Real-Time Visual SLAM for Dynamic Environments using Hybrid Segmentation and Optical Flow

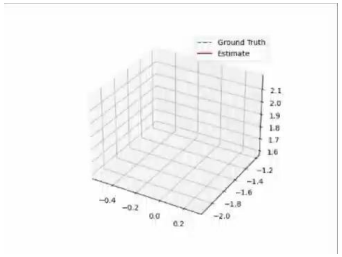
Group 4
Peng-Chen Chen, Yung-Ching Sun, Chi Zhang, Hao Yin

Motivation

Static visual SLAM

Most SLAM methods are based on static environment assumption

Input RGBD

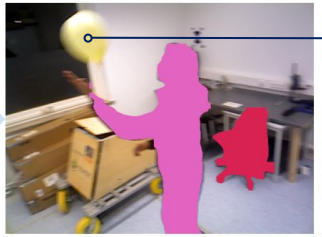


Inaccuracy in dynamic scenes

Semantic-based dynamic SLAM

· Unlabeled dynamic objects can be ignored

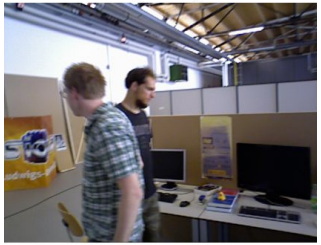
Input RGBD



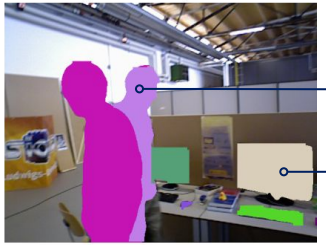
Semantic segmentation

Unlabeled dynamic object

· Segment regions contain both static and dynamic parts



Input RGBD



Semantic segmentation

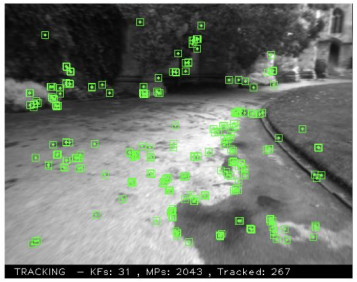
Dynamic

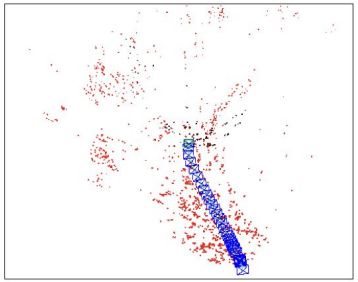
Static

A real-time pipeline to precisely identify dynamic regions?

Related Work

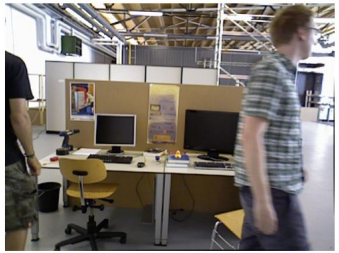
ORB-SLAM3 [1]

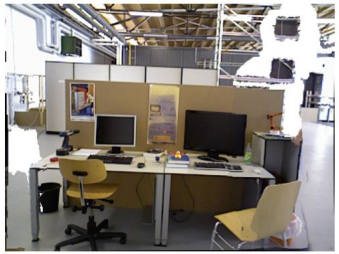




Feature-based visual SLAM

DynaSLAM [2]



