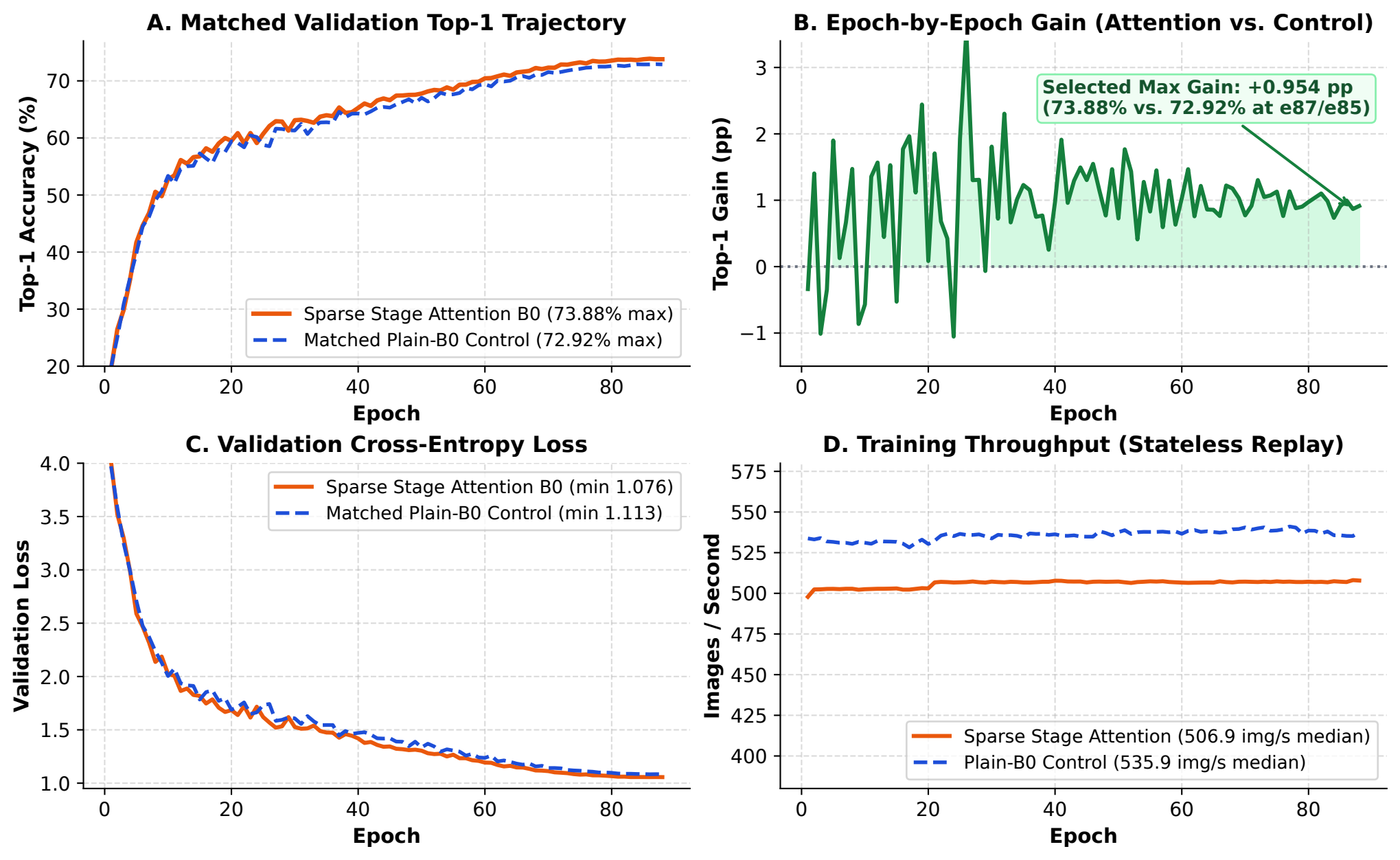

A. Matched Validation Top-1 Trajectory
Top-1 Accuracy (%)
Epoch
Sparse Stage Attention B0 (73.88% max)
Matched Plain-B0 Control (72.92% max)
B. Epoch-by-Epoch Gain (Attention vs. Control)
Top-1 Gain (pp)
Epoch
Selected Max Gain: +0.954 pp
(73.88% vs. 72.92% at e87/e85)
C. Validation Cross-Entropy Loss
Validation Loss
Epoch
Sparse Stage Attention B0 (min 1.076)
Matched Plain-B0 Control (min 1.113)
D. Training Throughput (Stateless Replay)
Images / Second
Epoch
Sparse Stage Attention (506.9 img/s median)
Plain-B0 Control (535.9 img/s median)