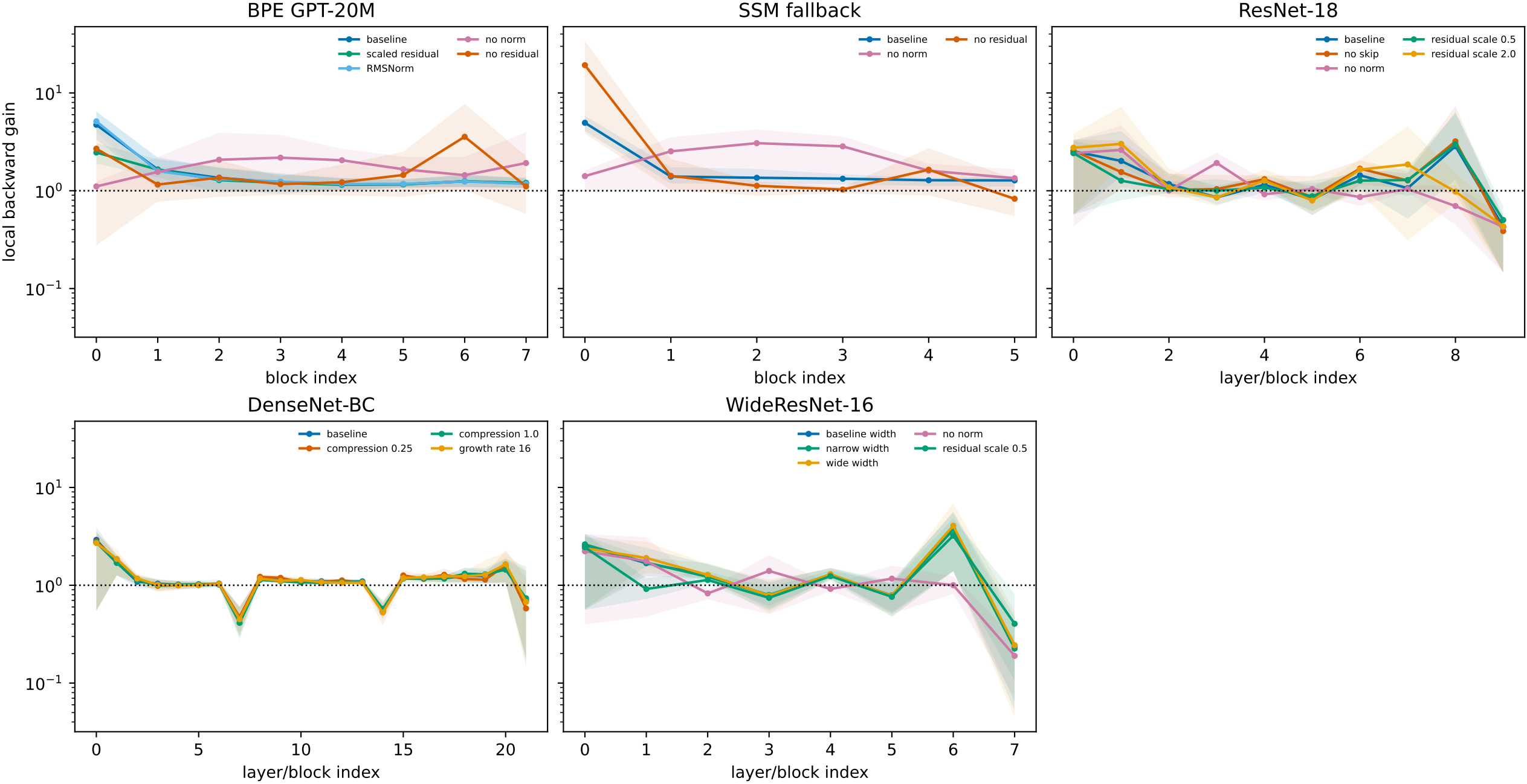
BPE GPT-20M
baseline
scaled residual
RMSNorm
no norm
no residual
local backward gain
block index
SSM fallback
baseline
no norm
no residual
block index
ResNet-18
baseline
no skip
no norm
residual scale 0.5
residual scale 2.0
layer/block index
DenseNet-BC
baseline
compression 0.25
compression 1.0
growth rate 16
layer/block index
WideResNet-16
baseline width
narrow width
wide width
no norm
residual scale 0.5
layer/block index