UL-VIO: Ultra-lightweight Visual-Inertial Odometry with Noise Robust Test-time Adaptation

Jinho Park[1], Se Young Chun[2], Mingoo Seok[1]

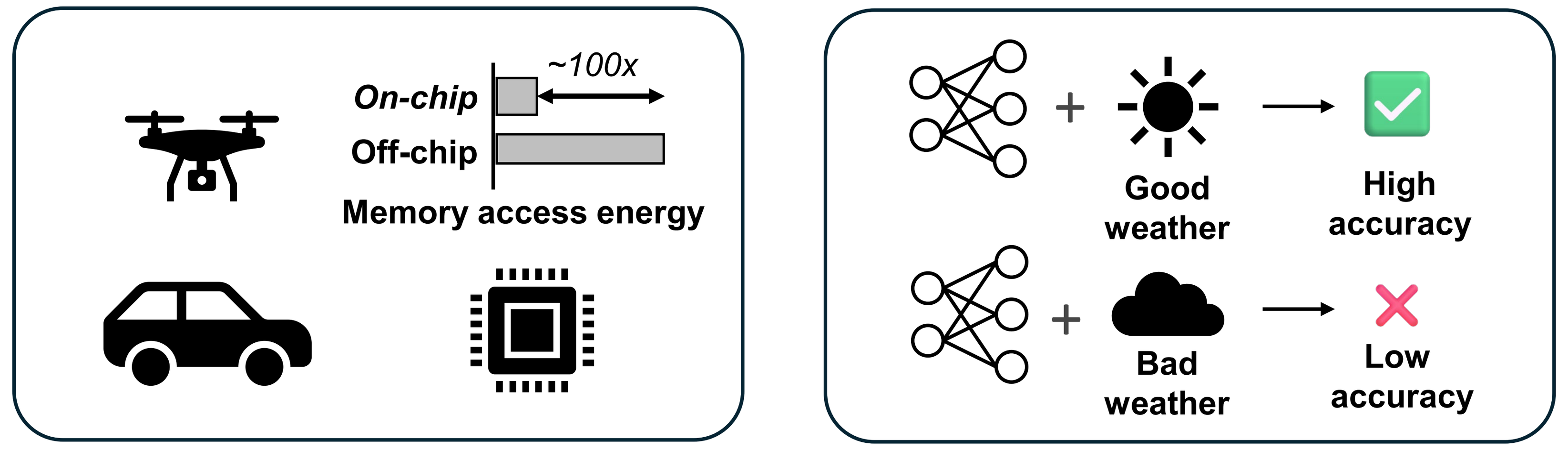
[1]Columbia University, [2]Seoul National University

