Method	FLOPs	Δ top-1	Δ top-5
RESNET-50 @ ILSVRC12 dataset			
ThiNet	2.25	-1.87	-1.12
Channel pruning for Accelerating VDNN	2.00	–	-1.40
Soft filter pruning	1.72	-1.54	-0.81
Discrimination-aware Channel Pruning	2.25	-1.06	-0.61

Comparisons of top-1 and top-5 accuracies for ResNet-50 on ILSVRC-12 validation set. Pre-trained ResNet-50 has 76.15% top-1 and 92.87% accuracies.

Method	FLOPs	Δ top-1	Δ top-5
RESNET-18 @ ILSVRC12 dataset			
Pruning filters for Efficient Convnets	1.72	-3.18	-1.85
Network Slimming	1.39	-1.77	-1.29
Discrimination-aware Channel Pruning	1.89	-2.29	-1.38
Channel Gating NN	1.61	-1.62	-1.03
Feature Boosting and Suppression	1.98	-2.54	-1.46

Table 1: Comparisons of top-1 and top-5 accuracies for ResNet-18 on ILSVRC-12 validation set. Pre-trained ResNet-18 has 69.76% top-1 and 90.36% top-5 accuracies.

Method	**Δ top-5 errors (%)**		
	3×	4×	5×
VGG-16 @ ILSVRC12 dataset			
Pruning filters for Efficient Convnets	—	-8.6	-14.6
Perforated CNNs	-3.7	-5.5	—
Network Slimming	-1.37	-3.26	-5.18
Runtime Neural Pruning	-2.32	-3.23	-3.58
Channel Pruning for Accelerating VDNN	0.0	-1.0	-1.7
AutoML Compression	—	—	-1.4
ThiNet-Conv	-0.37	—	—
Feature Boosting and Suppression	-0.04	-0.52	-0.59

Table 2: Comparisons of top-5 error rate for VGG-16 on ILSVRC-12 validation set under 3×, 4× and 5× FLOPs reduction. Results from Channel Pruning for Accelerating VDNN only show numbers with one digit after the decimal point.

Model	MUSCO	Tucker2-iter
AlexNet	-0.81	-4.2
VGG-16	-0.15	-2.8
YOLOv2	-0.19	-3.1
Tiny YOLOv2	-0.10	-2.7

Table 3: Quality drop after iterative compression and one-time compression. For AlexNet and VGG-16 metric is Δ Top-5 accuracy, for YOLO - Δ mAP

Model	FLOPs	mAP
FASTER R-CNN C4 (RESNET-50) @ VOC2007		
Used baseline	1.0×	75.0
Tucker2-iter (nx, 1.4)	1.17×	76.8(+1.8)
MUSCO(nx, 1.4, 2)	**1.39×**	**77.0(+2.0)**
MUSCO(nx, 1.4, 3)	**1.57×**	**75.4(+0.4)**
Tucker2-iter (nx, 3.16)	1.49×	75.0(+0.0)

Table 4: Comparison of Faster R-CNN (with ResNet-50 backbone) compressed models on Pascal VOC2007 evaluation dataset.

Model	FLOPs	mAP	mAP.50
FASTER R-CNN FPN (RESNET-50) @ COCO2014			
Original	1.0×	37.7	59.1
Tucker2-iter(vbmf, 0.7)	**1.2×**	**36.3(-1.4)**	**57.3(-1.8)**
MUSCO(vbmf, 0.7, 2)	**1.7×**	**36.2(-1.5)**	**57.1(-2.0)**
MUSCO(nx, 3, 4)	**1.8×**	**35.4(-2.3)**	**56.2(-2.9)**
Tucker2-iter(vbmf, 0.9)	2.0×	33.8(-3.9)	54.0(-5.1)

Table 5: Comparisom of Faster R-CNN (with ResNet-50 backbone) compressed models on COCO2014 dataset. MUSCO (vbmf, 0.7, 2) corresponds to the two-iteration compression with automatically selected ranks using GAS of EVBMF and rank weakening with weakeinig factor equals 0.7.