

ShakeDrop Regularization

Yoshihiro Yamada, Masakazu Iwamura and Koichi Kise
Osaka Prefecture University



Contributions

- We propose a new powerful regularization method, **ShakeDrop**, for improvements of ResNets architectures and achieved **state-of-the-art results** on image classification datasets, CIFAR-10/100 (as of Oct. 2017)
- We confirmed **ShakeDrop** stabilizes learning strongly disturbed by **multiplying even a negative factor** by regarding **StochasticDepth [1] mechanism as a probabilistic switch of two network architectures**

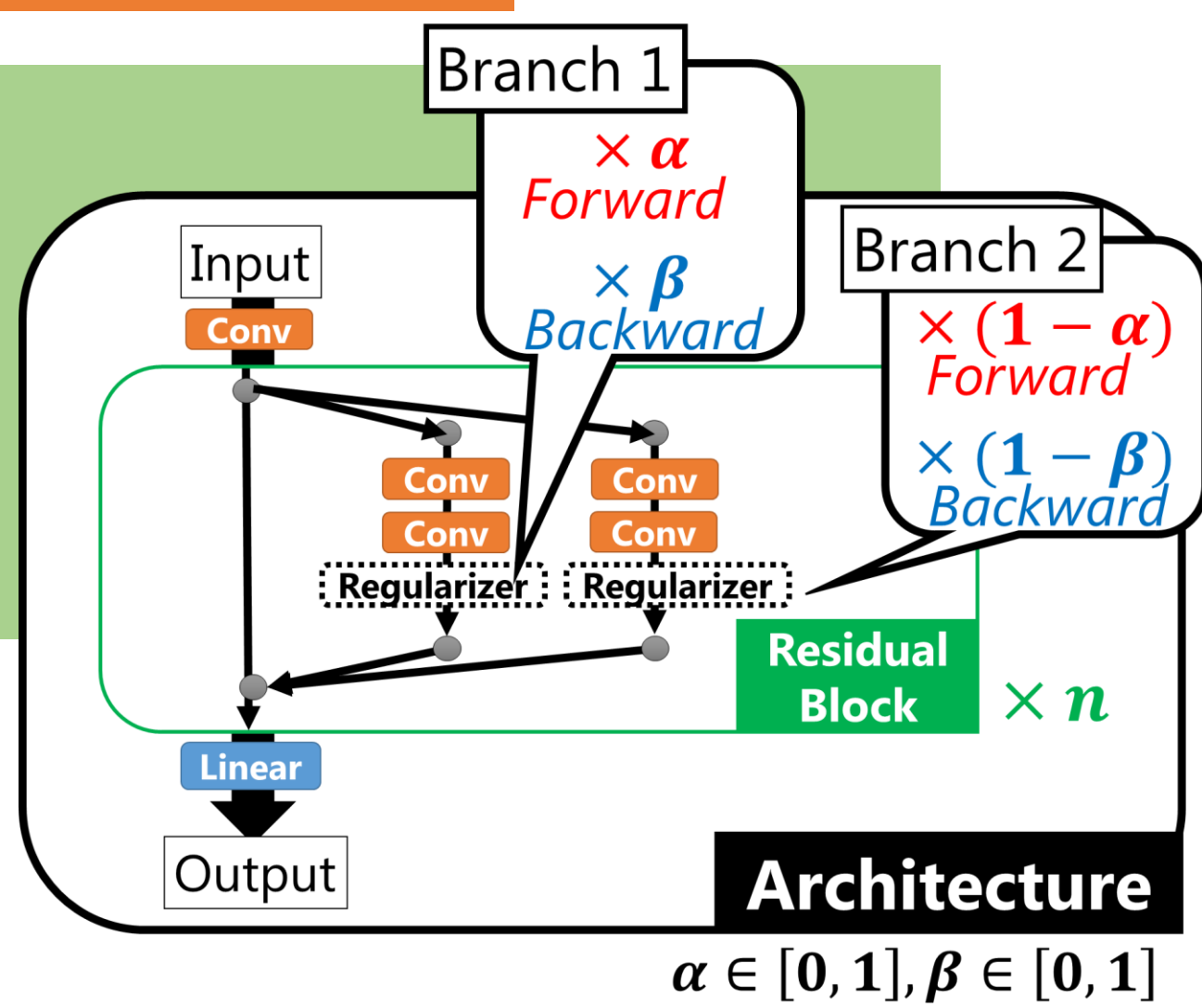
Regularization methods	1-branch networks			2-branch networks
	ResNet	WideResnet	PyramidNet	ResNeXt
StochasticDepth [1]	☹	☹	☹	☹
Shake-Shake [2]	-	-	-	😊
ShakeDrop (ours)	😊	😊	😊	😊

