

Rethinking ImageNet pre-training

Summary

1. Reports competitive results on object detection and instance segmentation using standard models trained from random initialization

2. 2 tricks:

1. Argues that Batch Normalization can not be used due to reduced batch size because of mem constraints.

• They tried 2 other normalization scheme:

• Group Normalization, whose accuracy is insensitive to batch size.

• Synchronized BN: the batch statistics is computed across multiple GPUs.

2. Train for longer