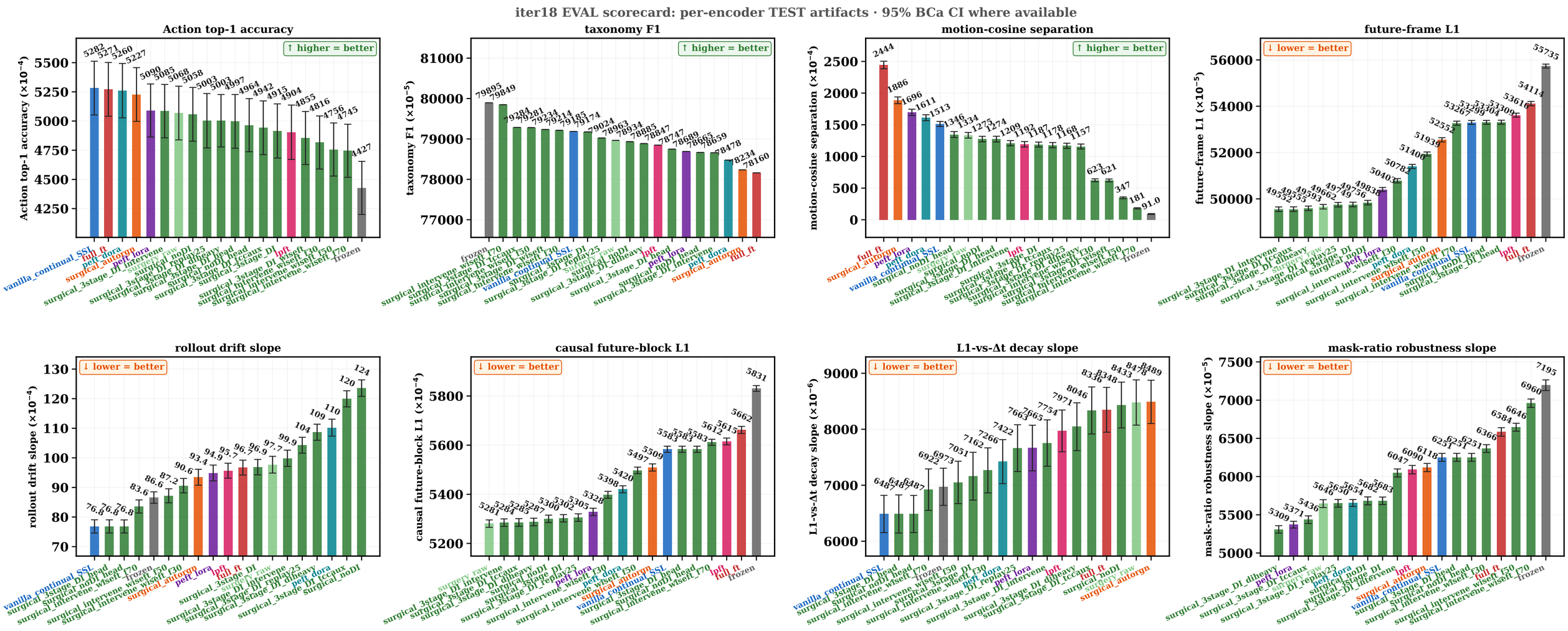
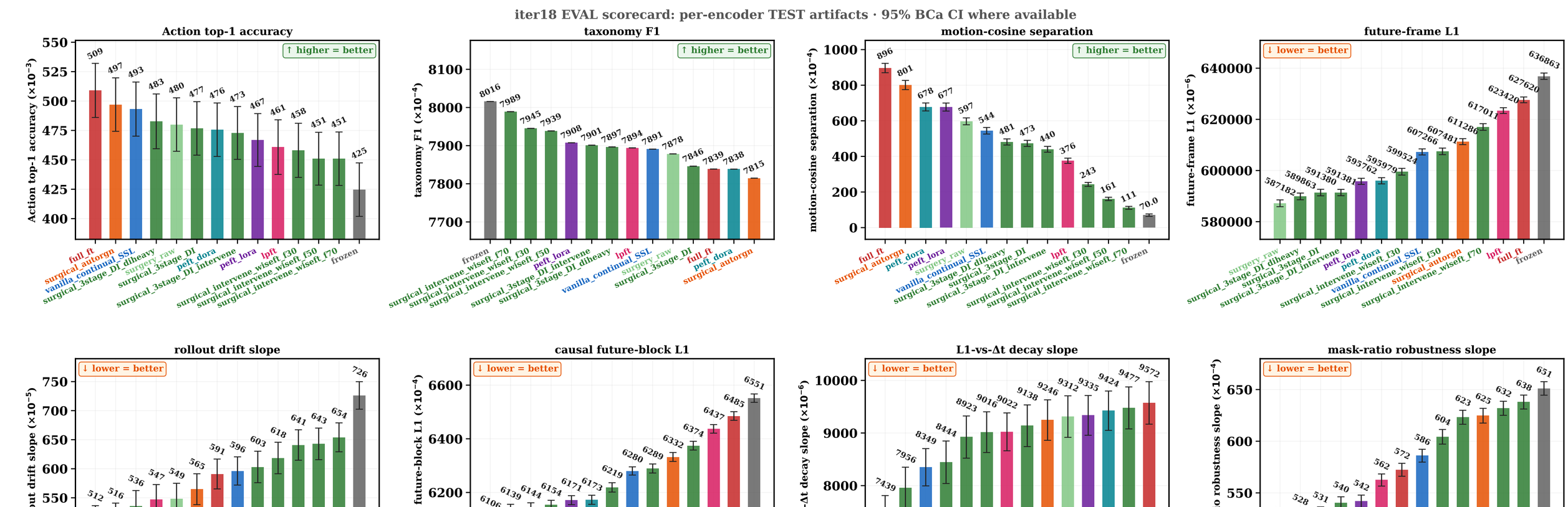

iter18 EVAL scorecard: per-encoder TEST artifacts · 95% BCa CI where available
Action top-1 accuracy
taxonomy F1
motion-cosine separation
future-frame L1
rollout drift slope
causal future-block L1
L1-vs-Δt decay slope
mask-ratio robustness slope
frame-order sensitivity
free-running exposure-bias gap
Arrow-of-Time accuracy
temporal-order (frame-permutation) acc
playback-pace accuracy
TCC cycle-back error
TCC Kendall τ
↑ higher = better
↓ lower = better


iter18 EVAL scorecard: per-encoder TEST artifacts · 95% BCa CI where available
Action top-1 accuracy
taxonomy F1
motion-cosine separation
future-frame L1
rollout drift slope
causal future-block L1
L1-vs-Δt decay slope
mask-ratio robustness slope
frame-order sensitivity
free-running exposure-bias gap
Arrow-of-Time accuracy
temporal-order (frame-permutation) acc
playback-pace accuracy
TCC cycle-back error
TCC Kendall τ
↑ higher = better
↓ lower = better