Architectures

Comparison of Regularization methods

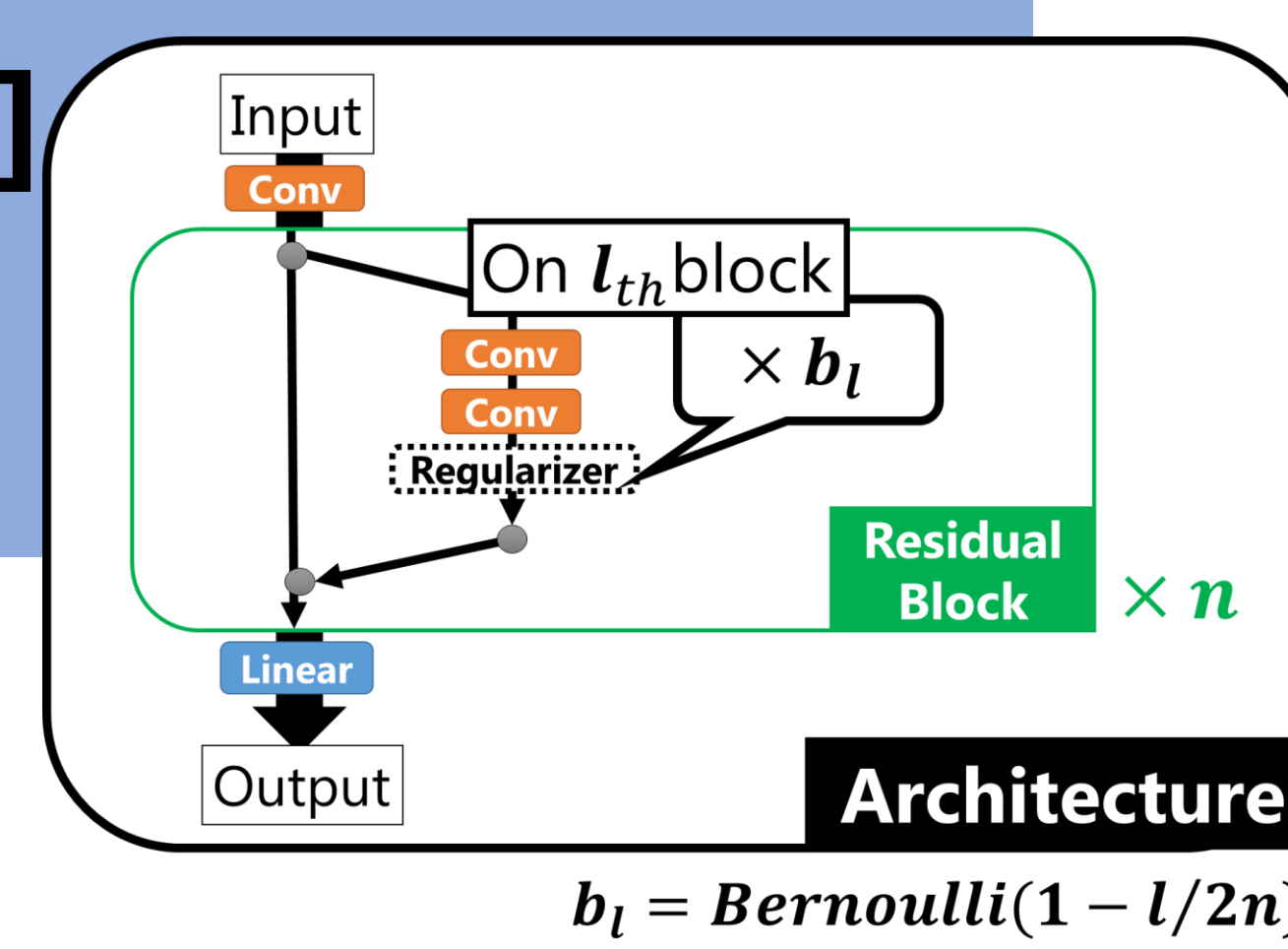
Shake-Shake [2]

