TRAINING A 5000×32×32×3 RGB DATASET ON NVIDIA TESLA V100 IN 10 MINUTES

Morteza Hosseini, Bharat Prakash, Hamed Pirsiavash, and Tinoosh Mohsenin
University of Maryland, Baltimore County, MD, USA

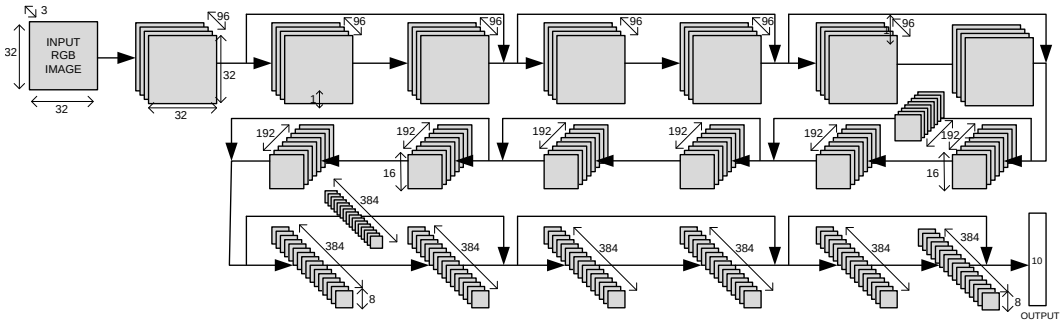

Figure 1: ResNet-20 architecture with filters increased by 6×.

1 APPROACH

- Neural Network Architecture: A ResNet-20 architecture (He et al., 2016) with filters augmented by 6× as depicted in Fig. 1 is adopted. Dropout layers (Srivastava et al., 2014) are inserted in between layers to minimize overfitting.
- Learning Rate: is scheduled such that, regardless of the machine that the CNN is being trained on, for the first 8 minutes, lr = 0.001, for the the next minute lr = 0.0001, and for the last minute lr = 0.00001. Batch size of 32 and ADAM optimizer are adopted.
- Data Augmentation: has been inspired by the works of Huang et al. (2017) that begins with aggressive vertical/horizontal shifts of 20%, zoom range of 20% and shear and rotation range of 15 degrees. All of the augmentation parameters gradually decrease through the training time until they all reach zero at the final epochs. For the filling mode of the margins of the shifted and rotated images, all 4 types of the *nearest*, *constant*, *wrap*, and *reflect* are randomly chosen at different epochs.
- System Specs: an Nvidia V100 GPU along with an Intel Broadwell CPU, Python 3.7, TensorFlow-gpu 2.4.1, CUDA 11.0.
- Best result: trained on a 10% subset of CIFAR-10 training set in 10 minutes, the best accuracy on the total test set of the CIFAR-10 is achieved to be 85.1%.

REFERENCES

Kaiming He et al. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on CVPR*, 2016.

Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.