Sponsored by: Semiconductor Research Corporation, CogniSense

Project page

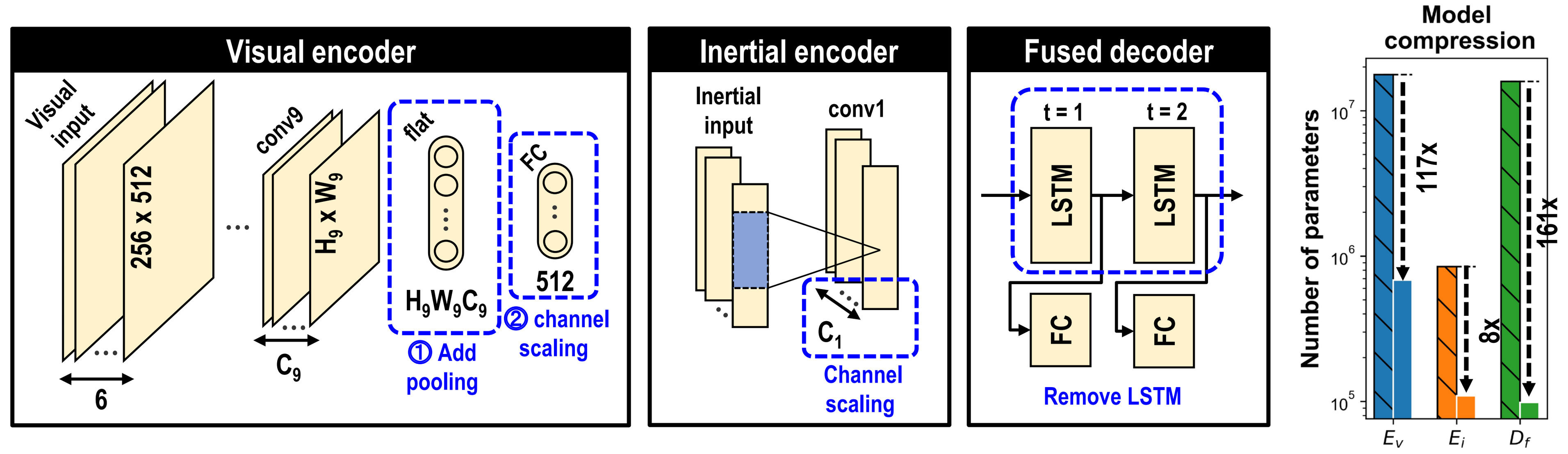
Linkedin

Motivation for Lightweight & Robust Network



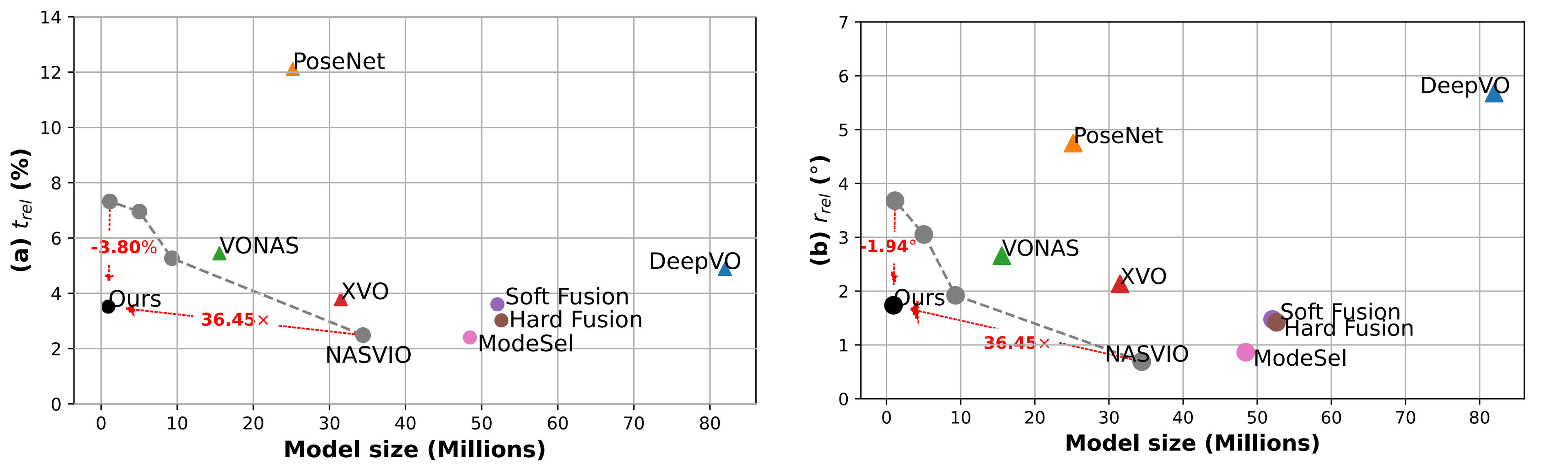
- Need for **lightweight** neural network to be hosted entirely by *on-chip* memory, which is of few MB in modern processors
- Robust to **environmental factors** and **noises** inducing distribution shifts