- For 2-branch networks
- High accuracy



StochasticDepth [1]

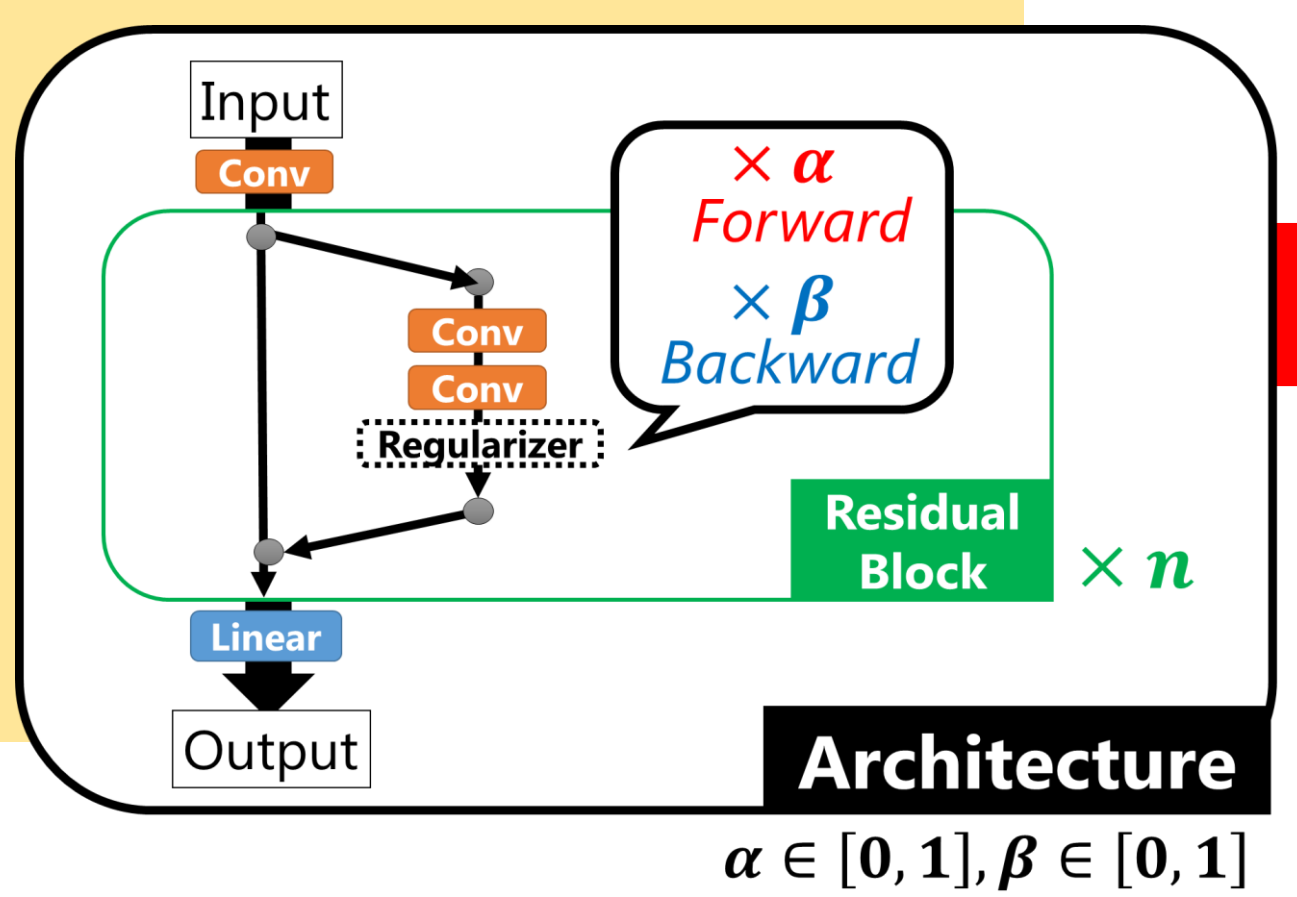
- For any-branch networks
- Low accuracy



