Appendix - You Look Twice: GaterNet for Dynamic Filter Selection in CNNs

1 Distribution of Gates

The amounts of filters in different residual units of ResNet-164 are different. To give a complete view of the gates that our gater is learning, we plot the same figure as Figure 2 in our paper with the number of gates on the Y-axis in Figure 1 below. A gate is an entry in the binary gate vector g. It corresponds to a filter in the backbone network ResNet-164. A gate is *always off* means that it is 0 for all the samples in the test set. A gate is *always on* means that it is 1 for all the test samples. And a gate is *input-dependent* means that it is 1 for some of the test samples and 0 for the others.
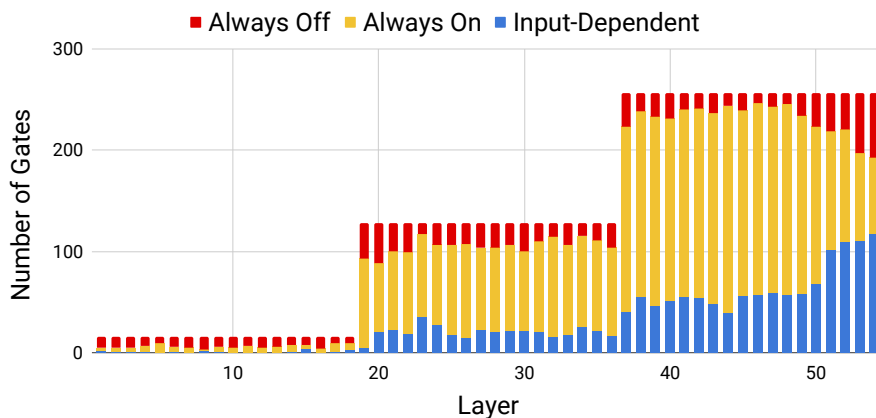
Figure 1: Number of the three type of gates for different layers in the backbone ResNet-164 of ResNet-164-Gated on Cifar-10 Test Set. There are totally 54 residual units in ResNet-164. Better viewed in color.

2 Scheduled Dropout

In the experiments of DenseNet-Gated, Shake-Shake-Gated and Inception-v4-Gated, scheduled dropout [1] similar to ScheduledDropPath in [2] are applied to the gate vector g. We start from a dropout rate of 0.0 and increase it gradually during training. The dropout rate reaches 0.05 at the end of training.

References

[1] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014.

[2] Barret Zoph, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le. Learning transferable architectures for scalable image recognition. In *CVPR*, June 2018.