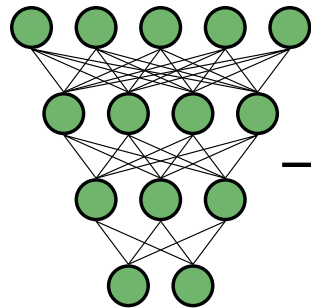
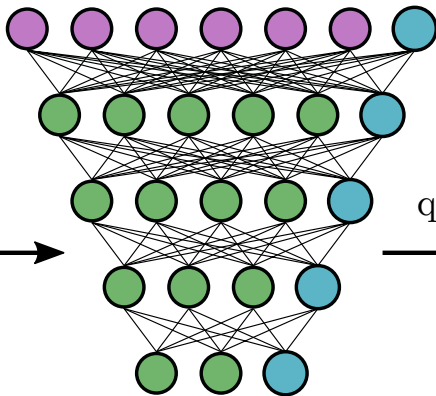


Baseline



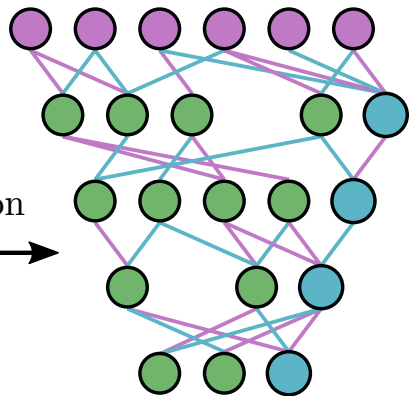
compound
scaling

Supernet



ternary
quantization
pruning

Subnet



sparsity

accuracy

efficiency