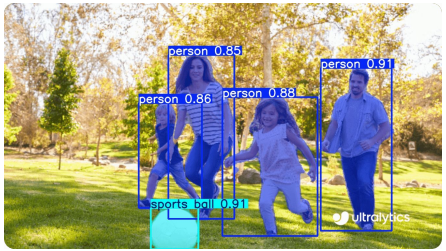


Segmentation of dynamic content using Mask R-CNN

Optical Flow

Motion of image objects

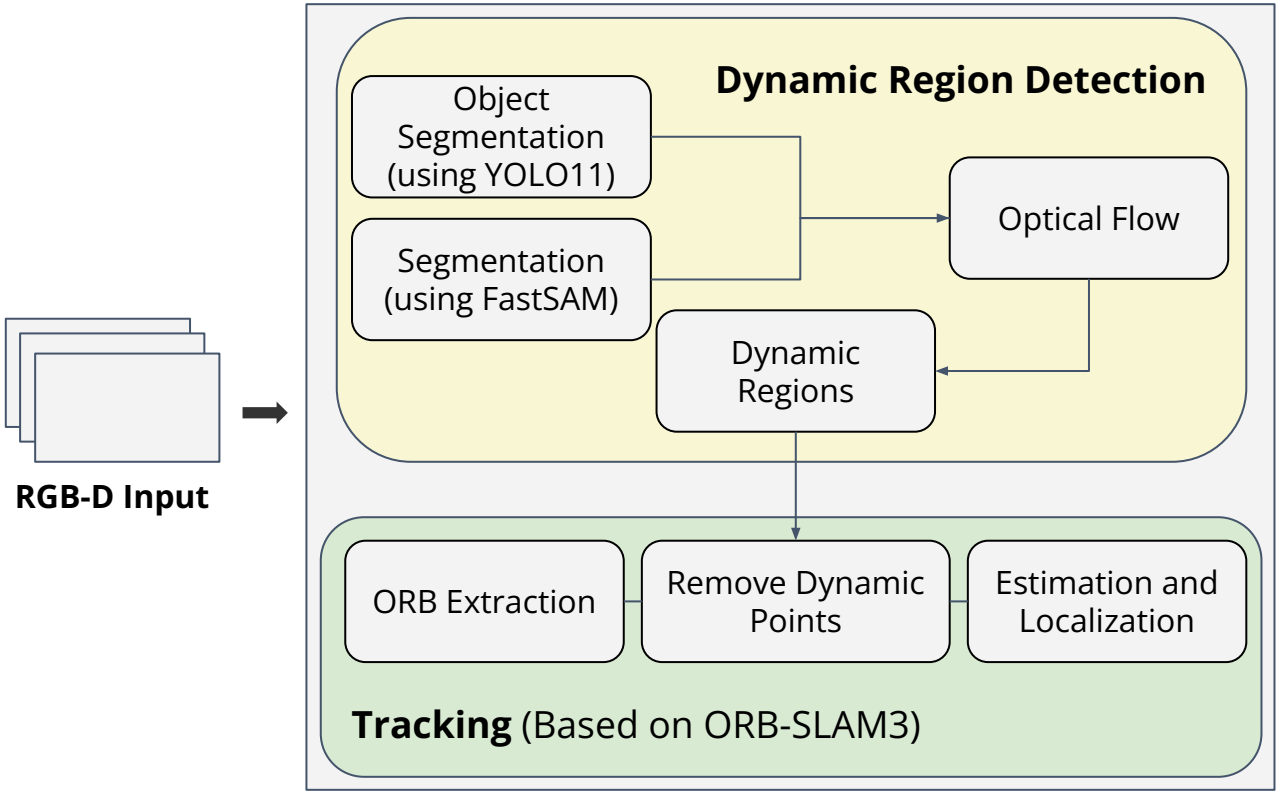
YOLOv11-seg [3]



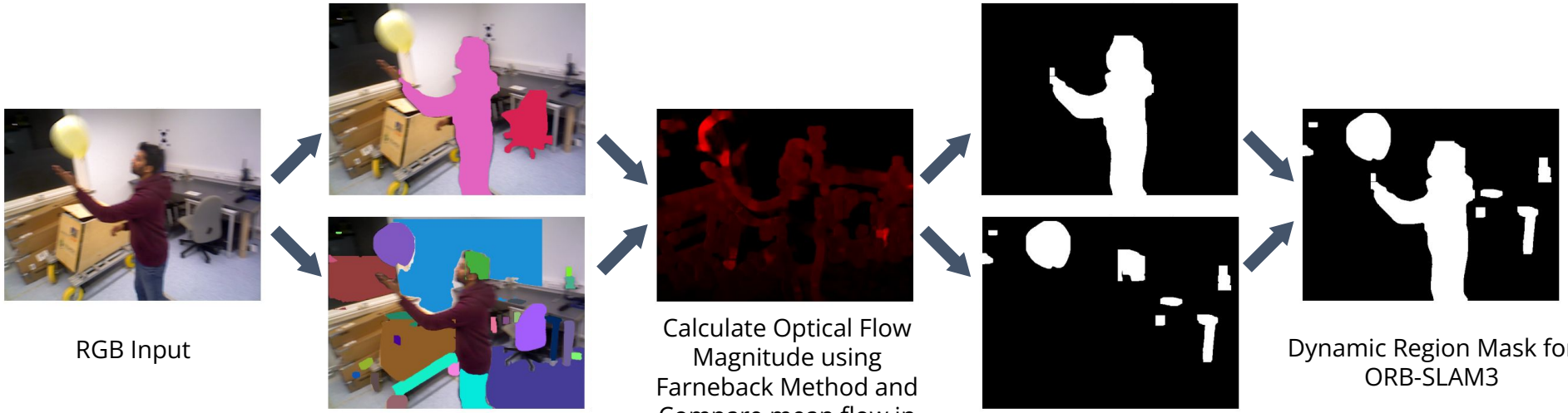
FastSAM [4]

