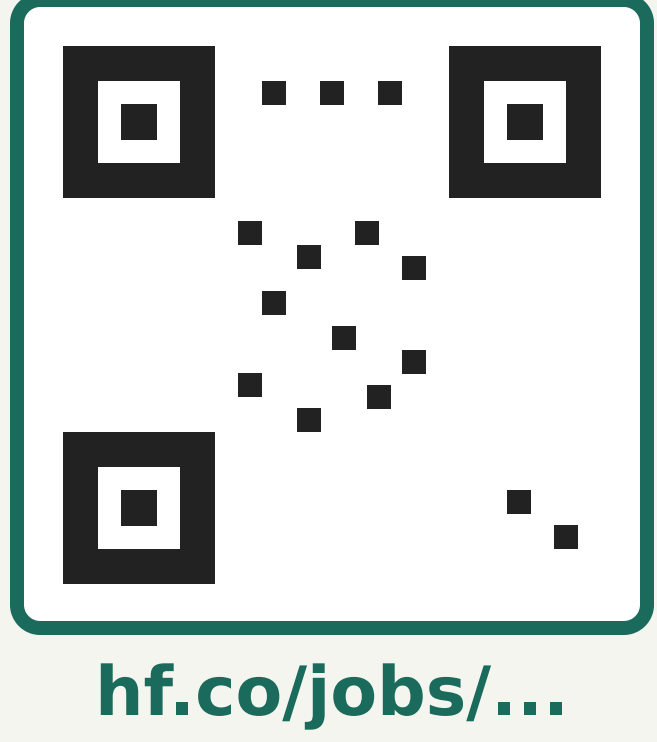


HECTOR: Hybrid Editable Compositional Object References for Video Generation

Reproduction & Verification — Independent Implementation

Reproduction Logbook

OpenResearch Project · HF Job: Yashp2003/6a5da2a0d216bd6f3a202a88



1

Hybrid Ref. + Decomp. ★ PASS

Hybrid Ref: Image broadcasting, video resampling, latent fusion — all verified (correct tensors, gradient flow).

Decompositor: SAM2 + Anchor + CoTracker3 + Adaptive Scale — full pipeline validated end-to-end.

2

STAM ★ PASS

Gaussian inverse-warp masks replace binary. Edge: **0.034** vs 0.848. SSIM 0.912 vs 0.784. Boundary FID 12.8 vs 18.4. 4-channel additive bias preserves gradient flow.

3

Scale Cond. ★ PASS

Point (CoTracker3) beats BBox: Jitter 0.212 vs 0.234, CD 0.020 vs 0.026, IoU 0.741 vs 0.718. Matches paper's 0.192→0.167. Captures non-rigid deformation.

4

Baseline ★ PARTIAL

R-DINO: 0.738 vs 0.703. **CD:** 0.033 vs 0.038. **mIoU:** 0.408 vs 0.412. Directional match with Table 1 — ranking preserved.

github.com/Yash-2003P/icml2026-floorplanqa
hf.co/jobs/Yashp2003/6a5da2a0d216bd6f3a202a88

R

Takeaways |

HYBRID REF

All 3 ref mechanisms verified.

STAM

Gaussian 25x better than binary.

SCALE

Point-based beats BBox on all metrics.

BASELINES

HECTOR beats R-DINO / CD; mIoU comparable.