SELF-IMPROVING SLAM IN DYNAMIC ENVIRONMENTS: LEARNING WHEN TO MASK

Paper 0654

Authors:
Adrian Bojko **Romain Dupont**
Mohamed Tamaazousti **Hervé Le Borgne**

Emails: adrian.bojko@cea.fr romain.dupont@cea.fr
mohamed.tamaazousti@cea.fr herve.le-borgne@cea.fr

Université Paris-Saclay, CEA, List

To mask or not to mask, that is the question.

Background

- **SLAM** : Simultaneous Localization and Mapping
- **Dynamic SLAM**: SLAM in Dynamic environments. Track and match image features that are **not** on dynamic objects.

Video frames → **Dynamic SLAM** = SLAM + dynamic object filtering → Camera trajectory + map

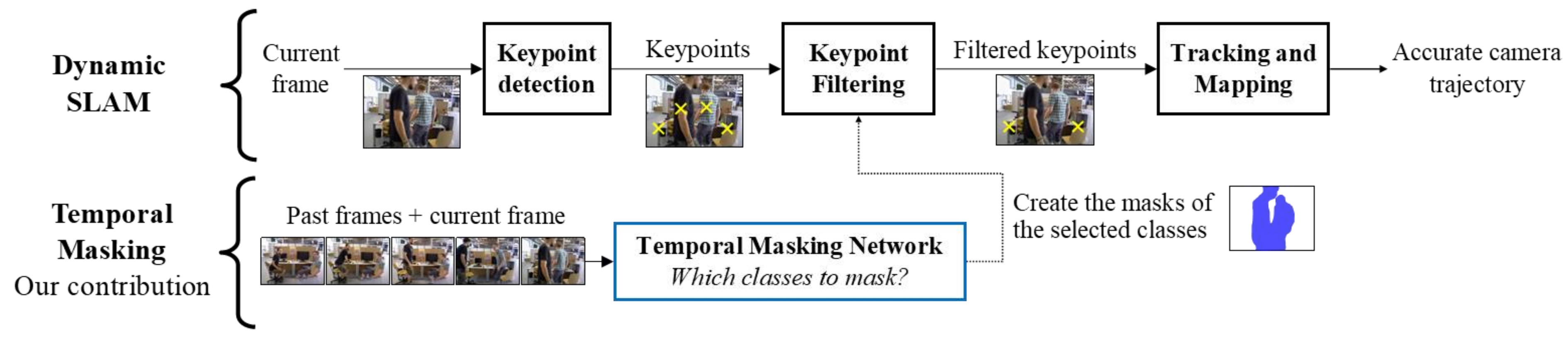
Question

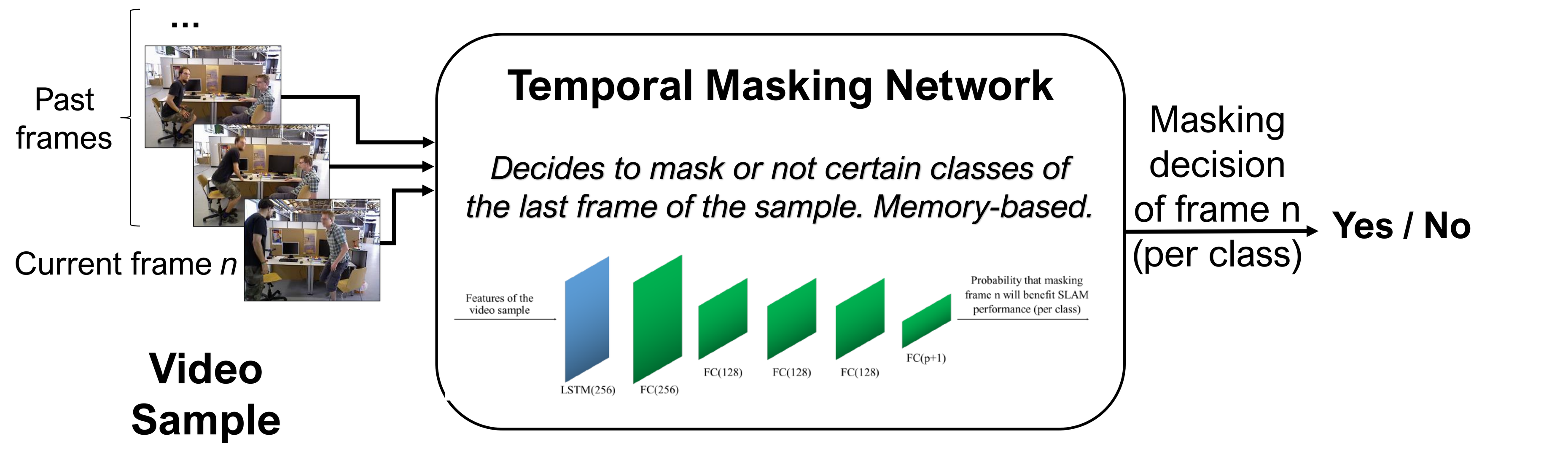
- Current Dynamic SLAM methods do not mask objects when needed, or mask them when it is not necessary. The problem is when to mask, not what to mask.
- The challenge is to mask classes of objects only when appropriate, *i.e.*, **when it improves SLAM performance,** without priors on motion.

