TrajPAC: Towards Robustness Verification of Pedestrian Trajectory Prediction Models

TrajPAC：行人轨迹预测模型的鲁棒性验证方法

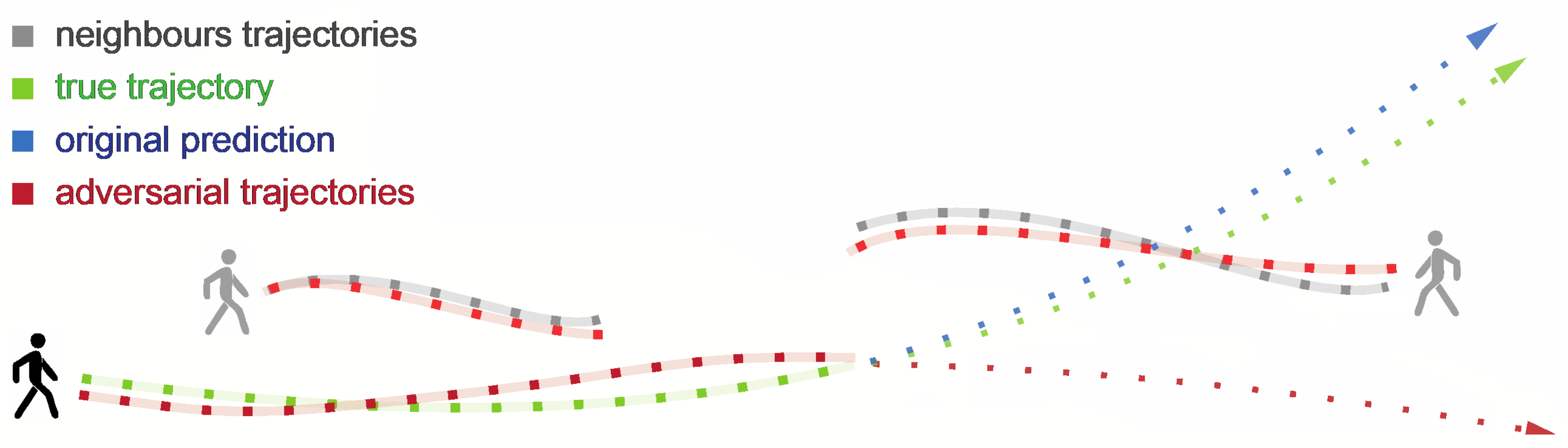
张亮*，徐星成，杨鹏飞，金高杰，黄承超，张立军

International Conference on Computer Vision (ICCV’23)

*Tel: 18811728908, E-mail: zhangliang@ios.ac.cn

Motivation

Robust pedestrian trajectory forecasting is crucial to developing safe autonomous vehicles. While existing methods have shown impressive results, their vulnerability to adversarial attacks poses significant security risks.



