JRDB-Traj: A Dataset and Benchmark for Trajectory Forecasting in Crowds

Saeed Saadatnejad[1], **Yang Gao**[1], **Hamid Rezatofighi**[2] **and Alexandre Alahi**[1]
[1]EPFL, [2]Monash University
{saeed.saadatnejad}@epfl.ch, {hamid.rezatofighi}@monash.edu

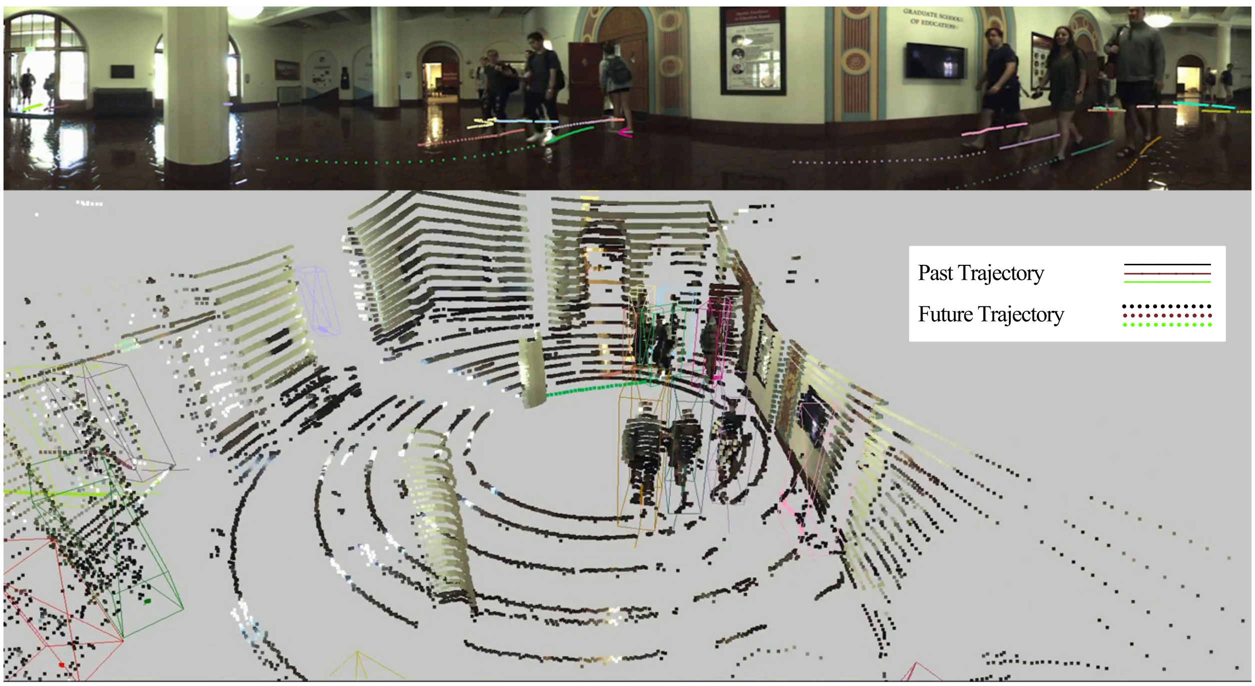

Figure 1: Trajectory Forecasting in JRDB-Traj dataset. The top figure displays the RGB image from the robot's perspective, while the bottom figure shows the corresponding 3D point cloud. Solid lines indicate the observed past trajectories and dots represent the ground-truth future trajectories.

