

## A Results

Table 5: Results of iterative  $\ell_1$ -**norm Filters Pruning** [26] on CIFAR-10 and CIFAR-100 datasets. The *SLR* column presents the result of pruned networks finetuned with small learning rate while *LLR* column shows the results of same networks finetuned with large learning rate.

| Model                      | #Param | % MACs(G)↓ | C10-SLR          | C10-LLR          | C100-SLR          | C100-LLR         |
|----------------------------|--------|------------|------------------|------------------|-------------------|------------------|
| Resnet-110 (baseline)      | 1.73M  | 0.00       | 94.01            | 94.01            | 72.35             | 72.35            |
| Resnet-110 #1              | 1.26M  | 23.08      | 93.51 $\pm$ 0.05 | 93.58 $\pm$ 0.05 | 69.70 $\pm$ 0.17  | 71.67 $\pm$ 0.03 |
| Resnet-110 #2              | 0.93M  | 38.46      | 92.91 $\pm$ 0.02 | 93.49 $\pm$ 0.05 | 68.47 $\pm$ 0.20  | 70.97 $\pm$ 0.43 |
| Resnet-110 #3              | 0.70M  | 50.00      | 92.65 $\pm$ 0.02 | 93.53 $\pm$ 0.12 | 66.89 $\pm$ 0.14  | 70.59 $\pm$ 0.19 |
| Resnet-110 #4              | 0.52M  | 57.69      | 92.11 $\pm$ 0.01 | 93.62 $\pm$ 0.03 | 66.19 $\pm$ 0.07  | 70.26 $\pm$ 0.26 |
| Resnet-110 #5              | 0.39M  | 65.38      | 91.51 $\pm$ 0.08 | 93.24 $\pm$ 0.16 | 65.44 $\pm$ 0.04  | 69.54 $\pm$ 0.07 |
| Resnet-110 Ensemble        | -      | -          | 93.82            | 94.01 $\pm$ 0.22 | 73.32             | 75.33 $\pm$ 0.14 |
| Resnet-56 (baseline)       | 0.85M  | 0.00       | 93.42            | 93.42            | 71.07             | 71.07            |
| Resnet-56 #1               | 0.66M  | 23.07      | 92.73 $\pm$ 0.06 | 93.23 $\pm$ 0.18 | 69.10 $\pm$ 0.15  | 69.83 $\pm$ 0.09 |
| Resnet-56 #2               | 0.52M  | 30.77      | 92.24 $\pm$ 0.17 | 93.21 $\pm$ 0.08 | 67.92 $\pm$ 0.05  | 69.58 $\pm$ 0.13 |
| Resnet-56 #3               | 0.42M  | 46.15      | 91.64 $\pm$ 0.19 | 92.90 $\pm$ 0.16 | 66.76 $\pm$ 0.03  | 69.50 $\pm$ 0.25 |
| Resnet-56 #4               | 0.35M  | 53.85      | 91.10 $\pm$ 0.27 | 92.74 $\pm$ 0.22 | 65.205 $\pm$ 0.05 | 69.49 $\pm$ 0.08 |
| Resnet-56 #5               | 0.29M  | 61.54      | 90.35 $\pm$ 0.36 | 92.31 $\pm$ 0.26 | 64.91 $\pm$ 0.14  | 69.03 $\pm$ 0.24 |
| Resnet-56 Ensemble         | -      | -          | 93.25            | 94.29 $\pm$ 0.02 | 71.37             | 74.23 $\pm$ 0.21 |
| VGG-16 (baseline)          | 14.99M | 0.00       | 94.23            | 94.23            | 73.24             | 73.24            |
