

1. Appendix

1.1. Pruned Structure Comparison

We included a diagram in the Appendix and also here(Fig.1) to compare the final pruned structure with and without our reorganization framework SPSRC. We conducted this analysis on ResNet50 on ImageNet. As seen, the final obtained structure is quite different in many layers.

1.2. Choice of Pruning Configuration

In order to study thoroughly which layer to prune and how much to prune in a certain layer, we analyze norm distribution for each layer after reorganization. Through our experiments, we found that layers that have more left-skewed norm distribution are more robust to pruning. For those more sensitive convolutional layers, we either skip them or decrease the pruning ratio when pruning the entire model. Figures 2, 3, 4 demonstrate the norm distribution of several example layers on three different models on CIFAR10.

MODEL	TOP-1% \uparrow	FLOPS(M) \downarrow	PARAMS(M) \downarrow
DENSE	93.72	260.57	1.73
L1	<u>93.3</u>	160.128	1.17
SPECTRAL(OURS)	93.56	160.128	1.17
NUCLEAR(OURS)	93.29	160.128	1.17
FROBENIUS(OURS)	93.12	160.128	1.17
GAL - 0.5	<u>92.55</u>	<u>134.19</u>	0.95
SPECTRAL (OURS)	<u>93.00</u>	113.76	0.96
NUCLEAR (OURS)	93.03	113.76	0.96
FROBENIUS (OURS)	92.95	113.76	0.96

Table 1. Results on CIFAR-10 with ResNet110. We separated the table with dashline based on compression ratios.

1.3. ResNet110

In Table 1, we demonstrated results for ResNet110 of two different pruning configurations, retrained for 50 and 80 epochs respectively after pruning. As shown in Table 1, SPSRC is higher in accuracy than all other methods with similar compression ratios with both pruning configurations.

