1.核心思想

核心思想有两点：

1. 借鉴resnext分组卷积思想，但不同的是采用1x1卷积核

2. 进行通道清洗，加强通道间的信息流通，提高信息表示能力。

此外本篇论文中也采取了mobilenet的depthwise separasable convolution的方式。

1.1 逐点群卷积pointwise group convolution

这个就是采用resnext的思想，将通道分组，每组分别进行卷积操作，然后再把结果进行concat。但是不同于resnext的是，shufflenet采用的是1x1卷积核。

1.2 通道清洗channel shuffle

通道shuffle就是在分组卷积后得到的feature map下不直接进行concat，先将每组feature map按通道打乱，重新concat。

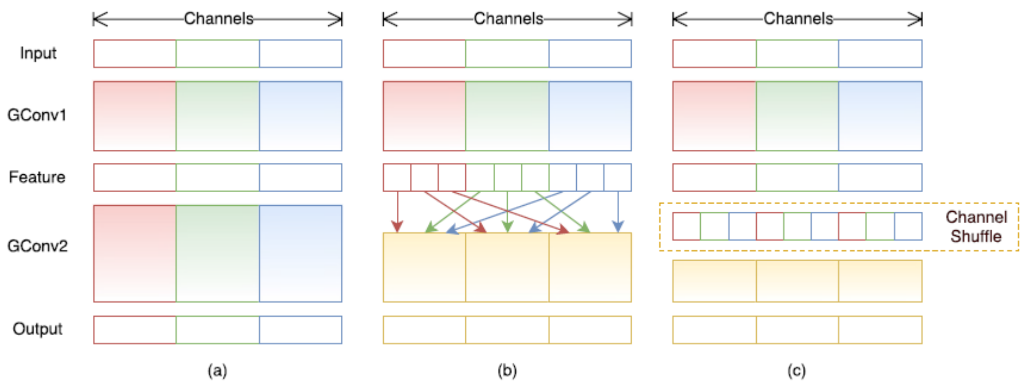


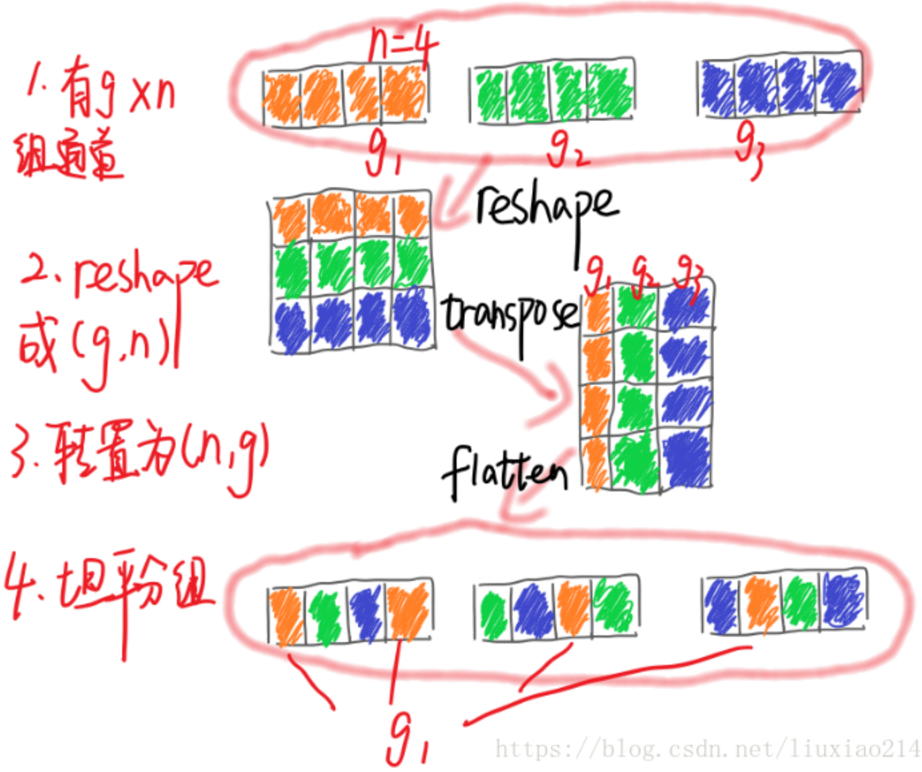
Figure 1. Channel shuffle with two stacked group convolutions. GConv stands for group convolution. a) two stacked convolution layers with the same number of groups. Each output channel only relates to the input channels within the group. No cross talk; b) input and output channels are fully related when GConv2 takes data from different groups after GConv1; c) an equivalent implementation to b) using channel shuffle.

