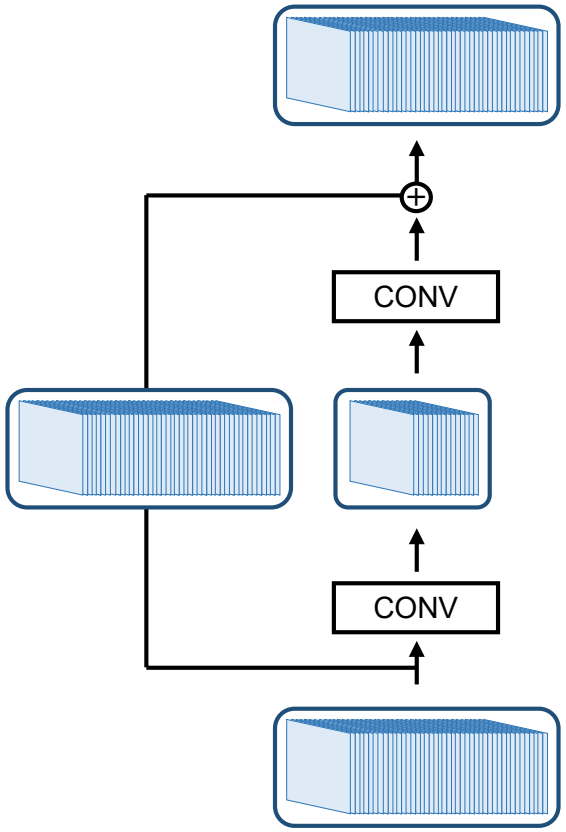
Slimming ResNet by Slimming Shortcut

Donggyu Joo, Doyeon Kim, and Junmo Kim

SIIT Lab. KAIST

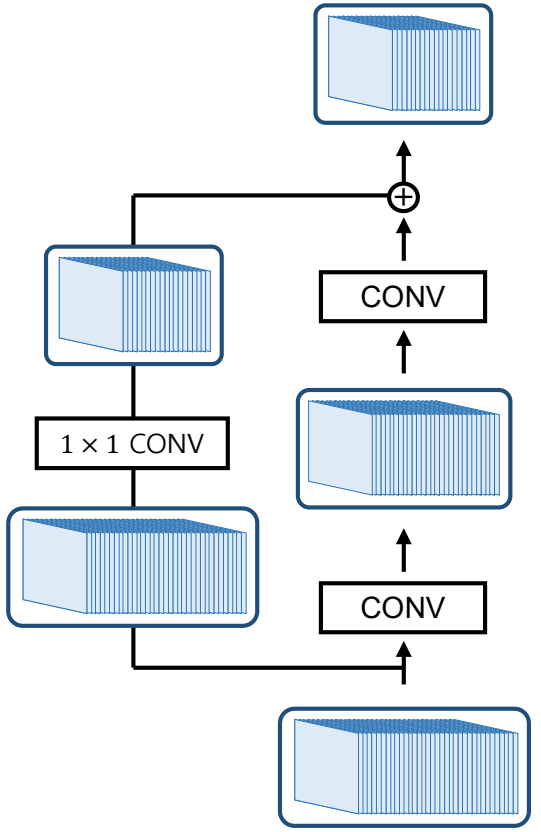
Statistical Inference and Information Theory Laboratory

