Supplementary: Torque based Structured Pruning for Deep Neural Network

1. Supplementary

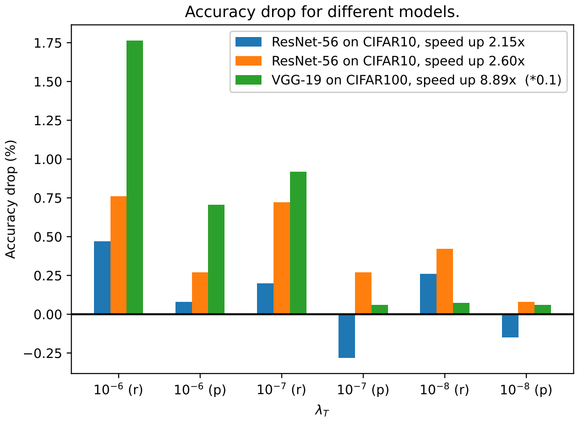

Figure 1. weight matrix of Layer 14 of VGG-19 trained on CIFAR-100. Each element in this matrix is a 2D kernel.

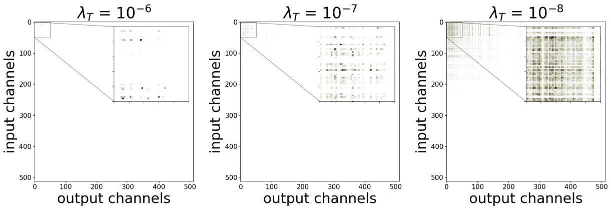

Figure 2. weight matrix of Layer 14 of VGG-19 trained on CIFAR-100. Each element in this matrix is a 2D kernel.

In Figure 1, we provide our analysis when training our model with different values of λ_T for both (r) and (p). Specifically, we experiment with three different $\lambda_T = 10^{-6}$, $\lambda_T = 10^{-7}$ and $\lambda_T = 10^{-8}$ for different model architectures at different speedup. (Note: we have scaled the values of VGG-19 on CIFAR100 by 0.1 in order for easier insight). As already mentioned in the main paper, all above models perform exceptionally well intialized with pre-trained model as compred to randomly initialized. Additionally, for most of the networks we observe $\lambda_T = 10^{-7}$ and $\lambda_T = 10^{-8}$ performance to be superior than $\lambda_T = 10^{-6}$.

We especially noticed that when the VGG-19 model is randomly initialized and trained with $\lambda_T = 10^{-6}$ on CIFAR-100 dataset, we see significant drop in performance. On further analysis of the VGG-19 layers trained with different λ_T, we observed at high value of λ_T, the constraint is too strong and the model is unable to recover from such strong damage. Figure 2 shows layer-14 of VGG-19 after torque training process.