Model Compression



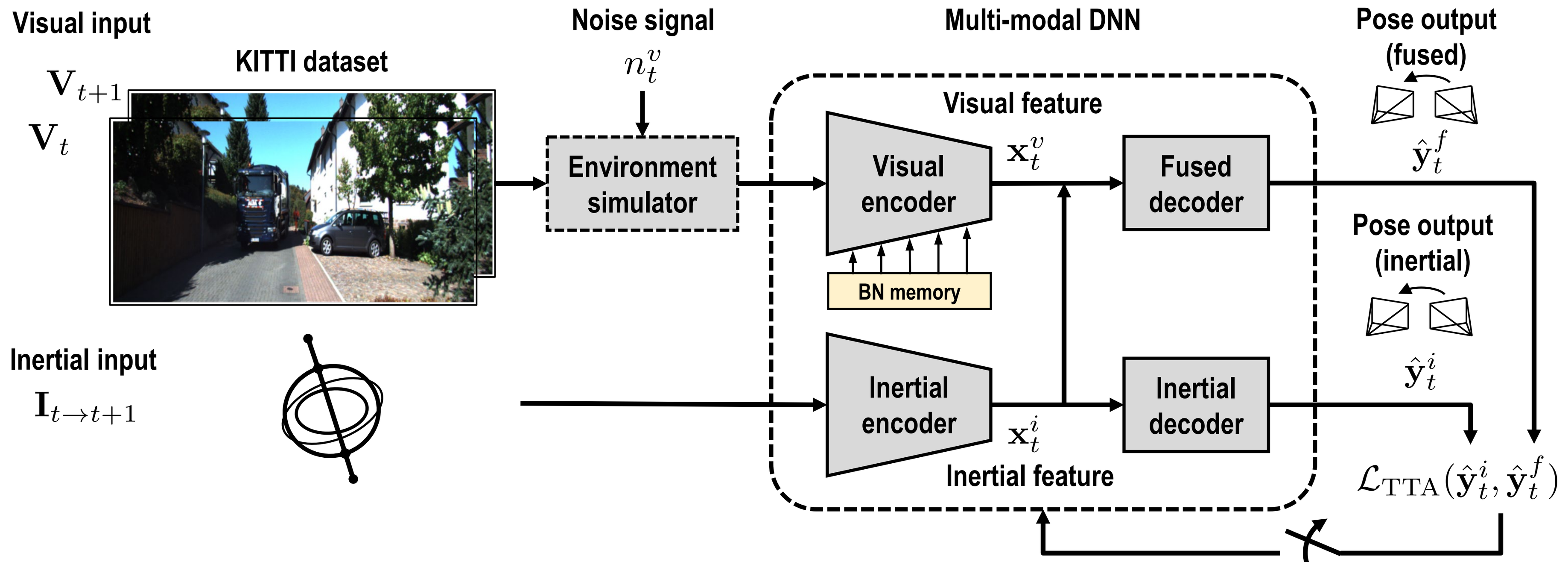
- **Channel scaling** – model parameter quadratically proportional to channel size in E_v, E_i, D_f
- Addition of **AvgPool** prior to FullyConnected (FC) to reduce feature size in E_v
- **Replace LSTM** w/ FC layer to reduce the model size in D_f

Model Compression (results)



