Table 1: Our training routines (under the double line) exceed previous SOTA or improve existing methods when combined. Accuracy on Tiny-ImageNet-LT (0.1) and iNaturalist, using SwinV2 and ConvNeXt. The error term corresponds to one standard error over 5 trials.

Method	Tiny-ImageNet (0.1) SwinV2	Tiny-ImageNet (0.1) ConvNeXt	iNaturalist SwinV2	iNaturalist ConvNeXt
ERM	52.2 ± 0.2	52.0 ± 0.3	68.4 ± 0.2	68.1 ± 0.1
ERM + Batch	52.4 ± 0.1	52.4 ± 0.2	68.8 ± 0.3	68.5 ± 0.1
ERM + Aug	53.1 ± 0.3	52.8 ± 0.1	70.8 ± 0.2	70.2 ± 0.3
ERM - Tuned	53.4 ± 0.2	53.1 ± 0.3	71.2 ± 0.3	70.9 ± 0.2
Reweighting	52.8 ± 0.4	52.3 ± 0.1	69.5 ± 0.4	69.3 ± 0.2
Resampling	52.5 ± 0.3	52.1 ± 0.2	69.0 ± 0.3	68.8 ± 0.1
Focal Loss	53.5 ± 0.1	53.1 ± 0.4	70.9 ± 0.4	70.8 ± 0.3
LDAM-DRW	54.2 ± 0.2	53.4 ± 0.3	71.5 ± 0.2	71.3 ± 0.3
M2m	54.3 ± 0.4	53.9 ± 0.2	72.5 ± 0.2	72.1 ± 0.4
MiSLAS	54.1 ± 0.3	53.4 ± 0.1	72.8 ± 0.1	72.4 ± 0.3
SAM-A	54.7 ± 0.4	53.9 ± 0.2	72.1 ± 0.3	72.2 ± 0.4
Joint-SSL	54.3 ± 0.2	53.7 ± 0.3	72.0 ± 0.2	71.7 ± 0.1
Joint-SSL + SAM-A + Smoothing	54.8 ± 0.1	54.1 ± 0.4	73.1 ± 0.4	72.6 ± 0.3
Joint-SSL + SAM-A + M2m	55.0 ± 0.2	54.3 ± 0.3	73.1 ± 0.1	72.9 ± 0.2

Table 2: **SAM-A, our modified label smoothing, and small batch sizes improve performance on class-imbalanced tabular datasets.**

Method	Otto	Adult	CoverType
XGBoost	82.7	87.5	96.9
MLP	83.0	87.4	97.5
ResNet	82.5	87.4	97.5
FT-Transformer	82.3	87.3	97.5
MLP w/ SAM-A + Smoothing	**83.2**	**87.6**	**97.6**

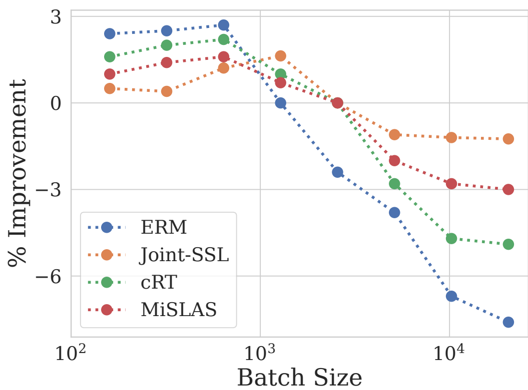

Figure 1: **Our Joint-SSL method acts as an alternative regularizer, mitigating the overfitting of minority classes in large batch sizes** We plot the percent improvement in accuracy over the baseline batch size for imbalance training (=0.01) as a function of batch size for different imbalance training methods. Joint-SSL training yields a flatter line, indicating insensitivity to batch size. ResNet-50 on CIFAR-100.

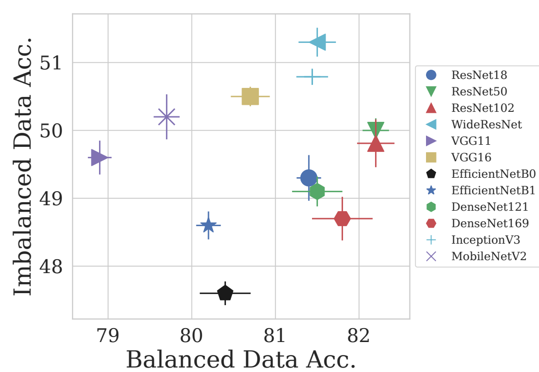

Figure 2: **Performance on balanced and imbalanced datasets is virtually uncorrelated across a wide variety of architectures** (Pearson correlation coefficient 0.14). We plot the imbalance accuracy vs. the balanced accuracy. Experiments were conducted on CIFAR-100 with an imbalanced train ratio of 0.001. Error bars represent one standard error over 5 trials.

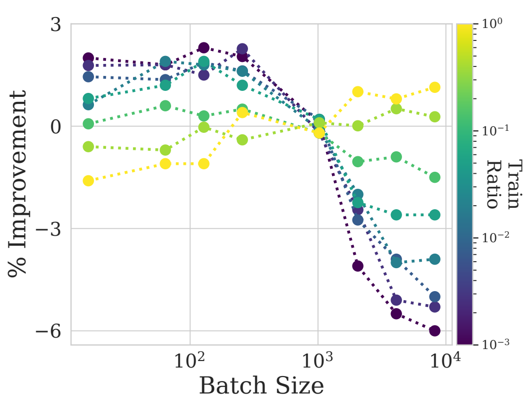

Figure 3: **Imbalanced data prefers small batch sizes - Swin Transformer v2** We plot the percent improvement in accuracy over the baseline batch size of 1024 for different train ratios as a function of batch size. Positive values indicate higher accuracy than the baseline. Balanced training sets yield flatter lines, indicating insensitivity to batch size - CIFAR-100.

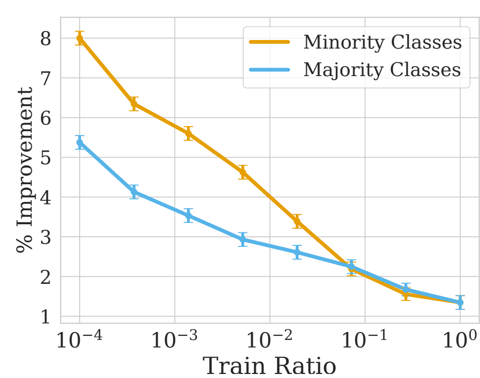

Figure 4: **Augmentations yield far bigger improvements on minority classes - Swin Transformer V2**. We compare the percent improvement in test accuracy of TrivialAugment compared to training without any augmentation as a function of the training ratio. Error bars represent one standard error over 5 trials. Experiments conducted on CIFAR-100.