Stabilize

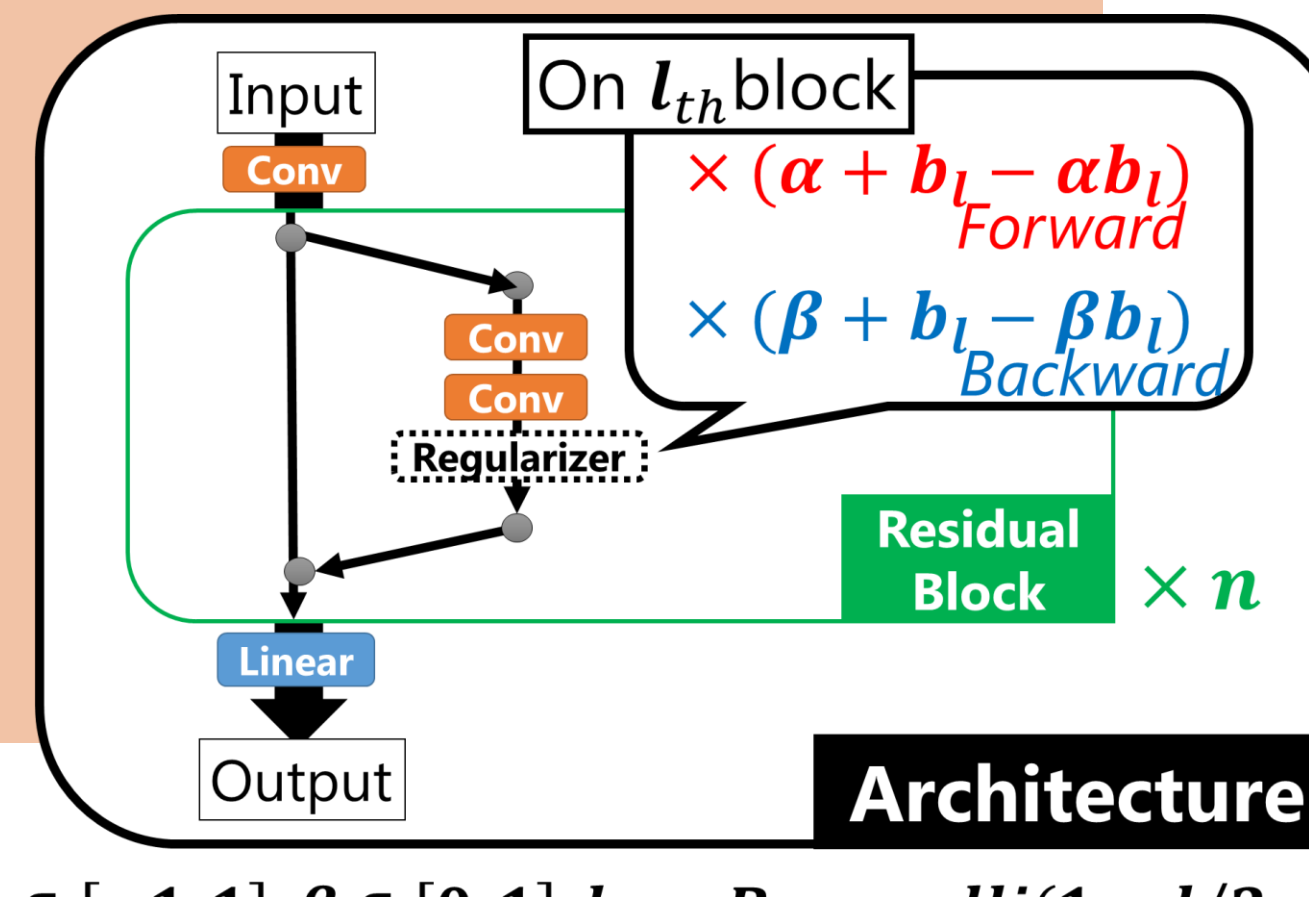
1-branch Shake

- (intermediate method)
- For any-branch networks
- Low accuracy