Segmentation models

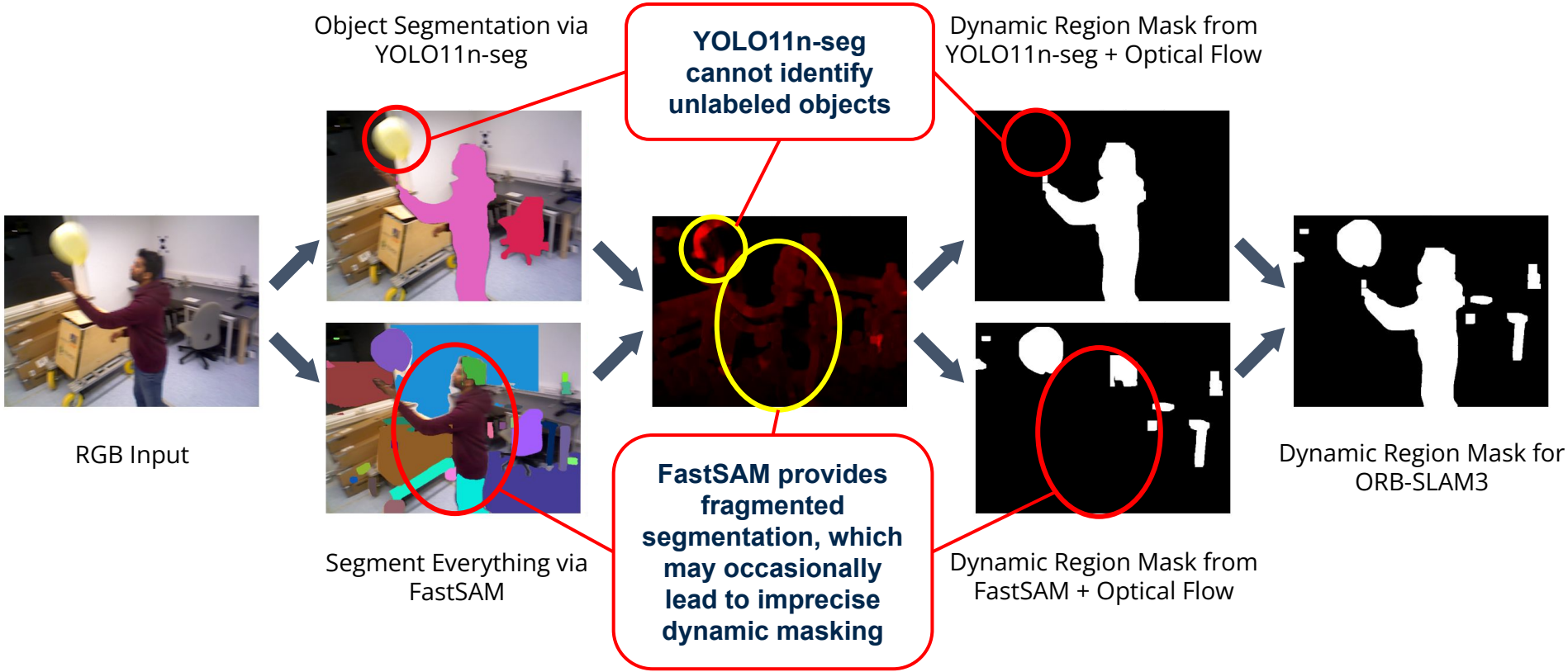
System Overview



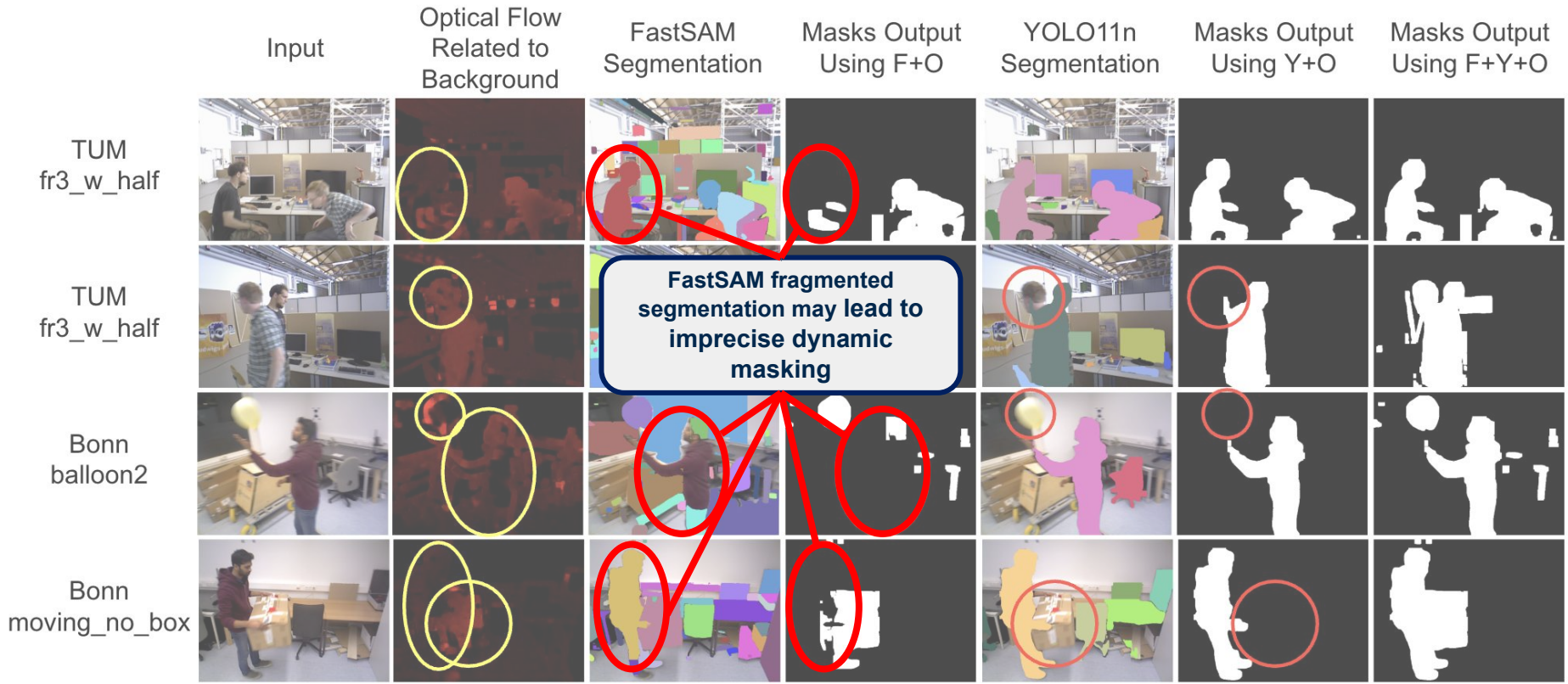
