

Model Pruning as Optimization

Andrew Jenkins

Pruning – Biological & Machine

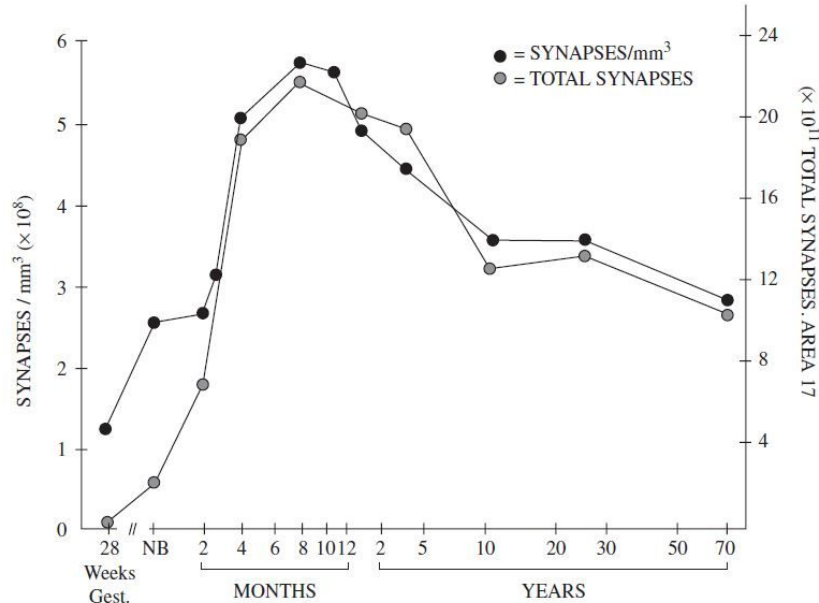
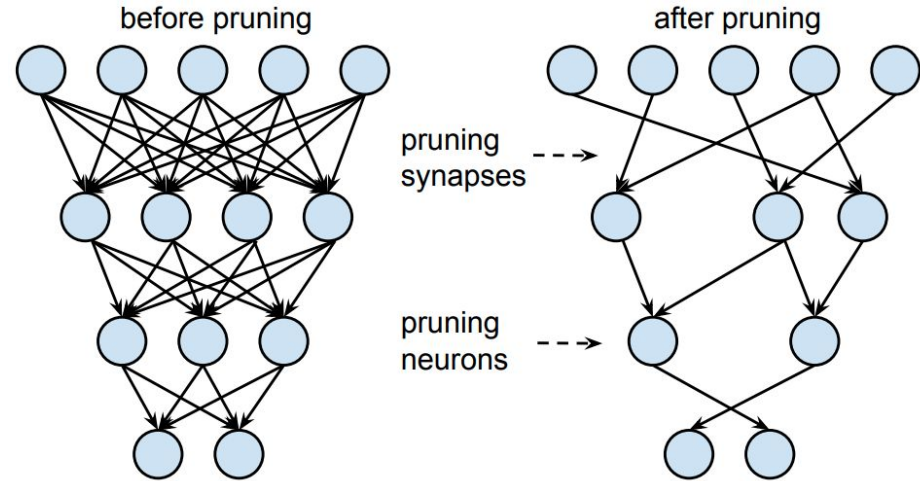


Figure 1.2: Synaptic density and total synapses in visual cortex as a function of age.⁵

⁵Reproduced from Huttenlocher (1990).



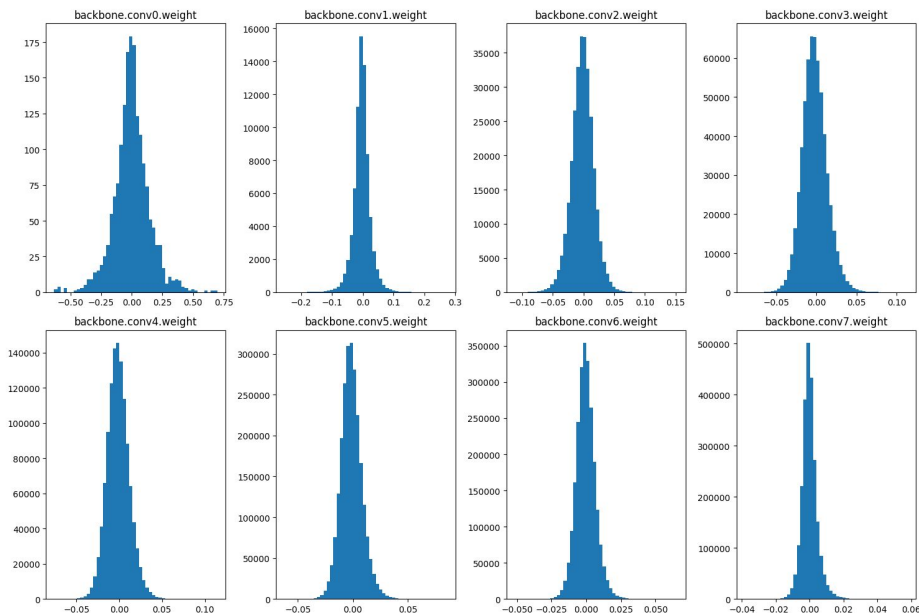
Algorithm/Optimization

Magnitude-based pruning to decrease model size without losing accuracy

Implementation

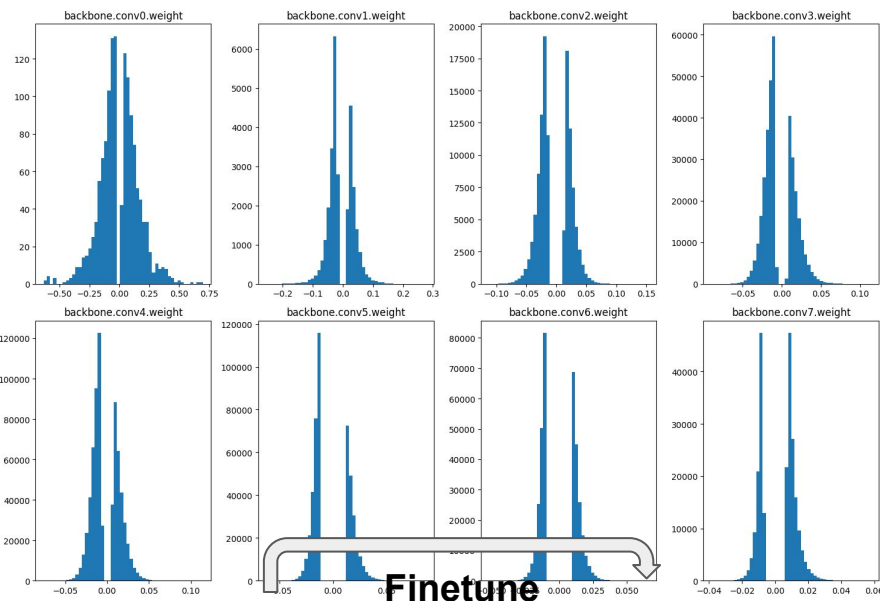
Used CIFAR10 & VGG16

Pretrained



Accuracy: 92.95%
params: 9,228,362
Size (MiB): 35.20

Pruned



Accuracy: 83.81%
params: 2,280,278
Size (MiB): 8.70

Accuracy: **91.95%**
params: 2,280,278
Size (MiB): 8.70