AdamW / CUDA 1.5 -Relative median time per batch 1.0 -0.5 -0.0