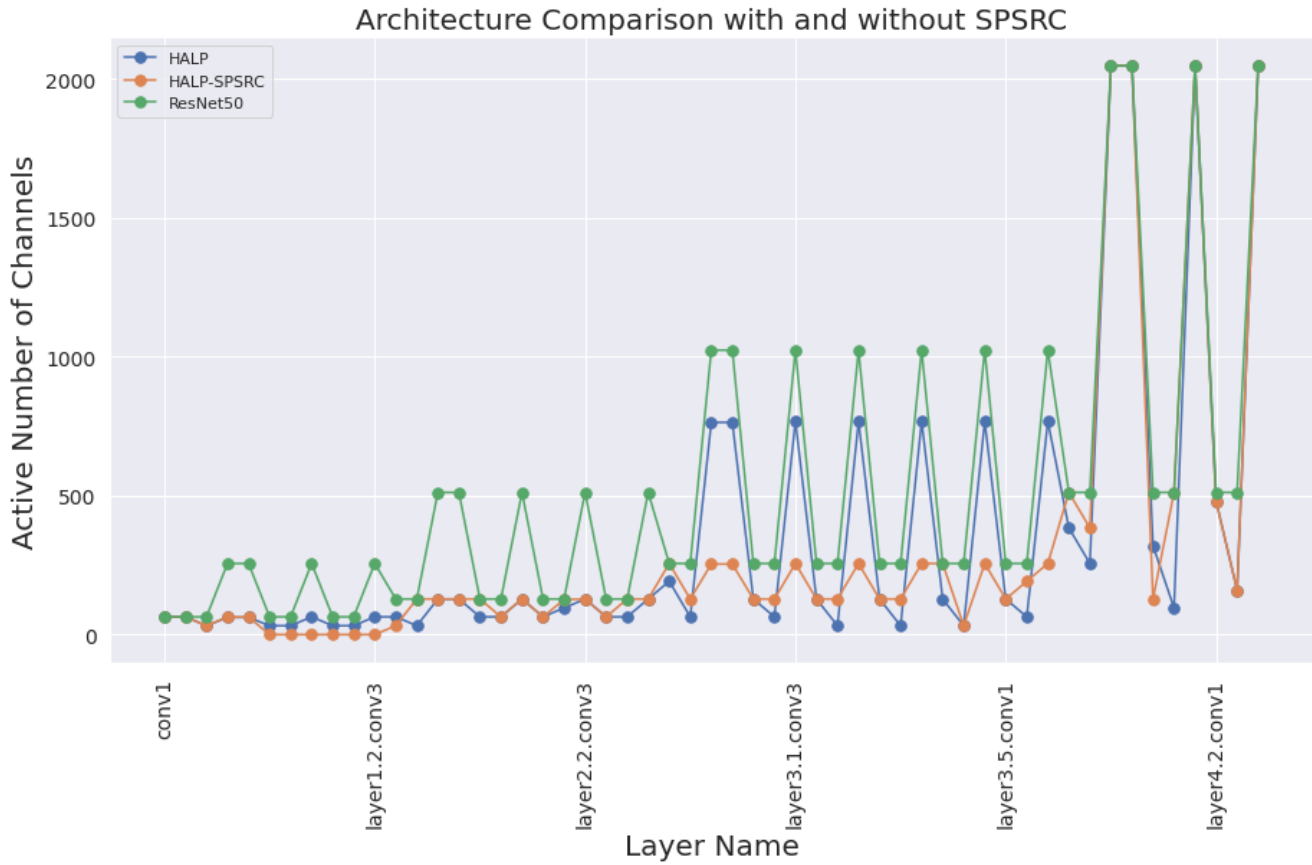


Figure 1. Pruned Structure Comparison on ResNet50-ImageNet

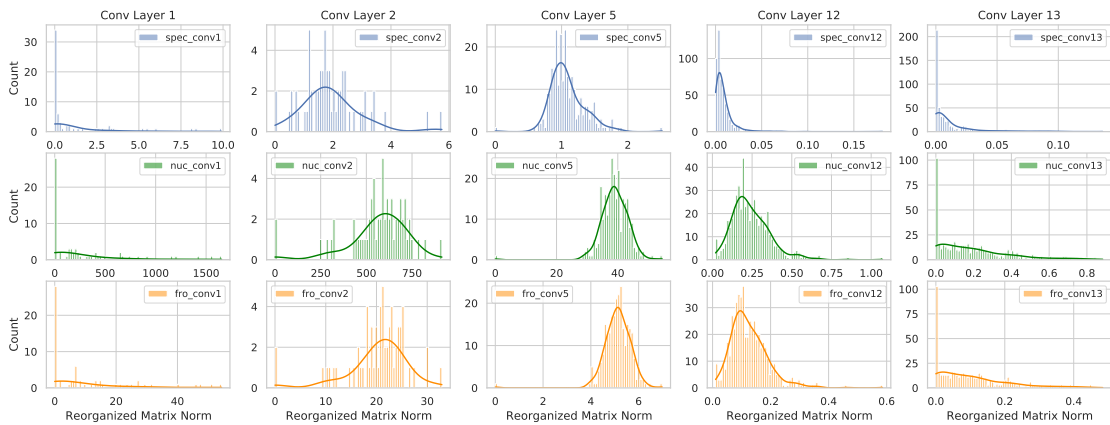


Figure 2. Norm histogram and density distribution of VGG16 on CIFAR10 on certain layers. As discussed above, layer 2 and layer 5 here are more sensitive to pruning thus are skipped when pruning the entire model.

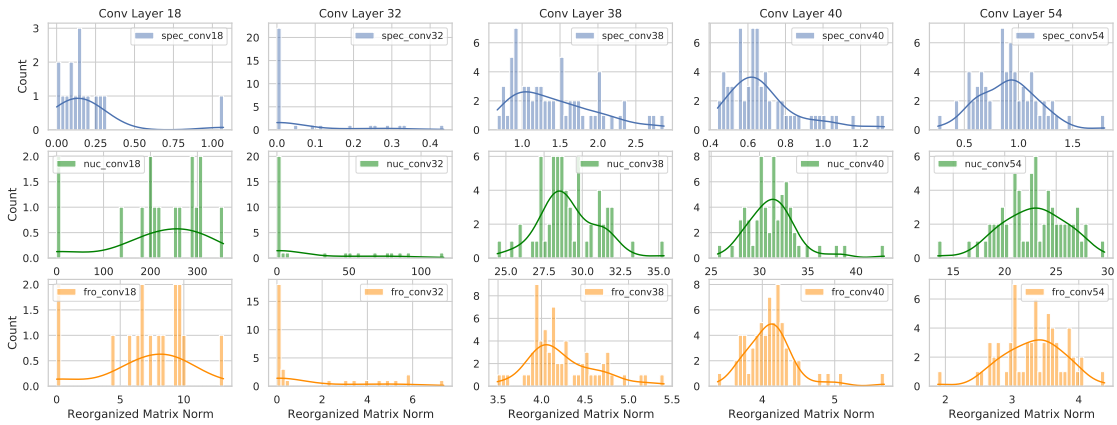


Figure 3. Norm histogram and density distribution of ResNet56 on CIFAR10 on certain layers. As discussed above, layer 18 and layer 54 here are more sensitive to pruning thus are skipped when pruning the entire model.

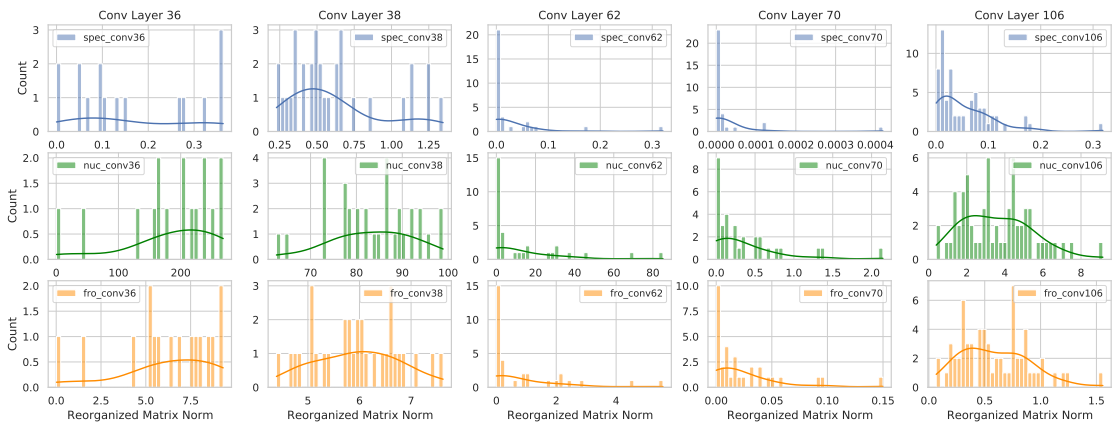


Figure 4. Norm histogram and density distribution of ResNet110 on CIFAR10 on certain layers. As discussed above, layer 36 and layer 38 here are more sensitive to pruning thus are skipped when pruning the entire model.