如何进行shuffle:

对于一个卷积层分为g组，

1. 卷积后一共得到g×n个输出通道的feature map；
2. 将feature map 进行 reshape为(g,n);
3. 进行转置为(n,g)；

4. 对转置结果flatten，再分回g组作为下一层的输入。



2.网络结构

2.1 shuffle unit

下图中，a是标准的残差结构，不过是3x3卷积核使用了mobilenet中的depthwise convolution操作；

b是在a的基础上加了本文的通道shuffle操作，先对1x1卷积进行分组卷积操作，然后进行channel shuffle；

c是在旁路加了一步长为2的3x3的平均池化，并将前两者残差相加的操作改为了通道concat，增加了通道数量。

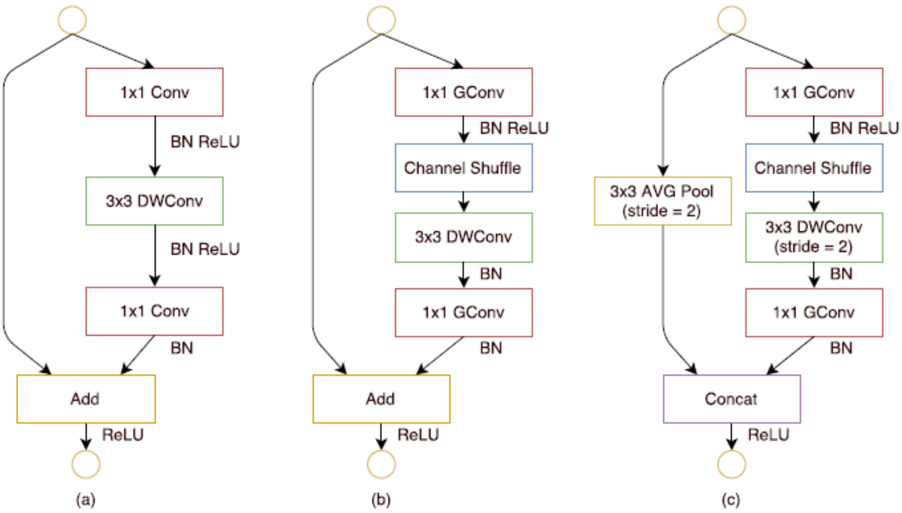


Figure 2. ShuffleNet Units. a) bottleneck unit [9] with depthwise convolution (DWConv) [3, 12]; b) ShuffleNet unit with pointwise group convolution (GConv) and channel shuffle; c) ShuffleNet unit with stride = 2.

2.2 结构

Layer	Output size	KSize	Stride	Repeat	Output channels (g groups)				
					$g = 1$	$g = 2$	$g = 3$	$g = 4$	$g = 8$
Image	224×224				3	3	3	3	3
Conv1	112×112	3×3	2	1	24	24	24	24	24
MaxPool	56×56	3×3	2						
Stage2	28×28		2	1	144	200	240	272	384
	28×28		1	3	144	200	240	272	384
Stage3	14×14		2	1	288	400	480	544	768
	14×14		1	7	288	400	480	544	768
Stage4	7×7		2	1	576	800	960	1088	1536
	7×7		1	3	576	800	960	1088	1536
GlobalPool	1×1	7×7							
FC					1000	1000	1000	1000	1000
Complexity					143M	140M	137M	133M	137M

Table 1. ShuffleNet architecture. The complexity is evaluated with FLOPs, i.e. the number of floating-point multiplication-adds. Note that for Stage 2, we do not apply group convolution on the first pointwise layer because the number of input channels is relatively small.

4.3 实验结果

1、评估逐点组卷积：分组的效果均比没有分组的效果好，但是某些模型随着组数增加，性能有下降，这就是通道间失去联系带来的问题；

Model	Complexity (MFLOPs)	Classification error (%)				
		$g = 1$	$g = 2$	$g = 3$	$g = 4$	$g = 8$
ShuffleNet $1\times$	140	33.6	32.7	32.6	32.8	**32.4**
ShuffleNet $0.5\times$	38	45.1	44.4	43.2	**41.6**	42.3
ShuffleNet $0.25\times$	13	57.1	56.8	55.0	54.2	**52.7**

Table 2. Classification error vs. number of groups g (*smaller number represents better performance*)

2、评估channel shuffle，shuffle会比没有shuffle效果好，而且对于组数越大，效果越好，说明了shuffle的重要性，也说明了上图中组数增加性能下降的问题。

Model	Cls err. (%, no shuffle)	Cls err. (%, shuffle)	Δ err. (%)
ShuffleNet 1x ($g = 3$)	34.5	**32.6**	1.9
ShuffleNet 1x ($g = 8$)	37.6	**32.4**	5.2
ShuffleNet 0.5x ($g = 3$)	45.7	**43.2**	2.5
ShuffleNet 0.5x ($g = 8$)	48.1	**42.3**	5.8
ShuffleNet 0.25x ($g = 3$)	56.3	**55.0**	1.3
ShuffleNet 0.25x ($g = 8$)	56.5	**52.7**	3.8

Table 3. ShuffleNet with/without channel shuffle (*smaller number represents better performance*)

3、与mobilenet的比较

Model	Complexity (MFLOPs)	Cls err. (%)	Δ err. (%)
1.0 MobileNet-224	569	29.4	-
ShuffleNet 2$\times$ ($g = 3$)	524	**26.3**	3.1
ShuffleNet 2$\times$ (with *SE*[13], $g = 3$)	527	**24.7**	4.7
0.75 MobileNet-224	325	31.6	-
ShuffleNet 1.5$\times$ ($g = 3$)	292	**28.5**	3.1
0.5 MobileNet-224	149	36.3	-
ShuffleNet 1$\times$ ($g = 8$)	140	**32.4**	3.9
0.25 MobileNet-224	41	49.4	-