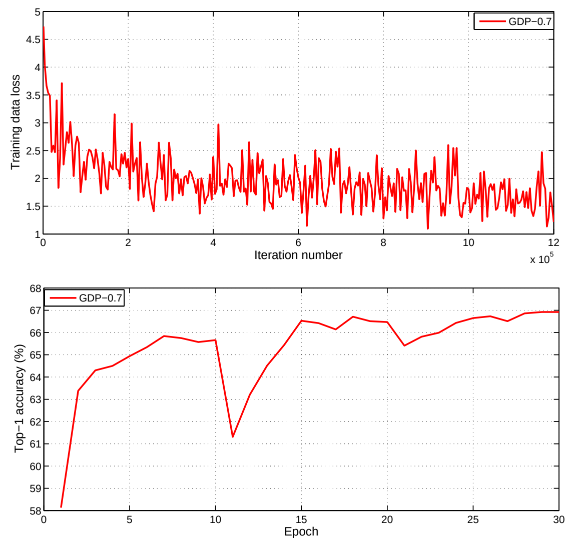

Figure 1: The training loss and testing Top-1 accuracy using GDP to prune VGG-16, when β is set to be 0.7.

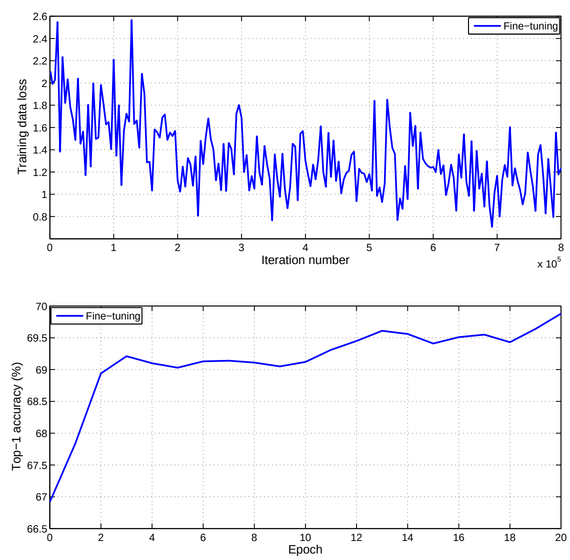

Figure 2: The training loss and testing Top-1 accuracy for fine-tuning the pruned network, which is pruned by GDP at the β setting to be 0.7.

Fig. 1 presents the changed process of training loss and testing Top-1 accuracy using GDP to prune VGG-16. After 30 epochs, we achieve 66.92% Top-1 accuracy. To further improve the classification accuracy of the pruned network by GDP, we fine-tuning the pruned network and achieve 69.88% Top-1 accuracy, which is shown in Fig. 2.

Table 1: FLOPs comparison of GDP and GDP-D, when β is set to be 0.7. FLOPs% is the percentage of the remaining FLOPs. The Top-1 accuracy is 68.87% using GDP-D and 69.88% using GDP.

Model	Layer	FLOPs	FLOPs% GDP-D	FLOPs% GDP
VGG-16	Conv1_1	89.91M	56.25%	56.25%
	Conv1_2	1.85B	33.44%	42.24%
	Conv2_1	926.45M	32.97%	41.63%
	Conv2_2	1.85B	54.21%	54.21%
	Conv3_1	925.65M	51.12%	51.50%
	Conv3_2	1.85B	51.55%	52.75%
	Conv3_3	1.85B	98.44%	98.05%
	Conv4_1	925.25M	58.79%	49.02%
	Conv4_2	1.85B	35.94%	12.60%
	Conv4_3	1.85B	38.69%	12.56%
	Conv5_1	462.52M	46.73%	42.52%
	Conv5_2	462.52M	55.23%	79.35%
	Conv5_3	462.52M	50.56%	87.52%
	Total	15.36B	51.16%	48.03%

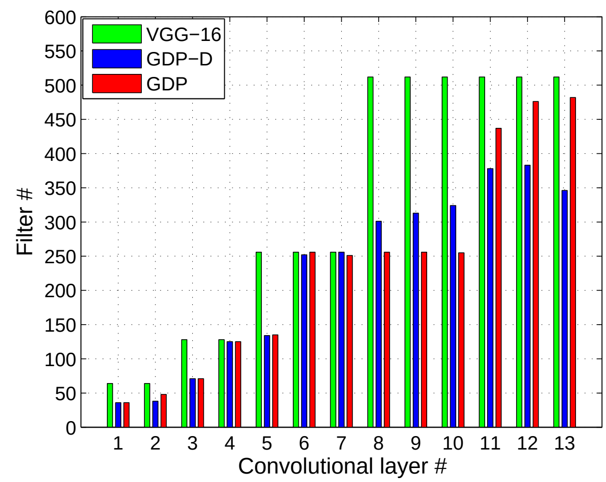

Figure 3: The number of remaining filters in each layer when using GDP-D and GDP to prune VGG-16 at the β setting to be 0.7.

Tab. 1 shows the results of pruning filters in each layer using GDP and GDP-D (*i.e.*, global pruning without dynamic updating). Comparing to GDP-D, GDP not only achieves a higher Top-1 accuracy (69.88% *vs.* 68.87%) and also tends to prune more filters in the layers with high computation complexity to reduce more total FLOPs, when 30% filters are pruned (*i.e.*, β is set to be 0.7). More specifically, as shown in Tab. 1 and Fig. 3, GDP tends to prune more filters on the middle layers (*e.g.*, Conv4_1, Conv4_2 and Conv4_3), while GDP-D tends to prune more filters on the last layers (*e.g.*, Conv5_2, Conv5_3).