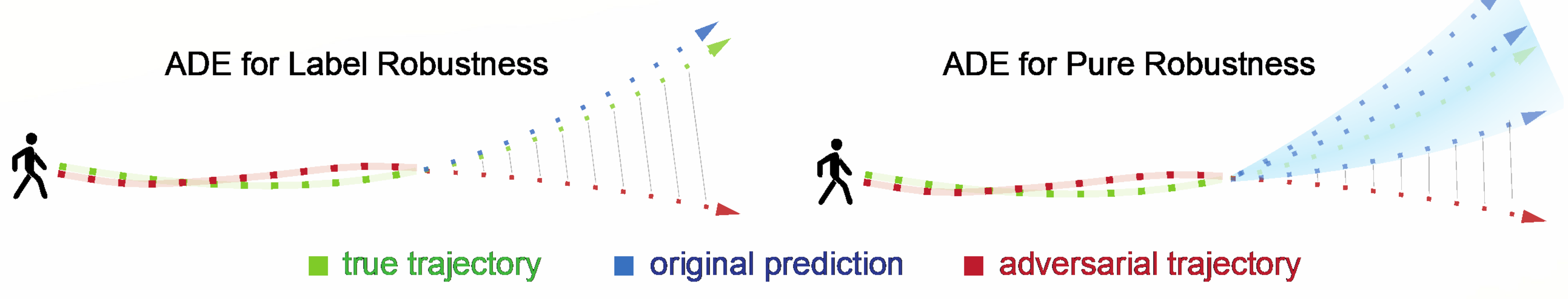
Methodology

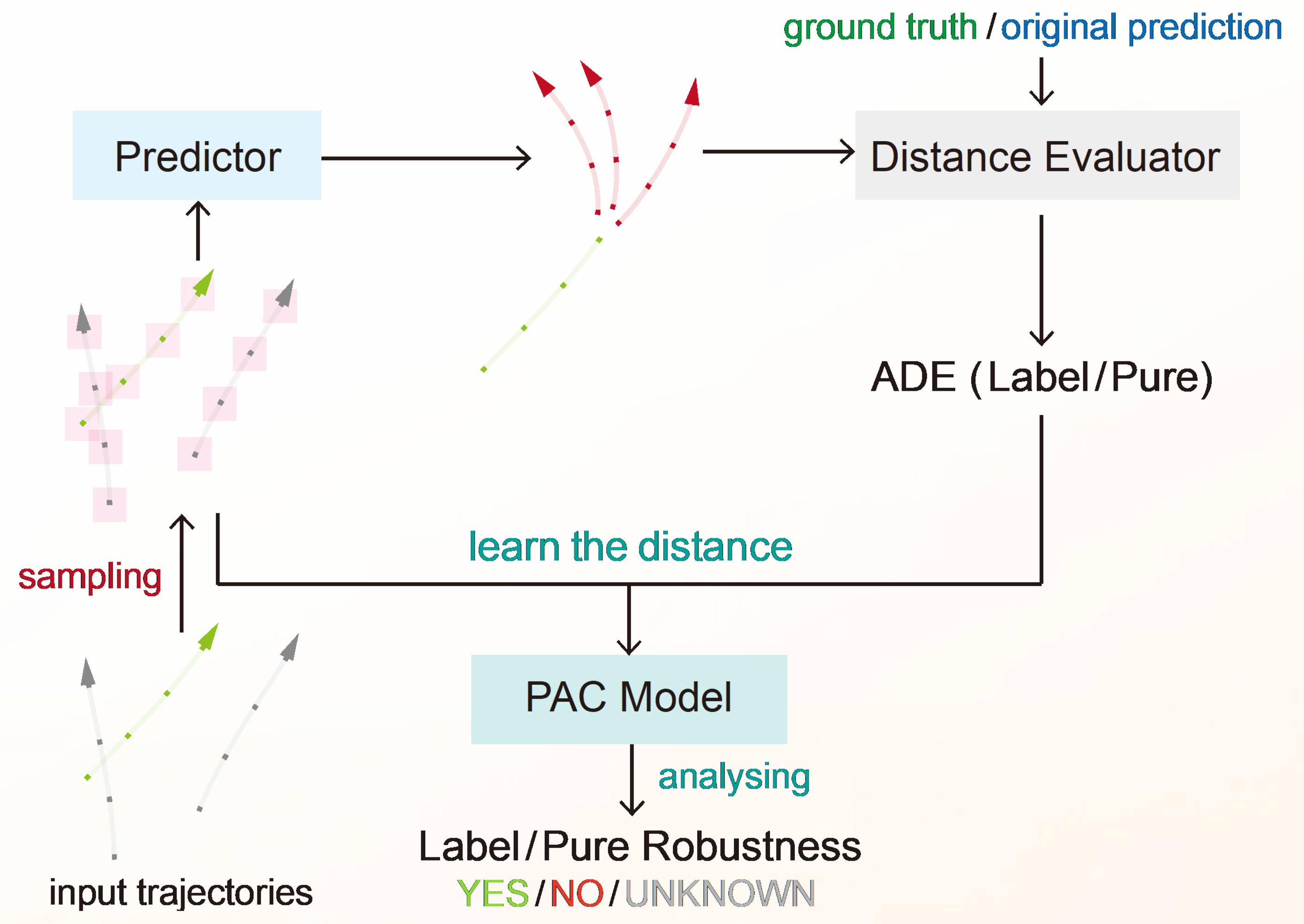
We present two formal definitions of robustness in trajectory prediction:

- Label Robustness: quantifies robustness in prediction accuracy after attacks
- Pure Robustness: measures robustness in prediction stability after attacks

We propose TrajPAC, a framework with three basic functions:

- Verifying the robustness of trajectory forecasting models
- Identifying potential counterexamples
- Providing interpretable analyses of the original methods