Abstract

Predicting future trajectories is critical in autonomous driving, especially in preventing accidents involving humans, where a predictive agent's ability to anticipate in advance is of utmost importance. Trajectory forecasting models, employed in fields such as robotics, autonomous vehicles, and navigation, face challenges in real-world scenarios, often due to the isolation of model components. To address this, we introduce a novel dataset for end-to-end trajectory forecasting, facilitating the evaluation of models in scenarios involving less-than-ideal preceding modules such as tracking. This dataset, an extension of the JRDB dataset, provides comprehensive data, including the locations of all agents, scene images, and point clouds, all from the robot's perspective. The objective is to predict the future positions of agents relative to the robot using raw sensory input data. It bridges the gap between isolated models and practical applications, promoting a deeper understanding of navigation dynamics. Additionally, we introduce a novel metric for assessing trajectory forecasting models in real-world scenarios where ground-truth identities are inaccessible, addressing issues related to undetected or over-detected agents. Researchers are encouraged to use our benchmark for model evaluation and benchmarking. The leaderboard and code are publicly available.

1 Introduction

The ability to predict future events is widely regarded as a fundamental aspect of intelligence [3]. This predictive capability assumes paramount significance in the context of autonomous driving, where precise predictions play a pivotal role in preventing accidents involving humans. A predictive agent possesses the foresight to anticipate the agents' actions few seconds in advance, enabling it to make informed decisions about when to stop or proceed safely. Trajectory forecasting models are aimed at predicting the future positions of agents based on a sequence of past observed locations. These models have been used in socially-aware robotics [4], autonomous vehicles [16], and navigation [9].

With the successful development of deep learning, autonomous driving algorithms have undergone a significant transformation, involving a complex series of interconnected tasks such as object detection, tracking, and trajectory forecasting. In the realm of industry solutions, it is common to deploy standalone models for each of these tasks, often evaluating and comparing them independently without considering their interdependencies. In other words, it assumes that previous modules function optimally. In practice, early modules may exhibit imperfections that can result in less-than-ideal trajectory forecasting.

We present a novel dataset designed for end-to-end trajectory forecasting in order to study the performance given non-perfect previous modules such as tracking. This is an extension of JRDB [12] dataset viewed from the perspective of a robot navigating within a dynamic environment. The task is to predict the future positions of all agents within the scene relative to the robot, leveraging raw sensory input data, point clouds and images. By focusing on end-to-end trajectory forecasting, our goal is to bridge the gap between isolated models and practical applications, fostering a deeper understanding of real-world navigation dynamics.

Notably, assessing the future trajectory prediction performance for multiple agents poses a challenge. Common metrics such as Average Displacement Error (ADE) and Final Displacement Error (FDE) cannot be employed because complete and accurate observed trajectories are unavailable. In essence, we lack the associated identities (IDs) required to calculate the displacement error accurately. Moreover, the inclusion of detection and tracking models can lead to instances where agents are either not detected or are over-detected, subsequently affecting the input data provided to the forecasting model. To address these issues, we introduce a novel, comprehensive metric for evaluating trajectory forecasting models in a two-step process involving matching and measuring displacement.

We made our benchmark publicly accessible, inviting researchers to submit their trajectory prediction models for evaluation and benchmarking against this new metric. The leaderboard and code are publicly available.

2 The JRDB-Traj Dataset

JRDB [12], JRDB-Act [5] and JRDB-Pose [19] previously introduced annotations including 2D and 3D bounding boxes with tracking IDs, action labels, social groups and body pose. We leverage the 3D bounding box annotations and make the trajectories using the center of the bounding box of the person on the ground.

2.1 Splits

We follow the official splits of JRDB [12] to create training, validation, and testing splits from the 54 captured sequences, with each split containing an equal proportion of indoor and outdoor scenes as well as scenes captured using a stationary or moving robot. All frames from a scene appear strictly in one split. The videos and point clouds for the last five seconds of the test are hidden.

2.2 Evaluation Metrics

Assessing trajectory forecasting performance in the absence of ground-truth IDs necessitates the establishment of associations between predicted and ground-truth trajectories in future frames, followed by their distance measurement—a standard approach in detection and tracking evaluations. Therefore, we report two prevalent detection and tracking metrics in these future frames:

1. IDF1 [15]: This is the ratio of correctly identified detections over the average number of ground-truth and computed detections.

