The Myth of the Pyramid

Supplementary Material

A. Filters by Layer for Models using Templates

In order to facilitate the reproducibility of experiments, we present in this section the values for each layer obtained by applying templates to the four models tested in our work in Section 3. In the case of ResNet we set the value of filters at the level of each layer inside residual modules, as presented in Table 10. VGG design consists of simple layers, so we change filters in each of them (Table 7). For the rest of architectures, we set the value of filters at the level of modules (Tables 8 and 9). The last layers of all models are fully connected, and the dataset imposes the number of neurons. We add the schematics of the templates as a visual aid in figure 2.

Figure 2. Schematic distribution of filters per layer in templates. Base distribution is the original pyramidal distribution. Templates follow a smoother transition in the number of filters between layers. Number of filters does not match between models but they are adjusted to fit the original number of FLOPS of the base model.

Table 7. VGG19 with the original distribution of filters and five templates. All models count similar number of FLOPs.

Template	Filter Values
Base (Original values)	64, 64, 128, 128, 256, 256, 256, 256, 512, 512, 512, 512, 512, 512, 512, 512
a	64, 99, 133, 168, 203, 238, 272, 307, 342, 377, 411, 446, 481, 516, 550, 585
b	153, 153, 153, 153, 153, 153, 153, 153, 153, 153, 153, 153, 153, 153, 153, 153
c	165, 158, 152, 145, 138, 131, 125, 118, 111, 104, 98, 91, 84, 77, 71, 64
d	64, 105, 146, 186, 227, 268, 308, 349, 390, 343, 297, 250, 204, 157, 111, 64
e	175, 161, 147, 133, 120, 106, 92, 78, 64, 80, 96, 112, 127, 143, 159, 175

Table 8. Original distribution of filters for MobileNet2 after applying five templates.

Template	Filter Values
Base (Original values)	16, 24, 32, 64, 96, 160, 320, 1280
a	16, 36, 56, 76, 97, 117, 137, 157
b	61, 61, 61, 61, 61, 61, 61, 61
c	77, 68, 60, 51, 42, 33, 25, 16
d	16, 38, 59, 80, 102, 73, 45, 16
e	92, 73, 54, 35, 16, 41, 67, 92

Table 9. Distribution of filters for MNASNet showing the original design and filters from five templates.

Template	Filter Values
Base (Original values)	32, 16, 24, 40, 80, 96, 192, 320, 1280
a	32, 46, 60, 73, 87, 101, 114, 128, 142
b	76, 76, 76, 76, 76, 76, 76, 76, 76
c	102, 93, 84, 76, 67, 58, 50, 41, 32
d	32, 49, 66, 84, 101, 84, 66, 49, 32
e	141, 114, 86, 59, 32, 59, 86, 114, 141

Table 10. Original distribution of filters for ResNet50 and five templates. All models count similar number of FLOPs. Filter redistribution is made at the lever of layers within modules. Expansion layers within modules in the same block are kept with equal filters to fit residual connections.

Template	Filter Values
Base (Original values)	64, [[64,64,256], [64,64,256], [64,64,256]], [[128,128,512], [128,128,512], [128,128,512], [128,128,512]], [[256,256,1024], [256,256,1024], [256,256,1024], [256,256,1024], [256,256,1024], [256,256,1024]], [[512,512,2048], [512,512,2048], [512,512,2048]]
a	64, [[64,73,256], [83,92,256], [102,111,256]], [[120,130,480], [139,148,480], [158,167,480], [177,186,480]], [[195,205,780], [214,224,780], [233,242,780], [252,261,780], [271,280,780], [289,299,780]], [[308,317,1232], [327,336,1232], [346,355,1232]]
b	64, [[123,123,492], [123,123,492], [123,123,492]], [[123,123,492], [123,123,492], [123,123,492], [123,123,492]], [[123,123,492], [123,123,492], [123,123,492], [123,123,492], [123,123,492], [123,123,492]], [[123,123,492], [123,123,492], [123,123,492]]
c	64, [[134,132,536], [129,127,536], [125,123,536]], [[120,118,480], [116,114,480], [111,109,480], [107,105,480]], [[102,100,408], [98,96,408], [93,91,408], [89,87,408], [84,82,408], [80,78,408]], [[75,73,300], [71,69,300], [66,64,300]]
d	64, [[64,76,256], [88,99,256], [111,123,256]], [[134,146,536], [158,170,536], [182,193,536], [205,217,536]], [[228,240,912], [252,239,912], [227,214,912], [202,189,912], [177,164,912], [152,139,912]], [[127,114,508], [102,89,508], [77,64,508]]
e	64, [[144,139,576], [134,129,576], [124,119,576]], [[114,109,456], [104,99,456], [94,89,456], [84,79,456]], [[74,69,296], [64,69,296], [75,80,296], [85,91,296], [96,101,296], [107,112,296]], [[117,123,468], [128,133,468], [139,144,468]]

