Learning to Optimize State Estimation in Multi-agent Reinforcement Learning-based Collaborative Detection

Tianlong Zhou[1], Tianyi Shi[1], Hongye Gao[1], Weixiong Rao[1]
[1]Tongji University, Shanghai, China

TONGJI UNIVERSITY

MoDI Mobile and Data Intelligence Lab

Introduction

- **Motivation**: Collaborative detection has attracted many researchers for its wide range of application scenarios, such as search-and-rescue (SAR), environmental monitoring, and active perception. As a basic component of collaborative detection, existing approaches often solve the state estimation problem with classic Bayesian filters, which impractically assume that the underlying state space model is fully known. Some recent learning-based works, i.e., KalmanNet, estimate target states alone but without estimation uncertainty and cannot make robust control decisions. To tackle such issues, we develop a multi-agent reinforcement learning (MARL) model, with a neural network-based state estimator, namely TWo-phase KALmaN Filter with Uncertainty quanTification (WALNUT).

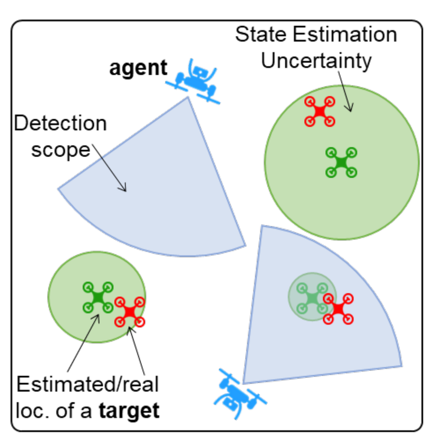

Fig. 1: Collaborative Detection.

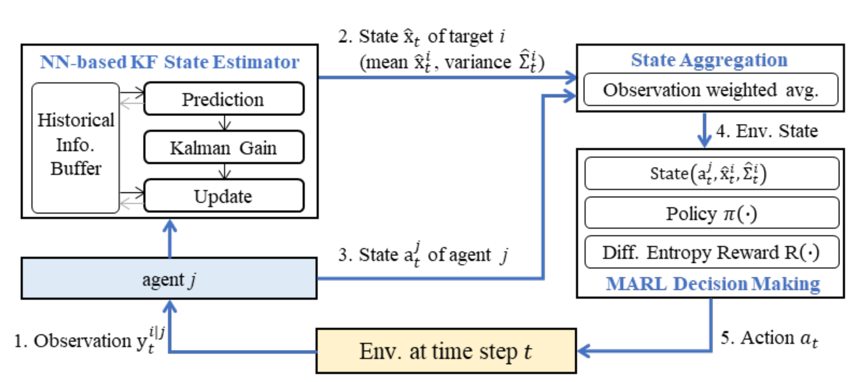

Fig. 2: Overall Framework.

- **Contributions**:
  - ➢ As an interpretable estimator, WALNUT explicitly gives both target state mean and covariance together in a data-driven fashion. By training a multi-task model, WALNUT works well on a partially known SS model even given sparse target observation samples.
  - ➢ We develop a MARL-based collaborative detection framework. The MARL model takes target state mean and covariance as input and then makes robust decisions for mobile agents to track movable targets.
  - ➢ We evaluate our work on both a synthetic environment and a real-world dataset.

Method

Overall Framework:
To solve the robust decision problem, the proposed framework first trains the state estimator WALNUT. Subsequently, the MARL decision maker will aggregate the estimated results from all agents to construct the environment state and make robust decisions by utilizing the well designed state.