Method: SLAM with Temporal Masking

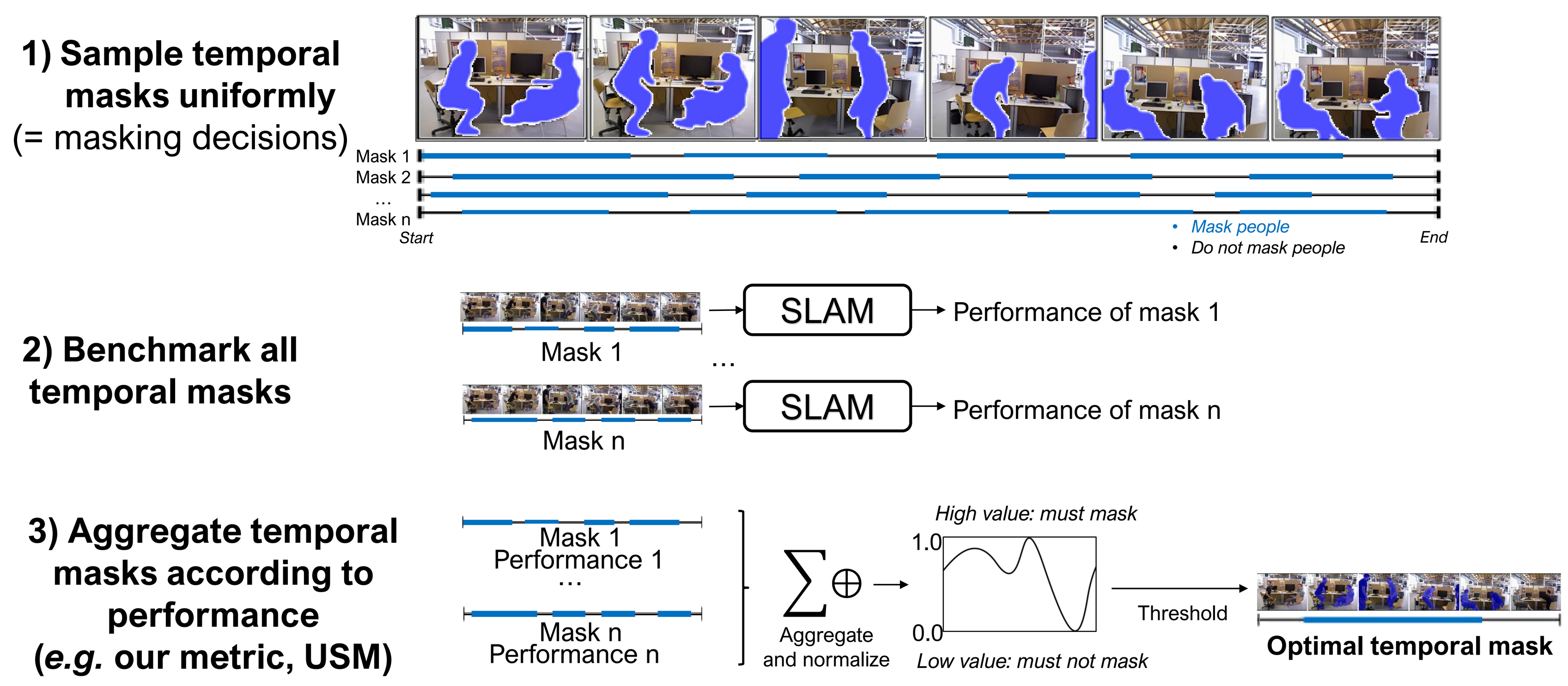
New paradigm: decide when to mask certain classes.

