St4RTrack: Tái tạo & theo dõi 3D

Nguyễn Phạm Phương Nam University of Information Technology **Hồ Ngọc Luật**

HCMC, VietNam

What?

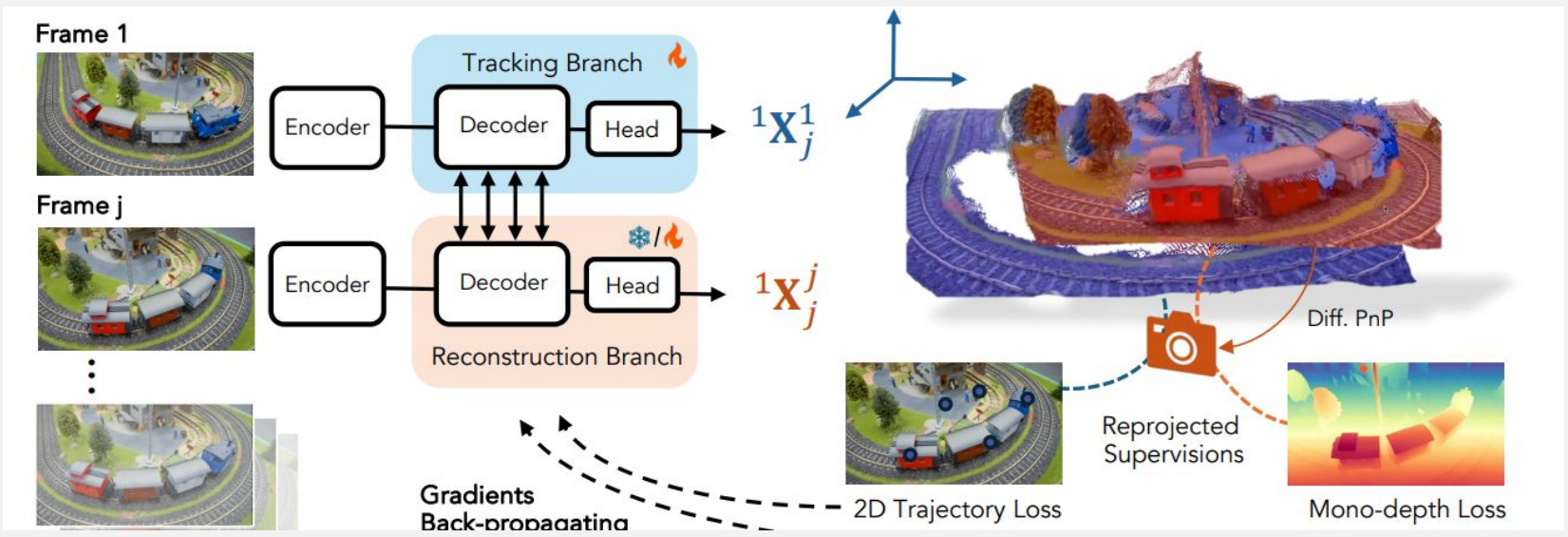
- Feed-forward: nhận (Frame 1, Frame j) ⇒ 2 pointmaps (world = Frame 1).
- Tracking: X_j^1 = điểm Frame 1 tại thời điểm j.
- Recon: X_j^j = hình học Frame j tại thời điểm j.
- Chuỗi theo j ⇒ world-frame 3D tracks dài hạn.

Why?

- Cảnh động ⇒ khó tách camera/scene motion; occlusion dễ gây outlier.
- World-frame tracking: tách camera motion vs scene motion.
- Robust TTA (đề xuất): PnP khả vi + reprojection

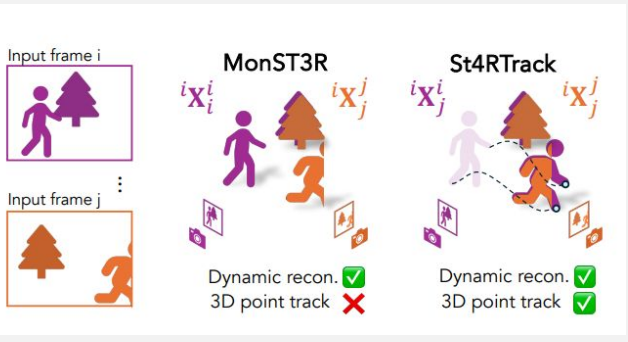
với mask m_n + uncertainty σ_n^2.

Overview of St4RTrack



1) Biểu diễn 4D thống nhất

Pointmap có thêm yếu tố thời gian: $^aX_t^b$ = nội dung frame b tại thời điểm t, trong hệ toạ độ của frame a.



Khác MonST3R: dự đoán "điểm frame 1 ở thời điểm j"
⇒ có correspondence theo thời gian (3D tracks)
⇒ world-frame tracking dài hạn.

2) Học & thích nghi (không nhãn 4D)

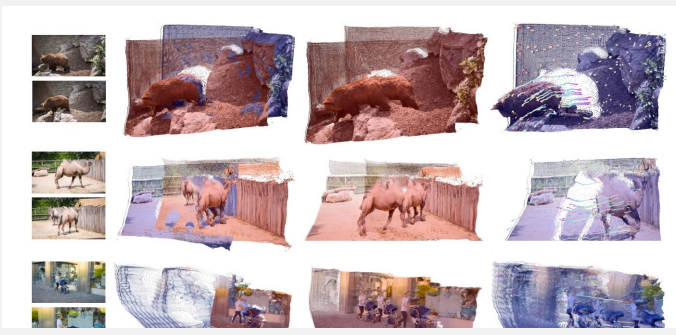
Pretrain (synthetic 4D):
- Tracking: mesh vertices (world).
- Recon: depth + camera GT.

Suy ra camera frame j:
- K (DUSt3R), pose (R□,T□): PnP+RANSAC (khả vi).

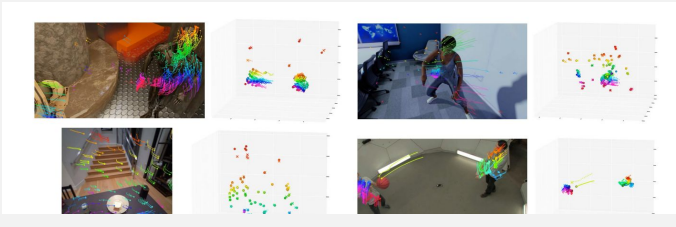
TTA / self-supervision (video thật, robust):
- mask m□: depth-consistency (z-buffer).
- uncertainty σ_n^2 : down-weight outlier.

- L_{traj} : reprojection vs CoTracker (weighted).
- L_{depth} : mono-depth vs MoGe (scale-invariant).
- L_{align}: 3D consistency (weighted).

3) Kết quả



Định tính: tích luỹ reconstruction/tracking
⇒ track dày đặc dài hạn trong world frame.



NII