Dynamic Region Masks Generation



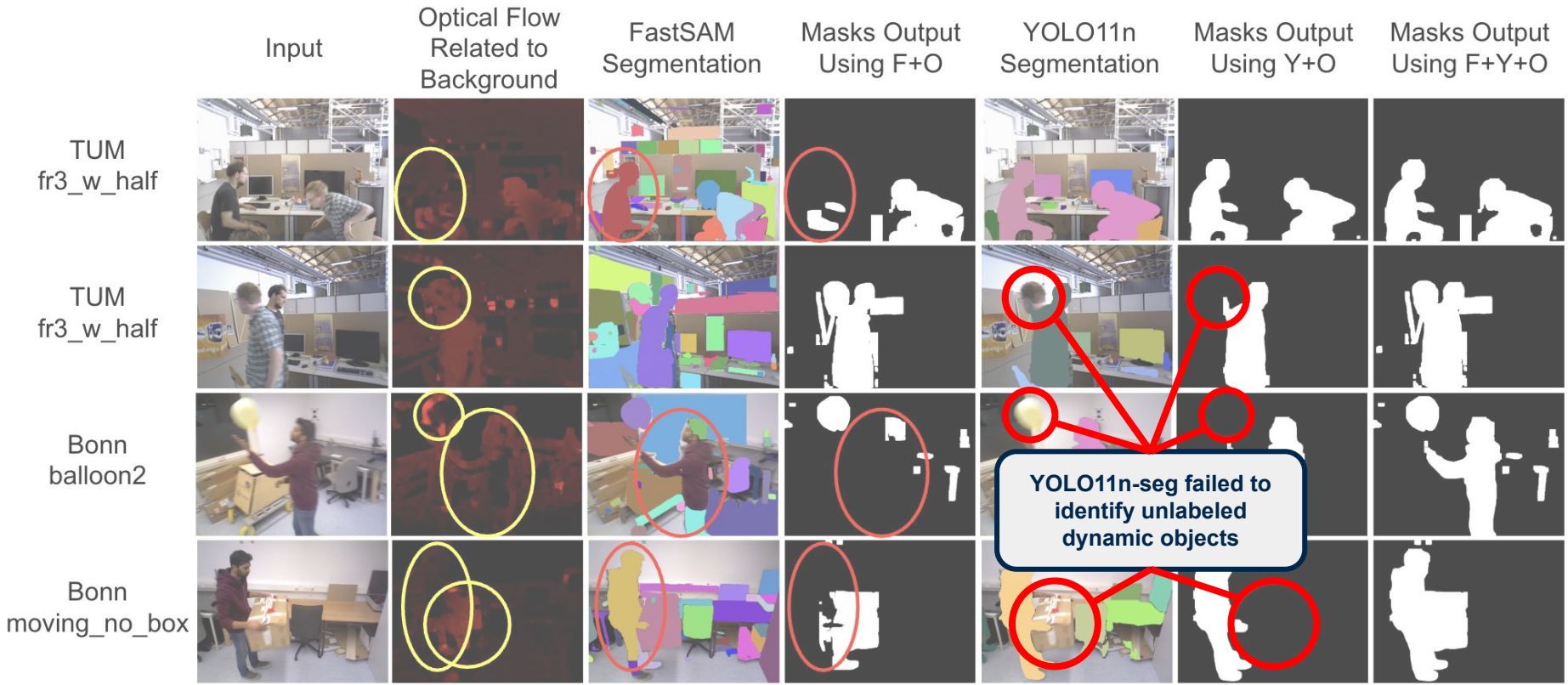
Why Do We Use Two Segmentation Models?