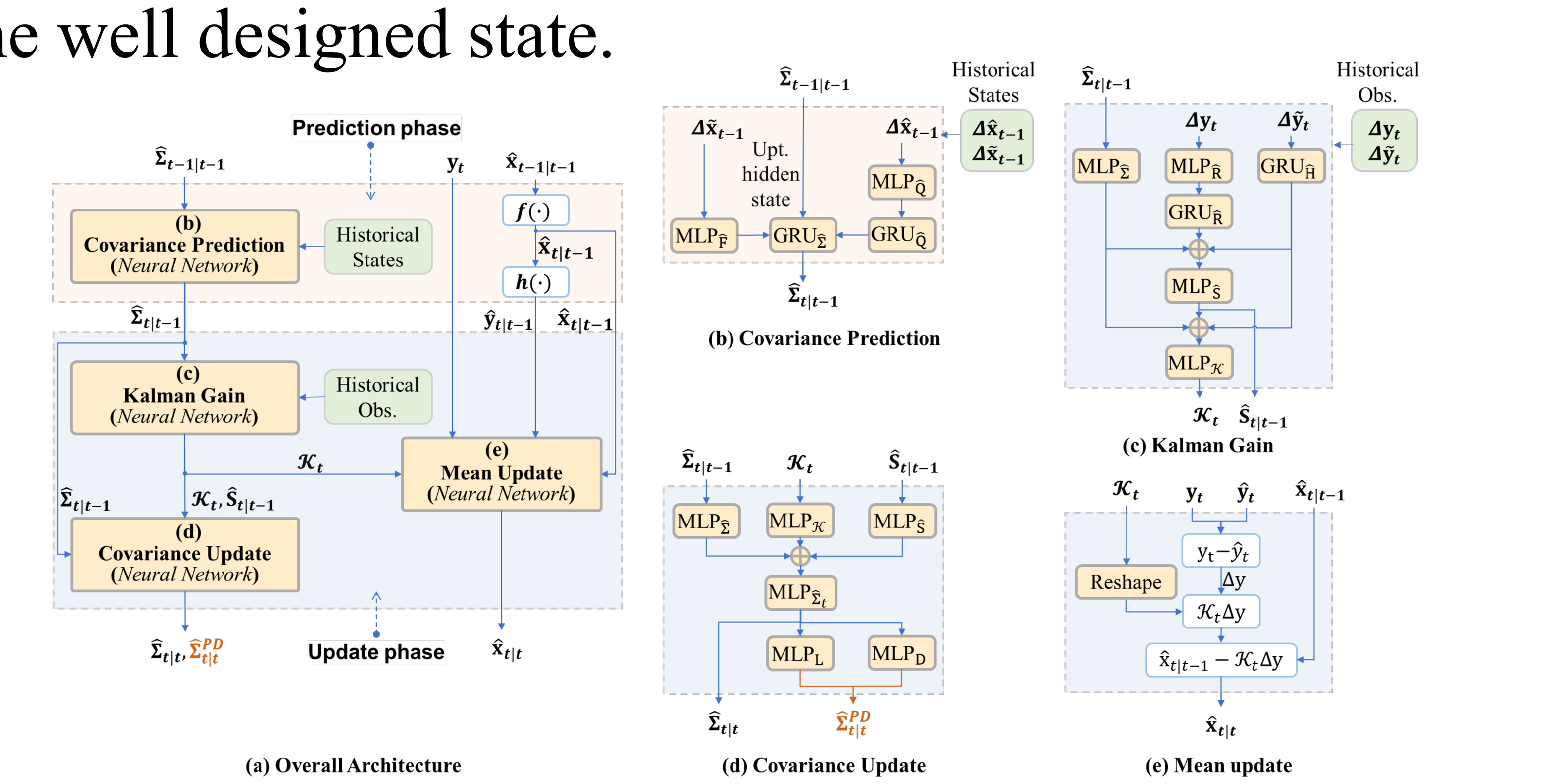

Fig. 3: (a) Overall framework of WALNUT. Four neural networks (NNs): (b) Covariance Prediction, (c) Kalman Gain, (d) Covariance update, and (e) Mean update.

WALNUT:
Following the prediction and update steps of EKF, WALNUT exploits neural networks (NNs) to perform state estimation (state mean and associate uncertainty). Each NN is designed to approximate the equations in EKF. Thanks to the developed NNs, WALNUT works well when given the partially known SS model and sparse observation samples.

MARL Decision Maker:
The decision maker can aggregate state estimations with theoretical guarantees. Using the aggregated environment state, the decision maker leverages the uncertainty estimated by WALNUT to construct a differential entropy reward for training a MARL policy. For further improvement, the policy also collects samples to fine-tune WALNUT in an online manner.