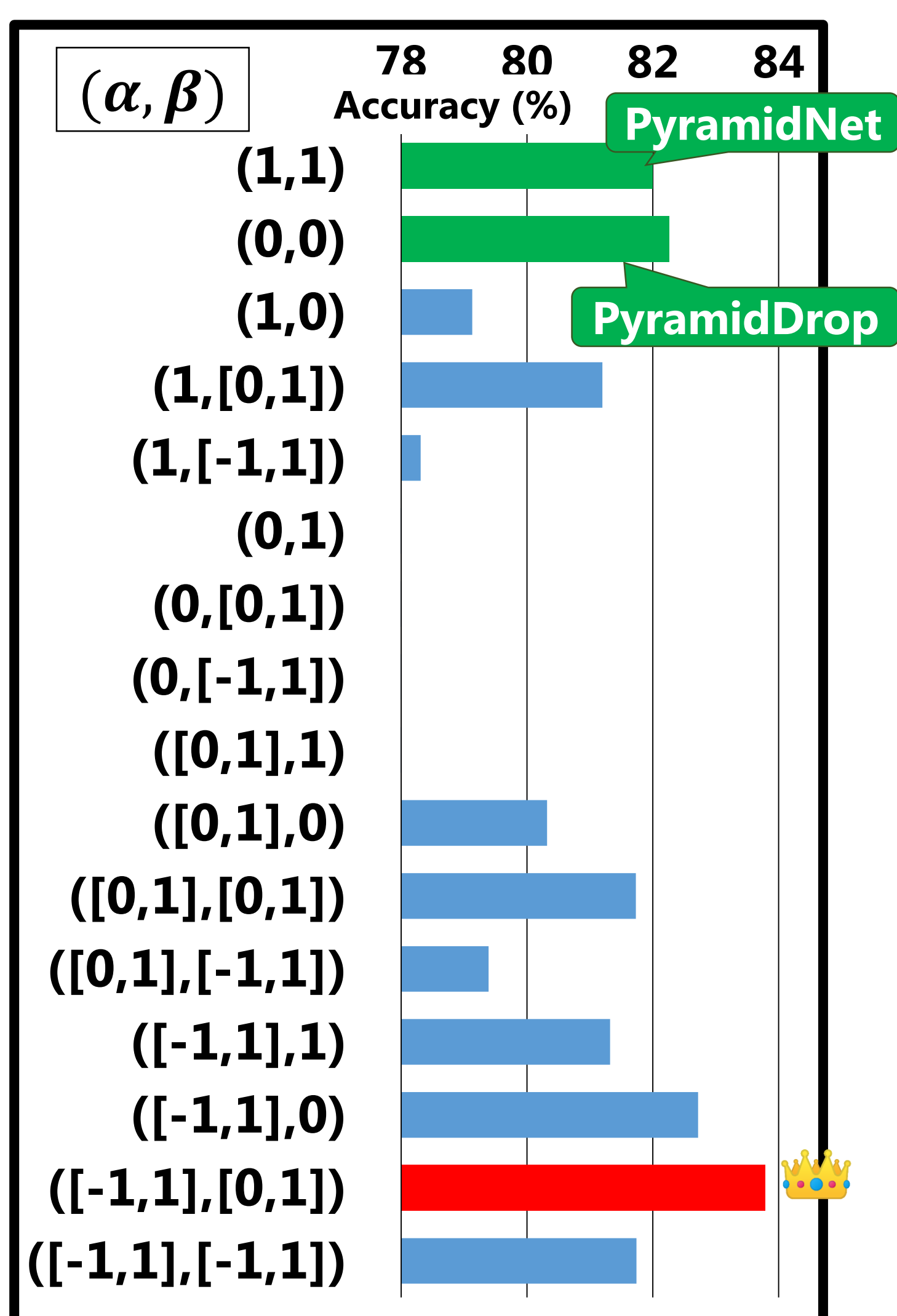
ShakeDrop (proposed method)

- For any-branch networks
- High accuracy



Experiments (image classification)

