

## A TRAINING ENHANCEMENTS ON EFFICIENTNET

The default settings of our experiments follow the training recipes from ResNet, which may not be optimal for EfficientNet training in Section 5.2. To this end, we import the following settings to our experiments as in the original EfficientNet training setups (Tan & Le, 2019): we change weight decay from  $1e-4$  to  $1e-5$ , and add Dropout (Srivastava et al., 2014), stochastic depth (Huang et al., 2016) and AutoAugment (Cubuk et al., 2019) to regularize the training process. Besides, we train models longer (*i.e.*, 200 epochs) to better cope with these training enhancements, and adopt the early stopping strategy to prevent the catastrophic overfitting issue in robustness (Wong et al., 2020). With these training enhancements, our EfficientNet-L1 gets further improved, *i.e.*, +1.7% for accuracy (from 80.5% to 82.2%) and +0.6% for robustness (from 58.0% to 58.6%).