2. OSPA-2 [14]: Optimal Sub-Pattern Matching (OSPA) [18] is a multi-object performance evaluation metric which includes the concept of miss-distance in tracking. OSPA-2 has been further adapted to detection and tracking tasks. It is a set-based metric that can directly capture a distance between two sets of trajectories without a thresholding parameter.

Furthermore, we propose End-to-end Forecasting Error (EFE) for assessing trajectory forecasting in real-world scenarios. In short, EFE determines the associations between predicted and ground-truth trajectories, measures their distances, and accounts for any mismatches in the number of trajectories. Importantly, EFE refrains from penalizing early terminations in predicted trajectories. In practical terms, if a model predicts a trajectory for an agent that extends beyond the scene boundaries, it does not contribute to error. A comprehensive explanation follows.

Let $\mathbf{X} = \{X_1^{\mathcal{D}_1}, X_2^{\mathcal{D}_2}, \dots X_m^{\mathcal{D}_m}\}$ and $\mathbf{Y} = \{Y_1^{\mathcal{D}_1}, Y_2^{\mathcal{D}_2}, \dots Y_n^{\mathcal{D}_n}\}$ be the sets of trajectories for prediction and the ground-truth, respectively. Note $\mathcal{D}_i$ represents the time indices which track i exists (having a state-value). Then, we calculate the time average distance of every pair of tracks $X_i^{\mathcal{D}_i}$ and $Y_j^{\mathcal{D}_j}$:

$$\underset{\sim}{\widetilde{d}}(X_i^{\mathcal{D}_i}, Y_j^{\mathcal{D}_j}) = \sum_{t \in \mathcal{D}_i \cup \mathcal{D}_j} \frac{d_O\left(\{X_i^t\}, \{Y_j^t\}\right)}{|\mathcal{D}_i \cup \mathcal{D}_j|}, \tag{1}$$

where $t \in \mathcal{D}_i \cup \mathcal{D}_j$ is the time-step when either or both track presents. Note that $\{X_i^t\}$ and $\{Y_j^t\}$ are singleton sets, i.e., $\{X_i^t\} = \emptyset$ or $\{X_i^t\} = x_i^t \in \mathbb{R}^2$ and $\{Y_j^t\} = \emptyset$ or $\{Y_j^t\} = y_j^t \in \mathbb{R}^2$. Therefore, $d_O\left(\{X_i^t\}, \{Y_j^t\}\right)$ can be simplified into the following distance function, :

$$d_O(\{X_i^t\},\{Y_j^t\}) = \begin{cases} d_c(x_i^t, y_j^t) & \text{if } |\{X_i^t\}| \wedge |\{Y_j^t\}| = 1, \\ c & \text{if } |\{X_i^t\}| = 0 \ \&\ |\{Y_j^t\}| \ != 0, \\ 0 & \text{Otherwise}, \end{cases} \tag{2}$$

where $d_c(x_i, y_i) := min(c, d(x_i, y_i))$ indicates the euclidean displacement error with the cutoff distance c.

Finally, we obtain the distance, EFE, between two sets of trajectory tracks, i.e., $\mathbf{X}$ and $\mathbf{Y}$ by:

$$EFE(\mathbf{X}, \mathbf{Y}) = \frac{1}{n}\left(\min_{\pi \in \Pi_n} \sum_{i=1}^{m} \underset{\sim}{\widetilde{d}}(X_i^{\mathcal{D}_i}, Y_{\pi_i}^{\mathcal{D}_{\pi_i}}) + c * (n - m)\right), \tag{3}$$

if $n \geq m$. Π_n is the set of all permutations of $\{1, 2, \dots, n\}$. Note that $\underset{\sim}{\widetilde{d}}$ reflects the localization errors of trajectories, whereas $c * (n - m)$ reflects the cardinality error (false and missed trajectories) and we put $c = 5$ meters as the threshold penalty. If $m > n$:

$$EFE(\mathbf{X}, \mathbf{Y}) = \frac{1}{m}\left(\min_{\pi \in \Pi_m} \sum_{i=1}^{n} \underset{\sim}{\widetilde{d}}(X_{\pi_i}^{\mathcal{D}_{\pi_i}}, Y_i^{\mathcal{D}_i}) + c * (m - n)\right). \tag{4}$$

