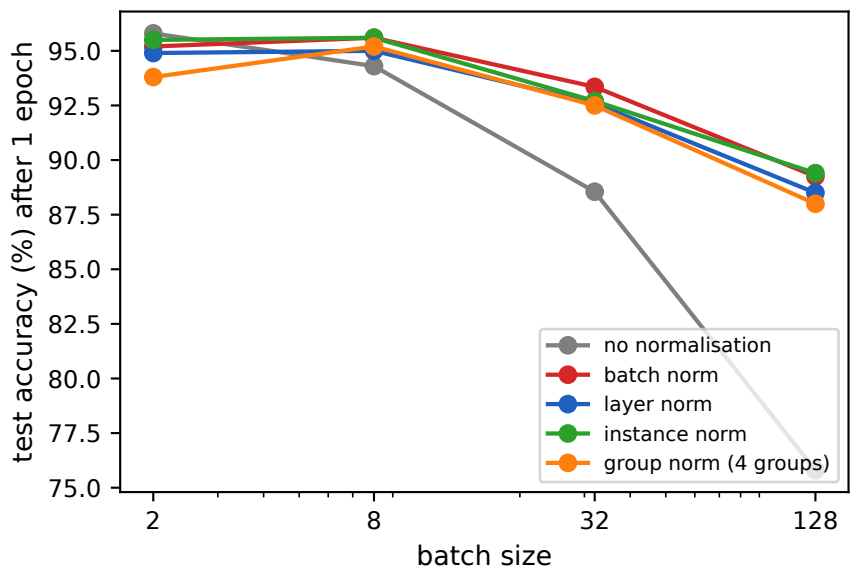
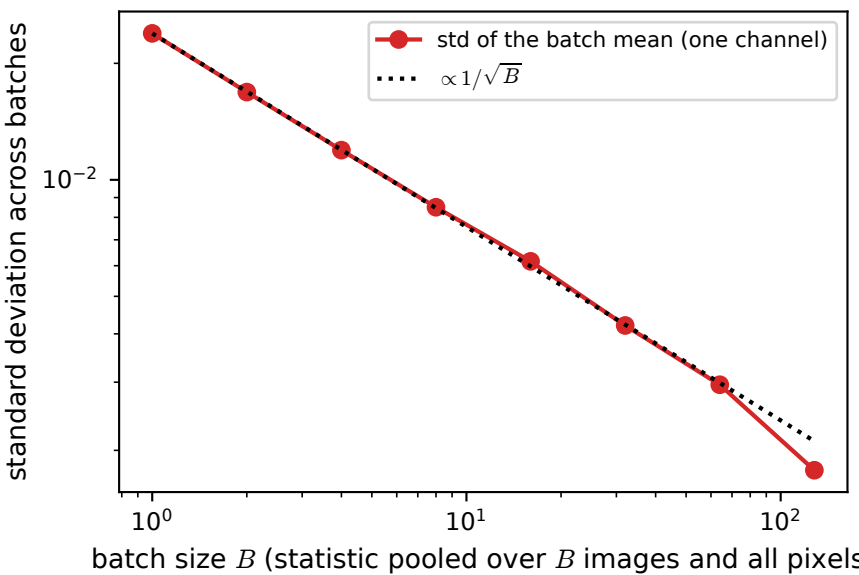

test accuracy (%) after 1 epoch
95.0
92.5
90.0
87.5
85.0
82.5
80.0
77.5
75.0
2
8
32
128
batch size
no normalisation
batch norm
layer norm
instance norm
group norm (4 groups)


standard deviation across batches
10^{-2}
10^0
10^1
10^2
batch size B (statistic pooled over B images and all pixels)
std of the batch mean (one channel)
$\propto 1/\sqrt{B}$