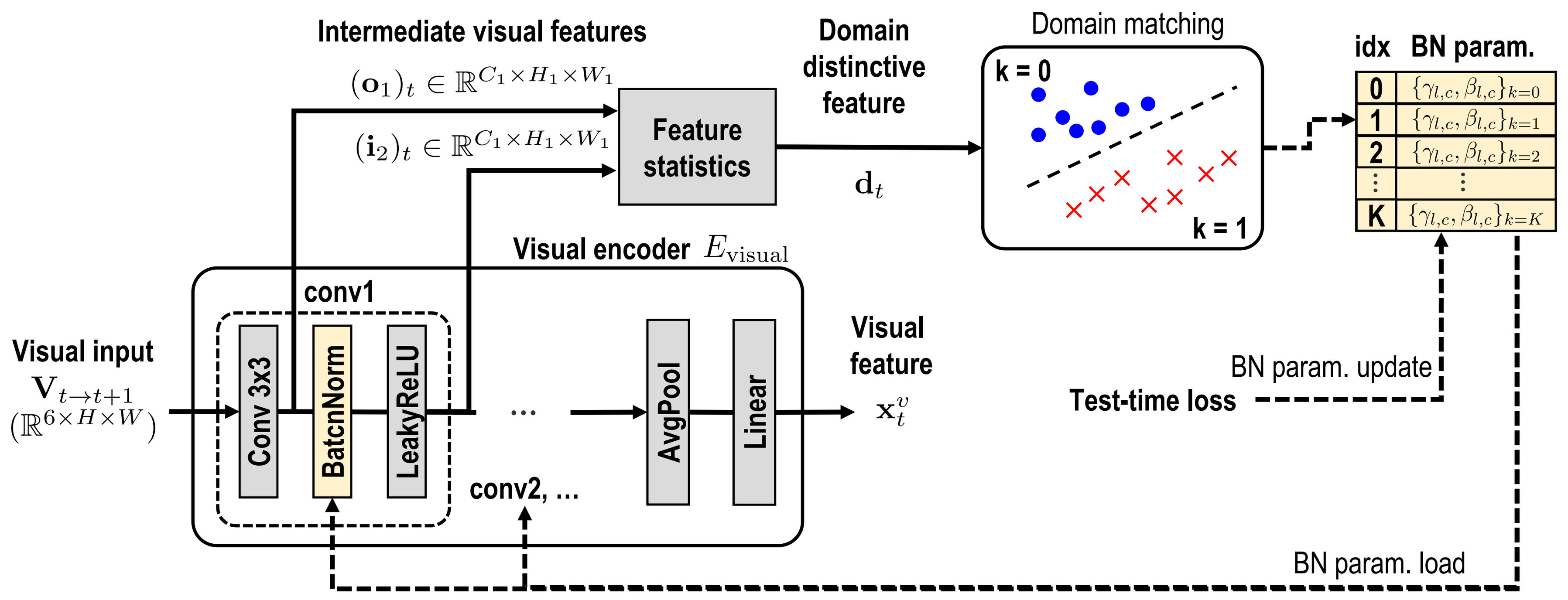
- We achieve **36x model size reduction** w/ a minute (1%) increase in absolute pose estimation error
- Latest **Apple A16** and **Qualcomm Snapdragon** CPUs possess only a **few MB** of *on-chip* memory

Multi-modal Consistency-based TTA



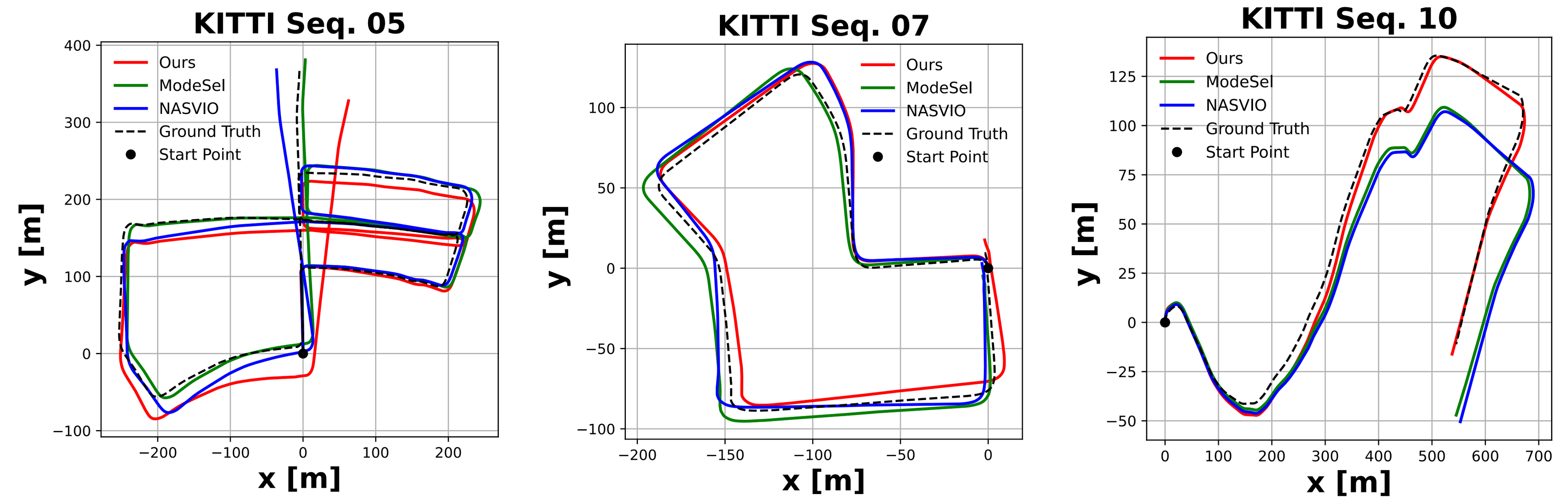
- **Visual encoder** deduces visual features from pairwise images
- **Inertial encoder** does so from the inertial input
- **Decoder** predicts pose transformation from fused features

