L16M32	20	17.7M	3.37	19.31	1.63
IGC-L450M2	20	19.3M	**3.30**	19.00	-
IGC-L32M26	20	24.1M	3.31	**18.75**	1.56

Summary

- Advantages
  - Small model （小）
  - Fast computation （快）
  - High accuracy （准）
  - Strong representation
- Drop-in replacement of regular convolutions
  - Interleaving
  - Group 1X1 convolution
- Applicable to
  - Image classification, detection, segmentation,
  - NLP
  - Text
  - …

References

[1] Jingdong Wang, Zhen Wei, Ting Zhang, Wenjun Zeng: Deeply-Fused Nets. CoRR abs/1605.07716 (2016)

[2] Liming Zhao, Jingdong Wang, Xi Li, Zhuowen Tu, Wenjun Zeng: On the Connection of Deep Fusion to Ensembling (Deep Convolutional Neural Networks with Merge-and-Run Mappings). CoRR abs/1611.07718 (2016)

[3] Ting Zhang, Guo-Jun Qi, Bin Xiao, Jingdong Wang: Interleaved Group Convolutions for Deep Neural Networks. ICCV (2017)

THANKS

会场休息