Experiments

TABLE 1: Baseline study in a collaborative detection environment.

		State				Uncertainty			Detection(%)
		MSE[dB]	%	MAE[dB]	%	MNLL	PICP	MPIW	Ratio
4a4t	ScalableMARL	-8.21 ± 1.80	0	-6.89 ± 0.74	0	15.83	0.48	5.00E-04	70.0 ± 0.0
	KalmanNet	-11.21 ± 1.71	36.5	-8.33 ± 0.65	20.9	-	-	-	82.4 ± 4.8
	TPKN	-11.45 ± 1.51	39.5	-8.81 ± 0.62	27.9	-	-	-	84.3 ± 3.3
	WALNUT	**-11.96 ± 1.84**	**45.7**	**-9.03 ± 0.74**	**31.1**	**-2.57**	**0.83**	**3.1E-5**	**85.1 ± 4.0**
4a2t	ScalableMARL	-13.09 ± 1.79	0	-9.01 ± 0.76	0	13.3	0.45	1.20E-04	70.0 ± 0.0
	KalmanNet	-15.81 ± 1.28	20.8	-10.00 ± 0.40	11	-	-	-	75.2 ± 9.1
	TPKN	-16.43 ± 1.53	25.5	-10.52 ± 0.47	16.8	-	-	-	82.1 ± 5.6
	WALNUT	**-16.85 ± 1.22**	**28.7**	**-10.86 ± 0.44**	**20.5**	**-3.58**	**0.82**	**3.8E-6**	**87.1 ± 4.6**
2a4t	ScalableMARL	-5.36 ± 2.08	0	-5.54 ± 0.97	0	15.67	0.58	1.70E-02	70.0 ± 0.0
	KalmanNet	-6.25 ± 1.34	16.6	-6.14 ± 0.54	10.8	-	-	-	71.5 ± 4.2
	TPKN	-6.95 ± 1.21	29.7	-6.67 ± 0.53	20.4	-	-	-	74.5 ± 3.6
	WALNUT	**-8.35 ± 1.23**	**55.8**	**-7.36 ± 0.54**	**32.9**	**-1.66**	**0.87**	**1.1E-4**	**78.8 ± 3.4**

TABLE 2: Comparison on Known and Unknown Training Maps in a 4a4t Collaborative Detection Setting.

	Test	Training: w/ obt.	Training: wo/ obt.
WALNUT	w/ obt.	-8.54 ± 1.13	-9.81 ± 1.58
	wo/ obt.	-10.74 ± 1.47	-11.96 ± 1.84
ScalableMARL	w/ obt.	-6.32 ± 1.19	-6.42 ± 1.36
	wo/ obt.	-8.06 ±1.07	-8.21 ± 1.80

TABLE 3: Effect of Partial Known SS Model.

α	0°	10°	20°
KF	-19.743 ± 0.863	-19.658 ± 0.814	-18.985 ± 0.528
KalmanNet	-21.337 ± 0.075	-21.219 ± 0.097	-20.321 ± 0.155
WALNUT-H	**-21.396 ± 0.071**	-20.569 ± 0.015	-19.033 ± 0.066
WALNUT	-21.393 ± 0.062	**-21.263 ± 0.085**	**-20.448 ± 0.126**

TABLE 4: Effect of Observation Noise.

ρ	0.5	1.0	2.0
KF	-23.470 ± 0.459	-19.386 ± 0.699	-16.364 ± 0.608
KalmanNet	-24.380 ± 0.156	-19.474 ± 0.172	-17.906 ± 0.065
WALNUT	**-24.494 ± 0.147**	**-20.858 ± 0.093**	**-18.000 ± 0.073**