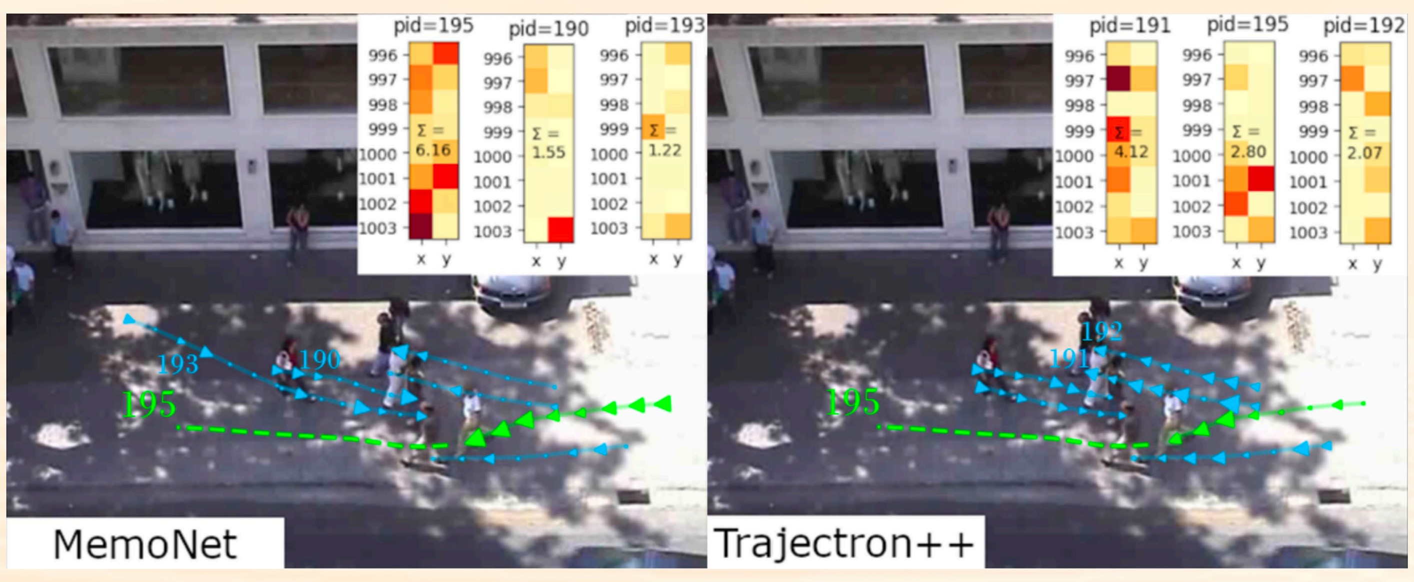




Evaluation & Results

- Good scalability, efficiency and soundness in robustness analysis of different trajectory prediction models.
- Comparable (and even better) performance to adversaries found by PGD.
- Identifying key features that contribute to the overall performance and robustness

Scene	ID	Label Robustness Traj++	Label Robustness Memo	Label Robustness AgentF	Label Robustness MID	Pure Robustness Traj++	Pure Robustness Memo	Pure Robustness AgentF	Pure Robustness MID
ETH	(4400, 79)	○	✓	✗	✗	○	✓	✗	✓
	(6490, 127)	✓	✓	✗	✗	✓	✓	✗	✓
	(10340, 257)	○	○	✗†	✗	○	✓	✗	✗
Hotel	(7550, 157)	✓	✓	✗	○	✓	✓	✓	✓
	(10530, 236)	✓	✓	○	✓	✓	✓	✓	✓
	(15030, 345)	✓	✓	✓	✓	✓	✓	✓	✓
Zara1	(4430, 69)	○	✓	✗	✗	✓	✓	✓	✓
	(6050, 102)	✓	✓	✗	✗	✓	✓	○	✓
	(8680, 142)	✗†	○	✗†	✓	○	✓	✓	✓
Zara2	(3400, 65)	✓	✓	✗	✓	✓	✓	✓	✓
	(7430, 141)	✓	✓	✗	✗	✓	✓	○	○
	(10030, 195)	✗	✓	✗	✗	✗	○	○	✓
Univ	(1840, 105)	✗	✗	✗†	✗	○	✓	○	✓
	(4820, 202)	✗†	✗	✗†	○	○	✓	○	✓
	(5250, 297)	✓	✓	✗	✓	○	✓	○	✓



In our subsequent work, we have presented a testing platform to comprehensively assess the robustness and generalizability of autonomous driving trajectory prediction models across various datasets, models, and attack methods.