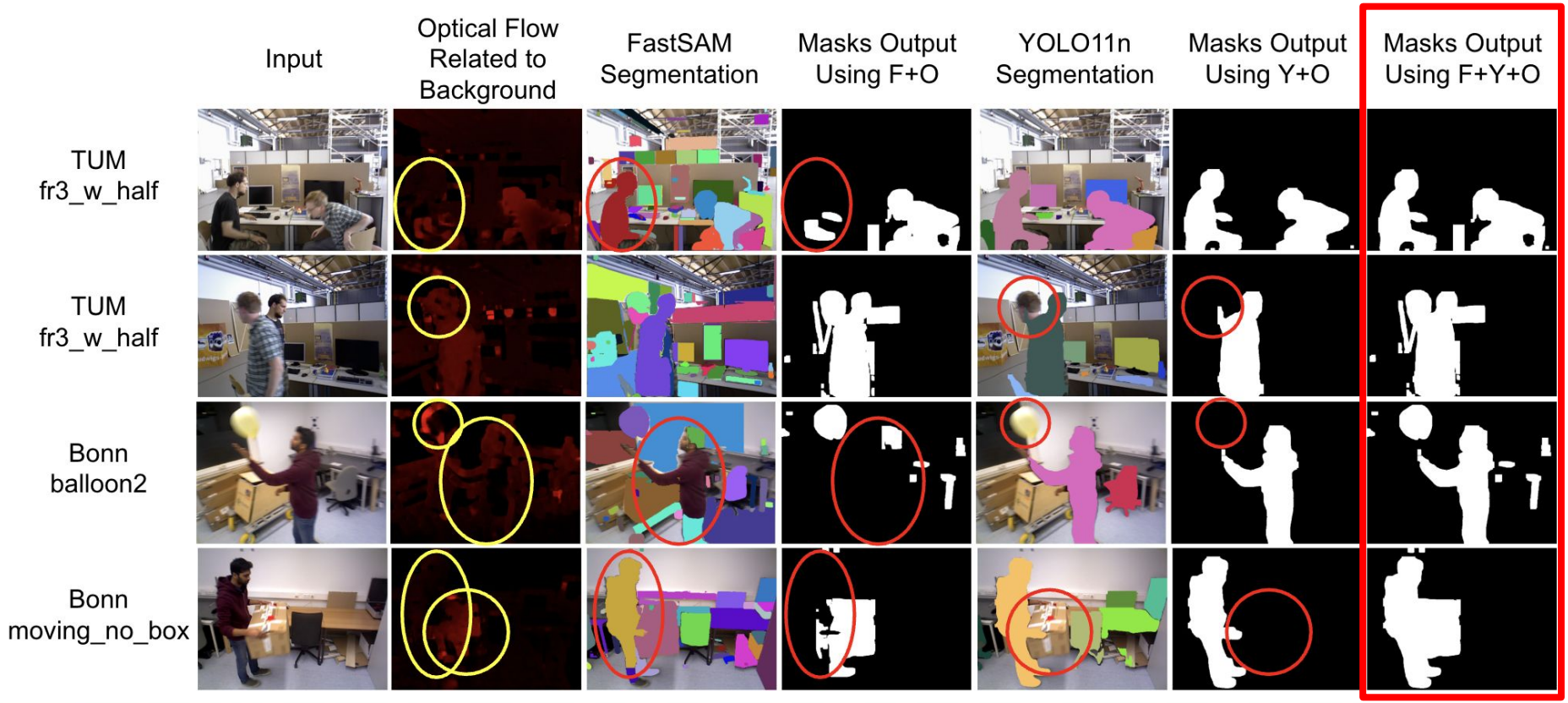
Hybrid Segmentation Enhance Dynamic Masking Reliability