Introduction & Objectives



Existing Pruning

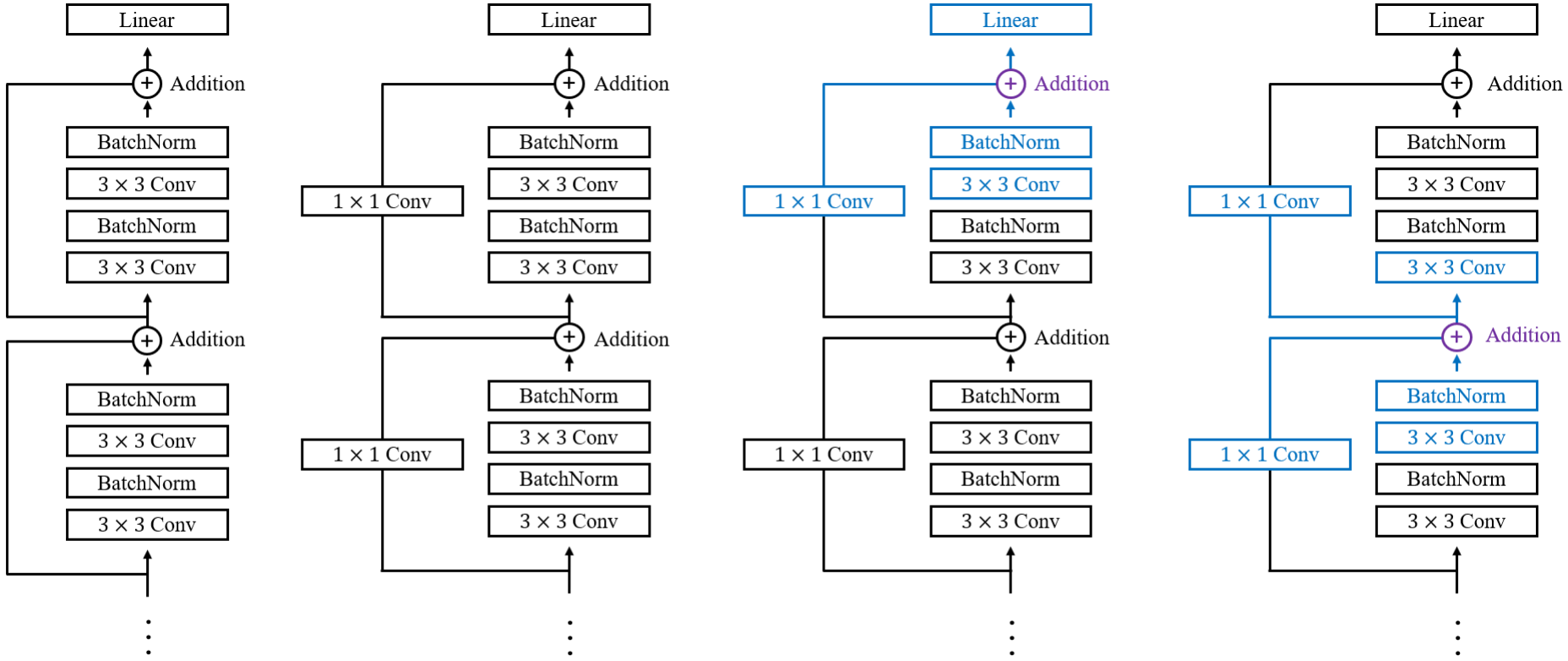
- ✓ Pruning is applied to only inside of residual block
- ✓ Unstable bottleneck-like structure is obtained



Proposed *SSPruning*

- ✓ Pruning is applied to both shortcut and block.
- ✓ We obtain the optimal structure without restriction

