

Supplementary Material: Model Compression Using Optimal Transport

Suhas Lohit Michael Jones
Mitsubishi Electric Research Laboratories, Cambridge, MA, USA
slohit@merl.com mjones@merl.com

1. Number of parameters in the teacher and student networks used in the paper

Datasets	Teacher architectures	# weights ($\times 10^6$)	Student architectures	# weights ($\times 10^6$)	% reduction in # weights
CIFAR-100 / SVHN	resnet110	1.74	resnet20	0.28	83.97
	resnet56	0.86			70.00
	resnet32x4	7.43	resnet8x4	1.23	83.41
	WRN-40-2	2.25	WRN-16-2	0.70	68.81
	vgg13	9.46	vgg8	3.97	58.10
ImageNet	ResNet-34	21.28	ResNet-18	11.17	48.00

Table 1. The number of parameters in the networks used in this paper and the compression achieved for every teacher-student pair.