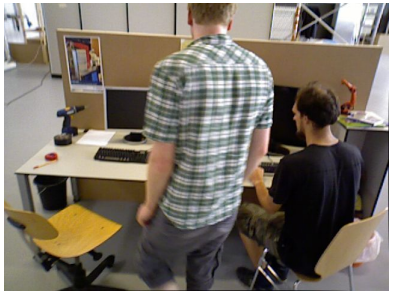
Hybrid Segmentation Enhance Dynamic Masking Reliability



Hybrid Segmentation Enhance Dynamic Masking Reliability

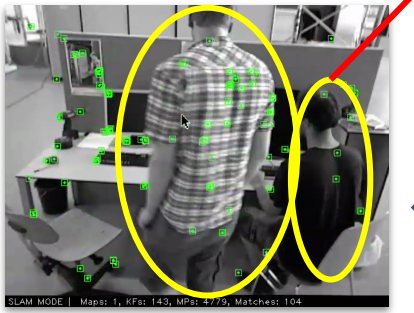


Tracking



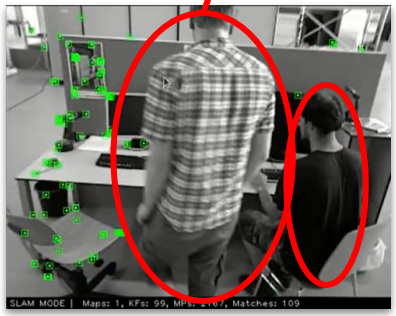
RGB Input

ORB Feature Extraction

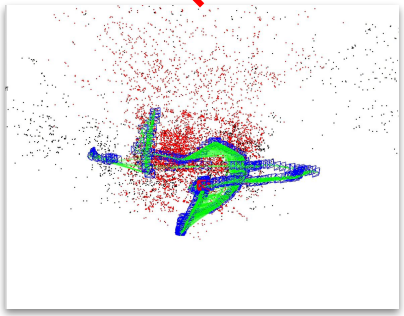


Dynamic Region Mask

Filter out the feature points on moving objects and estimate camera pose from static objects



