Supplementary material for "SADT: Combining Sharpness-Aware Minimization with Self-Distillation for Improved Neural Network Training"

Anonymous Author(s)
Affiliation
Address
email

1 Experiments

In this section, we present training behaviour for the VGG network and the custom network. In both case, we use the CIFAR100 dataset with batch size 2048 and each experiment runs for 200 epochs.
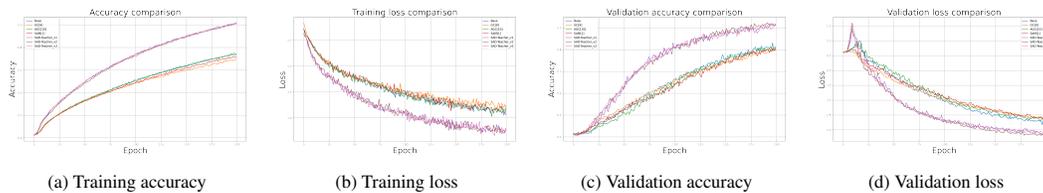

(a) Training accuracy   (b) Training loss   (c) Validation accuracy   (d) Validation loss

Figure 1: Training time comparison example: VGG model, CIFAR100, batch size 2048.

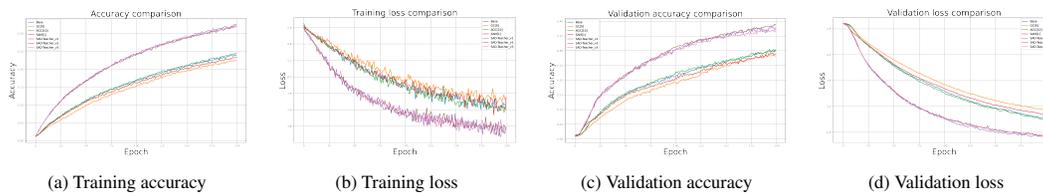(a) Training accuracy   (b) Training loss   (c) Validation accuracy   (d) Validation loss

Figure 2: Training time comparison example: Simple CNN model, CIFAR100, batch size 2048.