B. Accuracy of Models After Applying Templates

This appendix presents plots for the experimental results related to Section 4 including four CNN architectures (VGG, ResNet, MobileNetV2 and MnasNet) evaluated on six different datasets. Each architecture's original filter distribution was transformed using the proposed templates. Final models obtained from a particular architecture count similar FLOPs for making fair comparisons. Once the models are changed, they are trained from scratch. We observe that a different template provides the best improvement for each pair model-task. The difference between templates is more notorious when comparing models belonging to different categories (classical or highly-optimised).

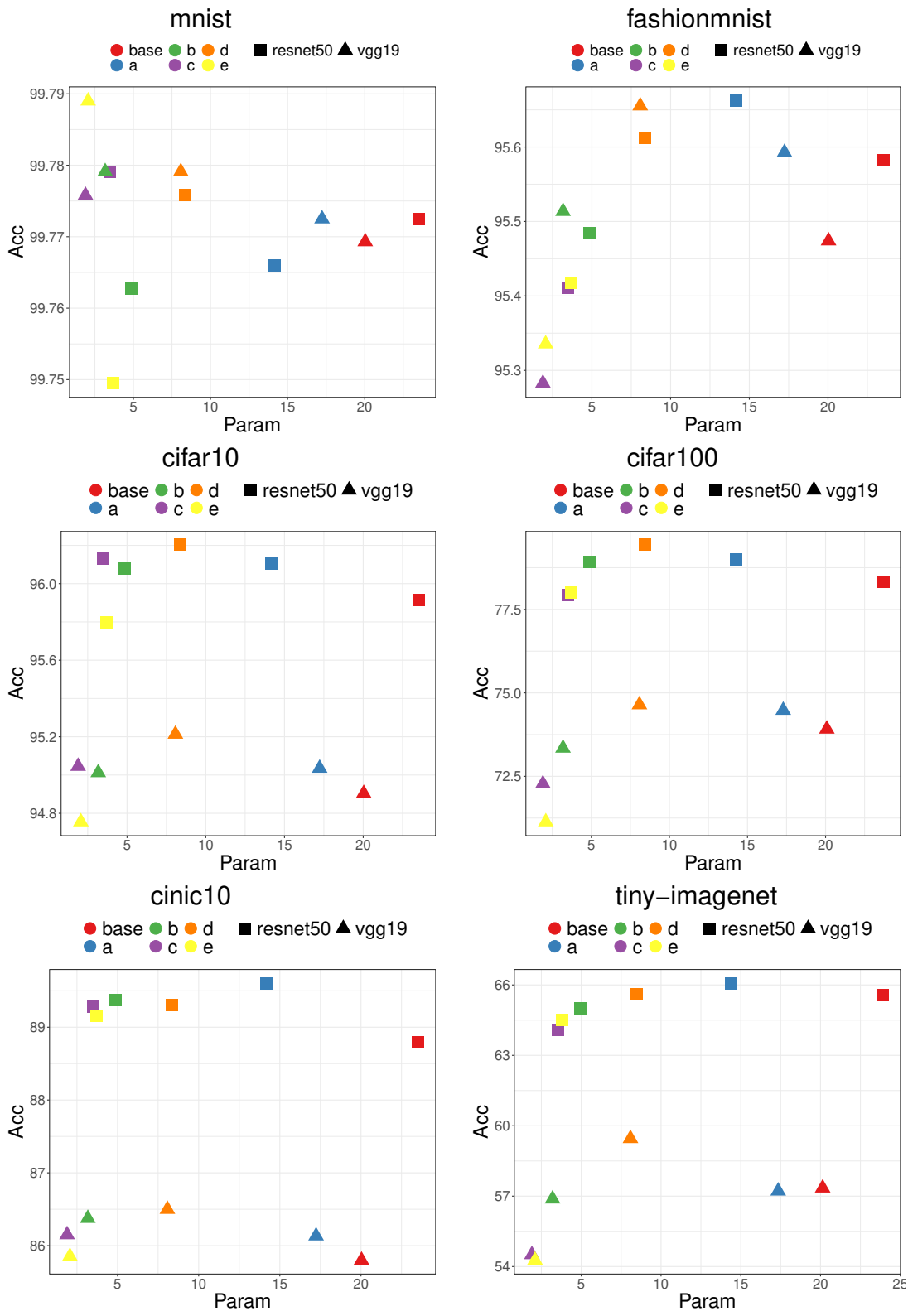

Figure 3. Accuracy of VGG and ResNet models after applying templates reported for several datasets. Base is the original distribution of filters. In many cases, templates outperform the base architecture. However, all of them use much fewer parameters than the base model. Note that models produced with templates from VGG have less than a third of FLOPs of those produced from ResNet.

