Network	# of Params	Output Feat. Dim.	Depth	Training Mem.	CIFAR-10	CIFAR-100
NiN [18]	-	-	-	-	8.81	35.68
All-CNN [27]	-	-	-	-	7.25	33.71
DSN [17]	-	-	-	-	7.97	34.57
FitNet [21]	-	-	-	-	8.39	35.04
Highway [29]	-	-	-	-	7.72	32.39
Fractional Max-pooling [4]	-	-	-	-	4.50	27.62
ELU [29]	-	-	-	-	6.55	24.28
ResNet [7]	1.7M	64	110	547MB	6.43	25.16
ResNet [7]	10.2M	64	1001	2,921MB	-	27.82
ResNet [7]	19.4M	64	1202	2,069MB	7.93	-
Pre-activation ResNet [8]	1.7M	64	164	841MB	5.46	24.33
Pre-activation ResNet [8]	10.2M	64	1001	2,921MB	4.62	22.71
Stochastic Depth [10]	1.7M	64	110	547MB	5.23	24.58
Stochastic Depth [10]	10.2M	64	1202	2,069MB	4.91	-
FractalNet [14]	38.6M	1,024	21	-	4.60	23.73
SwapOut v2 (width$\times$4) [26]	7.4M	256	32	-	4.76	22.72
Wide ResNet (width$\times$4) [34]	8.7M	256	40	775MB	4.97	22.89
Wide ResNet (width$\times$10) [34]	36.5M	640	28	1,383MB	4.17	20.50
Weighted ResNet [24]	19.1M	64	1192	-	5.10	-
DenseNet ($k = 24$) [9]	27.2M	2,352	100	4,381MB	3.74	19.25
DenseNet-BC ($k = 40$) [9]	25.6M	2,190	190	7,247MB	3.46	17.18
PyramidNet ($\alpha = 48$)	1.7M	64	110	655MB	4.58$\pm$0.06	23.12$\pm$0.04
PyramidNet ($\alpha = 84$)	3.8M	100	110	781MB	4.26$\pm$0.23	20.66$\pm$0.40
PyramidNet ($\alpha = 270$)	28.3M	286	110	1,437MB	3.73$\pm$0.04	18.25$\pm$0.10
PyramidNet (bottleneck, $\alpha = 270$)	27.0M	1,144	164	4,169MB	3.48$\pm$0.20	17.01$\pm$0.39
PyramidNet (bottleneck, $\alpha = 240$)	26.6M	1,024	200	4,451MB	3.44$\pm$0.11	16.51$\pm$0.13
PyramidNet (bottleneck, $\alpha = 220$)	26.8M	944	236	4,767MB	3.40$\pm$0.07	16.37$\pm$0.29
PyramidNet (bottleneck, $\alpha = 200$)	26.0M	864	272	5,005MB	**3.31**$\pm$0.08	**16.35**$\pm$0.24