BigTransfer (BiT), a set of pre-trained image models that can be transferred to obtain excellent performance on new datasets, even with only a few examples per class.

In essence, BiT combines large architectures and large datasets. The best performance across the models increases as the dataset size increases.

Another major factor is the number of epochs designated for pretraining. To ensure good results from pre-training ImageNet-21K as opposed to ImageNet, the dataset is much larger so as a result, we need more epochs and longer pre-training time for it to mantain accuracy.

GroupNorm and Weight Standardisation
BatchNorm performs worse when the number of images on each accelerator is too low. GroupNorm does not have this problem but does not scale well to large overall batch sizes. But when we combine GroupNom with Weight Standardisation, we see that GroupNorm scales well to large batch sizes, even outperforming BatchNorm.

Fine-tuning becomes cheap and efficient as we need only a smaller dataset to fine-tune the model.
Also, BiT has a relatively good configuration of the ideal hyperparameters without undergoing an expensive hyperparameter sweep.