Dictionary-based Visual Encoder Adaptation

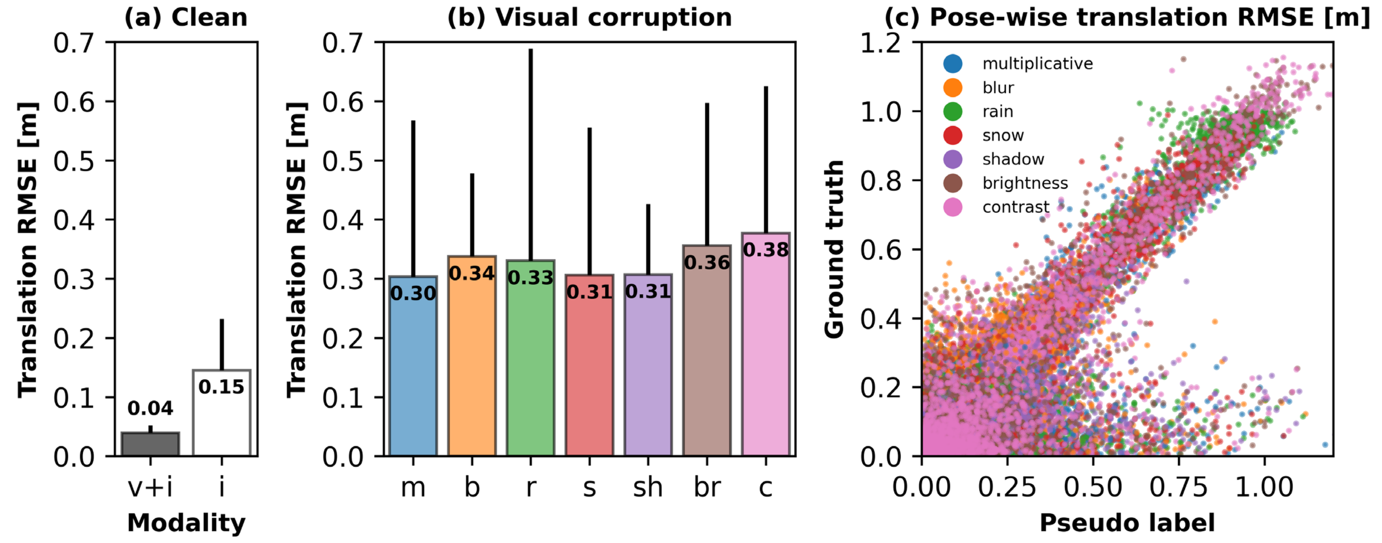


- Forward-path generate **visual feature** x^v, which will be used for decoder in the deeper layers
- First layer conv output $\mathbf{o}_1$ is used to generate **domain distinctive feature** $\mathbf{d}_t$
- Domain k is found using l_2 distance with the proxies $\{\mathbf{d}^k\}$
- **BatchNorm** parameter corresponding to the domain k is updated

