

Masterarbeit – Zeit-Effizientes Training von Convolutional Neural Networks

Jessica Bühler

13. Mai 2020

Übersicht I

1 PruneTrain

2 Large Batch

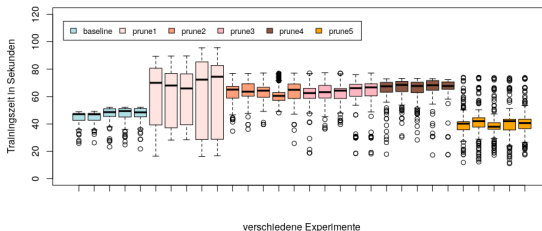


Abbildung: Übersicht über die Zeiten

prune1-prune4 sind Experimente ohne Anpassung der Batchgröße
währenddem Training
prune5 Anpassung der Batchgrösse währenddem Training

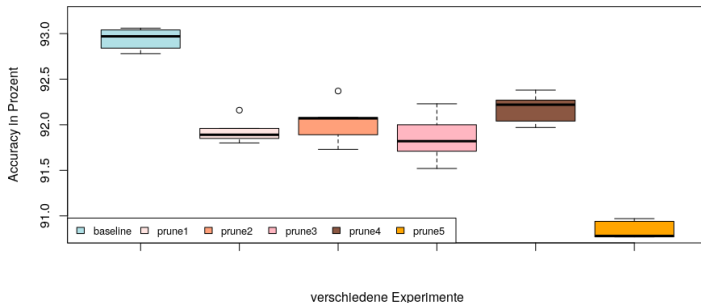
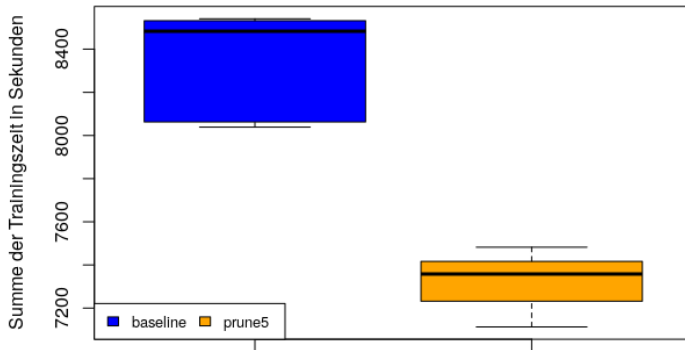


Abbildung: Übersicht über die Accuracy

Accuracy Verlust durch Prune5 2.6%

- Im PruneTrain Paper zeigt sich bereits beim Anwenden von PruneTrain ohne Anpassung der Batch Size ein Verringerung der Trainingszeit
- Dieses Ergebnis lässt sich leider nicht reproduzieren. Da dort mehrere GPU parallel benutzt werden, wird hier vorallem an der Kommunikation zwischen den GPU gespart wird.
- Beim Anpassen der Batchsize währenddem Pruning wird hier etwa 12% Prozent gespart.
- Wobei sich hier gezeigt hat, dass neben der Anpassung der Batchgrösse auch wichtig ist dass die richtige Lernrate genutzt wird.



Übersicht I

1 PruneTrain

2 Large Batch

Large Batch

- Large Batch sorgt bei einer grösseren Batchsize (4000 statt 256) dafür, dass weiterhin ein grosser Teil des Netzwerks geprunt werden kann.
- Bei Large Batch wird nicht der komplette Speicher genutzt. Hier sind noch weitere Untersuchungen fällig.
- bei Cifar10 konnte bei einer grossen Batch size keine Generalization Gap beobachtet werden -> Cifar100 oder ImageNet