BPB vs Training FLOPs at Compute Optimal Ratio (Space Patching)