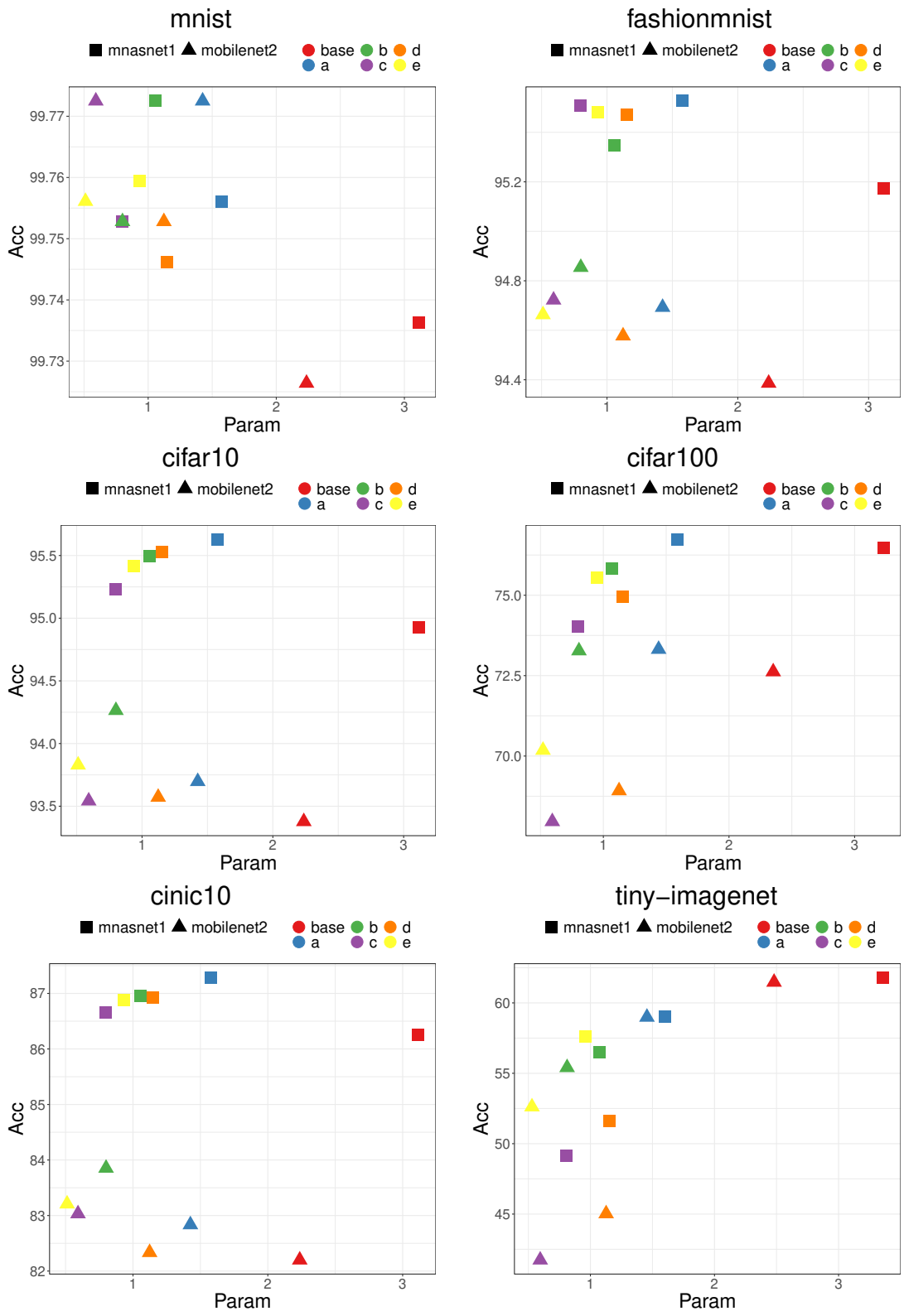

Figure 4. Parameter efficiency of MobileNetV2 and MnasNet models with templates reported for several datasets. Base is the original distribution of filters. In many cases, templates outperform the base architecture. However, all of them use much fewer parameters than the base model. Note that models produced with templates from MnasNet have more than 4X FLOPs of those produced from MobileNetV2.