| VGG-16 #1                  | 9.46M  | 0.00       | 94.13 $\pm$ 0.09 | 93.95 $\pm$ 0.02 | 71.39 $\pm$ 0.06  | 72.38 $\pm$ 0.11 |
| VGG-16 #2                  | 6.27M  | 0.00       | 94.09 $\pm$ 0.13 | 93.90 $\pm$ 0.03 | 70.48 $\pm$ 0.07  | 72.10 $\pm$ 0.1  |
| VGG-16 #3                  | 4.43M  | 0.00       | 94.09 $\pm$ 0.04 | 93.93 $\pm$ 0.04 | 69.73 $\pm$ 0.06  | 72.28 $\pm$ 0.11 |
| VGG-16 #4                  | 3.36M  | 0.00       | 94.03 $\pm$ 0.13 | 93.89 $\pm$ 0.10 | 69.09 $\pm$ 0.05  | 72.22 $\pm$ 0.19 |
| VGG-16 #5                  | 2.71M  | 0.00       | 93.88 $\pm$ 0.12 | 94.10 $\pm$ 0.09 | 68.37 $\pm$ 0.09  | 71.95 $\pm$ 0.04 |
| VGG-16 Ensemble            | -      | -          | 94.29            | 95.04 $\pm$ 0.07 | 72.86 $\pm$ 0.02  | 75.93 $\pm$ 0.06 |
| PreResnet-164 (baseline)   | 1.7M   | 0.00       | 95.06            | 95.06            | 76.35             | 76.35            |
| PreResnet-164 #1           | 1.09M  | 26.92      | 94.43 $\pm$ 0.06 | 94.92 $\pm$ 0.05 | 74.65 $\pm$ 0.04  | 76.20 $\pm$ 0.07 |
| PreResnet-164 #2           | 0.74M  | 46.15      | 93.71 $\pm$ 0.07 | 94.74 $\pm$ 0.14 | 73.17 $\pm$ 0.03  | 75.87 $\pm$ 0.03 |
| PreResnet-164 #3           | 0.54M  | 57.69      | 93.50 $\pm$ 0.01 | 94.66 $\pm$ 0.19 | 71.89 $\pm$ 0.01  | 75.15 $\pm$ 0.16 |
| PreResnet-164 #4           | 0.4M   | 65.38      | 92.61 $\pm$ 0.02 | 94.69 $\pm$ 0.17 | 70.50 $\pm$ 0.09  | 74.03 $\pm$ 0.68 |
| PreResnet-164 #5           | 0.31M  | 69.23      | 92.06 $\pm$ 0.11 | 94.15 $\pm$ 0.06 | 69.20 $\pm$ 0.04  | 73.99 $\pm$ 0.06 |
| PreResnet-164 Ensemble     | -      | -          | -                | 95.60 $\pm$ 0.04 | -                 | 79.19 $\pm$ 0.07 |
| WideResnet-16-8 (baseline) | 11.01  | 0.00       | 95.62            | 95.62            | 79.57             | 79.57            |
| WideResnet-16-8 #1         | 8.01   | 20.00      | 95.36 $\pm$ 0.01 | 95.18 $\pm$ 0.1  | 78.52 $\pm$ 0.04  | 78.19 $\pm$ 0.19 |
| WideResnet-16-8 #2         | 5.89   | 35.48      | 95.20 $\pm$ 0.02 | 95.25 $\pm$ 0.14 | 77.46 $\pm$ 0.14  | 77.81 $\pm$ 0.2  |
| WideResnet-16-8 #3         | 4.38   | 47.74      | 94.95 $\pm$ 0.01 | 95.08 $\pm$ 0.16 | 76.29 $\pm$ 0.02  | 77.43 $\pm$ 0.36 |
| WideResnet-16-8 #4         | 3.28   | 57.42      | 94.97 $\pm$ 0.01 | 95.08 $\pm$ 0.08 | 74.73 $\pm$ 0.03  | 76.95 $\pm$ 0.21 |
| WideResnet-16-8 #5         | 2.48   | 64.52      | 94.61 $\pm$ 0.09 | 94.91 $\pm$ 0.04 | 73.82 $\pm$ 0.1   | 76.46 $\pm$ 0.27 |
| WideResnet-16-8 Ensemble   | -      | -          | 95.63            | 95.79 $\pm$ 0.08 | 79.22 $\pm$ 0.03  | 80.45 $\pm$ 0.14 |