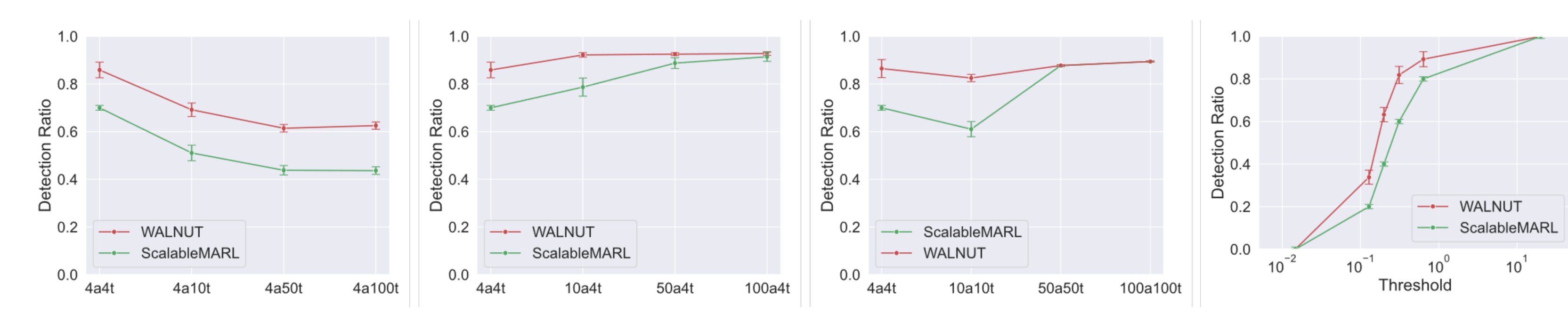

Fig. 4: Scalability Study. From left to right: (a) 4axt: 4 agents and various targets, (b) xa4t: various agents and 4 targets, (c) xaxt: various agents and targets, and (d) 4a4t with various thresholds.

Visualization

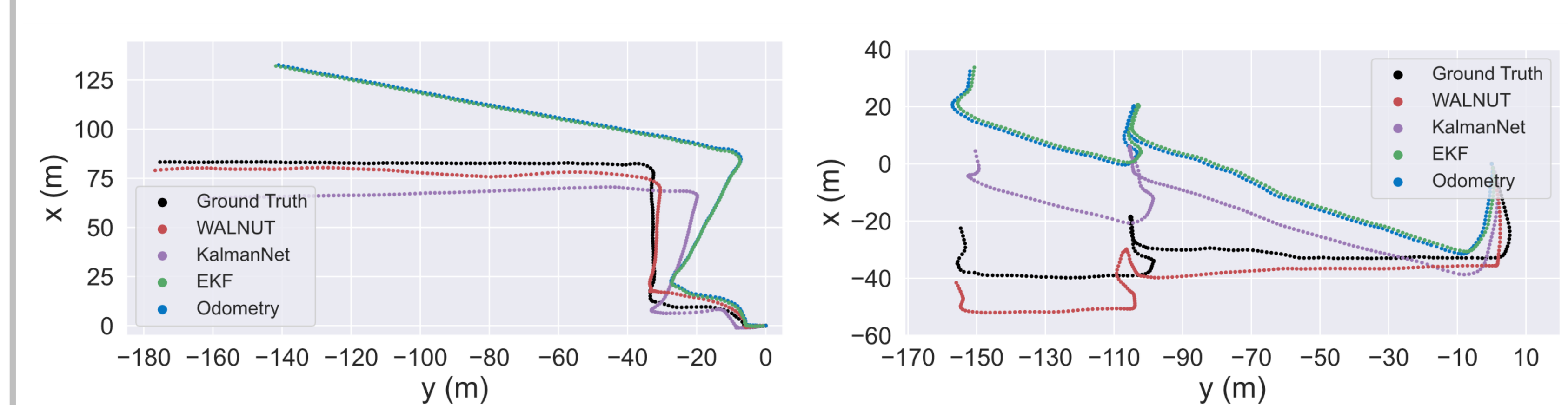

Fig. 5: Study on the real world NCLT dataset. The dataset utilizes ground robots to collect driving trajectories and sensor data on campus. We visualize the driving trajectories recovered by the baseline algorithm through observation with Ground Truth. From left to right: (a-b) Result of two test trajectories.