

St4RTrack: Tái tạo & theo dõi 3D

Nguyễn Phạm
Phương Nam

University of Information Technology

Hồ Ngọc Luật

HCMC, VietNam

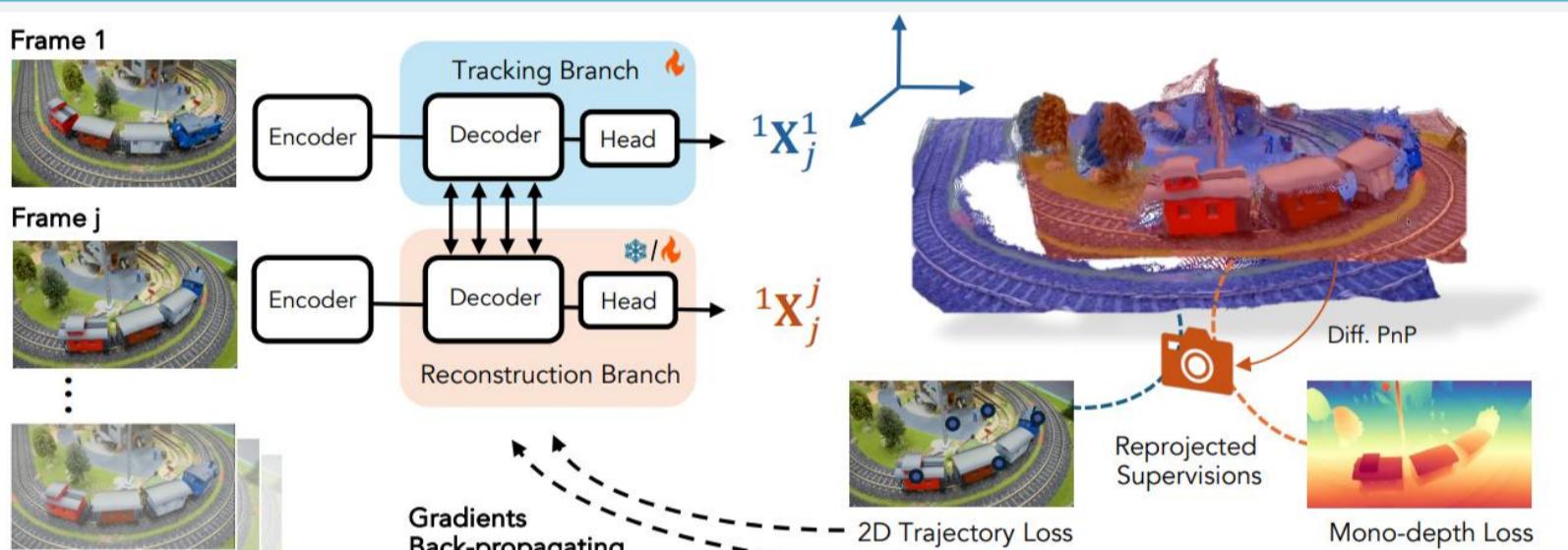
What?

- Feed-forward: nhận (Frame 1, Frame j) \Rightarrow 2 pointmap.
- Tracking: X^1_j = điểm Frame 1 tại thời điểm j.
- Recon: X^j_j = hình học Frame j tại thời điểm j.
- Ghép theo j \Rightarrow world-frame 3D tracks dài hạn.

Why?

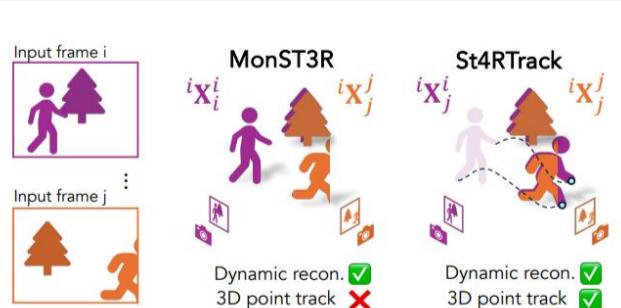
- Cảnh động \Rightarrow khó tách camera/scene motion.
- World-frame tracking: tách camera motion vs scene motion.
- Adapt video thật (không nhãn 4D): PnP khả vi + reprojection (2D tracks + mono-depth + 3D consistency).

Overview of St4RTrack



1) Biểu diễn 4D thống nhất

Pointmap có thêm yếu tố thời gian:
 aX^b_t = nội dung frame b tại thời điểm t,
trong hệ toạ độ của frame a.
Dự đoán: (X^1_j, X^j_j) .



Khác MonST3R: dự đoán “điểm frame 1 ở thời điểm j”
 \Rightarrow có correspondence theo thời gian (3D tracks)
 \Rightarrow 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_j, T_j) bằng PnP+RANSAC (khả vi).

TTA / self-supervision (video thật):

- L_{traj} : reprojection khớp 2D tracks (CoTracker).
- L_{depth} : khớp mono-depth (MoGe), scale-invariant.
- L_{align} : consistency 3D giữa X^1_j và X^j_j .

Khi test: thường chỉ fine-tune tracking branch.

3) Kết quả



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

