Running on CPU

--- (✗1 + ✗2) Final Scaling Benchmark (Forward + Backward Time) ---
D_MODEL=64, BATCH_SIZE=8

N_SEQ	ResNet-BK (ms)	Attention (ms)
64	20.576	7.254
256	52.015	24.901
1024	140.690	273.912
2048	325.164	1095.187

Benchmark graph saved to scaling_benchmark_final.png

Scaling Benchmark (Final): ResNet-BK (Analytic Grad) vs Attention (Autograd)
(B=8, D=64)