Table 6: Results of iterative **Weights pruning** [14] on CIFAR-10 and CIFAR-100 datasets. The *SLR* column presents the result of pruned networks finetuned with small learning rate while *LLR* column shows the results of same networks finetuned with large learning rate.

| Model                     | #Active Params (M) | C10-SLR          | C10-LLR          | C100-SLR         | C100-LLR         |
|---------------------------|--------------------|------------------|------------------|------------------|------------------|
| Resnet-56 (baseline)      | 0                  | 93.42            | 93.42            | 71.07            | 71.07            |
| Resnet-56 #1              | 0.66               | 93.15 $\pm$ 0.02 | 93.36 $\pm$ 0.11 | 70.95 $\pm$ 0.05 | 70.40 $\pm$ 0.12 |
| Resnet-56 #2              | 0.52               | 93.18 $\pm$ 0.05 | 93.35 $\pm$ 0.17 | 70.78 $\pm$ 0.01 | 70.58 $\pm$ 0.19 |
| Resnet-56 #3              | 0.42               | 93.03 $\pm$ 0.02 | 93.36 $\pm$ 0.12 | 70.31 $\pm$ 0.06 | 70.51 $\pm$ 0.13 |
| Resnet-56 #4              | 0.35               | 92.78 $\pm$ 0.01 | 93.43 $\pm$ 0.08 | 69.91 $\pm$ 0.01 | 70.75 $\pm$ 0.04 |
| Resnet-56 #5              | 0.29               | 92.69 $\pm$ 0.02 | 93.41 $\pm$ 0.08 | 69.53 $\pm$ 0.07 | 70.46 $\pm$ 0.3  |
| Resnet-56 Ensemble        | -                  | 93.39            | 94.15 $\pm$ 0.04 | 71.18 $\pm$ 0.08 | 72.69 $\pm$ 0.02 |
| Resnet-110 (baseline)     | 0                  | 94.01            | 94.01            | 72.35            | 72.35            |
| Resnet-110 #1             | 1.27               | 93.90 $\pm$ 0.08 | 93.65 $\pm$ 0.1  | 72.30 $\pm$ 0.07 | 72.11 $\pm$ 0.24 |
| Resnet-110 #2             | 0.94               | 93.73 $\pm$ 0.05 | 93.86 $\pm$ 0.1  | 71.75 $\pm$ 0.02 | 72.09 $\pm$ 0.06 |
| Resnet-110 #3             | 0.69               | 93.65 $\pm$ 0.03 | 93.89 $\pm$ 0.03 | 70.96 $\pm$ 0.04 | 72.38 $\pm$ 0.25 |
| Resnet-110 #4             | 0.50               | 93.32 $\pm$ 0.07 | 93.79 $\pm$ 0.01 | 70.69 $\pm$ 0.02 | 72.09 $\pm$ 0.05 |
| Resnet-110 #5             | 0.36               | 93.02 $\pm$ 0.04 | 93.69 $\pm$ 0.17 | 68.90 $\pm$ 0.08 | 71.59 $\pm$ 0.30 |
| Resnet-110 Ensemble       | -                  | 93.98 $\pm$ 0.04 | 94.56 $\pm$ 0.04 | 72.51 $\pm$ 0.06 | 74.19 $\pm$ 0.02 |
| WideResnet-110 (baseline) | 11.01              | 95.62            | 95.62            | 79.57            | 79.57            |
| WideResnet-110 #1         | 8.05               | 95.55 $\pm$ 0.02 | 95.32 $\pm$ 0.03 | 79.32 $\pm$ 0.07 | 78.75 $\pm$ 0.15 |
| WideResnet-110 #2         | 5.97               | 95.60 $\pm$ 0.03 | 95.46 $\pm$ 0.07 | 79.17 $\pm$ 0.07 | 78.74 $\pm$ 0.18 |
| WideResnet-110 #3         | 4.44               | 95.65 $\pm$ 0.05 | 95.30 $\pm$ 0.02 | 77.86 $\pm$ 0.02 | 78.84 $\pm$ 0.13 |
| WideResnet-110 #4         | 3.34               | 95.65 $\pm$ 0.02 | 95.37 $\pm$ 0.05 | 78.02 $\pm$ 0.19 | 78.81 $\pm$ 0.17 |
| WideResnet-110 #5         | 2.53               | 95.47 $\pm$ 0.07 | 95.55 $\pm$ 0.05 | 77.92 $\pm$ 0.16 | 78.82 $\pm$ 0.11 |
| WideResnet-110 Ensemble   | -                  | 95.62 $\pm$ 0.03 | 95.86 $\pm$ 0.08 | 79.62 $\pm$ 0.28 | 80.26 $\pm$ 0.11 |

Table 7: Results of iterative  $\ell_1$ -**norm Filters Pruning** [26] on Tiny-Imagenet dataset. The *SLR* column presents the result of pruned networks finetuned with small learning rate while *LLR* column shows the results of same networks finetuned with large learning rate.

| Model                | #Param | SLR              | LLR              |
|----------------------|--------|------------------|------------------|
| Resnet-18 (baseline) | 11.01  | 67.22            | 67.22            |
| Resnet-18 #1         | 8.30   | 65.66 $\pm$ 0.21 | 66.46 $\pm$ 0.23 |
| Resnet-18 #2         | 6.17   | 64.04 $\pm$ 0.17 | 65.81 $\pm$ 0.14 |
| Resnet-18 #3         | 4.64   | 63.59 $\pm$ 0.10 | 65.37 $\pm$ 0.07 |
| Resnet-18 #4         | 3.52   | 61.91 $\pm$ 0.12 | 64.75 $\pm$ 0.13 |
| Resnet-18 #5         | 2.71   | 61.06 $\pm$ 0.32 | 64.70 $\pm$ 0.33 |
| Resnet-18 Ensemble   | -      | 67.63 $\pm$ 0.21 | 69.30 $\pm$ 0.11 |
| Resnet-34 (baseline) | 21.39  | 68.81            | 68.81            |
| Resnet-34 #1         | 15.97  | 68.18 $\pm$ 0.06 | 68.66 $\pm$ 0.18 |
| Resnet-34 #2         | 12.02  | 67.10 $\pm$ 0.10 | 68.20 $\pm$ 0.02 |
| Resnet-34 #3         | 9.13   | 66.41 $\pm$ 0.07 | 67.90 $\pm$ 0.12 |
| Resnet-34 #4         | 6.99   | 66.05 $\pm$ 0.19 | 67.02 $\pm$ 0.24 |
| Resnet-34 #5         | 5.40   | 64.93 $\pm$ 0.14 | 67.26 $\pm$ 0.08 |
| Resnet-34 Ensemble   | -      | 69.88 $\pm$ 0.11 | 71.31 $\pm$ 0.13 |