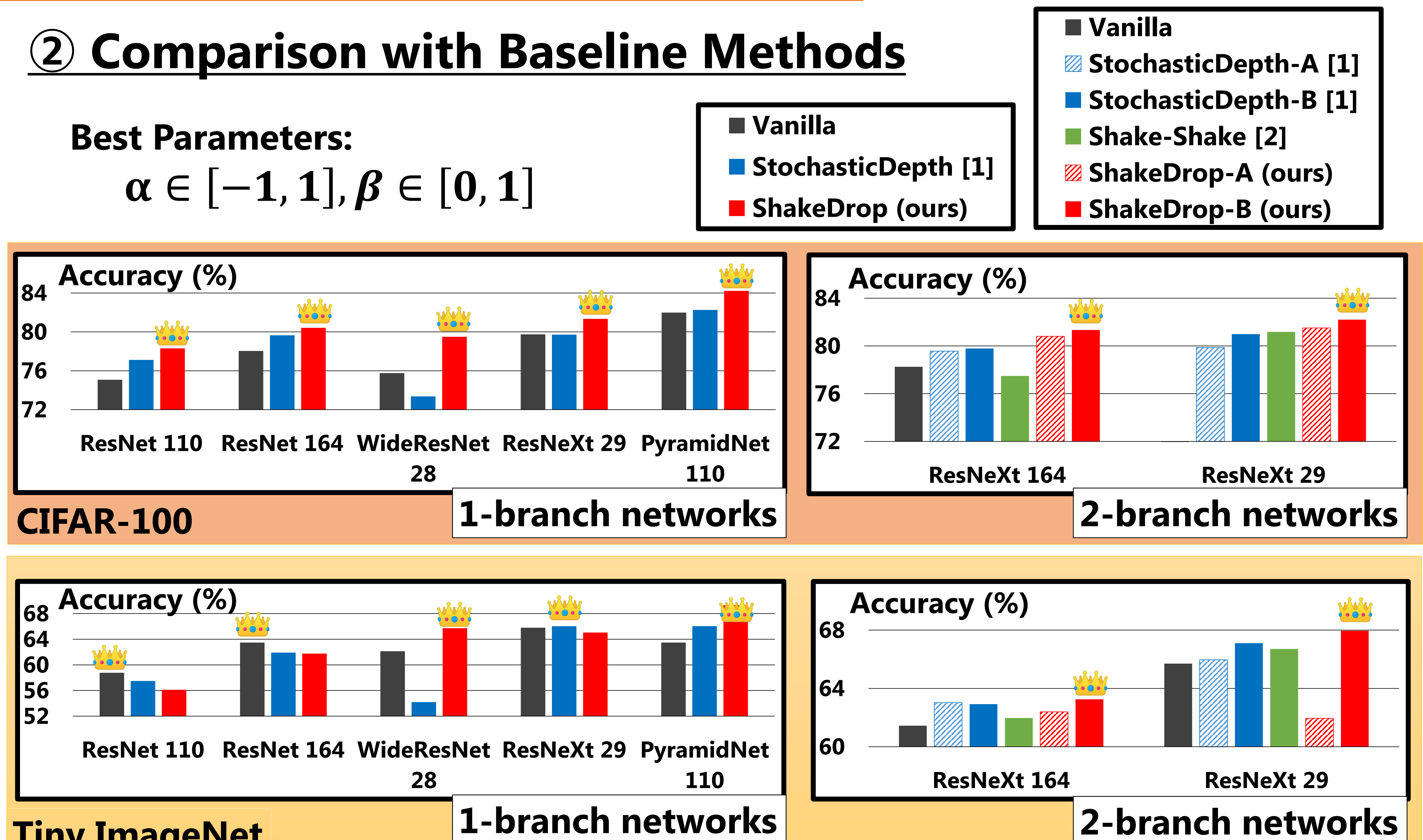
① Parameter Search



Best Parameters:
 $\alpha \in [-1, 1], \beta \in [0, 1]$

② Comparison with Baseline Methods

Best Parameters:
 $\alpha \in [-1, 1], \beta \in [0, 1]$



[1] Gao Huang, Yu Sun, Zhuang Liu, Daniel Sedra and Kilian Weinberger, "Deep Networks with Stochastic Depth," NIPS2016

[2] Xavier Gastaldi, "Shake-Shake Regularization of 3-branch Residual Networks," ICLR2017 workshop