Running on CPU

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

N_SEQ	ResNet-BK (ms)	Attention (ms)
64	24.612	5.079
256	74.383	27.184
1024	364.980	745.188
2048	807.485	2426.926

Benchmark graph saved to: scaling_benchmark.png

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

Log-Log Scaling Plot: O(N) vs O(N^2)

Legend:
- ResNet-BK (O(N)) - Teppei's Model
- Attention (O(N^2)) - Standard

X-axis: Sequence Length (N) [log scale]
Y-axis: Time per Batch (ms) [log scale]