SLAM with Temporal Masking
Includes Temporal Masking Network trained with automatically annotated dataset.





Automatic dataset annotation
Define masking decisions at a low cost for self-supervised training

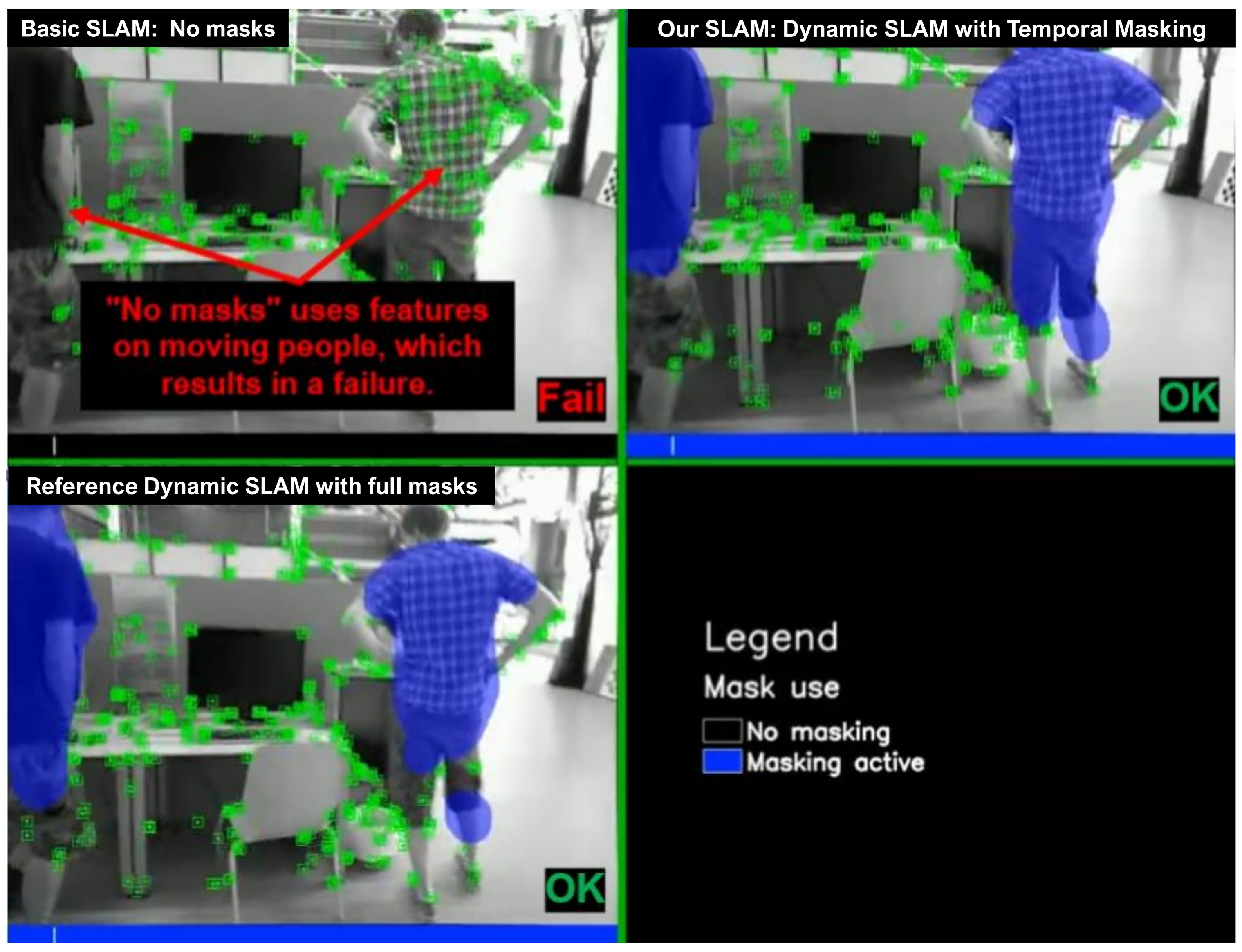