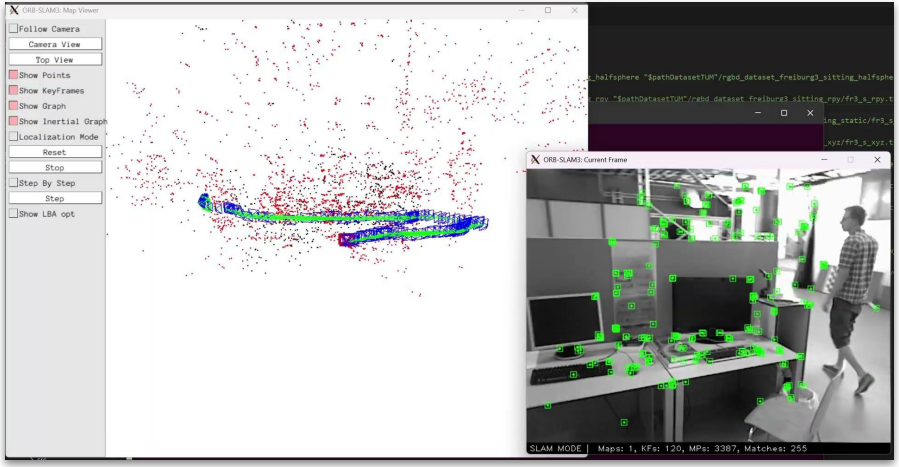
Remove Dynamic Feature Points

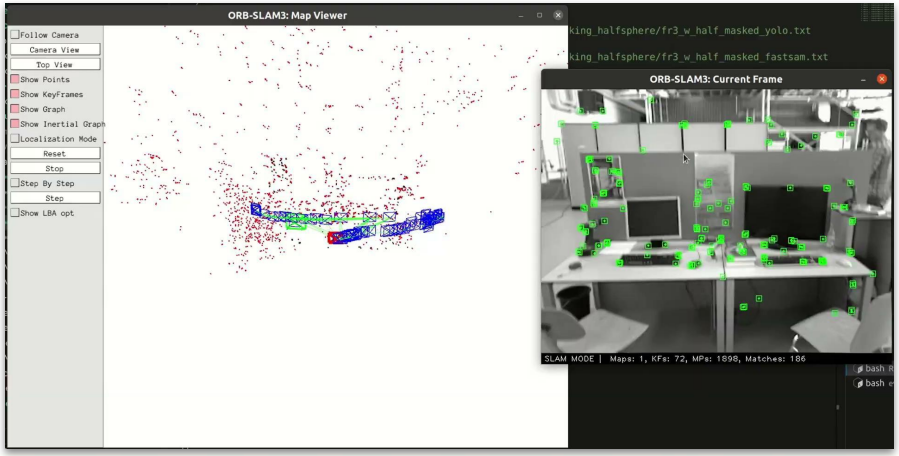


ORB-SLAM3

VS

Our Method





Experimental Setup: Datasets

- TUM RGB-D SLAM Dataset and Benchmark [5]
  - fr3_walking_halfsphere
  - fr3_walking_xyz
  - fr3_walking_static
- Bonn RGB-D Dynamic Dataset [6]

