Supplementary Material for Learning Filter Pruning Criteria for Deep Convolutional Neural Networks Acceleration

1 Different Filter Distribution

To comprehensively show the filter distribution of the network, we calculate the ℓ_2-norm of the filters of ResNet-18 on ILSVRC-2012. The kernel distribution estimate (KDE) of the norm distribution of all convolutional layers is shown in Figure 1. Clearly, the distribution of different layers is different from each other.

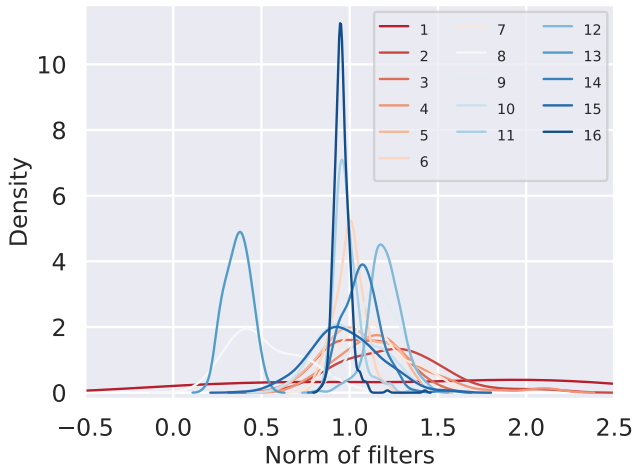

Figure 1: Norm distribution of all convolutional layers for ResNet-18 on ILSVRC-2012). The curves denote Kernel Distribution Estimate (KDE) of the norm distribution. The red and blue curves are layers near the input and the output of the network, respectively. The legend is the number of the index of the convolutional layers.

2 FLOP Loss

The pruning rate of every layer is sampled in the same way as Differentiable Criteria Sampler (DCS) in Sec.3.2.3. Several pruning rates are pre-defined for the sampling process, and the parameter for pruning rate sampling is denoted as β. The FLOP loss can be formulated as:

$$\mathcal{L}_{comp} = \begin{cases} \log\left(\mathbb{E}_{comp}(\theta, \beta)\right) & \mathbb{A}_{comp}(\theta, \beta) > (1+t) \times T \\ 0 & (1-t) \times T < \mathbb{A}_{comp}(\theta, \beta) < (1+t) \times T \\ -\log\left(\mathbb{E}_{comp}(\theta, \beta)\right) & \mathbb{A}_{comp}(\theta, \beta) < (1-t) \times T \end{cases} \quad (1)$$

where $\mathbb{E}_{comp}(\theta, \beta)$ is the expectation of the computation cost of the pruned network. We utilize the weighted sum of the sampling pruning rate as the *expected pruning rate* and use the *expected pruning rate* to calculate the expected computation cost. $\mathbb{A}_{comp}(\theta, \beta)$ indicates the actual computation cost of the pruned network. The pruning rate for calculating the actual computation cost is the sampled one. T is the target FLOPs and $t \in [0, 1]$ is the toleration ratio of the computational cost.

3 Criteria Visualization

The visualization of the pruning criteria for ResNet-110 is shown in 2. We find GM criteria is suitable for higher layers and ℓ_p-norm is suitable for lower layers.

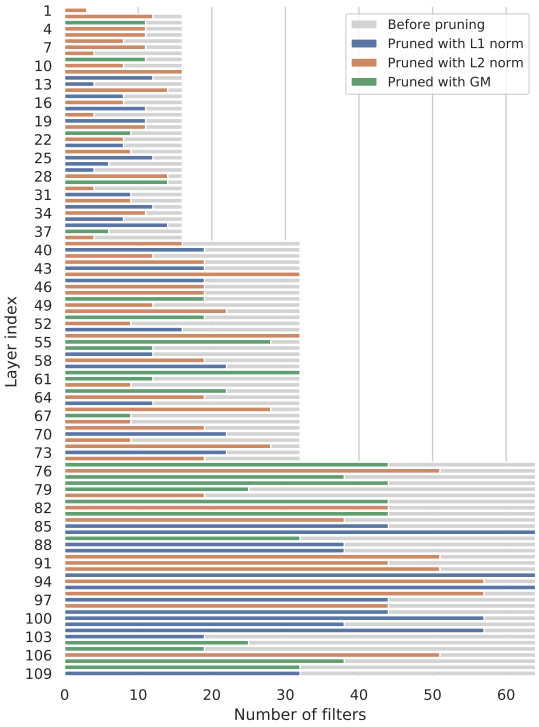

Figure 2: Visualization of the selected pruning criteria for each layer, and remaining filters for ResNet-110 on CIFAR-10.