Method 1. Shortcut Separation

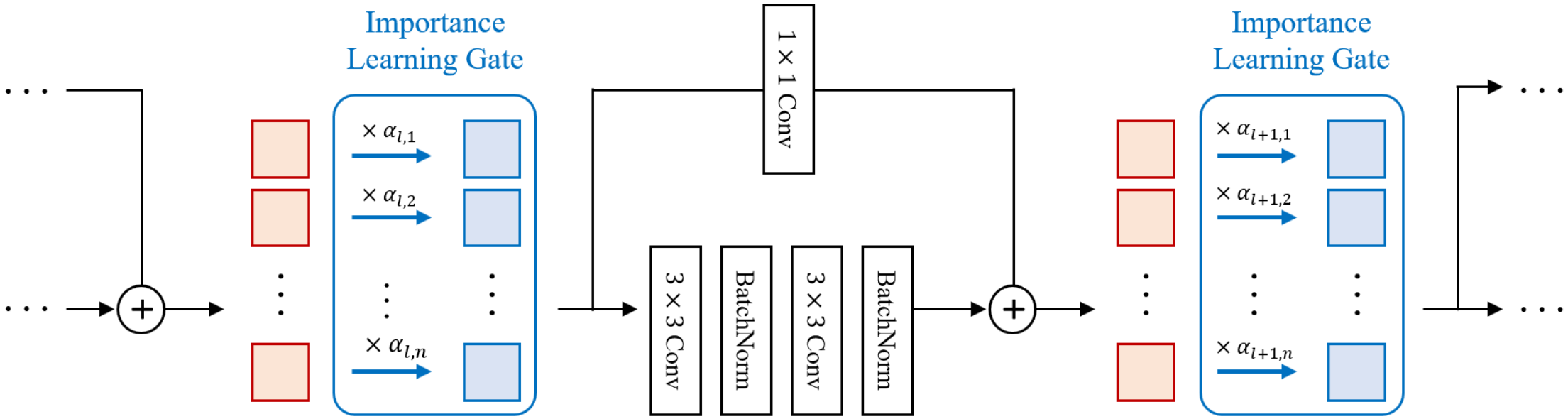


(a) Original ResNet architecture. All shortcuts are connected together.

(b) Shortcut Separation: Additional 1×1 convolutional is added at each shortcut to separate them.

(c, d) To remove the channels related to the addition layer (marked in purple), we only need to consider the channels in layers marked in blue.

Method 2. Importance Learning Gate (ILG)



- ✓ The importance learning gate (ILG) layer is applied after each addition layer.
- ✓ ILG is composed of n parameters $\alpha_1, \alpha_2, \cdots, \alpha_n \in R$ where n is the number of channels in the corresponding location. These n parameters are initialized as 1, and independently multiplied to the output of each corresponding channel.

Experiments

1. Comparative Experiments

- ✓ Datasets: CIFAR-10, Network: ResNet-32 and ResNet-110
- ✓ SSPruning-α: Result of the proposed method (α% of FLOPs are pruned from the original network).
- ✓ Our SSPruning achieves the state-of-the-art performance in several aspects.

Network	Method	Baseline Acc (%)	Pruned Acc (%)	Acc ↓ (%)	Flops ↓ (%)
ResNet-32	MIL [32]	92.33	90.74	1.59	31.2
	SFP [33]	92.63	92.08	0.55	41.5
	FPGM [12]	92.63	91.93	0.70	53.2
	SSPruning-50	**92.54**	**92.80**	**-0.26**	**50.1**
	SSPruning-60	**92.54**	**92.36**	**0.18**	**60.1**
ResNet-110	PFEC [34]	93.53	93.30	0.23	38.6
	SFP [33]	93.68	93.38	0.30	40.8
	NISP [29]	-	-	0.18	43.8
	FPGM [12]	93.68	93.85	-0.17	52.3
	SSPruning-50	**93.65**	**93.99**	**-0.34**	**50.1**
	SSPruning-60	**93.65**	**93.76**	**-0.11**	**60.1**

Experiments

2. Ablation on Shortcut Pruning

- ✓ Datasets: CIFAR-10, Network: ResNet-32
- ✓ Pruning of both shortcut and inblock outperforms traditional inblock pruning in various settings.

Experiment for SSPruning			
Flops ↓ (%)	Pruned Acc (%)		
	InBlock Only	Shortcut Only	Both (proposed)
50%	92.58	92.16	**92.80**
60%	92.08	91.57	**92.36**

(a)

Experiment for Random Selection			
Flops ↓ (%)	Pruned Acc(%)		
	InBlock Only	Shortcut Only	Both
50%	92.32	92.29	**92.51**
60%	91.84	91.85	**92.03**

(b)

THANK YOU

Slimming ResNet by Slimming Shortcut

ICPR 2020