Trajectory Results for UL-VIO



Motivation for Multi-modal Consistency



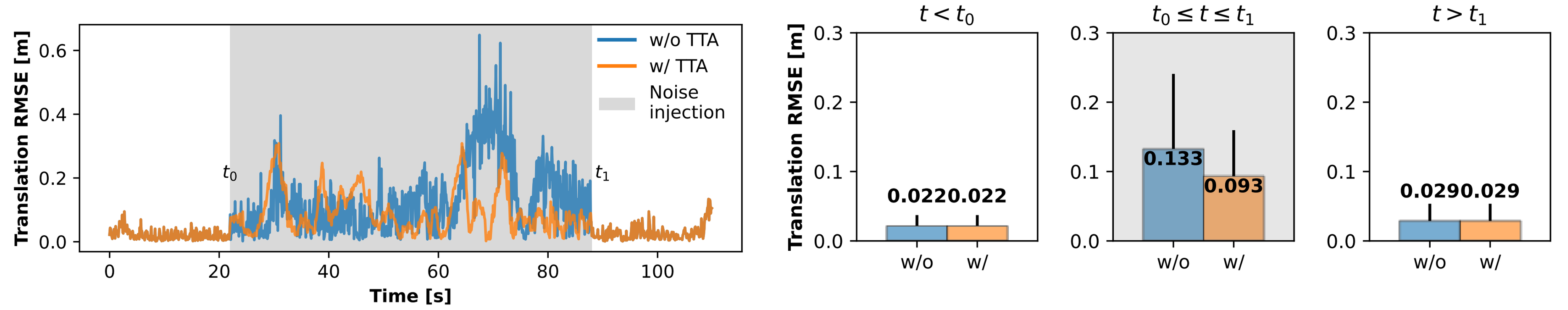
- While using **visual feature** along with inertial is preferred , inertial-only result become reliable under noise
- Inertial-only pose output (**pseudo label**) is highly correlated with the **ground-truth** target pose

Against Fine-tuned Baselines

Model		Average pose-wise t_{rmse} [m] Clean	Multi.	Blur	Rain	Snow	Shadow	Bright.	Cont.
Source		0.059	0.154	<u>0.261</u>	0.176	0.191	0.203	0.226	<u>0.250</u>
Fine-tuned with adver. noise (FT)	Multi.	0.099	<u>0.129</u>	0.394	0.227	0.372	0.192	0.299	0.331
	Blur	0.115	0.176	0.263	0.193	0.247	0.184	0.242	0.261
	Rain	0.289	0.325	0.372	**0.095**	0.394	0.311	0.525	0.531
	Snow	0.091	0.148	0.319	0.263	<u>0.183</u>	0.208	0.369	0.450
	Shadow	0.085	**0.112**	0.322	0.179	0.243	**0.121**	<u>0.221</u>	0.252
	Bright.	0.091	0.151	0.312	0.177	0.226	0.185	0.233	0.278
	Cont.	0.093	0.150	0.330	0.197	0.219	0.184	0.237	0.273
TTA (ours)		-	0.156	**0.230**	<u>0.143</u>	**0.172**	<u>0.155</u>	**0.193**	**0.212**

- Except for one case, e.g., multiplicative noise, our TTA method has **the best or 2nd best accuracy**

Continual TTA



- Translation RMSE quickly increases when the **visual corruption** is applied
- **TTA** mitigates the increase in the error to some extent

Continual TTA w/ Dynamic Noise Shifts

Time $t \longrightarrow$

Seq. / Noise	Seq. 05 Blur	Rain	Snow	Con.	Seq. 07 Blur	Rain	Snow	Con.	Seq. 10 Blur	Rain	Snow	Con.	Avg.
Baseline	0.118	0.121	**0.103**	0.166	0.127	0.153	0.110	0.191	0.137	0.134	**0.120**	0.167	0.137
TTA	**0.112**	**0.107**	0.110	**0.107**	**0.101**	**0.108**	**0.106**	**0.104**	**0.123**	**0.124**	0.121	**0.127**	**0.113**
***ddf* acc.**	97.9	100	100	100	98.2	100	100	100	98.8	100	100	100	99.6

- We inject different types of noise {blur, rain, snow, contrast} in a continuous manner for **KITTI**

Acknowledgements: This work was supported in part by COGNISENSE, one of seven centers in JUMP 2.0, a Semiconductor Research Corporation (SRC) program sponsored by DARPA. The work of SY Chun was supported by the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) [NO.RS-2021-II211343, Artificial Intelligence Graduate School Program (Seoul National University)] and by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) [No. NRF-2022M3C1A309202211]