Sample temporal masks **uniformly** among all possible temporal masks (space size: $2^{\text{sequence length}}$)
using a binary tree whose root-to-leaf iterations preserve sampling uniformity.

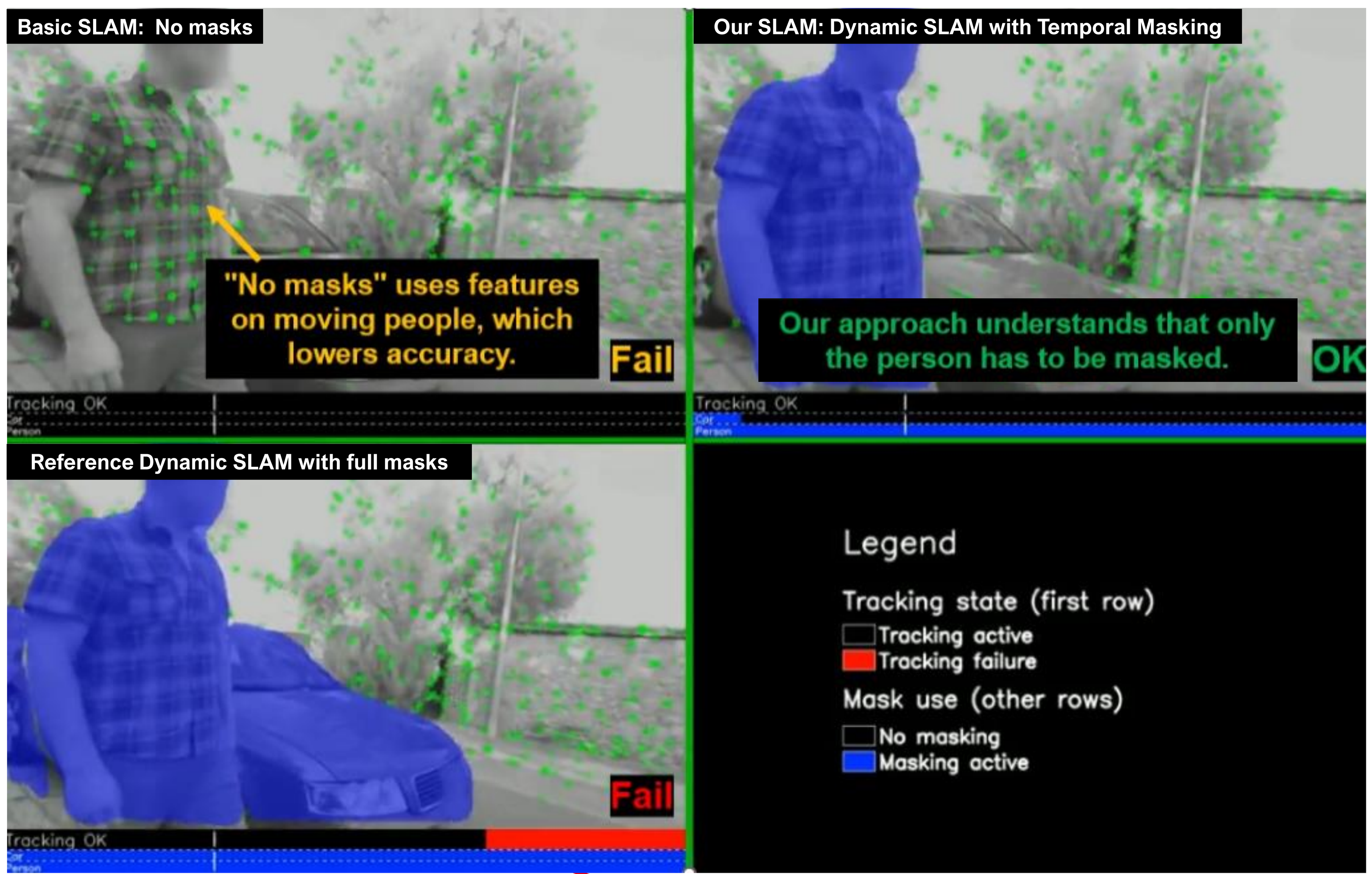
Additional contributions: USM Metric + ConsInv Dataset (150 sequences)