We further define $EFE(X, Y) = c$ if either X or Y is empty, and $EFE(\emptyset, \emptyset) = 0$. The code of the metrics and related details can be accessed in JRDB Toolkit.

Figure 2: Qualitative results of trajectory forecasting on JRDB-Traj dataset. Solid lines indicate the observed past trajectories and dots represent the predicted trajectories.

	EFE ↓	OSPA-2 ↓	IDF1 ↑
Zero-Velocity	2.981	3.082	49.431
PCENet [10]	3.939	3.961	23.813
OpenTraj [2]	3.498	3.579	37.127
SMEMO [11]	2.671	2.786	55.258
Social-LSTM [1]	2.646	2.76	54.673
Social-Pose [6]	2.558	2.675	56.298
HST [17]	2.506	2.627	58.392

Table 1: Quantitative evaluations of trajectory forecasting models on JRDB-Traj dataset.

2.3 Benchmark

In Table 1, we have evaluated the performance of some baselines on this dataset and task, starting by the simple Zero-Velocity, a simple yet competitive baseline. It repeats the last observed location of each agent as its future predicted locations.

We then evaluated the well-known Social-LSTM baseline [1]. Note that we forecast future trajectories by leveraging observed past trajectory estimates derived from detection and tracking algorithms. Here, we utilized the estimated detections by a pre-trained PiFeNet [8] and subsequently employed the Simpletrack [13] tracking algorithm to provide observed past trajectory estimates as inputs to all the aforementioned models. We employed the Social-LSTM code provided by TrajNet++ [7]. Nevertheless, it is essential to note that we predicted an agent's future trajectory when we had access to the last two observed data points within their trajectory. A qualitative example is available in Figure 2 with observed and predicted trajectories of all agents in the scene. Our code is available online: https://github.com/vita-epfl/JRDB-Traj.

We have also evaluated the performance of five other recent methods in trajectory forecasting, PCENet [10], OpenTraj [2], SMEMO [11], Social-Pose [6] and HST [17]. It shows that HST is currently the leader of this benchmark. Using our public benchmark, one can submit their model and evaluate their performance.

3 Conclusion

In this paper, we introduced JRDB-Traj, a new dataset and benchmark for trajectory forecasting from raw sensory inputs. We have also introduced EFE, a new metric for trajectory forecasting in crowds where ground-truth identities are inaccessible. We anticipate that this dataset will foster further research in this domain, bringing us closer to realizing a fully functional autonomous driving system suitable for practical applications. As future works, we will enrich this dataset with adding more fine-grained human motion (body keypoints) annotations to it.

ACKNOWLEDGMENT

The authors thank Simindokht Jahangard and Duy Tho Le for their valuable input.

References

[1] Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. Social lstm: Human trajectory prediction in crowded spaces. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pages 961–971, 2016. 4

[2] Javad Amirian, Bingqing Zhang, Francisco Valente Castro, Juan Jose Baldelomar, Jean-Bernard Hayet, and Julien Pettré. Opentraj: Assessing prediction complexity in human trajectories datasets. In *Proceedings of the asian conference on computer vision*, 2020. 4

[3] Andreja Bubic, D. Yves Von Cramon, and Ricarda Schubotz. Prediction, cognition and the brain. *Frontiers in Human Neuroscience*, 4:25, 2010. 2

