4-connected shift residual networks



ICCV 2019 – Neural Architects Workshop Andrew Brown, Pascal Mettes, Marcel Worring, University of Amsterdam

Network costs increasing!



- Increasing accuracy on ImageNet has come at increasing cost
 - Popular metrics: FLOPs and parameters
- Can we reduce cost without reducing accuracy?

Shift operation



- Shifts operations move input channels spatially
 - Different channels move in different directions
- Shifts are possible spatial convolution replacements
 - Spatial conv. \rightarrow shift + pointwise conv. (i.e. simple matrix multiplication)
- Shifts themselves are zero parameter, zero FLOP operations

Do shifts improve network cost?



- Shifts have shown improvements to compact networks
- Picture not clear for higher FLOP/accuracy networks

Which shift neighbourhood to use?



- Shifts move inputs but in which directions?
 - 8-connected shift: Left, right, up, down and diagonals
 - 4-connected shift: Left, right, up and down only

Applying shifts to ResNet



- First expt: replacement of spatial convolutions in ResNet residual blocks
 - 'Bottleneck' residual block design
 - 3×3 spatial convolution \rightarrow shift + point-wise convolution

Single shift results

	CI	FAR-100		ImageNet			
	#params	FLOPs	acc.	#params	FLOPs	acc.	
ResNet101 [10]	1078K	154M	74.9	44.6M	7.80G	77.6	
ResNet50 [10]	540K	78M	72.3	25.6M	4.09G	75.9	
8-connected shift	605K	85M	74.3	25.6M	4.41G	77.3	
4-connected shift	605K	85M	73.8	25.6M	4.41G	77.3	
8-connected shift (nO)	605K	85M	74.2	25.6M	4.41G	77.0	
4-connected shift (nO)	605K	85M	73.5	25.6M	4.41G	77.0	
No shift	605K	85M	58.4	25.6M	4.41G	61.2	

- Shifts give a large cost reduction
 - More than 40% in both parameters and FLOPs
- Single shift networks gives accuracy penalty BUT
 - Better than reducing network length

Single shift results: shift comparison

	CI	FAR-100		ImageNet			
	#params	FLOPs	acc.	#params	FLOPs	acc.	
ResNet101 [10]	1078K	154M	74.9	44.6M	7.80G	77.6	
ResNet50 [10]	540K	78M	72.3	25.6M	4.09G	75.9	
8-connected shift	605K	85M	74.3	25.6M	4.41G	77.3	
4-connected shift	605K	85M	73.8	25.6M	4.41G	77.3	
8-connected shift (nO)	605K	85M	74.2	25.6M	4.41G	77.0	
4-connected shift (nO)	605K	85M	73.5	25.6M	4.41G	77.0	
No shift	605K	85M	58.4	25.6M	4.41G	61.2	

• 4-connected shift performs as well as 8-connected on ImageNet

No shift results

	CI	FAR-100		ImageNet			
	#params	FLOPs	acc.	#params	FLOPs	acc.	
ResNet101 [10]	1078K	154M	74.9	44.6M	7.80G	77.6	
ResNet50 [10]	540K	78M	72.3	25.6M	4.09G	75.9	
8-connected shift	605K	85M	74.3	25.6M	4.41G	77.3	
4-connected shift	605K	85M	73.8	25.6M	4.41G	77.3	
8-connected shift (nO)	605K	85M	74.2	25.6M	4.41G	77.0	
4-connected shift (nO)	605K	85M	73.5	25.6M	4.41G	77.0	
No shift	605K	85M	58.4	25.6M	4.41G	61.2	

- No shift networks only one spatial convolution in very first layer
- Accuracy penalty suffered but surprisingly not so much!

Even more shifts!



- Add shifts to down- and up- sampling bottleneck convolutions
 - Idea is to allow larger receptive field within each block

Removing the bottleneck



- Now the spatial convolutions are gone, why use a bottleneck?
 - No longer a need to down-sample in each residual block
 - Flatten the channel structure
- Need to reduce length to reduce cost: 101 layers \rightarrow 35 layers

Multi-shift results: with bottleneck

	CIFAR-100				ImageNet			
	#params	FLOPs accuracy		#params	FLOPs	accuracy		
						wd: 4×10^{-5}	wd: 1×10^{-4}	
Baselines								
ResNet101 [10]	1078K	154M	74.9	44.6M	7.80G	77.6	77.4	
Multi-shifting								
8-connected	605K	85M	74.3	25.6M	4.41G	76.8	77.2	
4-connected	605K	85M	75.1	25.6M	4.41G	77.3	77.6	
Flattened architecture								
8-connected	1068K	162M	76.9	40.8M	7.72G	77.2	77.8	
4-connected	1068K	162M	77.5	40.8M	7.72G	77.8	78.4	

- Multi-shift networks match ResNet in accuracy!
 - ...but only for 4-connected shifts, not 8-connected shifts
 - Maintains >40% parameters and FLOPs reductions

Multi-shift results: without bottleneck

		CIFAR-100				ImageNet		
#par		s FLOPs accuracy		#params	FLOPs	accuracy		
						wd: 4×10^{-5}	wd: 1×10^{-4}	
Baselines								
ResNet101 [10]	1078K	154M	74.9	44.6M	7.80G	77.6	77.4	
Multi-shifting								
8-connected	605K	85M	74.3	25.6M	4.41G	76.8	77.2	
4-connected	605K	85M	75.1	25.6M	4.41G	77.3	77.6	
Flattened archited	ture							
8-connected	1068K	162M	76.9	40.8M	7.72G	77.2	77.8	
4-connected	1068K	162M	77.5	40.8M	7.72G	77.8	78.4	

- Multi-shift networks <u>without</u> bottleneck: beats ResNet in accuracy
- Again best performance (+0.8%) is for 4-connected shifts

Results in context



• Shifts can improve high accuracy CNNs!

Summary

- Studied variants of the shift operation
 - Compare 8- and 4- connected shift neighbourhoods

- Modified ResNet bottleneck residual blocks to include shifts
 - Consider both single and multiple shifts in each block

- Multi-4-connected shift variants can improve ResNet
 - 1st case: Improve costs by more than 40% at same accuracy
 - 2nd case: Improves ImageNet accuracy by +0.8% for ~same costs