Results

To mask or not to mask, our network shall learn.



Preventing drift in TUM RGB-D dataset



Preventing drift and excessive masking in ConsInv-Outdoors dataset

Later in the same sequence: failure of "Full Masks" due to excessive masking.

Mode	Dataset	Metric	Baselines: No masks	Baselines: Full masks	State of the Art: Optical flow	State of the Art: DynaSLAM	State of the Art: Slamantic	State of the Art: StaticFusion	Ours
RGB-D	TUM RGB-D	ATE RMSE (m) ↓	0.105	**0.019**	-	**0.019**	0.028	0.099	**0.019**
		Tracking Rate ↑	**96%**	**96%**	-	69%	**96%**	**96%**	**96%**
		USM ↑	0.55	**0.80**	-	0.57	0.76	0.54	**0.80**
Stereo	KITTI	ATE RMSE (m) ↓	2.59	2.67	-	2.74	2.70	-	**2.51**
		Tracking Rate ↑	**100%**	**100%**	-	**100%**	**100%**	-	**100%**
		USM ↑	0.80	0.80	-	0.79	0.80	-	**0.81**
Stereo	ConsInv-Outdoors	ATE RMSE* ↓	0.084	0.019	-	0.025	0.032	-	0.024
		Tracking Rate* ↑	100%	75%	-	74%	85%	-	88%
		USM ↑	0.61	0.81	-	0.80	0.82	-	**0.88**
Mono	ConsInv-Indoors-Dynamic	ATE RMSE* ↓	0.074	0.003	0.050	0.010	0.032	-	0.014
		Tracking Rate* ↑	94%	74%	74%	70%	84%	-	84%
		USM ↑	0.57	0.71	0.49	0.63	0.68	-	**0.75**
Mono	ConsInv-Extra-MeetingRoom (*domain shift*)	ATE RMSE* ↓	0.170	0.012	0.077	0.011	0.077	-	0.012
		Tracking Rate* ↑	96%	73%	62%	66%	86%	-	76%
		USM ↑	0.33	0.65	0.34	0.60	0.54	-	**0.67**
Mono	ConsInv-Extra-LivingRoom (*domain shift*)	ATE RMSE* ↓	0.091	0.012	0.012	0.020	0.016	-	0.013
		Tracking Rate* ↑	96%	82%	62%	71%	84%	-	85%
		USM ↑	0.51	0.73	0.55	0.60	0.69	-	**0.74**
Mono	ConsInv-Indoors-Static	Prevented false starts ↑	56%	**100%**	67%	**100%**	78%	-	**100%**

Comparison with the state of the art on various datasets in their preferred mode. Our method outperforms the state of the art. Unlike the USM, ATE RMSE (trajectory accuracy) and Tracking Rate (% of tracked frames) may be misleading in difficult scenarios.

Conclusion

With the proposed Temporal Masking paradigm, we overcame the current limits of Dynamic SLAM on real data, especially in difficult scenarios.

References: *DynaSLAM*: Bescos *et al.*, IEEE RA-L, 2018. | *Slamantic*: Schorghuber et al., ICCVW Proceedings, 2019. | *StaticFusion*: Scona *et al.*, ICRA, 2018.