[4] Changan Chen, Yuejiang Liu, Sven Kreiss, and Alexandre Alahi. Crowd-robot interaction: Crowd-aware robot navigation with attention-based deep reinforcement learning. In *International Conference on Robotics and Automation (ICRA)*, pages 6015–6022. IEEE, 2019. 2

[5] Mahsa Ehsanpour, Fatemeh Saleh, Silvio Savarese, Ian Reid, and Hamid Rezatofighi. Jrdb-act: A large-scale dataset for spatio-temporal action, social group and activity detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20983–20992, 2022. 2

[6] Yang Gao. Social-pose: Human trajectory prediction using input pose, 2022. 4

[7] Parth Kothari, Sven Kreiss, and Alexandre Alahi. Human trajectory forecasting in crowds: A deep learning perspective. *IEEE Transactions on Intelligent Transportation Systems*, 2021. 4

[8] Duy Tho Le, Hengcan Shi, Hamid Rezatofighi, and Jianfei Cai. Accurate and real-time 3d pedestrian detection using an efficient attentive pillar network. *IEEE Robotics and Automation Letters*, 8(2):1159–1166, 2022. 4

[9] Yuanfu Luo, Panpan Cai, Aniket Bera, David Hsu, Wee Sun Lee, and Dinesh Manocha. Porca: Modeling and planning for autonomous driving among many pedestrians. *IEEE Robotics and Automation Letters*, 3(4):3418–3425, 2018. 2

[10] Karttikeya Mangalam, Harshayu Girase, Shreyas Agarwal, Kuan-Hui Lee, Ehsan Adeli, Jitendra Malik, and Adrien Gaidon. It is not the journey but the destination: Endpoint conditioned trajectory prediction. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, pages 759–776. Springer, 2020. 4

[11] Francesco Marchetti, Federico Becattini, Lorenzo Seidenari, and Alberto Del Bimbo. Smemo: social memory for trajectory forecasting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 4

[12] Roberto Martin-Martin, Mihir Patel, Hamid Rezatofighi, Abhijeet Shenoi, JunYoung Gwak, Eric Frankel, Amir Sadeghian, and Silvio Savarese. Jrdb: A dataset and benchmark of egocentric robot visual perception of humans in built environments. *TPAMI*, 2021. 2

[13] Ziqi Pang, Zhichao Li, and Naiyan Wang. Simpletrack: Understanding and rethinking 3d multi-object tracking. In *European Conference on Computer Vision (ECCV)*, pages 680–696. Springer, 2022. 4

[14] Hamid Rezatofighi, Tran Thien Dat Nguyen, Ba-Ngu Vo, Ba-Tuong Vo, Silvio Savarese, and Ian Reid. How trustworthy are performance evaluations for basic vision tasks? *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2023. 3

[15] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In *European Conference on Computer Vision (ECCV)*, pages 17–35. Springer, 2016. 3

[16] Saeed Saadatnejad, Mohammadhossein Bahari, Pedram Khorsandi, Mohammad Saneian, Seyed-Mohsen Moosavi-Dezfooli, and Alexandre Alahi. Are socially-aware trajectory prediction models really socially-aware? *Transportation Research Part C: Emerging Technologies*, 2022. 2

[17] Tim Salzmann, Hao-Tien Lewis Chiang, Markus Ryll, Dorsa Sadigh, Carolina Parada, and Alex Bewley. Robots that can see: Leveraging human pose for trajectory prediction. *IEEE Robotics and Automation Letters*, 2023. 4

[18] Dominic Schuhmacher, Ba-Tuong Vo, and Ba-Ngu Vo. A consistent metric for performance evaluation of multi-object filters. *IEEE transactions on signal processing*, 56(8):3447–3457, 2008. 3

[19] Edward Vendrow, Duy Tho Le, Jianfei Cai, and Hamid Rezatofighi. Jrdb-pose: A large-scale dataset for multi-person pose estimation and tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 2