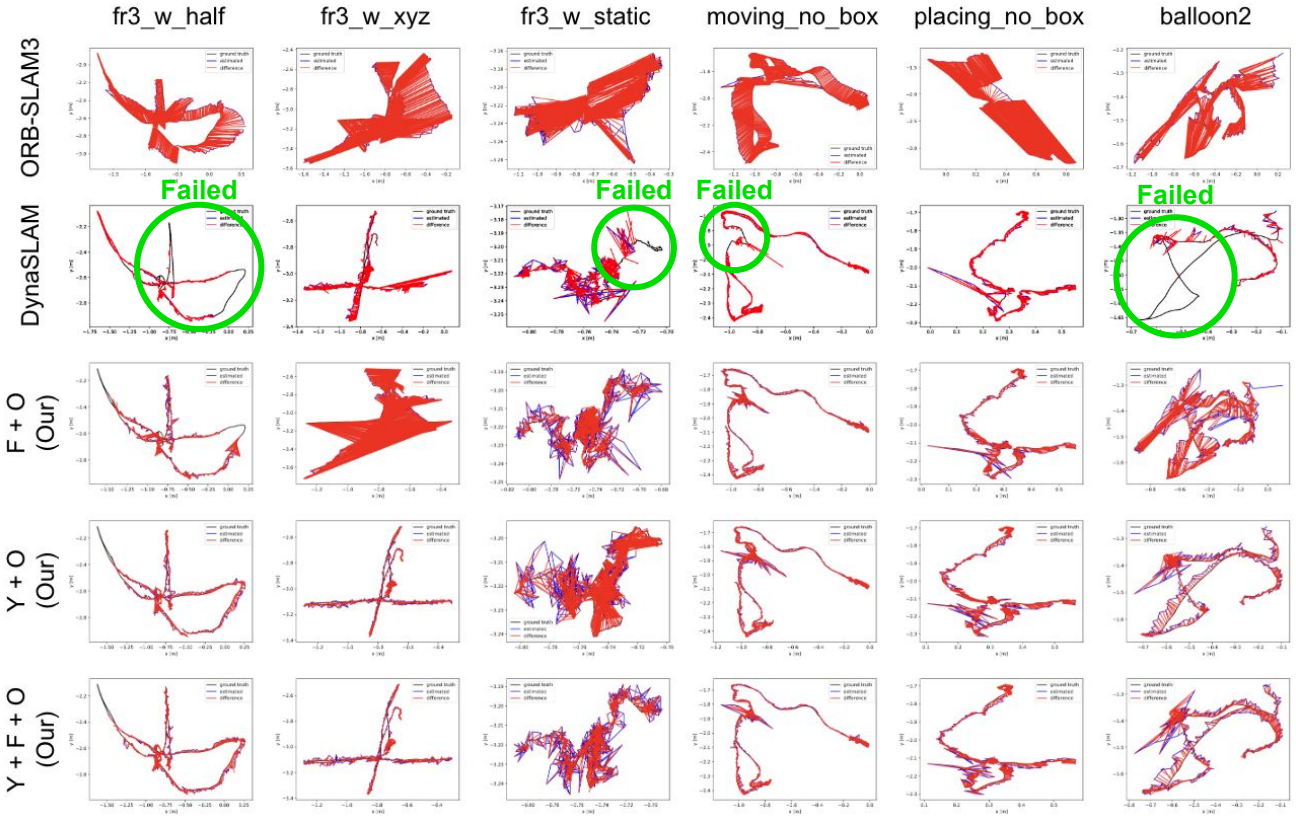
balloon2

moving_nonobstructing_box

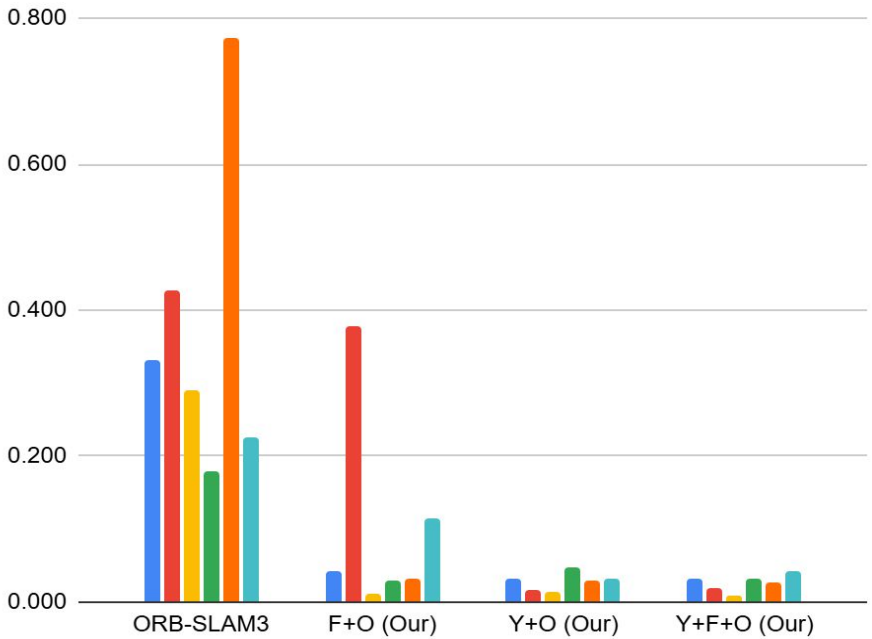
placing_nonobstructing_box

Trajectory Comparison

groundtruth camera trajectory
estimated camera trajectory
difference



Absolute Trajectory Error (ATE) RMSE Results



Best

Second Best

Dataset	ORB-SLAM3	F+O (Our)	Y+O (Our)	Y+F+O (Our)
	ATE [m]	ATE [m]	ATE [m]	ATE [m]
fr3_w_half	0.332	0.043	0.031	0.031
fr3_w_xyz	0.427	0.378	0.018	0.017
fr3_w_static	0.291	0.011	0.015	0.008
moving_no_box	0.179	0.030	0.049	0.025
placing_no_box	0.772	0.032	0.031	0.027
balloon2	0.227	0.114	0.032	0.037

F: FastSAM everything model, Y: YOLO11n segmentation model, O: Optical Flow, ATE: Absolute Trajectory Error

Computational Time Comparison

	ORB-SLAM3	DynaSLAM	Y+O (Our)	F+O (Our)	Y+F+O (Our)
Platform	i9 + RTX4060	i9 + Tesla M40	i9 + RTX4060	i9 + RTX4060	i9 + RTX4060
Inference Time [ms]	-	195	11.213	15.088	28.05
Tracking Time [ms]	10-16	333.68	11-17	11-17	11-17
Real-time	✅	❌	✅	✅	✅

F: FastSAM everything model, Y: YOLO11n segmentation model, O: Optical Flow

Conclusions

- Proposed a lightweight pipeline combining YOLOv11, FastSAM, and optical flow for dynamic region masking.
- YOLO11 + FastSAM + Optical Flow achieves great performance in most sequences, with robust tracking under occlusion and motion.
- Our approach shows high performance with real-time capability in experimental results.
- **Future Work**
  - Apply and evaluate our methods in real-world scenarios.
  - Render scene map and integrate semantic segmentation for better scene understanding.
  - Incorporate depth information to improve dynamic detection.

References

[1] Campos, Carlos, et al. "ORB-SLAM3: An accurate open-source library for visual, visual–inertial, and multimap slam." IEEE transactions on robotics 37.6 (2021): 1874-1890.
[2] Bescos, Berta, et al. "DynaSLAM: Tracking, mapping, and inpainting in dynamic scenes." IEEE robotics and automation letters 3.4 (2018): 4076-4083.
[3] G. Jocher and J. Qiu, “Ultralytics YOLO11,” 2024.
[4] X. Zhao, W. Ding, Y. An, Y. Du, T. Yu, M. Li, M. Tang, and J. Wang, “Fast segment anything,” 2023.
[5] Sturm, Jürgen, et al. "A benchmark for the evaluation of RGB-D SLAM systems." 2012 IEEE/RSJ international conference on intelligent robots and systems. IEEE, 2012.
[6] Emanuele Palazzolo, Jens Behley, Philipp Lottes, Philippe Giguère, Cyrill Stachniss, “ReFusion: 3D Reconstruction in Dynamic Environments for RGB-D Cameras Exploiting Residuals”, arXiv, 2019. PDF

UNIVERSITY OF MICHIGAN

Thank you!